



Comenius University in Bratislava
Faculty of Mathematics, Physics and
Informatics



Mgr. Júlia Pukancová

PhD. Thesis Summary

(Autoreferát dizertačnej práce)

Direct Approach to ABox Abduction in Description Logics

for the academic title philosophiae doctor

in the field of doctoral study: Computer Science, Informatics

Bratislava, 2018

The dissertation was conducted within the internal doctoral program at Department of Applied Informatics at Faculty of Mathematics, Physics and Informatics of Comenius University in Bratislava.

Applicant: Mgr. Júlia Pukancová
Department of Applied Informatics
Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava
Mlynská dolina
842 48 Bratislava

Advisor: doc. PhDr. Ján Šefránek, CSc.
Department of Applied Informatics
Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava
Mlynská dolina
842 48 Bratislava

Computer Science, Informatics

Field committee chair:
.....

1 Introduction

Ontologies are formal structures capturing hierarchies between classes relevant for a specific domain. They are used to describe semantics of data in many application domains: database management and querying, multi-agent systems, biomedical information systems, e-commerce applications, e-learning, etc. (Staab & Studer 2004).

Description logics (Baader et al. 2003) are a family of languages used as representation formalism for ontologies. The main advantage of description logics is that they enable to reason over ontologies, including checking for ontology consistency and derivation of consequences. In fact, description logics constitute a decidable fragment of first order logic.

In the research area of description logics, the standard reasoning problems that are most intensively investigated include consistency checking, subsumption checking, and instance retrieval. A number of well optimized reasoners were developed (Sirin et al. 2007, Shearer et al. 2008, Steigmiller et al. 2014, Haarslev et al. 2012, Horrocks 1998*a,b*). The description logics reasoners are mainly focused on these standard problems, and they are often highly optimized employing techniques such as tableau caching and incremental reasoning (Tsarkov et al. 2007).

Abductive reasoning, originally introduced by Peirce (1878), is a form of non-monotonic reasoning which is used to derive an explanation for an observed phenomenon. Given a theory Γ and an observation O that is not entailed by the theory (i.e. $\Gamma \not\models O$), we are looking for an explanation \mathcal{E} such that $\Gamma \cup \mathcal{E} \models O$. That means, with extending the theory Γ by the explanation \mathcal{E} , the observation O is entailed. A classical example of abductive reasoning is with having the rule *when it rains, the grass is wet* and the observation *the grass is wet*. Intuitively, this observation is explained by the assumption that *it rains*.

In the area of ontologies, abduction was investigated in context of a number of application domains. This type of reasoning is applied for example in diagnostic problems (Hubauer et al. 2011), ontology debugging (Wei-Kleiner et al. 2014), semantic match-making (Colucci et al. 2005), multimedia interpretation (Petasis et al. 2013, Kaya et al. 2007).

While there is a number of applications for abduction, there is not yet a large number of works dedicated to abduction in descriptions logics. Elsenbroich et al. (2006) established a categorization of the main abduction problems. In their work, they defined four types of abduction according to a class of observations and a class of explanations. Namely, the four types are concept abduction, knowledge base abduction, ABox abduction, and TBox abduction. Since the knowledge in the description logic knowledge base is split into the extensional part (ABox) and the intensional part (TBox), obviously in ABox abduction the class of observations and the class explanations are restricted to the data, and in TBox abduction to TBox axioms.

Approaches to solve ABox abduction in description logics algorithmically are split

in the two main groups. On the one hand, some works are based on a translation to another formalism. Klarman et al. (2011) utilizes standard techniques for translation of a description logic knowledge base to the modal logic and first order logic. The approach of Du et al. (2012) is built on the idea to use existing Prolog abduction solver, and so they dealt with a reduction to the logic programming. The works from the other group are based on the solving the abduction problem through standard reasoning techniques for description logics (Halland & Britz 2012a, Ma et al. 2012).

