Learning Description Logic Concepts:  
When can Positive and Negative Examples be Separated?  

Maurice Funk¹, Jean Christoph Jung¹, Carsten Lutz¹, Hadrien Pulcini² and Frank Wolter²  
¹Department of Computer Science, University of Bremen  
²Department of Computer Science, University of Liverpool

Abstract

Learning description logic (DL) concepts from positive and negative examples given in the form of labeled data items in a KB has received significant attention in the literature. We study the fundamental question of when a separating DL concept exists and provide useful model-theoretic characterizations as well as complexity results for the associated decision problem. For expressive DLs such as ALC and ALCQI, our characterizations show a surprising link to the evaluation of ontology-mediated conjunctive queries. We exploit this to determine the combined complexity (between EXPTime and NEXPTime) and data complexity (second level of the polynomial hierarchy) of separability. For the Horn DL EL, separability is EXPTime-complete both in combined and in data complexity while for its modest extension ELI it is even undecidable. Separability is also undecidable when the KB is formulated in ALC and the separating concept is required to be in EL or ELI.

1 Introduction

An important challenge for adopting ontologies in practical applications is the knowledge acquisition bottleneck, that is, the significant time and effort it takes to build the required ontologies. As a promising approach to help overcoming this difficulty, the varied field of ontology learning has received a lot of attention in the last two decades, see [Lehmann and Völker, 2014] for a recent overview. A prominent line of research within ontology learning is concept learning, also called concept induction, where the aim is to learn a structured class description (a concept) formulated in a relevant ontology language from positive and negative examples, given an already available ontology that contains background knowledge [Lehmann et al., 2014]. Applications of concept learning include the support of ontology development and the construction of concept based classifiers [Bühmann et al., 2016; Sarker et al., 2017].

Since description logics (DLs) form the basis for the OWL family of ontology languages, DL concept learning has received particularly much attention [Lehmann and Hitzler, 2010; Lisi, 2012; Tran et al., 2014; Fanizzi et al., 2018]. The precise formulation is as follows. Given a knowledge base (KB) $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ where $\mathcal{T}$ is a TBox that is formulated in a DL $\mathcal{L}_T$ and serves as an ontology providing background knowledge and $\mathcal{A}$ is an ABox in which individuals from designated sets $P$ and $N$ serve as positive and negative examples, respectively, find a concept $C$ formulated in a DL $\mathcal{L}_S$ that separates the positive from the negative examples, that is, $\mathcal{K} \models C(a)$ for all $a \in P$ and $\mathcal{K} \not\models C(a)$ for all $a \in N$. In addition to separation, one also wants to achieve that the learned concept $C$ generalizes the positive examples in a meaningful way, classifying new examples accordingly.

As a prominent system for DL concept learning, we mention DL LEARNER. It encompasses several algorithms that support a range of DLs, including expressive ones such as ALC and ALCQI. Horn DLs such as EL, and even full OWL 2 [Bühmann et al., 2018; Bühmann et al., 2016]. Like competing systems such as DL-FOIL, YINYANG, and pFOIL-DL [Fanizzi et al., 2018; Iannone et al., 2007; Straccia and Mucci, 2015], DL LEARNER uses a carefully crafted refinement operator along with various heuristics to learn concepts that provide an as good as possible generalization of the given examples, avoiding overfitting. The study of such operators originated in [Badea and Nienhuys-Cheng, 2000], see also [Lehmann and Haase, 2009; Lehmann and Hitzler, 2010]. If possible, refinement operators are designed so that the resulting algorithm terminates on any input and is complete in the sense that whenever there is a concept that separates the positive and negative examples in the input, then such a concept is indeed learned.

The aim of this paper is to investigate the fundamental question of when a separating concept exists for a learning instance $(\mathcal{K}, P, N)$, considering the most popular choices of DLs for the TBox language $\mathcal{L}_T$ and the separation language $\mathcal{L}_S$. Our main contributions are model-theoretic characterizations that give important insight into when this is the case and a precise analysis of the computational complexity of separability viewed as a decision problem, which we refer to as $(\mathcal{L}_T, \mathcal{L}_S)$ concept separability and as $\mathcal{L}$ concept separability when $\mathcal{L}_T = \mathcal{L}_S = \mathcal{L}$. We also consider concept definability, the special case of concept separability in which $P \cup N$ comprises all individuals from $\mathcal{A}$. In fact, all our complexity results hold for both separability and definability.

We believe that these problems are relevant both from a practical and from a theoretical perspective. In fact, complex-
ity lower bounds for concept separability point to an inherent complexity that no practical system that aims for completeness can avoid. Undecidability results even mean that there can be no practical learning system that is both terminating and complete. From the viewpoint of machine learning theory, a separating concept corresponds to a consistent hypothesis and understanding the existence of such a hypothesis is considered the most fundamental problem for exploring the version space which in our setup is the space of all separating concepts [Hirsh et al., 2004]. The associated decision problem is often called the consistency problem, and it is known to be closely related to PAC learnability [Pitt and Valiant, 1988; Kietz, 1993].

We cover the expressive DLs ALC, ALCI, ALCQ, and ALCQI as well as the Horn DLs EL and ELI. For the former, overfitting is a risk because the disjunction operator available in such DLs enables the construction of separating concepts that do not provide the desired generalization of the positive examples. Nevertheless, most practical systems such as DL LEARNER work with expressive DLs and avoid overfitting by using appropriate refinement operators. Horn DLs do not admit disjunction and therefore are not prone to overfitting. On the other hand, they provide less separating power and, as we show, tend to incur higher computational (worst-case) cost for learning.

