Multiple Strategies Detection in Ontology Mapping

Jie Tang
Department of Computer, Tsinghua University, P.R.China
Knowledge Engineering Group, Department of Computer, Tsinghua University, P.R.China, 100084
+86-010-62789831
j-tang02@mails.tsinghua.edu.cn

Bang-Yong Liang
Department of Computer, Tsinghua University, P.R.China
Knowledge Engineering Group, Department of Computer, Tsinghua University, P.R.China, 100084
+86-010-62789831
liangby97@mails.tsinghua.edu.cn

Juan-Zi Li
Department of Computer, Tsinghua University, P.R.China
Knowledge Engineering Group, Department of Computer, Tsinghua University, P.R.China, 100084
+86-010-62789831
ljz@keg.cs.tsinghua.edu.cn

ABSTRACT

Ontology mapping is the task of finding semantic relationships between entities (i.e. concept, attribute and relation) of two ontologies. In the existing literatures, many (semi-)automatic approaches have found considerable interest by combining several mapping strategies (namely multi-strategy mapping). However, experiments show that multi-strategy based mapping does not always outperform its single-strategy counterpart. We here mainly consider the following questions: For a new, unseen mapping task, should one use a multi-strategy or a single-strategy? And if the task is suitable for multi-strategy, then which strategies should be selected in the combined scenario? This paper proposes an approach of multiple strategies detection for ontology mapping. The results obtained so far show that multi-strategy detection improves both on precision and recall significantly.

Categories and Subject Descriptors

D.2.12 [**SOFTWARE ENGINEERING**]: Interoperability –*Data mapping.*

General Terms: Algorithms, Measurement

Keywords: Ontology Mapping, Semantic Web, Multi-strategy detection

1. INTRODUCTION

Ontology mapping aims to find semantic relationships between entities (i.e. concept, attribute, relation and instance) of two ontologies. In the existing literatures, many automatic approaches have addressed the ontology mapping by exploiting various types of information in ontology, e.g. entity names, taxonomy structures, and constraints as well as characteristics of entities' instances. Using each clue of the available information, an independent strategy can be developed. To achieve high accuracy for a large variety of ontologies, a single strategy (e.g., name based strategy) may be not successful. Hence, to combine different approaches is an effective way. For this purpose, many composite approaches combining multiple mapping algorithms are proposed [1, 3, 4]. However, experiments show that composite approach does not always outperform the single strategy algorithm [2].

In the previous work [3], we present RiMOM, a system that combines multiple strategies for ontology mapping. In this paper, we introduce an approach of multi-strategies detection into RiMOM to automatically detect the optimal composition of multiple strategies for a new mapping task. We call this new version of RiMOM as iRiMOM.

WWW 2005, May 10--14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

2. Problem statement

This section introduces the basic definitions in the mapping process/algorithms and the problem of multiple strategy detection.

2.1 Ontology

The underlying data models in our process are ontologies. The definition of ontology can be written as a six tupel:

$$O = \{C, Attr, H^C, REL, A^O, I\}$$

An ontology O is defined by a set of concepts C and a corresponding hierarchy $H^C \in C \times C$, each of which denotes a taxonomical relation. $Attr$ is a set of data properties for concepts C. Relations $REL \in C \times C$ is a set of non-taxonomical relations between concepts. I denotes a set instances.

2.2 Ontology Mapping

Ontology mapping is defined as: Given two ontologies O_1 and O_2, mapping from ontology O_1 to another O_2 means for each entity in ontology O_1, we try to find a corresponding entity, which has the same intended meaning, in ontology O_2.

Formally, a mapping function can be defined as:

$$Map(e_{i1}, e_{i2}, O_1, O_2) = f$$

with $e_{i1} \in O_1$, $e_{i2} \in O_2$:$e_{i1} \xrightarrow{f} e_{i2}$. e_{i1} denotes a entity, $e_{i1} \in C \cup Attr \cup REL$. f can be one of the mapping types (e.g. equivalentClass) or null. For short, we write the function as $Map(e_{i1},e_{i2})$. We use the notation $Map(O_1,O_2)$ to indicate all entity mappings from O_1 to O_2.

Each available clue in the ontologies can be exploited to develop a mapping strategy. We have developed six strategies in iRiMOM: Instance based strategy, Name based strategy, Entity description based strategy, Name path based strategy, Taxonomy context based strategy and Constraints based strategy [3]. Outputs of the strategies are combined by:

$$Map(e_{i1}, e_{i2}) = \sum_{k=1\ldots n} w_k \sigma(Map_k(e_{i1}, e_{i2})) \Big/ \sum_{k=1\ldots n} w_k$$

where w_k is the weight for individual strategy, and σ is a sigmoid function. It is defined as $\sigma(x) = 1/(1+e^{-5(x-0.5)})$, where x is the predicting value by individual strategy.

2.3 Multiple strategies Detection

We define a strategy as S, and define multi-strategy as a collection: $\{S_j\}$. The notation <$Map_j(e_{i1},e_{i2})$, S_j> denotes the mapping discovered by S_j for e_{i1} and e_{i2}. <$Map_{\{j\}}(e_{i1},e_{i2})$, $\{S_j\}$> denotes the composite mapping determined by multiple strategies $\{S_j\}$ for e_{i1} and e_{i2}. Given two ontologies O_1, O_2 and multi-strategy algorithms $\{S_j\}$, the target of multi-strategy detection is to detect the most appropriate strategy combination so as to result into mappings, i.e. $Map^*(O_1,O_2)$, with minimum error.

$$Map^*(O_1,O_2) = \arg_{Map} \min Err(Map_{\{j\}} \mid \{S_j\}) \quad (1)$$

$Err(Map_{\{j\}}(e_{i1},e_{i2})|\{S_j\})$ is the mapping error by using $\{S_j\}$.