2 Goals

Halland & Britz (2012a) proposed to solve abduction for description logics directly, by extending the standard tableau reasoning techniques (Baader et al. 2003) for description logics and exploiting the minimal hitting set algorithm (Reiter 1987). They assumed higher effectiveness thanks to avoiding a translation into another formalism and thanks to relying on the standard tableau optimization techniques. However they focused mainly on the required adjustment of the tableau algorithm and used the minimal hitting set algorithm as a black box. Completeness of their proposal was not proven and their proposal was also not implemented.

In this work, we build on top of the proposal of Halland and Britz, addressing the above mentioned issues. We focused on the following goals:

- increasing the expressivity of the underlying description logic and extending the classes of observations and explanations,
- formally proving both soundness and completeness,
- implementation using an existing tableau reasoner for description logics,
- empirical evaluation.

3 Results

In this thesis, we have focused on ABox abduction over description logics. We have proposed a sound and complete algorithm for the description logic \mathcal{ALCHO} , based on the minimal hitting set algorithm and the tableau algorithm for description logics. To deal with multiple observations, we have proposed two approaches. The first approach, so called splitting approach, is based on the solving the abduction problem for each observations from an input set of observations. To obtain the result for the original input set, all the respective solutions are combined together, so that they explain the original input set. The second approach is based on a reduction of the input set of observations into one

single observation (i.e. into one concept assertion). The algorithm then simply computes the explanations for the reduced observation as in the case of single observation.

Our algorithm features a number of optimizations in the minimal hitting set algorithm. We have also provided an implementation and an empirical evaluation. The implementation exploits the Pellet reasoner, which belongs to the number of efficient reasoners for description logics.

Our approach addressed the issues spread among the current solutions – to our best knowledge, no other approach have in the same time proposed, developed, implemented, and empirically evaluated a sound and complete direct ABox abduction algorithm for the class of observations of any ABox assertions and the class of explanations of any atomic or negated atomic assertions for the description logic *ALCHO*.

The empirical evaluation of our algorithm was provided for three ontologies in six experiments. First we tested single observations without and with reflexive assertions amongst the explanations. Then we tested multiple observations without and with reflexive assertions amongst the explanations through both approaches – the splitting approach and the reduction approach.

A particular strength of our proposal is in its completeness w.r.t. any given length of explanations. It means, when the algorithm runs up to the depth l , it assures to find all the explanations with the maximal length l . Usually, a high number of explanations were found already for a low maximal length of explanations. We assume, that in many such cases computing the explanations with higher maximal length may not even be desired, as the number of explanations would simply be too high. Moreover, minimal explanations are generally considered as preferred one. The algorithm finds the explanations iteratively with increasing the length, so naturally most of the minimal explanations are found in the first steps.

The evaluation supported the necessity of the limitation on the length of explanations also regarding to execution time and search space. In a case of single observations, the maximal length was set to 5, in a case of the multiple observations, the maximal length was set to 3. The size of the search space grows exponentially with increasing the length of explanations. Thus, also the time grows exponentially, which was showed by the evaluation. The Java heap space memory was exceeded already for the length 3 in the case of multiple observations and two of the three ontologies. However, our results from the evaluation also show how the search space is reduced thanks to the optimizations in the minimal hitting set algorithm.

For the future work, we plan to extend our algorithm for more expressive description logics. The class of explanations is currently restricted to atomic and negated atomic ABox assertions, and so we would like to consider also complex assertions. As in such a case the search space would be infinite, this class of explanations needs to be restricted in some way. One of the most common ways to restrict explanations in abduction is to define abducibles, i.e. a set of assertions, that are potential explanations. We assume this would

allow to operate with the size of the search space, and so more interesting experiments can be conducted. The restriction on abducibles seems to be important also from the point of view of practical use, as indicated also by our evaluation. In our opinion, this restriction is a realistic constraint that can help to boost effectivity as many times the user can rule out a number of uninteresting assertions with respect to the desired explanations.

Currently, our approach computes the minimal explanations regarding to subset minimality. We plan to consider also other types of preferences amongst explanations, such as semantic minimality. Semantic minimality compares explanations with respect to entailment, namely one explanation should not imply the other. We would like to implement also this restriction on our class of explanations, as we consider this to be an interesting extension.

