



KR4HC 2010

2nd International Workshop on
Knowledge Representation for Health Care

Lisbon, Portugal

August 17, 2010

Preface

As computerized health-care support systems are rapidly becoming more knowledge intensive, the representation of medical knowledge in a form that enables reasoning is growing in relevance and taking a more central role in the area of Medical Informatics. In order to achieve a successful decision-support and knowledge management approach to medical knowledge representation, the scientific community has to provide efficient representations, technologies, and tools to integrate all the important elements that health care providers work with: electronic health records and health-care information systems, clinical practice guidelines and standardized medical technologies, codification standards, etc.

Synergies to integrate the above mentioned elements and types of knowledge must be sought both in the medical problems (e.g., prevention, diagnosis, therapy, prognosis, etc.) and also in the Computer Science and Artificial Intelligence technologies (e.g., natural language processing, digital libraries, knowledge representation, knowledge integration and merging, decision support systems, machine learning, e-learning, etc).

The second KR4HC workshop will focus on electronic patient data. After many years of promise, we finally begin to see a widespread deployment of electronic patient records and dossiers.

Fifteen papers out of nineteen papers submitted (79%) were selected to be part of the KR4HC 2010 workshop after a peer review process that involved three expert reviewers per paper. The accepted papers are of seven different countries: Austria, Canada, France, Germany, Italy, The Netherlands, and Spain.

An extended version of the best papers in the Workshop will be published by Springer as Lecture Notes in Artificial Intelligence.

Annette ten Teije
David Riaño
Mor Peleg
Silvia Miksch

Previous events:

- ✦ First International Workshop on Knowledge Representation for Health Care (KR4HC 2009), Verona, Italy.
- ✦ Workshop on Knowledge Management for Healthcare Processes (K4Help 2008), Patras, Greece.
- ✦ Workshop on From Knowledge to Global Health Care 2007, Amsterdam, The Netherlands.
- ✦ Special Track on Machine Learning and Management of Health Care Procedural Knowledge 2007, Maribor, Slovenia.
- ✦ Workshop on AI Techniques in Healthcare: Evidence-Based Guidelines and Protocols 2006, Riva del Garda (Trentino), Italy.
- ✦ Symposium on Computerized Guidelines and Protocols (CGP 2004), Prague, Czech Republic.
- ✦ Computer-Based Support for Clinical Guidelines and Protocols (EWGLP 2000), Leipzig, Germany.

Organization

Silvia Miksch, Danube University Krems, Krems, Austria
Mor Peleg, University of Haifa, Haifa, Israel
David Riaño, Universitat Rovira i Virgili, Tarragona, Spain
Annette ten Teije, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Program Committee

Syed Sibte Raza Abidi, Dalhousie University, Canada
Ameen Abu-Hanna, University of Amsterdam, The Netherlands
Roberta Annicchiarico, Santa Lucia Hospital, Italy
Luca Anselma, Università di Torino, Italy
Fabio Campana, CAD RMB, Italy
Paul de Clercq, University of Maastricht, The Netherlands
John Fox, University of Oxford, UK
Adela Grando, University of Edinburgh, UK
Robert Greenes, Harvard University, USA
Femida Gwadry-Sridhar, University of Western Ontario, Canada
Frank van Harmelen, Vrije Universiteit Amsterdam, The Netherlands
Tamás Hauer, CERN, Switzerland
Jim Hunter, University of Aberdeen, Scotland
David Isern, Unversitat Rovira i Virgili, Catalonia, Spain
Katharina Kaiser, Vienna University of Technology, Austria
Patty Kostkova, City University London, UK
Johan van der Lei, Rotterdam, The Netherlands
Peter Lucas, University Nijmegen, The Netherlands
Mar Marcos, Universitat Jaume I, Spain
Stefani Montani, Università del Piemonte Orientale, Alessandria, Italy
Silvana Quaglini, University of Pavia, Italy
Kitty Rosenbrand, Dutch Institute for Healthcare Improvement (CBO), The Netherlands
Yuval Shahr, Ben Gurion University, Beer-Sheva, Israel
Brigitte Seroussi, STIM, DPA/DSI/AP-HP, France
Andreas Seyfang, Vienna University of Technology, Austria
Robert Stevens, University of Manchester, UK
Maria Taboada, University of Santiago de Compostela, Spain
Paolo Terenziani, Univ. del Piemonte Orientale Amedeo Avogadro, Italy
Samson Tu, Stanford University, USA
Dongwen Wang, University of Rochester, USA
Jeremy Wyatt, National Institute of Clinical Excellence, UK

Sponsors



DEPARTMENT OF
INFORMATION AND KNOWLEDGE ENGINEERING



Information Systems Dept.
University of Haifa



UNIVERSITAT
ROVIRA I VIRGILI

BANZAI 萬歲!

Research Group on Artificial Intelligence
Department of Computer Science and Mathematics
Universitat Rovira i Virgili

vrije Universiteit amsterdam



DIPUTACIÓ DE
TARRAGONA

Table of Contents

Ontologies

Ontology-Based Retrospective and Prospective Diagnosis and Medical Knowledge Personalization. Romero-Tris C, Riaño D, Real F.....	1
Pre-coordination vs. Post-coordination: Different Ontology Engineering Requirements and Patterns. Schober D, Raufie D, Boeker M, Schulz S.	12
A Semantic Web Approach to Integrate Phenotype Descriptions and Clinical Data. Taboada M, Sobrido MJ, Colombo V, Pilo, B.....	20
Ontology-Based Knowledge Modeling to Provide Decision Support for Comorbid Diseases. Abidi S.	30

Medical Processes

Inducing Decision Trees from Medical Decision Processes. Torres P, Riaño D, López-Vallverdú JA.....	40
Data-Aware Monitoring for Healthcare Workflows Using Formal Methods. Ruan J, MacCaull W.	51

Patient Data, Records and Guidelines

Critiquing Knowledge Representation in Medical Image Interpretation using Structure Learning. Radstake N, Lucas PJF, Velikova M, Samulski M.	61
Linguistic and Temporal Processing for Discovering Hospital Acquired Infection from Patient Records. Hagège C, Marchal P, Gicquel Q, Darmoni S, Pereira S, Metzger MH.	71
A Clustered Markov Analysis of Patients Developing Sepsis. Gwadry-Sridhar F, Bauer M, Lewden B, Hamou A.....	81
Towards the Interoperability of Electronic Guidelines: an Experiment with OpenEHR Archetypes and a Chronic Heart Failure Guideline. Marcos M, Martínez-Salvador B.	90

Guidelines

Identifying Actions Described in Clinical Practice Guidelines Using Semantic Relations. Kaiser K, Seyfang A, Miksch S.....	99
Updating a Protocol-Based Decision-Support System's Knowledge Base: a Breast Cancer Case Study. Eccher C, Seyfang A, Ferro A, Miksch S.	109
Toward Probabilistic Analysis of Guidelines. Hommersom A.....	119
Representing Knowledge of Multiple Disorders by Object-Oriented Bayesian Networks. Quan W, Lappenschaar M, Lucas PJF, Hommersom AJ.	129
Temporal Relations in Causal Independence. van der Heijden M, Lucas PJF.....	139

Ontology-Based Retrospective and Prospective Diagnosis and Medical Knowledge Personalization

Cristina Romero-Tris, David Riaño, and Francis Real

Research Group on Artificial Intelligence,
Universitat Rovira i Virgili, Tarragona, Spain
{cristina.romero,david.riano,francis.real}@urv.net

Abstract. Computers can be helpful to support physicians in medical diagnosis and health care personalization. Here, a health care ontology for the care of chronically ill patients that was created and validated in the K4CARE project is used in prospective and retrospective diagnoses, and also in the personalization of medical knowledge. This paper describes the technical aspects of these three ontology-based tasks and the successful experiences in their application to deal with wrong diagnoses, comorbidities, missing data, and prevention.

1 Introduction

Medical diagnosis is defined as the identification of a disease by investigating the signs, symptoms and history of a patient. Diagnosis provides a solid basis for the treatment and prognosis of patients [1]. In medical diagnosis we make a distinction between the diagnostic procedure and the final diagnosis.

The *diagnostic procedure* consists on repeatedly gathering new information about the case to be diagnosed so that the physician can narrow down the list of possible diseases that the patient suffers from. In this iterative process, the physician studies the patient case and evaluates concepts such as the available means of assessment, the signs and symptoms, the feasible diseases, syndromes and social issues, as well as the patient current interventions. At some point of this process, the physician may have accumulated enough information to decide on a *final diagnosis*, which is saved in the patient's medical record.

In this context, Medical Informatics is faced with two challenges: the construction of efficient computer tools to help physicians during the diagnostic procedure (*prospective diagnosis*) and the construction of computer tools to validate final diagnoses (*retrospective diagnosis*).

The most promising approaches to these challenges are those which are based on an explicit representation of medical knowledge that could be interpreted by computers in such a way that these computers could guide or support health care professionals in prospective and retrospective diagnoses (e.g. [2, 3]). Our approach is similar to others applied in expert systems [4, 5] but with an orientation to ontologies rather than to rules.

With regard to medical knowledge representation, health care ontologies are defined as formal explicit specifications of shared conceptualizations, providing abstract models of some health care phenomenon by identifying the relevant concepts of that phenomenon [6–9]. According to this, *ontology-based prospective diagnosis* (OPD) can be defined as the process of using a health care ontology to help physicians in a diagnostic procedure, whereas *ontology-based retrospective diagnosis* (ORD) is defined as the application of a health care ontology to detect contradictions (or evidences) in already diagnosed cases. *Health care ontology personalization* is a particular case of OPD in which an ontology is continuously adapted to a particular case as new information is obtained, in such a way that only the knowledge that is relevant to the case remains in the personalized ontology. This approach differs from others like [10, 11] where the ontology is used to personalize a patient treatment but where a personalized ontology is not obtained.

In this paper we explain the methods and tools developed in the European research project K4CARE to perform OPD, ORD, and ontology personalization. The paper starts with a presentation of the health care ontology developed in the first phases of the project (see section 2) and it follows with sections on each of the uses of this ontology: ORD in section 3, OPD in section 4 and ontology personalization in section 5. The paper finishes with some case studies in section 6, and the final conclusions in section 7.

2 The Case Profile Ontology

K4CARE is a European-funded IST project whose main objective is to create, implement, and validate a knowledge-based health care model for the professional assistance to senior patients at home [12]. The project is centered in twenty-one diseases, two syndromes, and five social issues whose related knowledge is represented with a health care ontology [13] and several formal intervention plans described as SDA diagrams [14]. This health care ontology is called the Case Profile Ontology (CPO) and it is constructed on the hierarchies of classes about each one of the six root concepts in figure 1: *problem assessment* (or available tests), *sign and symptom*, *disease*, *syndrome and social issue*, and *intervention*.

The CPO incorporates international codification systems as ICD10-CM [15] for diseases, and ATC [16] for pharmacological treatments. The whole ontology contains 272 ICD10-CM codes, 177 ATC codes, 214 problem assessments, 317 signs and symptoms, 19 diseases, 2 syndromes, 5 social issues, and 174 interventions. All these concepts in the CPO are mutually related through the reversible properties *evaluates*, *hasSignAndSymptom*, *canBeCauseOf*, and *hasIntervention*.

All this knowledge is used to validate health care data for already diagnosed patients (i.e., ORD), to support physicians during diagnostic procedures (i.e., OPD), and to personalize medical knowledge (i.e., ontology personalization).

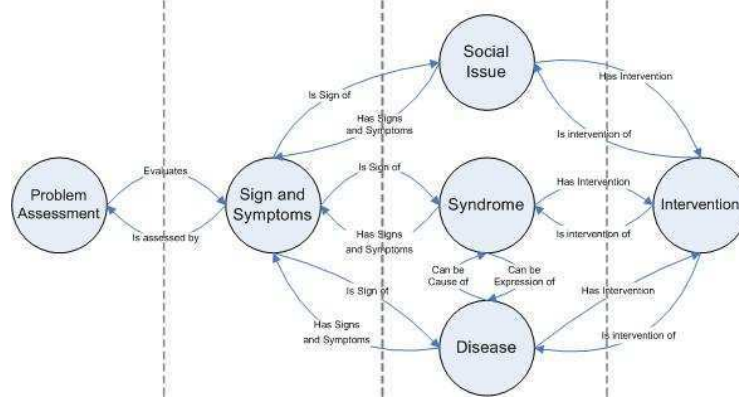


Fig. 1. Case Profile Ontology: root classes and properties.

3 CPO-Based Retrospective Diagnosis (ORD)

The knowledge in the CPO can be used to validate certain medical decisions. This validation is a reasoning process that utilizes the data of the cases under study and all the relationships between the classes of the CPO. Unlike other approaches that use a combination of rules and ontologies to reason [17, 18], our system is exclusively based on the CPO relationships to find feasible false positives and false negatives in the data.

In order to extract these false positives and false negatives, instead of defining instances and reasoning with them, here the properties that relate the terms of the CPO (see figure 1) support the following four sorts of health care reasoning:

- *Forward reasoning*: Provided the assessments of some signs and symptoms, determine the feasible social issues, syndromes, and diseases (property *isSignOf*) of a patient, and the set of interventions the patient requires (property *hasIntervention*).
- *Backward reasoning*: Knowing the interventions a patient is receiving, determine the social issues, syndromes, and diseases of this patient (property *isInterventionOf*), the signs and symptoms the patient should present (property *hasSignsAndSymptoms*), and get some recommendations on the appropriate means of assessment of the health and social conditions of the patient (property *IsAssessedBy*).
- *Inward reasoning*: From the actions on a patient (i.e., assessments and interventions) we can identify feasible signs and symptoms (property *Evaluates*), and also the diseases, syndromes, and social issues affecting that patient (property *IsInterventionOf*).
- *Outward reasoning*: From the diseases of a patient, we can foresee the possible syndromes the patient can develop (property *CanBeCauseOf*), suggest proper interventions (property *hasIntervention*), and be oriented on the signs and symptoms to look for (property *hasSignsAndSymptoms*).

4 CPO-Based Prospective Diagnosis (OPD)

In addition to ORD, the CPO can be used for OPD during a diagnostic procedure. The diagnostic procedure is a continuous approach to the identification of the health care problems that a particular patient has. As new information on the patient condition arrives, more accurate the final diagnosis can be. This process is represented in figure 2-A, which is explained in the following subsections.

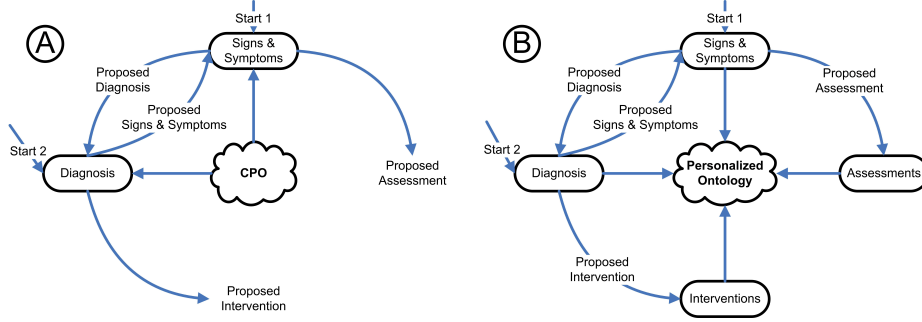


Fig. 2. CPO-Based Prospective Diagnosis and Personalization of Medical Knowledge

4.1 Supporting Physicians in Prospective Diagnosis

CPO-Based Prospective Diagnosis can start with the observation of a set of signs and symptoms (see start 1 in figure 2-A). From these data, the CPO can propose the feasible diagnoses (i.e., diseases, syndromes and social issues) and the recommended assessments. This information is analyzed by the physician who decides about the convenience of following all the indications, some of them, or none. For example, if asthenia and chest pain are observed for a patient, Chronic Obstructive Pulmonary Disease (COPD) and Heart Failure (HF) will be suggested as possible causes, and anamnesis and consultation will be the recommended assessments. At this point the physician can look for additional signs and symptoms (e.g., arrhythmia) to determine whether it is COPD or HF.

The process can also start when a set of diseases, syndromes or social issues are diagnosed (see start 2 in figure 2-A). In this case, the physician is informed of the signs and symptoms the patient should have. In this process, the physician is also recommended a set of interventions that the system finds appropriate for this patient. These recommendations may help the physician to decide whether the current treatment of the patient is both correct and complete or not.

Prospective Diagnosis can be applied in a continuous loop in which the physician finds out the signs and symptoms and obtains new possible diagnoses which, in their turn, can drive the process to other possible signs and symptoms, and so on. This loop (see figure 2-A) defines the ontology-based diagnostic procedure.

4.2 Fundamentals of CPO-Based Prospective Diagnosis

During a diagnostic process, the physician can accept some suggestions, reject others and add new data in the continuous loop described in the previous section.

The modification of the patient's signs and symptoms may have important consequences in the process because the perception of the patient diseases may change. This change is represented with a ranking of the feasible diseases that the patient may have according to the current set of signs and symptoms. This ranking has into account the relevance of signs and symptoms, which is not the same for all signs and symptoms. For example, *muscle weakness* is a sign of many diseases in the CPO, and therefore it contributes with very little information. However, *bradycardia* is specific for a small group of diseases, so if it observed, it is quite positive that the patient suffers from some disease of that group. Therefore, bradycardia is a more relevant sign than muscle weakness.

For each sign and symptom s in the CPO, we define d_s as the number of diseases affected of s (i.e., cardinality of property *isSignOf*), and w_s as the weight of s (see equation 1, where d is the number of diseases in the CPO).

$$w_s = \left(\frac{d - d_s}{d - 1} \right)^2 \quad (1)$$

Given a set of signs and symptoms S , their weights are used to calculate the relative weight w_i of a disease D_i (with signs and symptoms S_i), as equation 2 shows. These weights determine the position of the diseases in the ranking in such a way that if the diseases D_1 and D_2 obtain weights w_1 and w_2 respectively with $w_1 > w_2$, then D_1 is ranked higher than D_2 (i.e., the CPO provides more evidence that the patient has D_1 than D_2).

$$w_i = \sum_{s \in S \cap S_i} w_s / \sum_{s \in S_i} w_s \quad (2)$$

5 Personalization of Medical Knowledge

Personalization of medical knowledge is the process of adapting the CPO to a particular patient or patient prototype. The outcome of this process is a new ontology (the *personalized ontology*) only representing the health care knowledge that affects that patient or patient prototype. Figure 2-B shows this process.

The personalized ontology includes the new concept hierarchies *PatientCase* and *Background* to distinguish between the information that has been confirmed for the patient and the information that could affect the patient. So, for example, if the patient is diagnosed of HF but arrhythmia (one of the HF signs in the CPO) is not observed, then both concepts appear in *Background*, but *PatientCase* only refers to the observed HF as part of the patient condition, leaving arrhythmia as an unobserved sign or a sign that the patient could develop in the future.

In figure 2-B, when starting personalization the physician determines either the patient signs and symptoms (start 1) or the patient diagnosis (start 2).

This information is used to select the related knowledge in the CPO concerning assessments, signs and symptoms, diseases, syndromes, social issues, and interventions. Ontology properties are used for this purpose. At this point, the physician can confirm or reject part of this knowledge or incorporate a new one. All that knowledge will be part of the *Background*, but only the confirmed (and the incorporated) knowledge will be part of the *PatientCase* in the personalized ontology. The incorporation of this knowledge follows two steps: personalization of the CPO terminology, and personalization of the CPO properties.

5.1 Personalization of the CPO Terminology

All the health care terminology in the CPO is structured in the six main hierarchies of terms that were summarized in figure 1. For any two terms t_1, t_2 of the same hierarchy, if they belong to different branches they represent disjoint terms (i.e., $t_1 \cap t_2 = \emptyset$), but if they belong to the same branch then one of them includes the other one (i.e., $t_1 \subseteq t_2$). We now present an example of personalization: in figure 3, the hierarchy of interventions contains the disjoint terms *Vaccines* and *Psycholeptics*, and the term *BenzodiazepineDerivatives* which is a kind of psycholeptic (i.e., $BenzodiazepineDerivatives \subseteq Psycholeptics$).

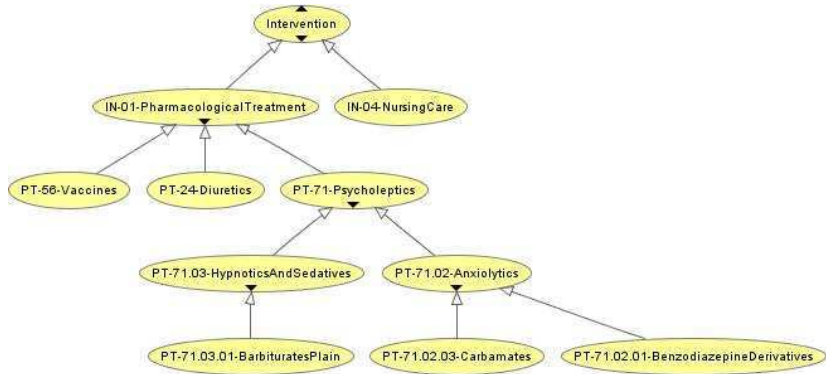


Fig. 3. Part of the Hierarchy of Interventions in the CPO.

When two terms t_1, t_2 that belong to the same branch ($t_1 \subseteq t_2$) appear in the description of a patient condition, this means that the patient has not only t_1 but also a t' which is a type of t_2 different from t_1 (i.e., $t' = t_2 - t_1$). This t' represents a new concept which is not represented by any of the CPO terms. The incorporation of this new term in the hierarchy of the personalized ontology causes this one to be different from the hierarchy of the CPO. This transformation of hierarchies is explained with the previous example about the interventions *BenzodiazepineDerivatives* and *Psycholeptics* in which a class representing all the

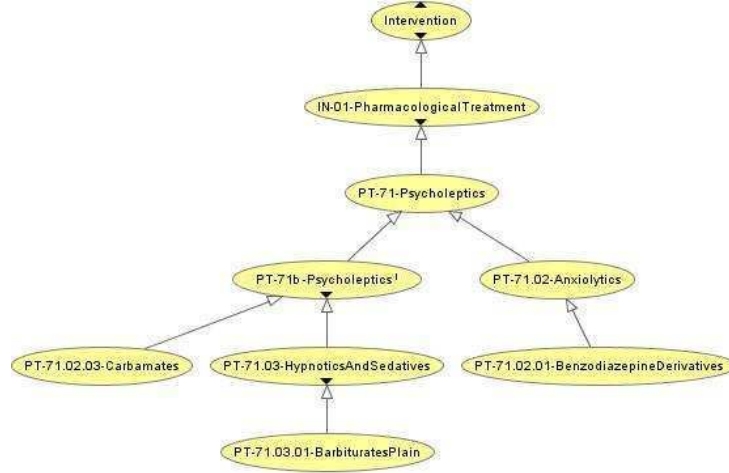


Fig. 4. Transformed Hierarchy on Interventions in the Personalized Ontology.

psycholeptics except *BenzodiazepineDerivatives* is introduced in the personalized ontology with the temporal name *Psycholeptics'* as a direct descendant of *Psycholeptics* (see figure 4). This new class maintains all the antecedents and descendants (except benzodiazepine) of psycholeptics in the CPO. In the personalized ontology, the excluded CPO sub-hierarchy *BenzodiazepineDerivatives* is maintained in the same position that it had in the CPO.

5.2 Personalization of the CPO Properties

The CPO properties observed in figure 1 are introduced in the personalized ontology attending to the following principles: (1) properties between pairs of confirmed terms keep their relationships unchanged; (2) properties between a rejected term and any other confirmed term are kept in the *Background* but not in the *PatientCase* knowledge of the personalized ontology; and (3) properties between terms incorporated by the physician are introduced as *PatientCase* but not as *Background* knowledge to avoid inconsistencies with the knowledge in the CPO.

6 Evaluation

All the introduced methods for ORD, OPD, and ontology personalization have been integrated in a tool. This tool offers an intuitive Graphic User Interface (GUI) to the physicians. For each new case, the GUI provides medical suggestions as the ones described in section 4.1 and a ranking of the feasible diseases of that patient, calculated with the formulas in section 4.2. The physician can use the GUI to accept or reject suggestions and also to add new information

till the patient case is well defined. At this point, the tool builds and stores a personalized ontology, according to the procedures explained in section 5.

The tool was used to evaluate the processes of personalization, ORD, and OPD. Tests on personalization aim to confirm that the personalized ontology fits the patient case. From the point of view of a particular patient case, there are three kinds of knowledge in the CPO: the validated knowledge (*PatientCase*), the knowledge that has to be validated (*Background*) and the knowledge that is considered useless because it has no relation with the patient case. As it is explained in section 5, only the two formers appear in the personalized ontology. Therefore, the objective of the first evaluation is to determine that, while useless knowledge disappear from the personalized ontology, the validated and the possibly validated knowledge remains. In order to do that, 19 different prototypes of patient were tested. Each prototype represented a patient case with only one of the 19 diseases represented in the CPO. Consequently, 19 personalized ontologies were obtained with the automatic methods explained in section 5. After comparing them with the CPO, the results were: (1) all the CPO knowledge related to the patient cases remains in the personalized ontology and (2) the average size of the personalized ontologies is 10.95% of the classes and 10.16% of the relationships of the complete CPO.

A dataset provided by the health care group SAGESSA with 916 patients assisted in 2009 was also analyzed with the tool. These patients had one or several chronic diseases among Hypertension, Diabetes, Heart Failure, and Ischemic Heart Disease. Table 1 shows the percentages to which the CPO is reduced after personalizing the knowledge about each sort of patient (i.e., table row).

Table 1. Ontology Size Reduction after Personalization.

Hypertension	Diabetes	Heart Failure	Ischemic HD	No. Pat.	%class	%relat
•				563	5.46%	3.87%
	•			118	9.77%	7.23%
		•		18	10.84%	10.29%
			•	30	6.71%	5.29%
•	•			125	12.25%	10.18%
•		•		13	12.67%	12.75%
•			•	19	9.11%	7.96%
	•	•		6	15.65%	15.52%
	•		•	7	13.99%	11.83%
		•	•	1	12.42%	12.38%
•	•	•		4	17.14%	17.83%
•	•		•	9	15.81%	14.24%
•		•	•	2	14.24%	14.84%
	•	•	•	0	17.14%	17.59%
•	•	•	•	1	18.63%	19.90%

The results show that the percentage of reduction of the CPO do not grow linearly with the number of diseases suffered by the patient. The reason is that the knowledge of two diseases is not mutually disjoint.

In order to evaluate OPD and ORD, we defined a test in three stages: determine the types of aid that the CPO can provide for decision support, introduce the developed tool in SAGESSA, and use the tool to retrospectively evaluate SAGESSA's databases. At the moment only the first stage has been completed.

During this stage we identified and tested the abilities of the tool in front of some problems as wrong diagnoses, comorbidities, missing data, related diseases, and prevention. In the next paragraphs we include some of the cases studied.

Wrong Diagnoses. A wrong diagnosis occurs when the physician diagnoses a patient with an incorrect condition. **Case 1:** The physician diagnosed COPD to patient *P1*. According to figure 2-A, the tool proposed the signs and symptoms related to COPD. The physician confirmed some of them (Asthenia, Chest-Pain, DecreasedExerciseTolerance, MalaiseAndFatigue, Tachycardia, Dyspnea, Tachypnea, Cyanosis, AbnormalXRyLung, and FluctuatingCourse), rejected others (AbnormalBacteriologicalExams, Cough, SleepApnea, Hyperventilation, Hypoventilation, IntercostalRetraction, PleuriticChestPain, Stridor, UseAccessoryMuscle, AbnormalThoraxExamination, Hypersomnia, Apnea, Bradypnea, AbnormalArterialBloodGas, AbnormalCoagulation, and AbnormalHemogram) and incorporated some new ones (AnginaPectoris, Arrhythmia, Palpitation, Edema, and AbnormalEKG).

With this information the system recalculated the possible diseases and provided the following ranking: Ischaemic Heart Disease -IHD (45.4%), COPD (38.95%) and HF (38.3%). From the resulting values the physician realized this is not a typical patient of any of these diseases and IHD ranks better. Finally the physician decided to change the original diagnosis (COPD) to IHD.

Comorbidities. Comorbidity is defined as the presence of one or more diseases in addition to a primary disease and the effects of such additional diseases. **Case 2:** The physician had the medical history of patient *P2*, who was diagnosed and started treatment for Diabetes months before. However, *P2* got worse and the physician observed some symptoms in the medical history such as Headache and Dizziness that were not related to Diabetes. Consequently, the physician was recommended some assessments that, when accepted, resulted into new signs and symptoms like Epistaxis and Palpitation. Old and new signs and symptoms were then the basis for the system to inform the physician that there was an evidence of Hypertension as an undetected comorbidity. Hypertension was confirmed.

Missing Data. When there is some missing data, the condition of the patient cannot be precisely defined. **Case 3:** The physician had scarce data about *P3*, the patient. Problem assessments determined Chest Pain, Rigidity, PseudobulbarPalsy, Ischuria, Abrasion, Urticaria, Arthralgia, Bulimia, and AbnormalSerumAnalysis. The system proposed Decubitus Ulcer (6.76%), Arthritis (5.73%), and IHD (5.71%). The physician suspected that some signs and symptoms were missing because the evidence of the diseases is very low. He decided to consider IHD and the system recommended him to look for and to confirm some of the additional signs and symptoms: DecreasedExerciseTolerance, MalaiseAndFatigue, AnginaPectoris, Arrhythmia, etc. The physician confirmed Arrhythmia and Abnormal ECG. Certainty on IHD increased to 46.02%.

Related Diseases and Prevention. Not all the existing diseases are independent. There are many diseases whose signs and symptoms, if not treated, can be cause of new diseases. Sometimes, when a patient is diagnosed with a disease, these other related diseases must be considered as likely to be developed by the patient in the future. **Case 4:** The physician diagnosed *P4* with Anaemia and registered the signs Anamnesis, Consultation, MedicalUltrasound, SedimentationRate, Geriatric, Transferrin, Fe, HeartRate, Cardiology, Hemogram, Ferritin, CRP, and Bilirubin. The system confirmed Anaemia as the most certain disease (53.94%), but it also warned us of two alternative diseases: IHD (30.54%) and HF (16.16%). The physician started measures to prevent them.

7 Conclusions

We have proposed the use of a health care ontology knowledge (CPO) to support physicians in three challenges: retrospective and prospective diagnoses, and health care knowledge personalization. A tool that integrates logical and numerical reasoning to guide physicians in the above mentioned challenges was implemented. In order to test the tool, physicians from SAGESSA utilized it to study multiple patients, among which this paper introduces some of them as case examples. The results obtained indicate that the tool does not only support physicians during the diagnosis procedure (*prospective diagnosis*), but it is also able to adapt all the knowledge available in the CPO to a particular patient or patient prototype (personalization), and it is also ready to detect inconsistencies in the medical history of the patients (*retrospective diagnosis*). This last application remains as immediate future work in which the tool will be installed in the computers of the health care group SAGESSA and used for the retrospective analysis of their databases.

The labour presented in this paper is based on the weights defined in [19] and it represents the first phase of a work consisting in the definition of an ontology-based reasoning process. In a second phase we will employ the database of SAGESSA to make an epidemiological study that provides the weight of each sign depending on the disease. The cases of the database will be used to determine the weights of the signs in the different diseases. With these new weights, equations 1 and 2 will be modified, as the reasoning strategy will remain the same.

This work has been partially funded by the K4CARE and HYGIA projects. The authors acknowledge the support of Dr. A. Collado, Dr. Ll. Colomès, and the SAGESSA group.

References

1. Critchley, M.: Medical Dictionary, Butterworths, London, 1986.
2. Isern, D., Moreno, M.: Computer-based execution of clinical guidelines: A review. *I. J. Medical Informatics* 77(12): 787-808 (2008)

3. Peleg, M., Tu, S., Bury, J., Ciccarese, P., Fox, J., Greenes, R.A., Hall, R., Johnson, P.D., Jones, N., Kumar, A., Miksch, S., Quaglini, S., Seyfang, A., Shortliffe, E.H., and Stefanelli, M. Comparing computer-interpretable guideline models: a case-study approach, *JAMIA* 10 (2003), pp. 5268.
4. Myers, J. D., The Background of INTERNIST-I and QMR. In: *A History of Medical Informatics*. Blum and Duncan (Ed.). New York: ACM press, 427-433, 1990.
5. Pople, H. E., CADUCEUS: An Experimental Expert System for Medical Diagnosis. In: *The AZ Business*, P. Winston and K. Prendergast, Eds. MIT Press, Cambridge, Mass., 1984.
6. Jovic, A., Prcela, M., Gamberger, D.: Ontologies in Medical Knowledge Representation. In: *29th Int. Conf. on Information Technology Interfaces*, 535-40, 2007.
7. Ceusters, W., Smith, B., Flanagan, J. Ontology and Medical Terminology: Why Description Logics Are Not Enough. In: *Towards Electronic Patient Record*, 2003.
8. Rodriguez, A., Mencke, M., Alor-Hernandez, G., et al. MEDBOLI: Medical Diagnosis Based on Ontologies and Logical Inference. In: *Int. Conf. on eHealth, Telemedicine, and Social Medicine*, 233-38, 2009.
9. Studer, R., Benjamins, R., Fensel, D.: Knowledge engineering: Principles and methods. *IEEE Trans. on Data and Knowledge Eng.* 25(1-2): 161-97, 1998.
10. Quaglini, S., Panzarasa, S., Giorgiani, T., Zucchella, C., Bartolo, M., Sinforiani, E., Sandrini, G.: Ontology-Based Personalization and Modulation of Computerized Cognitive Exercises. In: *AIME 2009, LNAI 5651*, 240-244.
11. Abidi, S. S. R., Chen, H.: Adaptable Personalized Care Planning via a Semantic Web Framework. In: *MIE 2006*.
12. Campana, F., Moreno, A., Riaño, D., Varga, L.: K4CARE: Knowledge-Based Homecare e-Services for an Ageing Europe. In: *Agent Technology and e-Health, Whitestain Series*, Birkhäuser Verlag Basel, 95-115, (2008).
13. Riaño, D., Real, F., Campana, F., Ercolani, S., Annicchiarico, R.: An Ontology for the Care of the Elder at Home. In: *Art. Int. in Medicine, LNAI 5651*, 235-239, 2009.
14. Riaño, D.: The SDA* Model: A Set Theory Approach. In: *20th IEEE Int. Workshop on CBMS*, 563-68, 2007.
15. WHO: International Classification of Diseases (ICD), <http://www.who.int/classifications/icd/en/>.
16. WHO: Anatomical Therapeutic Chemical Classification System (ATC), <http://www.whocc.no/atcddd/>.
17. Angele, J., Boley, H., de Bruijn, J., Fensel, D., Hitzler, P., Kifer, M., Krummenacher, R., Lausen, H., Polleres, A., Studer, R.: Web Rule Language (WRL). World Wide Web Consortium, September 2005. W3C Member Submission, <http://www.w3.org/Submission/WRL/>.
18. Eiter, T., Ianni, G., Polleres, A., Schindlauer, R., Tompits, H.: Reasoning with rules and ontologies. In: *Reasoning Web*, 93-127, 2006.
19. Dhyani, D., Ng, W. K., Bhowmick, S. S.: A Survey of Web Metrics. In: *ACM Computing Surveys* 34(4), 469-503, 2002.

Pre-coordination vs. post-coordination: different ontology engineering requirements and patterns

Daniel Schober, Djamila Raufie, Martin Boeker, Stefan Schulz

Medical Informatics Group, IMBI,
Freiburg University Medical Center, Germany
schober@imbi.uni-freiburg.de

Abstract. To what extent an ontology should provide readily useable pre-coordinated expressions? Pre-coordination is an issue in any major ontology engineering project. Reducing the incidence of pre-coordinated expressions shifts the focus to post-coordination, i.e. the restriction of the ontology to mainly provide primitive representational units, which then leaves the burden of assembling compositional expressions to the user. In this paper, different engineering requirements for ontology pre- and post-coordination are discussed in the context of description logics. On the one hand, pre-coordination can already be supported by equivalence statements drawing on Aristotelian definitions where a relatively low level of shared domain knowledge is required. On the other hand, to be user-friendly, ontologies for post-coordination require a comprehensive axiomatization of the domain using value restrictions and negations. These can be exploited to guide the user to adhere to preformulated design patterns and to avoid nonsensical, ambiguous, and idiosyncratic coordinations.

The ontology examples are available at:
<http://purl.org/biotop/src/pneumonia.zip>

1 Introduction

Pre-coordination is a general principle governing the design of terminological and ontological resources. According to Frege’s Principle [1], the meaning of a complex expression is fully determined by the meanings of its constituents, together with appropriate combination rules. Hence, for instance, the meaning of the expression “bacterial pneumonia” can be clearly derived from the meaning of its constitutional parts “pneumonia” and “bacteria”. Or the meaning of “pancreatitis” is fully derived from the meaning of “pancreas” (and its modified word stem “pancreat-”) and the meaning of the suffix “-itis”. However, not all composed linguistic expressions obey Frege’s principle: The meaning of “borderline disorder”, for instance, cannot be sufficiently derived from the meanings of “borderline” and “disorder”.

Lexicon, terminology, and ontology curators are continuously confronted with decisions about whether to include a complex term or a fully defined class (pre-coordination), or whether to leave the burden of constructing complex expressions to the user (post-coordination) [2]. A major reason for providing pre-coordinated constructs is the expected frequency of use. Hence “femur fracture”

is more likely to be found in a medical dictionary or ontology rather than, e.g. “fracture of the middle phalanx of the right index finger”.

It is a truism that pre-coordinating everything is doomed to failure, as a consequence of combinatorial explosion. Even admitting a certain degree of pre-coordination may multiply the editorial burden, lead to performance problems at reasoning time and, additionally, makes content retrieval difficult. A good example for this is SNOMED CT, a huge clinical terminology with more than 300,000 classes [3], with a high but unevenly distributed incidence of pre-coordination. SNOMED CT contains over 800 pre-coordinated codes for burns, including, e.g. “Third degree burn of wrist”, “Superficial burn of great toe”, “Phosphorus burn of skin”, which is however, still far from covering all possible combinations of burns degrees, skin regions, causes and complications. Here, providing just a few atomic building blocks, together with postcoordination rules would render the artifact smaller and efficient.

However, most biomedical terminologies in practical use are restricted to the provision of pre-coordinated terms, at the price of drastically reducing the level of detail available to the users. Exceptions to this rule are found with nomenclatures [4], which are combinatorial vocabularies organized along disjoint axes. Early SNOMED versions [5] and several nursing terminologies follow this principle. A formal foundation of this combinatorial architecture was first put into practice by the GALEN approach [6] and later taken up by SNOMED RT and CT [5]. As soon as an ontology supports both pre-coordination and post-coordination, intelligent systems (supporting fact retrieval, decision support etc.) need to be able to check for equivalence between existing pre- and user-generated post-coordinated expressions. Only a formal-logic grounding, supported by inference engines such as description logics classifiers can assure the detection of equivalences between pre-coordinated and post-coordinated expressions, which is a crucial task, e.g. for the re-use of coded clinical data in computerized medical records. But still widely used terminologies such as LOINC [7] allows to encode the same content in different ways without providing a formal framework for the detection of equivalences. It permits, e.g., to express “Weight at birth” by a single preexisting code (LOINC 8339-4), and in parallel, the post-coordination of new expression like:

(Weight | LOINC 3141-9, Weight circumstance | LOINC 8337-8, Birth | SNOMED F-88005).

SNOMED CT’s compositional grammar [8] supports the post-coordination of complex expressions such as “insertion of a left hip prosthesis”:

```
363704007 procedure site | = (24136001 | hip joint structure | :272741003 |
laterality | =7771000 | left | ) {363699004 | direct device | =304120007 |
total hip replacement prosthesis | , 260686004 | method | =257867005 |
insertion-action | }
```

Translated into description logics and submitted to a classifier equivalence or subsumption relations between such expressions and existing SNOMED CT classes can then be computed.

Motivated by an ongoing ontology engineering project in the field of infectious diseases and antibiotic therapy [9] we will analyze ontology design requirements regarding their use in a pre-coordinative vs. a post-coordinative context. We acknowledge that modern, ontology-based biomedical terminology systems are neither fully post-coordinative nor fully pre-coordinative and that there are good reasons for this. Our hypothesis is that different use cases pose quite distinct design principles and demands on shaping biomedical ontologies.

2 Methods

We use description logics (DL) [10] as being the standard representational formalism in most biomedical ontology projects like SNOMED CT [3], NCIT [11], and OBO Foundry [12]. Out of the different description logics dialects described in the W3C OWL 2 specification we will base our deliberations on the OWL 2 $\mathcal{RL}$ profile [13]. In particular, the constructors used are class inclusion ($\sqsubseteq$), class equivalence ($\equiv$), class disjointness using negation ($\neg$), existential quantification ($\exists$), value restriction ($\forall$), conjunction ($\sqcap$), and disjunction ($\sqcup$).

We will use the nosology of pneumonia as running example. A pneumonia is a disease characterized by the inflammation of lung tissue. Different types of pneumonia can be distinguished according to the following criteria:

1. Pathological/anatomical: the localization of the disease in the lung and its extension to certain tissues;
2. Course: acute or chronic;
3. Etiological: causes of the disease (infections, physical, chemical...);
4. Pre-existing conditions, which are complicated by pneumonia;
5. The environment in which an infections pneumonia was acquired (community or hospital);
6. Signs and symptoms.

2.1 Pre-coordination requirements

The focus of pre-coordination is the provision of axioms that state equivalences between atomic classes and compositional expressions. The meaning of a class is therefore fully defined in terms of a complex expression out of other classes. In the following examples several full definitions are given:

$$\text{Pneumonia} \equiv \text{Inflammation} \sqcap \exists \text{has-participant.LungTissue} \quad (1)$$

$$\text{AcutePneumonia} \equiv \text{Pneumonia} \sqcap \exists \text{bearer-of.AcutenessQuality} \quad (2)$$

$$\text{BacterialPneumonia} \equiv \text{Pneumonia} \sqcap \exists \text{has-agent.BacteriaPopulation} \quad (3)$$

$$\text{ViralPneumonia} \equiv \text{Pneumonia} \sqcap \exists \text{has-agent.VirusPopulation} \quad (4)$$

As the examples demonstrate, pre-coordinated expressions require, besides the equivalence operator $\equiv$, existential restriction ($\exists$) and conjunction ($\sqcap$), all of them belonging to the the OWL $\mathcal{EL}$ specification. However, if exclusively relying on axioms of this kind important ontologic aspects are missing like:

- Class delineation: Our example would perfectly fit in a world in which the classes `LungTissue` and `KidneyTissue` overlap and therefore a pneumonia is also located in a kidney.
- Cardinalities of role fillers (object properties), specifying, e.g., whether an instance of `AcutePneumonia` can have, besides acute, other qualities, e.g. chronic.
- Allowed values, e.g. whether the role `has-agent` in `BacterialPneumonia` can, additionally, be filled by, e.g. `VirusPopulation` or `ArthropodPopulation`.
- Toplevel categories, e.g. whether `Pneumonia` is a process, a pathological structure, or both.

The use of an ontology limited to pre-coordination axioms would therefore be restricted to use cases like retrieving information encoded by post-coordinated expressions by pre-coordinated queries (or vice-versa). For any further-reaching knowledge services, pre-coordination axioms must be embedded into an ontology that additionally supports:

- Taxonomic hierarchies structured by subclass relations, in order to allow inferences such as that every `BacterialPneumonia` is a `BacterialInflammation`;
- Mereotopological axioms, in order to allow inferences such as that every `Pneumonia` is located in some `Lung`, because every instance of `LungTissue` is part of some `Lung`;
- Disjointness statements, e.g. that processes cannot be structures, and therefore a pneumonia process is not the same as the underlying pathological structure, or that a population of viruses cannot be a population of bacteria.
- A strict categorization by an upper ontology.

2.2 Post-coordination requirements

The above picture changes if an ontology is primarily devised to support post-coordination to be performed by users. Such post-coordinated expressions are expected to be

- valid, i.e. they support the creation of meaningful compositions and prevent nonsensical coordinations, e.g. “left pancreatitis”, or “viral pneumonia caused by bacteria”;
- expressive, i.e. they guide the user to create unambiguous compositions;
- reliable, i.e. they support compositions that are consistent between different modelers.

This entails new demands for OWL ontologies, which must be addressed by elaborated ontology design patterns. In the following we will use the pneumonia example to illustrate such patterns. The pneumonia ontology we are developing is rooted in the BioTop upper ontology [14] and can be downloaded from <http://purl.org/biotop/src/pneumonia.zip>.

In BioTop, generalized localization is expressed by the relation `locus-of` (inverse `has-locus`), which relates a place with an entity which occurs, inheres, or

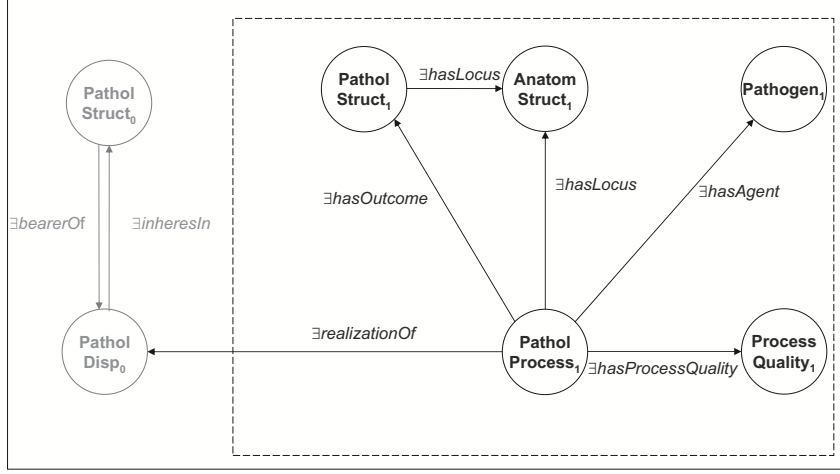


Fig. 1. Disease model distinguishing between pathological structure, disposition, and process. The rectangle includes classes necessary for describing an infectious disease: the disease process, its qualities (e.g. acute, chronic), the pathological structure which is being formed as an outcome of the process, and the anatomical structure where the process is located. The pathological process is a result from the action of the pathogen on the one hand, and from the realization of a pre-existing pathological disposition (bottom left), on the other hand. The latter inheres in pre-existing pathological structure (top left), which may be congenital or the outcome of a previous pathological process (not depicted).

is part of it. We can therefore restrict the localization of pneumonias by the following axiom:

$$\text{Pneumonia} \sqsubseteq \forall \text{has-locus}.(\exists \text{locus-of.LungTissue}) \quad (5)$$

Given an underlying anatomy model which, e.g. includes the axioms that only lungs have lung tissue, lungs are located in a thorax, and what is located in the lung can not be located in the abdomen or the extremities,

$$\text{LungTissue} \sqsubseteq \exists \text{has-locus.Lung} \quad (6)$$

$$\text{Lung} \sqsubseteq \exists \text{has-locus.Thorax} \quad (7)$$

$$\exists \text{has-locus.Thorax} \sqsubseteq \neg \exists \text{has-locus}(\text{Abdomen} \sqcup \text{Extremity}) \quad (8)$$

an application (e.g. which controls a dropdown menu in a GUI) can display exactly those terms which indicate sensible locations for pneumonia, whereas it hides all other anatomy terms. In a similar vein, the cause of bacterial pneumonia can be restricted by

$$\text{BacterialPneumonia} \sqsubseteq \forall \text{has-agent.BacteriaPopulation} \quad (9)$$

assuming that the class `BacteriaPopulation` is disjoint from, e.g. `VirusPopulation`, and others at the same level. In an ontology-enabled interface, the user would then only be offered subclasses of `BacteriaPopulation`, such as `MRSAPopulation` or `PneumococcusPopulation`.

