Evidential Nearest-Neighbors Classification for Inductive ABox Reasoning

Nicola Fanizzi  Claudia d’Amato  Floriana Esposito

Department of Computer Science
University of Bari

URSW 2009  Washington, D.C., October 26, 2009
Introduction & Motivation

Background
- Dempster-Shafer Theory
- Classification by the Nearest Neighbor Approach

Evidential Nearest Neighbor Procedure (ENN)

ENN for the Semantic Web
- Prediction of Class-Membership Assertions
- Prediction of Datatype Fillers
- Prediction of Relationships among Individuals

Conclusions and Future Works
In the SW context:

- purely deductive-based methods may fail when data sources are distributed and, as such, potentially incoherent.
- due to the OWA, 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.
- information is, most of the time, inherently uncertain.
  
  ↓

- a growing interest is being committed to alternative reasoning procedures also to deal with the various facets of uncertainty.
  - here, inductive reasoning is adopted.
Paper Contributions

- ABox reasoning based on the *evidence* and the analogical principle of the nearest-neighbor approach
  - extension of a framework for the *classification of individuals* through a prediction procedure *based on evidence theory and similarity*

- prediction of the values related to class-membership or datatype and object properties
Basics of Dempster-Shafer Theory…

- A **frame of discernment** $\Omega$ is defined as the set of all hypotheses in a certain domain.
- A **basic belief assignment** (BBA) is a function $m : 2^\Omega \mapsto [0, 1]$ verifying: $\sum_{A \in 2^\Omega} m(A) = 1$
  - Given a certain piece of evidence, the value of the BBA for a given set $A$ expresses a body of evidence 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.
...Basics of Dempster-Shafer Theory...

The BBA $m$ define a *body of evidence*, from which a *belief function* $Bel$ and a *plausibility function* $Pl$ can be derived as mappings from $2^\Omega$ 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:

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

- Analogously, the plausibility of $A$, denoted $Pl(A)$, *represents the amount of belief in $A$, if further information became available*. $Pl$ is defined as

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

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.
...Basics of Dempster-Shafer Theory

- The Dempster-Shafer *rule of combination* 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 $m_2$.

\[
\forall A \in 2^\Omega \quad m_{12}(A) = (m_1 \oplus m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C)
\]

Generally, the normalized version of the rule is used:

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

(and $m_{12}(\emptyset) = 0$)
Nearest Neighbor Classification

classes: \( a, b \) \quad k = 5

\[ \text{class}(x_q) \leftarrow ? \]
Nearest Neighbor Classification

classes: \( a, b \) \hspace{1cm} k = 5

\[ \text{class}(x_q) \leftarrow a \]
Let $X$ be the finite set of instances and $V \subseteq \mathbb{Z}$ a finite set of integers $V \subseteq \mathbb{Z}$ to be used as labels.

The training set is $\text{TrSet} = \{(x_1, v_1), \ldots, (x_M, v_M)\} \subseteq \text{Ind} \times V$ where $X = \text{Ind}(\mathcal{A})$ is the set of individual names occurring in the ontology.

The frame of discernment $\Omega$ w.r.t. the classification problem is the set of all possible classes.

$x_q$ is a new individual to be classified on the basis of its nearest neighbors in $\text{TrSet}$.

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 $\text{TrSet}$.

An appropriate metric $d$ is applied to ontology individuals (e.g. one of the measures in the family defined in [d’Amato et al. @ ESWC 2008]).
Each pair \((x_i, v_i) \in N_k(x_q)\) constitutes a distinct item of evidence w.r.t 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 Denoeux’95:

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

(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.
...Evidential Nearest Neighbor Procedure...

- Considering each training individual in $N_k(x_q)$, $k$ BBAs $m_j$ are obtained. These can be aggregated in the final belief:

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

Functions $\overline{Bel}$ and $\overline{Pl}$ can be derived from $\tilde{m}$

- $x_q$ is assigned the value in $V$ that maximizes the belief or plausibility:

$$v_q = \arg\max_{(x_i, v_i)\in N_k(x_q)} \overline{Bel}(\{v_i\}) \text{ or } v_q = \arg\max_{(x_i, v_i)\in N_k(x_q)} \overline{Pl}(\{v_i\})$$

- Selecting the hypothesis with the greatest degree of belief i.e. the most credible corresponds to a skeptical viewpoint
- Selecting the hypothesis with the lowest degree of doubt i.e. 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
It is possible to *combine* the two measures \( Bel \) and \( Pl \).

