Answering $\mathcal{EL}$ Queries in the Presence of Preferences

İsmail İlkan Ceylan[1]*, Thomas Lukasiewicz[2], and Rafael Peñaloza[3]**

[1] Theoretical Computer Science, TU Dresden, Germany
ceylan@tcs.inf.tu-dresden.de
[2] Department of Computer Science, University of Oxford, UK
Thomas.Lukasiewicz@cs.ox.ac.uk
[3] KRDB Research Centre, Free University of Bozen-Bolzano, Italy
rafael.penaloza@unibz.it

Conjunctive query (CQ) answering is an important reasoning task in description logics (DLs). Its goal is to retrieve the tuples of individuals that satisfy a conjunctive query; i.e., a finite set of atomic queries. These tuples are called answers. Clearly, a given CQ may have a considerable number of answers, specially if the set of individual names appearing in the ABox is large, as is the case for many existing DL ontologies. In order to manage all these answers in a structural manner, one can try to extend query answering with preference criteria, in such a way that the most preferred answers are returned first.

Possibilistic networks (PNs) have arisen as a way of representing conditional preferences over a finite set of events in a compact way [1]. The general idea is to provide a possibility degree to each conditional event which is proportional to the preference given to that event. We apply this idea to model the preferences of query answers indirectly, by modeling the preferences over the contexts that entail them. In a nutshell, we divide an $\mathcal{EL}$ knowledge base (KB) into *contexts*, and use a possibilistic network to describe the joint possibility distribution over these contexts. Our formalism is based on ideas previously presented for reasoning under probabilistic uncertainty described by a Bayesian network [3]. The preference of an answer to the query is the possibility degree of the best context that entails this answer. Dually, we also compute, given a query, the most preferred source; that is, the context with the highest degree that entails this query.

Similar to Bayesian networks [4], PNs are graphical models providing a compact representation of a discrete possibility distribution, through some independence assumptions [2]. A *possibility distribution* over a set Ω is a function $\mathsf{Pos} : \Omega \to [0,1]$ that intuitively provides a degree of how possible is an event $\omega \in \Omega$ to happen. This function is extended to sets $\Gamma \subseteq \Omega$ by defining $\mathsf{Pos}(\Gamma) = \sup_{\omega \in \Gamma} \mathsf{Pos}(\omega)$. The *product conditional distribution* which is defined by the equation $\mathsf{Pos}(\Gamma \cap \Theta) = \mathsf{Pos}(\Gamma \mid \Theta) \cdot \mathsf{Pos}(\Theta)$.

Possibilistic networks decompose a possibility distribution into a product of conditional probability distributions that depend on the structure of a graph. A

* Supported by DFG within the Research Training Group "RoSI" (GRK 1907).
** The work was developed while the author was still affiliated with TU Dresden and the author has been partially supported by the Cluster of Excellence "cfAED".

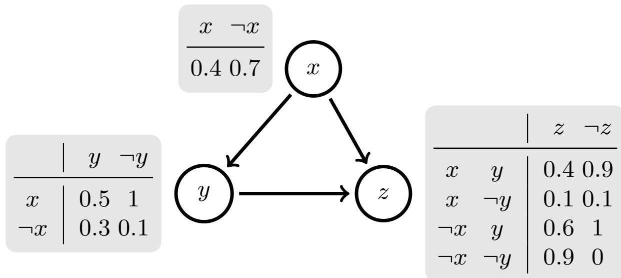

Fig. 1. A possibilistic network over $V_0 = \{x, y, z\}$

possibilistic network (PN) is a pair $\mathcal{P} = (G, \Phi)$, where $G = (V, E)$ is a DAG, and Φ contains a conditional possibility distribution $\mathsf{Pos}_{\mathcal{P}}(x \mid \mathsf{pa}(x))$ of every variable $x \in V$ given its parents $\mathsf{pa}(x)$ (see Figure 1). This PN defines the joint possibility distribution over the valuations of the variables in V

$$\mathsf{Pos}_{\mathcal{P}}(V) = \prod_{x \in V} \mathsf{Pos}_{\mathcal{P}}(x \mid \mathsf{pa}(x)).$$

Let V be a fixed but arbitrary finite set of propositional variables. A V*-context* is a propositional formula over V. A V*-GCI* is of the form $\langle C \sqsubseteq D : \varphi \rangle$ with C, D concepts and φ a V-context. A V*-TBox* is a finite set of V-GCIs. V*-assertions* are of the form $\langle C(a) : \varphi \rangle$ or $\langle r(a, b) : \varphi \rangle$ where $r \in \mathsf{N_C}$, $a, b \in \mathsf{N_I}$, C is a concept and φ is a V-context. A V*-ABox* is a finite set of V-assertions. A $\mathcal{PEL}$ KB is a tuple $\mathcal{K} = (\mathcal{P}, \mathcal{T}, \mathcal{A})$ where $\mathcal{P}$ is a PN over V, $\mathcal{T}$ is a V-TBox and $\mathcal{A}$ is a V-ABox.

