Axiom-based Probabilistic Description Logic

Martin Unold[1][a] and Christophe Cruz[1][b]
[1] *Laboratoire d'Informatique de Bourgogne – EA 7534 – 9 avenue Alain Savary – BP 47870, 21078 Dijon Cedex – France*
martin.unold@hs-mainz.de, christophe.cruz@ubfc.fr

Keywords: probabilistic description logic

Abstract: The paper proposes a new type of probabilistic description logics with a different interpretation of uncertain knowledge. The basic idea is that the probability of an axiom is not the probability of the axiom to be true in contrast to be false. Instead, it is the probability of the axiom to be true within the same knowledge base, i.e. in contrast to other axioms of the knowledge base to be true. The proposed description logic is evaluated with some sample knowledge bases and the results are discussed in the paper.

1 INTRODUCTION

The management of uncertainty in description logics has received a lot of attention, due to the development of the semantic web. There should be a level of trust for every piece of information found online and when combining this information to gain further implicitly stored knowledge, the involved uncertainty should be taken into account. Therefore probabilistic description logics have been developed in a couple of variants. Usually they see each probabilistic axiom within the same knowledge base separately from all others, i.e. the probability of that axiom should be valid in any scenario.

Nevertheless, in some cases, the probability is only estimated or the source of uncertainty allows only to state a general value for the uncertainty of all the information gained from that source. For example, the two statements

- A is true with a probability of 90%
- A is false with a probability of 10%

work perfectly together and would result in a consistent knowledge base. If the statements are changed to

- A is true with a probability of 90%
- A is false with a probability of 50%

they do not hold in a probabilistic knowledge base. Only some possibilistic approaches would be able to handle them. However, the two statements

- A is true with a probability of 90%
- A is false with a probability of 90%

would not work together in any case. Even though this looks like a huge contradiction, such cases might appear in several scenarios. For instance, when merging two crisp knowledge bases $\mathcal{K}_1$ and $\mathcal{K}_2$, $\mathcal{K}_1$ might contain one of these statements and $\mathcal{K}_2$ the other one. For sure, they are inconsistent together and therefore at least one of the knowledge bases must contain false information. If it is known, that approximately 90% of the knowledge in $\mathcal{K}_1$ and $\mathcal{K}_2$ is correct, respectively, one might not have any other choice than assigning 90% to each axiom of both knowledge bases to express the probability. The trust of 90% could be a result from earlier observations about the knowledge from the same source.

Generally, it might be a more realistic assumption that there exists only one correct world and not several possible worlds, each with a certain probability or possibility. Indeed, it is the knowledge about the correct world, which might be wrong sometimes, rather than the world itself being true only with a certain probability or possibility. For instance, a paternity testing has an accuracy of 99.99%. This value is determined by observations when applying the test to people with known paternity. Hence, a statement like "John is the father of Mary with a probability of 99,99%" does not state, that in 9999 of 10000 a parallel universes the statement is true and in 1 it is false. It rather states, that 9999 of 10000 comparable parenting testings are correct and 1 is incorrect. But there is only one world.

[a] https://orcid.org/0000-0003-2913-2421
[b] https://orcid.org/0000-0002-5611-9479

In such cases, the probability does not state the probability of a statement to be true in contrast to being false or unknown. It rather states the probability of a statement to be true in contrast to the other statements in the same environment or from the same source. This type of interpretation is neither handled by probabilistic nor by possibilistic description logics. This paper presents a different way of dealing with knowledge in such situations. It starts with a related work and a preliminaries section, where basically other ways of creating probabilistic or possibilistic knowledge bases are presented, respectively. The novel approach is then explained in section 4 and section 5 shows in which cases it is more useful. Finally, a conclusion summarizes the paper and gives an outlook to further developments.

2 RELATED WORK

A description logic is used as a formalism to represent knowledge about a special domain. It easily allows to link other information to knowledge in the semantic web, as explained in (Baader et al., 2005). Reasoning programs allow to infer implicitly stored information from explicit knowledge and inference rules. These programs can also decide if a specific part of the knowledge is logically consistent or not. Common decision problems also include those, that are typically used for queries on relational databases, such as instance or relation checking.

