A Graph Regularization Based Approach to Transductive Class-Membership Prediction

Pasquale Minervini, Claudia d’Amato, Nicola Fanizzi

Computer Science Department • University of Bari, Italy

November, 11, 2012 - Boston
Contents

1. Introduction & Motivations

2. Class-membership with Transductive Learning
   - Problem Definition
   - The approach

3. Experimental Evaluation

4. Conclusions
Deductive Reasoning, usually adopted in the SW context, may fail in presence of inconsistent and/or noisy knowledge bases.

Machine learning methods can be adopted to perform approximate and uncertain reasoning allowing to derive conclusions which are not derivable or refutable from the knowledge base.

**Issue:** unlabeled instances could be present (because of the OWA)
Focus

Goal: Class-membership (concept) prediction task

Focus on trasductive semi-supervised learning methods
- family of machine learning methods that use both labeled and unlabeled data for training a learning algorithm
- in trasductive setting, the learning algorithm only aims at estimating the class-membership for a given training set, without generalizing to individuals outside such set.

Motivation: automation of the knowledge acquisition process
- the acquisition of labeled (training) data for a learning task often requires the manual effort of human agents ⇒ the cost may render a fully labeled training set infeasible
- the acquisition of unlabeled data is relatively inexpensive
- labelled data are not always available
Transductive Class-Membership Prediction

Definition

Given:

- a \textit{target} concept \( C \);
- a set of training individuals \( \text{Ind}_C(\mathcal{K}) \) in a knowledge base \( \mathcal{K} \) partitioned in:
  - \( \text{Ind}_C^+(\mathcal{K}) = \{ a \in \text{Ind}_C(\mathcal{K}) \mid \mathcal{K} \models C(a) \} \) positive examples,
  - \( \text{Ind}_C^-(\mathcal{K}) = \{ a \in \text{Ind}_C(\mathcal{K}) \mid \mathcal{K} \models \neg C(a) \} \) negative examples,
  - \( \text{Ind}_C^0(\mathcal{K}) = \{ a \in \text{Ind}_C(\mathcal{K}) \mid \mathcal{K} \not\models C(a) \land \mathcal{K} \not\models \neg C(a) \} \) unlabeled examples;
- a \textit{cost function} \( \text{cost}(\cdot) : \mathcal{F} \mapsto \mathbb{R} \), specifying the \textit{cost} associated to a set of class-memberships assigned to training individuals by \( f \in \mathcal{F} \), where \( \mathcal{F} \) is a space of labelling functions of the form \( f : \text{Ind}_C(\mathcal{K}) \mapsto \{+1, -1\} \);

Find: a \textit{labelling function} \( f^* \in \mathcal{F} \) \textbf{minimizing} the given cost function w.r.t. \( \text{Ind}_C(\mathcal{K}) \):

\[
f^* \leftarrow \arg \min_{f \in \mathcal{F}} \text{cost}(f).
\]

The function \( f^* \) can then be used to estimate the class-membership w.r.t. \( C \) for all training individuals \( a \in \text{Ind}_C(\mathcal{K}) \).
Graph-based semi-supervised approach

1. Choose/build a *target concept* $C$
2. Determine the training set $\text{Ind}_C(\mathcal{K})$ w.r.t. $C$ in $\mathcal{K}$ as given by positive, negative and unlabeled instances
3. Build the Nearest Neighbor (NN) Semantic Similarity graph
4. Define a cost over functions $f \in \mathcal{F}$ as a cost function
   - *finding a labeling function that is*
     - consistent with the given labels $\Rightarrow$ loss function as a measure of consistency with the given labels
     - changes smoothly between similar instances $\Rightarrow$ *Regularization by graph* $\Rightarrow$ measure of smoothness among the similarity graph as a regularizer
Building the NN-Semantic Similarity graph

The *Similarity graph* is built as a matrix $W$ where $W_{ij}$ is the similarity value between two training examples $x_i$ and $x_j$

