Machine Learning for Ontology Mining: Perspectives and Issues

Claudia d'Amato

Department of Computer Science
University of Bari

OWLED 2014 ◇ Riva del Garda, October 18, 2014

Contents

Introduction & Motivations

- In the SW, ontologies play a key role
- They are equipped with deductive reasoning capabilities
  - they may fail
    - on large scale
    - when data are incoherent/noisy
- **Idea:** exploiting Machine Learning methods for *Ontology Mining* related tasks

Ontology Mining: Definition

Ontology Mining

all activities that allow

- to discover hidden knowledge from ontological knowledge bases
- by possibly using only a sample of data

Machine Learning: Basics

Machine Learning (ML) methods

- focus on the development of methods and algorithms that can teach themselves to grow and change when exposed to new data

Special Focus on:

- (similarity-based) *inductive learning methods*
  - use specific examples to reach general conclusions
  - are known to be very efficient and fault-tolerant

Induction vs. Deduction

Deduction (Truth preserving)

Given:

- a set of general axioms
- a proof procedure

Draw:

- *correct and certain* conclusions

Induction (Falsity preserving)

Given:

- a set of examples

Determine:

- a *possible/plausible* generalization covering
    - the given examples/observations
    - new and not previously observed examples

Inductive Learning Approaches and Tasks I

Supervised (Learning from examples)

- Given a training set $\{(\mathbf{x_1}, y_1), \ldots (\mathbf{x_n}, y_n)\}$ where $\mathbf{x_i}$ are input examples and y_i the desired output, learn an unknown function f such that $f(x) = y$ for new examples
  - y having discrete values $\Rightarrow$ *Classification Problem*
  - y having continuos values $\Rightarrow$ *Regression Problem*
  - y having a probability value $\Rightarrow$ *Probability Estimation Problem*
- Supervised Concept Learning:
  - Given a training set of positive and negative examples for a concept,
  - *construct* a description that will accurately classify whether future examples are positive or negative.

Inductive Learning Approaches and Tasks II

Unsupervised (Learning from Observations)

- Given a set of observations $\{\mathbf{x_1}, \ldots \mathbf{x_n}\}$
  - discover hidden patterns in the data $\Rightarrow$ *Discovery*
  - for a concept/class/category, construct a description that is able to determine if a (new) example is an instance of the concept (positive example) or not (called negative example). $\Rightarrow$ *Concept Learning*
  - assess groups of similar data items $\Rightarrow$ *Clustering*

Semi-supervised learning

- is halfway between supervised and unsupervised learning
- training data is built up by both few labeled (i.e. with the desired output) and unlabeled data
- both kinds of data are used for solving the learning tasks (almost the same tasks as for the case of supervised learning)

Focus I

Exploitation of Inductive Learning for performing:

- approximate inductive instance retrieval
  - regarded as a classification problem $\Rightarrow$ (semi-)automatic ontology population
- automatic concept drift and novelty detection
  - regarded as a clustering (and successive concept learning) problem
- semi-automatic ontology enrichment
  - regarded as pattern discovery problem problem
    - exploiting the evidence coming from the data $\Rightarrow$ discovering hidden knowledge patterns in the form of relational association rules
    - existing ontologies can be straightforwardly extended with formal rules
    - new axioms may be suggested

Instance Retrieval as a Classification Problem

Issues & Solutions I

Focus: *Instance Retrieval* → finding the extension of a query concept

- Task *casted* as a classification problem
  - *assess the class membership of the individuals in a KB w.r.t. the query concept*

State of the art classification methods cannot be straightforwardly applied for the purpose, since

- they are generally applied to *feature vector* representation
  → *upgrade DL expressive representations*
- An implicit *Closed World Assumption* is made in ML
  → *cope with the Open World Assumption made in DLs*
- Classification: classes considered as *disjoint*
  → *cannot assume disjointness of all concepts*

Issues & Solutions II

Adopted Solutions:

- Defined new semantic similarity measures for DL representations
  - to cope with the high expressive power of DLs
  - to convey the underlying semantics of KB
  - to deal with the semantics of the compared objects (concepts, individuals, ontologies)
- Formalized a set of criteria that a similarity function has to satisfy in order to be defined *semantic [d'Amato et al. @ EKAW 2008]*
- Definition of the classification problem taking into account the OWA
- Multi-class classification problem decomposed into a set a smaller classification problems

Definition (Problem Definition)

Given:

- a populated ontological knowledge base $KB = (\mathcal{T}, \mathcal{A})$
- a query concept Q
- a training set with $\{+1, -1, 0\}$ as target values

Learn a classification function f such that: $\forall a \in Ind(\mathcal{A})$:

- $f(a) = +1$ if a is instance of Q
- $f(a) = -1$ if a is instance of $\neg Q$
- $f(a) = 0$ otherwise (unknown classification because of OWA)

Dual Problem

- given an individual $a \in Ind(\mathcal{A})$, tell concepts $C_1, \ldots, C_k$ in KB it belongs to
- the multi-class classification problem is *decomposed* into a set of *ternary classification problems* (one per target concept)

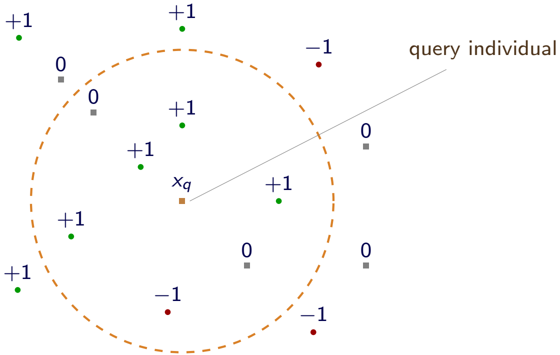
Developed methods

- relational K-NN for DL KBs *[d'Amato et al. @ ESWC 2008]*
- kernel functions for kernel methods to be applied to DLs KBs *[Fanizzi et al. @ JWS 2012, Bloehdorn and Sure @ ISWC'06]*
- REDUCE - grounded on Reduced Coulomb Energy Networks *[Fanizzi et al. @ IJSWIS 2009]*
- TERMITIS - grounded on the induction of Terminological Decision Trees *[Fanizzi et al. @ ECML/PKDD'10]*