There exists a huge variety of description logics with different expressiveness. An overview about this variety is shown in (Zolin, 2013). The expressiveness is determined by the amount of symbols allowed in the syntax of a particular description logic. In this section, the most common and the ones that are relevant in the following chapters are explained. A more detailed introduction to description logics can be found in (Baader et al., 2003).

The extension of a description logic by the management of uncertainty results in a probabilistic description logic. In contrast to a classical one, which is also called crisp. Regardless of the expressiveness, each type of description logic could be extended to a probabilistic one. When modeling a probabilistic description logic, the knowledge about the validity of the axioms is limited, i.e. it is unknown if an axiom is true or false, yet the axioms themselves are clearly true or false and nothing in between. An introduction to probabilistic logic can be found in (Nilsson, 1986).

Probability and fuzziness (also called vagueness) seem very similar on first sight, since both contain values in the range between 0 and 1 attached to crisp axioms in the knowledge base of a description logic. Nevertheless, they both work completely different and a detailed clarification about the differences is done in (Dubois and Prade, 2001). An introduction to both types of extensions of description logics is done in (Lukasiewicz and Straccia, 2008). The structure of a probabilistic description logic is rather the same as in the crisp case and the probabilistic interpretation maps all possible crisp interpretations to a certain probability.

A major problem with probabilistic description logics is the complexity (Lukasiewicz, 2008), in general cases computations require exponential time, because many possible worlds need to be taken into account. Hence, a lot of research focused on the reduction of complexity, such as (Klinov, 2008), or to find more efficient solutions for more specific cases, such as (Riguzzi et al., 2015). The latter approach assumes that all axioms within a knowledge base are independent of each other, while in this paper all axioms within a knowledge base are assumed to be highly dependent. In (Niepert et al., 2011) a log-linear model is used to do reasoning on uncertain knowledge.

One of the first attempts to add the management of uncertainty to description logics is done by (Koller et al., 1997), which was developed further by (Giugno and Lukasiewicz, 2002), where the complexity of reasoning algorithms has been analyzed. The development of reasoning programs has been subject of several studies (Lukasiewicz, 2007) and also the development of standards for the semantic web is of importance for applications. The management of uncertainty could be integrated to the semantic web by extensions of OWL or RDF. Examples are found in (Ding et al., 2006). Unfortunately, there is no W3C-standard available till now (Carvalho et al., 2017). Anyway, the solution for dealing with probabilities in description logics presented in this paper should be treated carefully when combining with other data, that use a different understanding of uncertainty.

3 PRELIMINARIES

Each description logic allows the existence of knowledge bases, where all the information about a specific domain is stored. Typically, the knowledge is divided into two parts, which are called the terminological axioms (TBox) and the assertional axioms (ABox). The TBox consists of general rules, such as "Each City has a Location". The ABox consists of specific information for single instances, such as "Paris is a City". The reasoner should then be able to infer the implicit information, that "Paris has a Location".

A knowledge base $\mathcal{K}$ is a collection of axioms. Usually, they are divided into two parts, so that a knowledge base becomes a tuple $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. It consists of a *TBox* $\mathcal{T}$ and an *ABox* $\mathcal{A}$. A TBox $\mathcal{T}$ is a set of general concept inclusions and an ABox $\mathcal{A}$ is a set of assertional axioms. A general concept inclusion is an expression of the format $C \sqsubseteq D$, where $C, D \in \mathbf{C}$ are concepts. An assertional axiom is either an expression of the format $a : C$ (concept assertion) or of the format $(a, b) : R$ (role assertion), where $a, b \in N_I$ are named individuals, $C \in \mathbf{C}$ is a concept and $R \in N_R$ is a role.