For expressive DLs, we start with initial characterizations in terms of (some form of) bisimulations and then improve them to more refined characterizations based on homomorphisms. Interestingly and unexpectedly to us, these establish a tight link between concept separability and the evaluation of ontology-mediated queries (OMQs) based on unions of rooted conjunctive queries [Calvanese et al., 2013; Bienvenu et al., 2014]. Here, ‘rooted’ means that queries have at least one answer variable and are connected. We use the OMQ connection to obtain complexity upper and lower bounds. In fact, \( \mathcal{L} \) concept separability is NExpTime-complete for \( \mathcal{L} \in \{ \text{ALC}, \text{ALCI}, \text{ALCQ}\} \) while ALCQI concept separability is only ExpTime-complete. This refers to combined complexity where all components of the learning instance are part of the input. We also study data complexity where the ABox is the only input while the TBox is fixed and thus of constant size. In all expressive DLs mentioned above, concept separability is \( \Sigma^P_2 \)-complete in data complexity.

The connection to OMQ evaluation does not extend to Horn DLs, for which we use characterizations based on products of universal models and simulations. Based on these, we show that \( (\mathcal{L}_T, \mathcal{E}L) \) concept separability is ExpTime-complete for \( \mathcal{L}_T \in \{ \mathcal{E}L, \mathcal{E}LI \} \), both in combined complexity and in data complexity. Rather surprisingly, we also prove that \( \mathcal{E}L \) concept separability is undecidable, thus ruling out terminating and complete learning systems. The proof is by a subtle reduction of a tiling problem. We finally consider \( (\mathcal{L}_T, \mathcal{L}_S) \) concept separability where \( \mathcal{L}_T \) is any of the expressive DLs mentioned above and \( \mathcal{L}_S \) is EL or ELI. These problems also turn out to be undecidable.

A stronger version of concept separability that is also considered in the literature requires that \( K \models \neg C(a) \) for all \( a \in N \), rather than only \( K \not\models C(a) \). Clearly, this is a meaningful notion only for DLs with negation. We present some first results also for this version of concept separability, giving model-theoretic characterizations for expressive DLs based on bisimulations and proving that for \( \mathcal{L} \in \{ \text{ALC}, \text{ALCI} \} \), \( \mathcal{L} \) concept separability is \( \text{ExPTIME}\)-complete in combined complexity and \( \text{coNP}\)-complete in data complexity.

The long version with appendix is available at http://www.informatik.uni-bremen.de/tki/research/papers.html.

2 Preliminaries

We introduce the basics of DLs as required for this paper, for full details see [Baader et al., 2017]. Let \( N_C \) be a set of concept names and \( N_R \) a set of role names, both countably infinite. A role is either a role name or an inverse role \( r^- \), \( r \) a role name. For uniformity, we identify \((r^-)^-\) with \( r \). An ALCQI concept is formed according to the syntax rule
\[
C, D ::= \top | A | \neg C | C \sqcap D | (\geq n \ r C)
\]
where \( A \) ranges over concept names, \( r \) over roles, and \( n \) over \( N \). We use \( \exists r.C \) as an abbreviation for \((\geq 1 \ r C), C \sqcup D \) for \((-\neg C \sqcap \neg D), \forall r.C \) for \( \neg \exists r.\neg C \), and \((\leq n \ r C) \) for \((-\geq n + 1 \ r C) \).

There are several fragments of ALCQI that are relevant for this paper. An \( \mathcal{E}L \) concept admits as constructors only the top concept \( \top \), conjunction \( \sqcap \), and existential restriction \( \exists r.C \) with \( r \) a role name (but not an inverse role). ALC concepts additionally admit negation \( \neg \cdot \). Further constructors are indicated by concatenation of a corresponding letter where \( Q \) stands for at least restrictions \((\geq n \ r C)\) and \( I \) for inverse roles. This explains the name ALCQI and allows us to refer to fragments such as EL and ALCCT. We refer to ALC and extensions thereof as expressive DLs and to EL and ELI as Horn DLs.

For any of the DLs \( \mathcal{L} \) introduced above, an \( \mathcal{L} \) TBox is a finite set of concept inclusions (CIs) \( C \subseteq D \), where \( C \) and \( D \) are \( \mathcal{L} \) concepts. Let \( \mathcal{N}_I \) be a countably infinite set of individual names. An ABox \( \mathcal{A} \) is a finite set of concept assertions \( A(a) \) and role assertions \( r(a, b) \) where \( A \in N_C, r \in N_R, \) and \( a, b \in \mathcal{N}_I \). We use \( \text{ind}(\mathcal{A}) \) to denote the set of all individual names that occur in \( \mathcal{A} \). An ABox is associated with a directed graph \( G_A = (\text{ind}(\mathcal{A}), \{(a, b) \mid r(a, b) \in \mathcal{A}\}) \). We use graph theoretic terminology when speaking about ABoxes, implicitly referring to the graph \( G_A \), speaking for example about \( a \) being reachable from \( b \) for \( a, b \in \text{ind}(\mathcal{A}) \). In technical constructions, we will sometimes use extended ABoxes in which concept assertions take the more general form \( C(a) \), \( C \) a concept of the DL under consideration. An \( \mathcal{L} \) knowledge base (KB) \((T, \mathcal{A})\) consists of an \( \mathcal{L} \) TBox \( T \) and an ABox \( \mathcal{A} \), and extended KBs admit extended ABoxes. We use \( \text{sub}(K) \) to denote the set of subconcepts of concepts that occur in \( K \).