Example: Nearest Neighbor Classification

query concept `HardWorker` $k = 7$

target values standing for the class values: $\{+1, 0, -1\}$

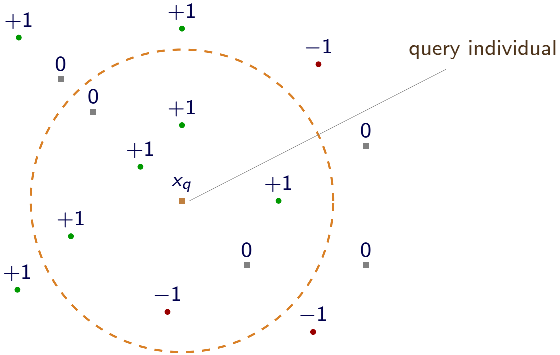


$class(x_q) \leftarrow ?$

Example: Nearest Neighbor Classification

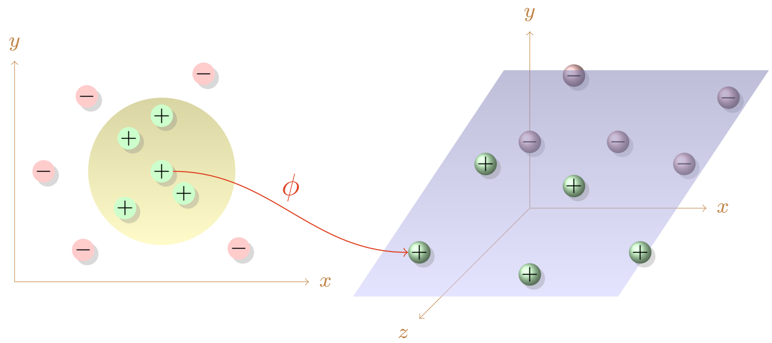
query concept `HardWorker` $k = 7$

target values standing for the class values: $\{+1, 0, -1\}$



$class(x_q) \leftarrow +1$

Example: Kernel Method Classification



On evaluating the Classifiers

Problem: How to evaluate the classification results

- Performance compared with a standard reasoner (PELLET)
- Registered *cases in which the reasoner did not return any result, differently from the classifier*
- **Behavior registered as mistake if precision and recall where used** while it could turn out to be a correct inference when judged by a human

Defined new metrics for evaluating the performances of the classifiers

- *To distinguish between inductively classified individuals and real mistakes additional indices have been considered.*

Additional Evaluation Parameters

- *match rate*: cases of match of the classification returns by both procedures.
- *omission error rate*: cases when our procedure cannot decide (0) while the reasoner gave a classification (± 1)
- *commission error rate*: cases when our procedure returned ± 1 while the reasoner gave the opposite outcome ∓ 1
- *induction rate*: cases when the reasoner cannot decide (0) while our procedure gave a classification (± 1)

Lesson Learnt from experiments I

Experiments performed on ontologies publicly available

- *Commission error* almost null on average
- *Omission error rate* almost null
- *Induction Rate* not null
    - **new knowledge (not logically derivable) is induced** $\Rightarrow$ it can be used for making the *ontology population task semi-automatic*
    - induced knowledge can be found $\Rightarrow$ *individuals are instances of many concepts* and they are *homogeneously spread* w.r.t. the several concepts.
- most of the time the most *effective* method $\Rightarrow$ relational K-NN
- the most scalable method $\Rightarrow$ kernel method embedding a DL kernel function

Lesson Learnt from experiments II

- Since inductive conclusions are not certain, probabilities for the classification results may be computed

$$\Downarrow$$

- Probability values may be ultimately used for completing ontologies with probabilistic assertions
  - enabling more sophisticate approaches to dealing with uncertainty

Concept Drift and Novelty Detection as a Clustering and successive Concept Learning Problem

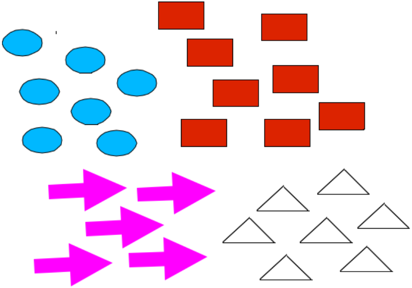
Concept Drift and Novelty Detection

- Ontologies evolve over the time.
  - *New assertions*
  - New concept definitions
- **Concept Drift**
  - the change of a known concept w.r.t. the evidence provided by new annotated individuals that may be made available over time
    - almost all `Workers` work for more than 10 hours per days ⇒ `HardWorker`
- **Novelty Detection**
  - isolated cluster in the search space that requires to be defined through new emerging concepts to be added to the KB
    - subset of `Workers` *employed* in a company ⇒ `Employ`
    - subset of `Workers` *working for* one or more companies ⇒ `Freelance`
- *FOCUS* : **(Conceptual) clustering methods for automatically discover them**

Basics on Clustering Methods

Clustering methods: unsupervised inductive learning methods that organize a collection of unlabeled resources into meaningful clusters such that

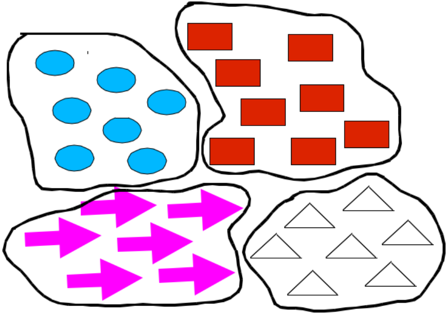
- intra-cluster *similarity* is high
- inter-cluster *similarity* is low



