Query Answering in Fuzzy DL-Lite with Graded Axioms

Gabriella Pasi and Rafael Peñaloza[(✉)]

IKR3 Lab, University of Milano-Bicocca, Milan, Italy
{gabriella.pasi,rafael.penaloza}@unimib.it

Abstract. Fuzzy DL-Lite has been studied as a means to answer queries w.r.t. vague or imprecise knowledge and facts. Existing approaches consider only the Zadeh semantics or are limited to precise terminological knowledge, where only facts are graded. We study the problem of answering conjunctive queries over fuzzy DL-Lite$_R$ ontologies which allow also graded axioms, and whose semantics is based on mathematical fuzzy logic. We show that for the Gödel t-norm, the degree of an answer is computable through repeated calls to a classical query answering engine. For non-idempotent t-norms, we show the difficulty in dealing with degrees, and provide some partial solutions.

1 Introduction

Description logics [3] are a well known family of knowledge representation formalisms characterised by their formal logic-based semantics, and their decidable (usually with relatively low complexity) reasoning tasks. Within these logics, the DL-Lite family of light-weight DLs [2] is specially interesting, as it allows for very efficient query answering in terms of data and combined complexity. It is also the formalism underlying the OWL 2 QL profile of the standard ontology language for the semantic web [10]. For this reason, they have been used for many practical applications in various knowledge domains.

As the semantics are based on classical first-order logic, in their standard form DLs are not able to handle vague or imprecise knowledge and facts. For that reason, fuzzy extensions of DLs have been devised. Although most work has focused on extending more expressive DLs and solving standard reasoning services [8], some efforts have been made towards answering queries in DL-Lite. Specifically, Straccia [20] studied the problem of computing the answers with highest degree on a query w.r.t. some background knowledge. This was followed by Pan et al. [17], who considered more complex queries to be answered. Other work considering query answering in fuzzy DLs includes [21]. These three works were based on the so-called Zadeh semantics, and had the limitation that only the facts in the ontology could be graded, but not the terminological knowledge.

Later, Turhan and Mailis studied the problem from the point of view of fuzzy logic [12], where the semantics are based on the properties of continuous triangular norms. They developed a technique for computing the satisfaction

degrees of queries when the semantics were based on the Gödel t-norm [14]. This technique, which is based on the construction of a classical query, was later implemented and shown to be effective in [15]. However, it still suffered from two main drawbacks: (i) it was only capable to handle the idempotent t-norm and (ii) terminological knowledge had to still be precise. The latter condition is essential for the correctness of their approach: their reduction is unable to keep track of the degrees used by the terminological axioms, as this would require an unbounded memory use.

In this paper, we tackle these limitations over the well-known DL-Lite$_R$ logic. Considering the t-norm-based semantics, we show that every consistent fuzzy DL-Lite$_R$ ontology has a canonical model, which can be homomorphically embedded in every other model, and present an infinitary construction for obtaining it, based on the ideas of classical DL-Lite$_R$. Using this fact, we develop a method for finding the degree of an answer to a conjunctive query, which is also capable of dealing with fuzzy terminological axioms w.r.t. the Gödel t-norm. Indeed, despite being more general, our approach is much simpler than previous methods. We show that to verify a lower bound d of the degree, it suffices to consider only the subontology with all axioms and assertions with a degree at least d, and answer a classical query over it. Thus, one can use any off-the-shelf ontology-based query answering engine available [23]. The cost of our simpler method is that, to find the specific degree, the data complexity jumps from AC^0 to LogSpace. Whether the algorithm can be further improved or the LogSpace upper bound is tight remains open.

Considering other continuous t-norms, we show through several examples that query answering needs some more meticulous analysis. While deciding consistency w.r.t. the product t-norm is still easy, finding the precise degrees of an answer becomes more involved. In the case of the Łukasiewicz t-norm, we recall the fact that for a minimal extension of DL-Lite$_R$ consistency is already NP-hard in combined complexity, and hence unlikely to be decidable through a simple reduction to classical DL-Lite$_R$.

2   Preliminaries

We briefly introduce the syntax and semantics of fuzzy DL-Lite$_R$. Let N_C, N_R, and N_I be three mutually disjoint sets whose elements are called *concept names*, *role names*, and *individual names*, respectively. The sets of DL-Lite$_R$ *concepts* and *roles* are built through the grammar rules

$$B ::= A \mid \exists Q \qquad\qquad C ::= B \mid \neg B$$
$$Q ::= P \mid P^- \qquad\qquad R ::= Q \mid \neg Q$$

where $A \in N_C$ and $P \in N_R$. Concepts of the form B and roles of the form Q are called *basic*, and all others are called *general*.

A *fuzzy DL-Lite$_R$ TBox* is a finite set of *fuzzy axioms* of the form $\langle B \sqsubseteq C, d \rangle$ and $\langle Q \sqsubseteq R, d \rangle$, where d is a number in $[0, 1]$. An axiom is *positive* if it does

not have negation on its right-hand side; note that negations can never occur on the left-hand side of an axiom. A *fuzzy DL-Lite$_R$ ABox* is a finite set of *fuzzy assertions* of the form $\langle B(a), d\rangle$ and $\langle P(a,b), d\rangle$, where $a, b \in N_I$. A *fuzzy DL-Lite$_R$ ontology* is a pair of the form $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ where $\mathcal{T}$ is a TBox and $\mathcal{A}$ is an ABox. In the remainer of this paper, we will mostly exclude the qualifiers "fuzzy," and "DL-Lite" and simply refer to axioms, ontologies, etc.

Table 1. The three main continuous t-norms and related operations

Name	$d \otimes e$	$d \Rightarrow e$	$\ominus d$
Gödel	$\min\{d, e\}$	$\begin{cases} 1 & d \leq e \\ e & \text{ow} \end{cases}$	$\begin{cases} 1 & d = 0 \\ 0 & \text{ow} \end{cases}$
Łukasiewicz	$\max\{d + e - 1, 0\}$	$\min\{1 - d + e, 1\}$	$1 - d$
Product	$d \cdot e$	$\begin{cases} 1 & d \leq e \\ e/d & \text{ow} \end{cases}$	$\begin{cases} 1 & d = 0 \\ 0 & \text{ow} \end{cases}$

The semantics of fuzzy DL-Lite$_R$ is based on fuzzy interpretations, which provide a *membership degree* or *truth degree* for objects belonging to the different concept and role names. To fully define this semantics in the presence of other constructors according to the theory of mathematical fuzzy logic, we need the notion of a triangular norm (or *t-norm* for short).

A *t-norm* $\otimes$ is a binary operator over $[0, 1]$ that is commutative, associative, monotonic, and has as neutral operator 1 [13]. The t-norm is used to generalize the logical conjunction to the interval $[0, 1]$. Every continuous t-norm defines a unique *residuum* $\Rightarrow$ where $f \otimes d \leq e$ iff $f \leq d \Rightarrow e$. The residuum interprets implications. With the help of this operation, it is also possible to interpret other logical operators such as negation ($\ominus d := d \Rightarrow 0$). The three basic continuous t-norms are the *Gödel*, *Łukasiewicz*, and *product* t-norms, which are defined, with their residua and negations in Table 1. These t-norms are the fundamental ones in the sense that every other continuous t-norm is isomorphic to the ordinal sum of copies of them [12,16]. Hence, as usual, we focus our study to these three t-norms.

