Constructing Metrics for Evaluating Multi-Relational Association Rules in the Semantic Web from Metrics for Scoring Association Rules

Minh Tran, Claudia d'Amato, Andrea G. B. Tettamanzi, Binh Nguyen

▶ To cite this version:

HAL Id: hal-02084977

https://hal.inria.fr/hal-02084977

Submitted on 30 Mar 2019

Constructing Metrics for Evaluating Multi-Relational Association Rules in the Semantic Web from Metrics for Scoring Association Rules

Tran Duc Minh
Thang Long University, Vietnam
tdminh2110@yahoo.com

Claudia d'Amato
University of Bari, Italy
claudia.damato@uniba.it

Andrea G. B. Tettamanzi
University of Côte d'Azur, CNRS, Inria, I3S, France
andrea.tettamanzi@unice.fr

Nguyen Thanh Binh
The University of Danang - University of Science and Technology, Vietnam
ntbinh@dut.udn.vn

Abstract—We propose a method to construct asymmetric metrics for evaluating the quality of multi-relational association rules coded in the form of SWRL rules. These metrics are derived from metrics for scoring association rules. We use each constructed metric as a fitness function for evolutionary inductive programming employed to discover hidden knowledge patterns (represented in SWRL) from assertional data of ontological knowledge bases. This new knowledge can be integrated easily within the ontology to enrich it. In addition, we also carry out a search for the best metric to score candidate multi-relational association rules in the evolutionary approach by experiment. We performed experiments on three publicly available ontologies validating the performances of our approach and comparing them with the main state-of-the-art systems.

Keywords—Evolutionary Algorithms; Description Logics; Pattern Discovery

I. Introduction

The Semantic Web [1] (SW) has undergone a steady and continuous development toward its primary objective, which is to associate meaning with the data and to exploit the data through intelligent processing techniques. We refer to assertional and intensional definitions provided by ontologies as an *ontological knowledge base*. Currently, we can use data mining techniques to exploit and extract hidden knowledge patterns from ontological knowledge bases (KBs), even in incomplete and noisy ontologies. These knowledge patterns are used to enrich an ontology at both terminological (TBox) and assertional (ABox) level. Level-wise generate-and-test methods related to these data mining techniques were proposed in the inductive logic programming (ILP) [2], [3] and a recent approach that exploits the resoning capabilities of the SW, like [4]. In addition, EDMAR [5], an evolutionary inductive programming approach, was recently proposed to discover hidden knowledge patterns in the form of multi-relational association rules (ARs) coded in SWRL [6].

In this paper, we construct a method to change metrics for scoring ARs into metrics for evaluating SWRL rules, with the aim to increase the number of metrics to assess the SWRL rules. New metrics will be used as fitness function in EDMAR and applied to three publicly available ontological KBs.

Basics and definitions are provided in Section II and III. Section IV describes how to convert metrics and defines some typical metrics. The EDMAR algorithm is shown in Section V. Section VI reports on experiments to find out the best metric. Section VII concludes.

II. Basics

We refer to ontological KBs in Description Logics (DLs) [7], which are a family of formal knowledge representation languages and are the theoretical foundation of OWL[1] (the standard representation language in the Semantic Web). A DL KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ consisting of a TBox $\mathcal{T}$ containing the terminological axioms and an ABox $\mathcal{A}$ containing the assertional axioms. It should be recalled that DLs adopt the *open-world assumption* (OWA - a statement may be true irrespective of whether or not it is known to be true). For more details concerning DLs see [7]. In the following, we address the general definition of relational AR with respect to a given ontological KB.

Definition 1 (Relational Association Rule). Given a populated ontological KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, a *relational association rule* r for $\mathcal{K}$ is a Horn-like clause of the form: $body \rightarrow head$, where (a) $body$ is a generalization of a set of assertions in $\mathcal{K}$ co-occurring together; (b) $head$ is a consequent that is induced from $\mathcal{K}$ and $body$.

Definition 2 (Problem Definition). Given $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, a minimum "frequency threshold", θ_f, and a minimum "fitness threshold", θ_{fit}, we desire to discover as many frequent and fit hidden patterns (w.r.t θ_f and θ_{fit}) as possible, in the form of relational ARs, that may induce new assertions for $\mathcal{K}$.

Intuitively, a frequent hidden pattern is a generalization of a set of concept/role assertions co-occurring reasonably often (w.r.t. a fixed frequency threshold) together, showing an underlying form of correlation that is exploited for obtaining new assertions.

---

[1]https://www.w3.org/OWL/.