As usual, the semantics of DLs is defined in terms of interpretations \( \mathcal{I} = (\Delta^\mathcal{I}, \mathcal{I}) \), where \( \Delta^\mathcal{I} \) is a non-empty set and \( \mathcal{I} \) maps each concept name \( A \in N_C \) to a subset \( A^\mathcal{I} \) of \( \Delta^\mathcal{I} \) and each role name \( r \in N_R \) to a binary relation \( r^\mathcal{I} \) on \( \Delta^\mathcal{I} \). We refer to [Baader et al., 2017] for details on how to extend \( \mathcal{I} \) to compound concepts. An interpretation \( \mathcal{I} \) satisfies a CI \( C \subseteq D \) if \( C^\mathcal{I} \subseteq D^\mathcal{I} \) and is a model of a TBox \( T \) if it satisfies all inclusions in \( T \). An interpretation is a model of an ABox \( \mathcal{A} \) if it satisfies all assertions in \( \mathcal{A} \), that is, \( a \in A^\mathcal{I} \) if
for all \((d, e) \in S: d \in A^2\) iff \(e \in A^3\).

<table>
<thead>
<tr>
<th>Forth</th>
<th>dual of [Forth]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back</td>
<td>dual of [QForth]</td>
</tr>
<tr>
<td>QBack</td>
<td>dual of [QForth]</td>
</tr>
</tbody>
</table>

Figure 1: Conditions on \(S \subseteq \Delta^2 \times \Delta^3\).

\(A(a) \in \mathcal{A}\) and \((a, b) \in \mathcal{r}^2\) if \(r(a, b) \in \mathcal{A}\). We thus make the standard names assumption. An interpretation is a model of a KB \(\mathcal{K}\) if it is a common model of its TBox and ABox and \(\mathcal{K}\) is \(\text{satisfiable}\) if it has a model. The assertion \(C(a)\) is a consequence of the KB \(\mathcal{K}\), in symbols \(\mathcal{K} \models C(a)\), if \(a \in C^2\) for all models \(\mathcal{I}\) of \(\mathcal{K}\).

We next recall model-theoretic characterizations of when elements in interpretations are indistinguishable by concepts formulated in one of the DLs \(\mathcal{L}\) introduced above. A pointed interpretation \(\mathcal{I}, d\) is an \(\mathcal{I}\) interpretation and \(d \in \Delta^2\). For pointed interpretations \(\mathcal{I}, d\) and \(\mathcal{J}, e\), we write \(\mathcal{I}, d \equiv_\mathcal{E} \mathcal{J}, e\) and say that \(\mathcal{I}, d\) and \(\mathcal{J}, e\) are \(\mathcal{E}\)-equivalent if \(d \in C^2\) iff \(e \in C^3\) for all \(\mathcal{E}\) concepts \(C\).

As for the model-theoretic characterizations, we start with \(\mathcal{ALC}\). A relation \(S \subseteq \Delta^2 \times \Delta^3\) is a bisimulation if conditions [Atom], [Forth] and [Back] from Figure 1 hold, where \(\mathcal{A}\) ranges over all concept names, \(r\) over all role names, \(\text{succ}_r^2(d) = \{d' \in \Delta^2 \mid (d, d') \in \mathcal{r}^2\}\), and ‘\(d'\) refers to swapping \(\mathcal{I}, d\) and \(\mathcal{J}, e\) in \(\mathcal{E}\)-bisimulations, as defined by \(\mathcal{E}\)-bisimulations, but with [Forth] and [Back] replaced by [QForth] and [QBack]. For \(\mathcal{ALE}\) and \(\mathcal{ALCQ}\), we define \(\sim_{\mathcal{ALE}}\) by replacing bisimulations with counting bisimulations, as defined by \(\mathcal{E}\)-bisimulations, but with [Forth] and [Back] replaced by [QForth] and [QBack]. For \(\mathcal{ALE}\), \(\mathcal{ALCQ}\), we define \(\sim_{\mathcal{ALE}}\) and \(\sim_{\mathcal{ALCQ}}\) analogously, but now demanding that in all conditions in Figure 1, \(r\) additionally ranges over inverse roles.

For \(\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}\), rather than characterizing \(\equiv_{\mathcal{L}}\) we consider the non-symmetric relation \(\mathcal{I}, d \leq_\mathcal{L} \mathcal{J}, e\) which holds if \(d \in C^2\) implies \(e \in C^3\) for all \(\mathcal{L}\) concepts \(C\). A relation \(S \subseteq \Delta^2 \times \Delta^3\) is an \(\mathcal{L}\)-simulation if \(\mathcal{I}\) to \(\mathcal{J}\) if it satisfies [AtomR] and [Forth] from Figure 1 where again \(\mathcal{A}\) ranges over all concept names and \(r\) over all role names. \(\mathcal{L}\)-simulations are defined in the same way, but with \(r\) ranging also over inverse roles. We write \(\mathcal{I}, d \preceq_\mathcal{L} \mathcal{J}, e\) if there exists an \(\mathcal{L}\)-simulation \(S\) from \(\mathcal{I}\) to \(\mathcal{J}\) with \((d, e) \in S\).

