A Proposal for Stable Semantic Metrics Based on Evolving Ontologies

Yinglong Ma[1,2], Xinyv Ma[1]
[1]*School of Computer Science and Technology*
North China Electric Power University
Beijing 102206, China
m_y_long@otcaix.iscas.ac.cn

Shaohua Liu[2], Beihong Jin[2]
[2]*Institute of Software*
Chinese Academy of Sciences
PBox8718, Beijing 100190, China
{jbh,ham_liu}@otcaix.iscas.ac.cn

Abstract

In this paper, we propose a set of semantic cohesion metrics for ontology measurement, which can be used to assess the quality of evolving ontologies in the context of dynamic and changing Web. We argue that these metrics are stable and can be used to measure ontology semantics rather than ontology structures. Measuring ontology quality with semantic inconsistencies caused by ontology evolutions, is mainly considered in this paper. The proposed semantic cohesion metrics are theoretically and empirically validated.

1. Introduction

Assessing ontology quality has become an important issue because assessing ontology can help ontology engineers to predict the design quality of ontologies, and select ontologies, and even repair and improve the design of ontologies. Many ontology measures have been proposed and some principal work also has been done to study the nature of measures for ontologies in general [3], [4], [5], [6]. These metrics and principles provide useful guides about what measurement methods are considered and how useful the method is. Although these measures are also applicable to assessing ontology quality, they have encountered some problems. First, most metrics still are based on structural notions without indeed taking into account the semantics such as subsumption, which leads to incomparable measurement results [7]. Second, most proposed metrics are not stable without considering possible additions of further axioms to ontology because they have not taken the open world assumption (OWA) into account, whereas OWA can indeed satisfy the requirements of ontologies in the context of dynamic and changing Web. Third, just because of changing and evolving characters of ontologies, consistent ontologies probably become inconsistent. But fewer metrics are considering measuring ontology inconsistency.

In this paper, we propose a set of stable semantic cohesion metrics to assess the quality of evolving ontologies in the context of dynamic and changing Web. We argue that these metrics are stable and can be used to measure ontology semantics rather than ontology structures. These metrics are Average Axiom Fanouts per Class (AAFC), Number of Minimally Inconsistent subsets (NMIS) and Average Value of Axiom Inconsistencies (AVAI). The proposed semantic cohesion metrics are validated by using Kitchenham et al.'s and Briand et al.'s frameworks [10], [8], and empirically validated by using a prototype implementing the metrics and algorithms presented in this paper.

2. Proposal of Semantic Cohesion metrics

An ontology can be regarded as a DL knowledge base[2]. A DL knowledge base consists of TBox and ABox. In the TBox, complex concept (i.e., class) descriptions can be built on atomic concepts by iteratively applying constructors such as $\sqcap$, $\sqcup$, $\neg$, $\forall R.C$ and $\exists R.C$, etc. An axiom expression is of the form $C \sqsubseteq D$ or $C \equiv D$, where C and D are concepts. An ABox is a set of assertions of the form $C(a)$ or $R(a, b)$, where R is a role, and a, b are individuals. An interpretation function $\cdot^{\mathcal{I}}$ maps every concept A to a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$, and maps every role R to a binary relation $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The syntax and semantics of DL based ontology in detail can refer [2] for the sake of space.

Definition 1: For the two knowledge bases $\mathcal{K}' = (\mathcal{T}', \mathcal{A}')$ and $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, if $\mathcal{K}'$ is the subset of $\mathcal{K}$, then $\mathcal{K}' \subseteq \mathcal{K}$, and the followings hold:
$\mathcal{K}' \subseteq \mathcal{K}$ iff $\mathcal{T}' \subseteq \mathcal{T}$ and $\mathcal{A}' \subseteq \mathcal{A}$.
$\mathcal{K}' \subset \mathcal{K}$ iff $\mathcal{T}' \subset \mathcal{T}$ and $\mathcal{A}' \subset \mathcal{A}$.

Definition 2: For a knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, $|\mathcal{K}| = |\mathcal{T} \cup \mathcal{A}| = |\mathcal{T}| + |\mathcal{A}|$.

Definition 3: For a knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, the set of defined concepts is denoted $\mathcal{C}_{\mathcal{K}} = \{C_1, \cdots, C_n\}$, where each $C_i (1 \leq i \leq n)$ is either an atomic concept or a complex concept defined in $\mathcal{K}$.

