Brushing-up DLs to cope with imperfect data (Extended Abstract)

Anni-Yasmin Turhan[1,2]

[1]Technische Universität Dresden, Nöthnitzer Str. 46, 01219 Dresden, Germany
[2]Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, Germany

Keywords
error-tolerant reasoning, defeasible reasoning, relaxed query answering

1. Introduction

A general view on reasoning in description logics (DLs) is supplied by the framework of ontology-mediated query answering. In recent years reasoning over data enriched by ontologies have become a strong focus of research. The corresponding DL reasoning problems are mainly considered with respect to the classical first-order semantics.

For logic-based applications where data is not curated, but generated automatically as, for instance as in situation recognition [1], noisy or erroneous data can clearly be an obstacle for reasoning under classical First-order semantics. In recent years several approaches have been investigated for reasoning in DLs that deal with this problem — often by changing the underlying semantics. Here we will discuss different reasoning problems using non-standard semantics, such as nonmonotonic or approximative semantics, that can preserve useful logical reasoning even in the presence of imperfect data. For logic-based applications where data is not curated, but generated automatically as for situation recognition [1], noisy or erroneous data can clearly be an obstacle for reasoning under classical First-order semantics. In recent years several approaches have been investigated for reasoning in DLs that deal with this problem —– often by changing the underlying semantics. Here we will discuss different reasoning problems using non-standard semantics, such as nonmonotonic or approximative semantics, that can preserve useful logical reasoning even in the presence of imperfect data.

An obvious obstacle to the use of logical reasoning when using classical semantics is that DL knowledge bases can turn inconsistent in the presence of contradicting information in the data as everything follows from an inconsistent knowledge base. An obstacle to reasoning that is perhaps less obvious is the clear cut semantics of classical query answering. In applications where the exact behaviour of the data sources were not known at the design time of the query or where an exact query is simply difficult to formulate for users, a bit of leeway for query answering can be very useful.

DL 2023: 36th International Workshop on Description Logics, September 2–4, 2023, Rhodes, Greece
anni-yasmin.turhan@tu-dresden.de (A. Turhan)
https://lat.inf.tu-dresden.de/~turhan/ (A. Turhan)
0000-0001-6336-335X (A. Turhan)

2. Methods for error-tolerant reasoning

There are several approaches that have been developed to regain logical reasoning even in the presence of contradictory information in the knowledge base. Most of these approaches abandon the classical first-order semantics and adopt a form of nonmonotonic semantics.

Defeasible DLs (DDLs) are a family of DLs, that use nonmonotonic semantics for reasoning. The idea is to augment classical knowledge bases with a component that stores *defeasible concept inclusions* (DCIs). DCIs state concept inclusions that can be overridden, if contradicting information occurs. In that way DDL knowledge bases can supply standard assumptions that, intuitively, hold for the typical instances of a concept. A popular approach for deciding defeasible subsumption in DDLs is use direct materialization, i.e. use the DCIs as material implication in conjunction with the (suspected) subsumee in the subsumption query. It is well-known [2, 3] that this approach suffers from *quantification neglect* as it does not propagate defeasible information to role-successors exhaustively.

An approach to characterize the semantics of DDLs is by the use of so-called *typicality models* [3, 4]. Here, the idea is to use (partial) copies of the canonical model of $\mathcal{EL}_\bot$ TBoxes that can be enriched with varying amounts of defeasible information from the DBox. More precisely, typicality models can be parameterized with two parameters. The first parameter is the *strength*, which essentially determines the set of subsets of the given DBox that augments the partial copies of the canonical model and thus determines the domain of the typicality interpretation. The selection of these sets can give rise to rational [3] or relevant [5] strength of reasoning. The second parameter is the *coverage* of reasoning, which determines whether the defeasible information is propagated to the potential subsumee from the subsumption query (*propositional coverage*) or whether this kind of information is propagated to all elements in the interpretation domain (*nested coverage*). While the propositional coverage is mainly investigated for legacy reasons as it inherits the properties of reasoning by direct materialization, nested coverage provides semantics that abandon defeasible information only, if it is overridden and thus can alleviate quantification neglect.

In [3, 5, 4, 6] we have devised reasoning methods for propositional coverage by computing minimal typicality models and for nested coverage by computing maximal typicality models. We have also related the resulting inference relations and have provided a complexity analysis. There are two extensions of this basic setting. The first extension is to lift the typicality model-based semantics to ABoxes and to provide an approach to decide instance checking for defeasible $\mathcal{EL}_\bot$ [4, 6]. The second extension is to admit the use of inverse roles for defeasible knowledge bases and lift typicality models for deciding defeasible subsumption accordingly [7, 8].

Closely related to defeasible DLs is reasoning under repair semantics. Here, the ABox is inconsistent with the TBox and consistent versions of the ABox are restored by deleting assertions. As there can be exponentially many repairs even for $\mathcal{EL}_\bot$, repair semantics comprises reasoning under one (brave) and under all repairs, similar to many well-established nonmonotonic logics. Reasoning under repair semantics is currently a very active research area with close relations also to belief revision and semi-ring provenance.

3. Methods for reasoning under relaxed semantics

In case the automatically gathered data is not fitting (the underlying schema assumed in) the query or if concept drift is encountered in longer running applications, the clear cut first-order semantics offers too little flexibility. Relaxing the query can be useful to address these mentioned effects as a relaxed query returns more than just the classical answers. In addition to the classical ones it also returns those answers that are similar to classical answers. Answering such queries is known in the database community as query answering under approximate semantics. The answer set of such queries contains the answer tuple and a numerical value indicating how similar the tuple is to a classical answer. To formulate the exact query answering problem for relaxed queries one needs a means to model the (dis)similarity of answers in addition to the query and the knowledge base. For such relaxed queries the knowledge base stays classical and does not need to be changed. By the use of individual similarity measures for each query, the "direction" of the relaxation can be used to model the user intent of the query. We have investigated two approaches to model and answer relaxed queries.

