Evidential Nearest-Neighbors Classification for Inductive ABox Reasoning

Nicola Fanizzi, Claudia d'Amato, and Floriana Esposito

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

Abstract. In the line of our investigation on inductive methods for Semantic Web reasoning, we propose an alternative way for approximate ABox reasoning based on the analogical principle of the nearest-neighbors. Once neighbors of a test individual are selected, a combination rule descending from the Dempster-Shafer theory can join together the evidence provided by the neighbor individuals. We show how to exploit the procedure for determining unknown class- and role-memberships or fillers for datatype properties which may be the basis for many further ABox inductive reasoning algorithms.

1 Introduction

In the context of reasoning in the Semantic Web (SW), a growing interest is being committed to alternative procedures extending the standard methods so that they can deal with the various facets of uncertainty related with Web reasoning [1]. Extensions of the classic probability measures [2] offer alternative ways to deal with inherent uncertainty of the knowledge bases (KBs) in the SW. Particularly, belief and plausibility measures adopted in the *Dempster-Shafer Theory of Evidence* [3] have been exploited as means for dealing with incompleteness [4] and also inconsistency [5], which may arise from the aggregation of data and metadata on a large and distributed scale. In this work we undertake again the inductive point of view. Indeed, in many SW domains a very large number of assertions can potentially be true but often only a small number of them is known to be true or can be inferred to be true. So far the application of combination rules related to the Dempster-Shafer theory has concerned the induction of metrics which are essential for all similarity-based reasoning methods [4]. One of the applications of such measures was related to the prediction of assertions through nearest neighbor procedures. Recently a general-purpose evidential nearest neighbor procedure based on the Dempster-Shafer combination rule has been proposed [6]. In this work this method is extended to the specific case of semantic KBs through a more epistemically appropriate combination procedure [7]. In the perspective of inductive methods, the need for a definition of a semantic similarity measure for *individuals* arises, that is a problem that so far received less attention in the literature compared to the measures for concepts.

Recently proposed dissimilarity measures for individuals in specific languages founded in *Description Logics* [8] turned out to be practically effective for the targeted inductive tasks [9], however they are still based on structural criteria so that they can hardly scale to more complex languages. We devised families of dissimilarity measures for semantically annotated resources, which can overcome the aforementioned limitations [10, 11]. Our measures are mainly based on the Minkowski's norms for Euclidean spaces induced by means of a method developed in the context of relational *machine learning* [12]. Namely, the measures are based on the degree of discernibility of the input individuals with respect to a given context [13] (or committee of features), which are represented by concept descriptions expressed in the language of choice.

The main contributions of this work regard the extension of a framework for the classification of individuals through a prediction procedure based on evidence theory and similarity. In particular we propose using Yager's rule of combination and exploiting the mentioned families of metrics defined for individuals in ontologies. This allows for measuring the confirmation of the truth of candidate assertions. The prediction of the values (related to class-membership or datatype and object properties) may have plenty of applications in uncertainty reasoning with ontologies.

The remainder of the paper is organized as follows. In the next section (§2), distance measures that shall be utilized for selecting neighbor individuals are introduced. Then (§3), the basics of the Dempster-Shafer theory and a nearest-neighbor procedure based on an alternative rule of combination are recalled. Hence (§4) we present the applications of the method to the problems of determining the class- or role-membership of individuals w.r.t. given query concepts / roles as well as the prediction of fillers for datatype properties. Relevant related work are discussed in (§5) and we conclude (§6) proposing extensions and applications of these methods in further works.

2 Dissimilarity Measures for Individuals

Since the reasoning method to be presented in the following is intended to be general purpose, no specific language will be assumed in the following for resources, concepts (classes) and their properties. It suffices to consider a generic representation that can be mapped to some Description Logic language with the standard model-theoretic semantics (see [8] for a thorough reference).

A *knowledge base* $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ comprises a *TBox* $\mathcal{T}$ and an *ABox* $\mathcal{A}$. $\mathcal{T}$ is a set of axioms concerning the (partial) definition of concepts (and roles) through class (role) expressions. $\mathcal{A}$ contains assertions (ground facts) concerning the world state. The set of the individuals occurring in $\mathcal{A}$ will be denoted with $\mathsf{Ind}(\mathcal{A})$. Each individual can be assumed to be identified by its own URI (it is useful in this context to make the *unique names assumption*).