Basics on Clustering Methods

Clustering methods: unsupervised inductive learning methods that organize a collection of unlabeled resources into meaningful clusters such that

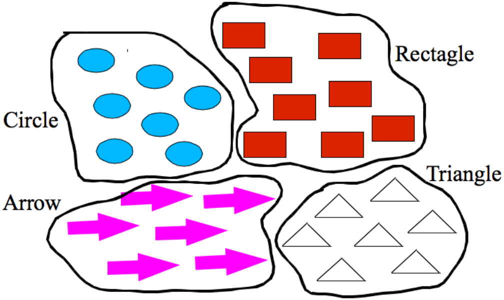
- intra-cluster *similarity* is high
- inter-cluster *similarity* is low



Basics on Clustering Methods

Clustering methods: unsupervised inductive learning methods that organize a collection of unlabeled resources into meaningful clusters such that

- intra-cluster *similarity* is high
- inter-cluster *similarity* is low



Clustering Individuals of An Ontology: Developed Methods

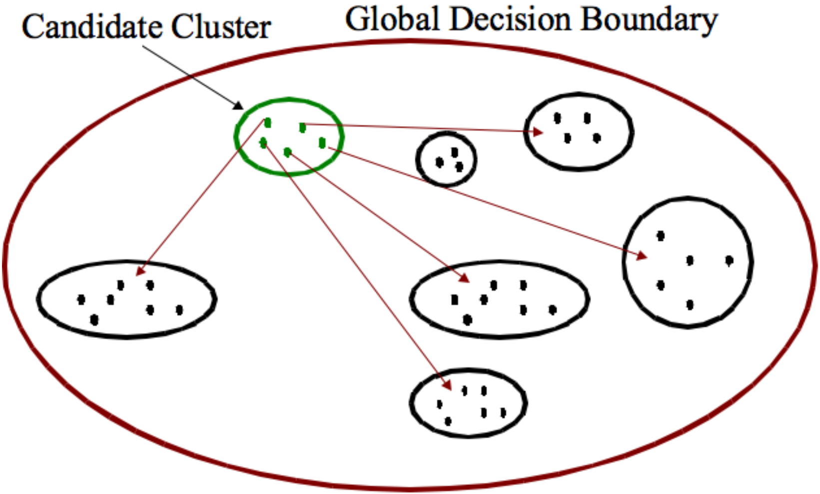
- KLUSTER *[Kietz & Morik, 94]*
- CSKA *[Fanizzi et al., 04]*
  - Produce a *flat output*
  - *Suffer from noise* in the data
- Similarity-based $\Rightarrow$ *noise tolerant*
  - Evolutionary Clustering Algorithm around Medoids *[Fanizzi et al. @ IJSWIS 2008]*
    - automatically assess the best number of clusters
  - k-Medoid (hierarchical and fuzzy) clustering algorithm *[Fanizzi et al. @ ESWC'08, Fundam. Inform. 2010]*
    - number of clusters required

Automated Concept Drift and Novelty Detection

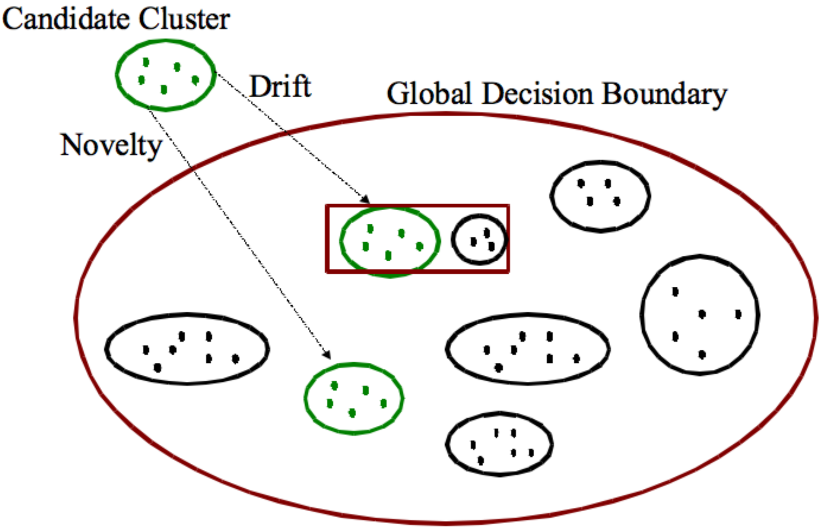
If *new annotated individuals are made available* they have to be integrated in the clustering model

1. Each individual is assigned to the closest cluster (measuring the distance w.r.t. the cluster medoids)
2. The entire clustering model is recomputed
3. The new instances are considered to be a *candidate* cluster
   - An *evaluation* of it is performed in order to assess its nature

Evaluating the Candidate Cluster: Main Idea 1/2



Evaluating the Candidate Cluster: Main Idea 2/2



Evaluating Concept Drift and Novelty Detection

- The *Global Cluster Medoid* is computed
  $\overline{m} := \mathrm{medoid}(\{m_j \mid C_j \in \mathsf{Model}\})$
- $d_{\max} := \max_{m_j \in \mathsf{Model}} d(\overline{m}, m_j)$
- if $d(\overline{m}, m_{CC}) \leq d_{\max}$ the CandCluster is a *Concept Drift*
  - CandCluster is **Merged** with the most similar cluster $C_j \in \mathsf{Model}$
- if $d(\overline{m}, m_{CC}) \geq d_{\max}$ the CandCluster is a *Novel Concept*
  - CandCluster is **added** to the model

Lesson Learnt from Experiments

Clustering algorithms

- applied on ontologies publicly available
- evaluated by the use of standard validity clustering indexes (e.g. Generalized Dunns index, cohesion index, Silhouette index)
- necessity of a domain expert/gold standard particularly for validating the concept novelty/drift

