Analogical Reasoning in Description Logics

Claudia d'Amato, Nicola Fanizzi, and Floriana Esposito

Dipartimento di Informatica, Università degli Studi di Bari,
Campus Universitario, Via Orabona 4, 70125 Bari, Italy,
{claudia.damato,fanizzi,esposito}@di.uniba.it

Abstract. This work presents a framework, founded on multi-relational *instance-based learning*, for inductive (memory-based) reasoning on knowledge bases expressed in Description Logics. The procedure, which exploits a relational dissimilarity measure based on the notion of Information Content, can be employed both to answer to class-membership queries and to predict assertions, that may not be logically entailed by the knowledge base. These tasks may be the baseline for other inductive methods for ontology construction and evolution. In a preliminary experimentation, we show that the method is sound. Besides it is actually able to induce new knowledge that might be acquired in the knowledge base.

1 Introduction and Motivation

Most of the research on formal ontologies has been focussing on methods based on deductive reasoning. However, important tasks that are likely to be provided by new generation knowledge-based systems, such as classification, construction, revision, population and evolution are likely supported also by inductive methods.

In order to support these tasks and overcome the inherent complexity of classic logic-based inference other forms of reasoning are being investigated, both deductive, such as *non-monotonic*, *paraconsistent* [10], *approximate* reasoning (see the discussion in [11]), *case-based reasoning* [7] and inductive-analogical forms such as inductive *generalization* [5] and *specialization* [9].

All these approaches require scalable and efficient reasoning. From this viewpoint, instance-based inductive methods [8] are particularly well suited. Indeed, they are known to be both very efficient and fault-tolerant compared to the classic logic-based methods. Particularly being this methods fault-tolerant, they can be suitably applied to shared knowledge bases coming from distributed sources and for this reason often characterized by the presence of noise.

Instance-based algorithms have been mainly applied to attribute-value representations to solve tasks such as classification, clustering, and case-based reasoning. Upgrading the algorithms to work on multirelational representations [8], namely on the concept languages used in the Semantic Web, founded in Description Logics (DLs) [1] (see Sect. 2), requires novel similarity measures that are suitable for such First Order Logic fragments. As pointed out in [4], most of these measures focus on the similarity of atomic concepts within hierarchies or simple ontologies. Yet, recently, dissimilarity measures for composite concept descriptions in DL have been proposed [6]. These measures (see Sect. 3) elicit the underlying semantics by querying the knowledge base (as hinted in [3]) for determining the extension of concept descriptions (estimated by their *retrieval* [1]). Such measures can be applied to assess the dissimilarity between concepts and/or between individuals.

An instance-based framework (Sect. 4) applicable to ontological knowledge has been devised. Exploiting a dissimilarity measure, it can derive inductively (by analogy) both consistent consequences from the knowledge base and also new assertions which were not previously logically derivable. Such a framework can be effectively used to semi-automatize the task of populating ontologies that are partially defined (in terms of assertions). For instance, classification can be performed even in absence of a definition for the target concept in the knowledge base by analogy with a set of training assertions on such a concept provided by an expert. In turn, this enables also other related (bottom-up) services such as classification, clustering, ontology construction and evolution. Specifically, we have worked on a classification procedure based on a *lazy learning* approach, namely a relational form of the *k-Nearest Neighbor* (k-NN) procedure (see [12]). The main idea is that similar individuals, by analogy, should likely belong to similar concepts. The adaptation to the context of DLs could not be straightforward. Indeed a theoretical problem has been posed by the *Open World Assumption* (OWA) that is generally made in the target context, differently from data mining settings where the *Closed World Assumption* (CWA) is the standard. Besides, in the standard k-NN multi-class setting, different classes are often assumed to be disjoint, which is not typical in a Semantic Web context.

The measure and the modified classification method have been implemented and some preliminary experimental results with real ontologies have been presented (Sect. 5). Moreover, the decision procedure could be further enriched with a non-parametric statistic test for controlling the degree of significance of the classification of new instances.

2 Representation and Logic Inference

The basics of $\mathcal{ALC}$ are briefly recalled. This logic adopts constructors supported by the standard Web ontology languages (see the DL handbook [1] for a thorough reference).