Similarity-based tasks, such as individual classification, retrieval, and clustering require language-independent measures for individuals whose definition can capture semantic aspects of their occurrence in the knowledge base [10, 11].

For our purposes, we need functions to assess the similarity of individuals. However individuals do not have an explicit syntactic (or algebraic) structure that can be compared (unless one resorts to language-specific notions [9], such as the *most specific concept* [8]). Focusing on the semantic level, the leading idea may be that, similar individuals should behave similarly w.r.t. the same concepts. A way for assessing the similarity of individuals in a knowledge base can be based on the comparison of their semantics along a number of dimensions represented by a set of concept descriptions (henceforth referred to as the *committee* or *context* [13]). Specifically, the measure may compare individuals on the grounds of their behavior w.r.t. a given context, say $\mathsf{C} = \{C_1, C_2, \ldots, C_m\}$, which stands as a group of discriminating relevant concepts (*features*) expressed in the considered language. We begin with defining the behavior of an individual w.r.t. a certain concept in terms of projecting it in this dimension: Given a concept $C_i \in \mathsf{C}$, the related *projection function* $\pi_i : \mathsf{Ind}(\mathcal{A}) \mapsto \{0, \frac{1}{2}, 1\}$ is defined:

$$\forall a \in \mathsf{Ind}(\mathcal{A}) \qquad \pi_i(a) = \begin{cases} 1 & \mathcal{K} \models C_i(a) \\ 0 & \mathcal{K} \models \neg C_i(a) \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

The case of $\pi_i(a) = \frac{1}{2}$ corresponds to the case when a reasoner cannot give the truth value for a certain membership query. This is due to the *Open World Assumption* normally made in Semantic Web reasoning. Hence, as in the classic probabilistic models, uncertainty may be coped with by considering a uniform distribution over the possible cases. Further ways to approximate these values in case of uncertainty are investigated in [4].

The discernibility functions related to the context w.r.t. which two input individuals are compared are defined as follows. Given a feature concept $C_i \in \mathsf{C}$, the related *discernibility function* $\delta_i : \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mapsto [0, 1]$ is defined as:

$$\forall (a, b) \in \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \qquad \delta_i(a, b) = |\pi_i(a) - \pi_i(b)|$$

The discernibility function δ_i assigns 0 if the two individuals a and b have the same behavior w.r.t. C_i, that is if they are both instance of C_i or both instance of $\neg C_i$ or nothing is known about this. This is because, if a and b have the same bahavior w.r.t. C_i then there are no other information for discriminating them.

Finally, a family of dissimilarity measures for individuals that is inspired to the Minkowski's metrics can be defined [10, 11]: Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a knowledge base. Given a context C and a related vector of weights $\boldsymbol{w}$, a family of dissimilarity measures $\{d_p^{\mathsf{C}}\}_{p \in \mathbb{N}}$, $d_p^{\mathsf{C}} : \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mapsto [0, 1]$ is defined as follows:

$$\forall (a, b) \in \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \qquad d_p^{\mathsf{C}}(a, b) = \left[\sum_{C_i \in \mathsf{C}} w_i \delta_i(a, b)^p \right]^{\frac{1}{p}}$$

The effect of the weights[1] is to normalize w.r.t. the other features involved. Obviously these measures are not absolute, then they should be also considered

[1] A possible way for determining the w_i is to assign a high value if the corresponding feature concept reflects high *information content*, low value otherwise (see [10] for more details).

w.r.t. the context of choice, hence comparisons across different contexts may not be meaningful. Larger contexts are likely to decrease the measures because of the normalizing factor yet these values is affected also by the degree of redundancy of the features employed. In other works the choice of the weights is done according to variance or entropy associated to the various concepts in the context [10, 11].

Compared to other proposed measures [14, 9, 15], the presented functions do not depend on the constructors of a specific language, rather they require only (retrieval or) instance-checking for computing the projections through class-membership queries to the knowledge base. The complexity of measuring the dissimilarity of two individuals depends on the complexity of such inferences (see [8], Ch. 3). Note also that the projections that determine the measure can be computed (or derived from statistics maintained on the knowledge base) before the actual distance application, thus determining a speed-up in the computation of the measure. This is very important for algorithms that massively use this distance, such as instance-based methods.

