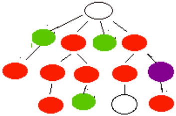
A Machine Learning Perspective for Ontology Mining

Claudia d'Amato

Department of Computer Science
University of Bari

ESSENCE 2017 ⋄ San Serolo, Venice - October 7, 2017

- **Ontologies ⇒ basic element for realizing the semantic interoperability**
  - on the Web and in other contexts



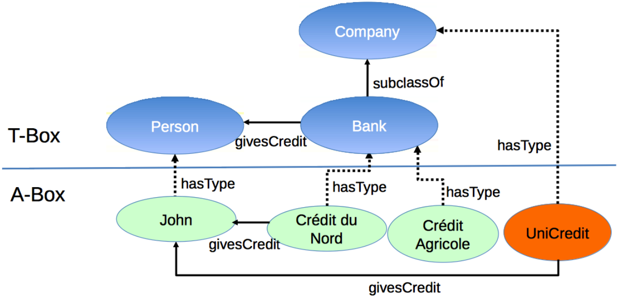
Examples of existing real ontologies

- Schema.org
- Gene Ontology
- Foundational Model of Anatomy ontology
- Financial Industry Business Ontology (by OMG Finance Domain Task Force)
- GoodRelations
- ...

Reasoning on Description Logics Ontologies

OWL adopted ⇒ **Description Logics** theoretical foundation

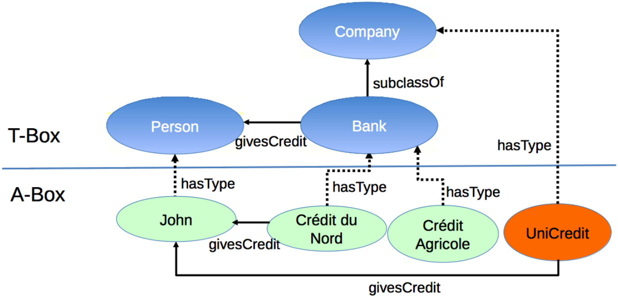
Ontologies are equipped with deductive reasoning capabilities ⇒ allowing to make explicit, knowledge that is implicit within them



Reasoning on Description Logics Ontologies

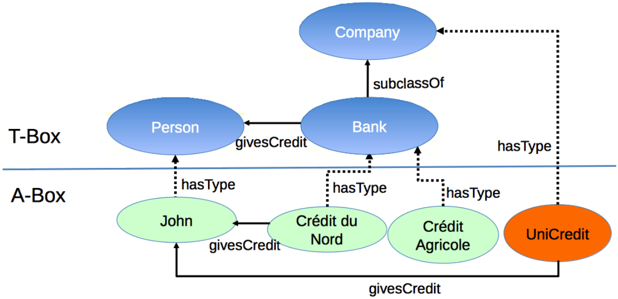
OWL adopted ⇒ **Description Logics** theoretical foundation

Ontologies are equipped with deductive reasoning capabilities ⇒ allowing to make explicit, knowledge that is implicit within them



Deduction:
"Crédit du Nord", "Crédit Agricole" are also `Company`

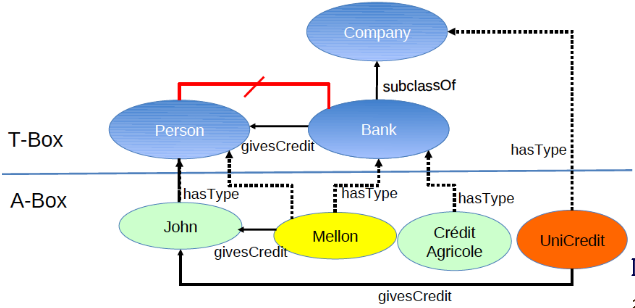
Reasoning on Description Logics Ontologies



Incompleteness

"*UniCredit*" is a `Bank`

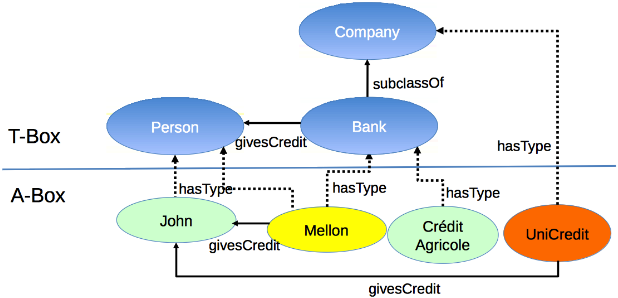
Reasoning on Description Logics Ontologies