The semantics of this logic is defined using multiple worlds. A *contextual interpretation* is a pair $(\mathcal{I}, \mathcal{W})$ where $\mathcal{I}$ is an $\mathcal{EL}$ interpretation and $\mathcal{W}$ is a valuation of the variables in V. $(\mathcal{I}, \mathcal{W})$ *satisfies* the axiom $\langle \lambda : \varphi \rangle$ ($(\mathcal{I}, \mathcal{W}) \models \langle \lambda : \varphi \rangle$), iff either (i) $\mathcal{W} \not\models \varphi$, or (ii) $\mathcal{I} \models \lambda$. It is a *model* of the $\mathcal{PEL}$ TBox $\mathcal{T}$ (resp. ABox $\mathcal{A}$) iff it satisfies all the axioms in $\mathcal{T}$ (resp. $\mathcal{A}$). A *possibilistic interpretation* is a pair $\mathfrak{P} = (\mathfrak{I}, \mathsf{Pos})$, where $\mathfrak{I}$ is a finite set of contextual interpretations and Pos is a possibility distribution over $\mathfrak{I}$. $\mathfrak{P}$ is a *model* of the $\mathcal{PEL}$ TBox $\mathcal{T}$ (resp. ABox $\mathcal{A}$) if every $(\mathcal{I}, \mathcal{W}) \in \mathfrak{I}$ is a model of $\mathcal{T}$ (resp. $\mathcal{A}$). $\mathfrak{P}$ is a *model* of the PN $\mathcal{P}$ if for every valuation $\mathcal{W}$,

$$\max_{(\mathcal{I}, \mathcal{W}) \in \mathfrak{I}} \mathsf{Pos}(\mathcal{I}, \mathcal{W}) = \mathsf{Pos}_{\mathcal{P}}(\mathcal{W}).$$

$\mathfrak{P}$ is a *model* of the $\mathcal{PEL}$ KB $\mathcal{K} = (\mathcal{P}, \mathcal{T}, \mathcal{A})$ iff it is a model of $\mathcal{T}$, $\mathcal{A}$, and $\mathcal{P}$.

Each possibilistic interpretation $\mathfrak{P} = (\mathfrak{I}, \mathsf{Pos})$ defines a possibility distribution $\mathsf{Pos}_{\mathfrak{P}}$ over all CQs given by $\mathsf{Pos}_{\mathfrak{P}}(\mathsf{q}) := \max_{(\mathcal{I}, \mathcal{W}) \in \mathfrak{I},\ \mathcal{I} \models \mathsf{q}} \{\mathsf{Pos}(\mathcal{I}, \mathcal{W})\}$. The *entailment degree* of q w.r.t. the $\mathcal{PEL}$ KB $\mathcal{K}$ is

$$\mathsf{Pos}_{\mathcal{K}}(\mathsf{q}) := \inf_{\mathfrak{P} \models \mathcal{K}} \{\mathsf{Pos}_{\mathfrak{P}}(\mathsf{q})\}.$$

These possibility distributions are extended to contexts in the obvious way, by setting $\mathsf{Pos}_{\mathfrak{P}}(\varphi) := \mathsf{Pos}_{\mathcal{P}}(\varphi) = \max_{\mathcal{W} \models \varphi} \mathsf{Pos}_{\mathcal{P}}(\mathcal{W})$. We can then define the con-

