Semantic Web Uncertainty Management

Volker Haarslev
Concordia University, Canada

Hsueh-leng Pai
Concordia University, Canada

Nematollaah Shiri
Concordia University, Canada

INTRODUCTION

Since the introduction of the *Semantic Web* vision (Berners-Lee, Hendler, & Lassila, 2001), attempts have been made for making Web resources more machine interpretable by giving them a well-defined meaning through semantic markups. One way to encode such semantic markups is to use ontologies. An *ontology* is "an explicit specification of a conceptualization" (Gruber, 1993). Informally, an ontology consists of a set of terms in a domain, relationships between the terms, and a set of constraints on the way in which those terms can be combined. By explicitly defining the relationships and constraints among the terms, the semantics of the terms can be better defined and understood.

Over the last few years, a number of ontology languages have been developed, most of which use *Description Logics* (DLs) (Baader, McGuinness, Nardi, & Schneider, 2003) as the foundation. The family of DLs is a subset of first-order logic (FOL) and is considered to be attractive as it keeps a good compromise between expressive power and computational tractability.

Uncertainty is a form of deficiency or imperfection in the information/data, where the truth of information is not established definitely. Uncertainty modeling and reasoning have been challenging issues for over two decades in many disciplines, such as database and artificial intelligence. Most of the information in the real world is uncertain or imprecise, for example, classifications of genes in bioinformatics, schema matching in information integration, finding best matches in a Web search, and so forth. Therefore, uncertainty management is essential for the success of many such applications and in particular DLs and the Semantic Web.

Despite its popularity, it has been realized that classical DLs are inadequate to model uncertainty. For example, in the medical domain, one might want to express that: "It is very likely that an obese person would have heart disease," where "obese" is a vague concept that may vary across regions and "likely" shows the uncertain nature of this information. Such an expression cannot be expressed using classical DLs.

The importance of incorporating uncertainty in DLs has been recognized by the knowledge representation community: "modeling primitives such as … fuzzy/probabilistic definitions" could be the next step for extension (Horrocks et al., 2000). For this, a number of frameworks have been proposed to incorporate uncertainty in DLs. This paper provides a survey of these proposals.

The rest of this paper is organized as follows. We first provide the background on the classical DL framework. We then study representative extensions of DLs with uncertainty. This follows by some possible research directions for incorporating uncertainty in the Semantic Web. We conclude with a summary and some remarks.

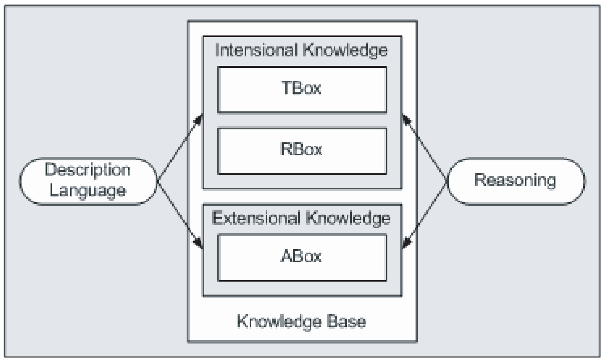
BACKGROUND

In this section, we review the basics of the classical DL framework, which provides facilities to represent knowledge bases and to reason about them.

As shown in Figure 1, the classical DL framework consists of three components:

1. **Description Language.** All description languages have elementary descriptions which include atomic

Figure 1. Classical DL framework



concepts (unary predicates) and atomic roles (binary predicates). Complex descriptions are built inductively from atomic ones using concept constructors. In this work, we focus on the description language $\mathcal{ALC}$ (Baader et al., 2003). Let C and D be concept descriptions. $\mathcal{ALC}$ includes atomic concepts A, atomic roles R, top/universal concept $\top$, bottom concept $\bot$, concept negation $\neg C$, concept conjunction $C \sqcap D$, concept disjunction $C \sqcup D$, role value restriction $\forall R.C$ (meaning $\forall y$: $R(x,y) \rightarrow C(y)$, for x in the domain), and role exists restriction $\exists R.C$ (meaning $\exists y$: $R(x,y) \wedge C(y)$, for x in the domain).

