Uncertain Reasoning for Creating Ontology Mapping on the Semantic Web

Miklos Nagy[1], Maria Vargas-Vera[2], and Enrico Motta[1]

[1] Knowledge Media Institute (KMi)
The Open University
Walton Hall, Milton Keynes
MK7 6AA, United Kingdom
mn2336@student.open.ac.uk,e.motta@open.ac.uk
[2] Department of Information Systems
Poznan University of Economics
al. Niepodleglosci 10, 60-967 Poznan, Poland
maria@kie.ae.poznan.pl

Abstract. Mapping ontologies with high precision on the Semantic Web is a challenging problem that needs to be addressed in various domains. One of the main problems with any mapping process, which needs to be applied on different domains is that it always has a certain degree of uncertainty associated with it. In this paper we introduce a method based on Dempster-Shafer theory that use uncertain reasoning over the possible mappings in order to select the best possible mapping without using any heuristic or domain specific rules.

1 Introduction

The problem of mapping two ontologies effectively and efficiently is a necessary precondition to integrate information on the Semantic Web. In recent years different research communities have proposed[1] a wide range of methods for creating such mappings. The proposed methods usually combine syntactic and semantic measures by introducing different techniques ranging from heuristics to machine learning. While these methods perform well in certain domains the quality of the produced mappings can differ from domain to domain depending on the specific parameters defined in the methods e.g. tuning similarity threshold.

We have developed a multi agent ontology mapping framework [2–4] in the context of Question-Answering over heterogeneous sources, where each agent can build mapping between a user's query and the ontology concepts. Our objective was to produce a ontology mapping method that does not depend on any fine tuned internal parameters for a specific domain or does not assume having large amount of data samples a-priory for machine learning or Bayesian probability assessment. Our hypothesis is that the correctness of different similarity mapping algorithms is always heavily dependent on the actual content and conceptual structure of these ontologies which are different even if two ontologies have been created on the same domain but with different purpose. Therefore from the mapping point of view these ontologies will always contain inconsistencies, missing or overlapping elements and different conceptualisation of the same terms, which introduces a considerable amount of uncertainty into the mapping process. In this paper we introduce a novel method how these uncertainties can be harnessed in order to improve the correctness of the mappings.

2 Similarity

In order to assess similarity we need to compare all concepts and properties from *Ontology*1 to all concepts and properties in *Ontology*2. Our similarity assessments, both syntactic and semantic produce a sparse similarity matrix where the similarity between C_n from *Ontology*1 and C_m in *Ontology*2 is represented by a particular similariy measure between the i and j elements of the matrix as follows:

$$
\begin{aligned}
&SIM := (s_{i,j})_{n\times m} \\
&1 \leq i \leq n \; and \; 1 \leq j \leq m
\end{aligned}
$$

where SIM represents a particular similarity assessment matrix, s is a degree of similarity that has been determined by a particular similarity e.g. Jaccard or semantic similarity measure. We consider each measure as an "expert" which assess mapping precision based on its knowledge. Therefore we assume that each similarity matrix is a subjective assessment of the mapping what needs to be combined into a coherent view. If combined appropriately this combined view provides a more reliable and precise mapping that each separate mapping alone. However one similarity measure or some technique can perform particularly well for one pair of concepts or properties and particularly badly for another pair of concepts or properties, which has to be considered in any mapping algorithm.

3 Belief over the mapping

In our ontology mapping method we assume that each expert carries only partial knowledge of the domain and can observe it from its own perspective where available prior knowledge is generally uncertain and subjective. In order to represent these subjective probabilities in our system we use the Dempster-Shafer theory of evidence [5], which provides a mechanism for modeling and reasoning uncertain information in a numerical way, particularly when it is not possible to assign belief to a single element of a set of variables. Missing data (ignorance) can also be modeled by Dempster-Shafer approach and additionally evidences from two or more sources can be combined using Dempster's rule of combination. The combined support, disbelief and uncertainty can each be separately evaluated. The main advantage of the Dempster-Shafer theory is that it provides a method for combining the effect of different learned evidences to establish a new belief by using Dempster's combination rule.

The following elements have been used in our system in order to model uncertainty:

Frame of Discernment(Θ) :finite set representing the space of hypothesizes. It contains all possible mutually exclusive context events of the same kind.

$$\Theta = \{H_1, ..., H_n, ...H_N\} \tag{1}$$

In our method Θ contains all possible mappings that have been assessed by the particular expert.

