Bayesian Description Logics

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

[1] Theoretical Computer Science, TU Dresden, Germany
[2] Center for Advancing Electronics Dresden
{ceylan,penaloza}@tcs.inf.tu-dresden.de

Abstract. We present Bayesian Description Logics (BDLs): an extension of Description Logics (DLs) with contextual probabilities encoded in a Bayesian network (BN). Classical DL reasoning tasks are extended to consider also the contextual and probabilistic information in BDLs. A complexity analysis of these problems shows that, for propositionally closed DLs, this extension comes without cost, while for tractable DLs the complexity is affected by the cost of reasoning in the BN.

1 Introduction

Description logic (DL) ontologies are usually composed of axioms that restrict the class of possible interpretations. As these are hard restrictions, DL ontologies can only encode absolute, immutable knowledge. For some application domains, however, knowledge depends on the situation (or context) in which it is considered. For example, the notion of a *luxury hotel* in a small rural center will be different from the one in a large cosmopolis. When building an ontology for hotels, it makes sense to contextualize the axioms according to location, and possibly other factors like season, type of weather, etc. Since these contexts refer to notions that are external to the domain of interest, it is not always desirable, or even possible, to encode them directly into the classical DL axioms.

To handle contextual knowledge, we label every axiom with the context in which it holds. For example, ⟨LuxuryHotel ⊑ ∃hasFeature.MeetingRoom : city⟩ states that in the context of a city, every luxury hotel has a meeting room. This axiom imposes no restriction in case the context is not a city: it might still hold, or not, depending on other factors.

Labeling the axioms in an ontology allows us to give a more detailed description of the knowledge domain. Reasoning in these cases can be used to infer knowledge that is guaranteed to hold in any given context. While the knowledge *within* this context is precise, there might be a level of uncertainty regarding the current context. To model this uncertainty, we attach a probability to each of the possible contexts. Since we cannot assume that the contexts are (probabilistically) independent, we need to describe the joint probability distribution
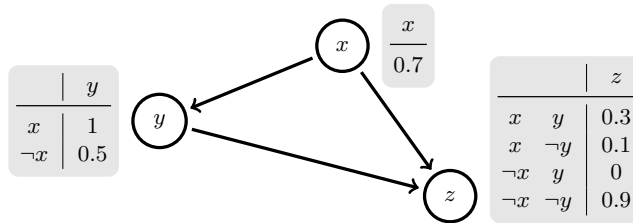
Fig. 1: The BN $\mathcal{B}_0$ over $V_0 = \{x, y, z\}$

over the space of all contexts. Thus, we consider knowledge bases that are composed of an ontology labeled with contextual information, together with a joint probability distribution over the space of contexts.

To represent the probabilistic component of the knowledge base, we use Bayesian networks (BNs) [12], a well-known probabilistic graphical model that allows for a compact representation of the probability distribution, with the help of conditional independence assumptions. For the logical component, we focus on classical DLs where the consequence relation is monotonic.

In previous work we have studied the Bayesian extension of the DL $\mathcal{EL}$ [5–7]. Here, we extend those ideas to arbitrary DLs and (monotonic) reasoning tasks.

2 Bayesian Description Logics

We now introduce Bayesian DLs. These logics extend classical DLs to handle uncertainty, expressed through a Bayesian network [12]. Formally, a *Bayesian network* (BN) is a pair $\mathcal{B} = (G, \Phi)$, where $G = (V, E)$ is a finite directed acyclic graph (DAG) whose nodes represent Boolean random variables,[3] and Φ contains, for every $x \in V$, a conditional probability distribution $P_{\mathcal{B}}(x \mid \pi(x))$ of x given its parents $\pi(x)$. If V is the set of nodes in G, we say that $\mathcal{B}$ is a BN *over* V.

Intuitively, $G = (V, E)$ encodes a series of conditional independence assumptions between the random variables. More precisely, every variable $x \in V$ is conditionally independent of its non-descendants given its parents. Thus, every BN $\mathcal{B}$ defines a unique joint probability distribution (JPD) over V given by

$$P_{\mathcal{B}}(V) = \prod_{x \in V} P_{\mathcal{B}}(x \mid \pi(x)).$$

A very simple BN is shown in Figure 1. Every node of the graph has an associated conditional probability distribution table. For the node x, the probability distribution is not conditional, since x has no parents in $\mathcal{B}_0$, and for z the table describes the conditional probability of z given its two parents x and y. From this network we can derive e.g. $P(x, \neg y, z) = P(z \mid x, \neg y) \cdot P(\neg y \mid x) \cdot P(x) = 0$.

---

[3] In their general form, BNs allow for arbitrary discrete random variables. We restrict w.l.o.g. to Boolean variables for ease of presentation.

We abstract from the specific DL used, and consider an arbitrary ontology language $\mathcal{L}$ that defines possibly infinite classes (i) $\mathfrak{A}$ of well-formed *axioms*; (ii) $\mathfrak{C}$ of *consequences*; and (iii) $\mathfrak{O}$ of *ontologies* such that every element of $\mathfrak{O}$ is a finite subset of $\mathfrak{A}$, and if $\mathcal{O} \in \mathfrak{O}$, then every subset of $\mathcal{O}$ is also an ontology in $\mathfrak{O}$. The semantics of this language is given by a class $\mathfrak{I}$ of *interpretations* and an *entailment relation* $\models \subseteq \mathfrak{I} \times (\mathfrak{A} \cup \mathfrak{C})$ that expresses which interpretations satisfy which axioms and consequences. We say that an interpretation $\mathcal{I}$ is a *model* of the ontology $\mathcal{O}$, denoted by $\mathcal{I} \models \mathcal{O}$ if $\mathcal{I} \models \alpha$ for all $\alpha \in \mathcal{O}$.[4] Finally, we say that $\mathcal{O}$ *entails* the consequence c, denoted by $\mathcal{O} \models c$ if every model of $\mathcal{O}$ satisfies c. Notice that the entailment relation is monotonic; i.e., if $\mathcal{O} \models c$ and $\mathcal{O} \subseteq \mathcal{O}' \in \mathfrak{O}$, then $\mathcal{O}' \models c$. Any DL with acyclic TBoxes, or with general ontologies is an ontology language of this kind; consequences in these languages are e.g. concept unsatisfiability, concept subsumption, or ontology inconsistency.

