Rough $\mathcal{EL}$ Classification

Rafael Peñaloza[1]* and Tingting Zou[2]

[1] Theoretical Computer Science, TU Dresden
Center for Advancing Electronics Dresden, Germany
`penaloza@tcs.inf.tu-dresden.de`
[2] College of Computer Science and Technology, Jilin University, China
`zoutingt@163.com`

Abstract. Rough Description Logics (DLs) have been studied as a means for representing and reasoning with imprecise knowledge. It has been shown that reasoning in rough DLs can be reduced to reasoning in a classical DL that allows value restrictions, and transitive and inverse roles. This shows that for propositionally closed DLs, the complexity of reasoning is not increased by the inclusion of rough constructors. However, applying such a reduction to rough $\mathcal{EL}$ yields an exponential time upper bound. We show that this blow-up in complexity can be avoided, providing a polynomial-time completion-based algorithm for classifying rough $\mathcal{EL}$ ontologies.

1 Introduction

Description Logics (DLs) [3] are a family of knowledge representation formalisms designed for expressing terminological knowledge in an unambiguous and well-understood manner. They have been successfully applied to model and reason with real-world knowledge domains, but possibly its biggest success so far is the designation of the DL-based language OWL as the standard ontology language for the semantic web, by the W3C.[3]

The DL $\mathcal{EL}$ is a light-weight logic that allows only for conjunction and existential restrictions as constructors. As it cannot express negations, $\mathcal{EL}$ is not propositionally closed. Despite its low expressivity, this logic and small extensions of it have been successfully used for representing knowledge from several domains, most prominently from the medical and biological fields. In fact, $\mathcal{EL}$ is the basic logic underlying large-scale ontologies like Snomed CT[4] or the Gene Ontology.[5] A prominent feature of this logic is its polynomial-time complexity of reasoning, which enables effective reasoning procedures. In fact, Snomed CT, which has approximately 300,000 axioms, can be classified by modern reasoners in less than seven seconds [15].

* Partly supported by DFG within the Cluster of Excellence 'cfAED'.
[3] `http://www.w3.org/TR/owl2-overview/`
[4] `http://www.ihtsdo.org/snomed-ct/`
[5] `http://www.geneontology.org`

In its classical form $\mathcal{EL}$, as all other classical DLs, lacks the capacity of modelling and reasoning with imprecise knowledge. This is in no way a small drawback, as imprecision is almost unavoidable in several knowledge domains, like those from the bio-medical fields. For example, even the notion of *species*, one of the mayor taxonomic ranks from biological classification is far from precise, or even being well-understood. Consider for instance the case of the *Ensatina* salamanders from North America. When seen independently, the Monterey Ensatina and the Large Blotched Ensatina form two different species, with their own characteristic traits; they can be easily distinguished as the former is completely brown in color, while the latter is black with large yellow blotches. However, there also exists a group of *intermediate* individuals, that mix the traits of both species, forming a gradual bridge between them; e.g., dark brown with lighter-brown blotches. It is not clear at which point these intermediate individuals stop being members of one species and start belonging to the other. Indeed, providing a satisfactory notion of when two individuals belong to the same species is a prominent open problem in biology [12].

The best-known approach for handling imprecision formally is through fuzzy logic [13]. Fuzzy extensions of DLs have been thoroughly studied during the last decade as a formalism for representing vague terminological knowledge [18,22]. However, it was recently shown that reasoning in expressive fuzzy DLs is either undecidable [6,10], or crisp, with truth degrees playing no role whatsoever [5]. Even for the inexpressive DL $\mathcal{EL}$, the extension to general fuzzy set based semantics usually yields intractable reasoning problems [7,9]. It can be argued that these negative complexity results arise from the high level of granularity provided by fuzzy semantics, where every number from the interval $[0, 1]$ can be used as a truth degree. In other words, it is possible to make arbitrarily small distinctions between elements of the domain. One can partially alleviate this problem by restricting to finitely many truth degrees [4,8]. In this case, the resources needed for reasoning are directly correlated with the size of the truth value space. This idea, however, adds the burden of deciding *a priori* the amount of degrees that will be needed and their relevant operations. It is thus desirable to obtain an intermediate formalism that allows for imprecise limitations of concepts, while avoiding the level of detail of fuzzy logics.

Rough sets were introduced in [19] as an alternative to fuzzy set theory [25] for dealing with imprecise notions. The main idea behind this formalism is to describe imprecise sets by allowing a class of *boundary* elements that can neither be stated to belong, nor to be outside, the set. More precisely, a set X without a clear distinction on its limits, is approximated using a set $\underline{X}$ of elements that are guaranteed to belong to X, and a set $\overline{X}$ of elements that might be members of X; this latter set is called the upper approximation of X. These sets are formally defined with the help of an *indiscernibility relation* that clusters together individuals sharing the same properties. The difference $\overline{X} \setminus \underline{X}$ is the set of boundary elements, which cannot be ensured to belong to X, nor to its complement.

