Similarity-based Learning Methods for the Semantic Web

Claudia d'Amato

Dipartimento di Informatica • *Università degli Studi di Bari*
Campus Universitario, Via Orabona 4, 70125 Bari, Italy

Trento, 15 Ottobre 2007

Contents

The Semantic Web

- **Semantic Web goal:** make the Web contents *machine-readable* and *processable* besides of human-readable
- **How to reach** the *SW goal*:
  - Adding meta-data to Web resources
  - Giving a *shareable and common semantics* to the meta-data by means of *ontologies*
- Ontological knowledge is generally described by the *Web Ontology Language (OWL)*
  - Supported by *well-founded semantics* of *DLs*
  - together with a series of available automated *reasoning services* allowing to derive logical consequences from an ontology

Motivations...

- The main approach used by inference services is *deductive reasoning*.
  - Helpful for computing class hierarchy, ontology consistency
- Conversely, tasks as *ontology learning, ontology population by assertions, ontology evaluation, ontology mapping* require inferences able to return *higher general conclusions w.r.t. the premises*.
- **Inductive learning methods**, based on *inductive reasoning*, could be effectively used.

...Motivations

- Inductive reasoning generates *conclusions* that are of *greater generality* than the premises.
- The starting *premises* are specific, typically *facts or examples*
- *Conclusions* have *less certainty* than the premises.
- The **goal** is to formulate plausible *general assertions explaining the given facts* and that are able to *predict new facts*.

Goals

- Apply ML methods, particularly *instance based learning methods*, to the SW and SWS fields for
  - **improving reasoning** procedures
  - **inducing new knowledge** not logically derivable
  - improving **efficiency** and **effectiveness** of: *ontology population*, *query answering*, *service discovery and ranking*
- Most of the instance-based learning methods require (dis-)similarity measures
  - **Problem:** Similarity measures for complex concept descriptions (as those in the ontologies) is a field not deeply investigated **[Borgida et al. 2005]**
- **Solution:** Define new measures for ontological knowledge
  - able to cope with the OWL high expressive power

The Representation Language...

- *DLs* is the *theoretical foundation* of *OWL* language
  - standard de facto for the knowledge representation in the SW
- Knowledge representation by means of Description Logic
  - $\mathcal{ALC}$ logic is *mainly considered* as satisfactory compromise between *complexity* and *expressive power*

...The Representation Language

- Primitive *concepts* $N_C = \{C, D, \ldots\}$: subsets of a domain
- Primitive *roles* $N_R = \{R, S, \ldots\}$: binary relations on the domain
- *Interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ where
  $\Delta^{\mathcal{I}}$: *domain* of the interpretation and $\cdot^{\mathcal{I}}$: *interpretation function*:

Name	Syntax	Semantics
top concept	$\top$	$\Delta^{\mathcal{I}}$
bottom concept	$\bot$	$\emptyset$
concept	C	$C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
full 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	$\forall R.C$	$\{x \in \Delta^{\mathcal{I}} \mid \forall y \in \Delta^{\mathcal{I}}((x, y) \in R^{\mathcal{I}} \rightarrow y \in C^{\mathcal{I}})\}$

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

least common subsumer is the most specific concept that subsumes a set of considered concepts

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 $\mathcal{A} \models D(a)$.

Classify Measure Definition Approaches

- **Dimension Representation**: feature vectors, strings, sets, trees, clauses...
- **Dimension Computation**: geometric models, feature matching, semantic relations, Information Content, alignment and transformational models, contextual information...
- Distinction: *Propositional* and *Relational* setting
  - analysis of computational models

Propositional Setting: Measures based on Geometric Model

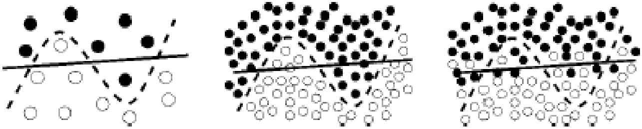
- **Propositional Setting**: Data are represented as n-tuple of fixed length in an n-dimentional space
- **Geometric Model:** objects are seen as *points in an n-dimentional space*.
  - The *similarity* between a pair of objects is considered *inversely related to the distance* between two objects points in the space.
  - Best known distance measures: *Minkowski* measure, *Manhattan* measure, *Euclidean* measure.
- Applied to vectors whose *features* are *all continuous*.

Kernel Functions

- *Similarity functions able to work with high dimensional feature spaces*.
- Developed jointly with **kernel methods**: efficient learning algorithms realized for solving classification, regression and clustering problems in high dimensional feature spaces.
  - **Kernel machine**: encapsulates the learning task
  - **kernel function**: encapsulates the hypothesis language
- Introduced in the field pattern recognition
  - Simplest goal: *estimate a function* using I/O training data able to correctly classify unseen examples (x, y)
  - y is determined such that (x, y) is in some sense similar to the training examples.
  - A similarity measure k is necessary

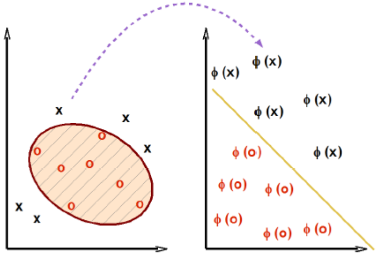
...Kernel Functions...

- *Possible Problem*: overfitting for small sample sizes
- *Intuition:* a "simple" (e.g., linear) function minimizing the error and that explains most of the data is preferable to a complex one (Occams razor).
    - *algorithms in feature spaces* target a *linear function* for performing the *learning task*.
- *Issue*: not always possible to find a linear function



...Kernel Functions

- *Solution:* mapping the initial feature space in a higher dimensional space where the learning problem can be solved by a linear function
- A *kernel function* performs such a *mapping implicitly*
  - **Any set that admits a positive definite kernel can be embedded into a linear space** [Aronsza 1950]



Similarity Measures based on Feature Matching Model

- **Features** can be of **different types**: binary, nominal, ordinal
- *Tversky's Similarity Measure*: based on the notion of *contrast model*
  - **common features** tend to **increase** the perceived similarity of two concepts
  - **feature differences** tend to **diminish** perceived similarity
  - feature *commonalities increase* perceived similarity *more than feature differences* can diminish it
  - it is assumed that *all features have the same importance*
- **Measures in propositional setting are not able to capture expressive relationships among data** that typically characterize most complex languages.

Relational Setting: Measures Based on Semantic Relations

- Also called **Path distance measures** [Bright,94]
- Measure the *similarity* value between single words (*elementary concepts*)
- concepts (words) are organized in a *taxonomy* using hypernym/hyponym and synoym links.
- the measure is a (weighted) *count of the links* in the path *between two terms* w.r.t. the most specific ancestor
  - terms with a **few links** separating them are semantically **similar**
  - terms with **many links** between them have **less similar** meanings
  - link counts are weighted because different relationships have different implications for semantic similarity.

Measures Based on Semantic Relations: WEAKNESS

- the similarity value is subjective due to the taxonomic ad-hoc representation
- the introduction of news term can change similarity values
- the similarity measures cannot be applied directly to the knowledge representation
  - it needs of an intermediate step which is building the term taxonomy structure
- only "linguistic" relations among terms are considered; there are not relations whose semantics models domain

Measures Based on Information Content...

- Measure semantic similarity of concepts in an *is-a* taxonomy by the use of notion of *Information Content (IC)* [Resnik,99]
- Concepts similarity is given by the shared information
  - The *shared information* is represented by a *highly specific super-concept* that subsumes both concepts
- *Similarity value* is given by the *IC of the least common super-concept*
  - *IC for a concept is determined* considering the probability that an instance belongs to the concept

...Measures Based on Information Content

- Use a criterion similar to those used in *path distance measures*,
- Differently from *path distance measures*, the use of probabilities **avoids the unreliability of counting edge** when changing in the hierarchy occur
- **The considered relation among concepts is only is-a relation**
  - **more semantically expressive relations cannot be considered**

Miscellaneous Approaches

- *Propositionalization and Geometrical Models*
- Path Distance and Feature Matching Approaches
- Feature Matching, Context-based and Information Content-based Approaches
- *Geometrical models* are largely used for their *efficiency*, but cab be applied only to *propositional* representations.
- *Idea*: focus the propositionalization problem
  - Find a way for transforming a multi-relational representation into a propositional representation.
  - Hence any method can be applied on the new representation rather than on the original one
  - *Hipothesis-driven distance* [Sebag 1997]: a method for building a distance on first-order logic representation by recurring to the propositionalization is presented

Relational Kernel Functions...

- Motivated by the necessity of solving real-world problems in an efficient way.
- Best known relational kernel function: the **convolution kernel** [Haussler 1999]
- *Basic idea*: the semantics of a composite object can be captured by a relation R between the object and its parts.
  - The kernel is composed of kernels defined on different parts.
- *Obtained by composing existing kernels by a certain sum over products*, exploiting the closure properties of the class of positive definite functions.

$$k(x, y) = \sum_{\overrightarrow{x} \in R^{-1}(x), \overrightarrow{y} \in R^{-1}(y)} \prod_{d=1}^{D} k_d(x_d, y_d) \qquad (1)$$

...Relational Kernel Functions

- The term "convolution kernel" refers to a class of kernels that can be formulated as shown in (1).
- Exploiting convolution kernel, string kernels, tree kernel, graph kernels etc.. have been defined.
- The **advantage of convolution kernels is that they are very general** and can be applied in several situations.
- **Drawback:** due to their generality, a significant amount of work is required to adapt convolution kernel to a specific problem
    - **Choosing R in real-world applications is a non-trivial task**

Similarity Measures for Very Low Expressive DLs...

- Measures for complex concept descriptions [Borgida et al. 2005]
  - A DL allowing only *concept conjunction* is considered (propositional DL)
- **Feature Matching Approach**:
  - features are represented by atomic concepts
  - An ordinary concept is the conjunction of its features
  - *Set intersection* and *difference* corresponds to the *LCS* and *concept difference*
- **Semantic Network Model** and **IC models**
  - The *most specific ancestor* is given by the *LCS*

...Similarity Measures for Very Low Expressive DLs

OPEN PROBLEMS in considering most expressive DLs:

- What is a *feature* in most expressive DLs?
  - i.e. $(\leq 3R), (\leq 4R)$ and $(\leq 9R)$ are three different features? or $(\leq 3R), (\leq 4R)$ are more similar w.r.t $(\leq 9R)$?
  - How to assess similarity in presence of role restrictions? i.e. $\forall R.(\forall R.A)$ and $\forall R.A$
- Key problem in *network-based measures*: how to assign a useful size for the various concepts in the description?
- *IC-based model*: how to compute the value $p(C)$ for assessing the IC?