We would also like to exploit other optimization techniques. We will investigate the implemented optimization techniques amongst existing effective reasoners for description logics. One of the most relevant optimization techniques for our algorithm is incremental reasoning (Kazakov & Klinov 2013, Cuenca Grau et al. 2010). Its relevance lies in the fact that our algorithm works with the input knowledge base extended with the observation, and the consistency is checked iteratively after adding and removing assertions. The reuse of the existing tableau is non-trivial (mainly in case of the assertion removing) and special techniques need to be applied. Incremental reasoning deals exactly with this problem.

Our implementation can be extended with other reasoners for description logics by exploiting OWL API (Horridge & Bechhofer 2011). Consecutively, an evaluation with the focus on the comparison of the particular reasoners can be conducted.

Interesting for our work is also to propose an ABox abduction algorithm for less expressive description logics. The lower expressivity is sufficient in many applications, and also highly efficient reasoners are implemented for these description logics, e.g. the ELK reasoner (Kazakov et al. 2014). Our algorithm can be also exploited with these reasoners.

From the technical point of view, the splitting approach, in which an ABox abduction problem with an observation $O = \{O_1, \dots, O_n\}$ into n is split into independent subproblems, opens space for parallelization. The solution for each O_i can be computed in parallel.

4 Publications

- Pukancová, J. (2015). Abductive diagnosis of diabetes mellitus with description logics, in ‘Proceedings of the Student Science Conference 2015’, Bratislava, Slovakia, April 22, 2015.
- Pukancová, J. & Homola, M. (2015), Abductive Reasoning with Description Logics: Use Case in Medical Diagnosis, in ‘Proceedings of the 28th International Workshop

on Description Logics (DL 2015), Athens, Greece, June 7-10, 2015’.

- Pukancová, J. & Homola, M. (2016), Tableau-based ABox abduction for description logics: Preliminary report, in ‘Proceedings of the 29th International Workshop on Description Logics, Cape Town, South Africa, April 22-25, 2016’.
- Pukancová, J. & Homola, M. (2017), Tableau-based ABox abduction for the *ALCHO* description logic, in ‘Proceedings of the 30th International Workshop on Description Logics, Montpellier, France, July 18-21, 2017’.

5 Citations

- López-Astorga, M., The role of disjunction in some alleged non-monotonic inferences. In: *Organon F*, Vol. 24, No. 1, 2017, s. 2-9.
- Bibichkov, I. - Sokol, V. - Shevchenko, O.: Ontological knowledge bases productivity optimization through the use of reasoner combination. In: *Easter-European Journal of Enterprise Technologies*, Vol. 5 (89), No. 2, 2017, s. 54.
- 2017 Del-Pinto, W. - Schmidt, R. A.: Forgetting-based abduction in *ALC*-ontologies. In: *1st Workshop on Second-Order Quantifier Elimination and Related Topics : CEUR Workshop Proceedings*, Vol. 2013. Aachen : CEUR-WS, 2017, S. 27-35.

6 Bibliography

Aliseda-Llera, A. (1997), *Seeking explanations: Abduction in logic, philosophy of science and artificial intelligence*, PhD thesis, PhD thesis: Stanford University, Department of Philosophy, Stanford.

Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D. & Patel-Schneider, P. F., eds (2003), *The Description Logic Handbook: Theory, Implementation, and Applications*, Cambridge University Press.

Colucci, S., Noia, T. D., Sciascio, E. D., Donini, F. M. & Mongiello, M. (2003), Concept abduction and contraction in description logics, in ‘Proceedings of the 2003 International Workshop on Description Logics (DL2003), Rome, Italy September 5-7, 2003’.

Colucci, S., Noia, T. D., Sciascio, E. D., Donini, F. M. & Mongiello, M. (2005), ‘Concept abduction and contraction for semantic-based discovery of matches and negotiation spaces in an e-marketplace’, *Electronic Commerce Research and Applications* 4(4), 345–361.

