Decidability and Complexity of Threshold Description Logics Induced by Concept Similarity Measures

Franz Baader   Oliver Fernández Gil
Theoretical Computer Science
Technische Universität Dresden, Germany
{franz.baader, oliver.fernandez}@tu-dresden.de

ABSTRACT

In a recent research paper, we have proposed an extension of the lightweight Description Logic (DL) $\mathcal{EL}$ in which concepts can be defined in an approximate way. To this purpose, the notion of a graded membership function m, which instead of a Boolean membership value 0 or 1 yields a membership degree from the interval $[0, 1]$, was introduced. Threshold concepts can then, for example, require that an individual belongs to a concept C with degree at least 0.8. Reasoning in the threshold DL $\tau\mathcal{EL}(m)$ obtained this way of course depends on the employed graded membership function m. The paper defines a specific such function, called *deg*, and determines the exact complexity of reasoning in $\tau\mathcal{EL}(deg)$. In addition, it shows how concept similarity measures (CSMs) $\sim$ satisfying certain properties can be used to define graded membership functions $m_\sim$, but it does not investigate the complexity of reasoning in the induced threshold DLs $\tau\mathcal{EL}(m_\sim)$. In the present paper, we start filling this gap. In particular, we show that computability of $\sim$ implies decidability of $\tau\mathcal{EL}(m_\sim)$, and we introduce a class of CSMs for which reasoning in the induced threshold DLs has the same complexity as in $\tau\mathcal{EL}(deg)$.

CCS Concepts

•Theory of computation → Description logics;

Keywords

Description Logics; Reasoning; Complexity; Similarity

1. INTRODUCTION

DLs are a well-investigated family of logic-based knowledge representation languages, which are frequently used to formalize ontologies for application domains such as biology and medicine. To define the important notions of such an application domain as formal concepts, DLs state necessary and sufficient conditions for an individual to belong to a concept. Once the relevant concepts of an application domain are formalized this way, they can be used in queries in order to retrieve new information from data. The DL $\mathcal{EL}$, in which concepts can be built using concept names as well as the concept constructors conjunction ($\sqcap$), existential restriction ($\exists r.C$), and the top concept ($\top$), has drawn considerable attention in the last decade since, on the one hand, important inference problems such as the subsumption problem are polynomial in $\mathcal{EL}$ [5, 1, 7]. On the other hand, though quite inexpressive, $\mathcal{EL}$ underlies the OWL 2 EL profile[1] and can be used to define biomedical ontologies, such as the large medical ontology SNOMED CT.[2]

Like all traditional DLs, $\mathcal{EL}$ is based on classical first-order logic, and thus its semantics is strict in the sense that all the stated properties need to be satisfied for an individual to belong to a concept. In applications where exact definitions are hard to come by, it would be useful to relax this strict requirement and allow for approximate definitions of concepts, where most, but not all, of the stated properties are required to hold. For example, in clinical diagnosis, diseases are often linked to a long list of medical signs and symptoms, but patients that have a certain disease rarely show all these signs and symptoms. Instead, one looks for the occurrence of sufficiently many of them. Similarly, people looking for a flat to rent or a bicycle to buy may have a long list of desired properties, but will also be satisfied if many, but not all, of them are met. In order to support defining concepts in such an approximate way, in [2] we have introduced a DL extending $\mathcal{EL}$ with threshold concept constructors of the form $C_{\bowtie t}$, where C is an $\mathcal{EL}$ concept, $\bowtie \in \{<, \leq, >, \geq\}$, and t is a *rational* number in $[0, 1]$. The semantics of these new concept constructors is defined using a graded membership function m that, given a (possibly complex) $\mathcal{EL}$ concept C and an individual d of an interpretation $\mathcal{I}$, returns a value from the interval [0,1], rather than a Boolean value from $\{0, 1\}$. The concept $C_{\bowtie t}$ then collects all the individuals that belong to C with degree $\bowtie t$, where this degree is computed using the function m. The DL $\tau\mathcal{EL}(m)$ is obtained from $\mathcal{EL}$ by adding these new constructors. There are, of course, different possibilities for how to define a graded membership function m, and the semantics of the obtained new logic $\tau\mathcal{EL}(m)$ depends on m.

In addition to introducing the family of DLs $\tau\mathcal{EL}(m)$, we have also defined a concrete graded membership function *deg*, which is obtained as a natural extension of the well-known homomorphism characterization of crisp membership

[1]see http://www.w3.org/TR/owl2-profiles/
[2]see http://www.ihtsdo.org/snomed-ct/

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SAC'17, April 3-7, 2017, Marrakesh, Morocco

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ISBN 978-1-4503-4486-9/17/04. . . $15.00
DOI: http://dx.doi.org/10.1145/3019612.3019715

and subsumption in $\mathcal{EL}$ [5]. It is proved in [2] that concept satisfiability and ABox consistency are NP-complete in $\tau\mathcal{EL}(deg)$, whereas the subsumption and the instance checking problem are co-NP complete (the latter w.r.t. *data complexity*). In addition, it is shown how a CSM $\sim$ that is equivalence invariant, role-depth bounded and equivalence closed[3] (see [10]) can be used to define a graded membership function $m_\sim$. In particular, the graded membership function *deg* can be obtained in this way, i.e., there is a standard CSM $\sim^*$ such that $m_{\sim^*} = deg$. However, the complexity of reasoning in the DLs $\tau\mathcal{EL}(m_\sim)$ for $\sim \neq \sim^*$ has not been investigated in [2].

