ISSN 1870-4069

Parameter Estimation for an ontology Evaluation Metric

Mireya Tovar Vidal, Gerardo Flores Petlacalco, Hugo Raziel Lasserre Chávez, Aimee Cecilia Hernández García, Emmanuel Santos Rodríguez, Irvin Yair Cabrera Moreno, Juan Carlos Flores Molina, Jesús Flores Castillo, Christian Martínez Cuamani

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación, Puebla,
Mexico

mtovar@cs.buap.mx, {gfloresp93, hugohp}@outlook.com, {zvok59, e.ss.rdz, yair.cb, fmjuancarlos, j.jikamaru}@gmail.com, cmc_2012@hotmail.com

Abstract. Ontologies are semantic resources that capture the knowledge of a particular domain, by means of the elements that comprise it. Despite the manner the ontologies are created (manually, automatically or semiautomatically), there still the fact that the process is not perfect and, therefore, an additional step for ontology quality validation is needed. The ontology evaluation task aims to measure the quality of these resources. Among other approaches, the corpus-based evaluation attempts to validate the ontology components by means of an external data source which usually is a collection of documents associated to the same domain of the ontology to be evaluated. In this paper, a metric that involves results of different methods for identifying semantic relations and concepts in the domain corpus is presented and for evaluating the quality of an ontology. So, an overall evaluation score is provided and a parameter estimation for the metric proposed is applied. The experimental results show a satisfactory performance, which it is considered interesting for the task of measuring the quality of ontologies of restricted domain.

Keywords. Ontology, ontology evaluation, metric, natural language processing.

1 Introduction

Daily, the information on the current web grows and much of this information is generated without a structure that can be understood by both machines and humans, which makes it a dificult task to process. The semantic web, proposed by Tim Berners-Lee [1], searches to give structure and knowledge to the conventional web. Ontologies are used to represent knowledge in a structured way on the semantic web.

An ontology is defined as "an explicit and formal specification of a shared conceptualization" [4].

This type of semantic resource, that allows to capture the explicit knowledge in data, is formed by concepts or classes, relations, instances, attributes, axioms, restrictions, rules, functions and events. Domain ontologies are a system of representation of the knowledge that it is possible to organize in taxonomic and non-taxonomic structures of concepts of some area or domain of specific knowledge. If the ontology has been designed manually or automatically, it is necessary to evaluate its quality.

In the literature the evaluation is classified depending on the form of evaluation used, which is: compare the ontology with a gold standard, apply the ontology in an application and evaluate the results, make comparisons against source data of the domain ontology, and finally evaluations made by humans to determine what criteria the ontology satisfies [2]. Gómez-Pérez [3] presents two terms for the ontology evaluation: verification and validation. The verification ensures that the definitions meet the requirements correctly. The validation ensures that the meaning of the definitions correctly model the phenomena of the world. In this work, a metric is applied to evaluate an ontology in the artificial intelligence domain and a parameter estimation for the ontology evaluation metric is applied. The aim of the experiment is to decide if the result of the automatic evaluation system is within the confidence interval of an estimator.

In this work, a metric is applied to evaluate an ontology in the artificial intelligence domain and a parameter estimation for the ontology evaluation metric is applied. The aim of the experiment is to decide if the result of the automatic evaluation system is within the confidence interval of a statistical estimator.

The purpose of statistics is to use the information contained in a sample to make inferences about the population from which the sample is taken [7]. Populations are characterized by numerical descriptive measures called parameters. The main objective of this investigation is to perform statistics to calculate the value of one parameter or more relevant parameters. Among the most important estimators are the mean, variance and standard deviation.

An estimator is expressed through a formula to calculate the value of an estimation based on the measurements contained in a sample. Two types of estimates will be used, the point estimate and the interval estimate. The point estimate includes the parameter estimate in a single value or point.

Interval estimators are also called confidence intervals. The upper and lower endpoints of the confidence interval are named as the upper and lower confidence limits. The probability that a confidence interval will include a fixed amount is called the confidence coefficient.