Note that the residuum always satisfies that $d \Rightarrow e = 1$ iff $d \leq e$, and that in the Gödel and product t-norms the negation is annihilating in the sense that it maps to 0 any positive value, while the negation of 0 is 1. In particular, this means that the negation is not involutive; that is, $\ominus \ominus d \neq d$ in general. From now on, unless specified explicitly otherwise, we assume that we have an arbitrary, but fixed, t-norm $\otimes$ which underlies the operators used. When the t-norm becomes relevant in the following sections, we will often use G, π, and L as prefixes to express that the underlying t-norm is Gödel, product, or Łukasiewicz, respectively, as usual in the literature.

We can now formally define the semantics of the logic. An *interpretation* is a pair $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set called the *domain*, and $\cdot^{\mathcal{I}}$ is the

interpretation function which maps every individual name $a \in N_I$ to an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$; every concept name $A \in N_C$ to a function $A^{\mathcal{I}} : \Delta^{\mathcal{I}} \to [0,1]$; and every role name $P \in N_R$ to a function $P^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \to [0,1]$. That is, concept names are interpreted as fuzzy unary relations and role names are interpreted as fuzzy binary relations over $\Delta^{\mathcal{I}}$. The interpretation function is extended to other constructors with the help of the t-norm operators as follows. For every $\delta, \eta \in \Delta^{\mathcal{I}}$,

$$(\exists Q)^{\mathcal{I}}(\delta) := \sup_{\delta' \in \Delta^{\mathcal{I}}} Q^{\mathcal{I}}(\delta, \delta') \qquad (\neg B)^{\mathcal{I}}(\delta) := \ominus B^{\mathcal{I}}(\delta) \qquad (\top)^{\mathcal{I}}(\delta) := 1$$

$$(P^-)^{\mathcal{I}}(\delta, \eta) := P^{\mathcal{I}}(\eta, \delta) \qquad (\neg Q)^{\mathcal{I}}(\delta, \eta) := \ominus Q^{\mathcal{I}}(\delta, \eta)$$

The interpretation $\mathcal{I}$ *satisfies* the axiom

- $\langle B \sqsubseteq C, d \rangle$ iff for every $\delta \in \Delta^{\mathcal{I}}$, $B^{\mathcal{I}}(\delta) \Rightarrow C^{\mathcal{I}}(\delta) \geq d$;
- $\langle Q \sqsubseteq R, d \rangle$ iff for every $\delta, \eta \in \Delta^{\mathcal{I}}$, $Q^{\mathcal{I}}(\delta, \eta) \Rightarrow R^{\mathcal{I}}(\delta, \eta) \geq d$.

It is a *model* of the TBox $\mathcal{T}$ if it satisfies all axioms in $\mathcal{T}$. $\mathcal{I}$ *satisfies* the assertion

- $\langle B(a), d \rangle$ iff $B^{\mathcal{I}}(a^{\mathcal{I}}) \geq d$;
- $\langle P(a, b), d \rangle$ iff $P^{\mathcal{I}}(a^{\mathcal{I}}, b^{\mathcal{I}}) \geq d$.

It is a *model* of the ABox $\mathcal{A}$ if it satisfies all axioms in $\mathcal{A}$, and it is a *model* of the ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ if it is a model of $\mathcal{T}$ and of $\mathcal{A}$.

We note that the classical notion of DL-Lite$_R$ [9] is a special case of fuzzy DL-Lite$_R$, where all the axioms and assertions hold with degree 1. In that case, it suffices to consider interpretations which map all elements to $\{0,1\}$ representing the classical truth values. When speaking of classical ontologies, we remove the degree and assume it implicitly to be 1.

For this paper, we are interested in answering conjunctive queries, which consider whether a combination of facts can be derived from the knowledge in an ontology. Let N_V be a set of *variables*, which is disjoint from N_I, N_C, and N_R. A *term* is an element of $N_V \cup N_I$; that is, an individual name or a variable. An *atom* is an expression of the form $C(t)$ (concept atom) or $P(t_1, t_2)$ (role atom). Let $\mathbf{x}$ and $\mathbf{y}$ denote vectors of variables. A *conjunctive query* (CQ) is a first-order formula of the form $\exists \mathbf{y}.\phi(\mathbf{x}, \mathbf{y})$ where ϕ is a conjunction of atoms which only use the variables from $\mathbf{x}$ and $\mathbf{y}$. Let $\mathsf{At}(\phi)$ denote the set of all atoms appearing in ϕ. The variables $\mathbf{y}$ are called *existential variables*, and those in $\mathbf{x}$ are *answer variables*. A *union of conjunctive queries* (UCQ) is a finite set of CQs that use the same answer variables.

Given the CQ $q(\mathbf{x}) = \exists \mathbf{y}.\phi(\mathbf{x}, \mathbf{y})$, the interpretation $\mathcal{I}$, and a vector of individuals $\mathbf{a}$ of the same length as $\mathbf{x}$, a *match* is a mapping π which assigns to each $a \in N_I$ the value $a^{\mathcal{I}}$; to each variable in $\mathbf{x}$ the corresponding element of $\mathbf{a}^{\mathcal{I}}$; and to each variable in $\mathbf{y}$ an element $\delta \in \Delta^{\mathcal{I}}$. We extend the match π to apply to assertions as follows: $\pi(B(t)) = B(\pi(t))$ and $\pi(P(t_1, t_2)) = P(\pi(t_1), \pi(t_2))$. The *degree* of the CQ $q(\mathbf{x})$ w.r.t. the match π is

$$q^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}, \pi(\mathbf{y})) := \bigotimes_{\alpha \in \mathsf{At}(\phi)} (\pi(\alpha))^{\mathcal{I}}.$$

That is, a match maps all the variables in the query to elements of the interpretation domain, where the vector $\mathbf{a}$ is used to identify the mapping of the answer variables. The degree of the query is the (fuzzy) conjunction of the degrees of the atoms under this mapping. From now on, $\Pi(\mathcal{I})$ denotes the set of all matches of $q(\mathbf{x})$ w.r.t. the interpretation $\mathcal{I}$.

