

Integrating the IAC Neural Network in Ontology Mapping

Ming Mao*
SAP Research
Palo Alto, CA 94304 USA
ming.mao@sap.com

Yefei Peng*
Yahoo!
Sunnyvale, CA 94089 USA
ypeng@yahoo-inc.com

Michael Spring
University of Pittsburgh
Pittsburgh, PA 15260 USA
spring@sis.pitt.edu

ABSTRACT

Ontology mapping seeks to find semantic correspondences between similar elements of different ontologies. This paper proposes a neural network based approach to search for a global optimal solution that best satisfies ontology constraints. Experiments on OAEI benchmark tests show it dramatically improves the performance of preliminary mapping results.

Categories and Subject Descriptors

D.2.12 [Software Engineering]: Interoperability – *Data mapping*; I.2.6 [Artificial Intelligence]: Learning – *Connectionism and neural nets*.

General Terms

Algorithms, Design, Experimentation.

Keywords

ontology mapping, interactive activation and competition (IAC) neural network, constraint satisfaction problem (CSP), PRIOR+

1. INTRODUCTION

The WWW now is widely used as an information exchange platform. However semantic interoperability in the WWW is still limited due to information heterogeneity problem. Ontology as a shared understanding of a domain has been suggested as a solution. With the popularity of ontologies, *ontology mapping*, finding semantic correspondences between similar elements of different ontologies, has attracted many researchers' attention. A comprehensive survey of ontology mapping can be found in [3].

The characteristics of an ontology and its representation result in many kinds of constraints. For example, the hierarchical relations in RDFS [7] do not allow crisscross mappings, the axioms such as *owl:sameAs* and *owl:equivalentClass* in OWL [8] indicate an equivalent relation between different elements, and the rules in SWRL [9] would imply or assert some properties that are not directly available, etc. Therefore, how to find a global optimal configuration that best satisfies ontology constraints is intriguing. Currently most ontology mapping approaches simply validate constraints using isolate heuristic rules. Others include the similarity flooding (SF) [6] utilizes graph theory to verify different constraints and the GLUE [1] adopts relaxation labeling technique to optimize its mapping configuration. However SF can not deal with competitive constraints and GLUE needs the probability of distribution as their prior knowledge.

In this paper we propose using the interactive activation and

competition (IAC) neural network (NN) to search for a global optimal solution to best satisfy ontology constraints. Experimental result of our approach on OAEI benchmark tests is promising. It significantly boosts the performance of our preliminary results.

2. THE APPROACH

Before the discussion of the IAC neural network based constraint satisfaction approach, we briefly introduce how we generate the preliminary mapping results: First we measure three similarities of ontologies (i.e. edit distance based similarity, profile similarity and structure similarity). Then, we adaptively aggregate different similarities according to their harmonies, i.e., a measure that correlates to the reliability of the similarity. Finally based on the aggregated harmony of a mapping task, we selectively activate the IAC neural network to search for an optimal configuration that best satisfy ontology constraints. This paper focuses on the integration of the IAC neural network only. Detailed information about similarity generation and adaptive aggregation is in [4].

2.1 The IAC Neural Network

Usually an IAC neural network consists of a number of competitive *nodes* connected to each other. Each node represents a *hypothesis*. The *connection* between two nodes represents constraint between their hypotheses. If two hypotheses support each other, the connection between them is *positive* (i.e., activation); whereas if two hypotheses are against each other, the connection between them is *negative* (i.e., competition). Each connection is associated with a *weight*, which is proportional to the strength of the constraint. The *activation* of a node is determined by four sources: its initial activation, the input from its adjacent nodes, its bias and external input. The comprehensive introduction of the IAC neural network can be found in [5].

The common properties between the characteristic of ontology mapping and the mechanism of the IAC network motivate the work addressed in the paper. First, the constraints in ontology mapping are either interactive or competitive between mapping hypotheses. For example, for two mapping hypotheses, i.e., e_{1i} maps to e_{2j} and e'_{1i} maps to e'_{2j} , where e_{xy} is an element in ontology O_x , the constraint "if e_{1i} maps to e_{2j} is true, then e'_{1i} maps to e'_{2j} is true, where e'_{1i} and e'_{2j} are children of e_{1i} , and e_{2j} respectively" is interactive; whereas the constraint "if e_{1i} maps to e_{2j} is true, then e'_{1i} maps to e'_{2j} is false, where e'_{1i} is the parent of e_{1i} , and e'_{2j} is the children of e_{2j} " is competitive. Second, our preliminary mapping results estimate both linguistic and structure information of ontologies. Those information bring prior knowledge, i.e., the confidence of some mapping hypotheses, which is suitable to integrate into the IAC network as external inputs or bias of a node.

2.2 Its Implementation in Ontology Mapping

Figure 1 illustrates the implementation of IAC neural network in the context of ontology mapping. In the picture, a *node* (e_{1i}, e_{2j})

*This work was done when the authors studied at University of Pittsburgh.
Copyright is held by the author/owner(s).
WWW 2008, April 21–25, 2008, Beijing, China.
ACM 978-1-60558-085-2/08/04.