This document is divided into four sections; in the section 2 the proposed metric is described; the section 3 contains information about the experimental results; and finally, the conclusion and future work are explained in the section 4.

2 Proposed Metric

This work is a complement to the work done in [5, 6], in this a parameters estimation of the designed metric is presented.

First the methodology is presented after the parameters used are presented. The methodology considers the following phases:

1. **Automatic Pre-Processing of Information.** The concepts and ontological relationships are extracted from the ontology using Jena. After, the documents or sentences of the corpus of domain are associated with the concepts and relationships through an information retrieval system. In addition, natural language processing tools are used to pre-processing the text such as: elimination of punctuation symbols, parts of speech and others.
2. **Automatic Discovery of Candidate Terms and/or Ontological Relations.** In this phase the approaches used for the discovery of concepts and ontological relations in the corpus of domain are submitted, some approaches are: lexical-syntactic pattern, formal concept analysis, similarity, dependency analysis and latent semantic analysis. The purpose of this phase is to find evidence of the relation and concepts in the corpus.
3. **Evaluation of the Ontology.** In this phase, a metrics to evaluate the domain ontology is proposed.

In the third phase of the methodology, the quality of the ontology is determined usign the metric of Equation 1. The metric receives the O ontology as a parameter and the output is the result of the evaluation. The metric (M) is formed by the product of three matrices: $Matrix_C$, $Matrix_E$ and $Matrix_I$:

$$
\begin{aligned}
&M(O) = Matrix_C \; Matrix_E \; Matrix_I, \\
&Matrix_C = \begin{bmatrix} A(E_1, CI) \;...\; A(E_n, CI) \\ A(E_1, NT) \;...\; A(E_n, NT) \end{bmatrix}, \\
&A(E_i, R) = \frac{\sum_{i=1}^{|R|} Reliability(T_i)}{|R|}, \\
&Reliability(T_i) = \begin{cases} 1 \; If \; \alpha * qual(C_{i,1}) + \\ \quad \beta * qual(C_{i,2}) + \\ \quad \gamma * qual(R_i) > 0.75 \\ 0 \text{ otherwise}, \end{cases} \\
&qual(R) = \frac{|E_i(R)|}{|R|}, \\
&Matriz_E = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ ... & ... \\ a_n & b_n \end{bmatrix}, \\
&Matrix_I = \begin{bmatrix} d_1 \; d_2 \end{bmatrix}.
\end{aligned}
\tag{1}
$$

$Matrix_C$ contains the results of the measure of accuracy (A) of each approach E_i that was applied in phase 2 of the methodology to the domain corpus and to the semantic relationships of the ontology.

In equation 1, $T_i = (C_{i,1}, R_i, C_{i,2})$ is a triplet of ontology; $C_{i,1}$ and $C_{i,2}$ are concepts; and R_i is the ontological relation.

In the case of the quality of the semantic relation ($qual(R)$) we consider the measure of accuracy that, considering the total of relations proposed by the approach ($E_i(R)$) and the total ontology relation ($n = |R|$), where R are the class-inclusion (CI) or non-taxonomic relations (NT). Finally, the matrices $Matrix_E$ and $Matrix_I$ are matrices of external and internal coefficients respectively, which normalize the values between 0 and 1.

As a second part of the evaluation metric, parameter estimation is included. We take the metric value for a point estimate and a confidence interval estimation, for one of the parameters, e.g., mean.

In the case of the confidence interval, we consider the confidence coefficient of $0.95 = 1 - \alpha$. In the Equation 2 and in the Equation 3 the mean and variance are defined as unbiased estimators:

$$\overline{Y} = \mu = \frac{\sum_{i=1}^{n} Y_i}{n-1}, \tag{2}$$

$$S^2 = \sigma^2 = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n-1}. \tag{3}$$

The confidence interval observed for μ is obtained through equation 4:

$$\overline{Y} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}, \tag{4}$$

where $t_{\alpha/2}$ is determined by $n - 1$ degree of freedom. The t distribution has a density function very similar to the normal density.

