

Dynamic Bayesian Ontology Languages

İsmail İlkan Ceylan

Theoretical Computer Science
TU Dresden, Germany
ceylan@tcs.inf.tu-dresden.de

Rafael Peñaloza

KRDB Research Centre
Free University of Bozen-Bolzano, Italy
rafael.penaloza@unibz.it

Abstract

Many formalisms combining ontology languages with uncertainty, usually in the form of probabilities, have been studied over the years. Most of these formalisms, however, assume that the probabilistic structure of the knowledge remains static over time. We present a general approach for extending ontology languages to handle time-evolving uncertainty represented by a dynamic Bayesian network. We show how reasoning in the original language and dynamic Bayesian inferences can be exploited for effective reasoning in our framework.

Introduction

Description Logics (DLs) (Baader *et al.* 2007) are a well-known family of knowledge representation formalisms that have been successfully employed for encoding the knowledge of many application domains. In DLs, knowledge is represented through a finite set of axioms, usually called an *ontology* or knowledge base. In essence, these axioms are atomic pieces of knowledge that provide explicit information of the domain. When mixed together in an ontology, these axioms may imply some additional knowledge that is not explicitly encoded. Reasoning is the act of making this implicit knowledge explicit through an entailment relation.

Some of the largest and best-maintained DL ontologies represent knowledge from the bio-medical domains. For instance, the NCBO Bioportal¹ contains 420 ontologies of various sizes. In the bio-medical fields it is very common to have only uncertain knowledge. The certainty that an expert has on an atomic piece of knowledge may have arisen from a statistical test, or from possibly imprecise measurements, for example. It thus becomes relevant to extend DLs to represent and reason with uncertainty.

The need for probabilistic extensions of DLs has been observed for over two decades already. To cover it, many different formalisms have been introduced (Jaeger 1994; Lukasiewicz and Straccia 2008; Lutz and Schröder 2010; Klinov and Parsia 2011). The differences in these logics range from the underlying classical DL used, to the semantics, to the assumptions made on the probabilistic component. One of the main issues that these logics need to handle

is the representation of joint probabilities, in particular when the different axioms are not required to be probabilistically independent. A recent approach solves this issue by dividing the ontology into *contexts*, which intuitively represent axioms that must appear together. The probabilistic knowledge is expressed through a Bayesian network that encodes the joint probability distribution of these contexts. Although originally developed as an extension of the DL \mathcal{EL} (Ceylan and Peñaloza 2014b), the framework has been extended to arbitrary ontology languages with a monotone entailment relation (Ceylan and Peñaloza 2014a).

One common feature of the probabilistic extensions of DLs existing in the literature is that they consider the probability distribution to be static. For many applications, this assumption does not hold: the probability of a person to have gray hair increases as time passes, as does the probability of a computer component to fail. To the best of our knowledge, there is so far no extension of DLs that can handle evolving probabilities effectively.

In this paper, we describe a general approach for extending ontology languages to handle evolving probabilities. By extension, our method covers all DLs, but is not limited to them. The main idea is to adapt the formalism from (Ceylan and Peñaloza 2014a) to use *dynamic* Bayesian networks (DBNs) (Murphy 2002) as the underlying uncertainty structure to compactly encode evolving probability distributions.

Given an arbitrary ontology language \mathcal{L} , we define its dynamic Bayesian extension $DB\mathcal{L}$. We show that reasoning in $DB\mathcal{L}$ can be seamlessly divided into the probabilistic computation over the DBN, and the logical component with its underlying entailment relation. In order to reduce the number of entailment checks, we compile a so-called context formula, which encodes all contexts in which a given consequence holds.

Related to our work are relational BNs (Jaeger 1997) and their extensions. In contrast to relational BNs, we provide a tight coupling between the logical formalism and the DBN, which allows us to describe evolving probabilities while keeping the intuitive representations of each individual component. Additionally, restricting the logical formalism to specific ontology languages provides an opportunity for finding effective reasoning algorithms.

Bayesian Ontology Languages

To remain as general as possible, we do not fix a specific logic, but consider an arbitrary *ontology language* \mathcal{L} consisting of two infinite sets \mathfrak{A} and \mathfrak{C} of *axioms* and *consequences*, respectively, and a class $\mathfrak{D} \subseteq \wp_{\text{fin}}(\mathfrak{A})$ of finite sets of axioms, called *ontologies*, such that if $\mathcal{O} \in \mathfrak{D}$, then $\mathcal{O}' \in \mathfrak{D}$ for all $\mathcal{O}' \subseteq \mathcal{O}$. The language \mathcal{L} is associated to a class \mathfrak{I} of *interpretations* and an *entailment relation* $\models \subseteq \mathfrak{I} \times (\mathfrak{A} \cup \mathfrak{C})$. An interpretation $\mathcal{I} \in \mathfrak{I}$ is a *model* of the ontology \mathcal{O} ($\mathcal{I} \models \mathcal{O}$) if $\mathcal{I} \models \alpha$ for all $\alpha \in \mathcal{O}$. \mathcal{O} *entails* $c \in \mathfrak{C}$ ($\mathcal{O} \models c$) if every model of \mathcal{O} entails c . Notice that the entailment relation is monotonic; i.e., if $\mathcal{O} \models c$ and $\mathcal{O} \subseteq \mathcal{O}' \in \mathfrak{D}$, then $\mathcal{O}' \models c$. Any standard description logic (DL) (Baader *et al.* 2007) is an ontology language of this kind; consequences in these languages are e.g. concept unsatisfiability, concept subsumption, or query entailment. However, many other ontology languages of varying expressivity and complexity properties exist. For the rest of this paper, \mathcal{L} is an arbitrary but fixed ontology language, with axioms \mathfrak{A} , ontologies \mathfrak{D} , consequences \mathfrak{C} , and interpretations \mathfrak{I} .