In DLs, concept descriptions are defined in terms of a set N_C of *primitive concept* names and a set N_R of *primitive roles*. The semantics of the concept descriptions is defined by an *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set, the *domain* of the interpretation, and $\cdot^{\mathcal{I}}$ is the *interpretation function* that maps each $A \in N_C$ to a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and each $R \in N_R$ to $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The *top* concept $\top$ is interpreted as the whole domain $\Delta^{\mathcal{I}}$, while the *bottom* concept $\bot$ corresponds to $\emptyset$. Complex descriptions can be built in $\mathcal{ALC}$ using the following constructors. The language supports *full negation*: given any concept description C, denoted $\neg C$, it amounts to $\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$. The *conjunction* of concepts, denoted with $C_1 \sqcap C_2$, yields an extension $C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$ and, dually, concept *disjunction*, denoted with $C_1 \sqcup C_2$, yields $C_1^{\mathcal{I}} \cup C_2^{\mathcal{I}}$. Finally, there are two restrictions on roles: the *existential restriction*, denoted with $\exists R.C$, and interpreted as the set $\{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} : (x, y) \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$ and the *value restriction*, denoted with $\forall R.C$, whose extension is $\{x \in \Delta^{\mathcal{I}} \mid \forall y \in \Delta^{\mathcal{I}} : (x, y) \in R^{\mathcal{I}} \rightarrow y \in C^{\mathcal{I}}\}$.

A *knowledge base* $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ contains a *TBox* $\mathcal{T}$ and an *ABox* $\mathcal{A}$. $\mathcal{T}$ is a set of concept definitions[1] $C \equiv D$, meaning $C^{\mathcal{I}} = D^{\mathcal{I}}$, where C is atomic (the concept name) and D is an arbitrarily complex description defined as above. $\mathcal{A}$ contains assertions on the world state, e.g. $C(a)$ and $R(a, b)$, meaning that $a^{\mathcal{I}} \in C^{\mathcal{I}}$ and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$. Moreover, normally the *unique names assumption* is made on the ABox individuals. These are denoted with $\mathsf{Ind}(\mathcal{A})$. In this context the most common inference is the semantic notion of *subsumption* between concepts:

Definition 2.1 (subsumption). *given two concept descriptions C and D, C* subsumes *D, denoted by $D \sqsubseteq C$, iff for every interpretation $\mathcal{I}$ it holds that $D^{\mathcal{I}} \subseteq C^{\mathcal{I}}$. When $D \sqsubseteq C$ and $C \sqsubseteq D$ then they are* equivalent*, denoted with $C \equiv D$.*

Semantically equivalent (yet syntactically different) descriptions can be given for the same concept. Nevertheless, equivalent concepts can be reduced to a normal form by means of rewriting rules that preserve their equivalence [1]:

Definition 2.2 (normal form). *A concept description D is in $\mathcal{ALC}$* normal form *iff $D = \bot$ or $D = \top$ or if $D = D_1 \sqcup \cdots \sqcup D_n$ $(\forall i = 1, \ldots, n, D_i \not\equiv \bot)$ and*

$$D_i = \bigsqcap_{A \in \mathsf{prim}(D_i)} A \sqcap \bigsqcap_{R \in N_R} \left[\forall R.\mathsf{val}_R(D_i) \sqcap \bigsqcap_{E \in \mathsf{ex}_R(D_i)} \exists R.E \right]$$

- $\mathsf{prim}(D_i)$ *is the set of (negated) primitive concepts occurring at the top level of D_i;*
- $\mathsf{val}_R(D_i)$ *is the conjunction $C_1^i \sqcap \cdots \sqcap C_n^i$ in the value restriction of role R, if any (otherwise $\mathsf{val}_R(D_i) = \top$);*
- $\mathsf{ex}_R(D_i)$ *is the set of concepts in the existential restrictions on R.*

and $\forall R \in N_R \; \forall E \in \mathsf{ex}_R(D_i) \cup \{\mathsf{val}_R(D_i)\}$ E is in normal form.
$\mathcal{L}$ will denote the set of concepts in normal form ($\mathcal{L} = \mathcal{ALC}/_{\equiv}$).

Another inference for reasoning with individuals requires finding the concepts which an individual belongs to, especially the most specific one:

Definition 2.3 (most specific concept). *Given an ABox $\mathcal{A}$ and an individual a, the* most specific concept *of a w.r.t. $\mathcal{A}$ is the concept C, denoted $\mathit{MSC}_{\mathcal{A}}(a)$, such that $\mathcal{A} \models C(a)$ and for any other concept D such that $\mathcal{A} \models D(a)$, it holds that $C \sqsubseteq D$, where $\models$ stands for the logical entailment.*

In a language endowed with existential (or numeric) restrictions, such as $\mathcal{ALC}$, the exact MSC may not be always expressed with a finite description [1], yet it may be approximated [5, 2].

