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 1998a,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 $\Gamma$ and an observation $O$ that is not entailed by the theory (i.e. $\Gamma \not\models O$), we are looking for an explanation $E$ such that $\Gamma \cup E \models O$. That means, with extending the theory $\Gamma$ by the explanation $E$, the observation $O$ is entailed. A classical example of abductive reasoning is with having the rule \textit{when it rains, the grass is wet} and the observation \textit{the grass is wet}. Intuitively, this observation is explained by the assumption that \textit{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 matchmaking (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 $\text{ALCHO}$.

The empirical evaluation of our algorithm was provided for three ontologies in six expe-
riments. First we tested single observations without and with reflexive assertions amongst
the explanations. Then we tested multiple observations without and with reflexive asser-
tions 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 eva-
luation. 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 (Kazakóv & Klíno 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 (Kazakóv 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 \( O = \{O_1, \ldots, O_n\} \) into \( n \) is split into independent subproblems, opens space for parallelization. The solution for each \( O_i \) can be computed in parallel.

4 Publications


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


5 Citations


6 Bibliography


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 $\text{ALCHO}$ 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