Uncertainty Reasoning through Similarity in Context

Claudia d'Amato Nicola Fanizzi

Dipartimento di Informatica
Università degli studi di Bari, *Italy*

2nd ARCOE Workshop @ ECAI2010, *Lisbon*, PT

Motivation

Reasoning with *Web ontologies*
expressed in standard representations based on *Description Logics*
difficult due to

- inherent incompleteness: OWA vs. CWA
- incoherence (+*noise*): heterogeneous and distributed sources

Various solutions investigated in the URSW community,
e.g modeling vague knowledge in terms of *probability* and *fuzziness*:
support to evolution ?

Idea

try inductive methods:
often efficient, noise-tolerant and incremental

In particular,

- methods based on **similarity** (or a notion of **distance**) proposed for many reasoning tasks, cast as inductive problems

In the literature: most of the measures for concept-similarity

- inductive techniques borrowed from *Machine Learning* require a notion of similarity between individuals

Outline

Survey of applications of similarity in context

1. Preliminaries
2. Contextual Metrics for Individuals
   - Similarity in Context
   - Family of Metrics
3. Inductive Instance Classification
   - Problem
   - k-Nearest Neighbor Procedure
4. Rough DLs
   - Rough Concept Approximations
   - Induced Indiscernibility Relation
   - Extensions
5. Conclusions and Outlook

Preliminaries I

Axioms in terms of a *vocabulary* of

- N_C set of primitive concept names
- N_R set of primitive role names
- N_I set of individual names

and language *constructors*

Semantics defined by interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$
where

- $\Delta^{\mathcal{I}}$ domain of the interpretation (non-empty)
- $\cdot^{\mathcal{I}}$ interpretation function that maps names to *extensions*
  each $A \in N_C$ to a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and
  each $R \in N_R$ to $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$

Then new concepts/roles defined using the language *constructors*

DL Knowledge Bases

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

TBox $\mathcal{T}$ set of axioms $C \sqsubseteq D$ (resp. $C \equiv D$)
meaning $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ (resp. $C^{\mathcal{I}} = D^{\mathcal{I}}$)
where C is atomic and D is a concept description

ABox $\mathcal{A}$ set of assertions — ground axioms
e.g. $C(a)$ and $R(a, b)$ stating:
a belongs to C and (a, b) belongs to R

- $\mathbf{Ind}(\mathcal{A})$ = set of individuals occurring in $\mathcal{A}$

Interpretations of interest (models) satisfy all the axioms in $\mathcal{K}$

Context & Similarity I

A context of reference must express the essential features for comparing domain objects.

- similarity is not merely a relation between objects but rather between the two in a given context (which is subject to changes) [Goldstone et al.,1997]
- the *task* also matters !

Context & Similarity II

In the following...

Context

Given a knowledge base $\mathcal{K}$,
a context $\mathbf{C}$ is a finite set of concept descriptions (*features*)

$$\mathbf{C} = \{F_1, F_2, \ldots, F_m\}$$

built on concepts and roles defined in $\mathcal{K}$

Learning the Context

- given a fitness / criterion function J for the task
- methods for finding contexts based on *distinguishability* proposed; stochastic search using
  - *Genetic Programming*
  - *Simulated Annealing*

Alternatively, since the metrics are based on weighted projections:

- consider as many features as possible (e.g. all defined concepts)
- find good choice for the weights $\vec{w}$
  - based on information (*entropy*)
  - based on variance

A Family of Metrics

Given a context **C** and a weight vector $\vec{w}$,
the *family* $\{d_p^{\mathsf{C}}\}_{p \in \mathbb{N}}$ *of functions* $d_p^{\mathsf{C}} : \mathbf{Ind}(\mathcal{A}) \times \mathbf{Ind}(\mathcal{A}) \mapsto [0, 1]$ is defined

$$d_p^{\mathsf{C}}(a, b) = \left[\sum_{i=1}^{m} w_i \mid \pi_i(a) - \pi_i(b) \mid^p \right]^{1/p}$$

where $\forall i \in \{1, \ldots, m\}$ the i-th projection function π_i:

$$\pi_i(a) = \begin{cases} 1 & \mathcal{K} \models F_i(a) \\ 0 & \mathcal{K} \models \neg F_i(a) \\ u_i & \text{(prior) } \textit{otherwise} \end{cases}$$