A tuple of individuals $\mathbf{a}$ is an *answer* of $q(\mathbf{x})$ to degree d w.r.t. the interpretation $\mathcal{I}$ (denoted by $\mathcal{I} \models q(\mathbf{a}) \geq d$) iff $q^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}) := \sup_{\pi \in \Pi(\mathcal{I})} q^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}, \pi(\mathbf{y})) \geq d$. It is a *certain answer* (or answer for short) of $q(\mathbf{x})$ over the ontology $\mathcal{O}$ to degree d (denoted as $\mathcal{O} \models q(\mathbf{a}) \geq d$) iff for every model $\mathcal{I}$ of $\mathcal{O}$ it holds that $\mathcal{I} \models q(\mathbf{a}) \geq d$. The set of certain answers of the query $q(\mathbf{x})$ w.r.t. $\mathcal{O}$ and their degree is denoted by $\mathsf{ans}(q(\mathbf{x}), \mathcal{O})$; that is,

$$\mathsf{ans}(q(\mathbf{x}), \mathcal{O}) := \{(\mathbf{a}, d) \mid \mathcal{O} \models q(\mathbf{a}) \geq d \text{ and for all } d' > d, \mathcal{O} \not\models q(\mathbf{a}) \geq d'\}.$$

It is important to keep in mind that the atoms in a CQ are not graded, but simply try to match with elements in the domain. The use of the truth degrees becomes relevant in the degree of the answers found.

A class of queries of special significance is that where the vector of answer variables $\mathbf{x}$ is empty. This means that the answer tuple of individuals must also be empty. In the classical setting, these are called *Boolean queries*, because they can only return a Boolean value: true if there is a match for the existential variables in every model, and false otherwise. In the fuzzy setting, the set of answers to such a query will only contain one element $((), d)$. Thus, in that case, we are only interested in finding the degree d, and call them *degree queries*. This degree is the tightest value for which we can find a satisfying matching. Formally, the ontology $\mathcal{O}$ *entails* the degree query $q()$ to degree d iff $\mathcal{O} \models q() \geq d$ and $\mathcal{O} \not\models q() \geq d'$ for all $d' > d$. Degree queries allow us to find the degree of a specific answer $\mathbf{a}$ without having to compute $\mathsf{ans}(q(\mathbf{x}), \mathcal{O})$: simply compute the degree of the degree query $q(\mathbf{a})$.

As typical in query answering for description logics, we consider two measures of complexity: *data complexity*, where only the size of the ABox is considered as part of the input, and *combined complexity* in which the size of the whole ontology (including the TBox) is taken into account.[1] For data complexity, it is relevant to consider sub-linear complexity classes. In particular, we consider AC^0 and LOGSPACE. For the formal definitions, see [4,18]; here we only mention briefly that evaluation of FO-queries over a database is in AC^0 on the size of the database [1] and AC^0 is strictly contained in LOGSPACE [11,19].

In classical DL-Lite, query answering w.r.t. an ontology is reduced to the standard problem of query answering over a database through a process known as query rewriting, and thus is in AC^0 w.r.t. data complexity. The main idea is to include in the query all the information that is required by the TBox, in such a way that only assertions from the ABox need to be considered. Since there are many possible choices to create the matches that comply with the TBox,

---

[1] Note that *combined complexity* does not include the query as part of the input, but only the ontology. This is in line with the terminology used in ontology-based query answering; e.g. [2]. It is typically only used in combination of simple fixed queries.

this method results in a UCQ. At this point, the ABox is treated as a database, which suffices to find all the certain answers. Similarly, a special UCQ can be used to verify that the ontology is *consistent*; that is, whether it is possible to build a model for this ontology. For the full details on how these query rewritings work in classical DL-Lite, see [9]. In terms of combined complexity, consistency can be decided in polynomial time; in fact, it is NLogSpace-complete [2].

3   The Canonical Interpretation

A very useful tool for developing techniques for answering queries in DL-Lite$_R$ is the canonical interpretation. We first show that the same idea can be extended to fuzzy ontologies, independently of the t-norm underlying its semantics.

Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be a DL-Lite ontology and assume w.l.o.g. that there are no axioms of the form $\langle \exists Q_1 \sqsubseteq \exists Q_2, d \rangle \in \mathcal{T}$; any such axiom can be substituted by the two axioms $\langle \exists Q_1 \sqsubseteq A, 1 \rangle, \langle A \sqsubseteq \exists Q_2, d \rangle$ where A is a concept name not appearing in $\mathcal{T}$. The *canonical interpretation* of $\mathcal{O}$ is the interpretation $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ over the domain $\Delta^{\mathcal{I}_{\mathsf{can}}} := N_I \cup N_N$, where N_N is a countable set of *constants* obtained through the following (infinite) process. Starting from the *empty* interpretation which sets $A^{\mathcal{I}_{\mathsf{can}}}(\delta) = 0$ and $P^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta) = 0$ for every $A \in N_C, P \in N_R$ and $\delta, \eta \in \Delta^{\mathcal{I}_{\mathsf{can}}}$, exhaustively apply the following rules:

- if $\langle A(a), d \rangle \in \mathcal{A}$ and $A^{\mathcal{I}_{\mathsf{can}}}(a) < d$, then update $A^{\mathcal{I}_{\mathsf{can}}}(a) := d$;
- if $\langle P(a, b), d \rangle \in \mathcal{A}$ and $P^{\mathcal{I}_{\mathsf{can}}}(a, b) < d$, then update $P^{\mathcal{I}_{\mathsf{can}}}(a, b) := d$;
- if $\langle A_1 \sqsubseteq A_2, d \rangle \in \mathcal{T}$ and $A_2^{\mathcal{I}_{\mathsf{can}}}(\delta) < A_1^{\mathcal{I}_{\mathsf{can}}}(\delta) \otimes d$, then update the value $A_2^{\mathcal{I}_{\mathsf{can}}}(\delta) := A_1^{\mathcal{I}_{\mathsf{can}}}(\delta) \otimes d$;
- if $\langle A \sqsubseteq \exists P, d \rangle \in \mathcal{T}$ and for every $\eta \in \Delta^{\mathcal{I}_{\mathsf{can}}}$, $P^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta) < A^{\mathcal{I}_{\mathsf{can}}}(\delta) \otimes d$ holds, then select a fresh element η_0 such that $P^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta_0) = 0$ and update $P^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta_0) := A^{\mathcal{I}_{\mathsf{can}}}(\delta) \otimes d$;
- if $\langle A \sqsubseteq \exists P^-, d \rangle \in \mathcal{T}$ and for every $\eta \in \Delta^{\mathcal{I}_{\mathsf{can}}}$ $P^{\mathcal{I}_{\mathsf{can}}}(\eta, \delta) < A^{\mathcal{I}_{\mathsf{can}}}(\delta) \otimes d$ holds, then select a fresh element η_0 such that $P^{\mathcal{I}_{\mathsf{can}}}(\eta_0, \delta) = 0$ and update $P^{\mathcal{I}_{\mathsf{can}}}(\eta_0, \delta) := A^{\mathcal{I}_{\mathsf{can}}}(\delta) \otimes d$;
- if $\langle \exists P \sqsubseteq A, d \rangle \in \mathcal{T}$ and $\exists \eta \in \Delta^{\mathcal{I}_{\mathsf{can}}}$ such that $A^{\mathcal{I}_{\mathsf{can}}}(\delta) < P^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta) \otimes d$, then update $A^{\mathcal{I}_{\mathsf{can}}}(\delta) := P^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta) \otimes d$;
- if $\langle \exists P^- \sqsubseteq A, d \rangle \in \mathcal{T}$ and $\exists \eta \in \Delta^{\mathcal{I}_{\mathsf{can}}}$ such that $A^{\mathcal{I}_{\mathsf{can}}}(\delta) < P^{\mathcal{I}_{\mathsf{can}}}(\eta, \delta) \otimes d$, then update $A^{\mathcal{I}_{\mathsf{can}}}(\delta) := P^{\mathcal{I}_{\mathsf{can}}}(\eta, \delta) \otimes d$;
- if $\langle Q_1 \sqsubseteq Q_2, d \rangle \in \mathcal{T}$ and $Q_2^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta) < Q_1^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta) \otimes d$, then update the value $Q_2^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta) := Q_1^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta) \otimes d$.

where the rules are applied in a fair manner; that is, an applicable rule is eventually triggered. The process of rule application is a monotone increasing function, and as such has a least fixpoint, which is the canonical interpretation $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$.