Definition 4: For a knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, let C and D be (complex) concepts in $\mathcal{T}$. We say C is directly subsumed by D iff either $C \sqsubseteq D$ is explicitly defined in $\mathcal{T}$ or $C \sqsubseteq D$ can be inferenced without using other specific subsumption axioms in $\mathcal{T}$.

Definition 5: For $C \in \mathcal{C}_{\mathcal{K}}$ of knowledge base $\mathcal{K}$, the axiom fanouts of C are denoted a set $AF_C = \{D_1, \cdots, D_m\}$, where for $D_i (1 \leq i \leq m \leq |\mathcal{C}_{\mathcal{K}}|)$, D_i is directly subsumed by C.

978-0-7695-3615-6/09 $25.00 © 2009 IEEE
DOI 10.1109/JCAI.2009.42

We introduce the definition because we want to make distinction between it and other definitions of fanouts of a class from structural ontology measurement such as [6], [9]. We will find a striking difference by using the example 3. If we only consider measuring the structure of ontology $\mathcal{K}$, then the set of fanouts of the class A is $\{F\}$.

Example 1: For a knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, let $\mathcal{T} = \{A \sqsubseteq B, A \equiv C \sqcup D, E \equiv C \sqcap D, F \sqsubseteq A\}$, and $\mathcal{A} = \emptyset$. We will find that C, D and F are directly subsumed by $C \sqcup D$, and $C \sqcap D$ is directly subsumed by $C \sqcup D$, that is, C, D, E and F are directly subsumed by A. So $AF_A = \{C, D, E, F\}$. Although E is subsumed by B but E is not directly subsumed by B because the axiom that E is subsumed by B can be inferenced by $E \sqsubseteq A$ and $A \sqsubseteq B$.

We propose a metric Average Axiom Fanouts per Class (AAFC). The AAFC value of a knowledge base $\mathcal{K}$ can be formulated as follows:

$$AAFC(\mathcal{K}) = \frac{\sum_{C \in \mathcal{C}_{\mathcal{K}}} |AF_C|}{|\mathcal{C}_{\mathcal{K}}|} \quad (1)$$

Example 2: Considering the example 1, we know that $AF_A = \{C, D, E\}$, $AF_B = \{A\}$, $AF_C = AF_D = AF_E = \emptyset$ and $\mathcal{C} = \{A, B, C, D, E\}$. So $AAFC(\mathcal{K}) = (3 + 1 + 0 + 0 + 0)/5 = 4/5$.

Definition 6: Let $\mathcal{K}$ be a knowledge base, and $\mathcal{K}'$ be any set of axioms and assertions in $\mathcal{K}$. A characteristic function $iv : 2^{\mathcal{K}} \rightarrow \{0, 1\}$ assigns a value to each $\mathcal{K}'$. For any set of axioms and assertions $\mathcal{K}' \subseteq \mathcal{K}$, the drastic inconsistency value of $\mathcal{K}'$ is defined as:

$$\text{iv}(\mathcal{K}')= \begin{cases} 0, if\ \mathcal{K}'\ is\ consistent\ or\ \mathcal{K}'\ is\ empty \\ 1, otherwise \end{cases}$$

Example 3: For $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, $\mathcal{T}$ contains axioms as follows: 1: $A \sqsubseteq B$, 2: $A \sqsubseteq \neg B$, 3: $C \sqcup \neg C \sqsubseteq C \sqcap \neg C$; $\mathcal{A}$ contains an individual assertion: 4: $A(a)$. For the sake of simplicity, we refer to the axioms and assertions by their numbers. There are only $\mathcal{K}'$=({1,2},{4}) and $\mathcal{K}''$=({3},$\emptyset$) such that $iv(\mathcal{K}') = 1$ and $iv(\mathcal{K}'') = 1$.

Definition 7: For a subset $\mathcal{K}'$ of knowledge base $\mathcal{K}$, $\mathcal{K}'$ is the minimally inconsistent subset of knowledge base $\mathcal{K}$ if the following conditions hold:

1. $iv(\mathcal{K}) = 1$
2. for every $\mathcal{K}''$ such that $\mathcal{K}'' \subset \mathcal{K}$, $iv(\mathcal{K}'') = 0$

The metric NMIS (Number of Minimally Inconsistent Subsets) is introduced, which can be formulated for $\mathcal{K}$:

$$NMIS(\mathcal{K}) = |SMIS| \quad (2)$$

where $SMIS$ refers to the set of all minimally inconsistent subsets (SMIS) in $\mathcal{K}$. Here, the algorithm $computeSMIS$ shown in Figure 1 can obtain the SMIS of a knowledge base $\mathcal{K}$.