For the rest of this paper, we consider $\mathcal{L}$ to be an arbitrary but fixed ontology language, with axioms $\mathfrak{A}$, ontologies $\mathfrak{O}$, consequences $\mathfrak{C}$, and interpretations $\mathfrak{I}$. As an example scenario we use subsumption in $\mathcal{EL}$ with general TBoxes. In the Bayesian ontology language $\mathcal{BL}$, ontologies are generalized by annotating every axiom with a context, which is defined as a set of literals belonging to a BN.

Definition 1 (KB). *Let V be a finite set of Boolean variables. A V-literal is either x or $\neg x$, where $x \in V$; a V-context is a consistent set of V-literals.*

A V-axiom is of the form $\langle \alpha : \kappa \rangle$ where $\alpha \in \mathfrak{A}$ and κ is a V-context. A V-ontology is a finite set O of V-axioms, such that $\{\alpha \mid \langle \alpha : \kappa \rangle \in \mathsf{O}\} \in \mathfrak{O}$. A V-consequence is a pair $\langle c : \kappa \rangle$, where $c \in \mathfrak{C}$ and κ is a V-context.

A $\mathcal{BL}$ knowledge base (KB) over V is a pair $\mathcal{K} = (\mathcal{B}, \mathsf{O})$ where $\mathcal{B}$ is a BN over V and O is a V-ontology.[5]

Intuitively, a V-axiom is only guaranteed to hold when the context κ is satisfied. The semantics of this logic is defined by extending interpretations to evaluate the random variables from the BN.

Definition 2 (interpretation). *Given a finite set of Boolean variables V, a V-interpretation is a pair $\mathrm{I} = (\mathcal{I}, \mathcal{V}^{\mathrm{I}})$ where $\mathcal{I} \in \mathfrak{I}$ and $\mathcal{V}^{\mathrm{I}} : V \to \{0,1\}$ is a valuation of the variables in V.*

If there is no danger of ambiguity we will usually omit the prefix V. The valuation $\mathcal{V}^{\mathrm{I}}$ is extended to contexts by defining $\mathcal{V}^{\mathrm{I}}(\neg x) = 1 - \mathcal{V}^{\mathrm{I}}(x)$ for all $x \in V$, and

$$\mathcal{V}^{\mathrm{I}}(\kappa) = \min_{\ell \in \kappa} \mathcal{V}^{\mathrm{I}}(\ell),$$

where $\mathcal{V}^{\mathrm{I}}(\emptyset) := 1$ for every context κ. The V-interpretation $\mathrm{I} = (\mathcal{I}, \mathcal{V}^{\mathrm{I}})$ is a *model* of the V-axiom $\langle \alpha : \kappa \rangle$ or the V-consequence $\langle c : \kappa \rangle$, denoted as $\mathrm{I} \models \langle \alpha : \kappa \rangle$ (respectively $\mathrm{I} \models \langle c : \kappa \rangle$), iff $\mathcal{V}^{\mathrm{I}}(\kappa) = 0$ or $\mathcal{I} \models \alpha$ (resp., $\mathcal{I} \models c$). It is a *model* of the V-ontology O iff it is a model of all the V-axioms in O. The ontology O *entails* $\langle c : \kappa \rangle$ if every model of O is a model of $\langle c : \kappa \rangle$. The idea is that the

---

[4] We use the infix notation for the relation $\models$.
[5] Unless stated otherwise we assume that $\mathcal{K}$ is over V in the rest of the paper.

restriction α is only required to hold whenever the context κ is satisfied. Thus, any interpretation that violates the context trivially satisfies the whole V-axiom.

Example 3. Let $V_0 = \{x, y, z\}$, and consider the $\mathcal{BEL}$ V_0-ontology

$$\mathsf{O}_0 := \{\, \langle A \sqsubseteq C : \{x, y\}\rangle,\ \langle A \sqsubseteq B : \{\neg x\}\rangle,\ \langle B \sqsubseteq C : \{\neg x\}\rangle \,\}.$$

The interpretation $\mathrm{I}_0 = (\mathcal{I}_0, \mathcal{V}_0)$ where $\mathcal{V}_0(\{x, \neg y, z\}) = 1$, and $\mathcal{I}_0 = (\{d\}, \cdot^{\mathcal{I}_0})$ with $A^{\mathcal{I}_0} = \{d\}$, and $B^{\mathcal{I}_0} = C^{\mathcal{I}_0} = \emptyset$ is a model of O_0, but not of $\langle A \sqsubseteq B : \{x\}\rangle$.

