Non-parametric Statistical Learning Methods for Inductive Classifiers in Semantic Knowledge Bases

Claudia d'Amato, Nicola Fanizzi, Floriana Esposito
Dipartimento di Informatica – Università degli studi di Bari
Via Orabona 4, 70125 Bari, Italy
{claudia.damato, fanizzi, esposito}@di.uniba.it

Abstract

This work concerns non-parametric approaches for statistical learning applied to the standard knowledge representations languages adopted in the Semantic Web context. We present methods based on epistemic inference that are able to elicit the semantic similarity of individuals in OWL knowledge bases. Specifically, a totally semantic and language independent semi-distance function is presented and from it, an epistemic kernel function for Semantic Web representations is derived. Both the measure and the kernel function are embedded into non-parametric statistical learning algorithms customized for coping with Semantic Web representations. Particularly, the measure is embedded into a k-Nearest Neighbor algorithm and the kernel function is embedded in a Support Vector Machine. The realized algorithms are used to performe inductive concept retrieval and query answering. An experimentation on real ontologies proves that the methods can be effectively employed for performing the target tasks and moreover that it is possible to induce new assertions that are not logically derivable.

1 Learning from Ontologies

The Semantic Web (SW) represents an emerging domain where business, enterprise and organization on the Web will have its own organizational model (an ontology), and knowledge intensive automated manipulations on complex relational descriptions are foreseen. Specific formal languages for knowledge representation have been designed for supporting a variety of applications in this context spanning from biological and geospatial fields to agents technology and service oriented architectures. *Description Logics* (DLs) [1], a family of languages that is endowed with well-founded semantics and reasoning services, have been adopted as the core of the ontology language (OWL[1]).

[1] http://www.w3.org/2004/OWL/

Most of the research on formal ontologies has been focused on methods based on deductive reasoning. Yet, such an approach may fail in case of noisy (and possibly inconsistent) data coming from heterogeneous sources. Inductive learning method could be effectively employed to overcome this problem. Nevertheless, the research on inductive methods and knowledge discovery applied to ontologic representations have received less attention [5, 17, 10, 9].

In this paper we propose two non-parametric statistical learning methods suited on DL representation, namely the Nearest Neighbor (henceforth NN) approach [18] and the kernel methods [19], in order to perform important inferences on semantic knowledge bases (KBs), such as concept retrieval [1] and query answering. Indeed, concept retrieval and query answering can be cast as classification problems, i.e. assessing the class-membership of the individuals in the KB w.r.t. some query concepts. Reasoning by analogy, similar individuals should likely belong to the extension of similar concepts. Moving from such an intuition, an instance-based framework (grounded on the NN approach and kernel methods) for retrieving resources contained in ontological KBs has been devised, to inductively infer (likely) consistent class-membership assertions that may not be logically derivable. As such, the resulting new (induced) assertions may enrich the KBs; indeed they can be suggested to the knowledge engineer that has only to validate them, thus making the ontology population task less time-consuming [2].

Both the NN approach and kernel methods are well known to be quite efficient and noise-tolerant and, differently from the parametric statistical learning methods, they allow the hypothesis (the model to be learnt) complexity to grow with the data. Moreover, both of them are grounded on the exploitation of a notion of (dis-)similarity. Specifically, the NN approach retrieves individuals belonging to query concepts, by analogy with the most similar training instances, namely on the grounds of the classification of the nearest ones (w.r.t. the dissimilarity measure). Kernel methods represent a family of statistical learning algorithms, in-

cluding the *support vector machines* (SVMs), that can be very efficient since they map, by means of a kernel function, the original feature space into a higher-dimensional space, where the learning task is simplified and where the kernel function implements a dissimilarity notion.

From a technical viewpoint, extending the setting of the NN and the kernel methods to the DL representation required to solve several issues: 1) a theoretical problem is posed by the *Open World Assumption* (OWA) that is generally made on the semantics of SW ontologies, differently from the typical database standard where the *Closed World Assumption* (CWA) is made; 2) both NN and kernel methods are devised for simple classifications where classes are assumed to be pairwise disjoint. This is quite unlikely in the SW context where an individual can be instance of more than one concept; 3) suitable metrics, namely (dis-)similarity measures and kernel functions, are necessary for coping with the high expressive power of DL representations; such definitions could not be straightforward.

