Towards Explaining DL Non-entailments by Utilizing Subtree Isomorphisms

Ivan Gocev[0000−0003−0894−6901], Georgios Meditskos[0000−0003−4242−5245], and Nick Bassiliades[0000−0001−6035−1038]

School of Informatics, Aristotle University of Thessaloniki, 54124, Greece
{ivangochev,gmeditsk,nbassili}@csd.auth.gr

Abstract. We present our current state of research on explaining non-entailments by finding isomorphic subtrees of $\mathcal{EL}$-description trees. Our approach extends the set of abducible axioms that consist only of concepts to role restrictions as well. We argue how our approach could find solutions to abduction problems in scenarios where other methods cannot, and illustrate this via an example comparing our approach to existing ones.

Keywords: Ontologies · Explanations · Abduction · Non-entailments

1 Introduction

There has been a significant amount of research done on explanations and ontology debugging in the world of OWL and Description Logics (DLs) [1, 2]. In particular, research has been focused on methods for explaining entailments (inferences) wrt. some background knowledge. One specific type of explanations are so-called justifications, which represent a minimal subset of an ontology that is sufficient enough for an entailment to hold [3]. Similarly, proofs have also been utilized for explaining DL entailments [4].

However, such methods fall short when explaining non-entailments, i.e. axioms that do not logically follow from a knowledge base. Some classical approaches are based on providing counter examples [5], or using abduction [10]. Recently, this topic has received more attention as seen in [6], where explanations in TBox abduction are formulated by mimicking justifications in ontologies, and [7] in which homomorphisms are used to solve abduction problems. Nonetheless, the problem of explaining non-entailments is infrequently investigated and there is still much to be done.

Explaining non-entailments could aid users when updating a terminology where some new changes can be found not to be entailed by the original knowledge base. For e.g., in medicine, explanations of non-entailments could be used to derive new relationships among drugs or symptoms of diseases. We propose that a combination of DLs and graph isomorphisms could yield a standardized approach. To this end, we investigate using subtree isomorphisms to solve abduction problems in DLs. In this paper, we briefly review some relevant work and analyze limitations of existing approaches. Then, we propose how certain gaps could be filled and illustrate our approach with an example.

2 Background and Motivation

Our focus is on the description logic $\mathcal{EL}$ based on two disjoint sets: N_C consisting of atomic concepts (or concept names) denoted by A_i, B_i, and N_R consisting of atomic roles (binary predicates or role names) denoted by r_i, s_i for $i \geq 0$. Complex $\mathcal{EL}$ concept descriptions C, D can be: the top concept ($\top$), atomic concepts ($A \in N_C$), conjunction ($C \sqcap D$), and role restriction ($\exists r.C$). A conjunction of atomic concepts is expressed as $\prod_{i=0}^{n} A_i$, for $n \geq 0$. Let $f(X) = \prod_{i=0}^{n} A_i \sqcap \exists r.(X)$, where X is an $\mathcal{EL}$ concept. Then, $f_0(f_1(f_2(\dots f_d(X)))) = f_0 \circ f_1 \circ f_2 \circ \dots \circ f_d(X)$ depicts nesting of role restrictions up to depth d, where $d \geq 0$. For $d = 0$ we have $f_0(X) = \bigsqcap_{i=0}^{n} A_i \sqcap \exists r_0.(X)$, for $d = 1$, $f_0(f_1(X)) = \bigsqcap_{i=0}^{n} A_i \sqcap \exists r_0.(\bigsqcap_{i=0}^{m} B_i \sqcap \exists r_1.(X))$, etc.

A TBox $\mathcal{T}$ represents terminological knowledge and is a finite set of general concept inclusions (GCIs) of the form $C \sqsubseteq D$ (we say "*C is subsumed by D*" with a meaning "C is included in D" or "D includes C"). We sometimes refer to concept inclusions (CIs) as axioms and denote them as $\alpha_1, \alpha_2, \dots, \alpha_n$. An observation, η, is a specific GCI (or axiom) that we are interested in. We write $\mathcal{T} \models \eta$ if an observation is entailed by a Tbox $\mathcal{T}$, and $\mathcal{T} \not\models \eta$ if an observation is not entailed by a Tbox $\mathcal{T}$.