The rules to be discovered (following Def. 2), will be represented in the Semantic Web Rule Language (SWRL) [6], which extends the set of OWL axioms with Horn-like rules.[2]

Definition 3 (SWRL Rule). Given a KB $\mathcal{K}$, a SWRL rule is an implication between an antecedent (body) and a consequent (head) of the form: $B_1 \wedge B_2 \wedge \ldots B_n \rightarrow H_1 \wedge \cdots \wedge H_m$, where $B_1 \wedge \cdots \wedge B_n$ is the rule body and $H_1 \wedge \cdots \wedge H_m$ is the rule head. Each $B_1, \ldots, B_n, H_1, \ldots H_m$ is called an *atom.*

An *atom* is a unary or binary predicate of the form $C(s)$, $R(s_1, s_2)$, where the predicate symbol C is a concept name in $\mathcal{K}$, R is a role name in $\mathcal{K}$, s, s_1, s_2 are *terms.* A *term* is either a variable (denoted by x, y, z) or a constant (denoted by a, b, c) standing for an individual name or data value.

The discovered rules can be generally called *multi-relational* rules since multiple binary predicates $R(s_1, s_2)$ with different role names of $\mathcal{K}$ could appear in a rule. The intended meaning of a rule is: whenever the conditions in the antecedent hold, the conditions in the consequent must also hold. A rule having more than one atom in the head can be equivalently transformed, due to the *safety condition* (see Def. 4), into multiple rules, each one having the same body and a single atom in the head. Therefore, we will consider, w.l.o.g., only SWRL rules (hereafter just "rules") with one atom in the head.

III. LANGUAGE BIAS

From Def. 3, we realize that we might generate a multitude of SWRL rules with respect to a given ontological KB. In order to avoid generating trivial rules (e.g., $\mathsf{Male}(x) \rightarrow \mathsf{Man}(x)$), redundant rules (e.g., $\mathsf{fatherOf}(y, x) \rightarrow \mathsf{Male}(y)$) or meaningless rules (e.g., $\mathsf{fatherOf}(y, x) \rightarrow \mathsf{parentOf}(u, v)$), we fix the language bias.

A *language bias* is a set of constraints giving a tight specification of the patterns worth considering, thus allowing to reduce the search space. Following [5], we are interested in rules having only atomic concepts and/or role names of $\mathcal{K}$ as predicate symbols, and individual names as constants. Only *connected* [8] and *non-redundant* [9] rules satisfying the *safety condition* [6] are considered.

Given an atom A, let $T(A)$ denote the set of all the terms occurring in A and let $V(A) \subseteq T(A)$ denote the set of all the variables occurring in A e.g. $V(C(x)) = \{x\}$ and $V(R(x, y)) = \{x, y\}$. Such notation may be extended to rules straightforwardly.

Definition 4 (Safety Condition). Given a KB $\mathcal{K}$ and a rule $r = B_1 \wedge B_2 \wedge \ldots B_n \rightarrow H$, r satisfies the *safety condition* if all variables appearing in the rule head also appear in the rule body; formally if: $V(H) \subseteq \bigcup_{i=1}^n V(B_i)$,

Definition 5 (Connected Rule). Given a KB $\mathcal{K}$ and a rule $r = B_1 \wedge B_2 \wedge \ldots B_n \rightarrow H$, r is *connected* if and only if every atom in r is transitively connected to every other atom in r.

Two atoms B_i and B_j in r, with $i \neq j$, are *connected* if they share at least a variable or a constant i.e. if $T(B_i) \cap T(B_j) \neq \emptyset$. Two atoms B_1 and B_k in r are

transitively connected if there exist in r, atoms $B_2, \ldots, B_{k-1}$, with $k \leq n$, such that, for all $i, j \in \{1, \ldots, k\}$ with $i \neq j$, $T(B_i) \cap T(B_j) \neq \emptyset$.

Definition 6 (Closed Rule). Given a KB $\mathcal{K}$ and a rule $r = B_1 \wedge B_2 \wedge \ldots B_n \rightarrow H$, r is *closed* if and only if every variable in r is closed.

Each variable $v_j \in \bigcup_{i=1}^n V(B_i), j \in \{1, \ldots, k\}$, with $k \leq n$, is closed if it appears at least twice in r.

