On extracting Rules for: enriching ontological knowledge bases, complementing heterogeneous sources of information, empowering the reasoning process

Claudia d'Amato

Department of Computer Science
University of Bari, Italy

September 18, 2014 - Dagstuhl Schloss
Neural-Symbolic Learning and Reasoning Seminar

Contents

1. Focussing on Ontological Knowledge Bases
2. Focussing on Heterogenous Sources of Information

Starting Observations

Focussing on Ontological Knowledge Bases

- LOD represents a tremendous source of assertional knowledge whose intentional formalization is coded in OWL ontologies
- Ontological knowledge bases are often not *complete*
  - i.e. missing concept and role assertions, disjointness axioms, relationships that instead occur in the reference domain

Research Idea

Idea: exploiting the evidence coming from the data for discovering hidden knowledge patterns to be used for

1. (semi-)automatize the completion of the assertional knowledge
2. extends existing ontologies with formal rules (while ensuring and maintaining the decidability of the reasoning operators)
3. suggesting knew knowledge axioms

Research Direction: discovering hidden knowledge patterns in the form of relational association rules

Related Works in the Semantic Web Context

- Galárraga et al. @ WWW'13
  - discovering of association rules for predicting new role assertions from an RDF data source (no reaoning 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.
  - More complex relationships can be expressed.

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

- For a given pattern all possible specializations are generated by applying the operators:

  [**Add a concept atom** :] adds an atom with a concept name as a predicate symbol and an *already appearing* variable as argument

  [**Add a role atom** :] adds an atom with a role name as a predicate symbol; *at least one variable already appears* in the pattern
- The Operators are applied so that always *connected and non-redundant rules* are obtained
  - Heuristics are applied for not generating redundant rules
- Additional operators for tanking into account constants could be similarly considered

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. $C(x) \wedge R(x, y) \wedge S(y, y)$

Pruning Conditions

- ($\mathcal{K} \cup$ extractedRule) *is unsatisfiable*
- headCoverage(extractedRule) $<$ *headCov-thr*
- confidence(extractedRule) - confidence(parentExtractedRule) $<$ *improvedConf-thr*
- extractedRule $\in$ setOfAlreadyExtractedRules
  - approximated check on rules with the same Support (extension)
- extractedRule $\in$ setOfInfrequentPattern
  - approximated check on Support (extension) of extractedRule contained in the Support (extension) of some infrequent pattern

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 to be Solved

- Experimental evaluation for accessing the effectiveness of the method: how to set up it?
- Develop a scalable algorithms for the purpose
- Improvement of the expressive power of the discovered rules
  - Manage role hierarchy, complex descriptions

Focus for Scalability

- 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

Research Idea

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

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

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

- *learning 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

Simple Motivating Example...

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

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

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

- 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**.
- Learning *Semantically Enriched Association Rules* for the purpose

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

Main Idea (inspired by the multi-relational setting)

1. the entity, i.e. the attribute/set of attributes, of primary interest for extracting association rules is determined
2. a view containing the attributes of interest w.r.t. the primary entity is built.

Building an Integrated Data Source: Formal Method

Choose:

1. the primary entity of interest in **D** or $\mathcal{K}$; set it as the first attribute A_1 (and primary key) in the table **T** to be built;
2. a (subset of) the attributes in **D** of interest for A_1; set them as additional attributes in **T**;
   - the corresponding values can be obtained as a result of a SQL query
3. a (subset of) concept names $\{C_1, \ldots, C_m\}$ in $\mathcal{K}$ of interest for A_1; set their names as additional attribute names in **T**
   - $\forall\ C_k \in \{C_1, \ldots, C_m\}$ and $\forall a_i$ in A_1, if $\mathcal{K} \models C_k(a_i)$ then $C_k := \mathit{true}$ in **T**, else if $\mathcal{K} \models \neg C_k(a_i)$ then $C_k := \mathit{false}$ in **T**, $C_k := \mathit{unknown}$ otherwise
4. a (subset of) role names $\{R_1, \ldots, R_t\}$ in $\mathcal{K}$ of interest for A_1; set their names as additional attribute names in **T**
   - $\forall\ R_l \in \{R_1, \ldots, R_t\}$ and $\forall a_i$ of A_1, if $\exists y \in \mathcal{K}\ s.t.\ \mathcal{K} \models R_l(a_i, y)$ then $R_l := \mathit{true}$ in **T**, else if $\forall y \in \mathcal{K}\ :\ \mathcal{K} \models \neg R_l(a_i, y)$ then $R_l := \mathit{false}$, $R_l = \mathit{unknown}$ otherwise
5. a (subset of) the datatype property names $\{T_1, \ldots, T_v\}$ in $\mathcal{K}$ of interest for A_1; set their names as additional attribute names in **T**
   - $\forall\ T_j \in \{T_1, \ldots, T_v\}$ and $\forall a_i$ of A_1, if $\mathcal{K} \models T_j(a_i, \mathit{dataValue}_j)$ then $T_j := \mathit{dataValue}_j$ in T_j in **T**, $T_j := \mathit{false}$ otherwise.