Most of the existing measures focus on concept (dis)similarity and particularly on the (dis)similarity of atomic concepts within hierarchies or simple ontologies (see the discussion in [4]). Conversely, for our purposes, a notion of dissimilarity between *individuals* is required. Recently, dissimilarity measures for specific DL concept descriptions have been proposed [7, 8]. Although they turned out to be quite effective for the inductive tasks of interest, they are partly based on structural criteria (a notion of normal form) which determine their main weakness: they are hardly scalable to deal with standard ontology languages. To overcome these limitations, a semantic pseudo-metrics [13] is exploited. This language-independent measure assesses the dissimilarity of two individuals by comparing them on the ground of their behaviors w.r.t. a committee of features (concepts), namely those defined in the KB or that can be generated to this purpose[2]. Since kernel functions implement a notion of dissimilarity, we have derived a kernel function from the semantic pseudo-metrics[13]. As for the function from which it is derived, the kernel is language independent and it is based on the semantics of the individual as epistemically elicited from the KB w.r.t. a number of dimensions, represented by a committee[3] of discriminant concept descriptions (features).

We have embedded the pseudo-metrics in a NN algorithm and the kernel function in a kernel machine (specifically a SVM). The obtained framework have been used for performing inductive concept retrieval and query answering with both approaches. We experimentally show that the methods perform concept retrieval and query answering comparably well w.r.t. a standard deductive reasoner. Moreover, we show that the proposed framework is sometimes able to induce new concept assertions that are not logically derivable, namely the reasoner does not give any reply if an individual is instance of a certain concept or not while our framework asserts that such an individual is instance of the same considered concept (or its negation).

In the next section the reference representation is briefly summarized. In Sect. 3 the basics of the NN approach and its extension to the SW setting is analyzed. In Sect. 4 the pseudo-metrics used for performing the NN search is presented. In Sect. 5 the basics of kernel methods and kernel functions are illustrated while in Sect. 6 the kernel derived from the pseudo-metrics is illustrated. In Sect. 7 the experimental evaluation of the proposed methods is discussed. Conclusions are drawn in Sect. 8.

2 Representation and Inference

We assume that concept descriptions are defined in terms of a generic sub-language based on OWL-DL that may be mapped to DLs with the standard model-theoretic semantics (see [1] for a thorough reference). 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 axioms that define concepts. $\mathcal{A}$ contains factual assertions concerning the resources, also known as individuals. The *unique names assumption* may be made on the ABox individuals, that are represented by their URIs. The set of the individuals occurring in $\mathcal{A}$ will be denoted with $\mathsf{Ind}(\mathcal{A})$.

As regards the inference services, like all other instance-based methods, our procedures may require performing *instance-checking* [1], which roughly amounts to determine whether an individual, say a, belongs to a concept extension, i.e. whether $C(a)$ holds for a certain concept C. This service is provided proof-theoretically by a reasoner. Note that because of the OWA, a reasoner may be unable to give a positive or negative answer to a class-membership query.

3 Query Answering as Nearest Neighbor Search

Query answering boils down to determining whether a resource belongs to a (query) concept extension. Here, an alternative inductive method is proposed for retrieving the resources that likely belong to a query concept. Such a method may also be able to provide an answer even when it may not be inferred by deduction.

In *similarity search* [21] the basic idea is to find the most similar object(s) to a query one (i.e. the one that has to be classified) w.r.t. a similarity (or dissimilarity) measure. We review the basics of the k-NN method applied to the SW context [8]. The objective is to induce an approximation for a discrete-valued target hypothesis function

[2]The choice of optimal committees may be performed in advance through randomized search algorithms [13].

[3]As for the pseudo-metrics [13], also for the kernel function, the feature set can be optimally generated by means of a simulated annealing procedure[13]

$h : IS \mapsto V$ from a space of instances *IS* to a set of values $V = \{v_1, \ldots, v_s\}$ standing for the classes (concepts) that have to be predicted. Note that normally $|IS| \ll |\mathsf{Ind}(\mathcal{A})|$ i.e. only a limited number of training instances is needed.

Let x_q be the query instance whose class-membership is to be determined. Using a dissimilarity measure, the set of the k nearest (pre-classified) training instances w.r.t. x_q is selected: $NN(x_q) = \{x_i \mid i = 1, \ldots, k\}$. A k-NN algorithm approximates h for classifying x_q on the grounds of the value that h assumes for the training instances in $NN(x_q)$, i.e. the k closest instances to x_q. Precisely, the value is decided by means of a weighted majority voting procedure, namely the value is given by the most *voted* class by the instances in $NN(x_q)$ weighted by the similarity of the neighbor individual. Formally, the estimation of the hypothesis function for the query individual is defined as:

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

where δ returns 1 in case of matching arguments and 0 otherwise, and, given a dissimilarity measure d, the weights are determined by $w_i = 1/d(x_i, x_q)$.

Note that the estimate function $\hat{h}$ is defined extensionally: the basic k-NN does not return an intensional classification model (namely a function or a concept definition), it merely gives an answer for the instances to be classified. It should be also observed that this setting assigns a value to the query instance which stands for one in a set of pairwise disjoint concepts (corresponding to the value set V). In a multi-relational setting such as the SW context this assumption cannot be made in general since an individual may be instance of more than one concept.