As an example language we use the DL \mathcal{EL} (Baader *et al.* 2005), which we briefly introduce here. Given two disjoint sets \mathbb{N}_C and \mathbb{N}_R , \mathcal{EL} *concepts* are built by the grammar rule $C ::= A \mid \top \mid C \sqcap C \mid \exists r.C$ where $A \in \mathbb{N}_C$ and $r \in \mathbb{N}_R$. \mathcal{EL} *axioms* and *consequences* are expressions of the form $C \sqsubseteq D$, where C and D are concepts. An *interpretation* is a pair $(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ where $\Delta^{\mathcal{I}}$ is a non-empty set and $\cdot^{\mathcal{I}}$ maps every $A \in \mathbb{N}_C$ to $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and every $r \in \mathbb{N}_R$ to $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. This function is extended to concepts by $\top^{\mathcal{I}} := \Delta^{\mathcal{I}}$, $(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}$, and $(\exists r.C)^{\mathcal{I}} := \{d \mid \exists e \in C^{\mathcal{I}}.(d, e) \in r^{\mathcal{I}}\}$. This interpretation *entails* the axiom (or consequence) $C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.

The Bayesian ontology language $\mathcal{B}\mathcal{L}$ extends \mathcal{L} by associating each axiom in an ontology with a context, which intuitively describes the situation in which the axiom is required to hold. The knowledge of which context applies is uncertain, and expressed through a Bayesian network (Ceylan and Peñaloza 2014a).

Briefly, a *Bayesian network* (Darwiche 2009) is a pair $\mathcal{B} = (G, \Phi)$, where $G = (V, E)$ is a finite directed acyclic graph (DAG) whose nodes represent Boolean random variables, and Φ contains, for every $x \in V$, a conditional probability distribution $P_{\mathcal{B}}(x \mid \pi(x))$ of x given its parents $\pi(x)$. Every variable $x \in V$ is conditionally independent of its non-descendants given its parents. Thus, the BN \mathcal{B} defines a unique joint probability distribution (JPD) over V :

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

Let V be a finite set of variables. A *V-context* is a consistent set of literals over V . A *V-axiom* is an expression of the form $\langle \alpha : \kappa \rangle$ where $\alpha \in \mathfrak{A}$ is an axiom and κ is a V -context. A *V-ontology* is a finite set \mathcal{O} of V -axioms, such that $\{\alpha \mid \langle \alpha : \kappa \rangle \in \mathcal{O}\} \in \mathfrak{D}$. A $\mathcal{B}\mathcal{L}$ *knowledge base* (KB) over V is a pair $\mathcal{K} = (\mathcal{O}, \mathcal{B})$ where \mathcal{B} is a BN over V and \mathcal{O} is a V -ontology.

We briefly illustrate these notions over the language $\mathcal{B}\mathcal{E}\mathcal{L}$, an extension of the DL \mathcal{EL} , in the following Example.

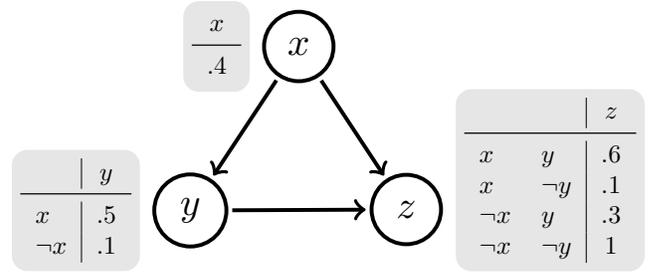


Figure 1: The BN \mathcal{B}_1 over the variables $V = \{x, y, z\}$

Example 1. Consider the $\mathcal{B}\mathcal{E}\mathcal{L}$ KB $\mathcal{K}_1 = (\mathcal{B}_1, \mathcal{O}_1)$ where

$$\begin{aligned} \mathcal{O}_1 = & \{ \langle \text{Comp} \sqsubseteq \exists \text{use.Mem} \sqcap \exists \text{use.CPU} : \emptyset \rangle, \\ & \langle \exists \text{use.FailMem} \sqsubseteq \text{FailComp} : \{x\} \rangle, \\ & \langle \exists \text{use.FailCPU} \sqsubseteq \text{FailComp} : \{x\} \rangle, \\ & \langle \exists \text{use.FailMem} \sqcap \exists \text{use.FailCPU} \sqsubseteq \text{FailComp} : \{\neg x\} \rangle, \\ & \langle \text{Mem} \sqsubseteq \text{FailMem} : \{y\} \rangle, \langle \text{CPU} \sqsubseteq \text{FailCPU} : \{z\} \rangle \}, \end{aligned}$$

and \mathcal{B}_1 is the BN shown in Figure 1.

This KB represents a computer failure scenario, where x stands for a critical situation, y represents the memory failing, and z the CPU failing.

The *contextual semantics* is defined by extending interpretations to evaluate also the variables from V . A *V-interpretation* is a pair $\mathbb{I} = (\mathcal{I}, \mathcal{V}^{\mathbb{I}})$ where $\mathcal{I} \in \mathfrak{I}$ and $\mathcal{V}^{\mathbb{I}}$ is a propositional interpretation over the variables V . The V -interpretation $\mathbb{I} = (\mathcal{I}, \mathcal{V}^{\mathbb{I}})$ is a *model* of $\langle \alpha : \kappa \rangle$ ($\mathbb{I} \models \langle \alpha : \kappa \rangle$), where $\alpha \in \mathfrak{A}$, iff $(\mathcal{V}^{\mathbb{I}} \not\models_p \kappa)$ ² or $(\mathcal{I} \models \alpha)$.

It is a *model* of the V -ontology \mathcal{O} iff it is a model of all the V -axioms in \mathcal{O} . It *entails* $c \in \mathfrak{C}$ if $\mathbb{I} \models c$. The intuition behind this semantics is that an axiom is evaluated to true by all models provided it is in the right context.