- a NN graph, for each instance $x_i$, contains similarity the value only for the $k$ most similar instances (the others are set to 0)
- employed a family of similarity measures between in individuals in a DL knowledge base [*d’Amato et al. @ URSW’09*]
The Family of Similarity Measure

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a DL knowledge base. Given a set of concept descriptions $F = \{F_1, F_2, \ldots, F_m\}$, corresponding weights $w_1, \ldots, w_m$, and $p > 0$, a family of dissimilarity functions $d_p^F: \text{Ind}(\mathcal{A}) \times \text{Ind}(\mathcal{A}) \mapsto [0, 1]$ is defined by:

$$d_p^F(a, b) := \frac{1}{|F|} \left[ \sum_{i=1}^{|F|} w_i \left| \delta_i(a, b) \right|^p \right]^{1/p},$$

where the dissimilarity function $\delta_i \ (i \in \{1, \ldots, m\})$ is defined by:

$$\forall a, b \in \text{Ind}(\mathcal{A}): \quad \delta_i(a, b) = \begin{cases} 
0 & F_i(a) \in \mathcal{A} \land F_i(b) \in \mathcal{A} \\
1 & F_i(a) \in \mathcal{A} \land \neg F_i(b) \in \mathcal{A} \lor \\
1/2 & \text{otherwise.}
\end{cases}$$
Quadratic Cost Criteria: Preliminaries

1. Original label space \( \{-1, +1\} \) relaxed to \([-1, +1]\) allows to express the confidence associated to a labeling
   - The labeling function space \( \mathcal{F} \) is relaxed to functions of the form \( f : \text{Ind}_C(K) \mapsto [-1, +1] \)
   - Labeling functions can be represented as vectors \( y \in [-1, +1]^n \) where \( n = |\text{Ind}_C(K)| \)

2. Let \( \hat{y} \in [-1, +1]^n \) be a possible labeling for \( n \) instances
   - \( \hat{y} \) be seen as a \( (l + u) = n \) dimensional vector, where the first \( l \) indices refer to labeled instances, and the last \( u \) to unlabeled instances: \( \hat{y} = [\hat{y}_l, \hat{y}_u] \)
Quadratic Cost Criteria...

- **Consistency** of $\hat{y}$ w.r.t. the original labels can be formulated *in the form of a quadratic cost*

  \[
  \sum_{i=1}^{l} (\hat{y}_i - y_i)^2 = \|\hat{y}_l - y_l\|^2
  \]

- Labellings can be regularised w.r.t. the graph structure *alternatively as:*

  \[
  0.5 \sum_{i,j=1}^{n} W_{ij} (\hat{y}_i - \hat{y}_j)^2 = \hat{y}^T L \hat{y}
  \]

where $W$ is the semantic similarity graph and $L = D - W$, with $D$ the diagonal matrix s.t. $D_{ii} = \sum_j W_{ij}$, is the unnormalized graph Laplacian [Belkin et al. @ COLT’04]

\[
(D^{-0.5} \hat{y})^T L (D^{-0.5} \hat{y})
\]

[Zhou et al. @ ICML’05]
Putting consistency and regularization together two *quadratic cost criteria* are obtained:

1. **Regression on Graph (RG)**
   \[
   cost(\hat{y}) = \|\hat{y}_l - y_l\|^2 + \mu \hat{y}^T L \hat{y} + \mu \epsilon \|\hat{y}\|^2; \quad [\text{Belkin et al. @ COLT'04}]
   \]

2. **Consistency Method (CM)**
   \[
   cost(\hat{y}) = \|\hat{y}_l - y_l\|^2 + \mu (D^{-0.5} \hat{y})^T L (D^{-0.5} \hat{y}) + \|\hat{y}_u\|^2. \quad [\text{Zhou et al. @ ICML'5}]
   \]