3. Multi-Strategy Detection for Mapping

The key idea behind multi-strategy detection is based on the observation: the higher difference the results obtained by the strategies, the lower probability the combined results outperform the single one [3].

Multi-strategy detection exploits this observation. We first define the mapping error for strategies $\{S_j\}$:

$$\begin{aligned} Err(Map_{\{j\}} \mid \{S_j\}) &= \textstyle\sum_j Err(Map_j \mid S_j, Map_{\{j\}}) \\ &= \textstyle\sum_j (1 - p(Map_j \mid S_j, Map_{\{j\}})) \end{aligned} \quad (2)$$

where $Map_{\{j\}}$ denotes $Map_{\{j\}}(O_1,O_2)$ that obtained by combination of the multiple strategies $\{S_j\}$, and Map_j is the mapping obtained by single strategy S_j. $Err(Map_j|S_j, Map_{\{j\}})$ denotes the mapping error of strategy S_j. It is quantified by using $1-p(Map_j|S_j, Map_{\{j\}})$. $p(Map_j|S_j, Map_{\{j\}})$ captures how much the mapping Map_j is consistent with the mapping $Map_{\{j\}}$. Assuming that the entities' mappings are independent of each other for the given S_j, we have

$$\begin{aligned} & p(Map_j \mid S_j, Map_{\{j\}}) \\ &= \prod_{e_{i1}\to e_{i2}} p(Map_j(e_{i1},e_{i2}) \mid S_j, Map_{\{j\}}(e_{i1},e_{i2})) \end{aligned} \quad (3)$$

where $e_{i1} \to e_{i2}$ means one discovered mapping.

$p(Map_j(e_{i1},e_{i2})|S_j, Map_{\{j\}}(e_{i1},e_{i2}))$ is the probability of difference for (e_{i1},e_{i2}) between combined mapping and S_j's mapping. For short, it is rewritten as $p((e_{i1},e_{i2})|S_j, Map_{\{j\}})$. It is estimated by the degree of difference between S_j's score and the combined score, since each mapping has a score in the automatic mapping scenario, e.g. S_j's score on mapping $e_{i1} \to e_{i2}$ is 0.5 and combined score is 0.6, then $p((e_{i1},e_{i2})|S_j, Map_{\{j\}})$=(0.6-0.5)/0.6= 0.167.

Thus, by substituting equation (3) into equation (2), we obtain

$$\begin{aligned} & Err(Map_{\{j\}} \mid \{S_j\}) \\ &= \textstyle\sum_j (1 - \prod_{e_{i1}\to e_{i2}} p(Map_j(e_{i1},e_{i2}) \mid S_j, Map_{\{j\}}(e_{i1},e_{i2}))) \end{aligned} \quad (4)$$

4. Experiments and Discussions

We have evaluated multi-strategy detection on three data sets. Characteristics of these data sets are shown in table 1 [1].

Table 1. Ontologies in experiments

Ontologies		concepts	instances	Mapping
Course Catalog	*Cornell*	24	1526	34
	Washington	39	1912	37
Course Catalog II	*Cornell*	176	4360	54
	Washington	166	6957	50
Company Profiles	*Standard.com*	333	13634	236
	Yahoo.com	115	9504	104

We take the RiMOM as the baseline to test the effect of strategies detection. RiMOM combines all the strategies.

Table 2 shows the comparison between RiMOM and iRiMOM by using precision (*Pre*) and recall (*Rec*) as the evaluation metrics. {S} denotes the detected strategies, where N—name based strategy, P—name path based strategy, I—instance based strategy, T—taxonomy context based strategy.

Table 2. Experimental comparison

Data set	mapping	RiMOM		iRiMOM		
		Pre	*Rec*	*Pre*	*Rec*	{S}
Course Catalog I	*Cornell to Wash.*	88.2	88.2	97.1	97.1	NP
	Wash. To Cornell	92.1	94.6	94.7	97.3	PIT
Course Catalog II	*Cornell to Wash.*	78.3	87.0	83.9	96.3	NIT
	Wash. To Cornell	75.4	94.0	81.5	96.0	PIT
Company Profiles	*Standard to Yahoo*	81.0	85.0	82.3	91.2	NIT
	Yahoo to Standard	71.4	89.5	71.4	89.5	NPIT

On five mapping tasks of the three data sets, iRiMOM clearly outperforms RiMOM (vary from +1.3% to +8.9% on precision and from +2.7% to +9.3% on recall). Experiments also prove that multi-strategy itself is useful in ontology mapping. In the six mapping tasks, the selected strategies are at least the composite of two strategies.

5. Conclusions

This paper introduces multiple strategies detection into ontology mapping for the first time, and proposes an approach, called iRiMOM, to deal with the interoperability over various ontologies. Experiments show that using strategies detection, iRiMOM improves on precision and recall by +8.9% and +9.3%, respectively. Some of the future directions for our work include investigating methods to make the strategies detection applicable to other domains and to make the detection more efficient.

6. REFERENCES

[1] A. Doan, J. Madhavan, P. Domingos, and A. Halevy. Learning to map between ontologies on the semantic web. In Proceedings of the World-Wide Web Conference (WWW-2002), pages 662-673. ACM Press, 2002.

[2] I. Muslea. Active learning with multiple views. Ph.D. dissertation. (USC, 2002)

[3] J. Tang, B.Y. Liang, J.Z. Li. Risk Minimization based Ontology Mapping. AWCC 2004. Springer Verlag.

[4] H. Wache, T. Voegele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and A. Huebner. Ontology-based integration of information-a survey of existing approaches. In Proc. of IJCAI 2001 Workshop on Ontologies and Information Sharing.