ARTICLE IN PRESS
IJA:7937
International Journal of Approximate Reasoning ••• (••••) •••–•••
Contents lists available at ScienceDirect
International Journal of Approximate Reasoning
www.elsevier.com/locate/ijar

A distance-based framework for inconsistency-tolerant reasoning and inconsistency measurement in DL-Lite

Xiaowang Zhang [a,b,c], Kewen Wang [d], Zhe Wang [d], Yue Ma [e], Guilin Qi [f], Zhiyong Feng [a,b,*]

[a] School of Computer Science and Technology, Tianjin University, Tianjin, China
[b] Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin, China
[c] Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing, China
[d] School of Information and Communication Technology, Griffith University, BNE, Australia
[e] Laboratoire de Recherche en Informatique, University Paris Sud, Paris, France
[f] School of Computer Science and Engineering, Southeast University, Nanjing, China

ARTICLE INFO

Article history:
Received 28 September 2015
Received in revised form 7 August 2016
Accepted 8 August 2016
Available online xxxx

Keywords:
Semantic Web
Description logics
DL-Lite
Inconsistency handling
Inconsistency-tolerant reasoning
Inconsistency measurement
Distance-based semantics

ABSTRACT

In this paper, we present a distance-based framework for DL-Lite based on the notion of features. Within this framework, we propose a distance-based paraconsistent semantics for DL-Lite where meaningful conclusions can be rationally drawn even from an inconsistent knowledge base and we develop a distance-based inconsistency measurement for DL-Lite to provide more informative metrics which can tell the differences between axioms causing inconsistency and among inconsistent knowledge. Furthermore, we investigate several important logical properties (e.g., consistency preservation, closure consistency, splitting property etc.) of the entailment relation based on the new semantics and show its advantages in non-monotonic reasoning for DL-Lite. Finally, we show that our two distance-based inconsistency measures are basic inconsistency measures where some good properties hold such as Free Axiom Independence and Dominance of inconsistency etc.

1. Introduction

Inconsistency handling is an important issue in ontology (knowledge base, KB) management communities [44,38] since inconsistency is not rare in ontology applications and may be due to several reasons, such as errors in modeling, migration from other formalisms, ontology merging, and ontology evolution. However, as a logical foundation of Web Ontology Language (OWL) [5,4], description logic (DL) reasoning mechanism based on two-valued semantics faces problem when inconsistency occurs since DL is a fragment of predicate logic [9], which is referred to as the triviality problem [8,37]. That is, any conclusions, that are possibly irrelevant or even contradicting, will be entailed from an inconsistent DL ontology under the classical semantics.

In many practical ontology applications, there is a strong need for inferring (only) useful information from inconsistent ontologies. For instance, consider a simple DL KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ where $\mathcal{T} = \{Penguin \sqsubseteq Bird, Swallow \sqsubseteq Bird, Bird \sqsubseteq Fly, Fly \sqsubseteq \exists\, hasWing\}$ and $\mathcal{A} = \{Penguin(tweety), \neg Fly(tweety), Swallow(fred)\}$. The KB says that penguins are birds; swallows are birds; birds can fly; flying animals have wings; *tweety* is a penguin; *tweety* cannot fly; and *fred* is a swallow. Under the classical

* Corresponding author.
E-mail addresses: xiaowangzhang@tju.edu.cn (X. Zhang), zyfeng@tju.edu.cn (Z. Feng).

http://dx.doi.org/10.1016/j.ijar.2016.08.003

semantics for DLs, anything can be inferred from $\mathcal{K}$ since $\mathcal{K}$ is inconsistent (i.e., it has no model.). Intuitively, one might wish to still infer $Bird(fred)$ and $Fly(fred)$, while it is useless to derive both $Fly(tweety)$ and $\neg Fly(tweety)$ from $\mathcal{K}$.

