Inductive Query Answering and Concept Retrieval Exploiting Local Models

Claudia d'Amato
Computer Science Department
University of Bari
claudia.damato@di.uniba.it

Nicola Fanizzi
Computer Science Department
University of Bari
fanizzi@di.uniba.it

Floriana Esposito
Computer Science Department
University of Bari
esposito@di.uniba.it

Thomas Lukasiewicz
Computing Laboratory
University of Oxford
thomas.lukasiewicz@comlab.ox.ac.uk

Abstract

We present a classification method, founded in the instance-based learning *and the* disjunctive version space *approach, for performing approximate retrieval from knowledge bases expressed in Description Logics. It is able to supply answers, even though they are not logically entailed by the knowledge base (e.g. because of its incompleteness or when there are inconsistent assertions). Moreover, the method may also induce new knowledge that can be employed to make the ontology population task semi-automatic. The method has been experimentally tested showing that it is sound and effective.*

1 Introduction

In the perspective of the next generation *Semantic Web* (SW) [16], many important tasks that are to be supported through automated reasoning services, such as classification, construction, revision, population, are likely to be supported by inductive methods. The *Knowledge Bases* (KBs) are expressed resorting to standard ontology languages and inference services which are ultimately especially based on representation and reasoning in *Description Logics* (DLs) [1]. However, purely (deductive) logic methods may hardly scale to the extent of the SW. This has stimulated the investigation of inductive-analogical forms such as inductive *generalization* and *specialization* [3][12][13][10].

In this work we extend an inductive instance-based method based on machine learning techniques [11] focusing on the problem of the classification of semantically annotated resources in the Semantic Web, which is closely related to the (dual) *retrieval* problem [1]. Indeed, answering to a query, namely finding the extension of a query concept, can be cast as a problem of establishing the class membership of the semantically annotated individuals in a KB. Instance-based methods are known to be very efficient and fault-tolerant compared to the classic logic-based methods. Experiments in this direction have been illustrated in [4].

Specifically, we recall our relational form of the *Nearest Neighbor* (NN) approach [11] based on the idea that similar individuals, by analogy, should likely belong to similar concepts [4]. The NN approach consists in selecting k training instances (the neighborhood) that are most similar to the query instance x_q that has to be classified. The class of x_q is determined by applying a majority voting criterion to the selected neighbors. There are various by-products of this approach: 1) it can give a better insight in the specific domain of the knowledge base; 2) it may help populating knowledge bases which is time consuming and error-prone; 3) it may trigger the tasks concept induction or revision by means of supervised and unsupervised machine learning methods [12][13][10]. For instance, distance-based clustering can be employed in order to group instances and new concept may be induced for accounting for such groups [9]. Upgrading the standard algorithms based on the NN approach to cope with multi-relational representations [7], like the concept languages used in the SW, requires the solution of the following problems: 1) the definition of novel (pseudo-)metrics for assessing the similarity of individuals that are suitable for their representations; 2) coping with the *Open World Assumption* (OWA) that is generally made on the semantics of the representations adopted in this context, rather than the *Closed World Assumption* (CWA) that is the typical machine learning and database settings; 3) rather than in the standard multi-class learning, where classes are assumed to be disjoint, one cannot assume the non-disjointness of the classes (individuals can belong to more than one concept). These problems have been solved in [4].

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.34

IEEE computer society

The basic idea of NN classification has been further extended towards another direction [8], by borrowing another form of learning: the *disjunctive version space* approach [14]. In this alternative setting, the notion of neighborhood is based on class-membership queries (rather than on a similarity criterion) performed on a training set of individuals. Specifically, an individual is considered as belonging to the neighborhood of a positive example, when it is instance of the local model (a concept description) representing the positive example and that does not cover any negative example. Each local model can be regarded as a disjunction of features that separate a positive example from negative ones. Thus the membership to the neighborhood of positive examples (w.r.t. the a target concept) gives a criterion to decide on the membership of an individual to be inductively classified. The procedure is not necessarily crisp, since a number of mistakes may be tolerated, blaming them to the noise in the data (see [8] for more details).