Given a V -ontology \mathcal{O} , every propositional interpretation, or *world*, \mathcal{W} on V defines an ontology $\mathcal{O}_{\mathcal{W}} := \{\alpha \mid \langle \alpha : \kappa \rangle \in \mathcal{O}, \mathcal{W} \models_p \kappa\}$. Consider the KB \mathcal{K}_1 provided in Example 1: The world $\mathcal{W} = \{x, \neg y, z\}$ defines the ontology

$$\begin{aligned} \mathcal{O}_{\mathcal{W}} = & \{ \text{Comp} \sqsubseteq \exists \text{use.Mem} \sqcap \exists \text{use.CPU}, \\ & \exists \text{use.FailMem} \sqsubseteq \text{FailComp}, \\ & \exists \text{use.FailCPU} \sqsubseteq \text{FailComp}, \text{CPU} \sqsubseteq \text{FailCPU} \}. \end{aligned}$$

Intuitively, a contextual ontology is a compact representation of exponentially many ontologies from \mathcal{L} ; one for each world \mathcal{W} . The uncertainty in $\mathcal{B}\mathcal{L}$ is expressed by the BN \mathcal{B} , which is interpreted using multiple world semantics.

Definition 2 (probabilistic interpretation). 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 $P_{\mathbb{I}}(\mathbb{I}) > 0$ only for finitely many interpretations $\mathbb{I} \in \mathbb{I}$. It is a *model* of the V -ontology \mathcal{O} if every $\mathbb{I} \in \mathbb{I}$ is a model of \mathcal{O} . \mathcal{P} is *consistent* with the BN \mathcal{B} if for every valuation \mathcal{W} of the variables in V it holds that

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

²We use \models_p to distinguish propositional entailment from \models .

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

The fundamental reasoning task in \mathcal{BL} , probabilistic entailment, consists in finding the probability of observing a consequence c ; that is, the probability of being at a context where c holds.

Definition 3 (probabilistic entailment). Let $c \in \mathcal{C}$, and \mathcal{K} a \mathcal{BL} KB. The *probability of c* w.r.t. the probabilistic interpretation $\mathcal{P} = (\mathbb{I}, P_{\mathbb{I}})$ is $P_{\mathcal{P}}(c) := \sum_{(\mathcal{I}, \mathcal{W}) \in \mathbb{I}, \mathcal{I} \models c} P_{\mathbb{I}}(\mathcal{I}, \mathcal{W})$. The *probability of c* w.r.t. \mathcal{K} is $P_{\mathcal{K}}(c) := \inf_{\mathcal{P} \models \mathcal{K}} P_{\mathcal{P}}(c)$.

It has been shown that to compute the conditional probability of a consequence c , it suffices to test, for each world \mathcal{W} , whether $\mathcal{O}_{\mathcal{W}}$ entails c (Ceylan and Peñaloza 2014b).

Proposition 4. Let $\mathcal{K} = (\mathcal{B}, \mathcal{O})$ be a \mathcal{BL} KB and $c \in \mathcal{C}$. Then $P_{\mathcal{K}}(c) = \sum_{\mathcal{O}_{\mathcal{W}} \models c} P_{\mathcal{B}}(\mathcal{W})$.

This means that reasoning in \mathcal{BL} can be reduced to exponentially many entailment tests in the classical language \mathcal{L} . For some logics, this exponential enumeration of worlds can be avoided (Ceylan and Peñaloza 2014c). However, this depends on the properties of the ontological language and its entailment relation, and cannot be guaranteed for arbitrary languages.

Another relevant problem is to compute the probability of a consequence given some partial information about the context. Given a context κ , the conditional probability $P_{\mathcal{K}}(c | \kappa)$ is defined via the rule $P_{\mathcal{K}}(c, \kappa) = P_{\mathcal{K}}(c | \kappa) P_{\mathcal{B}}(\kappa)$, where

$$P_{\mathcal{K}}(c, \kappa) = \sum_{\mathcal{O}_{\mathcal{W}} \models c, \mathcal{W} \models_p \kappa} P_{\mathcal{B}}(\mathcal{W}).$$

For simplicity, in the rest of this paper we consider only unconditional consequences. However, it should be noted that all results can be transferred to the conditional case.

Example 5. Consider again the KB $\mathcal{K}_1 = (\mathcal{B}_1, \mathcal{O}_1)$ from Example 1 and the consequence $\text{Comp} \sqsubseteq \text{FailComp}$. We are interested in finding the probability of the computer to fail, i.e. $P_{\mathcal{K}_1}(\text{Comp} \sqsubseteq \text{FailComp})$. This can be computed by enumerating all worlds \mathcal{W} for which $\mathcal{O}_{\mathcal{W}} \models \text{Comp} \sqsubseteq \text{FailComp}$, which yields the probability 0.238.

As seen, it is possible to extend any ontological language to allow for probabilistic reasoning based on a Bayesian network. We now further extend this formalism to be able to cope with controlled updates of the probabilities over time.

Dynamic Bayesian Ontology Languages

With \mathcal{BL} , one is able to represent and reason about the uncertainty of the current context, and the consequences that follow from it. In that setting, the joint probability distribution of the contexts, expressed by the BN, is known and fixed. In some applications, see especially (Sadilek and Kautz 2010) this probability distribution may change over time. For example, as the components of a computer age, their probability of failing increases. The new probability depends on how likely it was for the component to fail previously, and the ageing factors to which it is exposed. We now extend \mathcal{BL} to

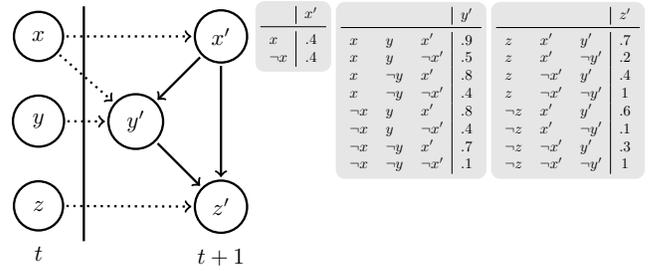


Figure 2: The TBN $\mathcal{B}_{\rightarrow}$ over the variables $V = \{x, y, z\}$

handle these cases, by considering *dynamic* BNs as the underlying formalism for managing uncertainty over contexts.

Dynamic BNs (DBNs) (Dean and Kanazawa 1989; Murphy 2002) extend BNs to provide a compact representation of evolving joint probability distributions for a fixed set of random variables. The update of the JPD is expressed through a two-slice BN, which expresses the probabilities at the next point in time, given the current context.