Inconsistency

Mellon cannot be a Person **and** a Bank

Reasoning on Description Logics Ontologies

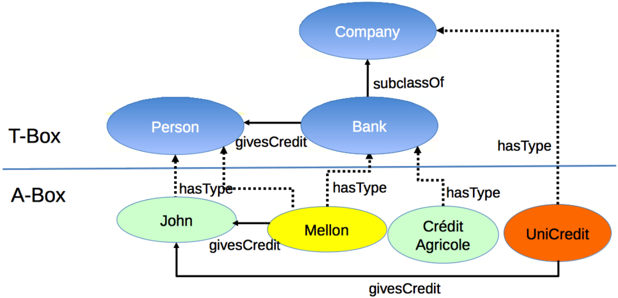


Noise

$\texttt{Person} \equiv \neg\texttt{Bank}$ missing

Reasoning on Description Logics Ontologies

Question: would it be possible to discover new/additional knowledge by exploiting *the evidence coming from the assertional data*?

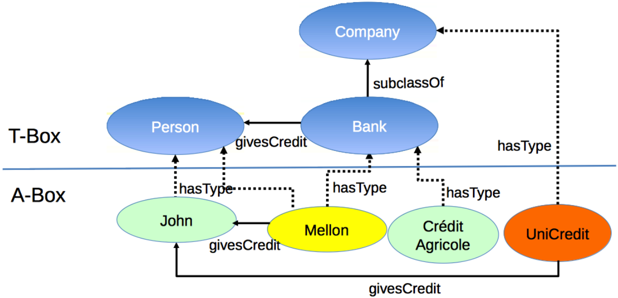


Noise

$\texttt{Person} \equiv \neg\texttt{Bank}$ missing

Reasoning on Description Logics Ontologies

Question: would it be possible to discover new/additional knowledge by exploiting *the evidence coming from the assertional data*?



Noise

$\texttt{Person} \equiv \neg\texttt{Bank}$ missing

Idea: exploiting **Machine Learning methods** for **Ontology Mining** related tasks *[d'Amato et al. @SWJ'10]*

Definition (Ontology Mining)

All activities that allow for

discovering hidden knowledge from
ontological knowledge bases

Definition (Ontology Mining)

All activities that allow for

discovering hidden knowledge from
ontological knowledge bases

Special Focus on:

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

Ontology Mining Tasks

- Instance Retrieval (Instance Level)
- Ontology Enrichment (Schema Level)
- Concept Drift and Novelty Detection (Ontology Dynamic)

from an inductive perspective

Focus on: similarity-based methods

Ontology Mining Tasks

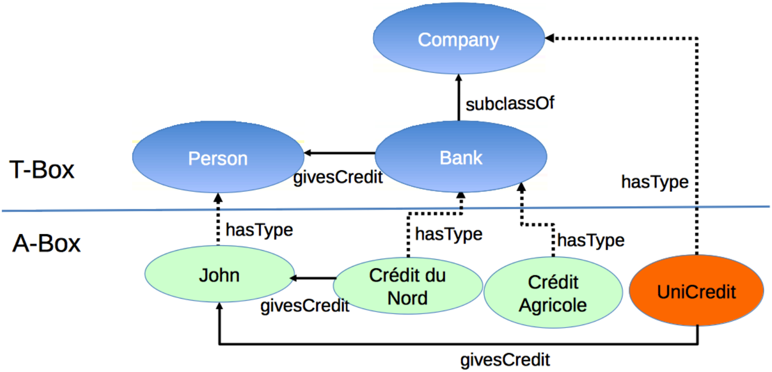
- **Instance Retrieval (Instance Level)**
- Ontology Enrichment (Schema Level)
- Concept Drift and Novelty Detection (Ontology Dynamic)