There are more complex patterns such as guiding the modelling of secondary pneumonias, i.e. pneumonias, which exist as complications of previous conditions. This requires a more sophisticated model of disease, according to the one proposed in [15], based on [16], cf. Fig. 1. Using this model we define a secondary disease as a pathological process which is the realization of a pre-existing disposition. This disposition inheres in a pathological structure, which may exist as a congenital disorder or as some outcome of a former pathological process. One example is a lung infarction as a cause of pneumonia, others are lung edema and bronchial carcinoma.

$$\begin{aligned} \text{Pneumonia} \sqsubseteq & \forall \text{realization-of.}(\text{PathologicalDisposition} \sqcap \\ & \forall \text{inheres-in.}(\text{LungInfarction} \sqcup \text{LungEdema} \sqcup \text{BronchialCarcinoma} \sqcup \dots)) \end{aligned} \quad (10)$$

A variation of this pattern is the one which refines a pneumonia process in terms of accompanying signs and symptoms, such as fever, chills, or cough. Again, the user would welcome a concise list of those signs and symptoms that are commonly attributed to the underlying disease, whereas numerous other conditions, which can, of course co-incide with a pneumonia (e.g. toothache, hernia,...) should be hidden. In these cases we consider the pneumonia a primary process, which produces a pathological structure which then bears the disposition for the signs and symptoms (the secondary processes), e.g.

$$\begin{aligned} \text{Pneumonia} \sqsubseteq & \exists \text{has-output.}(\text{PathologicalStructure} \sqcap \exists \text{bearer-of.} \\ & (\text{PathologicalDisposition} \sqcap \forall \text{has-realization.}(\text{Cough} \sqcup \text{Chills} \sqcup \text{Fever} \sqcup \dots)) \end{aligned} \quad (11)$$

A more intricate pattern is necessary when addressing the environment in which the pathogens have been acquired. Clinically, this distinction is essential, because the risk of being infected by multiple resistant bacteria is much higher inside health care institutions. Here we need to introduce a second process, `LungTissueExposureToPathogen`, the onset of which is prior to the pneumonia itself (incubation period). For the definition of hospital-acquired pneumonia, the criteria is that the organism in which this process takes place is located within a health care institution when this process initiates.

2.3 Discussion and Conclusion

Comparing the requirements for pre-coordination centered ontologies (i) vs. post-coordination centered ontologies (ii) we observe that post-coordination centered ontologies require more sophisticated design patterns in the form of axiomatic templates if the above requirements are to be met. A further analysis draws attention to the fact that the patterns needed for ontology guided post-coordination use different constructors as the patterns for pre-coordination.

Whereas pre-coordination centered ontologies can perfectly be supported by inexpressive description logics, such as OWL-2 $\mathcal{EL}$ (which is the case, e.g., with SNOMED CT) which is sufficient to express Aristotelian definitions, post-coordination requires careful usage of a broader range of constructors, *viz.* negation, disjunction and value restriction. The more guidance the ontology offers to the modelers in order to prevent them to deviate from the path of the preformulated design patterns, the more accurately its constraints need to be crafted. Attention must be paid to the fact that description logics based ontologies adhere to the open world assumption. In contrast to frame-based systems, e.g. CLIPS [17] all constraints on values need to be supported by a carefully devised architecture of disjoint categories: if there is nowhere stated that a bacteria cannot be a virus, a value constraint for bacterial pneumonia is of no use. This poses new challenges for intelligent user interfaces supported by the enforcement of constraints provided by the ontology. While procedures which leverage logically defined classes, their OWL representation and the identification of patterns in candidate terms for ontology inclusion are currently being investigated (e.g. http://obi-ontology.org/page/Quick_Term_Templates), principled ontological upper level constraints as found in upper and top level ontologies already provide valuable patterns. In all these areas better support by ontology editing tools would be welcome, i.e. in order to increase inter-annotator reliability, support for compositions that are consistent between different modelers should be provided.

Acknowledgements This work was supported by the DFG, grant agreement JA 1904/2-1, SCHU 2515/1-1, and by the EU 7th FP project DebugIT, grant agreement ICT-2007.5.2-217139.

References

1. Zoltán Gendler Szabó. Compositionality. In Edward N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Stanford: The Metaphysics Research Lab, plato.stanford.edu, 2003.
2. P.L. Elkin, M. Tuttle, K. Keck, K. Campbell, G. Atkin, and C.G. Chute. The role of compositionality in standardized problem list generation. In *Studies in Health Technology and Informatics*, volume 52, pages 660–444, 1998.
3. SNOMED *Clinical Terms*. Copenhagen, Denmark: International Health Terminology Standards Development Organisation (IHTSDO), 2010.
4. Josef Ingenerf and Wolfgang Giere. Concept-oriented standardization and statistics-oriented classification: Continuing the classification versus nomenclature controversy. *Methods of Information in Medicine*, 37(4/5):527–539, 1998.
5. Ronald Cornet and Nicolette de Keizer. Forty years of snomed: a literature review. *BMC Medical Informatics and Medical Decision Making*, 8:Suppl 1:S2, 2008.
6. Alan L. Rector, A. J. Glowinski, W. Anthony Nowlan, and Angelo Rossi Mori. Medical-concept models and medical records: An approach based on GALEN and PEN & PAD. *Journal of the American Medical Informatics Association*, 2(1):19–35, 1995.

7. S.M. Huff, R.A. Rocha, C.J. McDonald, G.J. DeMoor, J.E. DeMoor, and T. Fiers. Development of the logical observation identifier names and codes (LOINC) vocabulary. *Journal of the American Medical Informatics Association*, 5:276–292, 1998.
8. Kent Spackman and John Gutai. Compositional grammar for SNOMED CT expressions in HL7 Version 3. http://www.ihtsdo.org/fileadmin/user_upload/Docs_01/About_IHTSDO/Publications/CompositionalGrammar_20081223.pdf, 2008.
9. Christian Lovis, Douglas Teodoro, Emilie Pasche, Patrick Ruch, Dirk Colaert, and Karl Stroetmann. DebugIT: building a european distributed clinical data mining network to foster the fight against microbial diseases. Number 148 in *Studies in Health Technology and Informatics*, pages 50–59, 2009.
10. Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook. Theory, Implementation, and Applications. 2nd edition*. Cambridge, U.K.: Cambridge University Press, 2007.
11. F.W. Hartel, S. De Coronado, R. Dionne, G. Fragosio, and Golbeck J. Modeling a description logic vocabulary for cancer research. *Journal of Biomedical Informatics*, 38(2):114–129, 2005.
12. B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall, OBI Consortium, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.A. Sansone, R.H. Scheuermann, N. Shah, P.L. Whetzel, and S. Lewis. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11):1251–1255, 2007.
13. Boris Motik, Bernardo Cuenca Grau, Ian Horrocks, Zhe Wu, Achille Fokoue, and Carsten Lutz. Working Group OWL 2 Web Ontology Language document overview. available at: <http://www.w3.org/STR/owl2-overview>, May 2010.
14. Elena Beisswanger, Stefan Schulz, Holger Stenzhorn, and Udo Hahn. BioTop: An upper domain ontology for the life sciences – a description of its current structure, contents, and interfaces to OBO ontologies. *Applied Ontology*, 3(4):202–212, 2008.
15. Stefan Schulz, Djamila Raufie, and Martin Boeker. Scalable representations of diseases in biomedical ontologies. In Nigam Shah and Susanna Sansone, editors, *Proc. of the 2010 Bio-Ontologies SIG. Semantic Applications in Life Sciences*. Boston, MA, USA, July 9–10, 2010.
16. Richard H. Scheuermann, Werner Ceusters, and Barry Smith. Towards an ontological treatment of disease and diagnosis. In *Proc. of the 2009 AMIA Summit on Translational Bioinformatics*, pages 116–120. San Francisco, California, March 15–17, 2009.
17. Daniel Schober, Ulf Leser, Martin Zenke, and Jens Reich. GandrKB – ontological microarray annotation and visualization. *Bioinformatics*, 21(11):2785–2786, 2005.

A semantic web approach to integrate phenotype descriptions and clinical data

M. Taboada¹, M.J. Sobrido², V. Colombo¹, and B. Pilo³

¹ Department of Electronics and Computer Science, University of Santiago de Compostela, Spain (maria.taboada@usc.es)

² Fundacion Publica Galega de Medicina Xenomica, Santiago de Compostela, Spain

³ Section of Neurology, Hospital del Sureste, Arganda del Rey, Madrid, Spain

Abstract. Integrating phenotype descriptions from text-rich research resources, such as OMIM, and data from experimental and clinical practice is one of the current challenges to promote translational research. Exploring new technologies to uniformly represent biomedical information is needed to support integration of information drawn from disparate sources. Positive progress to integrate data requires to propose solutions supporting fully semantic translations. The Semantic Web is a promising technology, so international efforts, such as the OBO Foundry, are developing ontologies to support annotation and integration of scientific data. In this paper, we show an approach to get concordances between phenotype descriptions and clinical data, supported by knowledge adapters based on description logic and semantic web rules. This integration provides a valuable resource for researchers in order to infer new data for statistical analysis.

Keywords: OWL, semantic web rules, phenotype description.

1 Introduction

Investigation oriented to promote translational research, that is, the interchange of knowledge and data between basic research and clinical level, is an extensively recognized topic nowadays [1]. The proliferation of scientific literature and text-rich resources, such as the Online Mendelian Inheritance in Man (OMIM)⁴ [2], a database of human disease genes and phenotypes⁵, and the possibility of using them to reach efficient diagnoses and effective treatments, have increased their value as knowledge resources in translational research [3]. The latest advances in genetics and genomic recommend to explore human genetic disorder associations by correlating disorder phenotypes to gene mutations, on a genome-wide scale [4]. Different approaches have been developed to represent these disease phenome-genome associations, providing conceptual frameworks to prospect new disease common patterns. One of these approaches is the bipartite

⁴ <http://www.ncbi.nlm.nih.gov/omim>

⁵ Phenotypes are any observable characteristic or trait of an organism (e.g., *Juvenile cataracts* or *Tendon xanthomas*)

network of OMIM-based disorder-disease gene associations, a graph representing associations between all known genetic disorders and all known disease genes [4]. Other approaches address the annotation of phenotype descriptions. The Human Phenotype Ontology (HPO) contains over 8000 terms representing phenotypes anomalies and it has been used to annotate all entries of OMIM and to analyze randomized phenotypic networks [5]. The Phenotype and Trait Ontology (PATO) [6] consists of more than 2000 terms describing phenotypic qualities. It was designed to compose descriptions of phenotypes by combining qualities with phenotypes entities from other ontologies, such as HPO. It has been used to integrate phenotype ontologies across multiple species [7].

As well as annotating phenotype descriptions, ontologies can be used to give the resources needed to promote translational research. In fact, other types of tools demanded by both geneticists, whenever they find a new genetic variant, and clinicians, whenever they find a new combination of clinical manifestations, are open software suites to query about experimental and clinical data in order to retrieve and analyze data [8]. In this respect we can find some current initiatives, such as the Leiden Open Variation Database (LOVD) [9], a database to separately store sequence variant data and patient data, and the Universal Mutation Database (UMD) [10], a software to run queries across multiple databases via the web. These centralized databases try to solve the problem of interpreting unstructured data set produced by different experimental methods in disparate clinical settings. But, before clinical and research laboratories begin to deposit data massively, some obstacles must be addressed. First, representing phenotypes using free-text fields in databases hampers computational inference to automatically detect similar phenotype descriptions. So, in order to integrate phenotypes coming from different data sources, the use of a standard terminology to represent phenotypes is crucial in order to get interoperability. Second, phenotypes are the set of clinical manifestations of an organism. Owing to the huge variation of the underlying molecular pathways causing a disease, the phenotypes present in a disease may be very diverse. So, some authors propose to define phenotype templates to collect the characteristic data [8]. However, an ontology-based technology would provide a more open and flexible representation mechanism, facilitating the incorporation and interpretation of new phenotype characteristics at any time. Third, phenotypes from disparate data sources can be described in different granularity levels, so it is important to provide procedures to access to them efficiently. Traditional database querying using the Structure Query Language (SQL) to identify patients, whose phenotypes are described in different granularity levels, can lead to incomplete results. The SQL lacks the abstract query base and the notion of hierarchy, both of them required to deal with abstract phenotype descriptions. These limitations can be settled using the Semantic Web technology.

In this paper, we present a Semantic Web technology based approach that can be used to increase phenotype querying ability. Our work tries to solve the three database schema limitations commented above, by reusing standardized ontologies and terminologies to describe phenotypes in the Ontology Web

Language (OWL)[11]. In order to enable ontology reuse and data sharing, our approach is supported by knowledge adapters, that is, software elements based on the principles of UPML [12], an architecture for describing knowledge-based systems. Adapters are necessary to adjust the reusable ontologies to each other, in order to describe phenotypes, and to fit together phenotype descriptions and clinical data.

2 Materials

Owing to huge genetic variation and phenotype heterogeneity in each genetic disorder, data standards must be developed by experts in a future [8]. Taking into account this huge variability, the research strategy followed in this work was the study of a single disorder, which has location in one only gene. This approach allowed us to exclusively focus on phenotype descriptions, leaving to one side genetic mutations. The selected medical disorder was the Cerebrotendinous Xanthomatosis (CTX), a rare and inherited lipid-storage disease characterized clinically by progressive neurologic dysfunction, premature atherosclerosis, and cataracts [2]. The defect in CTX was shown to reside in only one gene: CYP27A1 [13].

2.1 Text-based knowledge source

Online Mendelian Inheritance in Man (OMIM) is a text-based knowledge source of human genes and related phenotypes. It provides textual descriptions of genes and phenotypes, and additional resources, such as the clinical synopses, which relate the inherited disorders described in OMIM to their clinical manifestations (e.g., *Juvenile cataracts* or *Tendon xanthomas* are manifestations of CTX disease).

2.2 Ontologies and terminologies

An ontology is a data model that represents a set of entities in some domain and the relationships among those entities. Examples of these entities in the medical domain include diseases or phenotypes. One of the convenience of using ontologies is the potential to apply reasoners (logical inference tools), which can infer new data to subsequently facilitate query answering and statistical analysis. As an example, consider a query to find patients presenting Xanthomatosis. We would expect this search to return both patients presenting Xanthomatosis and patients presenting Plane Xanthoma, Xanthoma of Eyelid, Tuberous Xanthoma or Xanthoma Tendinosum, because all of these are types of Xanthomatosis (Fig. 1).

The Open Biomedical Ontologies (OBO) Foundry⁶ initiative [14] provides a biological and biomedical ontology repository, including the Human Phenotype

⁶ <http://obofoundry.org/>

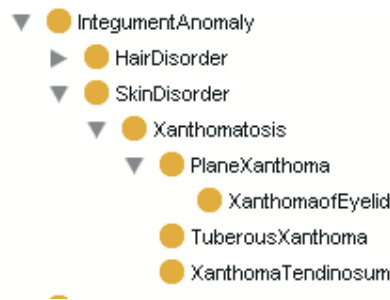


Fig. 1. 'Is-a' relationships of an ontology allowing subsumption reasoning.

Ontology (HPO) [5], developed to cover all phenotypic abnormalities collected in OMIM, and the Phenotype and Trait Ontology (PATO)[6], designed to cover phenotypic qualities, which are necessary to reach a complete description of phenotypes.

The International Health Terminology Standards Development Organization (IHTSDO) provides the SNOMED CT terminology ⁷, a large clinical terminological system. It contains formal definitions for clinical concepts using hierarchical relationships and non-hierarchical relationships. These relationships make concept semantics explicit, allowing automated classification and post-coordination.

The National Library of Medicine (NLM) ⁸ provides the UMLS Metathesaurus, which includes about 1.8 million concepts coming from over 150 source vocabularies. The clinical synopsis of OMIM is integrated in the UMLS.

The National Center for Biomedical Ontology (NCBO) supplies the BioPortal ⁹, a portal providing access to all ontologies and terminologies resources that are actively used in biomedical communities.

2.3 Patient Data

The patient data used for our current research comes from the first clinical study on CTX [15] carried out by a collaboration between Hospital Ramón y Cajal (in Madrid) and the Neurogenetic Unit from Fundación Pública Galega de Medicina Xenómica (in Santiago de Compostela). The study identified patients diagnosed with CTX between 1995 and 2008.

⁷ SNOMED CT Technical Reference Guide-July 2009 International Release

⁸ <http://umlsks.nlm.nih.gov>

⁹ <http://biportal.bioontology.org/>

3 Methods

Building ontologies to support integration of information drawn from disparate sources is not an easy task. First, we started paying attention to recognize available knowledge sources, such as ontologies and terminologies. Ideally, these resources would provide the required medical entities and relationships along with clear definitions. But, in many cases, these entities and relationships were not covered or they were not fully defined. In addition, adjusting the domain ontology and the data model to each other was necessary to increase phenotype querying ability.

Our approach is supported by knowledge adapters, that is, software elements based on the principles of UPML [12], an architecture for describing knowledge-based systems. UPML provides two types of adapters: refiners and bridges. Refiners are used to express the refinement of entities and relationships from the reused ontologies, and the set of mappings between the different reused terminologies. Our hunt for a phenotype ontology in the domain of CTX, that could be reused as a building block, showed that the Human Phenotype Ontology (HPO) [5], created for covering the phenotypic abnormalities collected in OMIM, was the most suitable. This ontology was developed in OBO and it is publicly available. The HPO provides a hierarchy of phenotype entities, covering a 60% of CTX phenotype descriptions in OMIM. So, we designed the remaining entities (40%) by reusing concepts from the UMLS Metathesaurus and SNOMED-CT, due to the following reasons:

1. The phenotype entities from the OMIM clinical synopsis are integrated in the UMLS; although some of them were loosely integrated (e.g. *Normal to slightly elevated plasma cholesterol*).
2. SNOMED CT has a high coverage in clinical findings and diseases.

Enhancing phenotype querying ability also requires that phenotype entities are described by computer-interpretable definitions. Two different methods to describe phenotypes can be found. On the one hand in medicine, the SNOMED CT community provides formal definitions of medical entities, enabling automated classification and allowing post-coordination; on the other hand in biology, the OBO foundry provides the EQ (Entity + Quality) method [6] for describing phenotypes using the PATO ontology and allowing ontological reasoning. We chose SNOMED CT to describe phenotypes due to it is built around a description logic backbone, it is integrated into the UMLS and it has a long and comprehensive reach with several hierarchies (e.g. the hierarchy *Clinical Findings* covers phenotypes entities, *Qualifier values* covers phenotypes qualities), and attribute relationships.

Describing phenotypes is achieved by a refiner specifying either (a) new entities extending the HPO with the aim of covering all descriptions in OMIM, or (b) adding attribute relationships coming from SNOMED CT, or (c) defining fresh entities or relationships, in case of no coverage. They are based on description logic, using the Web Ontology Language (OWL). On the other hand, a bridge is applied to explicitly model the relationships between the phenotype ontology

and the data model; and it is expressed by semantic web rules, using the Semantic Web Rule Language (SWRL) [16], a language to infer new phenotype knowledge about data instantiations. The Protégé-OWL [17] was selected as our application development environment due to the following reasons:

1. The technical support for designing OWL ontologies.
2. The available plug-in to create and execute SWRL rules.

4 Results

We are developing a prototype to test the use of semantic web in the relevant domain of CTX, using realistic datasets and ontologies. For it, our prototype is made of a patient phenotype management ontology, which is briefly described below.

4.1 The patient phenotype management ontology

Our framework consists of several components: a phenotype ontology, a qualifier ontology and a patient data ontology. The classes constituting the phenotype ontology were reused from the HPO and extended with SNOMED CT classes. First, we directly mapped all OMIM terms contained in the clinical synopsis section to the UMLS concepts using the Metamap tool. We restricted the mappings to the OMIM and SNOMED CT terminologies, in order to recover only the concepts from these terminologies. We also restricted the mappings to the semantic types *Disease or Syndrome*, *Pathologic Function*, *Anatomical abnormality* and *Finding* to minimize erroneous mappings. The resulting phenotype ontology contains over 150 SNOMED CT concepts and double the amount of the HPO phenotype descriptions for CTX.

The classes constituting the qualifier ontology were reused from the SNOMED CT hierarchies *Qualifier value*, *Observable entity* and all concepts needed to define the phenotype ontology. The patient data ontology models all fields included in the CTX database resulting from the national study described in [15].

4.2 Description logic-based ontology refinement

In SnomedCT, concepts in the hierarchy *Clinical Finding* represent the result of a clinical observation, assessment or judgment, and they include both normal and abnormal clinical states. An example of a clinical finding is the concept *serum cholesterol normal*. This hierarchy contains the sub-hierarchy of *Disease*. There are several attributes to *Clinical Finding* concepts, such as *finding site*, *interprets* or *has interpretation*. Table 1 shows a portion of the SNOMED CT definition for the concept *serum cholesterol normal*.

We refined the HPO concepts with the codes of the UMLS and SNOMED-CT terminologies plus the SNOMED CT definition attributes. We reused all defining attributes used in SNOMED CT for *Clinical Findings*. These attributes were represented in OWL such as *Functional Properties*.

Table 1. Partial definition of the SNOMED CT concept *serum cholesterol normal*.

<i>Serum cholesterol normal (finding)</i>
INTERPRETS serum cholesterol measurement (procedure)
HAS INTERPRETATION within reference range (qualifier value)

For the OMIM terms uncovered by SNOMED CT (e.g. *Normal to slightly elevated plasma cholesterol*), we manually modeled them reusing SNOMED CT concepts. For example, the phenotype *Normal to slightly elevated plasma cholesterol* was modeled by an OWL equivalent class resulting from the union of the two SNOMED CT classes *serum cholesterol normal (finding)* and *serum cholesterol borderline high (finding)*.

For the concepts incompletely defined by SNOMED CT, we manually modeled them reusing SNOMED CT defining attributes. For example, the concept *Juvenile Cataract* is described by the same defining attributes as the broader concept *Cataract* (Table 2). So, we refined *Juvenile Cataract* adding the knowledge necessary for interpreting it adequately (Table 3), that is, the presence of a cataract before expected time in a normal situation.

Table 2. Partial definition of the SNOMED CT concept *Juvenile Cataract (disorder)*.

<i>Juvenile Cataract (disorder)</i>
IS A Cataract (disorder)
ASSOCIATED MORPHOLOGY Abnormally opaque structure (morphologic abnormality)
FINDING SITE Structure of lens of eye (body structure)

Table 3. Refined definition of the SNOMED CT concept *Juvenile Cataract (disorder)*.

<i>Juvenile Cataract (disorder)</i>
IS A Cataract (disorder)
ASSOCIATED MORPHOLOGY Abnormally opaque structure (morphologic abnormality)
FINDING SITE Structure of lens of eye (body structure)
INTERPRETS age at first symptom (observable entity)
HAS INTERPRETATION Abnormally early (qualifier value)

4.3 SWRL-based adaptation

Clinical data abstractions are required to answer questions on phenotypes. For example, finding patients presenting juvenile cataracts demands to infer this

abnormality from the available data collected into the database (e.g. patients with cataracts at age below 30).

The defining attributes for *Clinical Findings* model the knowledge that interprets the result of some clinical observation. In order to fit together phenotype descriptions and clinical data, we need to add knowledge on *how* to interpret *Clinical Findings* (that is, phenotype descriptions) from clinical data. This knowledge bridge is expressed by the Semantic Web Rule Language (SWRL) in order to infer new phenotype knowledge about data instantiations. For example, the SWRL rule in Fig. 2 shows the knowledge required to infer the presence of the phenotype *Juvenile cataract* in patients presenting *cataracts* before their thirties.

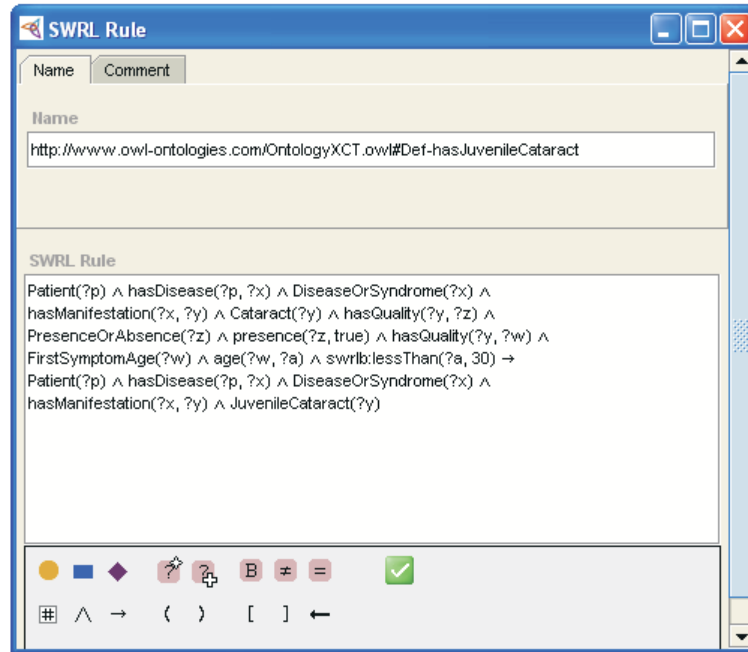


Fig. 2. SWRL rule to infer patients presenting juvenile cataract.

To have the ability to execute queries at different levels of abstraction, it is necessary to automatically populate the corresponding phenotype individuals by executing the SWRL rules. This is carried out in three steps, using the SWRL Tab provided by Protege-OWL: 1) Translating SWRL rules and relevant OWL knowledge to Jess, 2) running the Jess rule engine, and 3) translating asserted Jess knowledge to OWL knowledge. Using SWRL in this way supplies the ability to create queries with phenotypes at different levels of abstraction. For example, we can create queries about patients presenting cataracts or about patients

presenting juvenile cataracts, although datasources only contain data about the presence of cataracts.

5 Conclusions

To date, much of the semantic web prototypes have been generated by exactly mirroring the original data structure [1]. Positive progress to integrate data following this technology requires to propose solutions supporting fully semantic translations. In this paper, we described a semantic web approach that can be used to integrate phenotype descriptions and patient data from experimental and clinical practice. The approach reuses existing biomedical ontologies and terminologies, and it uses knowledge adapters to adjust the reusable ontologies to each other, and to fit together phenotype descriptions and patient data. Using the ontology language OWL and the rule language SWRL, we show how knowledge adapters can be constructed in order to increase phenotype querying ability.

Some limitations of our approach comes from using a young technology, such as the slowness of OWL stores and reasoners, or the lack of data sources in OWL [1]; whereas other barriers are the classical problems of designing knowledge-based systems, such as the bottleneck that involves knowledge acquisition to refine the existing ontologies and to bridge these to clinical data. Manual knowledge acquisition is an arduous, complex and expensive task. However, a common movement in the direction of providing complete, formal and interoperable representations covering the entire clinical domain can be perceived in the international communities, such as SNOMED CT or the OBO Foundry [18].

Acknowledgements This work has been funded by the Ministerio de Educación y Ciencia, through the national research project TermiMed (TIN2009-14159-C05-05).

References

1. Ruttenberg, A., et al.: Advancing translational research with the semantic web. *BMC Bioinformatics* **8**(3) (2007) S2
2. Amberger, A., Bocchini, C., Scott, A., Hamosh, A.: Mckusicks online mendelian inheritance in man (omim). *Nucleic Acids Res* **37** (2009) D793–96
3. Gudivada, R., Qu, X., Chen, J., Jegga, A., Neumann, E., Aronow, B.: Identifying disease-causal genes using semantic web-based representation of integrated genomic and phenomic knowledge. *J Biomed Inform* **41**(5) (2008) 717–29
4. Goh, K., Cusick, M., Valle, D., Childs, B., Vidal, M., Barabasi, A.: The human disease network. *Proc. Natl. Acad. Sci.* **104**(21) (2007) 8685–90
5. Robinson, P., Khler, S., Bauer, S., Seelow, D., Horn, D., Mundlos, S.: The human phenotype ontology: A tool for annotating and analyzing human hereditary disease. *The American Journal of Human Genetics* **83** (2008) 610–15
6. Washington, N., Haendel, M., Mungall, C., Ashburner, M., Westerfield, M., et al.: Linking human diseases to animal models using ontology-based phenotype annotation. *PLoS Biol* **7**(11) (2009) e1000247

7. Mungall, C., Gkoutos, G., Smith, C., Haendel, M., Ashburner, M., et al.: Integrating phenotype ontologies across multiple species. *Genome Biology* **11** (2010)
8. Kaput, J., Cotton, R., Hardman, L., Watson, M., Aqeel, A.A., Al-Aama, J., Al-Mulla, F., Alonso, S., Aretz, S., et al.: Planning the human variome project: the spain report. *Hum. Mutat.* **30**(4) (2009) 496–510
9. Fokkema, I., den Dunnen, J., Taschner, P.: Lovd: easy creation of a locus specific sequence variation database using an lsdb-in-a-box approach. *Hum. Mutat.* **26** (2005) 63–68
10. Beroud, C., Collod-Beroud, G., Boileau, C., Soussi, T., Junien, C.: Umd (universal mutation database): a generic software to build and analyze locus-specific databases. *Hum. Mutat.* **15** (2000) 86–94
11. McGuinness, D., van Harmelen, F.: Owl web ontology language overview. <http://www.w3.org/TR/owl-features/> (2004)
12. Fensel, D., Motta, E., van Harmelen, F., Benjamins, V., et al.: The unified problem-solving method development language upml. *Knowledge and Information Systems (KAIS): An International Journal* **5**(1) (2003) 83–131
13. Cali, J.J., Hsieh, C., Francke, U., Russell, D.: Mutations in the bile acid biosynthetic enzyme sterol 27-hydroxylase underlie cerebrotendinous xanthomatosis. *J. Biol. Chem.* **266** (1991) 7779–83
14. Smith, B., Ashburner, M., Rosse, C., Bard, J., et al.: The obo foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* **25** (2007) 1251–55
15. Pilo, B.: Xantomatosis Cerebrotendinosa en España: mutaciones, aspectos clínicos y terapéuticos. PhD thesis, Facultad de Medicina, University of Alcalá de Henares, Madrid, Spain (May 2009)
16. Horrocks, I., Patel-Schneider, P., Boley, H., Tabet, S., Grosof, B., Dean, M.: Swrl: a semantic web rule language combining owl and ruleml. <http://www.w3.org/Submission/SWRL/> (2004)
17. Knublauch, H., Ferguson, R., Noy, N., Musen, M. In: The Protege OWL plugin: an open development environment for Semantic Web applications. Volume 3298 of *Proc. of third international Semantic Web conference (ISWC)*. McIlraith S, Plexousakis D, van Harmelen F (Ed.), Springer, Hiroshima, Japan (2004) 229–43
18. Smith, B., Brochhausen, M.: Putting biomedical ontologies to work. *Methods Inf Med* **49** (2010) 135–40

Ontology-Based Knowledge Modeling to Provide Decision Support for Comorbid Diseases

Samina Raza Abidi

NICHE Research Group, Faculty of Computer Science, Dalhousie University, Canada
abidi@cs.dal.ca

Abstract: Handling comorbid diseases in a decision support framework is a challenging problem as it demands the synthesis of clinical procedures for two or more diseases whilst maintaining clinical pragmatics. In this paper we present a knowledge management approach for handling comorbid diseases by the systematic alignment of the Clinical Pathways (CP) of comorbid diseases. Our approach entails: (a) knowledge synthesis to derive disease-specific CP from evidence-based sources; (b) knowledge modeling to abstract medical and procedural knowledge from the CP; (c) knowledge representation to computerize the CP in terms of a CP ontology; and (d) knowledge alignment by aligning multiple CP to develop a unified CP knowledge model for comorbid diseases. We present the COMET system that provides decision support to handle comorbid cardiac heart failure and atrial fibrillation.

1. Introduction

Comorbidity is the existence of medical conditions concurrent with a primary condition in the same patient. Typically, chronic diseases are associated with comorbidities. For instance, Chronic Heart Failure (CHF) is one such chronic condition that is frequently associated with comorbidities such as Atrial Fibrillation (AF), diabetes, chronic lung disease and stroke. Handling comorbidities is a challenging problem as it demands the systematic synthesis of medical knowledge covering multiple conditions and applying this synthesized knowledge with respect to the patient's profile whilst maintaining clinical pragmatics.

In Canada, there is a significant care gap in the management of the cardiovascular diseases particularly the discrepancy between the applied care process versus the evidence-based care processes [1]. The application of evidence based clinical algorithms, which include Clinical Practice Guidelines (CPG) and Clinical Pathways (CP), at the point-of-care have enormous potential to reduce this care gap. CPG entail *medical knowledge* whereas CP entail *operational knowledge* about how to execute the CPG—i.e. it entails the institution-specific protocols specifying the actual sequencing, decisions and scheduling of clinical tasks, as per the CPG. However, despite the availability of large number of paper-based CPG and CP they are underutilized at the point-of-care due to the fact that paper-based CPG and CP are difficult to incorporate in active clinical practices.

Decision support to handle comorbidities is quite complex because it requires to reconcile/align the interventions recommended by multiple disease-specific CPG/CP

whilst ensuring clinical appropriateness, patient safety and task pragmatics. We argue that it is clinically prudent to handle comorbidities by applying CPG/CP to ensure evidence-based and standardized patient care.

In this paper, we present our knowledge management solution for clinical decision support to handle co-morbid diseases. We are particularly interested in primary care settings as they are the first point of contact with patients. We present the COMET (Co-morbidity Ontological Modeling & ExecuTion) system that is capable of handling three patient care scenarios: (i) patient has CHF; (ii) patient has AF; and (iii) patient develops a co-morbidity of either AF or CHF. COMET is designed to address the knowledge needs of General Practitioners (GP) and is based on semantic web technologies for knowledge representation and execution.

2. Solution Approach

The use of Semantic Web technologies (SWT) for clinical decision support and care planning is quite profound. The modus operandi is to model the medical knowledge in semantically-rich formalism, such as an ontology, and then execute the modeled knowledge based on patient data [2, 3, 4, 5, 6]. Semantically rich knowledge representation formalisms such as OWL have been used to describe concepts in the medical domain and to computerize CPG. More so, OWL based representation of CP allow their enactment in a clinical setting [7, 8, 9].

To provide decision support to handle comorbid diseases our approach is built on the individual disease-specific CP, whereby we aim to align multiple CP (for the comorbid diseases) along common clinical care activities to develop a unified knowledge model that encapsulates the medical and procedural knowledge to handle comorbid diseases. From a knowledge management perspective, there are two main approaches to align multiple CP to handle comorbidities: (a) Aligning CP at the knowledge modeling level; and (b) Aligning CP at the knowledge execution level [10]. In our work, we pursued CP alignment at the knowledge modeling level, whereby in a planned manner we aligned ontologically-modeled CP by establishing conceptual mapping between their common concepts to realize a comorbid CP.

Our approach leverages on semantic web technologies in terms of healthcare knowledge modeling, representation of healthcare knowledge using ontologies and the execution of the ontologically-modeled knowledge to provide CPG-based recommendations. Our solution approach entails four main aspects:

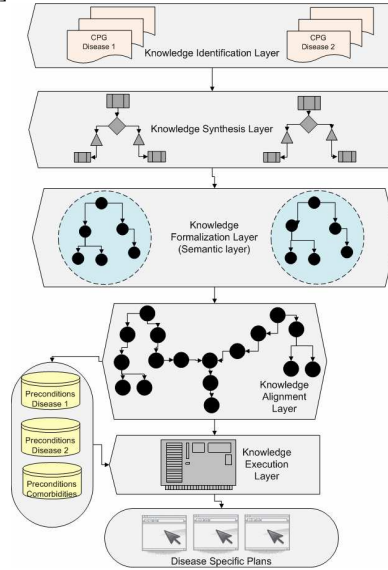


Figure 1: Our solution approach

1. *Knowledge identification and synthesis* involves the development of specialized CP for handling the diagnosis and management of (i) CHF, (ii) AF and (iii) co-morbid CHF-AF. This involves the derivation of clinically pragmatic workflow and recommendations from a large number of existing CPG for CHF and AF. We developed two new CP for CHF and AF that target the clinical needs of GP, especially those working in Nova Scotia.
2. *Knowledge modeling* involves the ontology-based modeling of the CHF and AF CP in order to semantically describe the diagnostic and treatment concepts in terms of clinical processes, tasks, decision-points, patient data, recommendations and information items. The knowledge modeling exercise resulted in an elaborate *CP ontology* that describes the CHF and AF diagnostic and treatment concepts and their interrelationships, and instantiates the CP for CHF and AF.
3. *Knowledge alignment* involves the synthesis of the individual CP of comorbid diseases to yield a comorbid knowledge model. Through knowledge modeling the individual ontologically-modeled CP for CHF and AF were aligned by establishing relationships between the care processes of CHF and AF, resulting in an ontologically-modeled co-morbid CHF-AF CP.
4. *Knowledge execution* involves translation of the modeled CP to deliver clinical decision support to GP to handle CHF, AF and comorbid CHF-AF.

3. Knowledge Synthesis: Development of CHF & AF CP

This phase involved the identification of relevant knowledge sources and then their synthesis to develop specialized CP for CHF and AF that target the decision needs of GP. The primary sources of knowledge were paper-based CPG [11, 12]. We also incorporated information from locally developed treatment protocols to account for the scheduling of treatment tasks and resources availability at GP clinics in Nova Scotia. We engaged domain experts as well to seek their experiential knowledge.

Given the complex nature of the CPG, a key challenge was to identify the essential task-specific heuristics in terms of decision logic, decision options, actions and sequence of the actions in accordance with general practice setting. We identified the essential do's and don'ts of practice by using class I and class IIa recommendations in the CPG. To incorporate initial clinical presentation and tests, we used the Boston Criteria [13] which uses a point score system for the diagnosis of CHF based on symptoms, physical and radiological findings. The knowledge synthesis exercise yielded algorithms for the diagnosis of CHF (shown in Figure 2) and AF.

In the final step of this phase, we developed two CP—one for CHF and another for AF. The CP developed were in line with the care procedures and standards at QEII hospital in Halifax (Canada). Development of the CHF and AF CP involved setting the systematic ordering and scheduling constraints—such as the sequencing, concurrence, branching and synchronization—between the various tasks specific heuristics distilled from the CPG, and relating them through decision and ordering constructs. Given the complexity of CHF and AF, each CP constituted multiple care plans corresponding to various patient care activities and interventions, such as the initial clinical assessment, diagnostic investigations, pre-treatment evaluation and correction of electrolytes, treatment plans and patient education.

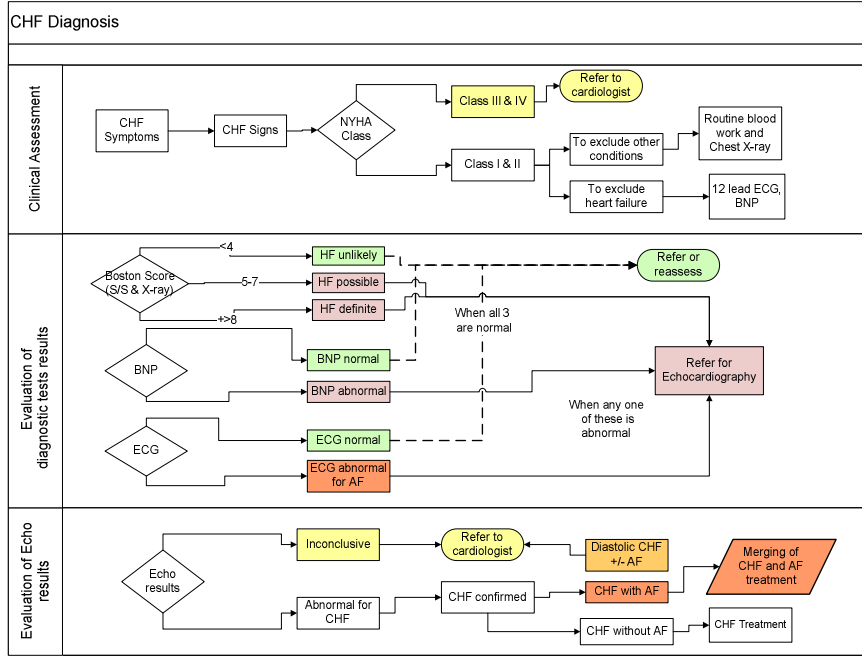


Figure 2: Algorithm to diagnose CHF derived from different knowledge sources.

4. Knowledge Representation: Developing the CP Ontology

In this phase, we conceptualized the CP knowledge in terms of main concepts, relationships between them and their restrictions in order to outline the dependencies between the care plans that are to be represented in the CP ontology. Figure 3 presents an example of conceptualization of the both the CP's declarative and procedural knowledge for the task specific heuristic “*When ACEI cannot be tolerated due to new or worsening cough, substitute ARB for ACEI*”. Here the relationship ‘has adverse effect’ is a declarative (factual) relationship between two objects, i.e. ‘ACEI’ (Medication) and ‘intolerance due to new or worsening cough’ (Adverse Effect), the relationship ‘is followed by’ is a procedural relationship between a fact ‘ACEI has adverse effect intolerance due to new or worsening cough’ and an action ‘Substitute ARB for ACEI’, and ‘is followed by task’ is a temporal relationship depicting dependency between the actions.

The key activity in this phase was the development of an

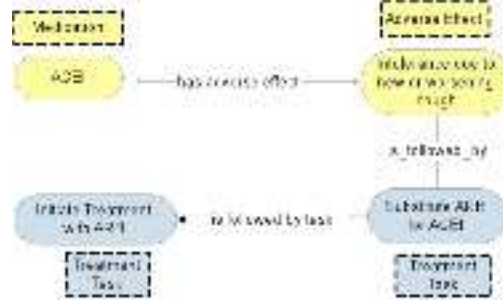


Figure 3: Conceptualization of a CP knowledge using declarative and procedural relationships

ontological model to represent the CP knowledge. The ontology is built in OWL using the ontology editor protégé. We provide a description of the structure of the CP ontology. For the purpose of clarity, class names will be written in UPPERCASE letters. The properties (i.e. relationship between the classes) will be *italicized* and the Individuals (instances) will be capitalized.

The main concepts are represented as a hierarchy of high-level classes as follows:

- PATIENT refers to individual patients who enter the system.
- CLINICAL_PATHWAY_ENTRY_POINT refers to points in the CP where a patient depending on his/her current clinical status may enter in the pathway.
- DIAGNOSTIC_CONCEPT refers to all the concepts related to diagnosis of CHF or AF, such as history, physical exam and tests results etc.
- MEDICATION refers to all the medication groups involved in the treatment of CHF or/and AF, such as ACEI, BB, Diuretics, calcium channel blockers, etc.
- TASK refers to diagnostic and therapeutic tasks in the CHF or/and AF CP.
- TREATMENT_CONSTRAINT refers to the different kinds of constraints on the treatment of CHF or/and AF, such as treatment contraindications, medication dosage, uptitration schedules, treatment monitoring etc.
- DECISION_OPTION refers to all the decision points in the CHF and AF CP.
- TEMPORAL_CONCEPT represents all time annotations, such as wait interval between two tasks, or the frequency of certain actions etc.
- STATUS refers to current clinical status of the patient.

The CP ontology is designed as a care flow model, whereby it models the patient's induction into the care pathway and captures his/her transition through various stages of diagnosis and treatment depending on whether the patient has a single disease or comorbidity. The care pathway is modeled through a series of properties that relate these main classes and sub-classes. Below are some exemplar properties.

PATIENT is related to CLINICAL_PATHWAY_ENTRY_POINT through an object property called *has_pathway_entry_point*, so that PATIENT is its domain and CLINICAL_PATHWAY_ENTRY_POINT is its range. PATIENT also has the datatype properties, *has_name*, *has_address*, *has_date_of_birth* to represent the patient's personal and demographic information.

DIAGNOSTIC_CONCEPT is regarded as an abstract class that subsumes more specific classes, i.e. CLINICAL_ASSESSMENT, PHYSICAL_EXAM, INVESTIGATION and INVESTIGATION_FINDING. These sub-classes are further decomposed into more specific sub-classes denoting concrete concepts such as SYMPTOM, ABDOMINAL_EXAM, PULSE_RATE, CHEST_X-RAY_FINDING.

MEDICATION has sub-classes corresponding to all the medication groups involved in the treatment of comorbid CHF and AF.

TASK has a datatype property called *has_description* to provide clinicians necessary evidence from the CPG to execute a particular task in a particular sequence. TASK has two main sub-classes; DECISION_MAKING_TASK or NON_DECISION_MAKING_TASK. DECISION_MAKING_TASK is related to DECISION_OPTION through object property *has_decision_option*. We ensured that for a TASK to be a DECISION_MAKING_TASK, it is necessary for it to have at least one *has_decision_option* relationship with DECISION_OPTION.

