Preferential Query Answering in the Semantic Web with Possibilistic Networks (Extended Abstract)

Stefan Borgwardt[1], Bettina Fazzinga[2], Thomas Lukasiewicz[3], Akanksha Shrivastava[3], and Oana Tifrea-Marciuska[3]

[1] Faculty of Computer Science, Technische Universität Dresden, Germany
stefan.borgwardt@tu-dresden.de
[2] ICAR, National Research Council (CNR), Italy
fazzinga@icar.cnr.it
[3] Department of Computer Science, University of Oxford, UK
akanksha.shrivstv@gmail.com, {thomas.lukasiewicz, oana.tifrea}@cs.ox.ac.uk

Abstract. In this paper, we explore how ontological knowledge expressed via existential rules can be combined with possibilistic networks (i) to represent qualitative preferences along with domain knowledge, and (ii) to realize preference-based answering of conjunctive queries (CQs). We call these combinations ontological possibilistic networks (OP-nets). We define skyline and k-rank answers to CQs under preferences and provide complexity (including data tractability) results for deciding consistency and CQ skyline membership for OP-nets. We show that our formalism has a lower complexity than a similar existing formalism.

Introduction

The abundance of information on the Web requires new personalized filtering techniques to retrieve resources that best fit users' interests and preferences. Moreover, the Web is evolving at an increasing pace towards the so-called Social Semantic Web (or Web 3.0), where classical linked information lives together with ontological knowledge and social interactions of users. While the former may allow for more precise and rich results in search and query answering tasks, the latter can be used to enrich the user profile, and it paves the way to more sophisticated personalized access to information. This requires new techniques for ranking search results, fully exploiting ontological and user-centered data, i.e., user preferences.

Conditional preferences are statements of the form "in the context of c, a is preferred over b", denoted $c\colon a \succ b$ [1, 7, 13]. Two preference formalisms that can represent such preferences are *possibilistic networks* and *CP-nets*.

Example 1. Bob wants to rent a car and (i) he prefers a new car over an old one, (ii) given he has a new car, he prefers it to be black over not black, and (iii) if he has an old car, he prefers it to be colorful over being black. We have two variables for car type (new (n) or old (o)) and car color (black (b) or colorful (c)),

	id	color	feature	type
t_1	s_1	b	f_1	o
t_2	s_2	c	f_2	n
t_3	s_3	c	f_2	o

specs

	id	name
t_7	f_1	ac
t_8	f_2	map
t_9	f_3	cd

feature

	vendor	price	specs
t_4	v_1	30	s_1
t_5	v_1	40	s_2
t_6	v_2	50	s_3

offer

	id	review
t_{10}	v_1	p
t_{11}	v_2	n

vendor

Fig. 1. Database D

T and C, respectively, such that $Dom(T) = \{n, o\}$ and $Dom(C) = \{b, c\}$. Bob's preferences can be encoded as $\top : n \succ o$, $n : b \succ c$, and $o : c \succ b$. In CP-nets [7], we have the following ordering of outcomes: $nb \succ nc \succ oc \succ ob$. That is, a new and colorful car is preferred over an old and colorful one, which is not a realistic representation of the given preferences. A more desirable order of outcomes for Bob would be $nb \succ oc \succ nc \succ ob$, which can be induced in possibilistic networks with an appropriate preference weighting in the possibility distribution. ■

We propose a novel language for expressing preferences over the Web 3.0 using possibilistic networks. It has lower complexity compared to a similar existing formalism called OCP-theories [9], which are an integration of Datalog+/− with CP-theories [13]. This is because deciding dominance in possibilistic networks can be done in polynomial time, while it is PSPACE-complete in CP-theories. Every possibilistic network encodes a unique (numerical) ranking on the outcomes, while CP-theories encode a set of (qualitative) total orders on the outcomes. Our framework also allows to specify the relative importance of preferences [1]. Possibilistic networks are also a simple and natural way of representing conditional preferences and obtaining rankings on outcomes, and can be easily learned from data [5]. We choose existential rules in Datalog+/− as ontology language for their intuitive nature, expressive power for rule-based knowledge bases, and the capability of performing query answering.

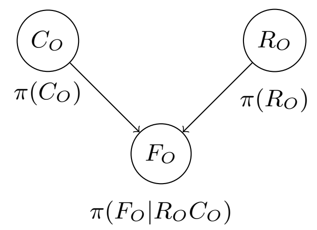
All details can be found in the full paper [6].