The abduction problem we are interested in is defined as following:

Definition 1. *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox, C_1 and C_2 concepts defined wrt. $\mathcal{T}$. An abduction problem is a tuple ($\mathcal{T}$, $C_2 \sqsubseteq C_1$), where $\mathcal{T}$ is called the background knowledge, $C_2 \sqsubseteq C_1$ the observation, and $\mathcal{T} \not\models C_2 \sqsubseteq C_1$. A solution to the abduction problem is a hypothesis $\mathcal{H}$ of the form:*

$$\mathcal{H} = \{\alpha | \alpha := (\bigsqcap_{i=0}^{n_1} A_i \sqsubseteq \bigsqcap_{i=0}^{n_2} B_i) \textit{ or } (f_0 \circ f_1 \circ \dots \circ f_{d_1}(X) \sqsubseteq g_0 \circ g_1 \circ \dots \circ g_{d_2}(X))$$
$$\textit{or } (\bigsqcap_{i=0}^{n_1} A_i \sqsubseteq f_0 \circ f_1 \circ \dots \circ f_{d_1}(X)) \textit{ or } (f_0 \circ f_1 \circ \dots \circ f_{d_1}(X) \sqsubseteq \bigsqcap_{i=0}^{n_1} A_i)\}$$

, and $\forall \alpha \in \mathcal{H}$, $\mathcal{T} \not\models \alpha$ and $\mathcal{T} \cup \mathcal{H} \models C_2 \sqsubseteq C_1$.

Consider some terminological knowledge $\mathcal{T}$, and an observation η s.t. $\mathcal{T} \not\models \eta$. A classical approach to explain this non-entailment is using *abduction* to generate a hypothesis $\mathcal{H}$, i.e. a "missing piece", such that when added to the terminological knowledge the observation becomes entailed. In the case that an observation should logically follow from a knowledge base, abduction allows us to find reasons why the observation is not entailed and fix the non-entailment. Dependent on the context of the observation, abduction could be used to explain why CIs are not entailed by some terminological knowledge (TBox abduction) [6], explain why assertions are not entailed by some assertive knowledge (ABox abduction) [8, 9], or a combination of both (knowledge base abduction) [10, 11]. Our focus is on the TBox, with the purpose of explaining concept inclusions.

A natural approach to abduction is to determine a set of possible abducibles [9, 11, 6], i.e. concepts or statements we could abduct. There exist common minimality criteria for constructing solutions to abduction problems, such as subset, size, and semantic minimality [8]. Still, these criteria do not necessarily provide useful information for why an observation is not entailed by a knowledge base. In particular, if we allow abduction of only certain CIs, for e.g. $A \sqsubseteq B$, and exclude CIs with role restrictions such as $A \sqsubseteq \exists r.B$, then a solution to the abduction problem may not be admissible. [6] highlights this challenge and tackles it by abducting all types of CIs found in predefined patters based on justifications. Graph morphisms are also promising in identifying relevant concepts and relations to abduct. [7] addresses this topic and explanations with low explanatory power have been successfully eliminated by utilizing graph homomorphisms. However, homomorphisms capture entailment through axioms such as $A \sqsubseteq B$ and do not explicitly include role restrictions in hypotheses. Our approach differs in that it also allows abduction of role restrictions relevant to the observation of interest, thus covering the cases in which a solution to the abduction problem is not admissible by homomorphisms.

To illustrate this, consider the following TBoxes:

$$\begin{aligned}\mathcal{T}_1 = \{&\exists\text{employment.ResearchPosition} \sqcap \exists\text{qualification.Diploma} \sqsubseteq \text{Researcher},\\ &\exists\text{writes.ResearchPaper} \sqsubseteq \text{Researcher}, \text{Doctor} \sqsubseteq \exists\text{qualification.PhD},\\ &\text{Professor} \equiv \text{Doctor} \sqcap \exists\text{employment.Chair},\end{aligned}$$

$$\begin{aligned}\mathcal{T}_2 = \{&\exists\text{employment.ResearchPosition} \sqcap \exists\text{qualification.Diploma}\\ &\sqcap \exists\text{writes.ResearchPaper} \equiv \text{Researcher}, \text{Doctor} \sqsubseteq \exists\text{qualification.PhD},\\ &\text{Professor} \equiv \text{Doctor} \sqcap \exists\text{employment.Chair},\end{aligned}$$

, s.t. $\mathcal{T}_1$ is originally taken from [7] and $\mathcal{T}_2$ is a modified version to exemplify how various CIs can impact the solution of the abduction problem.