Extending these alternative approaches to inductive classification, a distance-based NN algorithm is proposed [5], where distances are computed by building local models. Since local models are able to represent instances, such an approach should ensure more reliable results w.r.t. the classical NN approach. In this paper we experimentally show the reliability of the method grounded on local models.

The remainder of the paper is organized as follows. In §2, the basics of the $\mathcal{ALC}$ logic are summarized. The adaptation of the NN classification procedure to DL representations is recalled in §3, while the method for inducing local models is surveyed in §4. The measure based on local models is presented in §5. An experimental evaluation of the proposed method is discussed in §6. Conclusions are drawn in §7 with an outlook on future research.

2 Basics of the Representation and Inference

Description Logics (DLs) are a family of logics (fragment of first-order logic) of different expressive power (depending on the allowed constructors by the particular logic) and endowed by a set of reasoning services. Among the others, $\mathcal{ALC}$ logic has been considered a good compromise between expressive power and computational complexity.

DLs have been adopted as theoretical counterpart of the standard Ontology representation Language (OWL). 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 $C \equiv D$, meaning $C^{\mathcal{I}} = D^{\mathcal{I}}$, where C is atomic (the concept name) and D is an arbitrarily complex description defined as above (the cases of general axioms or cyclic definitions will not considered). $\mathcal{A}$ 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 *Given two concept descriptions C and D, D* subsumes *C, denoted by $C \sqsubseteq D$, iff for every interpretation $\mathcal{I}$ it holds that $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. When $C \sqsubseteq D$ and $D \sqsubseteq C$, they are* equivalent*, denoted with $C \equiv D$.*

Another important inference is *instance checking*, that is deciding whether an individual is an instance of a concept [1]. Conversely, it may be necessary to solve the *realization problem* that requires finding the concepts which an individual belongs to, especially the most specific one:

Definition 2.2 (most specific concept) *Given ABox $\mathcal{A}$ and an individual a, the* most specific concept *of a w.r.t. $\mathcal{A}$, denoted $\mathsf{MSC}_{\mathcal{A}}(a)$, is the concept C s.t. $\mathcal{A} \models C(a)$ and for any other concept D s.t. $\mathcal{A} \models D(a)$, it holds that $C \sqsubseteq D$.*

For many non-trivial DLs, such as $\mathcal{ALC}$, the exact most specific concept may not be always expressed with a finite (non-recursive) description [1], yet it may be approximated [3][2] (which may be satisfactory for inductive approaches). Generally an approximation of the most specific concept is considered up to a certain depth p.

3 Distance-based Classification of Resources in DL Knowledge Bases

Nearest Neighbor is a lazy-learning method [11, 6]. The learning phase is reduced to simply memorizing pre-classified training instances of the target concepts. The computational effort is devoted to the inductive classification phase, where a notion of (dis)similarity between instances is employed to classify a new (query) instance. Generally this method is employed to classify tuples of (discrete or numeric) features from some n-dimensional instance space: a database table. We will extend this setting to the more complex case of DL KBs. Such an approach can be used in the DL / SW context to classify individuals w.r.t. the concepts in an ontology (or any other query concept that could be defined on the ground of the concepts in the ontology), exploiting the analogy with its neighbors [4]. Specifically, given a KB, the classification method can be employed for assigning an individual with the concepts it is likely to belong to (realization problem [1]). Due to the inherent incompleteness of the DL knowledge bases, individuals are only partially described by the assertions occurring in the ABox. Inductive classification may induce new assertions, which cannot be inferred deductively.