By recurring to derivative, finding the minimum for **RG** (resp. **CM**) consists in solving a (possibly sparse) linear system whose *time complexity is nearly linear* in the number of non-zero entries in the coefficient matrix.
Experiments: Setup

**GOAL:** evaluation and comparison of *inductive* and *transductive methods* for class-membership prediction

- 20 random queries $C$ created for each ontology (Tab.)
  - $C$ (resp. $\neg C$) should contain at least 10 instances
  - Pellet Reasoner v2.3.0 employed for building the $TrSet$ $Ind_{C}(K)$
- **Trasductive Methods:** Regression Graph (RG), Consistency Method (CM), Label Propagation ((LP))
- **Inductive Methods:** Soft Margin SVM (SM-SVM), Laplacian SVM (LapSVM), k-NN ($k = \sqrt{l}$ where $l =$ num. labeled ex. (($\sqrt{l}$-NN))

<table>
<thead>
<tr>
<th>Ontology</th>
<th>Expressivity</th>
<th>#Axioms</th>
<th>#Indiv.</th>
<th>#Classes</th>
<th>#Obj.Prop.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioPax (Prot.)</td>
<td>$ALCHN(D)$</td>
<td>773</td>
<td>49</td>
<td>55</td>
<td>47</td>
</tr>
<tr>
<td>Family-Tree</td>
<td>$SROIF(D)$</td>
<td>2059</td>
<td>368</td>
<td>22</td>
<td>52</td>
</tr>
<tr>
<td>Leo</td>
<td>$ALCHIF(D)$</td>
<td>430</td>
<td>61</td>
<td>32</td>
<td>26</td>
</tr>
<tr>
<td>MDM0.73</td>
<td>$ALCHOF(D)$</td>
<td>1098</td>
<td>112</td>
<td>196</td>
<td>22</td>
</tr>
<tr>
<td>Wine</td>
<td>$SHOIN(D)$</td>
<td>1046</td>
<td>218</td>
<td>142</td>
<td>21</td>
</tr>
</tbody>
</table>
**Metrics**

**Match Rate**  Case of an individual that got the same label by the reasoner and the inductive classifier.

**Omission Error**  Case of an individual for which the inductive method could not determine whether it was relevant to the query concept or not while it was found relevant by the reasoner.

**Commission Error**  Case of an individual found to be relevant to the query concept while it logically belongs to its negation or vice-versa.

**Induction**  Case of an individual found to be relevant to the query concept or to its negation, while either case is not logically derivable from the knowledge base.
## Results