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

$$(\text{AGE}{=}[25, 34]) \wedge (\text{CITY} = \text{Madrid}) \Rightarrow (\text{HasChild} = \textit{true})$$

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

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

$$\text{SALARY}{=}[15000, 24999] \Rightarrow (\text{HasChild} = \textit{true}) \wedge (\text{Child} = \textit{true})$$

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

- Concepts (and roles) are managed without considering inclusion relationships among them $\Rightarrow$ saving of computational costs by explicitly treating this information
- Treatment of the individuals that are fillers of the considered roles explicitly
- Application of the (APRIORI) learning algorithm to a multi-relational representation (without building an intermediate propositional representation)

Exploiting Rules for Reasoning

- *Semantically enriched ARs* can be *exploited* when performing *reasoning* on ontological KBs (Tableaux generally used)
- **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

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)

$$IF\ \textsc{Age} = [0, 16]\ THEN\ \neg Parent \quad 0.99$$

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

Definition (Inference Problem)

Given: the following

- $\mathbf{D}$, $\mathcal{K}$, 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 $\mathbf{D}$

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

Basics on Tableaux Algorithm $\mathcal{ALC}$

To test the satisfiability of a concept E, the algorithm

- starts with ABox $\mathcal{A} = \{E(x_0)\}$ (with x_0 a new individual)
- applies to $\mathcal{A}$ the consistency preserving transformation rules (one for each operator) until no more rules apply
- The result could be:
  - all clashes $\Rightarrow$ the concept is unsatisfiable
  - $\mathcal{A}$ contains a model for E $\Rightarrow$ the concept is satisfiable
- Transformation rule for the disjunction ($\sqcup$-rule)
  - if a disjunctive concept has to be processed, two different ABoxes have to be considered, each one containing the assertion concerning one of the two disjoint concepts.
  - The original ABox is consistent if and only if one of the new ABoxes is consistent
  - the choice on one of the two to be processed *is non-deterministic*
- *ideal solution* for saving computation (*case of satisfiable concept*) $\Rightarrow$ directly choose the ABox containing a model

Data Driven Tableaux Algorithm: Differences

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

- 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.

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

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 $\textsc{Job} =$ Policemen (for x_8)) $\Rightarrow$ degreeOfMatch $= 1$ (note that lengthLeft $= 1$)
  - r_3: matchFound $= 2$ (because of $\textsc{Age} = [16, 25]$ and $Parent = True$ (for x_8)) $\Rightarrow$ degreeOfMatch $= 2$ (note that lengthLeft $= 2$)
  - r_4: matchFound $= 1$ (because of $\textsc{Job} =$ 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.

Issues to be Solved

- assessing the validity of the proposed data-driven Tablaeux algorithm when compared to the standard Tableaux algorithm
- assessing the validity of the data-driven ontology population procedure that is grounded on a rule-based classifier

Comparing the Tableaux Algorithms

- Compare/count the number of ABox expansions that are performed by the data-driven Tableaux algorithm when compared to the standard Tableaux algorithm
  - since the adopted heuristic aims to reduce the number of ABox expansions, the experimental evaluation should empirically show that the heuristic is able to significantly decreases the number of ABox expansion when the consistency of a (possibly complex) disjoint concept, that is already known to be consistent, is performed.
- Compare the execution time
  - since the data-driven Tableaux algorithm requires some additional computations (for instance of computing the degree of match) with respect to the standard Tableaux algorithm

Evaluating the Data Driven Ontology Population Procedure

- Evaluate the ability of the classifier to classify instances with respect to one or more concepts
  - the performances are compared with those of a standard deductive reasoner
  - mistakes, correct predicted class-memberships and the new induced class-memberships are measured.
- Evaluate the ability of the data-driven ontology population procedure to induce new knowledge, when compared with other inductive classifiers at the state of the art
  - final goal: to show that exploiting a hybrid source of information actually help to induce a larger (and/or more accurate) amount of new knowledge.

The End

That's all!

Questions ?