For example, the problem with the different species of Ensatina salamanders can be solved by stating that the intermediate individuals belong to the upper bounds of the sets of both species. This representation allows us to state properties of the intermediate individuals (e.g. that they have mixed traits from the border species) without providing a clear-cut division of these individuals into the two species.[6]

Although the combination of rough set theory with DLs is far from new (see e.g. [17] for some early work), interest in it has grown in the last few years [11,14,16,21]. Most of the work in this direction so far focuses on rough extensions of expressive DLs. The approach is to extend a description logic with two new constructors that describe the upper and lower approximations of concepts. The semantics of these constructors are based on equivalence relations that provide the indiscernibility relation from rough set theory. In [21] it was shown that these constructors can be modelled in classical DLs with the help of existential and value restrictions over a new transitive, symmetric and reflexive role ρ. Briefly, the role ρ describes the indiscernibility relation, and the value and existential restrictions can be used to describe the lower and upper approximations, respectively. This construction is useful for showing that the rough constructors do not increase the complexity of standard reasoning for sufficiently expressive DLs.

The reduction from [21], when applied to rough $\mathcal{EL}$, requires to extend the set of constructors to include value restrictions and inverse roles, among others. The extensions of $\mathcal{EL}$ with any of these constructors are known to be ExpTime-complete [1,2]. Thus, this approach yields an exponential-time upper bound for reasoning in rough $\mathcal{EL}$, in contrast to the polynomial-time complexity for classical $\mathcal{EL}$. In this paper we show that subsumption in rough $\mathcal{EL}$ is in fact PTime-complete, matching the known complexity for its classical logic.

The paper is divided as follows. We first provide a very brief introduction to the theory of rough sets, which will be useful for defining the syntax and semantics of rough $\mathcal{EL}$ in Section 3, where we also prove some basic properties of this logic. In Section 4, we describe a completion-based algorithm for deciding subsumption of rough $\mathcal{EL}$ concepts. As an added benefit, we obtain that *classifying* the full ontology needs only polynomial time.

2 Rough Sets

Rough sets were introduced in [19] as an alternative to fuzzy set theory [25] for dealing with imprecise notions. The main motivation in this formalism is to be able to *approximate* terms that defy a precise characterisation, with the help of an equivalence relation $\sim$, called the *indiscernibility relation*. Formally, the equivalence relation $\sim$ divides the universe into its equivalence classes, which form clusters, or *granules* of indiscernible elements. As usual, we will denote the equivalence class of an element x w.r.t. the relation $\sim$ by $[x]_\sim$. Intuitively, elements belonging to the same equivalence class cannot be distinguished through their perceivable characteristics, and hence cannot be divided by any given set.

[6] This description is not intended to provide a general solution of the species-problem.

Rough sets are also sometimes called granular sets in the literature and are one of the bases for granular computing [24].

Given a set X and an equivalence relation $\sim$, we can define its *best lower approximation*, denoted by $\underline{X}$, as the greatest union of equivalence classes contained in X; i.e., $\underline{X} := \bigcup_{[x]_\sim \subseteq X} [x]_\sim$. Likewise, its *best upper approximation* is the union of the equivalence classes of all elements of X; $\overline{X} := \bigcup_{x \in X} [x]_\sim$. Equivalently, we have

$$\underline{X} = \{x \mid [x]_\sim \subseteq X\}, \qquad \overline{X} = \{x \mid [x]_\sim \cap X \neq \emptyset\}.$$

The elements in $\underline{X}$ are those that can be clearly distinguished from any element not belonging to X, and hence are said to *surely* belong to X. The members of $\overline{X}$, on the other hand, are those indistinguishable from some element of X, and said to *possibly* belong to X. The elements in the *boundary* $\overline{X} \setminus \underline{X}$ of X are those for which the notion of *belonging to* X cannot be made precise, as they are indistinguishable from both, some members of X and some members of the complement of X.

From an informal point of view, it is possible to see rough sets as a three-valued membership function, where members of $\underline{X}$ *strongly belong* to X, the boundary elements *weakly belong* to X, and those in the complement of $\overline{X}$ do not belong to X. However, this description is overly simplistic, as the three-valued semantics are incapable of fully characterising the properties of the indiscernibility relation. In particular, the desired properties relating a three-valued conjunction with its three-valued implication are not satisfied by rough set conjunction and inclusion.

In the next section, we describe the combination of the description logic $\mathcal{EL}$ with the lower and upper-approximation constructors, whose semantics is based on rough sets. Afterwards, we describe a completion algorithm for deciding (classical) subsumption between rough $\mathcal{EL}$ concepts.

3 Rough $\mathcal{EL}$

The logic rough $\mathcal{EL}$ extends classical $\mathcal{EL}$ by allowing the *lower approximation* and *upper approximation* constructors $\bar{\cdot}$ and $\underline{\cdot}$ for expressing rough concepts. Formally, from two disjoint sets $\mathsf{N_C}$ and $\mathsf{N_R}$ of *concept* and *role* names, *rough $\mathcal{EL}$ concepts* are constructed using the following syntactic rule:

$$C \quad ::= \quad A \quad | \quad C_1 \sqcap C_2 \quad | \quad \exists r.C \quad | \quad \overline{C} \quad | \quad \underline{C} \quad | \quad \top,$$

where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$.

The semantics of this logic is based on interpretations that map concept names to subsets of a non-empty domain Δ, and role names to binary relations over Δ. To handle the rough concept constructors, these interpretations additionally have an indiscernibility (equivalence) relation.