$\mathcal{BL}$ generalizes $\mathcal{L}$: an ontology from $\mathcal{L}$ is a special case of a $\mathcal{BL}$ V-ontology, where all V-axioms are associated with the empty context; i.e., are of the form $\langle \alpha : \emptyset \rangle$. Since every valuation satisfies the empty context $\emptyset$, a V-interpretation $\mathrm{I} = (\mathcal{I}, \mathcal{V}^{\mathrm{I}})$ satisfies the V-axiom $\langle \alpha : \emptyset \rangle$ iff $\mathcal{I} \models \alpha$, and analogously for consequences $\langle c : \kappa \rangle$. For brevity, we use $\mathsf{O} \models c$ to denote that O entails $\langle c : \emptyset \rangle$. For a valuation $\mathcal{W}$ of the variables in V, we define the V-ontology containing the V-axioms that must be satisfied in any V-interpretation $\mathrm{I} = (\mathcal{I}, \mathcal{V}^{\mathrm{I}})$ with $\mathcal{V}^{\mathrm{I}} = \mathcal{W}$.

Definition 4 (restriction). *Let* $\mathcal{K} = (\mathcal{B}, \mathsf{O})$ *be a KB. The* restriction of O to *a valuation* $\mathcal{W}$ *of the variables in* V *is the* V*-ontology*

$$\mathsf{O}_{\mathcal{W}} := \{\langle \alpha : \emptyset \rangle \mid \langle \alpha : \kappa \rangle \in \mathsf{O}, \mathcal{W}(\kappa) = 1\}.$$

So far, our semantics have focused on the evaluation of the Boolean variables and the interpretation of concepts, ignoring the probabilistic information provided by the BN. To handle these probabilities, we introduce multiple-world semantics next. Intuitively, a V-interpretation describes a possible world; by assigning a probabilistic distribution over these interpretations, we describe the required probabilities, which should be consistent with the BN.

Definition 5 (probabilistic model). *A* probabilistic interpretation *is a pair* $\mathcal{P} = (\mathbb{I}, P_{\mathbb{I}})$, *where* $\mathbb{I}$ *is a set of* V*-interpretations and* $P_{\mathbb{I}}$ *is a probability distribution over* $\mathbb{I}$ *such that* $P_{\mathbb{I}}(\mathrm{I}) > 0$ *only for finitely many interpretations* $\mathrm{I} \in \mathbb{I}$. *It is a* model *of the* V*-ontology* O *(resp., of the* V*-consequence* $\langle c : \kappa \rangle$*) if every* $\mathrm{I} \in \mathbb{I}$ *is a model of* O *(resp., of* $\langle c : \kappa \rangle$*).* $\mathcal{P}$ *is* consistent *with the BN* $\mathcal{B}$ *if for every valuation* $\mathcal{W}$ *of the variables in* V *it holds that*

$$\sum_{\mathrm{I} \in \mathbb{I},\, \mathcal{V}^{\mathrm{I}} = \mathcal{W}} P_{\mathbb{I}}(\mathrm{I}) = P_{\mathcal{B}}(\mathcal{W}).$$

The probabilistic interpretation $\mathcal{P}$ *is a* model *of the KB* $(\mathcal{B}, \mathsf{O})$ *iff it is a (probabilistic) model of* O *and consistent with* $\mathcal{B}$.

As consequence of this semantics, probabilistic models preserve the probability distribution of $\mathcal{B}$ for subsets of literals; i.e., contexts. The proof follows from the fact that a context describes a partial valuation. Hence, the probability of a context κ is the sum of the probabilities of all valuations that extend κ. Formally, let $\mathcal{K} = (\mathcal{B}, \mathsf{O})$ be a KB, and κ a context. For every model $\mathcal{P}$ of $\mathcal{K}$ it holds that

$$\sum_{\mathrm{I} \in \mathbb{I},\, \mathcal{V}^{\mathrm{I}}(\kappa) = 1} P_{\mathbb{I}}(\mathrm{I}) = P_{\mathcal{B}}(\mathcal{V}^{\mathrm{I}}).$$

It is sometimes useful to consider a special kind of probabilistic interpretations, which we call *pithy*. These interpretations contain at most one V-interpretation for each valuation of the variables in V. Each of these V-interpretations provides the essential information associated to the corresponding valuation.

Definition 6 (pithy). *The probabilistic interpretation $\mathcal{P} = (\mathbb{I}, P_{\mathbb{I}})$ is called* pithy *if for every valuation $\mathcal{W}$ of the variables in V there exists at most one V-interpretation $\mathrm{I} = (\mathcal{I}, \mathcal{V}^{\mathrm{I}}) \in \mathbb{I}$ such that $\mathcal{V}^{\mathrm{I}} = \mathcal{W}$.*

We now study the main reasoning problems in $\mathcal{BL}$, and their complexity.

3 Reasoning in $\mathcal{BL}$

We have shown how to represent probabilistic knowledge using a $\mathcal{BL}$ KB. We now focus on reasoning with this knowledge. Recall that in the classical case an ontology $\mathcal{O}$ entails a consequence c iff every model of $\mathcal{O}$ is also a model of c. We generalize this notion to consider contexts and the probabilities from the BN.

Definition 7 (entailment). *Let $\langle c : \kappa \rangle$ be a V-consequence and $\mathcal{K}$ a $\mathcal{BL}$ KB. $\mathcal{K}$ entails $\langle c : \kappa \rangle$, denoted as $\mathcal{K} \models \langle c : \kappa \rangle$, if every probabilistic model of $\mathcal{K}$ is a model of $\langle c : \kappa \rangle$. For a probabilistic interpretation $\mathcal{P} = (\mathbb{I}, P_{\mathbb{I}})$, the* probability *of $\langle c : \kappa \rangle$ is $P_{\mathcal{P}}(\langle c : \kappa \rangle) := \sum_{\mathrm{I} \in \mathbb{I}, \mathrm{I} \models \langle c : \kappa \rangle} P_{\mathbb{I}}(\mathrm{I})$. The* probability *of $\langle c : \kappa \rangle$ w.r.t. $\mathcal{K}$ is*

$$P_{\mathcal{K}}(\langle c : \kappa \rangle) := \inf_{\mathcal{P} \models \mathcal{K}} P_{\mathcal{P}}(\langle c : \kappa \rangle).$$

We say that c is a positive entailment *in κ if $P_{\mathcal{K}}(\langle c : \kappa \rangle) > 0$, and a p-entailment in κ, $p \in (0, 1]$ if $P_{\mathcal{K}}(\langle c : \kappa \rangle) \geq p$. 1-entailments are also called* almost-sure.

