Analogical Reasoning in Description Logics

Claudia d'Amato, Nicola Fanizzi, Floriana Esposito

Department of Computer Science
University of Bari, Italy

Uncertainty Reasoning for the Semantic Web workshop
URSW 2006 @ ISWC 2006
November, 5th
Athens (GA) USA

Contents

Motivations

- Ontological knowledge
  - Result of a complex process of knowledge acquisition
  - Plays a key role for interoperability in the Semantic Web perspective
  - Is expressed by standard ontology mark-up languages which are supported by well-founded semantics of *Description Logics* (DLs)
- Need of services able to build knowledge bases automatically or semi-automatically
  - This can be done by the use of inductive inference services

Objectives...

- Induction of structural knowledge is known is ML
  e.g. Instance-based classification methods
  - This is generally applied on zero-order representations.
- *Goals*:
  - answering to class-memebership queries
  - performing instance retrieval through inductive analogical reasoning
  - predict assertions that may not be logically entailed by the KB (exploting the inductive analogical reasoning)
- *Problem* $\longrightarrow$ it is necessary to have a similarity/dissimilarity measure applicable to ontology languages

The Representation Language: $\mathcal{ALC}$ Logic

- Primitive *concepts* $N_C = \{C, D, \dots\}$: subsets of a domain
- Primitive *roles* $N_R = \{R, S, \dots\}$: binary relations on the domain

Name	Syntax	Semantics
top concept	$\top$	$\Delta^{\mathcal{I}}$
bottom concept	$\bot$	$\emptyset$
concept	C	$C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
concept negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
concept conjunction	$C_1 \sqcap C_2$	$C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$
concept disjunction	$C_1 \sqcup C_2$	$C_1^{\mathcal{I}} \cup C_2^{\mathcal{I}}$
existential restriction	$\exists R.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}}((x,y) \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}})\}$
universal restriction	$\forall R.C$	$\{x \in \Delta^{\mathcal{I}} \mid \forall y \in \Delta^{\mathcal{I}}((x,y) \in R^{\mathcal{I}} \rightarrow y \in C^{\mathcal{I}})\}$

Interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ where
$\Delta^{\mathcal{I}}$: *domain* of the interpretation and $\cdot^{\mathcal{I}}$: *interpretation function*.

Knowledge Base & Subsumption

$\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$

- *T-box* $\mathcal{T}$ is a set of definitions $C \equiv D$, meaning $C^{\mathcal{I}} = D^{\mathcal{I}}$, where C is the concept name and D is a description
- *A-box* $\mathcal{A}$ contains extensional assertions on concepts and roles e.g. $C(a)$ and $R(a, b)$, meaning, resp., that $a^{\mathcal{I}} \in C^{\mathcal{I}}$ and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$.

Subsumption

Given two concept descriptions C and D, C *subsumes* D, denoted by $C \sqsupseteq D$, iff for every interpretation $\mathcal{I}$, it holds that $C^{\mathcal{I}} \supseteq D^{\mathcal{I}}$

Other Inference Services

instance checking decide whether an individual is an instance of a concept

retrieval find all invididuals instance of a concept

realization problem finding the concepts which an individual belongs to, especially the most specific one, if any:

most specific concept

Given an A-Box $\mathcal{A}$ and an individual a, the *most specific concept* of a w.r.t. $\mathcal{A}$ is the concept C, denoted $\mathrm{MSC}_{\mathcal{A}}(a)$, such that $\mathcal{A} \models C(a)$ and $C \sqsubseteq D$, $\forall D$ such that $\mathcal{A} \models D(a)$.

Normal Form

D is in $\mathcal{ALC}$ *normal form* iff $D \equiv \bot$ or $D \equiv \top$ or if $D = D_1 \sqcup \cdots \sqcup D_n$ ($\forall i = 1, \ldots, n,\ D_i \not\equiv \bot$) with

$$D_i = \bigsqcap_{A \in \mathrm{prim}(D_i)} A \ \sqcap \bigsqcap_{R \in N_R} \left[\forall R.\mathrm{val}_R(D_i) \sqcap \bigsqcap_{E \in \mathrm{ex}_R(D_i)} \exists R.E \right]$$

where:

- $\mathrm{prim}(C)$ set of all (negated) atoms occurring at C's top-level
- $\mathrm{val}_R(C)$ conjunction $C_1 \sqcap \cdots \sqcap C_n$ in the value restriction on R, if any (o.w. $\mathrm{val}_R(C) = \top$);
- $\mathrm{ex}_R(C)$ set of concepts in the value restriction of the role R

For any R, every sub-description in $\mathrm{ex}_R(D_i)$ and $\mathrm{val}_R(D_i)$ is in normal form.

