Adding Threshold Concepts to the Description Logic $\mathcal{EL}$

Oliver Fernández Gil

Abstract

The main goal of this thesis is to extend the lightweight Description Logic $\mathcal{EL}$ with threshold concepts. Description Logics (DLs) are well-defined fragments of first-order logic (FOL) which use concept descriptions to define important notions of the domain by stating necessary and sufficient conditions for an individual to belong to the concept. Such concept descriptions are built inductively, starting from a set of concept names (unary predicates) and a set of role names (binary predicates) and using a set of concept constructors. The expressivity of a specific DL is determined by what sort of properties can be required and how they can be combined. In particular, in $\mathcal{EL}$ concepts can be built using the constructors *conjunction* ($\sqcap$), *existential restriction* ($\exists r.C$) and the *top concept* ($\top$). Then in $\mathcal{EL}$ we can, for instance, define the concept of a *happy-man* as:

$$\begin{aligned}&\mathsf{Human} \sqcap \mathsf{Male} \sqcap \mathsf{Healthy} \sqcap \mathsf{Handsome} \sqcap \\ &\exists \mathsf{spouse}.(\mathsf{Rich} \sqcap \mathsf{Intelligent} \sqcap \mathsf{Female}) \sqcap \\ &\exists \mathsf{child}.\mathsf{Male} \sqcap \exists \mathsf{child}.\mathsf{Female} \sqcap \exists \mathsf{friend}.\top\end{aligned} \tag{1}$$

Since $\mathcal{EL}$ is a fragment of FOL, concept descriptions are interpreted using the classical semantics of FOL, i.e., for an individual to belong to this concept it has to satisfy all the stated properties. However, maybe we would still want to call a man *happy* if most, though not all, of the properties hold.

We propose an extension of $\mathcal{EL}$ that allows us to define concepts in such an approximate way. The main idea is to use a membership degree function, which instead of giving a value in $\{0,1\}$ to evaluate the membership of an individual into a concept, gives a value in the interval [0..1]. Then from a classical $\mathcal{EL}$ concept description C, we can build the threshold concept $C_{\geq t}$ for t a rational number in [0..1], which collects all the individuals that belong to C with degree at least t. Additionally, we also allow the construction of lower threshold concepts of the form $C_{\leq t}$ and the use of strict inequalities. For example, an *unhappy* man could be required to belong to the $\mathcal{EL}$ concept (1) with a degree less than .2. Using the newly introduced constructors and defining their underlying semantics based on a graded membership function, we define the DL $\tau\mathcal{EL}(m)$ where m is a parameter of the logic representing the chosen function.

We then go further and define a particular membership degree function *deg* whose definition is a natural extension of the well-known homomorphism characterization for membership in $\mathcal{EL}$. Once *deg* has been defined, we investigate the computational properties of the threshold logic $\tau\mathcal{EL}(deg)$ obtained by using *deg*. Basically, we look at standard reasoning tasks in DLs, e.g., *satisfiability*, *subsumption*, ABox *consistency* and *instance checking*. It turns out that satisfiability and consistency are NP-complete problems, while subsumption and instance checking (data complexity) are coNP-complete. Furthermore, we extend our logic $\tau\mathcal{EL}(deg)$ to consider concept descriptions defined in a background TBox. Unfortunately, the presence of TBoxes apparently increases the computational complexity of the *satisfiability* and *subsumption* problems, namely, they become Π_2^P- and Σ_2^P-hard, respectively. These hardness results hold already with respect to acyclic TBoxes. Regarding upper bounds, we design a non-deterministic polynomial space algorithm that solves both problems, thus providing membership in PSPACE for both of them. Moreover, these PSPACE upper bounds carry over to reasoning with respect to acyclic knowledge bases.

The last part of the thesis is devoted to better understand the relationship between concept similarity measures and our threshold logic formalism. We describe how to use a variant of the relaxed instance query approach of Ecke et. al.[1] to turn a concept similarity measure (CSM) $\bowtie$ into a graded membership function $m_{\bowtie}$. In this way, we obtain a variety of logics $\tau\mathcal{EL}(m_{\bowtie})$ that could be useful in diverse scenarios according to the specific properties of their underlying similarity measures. Moreover, we show that the relaxed instance queries of Ecke et. al. can be expressed as instance queries w.r.t. threshold concepts

[1]Ecke, A., Peñaloza, R., Turhan, A.: Answering instance queries relaxed by concept similarity. In Proc. KR'14

of the form $C_{>t}$ in the resulting logic $\tau\mathcal{EL}(m_{\bowtie})$. Afterwards, we explore the computational complexity landscape of reasoning in such a big family of threshold logics. We obtain undecidability and decidability results, as well as more precise complexity results for logics induced by a particular class of measures satisfying certain properties.

Finally, we also look at CSMs that are instances of the *simi framework* introduced by Lehman and Turhan[2] and show that one such instance corresponds to our membership degree function *deg*.

[2]Lehmann, K., Turhan, A.: A framework for semantic-based similarity measures for $\mathcal{ELH}$-concepts. In Proc. JELIA'12, LNCS 7519.