Table 1. $\mathcal{PEL}$ reasoning problems and their complexity

Problem	data	KB	network	combined
p-entailment	P	P	NP-c	NP-c
top-k answer	P	P	Δ_2^p-c	Δ_2^p-c
conditional top-k answer	P	P	Δ_2^p-c	Δ_2^p-c
k most preferred worlds	P	P	coNP-c	coNP-c

ditional possibilities of a query given a context, and of a context given a query, using the standard product rule. Formally,

$$\mathsf{Pos}_{\mathcal{K}}(\mathsf{q} \wedge \varphi) = \mathsf{Pos}_{\mathcal{K}}(\mathsf{q} \mid \varphi)\mathsf{Pos}_{\mathcal{K}}(\varphi),$$
$$\mathsf{Pos}_{\mathcal{K}}(\mathsf{q} \wedge \varphi) = \mathsf{Pos}_{\mathcal{K}}(\varphi \mid \mathsf{q})\mathsf{Pos}_{\mathcal{K}}(\mathsf{q}),$$

where

$$\mathsf{Pos}_{\mathcal{K}}(\mathsf{q} \wedge \varphi) = \inf_{(\mathfrak{I},\mathsf{Pos})\models\mathcal{K}} \left\{ \max_{\mathcal{I}\models\mathsf{q},\ \mathcal{W}\models\phi} \mathsf{Pos}(\mathcal{I},\mathcal{W})\} \right\}.$$

We consider three main reasoning problems in this setting; namely, deciding p-entailment, retrieving the top-k answers to a query, and the k most preferred worlds entailing a given query. We formally define these problems next. The problem of *p-entailment* refers to deciding whether $\mathsf{Pos}_{\mathcal{K}}(\mathsf{q}) \geq p$ for some given $p \in (0, 1]$. The *top-k answer* problem consists in deciding whether a tuple $(\mathfrak{a}_1, \dots, \mathfrak{a}_k)$ of different answers to q w.r.t. $\mathcal{K}$ is such that (i) for all $i, 1 \leq i < k$, $\mathsf{Pos}_{\mathcal{K}}(\mathfrak{a}_i) \geq \mathsf{Pos}_{\mathcal{K}}(\mathfrak{a}_{i+1})$, and (ii) for every other answer $\mathfrak{a}$, $\mathsf{Pos}_{\mathcal{K}}(\mathfrak{a}_k) \geq \mathsf{Pos}_{\mathcal{K}}(\mathfrak{a})$. This problem can be generalized to consider additional contextual evidence; that is, verify whether $(\mathfrak{a}_1, \dots, \mathfrak{a}_k)$ are the top-k answers to q *given* the context φ. Finally, the *k most preferred worlds* problem is the problem of deciding whether a tuple of k valuations of the variables V $(\mathcal{W}_1, \dots, \mathcal{W}_k)$ is such that $\mathsf{Pos}_{\mathcal{K}}(\mathcal{W}_i \mid \mathsf{q}) \geq \mathsf{Pos}_{\mathcal{K}}(\mathcal{W}_{i+1} \mid \mathsf{q})$ holds for all $i, 1 \leq i < k$, and there exists no other valuation $\mathcal{W}$ such that $\mathsf{Pos}_{\mathcal{K}}(\mathcal{W} \mid \mathsf{q}) > \mathsf{Pos}_{\mathcal{K}}(\mathcal{W}_k \mid \mathsf{q})$.

The complexity of all these problems is summarized in Table 1, where network complexity refers to the complexity considering only the size of the PN as input, KB complexity considers the size of the ABox and TBox, while combined complexity considers the whole KB together with the PN and the query as the size of the input. As it can be seen, all the problems remain tractable w.r.t. data and KB complexity, but the complexity increases as soon as the PN or the query is considered part of the input. This corresponds to the behaviour exhibited by query answering in the classical $\mathcal{EL}$ [5]. The full details of these results can be found in the appendix.

Although all the complexity bounds are tight, they are all based on performing black-box query entailment tests on $\mathcal{EL}$ KBs. As future work we plan to adapt specific query answering techniques to produce effective algorithms that can be used in practice. We will also extend our framework to other kinds of standard and non-stardard reasoning tasks.

References

1. BenAmor, N., Dubois, D., Gouider, H., Prade, H.: Possibilistic Networks : A New Setting for Modeling Preferences. In: Proc. of SUM'14. LNCS, vol. 8720. Springer Verlag (2014)
2. Benferhat, S., Dubois, D., Garcia, L., Prade, H.: Possibilistic logic bases and possibilistic graphs. In: Proc. of UAI'99. Morgan-Kaufmann Publishers (1999)
3. Ceylan, İ.İ., Peñaloza, R.: The Bayesian Description Logic $\mathcal{BEL}$. In: Proc. of IJCAR'14. LNCS, vol. 8562. Springer Verlag (2014)
4. Darwiche, A.: Modeling and Reasoning with Bayesian Networks. Cambridge University Press (2009)
5. Rosati, R.: On conjunctive query answering in EL. In: Proc. of DL'07. CEUR Workshop Proceedings, vol. 250. CEUR-WS (2007)