Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 298 (2016) 22–43

FUZZY sets and systems

www.elsevier.com/locate/fss

Reasoning in Fuzzy Description Logics using Automata

Stefan Borgwardt [a,*], Rafael Peñaloza [b,*]

[a] Institute for Theoretical Computer Science, Technische Universität Dresden, Germany
[b] KRDB Research Centre, Free University of Bozen-Bolzano, Italy

Received 12 October 2014; received in revised form 29 April 2015; accepted 15 July 2015

Available online 21 July 2015

Abstract

Automata-based methods have been successfully employed to prove tight complexity bounds for reasoning in many classical logics, and in particular in Description Logics (DLs). Very recently, the ideas behind these automata-based approaches were adapted for reasoning also in fuzzy extensions of DLs, with semantics based either on finitely many truth degrees or the Gödel t-norm over the interval [0, 1]. Clearly, due to the different semantics in these logics, the construction of the automata for fuzzy DLs is more involved than for the classical case. In this paper we provide an overview of the existing automata-based methods for reasoning in fuzzy DLs, with a special emphasis on explaining the ideas and the requirements behind them. The methods vary from deciding emptiness of automata on infinite trees to inclusions between automata on finite words. Overall, we provide a comprehensive perspective on the automata-based methods currently in use, and the many complexity results obtained through them.

Keywords: Fuzzy description logics; Reasoning; Tree automata; Weighted automata

1. Introduction

The Web Ontology Language OWL 2[1] has been widely used to formulate many different ontologies, with a particular concern in the biomedical sciences and healthcare systems, as evidenced by the large ontology repository BioPortal.[2] This is due in no small part to the well-defined formal semantics of OWL 2 DL and its tractable profiles, which are provided by Description Logics (DLs) [1] of different expressivity, and the existence of many highly optimized reasoning systems for DLs.[3] OWL 2 DL and its profiles provide the flexibility for handling many different ontology development needs, and existing ontologies cover the whole range, from very large ontologies written over

* Corresponding authors. Tel.: +49 351 463 38231; fax: +49 351 463 37959.
E-mail addresses: stefborg@tcs.inf.tu-dresden.de (S. Borgwardt), rafael.penaloza@unibz.it (R. Peñaloza).

1 http://www.w3.org/TR/owl2-overview/.
2 http://bioportal.bioontology.org/.
3 http://www.w3.org/2001/sw/wiki/OWL/Implementations.

http://dx.doi.org/10.1016/j.fss.2015.07.013

a very inexpressive language, like SNOMED CT,[4] to more specialized ones describing complex interactions between concepts.

In description logics, knowledge is represented using *concepts* describing sets of objects, such as Fever and Male, and *roles* that draw connections between objects, such as hasSymptom and hasParent. Formally, concepts correspond to unary predicates from first-order logic, while roles are binary predicates. In DLs, complex concepts are usually built from two disjoint sets of *concept names* and *role names* using *concept constructors*. The specific DL used, and hence its expressivity, is determined by the choice of constructors allowed.

The ontology itself expresses the domain knowledge through a finite set of axioms that restrict the way in which concepts may be interpreted, and state some explicit knowledge about the individuals populating the domain. For example, the *general concept inclusion*

$$\forall \mathsf{hasParent}.\exists \mathsf{hasBloodType}.\mathsf{TypeO} \sqsubseteq \exists \mathsf{hasBloodType}.\mathsf{TypeO},$$

expresses that people whose parents both have blood type O must also have the same blood type. Similarly, the *assertions*

$$\mathsf{hasBloodType}(\mathtt{henry}, \mathtt{t}) \quad \text{and} \quad \mathsf{TypeA}(\mathtt{t})$$

state that the individual Henry has a blood type that can be classified as type A. Along with this explicitly represented knowledge, many implicit consequences can be derived from a given ontology. One of the main goals in DLs is to devise and implement efficient reasoning methods that can extract this implicit knowledge in an automated manner.

Since they attempt to model the real world, in biomedical ontologies it is common to find concepts such as HighConcentration (GALEN[5]) and LowFrequency (SNOMED CT) for which no precise definition can be given. For instance, there is no precise point at which a concentration stops being *normal* to become *high*. Fuzzy DLs have been introduced as knowledge representation formalisms capable of handling imprecise concepts, as well as imprecise knowledge. To achieve this, they introduce more than two degrees of truth, allowing for a more fine-grained analysis of interactions between such vague concepts. For example, given the knowledge

$$\exists \mathsf{hasDisease}.\mathsf{Influenza} \sqsubseteq \exists \mathsf{hasSymptom}.\mathsf{Fever} \sqcap \exists \mathsf{hasSymptom}.\mathsf{Headache}$$

that influenza induces the symptoms fever and headache, it makes sense to say that a lower degree of fever indicates a less severe case of influenza, which is modeled by using a lower degree of truth.

Fuzzy DLs are based on fuzzy set theory [2] and mathematical fuzzy logic [3]. Usually, the set of truth degrees is the interval $[0, 1]$ of real numbers, and thus a statement can be true to some intermediate degree, e.g. 0.5, instead of only completely true (1) or completely false (0). The logical constructors such as conjunction and implication are then generalized to this extended set of truth degrees. The oldest and most popular fuzzy semantics, the so-called *Zadeh* semantics, is based on the minimum function to interpret conjunction, the maximum function for disjunction, $1 - x$ for the negation $\neg x$, and the Kleene–Dienes-implication $\max\{1 - x, y\}$ for the implication $x \to y$. The latter is derived from the classical equivalence of $x \to y$ and $\neg x \vee y$. However, this choice of implication function has some unintuitive consequences [4]. This is particularly unfortunate in the context of description logics since they make heavy use of implications in the axioms ($\sqsubseteq$) and some constructors.

An alternative approach with a deeper formal motivation uses arbitrary *t-norms*, which satisfy only some basic properties, as interpretation functions for the conjunction [5]. The implication is then evaluated using the *residuum* of such a t-norm. The minimum function is a t-norm, also called the *Gödel t-norm*, but there are infinitely many others. Unfortunately, using t-norm based semantics for fuzzy DLs turned out to be problematic, with very inexpressive DLs losing their tractability [6], and slightly more expressive logics becoming undecidable [7–10]. Two possible ways to avoid this problem that have been identified are to restrict the semantics either to the (relatively) harmless Gödel t-norm or to a finite set of truth degrees.

Automata-based methods have proven to be very useful tools for finding tight bounds for the complexity of reasoning in DLs and other logics [11–13]. In a nutshell, automata provide an elegant and complexity-wise efficient formalism for handling complex, potentially infinite, models. Very recently, the ideas developed for the classical case

4 http://www.ihtsdo.org/snomed-ct/.
5 http://www.opengalen.org/.