The CP ontology captures a range of procedural rules identified during the earlier stage using OWL object properties, such as *has_pathway_entry_point*, *is_followed_by*

and *has_decision_option*. The property *is_followed_by* denotes a sequential relationship between objects such as two TASKS; DECISION_OPTION and a TASK; TREATMENT_CONSTRAINT and a TASK; TASK and a PATHWAY_ENTRY_POINT. Similarly, the property *has_decision_option* controls the procedural branching statements expressing the decision logic in the CP. Such a procedural rule is formalized by the *is_followed_by* relationship between DECISION_OPTION and TASK, where TASK is regarded as range for this property.

In summary, the CP ontology has 102 hierarchically arranged classes, 34 relations and more than 400 instances that cover the CP for CHF, AF and CHF+AF.

5. Knowledge Alignment: Developing a Co-morbid CP

From a knowledge modeling perspective, the requirements to handle comorbidities are (i) identifying common comorbid care activities; (ii) ensuring clinical tasks are not replicated; (iii) temporal relationships between the activities are clearly identified; (iv) preconditions for specific tasks are explicitly stated; (v) potential risks and harmful events when aligning the comorbid processes are noted; and (vi) care for comorbid diseases are coordinated to improve efficiency and patient safety.

We pursued the alignment of the CHF and AF CP at the knowledge modeling level—i.e. concepts and relationships were systematically aligned within the CP ontology to realize an instantiation of a comorbid CHF-AF CP. We explain below the alignment of the CHF and AF CP.

The class CLINICAL_ENTRY_POINT is instantiated by a number of plans to account for the various points during the patient care process. These plans are classified into: (i) discrete care plans that are valid only when a patient has either CHF or AF, and (ii) comorbid plans that are valid when patient has a concurrent illness.

Figure 4 gives a schematic of the alignment of the CHF and AF plans, where the last two care plans from the CHF pathways—i.e. ‘CHF entry point 4 - Pre-treatment electrolytes correction and assessment’ and ‘CHF entry point 5 - initiation of the treatment of heart failure’ are aligned with the comorbid plans—i.e. ‘CHF-AF entry point 1- Thromboprophylaxis for patients with AF and CHF’, and ‘CHF-AF entry point 2- Treatment for atrial fibrillation for patient with heart failure’.

To model the alignment of the treatment plans we established relationships between four main classes: PRE-TREATMENT_DECISION_TASK, TREATMENT, DRUG_ADMINISTRATION_DECISION_TASK and PHARMACOLOGICAL_DECISION_TASK. This also included the alignment of three main properties: *has_decision_option*, *is_followed_by* and *apply_to* which formed a compact model to evaluate preconditions such as signs of fluid overload or presence of any risks to comorbid treatments. These preconditions are modeled as instances of PRE-TREATMENT_DECISION_TASK. For example, the instances ‘Determine any contraindication to ACEI’, ‘Determine any contraindications to Digoxin’, or ‘Determine any risk factors of digitalis toxicity’ are responsible for checking the presence of any contraindication or serious risk associated with any medication, and ‘Determine presence of any signs of fluid over load’ is responsible to check if the patient needs treatment with diuretics. We related PRE-TREATMENT_DECISION_TASK to DECISION_OPTION through property

has_decision_option, such that the instances of *DECISION_OPTION* are available as options to the physician—i.e. decision options such as ‘ACEI is not contraindicated’ or ‘No risk factors associated for digitalis toxicity’ or ‘ACEI is contraindicated due to’ and ‘digitalis toxicity risk factors are present’. The last two instances of *DECISION_OPTION* are related through the property *apply_to_clinical_feature* to *CONTRAINDICATION* which is a sub-class of *TREATMENT_CONSTRAINT*.

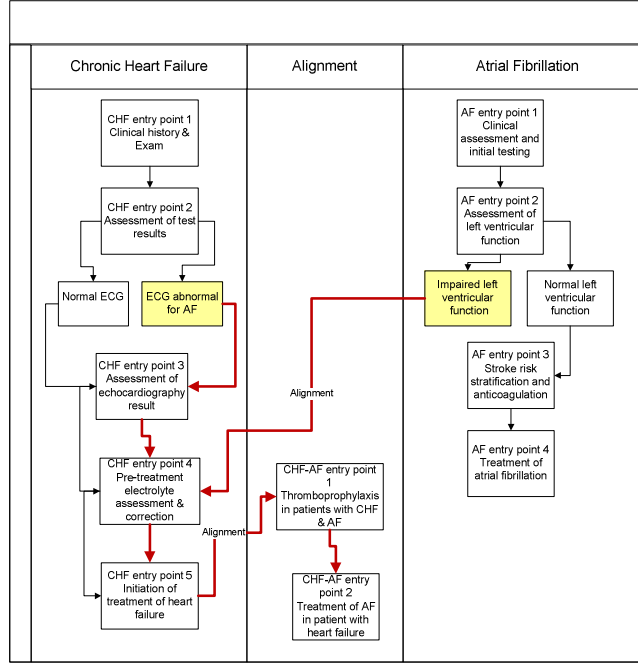


Figure 4: Aligning CHF and AF plans. The red arrows indicate the alignment between the plans of CHF and AF to handle comorbid CHF+AF.

In the CP ontology we modeled various therapy related constraints when handling comorbidities, such as: checking the presence of contraindications, potential risk factors, uptitration of drugs, informing the patient about side-effects and so on.

This phase yielded an ontologically-modeled CP for comorbid CHF and AF, that was derived by systematically aligning the independent CHF and AF CP.

6. Knowledge Execution to Handle Comorbidities

The COMET system executes the ontology-based CP to provide CP based recommendations to GP to handle CHF, AF and comorbid CHF+AF. Knowledge execution involved the traversal of the CP workflow, as modeled within the CP ontology, where each state of the workflow contains two elements: (i) actions to be performed whilst satisfying relevant constraints; and (ii) potential next state [9]. To execute the CP, we used the Resources, Properties, and Property-Values of the CP ontology. The potential property-values are either a pre-specified range, or of type

Resource. This allows the presentation of a range of property-values to the user—the selections by the user determine the next step in the CP.

COMET assists GP through a series of screen (web pages) to help collect the patient information and then provide CP-based recommendations and actions. Figure 5 shows a typical patient data input screen.

Figure 5: Screen to provide the patient's cardiac history, symptoms and CAD risk factors. The rightmost pane displays the next step: 'Perform physical exam'.

Based on the patient information, the patient is placed in a diagnostic class, say NYHA class II. Next, COMET recommends tests and based on the results prompts the GP to calculate a cumulative Boston criteria score. Given the high Boston score of the patient, COMET recommends echocardiography for confirmation of the diagnosis and the assessment of left ventricular.

COMET proceeds with initiation of heart failure therapy with inquiries to check the presence of any contraindications to medications. For uptitration of drugs, COMET presents a separate uptitration pathway in a new tab (see figure 7).

Figure 6: Screen showing the launch of comorbid CHF and AF pathway

Figure 7: Four different CP are simultaneously active, as illustrated by the 4 tabs. Two CP are the uptitration of drugs, whereas two are for managing the disease.

Suppose this patient complains of palpitations in addition to the above clinical features. This launches the comorbid CHF+AF CP (figure 6) and subsequently medication for the treatment of CHF and AF are concomitantly prescribed to the patient as shown in figure 7. In this manner, COMET traverses through all the active

CP (see the multiple tabs in figure 7), and at each step provides advice to monitor for treatment risks, contraindications, adverse events and prescription of assessments and medications. At the end of a CP COMET presents education material for the patient.

7. Evaluation

Evaluation of COMET was carried out in three stages: (i) Evaluation of the CP ontology for logical consistency; (ii) Internal validation based on clinical cases; and (iii) External validation by domain experts.

The CP ontology was evaluated for Consistency, Completeness and Conciseness [14]. We used an open source DL reasoner—Pellet to perform subsumption tests to derive concept satisfiability and consistency. The results conclude that the CP ontology is consistent and satisfiable. To evaluate the ontology for completeness, we instantiated the CHF and AF CP. Our results indicate that the ontological definitions of structural criteria such as necessary and sufficient conditions of a predicate, domain and range of relations, generalization and specialization of classes, etc., have adequate representational capacity to capture comorbid domain and procedural concepts. We used Pellet to check the conciseness of the CP ontology by running tests on the classes to compute the inferred class hierarchy. Our tests did not show any redundant arc in the ontology, hence the CP ontology was deemed to be concise.

The external validation of COMET was performed by one cardiologist and two GP. The cardiologist examined the medical knowledge and suggested some changes to the medical procedures, decision values and drug choices. We were able to readily incorporate these recommendations to our CP knowledge model. GP-1 recommended the addition of certain tests such as HbA1c (glycosylated hemoglobin) and thyroid-stimulating hormone to the CP. We were able to add these two tests to the individual list of class INVESTIGATION in the CP ontology. GP-2 was keen to use COMET to treat patients with milder symptoms of CHF and AF. The GP felt that COMET can be very beneficial to GP because it provides detailed information for each recommended task, which is necessary for clinical decision making for comorbid diseases.

8. Concluding Remarks

We have presented a knowledge management approach for handling comorbid diseases through (a) developing CP from various evidence based sources, (b) modeling the knowledge within individual disease-specific CP in terms of a CP ontology, and (c) *aligning* the ontologically-modeled CP of multiple co-morbid diseases to generate a unified knowledge model that contains procedures specific to the handling of comorbid diseases—these procedures are drawn from disease-specific CP. The comorbid CP model helps to (a) avoid duplication of intervention tasks, resources and diagnostic tests; (b) re- uses results of common activities; (c) ensures that different clinical activities, across different CP, are clinically compatible and their simultaneous application does not comprise patient safety; and (d) standardizes care across multiple institutions. We posit that alignment of CP is best pursued at the knowledge modeling level as it allows for a priori validation of clinical tasks for handling comorbid diseases, whereas alignment of CP at the knowledge execution

level may lead to inconsistencies in terms of clinical pragmatics. Our COMET system offers knowledge translation whereby GP are able to access evidence-based recommendations for the management of comorbid diseases at the point-of-care.

Acknowledgements: *This research has been supported by grants from Green Shield Canada Foundation. The author thanks Dr. Jafna Cox for his guidance in developing the clinical pathways and evaluating the system.*

References

1. Tremblay, G.J.L., Drouin, D., Parker, J., Monette, C., Cote, D.F., Reid, R.D.: The Canadian Cardiovascular Society and knowledge translation: Turning best evidence into best practice. *Can J Cardiol.*, 20(12), 119--1198 (2004)
2. Prcela, M., Gamberger, D., Jovic, A. (2008) Semantic Web Ontology Utilization for Heart Failure Expert System Design. In Andersen, S.K. et al. (Eds), *eHealth Beyond the Horizon-Get IT There*, IOS Press, Amsterdam, pp. 851--856
3. Colantonio, S., Martinelli, M., Moroni, D., Salvetti, O., Perticone, F., Sciacqua, A., Conforti, D., Gualtieri, A.: An Approach to Decision Support in Heart Failure. In Semeraro, G., Di Sciasco, E., Morbidoni, C., Stoermer, H. (Eds). In *Proceedings of SWAP 2007, the 4th Italian Semantic Web Workshop*, CEUR-WS.org, Bari, Italy (2007).
4. Casteliero, M.A., Diz, J.J.D.: Clinical Practice Guidelines: A case study of combining OWL-S, OWL and SWRL. *Knowledge-Based Systems*, 21(3), 257--255 (2008)
5. Dasmahapatra, S., Dupplaw, D., Hu, B., Lewis, H., Lewis, P., Poissonnier, M., Shadbolt, N.: Ontology-Based Decision Support for Multidisciplinary Management of Breast Cancer. In 7th International Workshop on Digital Mammography, Chapel Hill, NC (2004)
6. Abidi, S., Abidi, S.S.R., Butler, L., Hussain, S. (2009) Operationalizing prostate cancer clinical pathways: An ontological model to computerize, merge and execute institution-specific clinical pathways. In Riano, D. (Ed.) *Knowledge Management for Health Care Procedures*, ECAI 2008 Workshop K4Help 2008, LNAI 5626, Springer, Berlin, pp. 1-12
7. Alexandrou, D., Xenikoukakis, F., Mentzas, G.: SEMPATh: Semantic Adaptive and Personalized Clinical Pathways. In: *International Conference on E Health, Telemedicine and Social Medicine*, pp. 36-41. IEEE Computer Society (2009)
8. Ye, Y., Jiang, Z., Diao, X., Yang, D., Du, G. :An ontology-based hierarchical semantic modeling approach to clinical pathways workflow. *Computer in Biology and Medicine*, 39, 722--732 (2009)
9. Danyal, A., Abidi, S.R., Abidi, S.S.R. (2009) Computerizing Clinical Pathways: Ontology-Based Modeling and Execution. K.-P. In Adlassnig et al. (Eds). *Medical Informatics in a United and Healthy Europe*, IOS Press, pp. 643-647
10. Abidi, S.R., Abidi, S.S.R.: Towards the Merging of Multiple Clinical Protocols and Guidelines via Ontology-Driven Modeling. In: *12th Conf. on Artificial Intelligence in Medicine*, pp. 81-85. Springer Verlag (2009)
11. Arnold, J.M.O. et al.: Canadian Cardiovascular Society consensus conference recommendations on heart failure 2006: Diagnosis and Management (Special Article). *Can J Cardiol.*, 22(1), 23--45 (2006)
12. Hunt, S.A. et al.: ACC/AHA Guideline Update for Diagnosis and Management of Chronic Heart Failure in Adults-Summary Article. *Circulation*, 112, 1825--1852 (2005)
13. Yturralde, F.R., Gaasch, W.H.: Diagnostic Criteria for Diastolic Heart Failure. *Progress in Cardiovascular Diseases*, 47(5), 341-319 (2005)
14. Gómez-Pérez, A.: Ontology Evaluation. In: *Handbook of Ontologies*, pp. 251--271, Berlin: Springer (2004)

Inducing Decision Trees from Medical Decision Processes

Pere Torres, David Riaño, and Joan Albert López-Vallverdú

Research Group on Artificial Intelligence (Banzai),
Universitat Rovira i Virgili, Av. Països Catalans 26, 43007 Tarragona, Spain

Abstract. In medicine, decision processes are correct not only if they conclude with a right final decision, but also if the sequence of observations that drive the whole process to the final decision defines a sequence with a medical sense. Decision trees are formal structures that have been successfully applied to make decisions in medicine; however, the traditional machine learning algorithms used to induce these trees use information gain or cost ratios that cannot guarantee that the sequences of observations described by the induced trees have a medical sense. Here, we propose a slight variation of classical decision tree structures, provide four quality ratios to measure the medical correctness of a decision tree, and provide a machine learning algorithm to induce medical decision trees whose final decisions are both correct and the result of a sequence of observations with a medical sense. The algorithm has been tested with four medical decision problems, and the successful results discussed.

1 Introduction

The induction of decision trees (DT) is usually based on the concept of set of instances, where an instance is a description of a case on which a decision has been made [11]. However, very often, decisions are based not on the description of cases but on sequences of observations that drive the whole process to final decisions. This is the case in medicine, where decision processes as diagnoses or prescriptions use to be the consequence of consecutive questions and observations made during one or more encounters of a health care professional with the patient on which the decision has to be made.

In such sensitive settings, concluding with a final correct decision is as much important as asking the correct questions and ask them in a correct medical order [9]. For example, deciding about the application of a diuretic to a patient with hypertension will never be accepted as medically correct if a previous analysis does not confirm creatinine, sodium or potassium in blood.

Previous works have shown that DTs can extend their traditional role as classifiers and become a successful way of representing protocols about medical decision processes. In [9], for example, DTs have been created by experts in order to make decisions in a broad range of medical problems. Our hypothesis here is that medically correct DTs can also be automatically induced from health care databases whose instances describe previous medical decision processes.

This fact, defines a specific inductive learning paradigm that artificial intelligence and process management have faced under the respective names of *sequence learning* [14] and *workflow mining* [15]. However and in spite of the successful application of DTs in medicine [9, 10], we are not aware of works on sequence learning (or workflow mining) algorithms for the induction of DTs. Certainly there are alternatives to the traditional information gain approach [11] of DT induction algorithms (e.g., the use of cost functions [4], medical adherence ratios [5], or combinations of medical criteria [6]), but these approaches require the users to provide not only the set of instances to learn from, but also additional health care knowledge to guide the algorithm in the selection of the best next DT question or decision. Here, our viewpoint is to free the user from providing this knowledge, and make the learning algorithm to automatically exploit the knowledge that is embedded in the training instances. We are also aware of good results modeling decision processes with structures different than DTs (e.g., Petri Nets [15] or Bayesian Networks [7, 8]), however our concern here is to dispel the doubts about the possibility of automating the machine learning of medically correct DTs that represent medical decision processes.

Pros and cons of using DTs in real medical decision processes include, in the pros [10], that DTs are reliable, effective, accurate, easy to understand and to use, successfully applied in different areas of medical decision making, and easily validated by health care professionals. On the cons [3], good medical DTs require a learning process based on enough quality instances that are not always available in medical settings. Moreover, DTs may be sensitive to overfitting and very strict in the order of the questions, which means that the answer to one question must be known before progressing in the decision process. In medicine, not all the questions have an immediate answer or they may entail high health and economic costs [4, 6] that delay or block the decision processes represented in DTs. In spite of these drawbacks, artificial intelligence scientists and health care professionals agree that the use of DTs for medical decision making still pays off [3, 9, 10].

In this context, we propose a new inductive learning algorithm that captures the decision processes followed by health care professionals and combine them in a medical DT. These professional decision processes are found out from health care information systems where they are stored as instances of past decisions.

In section 2 we introduce a formalization of the concepts medical decision process (DP), medical decision tree (MDT), and define four complementary measures of quality to produce good medical DTs. In section 3 the new algorithm to induce medically correct MDTs is explained. The algorithm was tested with patient cases of 4 different medical decision problems taken from [9]. In section 4, the results of these tests are commented at the technical and medical levels.

2 Medical Decision Processes

In medicine, decision processes as diagnosis or prescription can be complex and multi-step [2]. Formally speaking, here we define a decision process (DP) as

a sequence of questions (or observations) that conclude with a final decision. Provided a finite set of questions Q and a finite set of feasible decisions D , the sequence $(q_{i_1}, q_{i_2}, \dots, q_{i_{n_i}}; d_i)$ represents a decision process for the patient case p_i that, after having asked the questions $q_{i_j} \in Q$ in the exact order of the sequence, the health care professional decides $d_i \in D$. Here, we consider that all the questions in a DP are different (i.e., the health care professional does not need to ask the same question twice) and also that not all the questions in Q are necessarily asked in a DP. The answer to a question q_{i_j} for patient p_i (i.e., $q_{i_j}(p_i)$) uses to have a direct influence on the next question(s) that health care professionals are more likely to ask (or to observe) about that patient.

We say that a DP on a concrete patient p_i is *medically correct* if there is not a medical evidence [13] that, for that patient:

- a) the questions must be asked in a different order,
- b) there are unavoidable questions missing in the sequence of questions,
- c) there are irrelevant questions in the sequence, and
- d) the final decision is wrong.

According to this definition, several medically correct DPs may exist for the same patient if none of them has a medical evidence against. For example, deciding whether a patient has *resistant hypertension* (RH) is a matter of observing that the patient has elevated blood pressure (observation $o1$) despite the application of an optimal three drug regimen (observation $o2$) that includes a diuretic (observation $o3$). The order in which these observations are made is irrelevant, therefore any DP containing these three observations must be considered medically correct (i.e., $(o1, o2, o3; RH)$, $(o2, o1, o3; RH)$, $(o3, o1, o2; RH)$, etc.).

2.1 Medical Decision Tree

A decision tree (DT) is a structure that describes decision processes that always start with the same question and concatenate questions in such a way that each possible answer to a question is followed by a new question or by a final decision. Therefore, a DT represents a set of decision processes that, in front of a given health care situation (i.e. patient condition) only one question or a decision is possible.

This is a hard restriction that puts some limitations to the use DTs to represent decision processes in real medical settings where, in front of a concrete health care situation, several questions or final decisions are possible (e.g., alternative treatments).

In order to overcome some of these limitations here, we define medical decision trees (MDTs) as extensions of DTs in which all the nodes can contain decisions and all decisions and questions in the tree have weights. Therefore, in front of a concrete health care situation, the MDT may recommend to continue the decision process with a new medically correct question or to conclude it with one among several medically correct final decisions.

MDTs describe a trade-off between how to provide concrete DPs for specific medical situations (i.e., what is the DP recommended for the current patient), and how to represent medical variability (i.e., what are the alternative DPs

for the current patient). As far as concreteness is concerned, if a concrete DP is demanded for a patient, the MDT is operated as it follows. Starting in the root node, for each visited node in the MDT, the weights of the answers to the question in the node and the weights of the decisions contained in the node are used to decide how to proceed: if the larger decision weight is greater than or equal to the result of adding the weights of all the answers, then the higher weighted decision is concluded, otherwise if the answer to the node question is (or it can be) known for the current patient, then the question is asked and the DP continues as the MDT indicates for that known answer. If the answer to the node question is not known (and the cost of knowing it is not affordable), then the answer with the highest weight is assumed and the MDT is continued as if that answer was the correct one¹. For example, the MDT in figure 1 after asking the edema distribution of a concrete patient, if the answer is regional the decision process proceeds through the left branch, if the answer is generalized through the right branch, but if the edema distribution of this patient is unknown, then the right branch (or right next question) is followed because it is weighted 111 while the left branch is only weighted 58. The way these weights are induced will be explained in the next sections.

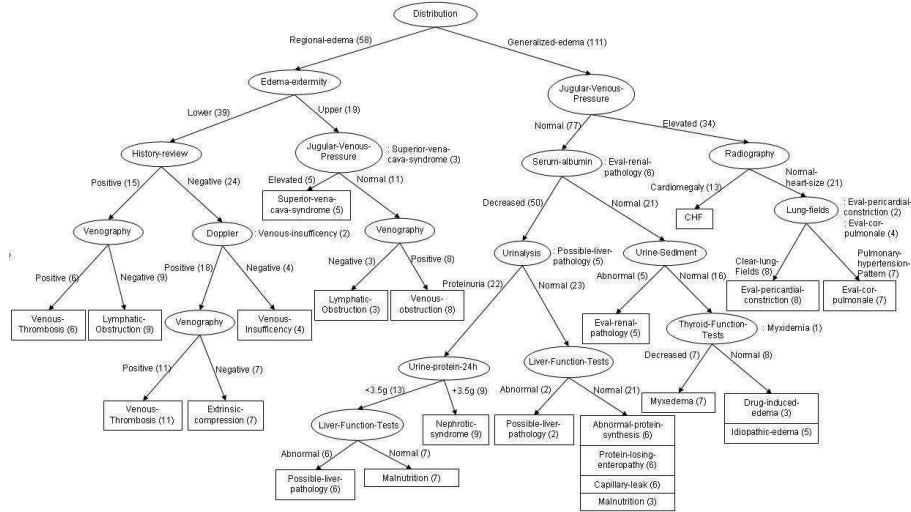


Fig. 1. MDT deciding on Edema cause

¹ This is the implemented solution to avoid the decision process to be blocked in the middle of a MDT waiting for an answer to a question when, for example, a final decision is required and we cannot wait to the conclusion of a medical test, or when a medical test cannot be applied to that patient, or when the resources of our health care center are limited (e.g. rural medicine).

As far as medical variability is concerned, MDTs are designed to capture medical DPs describing alternative solutions to the same medical problem. For example, for a patient p with $\text{distribution}(p) = \text{regional-edema}$, $\text{edema-extremity}(p) = \text{upper}$, and $\text{JVP}(p) = \text{elevated}$, the MDT in figure 1 describes alternative DPs as $(\text{distribution}, \text{edema-extremity}, \text{JVP}; \text{superior-vena-cava-syndrome})$ or $(\text{distribution}, \text{edema-extremity}; \text{superior-vena-cava-syndrome})$ with respective weights 5 and 3, but both possible.

In this work about the induction of MDTs from DPs, DPs that represent real medical actuations of health care professionals are taken as medically correct, whereas DPs obtained from the application of MDTs to concrete patients are taken as recommendations to health care professionals whose medical correctness depends on how close they are to DPs representing real medical actuations.

Given a DP $(q_{i_1}, q_{i_2}, \dots, q_{i_{n_i}}; d_i)$ on a patient p_i and a MDT, the application of this MDT to that patient defines a finite set of DPs of the form $(q_{\alpha_1}, q_{\alpha_2}, \dots, q_{\alpha_{n_\alpha}}; d_\alpha)$ in which the order of the questions may be different from the original DP (i.e., $q_{\alpha_j} \neq q_{i_j}$ for some $j(s)$), there may be some absent questions (i.e., $q_{i_j} \notin \{q_{\alpha_1}, q_{\alpha_2}, \dots, q_{\alpha_{n_\alpha}}\}$ for some $j(s)$), there may be some additional questions (i.e., $q_{\alpha_j} \notin \{q_{i_1}, q_{i_2}, \dots, q_{i_{n_i}}\}$ for some $j(s)$), or the final decision may be different (i.e., $d_\alpha \neq d_i$). These four situations give rise to four quality measures about the extent to which MDT recommendations are medically correct with respect to a set of (medically correct) real DPs.

2.2 Four Quality Measures of Medical Decision Trees

In order to measure the degree of medical correctness of a MDT in front of a set of medically correct DPs over a set of patients P , we define four quality measures: sequence resemblance, recall, conciseness, and decision accuracy.

Sequence resemblance measures the medical correctness of the sequences of questions that the MDT proposes for a set of patients in comparison to the DPs already followed for these patients. *Recall* (as the opposite of *Type I error*) evaluates the medically correct questions that the MDT asks when it is used with a set of patients. *Conciseness* (as the opposite of *Type II error*) rates the lack of questions asked by the MDT for a set of patients when these questions were not asked when the patients were assisted by health care professionals. Finally, *decision accuracy* calculates the similarity between the decision recommended by the MDT and the decisions taken by health care professionals.

These four measures are calculated as the average of the equations 1-4 for all the patients $p_i \in P$ that have followed a medically correct DP $(q_{i_1}, q_{i_2}, \dots, q_{i_{n_i}}; d_i)$ and for which the MDT recommends the DP $(q_{\alpha_1}, q_{\alpha_2}, \dots, q_{\alpha_{n_\alpha}}; d_\alpha)$; where $Q_i = \{q_{i_1}, q_{i_2}, \dots, q_{i_{n_i}}\}$, $Q_\alpha = \{q_{\alpha_1}, q_{\alpha_2}, \dots, q_{\alpha_{n_\alpha}}\}$, $\sigma(p_i)$ is a variation of Goodman & Kruskal's index [1] that defines C and D as the concordant and discordant queries in Q_α (with respect to Q_i). We say that the pair $(q_{\alpha_x}, q_{\alpha_y})$ is concordant with respect to Q_i if both queries are in Q_i and they are asked in the same relative order in both DPs (i.e., q_{α_x} is asked either before or after q_{α_y} in both DPs). Notice that two DPs with a sequence resemblance of 1.0 does not necessarily mean that they are equal, since this quality measure only considers the

queries that appear in both DPs. In equation 4, $\mu(d_\alpha|d_i)$ is the medical evaluation of deciding d_α when d_i should have been decided. In this work, we define $\mu(d_\alpha|d_i) = 1$ if $d_\alpha = d_i$, and 0 otherwise.

$$\sigma(p_i) = \frac{\#C}{\#C + \#D} \quad (1)$$

$$\rho(p_i) = 1 - \frac{\#(Q_i - Q_\alpha)}{n_i} \quad (2)$$

$$\kappa(p_i) = 1 - \frac{\#(Q_\alpha - Q_i)}{n_\alpha} \quad (3)$$

$$\delta(p_i) = \mu(d_\alpha|d_i) \quad (4)$$

2.3 Medical Correctness of a MDT with respect to a set of DPs

The previous four quality measures can be applied to analyze MDTs in terms of how close their decision processes are to a set of medical DPs. This analysis can help us to determine to what extent a MDT is applicable in a certain medical setting (i.e., is the procedure described by a MDT similar to the medical practice of a health care center?). If we have a set of DPs describing medically correct actuations of physicians in a center, the information contained in each one of these medically correct DPs $(q_{i_1}, q_{i_2}, \dots, q_{i_{n_i}}; d_i)$ define a medical case (i.e., patient) p_i with $q_{i_j}(p_i)$ the answer to question q_{i_j} ($j = 1, \dots, n_i$) for p_i . When the MDT is applied to this medical case p_i , one or more DPs of the form $(q_{\alpha_1}, q_{\alpha_2}, \dots, q_{\alpha_{n_\alpha}}; d_\alpha)$ are possible, each one with a quality vector $(\sigma(p_i), \rho(p_i), \kappa(p_i), \delta(p_i))$. The mean value of the four components is used as a measure of the medical similarity between the medically correct DP and the DP suggested by the MDT. Among all the DPs a MDT may suggest for p_i , the one with a greatest mean quality value is the best choice. The quality values of the best choices of a MDT for all the DPs in a set of DPs average a value that is taken as the medical correctness of that MDT with respect to that set of DPs.

2.4 Summary of MDT Basic Uses

MDTs as described in section 2.1 have the following basic uses:

1. On-line application of the MDT to a new patient: the weights in the decisions and in the answers of the tree are applied as it is explained in section 2.1 in order to progressively construct a DP for the new patient. This represents an application of MDTs as a decision support system.
2. Off-line supervision of an already conducted medical decision process: a MDT provides a set of alternative DPs for a medical problem on which a decision process has already been followed. As it is explained in section 2.3, the quality measures are used to identify the alternative which is more similar to the already conducted decision process. The quality of the chosen alternative

represents the medical support of the MDT to that decision process. Therefore, MDTs can be used to check medical correctness of medical actuations with respect to medical decisional models.

3. Validation of a MDT in a concrete medical setting: the quality measures can evaluate to what extent the application of a MDT in a medical setting is medically correct. If a representative set of DPs is available for a medical setting, the mean quality value calculated with the procedure explained in section 2.3 is a measure of the applicability of that MDT to that setting.

3 Induction of MDTs from Medical Decision Processes

Given a set of DPs $\{(q_{i_1}, q_{i_2}, \dots, q_{i_{n_i}}; d_i) : i = 1, \dots, m\}$ with questions in Q , decisions in D , and the length of the sequences of queries $n_i \in [0, \#Q]$, we call *immediate questions* the questions in $\{q_{i_1} : 1 \leq i \leq m, n_i > 0\}$ (i.e., the questions that are asked in the first place for some DP) and *immediate decisions* the decisions in $\{d_i : 1 \leq i \leq m, n_i = 0\}$ (i.e., the decisions that are made for some DP without any prior question).

A set of DPs as the one described in the previous paragraph is the input data of the algorithm to induce MDTs, in figure 2. The main idea of the algorithm is to make a node with all the immediate decisions in the input data. For each decision, the number of DPs in the data with this decision is taken as the weight of that decision. If there are not immediate questions in the data, this node is returned as the MDT. On the contrary, if there are immediate questions, the *best* question to be included in the new node is selected and used to partition the input data into as many disjoint data subsets as different observed answers to that question are in the data (see **Data'**). The algorithm is recursively called to construct a new MDT for each subset **Data'**. The new MDTs are connected to the node with branches that contain each respective answer together with a weight which is the cardinality of the corresponding **Data'**.

In this algorithm two are the steps that deserve special consideration: **Select_the_Best_Question** (in line 06), and **Select** the subset of DPs to be in **Data'** (in line 10). The best question among all the questions in a set of DPs is found with the following procedure: for each question q_i in the set of DPs, the number of these DPs for which q_i is asked in the j -th position (n_i^j) is calculated, then these values are used to detect the question q_k with $k = \arg \min_i \frac{\sum_j (j \cdot n_i^j)}{\sum_j n_i^j}$. This q_k is taken as the best question to be asked in the current node of the tree.

Once this q_k is found, the set of DPs in **Data** is transformed according to each one of the answers to that question. For each possible answer **a** to q_k , **Data'** is obtained from **Data** with the following rules: (1) if a DP in **Data** has an answer to q_k which is different to **a**, this DP does not belong to **Data'** (2) if a DP in **Data** has **a** as the answer to q_k , the question is removed from the sequence of questions of that DP, and the resulting DP is included in **Data'**, (3) if a DP in **Data** does not contain q_k but the sequence of questions is not empty, this DP is also in **Data'**, and (4) all the DPs in **Data** with an empty sequence of questions are not in **Data'**.

```

01 Algorithm MDT_Generator(Data)
02 begin
03   decisions := Data.Immediate_Decisions()
04   node := New_node(decisions)
05   if Data.Immediate_questions() is not empty
06     question := Data.Select_the_Best_Question()
07     node.Add_question(question)
08     answers := Data.Get_All_Possible_Answers(question)
09     for each a in answers do
10       Data' := Data.Select(question,a)
11       SubMDT := MTD_Generator(Data') --recursive call
12       node.Add_Branch(answer, SubMDT, Data'.Cardinality())
13     end for
14   end if
15   return node
16 end algorithm

```

Fig. 2. Tree generator algorithm

4 Testing the Induction of MDTs

Our first concern was to compare the medical correctness of the MDTs generated with our algorithm with the DTs generated with C4.5 [12], which is considered an efficient algorithm to induce decision trees, and compare their medical correctness. Three medical domains were used for this purpose: edema (ED), red eyes (RE), and preoperative (PO). For ED, 169 medically correct DPs deciding about 21 causes of edema were considered. For RE, 139 medically correct DPs about 21 possible medical causes of red eyes, and for PO, 204 medically correct DPs deciding whether a patient can proceed to surgery, postpone surgery, consult a specialist, or needs further assessment. The number of feasible questions (or observations) for each domain were 16, 14, and 10, respectively.

In all these decision domains, our algorithm was used to generate MDTs and the C4.5 algorithm (without pruning) to generate DTs, for all the DPs available. See, for example, the MDT obtained for ED in figure 1. Then, the medical correctness of each MDT and each DT was calculated with respect to the corresponding set of DPs, as section 2.3 explains. All the MDTs obtained with our algorithm reached the highest ratio possible 1.0 for sequence resemblance (σ), recall (ρ), conciseness (κ), and decision accuracy (δ), in all the cases. These good results provide evidence that our algorithm generates MDTs that represent all the alternative medical decision processes represented in a set of DPs. On the contrary, C4.5 obtained DTs with the quality measures that are show in table 1.

While our algorithm produces MDTs that ask all and only the questions in the medically correct DPs, the information gain approach (i.e. C4.5) proposes DTs that, in average, do not ask 63.26% of the questions that should have been asked (recall quality), and they ask 7.85% of questions that should not have been asked (conciseness quality). Moreover, in our approach, a sequence resemblance

Table 1. Measured qualities for C4.5 DTs

Domain	σ	ρ	κ	δ	mean
ED	1.0	0.3684	0.9194	1.0	0.8220
RE	1.0	0.4129	0.9763	1.0	0.8473
PO	1.0	0.3209	0.8689	1.0	0.7975
mean	1.0	0.3674	0.9215	1.0	0.8222
HT	1.0	0.9251	0.7295	1.0	0.9137

of 1.0 reflects that the DPs suggested by the MDT and the medically correct DPs are exactly the same (because recall and conciseness are also 1.0), whilst the sequence resemblance 1.0 of DTs is the result of comparing the sequential order of a small part of the questions in the DPs (because ρ and κ indicate that just an average 18.99% of the questions in the compared DPs are coincident).

Our algorithm was also used to detect whether there was any pattern describing the way that concrete physicians work. We considered 238 DPs in which a particular physician decided whether different patients with hypertension (HT) deserved medication or not. The MDT in figure 3 was obtained. It is interesting to observe that once the grade of hypertension is identified; if it is 3 (i.e. severe hypertension), medication is required; but if it is below 3 the same procedure that consists on progressively ask for the presence of a cardiovascular disease, diabetes, a target organ damage, and the number of cardiovascular risk factors is applied. The final decisions vary depending on the hypertension grade. The quality of this MDT is 1.0 for σ , ρ , κ , and δ . Comparatively, a DT was also produced with C4.5 that reached the quality ratios in row HT of table 1.

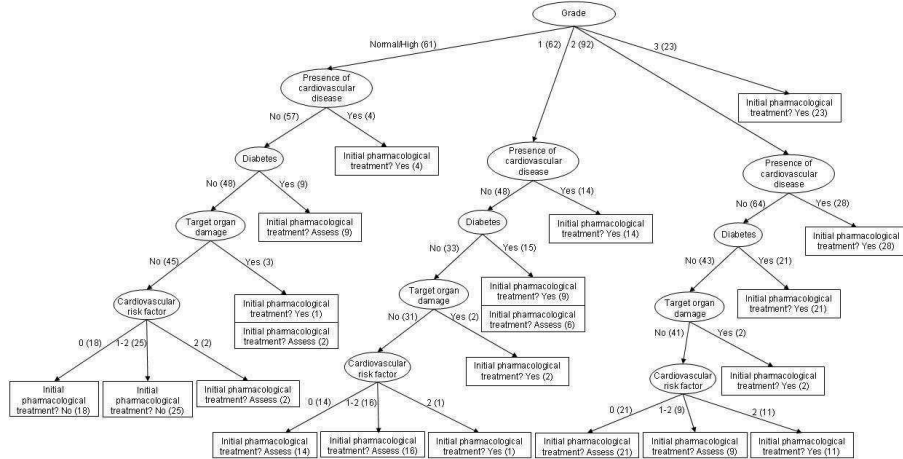


Fig. 3. MDT deciding on the convenience of pharmacological treatment in hypertension

In order to determine the degree of degradation of the quality of the DTs and MDTs generated, we used a percentage $p\%$ to randomly divide the set of DPs into two groups: one with $p\%$ of the DPs was used to construct the trees, and the other one with the remaining DPs to obtain the quality ratios of these trees. The percentage was progressively decreased till a sensitive reduction of the quality ratios of the obtained trees was observed for $p\%=30\%$ (see these quality ratios in table 2, where the values $a \pm b$ mean that the quality of a C4.5 DT was a , and our MDT was b above or below a depending on the sign).

Table 2. Measured qualities with cross validation

Domain	σ	ρ	κ	δ
ED	0.8911+0.0630	0.3588+0.5259	0.8045+0.0883	0.6897+0.1543
RE	0.5829+0.3552	0.4344+0.4326	0.8924-0.0202	0.8462-0.1039
PO	0.8784-0.0263	0.3025+0.5990	0.7724+0.1290	1.0000+0.0000
HT	0.9789-0.1653	0.9842-0.1853	0.5552+0.2722	0.9895+0.0105

The results show that the mean reductions of σ , ρ , κ , and δ for our algorithm were 11.05%, 13.70%, 12.66%, and 10.34%, and for the C4.5 algorithm 16.72%, -1.31%, 11.74%, and 11.87%, respectively. We also observed that, for ρ , MDTs performed above 50% better than DTs in ED and PO (and 43.26% in RE). For δ , MDTs were slightly better than DTs in ED (15.43%) and HT (1.05%), the same in PO, and slightly worst in RE (10.39%). It is worth noticing that, for ρ , the MDTs obtained with 30% of DPs still performed much better than the DTs obtained with all the DPs, in ED (51.62%), RE (45.41%), and PO (58.05%).

5 Conclusions and Future Work

In spite that the new algorithm to induce MDTs obtain trees that are much better than C4.5 in terms of asking all the important questions and not asking medically irrelevant questions, and still being as competitive as C4.5 in terms of decision accuracy (i.e. δ), there are still some aspects to improve. The most important one is the modification of the MDT structure to permit the representation of several, different, and medically correct questions at the same point of a DP. This non-deterministic decision behavior is frequently observed in medicine and MDTs should be able to represent it. Other future extensions of this work are the extension of tests to other medical domains, and the comparison of our MDTs with other successful decisional structures as Petri Nets [15] and Bayesian Networks [7, 8].

The authors want to acknowledge the help provided by Dr. A. Collado and Dr. Ll. Colomès, members of the SAGESSA Health Care Group.

References

1. Campello, R.J.G.B., Hruschka, E.R.: On comparing two sequences of numbers and its applications to clustering analysis *Information Sciences* **179** (2009) 1025–1039
2. Fauci, A.S., Braunwald, E., Kasper, D.L., Hauser, S.L., Longo, D.L., Jameson, J.L., and Loscalzo, J. (Eds.): *Harrison's Principles of Internal Medicine*, 17th Edition McGraw Hill (2008)
3. Kokol, P., Zorman, M., Stiglic, M.M., Malcic, I.: The limitations of decision trees and automatic learning in real world medical decision making. *Procs of the 9th World Congress on Medical Informatics* (1998) 529-533
4. Ling, C.X., Sheng, V.S., Yang, Q.: Test Strategies for Cost-Sensitive Decision Trees. *IEEE transaction on knowledge and data engineering*, **18**(8): (2006) 1055–1067
5. López-Vallverdú, J.A., Riaño, D., Collado, A.: Increasing acceptability of decision trees with domain attributes partial orders. *Proc. of the 20th IEEE CBMS, Maribor, Slovenia* (2007) 569–574
6. López-Vallverdú, J.A., Riaño, D., Bohada, J.A.: Health-Care criteria to improve the induction of decision mechanisms in medicine. *BMI* (submitted to)
7. Mani S., Aliferis C.: A Causal Modeling Framework for Generating Clinical Practice Guidelines from Data. *Artificial Intelligence in Medicine, LNCS 4594* (2007) 446–450
8. Mani S., Valtorta M., McDermott S.: Building Bayesian Network Models in Medicine: The MENTOR Experience. *Applied Intelligence* **22** (2005) 93-108
9. Mushlin S.B., Greene H.L.: *Decision Making in Medicine: An Algorithmic Approach*. Mosby Elsevier, 3rd Ed., 2010
10. Podgorelec V., Kokol P., Stiglic B., Rozman I.: Decision trees: an overview and their use in medicine. *J Med Syst.* **26**(5) (2002) 445–463
11. Quinlan, J.R.: *Induction of Decision Trees* *Machine Learning* **1** (1986) 81–106
12. Quinlan, J.R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann (1993)
13. Sackett, D.L.: Evidence based medicine: what it is and what it isn't. *BMJ.* **312**(7023) (1996) 71-72
14. Sun, R., Giles, C.L.: Sequence Learning: From Recognition and Prediction to Sequential Decision Making. *IEEE Intelligent Systems* **16**(4) (2001) 67–70
15. van der Aalst, W.M.P., van Dongen, B.F, Herbst, J., Maruster, L., Schimm, G, Weijters, A.J.M.M.: Workflow mining: a survey of issues and approaches *Data and Knowledge Engineering* **47** (2003) 237–267

Data-Aware Monitoring For Healthcare Workflows Using Formal Methods

Ji Ruan and Wendy MacCaul

Centre for Logic and Information, StFX University, Canada, {jruan, wmaccaul}@stfx.ca

Abstract. This paper presents an approach to monitor healthcare workflows using a logic-based formal method. We introduce a monitoring architecture with workflows and knowledge bases, and propose a logical language, FO-LTL-K, to express temporal and knowledge properties to be monitored. We formalize some of the norms for palliative care using the proposed logic and characterize the complexity of the model checking problem.

Keywords: Data and knowledge integration, Temporal knowledge representations, Logic-based methods, Ontologies, Complexity.

1 Introduction

A workflow [17] is the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules. Healthcare workflows have been proposed to improve the efficiency of health services delivery [10, 6]. Using model checking, initial attempts have been made to verify the correctness of healthcare workflows at the design stage [7, 12]. Monitoring (which uses data generated during care and relevant background healthcare knowledge) is necessary to determine the correctness of healthcare workflow execution. In this paper we develop an approach to monitor healthcare workflows using a logic-based formal method.

Logic-based formal methods provide a solid theoretical foundation for modelling and verifying complex systems. Three kinds of logic relate to our work: First-order logic (FOL), Temporal logic (TL), and Description logic (DL). FOL [16] enables us to represent different healthcare data types and quantify over their contents. TL is a formalisation describing temporal changes. It is applied to the verification of computer programs, or more broadly, to reactive systems [11]. Healthcare data reflects the dynamics of healthcare systems, such as the changes of health status of patients and the delivery of care, therefore we must reason about temporal changes. DL [1] models individuals, concepts, and roles. It is used in Artificial Intelligence for formal reasoning on the concepts of an application domain (e.g., Medicine), and is of particular importance in providing a logical formalism for knowledge bases, also called ontologies.

This paper is organized as follows: section 2 introduces palliative care and a monitoring architecture; section 3 discusses the representation of healthcare knowledge and data; section 4 proposes a logical language, **FO-LTL-K**, to express the properties to be monitored, and uses it to formalize some palliative care norms; section 5 provides the complexity of the model checking problem; section 6 discusses related and future work.

2 Palliative Care and A Monitoring Architecture

Palliative care refers to the physical, psychological, spiritual and practical care given to patients and their families when they are dealing with the issues associated with serious illness. As patients are usually part of a family, when care is provided, the patient and family are treated as a unit. The main focus of palliative care is to ease the suffering of the patient and his or her family and to help them to cope with their difficulties.

According to [3], a palliative care workflow consists of the following essential processes: *Referral*, *Consultation*, *Intake*, *Therapeutic Encounter*, and *Discharge*. These processes can be further refined as sub-workflows. For example, a therapeutic encounter is defined as a sub-workflow with the following processes: *Assessment*, *Information Sharing*, *Decision Making*, *Care Planning*, *Care Delivery*, and *Confirmation*. It is depicted in Fig. 1, in which the black triangle and square represent the start and end of the workflow respectively, the arrows indicate the order of the processes to be carried out, and the double-lined squares indicate compound processes. These processes can be further refined to include a greater level of detail.

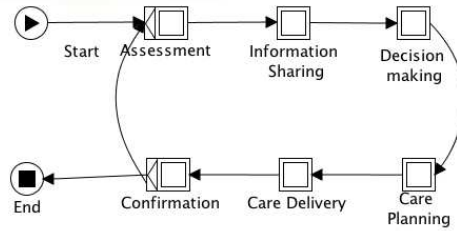


Fig. 1. A Therapeutic Encounter Workflow

Many stakeholders are involved in the processes of palliative care, e.g., Nurses, Physicians, Oncologists, Pharmacists, and Social workers. These stakeholders carry out the practical work and record various kinds of healthcare information. Their work is guided by certain requirements to ensure the quality of care and to prevent errors. In Canada, the norms of practice [3] are set by the Canadian Hospice Palliative Care Association to guide the process of palliative care. Norms are simple statements that provide a benchmark of normal or desired practice to which individuals or organizations providing palliative care can and should aspire. They are less rigid and less specific than standards, thus enabling individual organizations to develop their own standards of practice from the norms.