Consider first TBox $\mathcal{T}_1$. Here, we have an observation $\eta :=$ Professor $\sqsubseteq$ Researcher, s.t. $\mathcal{T}_1 \not\models \eta$ and $\mathcal{T}_2 \not\models \eta$. To remedy this, the following axioms need to be added to $\mathcal{T}_1$:

$$\mathcal{H}_1 = \{\text{Chair} \sqsubseteq \text{ResearchPosition}, \text{PhD} \sqsubseteq \text{Diploma}\} \tag{1}$$

The axioms in $\mathcal{H}_1$ are the "missing knowledge" for the observation η to become entailed. Such a hypothesis is an explanation for why an observation does not logically follow from a background knowledge, even though, in reality, it should.

$\mathcal{H}_1$ is produced by first representing the concepts in η as $\mathcal{EL}$-description trees (from now description trees or trees) denoted by T_1 and T_2 for Researcher and Professor respectively, which are merely a graphical representation of concept descriptions [12]. Next, a homomorphism $\varphi : T_1 \mapsto T_2$ is found. Since subsumption is characterized via homomorphisms between description trees, they are used to solve the abduction problem and effectively omit arbitrary hypotheses [7].

However, there are some cases in which homomorphisms cannot be used to find a solution to an abduction problem. Let us now consider TBox $\mathcal{T}_2$. A solution to the abduction problem now is the following:

$$\begin{aligned}\mathcal{H}_2 = \{&\text{Chair} \sqsubseteq \text{ResearchPosition}, \text{PhD} \sqsubseteq \text{Diploma},\\ &\text{Doctor} \sqsubseteq \exists\text{writes.ResearchPaper}\}\end{aligned} \tag{2}$$

In the second case the description trees of the concepts in η are not homomorphic due to structural differences, and mapping cannot be performed. Thus, a solution to the abduction problem is not admissible. Structural differences are expressed as axioms in the form of $A \sqsubseteq \exists r.B$, and are a key part in formulating solutions to abduction problems in cases such as this one.

Our main contribution is an approach for abduction of axioms that consist of concepts and role restrictions as well. We argue how *subtree isomorphisms* can extended abduction to include concepts as well as role restrictions in hypotheses, thus confronting the challenges discussed above.

3 Methodology

Our methodology will be described using the example in the previous section. We generate solutions to abduction problems by identifying isomorphic subtrees of concepts represented as description trees originally described in [12].

Definition 2. *A description tree is of the form $T = (V_T, E_T, v_0, l)$, where T is a tree with root v_0 whose nodes $v \in V_T$ are labeled with $l(v) \subseteq N_C$, and (directed) edges $vr\omega \in E_T$ are labeled with role names $r \in N_R$. The empty label corresponds to the top concept.*

If we denote the tree with root $T(v_i)$, and the corresponding concept C_T, we have $C_T = C_{T(v_0)}$ and $C_{T(v)} = \bigsqcap_{v \in V} l(v) \sqcap \bigsqcap_{vr\omega \in E} \exists r.C_{T(\omega)}$.

The vertex labels originally contain only atomic concepts and we extend them in the induced subtrees to role restrictions as well. To include role restrictions, we start by obtaining induced subtrees of description trees and join vertex labels.

Definition 3. *Given a description tree of the form $T = (V_T, E_T, v_0, l)$, a subtree $S = (V'_S, E'_S, v_0, l')$ is an induced subtree of T iff $V'_S \subseteq V_T$, $E'_S \subseteq E_T$, and $\forall v, u \in V'_S$ if $vru \in E'_S$ then $vru \in E_T$. The nodes are labeled with $l'(v) \subseteq N_C \cup \{\exists r.C\}$, where $r \in N_R$ and C is an $\mathcal{EL}$ concept, and (directed) edges are labeled with role names $r \in N_R$. An induced subtree is a description tree itself.*

Induced subtrees are formed by removing vertices from the original tree and embedding them as $\mathcal{EL}$ concepts within their parents' labels. For any $v \in V_T - \{u\}$, if $vru \in E_T$, then $l'(v) = l(v) \cup \{\exists r.(\sqcap_{u \in V_T} l(u))\}$.