Ontological Possibilistic Networks (OP-nets)

See [3, 8] for the basic notions regarding possibilistic networks and Datalog+/−. Let $O = (D, \Sigma)$ be a Datalog+/− *ontology*, where D is a database and Σ a finite set of *tuple-generating dependencies (TGDs)* and *negative constraints (NCs)*.

Example 2. Consider the database D in Fig. 1, modeling the domain of an online car booking system. Moreover, the dependencies

$$\begin{aligned}\Sigma = \{ & \mathit{offer}(V,P,S) \rightarrow \exists C,F,T\ \mathit{specs}(S,C,F,T),\\ & \mathit{offer}(V,P,S) \rightarrow \exists R\ \mathit{vendor}(V,R),\\ & \mathit{specs}(S,C,F,T) \rightarrow \mathit{color}(C) \wedge \mathit{type}(T),\\ & \mathit{specs}(S,C,F,T) \rightarrow \exists N\ \mathit{feature}(F,N),\\ & \mathit{offer}(V,P1,S) \wedge \mathit{offer}(V,P2,S) \rightarrow P1 = P2\ \}\end{aligned}$$



$\pi(specs(t_1))$	$\pi(specs(t_2))$	$\pi(specs(t_3))$
1	0.5	0.4

$\pi(vendor(t_{10}))$	$\pi(vendor(t_{11}))$
1	0.4

$\pi(\cdot\|\cdot)$	$\mathbf{t}_1\mathbf{t}_{10}$	$\mathbf{t}_1\mathbf{t}_{11}$	$\mathbf{t}_2\mathbf{t}_{10}$	$\mathbf{t}_2\mathbf{t}_{11}$	$\mathbf{t}_3\mathbf{t}_{10}$	$\mathbf{t}_3\mathbf{t}_{11}$
$feature(t_7)$	1	0.3	0.2	0.2	0.2	0.2
$feature(t_8)$	0.7	0.5	0.7	1	0.4	0.3
$feature(t_9)$	0.5	0.3	0.5	0.3	1	0.2

Fig. 2. Graph and possibility distribution for Example 3.

say that every offer must have a specification and a vendor and that there cannot be two equivalent offers from the same company with different prices. We denote by $\mathbf{t}_1$ the term $specs(s_1, b, f_1, o)$ and by t_1 the tuple (s_1, b, f_1, o). ■

Let now $\mathcal{X}_O$ be a finite set of variables, where each $X \in \mathcal{X}_O$ corresponds to a predicate from O, denoted $pred(X)$. The *domain* $Dom(X)$ consists of at least two ground atoms $p(c_1, \dots c_k)$ with $p = pred(X)$. An *outcome* $o \in Dom(\mathcal{X}_O)$ assigns to each variable an element of its domain, and can be seen as a conjunction of ground atoms. An *OP-net* is of the form (O, Γ), where Γ is a possibilistic network over $\mathcal{X}_O$, i.e., a collection of conditional possibility distributions $\pi(X_i|pa(X_i))$, where $pa(X_i)$ are the *parents* of X_i. Taken altogether, they define a joint possibility distribution over $Dom(\mathcal{X}_O)$. An outcome o *dominates* another outcome o' (written $o \succ o'$) if $\pi(o) > \pi(o')$. This relation can be decided in polynomial time.

Example 3. Consider the OP-net (O, Γ) given by the ontology O of Example 2 and the dependency graph and the conditional possibility distribution in Fig. 2. Here, we have $\mathcal{X}_O = \{C_O, R_O, F_O\}$ with the domains

$$\begin{aligned} Dom(C_O) &= \{specs(t_1), specs(t_2), specs(t_3)\}, \\ Dom(F_O) &= \{feature(t_7), feature(t_8), feature(t_9)\}, \\ Dom(R_O) &= \{vendor(t_{10}), vendor(t_{11})\}. \end{aligned}$$

The parents of F_O are $\{C_O, R_O\}$, which in turn do not depend on other variables. The distribution could either be learned or derived from explicit preferences; see Example 4 below. The possibilities of outcomes are then computed as

$$\pi(C_O R_O F_O) = \pi(F_O|C_O R_O) \otimes \pi(C_O) \otimes \pi(R_O).$$

The outcome o with $o(C_O) = specs(t_1)$, $o(R_O) = vendor(t_{10})$, $o(F_O) = feature(t_7)$ represents the conjunction $\mathbf{t}_1 \wedge \mathbf{t}_{10} \wedge \mathbf{t}_7$ and has the possibility 1. ■