[1] The cases of general axioms or cyclic definitions will not considered here.

3 A Dissimilarity Measure for $\mathcal{ALC}$

A measure of concept similarity can be derived from the notion of *Information Content* (IC) that, in turn, depends on the probability of an individual to belong to a certain concept. Now, differently from other works which assume that a probability distribution for the concepts in an ontology is known, here it is derived from the knowledge base, that is from the distribution that can be estimated therein [3].

In order to approximate this probability for a certain concept C, its extension is used with respect to the considered ABox. Namely, we chose the *canonical interpretation* $\mathcal{I}_{\mathcal{A}}$, i.e. the one adopting the set of individuals mentioned in the ABox as its domain and the identity as its interpretation function [1]. Now, given a concept C its probability is estimated by: $pr(C) = |C^{\mathcal{I}_{\mathcal{A}}}|/|\Delta^{\mathcal{I}_{\mathcal{A}}}|$. Finally, the information content of a concept computed, employing this probability: $IC(C) = -\log pr(C)$.

A measure of the concept dissimilarity is now formally defined [6]:

Definition 3.1. *Let $\mathcal{L}$ be the set of all concepts in $\mathcal{ALC}$ normal form. The* dissimilarity measure *f is a function $f : \mathcal{L} \times \mathcal{L} \mapsto R^+$ defined recursively as follows. For all $C, D \in \mathcal{L}$, with $C = \bigsqcup_{i=1}^{n} C_i$ and $D = \bigsqcup_{j=1}^{m} D_j$*

$$f(C,D) := f_{\sqcup}(C,D) = \begin{cases} 0 & \textit{if } C \equiv D \\ \infty & \textit{if } C \sqcap D = \bot \\ \max_{\substack{i \,\in\, [1,\,n] \\ j \,\in\, [1,\,m]}} f_{\sqcap}(C_i, D_j) & \textit{o.w.} \end{cases}$$

$$f_{\sqcap}(C_i, D_j) := f_P(\mathsf{prim}(C_i), \mathsf{prim}(D_j)) + f_{\forall}(C_i, D_j) + f_{\exists}(C_i, D_j)$$

$$f_P(\mathsf{prim}(C_i), \mathsf{prim}(D_j)) := \begin{cases} \infty & \textit{if } \mathsf{prim}(C_i) \sqcap \mathsf{prim}(D_j) \equiv \bot \\ \frac{IC(\mathsf{prim}(C_i) \sqcap \mathsf{prim}(D_j))+1}{IC(\mathsf{LCS}(\mathsf{prim}(C_i), \mathsf{prim}(D_j)))+1} & \textit{o.w.} \end{cases}$$

where LCS *computes the least common subsumer of the input concepts [1].*

$$f_{\forall}(C_i, D_j) := \sum_{R \in N_R} f_{\sqcup}(\mathsf{val}_R(C_i), \mathsf{val}_R(D_j))$$

$$f_{\exists}(C_i, D_j) := \sum_{R \in N_R} \sum_{k=1}^{N} \max_{p=1,\ldots,M} f_{\sqcup}(C_i^k, D_j^p)$$

where $C_i^k \in \mathsf{ex}_R(C_i)$ and $D_j^p \in \mathsf{ex}_R(D_j)$ and we suppose w.l.o.g. that $N = |\mathsf{ex}_R(C_i)| \geq |\mathsf{ex}_R(D_j)| = M$, otherwise the indices N and M are to be exchanged.

Now f has values in $[0, \infty]$. It is possible to derive a normalized dissimilarity measure as shown in the following.

Definition 3.2 (dissimilarity measure). *The* dissimilarity measure *d is a function $d : \mathcal{L} \times \mathcal{L} \mapsto [0, 1]$, such that given the concept descriptions in normal form $C = \bigsqcup_{i=1}^{n} C_i$ and $D = \bigsqcup_{j=1}^{m} D_j$, let*

$$d(C,D) := \begin{cases} 0 & \textit{if } f(C,D) = 0 \\ 1 & \textit{if } f(C,D) = \infty \\ 1 - 1/f(C,D) & \textit{otherwise} \end{cases}$$

A measure has been defined whose baseline (counts on the extensions of primitive concepts) depends on the semantics of the knowledge base, as conveyed by the ABox assertions. This is in line with the ideas in [3, 4], where semantics is elicited as a probability distribution over the domain Δ. By comparing concept descriptions reduced to the normal form, we have a structural definition of dissimilarity. However, since MSCs are computed from the same ABox assertions, reflecting the current knowledge state, this guarantees that structurally similar representations will be obtained for semantically similar concepts.

Let us recall that it is possible to calculate the most specific concept of an individual a w.r.t. the ABox, $\mathsf{MSC}(a)$ (see Def. 2.3) or at least its approximation $\mathsf{MSC}^k(a)$ up to a certain description depth k. In the following we suppose to have fixed this k to the depth of the ABox. Given two individuals a and b in the ABox, we consider $\mathsf{MSC}^k(a)$ and $\mathsf{MSC}^k(b)$ (supposed in normal form). Now, in order to assess the dissimilarity between the individuals, the f measure can be applied:

$$d(a,b) := d(\mathsf{MSC}^k(a), \mathsf{MSC}^k(b))$$

The computational complexity of the measure is strictly related to some reasoning services, namely retrieval and instance-checking (see [1], Ch. 3). Nevertheless, it is limited to primitive concepts only and it is linear in the number of individuals in the ABox (*data complexity*). In practical applications, these computations may be efficiently carried out exploiting the statistics that are maintained by the DBMSs. Besides, the counts that are necessary for computing the concept extensions could be estimated by means of the probability distribution over the domain.

4 Nearest Neighbor for Description Logics

Given an ontology, a classification method can be employed for assigning an individual with the concepts it is likely to belong to. These individuals are supposed to be partially described by assertions in the ABox, thus classification may induce new assertions by analogy, which cannot be inferred by deduction.

We briefly review the basics of the k-Nearest Neighbor method (*k-NN*) and propose how to exploit this classification procedure for inductive reasoning. It is considered a lazy approach to learning, since the learning phase is reduced to memorizing instances of the target concepts pre-classified by an expert. Then, during the classification phase, a notion of dissimilarity for the instance space is employed to classify a new instance in analogy with its neighbor.

Let x_q be the instance that must be classified. Using a dissimilarity measure (or any other distance function), the set of the k nearest pre-classified instances w.r.t. x_q is selected. The objective is to learn a discrete-valued target function $h : IS \mapsto V$ from a space of instances IS to a set of values $V = \{v_1, \ldots, v_s\}$ standing for the classes to assign. This should be adapted for the more complex context of DLs descriptions.

In its simplest setting, the k-NN algorithm approximates h for new instances x_q on the ground of the value that h assumes for the training instances in the neighborhood of x_q, i.e. the k closest instances to the new instance in terms of a dissimilarity measure. Precisely, it is assigned according to the value which is *voted* by the majority of instances in the neighborhood.

The problem with this formulation is that it does not take into account the (dis-)similarity among instances, except when selecting the instances to be included in the neighborhood. Therefore a modified setting is generally adopted, based on weighting the vote according to the distance of the query instance from the training instances:

$$\hat{h}(x_q) := \operatorname*{argmax}_{v \in V} \sum_{i=1}^{k} w_i \delta(v, h(x_i)) \tag{1}$$

where δ is the *Kronecker symbol*, a function that returns 1 in case of matching arguments and 0 otherwise, and, given a distance measure $w_i = 1/d(x_i, x_q)$ or $w_i = 1/d(x_i, x_q)^2$.

Note that the hypothesis function $\hat{h}$ is defined only extensionally, therefore the k-NN method does not return an intensional classification model (a function or a concept definition), it merely gives an answer for new query instances to be classified, employing the procedure above. It should be observed that a strong assumption of this setting is that it can be employed to assign the query instance to the class from a set of values which can be regarded as a set of pairwise disjoint concepts. This is a simplifying assumption that cannot be always valid. In our setting, indeed, an individual could be an instance of more than one concept.