2. **Knowledge Base (KB).** The KB is composed of both intensional knowledge and extensional knowledge (see Figure 1). The former includes the Terminological Box (TBox or $\mathcal{T}$) consisting of a set of terminological axioms that could be concept subsumptions $C \sqsubseteq D$ and/or concept definitions $C \equiv D$ (where C and D are concepts), and the Role Box (RBox or $\mathcal{R}$) consisting of a set of role axioms that could be role subsumptions $R \sqsubseteq S$ and/or role definitions $R \equiv S$ (where R and S are roles). The extensional knowledge includes the Assertional Box (ABox or $\mathcal{A}$) consisting of a set of assertions/facts that could be concept assertions a: C (i.e., a is an instance of concept C) and/or role assertions (a,b):R (i.e., individuals a and b are related through relationship R).
3. **Reasoning Component.** A DL framework is equipped with reasoning services which allows that implicit knowledge be derived from explicit knowledge.

DESCRIPTION LOGICS WITH UNCERTAINTY

In this section, we study existing frameworks for DLs with uncertainty. We first provide a classification of the approaches of these frameworks. We then study representative extensions of DLs with uncertainty.

Approaches to Extend Description Logics with Uncertainty

On the basis of their mathematical foundation and the type of uncertainty modeled, we can classify existing proposals of DLs with uncertainty into three approaches: fuzzy, probabilistic, and possibilistic.

The fuzzy approach, based on fuzzy set theory (Zadeh, 1965), deals with vagueness in the knowledge, where a proposition is true only to some degree. For example, the statement: "Jason is obese with degree 0.4" indicates Jason is slightly obese. Here, the value 0.4 is the degree of membership that Jason belongs to the fuzzy concept obese.

The probabilistic approach, based on classical probability theory, deals with the uncertainty due to lack of knowledge, where a proposition is either true or false, but one does not know for sure which one is the case. Hence, the certainty value associated with the proposition refers to the probability that the proposition is true. For example, one could say: "The probability that Jason would have heart disease, given that he is obese, lies in the range [0.8,1]."

Finally, the possibilistic approach, based on possibility theory (Zadeh, 1978), allows both certainty (necessity measure) and possibility (possibility measure) to be handled in the same formalism. For example, by knowing that "Jason's weight is above 80 kg," the proposition "Jason's weight is 80 kg" is necessarily true with certainty 1, while "Jason's weight is 90 kg" is possibly true with certainty 0.5.

Description Logics with Uncertainty—Current State

To incorporate uncertainty into DLs, each component of the DL framework needs to be extended (Figure 2). In what follows, we survey how the description language, the KB, and the reasoning component have been extended with uncertainty using fuzzy, probabilistic, and possibilistic approaches.

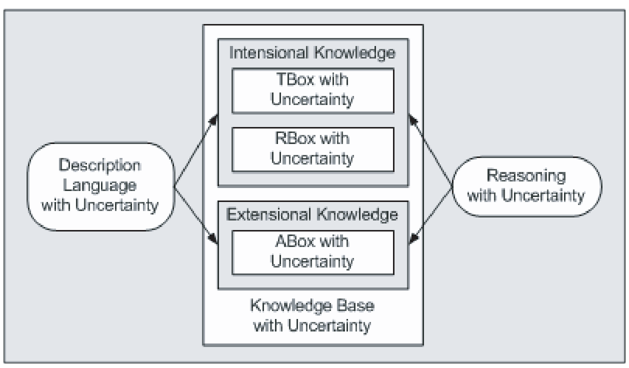
Description Languages with Uncertainty

The description languages contain a set of language constructors that serve as the building blocks of the description. In this section, we study how description languages have been extended using fuzzy and possibilistic approaches. To the best of our knowledge, no probabilistic extension of description languages has been proposed.

Fuzzy Description Languages

All existing proposals for fuzzy DL extend the semantics of the description language by fuzzifying their interpretation using fuzzy logic (Hölldobler, Khang, & Störr, 2002; Sánchez & Tettamanzi, 2004; Straccia, 2001, 2004a, 2004b,

Figure 2. DL Framework with uncertainty



2005a, 2005b; Tresp & Molitor, 1998). In general, a fuzzy interpretation $\mathcal{I}$ is a pair $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is the domain and $\cdot^{\mathcal{I}}$ is an interpretation function that maps language elements to some membership degree in [0,1]. For instance, the semantics of an atomic concept A is defined as $A^{\mathcal{I}}(a) \in [0,1]$ for all $a \in \Delta^{\mathcal{I}}$. That is, if individual a is an element of the domain, then the interpretation of the atomic concept A gives the membership degree that a belongs to A. The semantics of complex descriptions are defined in a straightforward way. For instance, the semantics of concept intersection is defined as $(C \sqcap D)^{\mathcal{I}}(a) = \min\{C^{\mathcal{I}}(a), D^{\mathcal{I}}(a)\}$, for all $a \in \Delta^{\mathcal{I}}$. That is, the certainty degree of concept C intersect D is the minimum of the certainty degrees of C and D.