Intuitively, $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ should be a minimal model of $\mathcal{O}$, which describes the necessary conditions of all other models of $\mathcal{O}$. Indeed, the first two rules ensure that the conditions imposed by the ABox are satisfied, while the remaining rules guarantee that all elements of the domain satisfy the positive axioms from the TBox, and each rule is as weak as possible in satisfying these constraints.

However, the construction of $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ does not take the negations into account. The effect of this is that $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ might not be a model of $\mathcal{O}$.

Example 1. Consider the fuzzy DL-Lite ontology $\mathcal{O}_0 = (\mathcal{T}_0, \mathcal{A}_0)$ where

$$\mathcal{T}_0 := \{\langle A_1 \sqsubseteq \neg A_2, 1\rangle\},$$
$$\mathcal{A}_0 := \{\langle A_1(a), 0.5\rangle, \langle A_2(a), 0.5\rangle\}.$$

Under the Gödel semantics, by application of the first rule, the canonical interpretation maps $A_1^{\mathcal{I}_{\mathsf{can}}}(a) = A_2^{\mathcal{I}_{\mathsf{can}}}(a) = 0.5$. However, this violates the axiom in $\mathcal{T}_0$, which requires that $A_1^{\mathcal{I}_{\mathsf{can}}}(a) \Rightarrow \ominus A_2^{\mathcal{I}_{\mathsf{can}}}(a) = 1$. That is, it requires that $A_1^{\mathcal{I}_{\mathsf{can}}}(a) < \ominus A_2^{\mathcal{I}_{\mathsf{can}}}(a)$, which is only possible when $A_1^{\mathcal{I}_{\mathsf{can}}}(a) = 0$ or $A_2^{\mathcal{I}_{\mathsf{can}}}(a) = 0$.

The issue is that the negative axioms may introduce inconsistencies, by enforcing upper bounds in the degrees used, which are not verified by the canonical interpretation. However, as long as there is a model, $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ is one.

Proposition 2. $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ *is a model of* $\mathcal{O}$ *iff* $\mathcal{O}$ *is consistent.*

It can be seen that $\mathcal{O}_0$ from Example 1 is inconsistent under the Gödel semantics. On the other hand, under the Łukasiewicz semantics, $\mathcal{O}_0$ is consistent which, by this proposition, means that $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ is a model of this ontology. This is easily confirmed by recalling that under the Łukasiewicz negation $\ominus 0.5 = 0.5$.

The name *canonical* comes from the fact that, as in the classical case, $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ can be homomorphically embedded in every model of $\mathcal{O}$. We show a similar result with the difference that in this case, the homomorphism needs to take into account the truth degrees from the interpretation function as well. This is described in the following proposition.

Proposition 3. *Let* $\mathcal{O}$ *be a consistent fuzzy DL-Lite ontology,* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ *a model of* $\mathcal{O}$, *and* $\mathcal{I}_{\mathsf{can}}(\mathcal{O}) = (\Delta^{\mathcal{I}_{\mathsf{can}}}, \cdot^{\mathcal{I}_{\mathsf{can}}})$ *its canonical interpretation. There is a function* ψ *from* $\Delta^{\mathcal{I}_{\mathsf{can}}}$ *to* $\Delta^{\mathcal{I}}$ *such that:*

1. *for each* $A \in N_I$ *and* $\delta \in \Delta^{\mathcal{I}_{\mathsf{can}}}$, $A^{\mathcal{I}_{\mathsf{can}}}(\delta) \leq A^{\mathcal{I}}(\delta)$; *and*
2. *for each* $P \in N_R$ *and* $\delta, \eta \in \Delta^{\mathcal{I}_{\mathsf{can}}}$, $P^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta) \leq P^{\mathcal{I}}(\delta, \eta)$.

The consequence of this proposition is that $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ is complete for existential positive queries, and in particular for conjunctive queries.

Corollary 4. *If* $\mathcal{O}$ *is a consistent fuzzy DL-Lite ontology, then for every CQ* $q(\mathbf{x})$ *and answer tuple* $\mathbf{a}$ *it holds that* $\mathcal{O} \models q(\mathbf{a}) \geq d$ *iff* $\mathcal{I}_{\mathsf{can}}(\mathcal{O}) \models q(\mathbf{a}) \geq d$.

Obviously, answering queries through $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ is impractical, because it is an infinite model constructed through an infinitary process. Additionally, we still have the burden to prove that the ontology is consistent, before trying to use Corollary 4 to answer queries. Fortunately, for the Gödel and product t-norms, we resort to existing results from the literature.

Given a fuzzy DL-Lite ontology $\mathcal{O} = (\mathcal{T}, \mathcal{A})$, let $\widehat{\mathcal{O}}$ be its *classical version* defined as $\widehat{\mathcal{O}} := (\widehat{\mathcal{T}}, \widehat{\mathcal{A}})$ with

$$\widehat{\mathcal{T}} := \{B \sqsubseteq C \mid \langle B \sqsubseteq C, d \rangle \in \mathcal{T}, d > 0\} \cup \{Q \sqsubseteq R \mid \langle Q \sqsubseteq R, d \rangle \in \mathcal{T}, d > 0\},$$
$$\widehat{\mathcal{A}} := \{B(a) \mid \langle B(a), d \rangle \in \mathcal{T}, d > 0\} \cup \{P(a, b) \mid \langle P(a, b), d \rangle \in \mathcal{T}, d > 0\}.$$

That is, $\widehat{\mathcal{O}}$ contains all the axioms and assertions from $\mathcal{O}$ which hold with a positive degree—note that any fuzzy axiom or assertion with degree 0 could be removed w.l.o.g. anyway. The following result is a consequence of work on more expressive fuzzy DLs [7].

Proposition 5. *Let $\mathcal{O}$ be a G-DL-Lite$_R$ or π-DL-Lite$_R$ ontology. Then $\mathcal{O}$ is consistent iff $\widehat{\mathcal{O}}$ is consistent.*

In those cases, consistency checking can be reduced to the classical case, without the need to modify the query or the basic formulation of the ontology. For the ontology $\mathcal{O}_0$ in Example 1, we have $\widehat{\mathcal{O}_0} = (\{A_1 \sqsubseteq \neg A_2\}, \{A_1(a), A_2(a)\})$, which is inconsistent in the classical case. We note that the example also shows that Proposition 5 does not hold for the Łukasiewicz t-norm.