Example 4: We revise the knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ in Example 3 by adding the axiom 5 and 6. Thus $\mathcal{T}$ contains axioms as follows: 1: $A \sqsubseteq B$, 2: $A \sqsubseteq \neg B$, 3: $C \sqcup \neg C \sqsubseteq C \sqcap \neg C$, 5: $D \sqsubseteq A$; $\mathcal{A}$ contains an individual assertion: 4: $A(a)$, 6: $A(a1)$. We will find that $\mathcal{K}$1=({1,2},{4}),$\mathcal{K}$2=({1,2},{6}) and $\mathcal{K}$3=({3},$\emptyset$) are the minimally inconsistent subsets. But such subsets as $\mathcal{K}$4=({1,2,5},{4}), $\mathcal{K}$5=({1,2},{4,6}) and $\mathcal{K}$6=({1,3},$\emptyset$) are not.

Input: the knowledge base $\mathcal{K}$
Output: SMIS

SMIS=$\emptyset$
for all $\mathcal{K}' \subseteq \mathcal{K}$, sorted by cardinality of $\mathcal{K}'$
  if there exists $\mathcal{K}'' \in$ SMIS such that $\mathcal{K}'' \subset \mathcal{K}'$ **then**
    go the next iteration
  if $\mathcal{K}'$ is inconsistent **then**
    move $\mathcal{K}'$ to SMIS
return SMIS

Figure 1. Algorithm $computeSMIS$

Input: the knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ and its SMIS
Output: AVAIvalue

AVAIvalue=0
if $\mathcal{K} = \emptyset$
  return AVAIvalue
for all $\alpha \in (\mathcal{T} \cup \mathcal{A})$
  i=0
  for all $MIS \in$ SMIS
    if $\alpha \in MIS$ **then**
      i=i+1
  AVAIvalue=AVAIvalue+i+1
AVAIvalue=AVAIvalue/($|\mathcal{K}| + 1$)
return AVAIValue

Figure 2. Algorithm $computeAVAI$

Definition 8: For the knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, its MISs are denoted $SMIS = \{MIS_1, MIS_2, \cdots, MIS_n\}$. Then inconsistency impact value of each axiom or assertion α in $\mathcal{K}$ can be defined by a function mapping $impv$: $(\mathcal{T} \cup \mathcal{A}) \rightarrow \mathbf{N}$, where

$$\text{impv}(\alpha)= \begin{cases} m+1, & if\ for\ all\ i_1,\ \cdots,\ i_j,\ \cdots,\ i_m, \\ & \alpha \in MIS_{i_j} \\ 1, & if\ for\ each\ MIS_k \in SMIS, \\ & \alpha \notin MIS_k \end{cases}$$

where $1 \leq j \leq m$, $1 \leq m, i_j \leq n$ and $1 \leq k \leq n$.

The metric AVAI (Average Value of Axiom Inconsistencies) is proposed, which is the ratio of the sum of inconsistency impact values of all axioms and assertions to the cardinality of a knowledge base $\mathcal{K}$:

$$AVAI(\mathcal{K}) = \frac{\sum_{\alpha \in (\mathcal{T} \cup \mathcal{A})} impv(\alpha)}{|\mathcal{K}| + 1} \quad (3)$$

The algorithm $computeAVAI$ shown in Figure 2 can compute the Average Value of Axiom Inconsistencies.

Example 5: Look back at the example 4. We find that $impv(1) = impv(2) = 3$, $impv(3) = impv(4) = impv(6) = 2$ and $impv(5) = 1$. Then $AVAI(\mathcal{K}) =$