Conceptual Clustering Step

Performed as a *supervised concept learning phase*

Definition (Problem Definition)

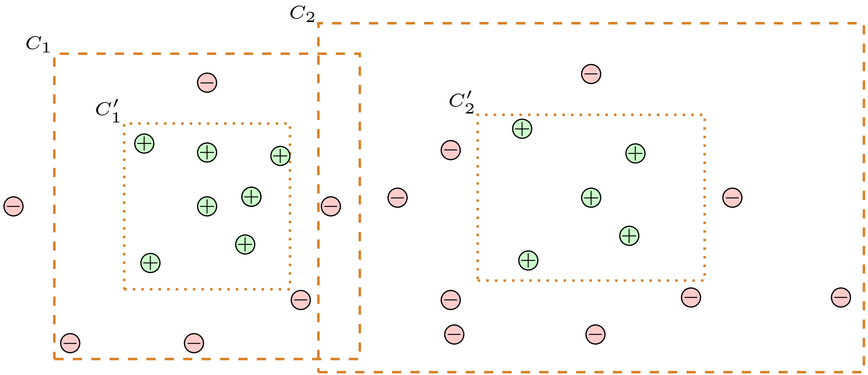
- *Given*
    - individuals in a cluster C as positive examples
    - the individuals in the other clusters as negative examples
    - The KB $\mathcal{K}$ as a background knowledge
- *Learn*
    - a DL concept description D so that
    - the individuals in the target cluster C are instances of D while those in the other clusters are not

The new descriptions could be used for *enriching* the ontology

Developed Methods for Supervised Concept Learning

- For DLs that allow for (approximations of) the msc and lcs, (e.g. $\mathcal{ALC}$ or $\mathcal{ALE}$):
  - given a cluster C_j,
    - $\forall a_i \in \mathrm{C}_j$ compute $M_i := msc(a_i)$ w.r.t. the ABox $\mathcal{A}$
    - let $MSCs_j := \{M_i | a_i \in \mathrm{node}_j\}$
  - C_j *intensional description* $lcs(MSCs_j)$
- Separate-and-conquer approach
  - YinYang *[Iannone et al. @ Appl. Intell. J. 2007]*
  - DL-FOIL *[Fanizzi et al. @ ILP 2008]*
  - DL-Lerner *[Lehmann and Hitzler @ MLJ 2010]*
- Divide-and-conquer approach
  - TermiTIS *[Fanizzi et al. @ ECML 2010]*

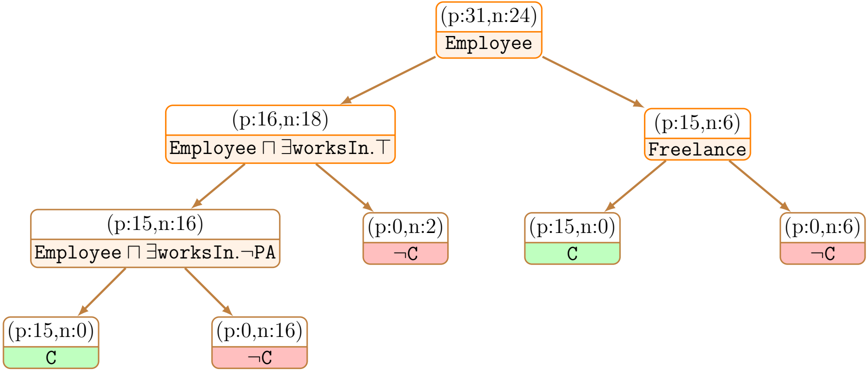
Separate and Conquer: Example



$C_1 = \texttt{MasterStudent}$ $\quad C'_1 = \texttt{MasterStudent} \sqcap \exists \texttt{worskIn}.\top$

$C_2 = \texttt{BachelorStudent}$ $\quad C'_2 = \texttt{BachelorStudent} \sqcap \exists \texttt{worskIn}.\top$

Divide and Conquer: Example



Examples of Learned Concept Descriptions with DL-FOIL

BIOPAX

induced:

```
Or( And( physicalEntity protein) dataSource)
```

original:

```
Or( And( And( dataSource externalReferenceUtilityClass)
ForAll(ORGANISM ForAll(CONTROLLED phys icalInteraction)))
protein)
```

NTN

induced:

```
Or( EvilSupernaturalBeing Not(God))
```

original:

```
Not(God)
```

FINANCIAL

induced:

```
Or( Not(Finished) NotPaidFinishedLoan Weekly)
```

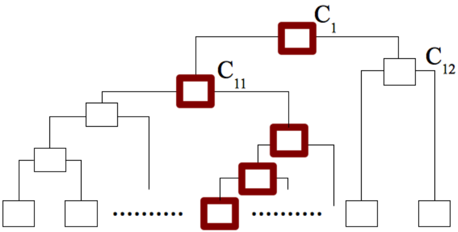
original:

```
Or( LoanPayment Not(NoProblemsFinishedLoan))
```

Clustering Individuals of An Ontology: Additional Usage

Realized hierarchical clustering algorithms whose dendrogrom (tree-structure) is exploited *[d'Amato et al. @ ESWC'08, IJSC 2010]*

- as an index for speeding up the resource retrieval
  - Obtained logaritmic complexity rather than linear complexity
- to improve the readability of the query results (e.g. from SPARQL-queries) and for performing a kind of faceted search
  - **Lesson Learnt:** Divisional *rather than* agglomerative methods should be employed