The classical NN approach can be formally defined as follows. Given a KB $\mathcal{K}$, let x_q be the instance that must be classified w.r.t. a concept Q, and let $\mathsf{Ind}(Q) \subseteq \mathsf{Ind}(\mathcal{A})$ be a set of training instances for Q. Using a (dis)similarity measure, the set of the k training instances that are more similar to x_q is selected. The objective of the method is to learn an estimate of a hypothesis function for the target concept membership $h : \mathsf{Ind}(Q) \mapsto V$ from a space of training instances $\mathsf{Ind}(Q)$ to a set of values $V = \{v_1, \ldots, v_s\}$ representing the multi-way classification to be decided. The algorithm approximates h for x_q on the ground of the value it assumes for the training instances in the neighborhood of x_q, i.e. the k closest training instances to x_q. Precisely, this value is estimated as the value which is (weighted) *voted* by the majority of instances in the neighborhood. Formally:

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

where where $\hat{h}$ is the estimated hypothesis function, δ returns 1 in case of matching arguments and 0 otherwise, and the weights w_i can be defined as $w_i = 1/d(x_i, x_q)$, given a dissimilarity function d.

An assumption made in this setting is that the values in V correspond to pairwise disjoint concepts. On the contrary, in a DL setting, an individual could be instance of more than one concept at the same time. In this general case, another classification procedure has to be adopted. A possible solution is the decomposition of the multi-way classification problems into smaller (binary/ternary) classification problems (one per target concept). Another problem is related to the CWA usually made in the knowledge discovery context, since, in the SW context, the OWA is usually made. To deal with the OWA, the absence of information on whether a training instance x belongs to the extension of a concept C_j should not be interpreted negatively, it should rather count as neutral information. Thus, a ternary value set $V = \{-1, 0, +1\}$ has to be adopted, where the values denote, respectively, membership, uncertainty, non-membership:

$$h(x) = \begin{cases} +1 & \mathcal{K} \vdash Q(x) \\ -1 & \mathcal{K} \vdash \neg Q(x) \\ \;\;0 & \mathcal{K} \not\vdash Q(x) \wedge \mathcal{K} \not\vdash \neg Q(x) \end{cases}$$

Since the procedure is based on a (weighted) majority vote, it 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 an answer, even when a purely logic-based approach would not be able to give the correct answer or any answer at all.

4 Building Intensional Local Models: The Disjunctive Version Space Approach

An alternative instance-based method for classifying resources in $\mathcal{ALC}$ KBs has been presented in [8], derived from the notion of *Disjunctive Version Space* [14]. Differently from the NN approach (see §3), this method determines the neighborhood of an instance w.r.t. a query concept by building intensional local models of the training instances. Specifically, an individual's neighborhood is determined by inducing a local definition for the query concept on the grounds of its examples ($\mathsf{E}^+_Q = \{x \in \mathsf{Ind}(Q) \mid \mathcal{K} \vdash Q(x)\}$) and counterexamples ($\mathsf{E}^-_Q = \{\bar{x} \in \mathsf{Ind}(Q) \mid \mathcal{K} \vdash \neg Q(\bar{x})\}$). The method can be summarized as follows.

For each positive training example $x \in \mathsf{E}^+_Q$, a *local hypothesis concept* H^x_Q is generated. Each H^x_Q has to be able to *cover* the positive example x (i.e. x has to be an instance of H^x_Q) but it does not have to cover any counterexample $\bar{x} \in \mathsf{E}^-_Q$. To fulfill such constraints, H^x_Q may be expressed as the conjunction of *maximally discriminating concepts* against each counterexample $D(x, \bar{x})$, for each $\bar{x} \in \mathsf{E}^-_Q$, namely:

$$H^x_Q = \bigsqcap_{\bar{x} \in \mathsf{E}^-_Q} D(x, \bar{x})$$

A maximally discriminating concept of an individual x against a counterexample $\bar{x}$ w.r.t. Q is a maximally specific concept (w.r.t. the subsumption ordering) x belongs to that does not cover $\bar{x}$ (i.e. $\bar{x}$ must not be an instance of $D(x, \bar{x})$). An approximation of $D(x, \bar{x})$ may be given in terms of a notion of *concept difference* between the MSC approximations related to the individuals involved:

$$D(x, \bar{x}) := \mathsf{MSC}^p(x) - \mathsf{MSC}^p(\bar{x})$$

where p is a fixed depth that may depend on the ABox depth (see §2), and the symbol $-$ denotes Teege's *difference operator* for DL descriptions [17]. In the particular case of $\mathcal{ALC}$ this difference is defined by:

$$D(x, \bar{x}) := \mathsf{MSC}^p(x) \sqcup \neg\mathsf{MSC}^p(\bar{x})$$

For each $x \in \mathsf{E}^+_Q$, the individual x_q to be classified will belong to x's neighborhood w.r.t. Q when x_q belongs to