There exist several proposals for inconsistency-tolerant reasoning DL KBs in the literature. These approaches usually fall into one of two different streams. The first one is based on the assumption that inconsistencies are caused by erroneous data and thus, they should be removed in order to obtain a consistent KB [26,36,13,16]. In most approaches in this stream, the task of repairing inconsistent ontologies is actually reduced to finding a maximal consistent subset of the original KB. A shortcoming of these approaches is similar to the so-called *multi-extension problem* in Reiter's default logic. That is, in many cases, an inconsistent KB may have several different maximal subsets that are consistent. The other stream, based on the idea of living with inconsistency, is to introduce a form of paraconsistent reasoning or inconsistency-tolerant reasoning by employing non-standard reasoning methods (e.g., non-standard inference and non-classical semantics). There are some strategies to select consistent subsets from an inconsistent KB as substitutes of the original KB in reasoning [45,21,31,27,20,53]. The Belnap's four-valued semantics has been successfully extended into DL [32] where two additional logical values besides "true" and "false" are introduced to indicate contradictory conclusions. Inference power of the four-valued semantics is further enhanced by a new quasi-classical semantics for DLs proposed by Zhang et al. [55], which is a generalization of Hunter's quasi-classical semantics for propositional logic. However, the reasoning capability of such paraconsistent methods is not strong enough for many practical applications. For instance, a conclusion, that can be derived from a consistent KB under the classical semantics, may become not derivable under their paraconsistent semantics. One limitation of existing approaches in the two approaches is mostly *coarse-grained* in the sense that they fail to fully utilize semantic information in the given inconsistent KB. For instance, when two interpretations make a concept unsatisfiable, one interpretation may be more reasonable than the other. But those existing approaches to paraconsistent semantics in DLs do not take this into account usually. Recently, there are some works in considering the priorities among different interpretations by introducing some preferred repair semantics [14,7,6,15]. However, those approaches are based on the assumption that TBoxes are consistent.

As an important approach to handling inconsistency, inconsistency measurement is providing some measures for the inconsistency of a KB so that we can compare different KBs and evaluate their quality of information and then choose one that is least inconsistent [17]. There exist many works in measuring inconsistency of KBs by applying minimal inconsistent sets [25], Shapley inconsistency values [22], partial Max-SAT solvers [50], four-valued semantics [34], MUS decomposition [23], closed set packing [24] and measuring inconsistency for prioritized KBs [40], stratified KBs [39], arbitrary KBs [35], probabilistic query answering [51], DL-Lite ontologies [56] by three-valued semantics. Most of those approaches to measuring inconsistency of KBs in DLs are either syntax-based (where the results are sensitive to the syntactical structure of KBs) [25,22,50,23,24] or multi-valued semantics (where the results are not so intuitive in characterizing facts in a classical logic) [34,56].

A distance-based semantics presented by Arieli [1] has been proposed to deal with inconsistent KBs in propositional logic, which is inspired from distance-based merging procedures in propositional logic [28,29], where those interpretations with minimal distances defined between interpretations are chosen as models so that those models could closely characterize the semantics. A distance-based inconsistency-measurement has been developed for DLs, in particular, for DL-Lite [33]. However, because the finiteness problem from the approach of [33] is dealt by considering only interpretations with finite domains, it is interesting to observe models with infinite domains since a propositional KB has a finite number of finite models) while, in DLs, a KB might have infinite number of models and a model might also be infinite.

To overcome these difficulties in the finiteness problem of DL, in this paper we first use the notion of *features* [46] and introduce a distance-based semantics for paraconsistent reasoning with DL-Lite. Features in DL-Lite are Herbrand interpretations extended with limited structure, which provide a novel semantic characterization for DLs. In addition, features also generalize the notion of *types* for TBoxes [30] to general KBs. Each KB in DL-Lite has a finite number of features and each feature is finite. This makes it possible to cast Arieli's distance-based semantics to DL.

In this paper, we present a distance-based framework for both inconsistency-tolerant reasoning and inconsistency measurement in DL-Lite ontologies. The DL-Lite [10] is a family of lightweight description logics (DLs), which form the logical foundation of OWL 2 QL, one of the three profiles of OWL 2 for Web ontology language recommended by W3C [11]. Following [47], we choose DL-Lite$^{\mathcal{N}}_{bool}$ [2], one of the most expressive members of the DL-Lite family, and define distance-based semantics for DL-Lite$^{\mathcal{N}}_{bool}$ in a way analogous to the model-based approaches in propositional logic. DL-Lite$^{\mathcal{N}}_{bool}$ [3], expressive enough to allow all boolean operators, generalizes the main DL-Lite dialects such as DL-Lite$_{core}$ and DL-Lite$_{\mathcal{F}}$. Though our paper mainly discusses DL-Lite$^{\mathcal{N}}_{bool}$, our proposed approach can be conveniently adapted to other DL-Lite dialects like DL-Lite$_{R}$ since types, a foundation of our distance-based semantics, can be adapted to them [47].

The main innovations and contributions of this paper can be summarized as follows.

- We introduce distance functions on *types* of DL-Lite$^{\mathcal{N}}_{bool}$ KBs, which avoids the problem of domain infiniteness and model infiniteness in defining the distance function in terms of models of KBs. We choose DL-Lite$^{\mathcal{N}}_{bool}$ [2], one of the most expressive members of the DL-Lite family, and define distance-based semantics for DL-Lite$^{\mathcal{N}}_{bool}$ in a way analogous to the model-based approaches in propositional logic.
- We develop a way of measuring types that are closest to a TBox, based on the new distance function on types, and the notion of *minimal model types* is introduced. This notion is also extended to *minimal model features* for KBs. We propose