Since outcomes are conjunctions of ground atoms, they may be inconsistent or equivalent w.r.t. Σ. An outcome o of (O, Γ) is *consistent* if the ontology $O_o = O \cup \{o(X) \mid X \in \mathcal{X}_O\}$ is consistent. Two outcomes o and o' are *equivalent*, denoted $o \sim o'$, if O_o and $O_{o'}$ have the same models. An *interpretation* $\mathcal{I}$ for

(O,Γ) is a total preorder over the consistent outcomes in $Dom(\mathcal{X}_O)$. It *satisfies* (or is a *model* of) (O,Γ) if it is compatible with the dominance and equivalence relations, i.e., for all consistent outcomes o and o', (i) if $o \prec o'$, then $(o,o') \in \mathcal{I}$ and $(o',o) \notin \mathcal{I}$, and (ii) if $o \sim o'$, then $(o,o'),(o',o) \in \mathcal{I}$. An OP-net is *consistent* if it has at least one consistent outcome and it has a model.

Encoding Preferences with OP-Nets

In [9], conditional preferences were generalized to Datalog+/− as follows. Let $Dom^+(X)$ be the set of all (possibly non-ground) atoms $p(t_1,\ldots,t_k)$ with $p = pred(X)$. An *ontological conditional preference* φ is of the form $v : \xi \succ \xi'$, where

- $v \in Dom^+(U_\varphi)$ for some $U_\varphi \subseteq \mathcal{X}_O$ is the context, and
- $\xi,\xi' \in Dom^+(X_\varphi)$ for some $X_\varphi \in \mathcal{X}_O - U_\varphi$.

A *ground instance* $v\theta : \xi\theta \succ \xi'\theta$ of φ is obtained via a substitution θ such that $v\theta \in Dom(U_\varphi)$ and $\xi\theta, \xi'\theta \in Dom(X_\varphi)$. Under suitable acyclicity conditions, one can construct an OP-net (O,Γ) that respects all ground instances of some given ontological conditional preferences.

Example 4. Consider the ontological conditional preference $specs(I,C,F,o)$: $vendor(V_1,p) \succ vendor(V_2,n)$, i.e., for an old car, it is preferable to have a vendor with positive feedback. One ground instance for this preference is $specs(t_1)$: $vendor(t_{10}) \succ vendor(t_{11})$. We could choose $\pi(vendor(t_{10})|specs(t_1)) = 1$ and $\pi(vendor(t_{11})|\ specs(t_1)) = \alpha < 1$ to encode this in an OP-net ■

Although possibilistic networks are less expressive than conditional preference theories (CP-theories) [3,13], they allow for a more compact encoding of conditional preferences over ground atoms and have lower complexity.

Query Answering under OP-Nets

The notions of skyline and k-rank answers are defined in the same way as for OCP-theories [9]. In a conjunctive query (CQ), a variable X of the OP-net may be used to annotate an atom over the predicate $pred(X)$. Hence, an *answer (tuple)* **a** to a CQ q w.r.t. an outcome o is an assignment of the distinguished variables that can be used to satisfy q in such a way that the marked atoms of q evaluate to the ones given by o. A *skyline answer* is an answer w.r.t. an undominated outcome of the OP-net. *CQ skyline membership* is the problem of deciding whether a given tuple is a skyline answer. Similarly, one can define *k-rank* answers as the k "most preferred" answers, i.e., those resulting from the outcomes with the highest possibilities.

Example 5. Consider the consistent OP-net (O,Γ) of Example 3 and the CQ $q(C,F,T,N) = \exists I\ specs(I,C,F,T) \wedge feature(F,N)$. Then, $\langle$b, f_1, o, ac$\rangle$ is the skyline answer under the consistent outcome $\mathbf{t}_1 \wedge \mathbf{t}_{10} \wedge \mathbf{t}_7$. The skyline answer for $q'(C,T) = \exists N\ q(C,f_2,T,N)$ is $\langle$c, n$\rangle$ with possibility $\pi(\mathbf{t}_2\mathbf{t}_{10}\mathbf{t}_8) = 0.5 \cdot 1 \cdot 0.7 = 0.35$, while the 2-rank answer is $\langle\langle$c, n$\rangle$, $\langle$c, o$\rangle\rangle$. Hence, if feature f_2 is mandatory, the offered new and colorful car is preferred over the old and colorful one, mainly due to positive feedback about vendor v_1. ■