One should assume that C represents a set of (possibly redundant) features that are able to discriminate individuals that are actually different. The choice of the concepts to be included (a *feature selection* problem [12]) may be crucial. Therefore, specific optimization algorithms founded in *randomized search* have been devised which are able to find optimal choices of discriminating contexts [10, 11]. However, the results obtained so far with knowledge bases drawn from ontology libraries showed that (a selection) of the primitive and defined concepts are often sufficient to induce sufficiently discriminating measures.

3 Evidence-Theoretic Nearest-Neighbor Prediction

In this section the basics of the theory of evidence and combination rules [3] are recalled then a nearest neighbor classification procedure based on the rule of combination [6] is extended in order to perform prediction of unobserved values (related to datatype properties or also class-membership).

3.1 Basics of the Evidence Theory

In the Dempster-Shafer theory, a *frame of discernment* Ω is defined as the set of all hypotheses in a certain domain. Particularly, in a classification problem it is the set of all possible classes. A *basic belief assignment* (BBA) is a function m that defines a mapping $m : 2^{\Omega} \mapsto [0, 1]$ verifying: $\sum_{A \in \Omega} m(A) = 1$. Given a certain piece of evidence, the value of the BBA for a given set A expresses a measure of belief that is committed exactly to A. The quantity $m(A)$ pertains only to A and does not imply any additional claims about any of its subsets. If $m(A) > 0$, then A is called a *focal element* for m.

The BBA m cannot be considered a proper probability measure: it is defined over 2^{Ω} instead of Ω and it does not require the properties of monotone measures [2]. The BBA m and its associated focal elements define a *body of evidence*, from which a *belief function* Bel and a *plausibility function* Pl can be derived as mappings from 2^Ω to $[0, 1]$. For a given $A \subseteq \Omega$, the *belief* in A, denoted $Bel(A)$, represents a measure of the total belief committed to A given the available evidence. *Bel* is defined as follows:

$$\forall A \in 2^\Omega \qquad Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \tag{1}$$

Analogously, the plausibility of A, denoted $Pl(A)$, represents the amount of belief that could be placed in A, if further information became available. *Pl* is defined as follows:

$$\forall A \in 2^\Omega \qquad Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \tag{2}$$

It is easy to see that: $Pl(A) = Bel(\Omega) - Bel(\bar{A})$. Moreover $m(\emptyset) = 1 - Bel(\Omega)$ and for each $A \neq \emptyset$: $m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} Bel(B)$. Using these equations, knowing just one function among m, *Bel*, and *Pl* allows to derive the others.

The Dempster-Shafer rule of combination [3] is an operation for pooling evidence from a variety of sources. This rule aggregates independent bodies of evidence defined within the same frame of discernment into one body of evidence. Let m_1 and m_2 be two BBAs. The new BBA obtained by combining m_1 and m_2 using the rule of combination, m_{12} is the orthogonal sum of m_1 and m_2. Generally, the normalized version of the rule is used:

$$\forall A \in 2^\Omega \setminus \{\emptyset\} \qquad m_{12}(A) = (m_1 \oplus m_2)(A) = \frac{\sum_{B \cap C = A} m_1(B)\, m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B)\, m_2(C)}$$

(and $m_{12}(\emptyset) = 0$) where the numerator $(1 - c)$ normalizes the values of the combined BBA w.r.t. the amount of conflict c between m_1 and m_2.

Different evidence fusion rules have been proposed [2]. A more epistemologically sound combination rule [7] for our purposes places the probability mass related to the conflict between the BBAs to the case of maximal ignorance.

$$\forall A \in 2^\Omega \qquad m_{12}(A) = \begin{cases} \sum_{B \cap C = A} m_1(B)\, m_2(C) & A \neq \Omega \wedge A \neq \emptyset \\ m_1(\Omega)\, m_2(\Omega) + c & A = \Omega \\ 0 & A = \emptyset \end{cases}$$

This means that the conflict between the two sources of evidence is not hidden, but it is explicitly recognized as a contributor to ignorance.