H_Q^x. From the other perspective, the *neighbor instance set* of x_q w.r.t. Q is defined as:

$$N_Q(x_q) := \{x \in \mathsf{Ind}(Q) \mid \mathcal{K} \vdash H_Q^x(x_q)\}$$

Once that the neighborhood has been selected, the inductive classification can be performed by applying the voting procedure summarized by Eq. 1. However, a different way for determining the weights has to be defined. This is because, so far no similarity measure has been explicitly introduced. A possible setting for the weight vector components is:

$$w_Q^v := \frac{\#(v, N_Q(x_q))}{\sum_{v \in V} \#(v, N_Q(x_q))}$$

where $\#(v, N_Q(x_q)) = |\{x \in N_Q(x_q) | h(x) = v\}|$ is the count of the neighbor instances of x_q voting for value v, given the query concept Q, over the total number of training individuals belonging to Q.

The complexity of the method mainly depends from the complexity of the *instance checking* operator that is used both for determining the counterexamples of the query concept Q and for computing the (approximation of) MCSs.

The presented method is suitable for logics endowed with a notion of difference and an approximation had to be made on the construction of the MSC's.

5 Metrics Based on Local Models

Moving from the NN approach for DL KBs (see §3) and the local model approach (see §4), we propose a distance measure based on the construction of local models. Since local models are able to characterize individuals, the goal is to set up a semantic distance measure that is able to capture the semantic differences between individuals of an ontology. Performing the NN classification by exploiting such a metric should ensure more reliable and precise results w.r.t. the mere disjunctive version space approach [8].

In [4], a semantic semi-distance measure for individuals has been introduced. Here, it has also been proved that such a measure is one of the most reliable currently available. This measure is based on the notion of *Hypothesis Driven Distance* [15]. In this paper, we propose an extension of such a measure that is grounded on local models. Following some ideas borrowed from machine learning [11], individuals are evaluated on the grounds of their behavior w.r.t. a context $\mathsf{F} = \{F_1, F_2, \ldots, F_m\}$ represented by a collection of $\mathcal{ALC}$ concepts, which stands as a group of discriminating *features*. In its simplest formulation, a family of totally semantic semi-distance functions for individuals, inspired by Minkowski's norms, was defined:

Definition 5.1 (family of measures) *Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a KB. Given a set of concepts $\mathsf{F} = \{F_1, F_2, \ldots, F_m\}$ and a corresponding tuple of weights $\overline{w} = \{\overline{w}_i\}_{i=1,\ldots,m}$, $p \in \mathbb{N}$, a family $\{d_p^{\mathsf{F}}\}_{p \in \mathbb{N}}$ of functions $d_p^{\mathsf{F}} : \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mapsto [0,1]$ is defined as follows: $\forall a, b \in \mathsf{Ind}(\mathcal{A})$*

$$d_p^{\mathsf{F}}(a,b) := \left[\sum_{i=1}^{m} \overline{w}_i \mid \pi_i(a) - \pi_i(b) \mid^p \right]^{1/p}$$

where projection function π_i *is defined by:*

$$\pi_i(a) = \begin{cases} 1 & \mathcal{K} \vdash F_i(a) \\ 0 & \mathcal{K} \vdash \neg F_i(a) \\ \frac{1}{2} & otherwise \end{cases}$$

