ReduCE: A Reduced Coulomb Energy Network Method for Approximate Classification

Nicola Fanizzi
Claudia d'Amato
Floriana Esposito

Dipartimento di Informatica
Università degli studi di Bari
Agenda

- Motivation
  - applications
- Inductive Inference: the learning problem
- RCE Networks
  - Learning
  - Approximate Classification of Individuals
- Experiments
- Conclusions & Outlook
Introduction: Motivation

- **Inductive Inference** on SeWeb knowledge bases through ML techniques
  - explicit knowledge models: new concepts
    - [ISWC04, Lehmann&Hitzler@ILP07] [DL-FOIL@ILP08]
  - implicit knowledge models: neural networks, kernel machines, probabilistic models
    - DL-kNN, DL-Kernels [ESWC2008; ISWC2008]

- **Focus**: Inductive methods for classification
  - often more **efficient** and noise-tolerant than standard logical methods
  - enable **approximation**
  - better exploitation of the (inherently incomplete or incoherent) available knowledge in Kbs
  - More **stability** wrt previously proposed methods
Introduction: Applications

Inductive Inference instance-checking exploited for

- approximate retrieval, subsumption, matchmaking, ...
- alternative methods for ontology population

- used for completing KBs
  - with induced assertions
  - or, also
    with probabilistic assertions
    enabling further sophisticated approaches to dealing with
    uncertainty in KBs
Learning Problem

- Given:
  - a target concept $Q$
  - A set of pre-classified individuals: examples
  - A knowledge base $\mathcal{K}$ as background knowledge
- Train a model $h_Q$ (*hypothesis*)

then use the learned model $h_Q$ to classify other individuals:

- Given $h_Q(x_q)$ and a query individual $x_q$
- Output an estimate for $h_Q(x_q)$
- and the likelihood of this assertion
Examples and Hypotheses

- a **limited** number of individuals for which the intended classification is known
  \[ e_q = \langle x_q, h_Q(x_q) \rangle \]

\[ \forall x_i \in TrSet: \quad h_Q(x_i) = \begin{cases} +1 & \mathcal{K} \models Q(x_i) \\ -1 & \mathcal{K} \models \neg Q(x_i) \\ 0 & \text{otherwise} \end{cases} \]

- \( h_Q \): the function to be approximated
- in our case a combination of **hyperspheres**
The Inductive Model: RCE Networks

- Category layer
  - $Q$
  - $\neg Q$

- Pattern layer (radii)
  - $\lambda_1$, $\lambda_2$, $\lambda_3$, ..., $\lambda_N$

- Input layer
  - $x_1$, $x_2$, ..., $x_d$

Weights:
- $w_{jk}$
- $a_{cj}$
Training the RCE network: basic algorithm

input
\[ TrSet = \{ \langle x_i, h_Q(x_i) \rangle \} \]: set of training examples

output
\[ w_{jk}, \lambda_j, a_{cj} : \text{RCE Network weights} \]

1. begin
2. initialize \( \epsilon \leftarrow \text{small parameter}; \lambda_{\text{max}} \leftarrow \text{max radius} \)
3. for \( j \leftarrow 1 \) to \(|TrSet|\) do
   (a) train weight: \( w_{jk} \leftarrow x_k \)
   (b) find nearest counterexample: \( \bar{x} \leftarrow \arg \min_{x \in C_j} d(x, x_j) \)
      where \( C_j = \{ x \in TrSet \mid h_Q(x_j) \neq h_Q(x) \} \)
   (c) set radius: \( \lambda_j \leftarrow \min[\max(d(\bar{x}, x_j), \epsilon), \lambda_{\text{max}}] \)
   (d) if \( h_Q(x_j) = +1 \) then \( a_{Qj} \leftarrow 1 \) else \( a_{-Qj} \leftarrow 1 \)
4. end
RCE Model Construction

1. Initial model
2. Add new clusters
3. Refine clusters
4. Merge similar clusters
5. Split overlapping clusters
6. Optimize final model
RCE Final Model

Prototypes (examples)
Measuring Similarity

- Derived from pseudo-distance [ESWC2008]