Let us consider a value set $V = \{C_1, \ldots, C_s\}$, of possibly overlapping concepts C_j $(1 \leq j \leq s)$ that may be assigned to a query instance x_q. If the classes were disjoint as in the standard setting, the decision procedure defining the hypothesis function is the same as in Eq. (1), with the query instance assigned the *single* class of the majority of instances in the neighborhood. In the general case, when the pairwise disjointness of the concepts cannot be assumed, one can adopt another classification procedure, decomposing the multi-class problem into smaller binary classification problems (one per target concept). Therefore, a simple binary value set ($V = \{-1, +1\}$) may be employed. Then, for each concept, a hypothesis $\hat{h}_j$ is computed, iteratively:

$$\hat{h}_j(x_q) := \operatorname*{argmax}_{v \in V} \sum_{i=1}^{k} \frac{\delta(v, h_j(x_i))}{d(x_q, x_i)^2} \qquad \forall j \in \{1, \ldots, s\} \tag{2}$$

where each function h_j $(1 \leq j \leq s)$ simply indicates the occurrence $(+1)$ or absence (-1) of the corresponding assertion in the ABox: $C_j(x_i) \in \mathcal{A}$. Alternately[2], h_j may return $+1$ when $C_j(x_i)$ can be inferred from the knowledge base $\mathcal{K}$, and -1 otherwise.

The problem with non-explicitly disjoint concepts is also related to the CWA usually made in the knowledge discovery context. That is the reason for adapting the standard setting to cope both with the case of generally non-disjoint classes and with the OWA which is commonly made in the Semantic Web context.

To deal with the OWA, the absence of information on whether a certain training instance x belongs to the extension of concept C_j should not be interpreted negatively, as shown before. Rather, it should count as neutral information. Thus, another value set has to be adopted for the h_j's, namely $V = \{-1, 0, +1\}$, where the three values denote,

[2] For the sake of simplicity and efficiency, this case will not be considered in the following, since instance checking may be computationally expensive (see [1], Ch. 3).

respectively, occurrence, absence and occurrence of the opposite assertion:

$$h_j(x) = \begin{cases} +1 & C_j(x) \in \mathcal{A} \\ -1 & \neg C_j(x) \in \mathcal{A} \\ 0 & o.w. \end{cases}$$

Occurrence can be easily computed with a lookup in the ABox, therefore the overall complexity of the procedure depends on the number $k \ll |\mathsf{Ind}(\mathcal{A})|$, that is the number of times the distance measure is needed.

Note that, being based on a majority vote of the individuals in the neighborhood, this procedure is less error-prone in case of noise in the data (i.e. incorrect assertions in the ABox), therefore it may be able to give a correct classification even in case of (partially) inconsistent knowledge bases.

Again, a more complex procedure may be devised by substituting the notion of occurrence (absence) of assertions in (from) the ABox with the one of derivability (non-derivability) from the whole knowledge base, i.e. $\mathcal{K} \vdash C_j(x)$, $\mathcal{K} \not\vdash C_j(x) \wedge \mathcal{K} \not\vdash \neg C_j(x)$ and $\mathcal{K} \vdash \neg C_j(x)$, respectively. Although this may exploit more information and turn out to be more accurate, it is also much more computationally expensive, since the simple lookup in the ABox must be replaced with a logical inference (instance checking).

5 Experiments

We present the outcomes of preliminary experiments carried out for testing the feasibility of the method illustrated in the previous section. In order to assess the appropriateness and validity of the method with the proposed similarity measure, we have applied it to a classification problem, using three datasets. Namely, we have chosen three different ontologies: the FSM ontology and the SURFACE-WATER-MODEL ontology from the Protégé library[3], and a handcrafted FAMILY ontology.

FSM is an ontology describing finite state machines. It is made up of 20 concepts (both primitives and defined), some of them are declared to be disjoint, 10 object properties, 7 datatype properties, 37 distinct individual names. About half of the individuals are instances of only a single class and are not involved in any property, while the other half are involved in properties.

SURFACE-WATER-MODEL is an ontology describing water quality models. It is based on the *Surface-water Models Information Clearinghouse* (SMIC) of the US Geological Survey (USGS). Namely, it is an ontology of numerical models for surface water flow and water quality simulation, which are applicable to lakes, oceans, estuaries, etc. These models are classified according to their availability, domain, dimensions, and characteristic types. It is made up of 19 concepts (both primitives and defined) without any specification about disjointness, 9 object properties, 115 distinct individual names; each of them is an instance of a single class and only some of them are involved in object properties.