Inspired by *Minkowski*'s norms; can be proven to be semi-distances

Inductive Classification

Instance checking as a

Learning Problem

given a query concept Q and
a query individual x_q

using S_Q sample of prototype training instances
with correct membership values
$h_Q(x_i) = v \in \{-1, 0, +1\} = V$

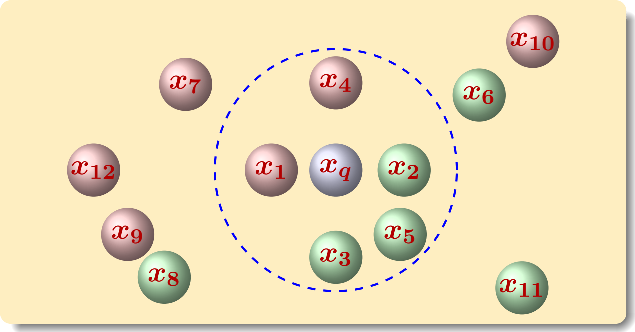
determine $\hat{h}_Q(x_q)$
i.e. *estimate* membership of x_q w.r.t. Q

We use well known non-parametric methods:

- k-NN, Parzen Windows — no ind. model, only rel. distances
- RBF Nets, SVMs ... — build an inductive model

k-Nearest Neighbor Procedure I

A sort of analogical reasoning [d'Amato et al.,2008-URSW I]



Selection of the $k = 5$ nearest neighbors.
green=positive ex., red=negative ex.

k-Nearest Neighbor Procedure II

Weighted majority vote:
given $NN_k(x_q) = \{x_1, \ldots, x_k\}$ of x_q's nearest neighbors w.r.t. d_p^{C}, the estimate of the membership hypothesis is

$$\hat{h}_Q(x_q) = \operatorname*{argmax}_{v \in V} \sum_{i=1}^{k} \overbrace{\gamma(d_p^{\mathsf{C}}(x_i, x_q))}^{\text{proximity weight}} \cdot \overbrace{\delta(v, h(x_i))}^{\text{vote}}$$

where:

- δ Kronecker indicator function
- γ decaying function
  e.g. $\gamma(x) = (1 - x)^b$ or $\gamma(x) = 1/x^b$ for some $b > 0$

Lessons Learned

Applying this and similar methods based on density estimates (RBF Networks, SVMs, . . .)

- build the inductive model *once* and classify efficiently *many times*
- may give an answer in case of uncertain class-membership (can be forced to do that)
- may provide an estimate of the likelihood of the answer
- experimentally: nearly sound and complete (few omission errors)
- measure used also in unsupervised tasks: e.g. *clustering* individual resources

