Roughening the $\mathcal{EL}$ Envelope

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. The $\mathcal{EL}$ family of description logics (DLs) has been successfully applied for representing the knowledge of several domains, specially from the bio-medical fields. One of its principal characteristics is that its reasoning tasks have polynomial complexity, which makes them suitable for large-scale knowledge bases. In their classical form, description logics cannot handle imprecise concepts in a satisfactory manner. Rough sets have been studied as a method for describing imprecise notions, by providing a lower and an upper approximation, which are defined through classes of indiscernible elements.

In this paper we study the combination of the $\mathcal{EL}$ family of DLs with the notion of rough sets, thus obtaining a family of rough DLs. We show that the rough extension of these DLs maintains the polynomial-time complexity enjoyed by its classical counterpart. We also present a completion-based algorithm that is a strict generalization of the known method for the DL $\mathcal{EL}^{++}$.

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 modelling and reasoning with real-world knowledge domains, but arguably its largest 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 lightweight 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, minor extensions of $\mathcal{EL}$ are the basic logics underlying large-scale ontologies like Snomed CT[4] or the Gene Ontology.[5] A prominent feature of these logics is

⋆ 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`

their polynomial-time complexity of reasoning, which enables effective reasoning procedures. In fact, modern reasoners are capable of classifying Snomed CT, which has approximately 300,000 axioms, in less than seven seconds [16].

In their classical form the members of the $\mathcal{EL}$ family, as all other classical DLs, lack 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. Moreover, these two groups of individuals are uncapable to interbreed, which is the minimal requirement for distinguishing elements of a species. 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. These intermediate individuals form also a chain of interbreeding relations that goes from the Monterey to the Large Blotched Ensatinas. It is thus unclear 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 problem in biology [11].

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 [19,23]. However, it was recently shown that reasoning in expressive fuzzy DLs is either undecidable [6,9], or must ignore the truth degrees [5]. Even for the inexpressive DL $\mathcal{EL}$, the extension to general fuzzy-set based semantics usually yields intractable reasoning problems [8]. 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 [4,7]. 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 [20] as an alternative to fuzzy set theory [26] 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 $\overline{X} \setminus \underline{X}$ are the 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.

In this paper we study rough $\mathcal{EL}^{++}$, a logic that combines the DL $\mathcal{EL}^{++}$ (without concrete domains) and rough set semantics. Although the combination of rough set theory with DLs is far from new (see e.g. [18] for some early work), interest in it has grown in the last few years [10,14,17,22]. 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 [22] 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 expressive DLs.

The reduction from [22], 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. This paper extends the results from [21].

2 Rough Sets

Rough sets were introduced in [20] as an alternative to fuzzy set theory [26] 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. Intuitively, elements belonging to the same equivalence class cannot be distinguished through their perceivable characteristics, and hence cannot be divided by a given set. Rough sets are also sometimes called granular sets in the literature and are one of the basis for granular computing [25].

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, members X and 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 cannot be enforced through the conjunction and implication of rough sets.

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 $\underline{\cdot}$ and $\overline{\cdot}$ for expressing rough concepts. Formally, from three mutually disjoint sets $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ of *concept*, *role*, and *individual* names, *rough* $\mathcal{EL}^{++}$ *concepts* are constructed using the following syntactic rule:

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

where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $a \in \mathsf{N_I}$.[6]

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 require an indiscernibility 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* $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$*, every role name* r *to a binary relation* $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$*, and every individual name* a *to an element* $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$*.*

As usual, we denote the equivalence class of an element $x \in \Delta^{\mathcal{I}}$ w.r.t. the relation $\sim_{\mathcal{I}}$ by $[x]_{\sim_{\mathcal{I}}}$. The interpretation function is extended to general rough $\mathcal{EL}^{++}$ concepts by setting:

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

Intuitively, the indiscernibility relation groups the elements 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 (i.e., can be detached) from every element *not* belonging to C. Clearly, for every interpretation $\mathcal{I}$ and concept C it holds that $\underline{C}^{\mathcal{I}} \subseteq C^{\mathcal{I}} \subseteq \overline{C}^{\mathcal{I}}$. The borderline cases, those elements belonging to $\overline{C}^{\mathcal{I}} \setminus \underline{C}^{\mathcal{I}}$, cannot be ensured to be, nor excluded from being instances of C through the equivalence relation.

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, and *role inclusion axioms* (RIs) of the form $r \circ s \sqsubseteq t$ or $r \sqsubseteq t$, where $r, s, t \in \mathsf{N_R}$. Their semantics is defined as follows. The interpretation $\mathcal{I}$ *satisfies* the GCI $C \sqsubseteq D$ if and only if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ holds. It satisfies the RI $r \circ s \sqsubseteq t$ (resp., $r \sqsubseteq t$) if $r^{\mathcal{I}} \circ s^{\mathcal{I}} \subseteq t^{\mathcal{I}}$ (resp., $r^{\mathcal{I}} \subseteq t^{\mathcal{I}}$). $\mathcal{I}$ is a *model* of the TBox $\mathcal{T}$ if it satisfies all the GCIs and RIs in $\mathcal{T}$.

Contrary to less expressive DLs such as $\mathcal{EL}$, it is possible to build inconsistent rough $\mathcal{EL}^{++}$ TBoxes, due to the presence of the bottom concept $\bot$. As a simple example, consider the GCI $\{a\} \sqsubseteq \bot$ that cannot be satisfied by any interpretation $\mathcal{I}$. Despite this situation, we still focus our attention to the problem of deciding subsumption between concepts, which can be used to solve all

[6] The logic $\mathcal{EL}^{++}$ allows also for concrete domains. In this paper we decided to exclude concrete domains to reduce the number of completion rules, and simplify the proofs. Including this constructor in the logic should not affect our complexity results.

other standard reasoning problems like concept satisfiability, or the instance problem [1].

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 and the TBox

$$\begin{aligned}
\mathsf{MontereyE} \sqcap \mathsf{LargeBlotchedE} &\sqsubseteq \bot \\
\exists \mathsf{interbreed}.\mathsf{MontereyE} &\sqsubseteq \overline{\mathsf{MontereyE}} \\
\exists \mathsf{interbreed}.\mathsf{LargeBlotchedE} &\sqsubseteq \overline{\mathsf{LargeBlotchedE}}
\end{aligned}$$

that describes usual desired properties of the notion of species; namely, that no individual may belong to two different species (first axiom), and that belonging to a species is characterized by the capacity of interbreeding with elements of that species (last two axioms). Consider now three salamanders a, b, c such that a and c belong to each of the limit species, and b can interbreed with both; i.e.,

$$\begin{aligned}
\{a\} &\sqsubseteq \mathsf{MontereyE} \\
\{c\} &\sqsubseteq \mathsf{LargeBlotchedE} \\
\{b\} &\sqsubseteq \exists \mathsf{interbreed}.\{\mathsf{a}\} \sqcap \exists \mathsf{interbreed}.\{\mathsf{c}\}.
\end{aligned}$$

From all these axioms, we can deduce that the salamander b belongs to the upper approximation of both limit species, and hence is an intermediate salamander. We could then deduce some further properties of $\{b\}$ in the presence of other axioms in the TBox.

If we had restricted the description to the classical definition of species through interbreeding, i.e., used the axioms $\exists \mathsf{interbreed}.\mathsf{MontereyE} \sqsubseteq \mathsf{MontereyE}$ and $\exists \mathsf{interbreed}.\mathsf{LargeBlotchedE} \sqsubseteq \mathsf{LargeBlotchedE}$ in place of the upper approximations as above, the TBox would be inconsistent as b would be a member of both species, which are specified to be disjoint. In this case, rough concepts provide a (partial) solution to the species problem.

As shown in [22], 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 [22] for full details). However, although transitive roles are a feature of $\mathcal{EL}^{++}$, 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-complete [1,2,24]. 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 concepts is 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 can sometimes be deduced, as shown next.

Theorem 4. *For all rough $\mathcal{EL}^{++}$ concepts C, D, E, D_1, D_2, 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* $C \sqsubseteq_{\mathcal{T}} \underline{E}$
4. *if* $C \sqsubseteq_{\mathcal{T}} \underline{D_1}$ *and* $C \sqsubseteq_{\mathcal{T}} \underline{D_2}$ *(respectively,* $C \sqsubseteq_{\mathcal{T}} D_2$*, or* $C \sqsubseteq_{\mathcal{T}} \overline{D_2}$*), then* $C \sqsubseteq_{\mathcal{T}} \underline{D_1 \sqcap D_2}$ *(resp.,* $C \sqsubseteq_{\mathcal{T}} D_1 \sqcap D_2$*, or* $C \sqsubseteq_{\mathcal{T}} \overline{D_1 \sqcap D_2}$*).*

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 $y \in E^{\mathcal{I}}$.
4. If $x \in C^{\mathcal{I}}$, then $[x]_{\sim_{\mathcal{I}}} \subseteq D_1^{\mathcal{I}}$. For the case where $C \sqsubseteq_{\mathcal{T}} \underline{D_2}$, it then follows that $[x]_{\sim_{\mathcal{I}}} \subseteq D_1^{\mathcal{I}} \cap D_2^{\mathcal{I}} = (D_1 \sqcap D_2)^{\mathcal{I}}$, and hence $x \in (\underline{D_1 \sqcap D_2})^{\mathcal{I}}$. If $C \sqsubseteq_{\mathcal{T}} D_2$, then $x \in D_2^{\mathcal{I}}$, and since $x \in [x]_{\sim_{\mathcal{I}}}$, it follows that $x \in (D_1 \sqcap D_2)^{\mathcal{I}}$. Finally, if $C \sqsubseteq_{\mathcal{T}} \overline{D_2}$, then $[x]_{\sim_{\mathcal{I}}} \cap D_2^{\mathcal{I}} \neq \emptyset$ and since $[x]_{\sim_{\mathcal{I}}} \subseteq D_1^{\mathcal{I}}$, it holds that $[x]_{\sim_{\mathcal{I}}} \cap D_1^{\mathcal{I}} = [x]_{\sim_{\mathcal{I}}}$. Thus, $[x]_{\sim_{\mathcal{I}}} \cap (D_1 \sqcap D_2)^{\mathcal{I}} \neq \emptyset$. □

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.

4 A Completion Algorithm

In this section, we describe an algorithm for deciding subsumption relations between concepts. To simplify the description, we will focus exclusively 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.

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	$A \sqsubseteq \overline{C}$	$\longrightarrow$	$\{A \sqsubseteq \overline{X}, X \sqsubseteq C\}$
NF10	$\bot \sqsubseteq E$	$\longrightarrow$	$\emptyset$

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

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

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

where $A, A', B \in \mathsf{BC}$, $C \in \mathsf{BC} \cup \{\bot\}$, 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. Moreover, as shown in [15], a correct handling of nominals requires to keep track of additional dependencies between basic concepts. This is done through a reachability relation $\rightsquigarrow_R$, where $A \rightsquigarrow_R B$ intuitively expresses that if A has a non-empty interpretation, then B must also be non-empty. This relation is built in parallel to the completion sets during the execution of the algorithm.

The algorithm uses a family of completion sets as data structure. In the following we will denote as $\mathsf{BC}_\mathcal{T}$ the set of all basic concepts that appear in the TBox $\mathcal{T}$, and analogously for $\mathsf{N_C}_\mathcal{T}$, $\mathsf{N_R}_\mathcal{T}$, and $\mathsf{N_I}_\mathcal{T}$. For every basic concept $A \in \mathsf{BC}_\mathcal{T}$ and every concept name $G \in \mathsf{N_C}_\mathcal{T}$, we store three completion sets $S^G(A)$, $\underline{S}^G(A)$, and $\overline{S}^G(A)$, and additionally a completion set $S^G(A, r)$ for every

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

role name $r \in \mathsf{N_R}_\mathcal{T}$. The members of the completion sets are all basic concepts or $\bot$. These sets will maintain the following invariants during the whole execution of the algorithm:

I1 if $B \in S^G(A)$, and $G \rightsquigarrow_R A$, then $A \sqsubseteq_\mathcal{T} B$
I2 if $B \in \overline{S}^G(A)$, and $G \rightsquigarrow_R A$, then $A \sqsubseteq_\mathcal{T} \overline{B}$
I3 if $B \in \underline{S}^G(A)$, and $G \rightsquigarrow_R A$, then $A \sqsubseteq_\mathcal{T} \underline{B}$
I4 if $B \in S^G(A, r)$, and $G \rightsquigarrow_R A$, then $A \sqsubseteq_\mathcal{T} \exists r.B$
I5 if $G \rightsquigarrow_R A$, then for every model $\mathcal{I}$ of $\mathcal{T}$, $G^\mathcal{I} \neq \emptyset$ implies $A^\mathcal{I} \neq \emptyset$.

The completion sets are initialised as

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

for basic concepts $A \in \mathsf{BC}_\mathcal{T}$, concept names $G \in \mathsf{N_C}_\mathcal{T}$, and role names $r \in \mathsf{N_R}_\mathcal{T}$. The reachability relation initially states only that $G \rightsquigarrow_R G$ and $G \rightsquigarrow_R \{a\}$ for every $G \in \mathsf{N_C}_\mathcal{T}$ and every $a \in \mathsf{N_I}_\mathcal{T}$. Obviously, this initialisation preserves all the invariants described above.

The completion rules from Table 2 are then applied to extend these sets. Before continuing to show correctness of this algorithm, we briefly explain these rules. The rules up to CR7 correspond to the completion rules for classical $\mathcal{EL}^{++}$ from [1] with the correct treatment of nominals adapted from [15]. The following rules up to CR15 consider the axioms containing rough concepts, as well as the consequences of crisp axioms when applied to rough concepts. The first two of those rules are a simple consequence of the properties of intersections of rough sets. We discuss the rule CR12 in more detail. Under the assumption that G is not empty, $G \rightsquigarrow_R A_2$ states that A_2 must also be non-empty. Additionally, $\{a\} \in \underline{S}^G(A_2)$ in particular implies that every member of A_2 must also belong to $\{a\}$, and hence A_2 must be equivalent to $\{a\}$. Consider now some element of A_1. $\{a\} \in \overline{S}^G(A_1)$ states that this element must be indiscernible from a and hence is indiscernible from an element of A_2. Thus, A_2 must be added to $\underline{S}^G(A_1)$. The rule CR11 follows from a similar but simpler argument.

The next six rules consider a cross-population of of the completion sets, following the properties of rough sets described in the previous section. Finally, the last two rules extend the reachability relation to keep information on which concept names should be interpreted as non-empty under the assumption that G is non-empty.

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.*

CR1	if $B_1 \in S^G(A), B_2 \in S^G(A)$, and $B_1 \sqcap B_2 \sqsubseteq C \in \mathcal{T}$, then add C to $S^G(A)$
CR2	if $B \in S^G(A)$ and $B \sqsubseteq \exists r.C \in \mathcal{T}$, then add C to $S^G(A, r)$
CR3	if $B \in S^G(A, r), C \in S^G(B)$, and $\exists r.C \sqsubseteq D \in \mathcal{T}$, then add D to $S^G(A)$
CR4	if $B \in S^G(A, r)$ and $\bot \in S^G(B)$, then add $\bot$ to $S^G(A)$
CR5	if $B \in S^G(A, r)$ and $r \sqsubseteq t \in \mathcal{T}$, then add B to $S^G(A, t)$
CR6	if $B \in S^G(A, r), C \in S^G(B, s)$, and $r \circ s \sqsubseteq t \in \mathcal{T}$, then add C to $S^G(A, t)$
CR7	if $\{a\} \in S^G(A_1) \cap S^G(A_2)$ and $G \rightsquigarrow_R A_2$, then add A_2 to $S^G(A_1)$
CR8	if $B_1 \in \underline{S}^G(A), B_2 \in \underline{S}^G(A)$, and $B_1 \sqcap B_2 \sqsubseteq C \in \mathcal{T}$, then add C to $\underline{S}^G(A)$
CR9	if $B_1 \in \underline{S}^G(A), B_2 \in \overline{S}^G(A)$, and $B_1 \sqcap B_2 \sqsubseteq C \in \mathcal{T}$, then add C to $\overline{S}^G(A)$
CR10	if $B \in \overline{S}^G(A)$ and $\bot \in \overline{S}^G(B)$, then add $\bot$ to $\underline{S}^G(A)$
CR11	if $\{a\} \in \overline{S}^G(A_1) \cap \overline{S}^G(A_2)$ and $G \rightsquigarrow_R A_2$, then add A_2 to $\overline{S}^G(A_1)$
CR12	if $\{a\} \in \overline{S}^G(A_1) \cap \underline{S}^G(A_2)$ and $G \rightsquigarrow_R A_2$, then add A_2 to $\underline{S}^G(A_1)$
CR13	if $B \in \underline{S}^G(A)$ and $\underline{B} \sqsubseteq C \in \mathcal{T}$, then add C to $\underline{S}^G(A)$
CR14	if $B \in \overline{S}^G(A)$, and $B \sqsubseteq \underline{C} \in \mathcal{T}$, then add C to $\underline{S}^G(A)$
CR15	if $B \in \overline{S}^G(A)$, and $B \sqsubseteq \overline{C} \in \mathcal{T}$, then add C to $\overline{S}^G(A)$
CR16	if $B \in \underline{S}^G(A)$ then add B to $S^G(A)$
CR17	if $B \in S^G(A)$ then add B to $\overline{S}^G(A)$
CR18	if $B \in \underline{S}^G(A)$ and $C \in S^G(B)$ then add C to $\underline{S}^G(A)$
CR19	if $B \in \overline{S}^G(A)$ and $C \in \overline{S}^G(B)$ then add C to $\overline{S}^G(A)$
CR20	if $B \in \overline{S}^G(A)$ and $C \in \underline{S}^G(B)$ then add C to $\underline{S}^G(A)$
CR21	if $\bot \in \overline{S}^G(A)$ then add $\bot$ to $\underline{S}^G(A)$
CR22	if $G \rightsquigarrow_R A$ and $B \in S^G(A, r)$, then $G \rightsquigarrow_R B$
CR23	if $G \rightsquigarrow_R A$ and $B \in \overline{S}^G(A)$, then $G \rightsquigarrow_R B$

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

Proof. Each of the completion sets contains only basic concepts that appear in $\mathcal{T}$. 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 for every basic concept, plus one additional completion set for each basic concept and role name. Thus, the number of completion sets is quadratic on the size of $\mathcal{T}$. Each application of a completion rule CR1–CR21 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. The reachability relation $\rightsquigarrow_R$ maps basic concepts, so it can have at most quadratically many elements. Each application of one of the last two rules adds a pair to this relation, and hence only quadratically many rule applications are possible.

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. $\square$

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 (i) $\{B, \bot\} \cap S^A(A) \neq \emptyset$, or (ii) there exists $a \in \mathsf{N}_{\mathsf{I}\mathcal{T}}$ such that $\bot \in S^A(\{a\})$. We prove first that the method is sound, by showing that rule applications preserve the invariants I1 to I5 described before.

Lemma 6. *The invariants* I1 *to* I5 *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 seven rules has been shown in [1,15]. 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 CR16 and CR17.

For the rule CR8, 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 CR9 can be treated analogously. Soundness of the rules treating nominals has been argued before, and of the remaining concept rules is a direct consequence of Theorem 4.

The last two rules simply transfer the assumption of non-emptiness to all existential successors, in the first case, and to all weak subsumers in the second case. This transfer preserves the invariant I5. $\square$

Since $A \rightsquigarrow_R A$, the first invariant entails that whenever $B \in S^A(A)$, the subsumption relation $A \sqsubseteq_{\mathcal{T}} B$ holds. Likewise, if $\bot \in S^A(A) \cup S^A(\{a\})$, the same invariant together with I5 yield that A must be interpreted as empty by every model of $\mathcal{T}$. If this is the case, then A is trivially subsumed by B.

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 described before. As usual, we show this by building, given concept names $A, B \in \mathsf{N}_{\mathsf{C}\mathcal{T}}$ not satisfying the conditions (i) nor (ii) above, a countermodel for the subsumption relation between A and B. The main idea is to have one domain element for each basic concept C appearing in $\mathcal{T}$, which can be reached from A through the relation $\rightsquigarrow_R$ (and thus, must have a non-empty interpretation in the countermodel). The interpretation function will include this element 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: from $C \sqsubseteq_{\mathcal{T}} \overline{D_1}$ and $C \sqsubseteq_{\mathcal{T}} \overline{D_2}$, and $x \in C^{\mathcal{I}}$, we can only deduce that there exist y_1 and y_2 such that $x \sim_{\mathcal{I}} y_i$ and

$y_i \in D_i^{\mathcal{I}}$ holds for $i \in \{1,2\}$. If we enforce $y_1 = y_2$, then it would follow that $x \in (\overline{D_1 \sqcap D_2})^{\mathcal{I}}$, but this is *not* a consequence of the two subsumptions. Thus, we need to treat the witnesses for C being subsumed by $\overline{D_1}$ and by $\overline{D_2}$ independently. Moreover, since nominals must be interpreted as singleton sets, we also need to identify all basic concepts that are subsumed by the same nominal. We also need to identify the auxiliary domain elements introduced for dealing with rough constructors, if they refer to the same nominal, or if they were generated by a conjunction of lower and upper approximations. We formalize these ideas next.

Lemma 7. *Let A, B be two concept names appearing in $\mathcal{T}$, and S^A the class completion sets for A obtained after the application of the completion rules has terminated. If $\{B, \bot\} \cap S^A(A) = \emptyset$ and $\bot \notin S^A(\{a\})$, for all $a \in \mathsf{N_I}_{\mathcal{T}}$, 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 set of relevant concepts

$$\mathcal{C} := \{C, C_u, C_D \mid C, D \in \mathsf{BC}_{\mathcal{T}}, A \rightsquigarrow_R C, D \in \overline{S}^A(C)\}.$$

Let $\bowtie$ be the relation on $\mathcal{C}$ where $x \bowtie y$ iff any of the following conditions hold:

1. exist $a \in \mathsf{N_I}_{\mathcal{T}}$, $C, D \in \mathsf{BC}_{\mathcal{T}}$ with $x = C$, $y = D$, and $\{a\} \in S^A(C) \cap S^A(D)$,
2. exist $a \in \mathsf{N_I}_{\mathcal{T}}$ and $C \in \mathsf{BC}_{\mathcal{T}}$ with $x = \{a\}$ and $y = C_{\{a\}}$,
3. exists $C \in \mathsf{BC}_{\mathcal{T}}$ with $x = C, y = C_u$ and $C \in \underline{S}^A(C)$, or
4. exist $C, D_1, D_2, E \in \mathsf{BC}_{\mathcal{T}}$ with $x = C_{D_1}$ $y = C_E$, $D_2 \in \underline{S}^A(C), D_1 \in \overline{S}^A(C)$, and $D_1 \sqcap D_2 \sqsubseteq E \in \mathcal{T}$.

Let $\ulcorner C \urcorner$ denote the equivalence class of C on the transitive, reflexive and symmetric closure of $\bowtie$. These equivalence classes form the interpretation domain; that is, $\Delta^{\mathcal{I}} := \{\ulcorner C \urcorner \mid C \in \mathcal{C}\}$.

The idea is to use the class $\ulcorner C \urcorner$ as a prototype individual belonging to the concept C. Recall that we have assumed that $\bot \notin S^A(A) \cup \bigcup_{a \in \mathsf{N_I}_{\mathcal{T}}} S^A(\{a\})$. From this assumption it follows that $\bot$ does not appear in any equivalence class.