In this context, an ordered triple of arbitrary pairwise disjoint sets (N_C, N_R, N_I) is called a signature. The sets contain named concepts, named roles and named individuals, respectively. The set of concepts $\mathbf{C}$ over a signature (N_C, N_R, N_I) is the smallest possible set, that fulfills some of the following conditions.

$$
\begin{aligned}
C \in N_C &\implies C \in \mathbf{C} \\
&\quad\ \top \in \mathbf{C} \\
&\quad\ \bot \in \mathbf{C} \\
C_1, C_2 \in \mathbf{C} &\implies C_1 \sqcup C_2 \in \mathbf{C} \\
C_1, C_2 \in \mathbf{C} &\implies C_1 \sqcap C_2 \in \mathbf{C} \\
C \in \mathbf{C} &\implies \neg C \in \mathbf{C} \\
C \in \mathbf{C}, R \in N_R &\implies \forall R.C \in \mathbf{C} \\
C \in \mathbf{C}, R \in N_R &\implies \exists R.C \in \mathbf{C}
\end{aligned}
$$

Each element $C \in \mathbf{C}$ is called a concept and it depends on the expressiveness of the description logic, if even more conditions must be fulfilled, i.e. more symbols are allowed. The description logic $\mathcal{ALC}$ allows only atomic roles and the introduced concepts. The basic version of $\mathcal{ALC}$ neither allows role inclusion axioms nor general concept inclusions, i.e. only assertional axioms are allowed. Description logics with more expressiveness could allow further roles, concepts and even an entirely different additional formalism.

So far, it is only a collection of symbols. The interesting part is the semantics of these axioms. The meaning of each axiom is expressed by stating the membership to each concept and role. These statements are done by an interpretation, which maps the signature to some elements of the universe of discourse, which is the domain of the knowledge, that is modeled in a particular knowledge base.

Formally, an interpretation $I = (\Delta^I, \cdot^I)$ over a signature (N_C, N_R, N_I) consists of a set Δ^I, which is called domain or universe of discourse, and a function $\cdot^I$, that maps ...

- ... each named individual $a \in N_I$ to an element of the domain $a^I \in \Delta^I$.
- ... each concept $C \in \mathbf{C}$ to a subset of the domain $C^I \subseteq \Delta^I$.
- ... each role $R \in \mathbf{R}$ to a set of tuples of domain elements $R^I \in \Delta^I \times \Delta^I$.

The interpretation function $\cdot^I$ must also fulfill the following conditions.

$$
\begin{aligned}
\delta \in \top^I &\iff \delta \in \Delta^I \\
\delta \in \bot^I &\iff \delta \notin \Delta^I \\
\delta \in (C_1 \sqcup C_2)^I &\iff \delta \in C_1^I \vee \delta \in C_2^I \\
\delta \in (C_1 \sqcap C_2)^I &\iff \delta \in C_1^I \wedge \delta \in C_2^I \\
\delta \in (\neg C)^I &\iff \delta \notin C^I \\
\delta \in (\forall R.C)^I &\iff \\
&\quad \forall \delta_0 \in \Delta^I, (\delta, \delta_0) \in R^I : \delta_0 \in C^I \\
\delta \in (\exists R.C)^I &\iff \\
&\quad \exists \delta_0 \in \Delta^I, (\delta, \delta_0) \in R^I : \delta_0 \in C^I
\end{aligned}
$$

Of course, the conditions depend on the existence of the corresponding type of concepts and roles, i.e. only those conditions must be fulfilled, where the corresponding type of concepts and roles are allowed by the description logic.

A knowledge base $\mathcal{K}$ is free of contradictions, if there exists an interpretation I, which is a model for all axioms of the knowledge base. An interpretation I is a model of a general concept inclusion $C \sqsubseteq D$ (denoted: $I \models (C \sqsubseteq D)$) if and only if $C^I \subseteq D^I$. An interpretation I is a model of a role assertion $(a, b) : R$ (denoted: $I \models ((a, b) : R)$) if and only if $(a^I, b^I) \in R^I$. An interpretation I is a model of a concept assertion $a : C$ (denoted: $I \models (a : C)$) if and only if $a^I \in C^I$. An interpretation I is a model of a TBox $\mathcal{T}$ (denoted: $I \models \mathcal{T}$) if and only if $I \models \phi$ for all general concept inclusions $\phi \in \mathcal{T}$. An interpretation is a model of an ABox $\mathcal{A}$ (denoted: $I \models \mathcal{A}$) if and only if $I \models \phi$ for all assertional axioms $\phi \in \mathcal{A}$. An interpretation I is a model of a knowledge base $\mathcal{K}$, if I is a model of all its boxes. Sometimes the phrase "in I holds" is used instead of "I is a model of".