Due to the associativity and commutativity of the operations involved, it is easy to prove that the resulting combination operator is associative and commutative, and admits the vacuous BBA (Ω unique focal set) as neutral element.

3.2 The Nearest Neighbors Procedure

Let us consider the finite set of instances X and a finite set of integers $V \subseteq \mathbb{Z}$ to be used as labels (which may correspond to disjoint classes or distinct attribute values). The available information is assumed to consist in a training set $\mathsf{TrSet} =$

$\{(x_1, v_1), \ldots, (x_M, v_M)\} \subseteq \mathsf{Ind} \times V$ of single-labeled instances (*examples*). In our case, $X = \mathsf{Ind}(\mathcal{A})$, the set of individual names occurring in the ontology.

Let x_q be a new individual to be classified on the basis of its nearest neighbors in TrSet. Let $N_k(x_q) = \{(x_{o(j)}, v_{o(j)}) \mid j = 1, \ldots, k\}$ be the set of the k nearest neighbors of x_q in TrSet sorted by a function $o(\cdot)$ depending on an appropriate metric d which can be applied to ontology individuals (e.g. one of the measures in the family defined in the previous section §2).

Each pair $(x_i, v_i) \in N_k(x_q)$ constitutes a distinct item of evidence regarding the value to be predicted for x_q. If x_q is close to x_i according to d, then one will be inclined to believe that both instances are associated to the same value, while when $d(x_q, x_i)$ increases, this belief decreases and that leads to a situation of almost complete ignorance concerning the value to be predicted for x_q.

Consequently, each $(x_i, v_i) \in N_k(x_q)$ may induce a BBA m_i over V which can be defined as follows [6]:

$$\forall A \in 2^V \qquad m_i(A) = \begin{cases} \lambda\sigma(d(x_q, x_i)) & A = \{v_i\} \\ 1 - \lambda\sigma(d(x_q, x_i)) & A = V \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $\lambda \in]0, 1[$ is a parameter and $\sigma(\cdot)$ is a decreasing function such that $\sigma(0) = 1$ and $\lim_{d\to\infty} \sigma(d) = 0$ (e.g. $\sigma(d) = \exp(-\gamma d^n)$ with $\gamma > 0$ and $n \in \mathbb{N}$). The values of the parameters can be determined heuristically.

Considering each training individual in $N_k(x_q)$ as an separate source of evidence, k BBAs m_j are obtained. These can be pooled by means of the rule of combination leading to the aggregated BBA m that synthesizes the final belief:

$$\bar{m} = \bigoplus_{j=1}^{k} m_j = m_1 \oplus \cdots \oplus m_k \tag{4}$$

In order to predict a value, functions $\overline{Bel}$ and $\overline{Pl}$ can be derived from $\bar{m}$ using the equations seen above, and the query individual x_q is assigned the value in V that maximizes the belief or plausibility:

$$v_q = \operatorname*{argmax}_{(x_i,v_i)\in N_k(x_q)} \overline{Bel}(\{v_i\}) \quad \text{or} \quad v_q = \operatorname*{argmax}_{(x_i,v_i)\in N_k(x_q)} \overline{Pl}(\{v_i\})$$

The former choice (select the hypothesis with the greatest degree of belief the most credible) corresponds to a *skeptical* viewpoint while the latter (select the hypothesis with the lowest degree of doubt the most plausible) is more *credulous*. The degree belief (or plausibility) of the predicted value provides also a way to compare the answers of an algorithm built on top of such analogical procedure. This is useful for tasks such as ranking, matchmaking, etc..

Finally, it is possible to combine the two measures *Bel* and *Pl* analogously to necessity (*Nec*) and possibility (*Pos*) in *Possibility Theory* (which can be considered a special case[2] of Dempster-Shafer theory). One can define a single

[2] Precisely, the body of evidence must contain *consonant* focal sets, i.e. when the set of focal elements is a nested family [2].

