Extending the Description Logic $\mathcal{EL}$ with Threshold Concepts Induced by Concept Measures

Franz Baader[1], **Oliver Fernández Gil**[1]

[1]Theoretical Computer Science, TU Dresden, Germany

{franz.baader, oliver.fernandez}@tu-dresden.de

Abstract

This extended abstract reports on a new family of knowledge representation formalisms, which is motivated by the fact that, in some applications, it would be more intuitive and convenient to define certain notions of the domain in an approximate way. For example, in clinical diagnosis, diseases are often linked to a long list of medical signs and symptoms, but patients that have a certain disease rarely show all of them. Instead, one looks for the occurrence of sufficiently many of these signs and symptoms. Classical logic-based knowledge representation formalisms would need to use large disjunctions to express such conditions, which are not only inconvenient to write and comprehend, but also hard to reason about. In a nutshell, we introduce novel representation formalisms extending classical Description Logics that can describe such concepts in a compact and easy-to-comprehend way and have better reasoning complexity than classical formalisms using large disjunctions. This extended abstract summarizes the work presented in (Baader and Fernández Gil 2024).

1 Description logics

In Description Logics (DLs) (Baader et al. 2017), one can define the important notions of an application domain as formal concepts, by stating necessary and sufficient conditions for an individual to belong to a concept. These conditions can be atomic properties required for the individual (expressed by concept names) or properties that refer to relationships with other individuals and their properties (expressed as role restrictions). In addition to a formalism for defining concepts, DLs provide their users with ways of stating terminological axioms in a so-called TBox. The simplest kind of TBoxes are called acyclic TBoxes, which consist of concept definitions without cyclic dependencies among the defined concepts. General TBoxes consist of general concept inclusions (GCI), which can be used to state inclusion constraints between concepts. Data (i.e., information about specific individuals) can be formulated in the ABox, which consists of concept assertions relating individuals to concepts and role assertions relating individuals with each other. Given a knowledge base consisting of a TBox and an ABox, one then wants to infer consequences such as implied subconcept–superconcept relationships between concepts (subsumption problem) or implied element relationships between individuals and concepts (instance problem).

In the DL $\mathcal{EL}$, concepts can be built using concept names as well as the concept constructors conjunction ($C \sqcap D$), existential restriction ($\exists r.C$), and the top concept ($\top$). The more expressive DL $\mathcal{ALC}$ is obtained from $\mathcal{EL}$ by adding negation ($\neg C$), and thus implicitly disjunction ($C \sqcup D$) and value restriction ($\forall r.C$). The DL $\mathcal{EL}$ has drawn considerable attention in the last two decades since, on the one hand, important inference problems are polynomial in $\mathcal{EL}$, even w.r.t. general TBoxes. On the other hand, $\mathcal{EL}$ underlies the OWL 2 EL profile[1] and can be used to define biomedical ontologies, such as the large medical ontology SNOMED CT[2]. In $\mathcal{EL}$ we can, for example, formalize the concept of a *good movie* as a movie that is uplifting, has a simple, but original plot, a likable, an evil, and a funny character, action and love scenes, an unobtrusive sound track, and a happy ending, by using the concept description C_{Movie}, which is defined as:

$$\begin{aligned}&\mathsf{Movie} \sqcap \mathsf{Uplifting} \sqcap \exists\mathsf{plot}.(\mathsf{Simple} \sqcap \mathsf{Original}) \sqcap\\ &\quad \exists\mathsf{character}.\mathsf{Likable} \sqcap \exists\mathsf{character}.\mathsf{Evil} \sqcap\\ &\exists\mathsf{character}.\mathsf{Funny} \sqcap \exists\mathsf{scene}.\mathsf{Action} \sqcap \exists\mathsf{scene}.\mathsf{Love} \sqcap\\ &\quad \exists\mathsf{sound}.\mathsf{Unobtrusive} \sqcap \exists\mathsf{ending}.\mathsf{Happy}.\end{aligned} \tag{1}$$

2 Adding threshold concepts to $\mathcal{EL}$

Like all traditional DLs, $\mathcal{EL}$ is based on classical first-order logic, and thus its semantics is strict in the sense that all the stated properties need to be satisfied for an individual to belong to a concept. In applications where exact definitions are hard to come by, it would be useful to relax this strict requirement and allow for approximate definitions of concepts, where most, but not all, of the stated properties are required to hold. For example, people looking for a movie to watch may have a long list of desired properties (such as the ones stated in the concept C_{Movie} in (1)), but will also be satisfied if many, but not all, of them are met.