Definition 6 (DBN). Let V be a finite set of Boolean random variables. A *two-slice BN* (TBN) over V is a pair (G, Φ) , where $G = (V \cup V', E)$ is a DAG containing no edges between elements of V , $V' = \{x' | x \in V\}$, and Φ contains, for every $x' \in V'$ a conditional probability distribution $P(x' | \pi(x'))$ of x' given its parents $\pi(x')$. A *dynamic Bayesian network* (DBN) over V is a pair $\mathcal{D} = (\mathcal{B}_1, \mathcal{B}_{\rightarrow})$ where \mathcal{B}_1 is a BN over V , and $\mathcal{B}_{\rightarrow}$ is a TBN over V .

A TBN over $V = \{x, y, z\}$ is depicted in Figure 2. The set of nodes of the graph can be thought of as containing two disjoint copies of the random variables in V . Then, the probability distribution at time $t + 1$ depends on the distribution at time t . In the following we will use V_t and x_t for $x \in V$, to denote the variables in V at time t .

As standard in BNs, the graph structure of a TBN encodes the conditional dependencies among the nodes: every node is independent of all its non-descendants given its parents. Thus, for a TBN \mathcal{B} , the conditional probability distribution at time $t + 1$ given time t is

$$P_{\mathcal{B}}(V_{t+1} | V_t) = \prod_{x' \in V'} P_{\mathcal{B}}(x' | \pi(x')).$$

We further assume the Markov property: the probability of the future state is independent from the past, given the present state.

In addition to the TBN, a DBN contains a BN \mathcal{B}_1 that encodes the JPD of V at the beginning of the evolution. Thus, the DBN $\mathcal{D} = (\mathcal{B}_1, \mathcal{B}_{\rightarrow})$ defines, for every $t \geq 1$, the unique probability distribution

$$P_{\mathcal{B}}(V_t) = P_{\mathcal{B}_1}(V_1) \prod_{i=2}^t \prod_{x \in V} P_{\mathcal{B}_{\rightarrow}}(x_i | \pi(x_i)).$$

Intuitively, the distribution at time t is defined by unraveling the DBN starting from \mathcal{B}_1 , using the two-slice structure of $\mathcal{B}_{\rightarrow}$ until t copies of V have been created. This produces a

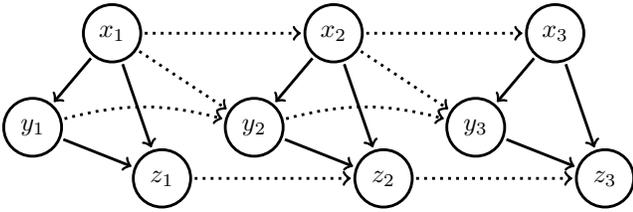


Figure 3: Three step unraveling $\mathcal{B}_{1:3}$ of $(\mathcal{B}_1, \mathcal{B}_{\rightarrow})$

new BN $\mathcal{B}_{1:t}$ encoding the distribution over time of the different variables. Figure 3 depicts the unraveling to $t = 3$ of the DBN $(\mathcal{B}_1, \mathcal{B}_{\rightarrow})$ where \mathcal{B}_1 and $\mathcal{B}_{\rightarrow}$ are the networks depicted in Figures 1 and 2, respectively. The conditional probability tables of each node given its parents (not depicted) are those of \mathcal{B}_1 for the nodes in V_1 , and of $\mathcal{B}_{\rightarrow}$ for nodes in $V_2 \cup V_3$. Notice that $\mathcal{B}_{1:t}$ has t copies of each random variable in V . For a given $t \geq 1$, we call \mathcal{B}_t the BN obtained from the unraveling $\mathcal{B}_{1:t}$ of the DBN to time t , and eliminating all variables not in V_t . In particular, we have that $P_{\mathcal{B}_t}(V) = P_{\mathcal{B}_{1:t}}(V_t)$.

The dynamic Bayesian ontology language \mathcal{DBL} is very similar to \mathcal{BL} , except that the probability distribution of the contexts evolves accordingly to a DBN.

Definition 7 (\mathcal{DBL} KB). A \mathcal{DBL} knowledge base (KB) is a pair $\mathcal{K} = (\mathcal{D}, \mathcal{O})$ where $\mathcal{D} = (\mathcal{B}_1, \mathcal{B}_{\rightarrow})$ is a DBN over V and \mathcal{O} is a V -ontology. Let $\mathcal{K} = (\mathcal{D}, \mathcal{O})$ be a \mathcal{DBL} KB over V . A *timed probabilistic interpretation* is an infinite sequence $\mathfrak{P} = (\mathcal{P}_t)_{t \geq 1}$ of probabilistic interpretations. \mathfrak{P} is a *model* of \mathcal{K} if for every t , \mathcal{P}_t is a model of the \mathcal{BL} KB $(\mathcal{B}_t, \mathcal{O})$.

In a nutshell, a DBN can be thought of as a compact representation of an infinite sequence of BNs $\mathcal{B}_1, \mathcal{B}_2, \dots$ over V . Following this idea, a \mathcal{DBL} KB expresses an infinite sequence of \mathcal{BL} KBs, where the ontological component remains unchanged, and only the probability distribution of the contexts evolves over time. A timed probabilistic interpretation \mathfrak{P} simply interprets each of these \mathcal{BL} KBs, at the corresponding point in time. To be a model of a \mathcal{DBL} KB, \mathfrak{P} must then be a model of all the associated \mathcal{BL} KBs.

Before describing the reasoning tasks for \mathcal{DBL} and methods for solving them, we show how the computation of all the contexts that entail a consequence can be reduced to the enumeration of the worlds satisfying a propositional formula.

Compiling Contextual Knowledge

From Proposition 4, we see that reasoning in \mathcal{BL} can be reduced to checking, for every world \mathcal{W} , whether $\mathcal{O}_{\mathcal{W}} \models c$. This reduces probabilistic reasoning to a sequence of standard entailment tests over the original language \mathcal{L} . However, each of these entailments might be very expensive. For example, in the very expressive DL \mathcal{SHOIQ} , deciding an entailment is already NEXPTIME-hard (Tobies 2000). Rather than repeating this reasoning step for every world, it makes sense to try to identify the relevant worlds *a priori*. We do this through the computation of a *context formula*.