Clearly, if $\mathcal{K} \models \langle c : \kappa \rangle$, then $P_{\mathcal{K}}(\langle c : \kappa \rangle) = 1$. The converse may not hold since the entailment relation might be violated in V-interpretations of probability zero.

Example 8. Consider the KB $\mathcal{K}_0 = (\mathcal{B}_0, \mathsf{O}_0)$, where $\mathcal{B}_0$ is the BN depicted in Figure 1 and O_0 the ontology from Example 3. Then, $P_{\mathcal{K}_0}(\langle A \sqsubseteq C : \emptyset \rangle) = 1$ and $P_{\mathcal{K}_0}(\langle C \sqsubseteq B : \{x, y\} \rangle) = 0$. Moreover, for any two concepts E, F, it holds that $P_{\mathcal{K}_0}(\langle E \sqsubseteq F : \{x, \neg y\} \rangle) = 1$ since $\langle E \sqsubseteq F : \{x, \neg y\} \rangle$ can only be violated in V-interpretations with probability 0. But, in general, $\mathcal{K}_0 \not\models \langle E \sqsubseteq F : \{x, \neg y\} \rangle$.

3.1 Probabilistic Entailments

We consider first the problem of computing the probability of an entailment, or deciding the properties of this probability. As an intermediate step, we show that it is possible w.l.o.g. to restrict reasoning to pithy models.

Lemma 9. *Let $\mathcal{K}$ be a KB. If $\mathcal{P}$ is a probabilistic model of $\mathcal{K}$, then there exists a pithy model $\mathcal{Q}$ of $\mathcal{K}$ such that for every consequence $\langle c : \kappa \rangle$ it holds that $P_{\mathcal{Q}}(\langle c : \kappa \rangle) \leq P_{\mathcal{P}}(\langle c : \kappa \rangle)$.*

Proof (Sketch). Let $\mathcal{W}$ be a valuation and $I, I' \in \mathbb{I}$ two V-interpretations such that $\mathcal{V}^I = \mathcal{V}^{I'} = \mathcal{W}$. Construct a new interpretation J as the disjoint union of I and I'. The probabilistic interpretation $(\mathbb{H}, P_{\mathbb{H}})$ with $\mathbb{H} = (\mathbb{I} \cup \{J\}) \setminus \{I, I'\}$ and

$$P_{\mathbb{H}}(H) := \begin{cases} P_{\mathbb{I}}(H) & H \neq J \\ P_{\mathbb{I}}(I) + P_{\mathbb{I}}(I') & H = J \end{cases}$$

is a model of $\mathcal{K}$. Moreover, $J \models \langle \alpha : \kappa \rangle$ iff both $I \models \langle \alpha : \kappa \rangle$ and $I' \models \langle \alpha : \kappa \rangle$. □

As we show next, the probability of a consequence can be computed by reasoning over the restrictions $O_{\mathcal{W}}$ of the V-ontolgy O.

Theorem 10. *Let $\mathcal{K} = (\mathcal{B}, O)$ be a KB and $\langle c : \kappa \rangle$ a consequence.*

$$P_{\mathcal{K}}(\langle c : \kappa \rangle) = 1 - P_{\mathcal{B}}(\kappa) + \sum_{\substack{O_{\mathcal{W}} \models c \\ \mathcal{W}(\kappa)=1}} P_{\mathcal{B}}(\mathcal{W}).$$

Proof. For every valuation $\mathcal{W}$ construct the V-interpretation $I_{\mathcal{W}}$ as follows. If $O_{\mathcal{W}} \models c$, then $I_{\mathcal{W}}$ is any model $(\mathcal{I}, \mathcal{W})$ of $O_{\mathcal{W}}$; otherwise, $I_{\mathcal{W}}$ is any model $(\mathcal{I}, \mathcal{W})$ of $O_{\mathcal{W}}$ that does not satisfy $\langle c : \kappa \rangle$, which must exist by definition. The probabilistic interpretation $\mathcal{P}_{\mathcal{K}} = (\mathbb{I}, P_{\mathbb{I}})$ with $\mathbb{I} = \{I_{\mathcal{W}} \mid \mathcal{W} \text{ a valuation of } V\}$ and $P_{\mathbb{I}}(I_{\mathcal{W}}) = P_{\mathcal{B}}(\mathcal{W})$ for all $\mathcal{W}$ is a model of $\mathcal{K}$ and

$$P_{\mathcal{P}_{\mathcal{K}}}(\langle c : \kappa \rangle) = \sum_{I_{\mathcal{W}} \models \langle c : \kappa \rangle} P_{\mathbb{I}}(I_{\mathcal{W}}) = \sum_{\mathcal{W}(\kappa)=0} P_{\mathbb{I}}(I_{\mathcal{W}}) + \sum_{\substack{\mathcal{W}(\kappa)=1, \\ I_{\mathcal{W}} \models \langle c : \kappa \rangle}} P_{\mathbb{I}}(I_{\mathcal{W}})$$

$$= 1 - P_{\mathcal{B}}(\kappa) + \sum_{\substack{O_{\mathcal{W}} \models c \\ \mathcal{W}(\kappa)=1}} P_{\mathcal{B}}(\mathcal{W}).$$