The goal of the present paper is to start filling this gap. Firstly, we will show that, for *computable* standard CSMs $\sim$, reasoning in $\tau\mathcal{EL}(m_\sim)$ can effectively be reduced to reasoning in the DL $\mathcal{ALC}$. Though the complexity of reasoning in $\mathcal{ALC}$ is known to be "only " PSpace [11], the complexity of the decision procedures for reasoning in $\tau\mathcal{EL}(m_\sim)$ obtained this way is non-elementary, due to the high complexity of the reduction function. Secondly, in order to obtain threshold DLs of lower complexity, we determine a class of standard CSMs definable using the *simi framework* of [10] such that reasoning in $\tau\mathcal{EL}(m_\sim)$ for a member $\sim$ of this class has the same complexity as reasoning in $\tau\mathcal{EL}(deg)$. Thirdly, we consider the problem of answering relaxed instance queries [8] using CSMs from this class. For the CSM $\sim^*$ corresponding to *deg*, it was shown in [2] that relaxed instance queries w.r.t. this CSM can be answered in polynomial time. We extend this result to all members of our class. This improves on the complexity upper bounds for answering relaxed instance queries in [8].

Due to the space constraints, we cannot include full proofs of all our results. They can be found in the technical report [3].

2. THE FAMILY OF DLs $\tau\mathcal{EL}(m_\sim)$

First, we introduce the DL $\mathcal{EL}$ and show how, up to equivalence, all $\mathcal{EL}$ concept descriptions over a finite vocabulary and with a bounded role depth can be effectively computed. This will be used later to show the decidability result mentioned in the introduction. Second, we recall the definition of graded membership functions and the induced threshold DLs as well as some additional definitions and results from [2]. Third, we recall how concept similarity measures can be used to define graded membership functions.

The Description Logic $\mathcal{EL}$

Let $\mathsf{N_C}$ and $\mathsf{N_R}$ be finite sets of *concept* and *role* names, respectively. The set $\mathcal{C}_{\mathcal{EL}}(\mathsf{N_C},\mathsf{N_R})$ of $\mathcal{EL}$ concept descriptions over $\mathsf{N_C}$ and $\mathsf{N_R}$ is inductively built from $\mathsf{N_C}$ using the concept constructors *conjunction* ($C \sqcap D$), *existential restriction* ($\exists r.C$), and *top* ($\top$). The semantics of $\mathcal{EL}$ concept descriptions is defined using standard first-order logic interpretations. An *interpretation* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ consists of a non-empty domain $\Delta^\mathcal{I}$ and an interpretation function $\cdot^\mathcal{I}$ that interprets concept names in $\mathsf{N_C}$ as subsets of $\Delta^\mathcal{I}$ and assigns binary relations over $\Delta^\mathcal{I}$ to role names in $\mathsf{N_R}$. This function is inductively extended to complex concept descriptions as follows.

$$\top^\mathcal{I} := \Delta^\mathcal{I}, \quad (C \sqcap D)^\mathcal{I} := C^\mathcal{I} \cap D^\mathcal{I},$$
$$(\exists r.C)^\mathcal{I} := \{x \in \Delta^\mathcal{I} \mid \exists y.((x,y) \in r^\mathcal{I} \wedge y \in C^\mathcal{I})\}.$$

Given two $\mathcal{EL}$ concept descriptions C and D, we say that C is *subsumed* by D (in symbols $C \sqsubseteq D$) iff $C^\mathcal{I} \subseteq D^\mathcal{I}$ for all interpretations $\mathcal{I}$. These two concepts are equivalent (in symbols $C \equiv D$) iff $C \sqsubseteq D$ and $D \sqsubseteq C$. In addition, C is *satisfiable* iff $C^\mathcal{I} \neq \emptyset$ for some interpretation $\mathcal{I}$.[4]

Information about specific individuals (represented by a set of individual names $\mathsf{N_I}$) can be stated in an ABox, which is a finite set of *assertions* of the form $C(a)$ or $r(a,b)$, where $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C},\mathsf{N_R})$, $r \in \mathsf{N_R}$, and $a, b \in \mathsf{N_I}$. An interpretation $\mathcal{I}$ is then extended to assign domain elements $a^\mathcal{I}$ to individual names a. We say that $\mathcal{I}$ satisfies an assertion $C(a)$ iff $a^\mathcal{I} \in C^\mathcal{I}$, and $r(a,b)$ iff $(a^\mathcal{I}, b^\mathcal{I}) \in r^\mathcal{I}$. Furthermore, $\mathcal{I}$ is a model of the ABox $\mathcal{A}$ (denoted as $\mathcal{I} \models \mathcal{A}$) iff it satisfies all the assertions of $\mathcal{A}$. The ABox $\mathcal{A}$ is *consistent* iff $\mathcal{I} \models \mathcal{A}$ for some interpretation $\mathcal{I}$. Finally, an individual a is an *instance* of C in $\mathcal{A}$ iff $a^\mathcal{I} \in C^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathcal{A}$.