In particular, Proposition 5 shows that deciding consistency of G-DL-Lite$_R$ and π-DL-Lite$_R$ ontologies is in AC^0 w.r.t. data complexity and in NLogSpace w.r.t. combined complexity. Thus adding truth degrees does not affect the complexity of this reasoning task. We now turn our attention to the task of query answering with the different semantics, starting with the idempotent case of the Gödel t-norm.

Before studying how to answer queries over fuzzy DL-Lite$_R$ ontologies and its complexity, we note that in the case that an ontology is classical—i.e., it uses only degree 1 in all its axioms—its canonical interpretation constructed as described in this section is equivalent to the classical canonical interpretation from [9]. This fact will be used in the following section.

4   Answering Queries over Gödel Ontologies

The Gödel semantics are very limited in their expressivity. On the one hand, $\ominus d \in \{0, 1\}$ for all $d \in [0, 1]$. This means that whenever we have an axiom of the form $\langle B \sqsubseteq \neg B', d \rangle$ or $\langle Q \sqsubseteq \neg Q', d \rangle$ with $d > 0$, we are in fact saying that for every element $\delta \in \Delta^{\mathcal{I}}$, if $B^{\mathcal{I}}(\delta) > 0$, then $B'^{\mathcal{I}}(\delta) = 0$—because $\ominus B'^{\mathcal{I}}(\delta) = 1$ (and similarly for role axioms). Thus, for this section we can assume w.l.o.g. that all negative axioms hold with degree 1. On the other hand, a positive axiom of the form $\langle B \sqsubseteq B', d \rangle$ requires that for every $\delta \in \Delta^{\mathcal{I}}$, $B'^{\mathcal{I}}(\delta) \geq \min\{B^{\mathcal{I}}(\delta), d\}$. That is, the only way to guarantee that an atom gets a high degree is to use axioms with a high degree. We use these facts to reduce reasoning tasks in this setting to the classical DL-Lite$_R$ scenario.

Consider a consistent G-DL-Lite$_R$ ontology $\mathcal{O}$. We can decide a lower bound for the degree of a CQ simply by querying a *cut* of $\mathcal{O}$. Given a value $d \in (0, 1]$,

the *d-cut* of $\mathcal{O}$ is defined as the sub-ontology $\mathcal{O}_{\geq d} := (\mathcal{T}_{\geq d}, \mathcal{A}_{\geq d})$ where

$$\mathcal{T}_{\geq d} := \{\langle \gamma, e \rangle \in \mathcal{T} \mid e \geq d\},$$
$$\mathcal{A}_{\geq d} := \{\langle \alpha, e \rangle \in \mathcal{A} \mid e \geq d\}.$$

That is, $\mathcal{O}_{\geq d}$ is the subontology containing only the axioms and assertions that hold to degree at least d. To show that d-cuts suffice for answering queries, we use the canonical interpretation.

Note that including new axioms or assertions to an ontology results in an update of the canonical interpretation which only increases the degree of some of the elements of the domain. More precisely, if $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ is the canonical interpretation of $\mathcal{O} = (\mathcal{T}, \mathcal{A})$, then the canonical interpretation of $\mathcal{O}' = (\mathcal{T} \cup \{\langle B \sqsubseteq C, d \rangle\}, \mathcal{A})$ is the result of applying the construction rules starting from $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$. Since $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ has already applied all the rules on axioms of $\mathcal{O}$ exhaustively, the only remaining rule applications will be based on the new axiom $\langle B \sqsubseteq C, d \rangle$ and new applications over $\mathcal{T}$. Under the Gödel semantics, all the updates increase the interpretation function up to the value d; that is, if $\cdot^{\mathcal{I}'_{\mathsf{can}}}$ is the interpretation function of $\mathcal{I}_{\mathsf{can}}(\mathcal{O}')$, the difference between $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ and $\mathcal{I}_{\mathsf{can}}(\mathcal{O}')$ is that there exist some elements such that $A^{\mathcal{I}_{\mathsf{can}}}(\delta) < A^{\mathcal{I}'_{\mathsf{can}}}(\delta) = d$, and similarly for roles. Moreover, if d_0 is the smallest degree appearing in the ontology $\mathcal{O}$, then its canonical interpretation uses only truth degrees in $\{0\} \cup [d_0, 1]$; that is, no truth degree in $(0, d_0)$ appears in $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$. With these insights we are ready to produce our first results. For the rest of this section, we always consider that the semantics is based on the Gödel t-norm; i.e., we have a G-DL-Lite$_R$ ontology.

Lemma 6. *Let $\mathcal{O}$ be a consistent G-DL-Lite$_R$ ontology, $q(\mathbf{x})$ a query, $\mathbf{a}$ a vector of individuals, and $d \in (0, 1]$. Then $\mathcal{O} \models q(\mathbf{a}) \geq d$ iff $\mathcal{O}_{\geq d} \models q(\mathbf{a}) \geq d$.*

Proof. Since $\mathcal{O}_{\geq d} \subseteq \mathcal{O}$, every model of $\mathcal{O}$ is also a model of $\mathcal{O}_{\geq d}$. Hence, if $\mathcal{O}_{\geq d} \models q(\mathbf{a}) \geq d$, then $\mathcal{O} \models q(\mathbf{a}) \geq d$.

For the converse, assume that $\mathcal{O}_{\geq d} \not\models q(\mathbf{a}) \geq d$. By Corollary 4, this means that $\mathcal{I}_{\mathsf{can}}(\mathcal{O}_{\geq d}) \not\models q(\mathbf{a}) \geq d$. That is, $q^{\mathcal{I}_{\mathsf{can}}}(\mathbf{a}^{\mathcal{I}_{\mathsf{can}}}) < d$. Let $\mathcal{I}_{\mathsf{can}}(\mathcal{O}) = (\Delta^{\mathcal{I}'_{\mathsf{can}}}, \cdot^{\mathcal{I}'_{\mathsf{can}}})$ be the canonical interpretation of $\mathcal{O}$. Recall that the difference between $\mathcal{O}$ and $\mathcal{O}_{\geq d}$ is that the former has some additional axioms with degrees smaller than d. As argued before, this means that the difference between $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ and $\mathcal{I}_{\mathsf{can}}(\mathcal{O}_{\geq d})$ are just some degrees, which are all smaller than d; that is, for every $A \in N_C$, $P \in N_R$, and $\delta, \eta \in \Delta^{\mathcal{I}'_{\mathsf{can}}}$, if $A^{\mathcal{I}'_{\mathsf{can}}}(\delta) \geq d$, then $A^{\mathcal{I}_{\mathsf{can}}}(\delta) \geq d$ and if $P^{\mathcal{I}'_{\mathsf{can}}}(\delta, \eta) \geq d$, then $P^{\mathcal{I}_{\mathsf{can}}}(\delta, \eta) \geq d$. By assumption, this means that $q^{\mathcal{I}'_{\mathsf{can}}}(\mathbf{a}^{\mathcal{I}'_{\mathsf{can}}}) < d$ and hence $\mathcal{I}_{\mathsf{can}}(\mathcal{O}) \not\models q(\mathbf{a}) \geq d$. Thus, $\mathcal{O} \not\models q(\mathbf{a}) \geq d$. $\square$

What this lemma states is that in order to find a lower bound for the degree of a query, one can ignore all the axioms and assertions that provide a smaller degree. However, one still needs to answer a query for a fuzzy ontology, for which we do not have any solution. The next lemma solves this issue.

