A New Dimension to Generalization: Computing Temporal $\mathcal{EL}$ Concepts from Positive Examples

Extended Abstract

Satyadharma **Tirtarasa**[1], Anni-Yasmin **Turhan**[1]

[1]*Technische Universität Dresden, Nöthnitzer Str. 46, 01219, Germany*

Abstract

The authoring of complex concepts or (instance) queries written in a description logic (DL) can be a difficult task. An established approach is to generate such concepts from positive examples is to employ the most specific concept (msc), which generalizes an ABox individual into a concept and the least common subsumer (lcs), which generalizes a collection of concepts into a single concept. These inferences are investigated for the $\mathcal{EL}$, but so far there are no methods for the 2-dimensional case such as temporalized DLs. We report in this abstract on our computation algorithms for the lcs and the msc in a temporalized DL: $\mathcal{EL}$ extended by the LTL operators next and global. We sketch the computation algorithms for both inferences—with and without the use of rigid symbols.

Keywords

generalization, temporal reasoning, learning from examples

Concepts written in a particular DL that occur in description logic knowledge bases or in queries usually model a notion from the application domain. When DL KBs or complex query concepts are built, it is necessary to formalize the intended meaning by authoring complex concepts. These can then be used in TBox axioms or can serve as instance queries. It can be difficult for users to come up with such a formalization. This task is even more difficult, if there are temporal aspects to be formalized in the intended concept. Often it is easier for a user to select example instances of a (temporalized) concept available in the data in the knowledge base. An approach to automatically generate complex concepts from a set of selected examples is the *bottom-up approach* [1] where the resulting concept is computed by applying two generalization inferences. The first is the *most specific concept* (MSC), which computes from an ABox individual a concept (in a given DL) that is the least w.r.t. subsumption. The second inference is the *least common subsumer* (LCS), which computes from a set of input concepts a concept that subsumes all input concepts and is the least concept w.r.t. subsumption. Applying the MSC to each example selected by the user and then applying the LCS to all of the resulting concepts, yields a concept description of the selected examples and is a first version of the intended (query) concept.

The inferences LCS and MSC are typically studied for DLs that do not admit disjunction, since in the presence of this concept constructor the LCS is simply the disjunction of the

input concepts and thus not informative. Both inferences, the LCS and the MSC, have been investigated for *unfoldable TBoxes*. If a cyclic or even general TBox is used, the LCS need not exist, since cyclic concepts cannot be expressed by finite concepts. For cyclic ABoxes the MSC need also not exist, due to the same reason [2]. Conditions for the existence of the LCS (and the MSC) in $\mathcal{EL}$ w.r.t. general TBoxes have been given in [3] and approximations that limit the role depth in concepts have been studied in [2, 4, 5].

Here we report on our paper recently presented at the SAC conference [6], where we provide a first study on the computation of the LCS and the MSC for temporalized DLs. Temporalized DLs have been intensively investigated in the last decade mainly motivated by ontology-based situation recognition, where the situation to be recognized is formalized as a temporal query. The task of situation recognition then boils down to answer a temporal query over a classical DL TBox and a sequence of ABoxes. For recent surveys on reasoning in this kind of settings see [7, 8]. Temporalized DLs are a kind of two dimensional DL, where two logic formalisms are combined [9, 10]. Two dimensional DLs usually have an object dimension, which corresponds to classical DLs and a context dimension [11] or even temporal evolution of context-sensitive domain [12]. Temporal DLs based on LTL have a linear context dimension.

Example-based learning for two dimensional DLs or temporalized DLs has so far not been studied in the literature. In case of temporalized DLs automated support for authoring temporal query concepts (or even temporal conjunctive queries) would be very helpful for situation recognition applications.

In our paper, we lift the LCS and MSC to the temporal setting. More precisely, we consider the logic $\text{LTL}_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$, which uses the temporal operators next ($\mathbf{X}$) and global ($\mathbf{G}$) from LTL "on top of" $\mathcal{EL}$ constructors. Our choice of temporal operators is motivated by the fact, that the operators until ($\mathbf{U}$) and finally ($\mathbf{F}$) imply a disjunction and would, if being used in the generalization inferences, simply enumerate the variants found in the temporal data. $\text{LTL}_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ *concepts* are defined by the following grammar:

$$C, D ::= A \mid C \sqcap D \mid \exists r.C \mid \mathbf{X} \, C \mid \mathbf{G} \, C \mid \top$$

$\text{LTL}_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ can express concepts that capture temporal aspects over the domain of interest. For example, consider $\text{LTL}_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ concept:

$$C_{\text{ex}} = Person \sqcap \exists isVaccinated.BioNTech \sqcap \mathbf{X}(\mathbf{G}(Vaccinated)) \sqcap \mathbf{G}(RiskGroup),$$

which describes persons that are vaccinated using BioNTech and that count from the next time point on as always vaccinated. The last conjunct expresses that these persons are always in the risk group from the initial time point on.