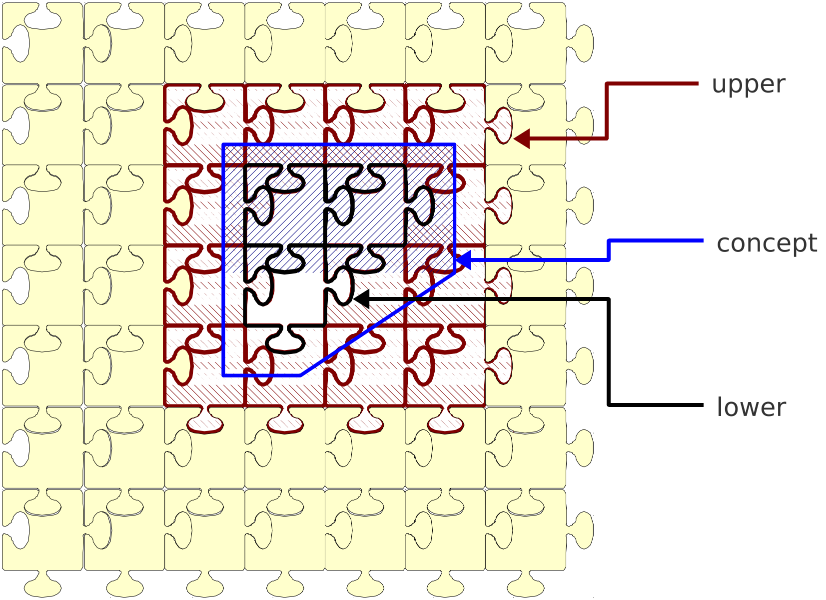
Rough DL

Recently Rough DLs introduced [Schlobach et al.,IJCAI2007] as a mechanism for modeling *vague* concepts by means of a *crisp* specification of its approximations

Approximations

Given an indiscernibility relation R,
the upper approximation of a concept C is
$\overline{C} = \{a \mid \exists b : R(a,b) \wedge b \in C\}$ (typical instances)
the lower approximation is
$\underline{C} = \{a \mid \forall b : R(a,b) \rightarrow b \in C\}$ (prototypical instances)

If R expressed in terms of the knowledge base then standard reasoners can be used


upper
concept
lower

Induced Indiscernibility Relation I

Idea induce an equivalence relation based on context

[Fanizzi et al.,URSW2008]

Given a context **C** and the related projection functions π_i
two individuals a and b are indiscernible w.r.t. **C** iff

$$\forall i \in \{1, \ldots, m\} \colon \; \pi_i(a) = \pi_i(b)$$

Indiscernibility relation R_{C} induced by **C** defined:

$$R_{\mathsf{C}} = \{(a, b) \in N_I \times N_I \mid \forall i \in \{1, \ldots, m\} \colon \; \pi_i(a) = \pi_i(b)\}$$

Induced Indiscernibility Relation II

R partitions N_I in equivalence classes.
given a generic individual a, the *induced concept* is $C_a = [a]_{\mathsf{C}}$

Extension of a C-definable concept corresponds to a combination (union) of equivalence classes.

Other concept descriptions, say D, may be approximated contextual upper and lower approximations of D w.r.t. C, defined:

$$\begin{aligned} \overline{D}^{\mathsf{C}} &= \{a \in N_I \mid C_a \sqcap D \not\models \bot\} \\ \underline{D}_{\mathsf{C}} &= \{a \in N_I \mid C_a \sqsubseteq D\} \end{aligned}$$

Extensions I

Using a notion of tolerance [Doherty et al.,2003]:

Definition (tolerance)

A tolerance function on a set U is a function $\tau : U \times U \mapsto [0, 1]$ such that $\forall a, b \in U$
(1) $\tau(a, a) = 1$ and
(2) $\tau(a, b) = \tau(b, a)$

Given τ on U and a threshold $\theta \in [0, 1]$,
a neighborhood function $\nu_\theta : U \mapsto 2^U$ is

$$\nu_\theta(a) = \{b \in U \mid \tau(a, b) \geq \theta\}$$

a.k.a. the θ-neighborhood of a

Extensions II

Consider $N_I = U$ and a metric d_p^{C}.

Equivalence relationships R_{C}^{θ} on N_I can be defined,
with classes made up of individuals within a degree of similarity,
controlled by θ:

$$[a]_{\mathsf{C}} = \nu_{\theta}(a)$$

Upper and lower approximation w.r.t. R_{C}^{θ} descend straightfowardly

Contribution and Outlook

An operative notion of similarity in context for reasoning with ontologies

Two applications:

1. inductive instance classification
2. Vague concept modeling (+reasoning) in Rough-DL

instance-based: naturally evolve

Ongoing / Future Work

- kernel methods based on similar settings
- ranking answers w.r.t. likelihood measure
- unsupervised tasks. outlier and novelty detection, clustering

time for questions

Offline

- Claudia d'Amato claudia.damato@di.uniba.it
- Nicola Fanizzi fanizzi@di.uniba.it