We propose a monitoring architecture to check whether the actual healthcare practice meets these norms. It has the following four components (Fig. 2): (1) a Workflow Management System (WfMS) which defines, creates and manages the execution of workflows through the use of a workflow engine able to interpret the process definition and interact with workflow participants; (2) a Data Management System (DMS)

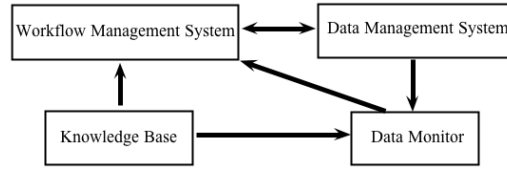


Fig. 2. A Monitoring Architecture

which stores and retrieves the information that is produced by the execution of workflows in WfMS; (3) a Knowledge Base (KB) which holds medical, organizational, and healthcare knowledge to support the execution of workflow; (4) a Data Monitor which monitors the data flow from the DMS and uses the knowledge from the KB to help healthcare stakeholders with decision-making.

3 Healthcare Knowledge and Data Representation

There are different kinds of information contained in a healthcare system. We classify them into two kinds in our architecture. The first is static, or changes infrequently over time—e.g., the knowledge of medication compatibility and the relationship of stakeholders in healthcare; the second is dynamic, hence may change frequently over time—e.g., the assessment result of pain may reduce from level 9 to level 3 due to proper treatment.

The first kind of information comes from the healthcare domain knowledge, and is maintained in the KB. There is a vast amount of knowledge related to healthcare, such as the knowledge of disease and human coping, the medication compatibility, and the responsibilities of different caregivers, etc. An ontology is a formal representation of a set of concepts within a domain and the relationships among those concepts. Extensive ontologies for the healthcare domain, such as the International Classification for Nursing Practices (ICNP) and the Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT), as well as numerous drug ontologies have been developed for the electronic exchange of clinical health information. Many ontologies are constructed using some fragment of DL. Efficient DL reasoners [15] are available, and are useful both in the development stage of the ontology (e.g., for consistency checking) and for querying the ontology. We use DL to encode knowledge related information to facilitate the use of DL reasoners.

The second kind of information is produced during the care delivery processes, and is managed in the DMS. A basic healthcare data unit is called a healthcare record, which has three required data fields: the patient's ID, a time stamp, and an electronic signature of the caregiver who provided the record. The time information is critical in healthcare since some tasks must be completed within a certain time limit, and some tasks should be repeated regularly. The granularity of time in palliative care is usually measured by hours or even days. According to the time stamp of each record, the healthcare records of a patient can be retrieved, and form a linear order. Thus, we use a linear temporal model to represent a patient's healthcare history. Apart from the three required data

```

<record>
<time>9am January 6, 2010</time>
<patientID> B9527 </patientID>
<pain>9</pain>
<weakness>6</weakness>
<medication> Med1 </medication>
<medication> Med2 </medication>
<actiontaken> Assessment; B9527</actiontaken>
<signature> Bob </signature>
</record>

```

Fig. 3. A healthcare record in XML format.

fields, each healthcare record will have other data fields, depending on what is intended to be captured at the moment. For example, it may include assessment results for pain, tiredness, depression, etc., or the medication that is prescribed, or other data. We use a standard XML format to encode and facilitate the exchange of such healthcare records.

Definition 1 (Healthcare Data Model). *A healthcare data model $\mathcal{M}$ is a finite sequence $r_1 r_2 \dots r_n$ where r_i is a healthcare record. A healthcare record contains tags and data. Data content is included in tags in the format: $\langle \text{tag} \rangle \text{content} \langle / \text{tag} \rangle$, and content items within the same tag can be separated using ‘;’. The whole record is wrapped by a pair $\langle \text{record} \rangle \langle / \text{record} \rangle$. In particular, patient’s ID, time stamp, and caregiver’s signature are tagged with $\langle \text{patientID} \rangle$, $\langle \text{time} \rangle$, and $\langle \text{signature} \rangle$ respectively.*

An example of a healthcare record in Fig. 3 shows (a) that the patient (ID ‘B9527’) underwent an assessment (which gives pain level 9 and weakness level 6) by caregiver Bob at 9am January 6, 2010, and (b) the prescribed medication (Med1 and Med2).

4 Specifications

Specifications express properties that should hold in a system. We propose a language called **FO-LTL-K**, which is a combination of a first-order linear temporal language and a description logic language. It is designed:

- to represent and quantify over the data contents in a health record, e.g., all the medications inside the tag $\langle \text{medication} \rangle$;
- to specify real-time requirements in different intervals, e.g., an assessment must be made within 3 hours (or 3 days);
- to specify temporal changes of data, e.g., in one assessment, the pain level is 4, but in the next assessment, it is 8;
- to specify healthcare knowledge expressions, e.g., medication Med1 is not compatible with medication Med3.

Definition 2 (Syntax). *The language **FO-LTL-K** is constructed from the alphabet:*

- Predicate symbols $P_1, P_2, \dots$, each of which is some fixed arity;
- Function symbols $F_1, F_2, \dots$, each of which is some fixed arity;
- Concept symbols (also called atomic concepts) $C_1, C_2, \dots$;
- Role symbols (also called atomic roles) $R_1, R_2, \dots$;
- Individual variables $x_1, x_2, \dots$; Individual constants $c_1, c_2, \dots$.

Terms t : Variables and constants are terms; $F(t_1, \dots, t_n)$ is a term if F is an n -ary function symbol and t_i is a term.

Concepts C, D : $C ::= C_i \mid \top \mid \perp \mid \neg C \mid C \sqcap D \mid \forall R.C \mid \exists R.C$, where, C_i is a concept symbol and R is a role symbol.

Formulas: $\varphi ::= t_1 = t_2 \mid P(t_1, \dots, t_n) \mid u_1 : C \mid (u_1, u_2) : R \mid \neg\varphi \mid \varphi \wedge \psi \mid \mathbf{X}\varphi \mid \varphi \mathbf{U} \psi \mid \forall x.\varphi(x)$, where t_i is a term, x is a variable, u_i is a constant or variable, P is a predicate, C is a concept, and R is a role. We use the following abbreviations for other connectives and operators: $C \sqcup D = \neg(\neg C \sqcap \neg D)$, $\varphi \vee \psi = \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi = \neg(\varphi \wedge \neg\psi)$, $\mathbf{F}\varphi = \top \mathbf{U} \varphi$, $\mathbf{G}\varphi = \neg\mathbf{F}\neg\varphi$, $\exists x.\varphi(x) = \neg\forall x.\neg\varphi(x)$.

A variable x is a free variable if it is not in the scope of any $\forall x$ or $\exists x$. A formula is a sentence if it does not have any free variables.

Intuitively, a temporal formula $\mathbf{X}\varphi$ means “in the next record, φ holds”, $\mathbf{F}\varphi$ means “in a future record, φ holds”, and $\varphi \mathbf{U} \psi$ means “ φ holds until ψ holds”. The precise meanings will be given in Definition 6. We introduce a formal definition of knowledge bases.

Definition 3 (Knowledge Base). A knowledge base $\mathcal{K}$ is a tuple $\langle \mathcal{T}, \mathcal{A} \rangle$, where $\mathcal{T}$ (the TBox) is a set of expressions of the form $C \sqsubseteq D$ ($R \sqsubseteq S$) and $C \equiv D$ ($R \equiv S$) for concepts C, D (and roles R, S), and $\mathcal{A}$ (the ABox) is a set of expressions of the form $(c_1 : C)$ and $((c_1, c_2) : R)$ for constants c_1, c_2 , concept C and role R .

Here are a few examples to show the intuitive meaning of knowledge expressions.

- Oncologist, Nurse, Physician, Caregiver, Manager : These are concepts that express the stakeholders associated with a care unit. $(Alice : \text{Nurse})$ asserts that Alice is a nurse, and $(Bob : \text{Manager})$ asserts that Bob is a manager. $\text{Nurse} \sqsubseteq \text{Caregiver}$ expresses that every nurse is a caregiver.
- HasSupervisor: This is a role relating two individuals. For example, $((Alice, Bob) : \text{HasSupervisor})$ asserts that Alice has a supervisor named Bob.

The following introduces the model that interprets the formulas in Definition 2.

Definition 4 (First-order Temporal Structure). A first-order temporal structure $\mathcal{H}$ is a tuple $\langle \Delta, \mathcal{I}, h_1 h_2 \dots h_n \rangle$, where Δ is a nonempty set that forms the domain, $\mathcal{I}$ is a knowledge interpretation, and $h_1 h_2 \dots h_n$ is a finite sequence of first-order interpretations. The interpretation $\mathcal{I}$ assigns every atomic concept C_i to a set $C_i^{\mathcal{I}} \subseteq \Delta$, every atomic role R to a set $R^{\mathcal{I}} \subseteq \Delta \times \Delta$, every constant c to an element $c^{\mathcal{I}} \in \Delta$, and every function F with n -arity to a function $F^{\mathcal{I}}$ from Δ^n to Δ , such that $F_i^{\mathcal{I}}(c_1^{\mathcal{I}}, \dots, c_n^{\mathcal{I}}) = c^{\mathcal{I}}$ whenever $F(c_1, \dots, c_n) = c$. Then $\mathcal{I}$ is extended to the concepts as follows:

$$\begin{aligned} \top^{\mathcal{I}} &= \Delta; & \perp^{\mathcal{I}} &= \emptyset \\ (\neg C)^{\mathcal{I}} &= \Delta \setminus C^{\mathcal{I}}; & (C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}} \end{aligned}$$

$$\begin{aligned}
(\forall R.C)^{\mathcal{I}} &= \{a \in \Delta^{\mathcal{I}} \mid \forall b.((a, b) \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}})\} \\
(\exists R.C)^{\mathcal{I}} &= \{a \in \Delta^{\mathcal{I}} \mid \exists b.((a, b) \in R^{\mathcal{I}} \text{ and } b \in C^{\mathcal{I}})\}
\end{aligned}$$

Each first-order interpretation h_i ($1 \leq i \leq n$) assigns an n -ary predicate symbol P to a relation P^{h_i} over Δ^n .

The reader will recall that we have partitioned all information into static knowledge and dynamic data. Therefore the knowledge interpretation $\mathcal{I}$ interprets the fragment of the language that does not change over time, i.e., concepts, roles, constants and functions; and the first-order interpretation h_i interprets the rest of the language.

The following defines a relation ' $\models$ ' between a first-order temporal structure and a knowledge base, which is sufficient to ensure that the two are consistent.

Definition 5 (Knowledge Base Satisfaction). Given a first-order temporal structure $\mathcal{H} = \langle \Delta, \mathcal{I}, h_1 h_2 \dots h_n \rangle$ and a knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, we say that $\mathcal{H}$ satisfies $\mathcal{K}$, written as $\mathcal{H} \models \mathcal{K}$, if the following hold:

- for all $C \sqsubseteq D$ in $\mathcal{T}$, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, and for all $C \equiv D$ in $\mathcal{T}$, $C^{\mathcal{I}} = D^{\mathcal{I}}$.
- for all $R \sqsubseteq S$ in $\mathcal{T}$, $R^{\mathcal{I}} \subseteq S^{\mathcal{I}}$, and for all $R \equiv S$ in $\mathcal{T}$, $R^{\mathcal{I}} = S^{\mathcal{I}}$.
- for all $t : C$ and $(t_1, t_2) : R$ in $\mathcal{A}$, $t^{\mathcal{I}} \in C^{\mathcal{I}}$ and $(t_1^{\mathcal{I}}, t_2^{\mathcal{I}}) : R^{\mathcal{I}}$

A first-order temporal structure $\mathcal{H} = \langle \Delta, \mathcal{I}, h_1 h_2 \dots h_n \rangle$ can be derived as follows.

1. The Δ consists of a finite set of strings that is used to specify the data contents;
2. The interpretation $\mathcal{I}$ is specified according to the setting of care, so that $\mathcal{H} \models \mathcal{K}$; this makes sure that the interpretation is consistent with the knowledge base;
3. The sequence of first-order interpretation $h_1 h_2 \dots h_n$ is derived from a healthcare data model (Definition 1). Suppose h_i is associated with the i -th record m_i in a corresponding healthcare data model $\mathcal{M}$. Thus the h_i s act as the states in a temporal system. Each predicate P that corresponds to a tag occurring in m_i is assigned to the set of data inside the tag, written as P^{h_i} , otherwise it is assigned to an empty set. For example, Fig. 3 shows a healthcare record m ; suppose m is associated with h , the tag $\langle \text{medication} \rangle$ is associated with a unary predicate P_1 , and no tag is associated with unary predicate P_2 , then $P_1^h = \{Med1, Med2\}$, $P_2^h = \{\}$.

Definition 6 (Semantics). Given a first-order temporal structure $\mathcal{H} = \langle \Delta, \mathcal{I}, h_1 h_2 \dots h_n \rangle$, and a knowledge base $\mathcal{K}$ that is satisfied by $\mathcal{H}$, the sentences in **FO-LTL-K** are interpreted as follows.

- $\mathcal{H}, h_i \models t_1 = t_2$ iff $t_1^{h_i} = t_2^{h_i}$;
- $\mathcal{H}, h_i \models P(t_1, \dots, t_n)$ iff $(t_1^{h_i}, \dots, t_n^{h_i}) \in P^{h_i}$;
- $\mathcal{H}, h_i \models \neg \varphi$ iff $\mathcal{H}, h_i \not\models \varphi$;
- $\mathcal{H}, h_i \models \varphi \wedge \psi$ iff $\mathcal{H}, h_i \models \varphi$ and $\mathcal{H}, h_i \models \psi$;
- $\mathcal{H}, h_i \models \mathbf{X} \varphi$ iff there exists h_{i+1} such that $\mathcal{H}, h_{i+1} \models \varphi$;
- $\mathcal{H}, h_i \models \varphi \mathbf{U} \psi$ iff there exists $j \geq i$ such that for all $i \leq k < j$, $\mathcal{H}, h_k \models \varphi$ and $\mathcal{H}, h_j \models \psi$;
- $\mathcal{H}, h_i \models c_1 : C$ iff $c_1^{\mathcal{I}} \in C^{\mathcal{I}}$;
- $\mathcal{H}, h_i \models (c_1, c_2) : R$ iff $(c_1^{\mathcal{I}}, c_2^{\mathcal{I}}) \in R^{\mathcal{I}}$;
- $\mathcal{H}, h_i \models \forall x. \varphi(x)$ iff for all $a \in \Delta$, $\mathcal{H}, h_i \models \varphi[a/x]$.

We now specify some of the norms in [3] using **FO-LTL-K**. Recall that the interpretation of norms depends on a local setting, each of which should have a reasonable time frame for actions to be completed. Therefore we consulted the caregivers in the Guysborough Antigonish Strait Health Authority (GASHA) for their interpretation of these norms.

- **Norm 1:** When language is a barrier, translators who understand the medical concepts and terminology facilitate information sharing.

This norm guides the second stage (*Information Sharing*) of a therapeutic encounter (see Fig. 1). When a patient is admitted into a palliative care program, some basic information, including the languages that he or she can speak, is entered into his or her healthcare record. This is captured in a unary predicate *PatientLanguage*. Suppose the official language of the palliative care program is kept in its knowledge base under the concept *OfficialLanguage*¹. Our specification of this norm takes time into consideration, requiring an action to take place within a constant time c , e.g., ‘3 hours’. It is represented as follows:

$$\forall t_1. (\neg \exists x. (PatientLanguage(x) \wedge x : OfficialLanguage) \wedge Time(t_1) \rightarrow \mathbf{X}(\exists t_2. (Time(t_2) \wedge (t_2 - t_1) \leq c \wedge ActionTaken(d, p))))$$

where, t_1 refers to the time stamp of a ‘current’ record, t_2 refers to the time stamp of the ‘next’ record due to the temporal operator **X**, the binary predicate *ActionTaken*(x, y) denotes “action x is taken to y ”, constant d denotes an action “Find a translator”, constant p denotes a patient, and the predicate *Time* is interpreted as a singleton set of the unique time stamp associated to a record².

- **Norm 2:** The patient family’s understanding of the shared information is assessed regularly.

There are at least 7 norms requiring regular assessments. The gap between two assessments depends on numerous factors, such as the severity of the issues, the availability of resources, etc. In palliative care, patients’ families often play a very important role as many patients choose to receive care at home. Suppose the knowledge base already has the following role definition: $IsFamilyOf \equiv IsFatherOf \sqcup IsMotherOf \sqcup IsChildOf$, we can express “ x is a family member of y ” implicitly using $(x, y) : IsFamilyOf$ instead of using $(x, y) : IsFatherOf \sqcup IsMotherOf \sqcup IsChildOf$. This norm can be specified as “the next assessment should be made within a time duration c ”, and is represented as:

$$\forall t_1 \forall x. (Time(t_1) \wedge ((x, p) : IsFamilyOf) \rightarrow \mathbf{F}(\exists t_2. (Time(t_2) \wedge (t_2 - t_1) \leq c \wedge (ActionTaken(e, x)))))$$

where the temporal operator **F** refers to a ‘future’ record, constant e denotes “assessment of the understanding of shared information”, and constant p denotes a patient.

¹ In comparison, the predicate *PatientLanguage* is interpreted over healthcare records because different patients may speak different languages.

² If $Time^h = \{c\}$, where c is the content of the single time stamp in the record associated with h , then $\mathcal{H}, h \models \forall x. (Time(x) \rightarrow x = c) \wedge \exists x. Time(x)$.

5 Complexity

A data monitor constructs a first-order temporal structure from a healthcare data model and a knowledge base, then updates itself when a new healthcare record is received. It checks or monitors the properties specified in **FO-LTL-K**, and generates alerts when necessary. The core of our monitoring method is the model checking problem $\mathcal{H}, h_i \models \varphi$ in **FO-LTL-K**. We examine the computational complexity of this problem.

Computational complexity theory studies the amount of computational resources needed, such as time and storage, to solve a problem. It is well-known that the complexity of the model checking problem (also called the satisfaction checking problem) of first-order logic is PSPACE-complete [18]. We show that the model checking problem of **FO-LTL-K** has the same complexity by reducing it to a model checking problem of first-order logic, using a syntactic translation TRS and a semantic translation TRSM. The main idea is to, (1) let TRS associate a natural number i with a formula φ in **FO-LTL-K** so the temporal information is embedded in $\text{TRS}(\varphi, i)$; (2) let TRSM transform the $h_1 h_2 \cdots h_n$ into a single h by augmenting each h_i with a time parameter i semantically. We show this in detail.

Syntactic Translation TRS. The syntactic translation TRS recursively maps a **FO-LTL-K** expression and a natural number t to a first-order expression as follows:

- For each constant c and variable x , $\text{TRS}(c, t) := c$, $\text{TRS}(x, t) := x$; for each n-arity function and term t_i , $\text{TRS}(F(t_1, \dots, t_n, t)) := F(\text{TRS}(t_1), \dots, \text{TRS}(t_n))$; for each concept symbol C_i , $\text{TRS}(C_i, t) := C_i$; for each role symbol R , $\text{TRS}(R, t) := R$;
- For each concept C , D and role R , $\text{TRS}(\neg C, t) := \neg \text{TRS}(C, t)$, $\text{TRS}(C \sqcap D, t) := \text{TRS}(C, t) \wedge \text{TRS}(D, t)$, $\text{TRS}(\forall R.C, t) := \forall y(R(x, y) \rightarrow \text{TRS}(C, t)(y))$, $\text{TRS}(\exists R.C, t) := \exists y(R(x, y) \wedge \text{TRS}(C, t)(y))$;
- For each predicate P and terms t_i , $\text{TRS}(P(t_1, \dots, t_n), t) := P(t_1, \dots, t_n, t)$; for each formula with concepts or roles: $\text{TRS}(u_1 : C, t) := \text{TRS}(C, t)(u_1)$, $\text{TRS}((u_1, u_2) : R, t) := R(u_1, u_2)$; for each formula $\forall x.\varphi$, $\text{TRS}(\forall x.\varphi, t) := \forall x.\text{TRS}(\varphi, t)$;
- For each formula with temporal modalities:
 - $\text{TRS}(\mathbf{X}(\varphi), t) := \exists t'((t' = t + 1) \wedge \text{TRS}(\varphi, t'))$;
 - $\text{TRS}(\varphi \mathbf{U} \psi, t) := \exists t'((t \leq t') \wedge \text{TRS}(\psi, t') \wedge \forall t''((t \leq t'') \wedge (t'' < t') \rightarrow \text{TRS}(\varphi, t'')))$;

Semantic Translation TRSM. The semantic translation TRSM maps a first-order temporal structure $\mathcal{H} = \langle \Delta, \mathcal{I}, h_1 \cdots h_n \rangle$ to a first-order structure $\text{TRSM}(\mathcal{H}) = \langle \Delta', h \rangle$:

- The domain $\Delta' = \Delta \cup \{1, \dots, n\}$;
- The interpretation h coincides with $\mathcal{I}$ on the constants and function symbols. For the new predicates
 - The concept and role symbols are associated with unary predicates and binary predicates, and h uses the interpretation from $\mathcal{I}$, e.g., $t \in C_1^{\mathcal{I}}$ iff $t \in C_1^h$.
 - Suppose P_1 is an n-arity predicate in the **FO-LTL-K** and P_2 is the corresponding (n+1)-arity predicate in the first-order language that is being mapped to using TRS; we require $(t_1, t_2, \dots, t_n, k) \in P_2^h$ iff $(t_1, t_2, \dots, t_n) \in P_1^{h_k}$ in $\mathcal{H}$.

We can show the following by an induction on the structure of φ .

Theorem 1. *Given a first-order temporal structure $\mathcal{H}$ and a **FO-LTL-K** sentence φ , we have: $\mathcal{H}, h_i \models \varphi$ iff $\text{TRSM}(\mathcal{H}, h_i) \models \text{TRS}(\varphi, i)$*

Since both translation functions TRS and TRSM are linear, and the model checking problem for first-order logic is PSPACE-complete [18], we have the following:

Theorem 2. *Given a first-order temporal structure $\mathcal{H}$ and a **FO-LTL-K** sentence φ , the model checking problem: $\mathcal{H}, h_i \models \varphi$, is PSPACE-complete.*

This result is not surprising, but at least it shows that the model checking problem of this logic is not computationally too expensive (e.g., exponential). Based on Definition 6, it is straightforward to develop an algorithm for model checking, that is, to determine if $\mathcal{H}, h_i \models \varphi$. For details of proofs and algorithms, see [13], where the application of DL reasoners to enhance the efficiency of the model checking is also discussed.

6 Related & Future Work

A number of frameworks (e.g., Asbru, EON, GLIF, GUIDE) [8] are available for the computer interpretation of clinical guidelines (also called medical guidelines). Clinical guidelines aim at guiding decisions and establishing criteria for diagnosis, management, and treatment of specific medical problems. They are based on an examination of current evidence within the paradigm of evidence-based medicine. Palliative care norms, as we have discussed, are more general, and are meant to be interpreted relative to the context of particular care settings. Model checking techniques have been used to verify the correctness of a variety of guidelines and to monitor their executions [2, 4]. While we also apply model checking techniques, our focus is different from these approaches, as our logical language is designed both to specify the norms and to take the advantages of the existing knowledge bases.

In [9], the authors provided an agent-based alarm management system for a palliative care unit. They showed how intelligent agents can continuously monitor the evolution of the health status of palliative patients using two kinds of alarms: basic alarms (e.g., *(Hunger < 3)* and *Extreme weakness : Dangerous weakness*), and evolution alarms (e.g., *Number of evaluations: 2. Δ Weakness > 2 : Fast weakness increase*). We can specify these alarms in our language. Moreover, our data model and language are richer for temporal and knowledge-based reasoning.

In [5], the authors proposed a tableau-based algorithm for the runtime monitoring of workflow constraints. Sample properties taken from runtime monitoring scenarios were expressed using a first-order linear temporal logic. The main difference with our work is that knowledge bases are an integral part of our architecture, and we allow description logic expressions to facilitate the use of existing knowledge bases.

For further research, we plan to implement data-aware monitoring for palliative care and explore the use of autonomous agents [14] in monitoring.

Acknowledgment. We thank the Natural Sciences and Engineering Research Council of Canada, the Atlantic Canada Opportunities Agency, and GASHA for their support. We also thank the anonymous reviewers for their comments.

References

- [1] F. Baader, D. Calvanese, D. L. McGuinness, and D. N. and P. F. Patel-Schneider. *The Description Logic Handbook: Theory, Implementation, Applications*. Cambridge University Press, Cambridge, UK, 2003.
- [2] A. Bottrighi, L. Giordano, G. Molino, S. Montani, P. Terenziani, and M. Torchio. Adopting model checking techniques for clinical guidelines verification. *Artificial intelligence in medicine*, 48:1–19, 2010.
- [3] F. Ferris et.al. *A Model to Guide Hospice Palliative Care: Based on National Principles and Norms of Practice*. Canadian Hospice Palliative Care Association, 2002.
- [4] P. Groot, A. Hommersom, P. J. Lucas, R.-J. Merk, A. ten Teije, F. van Harmelen, and R. Serban. Using model checking for critiquing based on clinical guidelines. *Artificial intelligence in medicine*, 46:19–36, 2009.
- [5] S. Halle and R. Villemare. Runtime monitoring of message-based workflows with data. In *Proceedings of the 2008 12th International IEEE Enterprise Distributed Object Computing Conference*, pages 63–72, USA, 2008. IEEE Computer Society.
- [6] K. Miller and W. MacCaull. Toward careflow management systems. *Journal of Emerging Technologies in Web Intelligence Special Issue, E-health: Towards System Interoperability through Process Integration and Performance Management*, 1(2):137–145, 2009.
- [7] K. Miller and W. MacCaull. Verification of careflow management systems with timed *BDI_{CTL}* logic. In *3rd International Workshop on Process-oriented Information Systems in Healthcare (ProHealth 09)*, 2009. Ulm, Germany.
- [8] Mor Peleg et. al. Comparing computer-interpretable guideline models: A case-study approach. *JAMIA*, 10:2003, 2002.
- [9] A. Moreno, D. Riano, and A. Valls. Agent-based alarm management in a palliative care unit. In *3rd Workshop on Agents Applied in Healthcare, IJCAI*, Edinburgh, 2004.
- [10] S. Panzarasa, S. Maddè, S. Quaglini, C. Pistarini, and M. Stefanelli. Evidence-based care-flow management systems: the case of post-stroke rehabilitation. *J. of Biomedical Informatics*, 35(2):123–139, 2002.
- [11] A. Pnueli. The temporal logic of programs. In *Proceedings of the Eighteenth IEEE Symposium on the Foundations of Computer Science*, pages 46–57, 1977.
- [12] F. Rabbi, H. Wang, and W. MacCaull. YAWL2DVE: An automatic translator for workflow verification. In *The 4th IEEE International Conference on Secure Software Integration and Reliability Improvement (SSIRI 2010)*, pages 53–59, 2010.
- [13] J. Ruan and W. MacCaull. Data-aware monitoring for healthcare workflows using formal methods, 2010. StFX CLI Technical Report, CLI-TR 01-2010.
- [14] J. Ruan, W. MacCaull, and H. Jewers. Enhancing patient-centered palliative care with collaborative agents. In *The Second International Workshop on Collaborative Agents REsearch and Development (CARE 2010)*, Toronto, Canada, 2010.
- [15] U. Sattler. A List of Description Logic Reasoners. <http://www.cs.man.ac.uk/~sattler/reasoners.html>. Last accessed May 07, 2010.
- [16] R. M. Smullyan. *First-order logic*. Dover Publications Inc., 1995.
- [17] W. van der Aalst and K. van Hee, editors. *Workflow Management: Models, Methods, and Systems*. The MIT Press, 2004.
- [18] M. Y. Vardi. The complexity of relational query languages (extended abstract). In *STOC '82: Proceedings of the fourteenth annual ACM symposium on Theory of computing*, pages 137–146, New York, NY, USA, 1982. ACM.

Critiquing Knowledge Representation in Medical Image Interpretation using Structure Learning

Niels Radstake¹, Peter J.F. Lucas¹, Marina Velikova¹, and Maurice Samulski²

¹Radboud University Nijmegen, Institute for Computing and Information Sciences,
{peter1,marinav}@cs.ru.nl

²Radboud University Nijmegen Medical Centre, Department of Radiology,
m.samulski@rad.umcn.nl

Abstract. Medical image interpretation is a difficult problem for which human interpreters, radiologists in this case, are normally better equipped than computers. However, there are many clinical situations where radiologist's performance is suboptimal, yielding a need for exploitation of computer-based interpretation for assistance. A typical example of such a problem is the interpretation of mammograms for breast-cancer detection. For this paper, we investigated the use of Bayesian networks as a knowledge-representation formalism, where the structure was drafted by hand and the probabilistic parameters learnt from image data. Although this method allowed for explicitly taking into account expert knowledge from radiologists, the performance was suboptimal. We subsequently carried out extensive experiments with Bayesian-network structure learning, for critiquing the Bayesian network. Through these experiments we have gained much insight into the problem of knowledge representation in medical image interpretation.

1 Introduction

The past decade has seen a transition in radiology from film-based storage of images to digitised, computer-based storage. The digitisation of medical images has offered a unique opportunity for adding computer-aided support to the traditional, human interpretation of medical images by radiologists. Although radiologists are well-trained for the task of image interpretation, there is room for improvement as misinterpretation of medical images is far from rare. Computers are successfully used in many areas of health care; however, it has been hard to match, and certainly surpass, the capabilities of expert radiologists in interpreting medical images. Medical images are noisy and patient specific, and, thus, computers have difficulty in coping with them.

The research described in this paper focuses on one of such hard medical image interpretation tasks: the interpretation of X-ray images of the breasts, usually called mammograms, for breast-cancer detection. Although there has been considerable progress in the last decade in computer-aided interpretation of mammograms, most of the improvement have come from new pattern recognition techniques which detect potentially suspicious breast regions. The proper

interpretation of mammograms, however, requires an approach similar to the one used by radiologists, who normally compare image parts and different images of the breasts to each other, i.e., they interpret potentially suspicious regions of the breasts in the context of all other available image information.

Bayesian networks have been used in our research as they permit integrating knowledge and information from different sources. In the early stages of our research, we therefore decided to design a Bayesian network that incorporated the most important image features of the two mammograms available for each breast; construction of its graph structure was guided by expert knowledge. This Bayesian network can, thus, be looked upon as a *knowledge representation* of mammogram interpretation: it offers a compact representation of how the features extracted from an image are interpreted in terms of breast tissue architecture and presence of masses (mammographic abnormality signs).

Extensive experimentation with the Bayesian network using image data, however, yielded disappointing results, which we did not fully understand. To gain more insight into the structure of the Bayesian network, we subsequently carried out work on learning Bayesian network structures, both restricted and unrestricted, from image data with the aim to improve the structure of the network. In this paper we discuss at length how structure learning succeeded in achieving this goal. The results can be seen as an application of learning for the purpose of knowledge critiquing.

The structure of the paper is as follows. In the next section, the issue of mammogram interpretation is reviewed. In Section 3 we briefly summarise the principles underlying Bayesian network representation and learning, and present the Bayesian network for mammogram interpretation. In Section 4, the different experimental methods and associated results are discussed. Lessons learnt are given attention too in Section 5.

2 Background

Mammography is the diagnostic procedure to detect breast cancer in the breasts using low-dose X-rays. The resulting mammograms are made using different projections, also called *views*. The most common views of the breast are *mediolateral oblique* (MLO) and *craniocaudal* (CC); see Figure 1. The MLO view is a 45° angled side view, showing a part of the pectoral muscles. The CC view is an projection of the breast from above with the nipple centered in the image.

Because a mammogram is a projection of the breast, its layers of breast tissue are superimposed. The X-ray attenuation, which is due to absorption and scattering of photons, describes the density of a region. This results in contrast, or whiteness, of a region, on the mammogram. The darker areas of the breast are non-dense and consist mainly of fatty tissue. The lighter areas are dense and contain lobules, ducts, and possibly masses.

The interpretation of mammograms by radiologists produces *regions of interest*, or *regions* for short. A region is also referred to as a *lesion* or an *abnormality*.

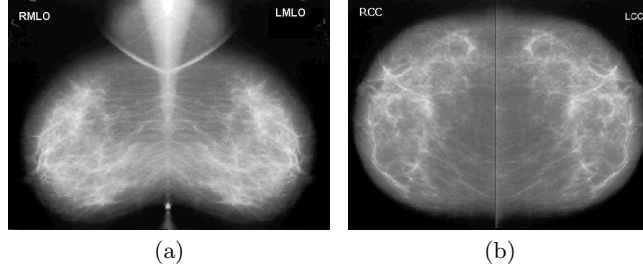


Fig. 1. Right and left breasts in (a) MLO view and (b) CC view.

2.1 Feature extraction and computer-aided detection

Studies have shown that radiologists fail to identify a significant number of cases with breast cancer, i.e., false negatives, due to misinterpretation. The reasons for these misses are unclear [3]. Audits have shown that abnormalities that are clearly visible in retrospect must have been overlooked or its signs were misinterpreted.

To increase the detection rate, *computer-aided detection* (CAD) systems are being developed. These systems use pattern recognition techniques to extract *features* in a mammogram, which are subsequently used to identify regions that are possibly suspicious. With such markings, the CAD system can assist the radiologist while analysing mammograms with the detection of breast abnormalities.

The CAD system [3] we have used employs four steps to classify regions: (1) mammogram segmentation into breast tissue, background, and the pectoral muscles; (2) initial detection of suspicious pixel-based locations; (3) extraction of regions and region-based features, and (4) classification of the extracted regions as cancerous or normal using a neural network classifier. Note that in practice most patients with breast cancer have one or two cancerous regions at most. The CAD system, however, often finds more, which then are false positives.

The region features used in this study can be categorised into two groups:

A. *Observed features* extracted from the image in step (3) mentioned above:

- The relative location of the region (**LocX** and **LocY**);
- The shortest distance of the region to the skin (**d2skin**);
- **Contrast**;
- The presence of radiodensity similar to that of adjacent tissue (**IsoDens**);
- **Spiculation**, indicating whether the region margin has a spiky pattern;
- The presence of a circumscribed lesion (**FocalMass**);
- Linear texture (**LinTex**), which is typical for normal breast tissue;
- Size of the region (**RegSize**).

B. *Calculated features*, computed from classifiers based on pixel- or region-based features:

- The malignancy pixel-based likelihood (**MassLik**);
- The false-positive level of a region (**FPLevel**), indicating the average number of normal regions in an image with the same or higher likelihood scores.

3 Bayesian Network Principles

3.1 Bayesian Networks for Knowledge Representation

A *Bayesian network* $\mathcal{B} = (G, P)$, BN for short, is a probabilistic graphical model that represents conditional independence assumptions in the form of an acyclic directed graph, ADG for short, G ; it is assumed that those conditional independences are obeyed by the associated joint probability distribution P , and P is then called Markov over G [8]. BNs have the virtue that they can be both manually constructed and learnt from data. Manual construction is usually guided by interpreting arcs in Bayesian networks as causal relationships. Initially in the research we constructed a Bayesian network based on available domain knowledge, shown in Figure 2 [4]. The BN incorporates the features described above and it is capable of interpreting MLO and CC features at the same time, allowing the integration of information from two views. Central to the BN model are the hidden variables **AbDensity** and **AbStruct**, indicating the presence of abnormal density and structure. The variable **Finding** represents the conclusion whether or not there is breast cancer.

3.2 Structure Learning

Structure learning is basically finding the Bayesian network graph, or *structure* as is often said, that fits the data best. The number of acyclic directed graphs is, however, more than exponential in the number vertices of the graph [9]. Exhaustively search for the best graph is, therefore, infeasible for most problems. However, for this problem, which concerns 11 variables, we are on the edge of what is still possible. Removing only 2 variables would make it feasible to carry out exhaustive search, although such would, of course, be very time consuming.

Structure learning is an optimisation problem, where a *score measure* is used to judge the fitness of the model, and a search method allows exploring the search space of acyclic directed graphs. The score measure always includes some measure of the likelihood of the data given the graph and its probabilistic parameters, $\Pr(D \mid G, P)$, where D are the data, or the marginalised likelihood $\Pr(D \mid G)$, where the parameters are marginalised out. In addition, they typically include the possibility to include a prior on the structure, $\Pr(G)$, and a penalty for unwanted complexity of graph structure. The two score measures used in this research are the *Bayesian score*, which is based on marginalised likelihood, and the *Bayesian Information Criterion* (BIC), which is likelihood based [5]. These measures take into account that graphs, though different in structure, may encode the same conditional independence assumptions, i.e., are *Markov equivalent* as is said.

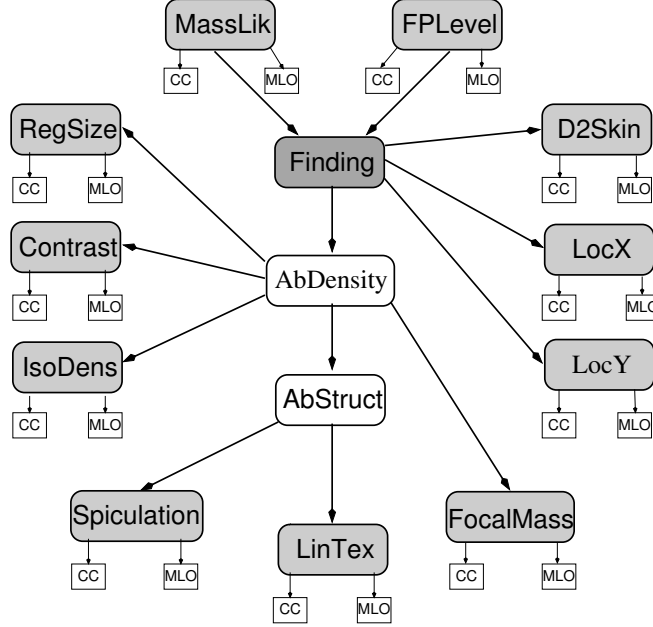


Fig. 2. Feature-based BN model for the interpretation of mammograms; indicated is that each feature node is linked to a corresponding CC and MLO feature [4].

As was argued above, use of exhaustive search is uncommon; more common is the use of greedy search, which searches the space of ADGs, or the space of equivalent classes of structures, called *essential graphs* (EG), i.e., greedy search in EG space. One of the first, and still popular, structure learning methods is the K2 algorithm introduced in [2]. It is a special case of a greedy ADG search algorithm; it minimises the search space by having an initial order on the nodes and by restricting the number of parents a node can have.

In addition to structure learning, we also explored three simpler BN structures for comparison: a fully disconnected graph (all variables are independent), naïve Bayes (NB), and tree-augmented network (TAN). In this order, the network structures are expected to give an increasingly better fitting model, although not as good as the BNs obtained through structure learning.

4 Structure Learning from Mammographic Data

Here, we describe the following set of experiments with actual mammographic data in order to explore various knowledge representation schemes: (i) learning different structures based on a hand-constructed expert sub-model, (ii) comparing the structures learnt based on the observed and calculated features, and (iii) modifying the greedy search algorithm for mammographic structure learning.

4.1 Data and Experimental Set-up

The image dataset used here was obtained from the Dutch breast cancer screening programme and contained data of 1063 cases, of which 383 are cancerous as confirmed by pathological reports. For each case, both the CC and MLO views were present. For each mammogram, the 5 most suspicious regions were selected. In the experiments, structures were learnt using separately the CC data and MLO data. The dataset is divided into a training set, used for learning the models, and a test set, used for scoring the models afterwards. These sets have an equal distribution of cancerous regions. The experiments described in this section were performed using the Bayes Net Toolboxes (BNT) [7], [6].

Most structure learning implementations work with discrete values for the variables. The features in the dataset were real-valued, so we discretised them using a built-in algorithm in [6], which finds an optimal number of bins according to a cost function based on Akaike’s criterion [1]. After additional structure testing with the discretised data, the final number of bins per variable varied between 2 and 7. For the K2 algorithm, the node ordering was derived from the expert model and other orderings were studied using the BIC score. For the greedy search algorithm an empty network was used as an initial structure.

4.2 Results

Learning structures based on an expert model. A sub-model containing 5 MLO variables was selected from the expert model (cf. Figure 3(a)), so that it was possible to perform exhaustive search. The Bayesian score of this sub-model is -48332 . The fitness of the models learnt was compared to the two reference models—fully disconnected and naïve Bayes—which do not consider any of the knowledge incorporated in the expert model. Next we present the structures learnt by various algorithms using the 5 variables from the expert sub-model.

The *TAN algorithm* learnt the structures using the fixed class node **Finding**. All possible TAN structures were learnt and one of them using **FPLevel** as a parent node to the remaining variables is shown in Figure 3(b), with a Bayesian score of -44915 .

For the *K2 algorithm* all possible node orderings were investigated. The model with the highest score is shown in Figure 3(c). For the initial structure of the *greedy search algorithm* (GS), different network structures have been used: (i) a fully disconnected structure, (ii) naïve Bayes, and (iii) a structure learnt using K2. They all resulted in a network structure equivalent to the structure learnt by K2. Finally, we learnt the optimal network structure using *exhaustive search*. All possible network structures with 5 variables (29281) were scored and the best performing model found with exhaustive search is chosen, which is the same model found using K2 and GS. The Bayesian score for this model is -43941 . These results imply that K2, GS, and exhaustive search algorithms have found the structure that fits better to the data than the TAN algorithm. Since the search space of all possible ADGs is relatively small and the model is very restricted, this is not surprising.

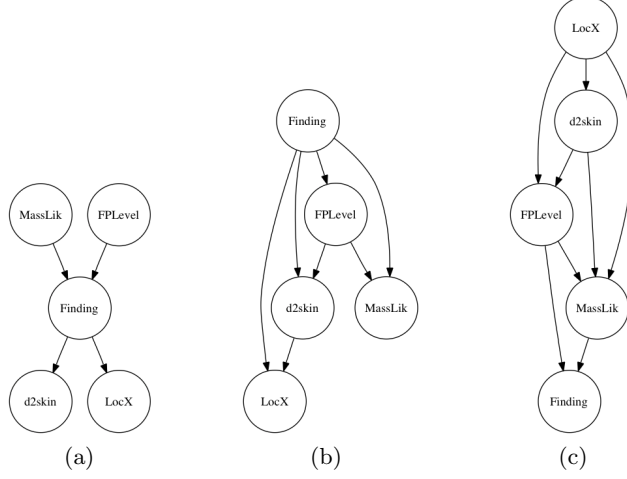


Fig. 3. Structures based on (a) expert sub-model, learnt by (b) TAN, and (c) K2, GS and exhaustive search.

The reference models—fully disconnected and naïve Bayes—yielded Bayesian scores of -49006 and -48077 , respectively, indicating a fit to the data worse than for the structure-learnt models.

When considering the resulting structures in comparison to the expert model in Figure 3(a), it can be observed that in the model learnt by K2, GS and exhaustive search **Finding** is conditioned on **MassLik** and **FPLevel**, as in the expert model, but the location variables **LocX** and **d2skin** are not conditioned on **Finding** and they have a direct causal relationship with the two calculated features. For the TAN model, fixing the class node to be **Finding** results in a structure where all the remaining features were conditioned on this node.

Influence of calculated features. In this experiment we investigated the influence of the calculated features, **MassLik** and **FPLevel**, on the network structures when learning models from data.

Models were learnt using all 11 variables (observed and calculated) and using only observed variables. The results are shown in Table 1. A closer look into the structures learnt revealed that **Finding** is conditioned only on **FPLevel** and not on the other variables. This means that only the **FPLevel** feature is taken into account and that the entire structure could be replaced by the very simple model: **FPLevel** $\rightarrow$ **Finding**. As **FPLevel** is the outcome of the neural network classifier of the CAD system, one would expect a strong dependence with **Finding**.

Another observation is that the features **LocX**, **LocY** and **d2skin**, describing the location of the region in the breast, are related in all learnt models. In some models, especially those learnt using CC data, these variables are independent of the other variables. It was also expected that the structures learnt would

Table 1. Bayesian scores ($\times 10^4$) using all and observed variables

Method	All variables		Observed variables	
	CC	MLO	CC	MLO
NB	-9.9787	-10.1710	-8.0157	-8.1824
TAN	-9.5442	-9.7866	-7.7521	-7.9304
K2	-9.4059	-9.6489	-7.5682	-7.7859
GS	-9.3385	-9.6023	-7.5612	-7.7728

reveal causal relationships between **Spiculation** and **LinTex** as these features are relatively complementary to each other: if linear texture is present, the region is not spiculated and vice versa. However, in only 25% of the learnt models this relation was present.

Modifying greedy search. In the previous experiments we observed that in all cases where **MassLik** and **FPLevel** variables are present, **Finding** becomes conditioned on **FPLevel**. Here we learnt structures using greedy search based on data without including **FPLevel** and **MassLik** in the learning process, but only in the scoring step of the algorithm. The greedy search algorithm started with a initial network structure G , which consisted of the nodes **Finding**, **LocX**, **LocY**, **d2skin**, **Contrast**, **Isodense**, **Spiculation**, **FocalMass**, **LinTexN**, and **RegSize** without arcs. For each step, it defined a set of neighborhood graphs $\mathbf{NG}$. A copy of this set $\mathbf{NG}'$ is made and each DAG in $\mathbf{NG}'$ is modified by adding **FPLevel** and **MassLik** as conditions on **Finding**: $\mathbf{FPLevel} \rightarrow \mathbf{Finding} \leftarrow \mathbf{MassLik}$. For each modified DAG, the score is computed. The (modified) graph with the highest score was selected and its (unmodified) original version was used for the next iteration. The search was stopped when there was no neighborhood network graph with a higher score than the current structure. For the CC data the Bayesian score was $-9.5461 (\times 10^4)$, whereas for the MLO data the Bayesian score was $-9.7830 (\times 10^4)$. The final networks are depicted in Figure 4.

For the three location variables we observed again a strong causal relationships, discovered for both views and for CC they were independent from the remaining features. The relationship $\mathbf{IsoDens} \rightarrow \mathbf{FocalMass}$ persisted in both view structures, indicating that the presence of focal mass is dependent on the presence of isodensity. Furthermore, **Spiculation** and **Contrast** appeared to be conditionally independent given **RegSize** for both view structures as indicated by the paths between these variables. This is an interesting result revealing that knowledge about the size of a region would determine the region spiculation and contrast features. We also note that for the MLO view the calculated features are the only determinants whether or not a finding is present, whereas the remaining features are independent of this converging node.