Thus, $P_{\mathcal{K}}(\langle c : \kappa \rangle) \leq 1 - P_{\mathcal{B}}(\kappa) + \sum_{O_{\mathcal{W}} \models c, \mathcal{W}(\kappa)=1} P_{\mathcal{B}}(\mathcal{W})$. Suppose now that the inequality is strict, then there exists a probabilistic model $\mathcal{P} = (\mathbb{J}, P_{\mathbb{J}})$ of $\mathcal{K}$ such that $P_{\mathcal{P}}(\langle c : \kappa \rangle) < P_{\mathcal{P}_{\mathcal{K}}}(\langle \alpha : \kappa \rangle)$. By Lemma 9, we can assume w.l.o.g. that $\mathcal{P}$ is pithy, and hence for every valuation $\mathcal{W}$ with $P_{\mathcal{B}}(\mathcal{W}) > 0$ there exists exactly one $J_{\mathcal{W}} \in \mathbb{J}$ with $\mathcal{V}^{J_{\mathcal{W}}} = \mathcal{W}$. We thus have

$$\sum_{J_{\mathcal{W}} \models \langle c : \kappa \rangle, \mathcal{W}(\kappa)=1} P_{\mathbb{J}}(J_{\mathcal{W}}) < \sum_{I_{\mathcal{W}} \models \langle c : \kappa \rangle, \mathcal{W}(\kappa)=1} P_{\mathbb{I}}(I_{\mathcal{W}}).$$

Since $P_{\mathbb{I}}(I_{\mathcal{W}}) = P_{\mathbb{J}}(J_{\mathcal{W}})$ for all $\mathcal{W}$, there exists a valuation $\mathcal{W}'$ with $I_{\mathcal{W}'} \models \langle c : \kappa \rangle$ but $J_{\mathcal{W}'} \not\models \langle c : \kappa \rangle$. Since $J_{\mathcal{W}'}$ is a model of $O_{\mathcal{W}'}$ it follows that $O_{\mathcal{W}'} \not\models c$. By construction, then we have that $I_{\mathcal{W}'} \not\models \langle c : \kappa \rangle$, which is a contradiction. □

Based on this theorem, we can compute the probability of an entailment as described in Algorithm 1. The algorithm simply verifies for all possible valuations $\mathcal{W}$, whether $O_{\mathcal{W}}$ entails c. Clearly, the **for** loop is executed $2^{|V|}$ times; once for each valuation of the variables in V. Each of these executions needs to compute

---

Algorithm 1 Probability of an entailment

Input: KB $\mathcal{K} = (\mathcal{B}, \mathsf{O})$, consequence $\langle c : \kappa \rangle$
Output: $P_{\mathcal{K}}(\langle c : \kappa \rangle)$
1: $P \leftarrow 0, Q \leftarrow 0$
2: **for all** valuations $\mathcal{W}$ **do**
3:     **if** $\mathcal{W}(\kappa) = 0$ **then**
4:         $Q \leftarrow Q + P_{\mathcal{B}}(\mathcal{W})$
5:     **else if** $\mathsf{O}_{\mathcal{W}} \models c$ **then**
6:         $P \leftarrow P + P_{\mathcal{B}}(\mathcal{W})$
7: **return** $1 - Q + P$

---

the probability $P_{\mathcal{B}}(\mathcal{W})$ and, possibly, decide whether $\mathsf{O}_{\mathcal{W}} \models c$. The former can be done in polynomial time on the size of $\mathcal{B}$, using the standard chain rule [12], while the complexity of the latter depends on the complexity of deciding entailments in language $\mathcal{L}$, which we denote with $\Bbbk$. Notice that the different valuations can be enumerated using only polynomial space, and the algorithm handles them independently. This yields the following result.

Theorem 11. *Deciding p-entailment in $\mathcal{BL}$ is in* $\mathrm{PSPACE}^{\Bbbk}$.

As a lower bound, unsurprisingly, p-entailment is at least as hard as deciding entailment in $\mathcal{L}$, and deciding probabilities from the BN. Since this latter problem is hard for the class PP [19], we get the following result.

Theorem 12. *Deciding p-entailment is both* PP-*hard and* $\Bbbk$-*hard.*

From these two theorems it follows that if the complexity of deciding entailments in $\mathcal{L}$ is at least PSPACE, then p-entailment in $\mathcal{BL}$ is in the same complexity class. This means that any DL at least as expressive as $\mathcal{ALC}$ (without a TBox), can be extended with our probabilistic semantics, without affecting the complexity of reasoning. For tractable DLs like the $\mathcal{EL}$ and DL-Lite families, the complexity increases to PP. If we restrict only in deciding positive or almost-sure entailments, then these bounds can be lowered.

Theorem 13. *The complexity of deciding positive entailment is between* NP *and* $\mathrm{NP}^{\Bbbk}$; *deciding almost-sure entailments is between* coNP *and* $co\mathrm{NP}^{\Bbbk}$.