Ontology enrichment as Pattern Discovery Problem

Starting Observations

Focussing on Ontological Knowledge Bases

- Ontological knowledge bases are often not *complete*
  - i.e. missing concept and role assertions, disjointness axioms, relationships that instead occur in the reference domain
- **Idea:** exploiting the evidence coming from the data for discovering hidden knowledge patterns to be used for
  - extending existing ontologies with formal rules
  - suggesting knew knowledge axioms
- **Research Direction:** discovering hidden knowledge patterns in the form of relational association rules *[d'Amato and Staab @ TR 2013]*

Related Works

- Galárraga et al. @ WWW'13
  - discovering of association rules for predicting new role assertions from an RDF data source (no reasoning capabilities and no TBox information exploited)
- Lisi @ IJSWIS 7(3), 2011
  - discovering of frequent patterns in the form of DATALOG clauses from an $\mathcal{AL} - Log$ KB at different granularity level w.r.t. the taxonomic ontology
- Völker & Niepert @ ESWC'11
  - association rules are learnt from RDF data (without any reasoning features) for inducing a schema ontology for them
- Józefowska, Lawrynowicz et al. @ TPLP 10(3), 2010
  - discovery of frequent patterns, in the form of conjunctive queries, from a combined DL KB plus rules
- Joshi, Hitzler et al. @ ESWC 2013
  - association rules are exploited for performing RDF data compression

Definition (Problem Definition)

Given:

- a populated ontological knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$
- a minimum "frequency threshold" (fr_thr)
- a minimum "head coverage threshold" (cov_thr)

Discover:

- all frequent hidden patterns, with respect to fr_thr, in the form of relational association rules that may induce new assertions for $\mathcal{K}$.

Definition (Relational Association Rule)

Given

- a populated ontological knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$,

a **relational association rule** r for $\mathcal{K}$ is a horn-like clause of kind

$$body \rightarrow head$$

where:

- *body* represents an abstraction of a set of assertions in $\mathcal{K}$ co-occurring with respect to fr_thr
- *head* represents a possibly new assertion induced from $\mathcal{K}$ and *body*

SWRL [Horrocks et al.@ WWW'04] is adopted as representation language.

- allows to extends the OWL axioms of an ontology with Horn-like rules
- The result is a KB with an enriched expressive power.

Discovering *SWRL* rules of the form:

$$C_1(x) \wedge R_1(x,y) \wedge \cdots \wedge C_n(z) \wedge R_l(z,a) \rightarrow R_k(y,z)$$
$$C_1(x) \wedge R_1(x,y) \wedge \ldots C_n(z) \wedge R_l(z,a) \rightarrow C_h(y)$$

C_i and R_i are concept and role names of the ontological KB

Examples:

- $Woman(x) \wedge hasWellPayedJob(x,y) \Rightarrow Single(x)$
- $Employ(x) \wedge worksAt(x,z) \wedge workForPrject(x,y) \wedge projectSupervisor(y,x) \Rightarrow CompanyManager(z,x)$

Language Bias (ensuring decidability)

- *safety condition* : all variables in the head must appear in the body
- *connection* : atoms share at least a variable or a constant
- interpretation under $DL - Safety$ condition: all variables in the rule bind only to known individuals in the ontology
- *Non Redundancy*: there are no atoms that can be derived by other atoms

The General Approach

- Inspired to the general framework for discovering frequent DATALOG patterns [*Dehaspe et al.*'99; *Goethals et al.*'02] where patterns are conjunctive DATALOG queries
- Grounded on a level-wise *generate-and-test* approach
  - Start: initial general pattern i.e. a concept name (jointly with a variable name) or a role name (jointly with variable names)
  - Proceed: at each level with
    - specializing the pattern by the use of suitable operators
    - evaluate the generated specializations for possible pruning
  - Stop: stopping criterion met
- *A rule is a list of atoms* (interpreted as a conjunction) where the *first one* represent *the head* [*Galarraga et al.@WWW*'13]

Pattern Specializations: Examples

Pattern to be Specialized $C(x) \wedge R(x, y)$

Non Redundant Concept D

Refined Patterns

1. $C(x) \wedge R(x, y) \wedge D(x)$
2. $C(x) \wedge R(x, y) \wedge D(y)$

Non Redundant Role S Fresh Variable z

Refined Patterns

1. $C(x) \wedge R(x, y) \wedge S(x, z)$
2. $C(x) \wedge R(x, y) \wedge S(z, x)$
3. $C(x) \wedge R(x, y) \wedge S(y, z)$
4. $C(x) \wedge R(x, y) \wedge S(z, y)$

Non Redundant Role S All Variables Binded

Refined Patterns

1. $C(x) \wedge R(x, y) \wedge S(x, x)$
2. $C(x) \wedge R(x, y) \wedge S(x, y)$
3. $C(x) \wedge R(x, y) \wedge S(y, x)$
4. $C(x) \wedge R(x, y) \wedge S(y, y)$

Exploitation of the Association Rules and Utility

Examples:

- (Semi-)automatic **ABox completion**
  - rules may fire new assertions
  - alternatively extracted rules may be used by a rule-based classifier
- **Ontology Enrichment**
  - A rule may suggest an inclusion axiom that is missing in the ontology e.g. $Car(x) \Rightarrow Vehicle(x)$
  - A rule may suggest a disjointness axiom axiom that is missing in the ontology $Man(x) \Rightarrow \neg Woman(x)$
- **Creating Ontology with Enriched expressive power**
  - discovered rules can be straightforwardly integrated with the existing ontology

Issues/Lessons Learnt

- Experimental evaluation for accessing the effectiveness of the method: how to set up it?
  - by considering smaller versions of ontologies, evaluate the correctness of predicted assertions when compared with the full ontology versions
- Develop a scalable algorithm for the purpose
  - investigate on *additional heuristics for reducing* the exploration of *the search space* and/or possible optimizations
  - (New) metrics for the evaluation of the *interestingness of the discovered rules* (potential inner and post pruning)
  - Set up/exploit *suitable data structures* i.e. Hash Table, RDB with indexes for *minimizing the usage of the reasoner* $\Rightarrow$ bottleneck
  - Alternative method for generating the rules by considering subsets of frequent patterns

Starting Obervations

Focussing on Heterogenous Sources of Information

- Available domain ontologies are increasing over the time
- Large amount of data stored and managed with RDBMS
- *Ontologies and RDB* may be used for complementing the knowledge for a given domain

Idea: exploiting the evidence coming from the data for **discoverying hidden KB patterns** *across heterogeneous sources* to be used for

1. possibly completing/complementing both sources of knowledge
2. empowering the reasoning process

Simple Motivating Example...

