Expert Systems with Applications 41 (2014) 3213–3222

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Optimizing ontology alignment through Memetic Algorithm based on Partial Reference Alignment

Xingsi Xue, Yuping Wang *, Aihong Ren

School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi, China

ARTICLE INFO

Keywords:
Ontology alignment
Clustering
Partial Reference Alignment
Memetic Algorithm

ABSTRACT

All the state of the art approaches based on evolutionary algorithm (EA) for addressing the meta-matching problem in ontology alignment require the domain expert to provide a reference alignment (RA) between two ontologies in advance. Since the RA is very expensive to obtain especially when the scale of ontology is very large, in this paper, we propose to use the Partial Reference Alignment (PRA) built by clustering-based approach to take the place of RA in the process of using evolutionary approach. Then a problem-specific Memetic Algorithm (MA) is proposed to address the meta-matching problem by optimizing the aggregation of three different basic similarity measures (Syntactic Measure, Linguistic Measure and Taxonomy based Measure) into a single similarity metric. The experimental results have shown that using PRA constructed by our approach in most cases leads to higher quality of solution than using PRA built in randomly selecting classes from ontology and the quality of solution is very close to the approach using RA where the precision value of solution is generally high. Comparing to the state of the art ontology matching systems, our approach is able to obtain more accurate results. Moreover, our approach's performance is better than GOAL approach based on Genetic Algorithm (GA) and RA with the average improvement up to 50.61%. Therefore, the proposed approach is both effective.

1. Introduction

With different communities and companies involved in the development of ontologies, it becomes a common situation that multiple ontologies co-exist in the same area for similar application purposes. However, because of human subjectivity, these ontologies may define one entity with different names or in different ways, raising so-called heterogeneity problem which poses as a barrier to semantic interoperability on the ontology level (Acampora, Loia, Salerno, & Vitiello, 2012). Addressing this problem requires to identify correspondences between the entities of various ontologies. This process is commonly known as ontology alignment which can be described as follows: Given two ontologies, each describing a set of discrete entities (which can be classes, properties, predicates, etc.), find the relationships (e.g., equivalence or subsumption) that hold between these entities (Euzenat & Valtchev, 2004).

Ontology alignment plays a key role in expansion and utilization of Semantic Web-based applications (Aumueller, Do, Massmann, & Rahm, 2005; Chen & Huang, 2010). However, it is highly impractical to align the ontologies manually when the size of ontologies is considerably large. Thus, numerous alignment systems have arisen over the years. Each of them could provide, in a fully automatic or semi-automatic way, a numerical value of similarity between elements from separate ontologies that can be used to decide whether those elements are semantically similar or not. Since none of the similar measures could provide the satisfactory result independently, most ontology alignment systems combine a set of different similar measures together by aggregating their aligning results. How to select weights and thresholds in ontology aligning process in order to aggregate various similar measures results to obtain a satisfactory alignment is called meta-matching (Euzenat & Shvaiko, 2007) which can be viewed as an optimization problem and be addressed by approaches like EA.

Nevertheless, most of the approaches based on EA optimize the parameters of meta-matching system with a prerequisite that a Reference Alignment (RA) between two ontologies to be aligned should be given in advance. Since the number of possible correspondences grows quadratically with the number of entities inside the ontology, the typical approach of manually constructing the reference alignment for large scale matching tasks is infeasible. Thus, a new approach utilizes the Partial Reference Alignment (PRA) (Ritze & Paulheim, 2011), which is a set of example mappings that could be provided by a domain expert in a reasonable amount of time, to determine the parameters of meta-matching system has been proposed. The most common way of constructing

* Corresponding author.
E-mail address: ywang@xidian.edu.cn (Y. Wang).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.11.021

PRA is achieved by randomly selecting a subset from the RA to emulate the process of creating example mappings by a domain expert. This may not be entirely correct, as the PRA obtained may not be a representative random sample of the full mapping (Ritze & Paulheim, 2011). To overcome this shortcoming, in this paper, we construct the PRA by a clustering-based approach to ensure the obtained PRA can represent the RA in semantic to a great extent, and then the MA, which is a kind of hybrid EA and extends a traditional GA with local optimization methods (e.g., hill climbing and simulated annealing) to reduce the likelihood of the premature convergence and improve the quality of solutions of problems (Acampora et al., 2012), is proposed to determine the optimal settings of the meta-matching system through the PRA.

The rest of the paper is organized as follows. Section 2 is devoted to discuss the related work; Section 3 introduces the basic definitions; Section 4 describes the syntactics and structure based clustering approach to build the PRA; Section 5 proposes the problem-specific MA; Section 6 shows the experimental results; finally, Section 7 draws conclusions and presents the further improvement.