Proof. To decide positive entailment, we can simply guess a valuation $\mathcal{W}$ and check in polynomial time with a $\Bbbk$-oracle that (i) $P_{\mathcal{B}}(\mathcal{W}) > 0$ and (ii) either $\mathcal{W}(\kappa) = 0$ or $\mathsf{O}_{\mathcal{W}} \models c$. The correctness of this algorithm is given by Theorem 10. Thus the problem is in $\mathrm{NP}^{\Bbbk}$. To show hardness, we recall that deciding, given a BN $\mathcal{B}$ and a variable $x \in V$, whether $P_{\mathcal{B}}(x) > 0$ is NP-hard [9]. Consider the KB $\mathcal{K} = (\mathcal{B}, \emptyset)$ and $c \in \mathfrak{C}$. It follows from Theorem 10 that $P_{\mathcal{B}}(x) > 0$ iff $P_{\mathcal{K}}(\langle c : \{\neg x\} \rangle) > 0$. Thus positive consequence is NP-hard. The bounds for almost-sure entailments can be shown analogously. $\square$

This theorem implies that if $\Bbbk$ contains Δ_2^P, then positive and almost-sure reasoning in $\mathcal{BL}$ is in the same complexity class as reasoning in $\mathcal{L}$. Moreover, if $\Bbbk$ is contained in PTIME, then these problems are NP-complete and coNP-complete, respectively. It can also be easily seen from Algorithm 1 that if $\Bbbk$ is in PTIME, then all these probabilistic reasoning problems are fixed-parameter tractable where $|V|$ is the parameter.[6]

Algorithm 1 shows that the probabilistic and the logical components of the KB can be decoupled while reasoning. This is an encouraging result as it means that one can apply the optimized methods developed for BN inference and for DL reasoning directly in $\mathcal{BL}$ without mayor modifications.

3.2   Contextual Entailments

We now turn our attention to deciding whether a consequence follows from all models of the KB in a classical sense; that is, whether $\mathcal{K} \models \langle c : \kappa \rangle$ holds. This problem is intractable already for the sublogic of $\mathcal{BEL}$ that allows only the conjunction constructor, even if restricted to the empty context.

Theorem 14. *Let $\mathcal{K}$ be a $\mathcal{BEL}$ KB and C, D two concepts. Deciding whether $\mathcal{K} \models \langle C \sqsubseteq D : \emptyset \rangle$ is coNP-hard.*

Proof. We present a reduction from validity of DNF formulas, which is known to be coNP-hard [8]. Let $\phi = \sigma_1 \vee \ldots \vee \sigma_n$ be a DNF formula where each σ_i is a conjunctive clause and let V be the set of all variables appearing in ϕ. For each variable $x \in V$, we introduce the concept names B_x and $B_{\neg x}$ and define the TBox $\mathsf{O}_x := \{\langle A \sqsubseteq B_x : \{x\}\rangle, \langle A \sqsubseteq B_{\neg x} : \{\neg x\}\rangle\}$. For every conjunctive clause $\sigma = \ell_1 \wedge \ldots \wedge \ell_m$ define the TBox $\mathsf{O}_\sigma := \{\langle B_{\ell_1} \sqcap \ldots \sqcap B_{\ell_m} \sqsubseteq C : \emptyset \rangle\}$. Let now $\mathcal{K} = (\mathcal{B}, \mathsf{O})$ where $\mathcal{B}$ is an arbitrary BN over V and $\mathsf{O} = \bigcup_{x \in V} \mathsf{O}_x \cup \bigcup_{1 \leq i \leq n} \mathsf{O}_{\sigma_i}$. It is easy to see that ϕ is valid iff $\mathcal{K} \models \langle A \sqsubseteq C : \emptyset \rangle$. $\square$

The main reason for this hardness is that the interaction of contexts might produce consequences that are not obvious at first sight. For instance, a consequence might follow in context κ not because the axioms from O_κ entail the consequence, but rather because any valuation satisfying κ will yield it. That is the main idea in the proof of Theorem 14; the axioms that follow directly from the empty context never entail the axiom $A \sqsubseteq C$, but if ϕ is valid, then this follows from all valuations. We obtain the following result.

Lemma 15. *Let $\mathcal{K} = (\mathcal{B}, \mathsf{O})$ be a KB. Then $\mathcal{K} \models \langle c : \kappa \rangle$ iff for every valuation $\mathcal{W}$ with $\mathcal{W}(\kappa) = 1$, it holds that $\mathsf{O}_\mathcal{W} \models c$.*

It thus suffices to identify all valuations that define V-ontologies entailing the consequence. To do this, we will take advantage of techniques developed in the area of axiom-pinpointing [4], access control [1], and context-based reasoning [2].

---

[6] A problem is fixed-parameter tractable if it can be solved in polynomial time, assuming that the parameter is fixed [13].

Notice that a consequence depends only on the V-ontology and not on the BN. For that reason, for the rest of this section we focus only on this part of the KB.

We can see every context κ as the conjunctive clause $\chi_\kappa := \bigwedge_{\ell \in \kappa} \ell$. In this view, the V-ontology O is a set of labeled axioms over the distributive lattice $\mathbb{B}$ of all Boolean formulas over the variables V modulo equivalence. Each formula ϕ in this lattice defines a sub-ontology O_ϕ containing all axioms $\langle \alpha : \kappa \rangle \in \mathsf{O}$ with $\chi_\kappa \models \phi$.

Using the terminology from [2], we are interested in finding a boundary for a consequence. Given a V-ontology O labeled over the lattice $\mathbb{B}$ and a consequence c, a *boundary* for c w.r.t. O is an element $\phi \in \mathbb{B}$ such that for every join-prime element $\psi \in \mathbb{B}$ it holds that $\psi \models \phi$ iff $\mathsf{O}_\psi \models c$ (see [2] for further details). Notice that the join-prime elements of $\mathbb{B}$ are exactly the valuations of variables in V. Using Lemma 15 we obtain the following result.

Theorem 16. *Let ϕ be a boundary for c w.r.t. O in $\mathbb{B}$. Then, for any context κ we have that $\mathcal{K} \models \langle c : \kappa \rangle$ iff $\chi_\kappa \models \phi$.*

Several methods have been developed for computing boundaries, either through a *black-box* approach that makes several calls to an external reasoner [1, 2, 15], or through a *glass-box* approach that modifies a pre-existing reasoner [3, 4, 14]. In general, it is easy to see that a boundary can be computed by performing exponentially many entailment tests. Once a boundary ϕ for c w.r.t. O has been computed, contextual entailment is decided by verifying whether $\chi_\kappa \models \phi$. This decision is in NP on $|V|$. Thus, overall we have the following.