Comparable results were obtained using a modification of the greedy search algorithm by not including **MassLik** and **FPLevel** in the learning process but adding them only afterwards to the learnt structure. The resulting structure for MLO differed from the one in Figure 4(a) by having the additional arc **RegSize**

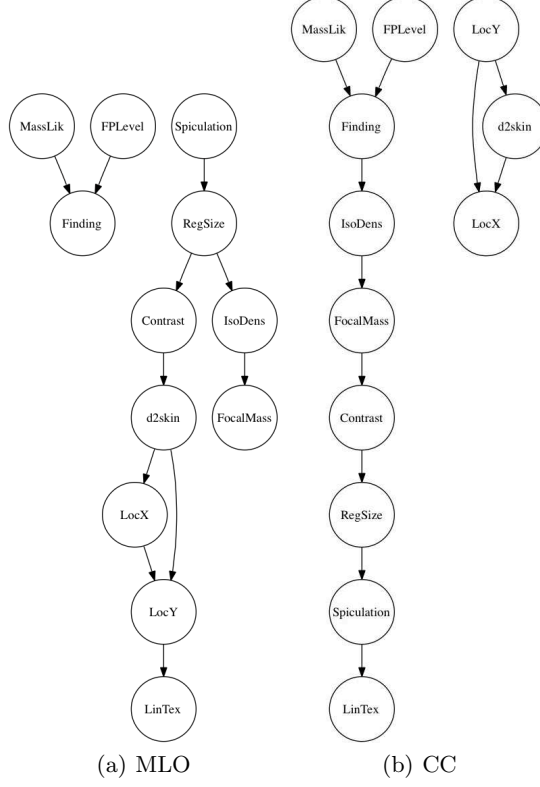


Fig. 4. ADG structures learnt by the modified GS algorithm.

→ **Finding**, with a Bayesian score of $-9.5819 (\times 10^4)$. For CC the structure included, in comparison with the one in Figure 4(b), the arc **Contrast** → **Finding** and excluded the arc **Finding** → **IsoDens**, with a Bayesian score of $-9.9670 (\times 10^4)$. These results demonstrated that using **MassLik** and **FPLevel** in the learning or scoring step of the structure building process make them the only causes of **Finding**, confirming the strong impact of the calculated features.

5 Conclusions and Discussion

Although comparisons between manually constructed and learnt BNs are standard practice, in particular in studies where the performances of structure-learning algorithms are compared, the purpose of the present research was different, namely to see whether structure learning could be effectively used as a source for critiquing a manually constructed BN. Thus, here learning methods

were used as a means to complement knowledge representation by hand. Such an approach may not always be useful, for example in cases where there is an easy conceptualisation of the problem domain available, or when data are not available. In addition, often representations obtained by machine learning are hard to understand, and structure learning of Bayesian networks is no exception to this general rule. However, this makes the combination of techniques from manual and automatic construction of Bayesian networks even more interesting.

In this research we dealt with a problem domain that is very hard from a conceptual point of view: the interpretation of medical images. Whereas in other domains it might be easier to construct manual models using knowledge engineering methods, in the domain of image interpretation it is not unlikely that mistakes are made in the conceptualisation. We carried out this study to find out whether structure learning could be of any help in this case, and a positive answer would only be arrived at if the results obtained had a clear meaning.

The results we achieved clearly show that structure learning results can be conceptually clear and of help in designing a Bayesian network for image interpretation. First, local interactions between variables in the structures learnt were revealed, where some of them were expected based on the domain knowledge, whereas others were novel and not obvious a priori. Second, the results also indicate that manual construction based on expert knowledge is a good start to build a Bayesian network for medical image interpretation, guiding us in the selection of the important factors playing a role in the domain and providing a good basis for comparison with the structures learnt. Finally, we observed that the inclusion of calculated features diminished the explanatory power of the remaining features and obscured their meaning in the problem of mammogram interpretation.

References

1. Akaike H., *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, vol. 19(6), pp. 716–723 (1974)
2. Cooper, G. F., Herskovits, E., *A Bayesian Method for the Induction of Probabilistic Networks from Data*. Machine Learning, vol. 9, pp. 309–347 (1992)
3. Van Engeland, S. *Detection of mass lesions in mammograms by using multiple views*. PhD thesis, Radboud University Nijmegen, The Netherlands (2006)
4. Ferreira, N., Velikova, M., Lucas, P. J. F. *Bayesian Modelling of Multi-View Mammography*. In Proc. of the ICML/UAI/COLT Workshop on Machine Learning for Health-Care Applications (2008)
5. Heckerman, D., *A Tutorial on Learning With Bayesian Networks*. Technical report MSR-TR-95-06. Microsoft Research (1995)
6. Leray, Ph., Francois, O., *BNT Structure Learning Package: Documentation and Experiments*. Technical Report, Laboratoire PSI - INSA Rouen (2004)
7. Murphy, K., *Bayesian Network Toolbox for Matlab* (2002)
8. Pearl, J., *Probabilistic Reasoning in Intelligent Systems*. San Mateo: Morgan Kaufmann (1988)
9. Robinson, R. W., *Counting unlabeled acyclic digraphs*. Combinatorial Mathematics V, vol. 622, pp. 28–43, Berlin: Springer (1977)

Linguistic and Temporal Processing for Discovering Hospital Acquired Infection from Patient Records

Caroline Hagège¹, Pierre Marchal¹, Quentin Gicquel², Stefan Darmoni³,
Suzanne Pereira⁴ and Marie-Hélène Metzger²

¹ Xerox Research Centre Europe, Meylan, France

² UCBL, Lyon, France

³ CISMEF, Rouen, France

⁴ VIDAL, Paris, France

Abstract. This paper describes the first steps of development of a rule-based system that automatically processes medical records in order to discover possible cases of hospital acquired infections (HAI). The system takes as input a set of patient records in electronic format and gives as output, for each document, information regarding HAI. In order to achieve this goal, a temporal processing together with a deep syntactic and semantic analysis of the patient records is performed. Medical knowledge used by the rules is derived from a set of documents that have been annotated by medical doctors. After a brief description of the context of this work, we present the general architecture of our document processing chain and explain how we perform our temporal and linguistic analysis. Finally, we report our preliminary results and we lay out the next steps of the project.

1 General Context

HAI designates any kind of infection acquired by patients during their stay at hospital. In France it is estimated that between 5% and 10% of patients in hospitals can suffer a HAI. It is also estimated that 30% of HAI cases are avoidable. These figures help to understand that a great effort is made to try to understand and when possible to prevent these cases. The work described here is part of the ALADIN project funded by the French government. An general overview of the project and development plans is presented in [2]. The work we report here consists in the developments of a temporal and linguistic processor that gathers possible indicators of risks of HAI in patient records.

2 From Patient Records to HAI Risk Detection

Our input is a set of patient records that have been previously anonymized⁵ in the partner hospitals. Document processing chain consists of two main automatic

⁵ Anonymizer presented in [2]

processing tasks which rely on manual annotation of patient records performed by medical doctors and described in [7]. These two tasks are:

1. Linguistic and temporal processing of the anonymized patient records
2. HAI risk detection

The first task is detailed below and first experiments regarding the second task are then related.

3 Linguistic Processing of Patient Records

3.1 Underlying linguistic analysis

Linguistic analysis is supported by a general dependency parser for French [4]. It takes as input raw text or XML text, which in this application is an anonymized patient record. It provides as output a syntactic representation of the text. This syntactic representation consists in a set of grammatical relations (dependencies) that are established between the linguistic units. Each linguistic unit consists of a set of linguistic features carrying morphological, syntactic and semantic information. Given the following sentence,

Le combicath réalisé à l'entrée retrouve 106 d'*Escherichia coli*
et 107 de *Moraxella catarrhalis*.
(Combicath realized at the admission shows 106 of *Escherichia coli*
and 108 of *Moraxella catarrhalis*)

the parser detects a SUBJECT relation between *combicath* and the verb *retrouve* and the verb *retrouve* has two coordinated OBJECTS represented by the numbers *106* and *107* which are modified respectively by *Escherichia coli* and *Moraxella catarrhalis*.

The aim of linguistic processing is twofold: Detect and filter in the text possible risk indicators and then verify if these indicators are in relation in order to state if a HAI risk is present or not in the document

3.2 Risk indicators detection

Risk indicators for HAI are a subset of medical terms that are defined according to the manually annotated documents. In the work described in this paper, we used the first 40 manually annotated patient records reporting HAI in an intensive care unit. By analyzing these reports, we distinguished the following classes of risk factors:

- INFECTIOUS AGENTS (which subsumes sub-classes BACTERIA, VIRUS and FUNGI)
- TEMPERATURE
- INFECTION
- MEDICAL INVASIVE DEVICES

This categorization will be extended to other classes of risk factors with the help of medical annotators definition in a next future. However, in the experiment related in this paper, we only rely on the above-mentioned categories as they may lead to the presence of an HAI.

The first step of development is capturing and typing properly in the patient records each element of these classes. These elements are either medical terms or more complex expressions that can involve non-contiguous lexical units. We constructed a specialized lexicon (finite-state transducer) which is composed with the general lexicon of our multi-purpose linguistic analyzer. This specialized lexicon consists of a set of medical terms with relevant associated information. For instance, *escherichia coli* is coded as a bacterium in our lexicon. During the first step of text processing, any occurrence of this string (together with possible variants with capital letters) will be *a-priori* marked with the information of being a potential risk indicator of the subclass BACTERIA. Regarding risk indicators that are not static lexical entries or terms, we take advantage of our linguistic processor to detect them. For instance, an expression like:

`la température monte à 39oC (temperature raises to 39oC)`

conveys information of the class TEMPERATURE which is a potential risk factor. We do not want however to simply code *39oC* as a risk factor because it can appear in a context like *temperature went down from 39oC to 37oC*. Using the syntactic relations provided, we set up a rule that states that if the keyword *fever* or *temperature* is either in a SUBJECT or MODIFIER syntactic relation with a predicate (noun or verb) expressing the idea of raising and if this predicate is in turn modified by a quantified expression whose number is higher than 37.5oC, then, we tag the subject or modifier lexical entry as a potential risk indicator. In this way, very different expressions like *Raising of temperature to 39oC*, *Fever reached 39.5oC*, etc. are detected as conveying a risk indicator while expressions like *temperature decreases from 39oC to 37oC* are not retained.

3.3 Filtering risk indicators

The presence of a term (simple or complex) or keyword in a text is not enough to state the existence of a risk indicator. We determine some contexts in which this is the case.

Take for instance the expression

`Escherichia Coli sensible à l'amoxicilline.`
`(Escherichia Coli sensitive to amoxicillin)`

In this expression two keywords that are potential risk indicators are present. First the bacteria name *Escherichia Coli*, second the antibiotic name *amoxicillin*. However, in this context the antibiotic name does not mean that *amoxicillin* has been administrated to the patient. It just further specifies the kind of bacteria as being sensitive to this kind of antibiotic. So, we do not want to consider *amoxicillin* as a risk indicator here, although in other contexts term *amoxicillin* has to be considered as a risk indicator (antibiotic).

Furthermore, the proper treatment of negation is also a way to filter out erroneous indicators. For instance in *Pneumonia without fever*, we do not want to catch *fever* as a risk indicator of class TEMPERATURE (only *Pneumonia* will be retained as an element of the INFECTION class) while for the expression *Pneumonia with fever* we want to have two risk indicators of classes INFECTION and TEMPERATURE.

We added to our linguistic processor some filters that consider these kinds of expressions and we also take advantage of the treatment of negation in the general parser in order to avoid the collection of wrong indicators that simple pattern-matching techniques would consider.

3.4 Relating risk indicators

Once the risk indicators are collected, we want to be able to relate them, since isolated risk indicators (as the administration of an antibiotic or the raise of the temperature) are not enough to conclude that there is a risk of HAI. The study of manually annotated patient records has shown that these HAI cases are determined through the concomitance of various risk indicators. This concomitance is supported by the existence of syntactic (and possibly distant) links between risk factors. As we have seen before, our linguistic processor computes syntactic links between linguistic units in texts. These links are tagged with the main grammatical functions (like *subject object complement* etc.) existing between words. Taking advantage of the grammatical links computed by the linguistic engine, we define a general relation which subsumes the main grammatical functions⁶. This relation called *LINK* corresponds to:

- any syntactic relation like SUBJECT, OBJECT and MODIFIER is a specialization of the LINK relation
- composition of any SUBJECT, OBJECT or MODIFIER relation involving the same verb is a LINK relation

We also admit commutativity and transitivity at one level between these LINK relations.

For instance, for the sentence cited in section 3.1

Le combicath réalisé à l'entrée retrouve 106 d'*Escherichia coli*
et 107 de *Moraxella catarrhalis*.
(Combicath realized at the admission shows 106 of *Escherichia coli*
and 108 of *Moraxella catarrhalis*)

the computed LINK relations enable to relate *combicath* (a medical device) with *Escherichia coli* and *Moraxella catarrhalis* which are bacteria.

We define a set of concomitance rules for risk indicators relying on these LINK relations. These rules are the very first attempt of capturing the annotations

⁶ A similar approach is taken in [13] where event extraction associated with the detection of weak signals of risk in the nuclear domain is described.

provided by the doctors. They have been manually derived from attested HAI suspicion cases and should be further validated by the doctors.

Here follows an example of a concomitance rule linking two risk indicators:

if i_1 LINK i_2
and $(i_1 \in (INF \cup IAGT \cup TEMP)) \wedge (i_2 \in ATB)$
then HAI suspicion.

where: i_1 and i_2 are detected risk indicators and $INF, IAGT, TEMP$ and ATB correspond respectively to classes INFECTION, INFECTIOUS AGENT, TEMPERATURE and ANTIBIOTIC

4 Temporal Processing of Patient Records

Linguistic processing described in the previous section extracts all text excerpts in which at least two HAI risk indicators which are in syntactic relations are detected

This is however not enough and we have to take into account another important element, the chronology.

4.1 The importance of temporal information in medical processing

The temporal dimension is of high importance in the processing of unstructured biomedical texts and more specifically in the processing of patient records. Reasoning about medical data cannot be performed properly if we do not take into account the chronology of events and facts that occurred in the patient history. [9] and [6] present general overviews of time-oriented systems in medicine. [8] presents the TimeText system which performs temporal processing of clinical discharge summaries and the evaluation of the temporal relations between events extracted by the system (*before*, *after* or *during* relations).

Our temporal processing has two main purposes. First, in the particular case of detection of HAI risks, time is one of the parameters to take into account in the definition of HAI. An infection is only considered as an HAI if it appears at least 48 hours after the admission in the hospital and if the patient has not undergone an operation. If the patient has been operated on, then the temporal constraints are different and HAI can occur in a time interval of one month from the operation date. For these reasons, we need to have access to precise time stamps. The second reason for having a precise temporal segmentation of patient records is the assumption that temporal blocks are the basic blocks in which we can correlate risk indicators collected during the linguistic processing. The scope of the sentence is not broad enough as we will see in section 5.1, and temporal blocks seem to be a good alternative for processing units.

4.2 Detecting coherent temporal units in patient records

We want to anchor temporally facts and events described in the patient record, so that we can check the temporal constraints if a suspicion of HAI occurs. Some

related work in this domain has been carried out. Besides already cited [8], [3] describes a system using a machine-learning approach for detecting temporal segments in medical discharge summaries. Temporal order relations are then established between these segments but precise temporal distance between the segments is not given and this information is necessary in our context of HAI detection. [10] also focuses on temporal processing of patient records, but, only a structured subpart of patient records is analyzed (triage notes)

Our goal is different from the works cited above, since we want to define coherent temporal blocks **and** anchor each block with precise time information, from which, if desired, temporal order relations can be calculated easily. Temporal granularity we considered for HAI detection is the day. This choice is motivated by our application needs (temporal constraints for HAI detection do not need a smaller scale).

Absolute dates As our input documents consist in de-identified patient records, all dates that could potentially help to identify a patient are anonymized. This subset of temporal expressions corresponds to absolute (or complete) dates, i.e. dates in which at least the field year is specified. So, expressions like *2010*, *03/02/2009* or *October, 2009* are considered as absolute dates and are anonymized. In the anonymization process, date of the first admission to the hospital is considered as T0 and any occurrence of absolute date found in the document is anonymized taking into account this reference and giving a positive (after) or negative (before) temporal delta expressed in days, months or years. By this means, the chronology is preserved even though dates are anonymized. For instance, after anonymization we have in documents expressions like:

Le patient est repris au Bloc le [T+8J]
(patient went back to the operation room on [T+8J])

which means that this event occurred 8 days after T0, i.e. the admission to the hospital.

Referential dates and their normalization Besides these absolute dates, patient records are full of temporal expressions that go beyond the scope of anonymization, since they do not carry any indication of the patient's identity. This is the case of all referential temporal expressions like *during the first week of his stay*, *today*, *for 4 days*. etc. Our linguistic engine tags and characterizes those expressions and normalizes them with respect to the same T0. Temporal expression processing follows the temporal characterization presented in [12] in which relative dates with reference to the moment of utterance (like *today* or *tomorrow*) versus relative dates with reference to another textual element (like *two days after operation*) are distinguished and processed accordingly. This approach is also in line with state-of-the-art work on temporal processing in NLP [1]. As a result, in a sentence like:

ablation VVC ce jour (CVC removal today)

knowing that the patient record is written when the patient leaves the service, and having the information that the stay at hospital was between T0 and T+10J (10 days of stay), the system is able to date the *CVC removal* at T0+10 days (normalized as [T+10J]).

Delimitation of temporal blocks in patient records Temporal expressions recognition and normalization makes it possible to divide patient records in homogeneous temporal blocks. During parsing, each time a temporal expression is found, it is normalized and compared with the current temporal anchor. If it is different, a new current temporal anchor corresponding to the last normalized value is created. Note that temporal processing is an extension of the linguistic processing (we use the Java API provided by the linguistic engine and take advantage of linguistic information calculated by the parser to segment and recognize the temporal expressions). Technically, a temporal anchor can be considered as a global variable that is always accessible during linguistic processing. Because we are considering days for temporal granularity, we perform a conversion when a temporal expression is expressed using hours using the result of integer division between value in hours and 24. For instance if *50 hours later* appears in a text, we add to the current temporal anchor the value **2 days**.

An example of temporal processing output of patient records in XML format is shown below. Each temporal block is enclosed in the TEMP tag and **val** attribute of this tag indicates the normalized temporal value.

```
<TEMPS id="14" val="T+5J">
Devant l'absence de germes dans les prélèvements pulmonaires,
cette antibiothérapie a été arrêtée à J5.
</TEMPS>
<TEMPS id="15" val="T+4J">
Extubation première le [T+4J]. Il n'y a pas eu de VNI en
post-extubation.
</TEMPS>
<TEMPS id="16" val="T+12J">
Par contre le patient s'est réencombré. Il a été nécessaire
de le réintuber le [T+12J]".
</TEMPS>
```

We performed a first evaluation on the XML temporal segmented documents on intensive care unit patient records coming from the hospital of Lyon. This evaluation is preliminary as it only was performed on 15 patient records. We obtained 71% of correctly delimited blocks with its appropriated time stamp.

4.3 Limits of our temporal processing

Incorrect results are due to the fact that a precise time stamp is not always the right way to represent the temporal information conveyed in the medical reports. Very often, temporal expressions are expressed as a time interval *between* [T+2J]

and $[T+4J]$. In this case our system anchors the temporal block with the last temporal expression of the interval. Other expressions like *examination at $[T+2J]$ will be repeated in two weeks* which is anchored at $[T+2J]$ is considered as correct. But we miss the representation of the (possible) repetition of the examination. Finally, some relative dates were not correctly normalized. For instance *Two days after the ablation...* is not normalized correctly if the reference (here the ablation which is described before in the medical record and possibly with another string) is not found.

Our approach is not suitable for a complete and exhaustive treatment of temporal information conveyed by medical records, however, for our specific purpose (HAI detection), as the events consisting in risk indicators (temperature raising, detection of bacteria, etc.) are mostly punctual events, it is good enough to be used as a filter for HAI detection.

5 HAI Risk Detection

Our current system was developed and first tested using patient records coming from an intensive care unit. Temporal and linguistic analysis are performed in parallel. As a result, we have at the same time information about risk factors, linguistic relations between these factors and the current temporal anchor. For each sentence, if HAI factors are found, if they are related according to one of the concomitance rules as described in section 3.4 and if the current time anchor is superior to a certain limit then the sentence is marked as suspect for HAI (the corresponding patient record will thus be considered as conveying HAI suspicion).

As final output, the system provides:

- a partition of the set of processed documents set which separates documents expressing an HAI suspicion from documents not expressing an HAI suspicion
- the extraction of textual information (sentences and HAI risk indicators) that have led to this classification

5.1 Evaluation

We performed a first and limited evaluation on patient records coming from an intensive care unit of the Lyon hospital. We selected 20 patient records annotated by doctors that had not been used for the system development. Half of these records do not contain suspicion of HAI and each document of the second half contains at least one occurrence of HAI. We analyzed these documents with our current system and obtained the following results: 70% of the documents carrying a HAI suspicion were considered correctly as suspicious. 90% of the non-suspicious document were correctly considered as non-suspicious. We thus obtained a 80% precision in stating if a patient record contains or does not contain HAI suspicion.

Although this evaluation does not scale enough to allow definitive conclusions, we can however make some first remarks by analyzing the errors are the following:

- Relying on a richer medical terminology could have improved the results. At this stage, we have coded in our linguistic processor medical terms, but this is not enough, and we want to integrate in our processing a lookup to existing medical terminologies provided by our partners in the project.
- The scope of the sentence is not broad enough to relate risk indicators. As explained before, the processing unit for our linguistic processor is the sentence. But often, HAI are described by more than one sentence. Risk indicators concerning the same HAI can be scattered in different sentences that are treated separately.

These preliminary results show however that this first simple approach can lead to already interesting results.

6 Next Steps

The first and quite obvious remark is that the sentence is not a sufficient scope for reasoning about risk indicators. The difficulty is to determine from the patient record how to delimit the right text block in which HAI occurrence has to be searched. As we cannot rely on the document structure to help us in this task (documents are unstructured text documents and from one service to another service structure can be very different), we will rely on the temporal blocks that have been delimited during the temporal processing as described in 4.2

A first look at other kinds of documents (patient records coming from the orthopaedics service) has demonstrated that the system that has been developed at this stage (using patient records from intensive care units) is not appropriate as such for orthopaedics. Medical terms and expressions building the set of HAI risk indicators are partly different. We will in the next steps take advantage of rich medical terminologies of project's partners and integrate terminology lookup in the whole processing chain. This will probably overcome the problem of the lexical and terminological needs of our system.

Finally, we want to go further and not only provide a simple marking of the documents (i.e. if the document has or has not a case of HAI) but also an explanation of how this conclusion has been taken. We want to provide for each analyzed document a log file in which risk indicators, rules that have been applied and temporal constraints are clearly stated.

7 CONCLUSION

We have presented an ongoing work in Natural Language Analysis of hospital patient records. The final goal of the project, which involves medical and NLP specialists, is to provide a tool that detects from patient records coming from

different services and different medical specialties cases of Hospital Acquired Infection in order to help services of public health to monitor this problem. We have developed a first prototype that processes patient records from intensive care units, and very first results are encouraging. The limits of our current system are however clear and the next steps will be dedicated to overcome these limits by integrating in the processing chain already existing rich medical terminologies, continue to improve the temporal analysis module, extend the current limits of textual units that serve as bases for risk detection reasoning and finally provide explicative reports indicating why a document is considered as carrying a HAI suspicion.

References

1. Saurí, R., Littman, J., Knippen, B., Gaizauskas, R., Setzer, A., Pustejovsky, J.: TimeML Annotation Guidelines Version 1.2.1. (2006)
2. Proux, D., Marchal, P., Segond, F., Kergoulay, I., Darmoni, S., Pereira, S., Gicquel, Q., Metzger, M-H.: Natural Language Processing to detect Risk Patterns related to Hospital Aquired Infections. Proceedings of RANLP 2009, Borovetz, Bulgaria. (2009) 865–881
3. Bramsen, P., Deshp, P., Lee, Y K., Barzilay R.: Finding Temporal Order in Discharge Summaries <http://www.scientificcommons.org/43547144> (2008)
4. Aït-Mokhtar, S., Chanod, J-P., Roux, C.: Robustness beyond Shallowness: Incremental Deep Parsing. Natural Language Engineering **8** (2002) 121–144
5. Hagège, C., Tannier, X.: XTM: A Robust Temporal Text Processor. Proceedings of CICLing 2008, Haifa, Israel (2008)
6. Zhou, L., Hripcsak G.: Temporal reasoning with medical data - A review with emphasis on medical natural language processing. Journal of Biomedical Informatics **40** (2007) 183–202
7. Dirieh Dibad, A., Sakji, S., Prieur, E., Pereira, S., Joubert, M., Darmoni, S.: Recherche d'information multiterminologique en contexte: étude préliminaire. Informatique et Santé **17** (2009) 101–112
8. Zhou, L., Parsons, S., Hripcsak, G.: The Evaluation of a Temporal Reasoning System in Processing Clinical Discharge Summaries. Journal of the American Medical Informatics Association **15**, N.1 (2008) 99–106
9. Combi, C., Shahr, Y.: Temporal Reasoning and Temporal Data Maintenance in Medicine: Issues and Challenges. Computers in biology and medicine, Elsevier, **27**, N. 5 (1997) 353–368
10. Irvine, A., Haas, S., Sullivan, T.: TN-TIES: A System for Extracting Temporal Information from Emergency Department Triage Notes AMIA Annual Symp. Proceedings, Published Online (2008)
11. Brun, C., Ehrmann, M.: Adaptation of a Named Entity System for the ESTER2 Evaluation Campaign Proceeding of IEEE NLP-KE 2009, Dalian, China (2009)
12. Ehrmann, M., Hagège, C.: Proposition de caractérisation et de typage des expressions temporelles en contexte. Actes de TALN 2009, Senlis, France (2009)
13. Capet, P., Delevallade, T., Nakamura, T., Tarsitano, C., Sandor, A., Voyatzi, S.: A Risk Assessment System with Automatic Extraction of Event Types. Proceedings of the 5th International Conference on Intelligent Information Processing - IIP 2008, Beijing, China (2008)

A Clustered Markov Analysis of Patients Developing Sepsis

Femida Gwadry-Sridhar¹, Michael Bauer², Benoit Lewden¹, Ali Hamou¹,

¹ I-THINK Research Lab, Lawson Health Research Institute, London, ON Canada

² University of Western Ontario, London, ON Canada

femida.gwadry-sridhar@lhsc.on.ca, bauer@csd.uwo.ca,

benoit.lewden@lawsonresearch.com, ali.hamou@sjhc.london.on.ca

Abstract. Sepsis is a significant cause of mortality and morbidity. There are now aggressive goal oriented treatments that can be used to help patients suffering from sepsis. By predicting which patients are more likely to develop sepsis, early treatment can potentially reduce their risks. However, diagnosing sepsis is difficult since there is no “standard” presentation, despite many published definitions of this condition.

In this work, data from a large observational cohort of patients – with elements collected at regular time periods – are observed in order to determine whether sepsis develops or not. A cluster analysis approach is used to form groups of correlated datapoints. This sequence of datapoints is then categorized on a per person basis and the frequency of transitioning from one grouping to another is computed. The result is posed as a Markov model which can accurately estimate the likelihood of a patient developing sepsis. A discussion of the implications and uses of this model is presented.

Keywords: Sepsis, Markov Chains, Cluster Analysis, Predictive Modelling.

1 Introduction

Sepsis is defined as an infection with systematic manifestations of an infection [7]. Severe sepsis is considered present when sepsis co-exists with sepsis-induced organ dysfunction or tissue hypoperfusion [7]. This can result in mortality and morbidity, especially when associated with shock and/or organ dysfunction [3]. Sepsis can be associated with increased hospital resource utilization, prolonged stays in intensive care units (ICU) and hospital wards, decreased long-term health related quality of life and an economic burden estimated at \$17 billion (USD) each year in the United States alone [5, 17, 19]. In Canada, there is limited data on the burden of severe sepsis; however, costs in Quebec may be as high as \$73 million (CAD) per year [13] which contributes to the estimate of total Canadian cost of approximately \$325 million (CAD) per year.

Patients with severe sepsis generally receive their care in an ICU. A multinational study of sepsis in teaching hospitals found that severe sepsis or septic shock is present or develops in 15% of ICU patients [1]. However, diagnosing sepsis is difficult since

there is no “standard” presentation despite many published definitions for sepsis [2, 14]. In the STAR registry [15] (containing a mix of teaching and community hospitals across Canada), the total rate for severe sepsis was 19%. Of these, 63% occurred after hospitalization.

The management of severe sepsis requires prompt treatment within the first six weeks of resuscitation [7]. Intensivists currently support the use of early goal-directed resuscitation of patients. This has shown to improve survival in patients presented to emergency rooms with septic shock [7].

Given the many advances in medicine today, there now exists aggressive goal oriented treatments that can be used to help sepsis and severe sepsis patients [4, 16, 18]. If researchers could predict which patients may be at risk for sepsis, treatment may be started early and potentially reduce the risk of mortality and morbidity. Therefore, methods that help with the early diagnosis of patients who either have or will develop sepsis within hospitals are in dire need and may result in a better prognosis if early interventions are initiated.

1.1 Analytic Techniques

A variety of analytical techniques can be used to assess relationships among a set of measured variables or quantities. Researchers commonly use univariate or multivariable regression models to estimate the association between prognostic variables and a clinical outcome (for instance, whether a patient has contracted sepsis or not). Multivariable regression models are frequently used in studies of clinical outcomes. These models can use both categorical and continuous variables, but the use of uncritical modeling techniques can lead to erroneous conclusions [9].

In this work, the determination of patients at risk of a clinical diagnosis is the primary aim – sepsis being the reference case. Many approaches exist for analyzing patients with and without sepsis [8]. These approaches generally use regression models, which assume the existence of an identifiable, singular set of variables (or elements) for prediction. Hence, if there exists multiple sets of variables that appear in several independent traits, then prediction with this method becomes difficult. The current literature illustrates some of the limitations with univariable and multivariable models [8]. In order to avoid such limitations, alternative approaches for the multifaceted nature of predicting sepsis need to be investigated.

In this paper, a Markov model is constructed to predict sepsis using patient data from 12 Canadian ICU’s. Our model differs fundamentally from those previously used for predicting an outcome. Outcomes are generated by the use of a k-means cluster analysis algorithm to group variables into correlated datapoints and then use these to define “states” in a Markov model. Transitions in the model were determined based on successive values of the datapoints for individual patients. The transition model assumes that patient’s risk changes dependent on their state of health.

Section 2 describes the approach in detail and depicts examples and implications of the proposed model. Section 3 describes the results and Section 4 explores the model’s repercussions and future initiatives. Section 5 concludes and summarizes this work.

Table 1. Sepsis Study: 25 elements collected to form a single datapoint

Variables	P value	Exp(B)
Anaerobia culture	0.122	0.317
Abdominal diagnosis	0.000	15.027
Blood diagnosis	0.000	3.574
Lung diagnosis	0.000	10.360
Other diagnosis	0.000	8.492
Urine diagnosis	0.000	7.280
Chest X-ray and purulent sputum	0.000	2.756
Gram negative culture	0.047	0.679
Gram positive culture	0.001	0.533
Heart rate > 90bpm	0.000	16.933
No culture growth	0.000	0.103
PaO ₂ /FiO ₂ < 250	0.000	12.305
pH < 7.30 or lactate > 1.5 upper normal with base deficit > 5	0.141	1.242
Platelets < 80 or 50% decrease in past 3 days	0.000	5.665
Respiratory rate > 19, PaCO ₂ < 32 or Mechanical ventilation	0.000	8.866
SBP < 90 or MAP < 70 or Pressure for one hour	0.000	9.963
Abdominal culture	0.259	1.872
Blood culture	0.000	2.311
Lung culture	0.724	0.932
Other site culture	0.614	0.869
Urine culture	0.100	1.450
Temperature < 36 or > 38	0.000	8.246
Urinary output < 0.5 mL/kg/hr	0.000	3.166
WBC > 12 or < 4 or > 10% bands	0.000	6.281
Yeast culture	0.011	0.492
Constant	0.000	0.000

2 Methods

2.1 Data Acquisition

We obtained data that was collected from 12 Canadian ICU's that were geographically distributed and included a mix of medical and surgical patients [15]. The study was approved by the University of Western Ontario Research Ethics Board and the need for informed consent was waived. Data was collected on all patients admitted to the ICU who had a stay greater than 24 hours or who had severe sepsis at the time of admission. Patients who did not receive active treatment were excluded. We screened 23000 patients for sepsis. Patients with confirmed sepsis were classified as "septic." It is normal practice in the ICU to treat patients with suspected sepsis as septic until blood cultures are available.

Hospitals routinely collect a minimum data profile on all eligible patients admitted to the ICU [6]. This includes demographic information, admission data, source of admission, diagnosis, illness severity, outcome and length of ICU and hospital

duration. Table 1 presents a summary of the elements collected in this data profile for this study. This table also shows the influence of each variable on sepsis output when analyzed with a simple logistic regression model. Illness severity scores were calculated using a validated formula from data obtained during the first 24 hours in the ICU [11, 12]. All patients were subsequently assessed on a daily basis for the presence of infection and severe sepsis. Hence, for patients with stays longer than 24 hours, repeated measures of the elements were collected and averaged on their daily condition. The characteristics of the patients have been described elsewhere [15].

2.2 Model Formulation

Our patient data is modeled in the following format:

$$\begin{array}{lcl} p_1 & : & d_{11}, d_{12}, \dots, d_{1n1} \\ p_2 & : & d_{21}, d_{22}, \dots, \dots, \dots, d_{2n2} \\ p_3 & : & d_{31}, d_{32}, \dots, \dots, \dots, d_{3n3} \\ \vdots & & \vdots \end{array}$$

where p_i represents a patient in our data set, and d_{ij} represents a datapoint, each consisting of measurements for the 25 elements identified in Table 1 (where i represents the patient and j represents the data collection element). A non-predetermined number of datapoints hence exists for each patient. These datapoints progressed time-wise from initial admittance to the ICU/ward until departure or death. The time period between datapoints typically fell on consecutive days, though this was not always the case (dependent on hospital protocols and staff availability).

Analysis of the collected data was especially challenging for several reasons:

- *The number of datapoints per patient varied. For instance, some patients had two or three datapoints, while for others there was a dozen or more.*
- *The time periods between datapoints for a patient occasionally varied.*
- *The conditions of patients when admitted were extremely diverse; some were severely ill, already showing signs of sepsis, while others showed none. Hence, the first datapoint of two separate patients would not correspond to the same stage of illness.*

Clustering. Unfortunately, due to these stated reasons, the datapoints across the patient dataset is not aligned. To address this, all datapoints were considered independently for an initial cluster analysis. In our previous work [8], cluster analysis had been used to group patients with and without sepsis based on their initial datapoint. This proved to be a useful approach for grouping patients. In this instance, the clustering would group similar datapoints as well – one could consider the datapoints within the same cluster as representing the same “state” of a patient. That is, similar datapoints would be clustered regardless of their position within a timeframe for an individual patient. Though not considered in this initial clustering algorithm, timeframe measurement data will be considered when such datapoints are placed within the Markov model.

A variety of different clustering sizes and algorithms were explored. Empirically, the number of clusters was chosen to be 12 – resulting in clusters that were neither too large to manage (over fitted), nor ones with too few data points to give an adequate determination of patient status due to lack of separation. A modified k-means clustering algorithm was used in this work due to its speed of execution on large datasets. This algorithm incorporates both the internal consistency of variables within the cluster (distance to the cluster’s centroid) and the external distance to all neighboring clusters as its heuristic. The k-means algorithm used is refinement-iterative and alternates between assignment (where each variable is assigned to a cluster) and updating (calculation of new means of the cluster centre). By doing so, clusters will eventually incorporate all variables within the closest proximity space.

During cluster generation, the algorithm is applied to every variable within the dataset. Variables are then ranked by influence and probability of defining a cluster and its location within the cluster field (internal consistency measure). The mathematical indicator used to create these groupings is defined as the orthogonal distance between clusters.

Since our datapoints were binary categorized, each binary element was normalized along its matrix length and squared. Furthermore, in the event of datapoints that were equidistant to multiple centroids, the internal consistency of each cluster was measured and the tighter field was chosen. This prevented the ballooning of the cluster sets. Based on this consistency measure, each cluster is labelled either sepsis or non-sepsis depending on the different distributions of the datapoints within.

Markov Model. Following the cluster computation, each individual patient within the dataset was analyzed. The individual data points, for instance, d_{11} , d_{12} , ..., d_{in1} , were considered separately. If d_{11} was in cluster C_i and d_{12} was in cluster C_j , then a transition from C_i to C_j was created. The frequency of each transition was also tracked and so the total number of transitions from C_i to any cluster, including C_i could be determined. This resulted in each cluster being considered its own “state” and revealing the probability of a transition from one cluster to another. Hence neighbouring datapoints will be represented by the transitions between states.

In order to complete the model, the following metrics were also recorded for each datapoint within a cluster: (a) whether the patient did or did not have sepsis; (b) whether this was the last datapoint associated with the patient or not and if it was the last data point what the outcome was – namely discharged or deceased.

3 Results

A Markov graph with 14 states – 12 states corresponding to the clusters and 2 additional states – one for patients that had been discharged and one for those that were deceased was generated. Since the full model is difficult to depict with all 14 states and transitions between the state pairs, Figure 1 illustrates a portion of the graph for 3 states: states #1, #5, and #6. The gray node represents the “deceased” state and the blue node represents the “discharged” state. What is not shown in the graph is the percent of patients that had been diagnosed with sepsis at the time their datapoints

were included. These are 1.3%, 35.0% and 90.18% for nodes #1, #5 and #6, respectively.

Note that a significant portion (57.8%) of the datapoints that were assigned to state #1, were followed by their adjacent datapoint which was also assigned to state #1. Given the very low percentage of patients diagnosed with sepsis within this cluster, this is likely a state indicative of patients being admitted, never developing sepsis symptoms, remaining for several days, and finally being discharged.

Node #5 can be considered representative of a patient developing sepsis (hidden or apparent, depending on element characteristics), or a patient recovering from it, as most patients have a 35% chance of sepsis in this state.

Node #6 represents a state which can be thought of as one characterizing a patient with sepsis, since 90.18% of the patients having datapoints in this cluster were diagnosed with sepsis. Interestingly, 25.9% return to state #1, where relatively few patients were identified with sepsis. It is also interesting to note that from state #6 no patients died or were discharged. This is likely due to aggressive treatment of sepsis when symptoms become severe and apparent.

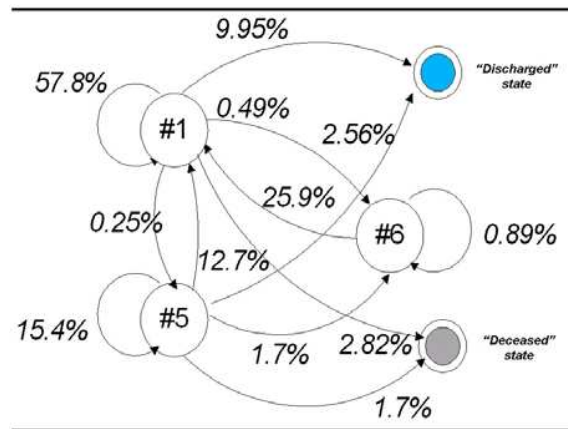


Fig. 1. Portion of Markov Graph

3.1 Further Implications

There are many clinical applications where this model can be used. Consider the following scenario where there exists a datapoint (the 25 element set) for a patient admitted to an ICU. This datapoint can be matched to a cluster, i.e., a "state". From this state in the model, the probability of the patient transitioning to a new state can be determined at the next "time point". For each "next state", there exists a probability that the patient will develop sepsis. Finally, the likelihood of the patient developing sepsis by summing over all possible transitions can easily be computed.

For example, assume that Figure 1 represents the entire Markov graph. If a patient's datapoint ends up in state #1, the patient's estimated probability of developing sepsis is calculated as follows:

$$\begin{aligned} \text{Probability of Sepsis} &= 0.578*0.0129 + 0.0025*0.3504 + 0.0049*0.9018 \\ &= 0.012648, \text{ or } 1.26\% \end{aligned}$$

This is only a one-step approximation – since it only considers a transition from state #1 to its adjacent states by multiplying the transition probabilities by the state sepsis ones. In this example, the one-step approximation illustrates the probability of the patient developing sepsis using paths of length one. In general, one could compute paths of any length (limited by the longest stay of any patient in the hospital in the dataset). Similarly, the probability of a patient not developing sepsis, becoming discharged or becoming deceased, can also be computed using the same algebraic technique.

4 Discussion

In this study, 4,196 patients were selected in order to test the accuracy of the proposed method. In total, 23,547 datapoints were available from the patient dataset. Testing the precision of the clustering was achieved by first randomly selecting data points and removing them from the dataset for testing. Cluster generation followed by training on this sub-sampled dataset, and testing on the removed datapoints, by matching them to the trained clusters. These datapoints were then analysed to see whether they were repatriated to a cluster that represented their end state (sepsis versus non-sepsis).

Not surprisingly, this indicated that the higher the sample size, the better the diagnostic precision. For sepsis patients, an 80.01% precision rate was achieved (when recursive training and testing was applied to the datapoints during cross-validation – in this instance, 1000 datapoints were swapped for each train/test run). For non-sepsis patients, a 94.69% precision rate was achieved, partly because non-sepsis patients represented a larger sample size throughout.

4.1 Future Work

The proposed technique will be thoroughly tested against other ICU patient cohorts in Europe. Clusters will be used in conjunction with decision trees models in order to identify which elements truly influence the development of sepsis versus non-sepsis in patients. Furthermore, the non-homogeneous feature-based classification ability of decision trees and the temporal break sequence modeling based on the Markov process should improve results significantly. *Kim and Oh's* [10] proposed algorithm will be modified and utilized in future works to further improve the model.

5 Conclusions

We have demonstrated the use of cluster analysis as an efficient and relatively swift method for identifying patients at risk – or not at risk – of a clinical condition, being sepsis. The reliability and the validity of these clusters increases based on sample size and based on testing across different datasets. Cluster analysis can be used as an effective tool to supplement the daily monitoring of patients. By utilizing the transitioning probabilities through a Markov model, clinicians can have a greater awareness regarding a patient who is no longer at risk of contracting sepsis. This gives the clinician the ability to make a more informed decision on treatment. Utilizing multi-faceted mathematical modeling is a useful clinical informatics approach to support clinical decision making, the utilization of evidence based treatment and efficiencies in health resource utilization.

References

1. Alberti C, Brun-Buisson C, Burchardi H, Martin C, Goodman S, Artigas A, et al. Epidemiology of sepsis and infection in ICU patients from an international multicentre cohort study. *Intensive Care Med.* 2002;28(2):108-21.
2. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Crit Care Med* 1992;20(6):864-74.
3. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med.* 2001;29(7):1303-10.
4. Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med.* 2001;344(10):699-709.
5. Brun-Buisson C, Doyon F, Carlet J, Dellamonica P, Gouin F, Lepoutre A, et al. Incidence, risk factors, and outcome of severe sepsis and septic shock in adults. a multicenter prospective study in intensive care units. French ICU Group for Severe Sepsis. *JAMA.* 1995;274(12):968-74.
6. Chen LM, Martin CM, Morrison TL, Sibbald WJ. Interobserver variability in data collection of the APACHE II score in teaching and community hospitals. *Crit Care Med.* 1999;27(9):1999-2004.
7. Dellinger RP, Levy MM, Carlet JM, Bion J, Parker MM, Jaeschke R, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Crit Care Med.* 2008;36(1):296-327.
8. Gwady-Sridhar F, Lewden B, Mequanint S, Bauer M. Comparison of analytic approaches for determining variables: a case study in predicting the likelihood of sepsis. In *HEALTHINF 2009. Proceedings of INSTICC; Jan 14-17 January 2009; Porto, Portugal; 2009;90-6.*
9. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15(4):361-87.
10. Kim SH, Oh SS. Decision-tree-based Markov model for phrase break prediction. *ETRI Journal.* 2007;29(4):527-9.

11. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13(10):818-29.
12. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100(6):1619-36.
13. Letarte J, Longo CJ, Pelletier J, Nabonne B, Fisher HN. patient characteristics and costs of severe sepsis and septic shock in Quebec. *J Crit Care* . 2002;17(1):39-49.
14. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med*. 2003;31(4):1250-6.
15. Martin CM, Priestap F, Fisher H, Fowler RA, Heyland DK, Keenan SP, et al. A prospective, observational registry of patients with severe sepsis: The Canadian Sepsis Treatment and Response Registry. *Crit Care Med*. 2009;37(1):81-8.
16. Minneci PC, Deans KJ, Banks SM, Eichacker PQ, Natanson C. Meta-analysis: the effect of steroids on survival and shock during sepsis depends on the dose. *Ann Intern Med*. 2004;141(1):47-56.
17. Pittet D, Rangel-Frausto S, Li N, Tarara D, Costigan M, Rempe L, et al. Systemic inflammatory response syndrome, sepsis, severe sepsis and septic shock: incidence, morbidities and outcomes in surgical ICU patients. *Intensive Care Med* 1995;21(4):302-9.
18. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med*. 2001;345(19):1368-77.
19. Salvo I, de CW, Musicco M, Langer M, Piadena R, Wolfler A, et al. The Italian SEPSIS Study: preliminary results on the incidence and evolution of SIRS, sepsis, severe sepsis and septic shock. *Intensive Care Med*. 1995;21(Suppl 2):244-249.

Towards the interoperability of electronic guidelines: an experiment with openEHR archetypes and a chronic heart failure guideline

M. Marcos* and B. Martínez-Salvador

Universitat Jaume I, Castellón, Spain,
Mar.Marcos@icc.uji.es, Begona.Martinez@icc.uji.es,
<http://www.keg.uji.es/>

Abstract. Clinical guidelines contain recommendations based on the best empirical evidence available at the moment. There is a wide consensus about the benefits of guidelines and about the fact that they should be deployed through clinical information systems, making them available during clinical consultations. However, one of the main obstacles to this integration is the interaction with the electronic health record system. With the aim of solving the interoperability problems of guideline systems, we have investigated the utilisation of the openEHR standardisation proposal in the context of one of the existing guideline representation languages. Concretely, we have designed a collection of archetypes to be used within a chronic heart failure guideline. The main contributions of our work include both the concrete set of archetypes that we have selected and the methodological approach that we have followed to obtain it.