The next lemma summarizes the model-theoretic characterizations for all relevant DLs [Lutz et al., 2011; Goranko and Otto, 2007]. An interpretation \(\mathcal{I}\) has \(\text{finite outdegree}\) if the directed graph \(G_{\mathcal{I}} = (\Delta^2, \bigcup_{r \in \mathcal{R}} I^r)\) has finite outdegree.

**Lemma 1** Let \(\mathcal{I}, d\) and \(\mathcal{J}, e\) be pointed interpretations and let \(\mathcal{E}\) have finite outdegree. Then

1. For \(\mathcal{L} \in \{\mathcal{ALC}, \mathcal{ALCI}, \mathcal{ALCQ}, \mathcal{ALCQI}\}\), \(\mathcal{I}, d \equiv_\mathcal{L} \mathcal{J}, e\) iff \(\mathcal{I}, d \leq_\mathcal{L} \mathcal{J}, e\).
2. For \(\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}\), \(\mathcal{I}, d \leq_\mathcal{L} \mathcal{J}, e\) iff \(\mathcal{I}, d \preceq_\mathcal{L} \mathcal{J}, e\).

**3 Separability and Definability**

We introduce concept separability and concept definability, the main notions studied in this paper. We also provide illustrating examples and give initial model-theoretic characterizations in all relevant DLs.

**Definition 1** Let \(\mathcal{L}_T, \mathcal{L}_S\) be DLs. An \(\mathcal{L}_T\) learning instance is a triple \((\mathcal{K}, P, N)\) with \(\mathcal{K} = (\mathcal{T}, \mathcal{A})\) an \(\mathcal{L}_T\) KB and \(P, N \subseteq \text{ind}(\mathcal{A})\) non-empty sets of positive and negative examples. An \(\mathcal{L}_S\) solution to \((\mathcal{K}, P, N)\) is an \(\mathcal{L}_S\) concept \(C\) such that

1. \(\mathcal{K} \models C(a)\) for all \(a \in P\)
2. \(\mathcal{K} \not\models C(a)\) for all \(a \in N\).

Any TBox language \(\mathcal{L}_T\) and separation language \(\mathcal{L}_S\) give rise to a decision problem of concept separability.

**Example 1** (1) Let \((\mathcal{K}, P, N)\) be defined by \(\mathcal{T} = \emptyset\),

\[
\mathcal{A} = \{\text{authorOf}(a_1, c_1) \mid i = 1, 2, 3\} \cup \{\text{IJApub}(c_1), \text{AIJpub}(c_2), \text{GraphicNovel}(c_3)\},
\]

\(P = \{a_1, a_2\}\), and \(N = \{a_3\}\). Then \(\exists\text{authorOf} . (\text{IJApub} \cup \text{AIJpub})\) is an \(\mathcal{EL}\) solution, but there is neither an \(\mathcal{EL}\) solution nor an \(\mathcal{EL}\) solution.

(2) Let \((\mathcal{K}, P, N)\) be as in (1), but with \(\mathcal{T}\) replaced by the \(\mathcal{EL}\) TBox

\[
\mathcal{T} = \{\exists\text{authorOf} . \text{IJApub} \subseteq \text{AIResearcher}, \exists\text{authorOf} . \text{AIJpub} \subseteq \text{AIResearcher}\}.
\]

Then \(\text{AIResearcher}\) is an \(\mathcal{EL}\) solution.

(3) Let \((\mathcal{K}, P, N)\) be defined by \(\mathcal{T} = \emptyset\),

\[
\mathcal{A} = \{\text{authorOf}(a, c_1), \text{authorOf}(a, c_2), \text{authorOf}(b, c_3)\},
\]

\(P = \{a\}\), and \(N = \{b\}\). Due to the standard names assumption, \(c_1\) and \(c_2\) are distinct objects and thus \(\{2\}\) \text{authorOf} \(\mathcal{T}\) is an \(\mathcal{EL}\) solution. However, there is no \(\mathcal{EL}\) solution even when \(\mathcal{T}\) is replaced with any \(\mathcal{EL}\) TBox.

(4) Let \((\mathcal{K}, P, N)\) be as in (3), but with \(P\) and \(N\) swapped. There is no \(\mathcal{EL}\) solution even when \(\mathcal{T}\) is replaced with any \(\mathcal{EL}\) TBox. Note that \(\{1\}\) \text{authorOf} \(\mathcal{T}\) is not a solution as the semantics does not rule out additional authorOf successors in concrete models.

Observe that if \(\mathcal{L}_S\) is closed under conjunction, then \((\mathcal{L}_T, \mathcal{L}_S)\) concept separability can be reduced in polynomial time to the special case of \((\mathcal{L}_T, \mathcal{L}_S)\) concept separability that admits only input instances \((\mathcal{K}, P, N)\) in which \(N\) is a
singleton. Indeed, if C is a solution to a learning instance (K, P, N), then it is a solution to (K, P, {a}) for all a ∈ N. If, conversely, C_a is a solution to (K, P, {a}) for all a ∈ N, then \( \prod_{a \in N} C_a \) is a solution to (K, P, N). In what follows, we will thus mostly restrict our attention to this special case.

We study both the combined complexity of concept separability and definability and the data complexity, where only the ABox of the KB is regarded as the input to the decision problem while the TBox is fixed and thus of constant size. Note that the sizes of P and N are dominated by the size of the ABox. When not making explicit which complexity measure we speak about, we mean combined complexity.

We now give the announced characterizations of \((L_S, L_T)\) concept separability, starting with the case that \(L_S\) is an expressive DL. The characterizations use the relations \(\sim_L\) from Lemma 1, \(L\) being the separation language. The corresponding relation for the TBox language is not used.