The standard deviation of θ is the square root of its variance. It is a measure of the amount of variation of a set of values.

3 Results

This section presents the experimental results of the implementation of the evaluation metric. To validate the phases of the methodology, an evaluation is carried out by domain experts and a baseline based on mutual information that measures the degree of co-relation of the semantic relations. Section 3.1 presents the amount of information evaluated by the domain experts and by the automatic system. Considering that the validation is done manually and this requires a high person-hour cost, the experts only evaluated a subset of the sentences from the corpus.

3.1 Description of the Dataset

The domain of ontology considered in the experiments is artificial intelligence (AI)[1] [8]. The ontology contains a number of concepts (C), class-inclusion (S)

[1] Ontologies and their corpus are available on the page: http://www.site.uottawa.ca/ azouaq/goldstandards.htm

relations, and non-taxonomic (R) relations (see Table 1). In Table includes information over the total of documents (D), the amount of sentences (O), total tokens or words (T) of these sentences, the vocabulary (V) of the sentences, the number of filtered sentences (Of) by the information retrieval system; the sentences reviewed by the experts to validate class-inclusion relations (OSE) and non-taxonomic relations (ORE).

Table 1. Dataset.

Domain	Ontology			Corpus						
	C	S	R	D	O	T	V	O_f	OSE	ORE
IA	276	205	61	8	475	11,370	1,510	415	312	110

3.2 Experimental Results for Class-Inclusion Relations

In Table 2, the results obtained from the approach using the accuracy criteria (A); the quality (C) in the prediction of the approach, according to three human experts (H_1, H_2 and H_3); and the baseline is presented. Table 2 also includes the results of the estimators: mean (μ), variance σ^2 and standard deviation σ. The standard deviation of the accuracy of the results of the approaches is presented in Fig. 1 and the Fig. 2 the standard deviation of the $Avg(H_i)$ is presented, too. As can be seen, the approach that is outside the limits of the standard deviation in both figures is FCA sfd_0, which suggests that it could be omitted as a parameter in the calculation of the ontology evaluation metric.

Table 2. Accuracy of the Ontology AI and the Quality of Predictions of Approaches for Class-Inclusion Relations.

Approach	A	$C(H_1)$	$C(H_2)$	$C(H_3)$	$Avg(H_i)$
1 LSP	88.78	89.76	84.39	88.29	87.48
2 Sim-cos	90.24	83.41	80.98	87.80	84.07
3 Sim-cos_u	98.05	90.24	86.83	95.61	90.89
4 FCA min	95.61	89.76	85.37	94.15	89.76
5 FCA sfd_0	100.00	92.20	88.78	97.56	92.85
6 LSA-cos	94.15	90.24	89.76	92.68	90.89
Baseline	56.00	57.00	51.00	55.00	54.00
μ	94.47	89.27	86.02	92.68	89.32
σ^2	19.01	9.06	10.15	15.53	9.72
σ	4.36	3.01	3.19	3.94	3.12

3.3 Experimental Results for Non-Taxonomic Relations

In Table 3, the results obtained from the approach using the accuracy criteria (A); the quality (C) in the prediction of the approach, according to three human

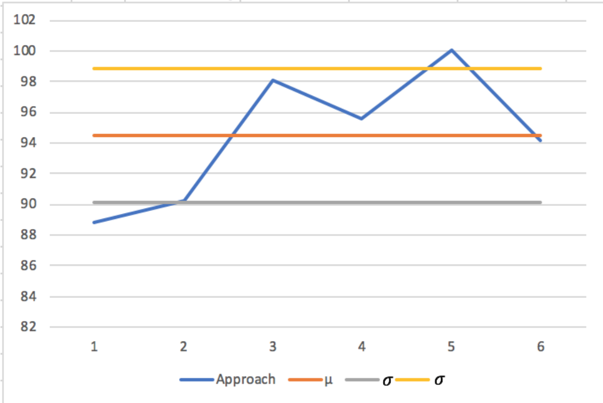

Fig. 1. Standard deviation (σ) of approaches (A) for class-inclusion relations.