In order to support defining concepts in such an approximate way, we introduce a family of DLs extending $\mathcal{EL}$ with threshold concept constructors of the form $C_{\bowtie t}$, where C is an $\mathcal{EL}$ concept, $\bowtie \in \{<, \leq, >, \geq\}$, and t is a *rational* number in $[0, 1]$. The semantics of these new concept constructors is defined using a graded membership function m that, given an $\mathcal{EL}$ concept C and an individual d of an interpretation $\mathcal{I}$, returns a value $m^{\mathcal{I}}(d, C)$ from the interval [0,1],

[1]See http://www.w3.org/TR/owl2-profiles/ .

[2]See http://www.ihtsdo.org/snomed-ct/ .

rather than a Boolean value from $\{0,1\}$. Threshold concepts are then interpreted as

$$(C_{\bowtie t})^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid m^{\mathcal{I}}(d, C) \bowtie t\}.$$

In this way we can, for instance, require a good movie to belong to the $\mathcal{EL}$ concept C_{Movie} in (1) with degree at least 0.8. If we employ the simple graded membership function that returns the percentage of the top-level conjuncts of C satisfied by d, then $(C_{\mathit{Movie}})_{\geq .8}$ is actually equivalent to a disjunction of 45 $\mathcal{EL}$ concepts, each of which is a conjunction of 8 of the 10 top-level conjuncts of C_{Movie}. Even for a large class of more complicated graded membership functions m, we can show that threshold concepts are equivalent to disjunctions of $\mathcal{EL}$ concepts or to conjunctions of negated $\mathcal{EL}$ concepts. However, such large $\mathcal{ALC}$ concepts are much harder to comprehend than the threshold concepts. In addition, using this translation for reasoning purposes does not yield good complexity results, due to the large sizes of these disjunctions and conjunctions as well as the high complexity (ExpTime) of reasoning in $\mathcal{ALC}$.

The DL $\tau\mathcal{EL}(m)$ is obtained from $\mathcal{EL}$ syntactically by adding the new threshold constructors $C_{\bowtie t}$ and semantically by interpreting these constructors using the membership function m. There are, of course, different possibilities for how to define a graded membership function m, and the semantics of the obtained new logic $\tau\mathcal{EL}(m)$ depends on m. Consequently, the complexity of reasoning in $\tau\mathcal{EL}(m)$ may also depend on which function m is used. Instead of investigating this complexity for a single, hand-crafted membership function, our goal was to obtain complexity results for a large class of graded membership functions.

3 Graded membership functions

Our main idea for constructing graded membership functions was to employ concept measures (CMs) $\sim$, which generalize equivalence $C \equiv D$ (concept similarity measures) or subsumption $C \sqsupseteq D$ (directional measures) between concepts C, D by returning a value in the interval $[0,1]$ rather than an element of $\{0,1\}$. One example of such a measure is the directional measure $\sim_{\mathsf{su}}$, which computes the percentage of the top-level conjuncts of C that subsume D. For instance, given $C = B \sqcap \exists r.\top$ and $D = A \sqcap \exists r.A$, we obtain $C \sim_{\mathsf{su}} D = 0.5$ since $D \sqsubseteq \exists r.\top$, but $D \not\sqsubseteq B$. From $\sim_{\mathsf{su}}$, we can obtain a concept similarity measure by combining the values $C \sim_{\mathsf{su}} D$ and $D \sim_{\mathsf{su}} C$. For example, we can take the average or the minimum as combining functions.

Concept measures $\sim$ can be used to define a graded membership function $m_\sim$ as follows.