Corollary 17. *Contextual entailment in $\mathcal{BL}$ can be decided in* ExpTime$^\Bbbk$.

Clearly, a boundary for c provides more information than necessary for deciding whether the axiom holds in a *given* context κ. It encodes *all* contexts that entail the desired axiom. We can use this knowledge to deduce the most likely context.

3.3   Most Likely Context

The problem of finding the most likely context for a consequence can be seen as the dual of computing the probability of this consequence. Intuitively, we are interested in finding the most likely explanation for an event; assuming that a consequence holds, we are interested in finding an explanation for it, in the form of a context, that has the maximal probability of occurring.

Definition 18 (most likely context). *Given a KB $\mathcal{K} = (\mathcal{B}, \mathsf{O})$ and $c \in \mathfrak{C}$, the context κ is called a* most likely context *for c w.r.t. $\mathcal{K}$ if (i) $\mathcal{K} \models \langle c : \kappa \rangle$, and (ii) for every context κ', if $\mathcal{K} \models \langle c : \kappa' \rangle$, then $P_\mathcal{B}(\kappa') \leq P_\mathcal{B}(\kappa)$.*

Notice that we are not interested in maximizing $P_\mathcal{K}(\langle c : \kappa \rangle)$ but rather $P_\mathcal{B}(\kappa)$. Indeed, these two problems can be seen as dual, since $P_\mathcal{K}(\langle c : \kappa \rangle)$ depends inversely, but not exclusively, on $P_\mathcal{B}(\kappa)$ (see Theorem 10).

Algorithm 2 computes the set of all most likely contexts for c w.r.t. $\mathcal{K}$, together with their probability. It maintains a value p of the highest known prob-

---
Algorithm 2 Compute all most likely contexts

Input: KB $\mathcal{K} = (\mathcal{B}, \mathsf{O})$, consequence c
Output: The set Λ of most likely contexts for c w.r.t. $\mathcal{K}$ and probability $p \in [0, 1]$
 1: $\Lambda \leftarrow \emptyset,\ p \leftarrow 0$
 2: $\phi \leftarrow \mathsf{boundary}(c, \mathsf{O})$ $\triangleright$ compute a boundary for c w.r.t. O
 3: **for all** contexts κ **do**
 4:   **if** $\chi_\kappa \models \phi$ **then**
 5:     **if** $P_\mathcal{B}(\kappa) > p$ **then**
 6:       $\Lambda \leftarrow \{\kappa\}$
 7:       $p \leftarrow P_\mathcal{B}(\kappa)$
 8:     **else if** $P_\mathcal{B}(\kappa) = p$ **then**
 9:       $\Lambda \leftarrow \Lambda \cup \{\kappa\}$
10: **return** Λ, p
---

ability for a context, and a set Λ with all the contexts that have probability p. The algorithm first computes a boundary for the consequence, which is used to test, for every context κ whether $\mathcal{K} \models \langle c : \kappa \rangle$. In that case, it compares $P_\mathcal{B}(\kappa)$ with p. If the former is larger, then the highest probability is updated to this value, and the set Λ is restarted to contain only κ. If they are the same, then κ is added to the set of most likely contexts. The number of contexts is exponential on $\mathcal{B}$, and for each of them we have to test propositional entailment, which is also exponential on $\mathcal{B}$. The most expensive step is to compute the boundary, whose complexity depends on the complexity of reasoning.

Theorem 19. *Algorithm 2 computes all most likely contexts for c w.r.t. $\mathcal{K}$ in exponential time, with a $\mathbb{k}$ oracle.*

In general, it is not possible to lower this exponential upper bound, since a simple consequence may have exponentially many most likely contexts. However, Algorithm 2 can be adapted to compute *one* most likely context in a more efficient way. The main idea is to order the calls in the **for** loop by decreasing probability. Once a context κ with $\chi_\kappa \models \phi$ has been found, it is guaranteed to be a most likely context and the algorithm may stop. This approach would still require exponential time in the worst case, even with a $\mathbb{k}$ oracle, since it requires the computation of the boundary.

4   Related Work

The amount of work on handling uncertain knowledge with description logics is too vast to cover in detail here. Many probabilistic description logics have been defined, which differ not only in their syntax but also in their use of the probabilities and their application. These logics were recently surveyed in [17]. We discuss here only those logics most closely related to ours.

  One of the first attempts for combining BNs and DLs was P-CLASSIC [16], which extended CLASSIC through probability distributions over the interpretation domain. The more recent PR-OWL [10] uses multi-entity BNs to describe

the probability distributions of some domain elements. In both cases, the probabilistic component is interpreted providing individuals with a probability distribution; this differs greatly from our multiple-world semantics, in which we consider a probability distribution over a set of classical DL interpretations.

Perhaps the closest to our approach are the Bayesian extension of DL-Lite [11] and DISPONTE [18]. The latter allows for so-called epistemic probabilities that express the uncertainty associated to a given axiom. Their semantics is based, as ours, on a probabilistic distribution over a set of interpretations. The main difference with our approach is that in [18], the authors assume that all probabilities are independent, while we allow a joint probability distribution through the BN. Another minor difference is that in DISPONTE it is impossible to deduce classical consequences, as we do. The logic described in [11] looks very similar to our framework. There is, however, a subtle but important difference. In our approach, an interpretation I satisfies an axiom $\langle \alpha : \kappa \rangle$ if $\mathcal{V}^{\mathrm{I}}(\kappa) = 1$ implies $\mathcal{I} \models \alpha$. In [11], the authors employ a closed-world assumption over the contexts, replacing this implication by an equivalence; i.e., $\mathcal{V}^{\mathrm{I}}(\kappa) = 0$ also implies $\mathcal{I} \not\models \alpha$.