To construct hypotheses we map induced subtrees as following:

Definition 4. *An isomorphism from an induced subtree $S_1 = (V'_1, E'_1, v_0, l'_1)$ to an induced subtree $S_2 = (V'_2, E'_2, \omega_0, l'_2)$ is a bijective mapping $\phi : V'_1 \mapsto V'_2$ such that $\phi(v_0) = \omega_0$ and:*

1. *$vr\omega \in E'_1 \Leftrightarrow \phi(v)r\phi(\omega) \in E'_2$*
2. *$\forall v \in V'_1$, and $\omega \in V'_2$ s.t. $\omega = \phi(v), \mathcal{T} \models \sqcap_{\omega \in V'_2} l'_2(\omega) \sqsubseteq \sqcap_{v \in V'_1} l'_1(v)$*

The first point in Definition 4 is a general notion of graph isomorphisms - preserving graph connectivity. On top of this notion, the semantics are added through mapping $\phi(v_0) = \omega_0$, and point 2 of Definition 4 capturing subsumption wrt. a TBox $\mathcal{T}$. If the labels of mapped vertices are not in a subsumption relation as in point 2, then we could potentially abduct that relation. We use this to formulate hypotheses.

Definition 5. *Let $\mathcal{T}$ be a Tbox, $S_1 = (V_1', E_1', v_0, l_1')$ and $S_2 = (V_2', E_2', v_0, l_2')$ induced subtrees of description trees T_1 and T_2 for $\mathcal{EL}$ concepts C_1 and C_2, respectively. Given an abduction problem $(\mathcal{T}, C_2 \sqsubseteq C_1)$ and isomorphism $\phi : V_1' \mapsto V_2'$, the hypothesis is defined as:*

$$\mathcal{H} = \{\textstyle\bigsqcap l_2'(\omega) \sqsubseteq \bigsqcap l_1'(v) \mid \omega = \phi(v) \text{ for } v \in V_1'\} \tag{3}$$

The hypothesis in Definition 5 coheres to Definition 1 - the abducted subsumption relations $\bigsqcap l_2'(\omega) \sqsubseteq \bigsqcap l_1'(v)$ are in fact axioms of the forms defined in Definition 1. This is due to the fact they we map vertices of induced subtrees that contain concepts and role restrictions in their labels $l_1'(v)$ and $l_2'(\omega)$.

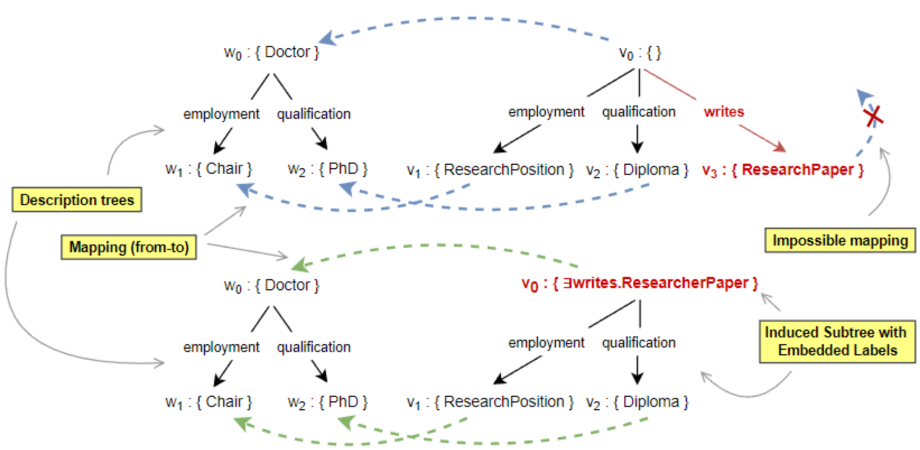

Fig. 1. Description trees of $\mathcal{EL}$ concepts from TBox $\mathcal{T}_2$ (upper part) and induced subtrees with updated labels (lower part). Dashed arcs are potential mappings.