For a probabilistic description logic, the set of possible worlds

$$W = \{I \mid I \text{ is an interpretation over } (N_C, N_R, N_I)\}$$

consists of all interpretations over a given signature (N_C, N_R, N_I). An element of this set $I \in W$ is called a possible world. A mapping $\pi : W \to [0,1]$ with

$$\sum_{I \in W} \pi(I) = 1$$

is called a probability distribution of possible worlds over (N_C, N_R, N_I). It has to be noted, that the definition of W and π is not dependent on a given interpretation or knowledge base, only an elementary description is necessary. This is important, especially when it comes to the development of algorithms, since there are a lot of possible worlds on first attempt.

An uncertain knowledge base will be introduced later to restrict π. Sometimes π is not defined as a probability distribution and the constraint $\sum_{I \in W} \pi(I) = 1$ is replaced by $\max_{I \in W} \pi(I) = 1$. With that constraint, one gets a possibility value for each world to be the true one.

A probabilistic knowledge base $\mathcal{K}$ is a collection of axioms. They could be probabilistic, but also crisp axioms, i.e. with a probability of 1. As in the crisp case, the knowledge base is divided into two parts $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. It consists of a *TBox* $\mathcal{T}$ and an *ABox* $\mathcal{A}$. A TBox $\mathcal{T}$ is a set of (probabilistic) general concept inclusions and an ABox $\mathcal{A}$ is a set of (probabilistic) assertional axioms. A probabilistic general concept inclusion is an expression of the format $p :: C \sqsubseteq D$, where $C, D \in \mathbf{C}$ are concepts and $p \in [0,1]$ is the probability of the axiom to be true. A probabilistic assertional axiom is either an expression of the format $p :: (a : C)$ (concept assertion) or of the format $p :: ((a,b) : R)$ (role assertion), where $a, b \in N_I$ are named individuals, $C \in \mathbf{C}$ is a concept, $R \in N_R$ is a role and $p \in [0,1]$ is the probability of the axiom to be true. For $p = 1$, an axiom has exactly the crisp meaning. An axiom with $p = 0$ is obsolete because it does not contain any information.

The set of concepts $\mathbf{C}$ is defined in exactly the same way as in the crisp case. Also the interpretation I is defined in exactly the same way. The only difference is, that it is called a possible world in the context of probabilistic description logics.

A probabilistic knowledge base $\mathcal{K}$ is free of contradictions, if there exists a probability distribution of possible worlds π, which is a model for all axioms of the probabilistic knowledge base. A probability distribution of possible worlds π is a model of a model of a probabilistic general concept inclusion $p :: C \sqsubseteq D$ (denoted: $\pi \models p :: C \sqsubseteq D$) if and only if $\sum_{I \models C \sqsubseteq D} \pi(I) = p$. It is a model of a probabilistic role assertion $p :: (a,b) : R$ (denoted: $\pi \models p :: (a,b) : R$) if and only if $\sum_{I \models (a,b):R} \pi(I) = p$. And it is a model of a probabilistic concept assertion $p :: a : C$ (denoted: $I \models p :: a : C$) if and only if $\sum_{I \models a:C} \pi(I) = p$. A probability distribution of possible worlds π is a model of a TBox $\mathcal{T}$ (denoted: $\pi \models \mathcal{T}$) if and only if $\pi \models p :: \phi$ for all (probabilistic) general concept inclusions $p :: \phi \in \mathcal{T}$. And it is a model of an ABox $\mathcal{A}$ (denoted: $\pi \models \mathcal{A}$) if and only if $\pi \models p :: \phi$ for all (probabilistic) assertional axioms $p :: \phi \in \mathcal{A}$. A probability distribution of possible worlds π is a model of a probabilistic knowledge base $\mathcal{K}$, if it is a model of all its boxes.