As shown in [9], $\mathcal{EL}$ concept descriptions C can be transformed into an equivalent *reduced form* C^r by applying the rewrite rule $C \sqcap D \longrightarrow C$ if $C \sqsubseteq D$ modulo associativity and commutativity of $\sqcap$ as long as possible, not only on the top-level conjunction of the description, but also under the scope of existential restrictions. Up to associativity and commutativity of $\sqcap$, equivalent $\mathcal{EL}$ concept descriptions have the same reduced form.

We denote the size of an $\mathcal{EL}$ concept description C with $\mathsf{s}(C)$. The *role depth* $\mathsf{rd}(C)$ of C is the maximal nesting of existential restrictions in C. As shown in [6], for finite sets $\mathsf{N_C}$ and $\mathsf{N_R}$ and a fixed bound k on the role depth, $\mathcal{C}_{\mathcal{EL}}(\mathsf{N_C},\mathsf{N_R})$ contains only finitely many equivalence classes of concept descriptions of role depth $\leq k$. The following lemma shows that finitely many representatives of these equivalence classes can be computed.

LEMMA 1. *For all $k \geq 0$ there exists a* finite *set $\mathcal{R}^k \subseteq \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C},\mathsf{N_R})$ consisting of $\mathcal{EL}$ concept descriptions in reduced form and of role depth $\leq k$ such that $C^r \in \mathcal{R}^k$ holds for all $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C},\mathsf{N_R})$ with $\mathsf{rd}(C) \leq k$, and this set can effectively be computed.*

PROOF (SKETCH). The lemma can be shown by induction on k. Concept descriptions of role depth $k = 0$ are conjunctions of concept names, where the empty conjunction corresponds to $\top$. The requirement to be reduced corresponds to the fact that each concept name occurs at most once in the conjunction. Thus,

$$\mathcal{R}^0 = \{ \bigsqcap_{A \in S} A \mid S \subseteq \mathsf{N_C} \},$$

which is obviously finite and, given $\mathsf{N_C}$, can easily be computed.

Up to equivalence, concept descriptions of role depth $\leq k$ for $k > 0$ are conjunctions of concept names $A \in \mathsf{N_C}$ and existential restrictions $\exists r.C$ for $r \in \mathsf{N_R}$ and $C \in \mathcal{R}^{k-1}$. The requirement to be reduced imposes the constraint that two different conjuncts $\exists r.C$ and $\exists r.D$ occurring in this conjunction satisfy that C, D are reduced and $C \not\sqsubseteq D$ (and thus

[3]In the following we will call a CSM satisfying these three properties a *standard* CSM.

[4]In $\mathcal{EL}$, all concept descriptions are satisfiable, but this is no longer the case for its extensions by threshold concepts introduced below.

also $C \not\equiv D$). Thus, for every role $r \in \mathsf{N_R}$ there are at most $|\mathcal{R}^{k-1}|$ conjuncts that are existential restrictions for r. Since by induction we know that $\mathcal{R}^{k-1}$ is finite, this implies that $\mathcal{R}^k$ is finite as well. In addition, since subsumption in $\mathcal{EL}$ is decidable and by induction $\mathcal{R}^{k-1}$ is computable, this also implies that $\mathcal{R}^k$ can effectively be computed. □

Extending $\mathcal{EL}$ with threshold concepts

In [2], $\mathcal{EL}$ is extended with *threshold concepts* $C_{\bowtie t}$, where C is an $\mathcal{EL}$ concept description, $\bowtie \in \{<, \leq, >, \geq\}$, and t is a rational number in $[0, 1]$. These threshold concepts can then be used like concept names when building complex concept descriptions such as $(\exists r.A)_{<1} \sqcap \exists r.(A \sqcap B)_{\geq .8} \sqcap B$. Note that the concept C occurring within the threshold operator must be an $\mathcal{EL}$ concept description, and thus nesting of these operators is not allowed. The semantics of the threshold operators is defined using a *graded membership function*, which is defined as follows.[5]

DEFINITION 1. A *graded membership function* m is a family of functions that contains for every interpretation $\mathcal{I}$ a function $m^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R}) \to [0, 1]$ satisfying the following conditions (for $C, D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$):

M1: $\forall \mathcal{I}\ \forall d \in \Delta^{\mathcal{I}} : d \in C^{\mathcal{I}} \Leftrightarrow m^{\mathcal{I}}(d, C) = 1$,

M2: $C \equiv D \Leftrightarrow \forall \mathcal{I}\ \forall d \in \Delta^{\mathcal{I}} : m^{\mathcal{I}}(d, C) = m^{\mathcal{I}}(d, D)$.

Intuitively, given an interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$, $m^{\mathcal{I}}(d, C) \in [0, 1]$ represents the degree to which d belongs to C in $\mathcal{I}$. The concept $C_{\bowtie t}$ then collects all the elements of $\Delta^{\mathcal{I}}$ that belong to C with degree $\bowtie t$, as measured by m. To be more precise, the formal semantics of threshold concepts is then defined as follows: $(C_{\bowtie t})^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid m^{\mathcal{I}}(d, C) \bowtie t\}$.

This way, a new family of DLs called $\tau\mathcal{EL}(m)$ is obtained, where m is a parameter indicating which function is used to obtain the semantics of threshold concepts.