Another issue is represented by the CWA usually made in the knowledge discovery context, opposite to the OWA characterizing the SW context. To deal with the OWA, the absence of information on whether a training instance x belongs to the extension of the query concept Q should not be interpreted negatively, as in the standard settings which adopt the CWA. Rather, it should count as neutral information. Thus, assuming the alternate viewpoint, the multi-class classification problem is transformed into a ternary one. Hence another value set has to be adopted, namely $V = \{+1, -1, 0\}$, where the three values denote: membership, non-membership, and unknown, respectively.

The task can be cast as follows: given a query concept Q, determine the membership of an instance x_q through the NN procedure (see Eq. 1) where $V = \{-1, 0, +1\}$ and the hypothesis function values for the training instances are determined as follows:

$$h_Q(x) = \begin{cases} +1 & \mathcal{K} \models Q(x) \\ -1 & \mathcal{K} \models \neg Q(x) \\ 0 & \textit{otherwise} \end{cases}$$

i.e. the value of h_Q for the training instances is determined by the entailment[4] of the corresponding assertion from the knowledge base.

Note that, being this procedure based on a majority vote of the individuals in the neighborhood, it is less error-prone in case of noise in the data (e.g. incorrect assertions) w.r.t. a purely logic deductive procedure. Therefore it may be able to give a correct classification even in case of inconsistent knowledge bases. At the same time, it should be noted that the inductive inference made by the procedure shown above is not guaranteed to be deductively valid. Indeed, inductive inference naturally yields a certain degree of uncertainty.

4 A Semantic Pseudo-Metric for Individuals

For the NN procedure, we intend to exploit a dissimilarity measure that totally depends on semantic aspects of the individuals in the knowledge base. The measure is based on the idea of comparing the semantics of the input individuals along a number of dimensions represented by a committee of concept descriptions. Indeed, on a semantic level, similar individuals should behave similarly w.r.t. the same concepts. Following the ideas borrowed from [20], totally semantic distance measures for individuals can be defined in the context of a KB. More formally, the rationale is to compare individuals on the grounds of their semantics w.r.t. a collection of concept descriptions, say $\mathsf{F} = \{F_1, F_2, \ldots, F_m\}$, which stands as a group of discriminating *features* expressed in the OWL-DL sub-language taken into account. In its simple formulation, a family of distance functions for individuals inspired to Minkowski's norms L_p can be defined as follows [13]:

Definition 4.1 (family of measures) *Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a KB. Given a set $\mathsf{F} = \{F_1, F_2, \ldots, F_m\}$ of concept descriptions, a family of dissimilarity functions $d_p^{\mathsf{F}} : \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mapsto [0, 1]$ is defined as follows:*

$$\forall a, b \in \mathsf{Ind}(\mathcal{A}) \quad d_p^{\mathsf{F}}(a, b) := \frac{1}{|\mathsf{F}|} \left[\sum_{i=1}^{|\mathsf{F}|} w_i \mid \delta_i(a, b) \mid^p \right]^{1/p}$$

where $p > 0$ and $\forall i \in \{1, \ldots, m\}$ the dissimilarity function δ_i is defined by: $\forall (a, b) \in (\mathsf{Ind}(\mathcal{A}))^2$

$$\delta_i(a, b) = \begin{cases} 0 & F_i(a) \in \mathcal{A} \wedge F_i(b) \in \mathcal{A} \\ 1 & F_i(a) \in \mathcal{A} \wedge \neg F_i(b) \in \mathcal{A} \textit{ or} \\ & \neg F_i(a) \in \mathcal{A} \wedge F_i(b) \in \mathcal{A} \\ 1/2 & \textit{otherwise} \end{cases}$$

[4] We use $\models$ to denote entailment, as computed through a reasoner.

or, model theoretically: $\forall(a,b) \in (\mathsf{Ind}(\mathcal{A}))^2$

$$\delta_i(a,b) = \begin{cases} 0 & \mathcal{K} \models F_i(a) \wedge \mathcal{K} \models F_i(b) \\ 1 & \mathcal{K} \models F_i(a) \wedge \mathcal{K} \models \neg F_i(b) \text{ or} \\ & \mathcal{K} \models \neg F_i(a) \wedge \mathcal{K} \models F_i(b) \\ 1/2 & \text{otherwise} \end{cases}$$

The model theoretic definition for the projections, requires the entailment of an assertion (instance-checking) rather than the simple ABox look-up; this can make the measure more accurate yet more complex to compute unless a KBMS is employed maintaining such information at least for the concepts in F.

It can be proved [13] that these functions have the standard properties for pseudo metrics (i.e. semi-distances [21]). This means that it cannot be proved that $d_p^{\mathsf{F}}(a,b) = 0$ iff $a = b$ (indiscernible case). anyway several methods have been proposed for avoiding this case [13].