It has to be noted, that there are also other options to define a probabilistic description logic. There is also the similar possibilistic logic, which allows to state possibilities and necessities for each axiom and the probability distribution is replaced by the maximum possibility for each world. These options are not part of further discussion at this point, since the proposal of this paper is another option, anyway.

4 PROPOSAL

In contrast to the probabilistic description logics introduced in the previous sections, the probability attached to an axiom $p :: \phi$ is interpreted differently in this proposal. Here, p is the probability of ϕ being true not in contrast to being false or unknown otherwise, but to be true as compared to other axioms of the same knowledge base.

An axiom-based probabilistic knowledge base is a tuple $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, that consists of probabilistic general concept inclusions $(p :: C \sqsubseteq D) \in \mathcal{T}$, probabilistic concept assertions $(p :: a : C) \in \mathcal{A}$ and probabilistic role assertions $(p :: (a,b) : R) \in \mathcal{A}$. To ease the notation, ϕ is denoted for any type of crisp axiom and $(p :: \phi) \in \mathcal{K}$ is used to state that an axiom is part of a knowledge base, although the axioms are members of $\mathcal{T}$ or $\mathcal{A}$.

The basic idea of this proposal is, that within an axiom-based probabilistic knowledge base, the ratio between correct and incorrect axioms should be according to their probability, respectively. Therefore, the sum of all remaining probabilities of all correct statements should be equal (or greater than) the sum of all probabilities of all incorrect statements, which is measured as confidence. A confidence function $\sigma : W \rightarrow \mathbb{R}$ assigns to every possible world $I \in W$ a value that states the confidence of the world to be the true one.

$$\sigma(I) = \frac{1}{|\mathcal{K}|} \cdot \left(\sum_{\substack{p::\phi \in \mathcal{K} \\ I \models \phi}} (1-p) - \sum_{\substack{p::\phi \in \mathcal{K} \\ I \not\models \phi}} p \right) \quad (1)$$

The value for confidence of an interpretation ranges from -1, which indicates a totally wrong interpretation, to 1. A value of around 0 or maybe a bit higher indicates an interpretation, that is very likely to be the true one. If the value is very high, i.e. close to 1, this is rather an indicator of a not well designed knowledge base, that is not restrictive enough.

As an example, consider the knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ with the following axioms.

$$\begin{aligned} \mathcal{T} &= \{0.75 :: \mathtt{C} \sqsubseteq \mathtt{D}\} \\ \mathcal{A} &= \{0.75 :: \mathtt{a} : \mathtt{C}, \\ &\quad 0.75 :: \mathtt{a} : \mathtt{C} \sqcup \mathtt{D}, \\ &\quad 0.75 :: \mathtt{a} : \neg\mathtt{D}\} \end{aligned}$$

The knowledge base consists of 4 axioms. To each a probability of 0.75 is assigned, which means that in average one of these axioms is false in the way of interpreting probability within this paper. It has to be noted, that each probability could be any value between 0 and 1. For the example knowledge base, there exist four different interpretations.

$$\begin{aligned} I_1 &: \quad \mathtt{a}^{I_1} \in \mathtt{C}^{I_1}, \mathtt{a}^{I_1} \in \mathtt{D}^{I_1} \\ I_2 &: \quad \mathtt{a}^{I_2} \in \mathtt{C}^{I_2}, \mathtt{a}^{I_2} \notin \mathtt{D}^{I_2} \\ I_3 &: \quad \mathtt{a}^{I_3} \notin \mathtt{C}^{I_3}, \mathtt{a}^{I_3} \in \mathtt{D}^{I_3} \\ I_4 &: \quad \mathtt{a}^{I_4} \notin \mathtt{C}^{I_4}, \mathtt{a}^{I_4} \notin \mathtt{D}^{I_4} \end{aligned}$$