from an inductive perspective

Introducing Instance Retrieval I

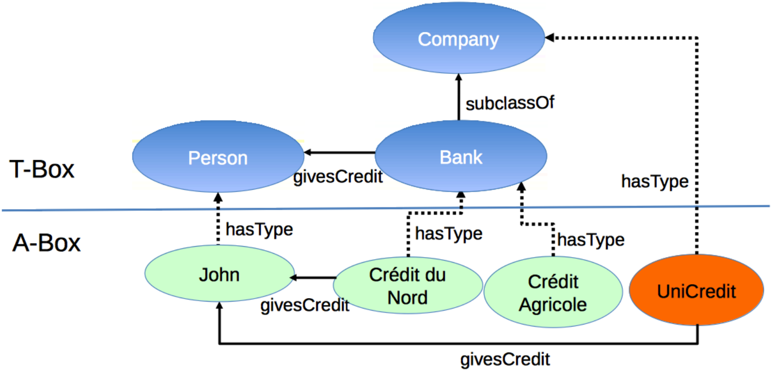
Instance Retrieval → Finding the extension of a query concept

- Instance Retrieval (Bank) = {"Crédit du Nord","Crédit Agricole"}



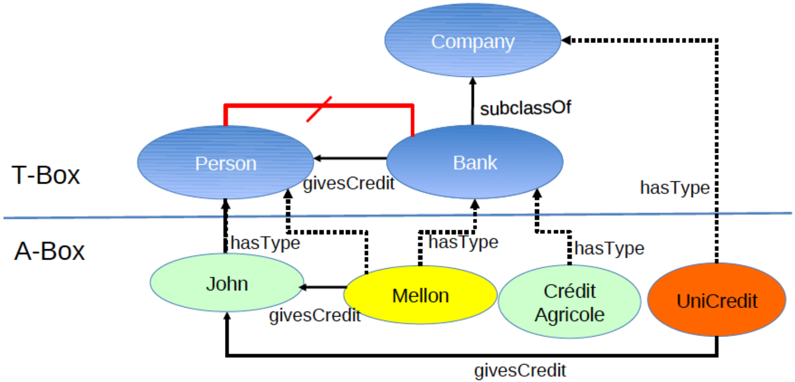
Introducing Instance Retrieval I

Problem: Instance Retrieval in incomplete/inconsistent/noisy ontologies



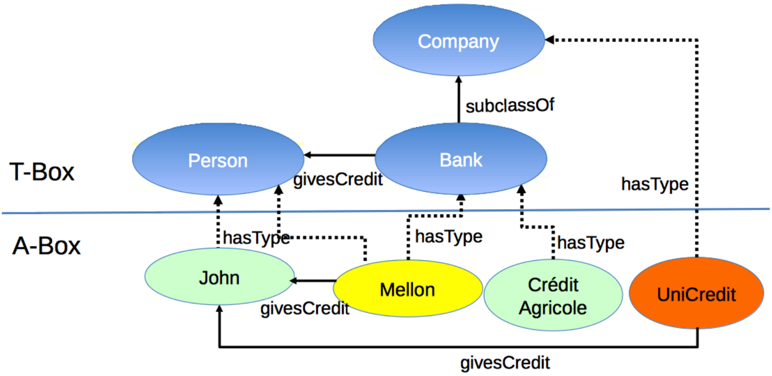
Introducing Instance Retrieval II

Problem: Instance Retrieval in incomplete/<u>inconsistent</u>/noisy ontologies



Introducing Instance Retrieval III

Problem: Instance Retrieval in incomplete/inconsistent/noisy ontologies



Issues & Solutions I

IDEA

Casting the problem as a Machine Learning **classification problem**

assess the class membership of individuals in a Description Logic (DL) KB w.r.t. the query concept

State of art classification methods cannot be straightforwardly applied

- generally applied to *feature vector* representation
  $\rightarrow$ *upgrade DL expressive representations*
- implicit *Closed World Assumption* made in ML
  $\rightarrow$ *cope with the Open World Assumption made in DLs*
- classes considered as *disjoint*
  $\rightarrow$ *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 deal with the semantics of the compared objects (concepts, individuals, ontologies)
  - to convey the underlying semantics of KB
- 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 $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$
- 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 $\mathcal{K}$ it belongs to
- the multi-class classification problem is *decomposed* into a set of *ternary classification problems* (one per target concept)

Developed methods

Pioneering the Problem

- relational K-NN for DL KBs *[d'Amato et al. ESWC'08]*

Improving the efficiency

- kernel functions for kernel methods to be applied to DLs KBs *[Fanizzi, d'Amato et al. @ ISMIS'06, JWS 2012; Bloehdorn and Sure @ ISWC'07]*