Theorem 1 Let \((K, P, \{b\})\) be an \(\text{ALCQI}\) learning instance. For \(L_S \in \{\text{ALC}, \text{ALCI}, \text{ALCQ}, \text{ALCQI}\}\), the following are equivalent:

1. \((K, P, \{b\})\) has an \(L_S\) solution;
2. there exists a model \(I\) of \(K\) of finite outdegree such that for all \(a \in P\) and all models \(J\) of \(K\), \(a \not\in L_S I, b\).

The proof of Theorem 1 is rather standard, based on Lemma 1 and compactness.

We now turn to the case where \(L_S\) is a Horn DL. The main difference to Theorem 1 is that a product construction is involved. Interestingly, our characterization is rather flexible regarding \(L_T\), which can even be \(\text{ALCQI}\). Let \(I_i, d_i, i \in I\), be a family of pointed interpretations. The (direct) product of pointed interpretations defined by

\[
\Delta \Pi I_i = \{d : I \rightarrow \bigcup_{i \in I} \Delta I_i \mid \text{for } i \in I : d(i) \in \Delta I_i\}
\]

\[
A \Pi I_i = \{d \in \Delta \Pi I_i \mid \text{for } i \in I : d(i) \in A I_i, A \in N_C\}
\]

\[
r \Pi I_i = \{\langle d, e \rangle \mid \text{for } i \in I : \langle d(i), e(i) \rangle \in r I_i, r \in N_R\}
\]

and \(d(i) = d_i\), for all \(i \in I\). We say that a set \(M\) of models of a KB \(K\) is \(L\) complete if for every \(L\) concept \(C\) and every \(a \in \text{ind}(K)\), \(K \models C(a)\) if and only if \(C(a)\) is satisfiable in all \(I \in M\). If, for example, \(K\) is an \(\text{ALCQI}\) KB, then the class of all ‘forest models’ of \(K\) of finite outdegree is well-known to be \(\text{ALCQI}\) complete.

Theorem 2 Let \((K, P, \{b\})\) be an \(\text{ALCQI}\) learning instance. For \(L_S \in \{\text{EL}, \text{ELT}\}\) and \(M\) a set of models of \(K\) that is \(L_S\) complete, the following are equivalent:

1. \((K, P, \{b\})\) has an \(L_S\) solution;
2. there exists a model \(I\) of \(K\) of finite outdegree such that \(\prod_{a \in P}(\prod_{J \in M(I, a)) \not\in L_S I, b}\).

The proof uses Lemma 1 and the fact that for any \(\text{ELT}\) concept \(C\) and family of pointed interpretations \(I_i, d_i\), \(d \in C^{\Pi I_i}\), if and only if \(d_i \in C^{\Delta}\) for all \(i \in I\).

4 Characterizations for Expressive DLs

We establish characterizations of concept separability that are more refined than the initial one presented in Theorem 1. The refined characterizations establish a suprising connection between concept separability and \((U)\text{CQ}\)-evaluation on KBs. They also provide the foundation for decision procedures for concept separability that we develop later on. We start with \(\text{ALCQI}\) and then proceed via \(\text{ALCQ}\) to \(\text{ALC}\) and \(\text{ALCQ}\), for which the characterizations are slightly more technical.

A homomorphism \(h\) from an ABox \(A\) to an interpretation \(I\) is a mapping \(h\) from \(\text{ind}(A)\) to \(\Delta I\) such that \(h(d) \in A\) implies \(\forall c, d \in A\) implies \(\forall c, h(d) \in A\). We write \(A, a \rightarrow I, b\) if there exists a homomorphism \(h\) from \(A\) to \(I\) with \(h(a) = b\).

Theorem 3 Let \((K, P, \{b\})\) be an \(\text{ALCQI}\) learning instance. Then the following conditions are equivalent:

1. \((K, P, \{b\})\) has an \(\text{ALCQI}\) solution;
2. there exists a model \(I\) of \(K\) such that for all \(a \in P\), \(A, a \not\rightarrow I, b\).

Theorem 4 Let \((K, P, \{b\})\) be an \(\text{ALCQ}\) learning instance. Then the following conditions are equivalent:

1. \((K, P, \{b\})\) has an \(\text{ALCQ}\) solution;
2. there exists a model \(I\) of \(K\) such that for all \(a \in P\), \(A, a \not\rightarrow I, b\).

For the adaptation of Theorems 3 and 4 to DLs without inverse roles. Let \(K = (T, A)\) be an \(\text{ALCQI}\) KB and \(a \in \text{ind}(A)\). We use \(A^\dagger\) to denote the set of assertions in \(A\) that use only individual names from the set \(\text{reach}_A(a)\) of individual names reachable from \(a\) in \(G_A\). Now \(A, a \rightarrow I, b\) means that there is a homomorphism \(h\) from \(A^\dagger\) to \(I\) with \(h(a) = b\) such that the extended KB \(K_{I,h}\), defined as

\((T, A \cup \{C(e) \mid C \in \text{sub}(K), e \in \text{reach}_A(a), h(e) \in C\})\),

is satisfiable. We define \(A, a \not\rightarrow I, b\) analogously, additionally requiring that \(h\) is locally injective for all role names \(r\) (but not necessarily for inverse roles).