Definition 7 (Redundant Rule). Given a KB $\mathcal{K}$ and a rule $r = B_1 \wedge B_2 \wedge \ldots B_n \rightarrow H$, r is a redundant rule if at least one atom in r is entailed by another atom in r with respect to $\mathcal{K}$, i.e., if, $\exists i \in \{0, 1, \ldots, n\}$, $\exists j \in \{0, 1, \ldots, n\}$, with $B_0 = H$, results: $B_j \models_{\mathcal{K}} B_i, i \neq j$

IV. METRICS FOR RULE EVALUATION

We notice that the generated rules satisfying the language bias in Section III are definitely not trivial, redundant and meaningless. However, the language bias might not guarantee the quality of generated rules. Therefore, we use metrics that are a data-driven approach for assessing the quality of rules. For each metric, a threshold is used to filter low-quality rules.

So far, only a few metrics introduced in [8] and [5] are used to evaluate the quality of SWRL rules. In this paper we propose a solution that is using metrics employed for evaluating the quality of ARs to assess the quality of SWRL rules. With this solution, we might increase the number of metrics to assess the SWRL rules and in addition also support other studies related to the evaluation of SWRL rules.

A. Computing Metrics for Association Rules

All metrics used to evaluate ARs are computed based on the frequency counts tabulated in Table I.

	Y	$\overline{Y}$	
X	f_{11}	f_{10}	f_{1+}
$\overline{X}$	f_{01}	f_{00}	f_{0+}
	f_{+1}	f_{+0}	N

Table I includes two itemsets X, Y, where X is called antecedent and Y is called consequent; $\overline{X}$ (resp. $\overline{Y}$) stands for an itemset that does not contain items in the itemset X (Y) for a transaction; f_{11} is the number of times transactions contain items in both X and Y; f_{1+} counts the number of transactions that contain items in X; f_{+1} counts the number of transactions that contain items in Y; N is the total number of transactions. From the four values of f_{11}, f_{1+}, f_{+1} and N, we can calculate the remaining values in the contingency table. For example, f_{10} is the number of transactions that contain items in X but not in Y and $f_{10} = f_{1+} - f_{11}$.

B. Computing Metrics for SWRL Rules

In order to build metrics for assessing SWRL rules derived from metrics for evaluating ARs, we also need to build a contingency table based on frequency counts. However, instead

---

[2]The result is a KB with an enriched expressive power. More complex relationships than subsumption can be expressed.

TABLE II.   A CONTINGENCY TABLE FOR THE SWRL RULE $B \rightarrow H$

	H	$\overline{H}$	
B	$E_H(r)$	f_{10}	$M_H(r)$
$\overline{B}$	f_{01}	f_{00}	f_{0+}
	$\Sigma_H(r)$	f_{+0}	N

of counting the number of transactions, we count the number of bindings of variables to individuals in the KB.

The entries in Table II are used to aid in the construction of the metrics for assessing SWRL rules. As in Table I, we only need to obtain the four values $E_H(r)$, $M_H(r)$, $\Sigma_H(r)$ and N, the remaining values being easy to compute based on these values.

Definition 8 (Binding). A binding is a mapping from a set of variables to a set of named individuals that occur in at least one assertion in the ABox.

Given a rule $r = B_1 \wedge \ldots \wedge B_n \rightarrow H$, let us denote:

- $\Sigma_H(r)$ the set of distinct bindings of the variables occurring in the head of r, formally: $\Sigma_H(r) = \{binding\ V(H)\}$

- $E_H(r)$ the set of distinct bindings of the variables occurring in the head of r provided the body and the head of r are satisfied, formally:
  $E_H(r) = \{$ *binding f of V(H) | there is a binding f' of $V(B_1 \wedge \cdots \wedge B_n)$ that extends f, such that $f'(B_1 \wedge \cdots \wedge B_n \wedge H)$ is satisfied*$\}$.
  Since the rules satisfy the safety condition, $V(H) \subseteq V(B_1 \wedge \cdots \wedge B_n)$

- $M_H(r)$ the set of distinct bindings of the variables occurring in the head of r also appearing as binding for the variables occurring in the body of r (since the rules satisfy the safety condition), formally:
  $M_H(r) = \{$ *binding f of V(H) | there is a binding f' of $V(B_1 \wedge \cdots \wedge B_n)$ that extends f, such that $f'(B_1 \wedge \cdots \wedge B_n)$ is satisfied*$\}$.

- Σ_i total number of individuals inside a KB.

- The rule head of r might be either a concept atom or a role atom. Therefore, N is a number defined as follows:
  $$N = \begin{cases} \Sigma_i, & \text{if } H \text{ is a concept atom;} \\ P^2_{\Sigma_i} = \Sigma_i(\Sigma_i - 1), & \text{if } H \text{ is a role atom;} \end{cases}$$
  This, implicitly, assumes a role cannot be reflexive.

