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 sophisticate 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 $h_{\mathrm{Q}}(x_q)$ and a query individual x_q
- Output an estimate for $h_{\mathrm{Q}}(x_q)$
  - and the likelihood of this assertion

Examples and Hypotheses

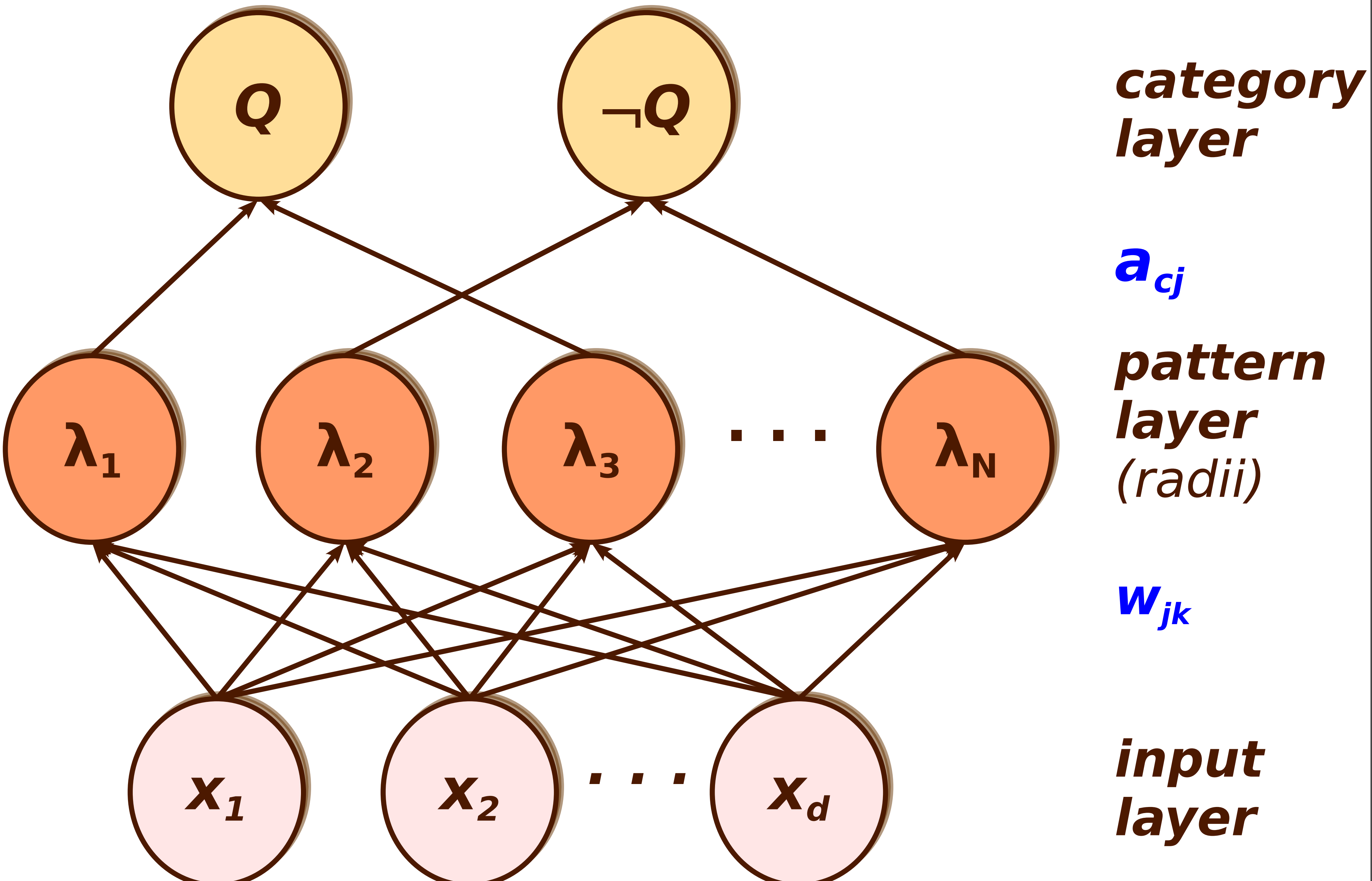
- a ***limited*** number of individuals
  for which the intended classification is known

$$e_q = \langle x_q, h_Q(x_q) \rangle$$

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

- h_Q: the function to be approximated
  - in our case a combination of hyperspheres