- Cuenca Grau, B., Halaschek-Wiener, C., Kazakov, Y. & Suntisrivaraporn, B. (2010), ‘Incremental classification of description logics ontologies’, *J. Autom. Reasoning* **44**(4), 337–369.
- Du, J., Qi, G., Shen, Y. & Pan, J. Z. (2011), Towards practical ABox abduction in large OWL DL ontologies, *in* ‘Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7-11, 2011’.
- Du, J., Qi, G., Shen, Y. & Pan, J. Z. (2012), ‘Towards practical ABox abduction in large description logic ontologies’, *Int. J. Semantic Web Inf. Syst.* **8**(2), 1–33.
- Du, J., Wang, K. & Shen, Y. (2015), Towards tractable and practical ABox abduction over inconsistent description logic ontologies, *in* ‘Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.’, pp. 1489–1495.
- Elsenbroich, C., Kutz, O. & Sattler, U. (2006), A case for abductive reasoning over ontologies, *in* ‘Proceedings of the OWLED*06 Workshop on OWL: Experiences and Directions, Athens, Georgia, USA, November 10-11, 2006’.
- Guo, Y., Pan, Z. & Heflin, J. (2005), ‘LUBM: A benchmark for OWL knowledge base systems’, *J. Web Sem.* **3**(2-3), 158–182.
- Haarslev, V., Hidde, K., Möller, R. & Wessel, M. (2012), ‘The RacerPro knowledge representation and reasoning system’, *Semantic Web* **3**(3), 267–277.
- Halland, K. & Britz, K. (2012a), Abox abduction in \mathcal{ALC} using a DL tableau, *in* ‘2012 South African Institute of Computer Scientists and Information Technologists Conference, SAICSIT ’12, Pretoria, South Africa, October 1-3, 2012’, pp. 51–58.
- Halland, K. & Britz, K. (2012b), Naïve ABox abduction in \mathcal{ALC} using a DL tableau, *in* ‘Proceedings of the 2012 International Workshop on Description Logics, DL-2012, Rome, Italy, June 7-10, 2012’, Sun SITE Central Europe (CEUR).
- Hladik, J. & Model, J. (2004), Tableau systems for \mathcal{SHIO} and \mathcal{SHIQ} , *in* V. Haarslev & R. Möller, eds, ‘Proc. of the 17th Int. Workshop on Description Logics (DL 2004)’, Vol. 104 of *CEUR Workshop Proceedings*, pp. 168–177.
- Horn, A. (1951), ‘On sentences which are true of direct unions of algebras’, *The Journal of Symbolic Logic* **16**(1), 14–21.
- Horridge, M. & Bechhofer, S. (2011), ‘The OWL API: A java API for OWL ontologies’, *Semantic Web* **2**(1), 11–21.

- Horrocks, I. (1998*a*), The FaCT system, *in* ‘Automated Reasoning with Analytic Tableaux and Related Methods’, Springer, pp. 307–312.
- Horrocks, I. (1998*b*), Using an expressive description logic: FaCT or fiction?, *in* ‘Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR’98), Trento, Italy, June 2-5, 1998.’, pp. 636–649.
- Horrocks, I., Kutz, O. & Sattler, U. (2006), The even more irresistible *SHOIQ*, *in* ‘Proceedings, Tenth International Conference on Principles of Knowledge Representation and Reasoning, Lake District of the United Kingdom, June 2-5, 2006’, pp. 57–67.
- Horrocks, I. & Sattler, U. (2007), ‘A tableau decision procedure for *SHOIQ*’, *Journal of automated reasoning* **39**(3), 249–276.
- Horrocks, I., Sattler, U. & Tobies, S. (1999), Practical reasoning for expressive description logics, *in* ‘Logic Programming and Automated Reasoning, 6th International Conference, LPAR’99, Tbilisi, Georgia, September 6-10, 1999, Proceedings’, Vol. 1705, Springer, pp. 161–180.
- Horrocks, I., Sattler, U. & Tobies, S. (2000), ‘Practical reasoning for very expressive description logics’, *Logic Journal of IGPL* **8**(3), 239–263.
- Hubauer, T., Legat, C. & Seitz, C. (2011), Empowering adaptive manufacturing with interactive diagnostics: A multi-agent approach, *in* ‘Advances on Practical Applications of Agents and Multiagent Systems - 9th International Conference on Practical Applications of Agents and Multiagent Systems, PAAMS 2011, Salamanca, Spain, 6-8 April 2011’, pp. 47–56.
- Kakas, A. C., Kowalski, R. A. & Toni, F. (1992), ‘Abductive logic programming’, *Journal of logic and computation* **2**(6), 719–770.
- Karp, R. M. (1972), Reducibility among combinatorial problems, *in* ‘Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York.’, pp. 85–103.
- Kaya, A., Melzer, S., Möller, R., Espinosa, S. & Wessel, M. (2007), Towards a foundation for knowledge management: Multimedia interpretation as abduction, *in* ‘Proceedings of the 2007 International Workshop on Description Logics (DL2007), Brixen-Bressanone, near Bozen-Bolzano, Italy, 8-10 June, 2007’.
- Kazakov, Y. & Klinov, P. (2013), Incremental reasoning in OWL \mathcal{EL} without bookkeeping, *in* ‘The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I’, pp. 232–247.