There are two kinds of metrics: symmetric and asymmetric metrics. In our work, we focus only on introducing asymmetric metrics that are suitable for assessing rules coded in SWRL, because SWRL rules have the general form $B \rightarrow H$ and values of an asymmetric metric for $B \rightarrow H$ and $H \rightarrow B$ may not be the same. We do not use symmetric metrics because these are only appropriate for evaluating itemsets of the form $B \wedge H$, which is equivalent to $H \wedge B$, without any distinction between antecedent and consequent.

Following [8], we redefine a few basic definitions of asymmetric metrics through the entries of Table II for use in the experiments (Sec. VI). These definitions are modified from

the classical ones (as given e.g. in [10]) to ensure they fit the case of SWRL rules.

Definition 9 (Rule Support). Given a rule $r = B_1 \wedge \ldots \wedge B_n \rightarrow H$, its support is the number of distinct bindings of the variables in the head, provided the body and the head of r are satisfied jointly, formally:

$$\text{supp}(r) = |E_H(r)|. \tag{1}$$

Definition 10 (Head Coverage for a Rule). Given the rule $r = B_1 \wedge \ldots \wedge B_n \rightarrow H$, its head coverage is the ration between the rule support and the distinct variable bindings from the head of the rule

$$\text{headCoverage}(r) = \frac{|E_H(r)|}{|\Sigma_H(r)|}. \tag{2}$$

Definition 11 (Rule Confidence). Given a rule $r = B_1 \wedge \ldots \wedge B_n \rightarrow H$, its confidence is defined as the ratio of the number of the rule support and the number of bindings in the rule body:

$$\text{conf}(r) = \frac{|E_H(r)|}{|M_H(r)|}. \tag{3}$$

Definition 12 (Rule PCA-Confidence). Given the rule $r = B_1 \wedge \ldots \wedge B_n \rightarrow H$, its PCA (Partial Completeness Assumption) confidence is defined as follows:

$$\text{pcaconf}(r) = \begin{cases} \frac{|E_H(r)|}{|M_H(r)|}, & \text{if } H \text{ is a concept atom;} \\ \frac{|E_H(r)|}{|P_H(r)|}, & \text{if } H \text{ is a role atom,} \end{cases} \tag{4}$$

where $P_H(r)$ is the set of distinct bindings of the variables occurring in the head of r provided that the body and the head of r are satisfied. In particular, this applies when a role atom is in the head of the considered rule. Formally: $P_H(r) = \{binding\ V(H) \mid \exists\ binding\ V(B_1 \wedge \cdots \wedge B_n) \cup v_{rng}(H') : B_1 \wedge \cdots \wedge B_n \wedge H'\}$, where

- H and H' are role atoms with the same predicate symbol R;

- $V(H) \subseteq V(B_1 \wedge \cdots \wedge B_n)$ since rules are connected and closed

- $v_{dom}(H) = v_{dom}(H')$ and $v_{rng}(H) \neq v_{rng}(H')$; with v_{dom} and v_{rng} standing for the domain and range variables respectively of the predicate symbol R

- $v_{rng}(H') \notin V(B_1 \wedge \cdots \wedge B_n)$.

Definition 13 (Certainty Factor for a Rule). Given a rule $r = B_1 \wedge \ldots \wedge B_n \rightarrow H$, its Certainty Factor [11] represents uncertainty in the rule and is defined as follows:

$$\text{cf}(r) = \begin{cases} \frac{\text{conf}(r) - \frac{|\Sigma_H(r)|}{N}}{1 - \frac{|\Sigma_H(r)|}{N}}, & \text{if } \text{conf}(r) > \frac{|\Sigma_H(r)|}{N} ; \\ \frac{\text{conf}(r) - \frac{|\Sigma_H(r)|}{N}}{\frac{|\Sigma_H(r)|}{N}}, & \text{if } \text{conf}(r) < \frac{|\Sigma_H(r)|}{N} ; \\ 0, & \text{if } \text{conf}(r) = \frac{|\Sigma_H(r)|}{N} . \end{cases} \tag{5}$$

Definition 14 (Added Value for a Rule). Given a rule $r = B_1 \wedge \ldots \wedge B_n \rightarrow H$, Added Value [12] for the rule r is defined as:

$$\text{av}(r) = \text{conf}(r) - \frac{|\Sigma_H(r)|}{N} \tag{6}$$

This metric is more meaningful when the amount of evidence is large, for it relies on probabilities.

C. Metrics for Evaluating Rule Precision

Definition 15 (Rule Precision). Given the rule $r = B_1 \wedge \ldots \wedge B_n \rightarrow H$, its *precision* is the ratio of the number of correct predictions made by r and the total number of correct and incorrect predictions (predictions logically contradicting $\mathcal{K}$), leaving out the predictions with unknown truth value.