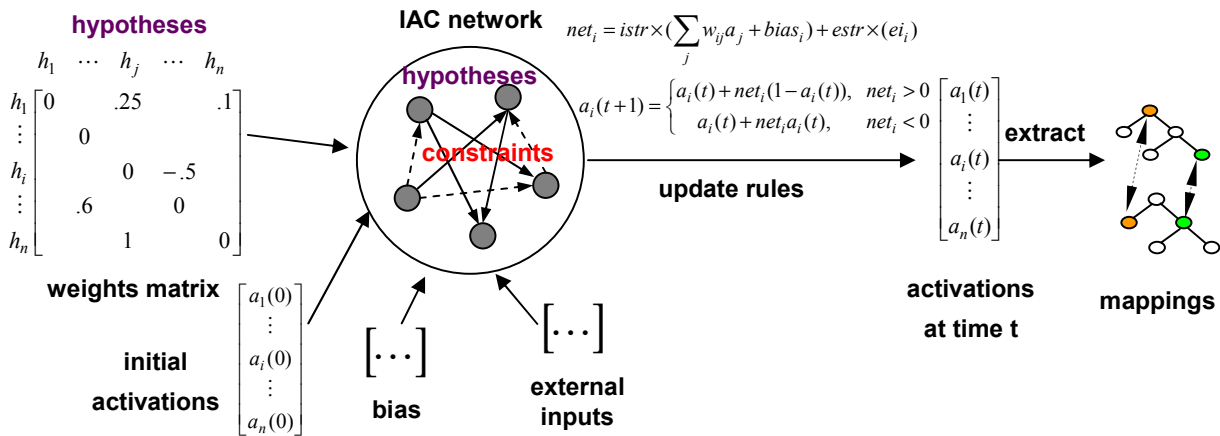


Figure 1. The IAC neural network in the context of ontology mapping

represents a hypothesis that indicates a mapping between e_{1i} and e_{2j} . The connections between nodes represent constraints between hypotheses. For example, the constraint that "only 1-to-1 mapping is allowed" results in a negative connection between nodes (e_{1i}, e_{2j}) and (e_{1i}, e_{2k}) , where $k \neq j$. Moreover, "two elements match if their children match", results in a positive connection between nodes (e_{1i}, e_{2j}) and (e_{1k}, e_{2l}) , where e_{1k} and e_{2l} are the children of e_{1i} and e_{2j} respectively. Currently we implemented 12 constraints. The detail of each constraint is emitted due to the space limit. The weights in weight matrix correspond to the prior probability or confidence of the constraint, which are currently set as 1 for positive constraints and -1 for negative constraints. The initial activation of each node is set to the aggregated similarity of (e_{1i}, e_{2j}) from previous processes. The bias of each node is set as 0. The external input is set to the reliability of each hypothesis. Currently the external input of unambiguous hypotheses is set as 10, otherwise as 0. The activation of a node can be updated with the rule illustrated in the picture, where a_i denotes the activation of node i , written as n_i , net_i denotes the net input of the node. Finally the network is stopped when its global goodness [5] is reached to some satisfaction.

3. EVALUATION

To evaluate our approach, we adopt OAEI benchmark tests [2], which include 1 reference ontology, dedicated to the very narrow domain of bibliography, and 50 test ontologies, 4 of them are real cases and others are handmade by discarding various information from the reference ontology. We follow OAEI evaluation criteria, calculating the precision, recall and f-measure of classes and properties over each test. The experiment methodology is: Given the preliminary mapping results, we first estimate the harmony of the aggregated similarity on each benchmark test. Then we selectively activate the IAC neural network when the harmony is less than a tentatively set number, e.g. 0.6. Currently the network is activated on 20 tests, i.e., 202, 209-210, 248-266, 302 and 303. Finally we extract mapping results using naive descendant extraction algorithm [2]. Experiment results in Figure 2 and Table 1 show the NN based approach improves the f-measure of 16 tests among 20 tests, especially as large as .37 on #254. The overall improvement of the approach on all 20 tests is 13%, 24%, and 19% for precision, recall, and f-measure respectively.

4. Conclusions and Future Work

In this paper we proposed an IAC neural network based approach

to find a global optimal solution that best satisfies ontology constraints. The experimental results show the approach dramatically improves the performance of preliminary mapping results on some tests. Future work may include exploring more complex constraints, optimizing weight matrix, implementing NN in parallel computing platforms to improve its efficiency etc.

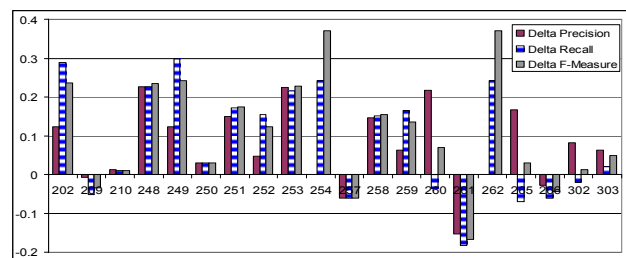


Figure 2. The performance of NN approach on each test

Table 1. The overall improvement of NN approach

H-Mean	Precision	Recall	F-Measure
Before NN	.76	.54	.63
After NN	.88	.67	.76
NN Improvement	13%	24%	19%

5. REFERENCES

- [1] Doan, A., J. Madhavan, et al. "Learning to Match Ontologies on the Semantic Web." *VLDB Journal* 12(4): 303-319.
- [2] Euzenat, J et al. The Proceeding of 2nd Ontology Matching Workshop. Busan, Korea. 2007.
- [3] Kalfoglou, Y. and M. Schorlemmer "Ontology mapping: the state of the art." *Knowledge Engineering Review* 18(1):1-31.
- [4] M. Mao, Y. Peng, et al. A Profile Propagation and Information Retrieval Based Ontology Mapping Approach, In Proceedings of Semantic, Knowledge and Grid. China 2007.
- [5] McClelland, J. and Rumelhart, D. Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises. The MIT Press. 1988.
- [6] Melnik, S., H. Garcia-Molina, et al. Similarity flooding: a versatile graph matching algorithm and its application to schema matching. In Proceedings of ICDE. 2002.
- [7] <http://www.w3.org/TR/rdf-schema/>
- [8] <http://www.w3.org/TR/owl-features/>
- [9] <http://www.daml.org/2003/11/swrl/>