Definition 1. *A* rough interpretation *is a tuple $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I}, \sim_\mathcal{I})$, where $\Delta^\mathcal{I}$ is a non-empty set called the* domain*, $\sim_\mathcal{I}$ is an equivalence relation on $\Delta^\mathcal{I}$, called* *the* indiscernibility relation*, and* $\cdot^{\mathcal{I}}$ *is the* interpretation function *mapping every concept name* A *to a subset* $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$*, and every role name* r *to a binary relation* $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$*.*

The interpretation function $\cdot^{\mathcal{I}}$ is extended to general rough $\mathcal{EL}$ concepts by setting:

- $(C_1 \sqcap C_2)^{\mathcal{I}} = C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$,
- $(\exists r.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}}.\ (x, y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$,
- $\overline{C}^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid [x]_{\sim_{\mathcal{I}}} \cap C^{\mathcal{I}} \neq \emptyset\}$,
- $\underline{C}^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid [x]_{\sim_{\mathcal{I}}} \subseteq C^{\mathcal{I}}\}$, and
- $\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$.

Intuitively, the indiscernibility relation groups the individuals of the domain that cannot be distinguished from each other, at the considered level of detail. The upper approximation $\overline{C}$ of a given concept C describes those individuals that cannot be excluded from belonging to C, as they are indistinguishable from some element belonging to this concept. Dually, the individuals $\underline{C}$ are those that are discernible from every individual *not* belonging to C. Clearly, for every interpretation $\mathcal{I}$ and concept C it holds that $\underline{C}^{\mathcal{I}} \subseteq \overline{C}^{\mathcal{I}}$. The borderline cases, those elements of $\overline{C}^{\mathcal{I}} \setminus \underline{C}^{\mathcal{I}}$, cannot be ensured to be, nor excluded from being instances of C.

The domain knowledge is described using a *TBox*: a finite set of GCIs of the form $C \sqsubseteq D$, where C, D are rough $\mathcal{EL}$ concepts. The interpretation $\mathcal{I}$ *satisfies* the GCI $C \sqsubseteq D$ if and only if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ holds. $\mathcal{I}$ is a *model* of the TBox $\mathcal{T}$ if it satisfies all the GCIs in $\mathcal{T}$.

As is the case for classical $\mathcal{EL}$, every rough $\mathcal{EL}$ TBox is consistent. Indeed, the interpretation $\mathcal{I} = (\{x\}, \cdot^{\mathcal{I}}, \{(x, x)\})$, where $A^{\mathcal{I}} = \{x\}$ and $r^{\mathcal{I}} = \{(x, x)\}$ for all $A \in \mathsf{N_C}$ and all $r \in \mathsf{N_R}$ satisfies every GCI, and hence is a model of every TBox. For that reason, we focus on the problem of deciding *subsumption* between concept names, and the more general problem of *classifying* the TBox.

Definition 2. *Let* $\mathcal{T}$ *be a TBox and* C, D *two rough* $\mathcal{EL}$ *concepts. We say that* C *is* subsumed *by* D *w.r.t.* $\mathcal{T}$*, denoted by* $C \sqsubseteq_{\mathcal{T}} D$*, if for every model* $\mathcal{I}$ *of* $\mathcal{T}$ *it holds that* $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$*.* Classification *is the problem of deciding, for every pair of concept names* A, B*, whether* $A \sqsubseteq_{\mathcal{T}} B$ *holds or not.*

Example 3. Consider once again the Ensatina salamanders. We can describe the class of *intermediate ensatina salamanders* as those that cannot be excluded from being Monterey Ensatina, nor from being Large Blotched Ensatina, by the axiom

$$\mathsf{IntermediateE} \sqsubseteq \overline{\mathsf{MontereyE}} \sqcap \overline{\mathsf{LargeBlotchedE}}.$$

We can also state that Large Blotched Ensatinas can be recognized by the blotches on their skin.

$$\mathsf{LargeBlotchedE} \sqsubseteq \exists \mathsf{hasFeature}.\mathsf{Blotches}.$$

From these two axioms, it is possible to deduce that intermediate ensatina salamanders are indistinguishable from some salamanders with skin blotches; i.e.,

$$\mathsf{IntermediateE} \sqsubseteq_{\mathcal{T}} \overline{\exists \mathsf{hasFeature.Blotches}}.$$

This conclusion gives us some information about the intermediate salamanders, despite being unable to clearly state whether they are Large Blotched Ensatinas or not.

As shown in [21], reasoning in rough DLs can be reduced to reasoning in a classical DL that allows value restrictions, inverse, and reflexive roles, and role inclusion axioms. Let ρ be a new role that does not appear in $\mathcal{T}$. If we restrict ρ to be reflexive, and include the role inclusion axioms $\rho \circ \rho \sqsubseteq \rho$ (transitivity), and $\rho^{-1} \sqsubseteq \rho$ (symmetry), then the concepts $\overline{C}$ and $\underline{C}$ are equivalent to the concepts $\exists\rho.C$ and $\forall\rho.C$, respectively (see [21] for full details). However, it is well known that extensions of classical $\mathcal{EL}$ with either value restrictions or inverse roles are already intractable; in fact reasoning in these extensions is ExpTime-hard [1,2,23]. Applying this reduction directly yields an ExpTime upper bound for the complexity of deciding subsumption of rough $\mathcal{EL}$ concepts. On the other hand, only one role name, namely ρ, is used in any of the possibly expensive constructors introduced by this reduction. As we will see in the following section, this limited use does help in improving the complexity, as the problem of deciding subsumption between rough $\mathcal{EL}$ concepts is in fact decidable in polynomial time.