- Let $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a kingship ontology

$$\mathcal{T} = \left\{ \begin{array}{lll} \mathsf{Person} \equiv \mathsf{Man} \sqcup \mathsf{Woman} & \mathsf{Man} \sqsubseteq \neg\mathsf{Woman} & \top \sqsubseteq \forall \mathsf{hasChild.Person} \\ \exists \mathsf{hasChild}.\top \sqsubseteq \mathsf{Person} & \mathsf{Parent} \equiv \exists \mathsf{hasChild.Person} & \mathsf{Mother} \equiv \mathsf{Woman} \sqcap \mathsf{Parent} \\ \mathsf{Father} \equiv \mathsf{Man} \sqcap \mathsf{Parent} & \mathsf{Grandparent} \equiv \exists \mathsf{HasChild.Parent} & \mathsf{Child} \equiv \exists \mathsf{HasChild}^{-}.\top \end{array} \right\}$$

$$\mathcal{A} = \left\{ \begin{array}{llll} \mathsf{Woman(alice)} & \mathsf{Man(xavier)} & \mathsf{hasChild(alice, claude)} & \mathsf{hasChild(alice, daniel)} \\ \mathsf{Man(bob)} & \mathsf{Woman(yoana)} & \mathsf{hasChild(bob, claude)} & \mathsf{hasChild(bob, daniel)} \\ \mathsf{Woman(claude)} & \mathsf{Woman(zurina)} & \mathsf{hasChild(xavier, zurina)} & \mathsf{hasChild(yoana, zurina)} \\ \mathsf{Man(daniel)} & \mathsf{Woman(maria)} & \mathsf{hasChild(daniel, maria)} & \mathsf{hasChild(zurina, maria)} \end{array} \right\}$$

- Let **D** be a job information database

ID	Name	Surname	Qualification	Salary	Age	City	Address
p001	*Alice*	Lopez	Housewife	0	60	Bari	Apulia Avenue 10
p002	*Robert*	Lorusso	Bank-employee	30.000	55	Bari	Apulia Avenue 10
p003	*Xavier*	Garcia	Policeman	35.000	58	Barcelona	Carrer de Manso 20
p004	*Claude*	Lorusso	Researcher	30.000	35	Bari	Apulia Avenue 13
p005	*Daniel*	Lorusso	Post Doc	25.000	28	Madrid	calle de Andalucia 12
p006	*Yoana*	Lopez	Teacher	34.000	49	Barcelona	Carrer de Manso 20
p007	*Zurina*	Garcia-Lopez	Ph.D student	20.000	25	Madrid	calle de Andalucia
p008	*Maria*	Lorusso	Pupil	0	8	Madrid	calle de Andalucia

...Simple Motivating Example

By jointly analyzing the available knowledge sources new additional information could be induced e.g.

- Women earning the highest amount of money are not mothers where:
  - information on being Woman and Mother comes from the ontology
  - information concerning the salary comes from the DB **D**.

Intended Directions: [d'Amato et al.@URSW III Ch.]

- Learning *Semantically Enriched Association Rules* *from both sources of knowledge* in an integrated way
- set up an effective *data-driven Tableaux algorithm* exploiting the evidence coming from the data for assessing the "most plausible model" for a given concept description

Building an Integrated Data Source: Main Idea

Construction of a unique table from D and $\mathcal{K}$

- State of the art implemented algorithms for learning Association Rules can be directly applied.
- No export of existing RDB has to be performed

Precondition/Assumption:

- dataset **D** and an ontological knowledge base $\mathcal{K}$ share (a subset of) common individuals
- a relation g that connects (some of) the individuals in $\mathcal{K}$ with (some of) the objects of **D** is available

Building an Integrated Data Source: Example

1. OBJECT primary entity
2. JOB, AGE selected attributes from **D**.
3. Person, Parent, Male, Female selected concept names from $\mathcal{K}$

Numeric attributes discretised

OBJECT	JOB	AGE	*Person*	*Parent*	*Male*	*Female*
x_1	Engineer	[36,45]	*true*	*true*	*true*	*false*
x_2	Policeman	[26,35]	*true*	*false*	*true*	*unknown*
x_3	Student	[16,25]	*true*	*false*	*true*	*false*
x_4	Student	[16,25]	*true*	*false*	*false*	*true*
x_5	Housewife	[26,35]	*true*	*true*	*false*	*true*
x_6	Clerk	[26,35]	*true*	*false*	*unknown*	*unknown*
x_7	Primary school teacher	[46,55]	*true*	*unknown*	*unknown*	*unknown*
x_8	Policeman	[16,25]	*true*	*true*	*unknown*	*unknown*
x_9	Student	[16,25]	*true*	*unknown*	*unknown*	*unknown*
...	...	...	...	...	...	...