In addition to this family of DLs, [2] introduces a concrete membership function *deg*, and investigates the computational properties of its corresponding DL $\tau\mathcal{EL}(deg)$. We show that *satisfiability* and *consistency* are NP-complete, whereas *subsumption* and *instance checking* (w.r.t. data complexity) are coNP-complete in $\tau\mathcal{EL}(deg)$ (Th. 5 and 6 in [2]). An important step towards obtaining these results was to characterize when an individual is an instance of a $\tau\mathcal{EL}(deg)$ concept description in an interpretation. This characterization generalizes the corresponding one for *crisp* membership in $\mathcal{EL}$, which is based on the representation of concepts and interpretations as graphs, and the existence of homomorphisms between these graphs. Since it is needed in Section 3, we briefly describe the general ideas behind it. In fact, it turns out that this characterization works for $\tau\mathcal{EL}(m)$ regardless of which graded membership function m is used.

$\mathcal{EL}$ *description graphs* are graphs where the nodes are labeled with sets of concept names and the edges are labeled with role names. As shown in [1, 5], interpretations can be represented as (arbitrary) $\mathcal{EL}$ description graphs and $\mathcal{EL}$ concept descriptions as $\mathcal{EL}$ *description trees*, i.e., as description graphs that are trees (whose root we will always denote as v_0). Description trees can be extended to $\tau\mathcal{EL}(m)$ by allowing the node labels also to contain elements of the form $C_{\bowtie t}$. For instance, the left-hand side of Figure 1 depicts

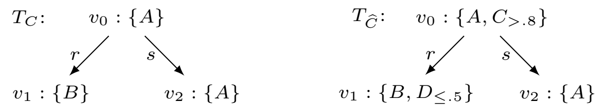

Figure 1: $\mathcal{EL}$ and $\tau\mathcal{EL}(m)$ description trees

the $\mathcal{EL}$ description tree corresponding to the $\mathcal{EL}$ concept description $A \sqcap \exists r.B \sqcap \exists s.A$, whereas the right-hand side shows the $\tau\mathcal{EL}(m)$ description tree corresponding to the $\tau\mathcal{EL}(m)$ concept description $A \sqcap C_{>.8} \sqcap \exists r.(B \sqcap D_{\leq .5}) \sqcap \exists s.A$.

Based on the definition of homomorphisms between $\mathcal{EL}$ description trees in [5], the notion of a τ-homomorphism ϕ from a $\tau\mathcal{EL}(m)$ description tree $\widehat{H}$ into an $\mathcal{EL}$ description graph $G_{\mathcal{I}}$ representing an interpretation $\mathcal{I}$ is defined in [2] to be a mapping from the nodes of $\widehat{H}$ to the nodes of $G_{\mathcal{I}}$ such that

1. the concept names occurring in the label set of a node v of $\widehat{H}$ are contained in the label set of its image $\phi(v)$;
2. if (v, w) is an edge with label r in $\widehat{H}$, then there is an edge $(\phi(v), \phi(w))$ with label r in $G_{\mathcal{I}}$;
3. if the label set of a node v of $\widehat{H}$ contains $C_{\bowtie t}$, then $m^{\mathcal{I}}(\phi(v), C) \bowtie t$.

Conditions 1 and 2 correspond to the classical definition of homomorphisms between $\mathcal{EL}$ description graphs. Using τ-homomorphisms, membership in $\tau\mathcal{EL}(m)$ concept descriptions can be characterized as follows.

THEOREM 1. *Let $\mathcal{I}$ be an interpretation with associated $\mathcal{EL}$ description graph $G_{\mathcal{I}}$, $d \in \Delta^{\mathcal{I}}$, and $\widehat{C}$ a $\tau\mathcal{EL}(m)$ concept description with associated $\tau\mathcal{EL}(m)$ description tree $T_{\widehat{C}}$. Then, $d \in \widehat{C}^{\mathcal{I}}$ iff there exists a τ-homomorphism ϕ from $T_{\widehat{C}}$ to $G_{\mathcal{I}}$ such that $\phi(v_0) = d$.*

If the interpretation $\mathcal{I}$ is finite and m is computable in polynomial time, then the existence of a τ-homomorphism can be checked in polynomial time. For the case $m = deg$ this fact as well as Theorem 1 were already shown in [2].

CSMs and graded membership functions

A concept similarity measure (CSM) is a function that maps pairs of concept descriptions to values in $[0, 1]$. Intuitively, the larger this value is the more similar the concept descriptions are. More formally, a CSM for $\mathcal{EL}$ concept descriptions over $\mathsf{N_C}$ and $\mathsf{N_R}$ is a mapping $\sim : \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R}) \times \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R}) \to [0, 1]$. Examples of such measures as well as properties these measures should satisfy can, e.g., be found in [12, 8, 10].

Definition 10 in [2] shows how a CSM $\sim$ can be used to define an associated graded membership function $m_\sim$:

$$m^{\mathcal{I}}_{\sim}(d, C) := \max\{C \sim D \mid D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R}) \text{ and } d \in D^{\mathcal{I}}\}.$$

To ensure that this definition yields a well-defined graded membership function, $\sim$ is required to be a *standard CSM*, which means that it needs to satisfy the following three properties:

- $\sim$ must be *equivalence invariant*, i.e., $C \equiv C'$ and $D \equiv D'$ implies $C \sim D = C' \sim D'$;

[5]Note that this definition corrects a typo in Def. 3 of [2].

