Superconcept Formation System–An Ontology Matching Algorithm for Sevice Discovery

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 Web applications, such as the service discovery. Because of their inherent heterogeneity, ontologies need to be matched before they can be made better use of. Most approaches in ontology matching belong to rule-based or learning-based ones, both with disadvantages. In this paper, we propose Superconcept Formation System (SFS), a learning-based matching algorithm combined with rule-based techniques. We explain our main idea, analyze the challenges, and demonstrate its promising prospect by preliminary experiment results.

Key words: Service Discovery, 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. For example, the service discovery in the Web requires a mutual understanding of underlying ontologies as a first step. While ontologies serve as a basis for solving the problem of reconciling different views of independently developed data sources, their inherent heterogeneity requires ontologies be matched before they can be made better use of.

Most ontology matching algorithms belong to either rule-based or learning-based ones. Both categories have their disadvantages. Briefly speaking, the former, [1] and [2] for example, have a problem in the way they treat different semantic aspects. In general, ontologies are characterized by the aspects of concept names, concept properties, and concept relationships. These aspects have different contributions in understanding ontologies' semantics. Therefore, it is essential to assign different weights accordingly. However, current research work has made use of human heuristic to define these weights. The main problem for learning-based matching algorithms, [3] for example, is the difficulty in getting enough and good-quality instance data.

2   Superconcept Formation System

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

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

Main Idea and Experiment Results In our opinion, a set of *superconcepts* will generate during the ontology matching. Each superconcept consist of a number of concepts that are from different ontologies and are equivalent with each other. We propose to design a n-dimension Euclidean space (vector) for concepts, 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 a set of superconcepts.

– We propose to supply with our system a set of training examples (equivalent concept pairs by manual matching); then we apply an Artificial 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.

Due to the limited space, 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 plateau-like pattern appeared**.

3   Conclusion

Ontology matching is a basis for Web applications, such as the service discovery. Both rule-based and learning-based algorithms have some disadvantages. We propose SFS, a learning-based algorithm integrated with rule-based techniques. We aim to learn weights for different semantic aspects, and we carry out the learning without the help from instance data. Our preliminary experiment results show a promising prospect. In the future work, we plan to perform more experiments, and compare the performance of our system with others'.

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