The semantics of $\text{LTL}_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ concepts is given using two dimensions: (1) an infinite temporal sequence of (2) classical DL interpretations. The LTL operators express the evolution of objects in the temporal dimension, while the $\mathcal{EL}$ constructors express relations to other elements within the classical interpretation (at the same time point).

We consider in our paper the generalization inferences only for concepts (w.r.t. unfoldable TBoxes and in case of the MSC also an ABox) as the LCS or MSC need not exist w.r.t. general TBoxes due to cyclic axioms. This means we can rewrite the ABoxes to incorporate the information from the TBox and then ignore the latter. Therefore, we can assume w.l.o.g.
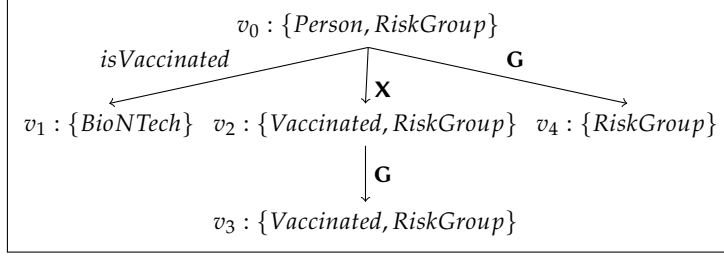
Figure 1: The LTL$_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ description tree that corresponds to normalized form of C_{ex}.

that the input knowledge base consists of only a sequence of ABoxes. As usual for temporal reasoning, we assume rigid domain and rigid individuals.

The bottom-up approach in our temporal setting means that the user selects a set of example individuals from the sequence of ABoxes (possibly from different time points). The MSC is applied to each of them and the resulting LTL$_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ concepts are generalized by the LCS into a single LTL$_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ concept. Our computation algorithms for the LCS and the MSC in LTL$_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ extend the approach from [1] to the temporal case. For ease of presentation we start with the LCS inference. The LCS computation algorithm from [1] performs three steps:

1. transform each input concept into *normal form*, that makes the interaction among the temporal operators explicit
2. represent each normalized concept as a tree-structure, called *description tree*,
3. build *cross-product* of all description trees and read-off the LCS concept from it.

The normalization of LTL$_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ concepts handles the interaction of $\mathbf{X}$ and $\mathbf{G}$. More precisely, it treats the temporal operators $\mathbf{X}$ and $\mathbf{G}$ as functional DL roles and therefore yields a one-dimensional DL concept. This DL concept can then be transformed into a tree. Furthermore, in order to obtain trees of equal height, the normalization performs tree padding. More precisely, it extends the shorter tree by adding $\mathbf{X}$ successors and propagates the information from the sub-tree rooted in the existential restriction for the $\mathbf{G}$ role onto the sub-tree (or rather the chain) rooted in existential restrictions for the $\mathbf{X}$ role. Consider again C_{ex}. This concept which is not normalized, since the effect of $\mathbf{G}(RiskGroup)$ is not propagated to the next timepoint, i.e. into the $\mathbf{X}$-concept. Figure 1 depicts a description tree for C_{ex} after normalization. Notice that the $\mathbf{G}$ information such as RiskGroup and Vaccinated are propagated and made explicit. As it turns out, the normalization can cause the concept to grow exponentially which is due to the interaction of the temporal operators.

Once the description trees are computed, their cross-product is constructed as usual and the LTL$_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ concept can be read-off from it. To be able to show correctness of the LCS computation algorithm, we need to give a characterization of subsumption for LTL$_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ concepts which is one of our main results of the paper. We show that a concept is subsumed by another concept if there exists a homomorphism between their description trees, provided that the trees are aligned. This alignment is crucial to make sure that the interaction between next and global are captured. By using this characterization we show the soundness, completeness and termination of our LCS computation algorithm for LTL$_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$.

To compute the MSC in $\mathcal{EL}$, the procedure from [1] performs the following steps:

1. construct a description graph representing the relational structure in the ABox (incorporating the TBox information),
2. tree-unravel the graph starting from the input individual to generate the description tree and read-off the MSC concept from it.