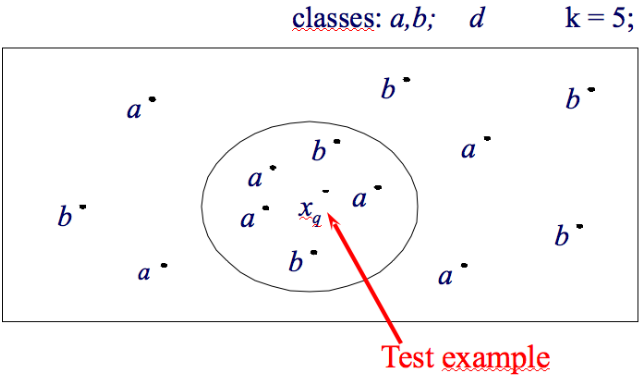
Rationale

- **Main Idea**: similar individuals, by analogy, should likely belong to similar concepts
- Realization of an inductive classification algorithm, applied to ontological knowledge, could be used for:
  - answering to class-membership queries
  - predicting new assertions (that may not be logically entailed by the KB)
  - making the *populating A-Box* task less time consuming, *using new information induced*

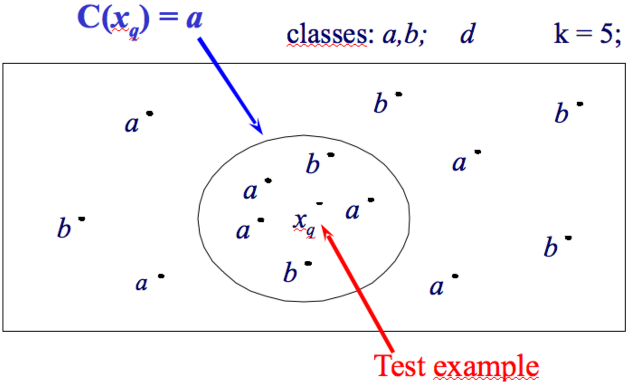
k-NN: Peculiarities

- Lazy Learning Algorithm
    - Learning phase consists in memorizing training example
- Classification results are given by analogy w.r.t. k selected training examples w.r.t. the query example
- Intermediate information and classification results are discarded after the classification of a query example

Classical k-NN algorithm...



...Classical k-NN algorithm



Classical k-NN algorithm: Characteristics

1. Generally applied to feature vector based representation
2. In classification phase it is assumed that each training and test example belong to a single class, so classes are considered to be disjoint
3. An implicit *Closed World Assumption* is made

Difficulties in applying k-NN to Ontological Knowledge

To apply k-NN for classifying individual asserted in an ontological knowledge base

1. It has to find a way for applying k-NN to more complex and expressive representations
2. It is not possible to assume disjointness of classes
   - concept hierarchy (through subsumption)
   - individuals in an ontology can belong to more than one class (concept).
3. The classification process has to cope with the *Open World Assumption* charactering DL reasoning and SW context

Customization to DLs

1. Use of a dissimilarity measure applicable to DL concepts *[d'Amato et all. @ SAC 2006]*
2. A new classification procedure has been adopted
   - the *multi-class* classification problem *decomposed* into smaller *binary classification problems* (one per target concept).
   - For each individual to classify w.r.t each class (concept), classification returns one of $\{-1, +1\}$
3. A third value 0 representing unknown information (uncertain classification) is added in the set of returned values $\{-1, 0, +1\}$
4. Hence a *weighted majority voting criterion* is applied

Details of the New k-NN Algorithm

- **Training Phase:** All training examples are memorized jointly with the classes to which they belong to and their MSCs
- **Classification:** Each example is classified applying the k-NN method for DLs, (adopting the cross validation procedure).