The measures strongly depend on F. Here, the assumption that the F represents a sufficient number of (possibly redundant) features that are able to discriminate really different individuals is implicitly made. Anyway, optimal feature can be learnt by the use of a randomized optimization procedure [13]. Nevertheless, it has been experimentally shown that good results could be obtained by using the very set of both primitive and defined concepts in the KB.

Of course these approximate measures become more and more precise as the knowledge base is populated with an increasing number of individuals.

5 Kernel Methods and Kernel Functions

Kernel methods [19] represent a family of statistical learning algorithms (including the *support vector machines* (SVMs)) that have been effectively applied to a variety of tasks, recently also in domains that typically require structured representations [14, 15]. They can be very efficient since they map, by means of a *kernel function*, the original feature space into a high-dimensional space, where the learning task is simplified. Such a mapping is not explicitly performed (*kernel trick*): the usage of a positive definite kernel function (i.e. a *valid* kernel) ensures that the embedding into a new space exists and that the kernel function corresponds to the inner product in this space [19].

Two components of the kernel methods have to be distinguished: the kernel machine and the kernel function. The kernel machine encapsulates the learning task, the kernel function encapsulates the hypothesis language. In this way, an efficient algorithm for attribute-value instance spaces can be converted into one suitable for structured spaces (e.g. trees, graphs) by merely replacing the kernel function.

Kernels functions are endowed with the closure property w.r.t. many operations, one of them is the convolution [16]: kernels can deal with compounds by decomposing them into their parts, provided that valid kernels have already been defined for them.

$$k_{\text{conv}}(x,y) = \sum_{\substack{\overline{x} \in R^{-1}(x) \\ \overline{y} \in R^{-1}(y)}} \prod_{i=1}^{D} k_i(\overline{x}_i, \overline{y}_i) \qquad (2)$$

where R is a composition relationship building a single compound out of D simpler objects, each from a space that is already endowed with a valid kernel. The choice of the function R is a non-trivial task which may depend on the particular application.

On the ground of this property several kernel functions have been defined: for string representations, trees, graphs and other discrete structures [14]. In [15], generic kernels based on type construction are formalized, where types are declaratively defined. In [6], kernels parametrized on a uniform representation are introduced. Specifically, a syntax-driven kernel definition, based on a simple DL representation (the *Feature Description Language*), is given.

Kernel functions for the SW representations have also been defined [11, 3]. Scecifically, in [11] a kernel for comparing $\mathcal{ALC}$ concept definitions is intrduced. It is based on the structural similarity of the AND-OR trees corresponding to the normal form of the input concepts. This kernel is not only structural, since it ultimately relies on the semantic similarity of the primitive concepts on the leaves, assessed by comparing their extensions through a set kernel. Moreover, the kernel is applied to couples of individuals, after having lifted them to the concept level through realization operators (actually by means of approximations of the most specific concept, see [1]). In [3], a set of kernels for individuals and for the various types of assertions in the ABox (on concepts, datatype properties, object properties) are presented. They should be composed for obtaining the final kernel; anyway, it is not really specified how such separate building blocks have to be integrated.

In this paper a new kernel function for SW representation is defined: the *DL-kernel*. Differently from those defined in [11, 3], the *DL-kernel* is totally semantic and language independent. Jointly with a SVM, the *DL-Link* is used for executing a classification task in order to perform inductive concept retrieval and query answering. Note that, as for the NN approach (see Sect. 3), the SVM assumes, in its general setting, the CWA and the disjointness of the classes w.r.t. which classification is performed. In order to cope with the OWA and the multi-class classification problem (since an individual can be instance of more than one concept) characterizing the SW context, the same setting modifications of the NN approach have been performed, namely an individual is classified w.r.t. each class (concept) by the use of a ternary set of value $V = \{+1, 0, -1\}$ where 0 represents the unknown information.

6 A Family of Epistemic Kernels

In this section, we propose a family of kernels, derived from the measure presented in Sect. 4, that can be directly applied to individuals. It is parameterized on a set of features (concepts) that are used for its computation. Similarly to KFOIL, a sort of dynamic propositionalization takes place. However, in this setting the committee of concepts which are used as dimensions for the similarity function are not due to be alternate versions of the same target concept but may vary freely, reflecting contextual knowledge. The form of the kernel function resembles that of the Minkowski's metrics for vectors of numeric features:

Definition 6.1 (DL-kernel) *Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a knowledge base. Given a set of concept descriptions $\mathsf{F} = \{F_1, F_2, \ldots, F_m\}$, a family of kernel functions $k_p^{\mathsf{F}} : \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mapsto [0,1]$ is defined as follows:*

$$\forall a, b \in \mathsf{Ind}(\mathcal{A}) \quad k_p^{\mathsf{F}}(a,b) := \frac{1}{|\mathsf{F}|} \left(\sum_{i=1}^{|\mathsf{F}|} \mid \sigma_i(a,b) \mid^p \right)^{\frac{1}{p}}$$

where $p > 0$ and $\forall i \in \{1, \ldots, m\}$ the simple similarity function *σ_i is defined: $\forall a, b \in \mathsf{Ind}(\mathcal{A})$*

$$\sigma_i(a,b) = \begin{cases} 1 & (\mathcal{K} \models F_i(a) \wedge \mathcal{K} \models F_i(b)) \vee \\ & \vee (\mathcal{K} \models \neg F_i(a) \wedge \mathcal{K} \models \neg F_i(b)) \\ 0 & (\mathcal{K} \models \neg F_i(a) \wedge \mathcal{K} \models F_i(b)) \vee \\ & \vee (\mathcal{K} \models F_i(a) \wedge \mathcal{K} \models \neg F_i(b)) \\ \frac{1}{2} & otherwise \end{cases}$$

The rationale for this kernel is that similarity between individuals is decomposed along with the similarity with respect to each concept in a given committee of features (concept definitions). Two individuals are maximally similar w.r.t. a given concept F_i if they exhibit the same behavior, i.e. both are instances of the concept or of its negation. Conversely, the minimal similarity holds when they belong to opposite concepts. Because of the OWA, sometimes a reasoner cannot assess the concept-membership, hence, since both possibilities are open, we assign an intermediate value to reflect such uncertainty.

It is also worthwhile to note that this is indeed a family of kernels parameterized on the choice of the feature set. The effectiveness and also the efficiency of the measure computation strongly depends on the choice of the feature committee (*feature selection*). Optimal features can be learnt by the use of randomized optimization procedures [13, 12].

The instance-checking is to be employed for assessing the value of the σ_i functions. Yet this is known to be computationally expensive (also depending on the specific DL language of choice). Alternatively, especially for largely populated ontologies which may be the objective of mining algorithms, a simple look-up may be sufficient.

If it is required that $k(a,b) = 1 \Leftrightarrow a = b$ even though the selected features are not able to distinguish the two individuals, one might make the *unique names assumption* on the individuals occurring in the ABox $\mathcal{A}$, and employ a special additional feature based on equality: $\sigma_0(a,b) = 1$ iff $a = b$ (and 0 otherwise). Alternatively, equivalence classes might be considered instead of mere individuals.

The most important property of a kernel function is its validity.

Proposition 6.1 (validity) *Given an integer $p > 0$ and a committee of features F, the function k_p^{F} is a valid kernel.*

This result can be assessed by proving the property by showing that the function can be obtained by composing simpler valid kernels through operations that guarantee the closure w.r.t. this property [16]. Specifically, since the similarity functions σ_i ($i = 1, \ldots, n$) correspond to *matching kernels*, the property follows from the closure w.r.t. sum, multiplication by a constant and kernel multiplication.

The intermediate value used in the uncertain cases may be chosen more carefully so to reflect the inherent uncertainty related to the specific features. An alternative choice that is being experimented is related to the balance between the number of known individuals that belong to the feature concept and those that certainly belong to its negation.

7 Experimental Evaluation

The measure presented in Sect. 3 has been integrated in the NN procedure (see Sect. 3) while the *DL-Kernel* (Sect. 6) has been embedded in a SVM from the LIBSVM library[5]. Both methods have been tested by applying them to a number of retrieval and query answering problems. In the following, the results of the experiments for each method are reported.

7.1 Experiments with the Feature Committee Measure

In order to assess the validity of the k-NN algorithm presented in Sect. 3 with the measure defined in Sect. 4, a number of OWL ontologies from different domains have been considered: SURFACE-WATER-MODEL (SWM), NEWTESTAMENTNAMES (NTN) from the Protégé library[6], Semantic Web Service Discovery dataset[7] (SWSD); an ontology generated by the Lehigh University Bench-

[5] Software downloadable at: `http://www.csie.ntu.edu.tw/~cjlin/libsvm`

[6] `http://protege.stanford.edu/plugins/owl/owl-library`

[7] `https://www.uni-koblenz.de/FB4/Institutes/IFI/AGStaab/Projects/xmedia/dl-tree.htm`

Table 1. Ontologies employed for the experiments.

Ontology	DL language	#conc.	#obj. prop.	#individuals
SWM	$\mathcal{ALCOF}(D)$	19	9	115
BIOPAX	$\mathcal{ALCHF}(D)$	28	19	323
LUBM	$\mathcal{ALR}_+\mathcal{HI}(D)$	43	7	555
NTN	$\mathcal{SHIF}(D)$	47	27	676
SWSD	$\mathcal{ALCH}$	258	25	732
FINANCIAL	$\mathcal{ALCIF}$	60	17	1000

mark[8] (LUBM); BioPax glycolysis ontology[9] (BioPax) and FINANCIAL ontology[10]. Tab. 1 summarizes details concerning these ontologies.