Table 1. Complexity of deciding consistency of OP-nets

Class	Comb.	ba-comb.	fp-comb.	Data
L, LF, AF	PSPACE	D_2^P	D^P	in AC^0
G	2EXP	EXP	D^P	P
WG	2EXP	EXP	EXP	EXP
S, SF	EXP	D_2^P	D^P	in AC^0
F, GF	EXP	D_2^P	D^P	P
WS, WA	2EXP	2EXP	D^P	P

Computational Complexity

We now analyze the computational complexity of the consistency and CQ skyline membership problems for OP-nets. We assume familiarity with the complexity classes AC^0, P, NP, co-NP, Δ_2^P, Σ_2^P, Π_2^P, Δ_3^P, PSPACE, EXP, and 2EXP. The class $D^P = NP \wedge co\text{-}NP$ (resp., $D_2^P = \Sigma_2^P \wedge \Pi_2^P$) is the class of all problems that are the intersection of a problem in NP (resp., Σ_2^P) and a problem in co-NP (resp., Π_2^P).

Following Vardi's taxonomy [12], the *combined complexity* is calculated by considering all the components, i.e., the database, the set of dependencies, and the query, as part of the input. The *bounded-arity combined (ba-combined) complexity* assumes that the arity of the underlying schema is bounded by a constant. For example, in description logics (DLs) [4], the arity is always bounded by 2. The *fixed-program combined (fp-combined) complexity* is calculated by considering the set of TGDs and NCs as fixed. Finally, for *data complexity*, we take only the size of the database into account.

Although CQ answering in Datalog+/− is undecidable in general, there exist many syntactic conditions that guarantee decidability. We refer the reader to [6] for a short overview of the classes of acyclic (A), guarded (G), and sticky (S) sets of TGDs, their "weak" counterparts WA, WG, and WS, linear TGDs (L), full TGDs (F), and the combinations AF, GF, SF, and LF.

Our complexity results for the consistency and the CQ skyline membership problems for OP-nets are compactly summarized in Tables 1 and 2, respectively. Compared to OCP-theories [9], we obtain lower complexities for L, LF, AF, G, S, F, GF, SF, WS, and WA in the fp-combined complexity (completeness for D^P and Δ_2^P, respectively, rather than PSPACE), and for L, LF, AF, S, F, GF, and SF in the ba-complexity (completeness for D_2^P and Δ_3^P, respectively, rather than PSPACE).

Theorem 6. *Let $\mathcal{T}$ be a class of OP-nets (O,Γ). If checking non-emptiness of the answer set of a CQ w.r.t. O is in a complexity class $\mathcal{C}$, then consistency in $\mathcal{T}$ is in $NP^{\mathcal{C}} \wedge co\text{-}NP^{\mathcal{C}}$ and CQ skyline membership in $\mathcal{T}$ is in $P^{NP^{\mathcal{C}}}$. If $\mathcal{C} = NP$ and we consider the fp-combined complexity, then consistency in $\mathcal{T}$ is in D^P and CQ skyline membership in $\mathcal{T}$ is in Δ_2^P.*

In particular, for $\mathcal{C} = PSPACE$, we obtain inclusion in PSPACE for both problems, and the same for any deterministic complexity class above PSPACE. For $\mathcal{C} = NP$, we get the classes D_2^P and Δ_3^P. The lower bounds PSPACE and above

Table 2. Complexity of deciding CQ skyline membership for OP-nets

Class	Comb.	ba-comb.	fp-comb.	Data
L, LF, AF	PSPACE	Δ_3^{P}	Δ_2^{P}	in AC^0
G	2EXP	EXP	Δ_2^{P}	P
WG	2EXP	EXP	EXP	EXP
S, SF	EXP	Δ_3^{P}	Δ_2^{P}	in AC^0
F, GF	EXP	Δ_3^{P}	Δ_2^{P}	P
WS, WA	2EXP	2EXP	Δ_2^{P}	P

follow from consistency and equivalence of outcomes being as powerful as checking entailment of arbitrary ground CQs. The remaining lower bounds for the (fp-/ba-)combined complexity hold already if only NCs are allowed, and are shown by reductions from variants of the validity problem for QBFs. For example, the problem of deciding, given a valid formula $\exists \mathbf{X} \forall \mathbf{Y}\, \varphi(\mathbf{X}, \mathbf{Y})$ where $\varphi(\mathbf{X}, \mathbf{Y})$ is a propositional 3-DNF formula, whether the lexicographically maximal satisfying truth assignment for $\mathbf{X} = \{x_1, \ldots, x_n\}$ maps x_n to *true* is Δ_3^{P}-complete [11].