This metric expresses the ability of a rule to perform correct predictions, but it is not able to take into account the induced knowledge, that is the *unknown* predictions. For this reason, the metrics proposed in [13] are also considered (for the evaluation in Sect. VI):

- *match rate*: number of predicted assertions in agreement with facts in the complete ontology, out of all predictions;

- *commission error rate*: number of predicted assertions contradicting facts in the full ontology, out of all predictions;

- *induction rate*: number of predicted assertions that are not known in the complete ontology, out of all predictions.

V. THE EDMAR ALGORITHM

Here, we use the EDMAR (Evolutionary Discovery of Multi-relational Association Rules) algorithm represented in [5]. All the proposed metrics in this paper are in turn used as fitness function in EDMAR algorithm.

VI. EXPERIMENTS AND RESULTS

We carry out a comparison of the best asymmetric metrics [14] with a combination of them. The comparative criteria are based on the number of discovered rules, the number of unknown facts and the number of correct predictions. This comparison is created with the desire to figure out a new metric that is better than the metrics in [14]. The idea of this solution is that the metric combined by the best metrics might obtain discovered rules which satisfies at least one of the best metrics but may not satisfy the remaining metrics.

The EDMAR algorithm is applied to the same publicly available ontologies used in [15]: Financial,[3] describing the banking domain; Biological Pathways Exchange (BioPAX),[4] describing biological pathway data; and New Testament Names Ontology (NTNMerged),[5] describing named entities in the New Testament. Details in Table V.

In order to assess the ability of the discovered rules to predict new assertional knowledge for a considered ontological KB, for each ontology in Table V, we construct three stratified samples by randomly removing respectively 20%, 30%, 40% of the concept assertions and eliminating role assertions or assertional axioms that involve the removed concept assertions, while the full ontology has been used as a testbed.

We carried out 30 runs of the EDMAR algorithm (cf. Sect. V) for each stratified sample and for each choice of fitness function using the following parameter setting:

---

[3]http://www.cs.put.poznan.pl/alawrynowicz/financial.owl.
[4]http://www.biopax.org/release/biopax-level2.owl.
[5]http://www.semanticbible.com/ntn/ntn-view.html.

$$
\begin{array}{rclrcl}
n & = & 5,000; & p_{\text{mut}} & = & 5\%; \\
\text{MAX\_GENERATIONS} & = & 200; & \theta_{\text{mut}} & = & 0.2; \\
\text{MAX\_RULE\_LENGTH} & = & 10; & \tau & = & \frac{1}{5} \\
\theta_{\text{fit}} & = & 0 & \theta_f & = & 1.
\end{array}
$$

Our experiments aimed at comparing the results obtained by the EDMAR algorithm using the five best asymmetric metrics evaluated by [14] (*HeadCoverage*, *Confidence*, *PCA-Confidence*, *Certainty Factor*, *Added Value*) and the metric $Combination_{metrics}(r)$ which is a combination of these five metrics ($Combination_{metrics}(r) = HeadCoverage(r) + Confidence(r) + PCA\text{-}Confidence(r) + Certainty\ Factor(r) + Added\ Value(r)$). These six metrics are used as fitness function in EDMAR algorithm and the comparison is based on the three following criteria:

1) The *number of the rules discovered* by the EA.
2) The *induction rate*: if it is positive, this means assertions are predicted that could not be inferred from the stratified version. The higher the induction rate, the more novel predictions (unknown facts) are induced for the KB.
3) The *number of correct predictions = number of predictions × precision*, where the *number of predictions* is the number of predicted assertions and *precision* is defined in Def. 15.

In order to compare the metrics according to these criteria, we applied the rules discovered from the stratified samples to the full ontology versions and collected all predictions, that is the head atoms of the instantiated rules. All predictions already contained in the stratified ontology versions have been discarded while the remaining predicted facts have been considered. A prediction is regarded as *correct* if it is contained/entailed by the full ontology version and as *incorrect* if it is inconsistent with the full ontology version. Results in Table IV have been averaged over 30 different runs with the above parameter setting and have been measured in terms of *precision* (Def. 15), *match rate*, *commission error rate*, and *induction rate* (Sect. IV-C).

Table III shows a comparison of the metrics according to the first criterion. The second and third criteria are used to compare the predictive power of the discovered rules and are shown in Table IV. From the observations in Tables III and IV, we might draw a few remarks:

- The EDMAR algorithm confirms its ability to discover accurate rules (*precision = match rate + commission rate + induction rate = 1*) and no contradicting knowledge is predicted (*commission error rate = 0*) on all versions of all considered ontologies.