Definition 8 (context formula). Let \mathcal{O} be a V -ontology, and $c \in \mathcal{C}$. A *context formula* for c w.r.t. \mathcal{O} is a propositional formula ϕ such that for every interpretation \mathcal{W} of the variables in V , it holds that $\mathcal{O}_{\mathcal{W}} \models c$ iff $\mathcal{W} \models_p \phi$.

The idea behind this formula is that, for finding whether $\mathcal{O}_{\mathcal{W}} \models c$, it suffices to check whether the valuation \mathcal{W} satisfies ϕ . This test requires only linear time on the length of the context formula. The context formula can be seen as a generalization of the pinpointing formula (Baader and Peñaloza 2010b) and the boundary (Baader *et al.* 2012), defined originally for classical ontology languages.

Example 9. Consider again the V -ontology \mathcal{O}_1 from Example 1. The formula $\phi_1 := (x \wedge (y \vee z)) \vee (\neg x \wedge y \wedge z)$ is a context formula for $\text{Comp} \sqsubseteq \text{FailComp}$ w.r.t. \mathcal{O}_1 . In fact, the valuation $\{x, \neg y, z\}$ satisfies this formula.

Clearly, computing the context formula must be at least as hard as deciding an entailment in \mathcal{L} : if we label every axiom in a classical \mathcal{L} -ontology \mathcal{O} with the same propositional variable x , then the boundary formula of c w.r.t. this $\{x\}$ -ontology is x iff $\mathcal{O} \models c$. On the other hand, the algorithm used for deciding the entailment relation can usually be modified to compute the context formula. Using arguments similar to those developed for axiom pinpointing (Baader and Peñaloza 2010a; Kalyanpur *et al.* 2007), it can be shown that for most description logics, computing the context formula is not harder, in terms of computational complexity, than standard reasoning. In particular this holds for any arbitrary ontology language whose entailment relation is EXPTIME-hard. This formula can also be compiled into a more efficient data structure like binary decision diagrams (Lee 1959). Intuitively, this means that we can compute this formula using the same amount of resources needed for only one entailment test, and then use it for verifying whether the sub-ontology defined by a world entails the consequence in an efficient way.

Reasoning in \mathcal{DBL}

Rather than merely computing the probability of currently observing a consequence, we are interested in computing the probability of a consequence to follow after some fixed number of time steps t .

Definition 10 (probabilistic entailment with time). Let $\mathcal{K} = (\mathcal{D}, \mathcal{O})$ be a \mathcal{DBL} KB and $c \in \mathcal{C}$. Given a timed interpretation \mathfrak{P} and $t \geq 1$, the *probability of c at time t* w.r.t. \mathfrak{P} is $P_{\mathfrak{P}}(c[t]) := P_{\mathcal{P}_t}(c)$. The *probability of c at time t* w.r.t. \mathcal{K} is $P_{\mathcal{K}}(c[t]) := \inf_{\mathfrak{P} \models \mathcal{K}} P_{\mathfrak{P}}(c[t])$.

We show that probabilistic entailment over a fixed time bound can be reduced to probabilistic entailment defined for BOLs by unravelling the DBN.

Lemma 11. Let $\mathcal{K} = (\mathcal{D}, \mathcal{O})$ be a \mathcal{DBL} KB, $c \in \mathcal{C}$, and $t \geq 1$. Then the probability of c at time t w.r.t. \mathcal{K} is given by

$$P_{\mathcal{K}}(c[t]) = \sum_{\mathcal{O}_{\mathcal{W}} \models c} P_{\mathcal{B}_t}(\mathcal{W})$$

Proof. (Sketch) A timed model \mathfrak{P} of \mathcal{K} is a sequence of probabilistic interpretations $\mathcal{P}_1, \mathcal{P}_2, \dots$, where each \mathcal{P}_i is a

model of the \mathcal{BL} KB $\mathcal{K}_i := (\mathcal{B}_i, \mathcal{O})$. We use this fact to show that

$$P_{\mathcal{K}}(c[t]) = \inf_{\mathfrak{P} \models \mathcal{K}} P_{\mathfrak{P}}(c[t]) = \inf_{\mathfrak{P} \models \mathcal{K}} P_{\mathcal{P}_t}(c) \quad (1)$$

$$= \inf_{\mathcal{P}_t \models \mathcal{K}_t} P_{\mathcal{P}_t}(c) = P_{\mathcal{K}_t}(c) \quad (2)$$

$$= \sum_{\mathcal{O}_{\mathcal{W}} \models c} P_{\mathcal{B}_t}(\mathcal{W}), \quad (3)$$

where (1) follows from Definition 10, (2) holds by definition, and (3) follows from Proposition 4. \square

Lemma 11 provides a method for computing the probability of an entailment at a fixed time t . One can first generate the BN \mathcal{B}_t , and then compute the probability w.r.t. \mathcal{B}_t of all the worlds that entail c . Moreover, using a context formula we can compile away the ontology and reduce reasoning to standard inferences in BNs, only.

Theorem 12. *Let $\mathcal{K} = (\mathcal{D}, \mathcal{O})$ be a DBL KB, $c \in \mathcal{C}$, ϕ a context formula for c w.r.t. \mathcal{O} , and $t \geq 1$. Then the probability of c at time t w.r.t. \mathcal{K} is given by $P_{\mathcal{K}}(c[t]) = P_{\mathcal{B}_t}(\phi)$.*

Proof. By Lemma 11 and the definition of a context formula, we have

$$P_{\mathcal{K}}(c[t]) = \sum_{\mathcal{O}_{\mathcal{W}} \models c} P_{\mathcal{B}_t}(\mathcal{W}) = \sum_{\mathcal{W} \models_p \phi} P_{\mathcal{B}_t}(\mathcal{W}) = P_{\mathcal{B}_t}(\phi),$$

which proves the result. \square

This means that one can first compute a context formula for c and then do probabilistic inferences over the DBN to detect the probability of satisfying ϕ at time t . For this, we can exploit any existing DBN inference method. One option is to do variable elimination over the t -step unraveled DBN $\mathcal{B}_{1:t}$, to compute \mathcal{B}_t . Assuming that t is fixed, it suffices to make $2^{|V|}$ inferences (one for each world) over \mathcal{B}_t and the same number of propositional entailment tests over the context formula. If entailment in \mathcal{L} is already exponential, then computing the probability of c at time t is as hard as deciding entailments.