Returning to the example, we discover from the corresponding description trees (T_1 and T_2, respectively) of concepts Researcher and Professor that are neither homomorphic nor isomorphic. This indicates that we cannot map every vertex from T_1 to T_2 wrt. Definition 4. The vertex v_3 from T_1 cannot be mapped to any vertex in T_2, because the structure will not be preserved - the type of role connecting v_0 to v_3 in T_1 does not connect any vertices in T_2 (Figure 1 (upper part)). However, if we have a look at the induced subtrees of T_1 and T_2, we will discover that they are isomorphic and we can find a mapping $\phi(v_0) =$

$\omega_0, \phi(v_1) = \omega_1, \phi(v_2) = \omega_2$ (Figure 1 (lower part)). Thus, from Definition 5 we obtain the hypothesis from eq. 2 and $\mathcal{T}_2 \cup \mathcal{H}_2 \models \text{Professor} \sqsubseteq \text{Researcher}$.

4 Conclusion

We presented work in progress of an approach for generating solutions to abduction problems, by identifying isomorphic subtrees of graphical representations of $\mathcal{EL}$ concepts. We do this by first generating the description trees of $\mathcal{EL}$ concepts. Further, the induced subtrees are constructed and vertex labels are joined accordingly. Finally, the hypothesis is formulated by discovering isomorphisms between the induced subtrees.

Currently, we are investigating further the use of subtree isomorphisms to obtain solutions to abduction problems. Regarding complexity, in a more general sense the subtree isomorphism problem is NP-complete. On the other hand, in the restricted case of $\mathcal{EL}$-description trees testing for existence of homomorphisms can be done in polynomial time [12], which may indicate that the subtree isomorphism problem could also be solved in polynomial time. This investigation is part of our future steps.

References

1. Schlobach, S. and Cornet, R., 2003, August. Non-standard reasoning services for the debugging of description logic terminologies. IJCAI (Vol. 3, pp. 355-362).
2. Kalyanpur, A., 2006. Debugging and repair of OWL ontologies. University of Maryland, College Park.
3. Kalyanpur, A., Parsia, B., Horridge, M. and Sirin, E., 2007. Finding all justifications of OWL DL entailments. In Proc. ISWC/ASWC (pp. 267-280). Springer.
4. Alrabbaa, C., Baader, F., Borgwardt, S., Koopmann, P. and Kovtunova, A., 2020. Finding Small Proofs for Description Logic Entailments: Theory and Practice (Extended Technical Report). arXiv preprint arXiv:2004.08311.
5. Alrabbaa, C., Hieke, W. and Turhan, A.Y., 2021. Counter Model Transformation for Explaining Non-Subsumption in EL. FCR@ KI (pp. 9-22).
6. Du, J., Wan, H. and Ma, H., 2017, February. Practical TBox abduction based on justification patterns. AAAI (Vol. 31, No. 1).
7. Haifani, F., Koopmann, P., Tourret, S. and Weidenbach, C., 2022. Connection-minimal Abduction in EL via Translation to FOL. Proc. IJCAR.
8. Calvanese, D., Ortiz, M., Simkus, M. and Stefanoni, G., 2013. Reasoning about explanations for negative query answers in DL-Lite. JAIR, 48, pp.635-669.
9. Koopmann, P., 2021. Signature-based abduction with fresh individuals and complex concepts for description logics. Description Logics.
10. Elsenbroich, C., Kutz, O. and Sattler, U., 2006. A Case for Abductive Reasoning over Ontologies. In OWLED (Vol. 216).
11. Koopmann, P., Del-Pinto, W., Tourret, S. and Schmidt, R.A., 2020, July. Signature-based abduction for expressive description logics. KR2020 (Vol. 17, No. 1, pp. 592-602).
12. Baader, F., Küsters, R. and Molitor, R., 1999, July. Computing least common subsumers in description logics with existential restrictions. IJCAI (Vol. 99, pp. 96-101).