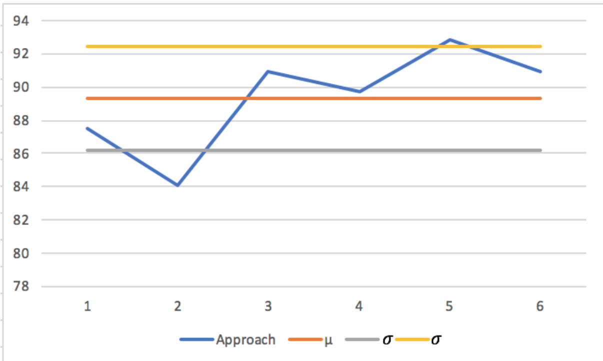

Fig. 2. Standard deviation (σ) of human averages (Avg for class-inclusion relations).

experts (H_1, H_2 and H_3); and the baseline is presented. Table 3 also includes the results of the estimators: mean (μ), variance σ^2 and standard deviation σ. The standard deviation of the accuracy of the results of the approaches is presented in Fig. 3 and the Fig. 4 the standard deviation of the $Avg(H_i)$ is presented, too.

3.4 Experimental Results of the Evaluation Metric and of the Automatic Evaluation System

The results of the approaches presented in Table 2 for class-inclusion relations and the results of Table 3 for non-taxonomic relations are used with the metric of evaluation. In Table 4 is shown the experimental results of the metric (M), where $M(S)$ is the result of the metric for the automated evaluation system, considering only the data validated by the experts, $M(H_i)$ with $i = 1, 2, 3$ and Avg is the average of the results obtained from the experts.

Table 3. Accuracy of the ontology AI and the quality of the predictions of approaches for non-taxonomic relations.

1 Approach	A	$C(H_1)$	$C(H_2)$	$C(H_3)$	$Avg(H_i)$
2 ADS	88.52	81.97	86.89	83.61	84.15
3 Sim-cos	93.44	86.89	88.52	85.25	86.89
4 Sim-cos_u	98.36	88.52	93.44	90.16	90.71
5 FCA min	93.44	80.33	88.52	85.25	84.70
6 FCA sfd_2	100.00	86.89	95.08	91.80	91.26
7 FCA sfd_3	95.08	81.97	90.16	90.16	87.43
8 LSA-cos	90.16	83.61	85.25	85.25	84.70
Baseline	48.00	51.00	46.00	52.00	50.00
μ	94.14	84.31	89.69	87.35	87.12
σ^2	16.91	9.72	12.27	10.47	8.45
σ	4.11	3.12	3.50	3.24	2.91

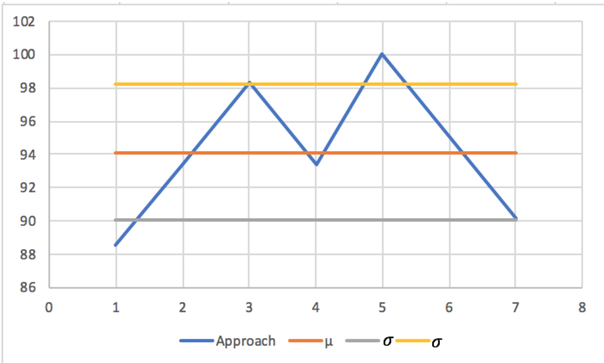

Fig. 3. Standard deviation (σ) of approaches (A) for non-taxonomic relations.

$M(A)$ is the result of the metric for the automated evaluation system considering all the domain corpus.

Table 4. Results of metric evaluation applied to AI domain ontology.