Theorem 5 Let \((K, P, \{b\})\) be an \(\text{ALC}\) (resp. \(\text{ALCQ}\)) learning instance. Then the following conditions are equivalent:

1. \((K, P, \{b\})\) has an \(\text{ALC}\) (resp. \(\text{ALCQ}\)) solution;
2. there exists a model \(I\) of \(K\) such that for all \(a \in P\), \(A, a \not\rightarrow I, b\) (resp. \(A, a \not\rightarrow I, b\)).

The following example illustrates the reason for why the characterization in Theorem 5 requires \(K_{I,h}\) to be satisfiable.
Example 2 Consider \( A = \{ r(c, a), A(a), A(b) \} \), \( P = \{ a \} \), \( N = \{ b \} \), \( T = \{ T \subseteq \forall r.B \} \), and let \( K = (T, A) \). Then \( K \models B(a) \) and \( K \not\models B(b) \) and so \( B \) is an \( ALC \) solution for \( (K, P, N) \). Let \( b : a \mapsto b \) be a homomorphism from \( A' \) to any model \( I \) of \( K \), but for any model \( I \) of \( K \) with \( b \in (\neg B)^I \) the extended KB \( K_{I,b} \) is not satisfiable.

We observe a slightly surprising consequence of the above characterizations. Concept separability is \( TBox \) anti-monotone in a DL \( \mathcal{L} \) if extending the TBox cannot result in a solution for a learning instance to become available, that is, for all \( \mathcal{L} \) learning instances \( (K_1, P, N) \) and \( (K_2, P, N) \) with \( T_1 \subseteq T_2 \) and \( A_1 = A_2 \), \( (K_2, P, N) \) has an \( \mathcal{L} \) solution if \( (K_2, P, N) \) has an \( \mathcal{L} \) solution.

Theorem 6 Concept separability is \( TBox \) anti-monotone in \( \mathcal{A}L\mathcal{C}1 \) and \( \mathcal{A}L\mathcal{C}Q\mathcal{I} \), but not in \( \mathcal{E}\mathcal{L}, \mathcal{E}C\mathcal{I}, \mathcal{A}L\mathcal{C}, \), and \( \mathcal{A}L\mathcal{C}Q \).

Note that \( TBox \) anti-monotonicity in \( \mathcal{A}L\mathcal{C}1 \) and \( \mathcal{A}L\mathcal{C}Q\mathcal{I} \) follows directly from Theorems 3 and 4. To show that concept separability is not \( TBox \) anti-monotone in \( \mathcal{A}L\mathcal{C} \) and \( \mathcal{A}L\mathcal{C}QI \) consider \( (K, P, N) \) from Example 2 and let \( K' = (\emptyset, A) \). Then \( (K', P, N) \) has an \( \mathcal{A}L\mathcal{C}QI \) solution (and thus no \( \mathcal{A}L\mathcal{C} \) solution), but \( (K, P, N) \) has an \( \mathcal{A}L\mathcal{C}Q \) solution. Example 1 (1) and (2) above shows that concept separability is not \( TBox \) anti-monotone in \( \mathcal{E}L \) and \( \mathcal{E}C\mathcal{I} \).

5 Complexity: Expressive DLs

We clarify the complexity of concept separability and concept definability in expressive DLs, which turn out to be \( \text{NE\textsc{X}TIME-complete} \) in \( \mathcal{A}L\mathcal{C}1 \), \( \mathcal{A}L\mathcal{C}1 \), and \( \mathcal{A}L\mathcal{C}Q \), and \( \text{EXP\textsc{TIME}-complete} \) in \( \mathcal{A}L\mathcal{C}Q\mathcal{I} \). In data complexity, they are \( \Sigma^P_2 \)-complete in all four DLs. In both lower and upper bound proofs, we exploit a connection between separability and (appropriate versions of) UCQ evaluation on KBs that is suggested by the characterizations in the previous section.

Theorem 7 \( \mathcal{A}L\mathcal{C}1 \) concept separability and \( \mathcal{A}L\mathcal{C}1 \) concept definability are \( \text{NE\textsc{X}TIME-complete} \).

The upper bound in Theorem 7 is proved by a polynomial time reduction to the complement of rooted UCQ evaluation on \( \mathcal{A}L\mathcal{C}1 \) KBs and the lower bound is proved by a polynomial time reduction from the complement of rooted UCQ evaluation on \( \mathcal{A}L\mathcal{C}1 \) KBs. Both problems are \( \text{CON\textsc{EX}TIME-complete} \) [Lutz, 2008] and rather well known, so we refrain from giving definitions and only mention that a CQ is rooted if it is connected and has at least one answer variable, and that a UCQ is rooted if every CQ in it is.

For the first reduction, let \( (K, P, \{ b \}) \) be an \( \mathcal{A}L\mathcal{C}1 \) learning instance with \( K = (T, A) \). For every \( a \in P \), let \( q_a \) be the maximal connected component \( A_a \) of \( A \) that contains \( a \) viewed as a connected CQ with only answer variable \( a \). Further, let \( q \) be the \( \text{UCQ} \bigvee_{a \in P} q_a \). Then \( (K, P, \{ b \}) \) has a solution iff \( K \not\models q(b) \). In fact, \( K \not\models q(b) \) iff there is a model \( I \) of \( K \) such that \( I \not\models q_a(b) \) for all \( a \in P \), that is, \( A_a, a \not\models T, b \). This is equivalent to \( a, a \not\models T, b \) as required by Point 2 of Theorem 3 since for all other connected components \( A' \) of \( A \), the identity is a homomorphism to \( T \).