- $\sim$ must be *role-depth bounded*, i.e., $C \sim D = C_k \sim D_k$ where $k > \min\{\mathsf{rd}(C), \mathsf{rd}(D)\}$ and C_k, D_k are the *restrictions of C, D to role depth k*, which are obtained from C, D by removing all existential restrictions occurring at role depth k;

- $\sim$ must be *equivalence closed*, i.e., the equivalence $C \equiv D$ iff $C \sim D = 1$ holds.

The first two conditions ensure that $m^{\mathcal{I}}_{\sim}(d, C)$ is well-defined, i.e., the maximum in the definition of this value really exists. In fact, these conditions imply that one can restrict the search for an appropriate $D \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$ to finitely many concept descriptions.

Lemma 2. *Let $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$ with $\mathsf{rd}(C) = k$. Then* $m^{\mathcal{I}}_{\sim}(d, C) = \max\{C \sim D \mid D \in \mathcal{R}^{k+1}$ *and* $d \in D^{\mathcal{I}}\}$.

Equivalence closedness is additionally needed to ensure that m satisfies the properties required in Definition 1.

3. REASONING IN $\tau\mathcal{EL}(m_{\sim})$

There is a great variety of standard CSMs and not all of them are well-behaved from a computational point of view. In particular, there are standard CSMs that are not computable. While non-computability of $\sim$ does not automatically imply that reasoning problems in $\tau\mathcal{EL}(m_{\sim})$ are undecidable, we were able to show [3] that there is a non-computable CSM $\sim$ such that the standard reasoning problems satisfiability, subsumption, consistency, and instance are undecidable in $\tau\mathcal{EL}(m_{\sim})$. Next, we will show that computability of $\sim$ implies decidability of these reasoning problems in $\tau\mathcal{EL}(m_{\sim})$.

Decidability

Let $\sim$ be a computable standard CSM. We show decidability of reasoning in $\tau\mathcal{EL}(m_{\sim})$ using an equivalence preserving and computable translation of $\tau\mathcal{EL}(m_{\sim})$ concept descriptions into $\mathcal{ALC}$ concept descriptions. Since the standard reasoning problems are decidable in $\mathcal{ALC}$ such an effective translation obviously yields their decidability in $\tau\mathcal{EL}(m_{\sim})$.

Recall that $\mathcal{ALC}$ [11] is obtained from $\mathcal{EL}$ by adding negation $\neg C$, whose semantics is defined in the usual way, i.e., $(\neg C)^{\mathcal{I}} := \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$. Obviously, negation together with conjunction also yields disjunction $C \sqcup D$. Since $\mathcal{EL}$ is a fragment of $\mathcal{ALC}$, it suffices to show how to translate threshold concepts $C_{\bowtie t}$ into $\mathcal{ALC}$ concept descriptions. In addition, we can concentrate on the case where $\bowtie \in \{\geq, >\}$ since $C_{<t} \equiv \neg C_{\geq t}$ and $C_{\leq t} \equiv \neg C_{>t}$.

Lemma 3. *Let $\bowtie \in \{\geq, >\}$, $t \in [0,1] \cap \mathbb{Q}$, and $C \in \mathcal{C}_{\mathcal{EL}}(\mathsf{N_C}, \mathsf{N_R})$ with $rd(C) = k$. Then*

$$C_{\bowtie t} \equiv \bigsqcup \{D \mid D \in \mathcal{R}^{k+1} \text{ and } C \sim D \bowtie t\}.$$

Proof. Let $\mathcal{I}$ be an interpretation and $d \in \Delta^{\mathcal{I}}$. By the semantics of threshold concepts and Lemma 2, we know that $d \in (C_{\bowtie t})^{\mathcal{I}}$ iff

$$m^{\mathcal{I}}_{\sim}(d, C) = \max\{C \sim D \mid D \in \mathcal{R}^{k+1} \text{ and } d \in D^{\mathcal{I}}\} \bowtie t.$$

Since $\bowtie \in \{\geq, >\}$, this is equivalent to saying that there is a $D \in \mathcal{R}^{k+1}$ such that $C \sim D \bowtie t$ and $d \in D^{\mathcal{I}}$. This is in turn equivalent to $d \in \bigcup \{D^{\mathcal{I}} \mid D \in \mathcal{R}^{k+1}$ and $C \sim D \bowtie t\}$. $\square$

Since $\mathcal{R}^{k+1}$ is finite, the disjunction on the right-hand side of the equivalence in the formulation of the lemma is finite, and thus this right-hand side is an admissible $\mathcal{ALC}$ concept description. This description can effectively be computed since $\mathcal{R}^{k+1}$ is computable by Lemma 1 and $\sim$ is computable by assumption.

Theorem 2. *If $\sim$ is a computable standard CSM, then satisfiability, subsumption, consistency and instance checking are decidable in $\tau\mathcal{EL}(m_{\sim})$.*

Since the cardinality of $\mathcal{R}^k$ increases by one exponent with each increase of k, this approach provides only a *non-elementary* bound on the complexity of reasoning in $\tau\mathcal{EL}(m_{\sim})$. We will now show that, for a restricted class of CSMs, one can obtain better complexity upper bounds.

Complexity

As shown in [2], there is a computable standard CSM $\sim^*$ such that $deg = m_{\sim^*}$, and the complexity of reasoning in $\tau\mathcal{EL}(deg)$ is NP/coNP-complete for the standard reasoning problems. We will now identify a class of standard CSMs $\sim$ such that the complexity of reasoning in the induced threshold DLs $\tau\mathcal{EL}(m_{\sim})$ is the same as in $\tau\mathcal{EL}(deg)$.