2. Related work

2.1. Evolutionary algorithm for ontology alignment

In recent years, numerous fully automatic or semi-automatic matching systems have been developed. Lately, the focus of matching systems is on meta-matching. Meta-matching does not use parameters from an expert, but selects those according to a training benchmark, which is a set of ontologies that have been previously aligned by an expert. One group of the meta-matching techniques is called heuristic meta-matching, where the most outstanding approaches are based on EA.

The ontology matching systems that make use of EA can be mainly divided into two categories. The first category tackles the ontology alignment problem as an optimization problem. The representative ontology matching system in this class is GAOM (Wang, Ding, & Jiang, 2006) which is developed by Wang. GAOM utilises GA, where each chromosome represents an alignment of two ontologies and is evaluated by a fitness function. Besides, MapPSO (Bock & Hettenhausen, 2012), which exploits the Particle Swarm Optimization (PSO) instead of GA, also adopts this idea. Recently, being inspired by GA, Acampora employs MA in the alignment problem (Acampora et al., 2012) to improve the performances of GA both in terms of quality of solutions and computational efficiency, and his approach also belongs to this category. Due to the different objective, it is not directly comparable to our approach. While the second class treats the ontology alignment problem as a meta optimization problem. The most notable ontology matching system in this category is GOAL (Martinez-Gil, Alba, & Aldana-Montes, 2008) which is proposed by Jorge Martinez-Gil. GOAL does not directly compute the alignment between two ontologies but it determines, through GA, the optimal weight configuration for a weighted average aggregation of several similarity measures by considering a reference alignment. The same idea is also developed in two more recent papers (Ginsca & Ifene, 2010; Naya, Romero, & Loureiro, 2010), the former proposes a GA based approach to find out how to aggregate different similarity metrics into a single measure, and the latter focus on optimizing the whole similarity aggregation step as a single unit, including the threshold which used for filtering the final alignment. However, all these methods have a drawback which affects strongly their applicability: In order to select the most suitable set of the weights, they require a reference alignment which is provided by domain experts for two ontologies. To overcome this drawback, recently, Acampora, Loia, and Vitiello (2013) propose to use the number of correspondences and the average of the confidence values of the correspondences to approximate the traditional evaluation metrics of alignment. In this way, it allows to directly optimize the set of weights for the ontologies under alignment without requiring a reference alignment. Similar to their work, our work also utilizes the MA in the whole similarity aggregation step of meta-matching system to optimize the ontology alignments. However, the main difference between the work in Acampora et al. (2013) and ours is the evaluation metrics of alignment, and the involvement of the domain experts in our approach is supposed to increase the adaptability of our approach in real application scenarios.

2.2. Semi-automatic matching system using PRA for tuning parameters

Despite the large body of work in the field of ontology matching and its predecessor schema matching, there is little work done which is focused on semi-automatic matching system using PRA for tuning parameters.

SAMBO (Lambrix & Liu, 2009) is the most notable matching system based on PRA which uses PRA as anchors to give hints for partitioning larger ontologies in a pre-processing step, as well as for filtering those incorrect mappings in a post-processing step. Another semi-automatic matching system exploiting PRA and applying machine learning methods is LSD (Doan, Domingos, & Halevy, 2001). It asks the user to provide the semantic mappings for a small set of data sources, then uses these mappings together with the sources to train a set of learners. ECOMatch (Ritze & Paulheim, 2011) also asks the user to provide example mappings instead of parameter settings, and then determines a suitable parameter setting based on those examples. However, so far, among those meta-matching systems that make use of the evolutionary algorithm, none have utilized the PRA for tuning the parameters.

In our approach, we first utilize a syntactics and structure based clustering algorithm to separate entities (classes and properties) of each ontology into a set of small clusters which are relatively independent in semantic, and then select entities from those clusters to ensure the selected sample entities can represent the original ontology in semantic to a great extent, finally we use the MA based on PRA to determine the optimal settings of the meta-matching system. The key differences between the existing approaches and ours are that: (1) existing approaches typically work in session mode and require constant users attention, while our approach works in batch mode once the users examples correspondences are given; (2) existing approaches determines the entities in PRA in a completely random way, while our approach selects those from different class clusters which can better represent the original ontology in semantic and therefore improves the opportunity of finding the better and even optimal parameter settings.

3. Preliminaries

3.1. Ontology and ontology alignment

There are many definitions of ontology over years. But the most frequently referenced one was given by Gruber in 1993 which defined the ontology as an explicit specification of a conceptualization. For convenience of the work in this paper, an ontology can be defined in Definition 1.

Definition 1. An ontology is a triple, $O = (C,P,I)$,

where: