Rough, Rougher, Roughest

Extending $\mathcal{EL}$ with a Hierarchy of Indiscernibility Relations

Rafael Peñaloza[1], Anni-Yasmin Turhan[2]

[1]*IKR3 Research Lab, University of Milano-Bicocca, Milan, Italy*
[2]*Knowledge Representation Group, Paderborn University, Germany*

Abstract

Rough sets use a partition of a base set induced by a given indiscernibility relation ρ. In practice such partitions can result from clustering the data. DLs with concept operators relying on a single ρ for upper and lower approximations of concepts have been investigated so far. The use of a hierarchy of such equivalence relations as it can result from hierarchical clustering methods is not explored yet in a DL setting. In this paper we extend the rough DL $\mathcal{EL}^{\rho}_{\bot}$ with a hierarchy of indiscernibility relations as and develop a decision procedure for testing subsumption.

Keywords

Rough logics, vagueness, subsumption

1. Introduction

Rough description logics extend usual description logics by concept operators that use rough sets to add a qualitative form of vagueness to concepts. In the field of rough sets, the domain is partitioned by an equivalence relation, the so-called the *indiscernibility relation* ρ, that groups indistinguishable elements into equivalence classes, also known as *granules*. Based on this partition, each set M is associated with two additional sets. One is the *lower approximation* $\underline{M}$, which contains all elements whose equivalence class is completely contained in M. The second set is the *upper approximation* $\overline{M}$, which contains all those elements that belong to an equivalence class that has an overlap with M. When applied to concepts, the lower approximation $\underline{C}$ models the "strong" or typical instances of C, while the upper approximation $\overline{C}$ models elements that are at least "close" or similar to instances of C.

Rough description logics have been defined as extensions of several classical DLs ranging from $\mathcal{EL}$ to $\mathcal{ALC}$. Reasoning in those rough DLs has mostly been investigated in relation to subsumption [1, 2, 3, 4, 5] or answering conjunctive queries [6]. Rough DLs are well-behaved in the sense that reasoning in them is usually of the same complexity as for their classical counter parts. Besides simply admitting a controlled form of vagueness in concept descriptions, there are other uses of rough DLs for ontology building and maintenance such as ontology engineering [7] and modeling concept drift [8].

Already rough sets alone have been used to structure data early on [9] as they can cluster the data, and the vagueness they introduce makes them resilient against incomplete or noisy data. There have been indiscernibility relations devised for different application domains in the literature. Varying the "degree" of indiscernibility, admits structuring the data into finer or coarser granules and thus considering the data on different levels of abstraction. There are also methods to obtain indiscernibility relations that give a hierarchy of granulations [7, 10, 11], that result in a linearly ordered set of indiscernibility relations.

In more general settings, clustering algorithms are a prime means to structure data as these algorithms group data items according to their homogeneity or proximity into clusters. There is a plethora of such methods and corresponding implementations readily available. A common and well-used type of clustering methods are the hierarchical clustering methods like the classical COBWEB algorithm [12]

DL 2024: 37th International Workshop on Description Logics, June 18–21, 2024, Bergen, Norway
rafael.penaloza@unimib.it (R. Peñaloza); turhan@uni-paderborn (A. Turhan)
https://rpenalozan.github.io/ (R. Peñaloza)
0000-0002-2693-5790 (R. Peñaloza)

and its variants. These clustering algorithms partition the data and effectively construct a dendrogram of the data; i.e., they result in a hierarchy of clusters. The corresponding partitions are then a linearly ordered set of equivalence relations.

The hierarchy of partitions obtained from hierarchical clustering or from indiscernibility relations motivates the extension of rough DLs by a finite, linearly ordered set of equivalence relations (being used as indiscernibility relations) $\rho_1, \ldots, \rho_n$. The results of clustering the data can be incorporated in the knowledge base by augmenting the ABox with the role assertions for pairs from the same cluster, i.e., from pairs related by ρ_i. Such an augmentation of the ABox could be achieved by a mapping commonly used in ontology-based data access (OBDA) [13]. The idea to incorporate an indiscernibility relation in the ABox by an ODBA mapping was already described in [14].

In this paper we extend the rough DL $\mathcal{EL}_\bot^\rho$ (originally introduced in [4]), which uses a single granulation by one indiscernibility relation, to $\mathcal{EL}_\bot^{\rho/\mathsf{lin}}$which uses a finite, linearly ordered set of indiscernibility relations, and as such admits the use of a hierarchy of granulations. In this initial study on $\mathcal{EL}_\bot^{\rho/\mathsf{lin}}$, we investigate subsumption in $\mathcal{EL}_\bot^{\rho/\mathsf{lin}}$and develop a decision procedure based on a completion algorithm that essentially computes canonical models for $\mathcal{EL}_\bot^{\rho/\mathsf{lin}}$TBoxes. This algorithm serves as a starting point for investigating ABox reasoning tasks for this logic. The paper is organised as follows: in the next section we introduce the rough DL $\mathcal{EL}_\bot^{\rho/\mathsf{lin}}$. In Section 3 we develop the reasoning algorithm based on a normal form and completion rules. In Section 4 we supply a brief discussion on possible extensions and we end the paper with conclusion and an outlook on future work.

2. The Logic $\mathcal{EL}_\bot^{\rho/\mathsf{lin}}$

We consider an extension of $\mathcal{EL}_\bot^\rho$ where several equivalence relations (representing indiscernibility at different levels of detail) are used. In our setting, these relations are totally ordered from the coarsest to the most finely-grained. More formally, given $n \geq 1$, we consider n equivalence relations $\sim_1, \ldots, \sim_n$ such that $\sim_i \subseteq \sim_{i+1}$ for all $1 \leq i < n$. That is, $\sim_1$ is the most fine-grained relation, while $\sim_n$ is the coarsest. Note that $\sim_i$ partitions each equivalence class of $\sim_{i+1}$ into (possibly) smaller classes.

Given a fixed but arbitrary $n \in \mathbb{N}$, the set of $\mathcal{EL}_\bot^{\rho/\mathsf{lin}}$*concepts* is constructed through the syntactic rule

$$C ::= A \mid \top \mid \bot \mid C \sqcap C \mid \exists r.C \mid \underline{C}_i \mid \overline{C}^i$$

where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $1 \leq i \leq n$. Concepts of the form $\underline{C}_i$ are called *lower approximation* of C w.r.t. $\sim_i$ and those of the form $\overline{C}^i$ are called *upper approximation* of C w.r.t. $\sim_i$. As usual, a $\mathcal{EL}_\bot^{\rho/\mathsf{lin}}$*TBox* (or *ontology*) is a finite set of *GCIs* of the form $C \sqsubseteq D$, where C and D are $\mathcal{EL}_\bot^{\rho/\mathsf{lin}}$concepts.

The semantics is based on *interpretations* of the form $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I}, \{\sim_i \mid 1 \leq i \leq n\})$ where $(\Delta^\mathcal{I}, \cdot^\mathcal{I})$ is a standard DL interpretation, and $\sim_1, \ldots, \sim_n$ are equivalence relations over $\Delta^\mathcal{I}$ such that $\sim_i \subseteq \sim_{i+1}$ holds for all $1 \leq i < n$. Given an element $\delta \in \Delta^\mathcal{I}$, we denote as $[\delta]_i$ the equivalence class of δ w.r.t. $\sim_i$, which is the class of all objects $\eta \in \Delta^\mathcal{I}$ such that $(\delta, \eta) \in \sim_i$. The interpretation function is extended to arbitrary $\mathcal{EL}_\bot^{\rho/\mathsf{lin}}$concepts as usual for $\top$, $\bot$, $\sqcap$, and $\exists$, while for the lower and upper approximations, we define

$$(\underline{C}_i)^\mathcal{I} := \{\delta \in \Delta^\mathcal{I} \mid [\delta]_i \subseteq C^\mathcal{I}\} \text{ and}$$
$$(\overline{C}^i)^\mathcal{I} := \{\delta \in \Delta^\mathcal{I} \mid [\delta]_i \cap C^\mathcal{I} \neq \emptyset\}.$$

An interpretation *satisfies* the GCI $C \sqsubseteq D$ (denoted as $\mathcal{I} \models C \sqsubseteq D$) iff $C^\mathcal{I} \subseteq D^\mathcal{I}$. The interpretation $\mathcal{I}$ is a *model* of the TBox $\mathcal{T}$ (denoted $\mathcal{I} \models \mathcal{T}$) iff $\mathcal{I} \models C \sqsubseteq D$ holds for all GCIs $C \sqsubseteq D \in \mathcal{T}$.

Note in particular that for every $\delta \in \Delta^\mathcal{I}$ and every i, with $1 \leq i < n$, it holds that $[\delta]_i \subseteq [\delta]_{i+1}$, and hence also $(\underline{C}_{i+1})^\mathcal{I} \subseteq (\underline{C}_i)^\mathcal{I} \subseteq (\overline{C}^i)^\mathcal{I} \subseteq (\overline{C}^{i+1})^\mathcal{I}$ for all concepts C. The following proposition is a consequence of these properties.

Proposition 1. *For all i, j with $1 \leq i \leq j \leq n$, all concepts C, and all interpretations $\mathcal{I}$ the following equivalences hold:*

1. (a) $\big(\underline{(\underline{C}_i)}_j\big)^{\mathcal{I}} = (\underline{C}_j)^{\mathcal{I}}$; (b) $\big(\underline{(\underline{C}_j)}_i\big)^{\mathcal{I}} = (\underline{C}_j)^{\mathcal{I}}$;
2. (a) $\big(\overline{(\overline{C}^i)}^j\big)^{\mathcal{I}} = (\overline{C}^j)^{\mathcal{I}}$; (b) $\big(\overline{(\overline{C}^j)}^i\big)^{\mathcal{I}} = (\overline{C}^j)^{\mathcal{I}}$;
3. $\big(\overline{(\underline{C}_j)}^i\big)^{\mathcal{I}} = (\underline{C}_j)^{\mathcal{I}}$; *and*
4. $\big(\underline{(\overline{C}^j)}_i\big)^{\mathcal{I}} = (\overline{C}^j)^{\mathcal{I}}$.

Proof. We prove only the claims 1. and 3.; the other two can be shown analogously.

For Claim 1.(a), $\delta \in (\underline{(\underline{C}_i)}_j)^{\mathcal{I}}$ iff $[\delta]_j \subseteq (\underline{C}_i)^{\mathcal{I}}$ iff (since $\sim_i \subseteq \sim_j$) $[\delta]_i \subseteq [\delta]_j \subseteq C^{\mathcal{I}}$ iff $\delta \in (\underline{C}_j)^{\mathcal{I}}$. Similarly for 1.(b), $\delta \in (\underline{(\underline{C}_j)}_i)^{\mathcal{I}}$ iff $[\delta]_i \subseteq (\underline{C}_j)^{\mathcal{I}}$ iff for every $\eta \in [\delta]_i$, it holds that $[\eta]_j \subseteq C^{\mathcal{I}}$ iff (since $\delta \sim_i \eta$ holds and implies that $\delta \sim_j \eta$ holds) $[\delta]_j \subseteq C^{\mathcal{I}}$ iff $\delta \in (\underline{C}_j)^{\mathcal{I}}$.

For Claim 3., $\delta \in (\overline{(\underline{C}_j)}^i)^{\mathcal{I}}$ iff $[\delta]_i \cap (\underline{C}_j)^{\mathcal{I}} \neq \emptyset$ iff there exists $\eta \in [\delta]_i$ such that $\eta \in (\underline{C}_j)^{\mathcal{I}}$ iff there is $\eta \in [\delta]_i$ with $[\eta]_j \subseteq C^{\mathcal{I}}$ iff (because $\delta \sim_i \eta$ holds and implies that $\delta \sim_j \eta$ holds) $[\delta]_j \subseteq C^{\mathcal{I}}$ iff $\delta \in (\underline{C}_j)^{\mathcal{I}}$. □

If $i = j$ Claims 1 and 2 from Proposition 1 cover idempotence of both kinds of approximations. This affects the design of the completion rules that treat propagation within the same level of roughness, that is w.r.t. one ρ_i. For $i < j$ the claims from Proposition 1 indicate how information is to be propagated or absorbed between different levels of roughness.

Other important properties which combine the approximation concept constructors for each given indiscernibility relation are the following, which were originally proven in [4].

Proposition 2. *For any three $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$ concepts C, D, E and all i with $1 \leq i \leq n$, the following properties hold:*

1. $\mathcal{T} \models \overline{C}^i \sqsubseteq D$ *iff* $\mathcal{T} \models C \sqsubseteq \underline{D}_i$; *and*
2. *if* $\mathcal{T} \models C \sqsubseteq \overline{D}^i$ *and* $\mathcal{T} \models D \sqsubseteq \underline{E}_i$, *then* $\mathcal{T} \models C \sqsubseteq \underline{E}_i$; *and*
3. *if* $\mathcal{T} \models C \sqsubseteq \overline{D}^i$ *and* $\mathcal{T} \models \underline{D}_i \sqsubseteq E$, *then* $\mathcal{T} \models C \sqsubseteq \underline{E}_i$.

These properties indicate how information is to be propagated within the same level of roughness. All of these properties will be useful when we design a reasoning algorithm for $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$ in the following section.

The rough DL $\mathcal{EL}_{\bot}^{\rho}$ is the special case of $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$ where $n = 1$; that is, where only one equivalence relation is used. Since $\mathcal{EL}_{\bot}$ is a particular case of $\mathcal{EL}_{\bot}^{\rho}$, where the GCIs $A \sqsubseteq \underline{A}_1$ and $\overline{A}^1 \sqsubseteq A$ are satisfied for all $A \in \mathsf{N_C}$, $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$ is obviously a generalisation of the classical DL $\mathcal{EL}_{\bot}$. As usual in these logics, we are mainly interested in deciding whether a consequence follows from an ontology; in this case, we consider the problem of deciding *subsumption* between two concept names. We say that $A \in \mathsf{N_C}$ is *subsumed* by $B \in \mathsf{N_C}$ w.r.t. the TBox $\mathcal{T}$ ($\mathcal{T} \models A \sqsubseteq B$) iff every model of $\mathcal{T}$ also satisfies the GCI $A \sqsubseteq B$.

3. Reasoning in $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$

We are interested in developing a reasoning algorithm capable of deciding subsumption relationships between concepts w.r.t. a given $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$ TBox. As this logic is an extension of $\mathcal{EL}_{\bot}^{\rho}$, we extend the known completion algorithm [4] to handle the new cases required by the multiple indiscernibility relations available.

As a first step, we need to limit the form that GCIs can take, requiring the TBox to comply with a *normal form*; that is, that all the axioms are of one of the forms

$$A_1 \sqcap A_2 \sqsubseteq C, \quad A \sqsubseteq \exists r.B, \quad \exists r.A \sqsubseteq C, \quad \underline{A}_i \sqsubseteq C, \quad A \sqsubseteq \underline{B}_i, \quad A \sqsubseteq \overline{B}^i,$$

Table 1
Normalisation rules. Where $A \in \mathsf{N_C} \cup \{\top\}$, C, D are complex concepts, and X is a new concept name.

NF1	$F \sqcap C \sqsubseteq E \longrightarrow \{C \sqsubseteq X, F \sqcap X \sqsubseteq E\}$
NF2	$\exists r.C \sqsubseteq E \longrightarrow \{C \sqsubseteq X, \exists r.X \sqsubseteq E\}$
NF3	$\underline{C}_i \sqsubseteq E \longrightarrow \{C \sqsubseteq X, \underline{X}_i \sqsubseteq E\}$
NF4	$\overline{C}^i \sqsubseteq E \longrightarrow \{C \sqsubseteq \underline{E}_i\}$
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	$A \sqsubseteq \underline{C}_i \longrightarrow \{A \sqsubseteq \underline{X}_i, X \sqsubseteq C\}$
NF9	$A \sqsubseteq \overline{C}^i \longrightarrow \{A \sqsubseteq \overline{X}^i, X \sqsubseteq C\}$
NF10	$\bot \sqsubseteq E \longrightarrow \emptyset$

where $A, B \in \mathsf{N_C} \cup \{\top\}$, $C \in \mathsf{N_C} \cup \{\top, \bot\}$, and $1 \leq i \leq n$.[1]

Any TBox $\mathcal{T}$ can be transformed into normal form applying the rules from Table 1—where **NF1** uses the commutativity of conjunction—until no rule can be applied anymore. The resulting TBox is a conservative extension of $\mathcal{T}$ which, importantly, is only polynomially larger than $\mathcal{T}$ as it is found after only a polynomial number of rule applications.

Our completion algorithm extends the ideas introduced in [4] to handle lower and upper approximation concepts. Briefly, the completion algorithm for $\mathcal{EL}_\bot^\rho$ preserves, for each concept name A appearing in a normalised TBox $\mathcal{T}$, a family of completion sets, which preserve the information of how the lower and upper approximations of other concept names relate to A. This information is needed for an adequate handling of the properties of these concept constructors. In the present case, we must extend this idea to differentiate between the available indiscernibility relations.

More formally, for each $A \in \mathsf{N_C} \cup \{\top\}$ appearing in the normalised TBox $\mathcal{T}$, and for each $1 \leq i \leq n$ we preserve two sets called $\overline{S}^i(A)$ and $\underline{S}_i(A)$. In addition, we keep track of a set $S(A)$ and for each role name $r \in \mathsf{N_R}$ appearing in $\mathcal{T}$ a set $S(A, r)$. Hence, for each such A, we keep $2n + \ell + 1$ many such completion sets, where ℓ is the number of role names in $\mathcal{T}$. With polynomially many As in the normalised TBox, the completion algorithm uses polynomially many completion sets.

The elements of each completion set all belong to $\mathsf{N_C} \cup \{\top, \bot\}$. The idea is that these sets are sound w.r.t. subsumption relations among simple concepts. Specifically, throughout the completion algorithm, the application of completion rules preserves the following invariants:

1. if $B \in \overline{S}^i(A)$ then $\mathcal{T} \models A \sqsubseteq \overline{B}^i$
2. if $B \in \underline{S}_i(A)$ then $\mathcal{T} \models A \sqsubseteq \underline{B}_i$
3. if $B \in S(A)$ then $\mathcal{T} \models A \sqsubseteq B$ and
4. if $B \in S(A, r)$ then $\mathcal{T} \models A \sqsubseteq \exists r.B$

for all $A \in \mathsf{N_C} \cup \{\top\}$, $B \in \mathsf{N_C} \cup \{\top, \bot\}$, $r \in \mathsf{N_R}$, and $1 \leq i \leq n$. These are essentially the same invariants that were used for $\mathcal{EL}_\bot^\rho$ in [4].

The completion sets are initialized with obvious tautologies; that is, at the beginning of the algorithm the sets are defined as

$$S(A) = \overline{S}^i(A) := \{A, \top\}, \qquad \underline{S}_i(A) := \{\top\}, \qquad S(A, r) := \emptyset$$

for all $A \in \mathsf{N_C} \cup \{\top\}, r \in \mathsf{N_R}, 1 \leq i \leq n$. Clearly this initialization preserves the invariants mentioned above. These sets are extended through application of the *completion rules* described in Table 2. As

[1]For brevity, we consider axioms of the form $A \sqsubseteq B$ as $\top \sqcap A \sqsubseteq B$.

Table 2
Completion rules for $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$.

CR1	if $\{B_1, B_2\} \subseteq S(A)$ 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 $\{B_1, B_2\} \in \underline{S}_i(A)$ and $B_1 \sqcap B_2 \sqsubseteq C \in \mathcal{T}$, then add C to $\underline{S}_i(A)$
CR5	if $B_1 \in \underline{S}_i(A), B_2 \in \overline{S}^i(A)$ and $B_1 \sqcap B_2 \sqsubseteq C \in \mathcal{T}$, then add C to $\overline{S}^i(A)$
CR6	if $B \in \underline{S}_i(A)$ and $\underline{B}_i \sqsubseteq C \in \mathcal{T}$, then add C to $\underline{S}_i(A)$
CR7	if $B \in \overline{S}^i(A)$ and $B \sqsubseteq \underline{C}_i \in \mathcal{T}$, then add C to $\underline{S}_i(A)$
CR8	if $B \in \overline{S}^i(A)$ and $B \sqsubseteq \overline{C}^i \in \mathcal{T}$, then add C to $\overline{S}^i(A)$
CR9	if $B \in \underline{S}_i(A)$, then add B to $S(A)$
CR10	if $B \in S(A)$, then add B to $\overline{S}^i(A)$
CR11	if $B \in \underline{S}_j(A)$ and $i < j$, then add B to $\underline{S}_i(A)$
CR12	if $B \in \overline{S}^i(A)$ and $i < j$, then add B to $\overline{S}^j(A)$
CR13	if $B \in \underline{S}_i(A)$ and $C \in S(B)$, then add C to $\underline{S}_i(A)$
CR14	if $B \in \overline{S}^i(A)$ and $C \in \overline{S}^i(B)$, then add C to $\overline{S}^i(A)$
CR15	if $B \in \underline{S}_i(A)$ and $C \in \underline{S}_i(B)$, then add C to $\underline{S}_i(A)$
CR16	if $B \in S(A, r)$ and $\bot \in S(B)$, then add $\bot$ to $S(A)$
CR17	if $B \in \overline{S}^i(A)$ and $\bot \in \overline{S}^i(B)$, then add $\bot$ to $\underline{S}_i(A)$
CR18	if $\bot \in \overline{S}^i(A)$, then add $\bot$ to $\underline{S}_i(A)$

usual for these kinds of algorithms, the rules are only applied if they add an element to one of the sets involved; that is, if the concept to be added is not already present in the set. The completion algorithm applies rules until no rule is applicable anymore; at that point, we say that the algorithm is *saturated.*

Note that this algorithm becomes saturated after at most polynomially many rule applications (in n and the size of $\mathcal{T}$). Indeed, there are $(2n + \ell + 1)m$ sets, where ℓ is the number of role names in $\mathcal{T}$ and m is the number of concept names in $\mathcal{T}$. Each of this sets contains at most $m + 2$ elements (the concept names in $\mathcal{T}$ plus $\top$ and $\bot$). Since each rule application adds one element to one of the sets, at most $(2n + \ell + 1)(m + 2)m$ rule applications are needed before reaching saturation. In addition, the conditions for the application of a rule require only a lookup between the sets and the GCIs in $\mathcal{T}$, which can also be performed in polynomial time. Thus, overall the algorithm needs only polynomial time to be saturated.

The result of the completion algorithm can be used to decide all the atomic subsumption relations entailed by the TBox $\mathcal{T}$. That is, for every $A, B \in \mathsf{N_C}$ we get that $\mathcal{T} \models A \sqsubseteq B$ iff $B \in S(A)$. Soundness is a consequence of the invariants described above.

Lemma 3. *The completion algorithm preserves the four invariants, throughout all rule applications.*

Proof. The proof is by induction on rule applications. The induction base is satisfied by the initialization. For rules without rough constructors (CR1-CR3 and CR16) soundness was shown already in [15].

For rules CR6 to CR15, CR17, and CR18 soundness is a consequence of Propositions 1 and 2. Since the rules CR11 and CR12 treat the interaction of different indiscernibility relations, we give a detailed proof of them. For CR11, suppose $\mathcal{T} \models A \sqsubseteq \underline{B}_i$ and $i < j$. For every model $\mathcal{I}$ and every $\delta \in \Delta^{\mathcal{I}}$, if $\delta \in A^{\mathcal{I}}$, then $\delta \in \underline{B}_j{}^{\mathcal{I}}$ and thus $[\delta]_j \subseteq B^{\mathcal{I}}$. Since from $i < j$ follows that $[\delta]_i \subseteq [\delta]_j$, we obtain $[\delta]_i \subseteq B^{\mathcal{I}}$ holds and thus $\delta \in \underline{B}_i{}^{\mathcal{I}}$. This implies $\mathcal{T} \models A \sqsubseteq \underline{B}_i$. The proof for CR12 is analogous.

The only remaining rules are CR4 and CR5. For the rule CR4, suppose that $\mathcal{T} \models A \sqsubseteq \underline{B_1}_i$ and

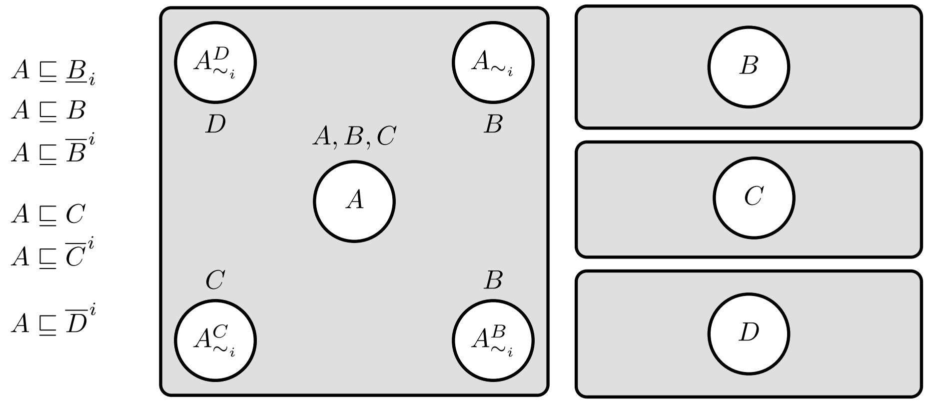

Figure 1: The construction of the model for the proof of Lemma 4. Each gray box is an equivalence class for $\sim_i$. The details of $[A]_{\sim_i}$ are given, relative to the derived subsumptions depicted on the left.

$\mathcal{T} \models A \sqsubseteq \underline{B_2}_i$. For every model $\mathcal{I}$ and every $\delta \in \Delta^{\mathcal{I}}$, if $\delta \in A^{\mathcal{I}}$ then $[\delta]_i \subseteq B_1^{\mathcal{I}} \cap B_2^{\mathcal{I}}$ and hence (as $\mathcal{I} \models B_1 \sqcap B_2 \sqsubseteq C$) $[\delta]_i \subseteq C^{\mathcal{I}}$, which implies $\mathcal{T} \models A \sqsubseteq \underline{C}_i$. Rule CR5 can be treated analogously. $\square$

For the converse direction—completeness—we follow the usual approach of building a sort of canonical model of $\mathcal{T}$ that serves as a counterexample for all the atomic subsumption relations which do *not* appear explicitly in the generated sets. The domain $\Delta^{\mathcal{I}}$ of the canonical model is composed of three kinds of elements. First, as usual for the $\mathcal{EL}$ family of DLs, it includes one domain element for each satisfiable concept name A appearing in $\mathcal{T}$, which stands for a standard instance representing that concept; i.e., it is a minimal representative of A. Hence, it will belong to each concept B that subsumes A w.r.t. $\mathcal{T}$. The two other kinds of domain elements handle the lower and upper approximations of named concepts in the interpretation domain. For the lower approximation, we include, for each $\sim_i$, with $1 \leq i \leq n$, an element $A_{\sim_i}$ that belongs to all concepts B such that $\mathcal{T} \models A \sqsubseteq \underline{B}_i$. In other words, $A_{\sim_i}$ keeps information about all the concept names B such that all objects indiscernible from instances of A are necessarily in B.

Dealing with the upper approximations requires a more nuanced construction, as a single element cannot fully witness the existence of indiscernible elements belonging to different concepts. We handle this with the help of different objects. Specifically, for each concept name B such that $\mathcal{T} \models A \sqsubseteq \overline{B}^i$, we create an element $A^B_{\sim_i}$ which is a representative instance of B (i.e., belongs to B and all its subsumers), but exists only through its connection to the representative of A. To handle the indiscernibility relations, these elements A, $A_{\sim_i}$, and $A^B_{\sim_i}$ all belong to the same $\sim_i$-equivalence class. As this is not a trivial structure, we explain it in more detail here. Note that $\mathcal{T} \models A \sqsubseteq \overline{B}^i$ means that every element of A must be associated (via $\sim_i$) with some element of B. In particular, the representative of A must have such an association as well. But we cannot connect A to the representative of B because the symmetry of $\sim_i$ would then entail that $B \sqsubseteq \overline{A}^i$, which is not necessarily a consequence of $\mathcal{T}$. We can also not choose only one representative, as we did for the lower approximations, because (again) we cannot guarantee that the representative belongs to other concepts that are not known subsumers of B. Figure 1 describes this intuition graphically. Each gray box is an equivalence class for $\sim_i$. There can be more elements than those shown, in each class, but the figure zooms into some relevant elements of $[A]_{\sim_i}$, given by the derivations shown at the left of the figure. Since $A \sqsubseteq \underline{B}_i$, the object $A_{\sim_i}$ belongs to the concept B. On the other hand, since A is subsumed by $\overline{B}^i$, $\overline{C}^i$, and $\overline{D}^i$, we create the three objects $A^B_{\sim_i}$, $A^C_{\sim_i}$, and $A^D_{\sim_i}$, respectively. Importantly, these objects belong to the concepts B, C, and D (respectively), but not to $[B]_{\sim_i}$, $[C]_{\sim_i}$, or $[D]_{\sim_i}$, represented as the three boxes on the right.

Before formalising this construction, we recall that $\bot$ requires a special treatment when it appears as a subsumer of a concept name. If $\mathcal{T} \models A \sqsubseteq \bot$, we know that every model makes A empty, and hence A is subsumed by all concepts. Rather than making all these relations explicit, we simply handle this special case separately.

Lemma 4. *Let A, B be two concept names appearing in the normalised $\mathcal{EL}_\bot^{\rho/lin}$ TBox $\mathcal{T}$, and $S(A)$ the set obtained after saturation of the completion algorithm. If $\{B, \bot\} \cap S(A) = \emptyset$, then $\mathcal{T} \not\models A \sqsubseteq B$.*

Proof. We build a model $\mathcal{I}$ of $\mathcal{T}$ such that $A^\mathcal{I} \not\subseteq B^\mathcal{I}$. The domain of this interpretation is

$$\Delta^\mathcal{I} := \{C, C_{\sim_i}, C^D_{\sim_i} \mid 1 \leq i \leq n, \text{ and } C, D \text{ are concept names appearing in } \mathcal{T}\}.$$

For each i, with $1 \leq i \leq n$, the equivalence relation $\sim_i$ is the transitive, symmetric, and reflexive closure of the relation

$$\{(C, C_{\sim_i}), (C, C^D_{\sim_i}) \mid C, D \text{ are concept names appearing in } \mathcal{T}\}.$$

Note that all objects in $\Delta^\mathcal{I}$ are of the form C, $C_{\sim_i}$, or $C^D_{\sim_i}$. By the definition of the equivalence relations $\sim_i$, for every $\delta \in \Delta^\mathcal{I}$ there exists some concept name C such that $\delta \sim_i C$. In particular, this means that every equivalence class of $\sim_i$ contains at least one concept name or, in other terms, that for every $\delta \in \Delta^\mathcal{I}$ there exists some $E \in \mathsf{N_C}$ such that $[\delta]_i = [E]_i$.

To define the interpretation function $\cdot^\mathcal{I}$, we set for each concept name C appearing in $\mathcal{T}$

$$\begin{aligned} C^\mathcal{I} := {} & \{D \mid C \in S(D)\} \cup \\ & \{D_{\sim_i} \mid C \in \underline{S}_i(D)\} \cup \\ & \{D^E_{\sim_i} \mid C \in S(E), E \in \overline{S}^i(D)\} \cup \\ & \{D^E_{\sim_i} \mid C \in \underline{S}_i(D), E \in \mathsf{N_C}\} \end{aligned}$$

and for each role name r

$$\begin{aligned} r^\mathcal{I} := {} & \{(C, D) \mid D \in S(C, r)\} \cup \\ & \{(C_{\sim_i}, D) \mid D \in S(E, r), E \in \underline{S}_i(C)\} \cup \\ & \{(C^E_{\sim_i}, D) \mid D \in S(E, r), E \in \overline{S}^i(C)\} \cup \\ & \{(C^E_{\sim_i}, D) \mid D \in S(F, r), F \in \underline{S}_i(C), E \in \mathsf{N_C}\}. \end{aligned}$$

By construction $A \in A^\mathcal{I}$ and since $B \notin S(A)$ we know that $A \notin B^\mathcal{I}$. It remains to show that this is indeed a model of $\mathcal{T}$. This is shown through a case distinction over the possible types of axioms admitted in the normal form. We show only the cases involving rough constructors.

[Case $\underline{C}_i \sqsubseteq D$] If $\delta \in (\underline{C}_i)^\mathcal{I}$, then by definition $[\delta]_i \subseteq C^\mathcal{I}$. Let $E \in \mathsf{N_C}$ be such that $[\delta]_i = [E]_i$. Then $E_{\sim_i} \in C^\mathcal{I}$ and hence $C \in \underline{S}_i(E)$. As the algorithm has finished, the rule CR6 is not applicable, this means that $D \in \underline{S}_i(E)$ and by CR9 $D \in S(E)$. Consider now an arbitrary $E^F_{\sim_i} \in [E]_i$. Since $D \in \underline{S}_i(E)$, by construction we know that $E^F_{\sim_i} \in D^\mathcal{I}$. Overall, this means that $\delta \in [E]_i \subseteq D^\mathcal{I}$, which proves the result.

[Case $C \sqsubseteq \underline{D}_i$] If $\delta \in C^\mathcal{I}$ and $[\delta]_i = [E]_{\sim_i}$ for some $E \in \mathsf{N_C}$, then by the rules CR9, CR10, and CR14 it follows that $C \in \overline{S}^i(E)$ which, by rule CR7 implies that $D \in \underline{S}_i(E) \subseteq S(E) \subseteq \overline{S}^i(E)$. Then, $[\delta]_i = [E]_i \subseteq D^\mathcal{I}$; that is, $\delta \in (\underline{D}_i)^\mathcal{I}$.

[Case $C \sqsubseteq \overline{D}^i$] As in the previous case, if $\delta \in C^\mathcal{I}$ with $[\delta]_i = [E]_i$, then $C \in \overline{S}^i(E)$. Rule CR8 then implies that $D \in \overline{S}^i(E)$ and hence $E^D_{\sim_i} \in D^\mathcal{I}$. By construction, $E^D_{\sim_i} \in [\delta]_i$, which implies that $[\delta]_i \cap D^\mathcal{I} \neq \emptyset$, and hence $\delta \in (\overline{D}^i)^\mathcal{I}$. □

Thus we have a decision procedure for subsumption in $\mathcal{EL}_\bot^\rho$. Overall, we get the main result from this paper.

Theorem 5. *Subsumption between concept names w.r.t. $\mathcal{EL}_\bot^{\rho/lin}$ TBoxes can be decided in polynomial time.*

Note that the completion algorithm can be used also to check TBox consistency and concept satisfiability. For the latter, we have from Lemma 4 that A is unsatisfiable w.r.t. $\mathcal{T}$ iff $\bot \in S(A)$. For the former, we can add the GCI $\top \sqsubseteq X$ and check whether X is unsatisfiable.

4. Discussions

Admitting Nominals. Extending $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$with nominals would effectively give a means to identify and address a particular granule in the TBox by using the concept $\overline{\{a\}}^{i}$. This might be useful for some applications. The subsumption algorithm for rough $\mathcal{EL}^{++}$ in [4] even admits nominals, so that a completion algorithm for $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$extended by nominals would simply need to combine the techniques. However, admitting nominals in the TBox would also admit to change the clustering results by GCIs that add a nominal to a granule such as

$$\{a\} \sqsubseteq \overline{C}^{i} \qquad \text{or} \qquad \overline{\{a\}}^{i} \equiv \overline{\{b\}}^{i},$$

or could remove individuals from a granule by disjointness axioms like $\overline{a}^{i} \sqcap \overline{C}^{i} \sqsubseteq \bot$. This is not compatible with the idea of having the granules populated by a mapping from the results of a clustering algorithm. The TBox could then even contradict such a clustering. Nevertheless, it might be useful to admit nominals and their approximations in the query language to reason over a knowledge base.

Partial Orders of Indiscernibility Relations. For this paper, we focused on a family of indiscernibility relations that form a total order, from the most fine-grained (least rough) to the roughest. A natural question is whether it is possible to relax the conditions to allow for *partial* orders between these equivalence relations. This remains an open problem at the time, yet we argue that it is as hard as the general case, where arbitrary equivalence relations (without any ordering between them) are chosen, even if we require the partial order to be connected. Indeed, if we have n arbitrary equivalence relations, we can always represent them as a connected partial order of $n+1$ relations where the new relation $\sim_0$ is contained in all others.

The reason why arbitrary classes of indiscernibility relations is problematic is that there is no way to predict the relationships between objects. For instance, we can have $\alpha \sim_1 \beta \sim_2 \gamma$ and there be no relation between α and γ. A construction akin to our completion algorithm would need to preserve at least 2^n different objects to keep track of this information. Yet, we still do not know whether another strategy could reduce the overall complexity to remain in polynomial time, or in a sub-exponential class.

5. Conclusions and Future Work

In this paper we have extended the $\mathcal{EL}$ family by another rough member, that admits a (linear) hierarchy of indiscernibility relations to be used in upper and lower approximation concepts. The resulting DL $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$can facilitate reasoning w.r.t. clustering results for data that vary in granularity. For the DL $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$, we have devised a subsumption algorithm based on completion rules. This algorithm runs in polynomial time and can also be employed to test satisfiability of concepts.

The next thing to investigate for $\mathcal{EL}_{\bot}^{\rho/\mathsf{lin}}$would be reasoning problems that answer queries over the TBox together with an ABox. Instance checking and answering of conjunctive queries are so far only studied to a small extent for rough DLs. Usually, completion algorithms for subsumption readily extend to algorithms for instance checking, while their extension to algorithms for answering conjunctive queries is more challenging.