Scaling on large datasets

- Statistical Relational Learning methods for large scale and data sparseness *[Huang et al. @ ILP'10, Minervini et a. @ ICMLA'15]*

Lesson Learnt from experiments I

Experiments performed on ontologies publicly available

Results compared with a standard deductive reasoner

Need for new metrics $\rightarrow$ **Defined** *to distinguish induced assertions from mistakes*

		REASONER		
		+1	0	-1
INDUCTIVE CLASSIFIER	+1	*M*	*I*	*C*
	0	*O*	*M*	*O*
	-1	*C*	*I*	*M*

M **Match Rate**
O **Ommission Error Rate**
C **Commission Error Rate**
I **Induction Rate**

Lesson Learnt from experiments II

- *Commission error* almost zero on average
- *Omission error rate* very low and only in some cases
  - Not null for ontologies in which disjoint axioms are missing
- *Induction Rate* not zero
  - **new knowledge (not logically derivable) induced** $\Rightarrow$ can be used for *semi-automatizing the ontology population task*

	match	commission	omission	induction
SWM	97.5 ± 3.2	0.0 ± 0.0	2.2 ± 3.1	0.3 ± 1.2
LUBM	99.5 ± 0.7	0.0 ± 0.0	0.5 ± 0.7	0.0 ± 0.0
NTN	97.5 ± 1.9	0.6 ± 0.7	1.3 ± 1.4	0.6 ± 1.7
FINANCIAL	99.7 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	0.2 ± 0.2

Ontology Mining Tasks

- Instance Retrieval (Instance Level)
- **Ontology Enrichment (Schema Level)**
- Concept Drift and Novelty Detection (Ontology Dynamic)

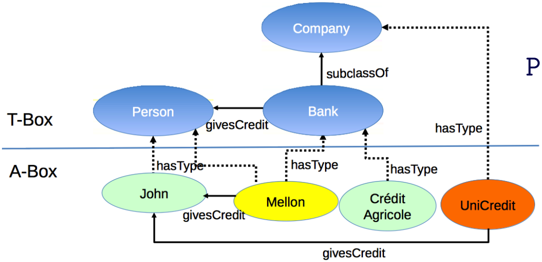
from an inductive perspective

Ontology Enrichment as a Disjointness Axioms Discovery Problem

Disjointness axioms often missing within ontologies

Problems:

- introduction of noise



Noise

$\texttt{Person} \equiv \neg\texttt{Bank}$ **missing**

- counterintuitive inferences

$\mathcal{K} = \{ JournalPaper \sqsubseteq Paper,\ ConferencePaper \sqsubseteq Paper,\ ConferencePaper(a)\ \}$

$\mathcal{K} \models JournalPaper(a)$?

Answer: Unknown

Observation: extensions of disjoint concepts do not overlap

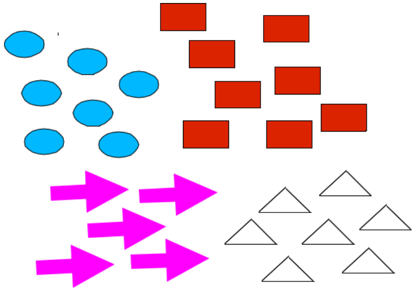
Question: would it be possible to *automatically capture* disjointness axioms by analyzing the data configuration/distribution?

Idea: Exploiting **(Conceptual) clustering methods** for the purpose

Basics on Clustering Methods

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

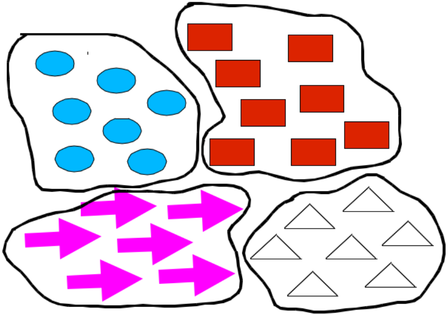
- intra-cluster *similarity* is high
- inter-cluster *similarity* is low



Basics on Clustering Methods

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

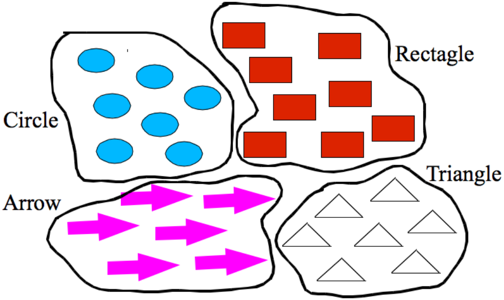
- intra-cluster *similarity* is high
- inter-cluster *similarity* is low



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



Observation: extensions of disjoint concepts do not overlap

Question: would it be possible to *automatically capture* them by analyzing the data configuration/distribution?