Keywords: Clinical guidelines, knowledge modelling, reusable components, openEHR archetypes, guideline representation languages.

1 Introduction

Clinical guidelines are defined by the U.S. Institute of Medicine as “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” [1]. Guidelines contain recommendations about different aspects of clinical practice, such as diagnosis tests or interventions to perform. These recommendations are based on the best empirical evidence available at the moment. Thus, the use of guidelines has been promoted as a means to control variations in care, reduce inappropriate interventions and deliver more cost-effective care, among others.

Despite some discrepancies, there is a wide consensus about the benefits of guidelines and about the fact that guidelines should be deployed through clinical information systems, making them available during clinical consultations [2].

* This work has been supported by Fundació Caixa Castelló-Bancaixa, through the research project P11B2009-38.

Current guideline systems include reminder systems and increasingly more complex systems representing the whole of guideline procedural knowledge. In any case, there must be some interaction with the clinical information system, in general, and with the electronic health record (EHR) system, in particular, to obtain and share all the relevant information.

In recent years, clinical guidelines have become the focus of many researchers in the areas of Artificial Intelligence and Medical Informatics. Significant contributions in these areas include a variety of languages for the representation of guidelines (see [3], [4]). Despite these efforts, the interaction with EHR systems remains as one of the main obstacles for the interoperability of guideline systems within clinical information systems [5]. Features such as the use of standards for shared EHRs, both for querying EHR data and generating EHR orders from guideline recommendations, are not directly supported in the current guideline representation languages.

One of the main initiatives with regard to EHR standards is the openEHR architecture [6]. It is the culmination of over 10 years of work at international level, aiming to harmonise and converge with other health standards. The concept of archetype plays a central role in openEHR. It was originally defined by Beale in 2001 [7]. An archetype is a formal (yet flexible), reusable, and composable model of a domain concept.

In this paper we take a step towards the interoperability of electronic guidelines, investigating the utilisation of openEHR archetypes jointly with one of the existing guideline representation languages. Concretely, we describe our experiences in designing a collection of openEHR archetypes intended for use within a guideline for the management of chronic heart failure modelled in PROforma.

2 Methodological aspects

We have carried out an experiment consisting in designing a collection of openEHR archetypes intended for use within a guideline for the management of chronic heart failure (CHF), concretely the guideline for the diagnosis and treatment of CHF developed by the European Society of Cardiology [8]. This guideline had been previously modelled in PROforma as part of a project aimed at the development of an electronic care plan for the treatment of comorbidities [9].

The motivation for this experiment is studying the feasibility of using archetypes as a means to solve the interoperability problems of guideline systems. A possible approach is viewing the guideline as a representation with archetype-enabled fragments in strategic points where interactions with the EHR system should occur, e.g. enquiry sources and actions in a PROforma model. This would require on one hand, designing a collection of archetypes suitable for the decision support tasks carried out in the guideline and, on the other hand, ensuring that the guideline is compliant with these archetypes. In this paper we concentrate on the former aspect.

2.1 Archetype repository

We have used as starting point the archetypes from the openEHR Clinical Knowledge Manager (CKM) [10], which is a web-based repository allowing for archetype search, browse and download. Archetypes in CKM have been created by independent domain experts, mainly clinicians and computer scientists, and then they have been released to the community as open source and freely available content. Before publication, archetypes undergo an iterative review process to ensure that they cover as many use-cases as possible and thus constitute a sensible maximal data set.

According to openEHR the main categories for the description of clinical concepts are: observation, evaluation, instruction and action. This categorisation originates from the way in which information is created during the care process, namely: an *observation* is created by an act of observation, measurement, or testing; an *evaluation* is obtained by inference from observations, using personal experience and/or published knowledge; an *instruction* is an evaluation-based instruction to be performed by healthcare agents; and an *action* is a record of the interventions that have occurred, instruction-related or not.

Within the observation category, in CKM we can find e.g. the archetypes `apgar score`, `blood pressure`, `body mass index` and `body weight`. We also find specialisations of the latter, namely `adjusted body weight` and `body weight at birth`. In the instruction category we can find e.g. `follow up instruction`, `healthcare service request` (with the specialisations `laboratory test request` and `referral request`), `imaging request`, `medication order` and so on.

2.2 Archetype methodology

Strictly speaking there is no documentation that can be used as a guide for archetype building and utilisation, probably due to the novelty of the openEHR approach. An exception is the methodology sketched by Leslie&Heard [11]. According to these authors the steps to be performed in archetype design are, for each subject/activity/task:

1. Identify clinical concepts. In this step all the different concepts involved in the subject must be identified, making clear whether is it a single concept or is it made up of multiple concepts. For this purpose the authors recommend using mind maps as a tool, both to identify individual concepts and detect and solve any overlap.
2. Identify existing archetypes. In this step it is necessary to investigate the archetypes available in repositories, in our case the openEHR CKM, and select the best candidates for reuse. In case the candidate (or candidates) is a maximal data set for the purpose under consideration, it should be used as it is; otherwise changes would be required in the selected archetype.
3. If necessary, create new archetypes. In case there is no archetype suitable for the concept in question, a new archetype should be created. The procedure for archetype creation, according to the authors:

- (a) Gather the content.
- (b) Organise the content, identifying e.g. purpose, data elements, and coding/terminology issues.
- (c) Choose the archetype class, namely: *entry*, for clinical concepts of the categories described above; *composition*, for documents; or *section*, for document parts.
- (d) Build the archetype, with the substeps: name the archetype, select the structure, add data types, add constraints, add meta data, and add terminology.
- (e) Publish the archetype.

Leslie&Heard methodology concentrates on the creation of new archetypes and does not provide indications for archetype modification in general and for archetype specialisation in particular. This is somehow surprising since archetypes are designed to reflect as many use-cases as possible, leaving room for the required specialisations.

For the purposes of our experiment, in which archetype specialisation is expected to play an important role, we have followed the methodology below:

1. Review the guideline to determine all the clinical concepts to be archetyped. In PROforma this roughly amounts to filtering the concepts included as enquiry sources, actions, and decision candidates.
2. Create a mind map to identify the separate clinical concepts and to detect possible overlaps, according to the indications of Leslie&Heard methodology.
3. Classify the above concepts into the clinical concept categories of openEHR.
4. For each concept, search the CKM repository for suitable archetypes in the corresponding category:
 - (a) if an archetype is found, determine whether it should be used as it is or whether it should be specialised.
 - (b) if no archetypes are found, specify a new archetype following Leslie&Heard methodology.
5. Create the required archetype specialisations and/or new archetypes using a specific-purpose tool (e.g. Ocean Informatics¹ open-source archetype editor).
6. Validate both archetype specialisations and new archetypes, with the help of clinical experts.

3 Results

Next we describe the results of applying the above methodology to the design of a collection of archetypes to be used as part of an electronic guideline for the management of CHF. Except for the two last steps, i.e. archetype specialisation/building and archetype validation, all the steps have been fulfilled.

¹ See <http://www.oceaninformatics.com/>.

Table 1. Summary of results.

Entry	Guideline concept(s)	Concept(s)	openEHR archetype(s) ^a
Observation ^b	angioedema	angioedema	✓...exam.v1 (C)
	cough	cough	✓...exam.v1 (C)
	ECG_etc_results	ECG results, Xray results, Nat. pept. results	✓...ecg.v1, ✓...imaging.v1, ~...lab.test.v1
	echocardio_results	echocardio results	✓...imaging.v1
	fluid_retention	fluid retention	✓...exam.v1 (C)
	postMI	postMI	✓...story.v1
	recentMI	recentMI	✓...story.v1
	state_revision_action	revision 24-48h. after treat.	✓...exam.v1 (C)
Evaluation ^c	ACEI_intolerant_decision	ACEI intolerant	✓...exclusion-medication.v1
	assess_CHF_action	assess CHF aethiology & type	✓...problem-diagnosis.v1
	atrial_fibrillation	atrial fibrillation	✓...problem-diagnosis.v1
	bronchial_hyperreactivity	bronchial hyperreactivity	✓...problem-diagnosis.v1
	CHF_decompensation	CHF decompensation	✓...problem-diagnosis.v1
	ECG_etc_results, echocardio_results	HF diagnosis	✓...problem-diagnosis.v1
	fluid_retention	fluid retention	✓...problem-diagnosis.v1
	hyperglycemia_unbalance	hyperglycemia unbalance	✓...problem-diagnosis.v1
	hypoglycemia_unbalance	hypoglycemia unbalance	✓...problem-diagnosis.v1
	improved_HF	improved HF	✓...problem-diagnosis.v1
	infection	infection	✓...problem-diagnosis.v1
	stage_decision	determine NYHA class	✓...problem-diagnosis.v1
	state_revision_action	revision 24-48h. after treat.	✓...clinical_synopsis.v1
	stenosis	renal artery stenosis	✓...problem-diagnosis.v1
	still_symptomatic	still symptomatic HF	✓...problem-diagnosis.v1
	symptoms_severity	symptoms severity	✓...problem-diagnosis.v1
Instruction ^d	ACEI_action	ACEI medication	✓...medication.v1
	aldosterone_antagonist_action	aldosterone antagonist medication	✓...medication.v1
	ARB_action	ARB medication	✓...medication.v1
	betablocker_action	betablocker medication	✓...medication.v1
	cardiac_glycosides_action	cardiac glycosides medication	✓...medication.v1
	CHF_fortnightly_follow_up_plan	14-days follow-up	✓...follow_up.v1
	CHF_non_pharma_treatment_plan	no pharmacological treat.	~...non_drug_therapy.v1
	diuretics_action	diuretics medication	✓...medication.v1
	ECG_Xray_Nat_peptides_action	ECG, Xray, Nat. pept. test	~...procedure.v1, ✓...imaging.v1, ~...request_lab_test.v1
	echocardio_action	echocardiography	✓...imaging.v1
	inotropic_support_action	inotropic support therapy	✓...medication.v1

^a Legend: ✓ archetypes used as they are, ~ requiring specialization, (C) requiring additional archetypes.

^b Observation names have the prefix openEHR-EHR-OBSERVATION.

^c Evaluation names have the prefix openEHR-EHR-EVALUATION.

^d Instruction names have the prefix openEHR-EHR-INSTRUCTION.

1. *Review of the guideline to determine all the clinical concepts.* In order to identify the clinical concepts necessary to support CHF management, and hence requiring archetypes, we have used as starting point the PROforma model of the guideline. As sketched before, we have focussed on the PROforma elements enquiry, decision and action. Thus, data sources within enquiries suggest a related concept, e.g. *echocardio results* led to the corresponding concept “echocardio results”. Similarly, decision candidates and actions suggest some related concept. In this step we have identified 33 guideline concepts in total.

An exception to the above general rule are the enquiries requesting an input from the physician where several options are acceptable according to the guideline. An example is *ARB_introduction*, which is a yes/no value that reflects the physician decision regarding the use of ARB drugs² as part of the therapy. This kind of enquiry sources do not correspond to established clinical concepts and therefore have not been considered for archotyping.

2. *Creation of a mind map to identify separate clinical concepts.* We have used a mind map to better visualise the concepts identified in the previous step, and also to make explicit the relationships among them. Here we have made use of on-line medical resources, in addition to the records of the interviews with the experts held during the modelling of the CHF guideline.

An important guiding principle is explained next. If as a result of guideline execution some piece of information is relevant enough to be stored in the EHR system, then there must be necessarily an archetype for it, even though the electronic guideline does not explicitly model it. A notable example is an evaluation concept to represent whether the patient has CHF or not, which was not modelled in the guideline.

Additionally, in this process it became apparent that many guideline concepts in fact correspond to two different clinical concepts, typically observations and evaluations. For example, for the guideline concept *echocardio results* two concepts must be considered: a concept to hold the results of an echocardiography and a concept to represent the medical assessment of these results. Moreover, in many cases the guideline refers to a concept about a medical assessment but the observations/tests on which the assessment is based are not explicitly mentioned. An example of this is the concept *hyperglycemia unbalance* which is based on a glucose blood test, among other things.

3. *Classification of clinical concepts into openEHR categories.* At this point we have determined the type of entry for each different clinical concept, in preparation for the next step. Obviously, we have taken into account the guiding principle described above, and also the twofold consideration of certain guideline concepts both as observations and evaluations.

In this step we have identified a total of 40 archetypable concepts. The two first columns of table 3 show the results of the process upto this step.

² Angiotensin receptor blockers.

4. *Search the CKM repository for suitable archetypes.* From the concepts of the previous step, we have used the search&browse utilities of CKM repository to identify the most suitable archetypes. With this aim, we have used the documentation part of archetypes, specially the sections on purpose, use and misuse. Table 3 shows the results of this step. For clarity issues the table does not include the category document/composition, and hence the concept “patient discharge” is not listed.

We have found more or less specific archetypes for all of our concepts. In total 15 different archetypes have been selected for use in the CHF guideline. For simplicity, this number only includes the entry categories (observation, evaluation, instruction and action) and the composition one. The most frequent archetypes are `openEHR-EHR-EVALUATION.problem-diagnosis.v1` and `openEHR-EHR-INSTRUCTION.medication.v1`. We consider that the majority of the archetypes could be used as they are, and that only 5 of them would require specialisation. Although these results still have to be validated with the help of medical experts, we regard them as a very positive finding concerning the usability of openEHR archetypes.

4 Conclusions and future work

There is a general consensus about the fact that clinical guidelines should be deployed using some computer support and that this support should be integrated within the clinical information system, to take full advantage of the potential benefits of guidelines. However, currently one of the main obstacles to this integration is the interaction with the EHR system.

With the aim of solving the interoperability problems of guideline systems, we have investigated the utilisation of the openEHR standardisation proposal in the context of one of the existing guideline representation languages. Concretely, we have designed a collection of archetypes to be used within a CHF guideline. The main contributions of our work include both the concrete set of archetypes that we have selected and the methodological approach that we have followed to obtain it.

Most of the work has consisted in the identification of the clinical concepts involved in the guideline and the selection of suitable openEHR archetypes. The task has been more complex than expected because of the mismatch between the guideline concepts and the clinical concepts available in the openEHR repositories. A plausible explanation for this is that in most cases electronic guidelines are modelled as independent objects without taking into account deployment issues such as the interaction with EHR systems. To solve this problem we advocate for a completely different approach, namely modelling guidelines using EHR standards as a guide.

On the tool side, the openEHR repository has proven to be very useful in our experiment. Its web-based platform provides very powerful search&browse utilities and includes all sort of descriptions including e.g. tabular views, conceptual maps, etc. In addition, the documentation view includes very useful sections

describing the purpose of the archetype, use or typical usage, and misuse or cases in which the archetype should not be used (with hints on which one to use instead). All this information refers to clinical aspects rather than to technical ones, which is of great help for guideline modellers without a clinical background. On the archetype side, a very positive finding is the high percentage of reuse.

As future work, we plan to proceed with the experiment by (1) validating the archetype design and specification with the help of medical experts; (2) using an archetype editor to actually create the required specialisations; and (3) developing an archetype-enabled guideline, from the archetypes themselves and an electronic CHF guideline. Although we have used as starting point a PROforma model of the CHF guideline, we have not yet ruled out other representation languages that might be suitable for the kind of interactions used in the openEHR framework.

References

1. Field, M., Lohr, K., eds.: Guidelines for clinical practice: from development to use. National Academy Press, Washington D.C., USA (1992)
2. Sonnenberg, F., Hagerty, C.: Computer-Interpretable Clinical Practice Guidelines. Where Are We and where Are We Going? IMIA Yearbook of Medical Informatics (2006) 145–158
3. Peleg, M., Miksch, S., Seyfang, A., Bury, J., Ciccarese, P., Fox, J., Greenes, R., Hall, R., Johnson, P., Jones, N., Kumar, A., Quaglini, S., Shortliffe, E., Stefanelli, M.: Comparing computer-interpretable guideline models: A case-study approach. Journal of the American Medical Informatics Association **10** (2003) 52–68
4. de Clercq, P., Blom, J., Korsten, H., Hasman, A.: Approaches for creating computer-interpretable guidelines that facilitate decision support. Artificial Intelligence in Medicine **31** (2004) 1–27
5. Chen, R., Georgii-Hemming, P., Ahlfeldt, H.: Representing a Chemotherapy Guideline Using openEHR and Rules. In Adlassnig, K.P. *et al.*, ed.: Medical Informatics in a United and Healthy Europe - Proc. of the 22nd International Congress of the European Federation for Medical Informatics (MIE-09), IOS Press (2009) 653–657
6. Beale, T., Heard, S.: Architecture Overview. <http://www.openehr.org/releases/1.0.2/architecture/overview.pdf> (April 2007)
7. Beale, T.: Archetypes. Constraint-based Domain Models for Future-proof Information Systems. Document from <http://www.openehr.org/publications/> (2001) Date of access: May 2010.
8. The Task Force for the Diagnosis and Treatment of Chronic Heart Failure of the European Society of Cardiology: Guidelines for the diagnosis and treatment of chronic heart failure: executive summary (update 2005). European Heart Journal **26** (2005) 1115–1140
9. Lozano, E., Marcos, M., Martínez-Salvador, B., Alonso, A., Alonso, J.: Experiences in the Development of Electronic Care Plans for the Management of Comorbidities. In: Knowledge Representation for Health-Care. Data, Processes and Guidelines. Volume 5943 of LNAI. Springer-Verlag (February 2010)
10. The openEHR Foundation: openEHR Clinical Knowledge Manager. <http://www.openehr.org/knowledge/> Date of access: May 2010.

11. Leslie, H., Heard, S.: Building an archetype. Presentation from <http://www.oceaninformatics.com/Media/docs/> (2008) Date of access: May 2010.

Identifying Actions Described in Clinical Practice Guidelines Using Semantic Relations

Katharina Kaiser^{1,2}, Andreas Seyfang¹, and Silvia Miksch^{1,3}

¹ Institute of Software Technology & Interactive Systems
Vienna University of Technology

Favoritenstraße 9-11/188, 1040 Vienna, Austria

² Center for Medical Statistics, Informatics, & Intelligent Systems
Medical University of Vienna

Spitalgasse 23, 1090 Vienna, Austria

³ Department of Information & Knowledge Engineering
Danube University Krems
Dr.-Karl-Dorrek-Straße 30, 3500 Krems, Austria

Abstract. Clinical practice guidelines are important instruments to support clinical care. We have analyzed guidelines according to their semantic relations in order to generate a formal representation. We used the UMLS Semantic Network as a basis for our analysis. We defined relations that will be used to automatically identify the control flow described in guidelines for generating a computer-interpretable format. A study showed that by using semantic relations we are able to discover a large part of the control flow.

1 Introduction

Clinical practice guidelines (CPGs) are important instruments to provide state-of-the-art clinical practice for the medical and clinical personnel [1]. It has been shown that integrating CPGs into clinical information systems can improve the quality of care [2]. Among many services they can offer are: summarizing patient data; providing alerts and reminders; retrieving and filtering information which is relevant to a specific decision; and weighing up the pros and cons of clinical options in a patient-specific way [3]. For integrating CPGs into such systems, the CPGs (i.e., documents in free, narrative text) have to be translated into a computer-interpretable format. Various of such guideline representation formats have been developed (see [4] for an overview and comparison), but the translation process is still a large bottle-neck. It's a complex and challenging task requiring both medical and computer science expertise. Several methods have been developed that describe systematic approaches (e.g., [5,6,7]) for guideline modelling. Some of them use intermediate formats to break the complexity of the modelling task into manageable tasks (e.g., the GEM format [8], the *Many-Headed Bridge between Guideline Formats (MHB)* [9]).

A CPG can be translated into a computer-interpretable format by representing it as an instance of an ontology. Thereby, concepts are identified and relations between them are established. In this work we show what relations are necessary for identifying text representing the control flow of a guideline. We thereby utilize the UMLS Semantic

Network [10], which provides not only semantic types, but also relations between them. Our hypothesis is that by using these relations we can identify activities to be performed in a guideline document.

Modelling CPGs in a computer-interpretable and -executable format is a great challenge. Automatically identifying actions and processes that describe the control flow can facilitate the modelling process by providing important information necessary for the modelling task and by reducing the workload for the involved persons. Thereby, the modelling process will still require manual interaction (e.g., to solve ambiguities and inconsistencies and to add knowledge necessary for execution).

In the following section we give a short overview on knowledge-based methods for guideline modelling and on using relations for identifying information. In Section 3 we explain the methods developed and resources used. Section 4 describes the evaluation of our methods and a discussion of its results. The final section contains our conclusions.

2 Background

First attempts to computerised CPGs have been implemented in the Protégé environment [11]. Thereby, the class structure has been defined according to the representation formalism (e.g., GLIF [12]). The according concepts are identified in the CPG and assigned to the classes. All implementations use a *model-centric approach* where no direct connection between the guideline text and the corresponding model is generated.

Other attempts have been made to support the modelling, maintenance, and shareability of computer-interpretable guidelines that rather stick on the original text representation. Moser and Miksch introduced prototypical patterns in clinical guidelines that can be used as means to reduce the gap between the information represented in clinical guidelines and the formal representation of these clinical guidelines in execution models [13]. They defined *structure patterns*, *temporal patterns*, and *element patterns*. Serban et al. [14] defined linguistic patterns found in CPGs that can be used to support both the authoring and the modelling process of guidelines. Furthermore, they defined an ontology out of these patterns and linked them with existing thesauri in order to use compilations of thesauri knowledge as building blocks for modeling and maintaining the content of a medical guideline [15]. Peleg and Tu developed visual templates that structure screening guidelines as algorithms of guideline steps used for screening and data collection and used them to represent the guidelines collected [16]. To support guideline implementers in standardising and implementing the action components of guideline recommendations, Essaihi et al. defined the Action Palette [17] — a set of medical action types that categorises activities recommended by clinical guidelines. The set of actions include: Prescribe, Perform therapeutic procedure, Educate/ Counsel, Test, Dispose, Refer/Consult, Conclude, Document, Monitor, Advocate, and Prepare. The intention is to develop commonly used services for each action type, thus easing workflow integration. In [18] we proposed a method to identify actions in otolaryngology CPGs using a subset of the *Medical Subject Headings* (MeSH) as a thesaurus.

In [19] SeReMeD is introduced. It is a method for automatically generating knowledge representations from natural language documents. The documents that were ap-

plied on this method were rather simple (i.e., X-ray reports). CPGs are substantial complex documents and the modelling task is particularly challenging.

3 Methods

In order to define the relations necessary we used the UMLS Semantic Network (SN) [10]. The SN offers amongst 135 semantic types also 54 semantic relationships between these types and therefore acts as an upper-level ontology. Together, there are 6,752 different relations. We then proceeded in four steps:

1. Analyzing CPG documents regarding actions and procedures to be performed within treatment.
2. Detection of relations that describe the actions and procedures.
3. Augmenting the initial relation set by specialism and broadening of semantic types.
4. Generation of a dictionary for identifying the correct relation type.

We started by analysing a protocol (i.e., a local adaption of a guideline) for natural birth that was derived from the guideline “Induction in labour” [20]. The document describes the labor and delivery management for risk-free births and consists of 120 sentences. 48 sentences include actions to be performed.

We used the MetaMap Transfer program (MMTx) [21] to find relevant concepts in the text and to assign a semantic type to each concept. In case an appropriate concept is assigned more than one semantic type, we chose the most specific one if more general were also applicable. If semantic types were not hierarchically linked, all remained as possible semantic types.

Then, we looked for appropriate relations between semantic types. In 47 sentences we detected 52 activities by means of 20 different semantic relations. For one sentence (‘In case of abnormal FHR, monitoring should be continuous.’) no relation could be identified although the text implies an action (‘monitoring should be continuous’). Table 1 shows all found relations together with the number of occurrences.

All relations containing the semantic type *Professional or Occupational Group* on the left hand side are thereby incomplete relations. That means that not both semantic types that are linked by the relationship are mentioned in the text. This is the result from the fact that actions and activities in CPGs are related to tasks the health care personnel has to perform (in our specific case: gynaecologists and obstetricians, who are assigned the semantic type *Professional or Occupational Group*). As most of the actions address these users directly, they are only implicitly referred to in the text. Furthermore, passive voice is frequently used, which also allows omission of the agent of the action. See Figure 1 for an example.

We also identified complete relations. See Figure 2 for an example.

The last four relations displayed in Table 1 are not contained in the SN. We generated these relations in order to be able to represent activities such as ‘*Fetal heart rate should be auscultated.*’, where ‘fetal heart rate’ is assigned the semantic type *Finding*.

In the next step we augmented our initial set of relations by generalization, specialization, and broadening of semantic types. Thereby, we expanded a relation to semantic types of the same *Semantic Collection* [22]. A semantic collection is a partition of the

Table 1. Relations detected in the guideline document.

Semantic Type	relationship	Semantic Type	Quantity
Professional or Occupational Group	performs	Health Care Activity	8
Professional or Occupational Group	performs	Therapeutic or Preventive Procedure	14
Professional or Occupational Group	performs	Diagnostic Procedure	2
Professional or Occupational Group	performs	Research Activity	1
Population Group	performs	Therapeutic or Preventive Procedure	1
Population Group	performs	Diagnostic Procedure	1
Professional or Occupational Group	interacts_with	Population Group	4
Therapeutic or Preventive Procedure	affects	Mental Process	1
Health Care Activity	affects	Mental Process	1
Professional or Occupational Group	uses	Intellectual Product	1
Professional or Occupational Group	uses	Pharmacologic Substance	3
Pharmacologic Substance	isa	Pharmacologic Substance	1
Health Care Activity	isa	Health Care Activity	2
Therapeutic or Preventive Procedure	isa	Health Care Activity	2
Therapeutic or Preventive Procedure	isa	Therapeutic or Preventive Procedure	2
Diagnostic Procedure	measures	Clinical Attribute	1
Professional or Occupational Group	analyzes	Body Substance	3
Professional or Occupational Group	analyzes	Finding	1
Professional or Occupational Group	analyzes	Organism Function	2
Professional or Occupational Group	analyzes	Organism Attribute	1

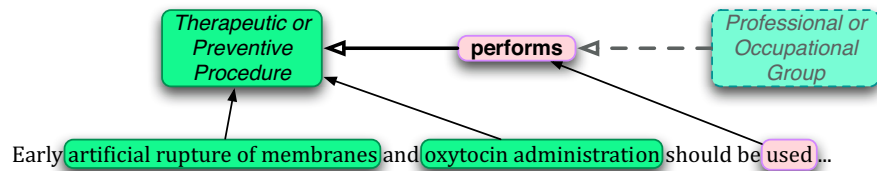


Fig. 1. An action describing an incompletely formulated semantic relation.

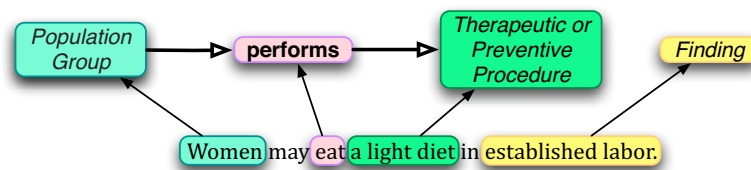


Fig. 2. An action describing a completely formulated semantic relation.

Table 2. Augmented relation set.

Semantic Type	relationship	Semantic Type
Group	performs	Health Care Activity
Professional or Occupational Group		Therapeutic or Preventive Procedure
Population Group		Diagnostic Procedure
Family Group		Laboratory Procedure
Age Group		Educational Activity
Patient or Disabled Group		Research Activity
Professional or Occupational Group	interacts_with	Age Group
		Population Group
		Family Group
		Group
		Patient or Disabled Group
Therapeutic or Preventive Procedure	affects	Professional or Occupational Group
		Mental Process
Professional or Occupational Group	uses	Intellectual Product
		Pharmacologic Substance
		Antibiotic
		Clinical Drug
		Drug Delivery Device
		Manufactured Object
Therapeutic or Preventive Procedure	uses	Medical Device
		Pharmacologic Substance
		Antibiotic
		Medical Device
Pharmacologic Substance	isa	Clinical Drug
		Drug Delivery Device
Antibiotic	isa	Pharmacologic Substance
Antibiotic	isa	Pharmacologic Substance
Antibiotic	isa	Antibiotic
Diagnostic Procedure	isa	Health Care Activity
Laboratory Procedure	isa	Health Care Activity
Therapeutic or Preventive Procedure	isa	Health Care Activity
Health Care Activity	isa	Health Care Activity
Therapeutic or Preventive Procedure	isa	Therapeutic or Preventive Procedure
Diagnostic Procedure	measures	Clinical Attribute
Diagnostic Procedure	measures	Organism Attribute
Laboratory Procedure	measures	Clinical Attribute
Laboratory Procedure	measures	Organism Attribute
Professional or Occupational Group	analyzes	Body Substance
		Finding
		Organism Function
		Organism Attribute
		Clinical Attribute
		Sign or Symptom

semantic network consisting of semantic types with structural similarity. For instance, the semantic collection ‘Group’ consists of the semantic types ‘Group’, ‘Age Group’, ‘Family Group’, ‘Patient or Disabled Group’, ‘Population Group’, and ‘Professional or Occupational Group’. All these semantic types have the same semantic relations. Altogether, there exist 28 semantic collections. We received 76 relations that can be used to cover activities and processes described in a CPG (see Table 2).

To identify a relation we need to know the relationship between two semantic types. As between two semantic types more than one relation can exist, we need to detect the appropriate relationship (i.e., the type of relation, e.g., performs, uses, interacts with). This is accomplished by means of the verbs in the particular sentence or clause. We only have seven different relationships. We assigned the verbs appearing in the action sentences to their particular relationship. Furthermore, we expanded our relationship-verb assignment by synonymous verbs from an online thesaurus [23] (for an example see Table 3).

Table 3. Verbs indicating a relation for relation type ‘perform’.

Verbs identified in the text

do, perform, eat, reserve, offer, register, maintain, evaluate, record, sign, use, consider, repeat, relieve, avoid, provide, insert

Synonyms from thesaurus

execute, discharge, accomplish, achieve, fulfill, start, complete, conduct, effect, dispatch, work, implement, require, undertake, order, arrange, keep, continue, discontinue, produce, install, load, push, drink, include, recommend, report, increase, reduce, treat, wait, initiate, carry out, carry off, ...

4 Evaluation

In order to show whether our semantic relations can be used also on other guideline documents, we performed an empirical study. We therefore used the guideline “Management of labor” [24]. Although the guideline covers the same application area, it is structured in a different way, uses different phrasing and wording. Furthermore, it contains lots of background information and literature references and it is almost twice as large (60 pages) as the document we used to define the set of semantic relations.

In a first step we had to identify all the actions and processes in the document that would be necessary to model the control flow. This was done by an expert for modelling guidelines using UMLS release 2010AA for identifying concepts and their semantic types. He also made an initial suggestion of the relation applicable to the action. For the whole document 95 sentences were identified containing actions or procedures to perform.

We then applied the MMTx program to the whole document to identify concepts and assign them semantic types. We further checked whether a relation between the

semantic types exists that indicate an action and verified whether the verb that should indicate the relationship was also available in our dictionary.

With our relations we could correctly identify 67 actions. 33 actions could not be identified because the action was not expressed as a relation (e.g., grammatically incomplete sentences) or because no UMLS concept could be identified and therefore, no semantic type was assigned (see Table 5 for examples). In only two sentences actions were identified that were not modelled by our expert. Thus, our method has a recall of 67% and a precision of 97% (see Table 4).

Table 4. Evaluation results.

COR	INC	POS	ACT	REC	PRE
67	2	100	69	67%	97%

COR = number of correctly identified actions by the method
 INC = number of incorrect identified actions by the method
 POS = number of actions according to the key target template
 ACT = number of actions identified by the method
 REC = COR/POS
 PRE = COR/ACT

Table 5. Examples for actions that could not be identified.

Sentence	Reason
Monitoring of fetal heart rate.	incomplete sentence
In the presence of preterm labor with bleeding, IV access is essential.	no relation identified for 'IV access is essential'
Do not start active pushing as soon as patient is fully dilated.	no UMLS concept found for 'active pushing'
Management of labor dystocia is especially important in the nulliparous woman to prevent unneeded Caesarean sections.	no relation identified for 'management ... is especially important ...'

In order to also include actions that were now erroneously omitted we have to consider incomplete sentences (e.g., headings, table items, lists) as well as sentences that contain an adjective instead of a verb to identify the relationship.

It is also to notice that the majority of relations appearing in the guideline are the same in the initial relation set. Only 15 actions were identified by nine relations that had been added to the initial relation set. These relations all contain semantic types that belong to the same Semantic Collection as the initial relation.

By using our dictionary for detecting the relationship we are able to omit the wrong identification of text that is not expressing an action. For instance, the sentence 'Vaginal examination is aimed at evaluation of cervical effacement ...' would be tagged with the

semantic type *Health Care Activity* for both ‘vaginal examination’ and ‘evaluation of cervical effacement’. But the verb phrase ‘is aimed at’ could indicate an intention and not the relationship ‘isa’, which is the only relationship between *Health Care Activity* and *Health Care Activity*.

5 Conclusions and Further Work

This work presents a method to identify actions and procedures to be performed during a treatment which are described in a guideline document. We used relations of the UMLS Semantic Network [10] to identify these actions in a guideline document. We defined a set of relations that are relevant for these kind of documents. With these relations we are able to identify a large part of activities that are contained in the control flow of such documents.

By assigning semantic information to text and linking it using semantic relations a further processing is facilitated. This can be used for guideline modelling into a computer-interpretable formalism. Manual and semi-automatic modelling will benefit from the information gathered by this method. Furthermore, this method can also support the processing of new versions of guidelines by identifying new or different relations.

Our next steps will be (1) the expansion of using the Semantic Network for identifying other information dimensions, such as effects, intentions, or parameters; (2) the definition and categorization of further relations to be able to differentiate between a single action, a decomposition of an action, or a selection of one or more alternative actions; (3) defining special patterns to identify actions that are either not identifiable by a relation or where the relation is identifiable by adjectives; and (4) a further processing to (automatically) transform the guideline document towards a computer-interpretable guideline (e.g., into the *Many-Headed-Bridge* (MHB) [9] formalism). This can promote the application of computer-interpretable CPGs.

Acknowledgement. This work is partially supported by “Fonds zur Förderung der wissenschaftlichen Forschung FWF” (Austrian Science Fund), grant TRP71-N23, and the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 2161.

References

1. Field, M.J., Lohr, K.N., eds.: Clinical Practice Guidelines: Directions for a New Program. National Academies Press, Institute of Medicine, Washington DC (1990)
2. Quaglini, S., Ciccarese, P., Micieli, G., Cavallini, A.: Non-compliance with guidelines: Motivations and consequences in a case study. In Kaiser, K., Miksch, S., Tu, S.W., eds.: Computer-based Support for Clinical Guidelines and Protocols. Proceedings of the Symposium on Computerized Guidelines and Protocols (CGP 2004). Volume 101 of Studies in Health Technology and Informatics., Prague, Czech Republic, IOS Press (2004) 75–87
3. Fox, J., Patkar, V., Chronakis, I., Begent, R.: From practice guidelines to clinical decision support: closing the loop. *Journal of the Royal Society of Medicine* **102**(11) (2009) 464–473

4. Peleg, M., Tu, S.W., Bury, J., Ciccarese, P., Fox, J., Greenes, R.A., Hall, R., Johnson, P.D., Jones, N., Kumar, A., Miksch, S., Quaglini, S., Seyfang, A., Shortliffe, E.H., Stefanelli, M.: Comparing computer-interpretable guideline models: A case-study approach. *Journal of the American Medical Informatics Association (JAMIA)* **10**(1) (Jan-Feb 2003) 52–68
5. Tu, S.W., Musen, M.A., Shankar, R., Campbell, J., Hrabak, K., McClay, J., Huff, S.M., McClure, R., Parker, C., Rocha, R., Abarbanel, R., Beard, N., Glasgow, J., Mansfield, G., Ram, P., Ye, Q., Mays, E., Weida, T., Chute, C.G., McDonald, K., Mohr, D., Nyman, M.A., Scheital, S., Solbrig, H., Zill, D.A., Goldstein, M.K.: Modeling guidelines for integration into clinical workflow. In Fieschi, M., Coiera, E., Li, Y.C.J., eds.: *Proceedings from the Medinfo 2004 World Congress on Medical Informatics*. Volume 107 of *Studies in Health Technology and Informatics*, AMIA, IOS Press (2004) 174–178
6. Svátek, V., Růžička, M.: Step-by-step mark-up of medical guideline documents. *International Journal of Medical Informatics* **70**(2-3) (July 2003) 319–335
7. Seyfang, A., Kaiser, K., Miksch, S.: Modelling clinical guidelines and protocols for the prevention of risks against patient safety. In: *Proceedings of the XXII International Conference of the European Federation for Medical Informatics*. Volume 150 of *Studies in Health Technology and Informatics*, Sarajevo, Bosnia and Herzegovina, IOS Press (2009) 633–637
8. Shiffman, R.N., Karras, B.T., Agrawal, A., Chen, R., Marenco, L., Nath, S.: GEM: a proposal for a more comprehensive guideline document model using XML. *Journal of the American Medical Informatics Association (JAMIA)* **7**(5) (2000) 488–498
9. Seyfang, A., Miksch, S., Marcos, M., Wittenberg, J., Polo-Conde, C., Rosenbrand, K.: Bridging the gap between informal and formal guideline representations. In Brewka, G., Coradeschi, S., Perini, A., Traverso, P., eds.: *European Conference on Artificial Intelligence (ECAI-2006)*. Volume 141 of *Frontiers in Artificial Intelligence and Applications*, Riva del Garda, Italy, IOS Press (2006) 447–451
10. McCray, A.T.: UMLS Semantic Network. In: *Proc. of the 13th Annual Symposium on Computer Applications in Medical Care (SCAMC'89)*. (1989) 503–507
11. Gennari, J.H., Musen, M.A., Fergerson, R.W., Grosso, W.E., Crubézy, M., Eriksson, H., Noy, N.F., Tu, S.W.: The Evolution of Protégé: An Environment for Knowledge-based Systems Development. *International Journal of Human Computer Studies* **58**(1) (2003) 89–123
12. Boxwala, A.A., Peleg, M., Tu, S.W., Ogunyemi, O., Zeng, Q., Wang, D., Patel, V.L., Greenes, R.A., Shortliffe, E.H.: GLIF3: A representation format for sharable computer-interpretable clinical practice guidelines. *Journal of Biomedical Informatics* **37**(3) (June 2004) 147–161
13. Moser, M., Miksch, S.: Improving clinical guideline implementation through prototypical design patterns. In Miksch, S., Hunter, J., Keravnou, E., eds.: *Proceedings of the 10th Conference on Artificial Intelligence in Medicine in Europe, AIME 2005*. Volume 3581 of *LNAI*, Aberdeen, UK, Springer Verlag (2005) 126–130
14. Serban, R., ten Teije, A., van Harmelen, F., Marcos, M., Polo-Conde, C.: Extraction and use of linguistic patterns for modelling medical guidelines. *Artificial Intelligence in Medicine* **39**(2) (2007) 137–149
15. Serban, R., ten Teije, A.: Exploiting thesauri knowledge in medical guideline formalization. *Methods Inf Med* **48** (2009) 468–474
16. Peleg, M., Tu, S.W.: Design patterns for clinical guidelines. *Artificial Intelligence in Medicine* **47**(1) (2009) 1 – 24
17. Essaihi, A., Michel, G., Shiffman, R.N.: Comprehensive categorization of guideline recommendations: Creating an action palette for implementers. In: *AMIA 2003 Symposium Proceedings*, AMIA (2003) 220–224
18. Kaiser, K., Akkaya, C., Miksch, S.: How can information extraction ease formalizing treatment processes in clinical practice guidelines? A method and its evaluation. *Artificial Intelligence in Medicine* **39**(2) (2007) 151–163

19. Denecke, K.: Semantic structuring of and information extraction from medical documents using the umls. *Methods of Information in Medicine* **47** (2008) 425–434
20. National Collaborating Centre for Women’s and Children’s Health: Induction of labour. Clinical guideline 70, National Institute for Health and Clinical Excellence (NICE), London (UK) (July 2008)
21. Aronson, A.R.: Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In: *Proceedings of the 25th Annual American Medical Informatics Association Symposium (AMIA’01)*, Washington, D.C., AMIA, Washington, D.C. (Nov. 3-7 2001) 17–21
22. Chen, Z., Perl, Y., Halper, M., Geller, J., Gu, H.: Partitioning the UMLS Semantic Network. *IEEE Transactions on Information Technology in Biomedicine* **6**(2) (June 2002) 102–108
23. Lindberg, C.A., ed.: *Oxford American Writer’s Thesaurus*. 2nd edn. Oxford University Press (2008)
24. Institute for Clinical Systems Improvement (ICSI): Management of labor. Clinical guideline, Institute for Clinical Systems Improvement (ICSI), Bloomington (MN) (May 2009)

Updating a Protocol-based Decision-Support System's Knowledge Base: a Breast Cancer Case Study

Claudio Eccher¹, Andreas Seyfang², Antonella Ferro³, and Silvia Miksch^{2,4}

¹ Fondazione Bruno Kessler, Trento, Italy
cleccher@fbk.eu

² Vienna University of Technology, Austria
seyfang@ifs.tuwien.ac.at

³ Medical Oncology, S. Chiara Hospital, Trento, Italy
antonella.ferro@apss.tn.it

⁴ Danube University Krems, Austria
silvia.miksch@donau-uni.ac.at, miksch@ifs.tuwien.ac.at

Abstract. Modelling clinical guidelines or protocols in a computer-executable form is a prerequisite to support their execution by Decision Support Systems. Progress of medical knowledge requires frequent updates of the encoded guideline model. Moreover, user perception of the decision process and user preferences regarding the presentation of choices require modifications of the model. In this paper, we describe these two maintenance requirements using a protocol for the medical therapy of breast cancer and the lessons learnt in the process. The protocol was modeled in Asbru and is used at the S. Chiara Hospital in Trento.

1 Introduction

In the context of Evidence Based Medicine Clinical Practice Guidelines (CPGs), defined as “...systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” [1], are the means for efficiently disseminating the ever increasing amount of available clinical knowledge. Clinical protocols, distillation of the knowledge available in books, articles, and CPGs adapted to the local resources and conventions at a specific site, can reduce the undesired variation in the provision of care and, in turn, improve the quality of care in a healthcare organization. A Computerized Decision-Support System (DSS) supporting guideline-based or protocol based-care in an automated fashion at the time and location of decision-making can promote and improve the compliance of clinicians with protocols, especially when integrated in the clinical workflow [2, 3]. In some fast-changing domains, moreover, computer support for protocol execution becomes even more important to efficiently spread the new knowledge and allow clinicians to provide evidence-based up-to-date treatments. The modeling of clinical protocols in machine processable form, although less time consuming than that of CPGs, is not trivial and requires a long phase of knowledge acquisition that must be carried out in close collaboration between domain experts (physicians) and knowledge engineers.

The problem is that the operational knowledge bases formalized in the knowledge acquisition phase must be regularly maintained as the clinical knowledge progresses

or in response to specific user requirements. This periodic activity can be required at relatively short intervals [4], as in the case of treatment protocols for breast cancer.

In this paper, we describe our experience of knowledge base updating within the Oncocure project, aimed to build a protocol-based DSS for supporting the provision of medical treatment to breast cancer patients at the Medical Oncology Unit (MOU) of the S. Chiara Hospital of Trento (Northern Italy). This work relies on the data ontology we defined in the course of the project to bridge the gap between a DSS and the legacy EPR used in the unit to manage oncologic patients [5].

2 Related work

While several papers in literature present tools, frameworks, or methodologies to support the process of formalizing textual CPGs in executable models and adapting them to local settings, the maintenance of multiple knowledge bases has not been adequately addressed by now.

In [6] a model-centric approach to guideline versioning for GLIF (Guideline Interchange Format) is proposed, in which GLIF ontology is extended by version information and a versioning tool was developed to support the modification of an existing model. Uruz, part of the Digital Electronic Guideline Library (Degel) framework [7], is a web-based markup tool that supports indexing and markup using any hierarchical guideline-representation format.

Intermediate languages have been proposed to facilitate the guideline maintenance process. In a document-centric approach, MHB (Many-Headed Bridge) [8] is an intermediate representation to bridge the gap between text-based representations and formal models (Asbru, GLIF, etc.). In the Oncocure project, the most important parts of the protocols are diagrams which cannot be processed yet in the DELT/A editor [9], which is used to annotate a protocol in MHB. Also, the initial modeling task appeared simple enough to be performed in one step, since the diagrams showed the decision pathways rather clearly. Only the further development and maintenance work described in this paper makes the utilization of such a tool appear desirable.

[10] presents the LASSIE methodology for automating part of the modeling process, by formalizing guidelines in several steps from the textual form to the Asbru language using a document-centric approach.

Besides dealing with changes during the development process of guidelines, other approaches focus on changes during the adaptation process. Peleg et al. [11] developed a model-based approach for adapting guidelines to local circumstances and encoding them in GLIF3 formalism. Terenziani et al. [12] proposed a context-adaptable approach to clinical guidelines.

But still, using the mentioned tools the modeling process is complex and labor intensive. Methods are needed to automate the most parts of the modeling task.

3 The Oncocure Project

The cancer care process carried out at the MOU is centered around periodic encounters with the patient, in which the oncologist decides the appropriate therapeutic strategy

on the basis of the patient visit and exam results. The decision is based on internal protocols prepared by the oncologist specialist in the corresponding cancer type.

The two-year Oncocure project, started in 2007, developed a prescriptive guideline-based DSS integrated with the legacy web-based oncologic electronic patient record (EPR) in use at the MOU, based on the Asbru model of the internal protocol for the administration of medical therapy to breast cancer patients [13]. The DSS wraps the Asbru interpreter to interface it with the EPR database on one side and with a Web-based user interface on the other, so that the DSS can use EPR data, display recommendation to and receive feedback from the user, in order to provide the oncologists with the most appropriate therapeutic strategy for the given disease in the presence of the specific tumor and patient conditions (tumor morphology, patient age, etc.).