$ENN_k(x_q, \mathsf{TrSet}, V)$

1. Compute the neighbor set $N_k(x_q) \subseteq \mathsf{TrSet}$.
2. **for each** $i \leftarrow 1$ **to** k **do**
   Compute m_i (Eq. 3)
3. **for each** $v \in V$ **do**
   Compute $\bar{m}$ (Eq. 4) and derive $\overline{Bel}$ and $\overline{Pl}$ (Eqs. 1–2)
   Compute the confirmation $\overline{C}$ (Eq. 5) from $\overline{Bel}$ and $\overline{Pl}$
4. Select $v \in V$ that maximizes $\overline{C}$ (Eq. 6).

Fig. 1. The evidence nearest neighbor procedure.

measure of *confirmation* C, ranging in $[-1, +1]$, by means of a simple one-to-one transformation [2]:

$$\forall A \subseteq \Omega \qquad C(A) = Bel(A) + Pl(A) - 1 \tag{5}$$

Hence, denoted with $\overline{C}$ the combination of $\overline{Bel}$ and $\overline{Pl}$, the resulting rule for predicting the uncertain value for the test individual can be written as follows:

$$v_q = \operatorname*{argmax}_{(x_i,v_i)\in N_k(x_q)} \overline{C}(\{v_i\}) \tag{6}$$

Summing up, the procedure is as reported in Fig. 1:

It is worthwhile to note that the complexity of the method is polynomial in the number of instances in the TrSet. If this set is compact and contains very prototypical individuals with plenty of related assertions, then the resulting predictions are likely to be accurate. Another source of complexity in the computations may be the number of values in V which may yield a large number of subsets $2^{|V|}$ for which BBAs are to be computed. However this depends also on the kind of problem that is to be solved (e.g. in class membership detection $|V| = 2$). Moreover what really matters in the number of focal sets for each BBA which may be much less than $2^{|V|}$.

4 Assertion Prediction

The utility of the presented procedure when applied to ontology reasoning can be manifold. In the following we propose its employment in the inductive prediction of unknown values related to class-membership and datatype / object property fillers. This feature may be easily embedded in an ontology management system in order to help the knowledge engineers elicit assertions which may be not be derived from the knowledge base yet they can be rather made in analogy with the others [9].

In the following, the symbol $\models\!\!\!\approx$ in expressions like $\mathcal{K} \models\!\!\!\approx \alpha$ will denote the derivation of the assertion α from the knowledge base $\mathcal{K}$ obtained through an alternative procedure (like the evidence nearest neighbor presented in the previous section).

4.1 Class-Membership

Let us suppose a (query) concept Q is given. In this case one may consider only examples made up of individuals with a definite class-membership leading to a binary problem with a set of values $V_Q = \{+1, -1\}$ denoting, resp., membership and non-membership w.r.t. the query concept. Alternatively, one may admit ternary problems with some labels set to 0 to explicitly denote an indefinite (uncertain) class-membership [9, 10]. We shall also consider the related training set $\mathsf{TrSet}_Q \subseteq \mathsf{Ind}(\mathcal{A}) \times V_Q$. The values of the labels v_i for the training examples can be obtained through deductive reasoning (instance-checking) or specific facilities made available by the knowledge management systems [16].

Now to predict the class-membership value v_q for some individual x_q w.r.t. Q, it suffices to call the procedure $\mathit{ENN}_k(x_q, \mathsf{TrSet}_Q, V_Q)$ and decide on the grounds of the returned value. Thus in a binary setting ($V_Q = \{+1, -1\}$), one will either conclude that $\mathcal{K} \approx Q(x_q)$ or $\mathcal{K} \approx \neg Q(x_q)$ depending on the value that maximizes $\overline{C}$ in Eq. 6 (resp., $v_q = +1$ or $v_q = -1$). Moreover the value of the confirmation function which determined the returned value v_q can be exploited for ranking the hits by comparing the strength of the inductive conclusions.

Adopting a ternary setting, it may turn out that the most likely value is $v_q = 0$ resulting in an uncertain case. One may force the choice among the values of $\overline{C}$ for $v_q = -1$ and $v_q = +1$, e.g. when the confirmation degree exceeds a some threshold.

The inductive procedure described above can be trivially exploited for performing the retrieval of a certain concept inductively. Given a certain concept Q, it would suffice to find all individuals $a \in \mathsf{Ind}(\mathcal{A})$ that are such that $\mathcal{K} \approx Q(a)$. The hits could be returned ranked by the respective confirmation value $\overline{C}(+1)$.