- Kazakov, Y., Krötzsch, M. & Simančík, F. (2014), ‘The incredible ELK’, *Journal of Automated Reasoning* **53**(1), 1–61.
- Klarman, S., Endriss, U. & Schlobach, S. (2011), ‘ABox abduction in the description logic \mathcal{ALC} ’, *Journal of Automated Reasoning* **46**(1), 43–80.
- Konolige, K. (1990), A general theory of abduction, in ‘Proceedings AAAI symposium on Automated Abduction, Stanford’, pp. 62–66.
- Ma, Y., Gu, T., Xu, B. & Chang, L. (2012), An ABox abduction algorithm for the description logic \mathcal{ALCI} , in ‘Intelligent Information Processing VI - 7th IFIP TC 12 International Conference, IIP 2012, Guilin, China, October 12-15, 2012. Proceedings’, pp. 125–130.
- Mayer, M. C. & Pirri, F. (1993), ‘First order abduction via tableau and sequent calculi’, *Logic Journal of the IGPL* **1**(1), 99–117.
- Pearl, J. (1987), Embracing causality in formal reasoning, in ‘Proceedings of the 6th National Conference on Artificial Intelligence. Seattle, WA, July 1987.’, pp. 369–373.
- Peirce, C. S. (1878), ‘Deduction, induction, and hypothesis’, *Popular science monthly* **13**, 470–482.
- Peraldi, I. S. E., Kaya, A., Melzer, S. & Möller, R. (2008), On ontology based abduction for text interpretation, in ‘Computational Linguistics and Intelligent Text Processing, 9th International Conference, CICLing 2008, Haifa, Israel, February 17-23, 2008, Proceedings’, pp. 194–205.
- Petasis, G., Möller, R. & Karkaletsis, V. (2013), BOEMIE: Reasoning-based information extraction, in ‘Proceedings of the 1st Workshop on Natural Language Processing and Automated Reasoning co-located with 12th International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR 2013), A Corunna, Spain, September 15th, 2013.’, pp. 60–75.
- Pukancová, J. & Homola, M. (2016), Tableau-based ABox abduction for description logics: Preliminary report, in ‘Proceedings of the 29th International Workshop on Description Logics, Cape Town, South Africa, April 22-25, 2016’.
- Pukancová, J. & Homola, M. (2017), Tableau-based ABox abduction for the \mathcal{ALCHO} description logic, in ‘Proceedings of the 30th International Workshop on Description Logics, Montpellier, France, July 18-21, 2017’.
- Reiter, R. (1987), ‘A theory of diagnosis from first principles’, *Artificial intelligence* **32**(1), 57–95.