A single measure of *confirmation* \( C \), ranging in \([-1, +1]\), can be defined [Klir’06]

\[
\forall A \subseteq \Omega \quad C(A) = Bel(A) + Pl(A) - 1
\]  

denoted with \( \overline{C} \) the combination of \( \overline{Bel} \) and \( \overline{Pl} \), *the resulting rule for predicting the value for* \( x_q \) *can be written as:*

\[
v_q = \arg\max_{(x_i, v_i) \in N_k(x_q)} \overline{C}([v_i])
\]
Evidential Nearest Neighbor Procedure

\[ \text{ENN}_k(x_q, \text{TrSet}, V) \]

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

Figure: The evidence nearest neighbor procedure.
Prediction of Class-Membership Assertions

Given:

- a (query) concept \( Q \)
- a set of values \( V_Q = \{+1, -1\} \) denoting, resp., membership and non-membership w.r.t. the query concept
- the values of the labels \( v_i \) for the training examples can be obtained through deductive reasoning (instance-checking)
- the related training set \( \text{TrSet}_Q \subseteq \text{Ind}(A) \times V_Q \)
- to predict the class-membership value \( v_q \) for some individual \( x_q \) w.r.t. \( Q \), it suffices to call \( \text{ENN}_k(x_q, \text{TrSet}_Q, V_Q) \)
  - the conclusion will be \( \mathcal{K} \models Q(x_q) \) or \( \mathcal{K} \models \neg Q(x_q) \) depending on the value that maximizes \( C \) (resp., \( v_q = +1 \) or \( v_q = -1 \))
  - the value which determined \( v_q \) can be exploited for ranking the hits by comparing the strength of the inductive conclusions
Prediction of Class-Membership Assertions: Extensions

- Ternary classification problems $V_q = \{-1, 0, +1\}$ are admitted, where 0 explicitly denote an indefinite (uncertain) class-membership [d’Amato et al. @ ESWC’08]
  - it can happened that the most likely value is $v_q = 0$
  - the choice could be forced (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 can be exploited for performing the inductive retrieval of a certain concept
  - given a concept $Q$, it would suffice to find all individuals $a \in \text{Ind}(A)$ that are s.t. $\mathcal{K} \models Q(a)$
  - the hits could be returned ranked by the respective confirmation value $\overline{C}(+1)$
Given a functional datatype property $P$ the problem is to predict the value of $P$ for a certain test individual $a$ that is in the domain of $P$.

$V_P$ correspond to the discrete and finite range of $P$ or to its restriction to the observed values for the training instances:

$$V_P = \{ v \in \text{range}(P) \mid \exists P(a, v) \in A \}$$

The training set will be $\text{TrSet}_P \subseteq \text{domain}(P) \times V_P$, where $\text{domain}(P) \subseteq \text{Ind}(A)$ is the set of individual names that have a known $P$-value in the KB.

To predict the value in $V_P$ of $P$ for some individual $a$ the procedure with $\text{ENN}_k(a, \text{TrSet}_P, V_P)$ has to be called.

Thus, if $v_q$ is the value that maximizes Eq. 5 then $\mathcal{K} \models P(a, v_q)$ can be written.
Different settings may be devised allowing for special value(s) denoting the case of a yet unobserved value(s) for that property.

The ENN procedure can be exploited for performing alternate forms of retrieval, e.g. finding all individuals with a certain value for the given property:

- given a certain value $v$, all individuals $a \in \text{Ind}(A)$ that are such that $\mathcal{K} \models P(a, v)$ have to be found.

- the hits could be returned ranked according to the respective confirmation value $\overline{C}(+1)$. 
Prediction of Relationships among Individuals

The **ENN** procedure *can be used to establish if a test individual is related through some object property with some others*

- the problem is decomposed into smaller ones aiming at verifying whether $\mathcal{K} \models R(a, b)$ holds:

for each $b \in \text{Ind}(A)$ do
  for each $a \in \text{Ind}(A)$ do
    $\text{TrSet} \leftarrow \{(x, v) \mid x \in \text{Ind}(A) \setminus \{a\}, \text{if } \mathcal{K} \models R(x, b), \text{then } v \leftarrow +1$
    else $v \leftarrow -1$
    $v_b^R \leftarrow \text{ENN}_k(a, \text{TrSet}, \{+1, -1\})$
    if $v_b^R = +1$ then
      return $\mathcal{K} \models R(a, b)$
    else
      return $\mathcal{K} \models \neg R(a, b)$
Prediction of Relationships among Individuals: Extension

- A ternary value set $V_R = \{-1, 0, +1\}$ could be alternatively considered
  - It allows for an explicit treatment of those individuals $a$ for which $R(a, b)$ is not derivable (or absent from the ABox)
  - A threshold of confirmation for accepting likely assertions can be used
Conclusions:

- Proposed and inductive method for approximate ABox reasoning based on the nearest-neighbors analogical principle and the theory of evidence
- Shown how to exploit the procedure for assertion prediction problems

Future works:

- Prediction of values for non-functional datatype properties
- Investigation on the possibility of considering infinite sets of values
- Setting up an extension of prediction procedure towards the consideration of sets of values instead of singletons
That’s all!

Questions ?