For each ontology, 30 queries were randomly generated by composition of (2 through 8) primitive or defined concepts in each knowledge base by means of the operators of the related OWL sub-language. We employed the simplest version of the distance (d_1^{F}) with the committee of feature F made by all concepts in the knowledge base.

The parameter k was set to $\sqrt{|\mathsf{Ind}(\mathcal{A})|}$, as advised in the instance-based learning literature. Yet we found experimentally that much smaller values could be chosen, resulting in the same classification.

The performance was evaluated comparing the classifier responses to those returned by a standard reasoner[11] as a baseline, and the following indices have been considered for the evaluation:

- *match rate*: number of cases of individuals that got exactly the same classification by both classifiers with respect to the overall number of individuals;
- *omission error rate*: amount of individuals for which inductive method could not determine whether they were relevant to the query or not (namely individuals classified as belonging to the class 0) while they were actually relevant (classified as $+1$ or -1 by the standard reasoner);
- *commission error rate*: amount of individuals (analogically) found to be relevant to the query concept, while they (logically) belong to its negation or vice-versa
- *induction rate*: amount of individuals found to be relevant to the query concept or to its negation, while either case is not logically derivable from the knowledge base

[8] http://swat.cse.lehigh.edu/projects/lubm/

[9] http://www.biopax.org/Downloads/Level1v1.4/biopax-example-ecocyc-glycolysis.owl

[10] http://www.cs.put.poznan.pl/alawrynowicz/financial.owl

[11] We employed PELLET v. 1.5.1. See http://pellet.owldl.com

Table 2. k-NN outcomes: averages $\pm$ standard deviations and [min,max] intervals.

	match	commission	omission	induction
SWM	93.3 ± 10.3 [68.7;100.0]	0.0 ± 0.0 [0.0;0.0]	2.5 ± 4.4 [0.0;16.5]	4.2 ± 10.5 [0.0;31.3]
BIOPAX	99.9 ± 0.2 [99.4;100.0]	0.2 ± 0.2 [0.0;0.06]	0.0 ± 0.0 [0.0;0.0]	0.0 ± 0.0 [0.0;0.0]
LUBM	99.2 ± 0.8 [98.0;100.0]	0.0 ± 0.0 [0.0;0.0]	0.8 ± 0.8 [0.0;0.2]	0.0 ± 0.0 [0.0;0.0]
NTN	98.6 ± 1.5 [93.9;100.0]	0.0 ± 0.1 [0.0;0.4]	0.8 ± 1.1 [0.0;3.7]	0.6 ± 1.4 [0.0;6.1]
SWSD	97.5 ± 3.7 [84.6;100.0]	0.0 ± 0.0 [0.0;0.0]	1.8 ± 2.6 [0.0;9.7]	0.8 ± 1.5 [0.0;5.7]
FINANCIAL	99.5 ± 0.8 [97.3;100.0]	0.3 ± 0.7 [0.0;2.4]	0.0 ± 0.0 [0.0;0.0]	0.2 ± 0.2 [0.0;0.6]

Tab. 2 reports the outcomes in terms of these indices. Preliminarily, it is important to note that, in each experiment, the commission error was low or absent. This means that the search procedure is quite accurate: it did not make critical mistakes i.e. cases when an individual is deemed as an instance of a concept while it really is an instance of a disjoint one. Also omission error is quite low, yet it more typical over all of the ontologies that were considered. A noteworthy difference was observed for the case of the SWS knowledge base for which we find the lowest match rate and the highest variability in the results over the various concepts.

The usage of all concepts in each ontology for the set F of d_1^{F} made the measure quite accurate, which is the reason why the procedure resulted quite conservative as regards inducing new assertions. In many cases, it matched rather faithfully the reasoner decisions. The cases of induction are interesting because they suggest new assertions which cannot be logically derived by using a deductive reasoner, yet they might be used to complete a knowledge base [2], e.g. after being validated by an ontology engineer.

7.2 Experimental Evaluation of the Epistemic Kernel

In order to experimentally assess the validity of the epistemic kernel (see Def. 6.1), the instance classification task has been performed on the ontologies detailed in Tab. 1.

The classification method was applied for all the individuals in each ontology. Specifically, for each ontology, the individuals were checked to assess if they were instances of the concepts in the ontology through the SVM method and the *DL-Kernel*[12] ($p = 2$) embedded in it. A similar experi-

[12] The feature set F for computing the epistemic kernel was made by all concepts in the considered ontology (see Def. 6.1).

Table 3. Results (average and std. deviation) of the experiments on concept classification with the *DL-kernel*.