FAMILY is an handcrafted ontology describing family relationships. It is made up of 14 concepts (both primitives and defined), some of them are declared to be disjoint,

[3] `http://protege.stanford.edu/download/ontologies.html`

5 object properties, 39 distinct individual names. Most of the individuals are instances of more than one concept, besides most of them are involved in more than one role. The ontology was intended to provide a more complex instance graph than in the other cases. Indeed, in the other ontologies there are assertions about instances only for some concepts (the others being defined intensionally), while in the FAMILY ontology every concept has at least one instance asserted in the ABox. The same happens for the role assertions; particularly there are some situations where roles assertions constitute a chain from an individual to another one, by means of other intermediate roles. So the FAMILY ontology likely describes a real-world state of affairs (the ABox having been modeled after a real family tree).

The proposed method was applied to each ontology, with the parameter k set to $\sqrt{|\mathsf{Ind}(\mathcal{A})|}$, as recommended in the literature based on specific experiments; namely, for every ontology, all individuals are classified to be instances of one or more concepts of the considered ontology. Specifically, we consider all the individuals in the ontology and for each of them the MSC is computed, thus the MSC list represents the set of training examples. Each example is classified applying the k-NN method for DLs, adopting the leave-one-out cross validation procedure. It is straightforward to point out that two elements are fundamental for getting good results with the method:

- the similarity measure has to be able to select really similar instances with respect to the example to be classified;
- the examples in the training set have to be meaningful for the one to be classified.

We intended to assess whether the method is able to classify instances correctly, i.e. assign the same concepts computed with instance checking. Additionally, it should also be able to induce by analogy new (previously unknown) class-membership assertions that cannot be logically inferred. Particularly, for each ontology and for each concept, three rates have been computed: *omission error rate*, *commission error rate*, *induction rate*. The omission error is related to completeness. It measures the amount of unlabelled individuals ($\hat{h}_j(x_q) = 0$) with respect to a certain concept (C_j) while it was to be classified as an instance of that class ($h_j(x_q) = 1$). The commission error is related to soundness. It measures the amount of individuals labelled as instances of the negation of the target concept ($\hat{h}_j(x_q) = -1$), while they belong to that concept ($h_j(x_q) = 1$) or vice-versa. The induction rate measures the amount of individuals labelled that were found to belong to a concept or its negation ($\hat{h}_j(x_q) = \pm 1$), while this information is not logically derivable from the knowledge base ($h_j(x_q) = 0$). Thus commission error (erroneous labelling) may be more harmful than omission error (no label assigned, also because of the OWA) for further inductive methods to be applied. A high induction rate means that the procedure was actually able to induce new assertions that are likely to be valid and can then be incorporated into the knowledge base.

By looking at Tab. 1 reporting the experimental outcomes, it is important to note that, for every ontology, the commission error was null. This means that the classifier has never made critical mistakes because no individual has been deemed as an instance of a concept while really it is an instance an disjoint class.

In particular, for the SURFACE-WATER-MODEL ontology, the predictive accuracy is 100% i.e. the omission error and induction rate are null. This means that for the SURFACE-WATER-MODEL ontology our classifier always assigns individuals to the

Table 1. Average results of the trials.

	omission error	*induction rate*	*commission error*
FSM	0	31	0
S.-W.-M.	0	0	0
FAMILY	50.93	16.85	0

correct concepts but it is also never able to induce new knowledge (indeed the induction rate is null). This is due to the fact that individuals in this ontology are all instances of the same concepts and roles, so computing their MSC, these are all very similar and so the amount of information they convey is very low.

For the same reasons, also for the FSM ontology, we have a maximal accuracy. However, differently from the previous ontology, the induction rate for the FSM ontology is not null. Since the induction rate represents assertions that are not logically deducible from the ontology and was induced by the classifier, these figures would be positive if this knowledge is correct. Particularly, in this case the increase of the induction rate has been due to the presence of some concepts that are declared to be mutually disjoint.

Results are different for the case of the FAMILY ontology, where the predictive accuracy is lower and there have been some omission errors. This is due to the way individuals are asserted as instances of the concepts. First of all, instances are more irregularly *spread* over the classes, that is they are instances of different concepts, which are sometimes disjoint. Specifically, there is a concentration of instances of some concepts. Hence the MSC approximations that were computed are very different, which reduces the possibility of matching significantly similar MSCs. Nevertheless our algorithm does not make any commission error and it is able to infer new knowledge.