O	$M(A)$	$M(S)$	$M(H_1)$	$M(H_2)$	$M(H_3)$	$M(Avg)$
AI	90.80%	94.31%	86.79%	87.86%	92.52%	89.05%

Now, we use the Equation 4 to calculate the confidence interval, assuming that the evaluations Y_i are normally distributed. Using the data in Table 4, $M(Avg) = \overline{Y} = 89.05$, $s = 3.05$, $n - 1 = 2$ degrees of freedom and using $t_{\alpha/2} = t_{0.025} = 4.303$, then we get $(81.49, 96.63)$ as the observed confidence interval for μ. The estimation of the parameter μ indicates that the results by the experts and by the automatic system are within the allowed limits that determine the mean value of the evaluation of the ontology.

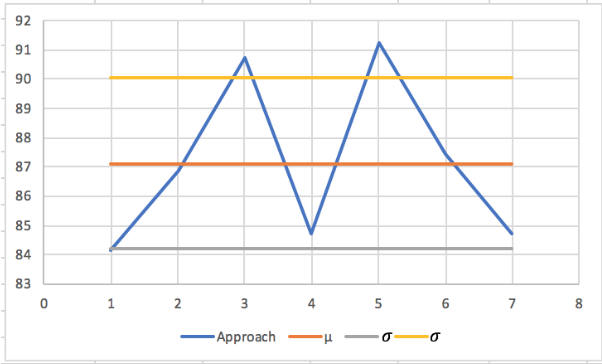

Fig. 4. Standard deviation (σ) of human averages ($Avg(H_i)$) for non-taxonomic relations.

4 Conclusions

This paper presents a study to determine the ideal parameters that indicate the importance of each approach used in the evaluation metric. Based on the experimental results, when using mean, variance and standard deviation how parameters and the validation of human experts, inference that some approaches can be eliminated in the system evaluation process. On the other hand, based on the mean as a parameter, the confidence interval is determined, on which the results of the evaluations can vary. It is observed that the automatic system restricted to the data evaluated by human experts and by the complete corpus is maintained within these limits, which implies that the automatic system provides an acceptable ontology evaluation result. In future work, we will develop other approaches for evaluating the ontology and extend the proposed metric. In addition to including other ontologies and their corpora in the experiments

Acknowledgment. This work is supported by the Sectoral Research Fund for Education with the CONACyT project 257357, and partially supported by the VIEP-BUAP project.

References

1. Berners-Lee, T., Hendler, J.: The semantic web. Scientific American 284, 34–43 (2001)
2. Brank, J., Grobelnik, M., Mladenić, D.: Automatic evaluation of ontologies. In: Kao, A., Poteet, S.R. (eds.) Natural Language Processing and Text Mining, pp. 193–219. Springer London (2007)
3. Gómez-Pérez, A.: Ontology evaluation. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies, pp. 251–273. Springer Berlin Heidelberg (2004)

4. Gruber, T.R.: Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In: Guarino, N., Poli, R. (eds.) Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic Publishers, Deventer, The Netherlands (1993)
5. Tovar Vidal, M., Pinto Avendaño, D., Montes Rendón, A., González Serna, J.G., Vilariño Ayala, D.: Evaluation of ontological relations in corpora of restricted domain. Computación y Sistemas 19(1) (2015)
6. Tovar Vidal, M., Pinto Avendaño, D.E., Montes Rendón, A., González Serna, J.G.: A metric for the evaluation of restricted domain ontologies. Computación y Sistemas 22(1) (2018)
7. Wackerly, D., Mendenhall, W., Scheaffer, R.: Estadística matemática con aplicaciones. Grupo Editorial Iberoamérica (2002)
8. Zouaq, A., Gasevic, D., Hatala, M.: Linguistic patterns for information extraction in ontocmaps. In: Blomqvist, E., Gangemi, A., Hammar, K., Suárez-Figueroa, M.d.C. (eds.) WOP. CEUR Workshop Proceedings, vol. 929. CEUR-WS.org (2012)