Definition 3.1 (family of similarity measures). Let $\mathcal{K} = \langle T, A \rangle$ be a knowledge base. Given a set of concept descriptions $F = \{F_i\}_{i=1}^{m}$ and a normalized vector of weights $w = (w_1, \ldots, w_m)^t$, a family of similarity functions

$$s_p^F : \text{Ind}(A) \times \text{Ind}(A) \to [0, 1]$$

is defined as follows:

\[ s_p^F(a, b) = \frac{1}{m} \left[ \sum_{i=1}^{m} w_i \left| \sigma_i(a, b) \right|^p \right]^{1/p} \]

where $p > 0$ and $\forall i \in \{1, \ldots, m\}$ the similarity function $\sigma_i$ is defined by:

\[ \sigma_i(a, b) = \begin{cases} 0 & \text{if } [\mathcal{K} \models F_i(a) \text{ and } \mathcal{K} \models \neg F_i(b)] \text{ or } [\mathcal{K} \models \neg F_i(a) \text{ and } \mathcal{K} \models F_i(b)] \\ 1 & \text{if } [\mathcal{K} \models F_i(a) \text{ and } \mathcal{K} \models F_i(b)] \text{ or } [\mathcal{K} \models \neg F_i(a) \text{ and } \mathcal{K} \models \neg F_i(b)] \\ \frac{1}{2} & \text{otherwise} \end{cases} \]
(Vanilla) Classification Procedure

input
\[ x_q: \text{query individual} \]
\[ TrSet: \text{set of training examples} \]
\[ \lambda_j: \text{parameters of the trained RCE network} \]

output
\[ \hat{h}_Q(x_q): \text{estimated classification} \]

1. begin
2. initialize \( k \leftarrow 0; \ N(x_q) \leftarrow \emptyset \)
3. for \( j \leftarrow 1 \) to \(|TrSet|\) do
   - if \( d(x_q, x_j) < \lambda_j \)
     then \( N(x_q) \leftarrow N(x_q) \cup \{x_j\} \)
4. if (\( \forall x, x' \in N(x_q): h_Q(x) = h_Q(x') \)) all share the same class
   then return \( h_Q(x) \), shared class of all \( x \in N_{set} \)
   else return 0 // uncertain case
5. end
Extensions

- generalizing the decision-making step:

\[ g(x_q) = \sum_{x_j \in N(x_q)} h_Q(x_j) \cdot s(x_j, x_q) \]

Then step 4. in the procedure becomes:

4. if \(|g(x_q)| > \theta\) then return \(\text{sgn}(g(x_q))\) else return 0

- likelihood:

\[
\ell(\hat{h}(x_q) = v \mid N(x_q)) = \frac{\sum_{j=1}^{k} \delta(v, h_Q(x_j)) \cdot s(x_q, x_j)}{\sum_{u \in V} \sum_{h=1}^{k} \delta(u, h_Q(x_h)) \cdot s(x_q, x_h)}
\]
Experiments: Ontologies

- For each ontology
- Satisfiable random query concepts (100) generated by composition (conjunction / disjunction) of NC primitive and defined concepts
  - NC randomly varying between 2 and 8

<table>
<thead>
<tr>
<th>ontology</th>
<th>DL language</th>
<th>#concepts</th>
<th>#object prop.</th>
<th>#data prop.</th>
<th>#individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWM</td>
<td>ALCOF(D)</td>
<td>19</td>
<td>9</td>
<td>1</td>
<td>115</td>
</tr>
<tr>
<td>BioPAX</td>
<td>ALCHF(D)</td>
<td>28</td>
<td>19</td>
<td>30</td>
<td>323</td>
</tr>
<tr>
<td>LUBM</td>
<td>ALR^+HI(D)</td>
<td>43</td>
<td>7</td>
<td>25</td>
<td>555</td>
</tr>
<tr>
<td>NTN</td>
<td>SHIF(D)</td>
<td>47</td>
<td>27</td>
<td>8</td>
<td>676</td>
</tr>
<tr>
<td>SWSD</td>
<td>ALCH</td>
<td>258</td>
<td>25</td>
<td>0</td>
<td>732</td>
</tr>
<tr>
<td>FINANCIAL</td>
<td>ALCIF</td>
<td>60</td>
<td>17</td>
<td>0</td>
<td>1000</td>
</tr>
</tbody>
</table>
Experiments

- Evaluation: for all query concepts and individuals:
  - comparison of inductive to deductive responses
    - returned by a standard reasoner (Pellet 2)

- Indices
  - match rate: identical classification
  - omission error rate: 0 vs. ±1
  - commission error rate: +1 vs. -1 or -1 vs. +1
  - induction rate: ±1 vs. 0

- Cross Validation:
  - individuals divided into training and test sets
  - rates averaged according to the 632+ bootstrap procedure
Table 2. Results of the first session with uncertainty threshold $\theta = .3$ and minimum ball radius $\epsilon = .1$: average values ± average standard deviations per query.