The confidence of each interpretation in $W = \{I_1, I_2, I_3, I_4\}$ can be calculated in the following way.

$$\begin{aligned} \sigma(I_1) &= \frac{1}{4}(0.25 + 0.25 + 0.25 - 0.75) = 0 \\ \sigma(I_2) &= \frac{1}{4}(-0.75 + 0.25 + 0.25 + 0.25) = 0 \\ \sigma(I_3) &= \frac{1}{4}(0.25 - 0.75 + 0.25 - 0.75) = -0.25 \\ \sigma(I_4) &= \frac{1}{4}(0.25 - 0.75 - 0.75 + 0.25) = -0.25 \end{aligned}$$

This solution shows that either I_1 or I_2 are very likely to be the correct worlds. It also gives a hint to which axioms might be the wrong ones in the knowledge base (here it is $\mathtt{C} \sqsubseteq \mathtt{D}$ or $\mathtt{a} : \neg\mathtt{D}$).

In a classical probabilistic description logic, this example would be considered inconsistent. Also, a possibilistic logic would either suffer from the drowning problem and remove all axioms from the knowledge base or if other strategies are used still not find a totally possible world.

There are several algorithms to handle probabilistic logics. The consideration of all possible worlds is very time consuming, since the amount of possible worlds increases exponentially in the amount of axioms in the knowledge base. Therefore some more efficient algorithms have been designed, they have disadvantages sometimes and might not work for an arbitrary knowledge base. For the proposed description logic, an optimization problem has to be solved, since the goal is to find the possible world with the highest confidence. Therefore the development of suitable algorithms, such as evolutionary algorithms, is subject to future work. Nevertheless, it might be easier to find an efficient solution for an optimization problem as compared to solve a huge linear system of equations, which is required for (classical) probabilistic description logics.

If all possible worlds have a negative confidence, the knowledge base might be inconsistent, i.e. the information within the axiom-based probabilistic knowledge base is wrong with respect to the probability values. A small negative value for the highest confidence might still be called consistent, since an optimal designed axiom-based probabilistic knowledge base should have a confidence value of $\sigma(I) \approx 0$.

5 EVALUATION

To test the quality of this approach, a crisp knowledge base is used, which is called ”the solution” in this section. It is a complete knowledge base, i.e. there is no further information inferable and it contains a lot of redundant knowledge. From this complete knowledge base, each axiom is changed to its negation with a probability of $(1-p)$ and $p \in [0,1]$ is assigned to each crisp axiom ϕ, so that it becomes a probabilistic axiom $p :: \phi$. The result is a damaged knowledge base, where the probability of an axiom to be false is p. By that, the knowledge base becomes a probabilistic one, that exactly meets the requirements to use the proposed approach.

For the requirements of this test, a simple random optimization algorithm is used, which tries to find a good solution by randomly assuming some axioms as true or false. It works in the following way.

```
maximumConfidenceApprox(KB) {
  I = empty();
  c = confidence(I);
  repeat {
    axiom = KB.getRandomAxiom();
    I.add|I.remove(axiom);
    if (I.consistent() & KB.confidence(I) > c)
      c = KB.confidence(I);
    else
      I.remove|I.add(axiom);
  } until c did not changed for all axioms
  return c;
}
```

For very small knowledge bases, an exact algorithm that determines the confidence for every possible world, is also used.

```
maximumConfidenceExact(KB) {
  c = -1;
  for (I subset KB) {
    if (I.consistent() & KB.confidence(I) > c)
      c = KB.confidence(I);
  }
  return c;
}
```

The aim of these algorithms is to find an interpretation I with the highest confidence value for the damaged knowledge base KB. The hope is, that this interpretation is also an interpretation for the correct solution. However, in many cases there exists another interpretation with a higher confidence value.

The algorithms are executed on knowledge bases with varying parameter $p \in (0.5, 1)$. The parameter p is only assigned to all assertional axioms of the knowledge base and not to general concept inclusions, since the assignment to general concept inclusions involves many other axioms at the same time and therefore it is more likely to be considered false if its probability is less than 1. Also, the value p is only displayed in a range down to 50%, since a knowledge base that contains only axioms that are more likely to be false than to be true is rather useless, anyway. The assumed correct solution, i.e. the solution before damaging the knowledge base, is always set to a confidence value of 0, since differences only appear randomly.