$$\hat{h}_j(x_q) := \operatorname*{argmax}_{v \in V} \sum_{i=1}^{k} \frac{\delta(v, h_j(x_i))}{d(x_q, x_i)^2} \qquad \forall j \in \{1, \ldots, s\}$$

where $V = \{-1, 0, +1\}$, and
$\delta(a, b) = 1$ if $a = b$; $\delta(a, b) = 0$ if $a \neq b$ and

$$h_j(x) = \begin{cases} +1 & C_j(x) \in \mathcal{A} & (\mathcal{K} \vdash C_j(x)) \\ -1 & \neg C_j(x) \in \mathcal{A} & (\mathcal{K} \vdash \neg C_j(x)) \\ 0 & \text{otherwise} & (\text{otherwise}) \end{cases}$$

The Information Content

- A measure of concept (dis)similarity can be derived from the notion of *Information Content* (IC)
- IC depends on the probability of an individual to belong to a certain concept
  - $IC(C) = -\log pr(C)$
- In order to approximate the probability for a concept C, it is possible to recur to its extension
  - $pr(C) = |C^{\mathcal{I}}|/|\Delta^{\mathcal{I}}|$
  - extension estimated by its retrieval wrt the ABox

Function Definition /I

$\mathcal{L} = \mathcal{ALC}/_{\equiv}$ the set of all concepts in $\mathcal{ALC}$ normal form
$\mathcal{I}$ canonical interpretation of A-Box $\mathcal{A}$

$f : \mathcal{L} \times \mathcal{L} \mapsto R^+$ defined $\forall C = \bigsqcup_{i=1}^{n} C_i$ and $D = \bigsqcup_{j=1}^{m} D_j$ in $\mathcal{L}_{\equiv}$

$$f(C, D) := f_{\sqcup}(C, D) = \begin{cases} 0 & C \equiv D \\ \infty & C \sqcap D \equiv \bot \\ \max_{\substack{i = 1, \ldots, n \\ j = 1, \ldots, m}} f_{\sqcap}(C_i, D_j) & \text{o.w.} \end{cases}$$

$$f_{\sqcap}(C_i, D_j) := f_P(\text{prim}(C_i), \text{prim}(D_j)) + f_{\forall}(C_i, D_j) + f_{\exists}(C_i, D_j)$$

Function Definition / II

$$f_P(\text{prim}(C_i), \text{prim}(D_j)) := \begin{cases} \infty & \text{if } \text{prim}(C_i) \sqcap \text{prim}(D_j) \equiv \bot \\ \frac{IC(\text{prim}(C_i) \sqcap \text{prim}(D_j))+1}{IC(LCS(\text{prim}(C_i),\text{prim}(D_j)))+1} & \text{o.w.} \end{cases}$$

$$f_\forall(C_i, D_j) := \sum_{R \in N_R} f_\sqcup(\text{val}_R(C_i), \text{val}_R(D_j))$$

$$f_\exists(C_i, D_j) := \sum_{R \in N_R} \sum_{k=1}^{N} \max_{p=1,\ldots,M} f_\sqcup(C_i^k, D_j^p)$$

where $C_i^k \in \text{ex}_R(C_i)$ and $D_j^p \in \text{ex}_R(D_j)$ and wlog.
$N = |\text{ex}_R(C_i)| \geq |\text{ex}_R(D_j)| = M$, otherwise exchange N with M

Dissimilarity Measure

The *dissimilarity measure* d is a function $d : \mathcal{L} \times \mathcal{L} \mapsto [0,1]$ such that, for all $C = \bigsqcup_{i=1}^{n} C_i$ and $D = \bigsqcup_{j=1}^{m} D_j$ concept descriptions in $\mathcal{ALC}$ normal form:

$$d(C,D) := \begin{cases} 0 & f(C,D) = 0 \\ 1 & f(C,D) = \infty \\ 1 - \frac{1}{f(C,D)} & otherwise \end{cases}$$

where f is the function defined previously

Measure Involving Individuals

Let c and d two individuals in a given A-Box.