- The combined metric (CM) allows EDMAR to discover the largest number of rules (Tab. III) along with the largest number of correct predictions (Tab. IV) in NTNMerged that is the ontology containing the largest number of individuals (see Table V). In addition, this metric comes up with rules that induce a large number of previously unknown facts (induction rate > 0).

We may see that the fitness function $Combination_{metrics}(r)$ of EDMAR outperforms each

TABLE III.   COMPARISON OF THE METRICS BY THE NUMBER OF
DISCOVERED RULES.
HEADCOVERAGE (H), CONFIDENCE (C), PCA-CONFIDENCE (P),
CERTAINTY FACTOR(CF), ADDED VALUE(AV) AND COMBINATION OF
FIVE METRICS (CM)

Ont.	S.	Total number of discovered rules by metric ± stdev					
		H	C	P	CF	A	CM
Financial	20%	26 ± 4	25 ± 4	25 ± 3	25 ± 3	26 ± 3	25 ± 4
	30%	25 ± 3	25 ± 4	25 ± 4	26 ± 3	24 ± 4	26 ± 5
	40%	23 ± 3	23 ± 3	22 ± 4	23 ± 4	21 ± 3	22 ± 4
Biopax	20%	129 ± 13	122 ± 12	130 ± 10	118 ± 9	119 ± 9	119 ± 8
	30%	128 ± 9	130 ± 13	130 ± 9	117 ± 8	110 ± 9	117 ± 8
	40%	129 ± 11	136 ± 11	133 ± 8	124 ± 9	122 ± 7	126 ± 7
NTNMerged	20%	1,157 ± 168	1,345 ± 423	1,418 ± 492	671 ± 36	656 ± 34	2,118 ± 1,082
	30%	1,052 ± 353	947 ± 238	1,017 ± 370	743 ± 45	728 ± 48	1,719 ± 840
	40%	1,088 ± 181	1,223 ± 177	1,295 ± 357	687 ± 38	664 ± 34	1,527 ± 497

of the five best metrics and is the best choice as an optimization criterion (i.e., fitness function) for EDMAR.

In addition, we also notice that the rate of removing concept assertions, does not affect the number of discovered rules from ontological knowledge bases (Tab. III).

We also compared the experimental performance of the EDMAR algorithm using $Combination_{metrics}(r)$ with the two level-wise generate-and-test algorithms RARD [4] and AMIE [8], which are closest to it in purpose. Table VI reports the number of rules discovered by each system given each knowledge base sample. We did not compare the second and third criteria because the maximum length of a rule in EDMAR is different from RARD and AMIE. We may remark the following:

- The metric $Combination_{metrics}(r)$ discovers more rules than RARD from NTNMerged, but fewer from the Financial and Biopax KBs. One reason is that EDMAR does not generate *open* rules as RARD; in addtion, the maximum length of a rule in EDMAR is 10 atoms while in RARD is 3 atoms. Another reason is that the number of individuals in Financial and Biopax is less than that of NTNMerged.

- The metric $Combination_{metrics}(r)$ is superior to AMIE because it discovers more rules than AMIE from all three KBs, although the rules discovered by AMIE can be inconsistent with KBs and the maximum length of a rule in AMIE is 3 atoms.

VII.   CONCLUSIONS

We have shown a method to convert metrics for scoring ARs into metrics for assessing SWRL rules. In addition, we figured out the best metric to score candidate multi-relational association rules in the evolutionary approach by comparison experiment.

Future work might focus on developing other metrics suited for scenarios based on the OWA.

REFERENCES

[1] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, 2001.

[2] P. Reiser and P. Riddle, "Scaling up inductive logic programming: An evolutionary wrapper approach," *Applied Intelligence*, vol. 15, no. 3, pp. 181–197, 2001.

[3] F. Divina, *Hybrid Genetic Relational Search for Inductive Learning*. PhD thesis, Vrije Universiteit Amsterdam, 2004.

[4] C. d'Amato, S. Staab, A. Tettamanzi, D. M. Tran, and F. Gandon, "Ontology enrichment by discovering multi-relational association rules from ontological knowledge bases," in *Proc. of SAC 2016*, ACM, 2016.

[5] D. M. Tran, C. d'Amato, B. T. Nguyen, and A. G. B. Tettamanzi, "An evolutionary algorithm for discovering multi-relational association rules in the semantic web," in *GECCO*, pp. 513–520, ACM, 2017.

[6] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean, "SWRL: A semantic web rule language combining OWL and RuleML," 2004.