In addition to the semantic extension, syntactic extensions to the description language are also proposed. The *manipulators/modifiers* (Hölldobler et al., 2002; Straccia, 2005a, 2005b; Tresp & Molitor, 1998) are unary operators that can modify the membership functions of the concepts they are applied to (e.g., "mostly," "very"), whereas *fuzzy quantifiers* (Sánchez & Tettamanzi, 2004) allow expressing vague quantities (e.g., "about 2") and quantity intervals (e.g., "roughly between 1 and 3").

Possibilistic Description Languages

Hollunder (1994) proposed a possibilistic extension to the description language. The idea is to keep the original description language syntax, while changing its interpretation using possibility theory. More specifically, the *possibility measure* Π for a concept (or event) C induced by a possibility distribution π on a set of interpretations $\Omega_{\perp}$ is defined as $\Pi(C) = sup\{\pi(\omega) \mid \omega \in \Omega_{\perp} \text{ and } \omega \models C\}$, characterizing the extent to which C is possible. On the other hand, the *necessity measure* N for a concept C is defined as $N(C) = inf\{1 - \pi(\omega) \mid \omega \in \Omega_{\perp} \text{ and } \omega \not\models C\}$, characterizing the extent to which this event is necessary or certain to occur. For example, the possibility that C and D are occurring at the same time is no more than the minimum of their possibilities, that is, $\Pi(C \sqcap D) \leq min\{\Pi(C), \Pi(D)\}$, but the necessity that both of them occur is $N(C \sqcap D) = min\{N(C), N(D)\}$.

Knowledge Base with Uncertainty

The fuzzy, probabilistic, and possibilistic extensions of KBs are defined in a similar way as the classical case, except that the interpretation $\mathcal{I}$ corresponds to the extension considered. An interpretation $\mathcal{I}$ *satisfies* (or is a *model* of) a KB Σ, denoted $\mathcal{I} \models \Sigma$, if it satisfies each element of Σ. Also, Σ is *satisfiable* if there exists an interpretation $\mathcal{I}$ that satisfies Σ. In this section, we study how each component of the KB can be extended with uncertainty.

Fuzzy Knowledge Base

Each component of the KB (i.e., the TBox, RBox, and ABox) has been extended with fuzzy logic.

For the TBox, two approaches are proposed. The first approach keeps the syntax the same as classical terminological axioms while extending only the semantics using fuzzy logic. Examples of this approach include Hölldobler et al. (2002), Sánchez & Tettamanzi (2004), and Straccia (2001, 2004a, 2004b, 2005b). A fuzzy interpretation $\mathcal{I}$ satisfies a fuzzy concept inclusion $C \sqsubseteq D$ if for all $a \in \Delta^{\mathcal{I}}$, we have that $C^{\mathcal{I}}(a) \leq D^{\mathcal{I}}(a)$. That is, the certainty of a subconcept C (e.g., *VeryTall*) is no more than the certainty value of a super-concept D (e.g., *Tall*). In the second approach, the TBox is extended both syntactically and semantically (Straccia, 2005a). Let C and D be concepts, op$\in\{\geq,\leq,>,<\}$, and $\alpha \in [0,1]$. A fuzzy terminological axiom is expressed as $\langle C \sqsubseteq D \text{ op } \alpha\rangle$. For example, $\langle C \sqsubseteq D \geq 0.8\rangle$ indicates "the certainty that C is subsumed by D is at least 0.8."

Extension to RBox with fuzzy logic is proposed in Straccia (2005b). It is similar to the TBox counterparts, except that we have role axioms instead of terminological axioms.

In terms of the ABox, a fuzzy assertion can be represented as $\langle X \text{ op } \alpha\rangle$, where X is either a concept assertion $a{:}C$ or a role assertion $(a,b){:}R$, op$\in\{\geq,\leq,>,<,=\}$, and $\alpha \in [0,1]$ (Hölldobler et al., 2002; Sánchez & Tettamanzi, 2004; Straccia, 2001, 2004a, 2004b, 2005a, 2005b; Tresp & Molitor, 1998). For example, $\langle a{:}C \geq 0.5\rangle$ means "the certainty that a is an instance of concept C is at least 0.5."

Probabilistic Knowledge Base

Several proposals extend TBox and ABox using probability theory as the mathematical basis.