Lemma 7. *Let $\mathcal{O}$ be a consistent G-DL-Lite$_R$ ontology such that $\mathcal{O}_{\geq d} = \mathcal{O}$ for some $d > 0$. Then, $\mathcal{O} \models q(\mathbf{a}) \geq d$ iff $\widehat{\mathcal{O}} \models q(\mathbf{a})$.*

Algorithm 1: Compute the degree of an answer to a query

Data: Ontology $\mathcal{O}$, query q, answer $\mathbf{a}$, $\mathcal{D} = \{d_0, d_1, \ldots, d_{n+1}\}$
Result: The degree of $q(\mathbf{a})$ w.r.t. $\mathcal{O}$

1 $i \leftarrow n + 1$
2 $\mathcal{N} \leftarrow \widehat{\mathcal{O}}_{\geq 1}$
3 **while** $\mathcal{N} \not\models q(\mathbf{a})$ **and** $i > 0$ **do**
4 | $i \leftarrow i - 1$
5 | $\mathcal{N} \leftarrow \widehat{\mathcal{O}}_{\geq d_i}$
6 **return** d_i

Proof. Every model of $\widehat{\mathcal{O}}$ is also a model of $\mathcal{O}$, with the additional property that the interpretation function maps all elements to $\{0, 1\}$. If $\mathcal{O} \models q(\mathbf{a}) \geq d > 0$, then for every model $\mathcal{I}$ of $\widehat{\mathcal{O}}$ it holds that $q^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}) \geq d > 0$ and thus $q^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}) = 1$, which means that $\widehat{\mathcal{O}} \models q(\mathbf{a})$.

Conversely, if $\widehat{\mathcal{O}} \models q(\mathbf{a})$, the canonical interpretation $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ must be such that $q^{\mathcal{I}_{\mathsf{can}}}(\mathbf{a}^{\mathcal{I}_{\mathsf{can}}}) > 0$; but as argued before, since $\mathcal{O}$ only has axioms and assertions with degrees $\geq d$, it must be the case that all degrees of $\mathcal{I}_{\mathsf{can}}(\mathcal{O})$ are in $\{0\} \cup [d, 1]$, and hence $q^{\mathcal{I}_{\mathsf{can}}}(\mathbf{a}^{\mathcal{I}_{\mathsf{can}}}) \geq d$. This implies, by Corollary 4 that $\mathcal{O} \models q(\mathbf{a}) \geq d$. □

These two lemmas together provide a method for reducing bound queries in G-DL-Lite$_R$ to query answering in classical DL-Lite$_R$.

Theorem 8. *If $\mathcal{O}$ is a consistent G-DL-Lite$_R$ ontology and $d > 0$, then it holds that $\mathcal{O} \models q(\mathbf{a}) \geq d$ iff $\widehat{\mathcal{O}}_{\geq d} \models q(\mathbf{a})$.*

This means that we can use a standard ontology-based query answering system to answer fuzzy queries in DL-Lite$_R$ as well. Note that the approach proposed by Theorem 8 can only decide whether the degree of an answer to a query is at least d, but it needs the value d as a parameter. If, instead, we are interested in computing the degree of an answer, or $\mathsf{ans}(q(\mathbf{x}), \mathcal{O})$, we can proceed as follows.

Since the TBox $\mathcal{T}$ and the ABox $\mathcal{A}$ which compose the ontology $\mathcal{O}$ are both finite, the set $\mathcal{D} := \{d \mid \langle \alpha, d \rangle \in \mathcal{T} \cup \mathcal{A}\}$ of degrees appearing in the ontology is also finite; in fact, its size is bounded by the size of $\mathcal{O}$. Hence, we can assume that $\mathcal{D}$ is of the form $\mathcal{D} = \{d_0, d_1, \ldots, d_n, d_{n+1}\}$ where $d_0 \geq 0, d_{n+1} \leq 1$ and for all $i, 0 \leq i \leq n, d_i < d_{i+1}$. In order to find the degree of an answer $\mathbf{a}$ to a query q, we proceed as follows: starting from $i := n + 1$, we iteratively ask the query $\mathcal{O}_{\geq d_i} \models q(\mathbf{a})$ and decrease i until the query is answered affirmatively, or i becomes 0 (see Algorithm 1). In the former case, d_i is the degree for $q(\mathbf{a})$; in the latter, the degree is 0—i.e., $\mathbf{a}$ is not an answer of q.

During the execution of this algorithm, each classical query needed at line 3 can be executed in AC^0 (and in particular in LOGSPACE) in the size of the data; i.e., the ABox [2]. The iterations in the loop do not affect the overall space used, as one can simply produce a new query every time and clean up the previous information. Overall, this means that the degree of an answer can be computed in LOGSPACE in data complexity.

Corollary 9. *The degree of an answer* **a** *to a query q w.r.t. the G-DL-Lite$_R$ ontology $\mathcal{O}$ is computable in* LogSpace *in data complexity.*

Computing $\mathsf{ans}(q(\mathbf{x}), \mathcal{O})$ is, however, a more complex task. Although we can follow an approach similar to Algorithm 1, where the answers to $q(\mathbf{x})$ are computed for each ontology $\widehat{\mathcal{O}}_{\geq d_i}$, in order to assign the appropriate degree to each answer, we need to either keep track of all the answers found so far, or add a negated query which excludes the answers with a higher degree. In both cases, we end up outside the realm of LogSpace complexity. On the other hand, the whole set of answers $\mathsf{ans}(q(\mathbf{x}), \mathcal{O})$ will usually contain many answers that hold with a very low degree, which may not be of much interest to the user making the query. When dealing with degrees, a more meaningful task is to find the k answers with the highest degree, for some natural number k; i.e., the *top-k answers* of q.

Algorithm 1 once again suggests a way to compute the top-k answers. As in the algorithm, one starts with the highest possible degree, and expands the classical ontology by including the axioms and assertions with a lower degree. The difference is that one stops now when the query returns at least k tuples as answers. At that point, the tuples found are those with the highest degree for the query. As before, each of these queries can be answered in AC^0 in data complexity, which yields a LogSpace upper bound for answering top-k queries in data complexity.