4 The breast cancer treatment protocols

The internal breast cancer treatment protocols are prepared by the breast cancer specialist of the MOU. In contrast to guidelines, which are mostly in textual form, the internal protocols are more structured, since their aim is to constitute an easy-to-use treatment handbook. The first step of the guideline formalization process, in fact, is made by the oncologists who, combining the knowledge in CPGs issued by national and international authorities and recommendations issued by periodic breast cancer conferences (e.g., the S. Gallen Consensus Conference), produces an informal conceptual model mainly constituted by “box-and-arrow” diagrams with short explanation text (Figure 1). In the Oncocure Project we followed a model-centric approach [14] without losing contact with the document structure. To produce a valid Asbru plan hierarchy keeping precise references to the source document we used a collaborative tool based on Semantic MediaWiki technology [15], which allowed web-based collaborative model editing, mixing formal annotations and informal content (text and images). The exported ‘skeletal’ model was then refined into a fully functional model using the DELT/A editor.

5 The DSS knowledge bases

In a preceding paper [5] we presented an ontology of parameters to bridge the gap between the Asbru-based DSS and the legacy EPR used by the oncologists at the point of care. We identified three kinds of parameters required by the model:

- *primitive parameters*, i.e. data directly available in the EPR database (e.g., the number of metastatic lymph nodes, the percentage of estrogen receptor positive cells);
- *abstractions*, i.e. parameters that can be computed without human intervention from lower-level data through clearly defined rules. The abstractions can be higher level concepts used in the protocols (e.g., the hormone responsiveness, the patient risk level), taxonomic abstractions (e.g. in the TNM classification system a T1a tumor is-a T1 tumor), or temporal abstractions (e.g., taxanes in previous treatments);
- *‘holistic’ parameters*, i.e. parameters to ask to the doctor, because it is not possible to define a computing algorithm with certainty or, even though the parameters were broken down into elementary findings, the resulting elements would themselves require subjective judgment by the oncologist (e.g., tumor aggressiveness).

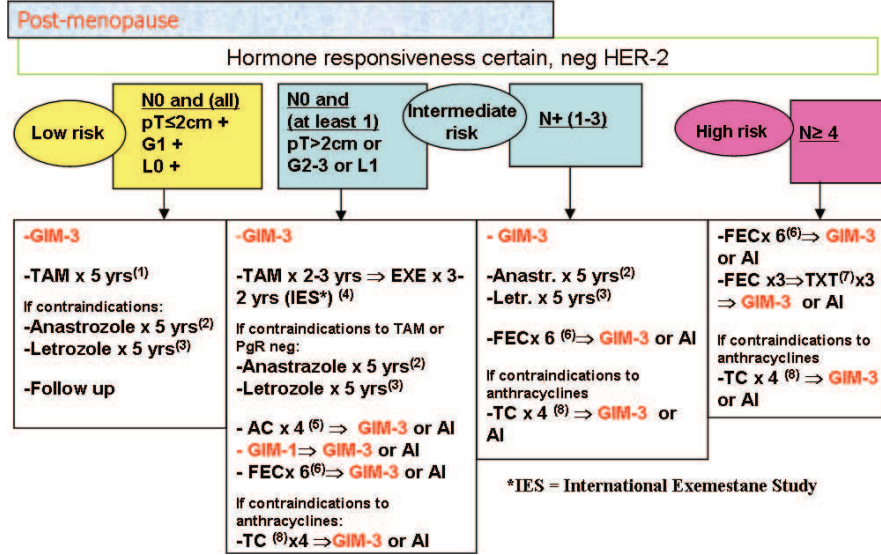


Fig. 1. A protocol informal diagram representing the treatments (in the boxes below the thin vertical arrows) recommended for the eligibility conditions defined above the arrows. Her-2 is the status of Her2 (ErbB-2) receptors. N, G, L, and pT are the lymph node involvement, the histologic grade, the lymphovascular invasion and the tumor size, respectively. In the therapy boxes, the acronyms TAM and EXE stand for Tamoxifen and Exemestane. AC, FEC and TC stand for the chemotherapy combinations Adriamycin-Cyclophosphamide, Fluorouracyl-Epirubicine-Cyclophosphamide and Taxotere-Cyclophosphamide. AI is the acronym for aromatase inhibitors (Anastrozole and Letrozole). GIM-1 and GIM-3 are study protocols now closed.

Although the abstractions could be performed by the Asbru data abstraction module, we opted to move most of them in an external rule engine mainly because the abstractions are, in general, rather stable, while the rules and parameters used for their computation may require more frequent revisions. This choice makes the maintenance of the guideline model easier.

The above distinction between different types of parameters and the strategies for dealing with the various types guided the implementation of the DSS architecture, based on three interrelated knowledge bases neatly separated from the information model:

- The XML-based minimum dataset for treating the breast cancer, i.e. the definition of all the parameters (divided in primitive parameters, abstractions and holistic parameters) needed by the system to execute the model: name, type, allowed values, range, etc. The breast cancer dataset is the instance of a general XSD schema, defined to allow the representation of any cancer domain.
- The rules to compute the abstractions from lower level data. In the current version the rules are written in CLIPS [16] and executed by the freely-available forward-chaining CLIPS engine.

- The Asbru model executed by the Asbru interpreter, which contains the hierarchical network of plans, the definition of the conditions connecting plan transitions, and the declaration of the parameters used in the conditions.

The minimum dataset is the reference for managing the patient Virtual Medical Record that stores the actual values of parameters produced and used by the DSS and represents the patient history related to the breast cancer disease.

6 Knowledge base maintenance

6.1 First model version

In Europe one of the main sources of breast cancer treatment recommendations is the International Consensus Conference held every two years in St. Gallen (Switzerland), in which breast cancer experts try to reach a consensus on the best patient treatment based on available scientific evidence from randomized clinical trials and their biological relevance. The first knowledge model logic developed in the course of the Oncocure project was based on the internal protocol derived mainly from the 10-th St. Gallen 2007 Consensus Conference report for the treatment of early breast cancer [17].

Dataset and abstraction rules In the first version of the breast cancer minimum dataset model we defined two abstractions related to St. Gallen 2007 recommended thresholds for choosing an adjuvant treatment: a three value parameter representing the hormone responsiveness of the tumor (certain, incomplete and absent) and a parameter representing the patient risk level that determines the choice of the treatment modality. Actually, the protocol diagrams (Figure 1) indicated a three-value risk level (low, intermediate and high). For the sake of simplicity, however, the modeler defined a boolean low risk level abstraction and combined it with the number of metastatic regional lymph nodes in the Asbru filter preconditions to obtain the eligibility conditions for the different treatments for each combination of hormone receptor and Her2 receptor status. Hence, in the CLIPS knowledge base we modeled a set of rules to compute the risk level as a logical combination of lower level parameters: tumor size (T), patient age, histologic tumor grade (G) and lymphovascular invasion (L). Another set of rules was defined to compute the hormone responsiveness as a logical combination of estrogen and progesterone receptor status, T, G, L and the Mib1 proliferation index⁵. After discussing with the oncologist her decision-making process, it resulted that G and L are parameters of secondary importance she used only to 'refine' the main treatment modality decision (chemotherapy, hormone therapy, or both). Therefore, we defined rules for assigning default values to G and L in the case they are missing in the database, in order to recommend the treatment of lowest impact for the patient. This 'precautionary' principle was also implemented for the Mib1 index used to compute hormone responsiveness.

Asbru model We organized the Asbru plans firstly following the protocol diagrams and secondly aiming to the simplicity of representation. Hence we grouped plans organizing equivalent therapies at the same level without making distinction between treatment modalities.

⁵ Expressed as a percentage, Mib1 indicates the proliferation activity in breast cancer.

6.2 Customization of the knowledge

Some modeling choices in the knowledge bases determine the information displayed in the user interface. For justifying the decision recommended by the system all the parameters are shown, including the abstractions. Asbru plans children of the same parent that are ready to switch to the active state are displayed as a set of equivalent recommended alternatives.

After the system was implemented the oncologists tested the correctness of the model performing simulations in our laboratory with real patient data from an anonymous copy of the EPR. After about ten simulations the oncologist realized that, although the model was logically correct, some modeling choices made by the knowledge engineer had unanticipated consequence on the oncologist's perception of the information displayed by the system. In fact, the modeler privileged the elegance of the representation and the efficiency of execution and did not consider user preferences that, in any case, were not expressed by the oncologist till she tried to use the working system. Consequently a request followed for modifying the display of data and recommendations, which required a process of knowledge customization that consisted in modifying abstractions and Asbru plan hierarchy.

Abstractions. The main critique the oncologist made was that the boolean low level risk abstraction displayed in the user interface, though logically correct, did not represent well the clinical reality, in that the oncologist wanted to see the St. Gallen risk levels to understand the clinical situation of the patient. Hence, we firstly customized the abstractions by redefining the old boolean risk value to a three-level risk both in the breast cancer dataset and in the domain parameter section of the Asbru model. Accordingly, we modified the algorithm to compute the risk level in the CLIPS knowledge base including also the number of regional lymph nodes with metastasis.

Asbru model. The second criticism of the first model was that the modeler grouped set of equivalent treatments as they were represented in the diagrams (see Figure 1), while the oncologist 'thinks' in terms of therapies organized according to the treatment modality (hormone therapies, chemotherapies, possibly concomitant with herceptin, or combination of chemotherapy followed by hormone therapy) and found more convenient if the system presented a hierarchical choice between different therapy modalities first, and then the detail of the (combination of) drugs constituting each treatment. We performed such a customization by regrouping treatment plans under heading plans representing concepts such as hormone therapy alone, chemotherapy alone, chemotherapy followed by hormone therapy, etc.

The resulting model was logically equivalent to the original. However, the grouping of plans and abstractions reflects the oncologist's perception and to a lesser extent that of the computer scientist

6.3 Progress of the clinical knowledge

The oncological treatment domain in general, and breast cancer in particular, is changing fast. As the clinical knowledge evolves following huge clinical trials, eligibility

conditions are continuously modified and new drugs or combination of drugs are recommended. This leads to the need to update the model of the knowledge accordingly.

The 11-th St. Gallen Consensus meeting, held in March 2009, introduced new complexities for the treatment of early breast cancer. According to the authors of the 2009 edition report [18], in fact, “[...] a fundamentally different approach from that used in previous consensus reports” was adopted. In particular, the 2009 consensus panel clarified the indications for the adjuvant systemic treatments for early breast cancer by refining the algorithms for choosing between the types of treatment modalities available (endocrine therapy, anti-HER2 therapy, chemotherapy), and identifying new ‘thresholds for indication’ of each type of systemic treatment.

We organized three meetings with the oncologist to analyze the new St. Gallen recommendations and compare them with the old ones, in order to define a new formalization including the changes. Again, we identified two main modifications needed to include these indications into the knowledge bases of the DSS: redefinition of some abstractions and regrouping of (combination of) drugs for specific patient conditions.

Update of the system knowledge base required changes in the three knowledge models of the system (the breast cancer minimum dataset, the CLIPS rules and the Asbru model) but thanks to the modularization of our architecture with the rigid separation between the knowledge and the information model this was not a problem at all.

Abstractions. To take into account the St. Gallen 2009 recommendations, which focused mainly on the definition of new thresholds for recommending each type of systemic treatment modality, we modeled the following changes:

- To determine the hormone responsiveness of a tumor the report recommended to change the threshold on the estrogen and progesterone receptor percentages from 70% to 50%. Consequently, in the CLIPS knowledge base we modified the old algorithms for calculating the three-value hormone responsiveness parameter (absent, uncertain, complete) to include the new thresholds.
- To take into account the new St. Gallen thresholds we changed the three-level risk value to a five-level risk value: *low* level, *high* level, and three intermediate levels - *medium-low*, *medium* and *medium-high*. Hence we modified the definition of the risk level both in the breast cancer dataset and in the domain parameter section of the Asbru model.
- We defined a more complex algorithm to compute the risk level that, in addition to the parameters above, includes the Mib-1 proliferation index, not considered before. We rewrote the necessary rules in the CLIPS knowledge base and updated the Asbru model changing the definition of eligibility conditions (i.e. filter preconditions) in which the risk level parameter was used.
- According to the new recommendations, grade G and Mib-1 have become primary factors to assess the risk level and, consequently, to choose the appropriate treatment. According to the oncologist, if these values are unreliable or missing, she has to demand a new pathological examination of the cancer tissue. Therefore, when G or Mib-1 are missing default values cannot be assigned to them anymore and we had to remove from the CLIPS knowledge base the corresponding default assignment rules.

Asbru model. The new St. Gallen recommendations substantially did not introduced new (combinations of) drugs for breast cancer treatment but, differently from the 2007 recommendations, the final 2009 report identified a “gray zone” of certainly and partially hormone-responsive Her2-negative tumors for which the decision is particularly difficult because they present both low and high risk characteristics. The indications are to split the set of treatments in first, second and third (currently the last) generations based on the combination of drugs used (new generation drugs being more specific and potent), and associate the three new intermediate risk values defined above with the treatment generations, recommending the newest combination of drugs for the highest risk levels. Following this indication, we modified the Asbru model by regrouping plans representing medical treatments in first, second and third generation treatments and redefining the required filter preconditions for each group. Lastly, since the study protocols were closed (GIM1 and GIM3 in Figure 1), we removed them from the model.

7 Lessons learnt and next steps

Need for customization. While it is clear that the model must always cover the domain experts’ perception to allow communicate the modeling process to them, it was surprising how much the interactive setting of choosing the treatment steps left its traces in the model. One explanation for this is the fact that in this application, all decisions are taken by the physician adding her own knowledge, part of which is not in the model. While there are clear rules about the preferences for certain treatments in each case, the actual choice depends on factors such as patient agreement and subjective suitability for certain treatment options.

Need for links between the original and the model. The decision trees in the protocol appear to be very simple. The original model followed this simple organization in the most straightforward fashion, as seen by a computer scientist. However, after restructuring the model to meet the user requirements, and later adapting it for changes in the original protocol, these connections became harder to recognize, and maintenance effort was increased.

Need for on-line editing. The protocol was modeled in a joint effort between Vienna and Trento. To allow modular, web-based editing of the model, we adapted Semantic MediaWiki for Asbru. However, without an easy conversion facility in both directions, i.e., also from Asbru to the Wiki, the different representations came out of sync in the maintenance process. The distributed authoring environment under development at Vienna University of Technology will provide such functionality. The Oncocure protocol will help to evaluate these tools.

Need for advanced techniques for linking to the original. The original protocol consists mainly of a series of diagrams with few text. Traditional linking tools mostly focus on text sources. Also, adding links into the source files, like DELT/A does, requires the user to first edit the linked file to reflect the changes. The next generation of modelling and annotation tools must therefore provide means to store links to highly structured documents outside of them.

8 Conclusion

The medical knowledge changes over the time, often at a fast pace. The knowledge base of a Decision-Support System, which formalizes the part of this knowledge of interest, must follow these changes. But the progress in medical knowledge is not the only source of changes. In fact, the knowledge base content and structure can impact on the way the information is displayed and the integration of the system in a clinical setting requires to take into account the user requirements and customize the knowledge bases accordingly. The periodic maintenance of the DSS knowledge base, consequently, is a difficult but necessary task if the system has to be successfully implemented in the clinical routine.

In this paper we have presented a case study regarding the maintenance of the knowledge base of a guideline-based DSS executing a breast cancer medical treatment protocol formalized in Asbru to give support to the oncologist in the therapy decision making. In our experience, the knowledge maintenance can be divided in two distinct tasks: the customization of the knowledge base to the user's preferences and the update to follow the progress of the medical knowledge. The first task is user-dependent and must be done when the user, after a period of system use, requires modifications that involve the modeling choices made to built the knowledge-base. We have given some example form a real case. The second task depends on the availability of new indications in literature or issued by national and international committees, and in general can be periodically scheduled following important events, such as consensus conferences and international symposia on the specific disease. The two phases are not independent, however, since a progress in the knowledge can lead also to the customization of the new knowledge for the user's preferences.

Separating different aspects of the knowledge in distinct knowledge bases, such as data abstraction and plan hierarchy in our case, increases transparency and eases the maintenance work. However, special care must be taken to keep the different parts of the knowledge base in sync in the change process. Without tools performing the required checks automatically, this is an additional source of error.

In our case, we were able to stay up-to-date both with oncologist's preferences and progress in medical knowledge. The fast adaptation of knowledge based increases user acceptance of our system. However, more advance editing tools would improve the maintenance by reducing labor cost and errors at the same time. Thus, our next steps will focus on the further development of our modeling tools.

References

1. Field, M., Lohr, K., eds. In: Clinical practice guidelines: directions for a new program. Institute of Medicine, National Academy Press, Washington DC (1990)
2. Sonnenberg, F., Hagerty, C.: Computer interpretable guidelines: where are we and where are we going? 2006 IMIA Yearbook of Medical Informatics, Methods Inf Med **45**(Suppl 1) (2006) S145–158
3. Bates, D., Kuperman, G., Wang, S., Gandhi, T., Kittler, A., Volk, L., Spurr, C., Khorasani, R., Tanasi-jevic, M., Middleton, B.: Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc **10** (2003) 523–530

4. Seyfang, A., Martinez-Salvador, B., Serban, R., Wittenberg, J., Miksch, S., Marcos, M., ten Teije, A., Rosenbrand, K.: Maintaining Formal Models of Living Guidelines Efficiently. In Bellazzi, R., Abu-Hanna, A., Hunter, J., eds.: Proc. of 11th Conference on Artificial Intelligence in Medicine (AIME 2007). Volume 4594 of LNCS. Springer (2007) 441–445
5. Eccher, C., Seyfang, A., Ferro, A., Stankevich, S., Miksch, S.: Bridging an Asbru Protocol to an Existing Electronic Patient Record. In Riaño, D., ten Teije, A., Miksch, S., Peleg, M., eds.: Knowledge Representation for Health-Care: Data, Processes and Guidelines. Revised Selected Paper of the KR4HC 2009 Workshop. Volume 5943 of LNAI. Springer (2010) 14–25
6. Peleg, M., Kantor, R.: Approaches for guideline versioning using GLIF. In Musen, M., ed.: Proc. of the 2003 AMIA Symposium, AMIA (2003) 509–513
7. Shahar, T., Young, O., Shalom, E., Galperin, M., Mayaffit, A., Moskovitch, R., Hessing, A.: A framework for a distributed, hybrid, multiple-ontology clinical-guideline library, and automated guideline-support tools. *J Biomed Inform* **37** (2004) 325–344
8. Seyfang, A., Miksch, S., Marcos, M., Wittenberg, J., Polo-Conde, C., Rosenbrand, K.: Bridging the gap between informal and formal guideline representations. In Brewka, G., Coradeschi, S., Perini, A., Traverso, P., eds.: 17th European Conference on Artificial Intelligence, IOS Press (2006) 447–451
9. Votruba, P., Miksch, S., Kosara, R.: Facilitating Knowledge Maintenance of Clinical Guidelines and Protocols. In Fieschi, M., Coiera, E., J., L., eds.: MedInfo 2004, AMIA (2004) 57–61
10. Kaiser, K., Miksch, S.: Versioning computer-interpretable guidelines: Semi-automatic modeling of 'living guidelines' using an information extraction method. *Artif Intell Med* **46** (2009) 55–66
11. Peleg, M., Wang, D., Fodor, A., Keren, S., Karnieli, E.: Lessons Learned from Adapting a Generic Narrative Diabetic-Foot Guideline to an Institutional Decision-Support System. In ten Teije, A., Miksch, S., Lucas, P., eds.: Computer-based Medical Guidelines and Protocols: A Primer and Current Trends. Volume 139 of Stud Health Technol Inform. IOS Press (2008) 243–252
12. Terenziani, P., Montani, S., Bottrighi, A., Torchio, M., Molino, G., G., C.: A context-adaptable approach to clinical guidelines. In Fieschi, M., Coiera, E., Li, J., eds.: Proc. from the Medinfo 2004, AMIA (2004) 169–173
13. Eccher, C., Seyfang, A., Ferro, A., Miksch, S.: Embedding Oncologic Protocols into the Provision of Care: The Oncocure Project. In Adlassing, K., Blobel, B., Mantas, J., Masic, I., eds.: Medical Informatics in a United and Healthy Europe: Proceedings of MIE 2009. Volume 150 of Stud Health Technol Inform. IOS Press (2009) 663–667
14. Leong, T.Y., Kaiser, K., Miksch, S.: Free and Open Source Enabling Technologies for Patient-Centric, Guideline-Based Clinical Decision Support: A Survey. 2006 IMIA Yearbook of Medical Informatics, *Methods Inf Med* **46**(1) (2007) 74–86
15. Eccher, C., Ferro, A., Seyfang, A., Rospocher, M., Miksch, S.: Modeling Clinical Protocols Using Semantic Media Wiki: The Case of the Oncocure Project. In Riaño, D., ed.: Knowledge Management for Health Care Procedures. Revised Selected Papers of the K4Help 2008 Workshop. Volume 5626 of LNAI. Springer (2008) 663–667
16. CLIPS: A Tool for Building Expert Systems: <http://clipsrules.sourceforge.net/index.html>. Last visited 2009, September 28.
17. Harbeck, N., Jakesz, R.: St. Gallen 2007: Breast Cancer Treatment Consensus Report. *Breast Care* **2** (2007) 130–134
18. Goldhirsch, A., Ingle, J., Gelber, R., Coates, A., Thrlimann, B., Senn, H., Panel members: Thresholds for therapies: highlights of the st gallen international Expert Consensus on the Primary Therapy of Early Breast Cancer 2009. *Ann Oncol* **20**(8) (2009) 1319–1329

Toward Probabilistic Analysis of Guidelines

Arjen Hommersom

Radboud University Nijmegen
Institute for Computing and Information Sciences
Nijmegen, The Netherlands
`arjenh@cs.ru.nl` *

Abstract. In the formal analysis of health-care, there is little work that combines probabilistic and temporal reasoning. On the one hand, there are those that aim to support the clinical thinking process, which is characterised by trade-off decision making taking into account uncertainty and preferences, i.e., the process has a probabilistic and decision-theoretic flavour. On the other hand, the management of care, e.g., guidelines and planning of tasks, is typically modelled symbolically using temporal, non-probabilistic, methods. This paper proposes a new framework for combining temporal reasoning with probabilistic decision making. The framework is instantiated with a guideline modelling language combined with probabilistic pharmacokinetics and applied to the treatment of diabetes mellitus type 2.

1 Introduction

Clinical guidelines are highly structured documents providing appropriate standards of care. Recommendations of clinical guidelines are based around actions, also sometimes referred to as tasks or interventions, that physicians are advised to perform when treating specific groups of patients. If there is only one possible action in the care pathway, the situation is relatively easy. For many diseases, however, there are at least several actions that have to be performed. In order to model such care pathways, expressive formalisms were developed such as Asbru, *PROforma*, and GLARE (see [2]). These languages are best characterised by the term ‘task-network models’, as they model the guideline as a network of component tasks that unfold over time [11].

In order to ensure quality of these guidelines, several symbolic analysis approaches have been proposed, such as simulation and verification, e.g., using formal methods for checking that resulting guidelines comply to certain quality criteria (see, e.g., [7]). Whereas the recommendations can be looked upon as symbolic and logical entities, the underlying knowledge, typically medical evidence, involves uncertainty. So far, this aspect of a guideline has been largely ignored, but is required for a deeper analysis of the quality of guidelines.

* Visiting researcher at the Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, BE-3001 Heverlee, Belgium.

The representation of uncertainty is particularly important if one wants to support the guideline development process. As far as we are aware, there are no knowledge representation formalisms specifically designed for this task. Indeed, the main challenge for such a representation is to integrate knowledge underlying guidelines with medical decision making taking into account uncertainty derived from the evidence. A complicating factor is that the current process of developing recommendations from uncertain knowledge is hardly clear. For example, the ‘*The guidelines manual*’ by the British NHS [10] states the following about recommendations: “If (the guideline) combines consideration of several possible interventions, it may include discussion of the position of an intervention within a pathway of care”. There is no way to decide *how* to determine the position of the intervention within a pathway of care. In this paper, a language is introduced for exploring different possible combinations of treatments and to consider their outcomes. This also opens up a new possibility for *personalisation* of medicine as different treatment options can be explored for specific patients.

This paper is organised as follows. In Section 2, we will introduce the probabilistic framework. The implementation of this framework is discussed in Section 3 and applied to the management of diabetes mellitus type 2 in Section 4. Finally, in Section 5, we conclude and discuss future work.

2 Probabilistic Clinical Model

In this section, the probabilistic model underlying our analysis is introduced. First, we discuss the model from a symbolic point of view, after which it is refined with probabilistic aspects.

2.1 Histories

As a point of departure, we look at care as a process that can be modelled as sequences of possible actions and conditions, that we previously called *histories*. An elaborate treatment can be found in [6], where histories were studied from a more general and theoretical point of view. In short, histories describe sequences of the state of a patient and interventions being performed on that patient. Expectations extend histories with possible states in the future, constrained by clinical management and (patho-)physiological processes. Formally, we define a history H as a sequence of triples $\{(p_k, i_k, t_k)\}_{k=0}^{k=n}$, where p_k is a (patient) state, i_k a description of interventions, and t_k a time-point from a partially ordered set. We denote s_k for the *state* consisting of the combination of the patient state and the performed interventions, i.e., $s_k = (p_k, i_k)$. In this paper, the complete state is modelled as a set of attribute-value pairs $\langle a_i, v_i \rangle$, where a_i is an attribute and v_i the value of that attribute.

The set of all histories is called $\mathcal{H}$. An expectation E extends a history to other possible histories, i.e., it is a function $E : \mathcal{H} \rightarrow \mathcal{P}(\mathcal{H})$, where $\mathcal{P}(\mathcal{H})$ denotes the power set of $\mathcal{H}$, together with some boundary condition (see [6, Chapter

7]) ensuring that these expectations are sound. Sound expectations are step-functions E_s , for which holds that if $H' \in E_s(H)$, for some H that contains n triples, then $H' = H \cup \{(p_{n+1}, i_{n+1}, t_{n+1})\}$, where $t_{n+1} > t_n$, i.e., E_s extends H with information at a later time-point and at no other time-point.

2.2 Probabilistic histories

The intent of expectations is to describe possible continuations of the history. Not all of these continuations are equally likely, however. In order to model this uncertainty, we define a *probabilistic expectation* as a function E_p which associates each possible expectation with a probability, i.e.,

$$E_p : \mathcal{H} \rightarrow \mathcal{P}(\mathcal{H} \times [0, 1])$$

such that E_p is a step function, i.e., it extends H with just one new time-point in the future. Furthermore, the set of possible expectations is complete, i.e., if $E_p(H) = \{(h_1, P_1), \dots, (h_n, P_n)\}$, then $\sum P_i = 1$.

In the probabilistic model, we assume (i) time-invariance, i.e., the probability of the expectation only depends on states, but not on the times and (ii) the expectations are *Markovian*, that is, expectations depend on the present (the last state), but not on the past. Note that there are no restrictions to prevent the embedding of all information about the past in the last state of the history, except that it will impact the complexity of reasoning. If these two assumptions are combined, expectations only depend on the final state (s_n) of a history, thus expectation can be described by a transition relation:

$$s_n \xrightarrow{P(s_n, s_{n+1})} s_{n+1}$$

which yields a new history $h_{n+1} = \{\dots, (s_n, t_n), (s_{n+1}, t_{n+1})\}$, with $t_{n+1} > t_n$, as one of the expectations of the history h_n .

To further refine the model, the transition relation is then decomposed according to the state decomposition, i.e., the transition of going from v to v' is given for all attribute-value $\langle a, v \rangle$ pairs such that:

$$s_n, \langle a, v \rangle \xrightarrow{P(c(a, v, v', s_n))} \langle a, v' \rangle$$

where $c(a, v, v', s_n)$ models a random (choice) variable, which is assumed to be independent of choices for other attributes in state s_n . The complete transition probability is defined as the conjunction of each of the choices. From this, and the fact that the choices are independent of each other, we obtain:

$$P(s_n, s_{n+1}) = P\left(\bigwedge_{i=0}^m c(a_i, v_i, v'_i, s_n)\right) = \prod_{i=1}^m P(c(a_i, v_i, v'_i, s_n))$$

where $s_n = \{\langle a_i, v_i \rangle\}_{i=0}^m$ and $s_{n+1} = \{\langle a_i, v'_i \rangle\}_{i=0}^m$.

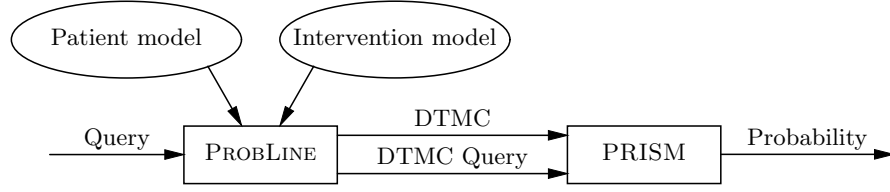


Fig. 1. General framework of PROBLINE. From a patient and intervention model it generates a discrete-time Markov chain (DTMC), which is used to compute the answer to a query.

3 ProbLine

3.1 General framework

Probabilistic histories act as a framework for a software tool tool called PROBLINE. Fig. 1 provides an overview of the approach. PROBLINE provides an interface for a patient model as well as an intervention model and answers probabilistic queries in the form $P(a_i = v_i \mid t)$, where $\langle a_i, v_i \rangle \in p_i$ and t a point in time, i.e., the probability that a patient attribute a_i has value v_i at time t . PROBLINE, as presented in this paper, is fully implemented and runs on the YAP system, which is a recent high-performance implementation of Prolog.

The intervention model and patient model describe probabilistic transitions as formally discussed in the probabilistic clinical model. As the system makes no other assumptions, different knowledge representation formalisms could be used to construct intervention and patient models.

Probabilistic histories can be interpreted as infinite discrete-time Markov chains (DTMC), i.e., a DTMC with a potentially infinite number of states. As all probabilistic queries are given a time t , and given that there are a finite number of transitions from a given state, the query can be computed on the basis of a finite DTMC. The computation of this finite DTMC that can answer the query is handled by the tool, which then calls the probabilistic symbolic model checking tool PRISM [9] to compute the answer to the query. This model checker incorporates state-of-the art symbolic data structures and algorithms for computing probabilistic queries of complex models.

3.2 Patient model

As far as we are aware, there are no dedicated knowledge representations for patient models. We therefore use the following basic representation. First, we can declare new patient attributes as follows:

- `patient_attribute(+A, +V)`
Declares a new attribute A with an initial value V .
- `patient_attribute(+A, +V, +P)`
Declares a new attribute A with possible initial values V (i.e., V is a list) and a list with probabilities P with a initial probability for each value in V .

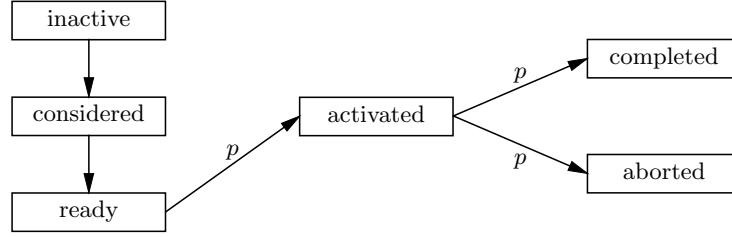


Fig. 2. Simplified Asbru state chart of [1]. The state transitions that might be probabilistic are indicated by p . The remaining two transitions are technical in the sense that they are independent of any user interaction. Conditions to go from one state to the other are not shown in this figure.

The transition system can then be modelled using the following two primitives:

- **choice**($+A, +V, +V', +S, +P$)
This corresponds to the choice operator as discussed in the previous section, where A is an attribute, V is a value, V' is an updated value, S is the state, and P is the probability of making this choice.
- **patient_av**($+S, +A, -V$)
Provides the value V of an attribute A in state S .

A graphical representation, for example a flowchart, might be more appropriate in practice, and can be built on top of these predicates.

3.3 Intervention model

As said, PROBLINE is not restricted to a specific knowledge representation. Nevertheless, in order to illustrate its capabilities, we instantiate the intervention model with a computer-interpretable guideline modelling language. In this work, we have chosen a small subset of Asbru and modelled its state-chart semantics [1]. A simplified version of the state chart semantics of Asbru is given in Fig. 2. In the original semantics there are non-deterministic choices to go from one state to the other, e.g., whether or not to abort a task if the abort condition holds. In this model, we include the possibility of probabilistic transitions between states. In particular, the transition from ready to activated, and from activated to some terminated state, i.e., completed or aborted state, is a probabilistic transition. These probabilities model the chance that a physician acts if a treatment is allowed to start or could be terminated.

The representation of tasks, in Asbru they are called plans, is implemented with the following four predicates:

- **plan_body**($+N, +T, +W, +C$)
Defines a new plan with name N , with a body type T (currently either ‘sequential’ or ‘parallel’). This plan has a list of children C and has a wait-for condition W in order to model optional and mandatory plans for N .

- `{abort,complete,filter}_condition(+N, +S)`:
Specifies in which state S the abort/complete/filter condition is true (see [1]).
If it is not specified for a state S , then it is ‘false’ by default because of the usual Prolog semantics (negation as failure). These conditions influence the state transitions in the Asbru semantics, e.g., a plan can only abort if the `abort_condition` holds. Note that arbitrary Prolog programs can be used to specify when such a condition holds.

The Asbru semantics as mentioned is implemented as a module of `PROBLINE` and could be extended with other features of the language, e.g., retry conditions and time-annotations.

4 Diabetes mellitus type 2

In the previous sections, the probabilistic framework of `PROBLINE`, including its syntax and semantics, have been discussed. In this section, we apply this framework to the management of diabetes mellitus type 2.

4.1 Management of the disease

It is well known that diabetes type 2 is a complicated disease: various metabolic control mechanisms are deranged and many different organ systems may be affected by the disorder. Patho-physiologically, there are two main phenomena, namely, insufficient secretion of the hormone insulin due to a decreased production of insulin by *B cells* in the Langerhans islets of the *pancreas*, and insulin resistance in liver, muscle and fat tissue. For the individual patient, there is a lot of uncertainty to which extent these phenomena occur, which makes it difficult to predict whether an intervention will be effective. Not only is the underlying physiology uncertain, choosing which drug to give a patient is also based on other considerations such as availability, cost, safety, tolerability, and convenience. Personalised medicine promises a path for individually optimised treatment choices, but realising this promise will require a more comprehensive characterisation of disease and drug response.

For diabetes, quite a lot of known about the effect of drugs. In this paper, we will focus on a well-known biguanide drug called metformin, which is commonly prescribed as the primary oral anti-diabetic. The dosage that we will consider is 2,000 mg/day, which is for many patients optimal [13]. The efficacy of metformin 2,000 mg/day has also been estimated [3]: the fasting plasma glucose (FPG) will be lowered up to 86 mg/dL +/- 10 mg/dL (95% confidence interval). The steady-state situation is not reached before 8 days and a linear reduction of the FPG seems appropriate [8]. Recently, Shu et al. [14] integrated diverse data supporting the hypothesis that genetic variations in the encoding of a protein called organic cation transporter 1 (OCT1) affects the response to metformin. These variants of OCT1 lead to half of the bio-availability of metformin. It was also estimated that about 20% of the Caucasian population carries one of these mutations. No cases are known with multiple mutations, nor is it common for other ethnicities.

```

plan_body(metformin, sequential, [], []).
abort_condition(metformin,S) :-
    patient_av(S, glycemia, hyper),
    patient_av(S, metformin_app, T),
    T >= 10.
complete_condition(metformin,S) :- patient_av(S, glycemia, normo).
filter_condition(metformin,_).

plan_body(treatment, sequential, [metformin], [metformin]).
abort_condition(treatment, _) :- false.
complete_condition(treatment,_).
filter_condition(treatment,_).

```

Fig. 3. Model of metformin application in Asbru. The treatment will only be aborted if the time $T \geq 10$, where the time granularity is in days.

4.2 Probabilistic model of metformin pharmacokinetics

Given the information above, we estimate a probability distribution for a variable `max_reduction` such that it is a discretised normal variable with mean 86 (mg/dL) and standard deviation 5 as 95% of the population are within 2 standard deviations (i.e., 10 mg/dL) for a normal distribution. Of course, if the raw data is available, then other distributions could be used instead. Here we are limited to the published mean and 95% confidence interval, for which we assume that a bell-shaped distribution is appropriate. Then, to describe expected probabilities of reaching normoglycemia, we have deterministic transitions:

```

choice(glycemia, hyper, normo, S, 1) :-
    patient_av(S, baselineFPG, FGL),
    patient_av(S, time_metformin_application, T),
    patient_av(S, oct1_variant, Oct),
    patient_av(S, max_reduction, Max),
    expected_normo(FGL, AdT, Max, Oct).

```

where `FPG` is a FPG at baseline, `T` is the time that metformin is applied, and `Oct` is a binary variable that is true if the patient has a variation of OCT1 affecting the efficacy of metformin. The predicate `expected_normo` computes the expected FPG based on these parameters and returns true if the expected FPG is less than 110 mg/dL, which is commonly defined as normoglycemia.

4.3 Experiments

Using PROBLINE, we can now answer a number of question surrounding the treatment with metformin. We provide a model of metformin application in Asbru in Fig. 3 where transition probabilities are set to 0.5 (see Fig. 2).

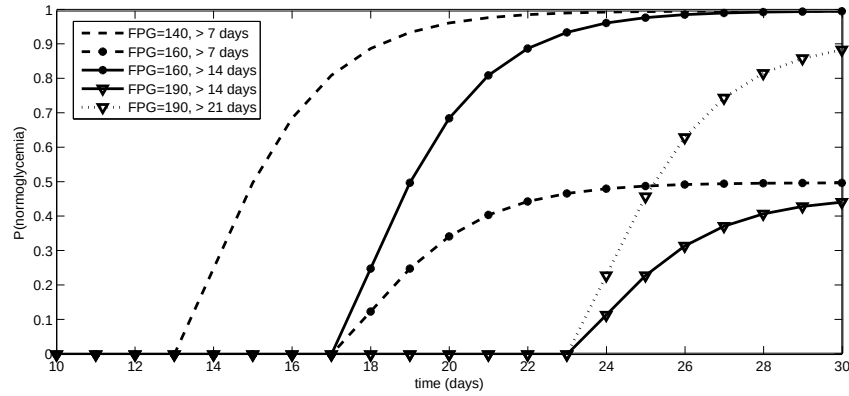


Fig. 4. Probabilistic simulation of metformin application to patients with different FPG at baseline. Time of metformin application is varied as well.

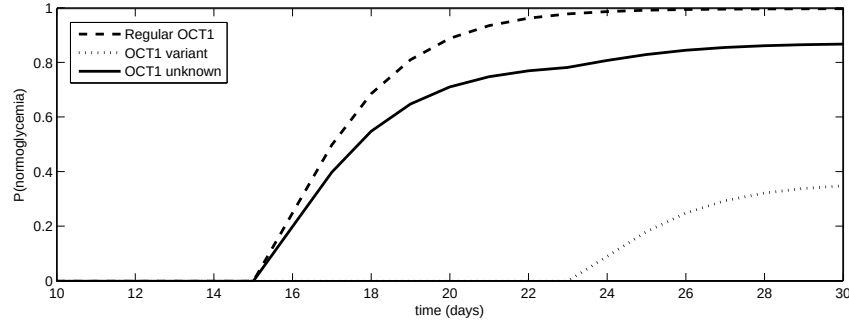


Fig. 5. Probabilistic simulation of metformin application to patients with or without a variation in the OCT1 protein.

Question 1: How long should metformin be applied before it can be decided to stop the treatment? There is a trade-off here: if the treatment with oral anti-diabetics is stopped too early then patients may be injecting themselves with insulin for no good reason; if the treatment is stopped too late, then patients who need treatment with insulin are not treated appropriately. In Fig. 4, we plot a number of dose-response curves for different patients (without OCT1 variations). For people with an initial low fasting plasma glucose, the effect of treatment is relatively quick, whereas people with an initial high fasting plasma glucose, the effect is much slower and might not be effective at all even after prolonged treatment.

Question 2: What improvement could we gain using genetic information? As mentioned, it is hypothesised that the OCT1 protein plays a key role in the

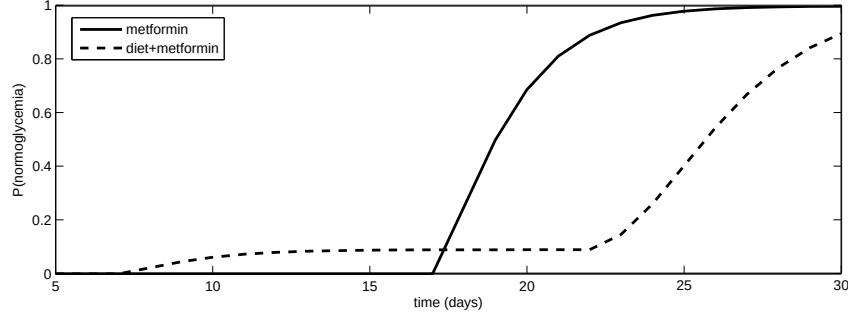


Fig. 6. Comparing metformin with and without diet in patients with normal OCT1.

efficacy of metformin. This suggests that it would be useful to test whether a patient has a variation in this gene before treatment. In Fig. 5, two patients are plotted with the same FPG at baseline but with different OCT1 proteins. On average, patients in this population (baseline FPG=150) have a good chance that metformin is effective. However, for the patients with the OCT1-variant, the chance that metformin is effective is rather small and it might be better to prescribe an alternative drug. In the end, such pharmacogenetics could be used for the personalisation of treatments [12].

Question 3: Should we try diet before metformin? In the management of diabetes, one of the first things that is recommended is a diet. However, the efficacy is low [5]. If we include this information, then we can ask PROBLINE to compare a treatment with or without a diet. Fig. 6 shows the difference between the two, where it seems that diet is of little benefit. Of course, there could be other reasons to recommend a diet, such as improved health or less side-effects of the treatment. The point is that PROBLINE allows for the exploration of different treatments taking into account uncertainty.

5 Conclusions

In this paper, we presented a new method for the analysis of care processes taking into account uncertainty. The system implementing this theory, PROBLINE, can handle typical task-network representations of guidelines and flowchart-like patient models. We presented a case-study in the treatment of diabetes mellitus type 2 illustrating the strength of this approach.

This work can be extended in several ways. In this paper, we introduced the core of the language consisting of reasoning with the dynamic aspects of guidelines, taking into account uncertainty. In future work, we would like to extend the language with other types of knowledge derived from, for example, medical ontologies and Bayesian networks. To accomplish this, a more powerful probabilistic logical language is required, e.g., using one of the recently developed logics in the field of statistical relational learning (see [4]).

References

1. M. Balser, C. Duelli, and W. Reif. Formal semantics of Asbru - an overview. In *Proceedings of the International Conference on Integrated Design and Process Technology*, Pasadena, June 2002. Society for Design and Process Science.
2. P. de Clercq, K. Kaiser, and A. Hasman. Computer-interpretable guideline formalisms. In A ten Teije, S. Miksch, and P.J.F. Lucas, editors, *Computer-based Medical Guidelines and Protocols: A Primer and Current Trends*, pages 22–43. IOS Press, 2008.
3. A.J. Garber, T.G. Duncan, A.M. Goodman, D.J. Mills, and J.L. Rohlf. Efficacy of metformin in type ii diabetes: results of a double-blind, placebo-controlled, dose-response trial. *Am J Med*, 103(6):491–507, 1997.
4. L. Getoor and B. Taskar, editors. *Introduction to Statistical Relational Learning*. Adaptive Computation and Machine Learning. MIT Press, 2007.
5. L.S. Hermann, B. Scherstén, P.O. Bitzén, T. Kjellström, F. Lindgärde, and A. Melander. Therapeutic comparison of metformin and sulfonylurea, alone and in various combinations. *Diabetes Care*, 16(10):1100–1109, 1994.
6. A.J. Hommersom. *On the Application of Formal Methods to Clinical Guidelines*. PhD thesis, University of Nijmegen, 2008.
7. A.J. Hommersom, P.C. Groot, M. Balser, and P.J.F. Lucas. Formal methods for verification of clinical practice guidelines. In A ten Teije, S. Miksch, and P.J.F. Lucas, editors, *Computer-based Medical Guidelines and Protocols: A Primer and Current Trends*, pages 63–80. IOS Press, 2008.
8. Y. Hong, S. Rohatagi, Habtemariam. B., J.R. Walker, S.L. Schwartz, and D.E. Mager. Population exposure-response modeling of metformin inpatients with type 2 diabetes mellitus. *J Clin Pharmacol*, 48(6):696–707, 2008.
9. M. Kwiatkowska, G. Norman, and D. Parker. PRISM: Probabilistic symbolic model checker. In T. Field, P. Harrison, J. Bradley, and U Harder, editors, *Proceedings TOOLS’02*, volume 2324 of *LNCIS*, pages 200–204. Springer, 2002.
10. NHS. The guidelines manual. http://www.nice.org.uk/media/68D/3C/The_guidelines_manual_2009_-_Chapter_9_Developing_and_wording_guideline_recommendations.pdf.
11. M. Peleg, S. Tu, J. Bury, P. Ciccarese, and J. Fox. Comparing computer-interpretable guideline models: a case-study approach. *Journal of the American Medical Informatics Association*, 10(1):52–68, 2003.
12. M.L. Reitman and E.E. Schadt. Pharmacogenetics of metformin response: a step in the path toward personalized medicine. *J Clin Invest*, 117(5), 2007.
13. J.H.B. Scarpello and H.C.S. Howlett. Metformin therapy and clinical uses. *Diab Vasc Dis Res*, 5(3):157–167, 2008.
14. Y. Shu, S.A. Sheardown, C. Brown, R.P. Owen, S. Zhang, R.A. Castro, A.G. Ianculescu, L. Yue, J.C. Lo, E.G. Burchard, C.M. Brett, and K.M. Giacomini. Effect of genetic variation in the organic cation transporter 1 (OCT1) on metformin action. *J Clin Invest*, 117:1422–1431, 2007.

Representing Knowledge of Multiple Disorders by Object-Oriented Bayesian Networks

Wenyun Quan, Martijn Lappenschaar, Peter J.F. Lucas, and
Arjen J. Hommersom

Radboud University Nijmegen
Institute for Computing and Information Sciences
{w.quan,mlappens,peterl,arjenh}@cs.ru.nl

Abstract. Multimorbidity in the elderly is becoming a significant health-care problem of the modern, western society. Unfortunately, medical knowledge whether contained in clinical guidelines or medical textbooks, is mostly organized around single disorders, rendering it unlikely that the elderly patient with multiple disorders receives appropriate treatment. In the research described in this paper, we use an object-oriented representation formalism to get a grip on the clinical knowledge contained in multiple clinical guidelines, with the aim of obtaining insight into the nature of the interaction between the various aspects of the conditions. As an experiment, we analyzed clinical guidelines on dementia and hypertension and developed an object-oriented probabilistic representation formalism that allowed us to represent such knowledge. We conclude that the object-oriented probabilistic representations offer a promising method for making steps ahead in this area.