Idea: Exploiting **(Conceptual) clustering methods** for the purpose

Definition (Problem Definition)

Given

- an ontological knowledge base $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$
- a set of individuals $\mathbf{I} \subseteq \mathsf{Ind}(\mathcal{A})$

Find

- n pairwise disjoint clusters $\{\mathbf{C}_1, \ldots, \mathbf{C}_n\}$
- for each $i = 1, \ldots, n$, a concept description D_i that describes $\mathbf{C}_i$, such that:
    - $\forall a \in \mathbf{C}_i : \ \mathcal{K} \models D_i(a)$
    - $\forall b \in \mathbf{C}_j, j \neq i : \ \mathcal{K} \models \neg D_i(b)$.
- Hence $\forall D_i, D_j, i \neq j : \ \mathcal{K} \models D_j \sqsubseteq \neg D_i$.

Clustering Individuals of An Ontology: Developed Methods

Purely Logic-based

- KLUSTER *[Kietz & Morik, 94]*
- CSKA *[Fanizzi et al., 04]*
  - Produce a *flat output*
  - *Suffer from noise* in the data

Similarity-based ⇒ *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.'10]*
  - number of clusters required

Developed Methods for:

Supervised Concept Learning

- Separate-and-conquer approach
  - YinYang *[Iannone et al. @ Appl. Intell. J. 2007]*
  - DL-FOIL *[Fanizzi et al. @ ILP 2008]*
  - DL-Learner *[Lehmann et al. @ MLJ 2010, SWJ 2011]*
- Divide-and-conquer approach
  - TermiTIS *[Fanizzi et al. @ ECML 2010, Rizzo et al. @ ESWC 2015]*
- can be exploited for learning intentional cluster descriptions - do not tackle the problem of learning disjointness axioms

Learning Disjointness Axioms

- Statistical-based approach
  - NAR - exploiting negative association rules *[Fleischhacker et al. @ OTM'11]*
  - PCC - exploiting Pearson's correlation coeff. *[Völker at al.@JWS 2015]*
- do not exploit any background knowledge

Terminological Cluster Tree

Defined a method for eliciting disjointness axioms *[Rizzo et.al.@ESWC'17]*

- solving a clustering problem via learning Terminological Cluster Trees
- providing a concept description for each cluster

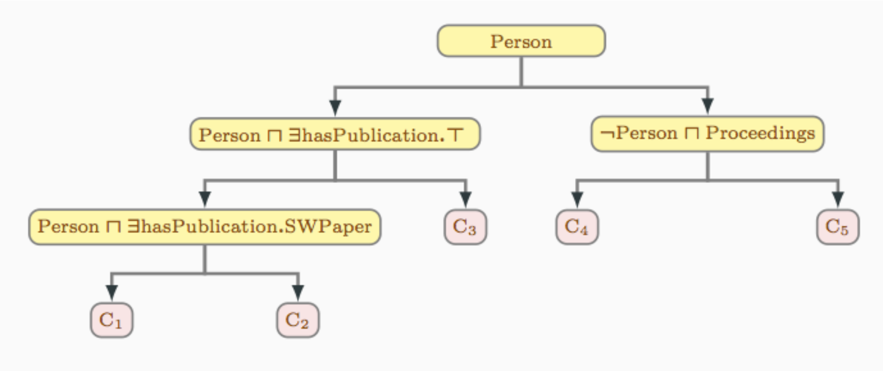
Definition (Terminological cluster tree (TCT))

A binary logical tree where

- a node stands for a cluster of individuals $\mathbf{C}$
- each inner node contains a description D (over the signature of $\mathcal{K}$)
- each departing edge corresponds to positive (left) and negative (right) examples of D

Example of TCT

Given $\mathbf{I} \subseteq \mathsf{Ind}(\mathcal{A})$, an example of TCT describing individuals in the Semantic Web research community



Collecting Disjointness Axioms

Given a TCT **T**:

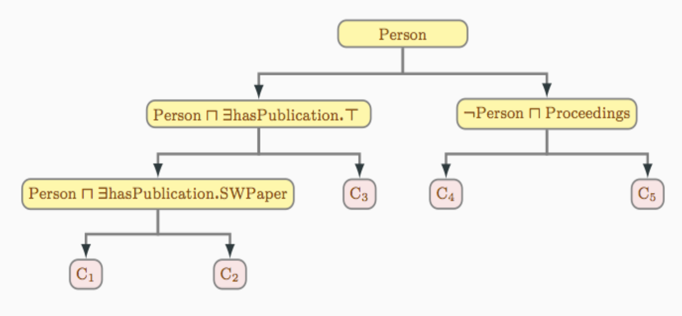
Step I:

- Traverse the **T** to collect the concept descriptions describing the clusters at the leaves
- A set of concepts **CS** is obtained

Step II:

- A set of candidate axioms **A** is generated from **CS**:
  - an axiom $D \sqsubseteq \neg E$ ($D, E \in \mathbf{CS}$) is generated if
    - $D \not\equiv E$ (or $D \not\sqsubseteq E$ or viceversa)
    - $E \sqsubseteq \neg D$ has not been generated

Collecting Disjointness Axioms: Example



CS = { Person, Person $\sqcap$ $\exists$hasPublication.$\top$, $\neg$(Person $\sqcap$ $\exists$hasPublication.$\top$), Person$\exists$hasPublication.SWPaper, $\neg$Proceedings, $\neg$Person $\sqcap$ Proceedings, $\cdots$ }

Axiom1: $Person \sqcap \exists hasPublication.SWPaper \sqsubseteq \neg(\neg Proceedings)$
Axiom2: $\cdots$

Inducing a TCT

Given the set of individuals **I** and $\top$ concept

Divide-and-conquere approach adopted

- **Base Case:** test the STOPCONDITION
  - the cohesion of the cluster **I** exceeds a threshold ν
    - distance between *medoids* below a threshold ν
- **Recursive Step** (STOPCONDITION does not hold):
  - a set **S** of refinements of the current (parent) description C generated
  - the BESTCONCEPT $E^* \in$ **S** is selected and installed as *current node*
    - the one showing the *best cluster separation* $\Leftrightarrow$ with max distance between the *medoids* of its positive P and negative N individuals
  - **I** is SPLIT in:
    - $I_{left} \subseteq$ **I** $\leftrightarrow$ individuals with the smallest distance wrt the *medoid* of P
    - $I_{right} \subseteq$ **I** $\leftrightarrow$ individuals with the smallest distance wrt the *medoid* of N

Note: *Number of clusters not required* - obtained from data distribution

Lesson Learnt from experiments I

Experiments performed on ontologies publicly available

- Goal I: Re-discover a target axiom (existing in $\mathcal{K}$)
  - Setting:
    - A copy of each ontology is created removing a target axiom
    - Threshold $\nu = 0.9, 0.8, 0.7$
    - Metrics # discovered axioms and #cases of inconsistency
  - Results:
    - target axioms rediscovered for almost all cases
    - *additional* disjointness *axioms discovered* in a significant number
    - limited number of inconsistencies found

Ontology	TCT 0.9		TCT 0.8		TCT 0.7	
	#inc.	#ax's	#inc.	#ax's	#inc.	#ax's
BioPax	2	53	2	53	3	52
NTN	10	70	9	73	10	75
Financial	0	125	0	126	0	127
GeoSkills	2	345	1	347	4	347
Monetary	0	432	0	432	0	433
DBPedia3.9	45	45	44	44	43	43

Lesson Learnt from experiments II

Goal II:

- Re-discover randomly selected target axioms added according to the **Strong Disjointness Assumption** *[Schlobach et al. @ ESWC 2005]*
  - two sibling concepts in a subsumption hierarchy considered as disjoint
- comparative analysis with statistical-based methods *[Völker at al. @ JWS 2015, Fleischhacker et al. @ OTM'11]*
  - PCC - based on *Pearson's correlation coefficient*
  - NAR - exploiting *negative association rules*
- Setting:
  - A copy of each ontology created removing 20%, 50%, 70% of the disjointness axioms
    - The copy used to induce TCT - $\nu = 0.9, 0.8, 0.7$ - # Run: 10 times
  - **Metrics**: rate of **rediscovered** target axioms, #cases of inconsistency, # addional discovered axioms