Corollary 10. *Top-k queries over consistent G-DL-Lite$_R$ ontologies can be answered in* NLogSpace *in data complexity.*

5   Non-idempotent t-Norms

We now move our attention to the t-norms that are not idempotent; in particular the product and Łukasiewicz t-norms. Unfortunately, as we will see, the correctness of the reductions presented in the previous section relies strongly on the idempotency of the Gödel t-norm, and does not transfer directly to the other cases. However, at least for the product t-norm, it is still possible to answer some kinds of queries efficiently.

First recall that Proposition 5 holds for the product t-norm as well. Hence, deciding consistency of a π-DL-Lite$_R$ ontology remains reducible to the classical case and thus, efficient. We now show with simple examples that the other results do not transfer so easily.

Example 11. Let $\mathcal{O}_1 := (\mathcal{T}_1, \mathcal{A}_1)$ with $\mathcal{T}_1 := \{\langle A_i \sqsubseteq A_{i+1}, 0.9\rangle \mid 0 \leq i < n\}$ and $\mathcal{A}_1 := \{\langle A_0(a), 1\rangle\}$. Note that $\mathcal{O}_1 = \mathcal{O}_{1 \geq 0.9}$, but the degree for the query $q() = A_n(a)$ is 0.9^n which can be made arbitrarily small by making n large.

Similarly, it is not possible to find the top-k answers simply by layering the d-cuts for decreasing values of d until enough answers can be found.

Example 12. Let $\mathcal{O}'_1 := (\mathcal{T}_1, \mathcal{A}'_1)$, where $\mathcal{A}'_1 := \mathcal{A}_1 \cup \{\langle A_n(b), 0.85\rangle\}$ and $\mathcal{T}_1, \mathcal{A}_1$ are as in Example 11. The top answer for $q(x) = A_n(x)$ is b with degree 0.85, but from $\mathcal{O}'_{1 \geq 0.9}$ we already find the answer a, which is not the top one.

The main point with these examples is that, from the lack of idempotency of the t-norm $\otimes$, we can obtain low degrees in a match which arises from combining several axioms and assertions having a high degree. On the other hand, the product behaves well for positive values in the sense that applying the t-norm to two positive values always results in a positive value; formally, if $d, e > 0$, then $d \otimes e > 0$. Thus, if we are only interested in knowing whether the result of a query is positive or not, there is no difference between the Gödel t-norm and the product t-norm.

Definition 13. *A tuple* **a** *is a* positive answer *to the query* $q(\mathbf{x})$ *w.r.t. the ontology* $\mathcal{O}$ *(denoted* $\mathcal{O} \models q(\mathbf{a}) > 0$*) iff for every model* $\mathcal{I}$ *of* $\mathcal{O}$ *it holds that* $q^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}) > 0$.

Theorem 14. *If* $\mathcal{O}$ *is a consistent* π-DL-Lite$_R$ *ontology, then* $\mathcal{O} \models q(\mathbf{a}) > 0$ *iff* $\widehat{\mathcal{O}} \models q(\mathbf{a})$.

Proof. Every model of $\widehat{\mathcal{O}}$ is also a model of $\mathcal{O}$, with the additional property that the interpretation function maps all elements to $\{0, 1\}$. If $\mathcal{O} \models q(\mathbf{a}) > 0$, then for every model $\mathcal{I}$ of $\widehat{\mathcal{O}}$ it holds that $q^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}) > 0$ and thus $q^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}) = 1$, which means that $\widehat{\mathcal{O}} \models q(\mathbf{a})$.

Conversely, if $\widehat{\mathcal{O}} \models q(\mathbf{a})$, then the canonical interpretation is such that $q^{\mathcal{I}_{\mathsf{can}}}(\mathbf{a}^{\mathcal{I}_{\mathsf{can}}}) > 0$, and hence for every model $\mathcal{I}$ it also holds that $q^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}) > 0$. □

This means that, for the sake of answering positive queries over the product t-norm, one can simply ignore all the truth degrees and answer a classical query using any state-of-the-art engine. In particular, this means that positive answers can be found in AC^0 in data complexity just as in the classical case.

We now briefly consider the Łukasiewicz t-norm, which is known to be the hardest to handle due to its involutive negation and nilpotence, despite being in many cases the most natural choice for fuzzy semantics [6]. As mentioned already, Proposition 5 does not apply to the Łukasiewicz t-norm. That is, there are consistent L-DL-Lite$_R$ ontologies whose classical version is inconsistent (see Example 1). As a result, there is currently no known method for deciding consistency of these ontologies, let alone answering queries. The culprits for this are the involutive negation, which is weaker than the negation used in the other two t-norms, but also the nilpotence, which may combine positive degrees to produce a degree of 0. The latter also means that, even if one could check consistency, it is still not clear how to answer even positive queries.

Example 15. Consider the ontology $\mathcal{O}_2 := (\mathcal{T}_2, \mathcal{A}_2)$ where

$$\mathcal{T}_2 := \{\langle A_0 \sqsubseteq A_1, 0.5\rangle, \langle A_1 \sqsubseteq A_2, 0.5\rangle\}$$
$$\mathcal{A}_2 := \{\langle A_0(a), 1\rangle\}.$$

Note that $\mathcal{O}_2$ is consistent, but there is a model $\mathcal{I}$ (e.g., the canonical interpretation) of this ontology which sets $A_2^{\mathcal{I}}(a^{\mathcal{I}}) = 0$. Hence, a is not a positive answer to the query $q(x) = A_2(x)$ even though it is an answer of $q(x)$ over $\widehat{\mathcal{O}_2}$.

Importantly, if we extend DL-Lite$_R$ with the possibility of using conjunctions in the right-hand side of axioms—which becomes only syntactic sugar in the classical case—one can show following the ideas from [5,6] that deciding consistency of a Ł-DL-Lite$_R$ ontology is NP-hard in combined complexity, going beyond the NLogSpace upper bound for the case considered in this work.

6   Conclusions and Future Work

We have introduced the first method for answering conjunctive queries over fuzzy DL-Lite$_R$ ontologies where also the axioms in the TBox are allowed to constrain the admissible truth degrees, and the semantics is based in the notions of mathematical fuzzy logic, with an underlying t-norm an operators.

We have shown that for the Gödel t-norm dealing with the truth degrees adds only a linear overhead to classical query answering, and can be solved through repeated calls to a classical QA engine, which can be chosen from any of the many efficient existing solutions. Moreover, checking consistency of an ontology is easily reduced to the classical case. Technically, this greatly improves the results from [14], where the query needed to be translated into a new one to handle the different degrees (and the TBox must be classical). In our case, we can take any existing solver, and use it without modification. The drawback is that, by not providing a direct translation, the data complexity of the query answering method jumps from AC^0 to LogSpace; consistency checking is still in AC^0 w.r.t. data complexity and in NLogSpace in combined complexity. Finding the k answers with the highest truth degree remains within the same complexity bounds. Importantly, the LogSpace upper bounds are not necessarily tight. It remains to be seen whether answering w.r.t. G-DL-Lite$_R$ ontologies is in fact LogSpace-hard, or can be solved in AC^0. On the other hand, if one is only interested in deciding whether the degree is greater or equal to a given bound, the complexity lowers to that of the classical case again.

