Semantic Web Uncertainty Management

Volker Haarslev
Concordia University, Canada

Hsueh-Ieng 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, p. 199). 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, p. 3). 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.

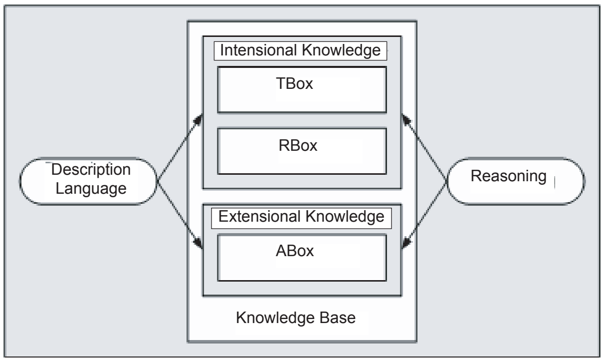
BACKGROUND

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.

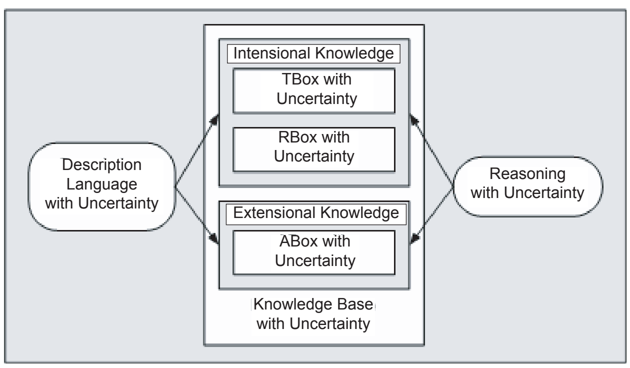
Description Languages with Uncertainty

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;

Figure 2. DL Framework with uncertainty



Related Content