Evidence:available certain fact and is usually a result of observation. Used during the reasoning process to choose the best hypothesis in Θ. We observe evidence for the mapping if the expert detects that there is a similarity between C_n from O_1 and C_m in O_2.

Belief mass function (m): is a finite amount of support assigned to the subset of Θ. It represents the strength of some evidence and

$$\sum_{A \subseteq \Theta} m_i(A) = 1 \tag{2}$$

where $\mathrm{m}_i(\mathrm{A})$ is our exact belief in a proposition represented by A that belongs to expert i. The similarity algorithms itself produce these assignment based on different similarity measures. In practice we assess up to 8 inherited hypernyms similarities with different algorithms (considered as experts) which can be combined based on the combination rule in order to create a more reliable mapping. Once the combined belief mass functions have been assigned the following additional measures can be derived from the available information.

Belief: amount of justified support to A that is the lower probability function of Dempster, which accounts for all evidence E_k that supports the given proposition A.

$$belief_i(A) = \sum_{E_k \subseteq A} m_i(E_k) \tag{3}$$

An important aspect of the mapping is how one can make a decision over how different similarity measures can be combined and which nodes should be retained as best possible candidates for the match. To combine the qualitative similarity measures that have been converted into belief mass functions we use the Dempster's rule of combination and we retain the node where the belief function has the highest value.

Dempster's rule of combination:Suppose we have two mass functions $m_i(E_k)$ and $m_j(E_{k'})$ and we want to combine them into a global $m_{ij}(A)$. Following Dempster's combination rule

$$m_{ij}(A) = m_i \oplus m_j = \sum_{E_k E_{k'}} m_i(E_k) * m_j(E_{k'}) \tag{4}$$

where i and j represent two different experts.

The belief combination process is computationally very expensive and from an engineering point of view, this means that it not always convenient or possible to build systems in which the belief combination process is performed globally by a single unit. Therefore, applying multi agent architecture is an alternative and distributed approach to the single one and in this case there is no more a single agent having the global view of the system, but each agent has partial view of it. This allows that the computational load can be divided among the agents of the group. Our algorithm takes all the concepts and its properties from the different external ontologies and assesses similarity with all the concepts and properties in the query graph.

4 Conclusions

Inconsistency and incompleteness are important problems that affect the Semantic Web therefore ontology mapping systems that operate in this environment should have the appropriate mechanisms to cope with these issues. The main contribution of our research is the use of Dempster-Shafer theory for assessing whether similar terms in different ontologies refer to the same or similar concepts. Our preliminary results have shown that using Dempster-Shafer theory is a promising approach and needs to be investigated further in ontology mapping context since in this form and context has not been done so far. We believe that this is because Dempste-Shafer combination rules can be unfeasible in domains with large number of variables. In our future research we will investigate how these optimization methods can be adapted and applied in our scenario with a dynamic multi agent environment where each agent has partial knowledge of the domain.

Acknowledgements: This research project has been supported by a Marie Curie Transfer of Knowledge Fellowship of the European Community's Sixth Framework Programme under the contract number MTKD-CT-2004-509766 (enI-RaF).

References

1. Choi N.,Song I-Y., Han H.(2006). A survey on ontology mapping. SIGMOD Record 35(3): 34-41.
2. Vargas-Vera M., and Motta E. (2004). An Ontology-driven Similarity Algorithm. *KMI-TR-151*, Knowledge Media Institute, The Open University, UK.
3. Vargas-Vera M.,Motta E., and Domingue J. (2003). AQUA: An Ontology-Driven Question Answering System. *AAAI Spring Symposium, New Directions in Question Answering*, Stanford, USA.: AAAI Press.
4. Nagy M.,Vargas-Vera M., and Motta E. (2005) Multi-agent Ontology Mapping Framework in the AQUA Question Answering System. the Fourth International Mexican Conference on Artificial Intelligence (MICAI-2005), Lecture Notes in Artificial Intelligence LNAI 3789, Gelbukh, A de Albornoz and H. Terashima (Eds), pp. 70-79, Monterrey Mexico.
5. Shafer G. (1976). A Mathematical Theory of Evidence.: Princeton University Press.