4.2 Datatype Fillers

In this case, let us suppose a certain (functional) datatype property P is given and the problem is to predict its value for a certain test individual a (which has to be supposed to be in its domain). The set of values V_P may correspond to the (discrete and finite) range of the property or to its restriction to the observed values for the training instances: $V_P = \{v \in \mathit{range}(P) \mid \exists P(a, v) \in \mathcal{A}\}$. Different settings may be devised allowing for some special value(s) denoting the case of a yet unobserved value(s) for that property.

The related training set will be some $\mathsf{TrSet}_P \subseteq \mathit{domain}(P) \times V_P$, where $\mathit{domain}(P) \subseteq \mathsf{Ind}(\mathcal{A})$ is the set of individual names that have a known P-value in the knowledge base. Differently from the previous problem, datatype properties generally do not have a specific intensional definition in the knowledge base (except for the specification of domain and range), hence a mere look-up in the ABox should suffice to determine the TrSet.

Now to predict the value in V_P of the datatype property P for some individual a, the method requires calling the procedure with $\mathit{ENN}_k(a, \mathsf{TrSet}_P, V_P)$. Thus in this setting, if v_q is the value that maximizes Eq. 6 then we can write $\mathcal{K} \approx P(a, v_q)$. Also in this case the value of the confirmation function which determined choice of the value v_q can be exploited for comparing the strength of an inductive conclusion to others.

In case of special settings with dummy values indicating unobserved values, when these are found to be the most credible among the others, a knowledge engineer should be contacted for the necessary changes to the ontology.

The inductive procedure described above can be trivially exploited for performing alternate forms of retrieval, e.g. finding all individuals with a certain value for the given property. Given a certain value v, it would suffice to find all individuals $a \in \mathsf{Ind}(\mathcal{A})$ that are such that $\mathcal{K} \approx P(a, v)$. Again, the hits could be returned ranked according to the respective confirmation value $\overline{C}(+1)$.

The limitation of treating only functional datatype properties may be overcome by considering a different way to assign the probability mass to BBAs than Eq. 3, including subsets of all possible values. Examples are to be constructed accordingly (labels will be chosen in 2^{V_P}). Alternatively, more complex frames of discernment, e.g. $\Omega' = 2^{\Omega}$, so consider sets of values as possible fillers of the property. In all such settings the computation of the BBAs and descending measures would become of course much more complex and expensive, yet clever solutions (or approximations) proposed in the literature [6] may contribute to mitigate this problem.

4.3 Relationships among Individuals

In principle, a very similar setting may be used in order to establish the possibility that a certain test individual is related through some object property with some other individual [17, 18].

Since the set $\mathsf{Ind}(\mathcal{A})$ is finite (the target is not discovering relations with unseen individuals), one may want to find all individuals that are related to a test one through some object property, say R. The problem can be decomposed into smaller ones aiming at verifying whether $\mathcal{K} \approx R(a, b)$ holds:

for each $b \in \mathsf{Ind}(\mathcal{A})$ **do**
  for each $a \in \mathsf{Ind}(\mathcal{A})$ **do**
    $\mathsf{TrSet} \leftarrow \{(x, v) \mid x \in \mathsf{Ind}(\mathcal{A}) \setminus \{a\}, \textit{if } \mathcal{K} \models R(x, b) \textit{ then } v \leftarrow +1 \textit{ else } v \leftarrow -1\}$
    $v_b^R \leftarrow \mathit{ENN}_k(a, \mathsf{TrSet}, \{+1, -1\})$
    if $v_b^R = +1$ **then**
      return $\mathcal{K} \approx R(a, b)$
    else
      return $\mathcal{K} \approx \neg R(a, b)$

Note that, in the construction of the training sets, the inference $\mathcal{K} \models R(x, b)$ may turn out to be merely an ABox lookup operation for the given assertions (when roles are not intensionally defined in a proper RBox). Conversely, if an RBox is available (sometimes as a subset of the TBox) the values of the label for the training examples can be obtained through deductive reasoning (instance-checking) or the mentioned facilities made available by advanced reasoners or knowledge management systems [16].