5   Conclusions

We have introduced probabilistic extensions of DLs. Our basic assumption is that we have *certain* knowledge, which depends on an *uncertain* situation, or context. In practical terms, this means that every axiom is associated to a context with the intended meaning that, if the context holds, then the axiom must be true. Uncertainty is represented through a BN that encodes the probability distribution of the contexts. The advantage of using Bayesian networks relies in their capacity of describing conditional independence assumptions compactly.

We have studied the complexity of reasoning in this probabilistic logic. We have shown that for expressive DLs, the complexity of reasoning is not affected by the introduction of the contextual and probabilistic components. For tractable DLs the complexity of reasoning increases, which was expected since handling BNs is intractable in general. Moreover, we have shown that reasoning can be decoupled between the probabilistic and the logical components.

There are several directions for future work. One of them is to tighten the complexity bounds for inexpressive DLs, in particular for the DL-Lite family, as was done for $\mathcal{EL}$ [6]. We also plan to develop more goal-directed algorithm for reasoning in these logics. We will also study the possibility of using graphical models for other kinds of reasoning, like preferences, and considering different restrictions on the underlying graph. Finally, we will consider problems that tighten the relationship between the probabilistic and the logical components. One of such problems would be to update the BN according to evidence attached to the ontology.

References

1. Baader, F., Knechtel, M., Peñaloza, R.: A generic approach for large-scale ontological reasoning in the presence of access restrictions to the ontology's axioms. In:

Proc. 8th International Semantic Web Conference (ISWC 2009). LNCS, vol. 5823, pp. 49–64. Springer (2009)
2. Baader, F., Knechtel, M., Peñaloza, R.: Context-dependent views to axioms and consequences of semantic web ontologies. J. of Web Semantics 12–13, 22–40 (2012)
3. Baader, F., Peñaloza, R.: Automata-based axiom pinpointing. J. of Automated Reasoning 45(2), 91–129 (August 2010)
4. Baader, F., Peñaloza, R.: Axiom pinpointing in general tableaux. J. of Logic and Computation 20(1), 5–34 (2010)
5. Ceylan, İ.İ.: Context-Sensitive Bayesian Description Logics. Master's thesis, Dresden University of Technology, Germany (2013)
6. Ceylan, İ.İ., Peñaloza, R.: The Bayesian Description Logic $\mathcal{BEL}$. In: Proc. 7th International Conference on Automated Reasoning (IJCAR 2014) (2014), to appear.
7. Ceylan, İ.İ., Peñaloza, R.: Reasoning in the Description Logic $\mathcal{BEL}$ using Bayesian Networks. In: Proc. 4th International Workshop on Statistical Relational AI (StarAI 2014) (2014), to appear.
8. Cook, S.A.: The complexity of theorem-proving procedures. In: Proc. Third Annual ACM Symposium on Theory of Computing. pp. 151–158. STOC '71, ACM, New York, NY, USA (1971), http://doi.acm.org/10.1145/800157.805047
9. Cooper, G.F.: The computational complexity of probabilistic inference using bayesian belief networks (research note). Artif. Intel. 42(2-3), 393–405 (Mar 1990)
10. da Costa, P.C.G., Laskey, K.B., Laskey, K.J.: Pr-owl: A bayesian ontology language for the semantic web. In: Uncertainty Reasoning for the Semantic Web I, URSW 2005-2007. LNCS, vol. 5327, pp. 88–107. Springer (2008)
11. d'Amato, C., Fanizzi, N., Lukasiewicz, T.: Tractable reasoning with bayesian description logics. In: Proc. Second International Conference on Scalable Uncertainty Management (SUM 2008). LNCS, vol. 5291, pp. 146–159. Springer (2008)
12. Darwiche, A.: Modeling and Reasoning with Bayesian Networks. Cambridge University Press (2009)
13. Downey, R.G., Fellows, M.: Parameterized Complexity. Monographs in Computer Science, Springer (1999)
14. Fu, W., Peñaloza, R.: Adding context to tableaux for DLs. In: Kazakov, Y., Lembo, D., Wolter, F. (eds.) Proceedings of the 2012 International Workshop on Description Logics (DL'12). CEUR-WS, vol. 846. Rome, Italy (2012)
15. Kalyanpur, A., Parsia, B., Horridge, M., Sirin, E.: Finding all justifications of owl dl entailments. In: Proc. 6th International Semantic Web Conference (ISWC 2007). Lecture Notes in Computer Science, vol. 4825, pp. 267–280. Springer (2007)
16. Koller, D., Levy, A.Y., Pfeffer, A.: P-classic: A tractable probablistic description logic. In: Proc. 14th National Conference on Artificial Intelligence (AAAI-97). pp. 390–397. AAAI Press (1997)
17. Lukasiewicz, T., Straccia, U.: Managing uncertainty and vagueness in description logics for the semantic web. J. of Web Semantics 6(4), 291–308 (2008)
18. Riguzzi, F., Bellodi, E., Lamma, E., Zese, R.: Epistemic and statistical probabilistic ontologies. In: Proc. 8th Int. Workshop on Uncertainty Reasoning for the Semantic Web (URSW-12). vol. 900, pp. 3–14. CEUR-WS (2012)
19. Roth, D.: On the hardness of approximate reasoning. Artif. Intel. 82(1-2), 273–302 (1996)