For the second reduction, let \( K = (T, A) \) be an \( \mathcal{A}L\mathcal{C}1 \) KB, \( q(x) \) a unary rooted CQ, and \( a \in \text{ind}(A) \). Let \( A_q \) be \( q \) viewed as an ABox where now \( x \) is an individual name, and let \( A' \) be the disjoint union of \( A \) and \( A_q \). Let further \( K' = (T, A') \), \( P = \{ x \} \), and \( N = \{ a \} \). Using similar arguments as above, one can show that \( K \not\models q(a) \) iff \( (K', P, N) \) has a solution.

In the case of \( \mathcal{A}L\mathcal{C}QI \), separability closely corresponds to a version of unary rooted (U)CQ evaluation that is based on locally injective homomorphisms from the CQ to models of the KB. We introduce this problem in the appendix and prove that, remarkably, it is only \( \text{EXP\textsc{TIME}-complete} \). The intuitive reason is that the \( \text{CON\textsc{EX}TIME}\)-lower bound for CQ evaluation on \( \mathcal{A}L\mathcal{C}1 \) KBs requires to ‘fold’ the CQ in exponentially many different ways, which can only be done with locally non-injective homomorphisms. In \( \mathcal{A}L\mathcal{C}QI \), concept separability is thus no harder than standard reasoning problems such as concept satisfiability w.r.t. a TBox—these are \( \text{EXP\textsc{TIME}-complete} \) for all expressive DLs considered in this paper [Baader et al., 2017].

Theorem 8 \( \mathcal{A}L\mathcal{C}QI \) concept separability and \( \mathcal{A}L\mathcal{C}QI \) concept definability are \( \text{EXP\textsc{TIME}-complete} \).

For the case without inverse roles, we also introduce corresponding versions of (U)CQ evaluation, based on the kinds of homomorphism used in Theorem 5. These turn out to be \( \text{CON\textsc{EX}TIME-complete} \). The lower bound is proved by a reduction of a suitable version of the tiling problem and crucially exploits the satisfiability condition used in the homomorphisms in Theorem 5. We work with a single positive example in the case of \( \mathcal{A}L\mathcal{C} \) while in \( \mathcal{A}L\mathcal{C}Q \) the number of positive examples is not bounded by a constant.

Theorem 9 For \( \mathcal{L} \in \{ \mathcal{A}L\mathcal{C}, \mathcal{A}L\mathcal{C}Q \} \), \( \mathcal{L} \) concept separability and \( \mathcal{L} \) concept definability are \( \text{N\textsc{EX}TIME-complete} \).

We close this section with clarifying data complexity.

Theorem 10 For \( \mathcal{L} \in \{ \mathcal{A}L\mathcal{C}, \mathcal{A}L\mathcal{C}Q, \mathcal{A}L\mathcal{C}1, \mathcal{A}L\mathcal{C}Q1 \} \), \( \mathcal{L} \) concept separability and \( \mathcal{L} \) concept definability are \( \Sigma^P_2 \)-complete in data complexity.

The upper bounds are proved by a ‘guess-coguess and check’ style procedure. The lower bounds are proved by reduction of the problem to decide whether for a given undirected graph \( G \) and \( k \geq 1 \), there is a 2-coloring of \( G \) that does not generate a monochromatic \( k \)-clique [Rutenburg, 1986]. The proof requires two positive examples.

6 Complexity: Horn DLs

We study concept separability and concept definability in the case that the separation language is one of the Horn DLs \( \mathcal{E}L \) and \( \mathcal{E}C\mathcal{I} \). As the TBox language, we consider both Horn DLs and expressive DLs. It turns out that these problems tend to be
undecidable, notable exceptions being the cases that the TBox language is $\mathbf{EL}$ or $\mathbf{ELI}$ and the separation language is $\mathbf{EL}$. We first refine the characterization from Theorem 2. For $L \in \{\mathbf{EL}, \mathbf{ELI}\}$, a model $U$ of a knowledge base $K$ is $L$ universal if $K \models C(a)$ iff $U \models C(a)$, for all $L$ concepts $C$ and $a \in \text{ind}(A)$.

**Theorem 11** Let $(K, P, \{b\})$ be an $\mathbf{ELI}$ learning instance, $\mathcal{L}_S \in \{\mathbf{EL}, \mathbf{ELI}\}$, and $U$ an $\mathcal{L}_S$ universal model for $K$ of finite outdegree. Then the following are equivalent:

1. $(K, P, \{b\})$ has an $\mathcal{L}_S$ solution,
2. $\Pi_{a \in P}(U, a) \not\models_{\mathcal{L}_S} U, b$.

The absence of disjunction in Horn DLs forces the product in Point 2, which captures the commonalities of the positive examples and achieves true generalization, avoiding overfitting. It is, however, also responsible for the high complexity and even undecidability of concept separability in Horn DLs. This is in fact rather surprising given that standard reasoning problems such as subsumption are only PTIME-complete in $\mathbf{EL}$ and EXPTIME-complete in $\mathbf{ELI}$.

We start with the separation language $\mathbf{EL}$, considering both $\mathbf{EL}$ and $\mathbf{ELI}$ as the TBox language.

**Theorem 12** Let $\mathcal{L}_T \in \{\mathbf{EL}, \mathbf{ELI}\}$. Then $(\mathcal{L}_T, \mathbf{EL})$ concept separability and definability are EXPTIME-complete, both in combined complexity and in data complexity.

