Weight-based consistent query answering over inconsistent $\mathbf{SHIQ}$ knowledge bases

Jianfeng Du · Guilin Qi · Yi-Dong Shen

Abstract Non-standard query mechanisms that work under inconsistency are required in some important description logic (DL)-based applications, including those involving an inconsistent DL knowledge base (KB) whose intensional knowledge is consistent but is violated by its extensional knowledge. This paper proposes a weight-based semantics for querying such an inconsistent KB. This semantics defines an answer of a conjunctive query posed upon an inconsistent KB as a tuple of individuals whose substitution for the variables in the query head makes the query body entailed by any subbase of the KB consisting of the intensional knowledge and a weight-maximally consistent subset of the extensional knowledge. A novel computational method for this semantics is proposed, which works for extensionally reduced $\mathbf{SHIQ}$ KBs and conjunctive queries without non-distinguished variables. The method first compiles the given KB to a propositional program; then, for any given conjunctive query, it reduces the problem of computing all answers of the given query to a set of propositional satisfiability (SAT) problems with $PB$-constraints, which are then solved by SAT solvers. A decomposition-based framework for optimizing the method is also proposed. The feasibility of this method is demonstrated in our experiments.

Keywords Semantic Web · Description logics · Query answering · Inconsistency-tolerant reasoning · Weight-based semantics
1 Introduction

Description logics (DLs) [5] are mostly decidable fragments of first-order logic. They provide logical foundations for the highly evolving Semantic Web. In particular, among the three species of the standard Web Ontology Language (OWL) [40] for the Semantic Web, namely OWL Lite, OWL DL and OWL Full, the former two, respectively, correspond to DLs $\mathcal{SHIF}(D)$ and $\mathcal{SHOIN}(D)$ [27]. A DL knowledge base (KB) consists of a TBox, an RBox and an ABox, where the TBox and the RBox store the intensional knowledge, while the ABox stores the extensional knowledge.

Logical contradictions can easily be introduced in a DL KB and make the KB inconsistent (i.e., have no models). The standard query mechanism in DLs returns anything that is satisfied by all models of an inconsistent KB, thus is meaningless. To solve this problem, one may repair the KB to render it consistent whenever some logical contradictions appear. But this approach is no panacea and some important applications, such as virtual data integration [45] and ontology population [14], require non-standard query mechanisms that work under inconsistency.

Data integration has been recognized as a crucial application in the Semantic Web [46]. In the scenario of virtual data integration [45], data distributed in different sources are mapped to an integrated view in a mediator. Since the data in different sources may often be heterogeneous, they are usually standardized using the terminology in the mediator. From the DL point of view, the knowledge stored in the mediator can be treated as the union of a TBox and an RBox, while the data in different sources as a whole can be treated as an ABox. However, the ABox may violate the constraints specified in the union of the TBox and the RBox, and it can hardly be repaired because its update reflects updates of the original data distributed in multiple sources. Hence, in this scenario, non-standard query mechanisms that work under inconsistency are needed.

In the scenario of ontology population [14], a DL KB is enriched by adding instance assertions. There are many research results on ontology population. To name a few, KIM [41] and Text2Onto [15] are frameworks that integrate algorithms for ontology population from textual data, including information extraction algorithms (e.g., [23]) that assign annotations carrying some semantics to regions of the data, as well as co-reference algorithms (e.g., [49]) that identify annotated individuals in multiple places. From the DL point of view, the course of information extraction adds concept or role assertions, whereas the course of co-reference adds (in)equality assertions. The populated DL KB, however, may become inconsistent due to potential conflicts between the new data and the original data. Directly repairing a populated DL KB may not be appropriate since choosing the best repair plan among a set of candidates may require massive human efforts. We had better use non-standard query mechanisms to directly work on the populated DL KB.

To provide a reasonable query mechanism that works under inconsistency, we propose a weight-based semantics for query answering over an inconsistent DL KB. We assume that the union of the TBox and the RBox of the KB is consistent, and every instance assertion in the ABox of the KB has a positive weight. We simply call a subset of the ABox consistent if it is consistent with the union of the TBox and the RBox and further call it weight-maximally consistent if the total weight of assertions in it is maximal among all consistent subsets of the ABox. The weight-based semantics defines an answer of a conjunctive query as a tuple of individuals whose substitution for the variables in the query head makes the query body entailed by any subbase of the given KB that consists of the TBox, the RBox and a weight-maximally consistent subset of the ABox. In a number of applications, such as virtual data integration and ontology population, the TBox and the RBox are fixed and well prepared,