Learning Semantically Enriched ARs

Given the integrated data source

- an APRIORI-like algorithm can be used to discover the set of frequent items
- the association rules are extracted

Example of extracted rules

#	RULE	Confidence
r1	$(\textsc{Age}{=}[16,25]) \wedge (\textsc{Job} = \textit{Student}) \Rightarrow (\textit{Parent} = \textit{false})$	0.98
r2	$(\textsc{Job}{=}\textit{Policeman}) \Rightarrow (\textit{Male} = \textit{true})$	0.75
r3	$(\textsc{Age}{=}[16,25]) \wedge (\textit{Parent} = \textit{true}) \Rightarrow (\textit{Female} = \textit{true})$	0.75
r4	$(\textsc{Job}{=}\textit{Primary school teacher}) \Rightarrow (\textit{Female} = \textit{true})$	0.78
r5	$(\textsc{Job}{=}\textit{Housewife}) \wedge (\textsc{Age} = [26,35]) \Rightarrow (\textit{Parent} = \textit{true}) \wedge (\textit{Female} = \textit{true})$	0.85

Exploitation of the Association Rules

Performing Data Analysis

- rule suggests the average age of being a
  parent in Madrid that could be different in other geographical areas, e.g.

$$(\textsc{Age}=[25,34]) \wedge (\textsc{City} = \text{Madrid}) \Rightarrow (\text{HasChild} = \textit{true})$$

Data completion (both in $\mathcal{K}$ and D)

- rule may allow some individuals to be asserted as instance of the concept Worker in $\mathcal{K}$(when not known) e.g.

$$\textsc{Salary}=[15000, 24999] \Rightarrow (\text{Worker} = \textit{true})$$

Ontology Enrichment

- rule may suggest a disjointness axiom (if absent in $\mathcal{K}$but extensionally provided) e.g.

$$(\text{Woman} = \textit{true}) \Rightarrow (\text{Man} = \textit{false})$$

Open Issues

- Evaluate the ability of the data-driven ontology population procedure to induce new knowledge w.r..t existing inductive classifiers
  - final goal: showing that hybrid sources of information actually help to induce a larger (and/or more accurate) amount of new knowledge.
- Concepts (and roles) inclusions are not taken into account $\Rightarrow$ saving of computational costs by explicitly treating this information
- Explicitly consider individuals that are role fillers
- Application of the discovery algorithm directly to a multi-relational representation

Exploiting Rules for Reasoning

- *Semantically enriched ARs* can be *exploited* when performing *deductive reasoning* on ontological KBs
- **Goals:**
  - reduce the computational effort for finding a model for a given (satisfiable) concept
  - suppling the most the plausible model (that best fits the available data)

Idea:

- set up an *heuristic exploiting the evidence coming from the data*
  - codified by the semantically enriched ARs
- *to be used when random choices occur*
  - e.g. when processing a concepts disjunction
  - *ideal solution* for saving computation (*case of satisfiable concept*) $\Rightarrow$ directly choose the ABox containing a model

Motivating Example

Example

Given an individual x, which is known to be a *Person*, a *high school student*, and has the property of being 15 years old.

Decide on whether x is instance of the concept *Parent* or not, while no information allows to infer neither x is a *Parent* nor x is $\neg$*Parent*.

Given the *semantically enriched association rule* (with high degree of confidence)

$$(\textsc{Age} = [0, 16]) \Rightarrow (\neg \mathit{Parent}) \quad 0.99$$

it can be exploited to conclude (with high confidence) that x is not a *Parent*.

Definition (Inference Problem)

Given: the following

- **D**, $\mathcal{K}$, the set R of semantically enriched ARs,
- a (possibly complex) concept E of $\mathcal{K}$,
- the individuals $x_1, \ldots, x_k \in \mathcal{K}$ that are instances of E,
- the grounding g of Σ on **D**

Determine: the model $\mathcal{I}_r$ for E representing the **most plausible model** given the $\mathcal{K}$, **D**, g and R.

Intuition:

- the most plausible model $\mathcal{I}_r$ for E is the one on top of the ranking of the possible models $\mathcal{I}_i$ for E
- Such a ranking is built according to the degree up to which the models are compliant with the set R of ARs and $\mathcal{K}$.

Data Driven Tableaux Algorithm: Differences with the Standard Tableaux algorithm

1. *the starting model* for the inference process is given by *the set of all attributes (and corresponding values) of the unified tabular representation* that are related to the individuals $x_1, \ldots, x_k$ that *are instances* of E,
2. a heuristic is adopted for performing the $\sqcup$-rule
3. the most plausible model for the concept E and the individuals $x_1, \ldots, x_k$ is built w.r.t. $\mathcal{K}$, $\mathbf{D}$ and R
4. The obtained model is a *mixed model*, namely a model containing both information from R and $\mathcal{K}$

Data Driven Tableaux Algorithm I

- For each individual $x_i \in \{x_1, \ldots, x_k\}$ that is instance of E, all attribute names A_i in the unified tabular representation $\mathbf{T}$ that related to x_i are selected jointly with the corresponding values a_i
- The assertions $A_i(a_i)$ are added to $\mathcal{I}_r$
    - For simplicity and without loss of generality, a single individual x will be considered
- Once the initial model $\mathcal{I}_r$ is built, all deterministic expansion rules, namely all but $\sqcup$-rule, are applied to $\mathcal{I}_r$ following the standard Tableaux algorithm.
- For the case of the $\sqcup$-rule, a heuristic is adopted.

Data Driven Tableaux Algorithm II

Let $C \sqcup D$ be the disjunctive concept to be processed by $\sqcup$-rule. The choice on C rather than D (or vice versa) is driven by:

1. *Select* the ARs in R containing C (resp. D) or its negation in the *knowledge items* of the right hand side
2. *Consider* the left hand side of each selected rule
3. *Compute* the degree of match between the left hand sides and the model under construction $\mathcal{I}_r$,
    - *Count* the number of (both data and semantic) items in the left hand side that are in $\mathcal{I}_r$
    - averaging this number w.r.t. the length of the left hand side of the rule
    - Items with uncertain (*unknown*) values are not considered
    - The degree of match for the rules whose (part of the) left hand side is contradictory w.r.t. the model $\mathcal{I}_r$ is set to 0
4. *Discard* rules with 0 degree of match

Data Driven Tableaux Algorithm III

5. *Compute* the weighted confidence value $weightedConf = ruleConfidence * degreeOfMatch$ for each of the remaining rules
6. *Discard* rules with $weightedConf$ below a given threshold
7. *Select* the rule with the highest $weightedConf$ (In case of multiple rules, a random choice is performed.
8. If the chosen rule contains $C = true$ (resp. $D = true$) in the right hand side $\Rightarrow$ *extend* the model under construction $\mathcal{I}_r$ with $C(x)$ (resp. $D(x)$) (x is the individual under consideration)
9. If the chosen rule contains $C = false$ (resp. $D = false$) in the right hand side $\Rightarrow$ *extend* the model under construction $\mathcal{I}_r$ with $D(x)$ (resp. $C(x)$).