For the lower bound, we adapt a proof used in [Harel et al., 2002] to prove that the simulation problem between $\mathcal{I}, d$ and $\mathcal{J}, e$ is EXPTIME-hard when $\mathcal{I}$ is a synchronized product of given interpretations and $\mathcal{J}$ is an explicitly given interpretation. Remarkably, the lower bound holds already for empty TBoxes. The number of positive examples is not bounded by a constant. For the upper bound, we build on Theorem 2 of Theorem 11, constructing the product of (exponential size) $\mathbf{EL}$ universal models of $\mathbf{ELI}$ knowledge bases and then deciding $\mathbf{EL}$ similarity in polynomial time. The upper bound easily extends to the case where the learned concept can use only symbols from a given signature.

Surprisingly, separability and definability even become undecidable when we extend $\mathbf{EL}$ with inverse roles. This implies that no learning algorithm can be complete and terminating.

**Theorem 13** $\mathbf{ELI}$ concept separability and definability are undecidable.

The proof is by a rather subtle reduction of the tiling problem of rectangles of unbounded size, borrowing and extending some ideas from the proof given in [Botoeva et al., 2019] that the CQ entailment problem between $\mathbf{ALC}$ KBs is undecidable. It requires only two positive examples.

We remark that $\mathbf{ELI}$ concept separability is closely related to the problem of learning conjunctive queries in the context of $\mathbf{ELI}$ KBs (‘query by example’). That problem was stated to be 2EXPTIME-complete in [Gutiérrez-Basulto et al., 2018], but the proof of the upper bound turns out to be incorrect. In the appendix, we use a minor variation of our proof of Theorem 13 to show that the problem is actually undecidable. Another related problem is the existence of most specific concepts (MSCs) and least common subsumers (LCSs) [Baader, 2003; Lutz et al., 2010; Zarrieß and Turhan, 2013]. In fact, if $(K, P, \{b\})$ is a learning instance, MSCs $M_a$ exist of all $a \in P$, and a LCS $C$ of the $M_a$ exists, then $(K, P, \{b\})$ has a solution if $K \not\models C(b)$. We conjecture that a variation of the proof of Theorem 13 can be used to show that in $\mathbf{ELI}$, the existence of the LCS is undecidable in the presence of a TBox.

We now consider the case where the TBox is formulated in an expressive DL and the separation language is a Horn DL. For example, such a setup is natural if avoiding overfitting is a concern. Also here, separability turns out to be undecidable.

**Theorem 14** Let $\mathcal{L}_T \in \{\mathbf{ALC}, \mathbf{ALCQ}, \mathbf{ALCQI}, \mathbf{ALCQIg}\}$ and $\mathcal{L}_S \in \{\mathbf{EL}, \mathbf{ELI}\}$. Then $(\mathcal{L}_T, \mathcal{L}_S)$ concept separability and $(\mathcal{L}_T, \mathcal{L}_S)$ concept definability are undecidable.

The proof is by reduction from the already mentioned CQ entailment problem between $\mathbf{ALC}$ KBs which undecidable already for (directed or undirected) tree-shaped CQs.

### 7 First Observations on Strong Separability

There is an alternative notion of concept learning that has been proposed in the literature [Lehmann et al., 2014; Badea and Nienhuys-Cheng, 2000; Fanizzi et al., 2018]. An $\mathcal{L}$ concept $C$ is a strong $L$ solution to a learning instance $(K, P, N)$ if $K \models C(a)$ for all $a \in P$ and $K \not\models C(a)$ for all $a \in N$.

We then define strong $L$ concept separability in the expected way. While we leave a thorough investigation for future work, we observe the following characterization.

**Theorem 15** For $\mathcal{L}_S \in \{\mathbf{ALC}, \mathbf{ALCIT}, \mathbf{ALCQ}, \mathbf{ALCQIT}\}$, an $\mathbf{ALCQI}$ learning instance $(K, P, N)$ has a strong $\mathcal{L}_S$ solution iff for all models $\mathcal{I}$ and $\mathcal{J}$ of $K$, all $a \in P$ and all $b \in N$, $\mathcal{I}, a \not\in \mathcal{L}_S \mathcal{J}$, $b$.

Based on this result, one can show that for $\mathbf{ALC}$ and $\mathbf{ALCIT}$ there is a tight link between strong concept separability and KB satisfiability (rather than to UCQ-evaluation on KBs) which can be used to prove the following.

**Theorem 16** For $\mathcal{L} \in \{\mathbf{ALC}, \mathbf{ALCIT}\}$, strong $\mathcal{L}$ concept separability is EXPTIME-complete in combined complexity and CONP-complete in data complexity.

Note the drop in complexity compared to the non-strong version of concept separability, from NEXP-TIME to EXP-TIME and from $\Sigma^P_2$ to CONP. For $\mathbf{ALCQ}$ and $\mathbf{ALCQIT}$, the complexity of strong $\mathcal{L}$ concept separability remains open.

### 8 Discussion

This paper provides characterizations and complexity results for concept separability in many important DLs. It would be interesting to further add other popular DL features such as role hierarchies, transitive roles, and nominals, and it might also be relevant to consider cases of $(\mathcal{L}_T, \mathcal{L}_S)$ concept separability in which $\mathcal{L}_T$ is less expressive than $\mathcal{L}_S$. We also plan to investigate strong separability and in how the characterizations given in this paper can be used or extended to analyse the version space beyond emptiness and whether they can also be used to analyse or craft refinement operators.

**Acknowledgements**

Lutz was supported by DFG CRC 1320 Ease.
References