This simple setting makes a sort of *closed-world assumption* in the decision of the induced assertions descending from the adoption of the binary value set and the composition of the TrSet. A more cautious setting would involve a ternary value set $V_R = \{-1, 0, +1\}$ which allows for an explicit treatment of those individuals a for which $R(a, b)$ is not derivable (or just absent from the ABox). The final decision on the induced conclusion has to consider also this new possibility (e.g. using a threshold of confirmation for accepting likely assertions).

5 Related Work

The proposed method is related to those approaches devised to offer alternative ways of reasoning with ABoxes for eliciting hidden knowledge (regularities) in order to complete and populate the ontology with likely assertions even in the occurrence of incorrect parts, supposing this kind of *noise* is not systematic.

The tasks of ontology completion and population have often been tackled through formal methods (such as *formal concept analysis* [19]). Discovering new assertions (and related probabilities in a classical setting) is another related task for eliciting hidden knowledge in the ontologies. In [18] a machine learning method is proposed to estimate the truth of statements by exploiting regularities in the data. In [17] another statistical learning method for OWL-DL ontologies is proposed, combining a latent relational graphical model with Description Logic inference in a modular fashion. The probability of unknown role-assertions can be inductively inferred and known concept-assertions can be analyzed by clustering individuals.

Similarity-based reasoning with ontologies is the primary aim of this work which follows a number of related methods founded on dissimilarity measures for individuals in knowledge bases expressed in Description Logics [9, 10]. Mostly, they adopt some alternate form of the classic Nearest-Neighbor lazy learning scheme [12] in order to draw inductive conclusions that often cannot be deductively entailed by the knowledge bases.

Similar approaches based on lazy learning have been proposed that adopt generalized probability theories such as the Dempster-Shafer. In [6], which was a source of inspiration for this paper, the standard rule of combination is exploited in an evidence-theoretic classification procedure where labels were not assumed to be mutually exclusive. Rules of combination had been used in [4] in order to learn precise metrics to be exploited in a lazy learning setting like those mentioned above.

One of the most appreciated advantages of performing inductive ABox reasoning through these methods is that they can naturally handle inconsistent (and inherently incomplete) knowledge bases, especially when inconsistency is not systematic. In [5] a method for dealing with inconsistent ABoxes populated through information extraction is proposed: it constructs ad hoc belief networks for the conflicting parts in an ontology and adopts the Dempster-Shafer theory for assessing the confidence of the resulting assertions.

6 Concluding Remarks and Outlook

In the line of our investigation of inductive methods for Semantic Web reasoning, we have proposed an alternative way for approximate ABox reasoning based on the nearest-neighbors analogical principle. Once neighbors of a test individual are selected through some distance measures, a combination rule descending from the Dempster-Shafer theory can fuse the evidence provided by the various neighbor individuals. We have shown how to exploit the procedure for assertion prediction problems such as determining unknown class- or role-memberships as well as attribute-values which may be the basis for many ABox inductive reasoning algorithms. The method is being implemented so to allow an extensive experimentation on real ontologies.

Special settings to accommodate cases of uncertain or unobserved values are to be investigated. One promising extension of the method concerns the possibility of considering infinite sets of values V following the studies [20, 2]. This would allow dealing with domains where the total amount of values is unknown (also due to the inherent nature of the Semantic Web). Moreover the predicted values often need not to be exclusive. Hence the prediction procedure would require an extension towards the consideration of sets of values instead of singletons.

As necessity and possibility measures are related to the belief measures (see note 2 at page 6) a natural extension may be towards the possibilistic theory and its calculus which is, in general, different from the Dempster-Shafer theory and calculus. Further possible extensions concern all other monotone measures such as the Sugeno λ-measures [2]. The extension towards the Possibility Theory is interesting also because of its parallelism with modal logics [20] and possibilistic extensions of Description Logics [21].