ONTOLOGY	match	induction	om. error	com. error
SWM	86.18 ± 17.55	7.98 ± 16.08	5.83 ± 4.64	0.00 ± 0.00
NTN	90.95 ± 13.14	3.99 ± 11.55	4.76 ± 7.48	0.30 ± 1.43
BIOPAX	92.05 ± 11.34	0.55 ± 2.70	0.00 ± 0.00	7.41 ± 11.38
LUBM	92.95 ± 12.11	3.53 ± 12.24	3.52 ± 4.90	0.00 ± 0.00
FINANCIAL	97.24 ± 6.75	0.32 ± 0.15	0.02 ± 0.08	2.42 ± 6.75
SWSD	98.68 ± 4.06	0.00 ± 0.00	1.32 ± 4.06	0.00 ± 0.00

Table 4. Results (average and standard deviation) of the experiments for performing query answering with the *DL-kernel*.

ONTOLOGY	match	induction	om. error	com. error
SWM	82.31 ± 21.47	9.11 ± 16.49	8.57 ± 8.47	0.00 ± 0.00
NTN	80.38 ± 17.04	8.22 ± 16.87	9.98 ± 10.08	1.42 ± 2.91
BIOPAX	84.04 ± 14.55	0.00 ± 0.00	0.00 ± 0.00	15.96 ± 14.55
LUBM	76.75 ± 19.69	5.75 ± 5.91	0.00 ± 0.00	17.50 ± 20.87
FINANCIAL	97.85 ± 3.41	0.42 ± 0.23	0.02 ± 0.07	1.73 ± 3.43
SWSD	97.92 ± 3.79	0.00 ± 0.00	2.09 ± 3.79	0.00 ± 0.00

mental setting has been considered in [3] with an exemplified version of the GALEN Upper Ontology[13]. There, the ontologies have been randomly populated and only seven concepts have been considered while no roles have been taken into account[14]. Differently from this case, we did not apply any changes on the considered ontologies.

The performance of the classifier induced by the SVM was evaluated by comparing its responses to those returned by a standard reasoner[15] used as baseline. The experiment has been performed by adopting the ten-fold cross validation procedure. The average measures obtained over all the concepts in each ontology are reported in Tab. 3, jointly with their standard deviation.

By looking at the table, it is important to note that, for every ontology, the commission error is almost null. This means that the classifier did not make critical mistakes, i.e. cases when an individual is deemed as an instance of a concept while it really is known to be an instance of another disjoint concept. In the same time it is important to note that a very high match rate is registered for every ontology. Particularly, by considering both Tab. 3 and Tab. 1, it is interesting to observe that the match rate increases with the increase of the complexity of the considered ontology. This is because the performance of a statistical method improves with the augmentation of the set of the available examples, that means that there is more information for better separating the example space.

Almost always the SVM-based classifier is able to induce new knowledge. However, a more conservative behavior w.r.t. the previous experiment has been also registered, indeed the omission error rate is not null (even if it is very close to 0). To decrease the tendency to a conservative behavior of the method, a threshold could be introduced for the consideration of the "unknown" (namely labeled with 0) training examples.

Another experiment regarded testing the SVM-based method when performing inductive concept retrieval w.r.t. new query concepts built from the considered ontology. The method has been applied to perform a number of retrieval problems applied to the considered ontologies again using the chosen SVM and the *DL-kernel* function. The experiment was quite intensive, involving the classification of all the individuals in each ontology. Specifically, the individuals were checked through the inductive procedure to assess whether they were retrieved as instances of a query concept. A number of 20 queries were randomly generated by applying the available constructors to primitive and/or defined concepts and roles from each ontology. The generated concepts had also to be satisfiable (yet they may yield no instance from the logic based retrieval). Like for the previous experiment, a ten-fold cross validation was performed for each dataset. The outcomes are reported in Tab. 4, from which it is possible to observe that the behavior of the classifier on these concepts is not very dissimilar with respect to the outcomes of the previous experiment reported in Tab. 3. These queries were expected to be harder than the previous ones which correspond to the very primitive or defined concepts for the various ontologies. Specifically, the commission error rate was low for all but two ontologies (BIOPAX and LUBM) for which some very difficult queries were generated which raised this rate beyond 10% and consequently also the standard deviation values.

By comparing these results with those obtained by the use of the NN approach (see Tab. 2) it is possible to assert that both methods showed high accuracy in performing the classification task, even if the NN method has resulted to be more accurate when a lower number of instances are available for performing the learning task.

8 Conclusions and Outlook

With the aim of going beyond the limitations of classic logic-based methods, we investigated on the application of non-prametric statistical multi-relational learning methods in the context of the SW representations. Specifically, a family of semi-distance functions and kernel functions have been defined for OWL descriptions. They have been inte-

[13] `http://www.cs.man.ac.uk/~rector/ontologies/simple-top-bio`

[14] Due to the lack of information for replicating the ontology used in [3], a comparative experiment with the proposed kernel framework cannot be performed.