Clearly, the subsumption relation $\sqsubseteq_{\mathcal{T}}$ is transitive; that is, if $C \sqsubseteq_{\mathcal{T}} D$ and $D \sqsubseteq_{\mathcal{T}} E$, then also $C \sqsubseteq_{\mathcal{T}} E$ holds. Due to the properties of lower and upper approximations, some additional subsumption relations, which are not necessarily obvious at first sight, can sometimes be deduced, as shown next.

Theorem 4. *Let C, D, E be rough $\mathcal{EL}$ concepts. The following properties hold:*

1. $\overline{C} \sqsubseteq_{\mathcal{T}} D$ *iff* $C \sqsubseteq_{\mathcal{T}} \underline{D}$
2. *if* $C \sqsubseteq_{\mathcal{T}} \overline{D}$ *and* $D \sqsubseteq_{\mathcal{T}} \underline{E}$, *then* $C \sqsubseteq_{\mathcal{T}} \underline{E}$
3. *if* $C \sqsubseteq_{\mathcal{T}} \underline{D}$ *and* $\underline{D} \sqsubseteq_{\mathcal{T}} E$, *then* $C \sqsubseteq_{\mathcal{T}} \underline{E}$

Proof. Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \sim_{\mathcal{I}})$ be a model of $\mathcal{T}$, and $x \in \Delta^{\mathcal{I}}$.

1. ($\Leftarrow$) If $x \in \overline{C}^{\mathcal{I}}$, then there exists a $y \in [x]_{\sim_{\mathcal{I}}} \cap C^{\mathcal{I}}$. By assumption, $y \in \underline{D}^{\mathcal{I}}$. Thus, $x \in [y]_{\sim_{\mathcal{I}}} \subseteq D^{\mathcal{I}}$.

   ($\Rightarrow$) Let $x \in C^{\mathcal{I}}$. We must prove that $[x]_{\sim_{\mathcal{I}}} \subseteq D^{\mathcal{I}}$. Let $y \sim_{\mathcal{I}} x$. Then, $y \in \overline{C}^{\mathcal{I}}$, and thus, by assumption, $y \in D^{\mathcal{I}}$.
2. Let $x \in C^{\mathcal{I}}$. By assumption, we know that there exists $z \in [x]_{\sim_{\mathcal{I}}} \cap D^{\mathcal{I}}$, and thus $z \in \underline{E}^{\mathcal{I}}$; i.e., $[x]_{\sim_{\mathcal{I}}} = [z]_{\sim_{\mathcal{I}}} \subseteq E^{\mathcal{I}}$. Hence $x \in \underline{E}^{\mathcal{I}}$.
3. If $x \in C^{\mathcal{I}}$, then by assumption it holds that $[x]_{\sim_{\mathcal{I}}} \subseteq D^{\mathcal{I}}$. Let $y \sim_{\mathcal{I}} x$. Then $[y]_{\sim_{\mathcal{I}}} = [x]_{\sim_{\mathcal{I}}} \subseteq D^{\mathcal{I}}$, and hence $y \in \underline{D}^{\mathcal{I}}$, and by assumption $y \in E^{\mathcal{I}}$. □

In the following section we will exploit these properties to build a completion-based algorithm that classifies a TBox and can be used to decide which subsumption relations hold.

NF1	$A \sqcap C \sqsubseteq E$	$\longrightarrow$	$\{C \sqsubseteq X, A \sqcap X \sqsubseteq E\}$
NF2	$\exists r.C \sqsubseteq E$	$\longrightarrow$	$\{C \sqsubseteq X, \exists r.X \sqsubseteq E\}$
NF3	$\underline{C} \sqsubseteq E$	$\longrightarrow$	$\{C \sqsubseteq X, \underline{X} \sqsubseteq E\}$
NF4	$\overline{C} \sqsubseteq E$	$\longrightarrow$	$\{C \sqsubseteq \underline{E}\}$
NF5	$C \sqsubseteq D$	$\longrightarrow$	$\{C \sqsubseteq X, X \sqsubseteq D\}$
NF6	$A \sqsubseteq E \sqcap F$	$\longrightarrow$	$\{A \sqsubseteq E, A \sqsubseteq F\}$
NF7	$A \sqsubseteq \exists r.C$	$\longrightarrow$	$\{A \sqsubseteq \exists r.X, X \sqsubseteq C\}$
NF8	$A \sqsubseteq \underline{C}$	$\longrightarrow$	$\{A \sqsubseteq \underline{X}, X \sqsubseteq C\}$
NF9	$A \sqsubseteq \overline{C}$	$\longrightarrow$	$\{A \sqsubseteq \overline{X}, X \sqsubseteq C\}$

Table 1. Normalisation rules, where $A \in \mathsf{BC}, C, D \notin \mathsf{BC}$ and X is a new concept name

4 A Completion Algorithm

In this section we describe an algorithm for deciding subsumption relations between concepts. To simplify the description, we will focus solely on subsumption between concept *names*. Notice that subsumption between complex rough $\mathcal{EL}$ concepts C, D can be reduced to this problem by adding the two axioms $A \sqsubseteq C$ and $D \sqsubseteq B$, where A, B are two new concept names, to $\mathcal{T}$ and then deciding whether $A \sqsubseteq_{\mathcal{T}} B$ holds. Thus, restricting to concept name subsumption results in no loss of generality (see [1] for details).