[7] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, eds., *The Description Logic Handbook: Theory, Implementation, and Applications*, Cambridge Univ. Press, 2003.

[8] L. Galárraga, C. Teflioudi, K. Hose, and F. Suchanek, "AMIE: Association rule mining under incomplete evidence in ontological knowledge bases," in *WWW '13*, pp. 413–422, ACM, 2013.

[9] J. Józefowska, A. Lawrynowicz, and T. Lukaszewski, "The role of semantics in mining frequent patterns from knowledge bases in description logics with rules," *Theory and Practice of Logic Programming*, vol. 10, no. 3, pp. 251–289, 2010.

[10] R. Agrawal, T. Imielinski, and A. N. Swami, "Mining association rules between sets of items in large databases," in *Proc. of the Int. Conf. on Management of Data*, pp. 207–216, ACM Press, 1993.

[11] L. M. Fu and E. H. Shortliffe, "The application of certainty factors to neural computing for rule discovery," in *IEEE TRANS. On Neural Networks*, pp. 647–657, 2000.

[12] S. Sahar and Y. Mansour, "An empirical evaluation of objective interestingness criteria," in *SPIE Conference on Data mining and Knowledge Discovery*, pp. 63–74, 1999.

[13] N. Fanizzi, C. d'Amato, and F. Esposito, "Learning with kernels in description logics," in *ILP 2008* (F. Zelezný and N. Lavrac, eds.), pp. 210–225, Springer, 2008.

[14] M. Duc Tran, C. d'Amato, B. T. Nguyen, and A. G. B. Tettamanzi, "Comparing rule evaluation metrics for the evolutionary discovery of multi-relational association rules in the semantic web," in *Genetic Programming*, (Cham), pp. 289–305, Springer International Publishing, 2018.

[15] C. d'Amato, A. G. B. Tettamanzi, and T. D. Minh, "Evolutionary discovery of multi-relational association rules from ontological knowledge bases," in *Knowledge Engineering and Knowledge Management*, (Cham), pp. 113–128, Springer International Publishing, 2016.

TABLE IV. AVG (± ST.DEV.) PERFORMANCE ON EACH ONTOLOGY OF HEADCOVERAGE (H), CONFIDENCE (C), PCA-CONFIDENCE (P), CERTAINTY FACTOR(CF), ADDED VALUE(AV) AND COMBINATION OF FIVE METRICS (CM)

precision = match rate + commission rate + induction rate

	Ont.	Samp.	Match Rate	Com. Rate	Ind. Rate	Total # Predictions
H	Financial	20%	0.855 ± 0.033	0	0.145 ± 0.033	47,232 ± 36,777
		30%	0.864 ± 0.044	0	0.136 ± 0.044	25,456 ± 34,174
		40%	0.861 ± 0.044	0	0.139 ± 0.044	23,207 ± 30,133
	BioPAX	20%	0.567 ± 0.031	0	0.433 ± 0.031	84,035 ± 15,018
		30%	0.591 ± 0.03	0	0.409 ± 0.03	85,499 ± 11,660
		40%	0.58 ± 0.027	0	0.42 ± 0.027	90,856 ± 14,048
	NTNMerged	20%	0.572 ± 0.026	0	0.428 ± 0.026	2,311,624 ± 287,858
		30%	0.564 ± 0.039	0	0.436 ± 0.039	2,314,346 ± 458,522
		40%	0.621 ± 0.027	0	0.379 ± 0.027	2,345,588 ± 357,565
C	Financial	20%	0.848 ± 0.045	0	0.152 ± 0.045	43,151 ± 44,254
		30%	0.860 ± 0.038	0	0.140 ± 0.038	27,589 ± 41,184
		40%	0.858 ± 0.051	0	0.142 ± 0.051	33,795 ± 41,880
	BioPAX	20%	0.574 ± 0.036	0	0.426 ± 0.036	79,454 ± 14,019
		30%	0.584 ± 0.027	0	0.416 ± 0.027	88,879 ± 12,890
		40%	0.582 ± 0.023	0	0.418 ± 0.023	96,884 ± 13,782
	NTNMerged	20%	0.618 ± 0.042	0	0.382 ± 0.042	1,437,868 ± 253,206
		30%	0.581 ± 0.036	0	0.419 ± 0.036	1,164,306 ± 167,173
		40%	0.670 ± 0.030	0	0.330 ± 0.030	1,557,516 ± 280,666
P	Financial	20%	0.859 ± 0.055	0	0.141 ± 0.055	41,350 ± 46,196
		30%	0.850 ± 0.055	0	0.150 ± 0.055	32,812 ± 41,501
		40%	0.859 ± 0.043	0	0.141 ± 0.043	29,762 ± 35,582
	BioPAX	20%	0.571 ± 0.028	0	0.429 ± 0.028	89,486 ± 11,303
		30%	0.584 ± 0.023	0	0.416 ± 0.023	92,392 ± 13,878
		40%	0.587 ± 0.027	0	0.413 ± 0.027	91,849 ± 11,960
	NTNMerged	20%	0.609 ± 0.046	0	0.391 ± 0.046	2,130,947 ± 380,546
		30%	0.588 ± 0.043	0	0.412 ± 0.043	1,409,235 ± 286,439
		40%	0.670 ± 0.042	0	0.330 ± 0.042	1,727,343 ± 262,891