Definition 1. *Let $\sim$ be a CM. Then, for all interpretations $\mathcal{I}$, the function $m^{\mathcal{I}}_\sim : \Delta^{\mathcal{I}} \times \mathcal{C}_{\mathcal{EL}} \to [0,1]$ is defined as*[3]

$$m^{\mathcal{I}}_\sim(d, C) := \max\{C \sim D \mid D \in \mathcal{C}_{\mathcal{EL}} \text{ and } d \in D^{\mathcal{I}}\}.$$

To ensure that this construction yields a well-defined graded membership function $m_\sim$, the concept measure $\sim$ needs to satisfy additional properties, which we collect under the name *standard measures*. In particular, these conditions ensure that the maximum employed in Definition 1 always exists.

[3]The symbol $\mathcal{C}_{\mathcal{EL}}$ denotes the set of all $\mathcal{EL}$ concepts.

There exists a vast number of standard measures, and we were able to show that there are infinitely many standard CMs $\sim$ such that $m_\sim$ is not computable and reasoning in $\tau\mathcal{EL}(m_\sim)$ is undecidable. We also identified a large class of computable standard CMs $\sim$ such that reasoning in $\tau\mathcal{EL}(m_\sim)$ is decidable. However, our proof of this result is based on the translation of the concepts of $\tau\mathcal{EL}(m_\sim)$ into the decidable DL $\mathcal{ALC}$ mentioned above, which may cause a non-elementary blow-up in the worst case. To obtain logics of the form $\tau\mathcal{EL}(m_\sim)$ for which reasoning has lower complexity, we introduce a restricted class of standard CMs, called *simi-d*, which is based on certain directional instances of the $simi$ framework of (Lehmann and Turhan 2012).

4 Main contributions

Our main contributions are results that determine the computational complexity of reasoning in threshold DLs of the form $\tau\mathcal{EL}(m_\sim)$ for $\sim \in$ *simi-d*. We consider the standard inference problems, i.e., concept *satisfiability*, *subsumption*, ABox *consistency* and *instance* checking. In contrast to what is the case for pure $\mathcal{EL}$, this complexity depends on which kind of TBox formalism is used. For the case without TBox, we obtain the following results.

Theorem 1. *Let $\sim \in$ simi-d. In $\tau\mathcal{EL}(m_\sim)$, satisfiability and consistency are NP-complete, whereas subsumption is coNP-complete. The instance problem is PSpace-complete.*

Unlike the situation for classical DLs, defining appropriate notions of acyclic and general TBoxes is already a nontrivial task for logics of the form $\tau\mathcal{EL}(m_\sim)$. Therefore, we have to omit the rather involved definitions from this extended abstract. We continue with acyclic TBoxes.

Theorem 2. *Let $\sim \in$ simi-d. In $\tau\mathcal{EL}(m_\sim)$, satisfiability, subsumption, consistency and instance w.r.t. acyclic $\tau\mathcal{EL}(m_\sim)$ TBoxes are PSpace-complete problems.*

Similar to the case of $\mathcal{ALC}$, the complexity increases to ExpTime for general TBoxes.

Theorem 3. *Let $\sim \in$ simi-d. In $\tau\mathcal{EL}(m_\sim)$, satisfiability, subsumption, consistency and instance w.r.t. general $\tau\mathcal{EL}(m_\sim)$ TBoxes are ExpTime-complete problems.*

For the instance problem, it is also interesting to consider *data complexity*, measured in the size of the ABox.

Theorem 4. *Let $\sim \in$ simi-d. In $\tau\mathcal{EL}(m_\sim)$, instance checking is coNP-complete w.r.t. general $\tau\mathcal{EL}(m_\sim)$ TBoxes. The lower bound already holds for the case of the empty TBox.*

References

Baader, F., and Fernández Gil, O. 2024. Extending the description logic $\mathcal{EL}$ with threshold concepts induced by concept measures. *Artif. Intell.* 326:104034.

Baader, F.; Horrocks, I.; Lutz, C.; and Sattler, U. 2017. *An Introduction to Description Logic*. Cambridge University Press.

Lehmann, K., and Turhan, A.-Y. 2012. A framework for semantic-based similarity measures for $\mathcal{ELH}$-concepts. In *Proc. of Logics in Art. Intell. - 13th European Conference, JELIA 2012*, volume 7519 of *LNCS*, 307–319. Springer.