The weights $\overline{w}_i$ should reflect the impact of the single feature concept w.r.t. the overall dissimilarity. This can be given by the quantity of information conveyed by a feature, measured as its entropy. Namely, the extension of a feature F w.r.t. the whole domain of objects may be probabilistically quantified as $P_F = |F^{\mathcal{I}}|/|\Delta^{\mathcal{I}}|$ (w.r.t. the canonical interpretation $\mathcal{I}$). This can be roughly approximated with: $P_F = |\mathsf{retrieval}(F)|/|\mathsf{Ind}(\mathcal{A})|$. Considering also the probability $P_{\neg F}$ related to its negation and that related to the unclassified individuals (w.r.t. F), denoted P_U, a possible entropic measure for the feature is (see [4] for more details):

$$H(F) = -\left(P_F \log(P_F) + P_{\neg F} \log(P_{\neg F}) + P_U \log(P_U)\right)$$

The measures require membership queries by performing instance checking [1]. As an alternative, especially when a good number of assertions are available in the ABox, the measures can be approximated by defining the functions π_i based on a simple ABox look-up.

Here, the assumption made is that the feature-set F may represent a sufficient number of (possibly redundant) features that are able to really discriminate different individuals. The measure is not so useful when it is based on a concise set of features, e.g. the measure get rather coarse if each example is covered by a single feature F_i. The granularity of the measures increase with the redundancy of F and more precisely with the increasing of the number and diversity of feature F_i. At the same time, if all individuals are instances of the same features, this behavior does not raise any information for determining dissimilarity.

Considering the importance that features have in determining the dissimilarity between individuals, a new way for feature generation is proposed. Rather than simply considering the concepts defined in the KB, features are generated by building local models on the ground of the disjunctive version space approach (see §4). Given some concept Q, let us consider the set of its positive examples E_Q^+. For each $x_i \in \mathsf{E}_Q^+$, a local hypothesis concept $H_Q^{x_i}$ is generated, by considering the set of counterexamples E_Q^- w.r.t. Q, as seen in §4. Hence, the feature F_j is defined as the union of all hypothesis $H_Q^{x_i}$ built from all positive examples of Q:

$$F_j = \bigsqcup_{x_i \in E_Q^+} H_Q^{x_i}$$

If $\mathsf{C} = \{C_1, C_2, \ldots, C_m\}$ is the set of all concepts in the KB (or part of them), by iterating the feature generation process presented above for each concept, a feature set $\mathsf{F} = \{F_1, F_2, \ldots, F_m\}$ made by local models is obtained. This set would define an induced semi-distance measure according to Def. 5.1. Then one may apply the NN procedure with the metric parametrized on the constructed features, to perform approximate query answering on an ontology.

The presented feature construction procedure allows to describe training instances in a more precise way than simply considering a selection of concepts in the KB. Consequently, it should ensure a more precise evaluation of the dissimilarity between individuals and hence more reliable classification results. A tradeoff between the number of features employed and the computational effort required for computing the projection functions is likely to be found.

6 Experiments

Experiments have been carried out for testing the feasibility of the NN procedure illustrated in §3 exploiting the dissimilarity measure based on local models presented in §5. The method was tested on some ontologies online available[1] and summarized on Tab. 1. It shows that they are generally represented with more expressive logics than $\mathcal{ALC}$. This affected the construction of the *MSC* approximations, which turned out to be more general than those that could be produced in the original DLs.

The proposed method was applied to each test ontology, for inducing the features for the metric (p parameter was set equal to 1) subsequently used for classifying all individuals in an ontology w.r.t. each concept therein: each individual was considered for determining if it belonged to the the considered concept ($+1$) or not (-1), or it was neutral (0 corresponding to an unknown answer). Specifically, for each training individual, an MSC approximation was pre-computed so to build the sets of examples and counterexamples w.r.t. each query concept. A cross-validation experimental design has been adopted.

We intended to assess whether our inductive method was able to classify instances correctly and also whether it is able to induce new (previously unknown) class-membership assertions that cannot be logically inferred. Its performance was compared to the relevance determined by an expert, whose role was played by a reasoner[2].

