Name : Kalpana
Name of the Supervisor: Dr. Monica Mehrotra
Department of Computer Science, Jamia Millia Islamia
Thesis Title: Enhancing Ontology Alignment with Machine Learning for Better Interoperability of Semantic Web

Abstract of Findings

In our study, we have explored the possibility of selecting best similarity technique out of the available lot, by using two very powerful tools of machine learning such as Neural Networks and Bayesian networks.

Our findings in this study have established the fact that alignments accuracy between the entities to be mapped may increase, only when all the different aspects of the characteristics of candidate entities are explored as a whole. We have considered all four aspects of similarities between two candidate entities i.e. string, semantic, structure and instance similarity. In doing so all four type of similarity values are aggregated and put in as one input. Total 18 similarity measures have been utilized for this purpose. Although it is quite difficult to compute instance based similarity values, our input sample includes few of them.

The first phase of our study is heavily based on machine learning. In our approach, we train Neural Networks and Bayesian Networks separately in two different experiments to find the most effective and important similarity measures among the available ones. Training of neural network is performed using feed forward back propagation algorithm which after training updates the weights of input nodes that represent individual similarity techniques. Similarly a Bayesian Network in which nodes represent individual

similarity techniques, after learning suggests the combination of similarity techniques which are most effective for achieving accurate alignments.

The Experiment I is performed to find the final alignments using the combination of five measures, LCS, Smith-Waterman, SMOA, Lin and Wu & Palmer, for which weights computed by the Neural Networks program were highest and the Experiment II is performed using the measures, Jaccard, Levenshtein, SMOA, Jaro-Winkle and Wu & Palmer as suggested by Bayesian networks.

In the second phase we check and compare the outcome of the two experiments performed in phase II by aligning the test data after training a neural network. Levenberg-Marqudt algorithm is used to train the neural network which is designed to establish the alignment or non-alignment of the candidate entities, when aggregated similarity achieved from the suggested similarity techniques combination by Neural and Bayesian Networks in fed as input.

In the third phase a statistical analysis is performed to finally establish the confidence interval of the accuracy for both the methods at 95% confidence level. The analysis shows that the set of methods suggested by Bayesian Networks gives better accuracy (highest 81.14%) in one or two experiments, but the range of accuracy varies largely i.e.22%. On the other hand the set of methods suggested by Neural Networks performs slightly low in accuracy (highest 77.65%) but the accuracy varies narrowly i.e. 17% in different experiments. Hence the set of methods given by Neural Networks is more stable.