The extension of this method to the temporal setting needs to construct a description graph from a sequence of ABoxes. The re-occurrence of an individual in the next ABox of the temporal sequence is modeled by connecting the individual and its re-occurrence by $\mathbf{X}$ reflecting the temporal information. For cyclic ABoxes, an exact solution need not exist already for $\mathcal{EL}$, but the construction can be approximated by a bound on the role-depth. In that case, the tree-unraveling the description graphs (up to the bound) can be done as in the case of $\mathcal{EL}$.

In order to show the correctness of our MSC algorithm, we develop a characterization of the instance relationship in $\text{LTL}_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$. In particular, we show that an individual is an instance of a concept (at certain time point) if there exists a homomorphism between the description tree of the concept and the graph by mapping the root of the tree to the individual in the graph. With our graph construction, we can treat the two temporal operators as standard roles in the graph representation when traversing. Using this characterization, we show that our MSC algorithm is sound, complete, and terminating, if the MSC exists w.r.t. the given ABox. While the traversing is linear in the size of the graph, it is possible that the size of the graph is exponentially larger than the initial one due to normalization.

The algorithms for generalization inferences in $\text{LTL}_{\mathcal{EL}}^{\mathbf{X},\mathbf{G}}$ that we have described so far handle non-rigid symbols only. In our paper [6], we also investigate the case of rigid concepts and roles. While rigid names often cause an exponential blow-up for or even undecidability of reasoning in two-dimensional DLs and temporal DLs, this is not the case in our lightweight temporal DL. In fact, it is well-known that rigid concepts can be readily expressed by using $\mathbf{G}$ at the initial time point. Although incorporating rigid roles is more complex, we show that they also can be simulated. Consider there exists $r(a, b)$ in the sequence of ABoxes and r is rigid. Then, we have to introduce $\mathbf{G}\, \exists r.C$ where C is the $\mathbf{G}$ conjunct in the MSC of b. Since the process of normalizing a concept might cause an exponential blow-up already, the computational cost of simulating rigid concepts and roles is not adding to the overall complexity.

Acknowledgments

---

[1] https://perspicuous-computing.science, project ID 389792660

References

[1] F. Baader, R. Küsters, R. Molitor, Computing least common subsumers in description logics with existential restrictions, in: Proc. of the 16th Int. Joint Conf. on Artificial Intelligence (IJCAI'99), 1999.

[2] R. Küsters, R. Molitor, Approximating most specific concepts in description logics with existential restrictions, AI Commun. 15 (2002).

[3] B. Zarrieß, A.-Y. Turhan, Most specific generalizations w.r.t. general $\mathcal{EL}$-TBoxes, in: Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13), AAAI Press, 2013.

[4] R. Peñaloza, A.-Y. Turhan, A practical approach for computing generalization inferences in $\mathcal{EL}$, in: Proceedings of the 8th European Semantic Web Conference (ESWC'11), LNCS, Springer-Verlag, 2011.

[5] A. Ecke, R. Peñaloza, A.-Y. Turhan, Computing role-depth bounded generalizations in the description logic $\mathcal{ELOR}$, in: Proceedings of the 36th German Conference on Artificial Intelligence (KI 2013), LNAI, Springer, 2013, pp. 49–60.

[6] S. Tirtarasa, A.-Y. Turhan, Computing Generalizations of Temporal $\mathcal{EL}$ Concepts with Next and Global, in: Proceedings of the 37th Annual ACM Symposium on Applied Computing (SAC '22), Association for Computing Machinery, 2022.

[7] A. Artale, R. Kontchakov, A. Kovtunova, V. Ryzhikov, F. Wolter, M. Zakharyaschev, Ontology-mediated query answering over temporal data: A survey, in: 24th Int. Sympos. on Temp. Representation and Reasoning, TIME, 2017.

[8] F. Baader, S. Borgwardt, P. Koopmann, V. Thost, A.-Y. Turhan, Semantic technologies for situation awareness, KI – Künstliche Intelligenz (2020).

[9] F. Baader, S. Ghilardi, C. Lutz, LTL over description logic axioms, ACM Trans. Comput. Log. 13 (2012).

[10] K. Schild, Combining terminological logics with tense logic, in: Progress in Artificial Intelligence, 6th Portuguese Conference on Artificial Intelligence, EPIA '93, volume 727 of *Lecture Notes in Computer Science*, Springer, 1993.

[11] S. Böhme, M. Lippmann, Decidable description logics of context with rigid roles, in: Frontiers of Combining Systems - 10th International Symposium, FroCoS„ volume 9322 of *Lecture Notes in Computer Science*, Springer, 2015.

[12] S. Tirtarasa, Temporal properties over contextualized description logics, in: Proceedings of the 33rd International Workshop on Description Logics (DL 2020), volume 2663 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020.