Rough DLs have been employed for instance unification [16]. A similar task is solved in entity resolution. Here, sometimes also information on non-equivalence of entities is given to avoid unification. Likewise, some applications of rough sets consider a discernibility relation in addition to the indiscernibility relation. It would be interesting to extend rough DLs by a discernibility relation that can express that two elements are not members of the same equivalence class. As such a relation introduces a form of negation, it is not immediately clear how to extend the reasoning algorithms.

Acknowledgments

Partially supported by the MUR for the Department of Excellence DISCo at the University of Milano-Bicocca and under the PRIN project PINPOINT Prot. 2020FNEB27, CUP H45E21000210001; and by the European Union – Next Generation EU within the project NRPP M4C2, Investment 1.,3 DD. 341 - 15 march 2022 – FAIR – Future Artificial Intelligence Research – Spoke 4 - PE00000013 - D53C22002380006

References

[1] C.-J. Liau, On rough terminological logics, in: Proc. of the 4th Int. Workshop on Rough Sets, Fuzzy Sets and machine Discovery (RSFD'96), 1996, pp. 47–54.

[2] S. Schlobach, M. C. Klein, L. Peelen, Description logics with approximate definitions - precise modeling of vague concepts, in: Proc. of 19th Int. Joint Conference on Artificial Intelligence (IJCAI 2007), 2007, pp. 557–562.

[3] C. M. Keet, Rough subsumption reasoning with rOWL, in: Proc. of the 2011 Ann. Conf. of the South African Inst. of Computer Scientists and Information Technologists, SAICSIT 2011, ACM, 2011, pp. 133–140.

[4] R. Peñaloza, T. Zou, Roughening the $\mathcal{EL}$ envelope, in: Proc. of Int. Symposium on Frontiers of Combining Systems (FroCoS 2013), volume 8152 of *LNCS*, Springer, 2013, pp. 71–86.

[5] C. d'Amato, N. Fanizzi, F. Esposito, T. Lukasiewicz, Representing uncertain concepts in rough description logics via contextual indiscernibility relations, in: Int. Workshop on Uncertainty Reasoning for the Semantic Web, volume 7123 of *LNCS*, Springer, 2013, pp. 300–314.

[6] R. Peñaloza, V. Thost, A.-Y. Turhan, Query answering for rough $\mathcal{EL}$ ontologies, in: M. Thielscher, F. Toni (Eds.), Proceedings of the 16th International Conference on Principles of Knowledge Representation and Reasoning (KR'18), AAAI Press, 2018.

[7] C. M. Keet, From granulation hierarchy to granular perspective, in: The 2009 IEEE International Conference on Granular Computing (GrC), IEEE Computer Society, 2009, pp. 306–311. doi:10.1109/GRC.2009.5255108.

[8] N. Fanizzi, C. d'Amato, F. Esposito, Conceptual clustering and its application to concept drift and novelty detection, in: S. Bechhofer, M. Hauswirth, J. Hoffmann, M. Koubarakis (Eds.), The Semantic Web: Research and Applications, 5th European Semantic Web Conference, ESWC, Proceedings, volume 5021 of *Lecture Notes in Computer Science*, Springer, 2008, pp. 318–332. doi:10.1007/978-3-540-68234-9_25.

[9] Z. Pawlak, Reasoning about data - A rough set perspective, in: Proc. of First Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC'98), volume 1424 of *LNCS*, Springer, 1998, pp. 25–34.

[10] C. M. Keet, Granulation with indistinguishability, equivalence, or similarity, in: 2007 IEEE International Conference on Granular Computing (GrC 2007), IEEE Computer Society, 2007, pp. 11–16. doi:10.1109/GRC.2007.29.

[11] S. Hirano, S. Tsumoto, Hierarchical clustering of non-euclidean relational data using indiscernibility-level, in: G. Wang, T. Li, J. W. Grzymala-Busse, D. Miao, A. Skowron, Y. Yao (Eds.), Rough Sets and Knowledge Technology, Third International Conference, RSKT, Proceedings, volume 5009 of *Lecture Notes in Computer Science*, Springer, 2008, pp. 332–339. doi:10.1007/978-3-540-79721-0_47.

[12] D. H. Fisher, Knowledge acquisition via incremental conceptual clustering, Mach. Learn. 2 (1987) 139–172. doi:10.1007/BF00114265.

[13] G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, M. Zakharyaschev, Ontology-based data access: A survey, in: J. Lang (Ed.), Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, ijcai.org, 2018, pp. 5511–5519. URL: https://doi.org/10.24963/ijcai.2018/777. doi:10.24963/IJCAI.2018/777.

[14] C. M. Keet, Ontology engineering with rough concepts and instances, in: Proc. of 17th International

Conference on Knowledge Engineering and Management by the Masses EKAW 2010, volume 6317 of *LNCS*, Springer, 2010, pp. 503–513.
[15] F. Baader, S. Brandt, C. Lutz, Pushing the $\mathcal{EL}$ envelope, in: Proc. of 18th Int. Joint Conference on Artificial Intelligence (IJCAI 2005), 2005.
[16] M. C. Klein, P. Mika, S. Schlobach, Rough description logics for modeling uncertainty in instance unification, in: Proc. of 3rd ISWC Workshop on Uncertainty Reasoning for the Semantic Web, volume 327 of *CEUR Workshop Notes*, 2007.