	Ont.	Samp.	Match Rate	Com. Rate	Ind. Rate	Total # Predictions
CF	Financial	20%	0.877 ± 0.038	0	0.123 ± 0.038	31,656 ± 45,045
		30%	0.852 ± 0.057	0	0.148 ± 0.057	48,568 ± 45,051
		40%	0.857 ± 0.039	0	0.143 ± 0.039	31,068 ± 36,044
	BioPAX	20%	0.556 ± 0.026	0	0.444 ± 0.026	80,361 ± 8,700
		30%	0.581 ± 0.023	0	0.419 ± 0.023	78,933 ± 11,147
		40%	0.564 ± 0.035	0	0.436 ± 0.035	84,476 ± 12,647
	NTNMerged	20%	0.565 ± 0.01	0	0.435 ± 0.01	1,039,112 ± 179,322
		30%	0.535 ± 0.014	0	0.465 ± 0.014	1,424,334 ± 180,205
		40%	0.557 ± 0.018	0	0.443 ± 0.018	2,110,928 ± 423,539
A	Financial	20%	0.859 ± 0.041	0	0.141 ± 0.041	33,358 ± 31,445
		30%	0.858 ± 0.041	0	0.142 ± 0.041	29,866 ± 31,123
		40%	0.859 ± 0.041	0	0.141 ± 0.041	29,870 ± 44,276
	BioPAX	20%	0.549 ± 0.032	0	0.451 ± 0.032	83,666 ± 11,663
		30%	0.578 ± 0.029	0	0.422 ± 0.029	78,059 ± 9,368
		40%	0.579 ± 0.02	0	0.421 ± 0.02	84,483 ± 10,376
	NTNMerged	20%	0.563 ± 0.012	0	0.437 ± 0.012	966,840 ± 204,430
		30%	0.541 ± 0.014	0	0.459 ± 0.014	1,324,518 ± 282,410
		40%	0.566 ± 0.014	0	0.434 ± 0.014	1,632,633 ± 218,033
CM	Financial	20%	0.861 ± 0.046	0	0.139 ± 0.046	37,766 ± 44,986
		30%	0.876 ± 0.043	0	0.124 ± 0.043	28,245 ± 33,875
		40%	0.855 ± 0.053	0	0.145 ± 0.053	38,415 ± 42,710
	BioPAX	20%	0.555 ± 0.024	0	0.445 ± 0.024	82,398 ± 11,717
		30%	0.576 ± 0.028	0	0.424 ± 0.028	83,524 ± 13,986
		40%	0.576 ± 0.026	0	0.424 ± 0.026	90,496 ± 12,094
	NTNMerged	20%	0.664 ± 0.06	0	0.336 ± 0.06	3,254,323 ± 751,324
		30%	0.628 ± 0.061	0	0.372 ± 0.061	3,043,882 ± 627,598
		40%	0.672 ± 0.055	0	0.328 ± 0.055	3,404,974 ± 491,617

TABLE V. KEY FACTS ABOUT THE ONTOLOGICAL KBS USED.

Ont.	# Con.	# Rol.	# Indiv.	# Declared Assertions	# Decl.+Derived Assertions
Finan	59	16	1000	3359	3814
Bio	40	33	323	904	1671
NTM	47	27	695	4161	6863

TABLE VI. COMPARISON OF THE NUMBER OF DISCOVERED RULES.

Ontology	Samp.	# The total number of rules discovered		
		CM	RARD	AMIE
Financial	20%	25 ± 4	177	2
	30%	26 ± 5	181	2
	40%	22 ± 4	180	2
Biopax	20%	119 ± 8	298	8
	30%	117 ± 8	283	8
	40%	126 ± 7	272	0
NTNMerged	20%	2,118 ± 1,082	243	1,129
	30%	1,719 ± 840	225	1,022
	40%	1,527 ± 497	239	1,063