The case of other continuous t-norms is, unfortunately, not as clear as Gödel. Although for any t-norm that is not nilpotent (that is, anyone which behaves as the Gödel or the product t-norms at the beginning of the interval $[0,1]$) ontology consistency trivially reduces to the classical case as in Proposition 5, it is not clear how to handle other reasoning problems, and in particular query answering efficiently. It is important to note that as soon as the t-norm is not idempotent, query-rewriting techniques cannot work as usual; in fact, query rewritings are agnostic to the number of times a single axiom or assertion is used in a derivation, which affects the final result when degrees are accumulated. For t-norms with nilpotent elements, and in particular for the Łukasiewicz t-norm, even the question of consistency is open. In fact, for a minimal extension of Ł-DL-Lite$_R$ consistency can be shown to be NP-hard in combined complexity.

We emphasise that, despite both using degrees in $[0, 1]$ and similar operators, π-DL-Lite$_R$ ontologies are *not* probabilistic ontologies. In particular, in the context of query answering, even if we consider only the assertions in the ABox, the product semantics produce very different results to probabilistic databases [22]. The main difference is that a probabilistic fact spawns two possible worlds— one in which the fact is true, and one where it is false—which pushes the data complexity of query answering to #P-hard. In π-DL-Lite$_R$, graded facts only provide a truth degree, and cannot be considered possible worlds.

In the literature of fuzzy query answering, a usual task is to answer *threshold queries*, where each of the conjuncts in the CQ is associated with a lower truth degree. The technique introduced in this paper is unable to handle this task, and new techniques will need to be developed for it. The most promising approach for the Gödel semantics is a new rewriting, which uses graded facts in the database. We will explore this road in more detail in the future.

References

1. Abiteboul, S., Hull, R., Vianu, V.: Foundations of Databases. Addison Wesley, Boston (1994)
2. Artale, A., Calvanese, D., Kontchakov, R., Zakharyaschev, M.: The DL-Lite family and relations. J. Artif. Intell. Res. **36**, 1–69 (2009). https://doi.org/10.1613/jair.2820
3. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications, 2nd edn. Cambridge University Press, Cambridge (2007)
4. Boppana, R.B., Sipser, M.: The complexity of finite functions. In: van Leeuwen, J. (ed.) Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity, pp. 757–804. Elsevier and MIT Press, Amsterdam, Cambridge (1990)
5. Borgwardt, S., Cerami, M., Peñaloza, R.: Many-valued horn logic is hard. In: Lukasiewicz, T., Peñaloza, R., Turhan, A. (eds.) Proceedings of the First Workshop on Logics for Reasoning about Preferences, Uncertainty, and Vagueness, PRUV 2014. CEUR Workshop Proceedings, vol. 1205, pp. 52–58. CEUR-WS.org (2014). http://ceur-ws.org/Vol-1205/00010052.pdf
6. Borgwardt, S., Cerami, M., Peñaloza, R.: The complexity of fuzzy EL under the Łukasiewicz t-norm. Int. J. Approx. Reason. **91**, 179–201 (2017). https://doi.org/10.1016/j.ijar.2017.09.005
7. Borgwardt, S., Distel, F., Peñaloza, R.: The limits of decidability in fuzzy description logics with general concept inclusions. Artif. Intell. **218**, 23–55 (2015). https://doi.org/10.1016/j.artint.2014.09.001
8. Borgwardt, S., Peñaloza, R.: Fuzzy description logics – a survey. In: Moral, S., Pivert, O., Sánchez, D., Marín, N. (eds.) SUM 2017. LNCS (LNAI), vol. 10564, pp. 31–45. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67582-4_3
9. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The DL-Lite family. J. Autom. Reason. **39**(3), 385–429 (2007)
10. Cuenca Grau, B., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Sattler, U.: OWL 2: the next step for OWL. J. Web Semant. **6**, 309–322 (2008)

11. Furst, M.L., Saxe, J.B., Sipser, M.: Parity, circuits, and the polynomial-time hierarchy. Math. Syst. Theory **17**(1), 13–27 (1984). https://doi.org/10.1007/BF01744431
12. Hájek, P.: Metamathematics of Fuzzy Logic, Trends in Logic, vol. 4. Kluwer (1998). https://doi.org/10.1007/978-94-011-5300-3
13. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms, Trends in Logic, vol. 8. Springer, New York (2000). https://doi.org/10.1007/978-94-015-9540-7
14. Mailis, T., Turhan, A.-Y.: Employing $DL\text{-}lite_R$-reasoners for fuzzy query answering. In: Supnithi, T., Yamaguchi, T., Pan, J.Z., Wuwongse, V., Buranarach, M. (eds.) JIST 2014. LNCS, vol. 8943, pp. 63–78. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-15615-6_5
15. Mailis, T.P., Turhan, A., Zenker, E.: A pragmatic approach to answering CQs over fuzzy DL-lite-ontologies - introducing flite. In: Calvanese, D., Konev, B. (eds.) Proceedings of the 2015 International Workshop on Description Logics (DL 2015). CEUR Workshop Proceedings, vol. 1350. CEUR-WS.org (2015). http://ceur-ws.org/Vol-1350/paper-56.pdf
16. Mostert, P.S., Shields, A.L.: On the structure of semigroups on a compact manifold with boundary. Ann. Math. **65**(1), 117–143 (1957)
17. Pan, J.Z., Stamou, G.B., Stoilos, G., Thomas, E.: Expressive querying over fuzzy DL-lite ontologies. In: Proceedings of the 2007 International Workshop on Description Logics (DL 2007). CEUR Workshop Proceedings, vol. 250. CEUR-WS.org (2007). http://ceur-ws.org/Vol-250/paper_47.pdf
18. Papadimitriou, C.H.: Computational Complexity. Addison Wesley, Boston (1994)
19. Reingold, O.: Undirected connectivity in log-space. J. ACM **55**(4), 17:1–17:24 (2008). https://doi.org/10.1145/1391289.1391291
20. Straccia, U.: Towards top-k query answering in description logics: the case of DL-lite. In: Fisher, M., van der Hoek, W., Konev, B., Lisitsa, A. (eds.) JELIA 2006. LNCS (LNAI), vol. 4160, pp. 439–451. Springer, Heidelberg (2006). https://doi.org/10.1007/11853886_36
21. Straccia, U.: Top-k retrieval for ontology mediated access to relational databases. Inf. Sci. **198**, 1–23 (2012). https://doi.org/10.1016/j.ins.2012.02.026
22. Suciu, D., Olteanu, D., Ré, C., Koch, C.: Probabilistic Databases. Synthesis Lectures on Data Management. Morgan & Claypool Publishers (2011). https://doi.org/10.2200/S00362ED1V01Y201105DTM016
23. Xiao, G., et al.: Ontology-based data access: a survey, pp. 5511–5519. ijcai.org (2018). https://doi.org/10.24963/ijcai.2018/777