A probabilistic TBox contains a set of classical terminological axioms and a set of probabilistic terminological axioms. The probabilistic information can either be embedded as part of the terminological axiom (Giugno & Lukasiewicz, 2002; Heinsohn, 1994; Jaeger, 1994) or be stored in Bayesian networks (Koller, Levy, & Pfeffer, 1997; Staker, 2002). For lack of space, we describe only the first approach here. A probabilistic terminological axiom is an expression of the form $P(C|D) \in [l,u]$, where C and D are concepts and $0 \leq l \leq u \leq 1$. This states that: "if an individual a is known to belong to concept D, then the probability that a belongs to concept C lies in $[l,u]$."

A probabilistic ABox (Giugno & Lukasiewicz, 2002; Jaeger, 1994) contains assertions of the form $P(a{:}C) \in [l,u]$, where C is a concept, a is an individual, and $l, u \in [0,1]$. Intuitively, this asserts that "the probability that an individual a belongs to concept C lies in $[l,u]$." The probabilistic assertions can also be applied to roles.

Possibilistic Knowledge Base

In Hollunder (1994), a possibilistic extension to TBox and ABox is proposed. Let X be an axiom or assertion, $\prod_\alpha$ be a possibility degree, and N_α be a necessity degree, where $\alpha \in (0,1]$. A possibilistic terminological axiom is either in the form $\langle X, \prod_\alpha \rangle$, meaning "$X$ is possibly true with degree at least α," or $\langle X, N_\alpha \rangle$, meaning "$X$ is necessarily true with degree at least α."

REASONING WITH UNCERTAINTY

The reasoning component provides inference services that enable implicit knowledge to become explicit. In this section, we study fuzzy, probabilistic, and possibilistic reasoning procedures.

Fuzzy Reasoning

For fuzzy reasoning, two main approaches are proposed. The first approach (Straccia, 2004a) transforms fuzzy DL into classical DL. The main advantage is that one can directly use existing reasoners developed for classical DL. The problem is that existing reasoners do not consider certainty values in fuzzy DL as something special, and hence no additional optimization techniques may be applied to inferences with certainty values involved.

The second approach extends existing reasoning procedure so that it takes into account the presence of certainty values during the reasoning process; hence, a reasoner has to be built from scratch. There are two variants of this approach. In the first one, certainty values are dealt with within the inference rules (Hölldobler et al., 2002; Straccia, 2001, 2004b). The basic idea is that, given a fuzzy ABox, a set of inference rules are applied to transform the ABox into simpler and satisfiability preserving equivalence, so that the implicit knowledge becomes explicit. For example, if we know that $\langle a{:}C \sqcap D \geq 0.5 \rangle$, then we could infer that $\langle a{:}C \geq 0.5 \rangle$ and $\langle a{:}D \geq 0.5 \rangle$. The inference rules are applied until either all branches in the extended ABox contain a contradiction/clash (meaning no model can be built), or there exists a clash-free completion of ABox (meaning the ABox is satisfiable). The second approach relies on integer programming and linear optimization to deal with certainty values (Straccia, 2005b). Here, a set of completion rules are applied to generate new assertions together with a set of constraints in the form of inequations over variables with values in [0,1]. The inference rules are applied until either the extended ABox contains a clash, or no rule can be further applied. If there is a clash, the ABox is unsatisfiable. Otherwise, an optimization technique is applied to solve the system of inequations to determine the satisfiability of the KB or the tightest bound such that an assertion is true.

Probabilistic Reasoning

The probabilistic reasoning procedure depends on how the probabilistic information is represented in the KB. In case Bayesian networks are used to express the probabilistic information, the inference procedure developed for Bayesian networks can be directly applied (Ding & Peng, 2004; Koller et al., 1997; Staker, 2002). On the other hand, if the probabilistic information is embedded in the KB (Baader et al., 2003, Giugno & Lukasiewicz, 2002), inference procedures similar to the fuzzy/linear optimization approach are applied to find the optimal bound for which, given a KB, a conditional probability $P(C|D)$ is satisfied.

Possibilistic Reasoning

The possibilistic reasoning is not well explored. No concrete calculus is provided in Hollunder (1994) for inferences on a given KB.

FUTURE TRENDS

In this section, we outline some possible directions for uncertainty management in the Semantic Web. First, it would be useful to have a generic framework that could handle various forms of uncertainty under a unifying umbrella. The first attempt has been made by Haarslev, Pai, and Shiri (2005). However, more work is required in this direction. For practical reasons, we also need to develop efficient tools that support DL with uncertainty. Finally, to handle real-life applications, more expressive fragments of DL (e.g., $\mathcal{SHOIN}$) should be extended with uncertainty.