The previous argument only works assuming a fixed time point t . Since it depends heavily on computing \mathcal{B}_t (e.g., via variable elimination), it does not scale well as t increases. Other methods have been proposed for exploiting the recursive structure of the DBN. For instance, one can use the algorithm described in (Vlasselaer *et al.* 2014) that provides linear scalability over time. The main idea is to compile the structure into an arithmetic circuit (Darwiche 2009) and use forward and backward message passing (Murphy 2002).

While computing the probability of a consequence at a fixed point in time t is a relevant task, it is usually more important to know whether the consequence can be observed within a given time limit. In our computer example, we would be interested in finding the probability of the system failing within, say, the following twenty steps.

Abusing of the notation, we use the expression $\sim c$, $c \in \mathcal{C}$, to denote that the consequence c does not hold; i.e., $\mathcal{I} \models \sim c$ iff $\mathcal{I} \not\models c$. Thus, for example, $P_{\mathcal{K}}(\sim c[t])$ is the probability

of c not holding at time t . To find the probability of observing c in the first t time steps, one can alternatively compute the probability of not observing c in any of those steps. Formally, for a timed interpretation \mathfrak{P} and $t \in \mathbb{N}$, we define

$$P_{\mathfrak{P}}(c[1:t]) := 1 - P_{\mathfrak{P}}(\sim c[1], \dots, \sim c[t]).$$

Definition 13 (time bounded probabilistic entailment). *The probability of observing c in at most t steps w.r.t. the DBL KB \mathcal{K} is $P_{\mathcal{K}}(c[1:t]) := \inf_{\mathfrak{P} \models \mathcal{K}} P_{\mathfrak{P}}(c[1:t])$.*

Just as before, given a constant $t \geq 1$, it is possible to compute $P_{\mathcal{K}}(c[1:t])$ by looking at the t -step unraveling of \mathcal{D} . More precisely, to compute $P_{\mathfrak{P}}(\sim c[1], \dots, \sim c[t])$, it suffices to look at all the valuations \mathcal{W} of $\bigcup_{i=1}^t V_i$ such that for all $i, 1 \leq i \leq t$, it holds that $\mathcal{O}_{\mathcal{W}^{(i)}} \models \sim c$. These valuations correspond to an evolution of the system where the consequence c is not observed in the first t steps. The probability of these valuations w.r.t. $\mathcal{B}_{1:t}$ then yields the probability of *not* observing this consequence. We thus get the following result.

Theorem 14. *Let $\mathcal{K} = (\mathcal{D}, \mathcal{O})$ be a DBL KB, $c \in \mathcal{C}$, ϕ a context formula for c w.r.t. \mathcal{O} , and $t \geq 1$. Then $P_{\mathcal{K}}(c[1:t]) = \sum_{\mathcal{W} \models_{\exists i} \mathcal{W}^{(i)} \models_p \phi} P_{\mathcal{B}_{1:t}}(\mathcal{W})$.*

Proof (Sketch). Using the pithy interpretations of the crisp ontologies $\mathcal{O}_{\mathcal{W}^{(i)}}$, we can build a timed interpretation \mathfrak{P}_0 such that $P_{\mathfrak{P}_0}(c[1:t]) = \sum_{\mathcal{W} \models_{\exists i} \mathcal{W}^{(i)} \models_p \phi} P_{\mathcal{B}_{1:t}}(\mathcal{W})$, in a way similar to Theorem 12 of (Ceylan and Peñaloza 2014b). The existence of another timed interpretation \mathfrak{P} such that $P_{\mathfrak{P}}(c[1:t]) < P_{\mathfrak{P}_0}(c[1:t])$ contradicts the properties of the pithy interpretations. Thus, we obtain that $P_{\mathfrak{P}_0}(c[1:t]) = \inf_{\mathfrak{P} \models \mathcal{K}} P_{\mathfrak{P}}(c[1:t]) = P_{\mathcal{K}}(c[1:t])$. \square

This means that the probability of observing a consequence within a fixed time-bound t can be computed by simply computing the context formula and then performing probabilistic *a posteriori* computations over the unraveled BN. In our running example, the probability of observing a computer failure in the next 20 steps is simply

$$P_{\mathcal{K}}(\text{Comp} \sqsubseteq \text{FailComp}[1:20]) = \sum_{\mathcal{W} \models_{\exists i} \mathcal{W}^{(i)} \models_p \phi} P_{\mathcal{B}_i}(\phi).$$

Thus, the computational complexity of reasoning is not affected by introducing the dynamic evolution of the BN, as long as the time bound is constant. Notice, however, that the number of possible valuations grows exponentially on the time bound t . Thus, for large intervals, this approach becomes unfeasible.

By extending the time limit indefinitely, we can also find the probability of *eventually* observing the consequence c (e.g., the probability of the system ever failing). The probability of *eventually* observing c w.r.t. \mathcal{K} is given by $P_{\mathcal{K}}(c[\infty]) := \lim_{t \rightarrow \infty} P_{\mathcal{K}}(c[1:t])$. Notice that $P_{\mathcal{K}}(c[1:t])$ is monotonic on t and bounded by 1; hence $P_{\mathcal{K}}(c[\infty])$ is well defined.

Observe that Theorem 14 cannot be used to compute the probability of *eventually* observing c since one cannot necessarily predict the changes in probabilities of finding worlds that entail the consequence c . Rather than considering these

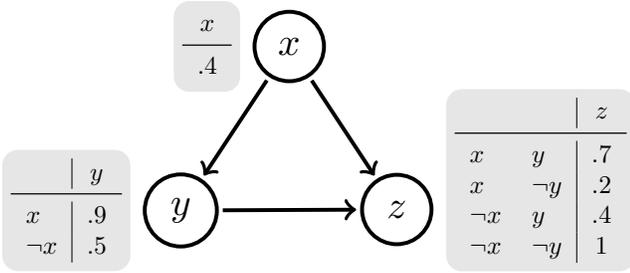


Figure 4: $P(V \mid \{x, y, z\}_t)$

increasingly large BNs separately, we can exploit methods developed for probability distributions that evolve over time. This will also allow us to extract more information from *DBL* KBs.