As a preprocessing step for the algorithm, we transform the TBox into an adequate normal form. We define the set of *basic concepts* as $\mathsf{BC} := \mathsf{N_C} \cup \{\top\}$; i.e., all concept names and the top concept are basic concepts. The TBox $\mathcal{T}$ is in *normal form*, if all its axioms are of one of the following forms:

$$A \sqsubseteq \exists r.B, \quad \exists r.A \sqsubseteq B, \quad A \sqcap A' \sqsubseteq B, \quad \underline{A} \sqsubseteq B, \quad A \sqsubseteq \underline{B}, \quad \text{or} \quad A \sqsubseteq \overline{B},$$ [7]

where $A, A', B \in \mathsf{BC}$ and $r \in \mathsf{N_R}$. The normalisation rules shown in Table 1 can be used to transform any TBox $\mathcal{T}$ into a TBox in normal form that preserves all the subsumption relations from $\mathcal{T}$. It is possible to show that these normalisation rules yield a normalised TBox in linear time. Notice in particular rule **NF4**, which takes advantage of the first property described in Theorem 4.

Our completion algorithm extends the methods described in [1] to appropriately handle the lower and upper approximations of concepts. The idea is to store the information of the subsumption relations using a collection of *completion sets*. The main difference with the classical approach is that we need to maintain special completion sets for the lower and upper approximations, in order to handle the special properties of these constructors.

The algorithm uses as data structure a family of completion sets. For each basic concept A appearing in the TBox $\mathcal{T}$, we store three completion sets $S(A), \underline{S}(A)$, and $\overline{S}(A)$, and additionally a completion set $S(A, r)$ for every role name r appearing in $\mathcal{T}$. The members of the completion sets are all basic concepts. These

[7] To simplify the description, we use the expression $\top \sqcap A \sqsubseteq B$ to represent axioms of the form $A \sqsubseteq B$.

CR1	if $B_1 \in S(A), B_2 \in S(A)$, and $B_1 \sqcap B_2 \sqsubseteq C \in \mathcal{T}$, then add C to $S(A)$
CR2	if $B \in S(A)$ and $B \sqsubseteq \exists r.C \in \mathcal{T}$, then add C to $S(A, r)$
CR3	if $B \in S(A, r), C \in S(B)$, and $\exists r.C \sqsubseteq D \in \mathcal{T}$, then add D to $S(A)$
CR4	if $B_1 \in \underline{S}(A), B_2 \in \underline{S}(A)$, and $B_1 \sqcap B_2 \sqsubseteq C \in \mathcal{T}$, then add C to $\underline{S}(A)$
CR5	if $B_1 \in \underline{S}(A), B_2 \in \overline{S}(A)$, and $B_1 \sqcap B_2 \sqsubseteq C \in \mathcal{T}$, then add C to $\overline{S}(A)$
CR6	if $B \in \underline{S}(A)$ and $\underline{B} \sqsubseteq C \in \mathcal{T}$, then add C to $\underline{S}(A)$
CR7	if $B \in \overline{S}(A)$, and $B \sqsubseteq \underline{C} \in \mathcal{T}$, then add C to $\underline{S}(A)$
CR8	if $B \in \overline{S}(A)$, and $B \sqsubseteq \overline{C} \in \mathcal{T}$, then add C to $\overline{S}(A)$
CR9	if $B \in \underline{S}(A)$ then add B to $S(A)$
CR10	if $B \in S(A)$ then add B to $\overline{S}(A)$
CR11	if $B \in \underline{S}(A)$ and $C \in S(B)$ then add C to $\underline{S}(A)$
CR12	if $B \in \overline{S}(A)$ and $C \in \overline{S}(B)$ then add C to $\overline{S}(A)$
CR13	if $B \in \overline{S}(A)$ and $C \in \underline{S}(B)$ then add C to $\underline{S}(A)$

Table 2. Completion rules for rough $\mathcal{EL}$

sets will maintain the following four invariants during the whole execution of the algorithm:

I1 if $B \in S(A)$, then $A \sqsubseteq_{\mathcal{T}} B$
I2 if $B \in \overline{S}(A)$, then $A \sqsubseteq_{\mathcal{T}} \overline{B}$
I3 if $B \in \underline{S}(A)$, then $A \sqsubseteq_{\mathcal{T}} \underline{B}$
I4 if $B \in S(A, r)$, then $A \sqsubseteq_{\mathcal{T}} \exists r.B$.

The completion sets are initialised as

$$S(A) = \overline{S}(A) := \{A, \top\}, \quad \underline{S}(A) := \{\top\}, \quad S(A, r) := \emptyset$$

for basic concepts A and role names r. Obviously, this initialisation preserves the invariants described above. The completion rules from Table 2 are then applied to extend these sets. To ensure termination, a rule is only applied if it adds new information; that is, if the basic concepts to be added to the completion sets by such rule application are not already in them. These rules are applied until the completion sets are *saturated*; i.e., until no rule is applicable anymore. We first show that this procedure terminates in polynomial time.