CONCLUSION

We studied existing frameworks for DL with uncertainty which can form a basis for uncertainty management in the Semantic Web. Based on the mathematical foundations and the types of uncertainty modeled, we classified these frameworks into fuzzy, probabilistic, and possibilistic approaches. We also studied how these approaches extend the components of the DL framework, including description language, KB, and reasoning procedures. It is our hope that this survey would foster further research in incorporating uncertainty in the next generation of Web.

ACKNOWLEDGMENTS

This work is supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and by ENCS, Concordia University. We also thank anonymous reviewers for their useful comments.

REFERENCES

Baader, F., McGuinness, D., Nardi, D., & Schneider, P. P. (2003). *The description logic handbook: Theory, implementation and applications*. UK: Cambridge University Press.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic Web. *Scientific American, 284*(5), 34-43.

Ding, Z., & Peng, Y. (2004). A probabilistic extension to ontology language OWL. *The 37th Hawaii International Conference on System Sciences*, Big Island, HI.

Giugno, R., & Lukasiewicz, T. (2002). P-$\mathcal{SHOQ}$(D): A probabilistic extension of $\mathcal{SHOQ}$(D) for probabilistic ontologies in the semantic Web. *The European Conference on Logics in Artificial Intelligence* (pp. 86-97). Cosenza, Italy.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199-220.

Haarslev, V., Pai, H. I., & Shiri, N. (2005). A generic framework for description logics with uncertainty. *Uncertainty Reasoning for the Semantic Web* (pp. 77-86). Galway, Ireland.

Heinsohn, J. (1994). Probabilistic description logics. *The 10th Annual Conference on Uncertainty in Artificial Intelligence* (pp. 311-318). Seattle, WA.

Hölldobler, S., Khang, T. D., & Störr, H.-P. (2002). A fuzzy description logic with hedges as concept modifiers. *The 3rd International Conference on Intelligent Technologies* (pp. 25-34). Hanoi, Vietnam.

Hollunder, B. (1994). An alternative proof method for possibilistic logic and its application to terminological logics. *The 10th Annual Conference on Uncertainty in Artificial Intelligence* (pp. 327-335). Seattle, WA.

Horrocks, I., Fensel, D., Broekstra, J., Decker, S., Erdmann, M., Goble, C., et al. (2000). *OIL: The ontology inference layer*. (Tech. Rep. No. IR-479). Vrije Universiteit Amsterdam.

Jaeger, M. (1994). Probabilistic reasoning in terminological logics. *The 4th International Conference on Principles of Knowledge Representation and Reasoning* (pp. 305-316). Whistler, Canada.

Koller, D., Levy, A. Y., & Pfeffer, A. (1997). P-CLASSIC: A tractable probabilistic description logic. *The 14th National Conference on Artificial Intelligence* (pp. 390-397). Providence, RI.

Sánchez, D., & Tettamanzi, A. G. B. (2004). Generalizing quantification in fuzzy description logics. *The 8th Fuzzy Days*, Dortmund, Germany.

Staker, R. (2002). Reasoning in expressive description logics using belief networks. *The International Conference on Information and Knowledge Engineering* (pp. 489-495). Las Vegas, NV.

Straccia, U. (2001). Reasoning within fuzzy description logics. *Journal of Artificial Intelligence Research, 14,* 137-166.

Straccia, U. (2004a). Transforming fuzzy description logics into classical description logics. *The 9th European Conference on Logics in Artificial Intelligence* (pp. 385-399). Lisbon, Portugal.

Straccia, U. (2004b). Uncertainty in description logics: A lattice-based approach. *The 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (pp. 251-258). Perugia, Italy.

Straccia, U. (2005a). Towards a fuzzy description logic for the semantic Web (preliminary report). *The 2nd European Semantic Web Conference* (pp. 167-181). Heraklion, Greece.

Straccia, U. (2005b). Fuzzy $\mathcal{ALC}$ with concrete domains. *The International Workshop on Description Logics* (pp. 96-103). Edinburgh, UK.

Tresp, C., & Molitor, R. (1998). A description logic for vague knowledge. *The 13th European Conference on Artificial Intelligence* (pp. 361-365). Brighton, UK.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8,* 338-353.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems, 1*(1), 3-28.

KEY TERMS

Description logics: A decidable subset of first order logic.

Inference: A logical conclusion derived by making implicit knowledge explicit.

Knowledge base: A collection of axioms and assertions.

Knowledge representation: A formalism used for coding the knowledge to be stored in a knowledge base.

Ontology: An explicit formal specification of conceptualization that consists of a set of terms in a domain and the relations among them.

Semantic Web: An extension of the current Web by giving Web resources well-defined meaning.

Uncertainty: A form of deficiency/imperfection in the information where the truth of information is not established definitely.