Keywords: Object-oriented Bayesian networks, Multinets, Comorbidity, Disease modeling.

1 Introduction

Recent epidemiological research in the Netherlands indicates that more than two third of all patients older than 65 years have two or more chronic disorders at the same time; this is referred to as the problem of *comorbidity* or *multimorbidity*. The difference between the terms ‘comorbidity’ and ‘multimorbidity’ is that the former considers multiple disorders from the viewpoint of one, possibly most significant disorder, whereas the latter takes all disorders to be equally important. Although both terms occur in the medical literature, there is a growing preference for use of the term ‘multimorbidity’. The Netherlands is similar to other western countries in that health-care funding is under increasing pressure due to the increasing size of the population of the elderly. However, it is not only the increasing size of the problem that worries policymakers. Another, possibly more significant, problem is that medical knowledge is organized around single disorders. If a patient has multiple disorders, this knowledge may no longer be fully applicable [1]. Thus, there are no guarantees that elderly patients receive appropriate treatment for their disorders, a conclusion that is hardly acceptable

given that the health-care field is increasingly expected to take the delivery of quality care seriously. At the moment there is no easy solution to the sketched problem of multimorbidity, as clinical practice guidelines are mostly concerned with the diagnosis and treatment of single disorders.

The question that drove the research discussed in this paper was, therefore, whether the exploitation of methods from probabilistic graphical models would allow obtaining more insight into the multimorbidity problem of clinical guidelines. Probabilistic graphical models support modeling knowledge of the interaction between multiple diseases, both in terms of manifestations and treatment actions and their effects, and also support the modeling of the temporal aspects of clinical guidelines. Thus, an analysis of a clinical guideline using probabilistic graphical models could, in principle, shed light on potential interactions. In particular object-oriented Bayesian networks (OOBNs) [3] offer suitable methods to carry out such an analysis.

For the research, we selected the guidelines on dementia and hypertension, common disorders in the elderly which share certain pathophysiology. By modeling central aspects of the guidelines as a network there should be a possibility to link them together. For establishing a diagnosis a static model often suffices, whereas modeling prognosis and treatment effects require a temporal dimension.

The paper is organized as follows. In Section 2 we present reduced models for both hypertension and dementia derived from clinical guidelines, and discuss the main interactions between them. We demonstrate in Section 3 how the knowledge of multiple disorders can be modeled in the framework of object-oriented Bayesian networks, based on the work of Koller and Pfeffer [3]. Furthermore, we introduce two extensions of OOBNs, i.e., qualitative OOBNs and dynamic OOBNs, to describe simultaneous and temporal interactions of comorbidity or multimorbidity. Thus, we conclude that object-oriented probabilistic graphical representations offer a suitable means for shedding light on the problem of multimorbidity. Further conclusions and future work are discussed in Section 4.

2 Medical Background

2.1 Clinical Guidelines

Clinical guidelines provide recommendations to physicians and patients, mostly about the management of one disease, and are developed by basing recommendations as much as possible on scientific evidence. European guidelines on hypertension [4] and dementia [7] were analyzed with the aim of constructing causal models that could be used as a start for detecting interactions.

The diagnostic processes are supported by indicating relevant symptoms, risk factors, signs obtained by physical examination and laboratory investigations to decrease the uncertainty in the diagnosis. Guidelines on hypertension now emphasize that the diagnosis and management of hypertension should be related to the quantification of total cardiovascular risk. The co-presence of a metabolic syndrome worsens the prognosis and common causalities of hypertension are

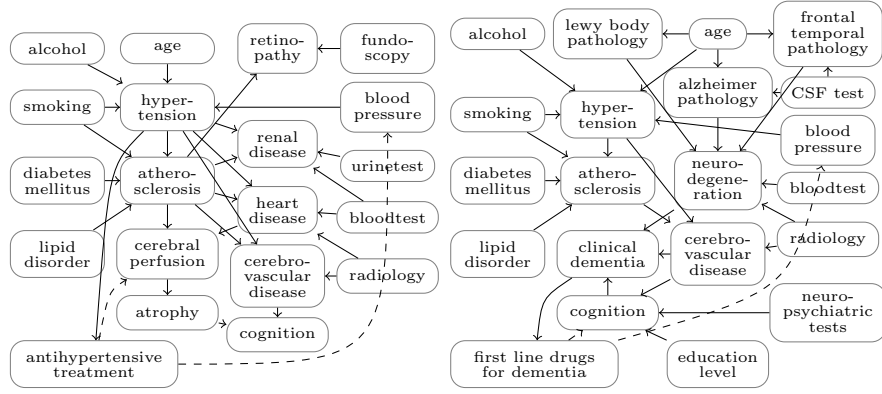


Fig. 1. Simplified graphical models of hypertension (left) and dementia (right), extracted from the respective clinical guidelines.

cerebral vascular disease, heart disease, renal disease and retinopathy. In the absence of specific neuro-pathological confirmation, the etiological diagnosis of a dementia syndrome can only be made in terms of probability. The diagnosis can be supported by medical history and comorbidity, neuro-physical examination, assessment of neuro-psychological functioning, blood tests, neuro-imaging, and cerebrospinal fluid analysis. Each of these so called biomarkers relates with a certain probability to a stage of the disease [2] and helps differentiating between the different types of dementia. Treatments are supported by providing advice through a stepwise approach with regular control of the patient's symptoms and signs, until a proper balance in the patient's condition is achieved.

We represent this knowledge in simplified graphical models due to space consideration (see Figure 1). Most of vertices are derived from guidelines, but some of the pathophysiological processes (atherosclerosis and neurodegeneration) and some interactions (therapeutic side effects) are derived from other medical knowledge such as scientific medical journals. To represent an associations between vertices, we draw an arc from the vertex of causality to the vertex of effect. The arcs in Figure 1 represent causalities between patient characteristics, pathophysiology and diagnostic processes, as where the dotted arcs represent effects of treatments.

2.2 The Complexity of Modeling Multiple Disorders

In case of the elderly, the physician often deals with multimorbidity, and then several guidelines need to be applied. Eventually this can lead to conflicts, e.g. treatment of one disease can reduce the efficacy of a treatment of another disease or worsen its outcome. In particular, there can be interactions between drugs. Treatment should focus on a common pathways rather than administering drugs for symptom relief of the diseases separately, since many side effects of drug treatment can blur out the total picture.

Interactive variable(s)	Type of action	Effect on
Hypertension	influence	Dementia
Antihypertensive treatment	influence	Brain Perfusion
Metabolic syndrome + hypertension	synergy	Dementia
Cholinesterase inhibitors + antihypertensive drugs	synergy	Blood pressure
NMDA-antagonist + antihypertensive drugs	synergy	Blood pressure

Table 1. Interactions between hypertension and dementia.

Three major problems occur when modeling multiple disorders. Firstly, some symptoms, signs, risk factors, pathophysiology, and so on, may be present in more than one disorder. For example, hypertension and dementia share common risk factors such as alcohol use, smoking, diabetes mellitus and lipid disorders. Secondly, the influence of pathophysiology changes over time. For example, hypertension is probably a midlife risk factor for dementia, initiating or advancing the onset in late life, whereas it is probably a protective factor by maintaining a high perfusion of the brain [6]. Finally, there can be interactions between different models, e.g. drugs for treating dementia have an effect on the blood pressure and may conflict with antihypertensive treatment.

All together there is a need for a model-based approach that can deal with complex structures, time evolution and probabilistic interactions. Therefore, we employ qualitative and dynamic OOBNs as possible solutions for the above-mentioned problems. OOBNs allow one modularize the representation of relevant medical knowledge. Interactions between variables can be represented using qualitative probabilistic networks, QPNs for short [8]. Interactions are cause-effect influences or synergies. An influence exists when the presence of a certain value of a variable makes the presence of a certain value of another variable more or less likely. In case of a synergy the presence of certain values of two variables makes a certain variable of a third variable more or less likely. Table 1 shows some possible influences and synergies between hypertension and dementia.

3 Dynamic Qualitative Object-Oriented Bayesian Networks

In this section, the basics of employing object orientation for building Bayesian network models are briefly described and illustrated by means of examples from dementia and hypertension.

3.1 Clinical Bayesian Networks

Bayesian networks are looked upon as an effective framework for knowledge representation and reasoning under uncertainty [5]. Formally, a *Bayesian network*, or BN, is a tuple $\mathcal{B} = (G, X, P)$, where

- $G = (V, E)$ is a directed acyclic graph (DAG) with a set of vertices V and an edge relation $E \subseteq V \times V$;

- X denotes a set of variables, where each vertex $v \in V$ corresponds one-to-one to a variable X_v in X , with the variable taking on values from its domain;
- The joint probability distribution P of X is defined such that all independences that can be derived from the DAG G using *d-separation* are obeyed by P [5].

A *clinical* Bayesian network represents the probabilistic relationships between diseases, symptoms, signs and interventions, such as tests and treatments. The formalism quite naturally allows for representing the uncertainties involved in these relationships. To develop methods for identifying disease interactions, and possibly conflicting clinical interventions, from combining clinical guideline BN translations, we need to be able to handle very complex models. Object-oriented Bayesian networks [3], OOBNs for short, allow us to easily construct large and complex models by exploiting ideas from object orientation in the construction process.

3.2 Object-Oriented Bayesian Networks

In order to model a probabilistic network for multiple diseases by combining the corresponding Bayesian networks, we need to study the relations between the networks. Often we can find vertices representing same variables in different probabilistic graphs for disorders. Those variables describe normally the same object. Basically, OOBNs allow us to define probabilistic models over multiple objects. Therefore we introduce an approach based on the theory of OOBNs from Koller and Pfeffer [3] in this section.

Object. An *object* in OOBNs is any entity composed of *input attributes* A_I and *value attributes* A_V , where input attributes are parameters to the object. Value attributes are divided into *encapsulated attributes* A_E and *output attributes* A_O , which are part of the specification of the object itself. For instance, the attributes Age, Gender and Lifestyle of a patient could act as input attributes for a value attribute Disease, while the encapsulated attribute Pathophysiology together with the output attributes Symptom and Sign describe Disease itself. The encapsulated attributes are d-separated from the rest of the network by the inputs and outputs of the object. There are two types of objects. A *simple object* is composed of a set of input attributes A_I and a single output attribute a_O of basic types, e.g., Boolean, Integer, etc. A *complex object* is composed of input attributes A_I of basic or structured types, encapsulated attributes A_E and output attributes A_O which are themselves objects.

Object-oriented Bayesian network fragment. In order to model uncertainty about the possible values of an object, the object is associated with a probabilistic model. A *stochastic function* can be associated with an object and defines a distribution of the possible values of this object for each assignment of values to the object's inputs. In general, complex models often involve many similar objects or attributes of objects, whose stochastic functions are essentially

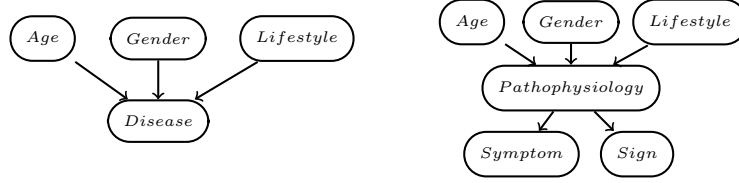


Fig. 2. A DAG of OONF Patient (left) and a DAG of OONF Disease (right).

identical. Thus, we use the notation of a class in order to use generic stochastic functions multiple times in defining similar objects.

A stochastic function is described by using recursive composition of *object-oriented Bayesian network fragments* (OONFs). An OONF of a set of input attributes A_I and a set of value attributes $A_V = A_E \cup A_O$ is a stochastic function over a DAG $G = (V, E)$ where the set of vertices V corresponds one-to-one to each element in the set of the attributes of OONF. For each value attribute $a_V \in A_V$, every parent of a_V is annotated with at least one input $a_I \in A_I$. Let $a_O(\mathcal{O})$ denote an output attribute of object $\mathcal{O}$ and $\mathcal{O}.p$ denote an attribute chain of $\mathcal{O}$, each input attribute $a_I \in A_I$ has the same type with the output attribute $a_O(\pi(a_I).p)$, where $\pi(a_I)$ denotes the set of parents of a_I in the DAG. An OONF over a simple object is a *simple OONF*.

To establish a probabilistic model representing uncertain knowledge of multimorbidity, we organize objects extracted from clinical guidelines for certain diseases. Patient and Medical Facility are considered as common objects in the clinical guidelines of disorders. A class Patient defines complex objects, which contain attributes Name, Address and Age etc. Figure 2 shows DAGs of the OONF for Patient and for Disease, where Disease -a complex object itself- is the value attribute of Patient. Therefore an OONF of Disease is defined inside the OONF of Patient with inputs Age, Gender, and Lifestyle. Since many common attributes of Patient, e.g., Name, Address, have no direct influence on diseases, they are not considered as inputs of the attribute Disease.

We can define the classes of common objects in the clinical guidelines with an object-oriented language, which together with the corresponding stochastic function define OONFs. As a subclass of Person, for instance, class Patient inherits its attributes from its superclass Person. Diseases are defined by each individual class which contains an encapsulated attribute Pathophysiology and output attributes Symptom and Sign. The relevant attributes for class Medical Facility and their relations are displayed in Figure 3.

Object-oriented Bayesian network. Up to this point we showed how to define a class for each individual object. The relations and interactions among objects are represented by an abstract object, or a *situation object* as it is called. An *object-oriented Bayesian network* (OONBN) is a single situation object without input, whose probabilistic properties are defined by an associated OONF.

An OOBN describing medical interventions consists of classes Patient, Medical Facility and the situation object Medical Treatment with an associated OONF. Figure 3 shows the constructed OOBN model for medical treatment. The *dotted lines* represent relations between attributes inside a class, and the *arcs* describe the probability dependency of the associated OONF between classes.

Inference in object-oriented Bayesian networks. The key to effective inference in OOBNs is the observation that complex objects are only used to group simple objects together into coherent units. A Bayesian network $\mathcal{B}$ of an OOBN contains a set of attributes for each simple object defined by the OOBN. The structure of a *multiply sectioned Bayesian network* (MSBN) [9] is a multiply sectioned DAG (MSDAG) with a hypertree organization, where each set of shared variables between subsets are d-sepset. Given such a BN $\mathcal{B}$, we can construct an MSBN which contains subnets for each object in the OOBN. Koller and Pfeffer [3] proved that a Bayesian network $\mathcal{B}$ constructed from an OOBN is an MSBN of hypertree structures. In a hypertree MSBN, all the message can be passed in the probabilistic inference process along the paths of the tree.

3.3 Representing Interactions of Multiple Disorders

Based on the clinical guidelines, the framework of OOBNs for representation and reasoning comorbidity or multimorbidity under uncertainty is introduced in the previous section. With summarizing the work of Koller [3] we demonstrated the implementation of their theory in our study case and yielded an abstract model to link the Bayesian networks of hypertension and dementia. Nevertheless, the (both simultaneous and temporal) interactions between multiple diseases are not displayed intuitively in the OOBN model. Therefore we extend our model in the framework of qualitative OOBNs with the theory of *qualitative probabilistic network* (QPN) [8] to represent simultaneous interactions in an OOBN, and in the framework of dynamic OOBNs to represent temporal interactions of diseases between OOBNs.

Class definitions of Person and Patient:

Class Person is	Class Patient is
Superclass nil;	Superclass Person;
Name: String;	Physiology:{
Address: String;	Medical Record:{diabetes mellitus,
Age: Integer;	dyslipidemia,etc};
Gender:{male,female};	Family History:{hypertension,
Education:{high,average,low};	dementia,etc};};
Lifestyle:{smoker,alcoholic,etc};	Disease: {Class Hypertension;
End	Class Dementia; etc};
	End

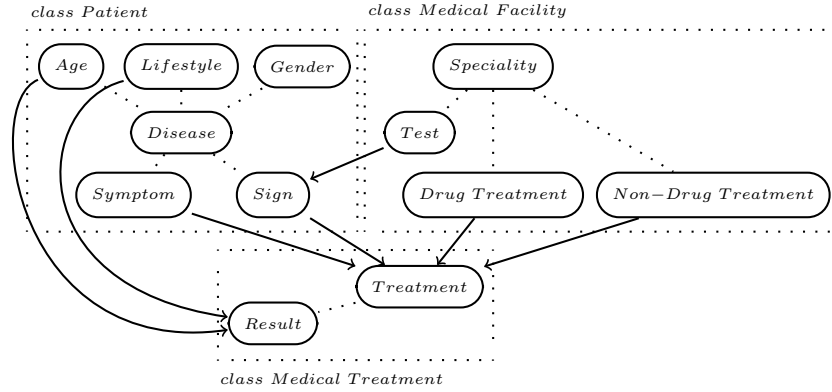


Fig. 3. An OOBN model of medical treatment.

Qualitative object-oriented Bayesian network. A *qualitative OOBN* is an extension of OOBNs by adding qualitative influences and synergies between objects and attributes from different probabilistic models. A *qualitative influence* expresses how the value of one variable influences the probability of observing values of another variable. A *additive synergy* expresses how the interaction between two variables influences the probability of observing the values of the third variable. A *product synergy* expresses how upon observation of a common child of two vertices, observing the value of one parent influences the probability of observing a value of the other parent. For instance, both antihypertensive treatment and first line drugs for dementia (cholinesterase inhibitors) negatively influences the blood pressure (by decreasing it), which is an important determinant in both disease models. A *negative additive synergy* expresses how the interaction between the treatment for hypertension and for dementia influences the probability of observing the values of blood pressure. Using both treatments together, the decreasing effects on the blood pressure add up, making the chance of a low blood pressure (and so hypotension) much more likely, and thus form a negative additive synergy on a high blood pressure. Figure 4 shows the negative additive synergy between these two treatments.

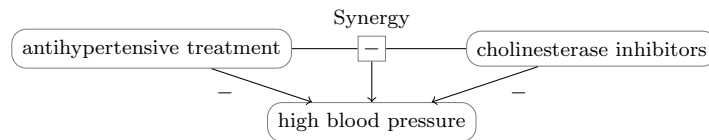


Fig. 4. Negative additive synergy between antihypertensive treatment and first line drugs for dementia.

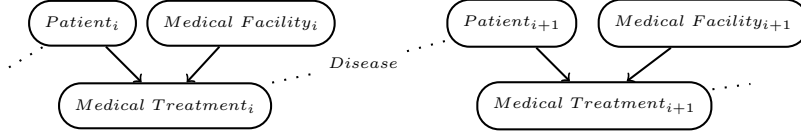


Fig. 5. A fragment of dynamic OOBN model with situation object Medical Treatment.

Dynamic object-oriented Bayesian network model. Since both hypertension and its treatment influence the onset and cause of dementia in the future, the consideration of time is therefore mandatory, i.e., it is not sufficient to model a static OOBN to deal with the problem of multimorbidity. For an elderly patient with hypertension for instance, often the cause of hypertension and the medical treatment for this disease can affect the probability of dementia in the future. In order to work with domains that evolve with time, we need a model for each unit of time. Dynamic models consider the cases that an event can cause another event in the future. A *time slice* is a local model, i.e., an OOBN in our representation formalism. The time slices are connected through temporal links to create a full model, i.e., a dynamic OOBN. A *dynamic OOBN* model is a repetitive temporal OOBN model, if the structures of the time slices together with conditional probabilities are identical and the temporal links are the same.

We illustrate this approach by discussing a dynamic OOBN model of medical interventions for dementia and hypertension. The OOBN we defined in the previous section can be considered as a local model in the dynamic OOBN. A time unit is determined by the period of treatment rather than a point in time. The dynamic link between the local models contains the complex attribute Disease. The values of Disease of treatment's result are the output values of the local model at current time, and are considered as the values of the Disease of Patient at the begin of the next time unit. The temporal links in Figure 5 are represented by dotted lines between local models. In a dynamic OOBN for hypertension and dementia, we can see the conflicts of medical interventions, e.g., antihypertensive treatment at time unit i changes the risk of dementia at time unit $i+1$. Vice versa, drugs for dementia alter antihypertensive treatment. Thus, conflicting medical interventions of multiple disorders can be represented in the dynamic model, and allow medical doctors to detect the future consequences of the current treatments in advance.

4 Discussion

The only way to make progress in the management of multiple disorders in elderly patients to guarantee the quality of the delivered care is, to offer means to pinpoint interactions between the various disorders and their management. Although clinical guidelines are mostly organized around single disorders, they can still be used as a starting point for such identification.

In this paper we provided a framework for modeling the essential disease knowledge from the guidelines using OOBNs. We showed that the models representing knowledge of clinical guidelines can be effectively linked together using OOBNs, yielding a tool for determining shared risks and pathophysiology. In addition, by adding qualitative influences and synergies the simultaneous interactions of multiple disorders can be represented in qualitative OOBN models. And by exploiting the temporal dimension offered by dynamic OOBNs, which allows monitoring disease progression and effects of therapeutic choices, it is possible to capture interactions in both from a temporal point of view.

Nevertheless, we did not provide the actual probabilistic parameters of the constructed qualitative and dynamic OOBNs. On one hand future work will focus on quantification of probabilities by adding known probabilistic evidence from scientific research and learning probabilities from patient datasets. Eventually, interactions between diseases will inherit probabilistic evidence from several pathways changing probabilities downwards to prognosis and therapeutic choices. On the other hand we will extend the formal definitions of qualitative influences and synergies in the framework of object-orientation, with a temporal dimension, to provide a formalism for modeling interactions of multiple disorders in terms of time. Following the identification of such interactions, disease management can be adjusted where probabilities deviate significant comparing to a single guideline.

References

1. Boyd, C.M. et al.: Clinical Practice Guidelines and Quality of Care for Older Patients, With Multiple Comorbid Diseases: Implications for Pay for Performance, *JAMA*, 294(6):716–724, 2005.
2. Jack, C., Knopman, D., Jagust, W., Shaw, L., Aisen, P., Weiner, M., Petersen, R., Trojanowski, Q.: Hypothetical Model of Dynamic Biomarkers of the Alzheimer’s Pathological Cascade, *Lancet Neurology*, 9:119-128 (2010)
3. Koller, D., Pfeffer, A.: Object-Oriented Bayesian Networks, In *Proceedings of Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, pages 302-313, Providence, Rhode Island, August 1-3 (1997)
4. Mancia, G., De Backer, G., Dominiczak, A., et al.: 2007 Guidelines for the management of arterial hypertension. The Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC), *European Heart Journal*, 28:1462-1536 (2007)
5. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann (1988)
6. Qiu, C., Winblad, B., Fratiglioni, L.: The Age-dependent Relation of Blood Pressure to Cognitive Function and Dementia, *Lancet Neurology*, 4:487-499 (2005)
7. Waldemar, G., Dubois, B., Emre, et al.: Recommendations for the diagnosis and management of Alzheimer’s disease and other disorders associated with dementia: EFNS guideline, *European Journal of Neurology*, 14:e1-e26 (2007)
8. Wellman, M.: *Fundamental Concepts of Qualitative Probabilistic Networks*, Elsevier Science Publishers B.V., *Artificial Intelligence*, 44:257-303 (1990)
9. Xiang, Y.: *Probabilistic Reasoning in Multiagent Systems: A Graphical Models Approach*, Cambridge University Press (2002)

Temporal Relations in Causal Independence

Maarten van der Heijden^{1,2} and Peter J.F. Lucas¹

¹ Institute for Computing and Information Sciences,
Radboud University Nijmegen, The Netherlands

² Department of Primary and Community Care,
Radboud University Nijmegen Medical Centre, The Netherlands

Abstract. Allen’s temporal interval algebra has been successfully used to model the temporal evolution of events in an abstract fashion. However, processes that can be described in this way often have associated uncertainty that cannot be expressed in the algebra. Temporal events can be seen as independent random variables that become dependent once we know their effect. This way of modelling is supported by the theory of causal independence, which is linked to the convolution theorem of probability theory. The convolution theorem allows us to relate the probability of the effect with the probability of the causes, constrained by Allen’s relations. We thus obtain a natural way to add uncertainty to Allen’s temporal interval algebra. Uncertain temporal relations are particularly important in the biomedical area where the heterogeneous nature of the data makes it necessary to exploit qualitative information for uncertainty reasoning.

1 Introduction

Modelling temporal evolution of processes is an important aspect of many real world applications of artificial intelligence. However, in many problem domains it is difficult to obtain detailed information about when particular events took place. For such situations, Allen’s interval algebra [1, 2] has shown to be a valuable tool, as it offers an algebra to express qualitative, rather than quantitative, relations between time intervals. A typical domain where one normally resorts to the use of qualitative temporal information is clinical medicine. Although for some data actual time stamps are known, more often available data is heterogeneous in nature, e.g. lab data collected at a particular time are recorded at the precision of hours and minutes, whereas for the symptoms and signs of a patient’s disorder it is only known on what day they appeared. It is therefore not surprising that clinical guidelines, which are used by health-care professionals to ensure quality of care, often provide only qualitative temporal information about when to do what.

Although qualitative temporal information as expressed by Allen’s algebra relieves the necessity to be precise about the time when events took place, whether events actually occur has to be known for certain. Unfortunately, the temporal evolution of a disease process, whether treated or not, is uncertain by its nature:

to adequately model disease processes it is desirable to be able to express this uncertainty, which lies outside the scope of Allen's algebra.

Bayesian networks [3] have become a popular type of model for representing and reasoning with uncertain knowledge. Dynamic Bayesian networks [4], of which HMMs are a special case, are typically used to model uncertain temporal progression, but this type of model requires temporal data beyond what is typically available in a clinical environment due to the heterogeneous nature of the data. Furthermore, in clinical medicine many variables are unobservable which complicates finding reliable probabilistic parameters. However it is often possible to accurately determine qualitative temporal relations between variables.

To illustrate the above, we will consider an aspect of chronic obstructive pulmonary disease, COPD for short, a chronic disorder that affects the airways and lung. Patients often suffer from *exacerbations*, i.e. an acute deterioration of respiratory health status, which is usually caused by an airway infection. Other possible causes are exposure to (cigarette) smoke or other irritant substances and small particle matter, usually a result of air pollution. The occurrence of these causes is generally independent, but some temporal relation may exist. In most cases a patient will first contract an airway infection (which not necessarily immediately causes an exacerbation) and may in a later stage suffer an exacerbation when exposed to air pollution (which in itself would not necessarily result in an exacerbation). Although the reverse order is possible, it is less likely because the effects of irritant substances are more transient than airway infections. This situation can be modelled by a causal independence model that takes into account the qualitative temporal information that infections usually precede other causes of exacerbations. Similar situations arise often in medicine, where partial information on ordering or temporal overlap may be available, but not exact times of onset.

In this paper we investigate how we can integrate these qualitative temporal relations with causal independence models to obtain a probabilistic temporal model as part of a Bayesian network. Concretely this means that we can construct constraints on multiple causes in causal independence models. Also a probability distribution over Allen's relations between causes can be specified to calculate the probability of the effect. These models are useful in a medical context where causal independence models are often employed and reasoning about disease progress, i.e. temporal reasoning, is a key element of disease management.

This paper is structured as follows, in the next section we discuss related work, followed by a section with preliminaries which are used in Section 4 to develop causal independence models with temporal relations. Section 5 contains some conclusions.

2 Related work

The use of causal principles to model temporal events has been studied before by other researchers. However, the approach taken by others is different. Here a brief summary is given of this related research.

One of the main results explored in this paper is the convolution theorem from probability theory [5]. In this paper, the basic principles behind this theorem are linked to causal independence modelling, yielding a novel method to represent uncertain, qualitative temporal information. However, probabilistic convolution in a different form has been used before by Tawfik and Neufeld to represent the probability distribution over the sum of two random variables [6]. Other related work is by Galán et al. [7], who applied convolution to compute the probability of a random variable taking a value at a given instant given the probability distribution of a cause occurring and the distribution over the delayed effect. The result is then used as a parameter for a noisy-OR. Our work differs from this since we consider the convolution over the complete time line, integrating the casual independence model with temporal constraints. Ryabov and Trudel [8] have studied probabilistic interval relations in the context of temporal reasoning formulated as constraint satisfaction problems. What we investigate in this paper is an embedding of Allen’s algebra in the framework of probabilistic reasoning using Bayesian networks. Temporal reasoning with uncertainty is also important in planning [9], but our work has a different focus as we use temporal reasoning to construct prognostic models.

3 Preliminaries

This section reviews the necessary building blocks for the theory that is developed in the next section, that is, Allen’s temporal interval algebra, causal independence models and the convolution theorem.

3.1 Temporal interval relations

Allen’s interval algebra offers a well-known, and frequently used, representation of qualitative temporal information [1, 2]. It consists of a set of relations that specify temporal associations between pairs of time intervals. There are 13 relations that are mutually exclusive and complete with respect to the basic qualitative relations between intervals; the following relations are distinguished: *before*, *meets*, *overlaps*, *starts*, *during*, *finishes*, *equals* and their inverses (where *equals* is its own inverse). In Figure 1 the relations are shown graphically.

3.2 Causal independence

In Bayesian networks [3], modelling the probabilistic influence of a set of causes $C_1, \dots, C_n$ on a common effect R by

$$P(R \mid C_1, \dots, C_n)$$

is hard because the number of parameters increases exponentially with the number of causes. To model this situation in a tractable way, as is illustrated in Figure 2, one can add intermediary vertices I_k between each cause vertex C_k

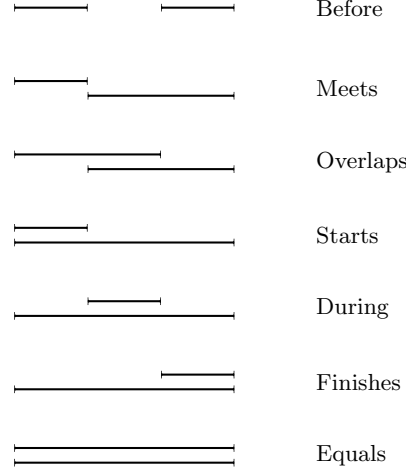


Fig. 1. Graphical representation of the seven basic relations that can hold between two time intervals. These relations and their inverses make up Allen's algebra.

and effect vertex R , and replace the problem of assessing the family of conditional probability distributions $P(R \mid C_1, \dots, C_n)$ by the potentially simpler problem of assessing a linear number of probabilistic parameters $P(I_k \mid C_k)$ and a function b which corresponds to the family of deterministic probability distributions $P(R \mid I_1, \dots, I_n)$. As the function b is again exponential in the number of arguments, the method does not automatically save space. However, if the function b can be decomposed into simpler functions, such as repeated binary functions $b'(I, I')$, then the method may yield a substantial saving in terms of the space required. This method is known as causal independence [10].

The most well known causal independence model of this kind is the noisy-OR, which uses a logical disjunction as the deterministic interaction function.

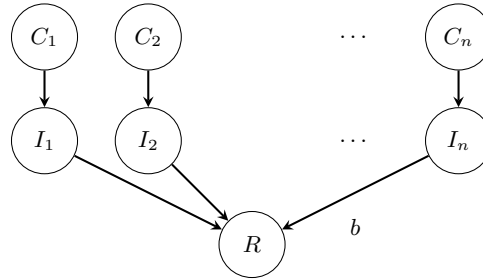


Fig. 2. Causal independence model. R results from the application of the Boolean function b on the output of the intermediate variables.

The resulting model should be interpreted as modelling the situation that each cause has some probability of occurring and the effect is produced if at least one of the causes is true.

In this paper, we will use causal independence in order to model the interaction between events that have temporal relations to each other. For example, we may have that I_1, I_2 , and I_3 represent intermediate events that through their causes C_1, C_2 , and C_3 yield the common effect R , where event I_1 happens before I_2 , which, in turn, happens before I_3 .

3.3 Convolution in probability theory

As in causal independence modelling we use a function b to model the interaction between random variables $I_1, \dots, I_n$, it is worth investigating probability distributions defined on functions of random variables, i.e. probability distributions of the form $P(b(I_1, \dots, I_n) = r)$. One well-known example of such a probability distribution is the one where a variable Z is defined as the sum of two other variables X and Y , i.e. $X + Y = Z$. In that case one can prove that the following holds [5]: $P(X + Y = z) = f_{X+Y}(z) = \sum_x f_X(x) f_Y(z - x)$ where f is the joint probability mass function of X and Y . Furthermore, if X and Y are independent, this joint probability can be written as

$$P(X + Y = z) = f_{X+Y}(z) = \sum_x f_X(x) f_Y(z - x)$$

which is known as the *convolution* of f_X and f_Y . The link to causal independence models is interesting. First, one should realise that

$$P_b(r \mid C_1, \dots, C_n) = P(b(I_1, \dots, I_n) = r \mid C_1, \dots, C_n)$$

where P_b means that probability distribution P is defined using the interaction function b using causal independence.

Theorem 1 *Let b be a Boolean function defined on independent random variables I and J then it holds that:*

$$P(b(I, J) = r \mid C) = \sum_i P(i \mid C_I) P(b(i, J) = r \mid C_J)$$

Proof. The (I, J) space determined by $b(I, J) = r$ can be decomposed as: $\bigcup_i \{I = i\} \cap \{J = j \mid b(i, j) = r\}$ where the expression $b(i, j) = r$ is a logical constraint on the Boolean values of the variable J . The sets $\{I = i\} \cap \{J = j \mid b(i, j) = r\}$ are mutually exclusive for each i . Hence,

$$\begin{aligned} P(b(I, J) = r \mid C) &= P\left(\bigcup_i \{I = i\} \cap \{J = j \mid b(i, j) = r\} \mid C\right) \\ &= \sum_i P(I = i, b(i, J) = r \mid C) \\ &= \sum_i P(I = i \mid C_I) P(b(i, J) = r \mid C_J) \end{aligned} \quad \square$$

In this paper we are interested especially in temporal causal independence models, which means that we compute the convolution over time. Given that convolution is an operation that calculates a sort of weighted average, it seems natural to allow taking the weighted average over only the time interval of interest, expressed by means of constraints. Conceptually this is what we develop in the next section. We use the fact that Allen's relations are Boolean functions, with which we can write a causal independence model as a convolution.

4 Causal independence and convolution with temporal relations

This section describes our main results, starting from a simple case and then expanding on the ideas developed. Given that we can use convolution to describe the probability distribution of a causal independence model, we investigate in which way we can incorporate the temporal relations from Allen's algebra. In the following, we assume time is discrete, which allows us to abstract away from some temporal imprecision (e.g. exact timing of patient complaints) by modelling it as a time point in an appropriate granularity.

4.1 Simple time model

We first look at a simple model of time, which limits Allen's algebra. Consider a discrete time model where we have a finite set of ordered time intervals $T = (\{i_1, \dots, i_n\}, <)$, where $i < i'$, if ' i before i' ' holds. We then define a probability distribution over this set representing the probability that result r was caused by some cause C in interval i . Initially we will ignore the fact that intervals have duration, and treat them as if they are indivisible. Now consider the case of two possible causes C_1 and C_2 , which have effect in intervals $i_1, i_2 \in T$, and a temporal Allen relation a that holds between the intervals. We can write the probability of the result r as the convolution:

$$P_a(r | C) = \sum_{i_1} P(i_1 | C_1) \sum_{i_2} P(a(i_1, i_2) = r | C_2)$$

where $a(i_1, i_2) = r$ means that r is caused by i_1 and i_2 subject to temporal relation a . Now, using F to denote distribution functions and f for probability mass or density functions, we can for example when F is the geometric distribution, a is *before*, C_1 is *infection* and C_2 is *smoke*, write the convolution for the probability of an exacerbation (the result r) as:

$$\begin{aligned} P_{before}(r | C) &= \sum_{i_1} f_{I_1}(i_1 | C_1) \sum_{i_2} f_{I_2}(before(i_1, i_2) = r | C_2) \\ &= \sum_{i_1} (1 - p_1)^{i_1} p_1 \sum_{i_2 > i_1} (1 - p_2)^{i_2} p_2 \end{aligned}$$

As a consequence of this choice of time model we have to restrict ourselves to a subset of Allen's algebra, since we cannot express any relation that involves overlap between the intervals. The convolution gives a probability distribution over times on which the result r can occur, constrained by the interval I_1 being either equal to, meeting with or before I_2 , or the converse relations for $a(I_2, I_1)$ (' i meets j ' if $i < j$ and there is no k such that $i < k < j$). The other relations, i.e. *overlaps*, *starts*, *during*, *finishes*, require taking into account the overlap between intervals which we cannot do with this model of time, as it violates the assumption of a total order on the time points (e.g. if i overlaps j , the intervals are incomparable in the ordering).

4.2 Causal effects within intervals

Instead of treating intervals as indivisible we can model the time intervals with explicit start and end points, that is an interval is described as the set of integer time points in $[s, e)$. To assign probabilities to the intervals, we define a distribution that depends on the length of the intervals, e.g. for discrete durations we can use a binomial distribution. In the causal independence model I is a random variable with durations as values, i.e. the value for interval $[s, e)$ is the length $e - s$. The assumption underlying this interpretation is that when the cause C is true the probability of C causing r at any single time point is p and the event occurs at most once in the interval. The probability of the result for different temporal relations then follows from the convolution, as only those intervals that satisfy the relation contribute to the probability of r . As an example, a time line with three time points t_1, t_2 , and t_3 has intervals $[t_1, t_2)$, $[t_2, t_3)$ and $[t_1, t_3)$, for the temporal relation *meets*, only $i_1 = [t_1, t_2)$ and $i_2 = [t_2, t_3)$ satisfy the temporal constraint and would thus give a nonzero probability in the convolution.

We can now formulate expressions for the probability of the effect r , given some Allen's relation between the causes. Take for example a binomial distribution over the interval length, then if $i_1 = [s_1, e_1)$ overlaps $i_2 = [s_2, e_2)$ and $L = e_1 - s_1$ we obtain:

$$\begin{aligned} P_{\text{overlaps}}(r \mid C) &= \sum_{i_1} f_{I_1}(i_1 \mid C_1) \sum_{i_2} f_{I_2}((s_2 > s_1 \wedge e_2 > e_1 \wedge s_2 < e_1) = r \mid C_2) \\ &= \sum_{i_1} \binom{L}{1} p_1 (1 - p_1)^{L-1} \sum_{\delta, \epsilon} \binom{L + \epsilon - \delta}{1} p_2 (1 - p_2)^{L + \epsilon - \delta - 1} \end{aligned}$$

with $0 < \delta < (e_1 - s_1)$ and $\epsilon > 0$. The probability of the result at some time point depends on the length of the first interval i_1 and all intervals i_2 that satisfy the temporal constraint. Interval i_2 is written in terms of i_1 and the translation parameters δ and ϵ , where δ shifts the start point of the interval and ϵ the end point. In this particular example, in order to model overlap between i_1 and i_2 , the start point of interval i_2 has to be later than the start of i_1 but earlier than the end of i_1 , which means that δ cannot be larger than the duration of i_1 . The

end of i_2 has to be later than the end of i_1 (otherwise i_2 would be during i_1), which means that ϵ has to be greater than 0, but is not bounded from above.

It turns out that by choosing different values for δ and ϵ we can construct the temporal relations. Specifically, for each temporal relation we can write simple constraints on δ and ϵ to obtain the convolution given Allen's relation a . These constraints are shown in Table 1. A further consequence is that given data on the possible causes, we can learn the parameter values for some domain, thereby providing some quantification of the qualitative Allen's algebra (the degree of overlap between intervals can be quantified by δ and ϵ).

δ	ϵ	$a(i_1, i_2)$
0	0	equals
0	< 0	started-by
> 0	0	finished-by
< 0	> 0	during
$e_1 - s_1$	> 0	meets
$> e_1 - s_1$	$> \delta - e_1 + s_1$	before
$> 0; < e_1 - s_1$	> 0	overlaps

Table 1. Constraints on the parameters δ – shifting the interval start point – and ϵ – shifting the end point. Given these constraints we obtain the convolution for the Allen's relations in column three.

For more than two possible causes we can write a recursive convolution:

$$P_a(r | C) = \sum_{i_1, \dots, i_n} P(i_1 | C_1) P(a(i_1, i_2) = r | C_2) \cdots \\ P(a(i_{n-2}, i_{n-1}) = r | C_{n-1}) P(a(i_{n-1}, i_n) = r | C_n)$$

The probability of the result depends on all possible combinations of intervals that satisfy the constraints of the interval relations. This enables us to construct causal independence models with more than two causes. Note that intervals that do not satisfy the constraints have probability zero. Propagating constraints to find the relations between e.g. i_1 and i_3 is not necessary as intervals with nonzero probability that satisfy $a(i_1, i_2)$ and $a(i_2, i_3)$ will also satisfy $a(i_1, i_3)$.

If the exact temporal relation between two intervals is not known, the uncertainty can be expressed as a disjunction of possible relations. As both the Allen's relations and the disjunction are binary relations, the disjunction can be written as a recursive convolution. However, a disjunction is a relatively imprecise way of specifying uncertainty and we may instead express uncertainty on the temporal relation with a probability distribution. For example we could ask a physician to rate the likelihood of possible orders of causes for exacerbations, which could result in for instance a high probability that an exacerbation is caused by an interval of infection during which the patient is exposed to smoke. Formally, we

obtain a probability weighted convolution over the temporal relations:

$$P(r \mid C) = \sum_a P(a) \left(\sum_{i_1} P(i_1 \mid C_1) \sum_{i_2} P(a(i_1, i_2) = r \mid C_2) \right)$$

Thus we calculate the convolution for each temporal relation a which has nonzero probability, if the relation is known the probability of that specific relation a is 1 and we regain the usual convolution.

4.3 Example: COPD exacerbations

Recall the example from the introduction where we presented the case of exacerbations of chronic obstructive pulmonary disease. Since exacerbations have great impact on patients, it is clinically relevant to try to prevent exacerbations. Thus, we are interested in predicting exacerbations using probabilistic models of possible causes. In the introduction we identified three possible causes of exacerbations, that is airway infection, (cigarette) smoke and air pollution. The most likely situation is that an exacerbation is caused by an airway infection, however a combination with other causes may also explain the exacerbation. Let us now consider this example in some more detail.

Since infections are the most common cause of exacerbations, let us assume that for an exacerbation (exa) caused by some combination of the causes, *infection* (inf) occurs first and then during the period that the patient has an infection, exposure to *smoke* (sm) and to *pollution* (pol) can occur. We thus have the temporal relations: ‘*sm du inf*’ and ‘*pol du inf*’ (the relation *during* is chosen purely for the sake of a simple example). We assume that all causes are present, and that the probability of an exacerbation is given by a binomial distribution over the length of the interval, with parameters $p_{inf} = 0.5$, $p_{sm} = 0.2$ and $p_{pol} = 0.1$. Note that we do not assume immediate consequences, the effect can occur somewhere in an interval in which the cause is true. Now consider for the sake of simplicity a short timespan of interest of 3 days (with a granularity of 1 day). The only intervals that satisfy the constraints are then $[1, 4)$ for *inf* and $[2, 3)$ for both *sm* and *pol*. Thus the probability of an exacerbation can be computed by multiplying the probability that an infection causes an exacerbation in the three day interval with the probability that an exacerbation is caused by exposure to smoke on one day and the probability of exposure to pollution on one day. We thus obtain (note we can drop the sum over intervals because there is only one solution for the constraints):

$$\begin{aligned} P(\text{exa} \mid \text{inf, sm, pol}) &= \sum_{i_1, i_2, i_3} f_{I_1}(i_1 \mid \text{inf}) f_{du(I_2, i_1)}(i_2 \mid \text{sm}) f_{du(I_3, i_1)}(i_3 \mid \text{pol}) \\ &= \binom{3}{1} (p_{inf})^1 (1 - p_{inf})^2 (p_{sm})^1 (1 - p_{sm})^0 (p_{pol})^1 (1 - p_{pol})^0 \\ &= (3 \cdot 0.5 \cdot 0.5^2) \cdot 0.2 \cdot 0.1 = .0075 \end{aligned}$$

That is, the contribution of infection is binomial, with one occurrence in three time points and for smoke and infection we obtain essentially a Bernoulli distribution because only one time point satisfies the constraints, hence the binomial coefficient is 1. The probability of this particular situation is quite low due to the fact that we assumed all causes to be present while the apriori probability for that is low. Allowing for example any combination of two causes to be sufficient to cause an exacerbation would result in higher probabilities (lack of space prevents us from showing this numerically).

5 Conclusions

As many temporal processes are essentially uncertain in nature, modelling uncertainty in this context is an important research topic. In the area of clinical medicine, however, it is hard to gather homogeneous temporal information and thus qualitative relations are used to reason about time. On many occasions, Allen's temporal interval algebra has been shown to be a valuable tool. However, this algebra does not allow coping with the uncertainty that comes with the medical domain. In this paper we have shown that temporal causal independence models can be combined with the Allen algebra, leading to a meaningful probabilistic semantics of qualitative temporal interval information.

References

1. Allen, J.: Maintaining knowledge about temporal intervals. *Communications of the ACM* (1983)
2. Allen, J.: Towards a general theory of action and time. *Artificial Intelligence* (1984)
3. Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann (1988)
4. Murphy, K.P.: Dynamic Bayesian Networks: Representation, Inference and Learning. PhD thesis, University of California, Berkeley (2002)
5. Grimmett, G., Stirzaker, D.: Probability and Random Processes. Oxford University Press (2001)
6. Tawfik, A., Neufeld, E.: Temporal reasoning and Bayesian networks. *Computational Intelligence* (2000)
7. Galán, S., Aguado, F., Díez, F., Mira, J.: NasoNet, modelling the spread of nasopharyngeal cancer with networks of probabilistic events in discrete time. *Artificial Intelligence in Medicine* (2002)
8. Ryabov, V., Trudel, A.: Probabilistic temporal interval networks. In: TIME '04: Proceedings of the 11th International Symposium on Temporal Representation and Reasoning. (2004)
9. Tsamardinos, I., Pollack, M., Ramakrishnan, S.: Assessing the probability of legal execution of plans with temporal uncertainty. In: Proceedings of ICAPS'03 Workshop on Planning under Uncertainty and Incomplete Information. (2003)
10. Heckerman, D.: Causal independence for knowledge acquisition and inference. In: Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence. (1993)