Lemma 5. *The rules from Table 2 can only be applied a polynomial number of times, and each rule application needs polynomial time.*

Proof. Each of the completion sets contains only concept names that appear in $\mathcal{T}$ and (possibly) $\top$. Thus, the size of each of these sets is linear on $\mathcal{T}$. For each concept name in $\mathcal{T}$ there are three such completion sets, plus one additional completion set for each role name. Thus, the number of completion sets is quadratic on the size of $\mathcal{T}$. Each rule application adds one concept name to one completion set, and never removes any. This means that there can be at most polynomially many rule applications, before no new concept name can be added to any completion set.

For testing the pre-condition of a rule application, we can simply explore all the completion sets, at most twice, and the set of axioms $\mathcal{T}$. This exploration needs in total polynomial time. □

When the algorithm terminates, we can read all the subsumption relations between concept names appearing in the TBox $\mathcal{T}$, by simply considering the elements appearing in the subsumption sets. More precisely, the subsumption relation $A \sqsubseteq_{\mathcal{T}} B$ holds iff $B \in S(A)$. We show first that the method is sound, by showing that rule applications preserve the invariants I1 to I4 described before.

Lemma 6. *The invariants* I1 *to* I4 *are preserved through all rule applications.*

Proof. As said before, the invariants are satisfied by the initialisation of the completion sets. Soundness of the first three rules has been shown in [1]. For the remaining rules, we take advantage of the properties of rough concepts. Recall that for every concept name A, it holds that $\underline{A} \sqsubseteq_{\mathcal{T}} A \sqsubseteq_{\mathcal{T}} \overline{A}$. This shows soundness of the rules CR9 and CR10.

For the rule CR4, let $A \sqsubseteq_{\mathcal{T}} \underline{B_1}$ and $A \sqsubseteq_{\mathcal{T}} \underline{B_2}$. Then for every interpretation $\mathcal{I}$ and every $x \in \mathcal{I}$ if $x \in A^{\mathcal{I}}$, then $[x]_{\sim_{\mathcal{I}}} \subseteq B_1^{\mathcal{I}} \cap B_2^{\mathcal{I}}$. Thus, $[x]_{\sim_{\mathcal{I}}} \subseteq C^{\mathcal{I}}$, which implies that $A \sqsubseteq \underline{C}$. Rule CR5 can be treated analogously.

Soundness of the remaining rules is a direct consequence of Theorem 4. □

It remains only to show completeness; i.e., that once the algorithm has terminated, all the subsumption relations are explicitly stated in the completion sets. As usual, this is shown by building a canonical model that serves as a countermodel for all the subsumption relations between concept names that do not appear in the completion sets. The main idea is to have one individual for each concept name C appearing in $\mathcal{T}$; the interpretation will include this individual in every basic concept D that subsumes C w.r.t. $\mathcal{T}$. However, we need to create additional auxiliary individuals to correctly deal with the upper and lower approximations of each of these concept names. We thus add an element C_u that will be interpreted to belong to all concept names D such that $\underline{D}$ subsumes C. For dealing with the upper approximations, the construction is slightly more complex, as different elements might be needed to witness the existence of an indiscernible element belonging to different concept names. For every concept name D such that $\overline{D}$ subsumes C, we create an element C_D that will belong to D, as well as all (named) subsumers of D. Intuitively, this element C_D will be the witness for C to be a member of $\overline{D}$. Obviously, all the elements of the form C_D will be interpreted to belong to the same equivalence class as C and C_u. We formalize these ideas next.

Lemma 7. *Let A, B be two concept names appearing in $\mathcal{T}$, and $S(A)$ the completion set obtained after the application of the completion rules has terminated. If $B \notin S(A)$, then $A \not\sqsubseteq_{\mathcal{T}} B$.*

Proof. We need to build a model $\mathcal{I}$ of $\mathcal{T}$ such that $A^{\mathcal{I}} \not\subseteq B^{\mathcal{I}}$. We start by defining the domain

$$\Delta^{\mathcal{I}} := \{C, C_u, C_D \mid C, D \text{ are concept names appearing in } \mathcal{T}\}.$$

The equivalence relation $\sim_{\mathcal{I}}$ is the transitive, reflexive and symmetric closure of $\{(C, C_u), (C, C_D) \mid C, D \in \mathsf{N_C} \text{ appear in } \mathcal{T}\}$; thus, the equivalence class of a concept name C is $[C]_{\sim_{\mathcal{I}}} := \{C, C_u, C_D \mid D \in \mathsf{N_C} \text{ appears in } \mathcal{T}\}$. It remains only to define the interpretation function $\cdot^{\mathcal{I}}$. For a concept name C, we set

$$\begin{aligned} C^{\mathcal{I}} := \{D \mid C \in S(D)\} \cup \{D_u \mid C \in \underline{S}(D)\} \cup \\ \{D_X \mid C \in S(X), X \in \overline{S}(D)\} \cup \{D_X \mid C \in \underline{S}(D), X \in \mathsf{N_C}\}, \end{aligned}$$

and for a role name r

$$\begin{aligned} r^{\mathcal{I}} := \{(C, D) \mid D \in S(C, r)\} \cup \{(C_u, D) \mid D \in S(X, r), X \in \underline{S}(C)\} \cup \\ \{(C_X, D) \mid D \in S(X, r), X \in \overline{S}(C)\} \cup \\ \{(C_X, D) \mid D \in S(Y, r), Y \in \underline{S}(C), X \in \mathsf{N_C}\}. \end{aligned}$$

Clearly, for the interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}}, \sim_{\mathcal{I}})$ it holds that $A \in A^{\mathcal{I}}$ but $A \notin B^{\mathcal{I}}$. It only remains to be shown that $\mathcal{I}$ is indeed a model of $\mathcal{T}$. We analyse the different cases, depending on the form of the axiom.