Due to the OWA, cases were observed when, it could not be (deductively) ascertained whether a resource was relevant or not for a given query. Therefore, we introduced the following indices for the evaluation [4]:

- *match rate*: number of cases of individuals that got exactly the same classification with both definitions;
- *omission error rate*: amount of individuals for which class-membership w.r.t. the given query could not determined using the induced definition, while they actually belong (do not belong) to the query concept;
- *commission error rate*: amount of individuals found not to belong to the query concept, while they actually belong to it and vice-versa;
- *induction rate*: amount of individuals found to belong or not to belong to the query concept, while either case is not logically derivable from the KB

Tab. 2 reports experimental results. For each ontology, the average rates are reported together with their standard deviation. By looking at the table, it is important to note that, for every ontology, the match rate is quite high especially for ontologies where more individuals were available. This hints that the method becomes more accurate with a growing number of individuals. The commission error was generally null except for three ontologies: FSM, BIOPAX, FINANCIAL. Not surprisingly these cases coincide with the ontologies for which the result show the most relevant variance. Indeed on careful analysis of the outcomes they have been probably caused by the lack of examples for some concepts in these ontologies. Also the omission error rate is generally almost null with much a low variance in all cases. Besides, it is possible to note that the inductive classifier was often able to induce new knowledge (not logically derivable) for the test instances. Interestingly, for some ontologies, such as SWM, TRAINS and NEWSPAPERS an increase of induction rate compensates the decay of match rate, which can be with an inductive classifier which is less cautious and tries to provide a positive or negative answer.

7 Conclusions and Outlook

A merely deductive approach to classifying semantically annotated Web resources may fall short with real ontologies integrating distributed knowledge sources across the Web. This was the reason for investigating forms of approximate classification which may be more responsive and robust. We have presented a classification procedure for KBs expressed in DLs, grounded on *instance-based learning* and the *disjunctive version space* approach. Moreover, it may serve to predict/suggest missing information about individuals in a KB. Besides, the procedure is robust to noise and only seldom made commission errors in the experiments that have been carried out so far. It may be the basis for advanced retrieval procedures. The most immediate activity is the test of the method on classifying instance w.r.t. randomly generated queries [4]. Future work will focus on making the method more language-independent, so that it can be applied to more expressive DL languages. Moreover,

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

[2] We employed PELLET 1.5.4, which is publicly available at `http://clarkparsia.com/pellet`

Table 1. Facts concerning the ontologies involved in the experiments.

	DL	#concepts	#obj. prop.	#datatype prop.	#individuals
FSM	$\mathcal{SOF}(\mathcal{D})$	20	10	7	37
NEWSPAPERS	$\mathcal{ALCF}(D)$	29	25	28	72
SWM	$\mathcal{ALCOF}(D)$	19	9	1	115
TRAINS	$\mathcal{ALC}$	44	7	0	250
BIOPAX	$\mathcal{ALCIF}(\mathcal{D})$	41	38	33	323
HDISEASE	$\mathcal{ALCIF}(\mathcal{D})$	1498	10	15	639
NTN	$\mathcal{SHIF}(\mathcal{D})$	47	27	8	676
FINANCIAL	$\mathcal{ALCIF}$	60	16	0	1000
XGENIA	$\mathcal{ALCHI}(\mathcal{D})$	49	42	1	1987

Table 2. Average results (and standard deviations) of the performance indices in the experiments.

ontology	match rate	comm. err. rate	omission rate	induction rate
FSM	80.82±18.26	18.64±18.68	00.00±00.00	00.54±02.41
NEWSPAPERS	75.48±17.01	00.00±00.00	05.01±03.39	19.51±13.98
SWM	68.11±23.13	00.00±00.00	03.01±01.37	28.88±23.37
TRAINS	84.70±21.93	00.00±00.00	01.88±02.36	13.42±20.05
BIOPAX	78.38±23.94	21.05±24.29	00.00±00.00	00.57±02.73
HDISEASE	98.37±06.02	00.00±00.00	00.17±00.36	01.46±05.95
NTN	93.29±08.30	00.12±00.55	01.57±01.80	05.02±06.69
FINANCIAL	91.23±19.12	08.42±19.14	00.02±00.07	00.33±00.14
XGENIA	97.88±03.66	00.00±00.00	00.28±00.63	01.84±03.74

we are studying the possibility of providing, together with each individual classification, an estimate of its probability.