Finally, we can show that tractability in data complexity for deciding consistency and CQ skyline membership for OP-nets carries over from classical CQ answering. Here, data complexity means that Σ and the variables and possibility distributions of Γ are both fixed, while D is part of the input.

Theorem 7. *Let $\mathcal{T}$ be a class of OP-nets (O, Γ) for which CQ answering in O is possible in polynomial time (resp., in AC^0) in the data complexity. Then, deciding consistency and CQ skyline membership in $\mathcal{T}$ is possible in polynomial time (resp., in AC^0) in the data complexity.*

The listed P-hardness results hold due to a standard reduction of propositional logic programming to guarded full TGDs. These results do not apply to WG, where CQ answering is data complete for EXP, and data hardness holds even for ground atomic CQs; however, data completeness for EXP can be proved similarly to the results for combined complexity above.

We want to emphasize that our complexity results are generic, applying also to Datalog+/− languages beyond the ones listed. Even more, they are valid for arbitrary preference formalisms for which dominance between two outcomes can be decided in polynomial time, e.g., combinations of Datalog+/− with rankings computed by information retrieval methods [10].

Interesting topics of ongoing and future research include the implementation and experimental evaluation of the presented approach, as well as a generalization based on possibilistic logic [3] to gain more expressivity and some new features towards non-monotonic reasoning [1]; moreover, an apparent relation between possibilistic logic and quantitative choice logic [2] may be exploited.

Acknowledgments. This work was supported by a Google European Doctoral Fellowship, by the UK EPSRC grants EP/J008346/1, EP/L012138/1, and EP/M025268/1, and by the grant PON02_00563_3470993 ("VINCENTE").

References

1. Ben Amor, N., Dubois, D., Gouider, H., Prade, H.: Possibilistic networks: A new setting for modeling preferences. In: Proc. SUM, pp. 1–7. Springer (2014)
2. Benferhat, S., Brewka, G., Le Berre, D.: On the relationship between qualitative choice logic and possibilistic logic. In: Proc. IPMU. pp. 951–957 (2004)
3. Benferhat, S., Dubois, D., Garcia, L., Prade, H.: On the transformation between possibilistic logic bases and possibilistic causal networks. Int. J. Approx. Reas. 29(2), 135–173 (2002)
4. Bienvenu, M., Ortiz, M.: Ontology-mediated query answering with data-tractable description logics. In: Reasoning Web, pp. 218–307. Springer (2015)
5. Borgelt, C., Kruse, R.: Operations and evaluation measures for learning possibilistic graphical models. Artif. Intell. 148, 385–418 (2003)
6. Borgwardt, S., Fazzinga, B., Lukasiewicz, T., Shrivastava, A., Tifrea-Marciuska, O.: Preferential query answering in the Semantic Web with possibilistic networks. In: Proc. IJCAI. AAAI Press (2016), to appear, available at https://db.tt/GNqy1xDC.
7. Boutilier, C., Brafman, R.I., Domshlak, C., Hoos, H.H., Poole, D.: CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. J. Artif. Intell. Res. 21, 135–191 (2004)
8. Calì, A., Gottlob, G., Lukasiewicz, T.: A general datalog-based framework for tractable query answering over ontologies. J. Web Semant. 14, 57–83 (2012)
9. Di Noia, T., Lukasiewicz, T., Martinez, M.V., Simari, G.I., Tifrea-Marciuska, O.: Combining existential rules with the power of CP-theories. In: Proc. IJCAI. AAAI Press (2015)
10. Joachims, T.: Optimizing search engines using clickthrough data. In: Proc. SIGKDD. pp. 133–142 (2002)
11. Krentel, M.W.: Generalizations of Opt P to the polynomial hierarchy. Theor. Comput. Sci. 97, 183–198 (1992)
12. Vardi, M.Y.: The complexity of relational query languages. In: Proc. STOC. pp. 137–146. ACM (1982)
13. Wilson, N.: Extending CP-nets with stronger conditional preference statements. In: Proc. AAAI. pp. 735–741. AAAI Press (2004)