Figure 1 depicts the difference between the correct solution and the one with the highest confidence value. It is determined by the exact algorithm, since the knowledge bases consist of only 100 and 200 axioms, respectively. Therefore the best solution is always at least as good as the correct one. The horizontal axis holds the value p for an axiom of the knowledge base to be true. For instance, if $p = 0.9$ for a knowledge base of 100 axioms, then in average 90 of these 100 axioms are correct and 10 are incorrect.

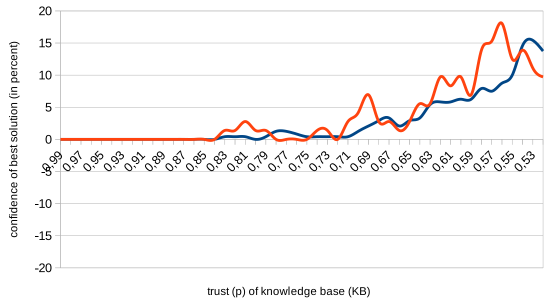

Figure 1: The difference of confidence (in percent) between best and correct solution for different probabilities in small knowledge bases of 100 (red) and 200 (blue) axioms.

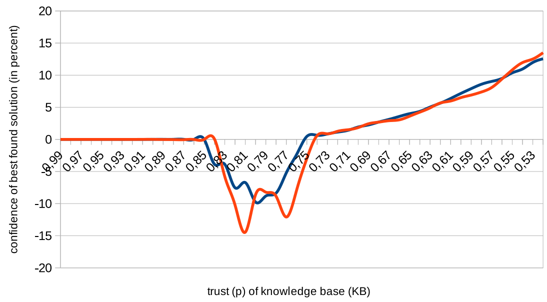

Figure 2: The difference of confidence (in percent) between best and correct solution for different probabilities in large knowledge bases of 10000 (red) and 20000 (blue) axioms.

A different result is depicted in figure 2, where larger knowledge bases with 10000 and 20000 axioms are used, respectively. The exact algorithm is not able to terminate within an appropriate time, thus the heuristic algorithm has to be used. One can notice, that it has more difficulties finding a good solution if there are many false axioms in the knowledge base, which results in a confidence value even worse than the one of the correct solution, although there might be even a solution with higher confidence than the correct one. Nevertheless, the results show that the approach might work well for knowledge bases with high trust, i.e. only a little amount of incorrect information. In this case, it can detect the wrong information quite well.

For large knowledge bases with little trust, the results show an increasing confidence for the best solution. Therefore it might be very likely that there exist even better solutions in many cases, which are simply not found by the algorithm. More sophisticated heuristics should be developed to improve the results. Nevertheless, unfortunately the correct world is very likely to not be found, since there are other worlds with higher confidence.

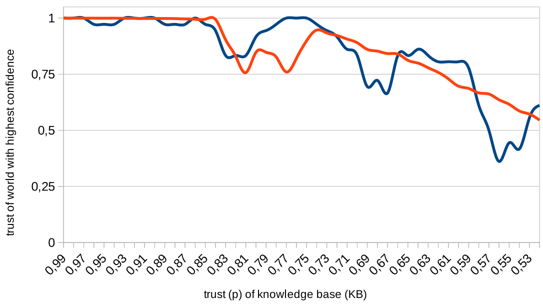

Figure 3: The level of trust for the best solution for different probabilities in knowledge bases of 100 (blue) and 10000 (red) axioms.

Another interesting aspect of this evaluation is shown in figure 3. It depicts the amount of correct information compared to the assumed correct solution for the solution with the highest confidence value. It clearly shows, that if a solution with higher confidence than the correct one is found, at least the trust of this new knowledge base is higher than for the starting point. Especially for an intermediate trust value of around 60% to 75%, the improvement in trust is noticeable. Probably the improvement is also better for higher trust values p, if a better algorithm for the optimization problem is used. An additional observation is, that the stability of values for large knowledge bases is higher as compared to smaller knowledge bases, i.e. the experimental results suffer from high volatility for small datasets.