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 and et al., editors, *Mathematical Problems from Applied Logic. New Logics for the XXIst Century*, volume 4 of *International Mathematical Series*. Kluwer/Plenum Publishers, 2005.

[3] W. Cohen and H. Hirsh. Learning the CLASSIC description logic. In P. Torasso and et al., editors, *Proc. of the 4th Int. Conf. on the Principles of Knowledge Representation and Reasoning*, pages 121–133. Morgan Kaufmann, 1994.

[4] C. d'Amato, N. Fanizzi, and F. Esposito. Query answering and ontology population: An inductive approach. In S. Bechhofer and al., editors, *Proceedings of the 5th European Semantic Web Conference, ESWC2008*, volume 5021 of *LNCS*, pages 288–302. Springer, 2008.

[5] C. d'Amato, N. Fanizzi, and F. Esposito. Approximate classification of semantically annotated web resources exploiting pseudo-metrics induced by local models. In *To appear Proc of the Int Conf on Web Intelligence*. ACM, 2009.

[6] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley, 2nd edition, 2001.

[7] W. Emde and D. Wettschereck. Relational instance-based learning. In L. Saitta, editor, *Proceedings of the 13th International Conference on Machine Learning, ICML96*, pages 122–130. Morgan Kaufmann, 1996.

[8] N. Fanizzi, C. d'Amato, and F. Esposito. Instance-based retrieval by analogy. In *Proceedings of the 22nd Annual ACM Symposium of Applied Computing, SAC2007*, volume 2, pages 1398–1402, Seoul, South Korea, 2007. ACM.

[9] N. Fanizzi, C. d'Amato, and F. Esposito. Conceptual clustering for concept drift and novelty detection. In S. Bechhofer and al., editors, *Proc. of the European Semantic Web Conf.*, volume 5021 of *LNCS*, pages 318–332. Springer, 2008.

[10] N. Fanizzi, C. d'Amato, and F. Esposito. DL-FOIL: Concept learning in Description Logics. In F. Zeleznÿ and N. Lavrač, editors, *Proceedings of the 18th International Conference on Inductive Logic Programming, ILP2008*, volume 5194 of *LNAI*, pages 107–121, Prague, Czech Rep., 2008. Springer.

[11] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning – Data Mining, Inference, and Prediction*. Springer, 2001.

[12] L. Iannone, I. Palmisano, and N. Fanizzi. An algorithm based on counterfactuals for concept learning in the semantic web. *Applied Intelligence*, 26(2):139–159, 2007.

[13] J. Lehmann and P. Hitzler. A refinement operator based learning algorithm for the $\mathcal{ALC}$ description logic. In H. Blockeel and et al., editors, *Proc of the Int. Conf. on Inductive Logic Programming, ILP2007*, volume 4894 of *LNCS*. Springer, 2008.

[14] M. Sebag. Delaying the choice of bias: A disjunctive version space approach. In L. Saitta, editor, *Proc. of the Int. Conf. on Machine Learning, ICML96*, pages 444–452. Morgan Kaufmann, 1996.

[15] M. Sebag. Distance induction in first order logic. In S. Džeroski and N. Lavrač, editors, *Proc. of the Int. WS on Inductive Logic Programming, ILP97*, volume 1297 of *LNAI*, pages 264–272. Springer, 1997.

[16] N. Shadbolt, T. Berners-Lee, and W. Hall. The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101, 2006.

[17] G. Teege. A subtraction operation for description logics. In P. Torasso and et al., editors, *Proc. of the Int. Conf. on Principles of Knowledge Representation and Reasoning*, pages 540–550. Morgan Kaufmann, 1994.