It is easy to see that every TBN defines a time-homogeneous Markov chain over a finite state space. More precisely, if \mathcal{B} is a TBN over V , then $\mathcal{M}_{\mathcal{B}}$ is the Markov chain, where every valuation \mathcal{W} of the variables in V is a state and the transition probability distribution given the current state \mathcal{W} is described by the BN obtained from adding \mathcal{W} as evidence to the first slice of \mathcal{B} . For example, the TBN $\mathcal{B}_{\rightarrow}$ from Figure 2 yields the conditional probability distribution given that $\{x, y, z\}$ was observed at time t depicted in Figure 4. From this, we can derive the probability of observing $\{x, y, z\}$ at time $t + 1$ given that it was observed at time t , which is $P(\{x, y, z\}_{t+1} \mid \{x, y, z\}_t) = 0.252$.

We extend the notions from Markov chains to TBNs in the obvious way. In particular, the TBN \mathcal{B} is *irreducible* if for every two worlds \mathcal{V}, \mathcal{W} , the probability of eventually reaching \mathcal{W} given \mathcal{V} is greater than 0. It is *aperiodic* if for every world \mathcal{W} there is an $n_{\mathcal{W}}$ such that for all $n \geq n_{\mathcal{W}}$, it holds that $P(\mathcal{W}_n \mid \mathcal{W}) > 0$. A distribution $P_{\mathcal{W}}$ over the worlds is *stationary* if $\sum_{\mathcal{V}} P(\mathcal{V} \mid \mathcal{W})P_{\mathcal{W}}(\mathcal{W}) = P_{\mathcal{W}}(\mathcal{V})$ holds for every world \mathcal{V} . It follows immediately that if \mathcal{B} is irreducible and aperiodic, then it has a unique stationary distribution (Harris 1956).

Given a TBN \mathcal{B} over V , let now $\Delta_{\mathcal{B}}$ be the set of all stationary distributions over the worlds of V . For a world \mathcal{W} , define $\delta_{\mathcal{B}}(\mathcal{W}) := \min_{P \in \Delta_{\mathcal{B}}} P(\mathcal{W})$ to be the smallest probability assigned by any stationary distribution of \mathcal{B} to \mathcal{W} . If $\delta_{\mathcal{B}}(\mathcal{W}) > 0$, then we know that, regardless of the initial distribution, in the limit we will always be able to observe the world \mathcal{W} with a constant positive probability. In particular, this means that the probability of eventually observing \mathcal{W} equals 1. Notice moreover that this results is independent of the initial distribution used.

We can take this idea one step forward, and consider sets of worlds. For a propositional formula ϕ , let

$$\delta_{\mathcal{B}}(\phi) := \min_{P \in \Delta_{\mathcal{B}}} \sum_{\mathcal{W} \models \phi} P(\mathcal{W}).$$

In other words, $\delta_{\mathcal{B}}(\phi)$ expresses the minimum probability of satisfying ϕ in any stationary distribution of \mathcal{B} . From the arguments above, we obtain the following theorem.

Theorem 15. *Let $\mathcal{K} = (\mathcal{D}, \mathcal{O})$ be a *DBL* KB over V with*

$\mathcal{D} = (\mathcal{B}_1, \mathcal{B}_{\rightarrow})$, $c \in \mathcal{C}$, and ϕ a context formula for c w.r.t. \mathcal{O} . If $\delta_{\mathcal{B}_{\rightarrow}}(\phi) > 0$, then $P_{\mathcal{K}}(c[\infty]) = 1$.

In particular, if $\mathcal{B}_{\rightarrow}$ is irreducible and aperiodic, $\Delta_{\mathcal{B}}$ contains only one stationary distribution, which simplifies the computation of the function δ . Unfortunately, such a simple characterization of $P_{\mathcal{K}}(c[\infty])$ cannot be given when $\delta_{\mathcal{B}_{\rightarrow}}(\phi) = 0$. In fact, in this case the result may depend strongly on the initial distribution.

Example 16. Let $V = \{x\}$, $\mathcal{O}_2 = \{A \sqsubseteq B : \{x\}\}$, and consider the TBN $\mathcal{B}'_{\rightarrow}$ over V defined by $P(x' \mid x) = 1$ and $P(x' \mid \neg x) = 0$. It is easy to see that any distribution over the valuations of V is stationary. For every initial distribution \mathcal{B} , if $\mathcal{K} = (\mathcal{D}, \mathcal{O}_2)$ where $\mathcal{D} = (\mathcal{B}, \mathcal{B}'_{\rightarrow})$, then $P_{\mathcal{K}}(A \sqsubseteq B[\infty]) = P_{\mathcal{B}}(x)$.

So far, our reasoning services have focused on predicting the outcome at future time steps, given the current knowledge of the system. Based on our model of evolving probabilities, the distribution at any time $t + 1$ depends only on time t , if it is known. However, for many applications it makes sense to consider evidence that is observed throughout several time steps. For instance, in our computer failure scenario, the DBN $\mathcal{B}_{\rightarrow}$ ensures that, if at some point a critical situation is observed (x is true), then the probability of observing a memory or CPU failure in the next step is higher. That is, the evolution of the probability distribution is affected by the observed value of the variable x .

Suppose that we have observed over the first t time steps that no critical situation has occurred, and we want to know the probability of a computer failure. Formally, let E be a consistent set of literals over $\bigcup_{i=1}^t V_i$. We want to compute the probability $P_{\mathcal{K}}(c[t] \mid E)$ of observing c at time t given the evidence E . This is just a special case of bounded probabilistic entailment, where the worlds are not only restricted w.r.t. the context formula but also w.r.t. the evidence E .

The efficiency of this approach depends strongly on the time bound t , but also on the structure of the TBN $\mathcal{B}_{\rightarrow}$. Recall that the complexity of reasoning in a BN depends on the tree-width of its underlying DAG (Pan *et al.* 1998). The unraveling of $\mathcal{B}_{\rightarrow}$ produces a new DAG whose tree-width might increase with each unraveling step, thus impacting the reasoning methods negatively.