The CSM $\sim^*$ inducing *deg* is an instance of the *simi framework* introduced in [10]. This framework can be used to define a variety of similarity measures between $\mathcal{EL}$ concepts satisfying certain desirable properties. Here, we introduce a fragment of *simi* that is sufficient for our purposes.

To construct a CSM $\sim$ using *simi*, one first defines a directional measure $\sim_d$, and then uses a fuzzy connector $\otimes$ to combine the values obtained by comparing the reduced concepts in both directions with $\sim_d$:

$$C \sim D := (C^r \sim_d D^r) \otimes (D^r \sim_d C^r),$$

where the fuzzy connector $\otimes$ is a *commutative* binary operator $\otimes : [0,1] \times [0,1] \to [0,1]$ satisfying certain additional properties (see [10]). The definition of $C \sim_d D$ (see Def. 3 in [10]) depends on several parameters:

- a function g that assigns to every $\mathcal{EL}$ atom (i.e., concept name or existential restriction) a weight in $\mathbb{R}_{>0}$;

- a discounting factor $w \in [0, 1)$;

- a primitive measure between concept names and between role names: $pm : (\mathsf{N_C} \times \mathsf{N_C}) \cup (\mathsf{N_R} \times \mathsf{N_R}) \to [0, 1]$.

Once these parameters are fixed, the induced directional measure $\sim_d$ is defined as follows: if $C \equiv \top$, then $C \sim_d D := 1$; if $C \not\equiv \top$ and $D \equiv \top$, then $C \sim_d D := 0$; otherwise, we use $top(C)$ and $top(D)$ to denote the set of $\mathcal{EL}$ atoms occurring in the top-level conjunction of C and D, and define

$$C \sim_d D := \frac{\sum_{C' \in top(C)} \left[g(C') \times \max_{D' \in top(D)} \left(simi_a(C', D') \right) \right]}{\sum_{C' \in top(C)} g(C')},$$

where $simi_a(A, B) := pm(A, B)$ for all $A, B \in \mathsf{N_C}$,
$simi_a(\exists r.E, \exists s.F) := pm(r, s)[w + (1-w)(E \sim_d F)]$, and
$simi_a(C', D') := 0$ in any other case.

If $\otimes$, g and pm can be computed in polynomial time, then the induced CSM $\sim$ can also be computed in polynomial time (see [10], Lemma 2). Moreover, all the CSMs obtained

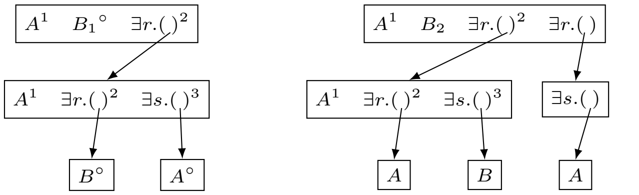

Figure 2: Computation of the directional measure

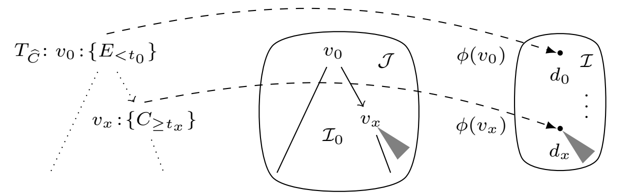

Figure 3: Polynomial bounded model construction

as instances of *simi* where g assigns 1 to atoms of the form $\exists r.C$ are standard CSMs (see [3]). One such instance of *simi* is $\sim^*$, where $\otimes = \min$, $w = 0$, g assigns 1 to all atoms, and pm is the *default* primitive measure pm_d assigning value 1 when $A = B$ ($r = s$), and 0 otherwise. We now define a class of instances of *simi* containing $\sim^*$.

Definition 2. The class *simi-mon* is obtained from *simi* by restricting the admissible parameters as follows:

- $\otimes$ is computable in polynomial time and monotonic w.r.t. $\geq$;[6]
- g is computable in polynomial time and assigns 1 to all atoms of the form $\exists r.C$;
- $pm = pm_d$ and w is arbitrary.

In the following we will show that, for all $\sim \in$ *simi-mon*, reasoning in $\tau\mathcal{EL}(m_\sim)$ is *not* harder than reasoning in $\tau\mathcal{EL}(deg)$. We start with illustrating some useful properties satisfied by CSMs in *simi-mon*.

Example 1. We consider a CSM $\sim$ whose definition deviates from the one of $\sim^*$ only in one place: we use $w = .5$. Consider

$$C := A \sqcap B_1 \sqcap \exists r.(A \sqcap \exists r.B \sqcap \exists s.A),$$
$$D := A \sqcap B_2 \sqcap \exists r.(A \sqcap \exists r.A \sqcap \exists s.B) \sqcap \exists r.\exists s.A.$$

Figure 2 basically shows the atoms in D chosen by max when computing $C \sim_d D$. The *superscripts* are used to denote the corresponding pairings for which the value is > 0. For instance, at the top level of C, A^1 means that A is paired with the top-level atom of D having the same superscript. The symbol $\circ$ on the left-hand side tells us that no match yielding a value > 0 exists. Now, removing the atoms without superscript in D yields the concept $Y := A \sqcap \exists r.(\exists r.\top \sqcap \exists s.\top)$. One can easily verify that $C \sim_d D = C \sim_d Y = 5/9$, and it is clear that C and D are both subsumed by Y.

