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

Contents

Introduction & Motivations

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

$$\Downarrow$$

- 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* Ω 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^{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:

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

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

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

...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.

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

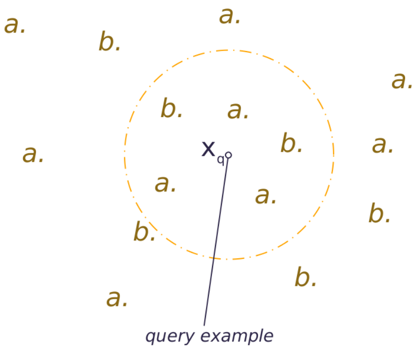
Generally, the normalized version of the rule is used:

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

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

Nearest Neighbor Classification

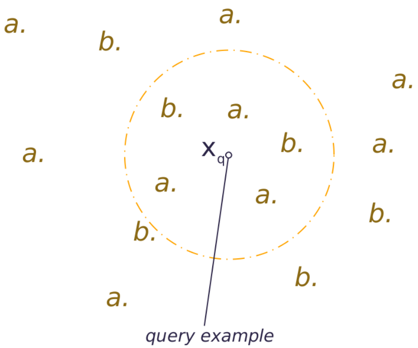
classes: a, b $\qquad k = 5$



$class(x_q) \leftarrow ?$

Nearest Neighbor Classification

classes: a, b $k = 5$



$class(x_q) \leftarrow$**a**

Evidential Nearest Neighbor Procedure...

- Let X be the finite set of instances and $V \subseteq \mathbf{Z}$ a finite set of integers $V \subseteq \mathbf{Z}$ to be used as labels
- The training set is $\mathsf{TrSet} = \{(x_1, v_1), \ldots, (x_M, v_M)\} \subseteq \mathsf{Ind} \times V$ where $X = \mathsf{Ind}(\mathcal{A})$ is the set of individual names occurring in the ontology
- The *frame of discernment* Ω 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 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
  - 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]**

...Evidential Nearest Neighbor Procedure...

- *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**:

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

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 $n \in \mathbf{N}$). 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:
$$\bar{m} = \bigoplus_{j=1}^{k} m_j = m_1 \oplus \cdots \oplus m_k$$
Functions $\overline{Bel}$ and $\overline{Pl}$ can be derived from $\bar{m}$
- x_q *is assigned the value in* V *that maximizes the belief or plausibility*:
$$v_q = \operatorname*{argmax}_{(x_i,v_i)\in N_k(x_q)} \overline{Bel}(\{v_i\}) \quad \text{or} \quad v_q = \operatorname*{argmax}_{(x_i,v_i)\in N_k(x_q)} \overline{Pl}(\{v_i\})$$
    - *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

...Evidential Nearest Neighbor Procedure...

- 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]**

$$\forall A \subseteq \Omega \qquad C(A) = Bel(A) + Pl(A) - 1 \tag{4}$$

- 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:

$$v_q = \operatorname*{argmax}_{(x_i,v_i)\in N_k(x_q)} \overline{C}(\{v_i\}) \tag{5}$$

...Evidential Nearest Neighbor Procedure

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

1. Compute the neighbor set $N_k(x_q) \subseteq \mathrm{TrSet}$.
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 $\overline{Bel}$ and $\overline{Pl}$ (Eqs. 1–2)

   Compute the confirmation $\overline{C}$ (Eq. 4) from $\overline{Bel}$ and $\overline{Pl}$
4. Select $v \in V$ that maximizes $\overline{C}$ (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 $\mathrm{TrSet}_Q \subseteq \mathrm{Ind}(\mathcal{A}) \times V_Q$
- *to predict the class-membership value* v_q *for some individual* x_q *w.r.t.* Q, it suffices to call $ENN_k(x_q, \mathrm{TrSet}_Q, V_Q)$
  - *the conclusion will be* $\mathcal{K} \approx Q(x_q)$ or $\mathcal{K} \approx \neg Q(x_q)$ depending on the value that maximizes $\overline{C}$ (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* $a \in \mathrm{Ind}(\mathcal{A})$ *that are s.t.* $\mathcal{K} \approx Q(a)$
  - the hits could be returned ranked by the respective confirmation value $\overline{C}(+1)$

Prediction of Datatype Fillers

- **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: $V_P = \{v \in range(P) \mid \exists P(a, v) \in \mathcal{A}\}$
- the training set will be $\mathrm{TrSet}_P \subseteq domain(P) \times V_P$, where $domain(P) \subseteq \mathrm{Ind}(\mathcal{A})$ 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 $ENN_k(a, \mathrm{TrSet}_P, V_P)$ has to be called
  - thus, if v_q is the value that maximizes Eq. 5 then $\mathcal{K} \models_{\approx} P(a, v_q)$ can be written

Prdition of Datatype fillers: Extensions

- 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 $a \in \mathrm{Ind}(\mathcal{A})$ that are such that $\mathcal{K} \models_{\approx} P(a, v)$ 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 $\mathcal{K} \approx R(a, b)$ holds:

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

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 & Future Work

Conclusions:

- Proposed and inductive method for approximate ABox reasoning based on the nearest-neighbors analogical principle and the teory 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 V
- setting up an extension of prediction procedure towards the consideration of sets of values instead of singletons

The End

That's all!

Questions ?