Conclusions

We have introduced a general approach for extending ontology languages to handle time-evolving probabilities with the help of a DBN. Our framework can be instantiated to any language with a monotonic entailment relation including, but not limited to, all the members of the description logic family of knowledge representation formalisms.

Our approach extends on ideas originally introduced for static probabilistic reasoning. The essence of the method is to divide an ontology into different contexts, which are identified by a consistent set of propositional variables from a previously chosen finite set of variables V . The probabilistic knowledge is expressed through a probability distribution over the valuations of V which is encoded by a DBN.

Interestingly, our formalism allows for reasoning methods that exploit the properties of both, the ontological, and

the probabilistic components. From the ontological point of view, we can use supplemental reasoning to produce a context formula that encodes all the possible worlds from which a wanted consequence can be derived. We can then use standard DBN methods to compute the probability of satisfying this formula.

This work represents first steps towards the development of a formalism combining well-known ontology languages with time-evolving probabilities. First of all, we have introduced only the most fundamental reasoning tasks. It is possible to think of many other problems like finding the most plausible explanation for an observed event, or computing the expected time until a consequence is derived, among many others.

Finally, the current methods developed for handling DBNs, although effective, are not adequate for our problems. To find out the probability of satisfying the context formula ϕ , we need compute the probability of each of the valuations that satisfy ϕ at different points in time. Even using methods that exploit the structure of the DBN directly, the information of the context formula is not considered. Additionally, with the most efficient methods to-date, it is unclear how to handle the evidence over time effectively. Dealing with these, and other related problems, will be the main focus of our future work.

Acknowledgments

İsmail İlkan Ceylan is supported by DFG within the Research Training Group “RoSI” (GRK 1907). Rafael Peñaloza was partially supported by DFG through the Cluster of Excellence ‘cfAED,’ while he was still affiliated with TU Dresden and the Center for Advancing Electronics Dresden, Germany.

References

- Franz Baader and Rafael Peñaloza. Automata-based Axiom Pinpointing. *J. of Automated Reasoning*, 45(2):91–129, 2010.
- Franz Baader and Rafael Peñaloza. Axiom Pinpointing in General Tableaux. *J. of Logic and Computation*, 20(1):5–34, 2010.
- Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the \mathcal{EL} envelope. In *Proc. of IJCAI’05*, pages 364–369. Morgan Kaufmann Publishers, 2005.
- Franz Baader, Diego Calvanese, Deborah L McGuinness, Daniele Nardi, and Peter F Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2nd edition, 2007.
- Franz Baader, Martin Knechtel, and Rafael Peñaloza. Context-Dependent Views to Axioms and Consequences of Semantic Web Ontologies. *J. of Web Semantics*, 12–13:22–40, 2012.
- İsmail İlkan Ceylan and Rafael Peñaloza. Bayesian Description Logics. In *Proc. of DL’14*, volume 1193 of *CEUR Workshop Proceedings*. CEUR-WS, 2014.
- İsmail İlkan Ceylan and Rafael Peñaloza. The Bayesian Description Logic \mathcal{BEL} . In *Proc. of IJCAR’14*, volume 8562 of *LNCS*, pages 480–494. Springer Verlag, 2014.
- İsmail İlkan Ceylan and Rafael Peñaloza. Tight Complexity Bounds for Reasoning in the Description Logic \mathcal{BEL} . In *Proc. of JELIA’14*, volume 8761 of *LNCS*, pages 77–91. Springer Verlag, 2014.
- Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009.
- Thoma Dean and Keiji Kanazawa. A model for reasoning about persistence and causation. *Computational intelligence*, pages 142–150, 1989.
- Theodore Edward Harris. The existence of stationary measures for certain Markov processes. In *Proc. of Berkeley Symposium on Mathematical Statistics and Probability*, volume 2. University of California Press, 1956.
- Manfred Jaeger. Probabilistic Reasoning in Terminological Logics. In *Proc. of KR’94*, pages 305–316. AAAI Press, 1994.
- Manfred Jaeger. Relational Bayesian Networks. In *Proc. of UAI’97*, pages 266–273. Morgan Kaufmann Publishers, 1997.
- Aditya Kalyanpur, Bijan Parsia, Matthew Horridge, and Evren Sirin. Finding All Justifications of OWL DL Entailments. In *Proc. of ISWC’07*, volume 4825 of *LNCS*, pages 267–280. Springer Verlag, 2007.
- Pavel Klinov and Bijan Parsia. Representing sampling distributions in P-SROIQ. In *Proc. of URSW’11*, volume 778 of *CEUR Workshop Proceedings*. CEUR-WS, 2011.
- Chang-Yeong Lee. Representation of Switching Circuits by Binary-Decision Programs. *Bell System Technical Journal*, 38(4), 1959.
- Thomas Lukasiewicz and Umberto Straccia. Managing uncertainty and vagueness in description logics for the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):291–308, November 2008.
- Carsten Lutz and Lutz Schröder. Probabilistic Description Logics for Subjective Uncertainty. In *Proc. of KR’10*, pages 393–403. AAAI Press, 2010.
- Kevin Patrick Murphy. *Dynamic bayesian networks: representation, inference and learning*. PhD thesis, University of California, Berkeley, 2002.
- Heping Pan, Daniel McMichael, and Marta Lendjel. Inference Algorithms in Bayesian Networks and the Probanet System. *Digital Signal Processing*, 8(4):231–243, 1998.
- Adam Sadilek and Henry Kautz. Recognizing multi-agent activities from GPS data. In *Proc. of AAAI’10*, pages 1134–1139, 2010.
- Stephan Tobies. The Complexity of Reasoning with Cardinality Restrictions and Nominals in Expressive Description Logics. *J. of Artificial Intel. Research*, 12:199–217, 2000.
- Jonas Vlasselaer, Wannes Meert, Guy Van Den Broeck, and Luc De Raedt. Efficient Probabilistic Inference for Dynamic Relational Models. In *Proc. of StarAI’14*, volume WS-14-13 of *AAAI Workshops*. AAAI Press, 2014.