- Schmidt-Schauß, M. & Smolka, G. (1991), ‘Attributive concept descriptions with complements’, *Artificial intelligence* **48**(1), 1–26.
- Shearer, R., Motik, B. & Horrocks, I. (2008), Hermit: A highly-efficient OWL reasoner., in ‘OWLED’, Vol. 432, p. 91.
- Sirin, E., Parsia, B., Cuenca Grau, B., Kalyanpur, A. & Katz, Y. (2007), ‘Pellet: A practical OWL-DL reasoner’, *Web Semantics: science, services and agents on the World Wide Web* **5**(2), 51–53.
- Staab, S. & Studer, R., eds (2004), *Handbook on Ontologies*, International Handbooks on Information Systems, Springer.
- Steigmiller, A., Liebig, T. & Glimm, B. (2014), ‘Konclude: system description’, *Web Semantics: Science, Services and Agents on the World Wide Web* **27**, 78–85.
- Tsarkov, D., Horrocks, I. & Patel-Schneider, P. F. (2007), ‘Optimizing terminological reasoning for expressive description logics’, *J. Autom. Reasoning* **39**(3), 277–316.
- Van Harmelen, F., Lifschitz, V. & Porter, B. (2008), *Handbook of knowledge representation*, Vol. 1, Elsevier.
- Wei-Kleiner, F., Dragisic, Z. & Lambrix, P. (2014), Abduction framework for repairing incomplete \mathcal{EL} ontologies: Complexity results and algorithms, in ‘Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.’, pp. 1120–1127.

7 Summary

Abduction is a non-monotonic reasoning method used to derive explanations for an observed phenomenon. By enriching the theory with the explanation, the observation is entailed. More definitions of abduction problem exist according to the specific classes of the observations and the explanations, usually implied by the particular formalism. The family of description logics is a core formalism for representing ontologies, currently widespread because of applications in database management and querying, multi-agent systems, biomedical information systems, etc. This thesis is focused on abduction over description logics, as it recently earned an interest because of applications such as diagnostics, ontology debugging, semantic matchmaking, multimedia interpretation, etc. We focus on ABox abduction, since we are particularly interested in explanations at the extensional level, corresponding to the ABox in description logics. The main goal was to develop an ABox abduction algorithm addressing some open issues that are not completely addressed by the current related works. We have developed an ABox abduction algorithm for the description logic *ALCHO* based on the tableau algorithm for description logics and the minimal hitting set algorithm. We have formally proven soundness and completeness of the proposal. We have provided an implementation with focus on optimization techniques. An extensive empirical evaluation was conducted.

8 Sumár

Abdukcia je nemonotónna inferenčná metóda, ktorá sa využíva na odvodzovanie vysvetlení pre pozorované javy. Rozšírením teórie o vysvetlenie docielime vyplývanie daného pozorovania. Existuje viacero definícií abduktívneho problému podľa konkrétnej triedy pozorovaní a vysvetlení, zvyčajne vyplývajúce z príslušného formalizmu. Rodina deskripčných logík je jadrovým formalizmom pre reprezentáciu ontológií. Ich súčasná popularita je daná mnohými aplikáciami napr. v databázových systémoch a dopytovaní, biomedicínskych informačných systémoch, multi-agentových systémoch, atď. Táto práca je zameraná na abdukciu nad deskripčnými logikami, ktorá si v poslednej dobe získala pozornosť z dôvodu využitia v aplikačných doménach ako diagnostika, debugovanie ontológií, sémantické párovanie produktov, interpretácia multimédií, a pod. Zameriavame sa na ABoxovú abdukciu, keďže našim hlavným záujmom sú vysvetlenia na extenzionálnej úrovni, čo korešponduje ABoxu v deskripčných logikách. Hlavným cieľom tejto práce bolo vyvinúť ABox abduktívny algoritmus adresujúci otvorené problémy, ktoré žiadna súčasná práca nerieši komplexne. Navrhli sme ABox abduktívny algoritmus pre deskripčnú logiku *ALCHO* založený na tablovom algoritme pre deskripčné logiky a na minimal hitting set algoritme. Formálne sme dokázali zdravosť a úplnosť nášho návrhu. Algoritmus bol implementovaný s dôrazom na optimalizačné metódy. Bola tiež vykonaná rozsiahla empirická evaluácia.