The first approach relaxes instance queries (sometimes also called concept queries) by the use of concept similarity measures (CSMs). Such CSMs should be well-behaved, i.e. should fulfil properties such as being symmetric and equivalence-invariant and can be constructed for $\mathcal{ELH}$ by the framework from [9] or [10]. We have developed methods for relaxed instance query answering in [11, 10].

More recently, we have investigated relaxed regular path queries. Regular path queries (RPQs) are specified by a non-deterministic finite automaton (NFA) and they retrieve a pair of nodes from a graph that is connected by a path labelled by a word from the language that the automaton accepts. To relax RPQs over graph data bases, Grahne and Thomo [12] have developed an approach that uses weighted transducers as a dissimilarity measure. We have lifted this approach in two ways in [13]. First, to answering relaxed RPQs over knowledge bases written in lightweight DLs and, second, to more expressive query types, i.e. two-way regular path queries. It has been shown in [14] that a finite and more precisely polynomial part of the universal model is sufficient to answer classical RPQs. Now, the essential idea to compute answer tuples under approximate semantics is to use again (a finite part of) the universal model and to perform the cross-product construction of this part, the weighted transducer and the NFA similar to [12] and then retrieve the shortest path for each answer tuple candidate. We have shown the correctness of this method and have supplied a complexity classification for computing the relaxed answers (that are below a given dissimilarity threshold) in [13].

The methods that were developed to make reasoning in DLs more robust against errors in the data can easily be transferred to other closely related formalisms such as existential rules or knowledge graphs. While some of these methods are already well-understood and mature, others are still in its infancy. In particular, methods for answering relaxed queries are missing for conjunctive queries and for expressive DLs.

Acknowledgments

This work was partially supported by the AI competence center ScaDS.AI Dresden/Leipzig and by the DFG through the Collaborative Research Center TRR 248[1].

References

[1] F. Baader, S. Borgwardt, P. Koopmann, V. Thost, A.-Y. Turhan, Semantic technologies for situation awareness, Künstliche Intell. 34 (2020) 543–550.

[2] P. A. Bonatti, M. Faella, I. M. Petrova, L. Sauro, A new semantics for overriding in description logics, Artif. Intell. 222 (2015) 1–48.

[3] M. Pensel, A.-Y. Turhan, Including quantification in defeasible reasoning for the description logic $\mathcal{EL}_{\bot}$, in: Proceedings of the 14th International Conference on Logic Programming and Nonmonotonic Reasoning - LPNMR, Springer, 2017, pp. 78–84.

[4] M. Pensel, A.-Y. Turhan, Reasoning in the defeasible description logic $\mathcal{EL}_{\bot}$—computing standard inferences under rational and relevant semantics, International Journal of Approximate Reasoning (IJAR) 103 (2018) 28–70.

[5] M. Pensel, A.-Y. Turhan, Making quantification relevant again - the case of defeasible $\mathcal{EL}_{\bot}$, in: Proceedings of the 4th International Workshop on Defeasible and Ampliative Reasoning (DARe-17), volume 1872 of *CEUR Workshop Proceedings*, 2017, pp. 44–57.

[6] M. Pensel, A Lightweight Defeasible Description Logic in Depth - Quantification in Rational Reasoning and Beyond, Ph.D. thesis, TU Dresden, Germany, 2019.

[7] I. de Camargo e Souza Câmara, A.-Y. Turhan, Deciding subsumption in defeasible $\mathcal{ELI}_{\bot}$ with typicality models, in: Logics in Artificial Intelligence - 18th European Conference, JELIA 2023, LNCS, Springer, 2023. To appear.

[8] I. de Camargo e Souza Câmara, Quantification in Description Logics of Typicality, Ph.D. thesis, University of São Paulo, Brazil, 2023. To Appear.

[9] K. Lehmann, A.-Y. Turhan, A framework for semantic-based similarity measures for $\mathcal{ELH}$-concepts, in: Proceedings of the 13th European Conference on Logics in Artificial Intelligence, (JELIA'12), volume 7519 of *LNCS*, Springer, 2012, pp. 307–319.

[10] A. Ecke, R. Peñaloza, A.-Y. Turhan, Similarity-based relaxed instance queries, J. Appl. Log. 13 (2015) 480–508.

[11] A. Ecke, R. Peñaloza, A.-Y. Turhan, Answering instance queries relaxed by concept similarity, in: Proceedings of the Fourteenth International Conference on Principles of Knowledge Representation and Reasoning (KR'14), AAAI Press, 2014.

[12] G. Grahne, A. Thomo, Regular path queries under approximate semantics, Ann. Math. Artif. Intell. 46 (2006) 165–190.

[13] O. Fernández Gil, A.-Y. Turhan, Answering regular path queries under approximate semantics in lightweight description logics, in: Proceedings of the Thirty-fourth AAAI Conference on Artificial Intelligence, (AAAI), AAAI Press, 2021.

[14] M. Bienvenu, M. Ortiz, M. Simkus, Regular path queries in lightweight description logics: Complexity and algorithms, J. Artif. Intell. Res. 53 (2015) 315–374.

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