The indiscernibility relation $\sim_{\mathcal{I}}$ is the transitive, reflexive and symmetric closure of $\{(\ulcorner C \urcorner, \ulcorner C_u \urcorner), (\ulcorner C \urcorner, \ulcorner C_D \urcorner) \mid C, D \in \mathsf{BC}_{\mathcal{T}}, A \rightsquigarrow_R C\}$; thus, the indiscernibility class defined by a basic concept C is

$$[C]_{\sim_{\mathcal{I}}} := \{\ulcorner C \urcorner, \ulcorner C_u \urcorner\} \cup \{\ulcorner C_D \urcorner \mid D \in \overline{S}^A(C)\}.$$

It remains only to define the interpretation function $\cdot^{\mathcal{I}}$. For a concept name $C \in \mathsf{N_C}_{\mathcal{T}}$, role name $r \in \mathsf{N_R}_{\mathcal{T}}$ and individual name $a \in \mathsf{N_I}_{\mathcal{T}}$, we set

$$\begin{aligned}
C^{\mathcal{I}} &:= \{\ulcorner D\urcorner \mid C \in S^A(D)\} \cup \{\ulcorner D_u\urcorner \mid C \in \underline{S}^A(D)\} \cup \\
&\quad\ \{\ulcorner D_X\urcorner \mid C \in S^A(X), X \in \overline{S}^A(D)\} \cup \{\ulcorner D_X\urcorner \mid C \in \underline{S}^A(D), D_X \in \mathcal{C}\}, \\
r^{\mathcal{I}} &:= \{(\ulcorner C\urcorner, \ulcorner D\urcorner) \mid D \in S(C,r)\} \cup \{(\ulcorner C_u\urcorner, \ulcorner D\urcorner) \mid D \in S(X,r), X \in \underline{S}(C)\} \cup \\
&\quad\ \{(\ulcorner C_X\urcorner, \ulcorner D\urcorner) \mid D \in S(X,r), X \in \overline{S}(C)\} \cup \\
&\quad\ \{(\ulcorner C_X\urcorner, \ulcorner D\urcorner) \mid D \in S(Y,r), Y \in \underline{S}(C), C_X \in \mathcal{C}\}, \text{ and} \\
a^{\mathcal{I}} &:= \ulcorner\{a\}\urcorner.
\end{aligned}$$

It can be seen that this interpretation function is well defined. It is a simple case analysis to show that, for every $C \in \mathsf{BC}_{\mathcal{T}}$, and every $D \in \mathsf{BC}_{\mathcal{T}} \cup \{\bot\}$ it holds that $\ulcorner C\urcorner \in D^{\mathcal{I}}$ iff $D \in S^A(C)$. Hence, we have that $\ulcorner A\urcorner \in A^{\mathcal{I}}$ but $\ulcorner A\urcorner \notin B^{\mathcal{I}}$. It only remains to be shown that $\mathcal{I}$ is indeed a model of $\mathcal{T}$. The proof is by case analysis, on the shape of the axiom, and the domain element.

$[\underline{C} \sqsubseteq D]$ Let $x \in \underline{C}^{\mathcal{I}}$; i.e., $[x]_{\sim_{\mathcal{I}}} \subseteq C^{\mathcal{I}}$ and let $E \in \mathsf{BC}_{\mathcal{T}}$ such that $[E]_{\sim_{\mathcal{I}}} = [x]_{\sim_{\mathcal{I}}}$. Then $\ulcorner E_u\urcorner \in C^{\mathcal{I}}$. By definition, this means that $C \in \underline{S}(E)$. Since the rule CR13 is not applicable, $D \in \underline{S}(E)$, and by rule CR16, $D \in S(E)$. Let now $\ulcorner E_F\urcorner \in [E]_{\sim_{\mathcal{I}}}$. Since $D \in \underline{S}(E)$, by definition $\ulcorner E_F\urcorner \in D^{\mathcal{I}}$. It thus follows that $[E]_{\sim_{\mathcal{I}}} \subseteq D^{\mathcal{I}}$ and since $x \in [E]_{\sim_{\mathcal{I}}}$, $x \in D^{\mathcal{I}}$.

$[C \sqsubseteq \underline{D}]$ Let $x \in C^{\mathcal{I}}$ and $E \in \mathsf{BC}_{\mathcal{T}}$ with $[x]_{\sim_{\mathcal{I}}} = [E]_{\sim_{\mathcal{I}}}$. Then, x is one of $\ulcorner E_u\urcorner, \ulcorner E\urcorner$, or $\ulcorner E_F\urcorner$ for some $F \in \mathsf{BC}_{\mathcal{T}}$. Since $x \in C^{\mathcal{I}}$, by definition we know that either $C \in \underline{S}^A(E)$, $C \in S^A(E)$, or $C \in S^A(F)$ and $F \in \overline{S}^A(E)$, depending on the shape of x. In any of the three cases, saturation of the rules CR16, CR17, and CR19, implies that $C \in \overline{S}(E)$. By rule CR14, it the 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 in the previous case, we know that $C \in \overline{S}(E)$, and from rule CR15 it follows that $D \in \overline{S}(E)$. Thus, $\ulcorner E_D\urcorner \in D^{\mathcal{I}}$. Since $\ulcorner E_D\urcorner \in [x]_{\sim_{\mathcal{I}}}$, this implies that $[x]_{\sim_{\mathcal{I}}} \cap D^{\mathcal{I}} \neq \emptyset$, and hence $x \in \overline{D}^{\mathcal{I}}$.