These properties can be generalized to all pair of concepts and measures in *simi-mon* (see [3] for complete proofs).

Lemma 4. *Let $\sim \in$ simi-mon. For all $\mathcal{EL}$ concept descriptions C and D, there exists an $\mathcal{EL}$ concept description Y such that:*

1. *$D \sqsubseteq Y$ and $\mathsf{s}(Y) \leq \mathsf{s}(C)$,*
2. *$C \sim_d D = C \sim_d Y$,*
3. *$C \sqsubseteq Y$.*

[6]Examples are *average* and all polynomially computable bounded *t-norms*.

We use these properties to show that, like $\tau\mathcal{EL}(deg)$ (see Lemma 4 in [2]), $\tau\mathcal{EL}(m_\sim)$ enjoys a polynomial model property if $\sim \in$ *simi-mon*.

Lemma 5. *Let $\sim \in$ simi-mon and $\widehat{C}$ a $\tau\mathcal{EL}(m_\sim)$ concept description. If $\widehat{C}$ is satisfiable, then there is a tree-shaped interpretation $\mathcal{J}$ such that $\widehat{C}^\mathcal{J} \neq \emptyset$ and $|\Delta^\mathcal{J}| \leq \mathsf{s}(\widehat{C})$.*

Proof (Sketch). Figure 3 outlines the description tree $T_{\widehat{C}}$ of a $\tau\mathcal{EL}(m_\sim)$ concept $\widehat{C}$, an interpretation $\mathcal{I}$ such that $d_0 \in \widehat{C}^\mathcal{I}$, and a corresponding τ-homomorphism ϕ obtained by applying Theorem 1. The tree in the middle represents the finite interpretation $\mathcal{J}$ we want to build. The construction of $\mathcal{J}$ starts with a base interpretation $\mathcal{I}_0$ that corresponds to $T_{\widehat{C}}$ (first ignoring labels of the form $C_{\bowtie t}$). Consequently, the identity mapping ϕ_{id} from $T_{\widehat{C}}$ to $G_{\mathcal{I}_0}$ satisfies Conditions 1 and 2 required for τ-homomorphisms. However, the third condition need not be satisfied since, for instance, $m_\sim^{\mathcal{I}_0}(v_x, C)$ could well be smaller than t_x. To fix this, $\mathcal{I}_0$ is extended into $\mathcal{J}$ by attaching to it a *tree-shaped* interpretation (the gray triangle in the figure) such that $m_\sim^\mathcal{J}(v_x, C) \geq t_x$. This interpretation can be extracted from $\mathcal{I}$ using the fact that $\phi(v_x) = d_x$ implies that $m_\sim^\mathcal{I}(d_x, C) \geq t_x$ (because ϕ is a τ-homomorphism). To be more precise, consider an $\mathcal{EL}$ concept description D such that

$$d_x \in D^\mathcal{I} \text{ and } m_\sim^\mathcal{I}(d_x, C) = C \sim D.$$

In principle, we could use the interpretation $\mathcal{I}_D$ having the description tree T_D as the one to be attached to v_x. However, we do not know anything about the size of D. This is where Lemma 4 comes into play. Instead of D it allows us to use the concept Y. Statement 1. of the lemma tells us that $\mathsf{s}(Y) \leq \mathsf{s}(C)$ and that d_x also belongs to $Y^\mathcal{I}$. Statement 2. shows that Y yields the same value as D in the directional measure, and Statement 3. can be used to show that this is also the case for $\sim$.

This approach can be applied to all threshold concepts of the form $C'_{>t}$ or $C'_{\geq t}$ occurring in $\widehat{C}$. For each such concept, the number of domain elements added to satisfy it is bounded by the size of C'. It remains to see why threshold concepts using $<$ or $\leq$, like $E_{<t_0}$ in the figure, are not violated. The reason is basically that they are satisfied in $\mathcal{I}$, and that everything occurring in T_Y also occurs in $\mathcal{I}$ (since $d_x \in Y^\mathcal{I}$) (see [3] for a detailed proof). □

Lemma 4 can also be used to show that, for a finite interpretation $\mathcal{I}$, the function $m_\sim^\mathcal{I}$ can be computed in polynomial time. Basically, the reason is that the lemma restricts the search for an appropriate concept D yielding the maximum to small concepts Y that have a strong resemblance to C.

Proposition 1. *Let $\sim \in$ simi-mon. For every finite interpretation $\mathcal{I}$, $d \in \Delta^\mathcal{I}$, and $\mathcal{EL}$ concept description C,*

$m^{\mathcal{I}}_{\sim}(d, C)$ *can be computed in time polynomial in the size of* $\mathcal{I}$ *and* C.

Together with this proposition, Lemma 5 yields a standard guess-and-check NP-procedure for satisfiability in $\tau\mathcal{EL}(m_\sim)$. Regarding the other reasoning tasks, the constructions introduced in [2] for $\tau\mathcal{EL}(deg)$ to provide appropriate bounded model properties for them can also be applied for $\tau\mathcal{EL}(m_\sim)$.