6 CONCLUSION

The paper presented a novel approach to deal with probabilistic description logics. By not interpreting each axiom separately, but a whole knowledge base together, it enables results for cases, where other approaches would consider the knowledge base as inconsistent. Also the computation time can be highly reduced as compared to probabilistic description logics by using optimization algorithms, which should be subject to further research.

The disadvantage of the presented approach is, that axioms can not be treated independently, i.e. the whole knowledge base must be kept together. Another problem is that the probabilities of all axioms are assumed to be independent of each other, i.e. there is no space for redundancy. Nevertheless, the assumption of having only one correct world instead of assuming the existence of several possible worlds seems more appropriate and realistic in many application scenarios. Therefore a further development of the presented idea seems useful.

For the evaluation, another interesting aspect might be to consider also T-Box axioms to be uncertain. Therefore another strategy for evaluation should be developed, e.g. a different or separate treatment of A-Box and T-Box axioms (Van Asch, 2013). For this purpose, it might also be useful to change or parameterize formula (1) to determine the confidence of an interpretation for a given knowledge base in a different way.

ACKNOWLEDGEMENTS

We would like to thank the Deutsche Forschungsgemeinschaft (DFG) and the agence nationale de la recherche (ANR) for the support of the TEXTELSEM project, in which this work has been done.

REFERENCES

Baader, F., Calvanese, D., McGuinness, D., Patel-Schneider, P., and Nardi, D. (2003). *The description logic handbook: Theory, implementation and applications*. Cambridge university press.

Baader, F., Horrocks, I., and Sattler, U. (2005). Description logics as ontology languages for the semantic web. In *Mechanizing mathematical reasoning*, pages 228–248. Springer.

Carvalho, R. N., Laskey, K. B., and Costa, P. C. (2017). Pr-owl–a language for defining probabilistic ontologies. *International Journal of Approximate Reasoning*, 91:56–79.

Ding, Z., Peng, Y., and Pan, R. (2006). Bayesowl: Uncertainty modeling in semantic web ontologies. In *Soft computing in ontologies and semantic web*, pages 3–29. Springer.

Dubois, D. and Prade, H. (2001). Possibility theory, probability theory and multiple-valued logics: A clarification. *Annals of mathematics and Artificial Intelligence*, 32(1-4):35–66.

Giugno, R. and Lukasiewicz, T. (2002). P-shoq (d): A probabilistic extension of shoq (d) for probabilistic ontologies in the semantic web. In *JELIA*, volume 2, pages 86–97. Springer.

Klinov, P. (2008). Pronto: A non-monotonic probabilistic description logic reasoner. In *European Semantic Web Conference*, pages 822–826. Springer.

Koller, D., Levy, A., and Pfeffer, A. (1997). P-classic: A tractable probablistic description logic. *AAAI/IAAI*, 1997(390-397):14.

Lukasiewicz, T. (2007). Probabilistic description logic programs. *International Journal of Approximate Reasoning*, 45(2):288–307.

Lukasiewicz, T. (2008). Expressive probabilistic description logics. *Artificial Intelligence*, 172(6-7):852–883.

Lukasiewicz, T. and Straccia, U. (2008). 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.

Niepert, M., Noessner, J., and Stuckenschmidt, H. (2011). Log-linear description logics. In *Twenty-Second International Joint Conference on Artificial Intelligence*.

Nilsson, N. J. (1986). Probabilistic logic. *Artificial intelligence*, 28(1):71–87.

Riguzzi, F., Bellodi, E., Lamma, E., and Zese, R. (2015). Probabilistic description logics under the distribution semantics. *Semantic Web*, 6(5):477–501.

Van Asch, V. (2013). Macro-and micro-averaged evaluation measures [[basic draft]]. *Belgium: CLiPS*, pages 1–27.

Zolin, E. (2013). Complexity of reasoning in description logics.