Why New Measures

- **Already defined** similalrity/dissimilalrity **measures cannot be** directly **applied to ontological knowledge**
  - They define similarity value between *atomic concepts*
  - They are defined for *representation less expressive* than ontology representation
  - They *cannot exploit all the expressiveness* of the *ontological* representation
  - **There are no measure for assessing similarity between individuals**
- **Defining new measures that are really semantic is necessary**

Similarity Measure between Concepts: Needs

- Necessity to have a measure really based on Semantics
- Considering [Tversky'77]:
  - common features tend to increase the perceived similarity of two concepts
  - feature differences tend to diminish perceived similarity
  - feature commonalities increase perceived similarity more than feature differences can diminish it
- The proposed similarity measure is:

Similarity Measure between Concepts

Definition [d'Amato et al. @ CILC 2005]: Let $\mathcal{L}$ be the set of all concepts in $\mathcal{ALC}$ and let $\mathcal{A}$ be an A-Box with canonical interpretation $\mathcal{I}$. The *Semantic Similarity Measure* s is a function

$$s : \mathcal{L} \times \mathcal{L} \mapsto [0, 1]$$

defined as follows:

$$s(C, D) = \frac{|I^{\mathcal{I}}|}{|C^{\mathcal{I}}| + |D^{\mathcal{I}}| - |I^{\mathcal{I}}|} \cdot \max\left(\frac{|I^{\mathcal{I}}|}{|C^{\mathcal{I}}|}, \frac{|I^{\mathcal{I}}|}{|D^{\mathcal{I}}|}\right)$$

where $I = C \sqcap D$ and $(\cdot)^{\mathcal{I}}$ computes the concept extension wrt the interpretation $\mathcal{I}$.

Similarity Measure: Meaning

- If $C \equiv D$ ($C \sqsubseteq D$ and $D \sqsubseteq C$)then $s(C, D) = 1$, i.e. the maximum value of the similarity is assigned.
- If $C \sqcap D = \perp$ then $s(C, D) = 0$, i.e. the minimum similarity value is assigned because concepts are totally different.
- Otherwise $s(C, D) \in]0, 1[$. The *similarity* value is *proportional* to the *overlapping* amount of the concept extetions *reduced by* a quantity representing how the two concepts are near to the overlap. This means considering similarity not as an absolute value but as weighted w.r.t. *a degree of non-similarity*.

Similarity Measure: Example...

Primitive Concepts: $N_C = \{\text{Female}, \text{Male}, \text{Human}\}$.
Primitive Roles:
$N_R = \{\text{HasChild}, \text{HasParent}, \text{HasGrandParent}, \text{HasUncle}\}$.
$\mathcal{T} = \{$ Woman $\equiv$ Human $\sqcap$ Female; Man $\equiv$ Human $\sqcap$ Male
Parent $\equiv$ Human $\sqcap$ $\exists$HasChild.Human
Mother $\equiv$ Woman $\sqcap$ Parent $\exists$HasChild.Human
Father $\equiv$ Man $\sqcap$ Parent
Child $\equiv$ Human $\sqcap$ $\exists$HasParent.Parent
Grandparent $\equiv$ Parent $\sqcap$ $\exists$HasChild.($\exists$ HasChild.Human)
Sibling $\equiv$ Child $\sqcap$ $\exists$HasParent.($\exists$ HasChild $\geq$ 2)
Niece $\equiv$ Human $\sqcap$ $\exists$HasGrandParent.Parent $\sqcup$ $\exists$HasUncle.Uncle
Cousin $\equiv$ Niece $\sqcap$ $\exists$HasUncle.($\exists$ HasChild.Human)$\}$.

...Similarity Measure: Example...

$\mathcal{A}$ = {Woman(Claudia), Woman(Tiziana), Father(Leonardo), Father(Antonio), Father(AntonioB), Mother(Maria), Mother(Giovanna), Child(Valentina), Sibling(Martina), Sibling(Vito), HasParent(Claudia,Giovanna), HasParent(Leonardo,AntonioB), HasParent(Martina,Maria), HasParent(Giovanna,Antonio), HasParent(Vito,AntonioB), HasParent(Tiziana,Giovanna), HasParent(Tiziana,Leonardo), HasParent(Valentina,Maria), HasParent(Maria,Antonio), HasSibling(Leonardo,Vito), HasSibling(Martina,Valentina), HasSibling(Giovanna,Maria), HasSibling(Vito,Leonardo), HasSibling(Tiziana,Claudia), HasSibling(Valentina,Martina), HasChild(Leonardo,Tiziana), HasChild(Antonio,Giovanna), HasChild(Antonio,Maria), HasChild(Giovanna,Tiziana), HasChild(Giovanna,Claudia), HasChild(AntonioB,Vito), HasChild(AntonioB,Leonardo), HasChild(Maria,Valentina), HasUncle(Martina,Giovanna), HasUncle(Valentina,Giovanna) }

...Similarity Measure: Example

$$
\begin{aligned}
s(\mathsf{Grandparent}, \mathsf{Father}) &= \frac{|(\mathsf{Grandparent} \sqcap \mathsf{Father})^{\mathcal{I}}|}{|\mathsf{Granparent}^{\mathcal{I}}| + |\mathsf{Father}^{\mathcal{I}}| - |(\mathsf{Grandarent} \sqcap \mathsf{Father})^{\mathcal{I}}|} \cdot \\
&\quad \cdot\ max\left(\frac{|(\mathsf{Grandparent} \sqcap \mathsf{Father})^{\mathcal{I}}|}{|\mathsf{Grandparent}^{\mathcal{I}}|}, \frac{|(\mathsf{Grandparent} \sqcap \mathsf{Father})^{\mathcal{I}}|}{|\mathsf{Father}^{\mathcal{I}}|}\right) = \\
&= \frac{2}{2+3-2} \cdot max\left(\frac{2}{2}, \frac{2}{3}\right) = 0.67
\end{aligned}
$$

Similarity Measure between Individuals

Let c and d two individuals in a given A-Box.
We can consider $C^* = \mathrm{MSC}^*(c)$ and $D^* = \mathrm{MSC}^*(d)$:

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

Analogously:

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

Similarity Measure: Conclusions...

- s is a *Semantic* Similarity measure
  - It uses only *semantic inference* (Instance Checking) for determining similarity values
  - It does *not make use of the syntactic structure* of the concept descriptions
  - It does *not add complexity besides of* the complexity of *used inference operator* (*IChk* that is PSPACE in $\mathcal{ALC}$)
- Dissimilarity Measure is defined using the set theory and reasoning operators
  - **It uses a numerical approach but it is applied to symbolic representations**

...Similarity Measure: Conclusions

- Experimental evaluations demonstrate that s works satisfying when it is applied between concepts
- s applied to individuals is often zero even in case of similar individuals
  - The MSC^* is so specific that often covers only the considered individual and not similar individuals
- The *new idea* is to measure the similarity (dissimilarity) of the subconcepts that build the MSC^* concepts in order to find their similarity (dissimilarity)
  - *Intuition*: Concepts defined by almost the same sub-concepts will be probably similar.

$\mathcal{ALC}$ 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.

Overlap Function

Definition [d'Amato et al. @ KCAP 2005 Workshop]:

$\mathcal{L} = \mathcal{ALC}/_{\equiv}$ 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 $\mathcal{L}_{\equiv}$

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

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

Overlap Function / II

$$f_P(\mathrm{prim}(C_i), \mathrm{prim}(D_j)) := \frac{|(\mathrm{prim}(C_i))^{\mathcal{I}} \cup (\mathrm{prim}(D_j))^{\mathcal{I}}|}{|((\mathrm{prim}(C_i))^{\mathcal{I}} \cup (\mathrm{prim}(D_j))^{\mathcal{I}}) \setminus ((\mathrm{prim}(C_i))^{\mathcal{I}} \cap (\mathrm{prim}(D_j))^{\mathcal{I}})|}$$

$$f_P(\mathrm{prim}(C_i), \mathrm{prim}(D_j)) := \infty \text{ if } (\mathrm{prim}(C_i))^{\mathcal{I}} = (\mathrm{prim}(D_j))^{\mathcal{I}}$$

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

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

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

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

where f is the function overlapping

Discussion

- If C $\equiv$ D (namely C $\sqsubseteq$ D e D $\sqsubseteq$ C) (semantic equivalence) $d(C, D) = 0$, rather d assigns the minimun value
- If $C \sqcap D \equiv \perp$ then $d(C, D) = 1$, rather d assigns the maximum value because concepts involved are totally different
- Otherwise $d(C, D) \in]0, 1[$ rather *dissimilarity is inversely proportional to the quantity of concept overlap*, measured considering the entire definitions and their subconcepts.

Dissimilarity Measure: example...

$C \equiv A_2 \sqcap \exists R.B_1 \sqcap \forall T.(\forall Q.(A_4 \sqcap B_5)) \sqcup A_1$
$D \equiv A_1 \sqcap B_2 \sqcap \exists R.A_3 \sqcap \exists R.B_2 \sqcap \forall S.B_3 \sqcap \forall T.(B_6 \sqcap B_4) \sqcup B_2$
where A_i and B_j are all primitive concepts.

$C_1 := A_2 \sqcap \exists R.B_1 \sqcap \forall T.(\forall Q.(A_4 \sqcap B_5))$
$D_1 := A_1 \sqcap B_2 \sqcap \exists R.A_3 \sqcap \exists R.B_2 \sqcap \forall S.B_3 \sqcap \forall T.(B_6 \sqcap B_4)$

$$f(C,D) := f_{\sqcup}(C,D) = \max\{\ f_{\sqcap}(C_1,D_1), f_{\sqcap}(C_1,B_2), f_{\sqcap}(A_1,D_1), f_{\sqcap}(A_1,B_2)\ \}$$

...Dissimilarity Measure: example...

For brevity, we consider the computation of $f_\sqcap(C_1, D_1)$.

$f_\sqcap(C_1, D_1) = f_P(\text{prim}(C_1), \text{prim}(D_1)) + f_\forall(C_1, D_1) + f_\exists(C_1, D_1)$
Suppose that $(A_2)^\mathcal{I} \neq (A_1 \sqcap B_2)^\mathcal{I}$. Then:

$$\begin{aligned} f_P(C_1, D_1) &= f_P(\text{prim}(C_1), \text{prim}(D_1)) \\ &= f_P(A_2, A_1 \sqcap B_2) \\ &= \frac{|I|}{|I \setminus ((A_2)^\mathcal{I} \cap (A_1 \sqcap B_2)^\mathcal{I})|} \end{aligned}$$

where $I := (A_2)^\mathcal{I} \cup (A_1 \sqcap B_2)^\mathcal{I}$

...Dissimilarity Measure: example...