$[C_1 \sqcap C_2 \sqsubseteq D]$ Let $x \in C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$ and $E \in \mathsf{BC}_{\mathcal{T}}$ such that $[x]_{\sim_{\mathcal{I}}} = [E]_{\sim_{\mathcal{I}}}$. If $x = \ulcorner E\urcorner$, then by definition $\{C_1, C_2\} \subseteq S^A(E)$, and from rule CR1 it follows that $D \in S^A(E)$ and hence $x = \ulcorner E\urcorner \in D^{\mathcal{I}}$. The case for $x = \ulcorner E_u\urcorner$ can be shown analogously using rule CR8. If $x = \ulcorner E_F\urcorner$ for some $F \in \overline{S}^A(E)$, then for each $i \in \{1,2\}$ it holds that $C_i \in S^A(F) \cup \underline{S}^A(E)$. The cases where both C_is belong to the same set are analogous to the cases for $\ulcorner E\urcorner$ and $\ulcorner E_u\urcorner$ shown before. For the remaining two cases assume w.l.o.g. that $C_1 \in S^A(F)$ and $C_2 \in \underline{S}^A(E)$. Then, by rule CR17 $C_1 \in \overline{S}^A(F)$. The definition of the relation $\bowtie$ then implies that $\ulcorner E_F\urcorner = \ulcorner E_D\urcorner$. From rules CR19 and CR9 it also follows that $D \in \overline{S}^A(E)$. These two facts together imply that $x = \ulcorner E_D\urcorner \in D^{\mathcal{I}}$.

The remaining cases can be treated in a similar way, following the arguments for the classical setting from [1,15]. The only additional difficulty arises in a case analysis for the shape of the domain elements, as the classes for C_u and C_D depend on the completion sets $\underline{S}$ and $\overline{S}$, which have a slightly different behaviour than S. □

This lemma shows that the algorithm is complete. In order to decide whether a concept name A is subsumed by $B \in \mathsf{N_C}_\mathcal{T}$, one needs only analyse the sets $S^A(A)$ and $S^A(\{a\})$ for all $a \in \mathsf{N_I}_\mathcal{T}$. If the goal is to classify the TBox $\mathcal{T}$, then this analysis has to be repeated for all concept names A, however, there is no need to recompute the completion sets; one run of the completion algorithm provides information on all the subsumption relations between concept names. We thus obtain the following result.

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.*

Since subsumption is already PTime-hard for classical $\mathcal{EL}$ [12], this theorem proves that the problem is PTime-complete.

5 Conclusions

We have studied rough $\mathcal{EL}^{++}$, a description logic that extends the lightweight DL $\mathcal{EL}^{++}$ to allow for lower and upper approximations from rough set theory. Rough DLs are presented as an alternative to fuzzy DLs for dealing with imprecise knowledge, in face to the recent negative complexity results for fuzzy description logics. Rough DLs allow for a less fine-grained treatment of vagueness, which reflects in a lower complexity of reasoning.

The logic we studied covers the logical basis for the OWL 2 EL profile of the standard ontology language for the semantic web OWL 2, except for the expression of concrete domains. 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 $\mathcal{EL}^{++}$ reasoners. We do not expect that adding p-admissible concrete domains to this formalism would negatively affect these complexity results.

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. The only difference is in the handling of nominals, where we adapt the method from [15] to obtain completeness. Unfortunately, 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 CR8 to CR21 and rule CR23 whenever the TBox uses only classical $\mathcal{EL}^{++}$ constructors. Our algorithm requires maintaining a higher number of completion sets and dealing with a

[8] Without the lower approximation constructor, the sets $\underline{S}$ are never populated.

larger variety of rules. Despite this, the structure of these completion sets and rules is very similar to the ones used in current implementations of $\mathcal{EL}^{++}$ reasoners. Thus, we do not expect that implementing them into a rough $\mathcal{EL}^{++}$ system would cause much trouble.

These polynomial-time complexity results give strength to the observation from [22] that rough constructors can be added to classical DLs with no additional cost in terms of complexity.

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 [17]. 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 intend to extend our algorithm to deal with concrete domains, hence covering the whole OWL 2 EL profile. Finally, we intend to implement the system and use it for applications that require the representation of imprecise knowledge.