Lesson Learnt from experiments III

- Results:
  - *almost all axioms rediscovered*
    - Rate decreases when larger fractions of axioms removed, *as expected*
  - *TCT outperforms PCC and NAR* wrt *additionally discovered axioms* whilst introducing limited inconsistency
    - TCT allows to express complex disjointness axioms
    - PCC and NAR tackle only disjointness between concept names

Exploiting the $\mathcal{K}$ as well as the data distribution improves **disjointness axioms discovery**

Ontology Mining Tasks

- Instance Retrieval (Instance Level)
- Ontology Enrichment (Schema Level)
- **Concept Drift and Novelty Detection** (Ontology Dynamic)

from an inductive perspective

Concept Drift and Novelty Detection

- Ontologies evolve over the time ⇒ *New assertions* added.
- **Concept Drift**
  - change of a concept towards a more general/specific one w.r.t. the evidence provided by new annotated individuals
    - almost all `Worker` 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 `Worker` *employed* in a company ⇒ `Employee`
    - subset of `Worker` *working for* several companies ⇒ `Free-lance`

Concept Drift and Novelty Detection

- Ontologies evolve over the time $\Rightarrow$ *New assertions* added.
- **Concept Drift**
  - change of a concept towards a more general/specific one w.r.t. the evidence provided by new annotated individuals
- **Novelty Detection**
  - isolated cluster in the search space that requires to be defined through new emerging concepts to be added to the KB

Idea: *automatically capturing* them by analyzing the data configuration/distribution

Research Direction

Exploiting **(Conceptual) clustering methods** for the purpose

Lesson Learnt from Experiments

Developed Methods

- Evolutionary Clustering 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.'10]*
  - number of clusters required

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**

Conclusions

Machine Learning methods

- could be usefully exploited for ontology mining
- suitable in case of incoherent/noisy KBs
- **can be 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

That's all!

Thank you

Nicola Fanizzi

Giuseppe Rizzo

Floriana Esposito

Refinement Operators

Downward refinement operators specializing a concept C

ρ_1 $C' = C \sqcap (\neg)A$;

ρ_2 $C' = C \sqcap (\neg)(\exists)R.\top$;

ρ_3 $C' = C \sqcap (\neg)(\forall)R.\top$;

ρ_4 $\exists R.C'_i \in \rho(\exists R.C_i) \land C'_i \in \rho(C_i)$;

ρ_5 $\forall R.C'_i \in \rho(\forall R.C_i) \land C'_i \in \rho(C_i)$.

Distance measure between individuals

Distance Function (adapted from [d'Amato et al.@ESWC2008]):

$$d_n^{\mathcal{C}} : \mathsf{Ind}(\mathcal{A}) \times \mathsf{Ind}(\mathcal{A}) \rightarrow [0, 1]$$

$$d_n^{\mathcal{C}}(a, b) = \left[\sum_{i=1}^{m} w_i \left[1 - \pi_i(a)\pi_i(b)\right]^n \right]^{1/n}$$

Context: a set of atomic concepts $\mathcal{C} = \{B_1, B_2, \ldots, B_m\}$

Projection Function:

$$\forall\ a \in \mathsf{Ind}(\mathcal{A})(\mathcal{A}) \qquad \pi_i(a) = \begin{cases} 1 & \text{if } \mathcal{K} \models B_i(a) \\ 0 & \text{if } \mathcal{K} \models \neg B_i(a) \\ 0.5 & \text{otherwise} \end{cases}$$

2nd Experiment - Outcomes

Ontology	*f*	TCT 0.9		TCT 0.8		TCT 0.7		PCC		NAR	
		#inc.	#ax's	#inc.	#ax's	#inc.	#ax's	#inc.	#ax's	#inc.	#ax's
BioPax	20%	235	3859	357	4235	365	4256				
	50%	125	3576	357	4176	432	4115	257	280	352	2990
	70%	125	3432	235	3875	417	4154				
NTN	20%	312	3128	343	3126	354	3124				
	50%	234	3023	234	3034	235	3034	32	957	376	3766
	70%	156	2987	176	2679	123	2675				
Financial	20%	76	165	87	325	96	276				
	50%	37	143	56	307	53	259	124	1112	542	5366
	70%	33	143	43	276	40	221				
GeoSkills	20%	234	14289	357	14297	432	14345				
	50%	231	14123	356	14154	417	14256	456	13384	456	13299
	70%	234	14122	358	14154	377	14187				
Monetary	20%	535	13456	573	13453	623	13460				
	50%	315	13236	432	13236	532	13236	543	13384	423	13456
	70%	247	13127	231	13127	312	13127				
DBPedia3.9	20%	1345	29730	1432	30143	1432	30567				
	50%	1346	29730	1431	30143	1433	30567	1243	30470	1243	30365
	70%	1343	19730	1432	30143	1432	30567				