In order to calculate $f_\forall$ it is important to note that

- There are two different role at the same level T and S
- So the summation over the different roles is made by two terms.

$$
\begin{aligned}
f_\forall(C_1, D_1) &= \sum_{R \in N_R} f_\sqcup(\mathrm{val_R}(C_1), \mathrm{val_R}(D_1)) = \\
&= f_\sqcup(\mathrm{val_T}(C_1), \mathrm{val_T}(D_1)) + \\
&+ f_\sqcup(\mathrm{val_S}(C_1), \mathrm{val_S}(D_1)) = \\
&= f_\sqcup(\forall Q.(A_4 \sqcap B_5), B_6 \sqcap B_4) + f_\sqcup(\top, B_3)
\end{aligned}
$$

...Dissimilarity Measure: example

In order to calculate $f_\exists$ it is important to note that

- There is only a single one role R so the first summation of its definition collapses in a single element
- N and M (numbers of existential concept descriptions w.r.t the same role (R)) are $N = 2$ and $M = 1$
  - So we have to find the max value of a single element, that can be semplifyed.

$$
\begin{aligned}
f_\exists(C_1, D_1) &= \sum_{k=1}^{2} f_\sqcup(\mathrm{ex}_\mathrm{R}(C_1), \mathrm{ex}_\mathrm{R}(D_1^k)) = \\
&= f_\sqcup(B_1, A_3) + f_\sqcup(B_1, B_2)
\end{aligned}
$$

Dissimilarity Measure: Conclusions

- Experimental evaluations demonstrate that d *works satisfying* both for concepts and individuals
- *However*, for complex descriptions (such as MSC^*), deeply nested subconcepts could increase the dissimilarity value
- **New idea:** differentiate the weight of the subconcepts wrt their levels in the descriptions for determining the final dissimilarity value
  - **Solve the problem:** *how differences in concept structure might impact concept (dis-)similarity? i.e. considering the series* $dist(B, B \sqcap A), dist(B, B \sqcap \forall R.A), dist(B, B \sqcap \forall R.\forall R.A)$ *this should become smaller since more deeply nested restrictions ought to represent smaller differences.*" **[Borgida et al. 2005]**

The weighted Dissimilarity Measure

Overlap Function Definition [d'Amato et al. @ SWAP 2005]:

$\mathcal{L} = \mathcal{ALC}/_{\equiv}$ 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 $\mathcal{L}_{\equiv}$

$$f(C,D) := f_{\sqcup}(C,D) = \begin{cases} |\Delta| & C \equiv D \\ 0 & C \sqcap D \equiv \bot \\ 1 + \lambda \cdot \max\limits_{\substack{i = 1,\ldots,n \\ j = 1,\ldots,m}} f_{\sqcap}(C_i, D_j) & \text{o.w.} \end{cases}$$

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

Looking toward Information Content: Motivation

- *The use of Information Content* is presented as *the most effective way for measuring complex concept descriptions* **[Borgida et al. 2005]**
- The necessity of considering concepts in normal form for computing their (dis-)similarity is argued **[Borgida et al. 2005]**
  - *confirmation* of the used approach in the previous measure
- **A dissimilarity measure for complex descriptions grounded on IC has been defined**
  - $\mathcal{ALC}$ concepts in *normal form*
  - based on the *structure and semantics* of the concepts.
  - *elicits the underlying semantics*, by querying the KB for assessing the *IC* of concept descriptions w.r.t. the KB
  - *extension for considering individuals*

Information Content: Defintion

- 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 wrt the considered ABox.
  - $pr(C) = |C^{\mathcal{I}}|/|\Delta^{\mathcal{I}}|$
- A function for measuring the *IC variation* between concepts is defined

Function Definition /I

[d'Amato et al. @ SAC 2006] $\mathcal{L} = \mathcal{ALC}/_{\equiv}$ 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 $\mathcal{L}_{\equiv}$

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

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

Function Definition / II

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

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

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

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

Dissimilarity Measure: Definition

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:

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

where f is the function defined previously

Discussion

- $d(C, D) = 0$ iff IC=0 iff C $\equiv$ D (semantic equivalence) rather d assigns the minimun value
- $d(C, D) = 1$ iff IC $\rightarrow \infty$ iff $C \sqcap D \equiv \bot$, rather d assigns the maximum value because concepts involved are totally different
- Otherwise $d(C, D) \in]0, 1[$ rather d tends to 0 if IC tends to 0; d tends to 1 if IC tends to infinity

$\mathcal{ALN}$ Normal Form

C is in $\mathcal{ALN}$ *normal form* iff $C \equiv \bot$ or $C \equiv \top$ or if

$$C = \bigsqcap_{P \in \mathrm{prim}(C)} P \sqcap \bigsqcap_{R \in N_R} (\forall R.C_R \sqcap \geq n.R \sqcap \leq m.R)$$

where:

$C_R = \mathrm{val}_R(C)$, $n = \mathrm{min}_R(C)$ and $m = \mathrm{max}_R(C)$

- $\mathrm{prim}(C)$ 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. $\mathrm{val}_R(C) = \top$);
- $\mathrm{min}_R(C)$ $= \max\{n \in N \mid C \sqsubseteq (\geq n.R)\}$ (always finite number);
- $\mathrm{max}_R(C)$ $= \min\{n \in N \mid C \sqsubseteq (\leq n.R)\}$ (if unlimited $\mathrm{max}_R(C) = \infty$)

Measure Definition / I

[Fanizzi et. al @ CILC 2006] $\mathcal{L} = \mathcal{ALN}/_{\equiv}$ the set of all concepts in $\mathcal{ALN}$ normal form $\mathcal{I}$ canonical interpretation of $\mathcal{A}$ A-Box $s : \mathcal{L} \times \mathcal{L} \mapsto [0,1]$ defined $\forall C, D \in \mathcal{L}$:

$$\begin{aligned} s(C,D) \;&:=\; \lambda[s_P(\mathrm{prim}(C), \mathrm{prim}(D)) \;+ \\ &+\; \frac{1}{|N_R|}\sum_{R \in N_R} s(\mathrm{val}_R(C), \mathrm{val}_R(D)) \;+ \frac{1}{|N_R|}\cdot \\ &\cdot\; \sum_{R \in N_R} s_N((\mathrm{min}_R(C), \mathrm{max}_R(C)), (\mathrm{min}_R(D), \mathrm{max}_R(D)))] \end{aligned}$$

where $\lambda \in]0,1]$ (let $\lambda = 1/3$),

Measure Defintion / II

$$s_P(\mathrm{prim}(C), \mathrm{prim}(D)) := \frac{|\bigcap_{P_C \in \mathrm{prim}(C)} P_C^{\mathcal{I}} \cap \bigcap_{Q_D \in \mathrm{prim}(D)} Q_D^{\mathcal{I}}|}{|\bigcap_{P_C \in \mathrm{prim}(C)} P_C^{\mathcal{I}} \cup \bigcap_{Q_D \in \mathrm{prim}(D)} Q_D^{\mathcal{I}}|}$$

$$s_N((m_C, M_C), (m_D, M_D)) := \frac{\min(M_C, M_D) - \max(m_C, m_D) + 1}{\max(M_C, M_D) - \min(m_C, m_D) + 1}$$

$$s_N((m_C, M_C), (m_D, M_D)) := 0 \text{ if } \min(M_C, M_D) > \max(m_C, m_D)$$

Similarity Measure: example...

Let $\mathcal{A}$ be the considered ABox

Person(Meg), ¬Male(Meg), hasChild(Meg,Bob), hasChild(Meg,Pat),
Person(Bob), Male(Bob), hasChild(Bob,Ann),
Person(Pat), Male(Pat), hasChild(Pat,Gwen),
Person(Gwen), ¬Male(Gwen),
Person(Ann), ¬Male(Ann), hasChild(Ann,Sue), marriedTo(Ann,Tom),
Person(Sue), ¬Male(Sue),
Person(Tom), Male(Tom)

and let C and D be two descriptions in $\mathcal{ALN}$ normal form:

$$C \equiv \text{Person} \sqcap \forall \text{marriedTo.Person} \sqcap \leq 1.\text{hasChild}$$

$$D \equiv \text{Male} \sqcap \forall \text{marriedTo.(Person} \sqcap \neg\text{Male)} \sqcap \leq 2.\text{hasChild}$$

...Similarity Measure: example...

In order to compute $s(C, D)$ let us consider:

- Let be $\lambda := \frac{1}{3}$
- $N_R = \{\text{hasChild}, \text{marriedTo}\} \rightarrow |N_R| = 2$

$$s(C, D) := \frac{1}{3}\Bigg[s_P(\text{prim}(C), \text{prim}(D)) + \frac{1}{2}\sum_{R \in N_R} s(\text{val}_R(C), \text{val}_R(D)) + \\ + \frac{1}{2}\sum_{R \in N_R} s_N((\text{min}_R(C), \text{max}_R(C)), (\text{min}_R(D), \text{max}_R(D)))\Bigg]$$

...Similarity Measure: example...

In order to compute s_P let us note that:

- prim(C) = Person
- prim(D) = Male

$$s_P(\{\text{Person}\}, \{\text{Male}\}) =$$

$$= \frac{|\{\text{Meg,Bob,Pat,Gwen,Ann,Sue,Tom}\} \cap \{\text{Bob,Pat,Tom}\}|}{|\{\text{Meg,Bob,Pat,Gwen,Ann,Sue,Tom}\} \cup \{\text{Bob,Pat,Tom}\}|} =$$

$$= \frac{|\{\text{Bob,Pat,Tom}\}|}{|\{\text{Meg,Bob,Pat,Gwen,Ann,Sue,Tom}\}|} = 3/7$$

...Similarity Measure: example...

To compute s for value restrictions, it is important to note that

- $N_R = \{\text{hasChild}, \text{marriedTo}\}$
- $val_{marriedTo}(C) = \text{Person}$ and $val_{hasChild}(C) = \top$
- $val_{marriedTo}(D) = \text{Person} \sqcap \neg\text{Male}$ and $val_{hasChild}(D) = \top$

$$s(\text{Person}, \text{Person} \sqcap \neg\text{Male}) + s(\top, \top) =$$

$$= \frac{1}{3} \cdot (s_P(\text{Person}, \text{Person} \sqcap \neg\text{Male}) + \frac{1}{2} \cdot (1+1) + \frac{1}{2} \cdot (1+1)) +$$

$$+ \frac{1}{3} \cdot (1+1+1) = \frac{1}{3} \cdot (\frac{4}{7} + 1 + 1) + 1 = \frac{13}{7}$$

...Similarity Measure: example

To compute s for number restrictions it is important to note that

- $N_R = \{\text{hasChild}, \text{marriedTo}\}$ $\quad \min(M_C, M_D) > \max(m_C, m_D)$
- $min_{marriedTo}(C) = 0; \quad max_{marriedTo}(C) = |\Delta| + 1 = 7 + 1 = 8$
  $min_{hasChild}(C) = 0; \quad max_{hasChild}(C) = 1$
- $min_{marriedTo}(D) = 0; \quad max_{marriedTo}(D) = |\Delta| + 1 = 7 + 1 = 8$
  $min_{hasChild}(D) = 0; \quad max_{hasChild}(D) = 2$

$$s_N(\ (m_{\text{hasChild}}(C), M_{\text{hasChild}}(C)), (m_{\text{hasChild}}(D), M_{\text{hasChild}}(D))) +$$

$$+\ s_N((m_{\text{marriedTo}}(C), M_{\text{marriedTo}}(C)), (m_{\text{marriedTo}}(D), M_{\text{marriedTo}}(D))) =$$

$$= \frac{min(M_{\text{hasChild}}(C), M_{\text{hasChild}}(D)) - max(m_{\text{hasChild}}(C), m_{\text{hasChild}}(D)) + 1}{max(M_{\text{hasChild}}(C), M_{\text{hasChild}}(D) - min(m_{\text{hasChild}}(C), m_{\text{hasChild}}(D)) + 1)} + 1 =$$

$$= \frac{min(1,2) - max(0,0) + 1}{max(1,2) - min(0,0) + 1)} + 1 = \frac{2}{3} + 1 = \frac{5}{3}$$

Relational Kernel Function: Motivation

- Kernel functions jointly with a kernel method.
- *Advangate:* 1) efficency; 2) the learning algorithm and the kernel are almost completely independent.
  - An efficient *algorithm for attribute-value* instance spaces *can be converted into one* suitable *for structured spaces* by merely *replacing the kernel function*.
- A **kernel function for $\mathcal{ALC}$ normal form concept descriptions** has been defined.
  - Based both on the *syntactic structure* (exploiting the *convolution* kernel **[Haussler 1999]** and on the *semantics*, derived from the ABox.