The Inductive Model: RCE Networks



Training the RCE network: basic algorithm

input

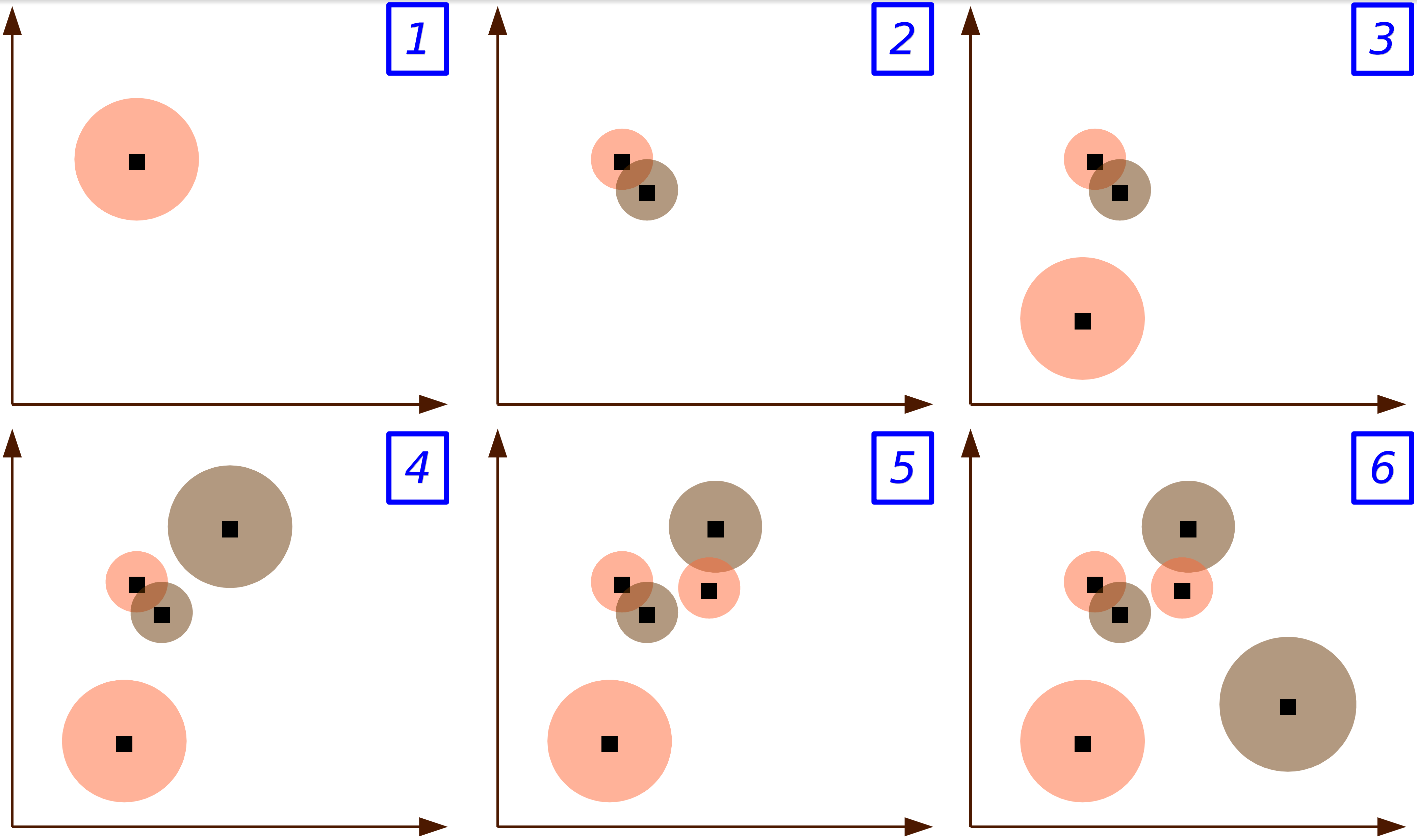
$TrSet = \{\langle x_i, h_Q(x_i)\rangle\}$: set of training examples