[15] We employed PELLET v. 1.5.1. See `http://pellet.owldl.com`

grated respectively in a NN algorithm and a SVM for inducing a statistical classifier working with the complex representations. The resulting classifiers could be tested on inductive retrieval and classification problems.

The peculiarity of learning with ABoxes that are naturally assumed to be interpreted with an open-world semantics required a peculiar assessment of the performance of the induced classifiers, since conclusions have to deal with the chance of uncertainty: some instances cannot be attributed to a class or to its negation.

Experimentally, it was shown that performance of both classifiers are not only comparable to a standard deductive reasoner, but they are also able to induce new knowledge, which is not logically derivable. Particularly, an increase in prediction accuracy was observed for ontologies that are homogeneously populated (similar to a database).

The induced classification results can be exploited for predicting or suggesting missing information about individuals, thus completing large ontologies. Specifically, it can be used to semi-automatize the population of an ABox. Indeed, the new assertions can be suggested to the knowledge engineer that has only to validate their inclusion. This constitutes a new approach in the SW context, since the efficiency of the statistical and numerical approaches and the effectiveness of a symbolic representation have been combined.

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, B. Ganter, B. Sertkaya, and U. Sattler. Completing description logic knowledge bases using formal concept analysis. In M. Veloso, editor, *IJCAI 2007, Proc. of the 20th Int. Joint Conf. on Artificial Intelligence*, pages 230–235, 2007.

[3] S. Bloehdorn and Y. Sure. Kernel methods for mining instance data in ontologies. In K. A. et al., editor, *In Proc. of the 6th Int. Semantic Web Conference*, volume 4825 of *LNCS*, pages 58–71. Springer, 2007.

[4] A. Borgida, T. Walsh, and H. Hirsh. Towards measuring similarity in description logics. In I. Horrocks, U. Sattler, and F. Wolter, editors, *Working Notes of the International Description Logics Workshop*, volume 147 of *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. Cumby and D. Roth. On kernel methods for relational learning. In T. Fawcett and N.Mishra, editors, *Proceedings of the 20th International Conference on Machine Learning, ICML2003*, pages 107–114. AAAI Press, 2003.

[7] 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.

[8] C. d'Amato, N. Fanizzi, and F. Esposito. Reasoning by analogy in description logics through instance-based learning. In G. Tummarello, P. Bouquet, and O. Signore, editors, *Proceedings of Semantic Web Applications and Perspectives, 3rd Italian Semantic Web Workshop, SWAP2006*, volume 201 of *CEUR Workshop Proceedings*, Pisa, Italy, 2006.

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

[10] 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.

[11] N. Fanizzi and C. d'Amato. A declarative kernel for $\mathcal{ALC}$ concept descriptions. In F. Esposito, Z. W. Raś, D. Malerba, and G. Semeraro, editors, *In Proceedings of the 16th International Symposium on Methodologies for Intelligent Systems, ISMIS2006*, volume 4203 of *Lecture Notes in Computer Science*, pages 322–331. Springer, 2006.

[12] N. Fanizzi, C. d'Amato, and F. Esposito. Evolutionary conceptual clustering of semantically annotated resources. In *Proceedings of the IEEE International Conference on Semantic Computing, ICSC2007*, pages 783–790. IEEE, 2007.

[13] N. Fanizzi, C. d'Amato, and F. Esposito. Induction of optimal semi-distances for individuals based on feature sets. In D. Calvanese, E. Franconi, V. Haarslev, D. Lembo, B. Motik, A.-Y. Turhan, and S. Tessaris, editors, *Working Notes of the 20th International Description Logics Workshop, DL2007*, volume 250 of *CEUR Workshop Proceedings*, Bressanone, Italy, 2007.

[14] T. Gärtner. A survey of kernels for structured data. *SIGKDD Explorations*, 5(1):49–58, 2003.

[15] T. Gärtner, J. Lloyd, and P. Flach. Kernels and distances for structured data. *Machine Learning*, 57(3):205–232, 2004.

[16] D. Haussler. Convolution kernels on discrete structures. Technical Report UCSC-CRL-99-10, Department of Computer Science, University of California – Santa Cruz, 1999.

[17] J.-U. Kietz and K. Morik. A polynomial approach to the constructive induction of structural knowledge. *Machine Learning*, 14(2):193–218, 1994.

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

[19] B. Schölkopf and A. Smola. *Learning with Kernels*. The MIT Press, 2002.

[20] M. Sebag. Distance induction in first order logic. In S. Džeroski and N. Lavrač, editors, *Proceedings of the 7th International Workshop on Inductive Logic Programming, ILP97*, volume 1297 of *LNAI*, pages 264–272. Springer, 1997.

[21] P. Zezula, G. Amato, V. Dohnal, and M. Batko. *Similarity Search – The Metric Space Approach*. Advances in database Systems. Springer, 2007.