Kernel Defintion/I

[Fanizzi et al. @ ISMIS 2006] Given the space X of $\mathcal{ALC}$ normal form concept descriptions, $D_1 = \bigsqcup_{i=1}^{n} C_i^1$ and $D_2 = \bigsqcup_{j=1}^{m} C_j^2$ in X, and an interpretation $\mathcal{I}$, the $\mathcal{ALC}$ *kernel* based on $\mathcal{I}$ is the function $k_{\mathcal{I}} : X \times X \mapsto \mathbb{R}$ inductively defined as follows.

disjunctive descriptions:

$$k_{\mathcal{I}}(D_1, D_2) = \lambda \sum_{i=1}^{n} \sum_{j=1}^{m} k_{\mathcal{I}}(C_i^1, C_j^2) \quad \text{with } \lambda \in]0,1]$$

conjunctive descriptions:

$$k_{\mathcal{I}}(C^1, C^2) = \prod_{\substack{P_1 \in \mathrm{prim}(C^1) \\ P_2 \in \mathrm{prim}(C^2)}} k_{\mathcal{I}}(P_1, P_2) \cdot \prod_{R \in N_R} k_{\mathcal{I}}(\mathrm{val}_R(C^1), \mathrm{val}_R(C^2)) \cdot$$

$$\cdot \prod_{R \in N_R} \sum_{\substack{C_i^1 \in \mathrm{ex}_R(C^1) \\ C_j^2 \in \mathrm{ex}_R(C^2)}} k_{\mathcal{I}}(C_i^1, C_j^2)$$

Kernel Definition/II

primitive concepts:

$$k_{\mathcal{I}}(P_1, P_2) = \frac{k_{\text{set}}(P_1^{\mathcal{I}}, P_2^{\mathcal{I}})}{|\Delta^{\mathcal{I}}|} = \frac{|P_1^{\mathcal{I}} \cap P_2^{\mathcal{I}}|}{|\Delta^{\mathcal{I}}|}$$

where k_{set} is the kernel for set structures **[Gaertner 2004]**. This case includes also the negation of primitive concepts using set difference: $(\neg P)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus P^{\mathcal{I}}$

Computing the kernel fucntion: Example...

Considered concept descriptions:

$C \equiv (P_1 \sqcap P_2) \sqcup (\exists R.P_3 \sqcap \forall R.(P_1 \sqcap \neg P_2))$
$D \equiv P_3 \sqcup (\exists R.\forall R.P_2 \sqcap \exists R.\neg P_1)$

Supposing:
$P_1^{\mathcal{I}} = \{a, b, c\}$, $P_2^{\mathcal{I}} = \{b, c\}$, $P_3^{\mathcal{I}} = \{a, b, d\}$, $\Delta^{\mathcal{I}} = \{a, b, c, d, e\}$
Disjunctive level:

$$
\begin{aligned}
k_{\mathcal{I}}(C, D) &= \lambda \sum_{i=1}^{2} \sum_{j=1}^{2} k_{\mathcal{I}}(C_i, D_j) = \\
&= \lambda \cdot (k_{\mathcal{I}}(C_1, D_1) + k_{\mathcal{I}}(C_1, D_2) + k_{\mathcal{I}}(C_2, D_1) + k_{\mathcal{I}}(C_2, D_2))
\end{aligned}
$$

where $C_1 \equiv P_1 \sqcap P_2$, $C_2 \equiv \exists R.P_3 \sqcap \forall R.(P_1 \sqcap \neg P_2)$,
$D_1 \equiv P_3$, $D_2 \equiv \exists R.\forall R.P_2 \sqcap \exists R.\neg P_1$.

...Computing the kernel fucntion: Example...

The kernel for the conjunctive level has to be compute for every couple C_i, D_j

$$
\begin{aligned}
k_{\mathcal{I}}(C_1, D_1) &= \prod_{P_1^C \in \text{prim}(C_1)} \prod_{P_1^D \in \text{prim}(D_1)} k_{\mathcal{I}}(P_1^C, P_1^D) \cdot k_{\mathcal{I}}(\top, \top) \cdot k_{\mathcal{I}}(\top, \top) \\
&= k_{\mathcal{I}}(P_1, P_3) \cdot k_{\mathcal{I}}(P_2, P_3) \cdot 1 \cdot 1 = \\
&= \frac{|\{a, b, c\} \cap \{a, b, d\}|}{a, b, c, d, e} \cdot \frac{|\{b, c\} \cap \{a, b, d\}|}{a, b, c, d, e} = \frac{2}{5} \cdot \frac{1}{5} = \frac{2}{25}
\end{aligned}
$$

No contribution comes from value and existential restrictions: the factors amount to 1 since $\text{val}_R(C_1) = \text{val}_R(D_1)) = \top$ and $\text{ex}_R(C_1) = \text{ex}_R(D_1) = \emptyset$ which make those equivalent to $\top$ too.

...Computing the kernel fucntion: Example...

The conjunctive kernel for C_1 and D_2 has to be computed. Note that there are no universal restrictions and $N_R = \{R\} \Rightarrow |N_R| = 1$ this means that all products on varying $R \in N_R$ can be simplified. Empty prim is equivalent to $\top$.

$$
\begin{aligned}
k_{\mathcal{I}}(C_1, D_2) &= [k_{\mathcal{I}}(P_1, \top) \cdot k_{\mathcal{I}}(P_2, \top)] \cdot k_{\mathcal{I}}(\top, \top) \cdot \sum_{\substack{E_C \in \mathrm{ex}_R(C_1) \\ E_D \in \mathrm{ex}_R(D_2)}} k_{\mathcal{I}}(E_C, E_D) \\
&= (3 \cdot 2) \cdot 1 \cdot [k_{\mathcal{I}}(\top, \forall R.P_2) + k_{\mathcal{I}}(\top, \neg P_1)] = \\
&= 6 \cdot [\lambda \sum_{\substack{C' \in \{\top\} \\ D' \in \{\forall R.P_2\}}} k_{\mathcal{I}}(C', D') + 2] = \\
&= 6 \cdot [\lambda \cdot (1 \cdot k_{\mathcal{I}}(\top, P_2) \cdot 1) + 2] = \\
&= 6 \cdot [\lambda \cdot (\lambda \cdot 1 \cdot 2/5 \cdot 1) + 2] = 12(\lambda^2/5 + 1)
\end{aligned}
$$

...Computing the kernel fucntion: Example...

$$
\begin{aligned}
k_{\mathcal{I}}(C_2, D_1) &= k_{\mathcal{I}}(\top, P_3) \cdot k_{\mathcal{I}}(\mathrm{val}_R(C_2), \top) \cdot \sum_{\substack{E_C \in \mathrm{ex}_R(C_2) \\ E_D \in \mathrm{ex}_R(D_1)}} k_{\mathcal{I}}(E_C, E_D) = \\
&= 3/5 \cdot k_{\mathcal{I}}(P_1 \sqcap \neg P_2, \top) \cdot k_{\mathcal{I}}(P_3, \top) = \\
&= 3/5 \cdot [\lambda(k_{\mathcal{I}}(P_1, \top) \cdot k_{\mathcal{I}}(\neg P_2, \top))] \cdot 3/5 = \\
&= 3/5 \cdot [\lambda(3/5 \cdot 3/5)] \cdot 3/5 = 81\lambda/625
\end{aligned}
$$

Note that the absence of the prim set is equivalent to $\top$ and, since one of the sub-concepts has no existential restriction the product gives no contribution.

...Computing the kernel fucntion: Example

Finally, the kernel function on the last couple of disjuncts

$$\begin{aligned}
k_{\mathcal{I}}(C_2, D_2) &= k_{\mathcal{I}}(\top, \top) \cdot k_{\mathcal{I}}(P_1 \sqcap \neg P_2, \top) \cdot \sum_{\substack{C'' \in \{P_3\} \\ D'' \in \{\forall R.P_2, \neg P_1\}}} k_{\mathcal{I}}(C'', D'') = \\
&= 1 \cdot 9\lambda/25 \cdot [(k_{\mathcal{I}}(P_3, \forall R.P_2) + k_{\mathcal{I}}(P_3, \neg P_1)] = \\
&= 9\lambda/25 \cdot [\lambda \cdot k_{\mathcal{I}}(P_3, \top) \cdot k_{\mathcal{I}}(\top, P_2) \cdot k_{\mathcal{I}}(\top, \top) + 1/5] = \\
&= 9\lambda/25 \cdot [\lambda \cdot 3/5 \cdot 2\lambda/5 \cdot 1 + 1/5] = \\
&= 9\lambda/25 \cdot [6\lambda^2/25 + 1/5]
\end{aligned}$$

By collecting the four intermediate results, the value for the computed kernel function on C and D can be computed:

$$k_{\mathcal{I}}(C, D) = 2/25 + 12(\lambda^2/5 + 1) + 81\lambda/625 + 9\lambda/25 \cdot [6\lambda^2/25 + 1/5]$$

Kernel function: Discussion

- The kernel function can be extended to the case of individuals/concept
- The kernel is *valid*
  - The function is symmetric
  - The function is closed under multiplication and sum of valid kernel (kernel set).
- Being the kernel valid, and *induced distance measure* (metric) can be obtained **[Haussler 1999]**

$$d_{\mathcal{I}}(C, D) = \sqrt{k_{\mathcal{I}}(C, C) - 2k_{\mathcal{I}}(C, D) + k_{\mathcal{I}}(D, D)}$$