Data Driven Tableaux Algorithm IV

10. If no rules are available for one of the two concepts, e.g. concept D, the concept for which some evidence, via existing rules, is available, i.e. C, will be chosen for expanding $\mathcal{I}_r$.
11. If no rule in R contains C (resp. D) or its negation in the right hand side ⇒ *Compute* the prior probability of C (resp. D) and perform the choice on its ground
    - computed by adopting a frequency-based approach e.g. $P(C) = |ext(C)|/|\mathcal{A}|$
    - The concept to be chosen for extending $\mathcal{I}_r$ is the one having the highest prior probability.

Data Driven Tableaux Algorithm: Example

Assume the enriched ARs discovered in a demographic domain

#	RULE	Confidence
r1	$(\text{AGE}=[16,25]) \wedge (\text{JOB} = \mathit{Student}) \Rightarrow (\mathit{Parent} = \mathit{false})$	0.98
r2	$(\text{JOB}=\mathit{Policeman}) \Rightarrow (\mathit{Male} = \mathit{true})$	0.75
r3	$(\text{AGE}=[16,25]) \wedge (\mathit{Parent} = \mathit{true}) \Rightarrow (\mathit{Female} = \mathit{true})$	0.75
r4	$(\text{JOB}=\mathit{Primary\ school\ teacher}) \Rightarrow (\mathit{Female} = \mathit{true})$	0.78
r5	$(\text{JOB}=\mathit{Housewife}) \wedge (\text{AGE} = [26,35]) \Rightarrow$ $(\mathit{Parent} = \mathit{true}) \wedge (\mathit{Female} = \mathit{true})$	0.85

and the model $\mathcal{I}_r$ under construction for the inference procedure

OBJECT	JOB	AGE	*Parent*	*Male*	*Female*
x_7	Primary school teacher	[46,55]	*unknown*	*unknown*	*unknown*
x_8	Policeman	[16,25]	*true*	*unknown*	*unknown*
x_9	Student	[16,25]	*unknown*	*unknown*	*unknown*

The reasoning process has to evaluate the expansion of $(\mathit{Male} \sqcup \mathit{Female})(x)$ w.r.t. $\mathcal{I}_r$ Application of the heuristic

Applying the Heuristic: Example I

- *Selection of the rules* having *Male* (resp. *Female*) in the right hand side $\Rightarrow$ r_2, r_3, r_4 and r_5.
- Computation of the degree of match
  - r_2: matchFound = 1 (because of JOB = Policemen (for x_8)) $\Rightarrow$ degreeOfMatch = 1 (note that lengthLeft = 1)
  - r_3: matchFound = 2 (because of $\text{AGE} = [16, 25]$ and $Parent = True$ (for x_8)) $\Rightarrow$ degreeOfMatch = 2 (note that lengthLeft = 2)
  - r_4: matchFound = 1 (because of JOB = PrimarySchoolTeacher (for x_7)) $\Rightarrow$ degreeOfMatch = 1 (note lengthLeft = 1)
  - r_5: matchFound = 0 (because no item matches the left hand side of r_5) $\Rightarrow$ degreeOfMatch = 0 (note lengthLeft = 2 since the left hand side of r_5 is made by two items)
- r_5 is *discarded* because of null degree of match

Applying the Heuristic: Example II

- For each of the remaining rules, *compute* the weighted confidence value
  - r_2: $weightedConf = ruleConfidence * degreeOfMatch = 0,75 * 1$
  - r_3: $weightedConf = 0.75 * 1 = 0.75$
  - r_4: $weightedConf = 0.78 * 1 = 0.78$
- *Filter out* rules with $weightedConf < thr$ (here 0.5) $\Rightarrow$ none of the above rules is discarded
- *Select* the rule with the highest $weightedConf \Rightarrow r_4$ is selected
- the right hand side of r_4 contains $Female \Rightarrow$ the model under construction $\mathcal{I}_r$ is enriched with $Female(x)$ (where x is the individual under consideration)
- this enriched model is considered for the application of the successive expansion rules, until the stopping criterion is met.

Open Issues

- Compare the Data-Driven and the Standard Tableaux Algorithms
  - number of performed ABox expansions
    - expected results: the heuristic decreases the ABox expansions when performing the consistency check of a consistent disjoint concept
  - execution time
    - since the data-driven Tableaux algorithm requires some additional computations (e.g. computing the degree of match)
- Formal proof that the model computed by the data-driven Tablaeux algorithm is the most plausible model w.r.t. a notion of plausibility
  - intuitively, since it is the most compliant one with the statistical regularities coming from the data it is also the most reliable model

Conclusions

Machine Learning methods

- could be usefully exploited for ontology mining
- suitable on large scale and in case of incoherent/noisy KBs
- can be also seen as an additional layer on top of deductive reasoning for realizing new/additional forms of approximated reasoning capabilities.

Future directions:

- Semi-Supervised Learning methods particularly appealing for LOD
- special focus on scalability issues

The End

That's all!

Questions ?




Nicola Fanizzi | Steffen Staab | Floriana Esposito | Luciano Serafini