References

1. Laskey, K., Laskey, K., Costa, P., Kokar, M., Martin, T., Lukasiewicz, T.: Uncertainty Reasoning for the World Wide Web. W3C Incubator Group. (2008) http://www.w3.org/2005/Incubator/urw3/XGR-urw3-20080331/.
2. Klir, G.: Uncertainty and Information. Wiley (2006)
3. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press (1976)
4. Fanizzi, N., d'Amato, C., Esposito, F.: Approximate measures of semantic dissimilarity under uncertainty. In da Costa, P., et al., eds.: Uncertainty Reasoning for the Semantic Web I. Volume 5327 of LNAI. Springer (2008) 355–372
5. Nikolov, A., Urea, V., Motta, E., de Roeck, A.: Using the Dempster-Shafer theory of evidence to resolve ABox inconsistencies. In da Costa, P., et al., eds.: Uncertainty Reasoning for the Semantic Web I. Volume 5327 of LNAI. Springer (2008) 143–160
6. Denoeux, T.: A k-nearest neighbor classication rule based on Dempster-Shafer theory. IEEE Transactions on Systems, Man and Cybernetics **25** (1995) 804–813
7. Yager, R.: On the Dempster-Shafer framework and new combination rules. Information Sciences **41** (1987) 93–137
8. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P., eds.: The Description Logic Handbook. Cambridge University Press (2003)

9. d'Amato, C., Fanizzi, N., Esposito, F.: Analogical reasoning in description logics. In da Costa, P., et al., eds.: Uncertainty Reasoning for the Semantic Web I. Volume 5327 of LNAI. Springer (2008) 336–354
10. d'Amato, C., Fanizzi, N., Esposito, F.: Query answering and ontology population: An inductive approach. In Bechhofer, S., et al., eds.: Proceedings of the 5th European Semantic Web Conference, ESWC2008. Volume 5021 of LNCS., Springer (2008) 288–302
11. Fanizzi, N., d'Amato, C., Esposito, F.: Metric-based stochastic conceptual clustering for ontologies. Information Systems **34** (2009) 725–739
12. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning – Data Mining, Inference, and Prediction. Springer (2001)
13. Goldstone, R., Medin, D., Halberstadt, J.: Similarity in context. Memory and Cognition **25** (1997) 237–255
14. Borgida, A., Walsh, T., Hirsh, H.: Towards measuring similarity in description logics. In Horrocks, I., Sattler, U., Wolter, F., eds.: Working Notes of the International Description Logics Workshop. Volume 147 of CEUR Workshop Proceedings., Edinburgh, UK (2005)
15. d'Amato, C., Staab, S., Fanizzi, N.: On the influence of description logics ontologies on conceptual similarity. In Gangemi, A., Euzenat, J., eds.: Proceedings of the 16th Knowledge Engineering Conference, EKAW2008. Volume 5268 of LNAI., Springer (2008) 48–63
16. Horrocks, I., Li, L., Turi, D., Bechhofer, S.: The Instance Store: DL reasoning with large numbers of individuals. In Haarslev, V., Möller, R., eds.: Proceedings of the 2004 Description Logic Workshop, DL 2004. Volume 104 of CEUR Workshop Proceedings., CEUR (2004) 31–40
17. Rettinger, A., Nickles, M.: Infinite hidden semantic models for learning with OWL DL. In d'Amato, C., et al., eds.: Proceedings of 1st ESWC Workshop on Inductive Reasoning and Machine Learning for the Semantic Web, IRMLeS09. Volume 474 of CEUR Workshop Proceedings., Heraklion, Greece (2009)
18. Tresp, V., Huang, Y., Bundschus, M., Rettinger, A.: Materializing and querying learned knowledge. In d'Amato, C., et al., eds.: Proceedings of 1st ESWC Workshop on Inductive Reasoning and Machine Learning for the Semantic Web, IRMLeS09. Volume 474 of CEUR Workshop Proceedings., Heraklion, Greece (2009)
19. Baader, F., Ganter, B., Sertkaya, B., Sattler, U.: Completing description logic knowledge bases using formal concept analysis. In Veloso, M., ed.: Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India (2007) 230–235
20. Harmanec, D., Klir, G., Wang, Z.: Modal logic interpretation of Dempster-Shafer theory: An infinite case. International Journal of Approximate Reasoning **14** (1996) 81–93
21. Qi, G., Pan, J., Ji, Q.: A possibilistic extension of description logics. In Calvanese, D., et al., eds.: Working Notes of the 20th International Description Logics Workshop, DL2007. Volume 250 of CEUR Workshop Proceedings., Bressanone, Italy (2007) 435–442