Semi-Distance Measure: Motivations

- Most of the presented measures are grounded on concept structures $\Rightarrow$ hardly scalable w.r.t. most expressive DLs
- **IDEA**: *on a semantic level, similar individuals should behave similarly w.r.t. the same concepts*
- Following HDD **[Sebag 1997]**: individuals can be compared on the grounds of their behavior w.r.t. a given set of hypotheses $\mathsf{F} = \{F_1, F_2, \ldots, F_m\}$, that is a collection of (primitive or defined) concept descriptions
  - F stands as a group of *discriminating features* expressed in the considered language
- As such, the new measure *totally depends on semantic* aspects of the individuals in the KB

Semantic Semi-Dinstance Measure: Definition

[Fanizzi et al. @ DL 2007] Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a KB and let $\mathsf{Ind}(\mathcal{A})$ be the set of the individuals in $\mathcal{A}$. Given sets of concept descriptions $\mathsf{F} = \{F_1, F_2, \ldots, F_m\}$ in $\mathcal{T}$, a *family of semi-distance functions* $d_p^{\mathsf{F}} : \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \mapsto \mathbb{R}$ is defined as follows:

$$\forall a, b \in \mathsf{Ind}(\mathcal{A}) \quad d_p^{\mathsf{F}}(a, b) := \frac{1}{m} \left[\sum_{i=1}^{m} \mid \pi_i(a) - \pi_i(b) \mid^p \right]^{1/p}$$

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

$$\forall a \in \mathsf{Ind}(\mathcal{A}) \quad \pi_i(a) = \begin{cases} 1 & F_i(a) \in \mathcal{A} \quad (\mathcal{K} \models F_i(a)) \\ 0 & \neg F_i(a) \in \mathcal{A} \quad (\mathcal{K} \models \neg F_i(a)) \\ \frac{1}{2} & \textit{otherwise} \end{cases}$$

Semi-Distance Measure: Discussion

- *More similar* the considered *individuals are*, more similar the project function values are $\Rightarrow d_p^{\mathsf{F}} \simeq 0$
- *More different* the considered *individuals are*, more different the projection values are $\Rightarrow$ the value of d_p^{F} will increase
- The measure complexity mainly depends from the complexity of the *Instance Checking* operator for the chosen DL
  - $\mathit{Compl}(d_p^{\mathsf{F}}) = |\mathsf{F}| \cdot 2 \cdot \mathit{Compl}(\mathsf{IChk})$
- **Optimal discriminating feature set could be learned**

Goals for using Inductive Learning Methods in the SW

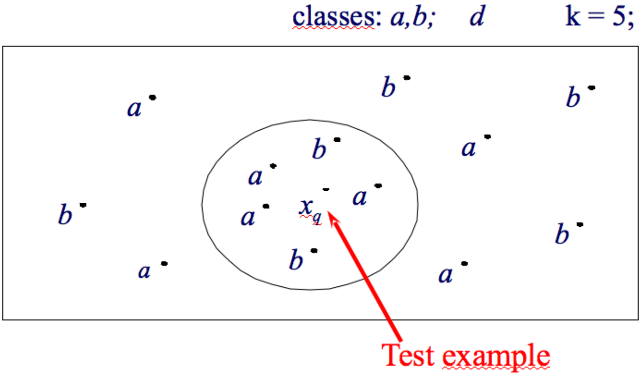
Instance-base classifier for

- Semi-automatize the A-Box population task
- Induce new knowledge not logically derivable
- Improve concept retrieval and query answearing inference service
- *Realized algorithms*
    - Relational K-NN
    - Relational kernel embedded in a SVM

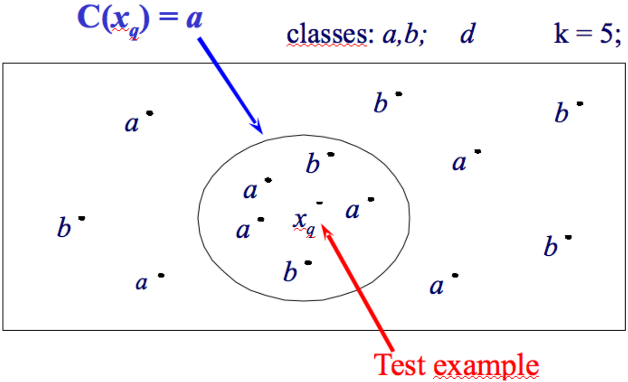
Unsupervised learning methods for

- Improve service discovery task
- Exploiting (dis-)similarity measures for improving the ranking of the retrieved services

Classical K-NN algorithm...



...Classical K-NN algorithm...



...Classical K-NN algorithm

- Generally applied to feature vector representation
- In classification phase it is assumed that each training and test example belong to a single class, so classes are considered to be disjoint
- 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 a most complex and expressive knowledge representation
2. It is not possible to assume disjointness of classes. 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 Semantic Web area

Choices for applying K-NN to Ontological Knowledge

1. To have similarity and dissimilarity measures applicable to ontological knowledge allows applying K-NN to this kind of knowledge representation
2. A new classification procedure is adopted, decomposing the multi-class classification problem into smaller binary classification problems (one per target concept).
   - For each individual to classify w.r.t each class (concept), classification returns {-1,+1}
3. A third value 0 representing unknown information is added in the classification results {-1,0,+1}
4. Hence a majority voting criterion is applied

Realized K-NN Algorithm...

[d'Amato et al. @ URSW Workshop at ISWC 2006]

- **Main Idea:** similar individuals, by analogy, should likely belong to similar concepts
    - for every ontology, all individuals are classified to be instances of one or more concepts of the considered ontology
- For each individual in the ontology MSC is computed
- MSC list represents the set of training examples

...Realized K-NN Algorithm

- Each example is classified applying the k-NN method for DLs, adopting the leave-one-out cross validation procedure.

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

where

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

Experimentation Setting

ontology	*DL*
FSM	$\mathcal{SOF}(\mathcal{D})$
S.-W.-M.	$\mathcal{ALCOF}(\mathcal{D})$
FAMILY	$\mathcal{ALCN}$
FINANCIAL	$\mathcal{ALCIF}$

ontology	*#concepts*	*#obj. prop*	*#data prop*	*#individuals*
FSM	20	10	7	37
S.-W.-M.	19	9	1	115
FAMILY	14	5	0	39
FINANCIAL	60	17	0	652

Measures for Evaluating Experiments

- **Performance evaluated** by *comparing the procedure responses to those returned by a standard reasoner* (Pellet)
- **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

Results (average±std-dev.) using the measure based on overlap.

	Match Rate	Commission Rate	Omission Rate	Induction Rate
FAMILY	.654±.174	.000±.000	.231±.173	.115±.107
FSM	.974±.044	.026±.044	.000±.000	.000±.000
S.-W.-M.	.820±.241	.000±.000	.064±.111	.116±.246
FINANCIAL	.807±.091	.024±.076	.000±.001	.169±.076

Results (average ± std-dev.) using the measure based in IC

	Match	Commission	Omission	Induction
FAMILY	.608±.230	.000±.000	.330±.216	.062±.217
FSM	.899±.178	.096±.179	.000±.000	.005±.024
S.-W.-M.	.820±.241	.000±.000	.064±.111	.116±.246
FINANCIAL	.807±.091	.024±.076	.000±.001	.169±.046

Experimentation: Discussion...

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

...Experimentation: Discussion...

SURFACE-WATER-MODEL and FINANCIAL 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*. This is mainly due to
  - some *concepts* that are declared to be *mutually disjoint*
  - some *individuals* are *involved in relations*

...Experimentation: Discussion

FAMILY Ontology

- Predictive Accuracy is not so high and Omission Error not null
  - Because 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 derivable*

Comparing the Measures

- The **measure based on IC** *poorly classifies concepts* that have *less information* in the ontology
  - *The measure based on IC is less able*, w.r.t. the measure based on overlap, *to classify concepts* correctly, when they have *few information* (instance and object properties involved);
- **Comparable behavior** when *enough information* is available
- **Inducted knowledge can be used for**
  - *semi-automatize ABox population*
  - *improving concept retrieval*

Experiments: Querying the KB exploiting relational K-NN

Setting

- 15 queries randomly generated by conjunctions/disjunctions of primitive or defined concepts of each ontology.
- **Classification of all individuals in each ontology w.r.t the query concept**
- Performance evaluated by comparing the procedure responses to those returned by a standard reasoner (Pellet) employed as a baseline.
- The *Semi-distance measure* has been used
  - *All concepts in ontology have been employed as feature set* F

Ontologies employed in the experiments

ontology	*DL*
FSM	$\mathcal{SOF}(D)$
S.-W.-M.	$\mathcal{ALCOF}(D)$
SCIENCE	$\mathcal{ALCIF}(D)$
NTN	$\mathcal{SHIF}(D)$
FINANCIAL	$\mathcal{ALCIF}$

ontology	*#concepts*	*#obj. prop*	*#data prop*	*#individuals*
FSM	20	10	7	37
S.-W.-M.	19	9	1	115
SCIENCE	74	70	40	331
NTN	47	27	8	676
FINANCIAL	60	17	0	652

Experimentation: Resuls

Results (average$\pm$std-dev.) using the semi-distance semantic measure

	match rate	*commission rate*	*omission rate*	*induction rate*
FSM	97.7 ± 3.00	2.30 ± 3.00	0.00 ± 0.00	0.00 ± 0.00
S.-W.-M.	99.9 ± 0.20	0.00 ± 0.00	0.10 ± 0.20	0.00 ± 0.00
Science	99.8 ± 0.50	0.00 ± 0.00	0.20 ± 0.10	0.00 ± 0.00
Financial	90.4 ± 24.6	9.40 ± 24.5	0.10 ± 0.10	0.10 ± 0.20
NTN	99.9 ± 0.10	0.00 ± 7.60	0.10 ± 0.00	0.00 ± 0.10

Experimentation: Discussion

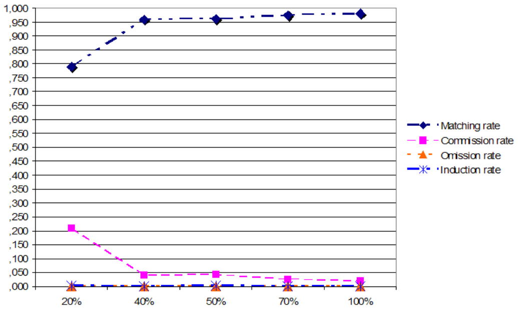
- Very low commission error: almost never the classifier makes critical mistakes
- Very high match rate 95%(more than the previous measures 80%) ⇒ Highly comparable with the reasoner
- Very low induction rate ⇒ Less able (w.r.t. previous measures) to induce new knowledge
- *Lower match rate for* FINANCIAL *ontology* as **data are not enough sparse**
- The **usage of all concepts for the set F made the measure accurate**, which is the reason why the procedure resulted conservative as regards inducing new assertions.

Testing the Effect of the Variation of F on the Measure

- *Espected result*: with an increasing number of considered hypotheses for F, the accuracy of the measure would increase accordingly.
- **Considered ontology:** *Financial* as is is the most populated
- Experiment repeated with an increasing percentage of concepts randomly selected for F from the ontology.
- Results confirm the hypothesis
- **Similar results for the other ontologies**

Experimentation: Results

% of concepts	*match*	*commission*	*omission*	*Induction*
20%	79.1	20.7	0.00	0.20
40%	96.1	03.9	0.00	0.00
50%	97.2	02.8	0.00	0.00
70%	97.4	02.6	0.00	0.00
100%	98.0	02.0	0.00	0.00



SVM and Relational Kernel Function for the SW

- A SMV is a classifier that, by means of kernel function, implicitly maps the training data into a higher dimensional feature space where they can be classified using a linear classifier
  - A SVM from the LIBSVM library has been considered
- *Learning Problem*: Given an ontology, classify all its individuals w.r.t. all concepts in the ontology **[Fanizzi et al. @ KES 2007]**
- *Problems to solve:* 1) Implicit CWA; 2) Assumption of class disjointness
- *Solutions:* Decomposing the classification problem is a set of ternary classification problems $\{+1, 0, -1\}$, for each concept of the ontology

Ontologies employed in the experiments

ontology	*DL*
PEOPLE	$\mathcal{ALCHIN}(D)$
UNIVERSITY	$\mathcal{ALC}$
FAMILY	$\mathcal{ALCF}$
FSM	$\mathcal{SOF}(D)$
S.-W.-M.	$\mathcal{ALCOF}(D)$
SCIENCE	$\mathcal{ALCIF}(D)$
NTN	$\mathcal{SHIF}(D)$
NEWSPAPER	$\mathcal{ALCF}(D)$
WINES	$\mathcal{ALCIO}(D)$

ontology	*#concepts*	*#obj. prop*	*#data prop*	*#individuals*
PEOPLE	60	14	1	21
UNIVERSITY	13	4	0	19
FAMILY	14	5	0	39
FSM	20	10	7	37
S.-W.-M.	19	9	1	115
SCIENCE	74	70	40	331
NTN	47	27	8	676
NEWSPAPER	29	28	25	72
WINES	112	9	10	188

Experiment: Results

Ontoly		*match rate*	*ind. rate*	*omis.err.rate*	*comm.err.rate*
People	*avg.*	**0.866**	**0.054**	**0.08**	**0.00**
	range	0.66 - 0.99	0.00 - 0.32	0.00 - 0.22	0.00 - 0.03
University	*avg.*	**0.789**	**0.114**	**0.018**	**0.079**
	range	0.63 - 1.00	0.00 - 0.21	0.00 - 0.21	0.00 - 0.26
FSM	*avg.*	**0.917**	**0.007**	**0.00**	**0.076**
	range	0.70 - 1.00	0.00 - 0.10	0.00 - 0.00	0.00 - 0.30
Family	*avg.*	**0.619**	**0.032**	**0.349**	**0.00**
	range	0.39 - 0.89	0.00 - 0.41	0.00 - 0.62	0.00 - 0.00
NewsPaper	*avg.*	**0.903**	**0.00**	**0.097**	**0.00**
	range	0.74 - 0.99	0.00 - 0.00	0.02 - 0.26	0.00 - 0.00
Wines	*avg.*	**0.956**	**0.004**	**0.04**	**0.00**
	range	0.65 - 1.00	0.00 - 0.27	0.01 - 0.34	0.00 - 0.00
Science	*avg.*	**0.942**	**0.007**	**0.051**	**0.00**
	range	0.80 - 1.00	0.00 - 0.04	0.00 - 0.20	0.00 - 0.00
S.-W.-M.	*avg.*	**0.871**	**0.067**	**0.062**	**0.00**
	range	0.57 - 0.98	0.00 - 0.42	0.00 - 0.40	0.00 - 0.00
N.T.N.	*avg.*	**0.925**	**0.026**	**0.048**	**0.001**
	range	0.66 - 0.99	0.00 - 0.32	0.00 - 0.22	0.00 - 0.03

Experiments: Discussion

- High matching rate
- Induction Rate not null $\Rightarrow$ new knowledge is induced
- For every ontology, the commission error is quite low $\Rightarrow$ the classifier does not make critical mistakes
  - Not null for UNIVERSITY and FSM ontologies $\Rightarrow$ They have the lowest number of individuals
  - There is not enough information for separating the feature space producing a correct classification
- **In general** *the match rate increases with the increase of the number of individuals in the ontology*
  - Consequently the commission error rate decreases
- **Similar results by using the classifier for querying the KB**

Why the Attention to Modeling Service Descriptions

- WS Technology has allowed uniform access via Web standards to software components residing on various platforms and written in different programming languages
- *WS major limitation:* their retrieval and composition still require manual effort
- *Solution*: augment WS with a semantic description of their functionality $\Rightarrow$ *SWS*
- **Choice:** DLs as representation language, *because*:
  - DLs are endowed by a formal semantics $\Rightarrow$ guarantee expressive service descriptions and precise semantics definition
  - DLs are the theoretical foundation of OWL $\Rightarrow$ ensure compatibility with existing ontology standards
  - *Service discovery* can be performed exploiting standard and non-standard DL inferences

DLs-based Service Descriptions

- **[Grimm et al. 2004]** *A service description is expressed by a set of DL-axioms* $D = \{S, \phi_1, \phi_2, ..., \phi_n\}$*, where the axioms* ϕ_i *impose restrictions on an atomic concept* S*, which represents the service to be performed*

$$
\begin{aligned}
D_r = \{ \quad & S_r \equiv \mathsf{Company} \sqcap \exists \mathsf{payment.EPayment} \sqcap \exists \mathsf{to}.\{\mathsf{bari}\} \sqcap \\
& \qquad \sqcap \exists \mathsf{from}.\{\mathsf{cologne,hahn}\} \sqcap \leq 1\ \mathsf{hasAlliance} \sqcap \\
& \qquad \sqcap \forall \mathsf{hasFidelityCard}.\{\mathsf{milesAndMore}\}; \\
& \{\mathsf{cologne,hahn}\} \sqsubseteq \exists\ \mathsf{from}^{-}.S_r \quad \}
\end{aligned}
$$

$KB =$
$\{\mathsf{cologne{:}Germany},\ \mathsf{hahn{:}Germany},\ \mathsf{bari{:}Italy},\ \mathsf{milesAndMore{:}Card}\}$

Introducing Constraint Hardness

- **[d'Amato et al. @ Sem4WS Workshop at BPM 2006]** In real scenarios a service request is characterized by some needs that *must* be satisfied and others that *may* be satisfied
- *HC* represent necessary and sufficient conditions for selecting requested service instances
- *SC* represent only necessary conditions.

Definition

Let $D_r^{HC} = \{S_r^{HC}, \sigma_1^{HC}, ..., \sigma_n^{HC}\}$ be the set of *HC* for a requested service description D_r and let $D_r^{SC} = \{S_r^{SC}, \sigma_1^{SC}, ..., \sigma_m^{SC}\}$ be the set of *SC* for D_r. The complete description of D_r is given by $D_r = \{S_r \equiv S_r^{HC} \sqcup S_r^{SC}, \sigma_1^{HC}, ..., \sigma_n^{HC}, \sigma_1^{SC}, ..., \sigma_m^{SC}\}$.

Modelling Service Descriptions: Example

$$
\begin{aligned}
D_r = \{ \quad & S_r \equiv \mathsf{Flight} \sqcap \exists \mathsf{from}.\{\mathsf{Cologne},\mathsf{Hahn},\mathsf{Frankfurt}\} \sqcap \exists \mathsf{to}.\{\mathsf{Bari}\} \sqcap \\
& \qquad \sqcap \forall \mathsf{hasFidelityCard}.\{\mathsf{MilesAndMore}\}; \\
& \{\mathsf{Cologne},\ \mathsf{Hahn},\ \mathsf{Frankfurt}\} \sqsubseteq \exists\ \mathsf{from}^{-}.S_r; \\
& \{\mathsf{Bari}\} \sqsubseteq \exists\ \mathsf{to}^{-}.S_r \qquad \}
\end{aligned}
$$

where

$$
\begin{aligned}
HC_r = \{ \quad & \mathsf{Flight} \sqcap \exists \mathsf{to}.\{\mathsf{Bari}\} \sqcap \exists \mathsf{from}.\{\mathsf{Cologne},\ \mathsf{Hahn},\ \mathsf{Frankfurt}\}; \\
& \{\mathsf{Cologne},\ \mathsf{Hahn},\ \mathsf{Frankfurt}\} \sqsubseteq \exists\ \mathsf{from}^{-}.S_r; \\
& \{\mathsf{Bari}\} \sqsubseteq \exists\ \mathsf{to}^{-}.S_r \qquad \} \\
SC_r = \{ \quad & \mathsf{Flight} \sqcap \forall \mathsf{hasFidelityCard}.\{\mathsf{MilesAndMore}\} \qquad \};
\end{aligned}
$$

$$
\begin{aligned}
KB = \{ \quad & \mathsf{Cologne},\mathsf{Hahn},\mathsf{Frankfurt}{:}\mathsf{Germany},\ \mathsf{Bari}{:}\mathsf{Italy}, \\
& \mathsf{MilesAndMore}{:}\mathsf{Card}\}
\end{aligned}
$$

Discovery and Matching Services

- *Service Discovery* is the task of locating service providers that can satisfy the requesters needs
- Discovery is performed by *matching a requested service description to the service descriptions of potential providers*
- The matching process (w.r.t. a KB) is expressed as a boolean function *match*(KB, D_r, D_p) which specifies how to apply DL inferences to perform the matching

The Matching Process

Let $D_r = \{S_r, \sigma_1, \ldots, \sigma_n\}$ be a requested service description and $D_p = \{S_p, \sigma_1, \ldots, \sigma_m\}$ a provided service description

- **Satisfiability of Concept Conjunction [Trastour 2001]**

$$KB \cup D_r \cup D_p \cup \{\exists x : S_r(x) \wedge S_p(x)\} \textit{ is consistent} \Leftrightarrow$$
$$\Leftrightarrow KB \cup D_r \cup D_p \cup \{i : S_r \sqcap S_p\} \textit{ is satisfiable}$$