Experiment II - Outcomes

Ontology	*f*	TCT – *standard mode*			TCT – *early stopping*		
		TCT 0.9	TCT 0.8	TCT 0.7	TCT 0.9	TCT 0.8	TCT 0.7
BioPax	20%	0.90 ± 0.12	0.76 ± 0.13	0.74 ± 0.13	0.80 ± 0.23	0.65 ± 0.23	0.70 ± 0.13
	50%	0.85 ± 0.13	0.74 ± 0.13	0.74 ± 0.13	0.63 ± 0.23	0.63 ± 0.23	0.63 ± 0.23
	70%	0.85 ± 0.13	0.74 ± 0.12	0.74 ± 0.14	0.69 ± 0.13	0.67 ± 0.13	0.66 ± 0.14
NTN	20%	0.99 ± 0.08	0.95 ± 0.06	0.95 ± 0.08	0.70 ± 0.15	0.67 ± 0.15	0.67 ± 0.14
	50%	0.97 ± 0.03	0.93 ± 0.10	0.93 ± 0.01	0.55 ± 0.13	0.54 ± 0.13	0.54 ± 0.15
	70%	0.90 ± 0.10	0.89 ± 0.11	0.89 ± 0.10	0.55 ± 0.13	0.55 ± 0.13	0.55 ± 0.13
Financial	20%	0.99 ± 0.08	0.99 ± 0.08	0.99 ± 0.08	0.60 ± 0.10	0.59 ± 0.11	0.59 ± 0.11
	50%	0.97 ± 0.03	0.97 ± 0.03	0.97 ± 0.03	0.56 ± 0.10	0.56 ± 0.10	0.56 ± 0.10
	70%	0.95 ± 0.05	0.95 ± 0.05	0.95 ± 0.05	0.56 ± 0.10	0.56 ± 0.10	0.56 ± 0.10
GeoSkills	20%	0.99 ± 0.08	0.99 ± 0.08	0.99 ± 0.08	0.70 ± 0.15	0.69 ± 0.11	0.69 ± 0.11
	50%	0.92 ± 0.10	1.00 ± 0.00	1.00 ± 0.00	0.65 ± 0.23	0.65 ± 0.23	0.65 ± 0.23
	70%	0.92 ± 0.10	0.92 ± 0.10	0.92 ± 0.10	0.65 ± 0.23	0.63 ± 0.22	0.62 ± 0.23
Monetary	20%	0.99 ± 0.08	1.00 ± 0.00	1.00 ± 0.00	0.65 ± 0.23	0.63 ± 0.20	0.62 ± 0.23
	50%	0.94 ± 0.13	1.00 ± 0.00	1.00 ± 0.00	0.63 ± 0.12	0.66 ± 0.15	0.65 ± 0.11
	70%	0.94 ± 0.13	0.91 ± 0.14	0.91 ± 0.13	0.62 ± 0.12	0.60 ± 0.13	0.60 ± 0.12
DBPedia3.9	20%	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.70 ± 0.12	0.68 ± 0.13	0.67 ± 0.12
	50%	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.65 ± 0.23	0.68 ± 0.13	0.64 ± 0.12
	70%	0.96 ± 0.08	0.90 ± 0.08	0.90 ± 0.08	0.65 ± 0.22	0.68 ± 0.13	0.64 ± 0.12

Example of axioms

Successfully discovered axioms

- ExternalReferenceUtilityClass $\sqcap \exists$TAXONREF.$\top$
  disjoint with
  xref
- Activity
  disjoint with
  Person $\sqcap \exists$nationality.United_states
- Person $\sqcap$ hasSex.Male ($\equiv$ Man)
  disjoint with
  SupernaturalBeing $\sqcap$ God ($\equiv$ God)