<table>
<thead>
<tr>
<th></th>
<th>Leo</th>
<th>BioPAX (Prot.)</th>
<th>MDM0.73</th>
<th>Wine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Match</td>
<td>Omission</td>
<td>Commission</td>
<td>Induction</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG</td>
<td>1 ± 0</td>
<td>0.004 ± 0.028</td>
<td>0.008 ± 0.039</td>
<td>0.002 ± 0.02</td>
</tr>
<tr>
<td>CM</td>
<td>1 ± 0</td>
<td>0.002 ± 0.02</td>
<td>0.013 ± 0.036</td>
<td>0.002 ± 0.02</td>
</tr>
<tr>
<td>LP</td>
<td>0.942 ± 0.099</td>
<td>0.002 ± 0.02</td>
<td>0.014 ± 0.051</td>
<td>0.002 ± 0.02</td>
</tr>
<tr>
<td>SM-SVM</td>
<td>0.963 ± 0.1</td>
<td>0 ± 0</td>
<td>0.026 ± 0.068</td>
<td>0.002 ± 0.02</td>
</tr>
<tr>
<td>LapSVM</td>
<td>0.978 ± 0.068</td>
<td>0 ± 0</td>
<td>0.026 ± 0.068</td>
<td>0.002 ± 0.02</td>
</tr>
<tr>
<td>√l-NN</td>
<td>0.971 ± 0.063</td>
<td>0 ± 0</td>
<td>0.026 ± 0.068</td>
<td>0.002 ± 0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Match</th>
<th>Omission</th>
<th>Commission</th>
<th>Induction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG</td>
<td>0.953 ± 0.063</td>
<td>0.003 ± 0.016</td>
<td>0.011 ± 0.032</td>
<td>0.015 ± 0.039</td>
</tr>
<tr>
<td>CM</td>
<td>0.953 ± 0.063</td>
<td>0.001 ± 0.009</td>
<td>0.013 ± 0.036</td>
<td>0.018 ± 0.04</td>
</tr>
<tr>
<td>LP</td>
<td>0.942 ± 0.065</td>
<td>0 ± 0</td>
<td>0.026 ± 0.046</td>
<td>0.033 ± 0.054</td>
</tr>
<tr>
<td>SM-SVM</td>
<td>0.793 ± 0.252</td>
<td>0 ± 0</td>
<td>0.174 ± 0.255</td>
<td>0.033 ± 0.054</td>
</tr>
<tr>
<td>LapSVM</td>
<td>0.915 ± 0.086</td>
<td>0 ± 0</td>
<td>0.052 ± 0.065</td>
<td>0.033 ± 0.054</td>
</tr>
<tr>
<td>√l-NN</td>
<td>0.944 ± 0.069</td>
<td>0 ± 0</td>
<td>0.023 ± 0.051</td>
<td>0.033 ± 0.054</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Match</th>
<th>Omission</th>
<th>Commission</th>
<th>Induction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RG</td>
<td>0.24 ± 0.03</td>
<td>0 ± 0.005</td>
<td>0.007 ± 0.017</td>
<td>0.5 ± 0.176</td>
</tr>
<tr>
<td>CM</td>
<td>0.242 ± 0.028</td>
<td>0 ± 0.005</td>
<td>0.005 ± 0.015</td>
<td>0.326 ± 0.121</td>
</tr>
<tr>
<td>LP</td>
<td>0.239 ± 0.035</td>
<td>0 ± 0.005</td>
<td>0.008 ± 0.021</td>
<td>0.656 ± 0.142</td>
</tr>
<tr>
<td>SM-SVM</td>
<td>0.235 ± 0.036</td>
<td>0 ± 0</td>
<td>0.012 ± 0.024</td>
<td>0.753 ± 0.024</td>
</tr>
<tr>
<td>LapSVM</td>
<td>0.238 ± 0.033</td>
<td>0 ± 0</td>
<td>0.009 ± 0.021</td>
<td>0.753 ± 0.024</td>
</tr>
<tr>
<td>√l-NN</td>
<td>0.241 ± 0.031</td>
<td>0 ± 0</td>
<td>0.006 ± 0.018</td>
<td>0.753 ± 0.024</td>
</tr>
</tbody>
</table>

C. d'Amato  | Transductive Class-Membership Prediction
Discussion

- **Family-Tree** ontology (not reported), provided $0.76 \pm 0.13$ match rates and $0.24 \pm 0.13$ induction rates for all but **LP** method, where the induction rates were $0.21 \pm 0.14$.

- In general, LapSVM outperformed the other two non-SSL SVM classification methods.

- Trasductive approaches generally outperform inductive approaches in terms of *commission error* and *match rate*.

- Trasductive approaches resulted more conservative than inductive approaches for MDMO.73 and WINE ontologies, showing:
  - highest omission rates
  - lowest induction rates

- The proposed **RG** and **CM** always outperform the **LP** adopted as a baseline trasductive methods.
Conclusions: A method for trasductive class-membership prediction based on graph-based regularization has been proposed. It relies on quadratic cost criteria whose optimization can be reduced to solve a (possibly sparse) linear system. Experimental evaluations showed the improvement of the trasductive approach over the inductive one particularly in terms of commission error and match rate.

Future Works:
- Deeply investigate on the correlation between the order of magnitude of unlabeled instances and the results of the proposed method.
That’s all!
Questions?