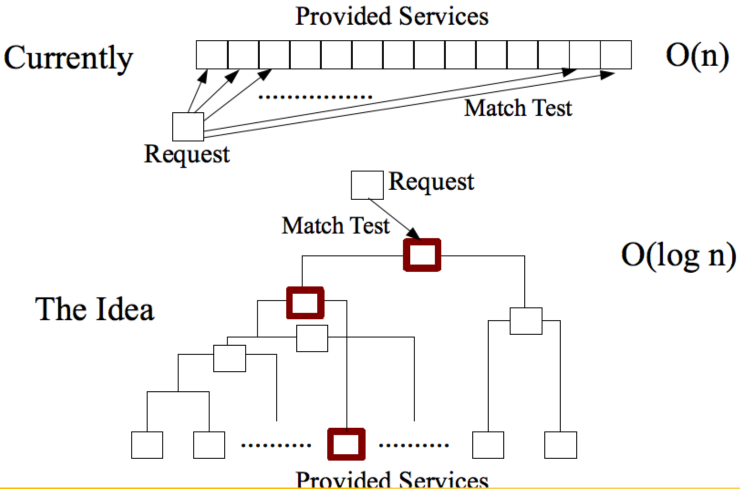
- **Entailment of Concept Subsumption [Paolucci 2002]**

$$KB \cup D_r \cup D_p \cup \{\exists x : S_r(x) \wedge S_p(x)\} \textit{ is consistent} \Leftrightarrow$$
$$\Leftrightarrow KB \cup D_r \cup D_p \cup \{i : S_r \sqcap S_p\} \textit{ is satisfiable}$$

- **Entailment of Concept Non-Disjointness [Grimm 2004]**

$$KB \cup D_r \cup D_p \models \exists x : S_r(x) \wedge S_p(x) \Leftrightarrow$$
$$\Leftrightarrow KB \cup D_r \cup D_p \cup \{S_r \sqcap S_p \sqsubseteq \bot\} \text{ is unsatisfiable}$$

Performing Service Matchmaking



Problems to Solve

- A *hierarchical agglomerative clustering method* is necessary in order to have a dendrogram (tree) as output of the clustering process
  - A *(dis-)similarity measure* applicable to *complex DL concept descriptions* is necessary for grouping elements
- A *conceptual clustering method* is necessary in order to generate *intensional cluster descriptions* of inner nodes
  - Availability of a *"good" generalization procedure*

Building intensional cluster descriptions

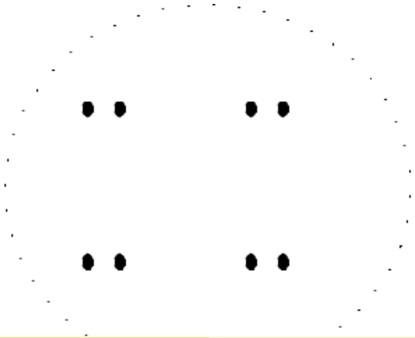
Possible generalization procedures

- *LCS*-$\mathcal{ALC}$ $\Rightarrow$ it could be too much specific (over-fitting)
- Approximating every $\mathcal{ALC}$ concept description to $\mathcal{ALE}$ description **[Brandt et al. 2002]** $\Rightarrow$ computing the *LCS*-$\mathcal{ALE}$
  - it could be too much general. Many TOP concepts could be generated, especially in presence of very simple concept descriptions
- Given an $\mathcal{ALC}$ T-Box and a set of $\mathcal{ALE}$-(T) concept descriptions, computing the *GCS* of such concept descriptions (namely the LCS-$\mathcal{ALE}$ computed w.r.t. the T-Box) **[Baader et al. 2004]** $\Rightarrow$ it seems to be the right compromise between the two solutions above

The hierarchical agglomerative clustering approach

Classical setting:

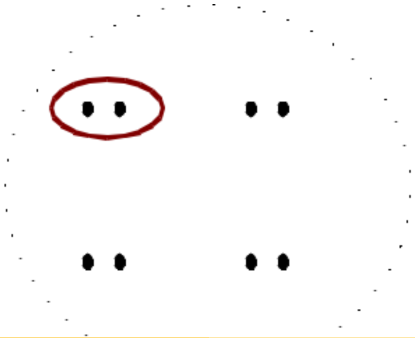
- Data are represented as feature vectors in an n-dimentional space
- Similarity is often measured in terms of geometrical distance



The hierarchical agglomerative clustering approach

Classical setting:

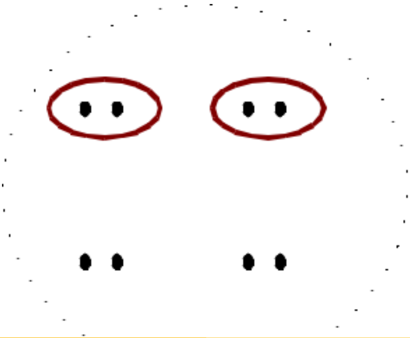
- Data are represented as feature vectors in an n-dimentional space
- Similarity is often measured in terms of geometrical distance



The hierarchical agglomerative clustering approach

Classical setting:

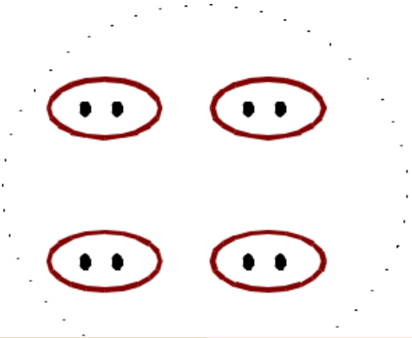
- Data are represented as feature vectors in an n-dimentional space
- Similarity is often measured in terms of geometrical distance



The hierarchical agglomerative clustering approach

Classical setting:

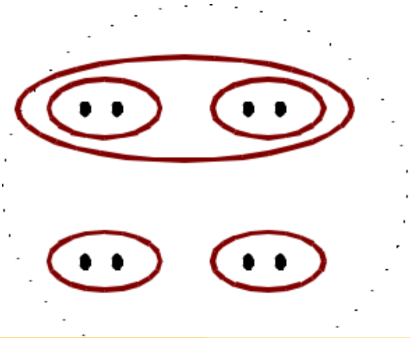
- Data are represented as feature vectors in an n-dimentional space
- Similarity is often measured in terms of geometrical distance



The hierarchical agglomerative clustering approach

Classical setting:

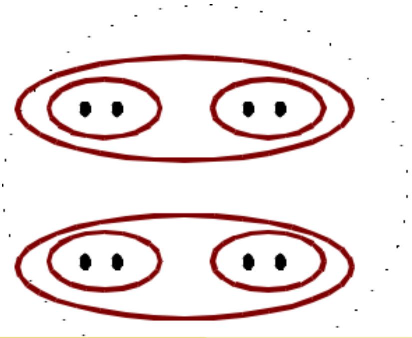
- Data are represented as feature vectors in an n-dimentional space
- Similarity is often measured in terms of geometrical distance



The hierarchical agglomerative clustering approach

Classical setting:

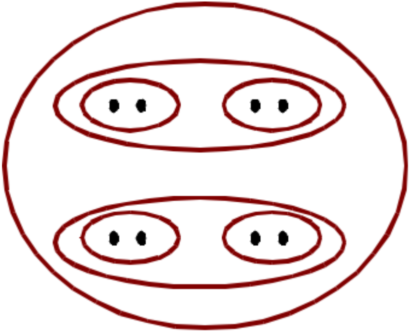
- Data are represented as feature vectors in an n-dimentional space
- Similarity is often measured in terms of geometrical distance



The hierarchical agglomerative clustering approach

Classical setting:

- Data are represented as feature vectors in an n-dimentional space
- Similarity is often measured in terms of geometrical distance



Single-link and Complete-link Algorithms

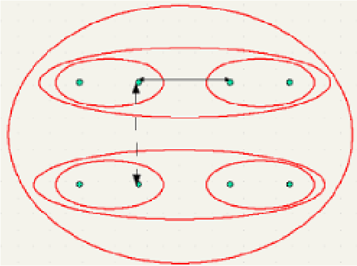


Figure B.1: Clustering process performed by the single-link algorithm. Cluster distances are given by the minimum distance among their elements.

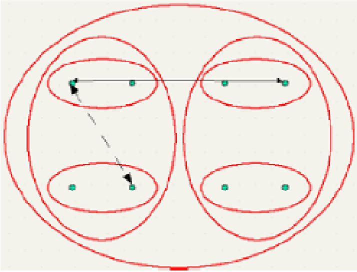


Figure B.2: Clustering process performed by the complete-link algorithm. Cluster distances are given by the maximum distance among their elements.

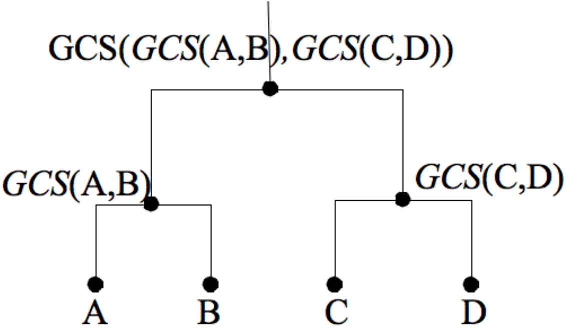
Realized clustering algorithm

DL-link algorithm [d'Amato et al. @ Service Matchmaking WS at ISWC 2007]

- Modified version of the *singl-link*, *complete-link* and *average link* algorithms
  - Able to cope with DL-based representations
  - Intentional cluster descriptions are given
  - Works directly with intentional cluster descriptions

DL-link Algorithm

- Output: binary tree (dendrogram) called **DL-Tree**
  - since, at every step, only two clusters are merged



Restructuring the DL-Tree

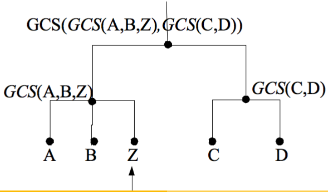
- Since redundant nodes do not add any information
  - If two (or more) children nodes of the DL-Tree have the same intentional description *or*
  - If a parent node has the same description of a child node
    - $\Rightarrow$ **a post-processing step is applied to the DL-Tree**

1. *If a child node is equal to another child node* $\Rightarrow$ one of them is deleted and their children nodes are assigned to the remaining node.
2. *If a child node is equal to a parent node* $\Rightarrow$ the child node is deleted and its children nodes are added as children of its parent node.
3. **The result of this flattening process is an n-ary DL-Tree**.

Updating the DL-Tree: e.g. a new service occurs

The DL-Tree has not to be entirely re-computed. **Indeed**:

1. The similarity value between **Z** and all available services is computed ⇒ the most similar service is selected.
2. **Z** is added as sibling node of the most similar service while
3. the parent node is re-computed as the GCS of the old child nodes plus **Z**.
4. In the same way, all the ancestor nodes of the new generated parent node are computed.



Service Discovery Evaluation

- hand-made service ontology: 256 concept descriptions, 96 service descriptions, 25 object properties
- Requested a service in the ontology (leaf node, inner node) and random queries
- Subsumption-based matching
- All services satisfying the request are returned

	Algorithm	*Metrics*	*Leaf Node*	*Inner Node*	*Random Query*
	DL-Tree based	*avg.*	**41.4**	**23.8**	**40.3**
		range	13 - 56	19 - 27	19 - 79
		avg. exc. time	266.4 ms.	180.2 ms.	483.5 ms.
	Linear	*avg.*	**96**	**96**	**96**
		avg. exc. time	678.2 ms.	532.5 ms.	1589.3 ms.