output

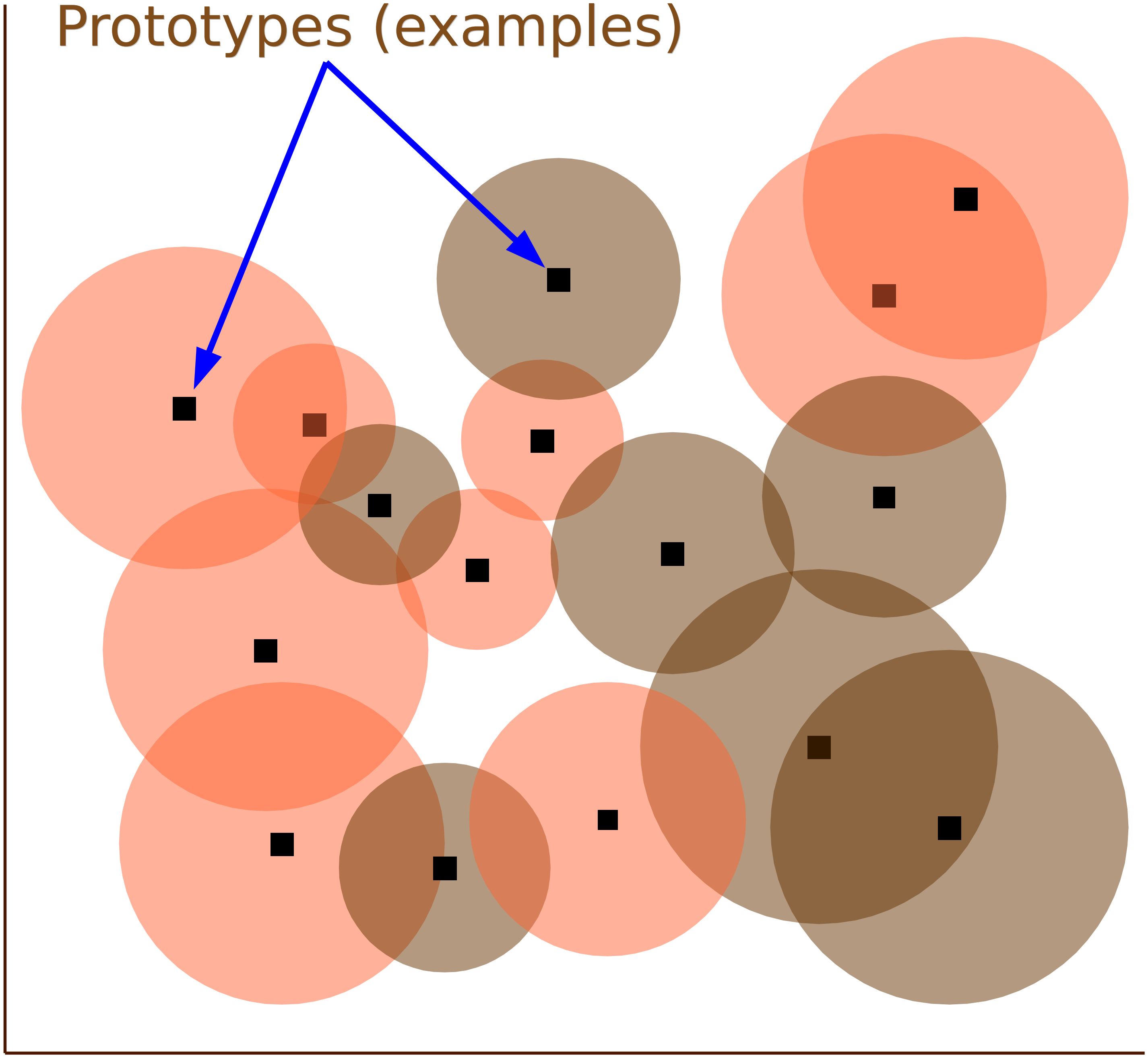
w_{jk}, λ_j, a_{cj}: RCE Network weights

1. **begin**
2. initialize $\epsilon \leftarrow$ small parameter; $\lambda_{\max} \leftarrow$ 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: $\lambda_j \leftarrow \min[\max(d(\bar{x}, x_j), \epsilon), \lambda_{\max}]$
   (d) **if** $(h_Q(x_j) = +1)$ **then** $a_{Qj} \leftarrow 1$ **else** $a_{\neg Qj} \leftarrow 1$
4. **end**

RCE Model Construction



RCE Final Model



Measuring Similarity

- Derived from pseudo-distance [ESWC2008]

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

$$s_p^{\mathsf{F}} : \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \longrightarrow [0, 1]$$

is defined as follows:
$\forall a, b \in \mathsf{Ind}(\mathcal{A})$

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

where $p > 0$ and $\forall i \in \{1, \ldots, m\}$ the similarity function σ_i is defined by:
$\forall a, b \in \mathsf{Ind}(\mathcal{A})$

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

(Vanilla) Classification Procedure

input

x_q: query individual

$TrSet$: set of training examples

λ_j: parameters of the trained RCE network

output

$\hat{h}_Q(x_q)$: 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** $\mathrm{sgn}(g(x_q))$ **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

ontology	DL language	#concepts	#object prop.	#data prop.	#individuals
SWM	$\mathcal{ALCOF}(\mathcal{D})$	19	9	1	115
BioPAX	$\mathcal{ALCHF}(\mathcal{D})$	28	19	30	323
LUBM	$\mathcal{ALR}^{+}\mathcal{HI}(\mathcal{D})$	43	7	25	555
NTN	$\mathcal{SHIF}(\mathcal{D})$	47	27	8	676
SWSD	$\mathcal{ALCH}$	258	25	0	732
Financial	$\mathcal{ALCIF}$	60	17	0	1000

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

Outcomes

Table 2. Results of the first session with uncertainty threshold $\theta = .3$ and minimum ball radius $\epsilon = .1$: average values $\pm$ average standard deviations per query.

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

- credulous
- method more stable than previous ones (KNN, Kernel Machines) [ESWC2008][ISWC2008]

Outcomes / 2

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

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

- more cautious
- stable results

Outcomes / 3

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

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

- good performance
- if individuals abound: choice of parametres 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 θ)
    - Expected to increase induction
- Use probability
  - ranking
  - addition to assertions

Questions ?

For offline contacts:

Nicola **Fanizzi** fanizzi@di.uniba.it
Claudia **d'Amato** claudia.damato@di.uniba.it
Floriana **Esposito** esposito@di.uniba.it

Other methods / systems

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