Theorem 3. *Let* $\sim \in$ *simi-mon. In* $\tau\mathcal{EL}(m_\sim)$*, satisfiability and consistency are in NP, whereas subsumption and instance checking (w.r.t. data complexity) are in coNP.*

In [2], satisfiability in $\tau\mathcal{EL}(deg)$ is shown to be NP-hard by reducing an NP-complete variant $\mathcal{V}$ of propositional satisfiability to it. However, the reduction introduces a *fresh* concept name for each propositional variable occurring in an instance of $\mathcal{V}$. Since in the present paper we assume that concept descriptions in $\tau\mathcal{EL}(m_\sim)$ are defined over a fixed finite vocabulary $\mathsf{N_C} \cup \mathsf{N_R}$, it is thus not possible to use the same reduction. In [3] we introduce a new reduction that shows that satisfiability in $\tau\mathcal{EL}(m_\sim)$ is NP-hard, even if only one concept name and one role name is available. However, for this result to hold we need additional restrictions on $\sim$. Let *simi-smon* be the subset of *simi-mon* whose measures are defined using a fuzzy connector that is *strictly monotonic* or has 1 as a unit. Since satisfiability can be reduced to the consistency, non-subsumption and non-instance problem, we thus obtain the following hardness results.

Proposition 2. *Let* $\sim \in$ *simi-smon. In* $\tau\mathcal{EL}(m_\sim)$*, satisfiability and consistency are NP-hard, whereas subsumption and instance checking are coNP-hard.*

In [2], it was shown for $\tau\mathcal{EL}(deg)$ that instance checking becomes polynomial if instead of arbitrary $\tau\mathcal{EL}(deg)$ concept descriptions one considers only threshold concepts of the form $C_{>t}$. We can show that this result holds not just for *deg*, but for all CSMs in our class *simi-mon*.

Proposition 3. *Let* $\sim \in$ *simi-mon. In* $\tau\mathcal{EL}(m_\sim)$*, the instance checking problem for threshold concepts of the form* $C_{>t}$ *can be decided in polynomial time.*

Since it was shown in [2] (Proposition 5) that computing instances of threshold concepts of the form $C_{>t}$ in a logic $\tau\mathcal{EL}(m_\sim)$ corresponds to answering so-called relaxed instance queries w.r.t. $\sim$ (see [8]), this also yields a polynomiality result for answering relaxed instance queries w.r.t. CSMs in *simi-mon*.

4. CONCLUSIONS

We have shown that the complexity results for reasoning in the threshold logic $\tau\mathcal{EL}(deg)$ of [2] can be extended to a large class of logics $\tau\mathcal{EL}(m_\sim)$ that are induced by appropriate concept similarity measures. Like in [2], we do not consider terminological axioms (TBoxes) in the present paper. In [4], reasoning w.r.t. acyclic TBoxes in $\tau\mathcal{EL}(deg)$ was considered. It would be interesting to see whether the results of [4], which surprisingly show that acyclic TBoxes increase the complexity, can also be extended to our logics $\tau\mathcal{EL}(m_\sim)$ for $\sim \in$ *simi-mon*.

5. ACKNOWLEDGMENTS

The second author was supported by DFG Graduiertenkolleg 1763 (QuantLA).

6. REFERENCES

[1] F. Baader. Terminological cycles in a description logic with existential restrictions. In *Proc. IJCAI'2003*, pages 325–330, 2003.

[2] F. Baader, G. Brewka, and O. Fernández Gil. Adding threshold concepts to the description logic $\mathcal{EL}$. In *Proc. FroCoS'15*, volume 9322 of *LNCS*, pages 33–48. Springer, 2015.

[3] F. Baader and O. Fernández Gil. Decidability and complexity of threshold description logics induced by concept similarity measures. LTCS-Report LTCS-16-07, Chair for Automata Theory, Institute for Theoretical Computer Science, TU Dresden, Germany, 2016. See http://lat.inf.tu-dresden.de/research/reports.html.

[4] F. Baader and O. Fernández Gil. Extending the description logic $\tau\mathcal{EL}(deg)$ with acyclic TBoxes. In *Proc. ECAI'16*, volume 285, pages 1096–1104. IOS Press, 2016.

[5] F. Baader, R. Küsters, and R. Molitor. Computing least common subsumers in description logics with existential restrictions. In *Proc. IJCAI'99*, pages 96–103, 1999.

[6] F. Baader, B. Sertkaya, and A. Turhan. Computing the least common subsumer w.r.t. a background terminology. *J. of Applied Logic*, 5(3):392–420, 2007.

[7] S. Brandt. Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and - what else? In *Proc. ECAI'04*, pages 298–302. IOS Press, 2004.

[8] A. Ecke, R. Peñaloza, and A.-Y. Turhan. Similarity-based relaxed instance queries. *J. Applied Logic*, 13(4):480–508, 2015.

[9] R. Küsters. *Non-Standard Inferences in Description Logics*, volume 2100 of *LNCS*. Springer, 2001.

[10] K. Lehmann and A.-Y. Turhan. A framework for semantic-based similarity measures for $\mathcal{ELH}$-concepts. In *Proc. JELIA'12*, volume 7519 of *LNCS*, pages 307–319. Springer, 2012.

[11] M. Schmidt-Schauß and G. Smolka. Attributive concept descriptions with complements. *Artif. Intell.*, 48(1):1–26, 1991.

[12] B. Suntisrivaraporn. A similarity measure for the description logic $\mathcal{EL}$ with unfoldable terminologies. In *Proc. INCoS'13*, pages 408–413. IEEE, 2013.