$[\underline{C} \sqsubseteq D]$ Let $x \in \underline{C}^{\mathcal{I}}$, then $[x]_{\sim_{\mathcal{I}}} \subseteq C^{\mathcal{I}}$. Let E be the concept name such that $[E]_{\sim_{\mathcal{I}}} = [x]_{\sim_{\mathcal{I}}}$. Then $E_u \in C^{\mathcal{I}}$. By definition, this means that $C \in \underline{S}(E)$. Since the rule CR6 is not applicable, $D \in \underline{S}(E)$, and by rule CR9, $D \in S(E)$. Let now $E_F \in [E]_{\sim_{\mathcal{I}}}$. Since $D \in \underline{S}(E)$, by definition $E_F \in D^{\mathcal{I}}$. Thus, $x \in [E]_{\sim_{\mathcal{I}}} \subseteq D^{\mathcal{I}}$.

$[C \sqsubseteq \underline{D}]$ Let $x \in C^{\mathcal{I}}$ and $[x]_{\sim_{\mathcal{I}}} = [E]_{\sim_{\mathcal{I}}}$. By definition and the rules CR9, CR10, and CR12, it holds that $C \in \overline{S}(E)$. By rule CR7, it then follows that $D \in \underline{S}(E)$ and hence also $D \in S(E) \cap \overline{S}(E)$. This implies that $[x]_{\sim_{\mathcal{I}}} = [E]_{\sim_{\mathcal{I}}} \subseteq D^{\mathcal{I}}$, and thus $x \in \underline{D}^{\mathcal{I}}$.

$[C \sqsubseteq \overline{D}]$ Let $x \in C^{\mathcal{I}}$ and $[x]_{\sim_{\mathcal{I}}} = [E]_{\sim_{\mathcal{I}}}$. As before, from rules CR9, CR10, and CR12, we derive that $C \in \overline{S}(E)$, and from rule CR8 it follows that $D \in \overline{S}(E)$. Thus, $E_D \in D^{\mathcal{I}}$. Since $E_D \in [x]_{\sim_{\mathcal{I}}}$, it follows that $[x]_{\sim_{\mathcal{I}}} \cap D^{\mathcal{I}} \neq \emptyset$, and hence $x \in \overline{D}$.

The cases for the axioms that do not use rough concepts can be shown analogously, following the arguments from [1], with an additional case analysis that arises from the new elements C_u and C_X, and the concepts they belong to. □

Combining all this lemmata, we obtain the desired results, as stated in the following theorem.

Theorem 8. *Subsumption of rough $\mathcal{EL}$ concept names w.r.t. TBoxes can be decided in polynomial time. Moreover, the TBox $\mathcal{T}$ can be classified in polynomial time.*

5 Conclusions

We have studied rough $\mathcal{EL}$, a description logic that extends the lightweight DL $\mathcal{EL}$ to allow for the lower and upper approximations from rough set theory. We have shown that subsumption of concept names w.r.t. rough $\mathcal{EL}$ TBoxes can be decided in polynomial time. This result was obtained by providing a completion-based algorithm capable of classifying the TBox in polynomial time. As an added benefit, our approach does not require including expensive constructors that damage the efficiency of DL reasoners.

Our algorithm is a direct extension from the one presented in [1] in that, when no rough constructors appear in the TBox, the algorithm behaves similarly. However, the cost of handling potential rough concepts is to double the space needed.[8] This unnecessary cost can be easily avoided by disallowing applications of rules CR4 to CR13 whenever the TBox uses only classical $\mathcal{EL}$ constructors. The additional rules and completion sets needed for our completion algorithm should not impose many problems for a prospective implementation.

These polynomial-time complexity results give strength to the observation from [21] that rough constructors can be added to classical DLs with no additional cost in terms of complexity. In fact, it has been shown in [20], that the polynomial upper bound still holds if the bottom concept, nominals, and role inclusion axioms are also allowed. Except for the absence of concrete domains, these constructors cover the whole DL $\mathcal{EL}^{++}$, the formalism underlying the OWL 2 EL profile of the standard ontology language for the semantic web OWL 2.

We should emphasize that in this paper we have considered only classical subsumption in a rough description logic. There exist other non-standard reasoning services that consider rough concepts in higher detail, as described in [16]. As presented in this paper, our completion algorithm is incapable of solving those reasoning tasks.

As part of our future work, we intend to study the complexity of rough-set-specific reasoning problems for rough $\mathcal{EL}$ and, if possible, extend our completion algorithm to handle them adequately. We also intent to study possible optimizations that would allow for a practical implementation of our approach. Another open question that may be of interest corresponds to extending the logic to allow also for rough role constructors.

References

1. F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. 19th Int. Joint Conf. on Artif. Intel. (IJCAI 2005)*, Edinburgh, UK, 2005. Morgan-Kaufmann Publishers.
2. F. Baader, S. Brandt, and C. Lutz. Pushing the el envelope further. In K. Clark and P. F. Patel-Schneider, editors, *Proceedings of the OWLED 2008 DC Workshop on OWL: Experiences and Directions*, 2008.

