Superconcept Formation System–An Ontology Matching Algorithm for Web Applications

Jingshan Huang and Michael N. Huhns

Computer Science and Engineering Department, University of South Carolina, USA
{huang27,huhns}@sc.edu

Abstract. Being a formal, declarative knowledge representation model, ontologies are the basis for many Web applications, such as Web services, E-business, and service-oriented computing. But ontologies from different sources are inherently heterogeneous, so they must be matched and reconciled before they can be used. Most approaches for ontology matching are either rule-based or learning-based, but both have disadvantages. In this paper, we propose Superconcept Formation System (SFS), a learning-based matching algorithm with rule-based enhancements. We explain SFS, analyze the challenges, and demonstrate its promising prospect by preliminary experimental results.

Key words: Web Application, Ontology Matching, Machine Learning

1 Introduction

As a formal and declarative knowledge representation model, ontologies are becoming increasingly important as a foundation for Web applications. With the Semantic Web gaining attention as the next generation of the Web, reconciling different views of independently developed and exposed data sources becomes a critical issue. While ontologies serve as a basis for solving this problem, their inherent heterogeneity requires ontologies be matched before they can be made better use of.

Most ontology matching algorithms are either rule-based or learning-based. Both types have their disadvantages. Briefly, the former (such as [1] and [2]) handle semantics imprecisely and heuristically. When ontologies are characterized by the typical aspects of concept names, concept properties, and concept relationships, then these aspects contribute differently to the meaning of a concept. It is thus essential to assign different weights to the aspects. However, current research has made use of humans to define these weights heuristically. For learning-based matching algorithms ([3] for example), the main problem is the difficulty in getting enough instance data of sufficient quality.

2 Superconcept Formation System

Overview Based on the above insights, we propose a learning-based matching algorithm combined with rule-based techniques. Our approach overcomes the

listed disadvantages: (1) we adopt machine learning techniques to avoid predefined weights; and (2) our learning technique is carried out based on schema information alone to avoid the difficulty in getting instance data. Notice that because schemas have many more varieties than instance data, our approach is significantly more challenging than most other learning-based ones.

Main Idea and Experimental Results In our approach, a set of *superconcepts* will be generated during ontology matching. Each superconcept consists of concepts from different ontologies that are equivalent to each other. We design an n-dimensional Euclidean space for concepts, where each dimension corresponds to one semantic aspect. We then calculate the dissimilarity between a pair of concepts, which is the weighted sum of dissimilarities from all corresponding dimensions. After we obtain all pairwise dissimilarities, we apply an Agglomerative Clustering Algorithm to generate the set of superconcepts.

- We propose to supply with our system a set of training examples (equivalent concept pairs by manual matching); then we apply a neural network technique to learn the weights for different semantic aspects.
- Observing that with the increase of dissimilarity threshold t, the number of superconcepts n decreases, we propose that in the evolution pattern of n with respect to t, if a plateau is discovered, then it is reasonable to assign a value within this plateau to t.

Please refer to `http://www.cse.sc.edu/~huang27/result.jpg` for our preliminary results. In summary, **all weights converged to certain values**, and **a curve with a plateau-like pattern emerged**.

3 Conclusion

Ontology matching is a basis for Web applications. Both rule-based and learning-based algorithms have disadvantages, so we propose SFS, a learning-based algorithm integrated with rule-based techniques. We learn weights for different semantic aspects and carry out the learning without the help from instance data. Our preliminary experimental results are promising. In the future, we plan to perform more experiments and compare the performance of our system with work elsewhere.

References

1. Huang, J., Zavala, R., Mendoza, B., and Huhns, M.N.: Reconciling Agent Ontologies for Web Service Applications. In: Proceedings of Multiagent System Technologies: Third German Conference (MATES 05). Springer Verlag, Berlin (2005).
2. Madhavan, J., Bernstein, P.A., and Rahm, E.: Generic Schema Matching with Cupid. In: Proceedings of the Twenty-seventh VLDB Conference. Roma, Italy (2001).
3. Doan, A., Madhavan, J., Dhamankar, R., Domingos, P., and Halevy, A.: Learning to match ontologies on the Semantic Web. In: The VLDB Journal, Vol. 12, No. 4. Springer Verlag, New York, NY (2003) 303–319.