Concluding, we have observed that the proposed method is able to induce new assertions in addition those that were already logically derivable from the knowledge base. Particularly, an increase in prediction accuracy was observed when the instances are homogeneously spread. Besides, the method confirmed its tolerance to noise as no commission error was observed.

6 Conclusions and Future Work

This paper explored the application of an instance-based learning method to relational representations such as DLs. A dissimilarity measure has been employed in the method that may be applied to predicting/suggesting missing information about a knowledge base individuals. The first experiments made showed that the method is effective, although its performance depends on the number (and distribution) of the available training instances. Besides, as expected, the procedure is robust to noise and never made commission errors in the experiments carried out.

Currently, we are investigating the application of statistical tests which are likely to augment the significance of the inductive conclusions drawn by the mere instance-based method. Besides, an instance-based method may be also suitable for the induction of missing values for (scalar or numeric) datatype properties of an individual as an estimate derived from the values of the datatypes for the surrounding individuals. The employed measure can be refined by introducing a weighting factor, useful for decreasing the impact of the similarity between nested sub-concepts in the descriptions on the determination of the overall value. Another natural extension may concern the definition of measures for more expressive DLs languages so to scale up the applicability of the method.

References

[1] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook*. Cambridge University Press, 2003.

[2] F. Baader and R. Küsters. Non-standard inferences in description logics: The story so far. In D. Gabbay, S. S. Goncharov, and M. Zakharyaschev, editors, *Mathematical Problems from Applied Logic. New Logics for the XXIst Century*, volume 4 of *International Mathematical Series*. Kluwer/Plenum Publishers, 2005.

[3] F. Bacchus. Lp, a logic for representing and reasoning with statistical knowledge. *Computational Intelligence*, 6:209–231, 1990.

[4] A. Borgida, T. Walsh, and H. Hirsh. Towards measuring similarity in description logics. In *Working Notes of the International Description Logics Workshop*, CEUR Workshop Proceedings, Edinburgh, UK, 2005.

[5] W. Cohen and H. Hirsh. Learning the CLASSIC description logic. In P. Torasso, J. Doyle, and E. Sandewall, editors, *Proceedings of the 4th International Conference on the Principles of Knowledge Representation and Reasoning*, pages 121–133. Morgan Kaufmann, 1994.

[6] C. d'Amato, N. Fanizzi, and F. Esposito. A dissimilarity measure for $\mathcal{ALC}$ concept descriptions. In *Proceedings of the 21st Annual ACM Symposium of Applied Computing, SAC2006*, volume 2, pages 1695–1699, Dijon, France, 2006. ACM.

[7] M. d'Aquin, J. Lieber, and A. Napoli. Decentralized case-based reasoning for the Semantic Web. In Y. Gil, V. Motta, E. Benjamins, and M. A. Musen, editors, *Proceedings of the 4th International Semantic Web Conference, ISWC2005*, number 3279 in LNCS, pages 142–155. Springer, 2005.

[8] W. Emde and D. Wettschereck. Relational instance-based learning. In L. Saitta, editor, *Proceedings of the Thirteenth International Conference, ICML96*, pages 122–130. Morgan Kaufmann, 1996.

[9] F. Esposito, N. Fanizzi, L. Iannone, I. Palmisano, and G. Semeraro. Knowledge-intensive induction of terminologies from metadata. In F. van Harmelen, S. McIlraith, and D. Plexousakis, editors, *ISWC2004, Proceedings of the 3rd International Semantic Web Conference*, volume 3298 of *LNCS*, pages 441–455. Springer, 2004.

[10] P. Haase, F. van Harmelen, Z. Huang, H. Stuckenschmidt, and Y. Sure. A framework for handling inconsistency in changing ontologies. In Y. Gil, V. Motta, E. Benjamins, and M. A. Musen, editors, *Proceedings of the 4th International Semantic Web Conference, ISWC2005*, number 3279 in LNCS, pages 353–367, Galway, Ireland, November 2005. Springer.

[11] P. Hitzler and D. Vrandečić. Resolution-based approximate reasoning for OWL DL. In Y. Gil, V. Motta, E. Benjamins, and M. A. Musen, editors, *Proceedings of the 4th International Semantic Web Conference, ISWC2005*, number 3279 in LNCS, pages 383–397, Galway, Ireland, November 2005. Springer.

[12] T. Mitchell. *Machine Learning*. McGraw-Hill, 1997.