[8] Without the lower approximation constructor, the sets $\underline{S}$ are never populated, but due to rule CR10, the sets $\overline{S}$ include all elements in S.

3. F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2nd edition, 2007.
4. F. Bobillo and U. Straccia. Finite fuzzy description logics and crisp representations. In *Uncertainty Reasoning for the Semantic Web II, International Workshops URSW 2008-2010 Held at ISWC and UniDL 2010 Held at FLoC, Revised Selected Papers*, volume 7123 of *Lecture Notes in Computer Science*, pages 99–118. Springer, 2013.
5. S. Borgwardt, F. Distel, and R. Peñaloza. How fuzzy is my fuzzy description logic? In B. Gramlich, D. Miller, and U. Sattler, editors, *Proceedings of the 6th International Joint Conference on Automated Reasoning (IJCAR'12)*, volume 7364 of *Lecture Notes in Artificial Intelligence*, pages 82–96, Manchester, UK, 2012. Springer-Verlag.
6. S. Borgwardt and R. Peñaloza. Undecidability of fuzzy description logics. In G. Brewka, T. Eiter, and S. A. McIlraith, editors, *Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning (KR 2012)*, pages 232–242, Rome, Italy, 2012. AAAI Press.
7. S. Borgwardt and R. Peñaloza. About subsumption in fuzzy $\mathcal{EL}$. In *Proceedings of the 2013 International Workshop on Description Logics (DL'13)*, CEUR-WS, Ulm, Germany, 2013. To appear.
8. S. Borgwardt and R. Peñaloza. The complexity of lattice-based fuzzy description logics. *Journal on Data Semantics*, 2(1):1–19, 2013.
9. S. Borgwardt and R. Peñaloza. Positive subsumption in fuzzy $\mathcal{EL}$ with general t-norms. In *Proc. 23rd Int. Joint Conf. on Artif. Intel. (IJCAI 2013)*, Beijing, China, 2013. AAAI Press. To appear.
10. M. Cerami and U. Straccia. On the (un)decidability of fuzzy description logics under Łukasiewicz t-norm. *Information Sciences*, 227:1–21, 2013.
11. C. d'Amato, N. Fanizzi, F. Esposito, and T. Lukasiewicz. Representing uncertain concepts in rough description logics via contextual indiscernibility relations. In *Uncertainty Reasoning for the Semantic Web II, International Workshops URSW 2008-2010 Held at ISWC and UniDL 2010 Held at FLoC, Revised Selected Papers*, volume 7123 of *Lecture Notes in Computer Science*, pages 300–314. Springer, 2013.
12. K. de Queiroz. Different species problems and their resolution. *BioEssays*, 27(12):1263–1269, 2005.
13. P. Hájek. *Metamathematics of Fuzzy Logic (Trends in Logic)*. Springer-Verlag, 2001.
14. Y. Jiang, J. Wang, S. Tang, and B. Xiao. Reasoning with rough description logics: An approximate concepts approach. *Information Sciences*, 179(5):600–612, 2009.
15. Y. Kazakov, M. Krötzsch, and F. Simančík. ELK reasoner: Architecture and evaluation. In *Proceedings of the OWL Reasoner Evaluation Workshop 2012 (ORE'12)*, volume 858 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2012.
16. C. M. Keet. Rough subsumption reasoning with rowl. In I. Brown, K. Sewchurran, and H. Suleman, editors, *Proc. of the 2011 Annual Conf. of the South African Inst. of Comp. Scientists and Inform. Tech. (SAICSIT 2011)*, pages 133–140. ACM, 2011.
17. C.-J. Liau. On rough terminological logics. In *Proc. of the 4th Intern. Workshop on Rough Sets, Fuzzy Sets and Machine Discovery*, pages 47–54, 1996.
18. T. Lukasiewicz and U. Straccia. Managing uncertainty and vagueness in description logics for the semantic web. *Journal of Web Semantics*, 6(4):291–308, 2008.
19. Z. Pawlak. Rough sets. *International Journal of Parallel Programming*, 11(5):341–356, 1982.

20. R. Peñaloza and T. Zou. Roughening the $\mathcal{EL}$ envelope. In *Proc. of the 2013 International Symposium on Frontiers of Combining Systems (FroCoS 2013)*, Lecture Notes in Computer Science, Nancy, France, 2013. Springer-Verlag. To appear.
21. S. Schlobach, M. C. A. Klein, and L. Peelen. Description logics with approximate definitions - precise modeling of vague concepts. In M. M. Veloso, editor, *Proc. 20th Int. Joint Conf. on Artif. Intel. (IJCAI 2007)*, pages 557–562, 2007.
22. U. Straccia. Reasoning within fuzzy description logics. *J. Artif. Intell. Res.*, 14:137–166, 2001.
23. D. Toman and G. Weddell. On reasoning about structural equality in xml: a description logic approach. *Theor. Comput. Sci.*, 336(1):181–203, May 2005.
24. Y. Yao. Perspectives of granular computing. In *Proceeding of the 2005 IEEE International Conference on Granular Computing*, volume 1, pages 85–90 Vol. 1, 2005.
25. L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.