Cohesion metric		Kitchenhanm et al's Framework				Briand et al's framework				Measurement property			
Cohesion metric Approach	Specific metric name	Entity	Attribute	Unit	ScaleType	Nonnegativity and Normalization	Null Value	Monotonicity	Cohesive Modules	Type	Level	Stable	Target
Yao's metrics	NOR	Ontology	Number of all root classes	Class	Interval	+	--	--	+	ST	S	Y	D, U
	NOL	Ontology	Number of all leaf classes	Class	Interval	+	--	--	+	ST	S	Y	D, U
	ADIT-LN	Ontology	Average depth of inheritance tree of all leaf classes	Depth of inheritance	Interval	+	--	--	+	SE	S	N	D, U
Orme's metric	AFRC	Ontology	Fanout	Fanout per root class	Continuous	+	+	+	+	SE	S	N	D, U
OntoQA	Coh	Ontology	Number of separate connected component in the instances	Connected components	Interval	+	+	--	+	ST	I	Y	D, U
Qi's metrics	d_{UR}	TBox	Number of unsatisfiable concepts	Concept	Continuous	+	--	+	+	SE	S	Y	D
	d_{RU}	TBox	Number of root unsatisfiable concepts	Concept	Continuous	+	--	+	+	SE	S	Y	D
	d_{IR}	TBox	Number of axioms in conflict	MIPS	Continuous	+	--	--	+	SE	S	Y	D
Our metrics	AAFC	TBox	Axiom fanout	Axiom fanout per class	Continuous	+	+	+	+	SE	S	Y	D, U
	NMIS	Ontology	Number of MIS	MIS	Interval	+	+	+	+	SE	S, I	Y	D
	AVAI	Ontology	Average value of axiom inconsistencies	Inconsistent impact value per axiom	Continuous	+	+	+	+	SE	S, I	Y	D

Figure 3. Comparison between different ontology cohesion metrics

$\frac{\sum_{\alpha \in (\mathcal{T} \cup \mathcal{A})} impv(\alpha)}{|\mathcal{K}|+1} = \frac{3+3+2+2+1+2}{6+1} = \frac{13}{7}$. The axioms 1 and 2 have the highest inconsistency impact value.

3. Measurement Validation and Analysis

Theoretical Validation and Analysis:

We theoretically validate our metrics using Kitchenham et al.'s and Briand et al.'s frameworks. According to the former framework, we specify the entities, the attribute, the unit, and the data scale. Our ontology cohesion metrics also are validated in accord with the specific criteria of cohesion measurement of Briand et al.'s framework. Because of lack of space, we only can summarize the validations of our semantic cohesion metrics in figure 3. It can be clearly seen that our ontology cohesion metrics are cohesion metrics indeed, and completely satisfy all cohesion properties of Briand et al's validation framework. Meanwhile, according to the differences among validation criteria and properties, our metrics are compared to the existing ontology cohesion metrics of Yao et al.[6], Orme et al.[9], Qi et al.[11] and Tartir et al.[5]. The specific comparison is shown in figure 3. In the figure, '+' ('–') represents that some metric satisfies (does not satisfy) some criteria of the validation framework. 'ST' and 'SE' represent that the metrics are used to measure ontology structure and ontology semantics, respectively. Similarly, 'S'—-Schema, 'I'—-Individual, 'Y'—-Yes, 'N'—-No, 'U'—-Users and 'D'—-Developers.

Preliminary Empirical Evaluation and Analysis:

We have developed a Java-based prototype implementing the algorithms presented above by using KAON2[1], which can completely satisfy our requirements of semantic reasoning. We use ontologies from Object[2], Koala[3], University[4] and Mini-tambis[5]. Considering the generality of ontologies, each of the four ontologies is modified twice (thus resulting in three versions per an original ontology), and form 12 ontologies. The 12 ontologies are used as the data sets to calculate the ontology cohesion metrics.

Then a panel of nineteen evaluators were assembled to assess the set of ontologies to determine cohesiveness of these ontologies. The evaluators have above 3.5 years average experience with ontology based systems or ontology representation. First, the evaluators were sent an electronic copy of each ontology, and then they rated the cohesiveness of each of the 12 ontologies on the following scale: 1) 0=Low; 2) 0.25=Moderate; 3) 0.5=Average; 4) 0.75=High; 5) 1.0=Excellent.

Interrater reliability can be used to address the consistency of the implementation of a rating system, and expressed as a real number in the range of [0,1]. We used SPSS software to compute interrater reliability over the human evaluators ratings, and to determine how well our evaluators agreed with one another. In our experiments, the interrater reliability is 0.9210, which indicates a consistent agreement between the evaluators.