<table>
<thead>
<tr>
<th>ontology</th>
<th>match rate</th>
<th>commission rate</th>
<th>omission rate</th>
<th>induction rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWM</td>
<td>83.99±01.06</td>
<td>00.00±00.00</td>
<td>04.80±00.47</td>
<td>11.21±00.75</td>
</tr>
<tr>
<td>BioPax</td>
<td>85.43±00.43</td>
<td>03.49±00.23</td>
<td>05.32±00.02</td>
<td>05.76±00.25</td>
</tr>
<tr>
<td>LUBM</td>
<td>89.77±00.26</td>
<td>00.00±00.00</td>
<td>06.68±00.21</td>
<td>03.55±00.06</td>
</tr>
<tr>
<td>NTN</td>
<td>86.71±00.32</td>
<td>00.08±00.00</td>
<td>05.48±00.21</td>
<td>07.73±00.33</td>
</tr>
<tr>
<td>SWSD</td>
<td>98.12±00.05</td>
<td>00.00±00.00</td>
<td>01.30±00.05</td>
<td>00.58±00.00</td>
</tr>
<tr>
<td>Financial</td>
<td>90.26±00.09</td>
<td>04.16±00.05</td>
<td>02.57±00.01</td>
<td>03.01±00.05</td>
</tr>
</tbody>
</table>

- credulous
- method more stable than previous ones (KNN, Kernel Machines)

[ESWC2008][ISWC2008]
Table 3. Results of the second session with uncertainty threshold $\theta = .7$ and minimum ball radius $\epsilon = .01$: average values ± average standard deviations per query.

<table>
<thead>
<tr>
<th>ontology</th>
<th>match rate</th>
<th>commission rate</th>
<th>omission rate</th>
<th>induction rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWM</td>
<td>93.52±00.58</td>
<td>00.00±00.00</td>
<td>06.19±00.59</td>
<td>00.29±00.05</td>
</tr>
<tr>
<td>BioPax</td>
<td>81.42±04.83</td>
<td>00.80±00.18</td>
<td>13.00±04.86</td>
<td>04.78±00.35</td>
</tr>
<tr>
<td>LUBM</td>
<td>91.59±00.24</td>
<td>00.00±00.00</td>
<td>07.80±00.23</td>
<td>00.62±00.02</td>
</tr>
<tr>
<td>NTN</td>
<td>83.78±01.51</td>
<td>00.00±00.00</td>
<td>14.23±02.31</td>
<td>01.99±00.83</td>
</tr>
<tr>
<td>SWSD</td>
<td>98.29±00.05</td>
<td>00.00±00.00</td>
<td>01.71±00.05</td>
<td>00.00±00.00</td>
</tr>
<tr>
<td>Financial</td>
<td>82.65±00.70</td>
<td>01.56±00.10</td>
<td>13.72±00.97</td>
<td>02.08±00.27</td>
</tr>
</tbody>
</table>

- more cautious
- stable results
Table 4. Results of the third session with uncertainty threshold $\theta = .5$ and minimum ball radius $\epsilon = .01$: average values ± average standard deviations per query.

<table>
<thead>
<tr>
<th>ontology</th>
<th>match rate</th>
<th>commission rate</th>
<th>omission rate</th>
<th>induction rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>SWM</td>
<td>94.24±00.83</td>
<td>00.00±00.00</td>
<td>05.26±00.86</td>
<td>00.51±00.24</td>
</tr>
<tr>
<td>BioPax</td>
<td>85.11±00.95</td>
<td>01.36±00.29</td>
<td>08.21±00.90</td>
<td>05.31±00.44</td>
</tr>
<tr>
<td>LUBM</td>
<td>97.49±00.74</td>
<td>00.00±00.00</td>
<td>02.47±00.73</td>
<td>00.04±00.02</td>
</tr>
<tr>
<td>NTN</td>
<td>86.85±00.24</td>
<td>00.00±00.00</td>
<td>06.57±00.74</td>
<td>06.58±00.63</td>
</tr>
<tr>
<td>SWSD</td>
<td>98.29±00.05</td>
<td>00.00±00.00</td>
<td>01.71±00.05</td>
<td>00.00±00.00</td>
</tr>
<tr>
<td>Financial</td>
<td>87.98±01.84</td>
<td>03.18±00.71</td>
<td>06.12±02.72</td>
<td>02.72±00.32</td>
</tr>
</tbody>
</table>

- good performance
- if individuals abound: choice of parameters via preliminary cross-validation
Conclusions & Outlook

- Similarity-based *parametrized* method for approximate classification in DLs

- Experiments:
  - competitive wrt previous methods
  - High match rate
  - Low induction rate
  - Some omission errors
  - Very limited commission errors
  - Low variance wrt to past inductive methods

- Improvements
  - efficient data structures
  - pre-determination of parameters
  - Pre-computation of prototypical ex's
    - Clustering medoids

- Extensions
  - ANNs, RBFNs
  - force binary response
    - (tweak $\theta$)
    - Expected to increase induction

- Use probability
  - ranking
  - addition to assertions
Questions ?

For offline contacts:
Nicola Fanizzi
Claudia d'Amato
Floriana Esposito

fanizzi@di.uniba.it
claudia.damato@di.uniba.it
esposito@di.uniba.it

Other methods / systems
http://lacam.di.uniba.it:8000/~nico/research/ontologymining.html