Consider $C^* = \mathrm{MSC}^*(c)$ and $D^* = \mathrm{MSC}^*(d)$:

Then:

$$d(c, d) := d(C^*, D^*) = d(\mathrm{MSC}^*(c), \mathrm{MSC}^*(d))$$

Analogously:

$$\forall c: \ d(c, D) := d(\mathrm{MSC}^*(c), D)$$

Experimentation Setting

- **FSM** (Protege Library) made by:
  - 20 concepts, 10 object properties, 37 individuals
- **Surface-Water-Model** (Protege Library) made by:
  - 19 concepts, 9 object properties, 115 individuals
- **Family** (handmade):
  - 14 concepts, 5 object properties, 39 individuals

Leave-one-out cross validation

Measures for Evaluating Experiments

- **Predictive Accuracy:** measures the number of correctly classified individuals w.r.t. overall number of individuals.
- **Omission Error Rate:** measures the amount of unlabelled individuals $C(x_q) = 0$ with respect to a certain concept C_j while they are instances of C_j in the KB.
- **Commission Error Rate:** measures the amount of individuals labelled as instances of the negation of the target concept C_j, while they belong to C_j or vice-versa.
- **Induction Rate:** measures the amount of individuals that were found to belong to a concept or its negation, while this information is not derivable from the KB.

Experimentation Evaluation

Average results of the trials

Ontologies	Predictive Accuracy	Omission Error	Induction Rate	Commission Error
FSM	100	0	31	0
S.-W.-M.	100	0	0	0
FAMILY	49.07	50.93	16.85	0

Experimentation: Discussion...

- for every ontology, the *commission error is null*; the classifier never makes critical mistakes
- **SURFACE-WATER-MODEL Ontology**: the classifier always assigns individuals to the correct concepts; it is never capable to induce new knowledge
  - Since individuals are all instances of a single concept and are involved in a few roles, so MSCs are very similar and thus the amount of information they convey is very low

...Experimentation: Discussion...

FSM Ontology

- The classifier always assigns individuals to the correct concepts
  - Because most of individuals are instances of a single concept
- Induction rate is not null so new knowledge is induced
  - mainly due to the presence of some concepts that are declared to be mutually disjoint
  - also because some individuals are involved in relations

...Experimentation: Discussion

FAMILY Ontology

- Predictive Accuracy is not so high and Omission Error not null
  - Since instances are more irregularly spread over the classes, so computed MSCs are often very different provoking sometimes incorrect classifications (weakness on k-NN algorithm)
- No Commission Error (but only omission error)
- The *Classifier* is able of *induce new knowledge* that is *not logically derivable*

Family Ontology Results

	Predictive Accuracy	Omission Error	Induction Rate	Commission Error
Female	75	25	30.76	0
Woman	75	25	35.89	0
Mother	0	100	30.76	0
Male	83.36	16.64	30.76	0
Man	83.36	16.64	33.33	0
Father	14.28	85.72	30.76	0
Human	100	0	2.56	0
Child	80.95	19.05	12.82	0
Sibling	0	100	0	0
Parent	37.5	62.5	12.82	0
Grandparent	50	50	5.12	0
Grandchild	37.5	62.5	12.82	0
Cousin	50	50	0	0
UncleAunt	0	100	0	0
average	49.07	50.93	16.85	0

Conclusions

- A classification method for inductive inference on DLs ABoxs has been presented
  - It can be used for predicting/suggesting missing information about individuals
- The proposed method is able to induce new assertions, in addition to the knowledge already derivable by a reasoner
  - An increase in predictive accuracy was observed when the instances are homogeneously spread
- A dissimilarity measure for $\mathcal{ALC}$ DL has been tested

Ongoing & Future Work

- Extend the classification method with a modified answering procedure grounded on statistical inference, in order to accept answers as correct with a high degree of confidence
- Define new similarity/dissimilarity measures for most expressive DL such as $\mathcal{ALCN}$
- *Applying* the Dissimilarity Measure in distance-based *clustering algorithms*

The End

That's all!

Questions?

- Claudia **d'Amato**

  claudia.damato@di.uniba.it
- Nicola **Fanizzi**

  fanizzi@di.uniba.it
- Floriana **Esposito**

  esposito@di.uniba.it