A criterion for Ranking Services

- Generally services selected by the matching process are returned in a flat list
- *Services selected by the matching* process, *have to be ranked* w.r.t. certain criteria (a total order would be preferable)
- *Ranking* procedure *based on* the use of a semantic *similarity measure for DL* concept descriptions.
  - **Provided services most similar to the requested service and satisfying both HC and SC of the request are ranked in the highest positions**
  - **Provided services less similar to the request and/or satisfying only HC are ranked in the lowest positions**

Ranking Services using Constraint Hardness

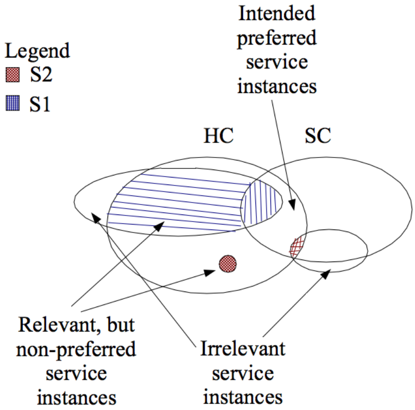
[d'Amato et al. @ Sem4WS Workshop at BMP 2006]

given:

$S_r = \{S_r^{HC}, S_r^{SC}\}$ service request;
S_p^i $(i = 1, .., n)$ provided services selected by *match*(KB, D_r, D_p^i);

for $i = 1, \dots, n$ **do**
compute $\bar{s}_i := s(S_r^{HC}, S_p^i)$
let be $S_r^{new} \equiv S_r^{HC} \sqcap S_r^{SC}$
for $i = 1, \dots, n$ **do**
compute $\bar{\bar{s}}_i := s(S_r^{new}, S_p^i)$
$s_i := (\bar{s}_i + \bar{\bar{s}}_i)/2$

Ranking Procedure: Rational



Ranking Services: Example...

$$\begin{aligned} D_r = \{ \quad & S_r \equiv \mathsf{Flight} \sqcap \forall \mathsf{operatedBy.LowCostCompany} \sqcap \exists \mathsf{to}.\{\mathsf{bari}\} \sqcap \\ & \qquad \sqcap \exists\ \mathsf{from}.\{\mathsf{cologne,hahn}\} \sqcap \forall \mathsf{hasFidelityCard.Card}; \\ & \{\mathsf{cologne,hahn}\} \sqsubseteq \exists\ \mathsf{from}^{-}.S_r \qquad \} \end{aligned}$$

where

$$\begin{aligned} HC_r = \{ \quad & \mathsf{Flight} \sqcap \exists \mathsf{to}.\{\mathsf{bari}\} \sqcap \exists\ \mathsf{from}.\{\mathsf{cologne,hahn}\} \\ & \{\mathsf{cologne,hahn}\} \sqsubseteq \exists\ \mathsf{from}^{-}.S_r \qquad \} \\ SC_r = \{ \quad & \mathsf{Flight} \sqcap \forall \mathsf{operatedBy.LowCostCompany} \sqcap \\ & \sqcap \forall \mathsf{hasFidelityCard.Card} \quad \}; \end{aligned}$$

...Ranking Services: Example...

$$D_p^l = \{ \quad S_p^l \equiv \mathsf{Flight} \sqcap \exists \mathsf{to.Italy} \sqcap \exists \mathsf{from.Germany};$$
$$\mathsf{Germany} \sqsubseteq \exists\ \mathsf{from}^-.S_p^l; \quad \mathsf{Italy} \sqsubseteq \exists\ \mathsf{to}^-.S_p^l \quad \}$$

where

$$HC_p^l = \{ \quad \mathsf{Flight} \sqcap \exists \mathsf{to.Italy} \sqcap \exists \mathsf{from.Germany};$$
$$\mathsf{Germany} \sqsubseteq \exists\ \mathsf{from}^-.S_p^l; \quad \mathsf{Italy} \sqsubseteq \exists\ \mathsf{to}^-.S_p^l \quad \}$$
$$SC_p^l = \{\}$$

$$D_p^k = \{ \quad S_p^k \equiv \mathsf{Flight} \sqcap \forall \mathsf{operatedBy.LowCostCompany} \sqcap \exists \mathsf{to.Italy} \sqcap$$
$$\sqcap \exists \mathsf{from.Germany};$$
$$\mathsf{Germany} \sqsubseteq \exists\ \mathsf{from}^-.S_p^k; \quad \mathsf{Italy} \sqsubseteq \exists\ \mathsf{to}^-.S_p^k \quad \}$$

where

$$HC_p^k = \{ \quad \mathsf{Flight} \sqcap \exists \mathsf{to.Italy} \sqcap \exists \mathsf{from.Germany};$$
$$\mathsf{Germany} \sqsubseteq \exists\ \mathsf{from}^-.S_p^k; \quad \mathsf{Italy} \sqsubseteq \exists\ \mathsf{to}^-.S_p^k \quad \}$$
$$SC_p^k = \{ \quad \mathsf{Flight} \sqcap \forall \mathsf{operatedBy.LowCostCompany}\};$$

$$KB = \{\mathsf{cologne,hahn{:}Germany,\ bari{:}Italy,\ LowCostCompany} \sqsubseteq \mathsf{Company}\ \}$$

...Ranking Services: Example...

Note that:

- S_p^l satisfies only *HC* of S_r
- S_p^k satisfies both *HC* and *SC* of S_r
- Suppose that $|(S_p^l)^{\mathcal{I}}| = 8$ and $|(S_p^k)^{\mathcal{I}}| = 5$ and *all instances satisfy* S_r.
    - Note that $S_p^k \sqsubseteq S_p^l$ then $(S_p^k)^{\mathcal{I}} \subseteq (S_p^l)^{\mathcal{I}} \Rightarrow |(S_r)^{\mathcal{I}}| = 8$.
    - $|(S_r^{HC} \sqcap S_p^l)^{\mathcal{I}}| = 8$ and that $|((S_r^{HC} \sqcap S_r^{SC}) \sqcap S_p^l)^{\mathcal{I}}| = |(S_r^{new} \sqcap S_p^l)^{\mathcal{I}}| = 0 \Rightarrow \overline{\overline{s_l}} = 0$,
- *SC* of S_p^k are subsumed by *SC* of S_r (namely by S_r^{SC})
    - Let us suppose that instances of S_p^k that satisfy both *HC* and *SC* of S_r, namely that satisfy $S_r^{new} \equiv S_r^{HC} \sqcap S_r^{SC}$ are 3.

...Ranking Services: Example...

$$\bar{s_l} := s(S_r^{HC}, S_p^l) = \frac{|(S_r^{HC} \sqcap S_p^l)^{\mathcal{I}}|}{|(S_r^{HC} \sqcup S_p^l)^{\mathcal{I}}|} \cdot \max\left(\frac{|(S_r^{HC} \sqcap S_p^l)^{\mathcal{I}}|}{|(S_r^{HC})^{\mathcal{I}}|}, \frac{|(S_r^{HC} \sqcap S_p^l)^{\mathcal{I}}|}{|(S_p^l)^{\mathcal{I}}|)}\right) = \frac{8}{8} \cdot \max\left(\frac{8}{8}, \frac{8}{8}\right) = 1$$

$$\bar{s_k} := s(S_r^{HC}, S_p^k) = \frac{|(S_r^{HC} \sqcap S_p^k)^{\mathcal{I}}|}{|(S_r^{HC} \sqcup S_p^k)^{\mathcal{I}}|} \cdot \max\left(\frac{|(S_r^{HC} \sqcap S_p^k)^{\mathcal{I}}|}{|(S_r^{HC})^{\mathcal{I}}|}, \frac{|(S_r^{HC} \sqcap S_p^k)^{\mathcal{I}}|}{|(S_p^k)^{\mathcal{I}}|)}\right) = \frac{5}{8} \cdot \max\left(\frac{5}{8}, \frac{5}{5}\right) = \frac{5}{8} = 0.625$$

$$\bar{\bar{s_l}} := s(S_r^{new}, S_p^l) = 0$$

$$\bar{\bar{s_k}} := s(S_r^{new}, S_p^k) = \frac{|(S_r^{new} \sqcap S_p^k)^{\mathcal{I}}|}{|(S_r^{new} \sqcup S_p^k)^{\mathcal{I}}|} \cdot \max\left(\frac{|(S_r^{new} \sqcap S_p^k)^{\mathcal{I}}|}{|(S_r^{new})^{\mathcal{I}}|}, \frac{|(S_r^{new} \sqcap S_p^k)^{\mathcal{I}}|}{|(S_p^k)^{\mathcal{I}}|)}\right) = \frac{3}{5} \cdot \max\left(\frac{3}{3}, \frac{3}{5}\right) = \frac{3}{5} = 0.6$$

...Ranking Services: Example...

$s_l = \frac{\overline{s_l} + \overline{\overline{s_l}}}{2} = \frac{1+0}{2} = 0.5$ $\qquad$ $s_k = \frac{\overline{s_k} + \overline{\overline{s_k}}}{2} = \frac{0.625+0.6}{2} = 0.6125$

1. S_p^k *Similarity Value* 0.6125
2. S_p^l *Similarity Value* 0.5

Exactly what we want!!!

Conclusions

- A set of semantic (dis-)similarity measures for DLs has been presented
  - Able to assess (dis-)similarity between complex concepts, individuals and concept/individual
- Experimentally evaluated by embedding them in some inductive-learning algorithms applied to the SW and SWS domanis
- Realized an instance based classifier (K-NN and SVM) able to outperform concept retrieval and induce new knowledge
- Realized a set of clustering algorithms for improving the service discovery task
- A new ranking services procedure has been proposed based on the exploitation of a (dis-)similarity measure and constraint hardness

Future Works...

- Extention of Similarity and Dissimilarity Measures for most expressive DL such as $\mathcal{ALCN}$
  - This could allow to cope with a wide range real life problems
- Explicitly treat roles contribution in assessing (dis-)similarity (currently only implicitly treated)
- Extension of the semi-distance measure for treating complex descriptions
  - Setting a method for determining the minimal discriminating feature set
- Make possible the applicability of the measures to concepts/individuals asserted in different ontologies (for using them in tasks such as: ontology matching and alignment)

...Future Works

- **The k-NN-based classifier** could be extended with different answering procedures grounded on statistical inference (non-parametric tests based on ranked distances) in order to accept answers as correct with a high degree of confidence.
- The k-NN-based classifier could be extended in a way such that the probability that an individual belongs to one or more concepts are given.
- **For clusters-based discovery process** an heuristic (for finding the most appropriate service) could be useful for the cases in which, at the same level, more than one branch satisfy the matching test
- An incremental clustering method would be investigated for up dating clusters when a new provided service is available

The End

That's all!

Thanks for your attention