References

1. Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. 19th Int. Joint Conf. on Artif. Intel. (IJCAI 2005)*, Edinburgh, UK, 2005. Morgan-Kaufmann Publishers.
2. Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the el envelope further. In Kendall Clark and Peter F. Patel-Schneider, editors, *Proceedings of the OWLED 2008 DC Workshop on OWL: Experiences and Directions*, 2008.
3. Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2nd edition, 2007.
4. Fernando Bobillo and Umberto Straccia. Finite fuzzy description logics and crisp representations. In *Uncertainty Reasoning for the Semantic Web II, International Workshops URSW 2008-2010 and UniDL 2010, Revised Selected Papers*, volume 7123 of *LNCS*, pages 99–118. Springer, 2013.
5. Stefan Borgwardt, Felix Distel, and Rafael Peñaloza. How fuzzy is my fuzzy description logic? In Bernhard Gramlich, Dale Miller, and Ulrike Sattler, editors, *Proceedings of the 6th International Joint Conference on Automated Reasoning (IJCAR'12)*, volume 7364 of *LNAI*, pages 82–96, Manchester, UK, 2012. Springer.
6. Stefan Borgwardt and Rafael Peñaloza. Undecidability of fuzzy description logics. In Gerhard Brewka, Thomas Eiter, and Sheila 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. Stefan Borgwardt and Rafael Peñaloza. The complexity of lattice-based fuzzy description logics. *Journal on Data Semantics*, 2(1):1–19, 2013.
8. Stefan Borgwardt and Rafael Peñaloza. Positive subsumption in fuzzy $\mathcal{EL}$ with general t-norms. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13)*, Beijing, China, 2013. AAAI Press. To appear.

9. Marco Cerami and Umberto Straccia. On the (un)decidability of fuzzy description logics under Łukasiewicz t-norm. *Information Sciences*, 227:1–21, 2013.
10. Claudia d'Amato, Nicola Fanizzi, Floriana Esposito, and Thomas 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 and UniDL 2010, Revised Selected Papers*, volume 7123 of *LNCS*, pages 300–314. Springer, 2013.
11. Kevin de Queiroz. Different species problems and their resolution. *BioEssays*, 27(12):1263–1269, 2005.
12. Christoph Haase. Complexity of subsumption in extensions of $\mathcal{EL}$. Master's thesis, Dresden University of Technology, Germany, 2007.
13. Petr Hájek. *Metamathematics of Fuzzy Logic (Trends in Logic)*. Springer, 2001.
14. Yuncheng Jiang, Ju Wang, Suqin Tang, and Bao Xiao. Reasoning with rough description logics: An approximate concepts approach. *Information Sciences*, 179(5):600–612, 2009.
15. Yevgeny Kazakov, Markus Kroetzsch, and Frantisek Simancik. Practical reasoning with nominals in the el family of description logics. In Gerhard Brewka, Thomas Eiter, and Sheila A. McIlraith, editors, *Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning (KR 2012)*, pages 264–274. AAAI Press, 2012.
16. Yevgeny Kazakov, Markus Krötzsch, and František 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.
17. C. Maria Keet. Rough subsumption reasoning with rowl. In Irwin Brown, Kosheek Sewchurran, and Hussein 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.
18. Churn-Jung Liau. On rough terminological logics. In *Proc. of the 4th Intern. Workshop on Rough Sets, Fuzzy Sets and Machine Discovery*, pages 47–54, 1996.
19. Thomas Lukasiewicz and Umberto Straccia. Managing uncertainty and vagueness in description logics for the semantic web. *Journal of Web Semantics*, 6(4):291–308, 2008.
20. Zdzislaw Pawlak. Rough sets. *International Journal of Parallel Programming*, 11(5):341–356, 1982.
21. Rafael Peñaloza and Tingting Zou. Rough $\mathcal{EL}$ classification. In *Proceedings of the 2013 International Workshop on Description Logics (DL'13)*, CEUR-WS, Ulm, Germany, 2013. To appear.
22. Stefan Schlobach, Michel C. A. Klein, and Linda Peelen. Description logics with approximate definitions - precise modeling of vague concepts. In Manuela M. Veloso, editor, *Proc. 20th Int. Joint Conf. on Artif. Intel. (IJCAI 2007)*, pages 557–562, 2007.
23. Umberto Straccia. Reasoning within fuzzy description logics. *J. Artif. Intell. Res.*, 14:137–166, 2001.
24. David Toman and Grant Weddell. On reasoning about structural equality in xml: a description logic approach. *Theor. Comput. Sci.*, 336(1):181–203, May 2005.
25. Yiyu Yao. Perspectives of granular computing. In *Proceeding of the 2005 IEEE International Conference on Granular Computing*, volume 1, pages 85–90 Vol. 1, 2005.
26. Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.