Next, we performed statistical analysis to check the correlation between the averaged evaluation ratings for each of the ontologies and the cohesion metric values. We used Pearson's correlation coefficient with the following hypotheses:

1. http://kaon2.semanticweb.org/
2. http://www.flacp.fujitsulabs.com/tce/ontologies/2004/03/object.owl
3. http://protege.stanford.edu/plugins/owl/owl-library/koala.owl
4. http://www.mindswap.org/ontologies/debugging/university.owl
5. http://www.mindswap.org/2005/debugging/ontologies/miniTambis.owl

Correlated metrics	Correlation	p-value
AAFC	0.601	<0.001
NMIS	0.871	<0.001
AVAI	0.860	<0.001

Figure 4. Statistical Cohesion analysis

$H_0 : \rho = 0$ (There is no correlation between the metrics values and the evaluators' values)

$H_1 : \rho \neq 0$ (There is correlation between the metrics values and the evaluators' values)

The correlation coefficient assumes between -1.0 and 1.0. The larger absolute value of the correlation coefficient means stronger correlation between the pair of variables. If the correlation coefficient is 0, the pair of variables are independent. The following scales proposed by Cohen [1]:1) <0.1 (trivial); 2) 0.10–0.30 (minor); 3) 0.30–0.50 (moderate); 4) 0.50–0.70 (large); 5) 0.70–0.90 (very large); 6) 0.90–1.0 (almost perfect). Another quantitative measures, p-values, are in relation to the hypothesis test of the correlation coefficient being zero. P-values are used in hypothesis tests to either reject or fail to reject a null hypothesis. A small p-value indicates that a null hypothesis is false. Figure 4 shows the Pearson correlation coefficients between cohesion metrics values and evaluators' values and p-values for the hypothesis that $H_0 : \rho = 0$.

Figure 4 reflects the results of comparing each of the ontology cohesion metrics with the evaluators' ratings of cohesiveness. The metrics AAFC, NMIS and AVAI have statistically significant large or very large correlations with the evaluators' ratings. So the three metrics can successfully indicate the cohesion of an ontology.

Repairing and Improving Design of ontologies:

For the metric AAFC, the higher AAFC value an ontology has, the wider range of general domain knowledge it represents, vice versa. The AAFC value of an ontology can help ontology developers to estimate if the ontology should be repaired such that ontology design can be balanced between general and detailed domain knowledge. For the metric NMIS, the more there are MISs in an knowledge base, the more the knowledge base is difficult to share. For the metric AVAI, the higher AVAI value an ontology has, the more inconsistent the ontology is. A coarse proposal for repairing of inconsistent ontologies is as follows: if we detect that an ontology has a higher value of AVAI or NMIS, we should modify the ontology and find out the axioms with the highest inconsistency impact value, and delete them from the ontology. The newly revised ontology will be iteratively repaired until the resulting ontology has no inconsistency.

4. Conclusion

Three stable ontology cohesion metrics are proposed, which are theoretically validated, and the results from our experiments indicate that a good correlation exists between evaluators' opinions of ontology cohesion and the cohesion measured by our cohesion metrics.

Acknowledgement

This work is supported by the Chinese National ”863” High-Tech Program under Grant (No.2006AA01Z231). This work also is supported by Young Teacher Science Development Foundation in NCEPU.

References

[1] Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Plublishing, 1988.

[2] Baader, F., Calvanese, D., et al. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge Press, 2003.

[3] Aldo Gangem and Carola Catenacci and Massimiliano Ciaramita and Jos Lehmann. ”A theoretical framework for ontology evaluation and validation”. Proceedings of SWAP2005, 2005.

[4] A. Lozano-Tello and A. Gomez-Perez. ”OntoMetric: A method to choose the appropriate ontology”. Journal of Database Management, 15(2):1–18, 2004.

[5] S. Tartir and I. Arpinar and M. Moore and A.P. Sheth and B. Aleman-Meza. ”OntoQA: Metric-based ontology quality analysis”. Proceedings of IEEE ICDM 2005 Workshop, 2005.

[6] Haining Yao and Anthony Orme and Letha Etzkorn. ”Cohesion Metrics for Ontology Design and Application”. Journal of Computer Science, 1(1):107-113, 2005.

[7] Denny Vrandecic and York Sure. ”How to Design Better Ontology Metrics”. Proceedings of ESWC2007, 2007, pp.311–325.

[8] Lionel Briand and Sandro Morasca and Victor Basili. ”Property-Based Software Engineering Measurement”. IEEE Transaction on Software Engineering, 22(1):68–86, 1996.

[9] Anthony Orme and Haining Yao and Letha Etzkorn. ”Indicating ontology data quality, stability, and completeness throughout ontology evolution”, Journal of Software Maintenance and Evolution, 19(1):49–75, 2007.

[10] B. Kitchenham and S. Pfleeger and N. Fenton. ”Towards a framework for software measurement validation”. IEEE Transaction on Software Engineering, 21(5):929-9-44, 1995.

[11] Guilin Qi and Anthony Hunter. Measuring Incoherence in Description Logic-based Ontologies. Proceedings of ESWC2007, 2007, pp.381–394.