Knowledge-Based Systems 000 (2017) 1–10

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Collaborative ontology matching based on compact interactive evolutionary algorithm

Xingsi Xue[a,b,*], Jianhua Liu[a,b]

[a] College of Information Science and Engineering, Fujian University of Technology, Fuzhou, Fujian, 350118, China
[b] Fujian Provincial Key Laboratory of Big Data Mining and Applications (Fujian University of Technology), Fuzhou, Fujian, 350118, China

ARTICLE INFO

Article history:
Received 11 May 2017
Revised 15 August 2017
Accepted 12 September 2017
Available online xxx

Keywords:
Semantic web
Ontology matching
Collaborative validation
Memetic algorithm

ABSTRACT

Ontology is the kernel technology of semantic web, which plays a prominent role for achieving interoperability across heterogeneous systems and applications by formally describing the semantics of data that characterize a particular application domain. However, different ontology engineers might have potentially opposing world views which could yield the different descriptions on the same ontology entity, raising so called ontology heterogeneous problem. Ontology matching, which aims at identifying the correspondences between the entities of heterogeneous ontologies, is recognized as an effective technology to solve the ontology heterogeneous problem. Due to the complexity of ontology matching process, ontology alignments generated by the automatic ontology matchers should be validated by the users to ensure their qualities, and the technology that makes multiple users collaborate with each other to help the automatic tool create high quality matchings in a reasonable amount of time is called collaborative ontology matching. Such a collaborative ontology matching poses a new challenge of how to reduce users' workload, but at the same time, increase their involvement's value. To address this challenge, in this paper, we propose a Compact Interactive Memetic Algorithm (CIMA) based collaborative ontology matching technology, which can reduce users' workload by adaptively determining the time of getting users involved, presenting the most problematic correspondences for users and helping users to automatic validate multiple conflict mappings, and increase user involvement's value by propagating the collaborative validation and decreasing the negative effect brought by the error user validations. The experimental results show that our proposal is able to efficiently exploit the collaborative validation to improve its non-interactive version, and the runtime and alignment quality of our approach both outperform state-of-the-art interactive ontology matching systems under different user error rate cases.

1. Introduction

Semantic inter-operability represents the capability of two or more systems to meaningfully and accurately interpret the exchanged data so as to produce useful results, which requires the meanings of any data must be specified in an appropriate detail in order to resolve the potential ambiguity. Ontology is the kernel technology of semantic web, which plays a prominent role for achieving interoperability across heterogeneous systems and applications by formally describing the semantics of data that characterize a particular application domain [1]. So far, ontologies have been applied in multiple domains of application such as information retrieval, medical diagnosis, e-Commerce, knowledge acquisition, bio-informatics and service-oriented computing [2,3]. However, different ontology engineers might have potentially opposing world views which could yield the different descriptions on the same ontology entity, raising so called ontology heterogeneous problem. Ontology matching, which aims at identifying the correspondences between the entities of heterogeneous ontologies, is recognized as an effective technology to solve ontology heterogeneous problem. Since the purely manual specification of semantic correspondences is highly impractical especially when the scale of the ontology is large, in recent years, many automatic ontology matching systems have been proposed. However, due to the complexity of the ontology matching process, ontology alignments generated by the automatic matching tools should be validated by the users to ensure their qualities [4]. To make multiple users collaborate with each other to help the automatic tool create high quality matchings in a reasonable amount of time is called collaborative ontology matching [5], but the research on it is still in its infancy.

* Corresponding author.
E-mail address: xxs@fjut.edu.cn (X. Xue).

https://doi.org/10.1016/j.knosys.2017.09.017
0950-7051/

Since users are rare and valuable resources, we can not expect them to validate the whole alignment especially in the context of large scale ontology matching. Therefore, collaborative ontology matching poses a new challenge of how to reduce users' workload, but at the same time, increase their involvement's value [6]. To address this challenge, we need to answer four questions: (1) when should we get a user involved in the automatic ontology matching process; (2) which specific candidate correspondences should be presented to users for validation; (3) how to deal with the inconsistency in user validations; and (4) how can the user validation be exploited to improve the ontology alignment. In this paper, we present a Compact Interactive Memetic Algorithm (CIMA) based collaborative ontology matching technique, which (1) uses the Compact Memetic Algorithm (CMA) to adaptively determining the time of getting users involved; (2) automatically determines the most problematic mappings for users to validate; (3) uses the consensus measure to measure the reliability of inconsistent user validations; and (4) utilizes the propagation gain measure to increase user involvement's value. Particularly, the novelty and innovation of our work are listed as follows:

- A novel compact interactive Memetic Algorithm is proposed to automatically search for the potential mappings, and reduce the users' workload by adaptively determining the time of getting users involved;
- Three techniques, i.e., ontology partition, problematic mapping determination and multiple conflict mappings validation, are presented to reduce the users' workload;
- Two measures, i.e., consensus measure and propagation gain measure, are respectively proposed to measure the reliability of inconsistent user validations and the propagation gain so that our approach can propagate the user validations to increase user involvement's value.

The rest of the paper is organized as follows: Section 2 describes the related works of this paper; Section 3 illustrates the ontology matching problem and similarity measure; Section 4 presents a CIMA based collaborative ontology matching technology; Section 5 presents the experimental studies and analysis; finally, Section 6 draws the conclusions and presents the future work.

2. Related work

To improve the ontology alignment's quality, it's necessary to ask users to examine the mappings produced by the automatic ontology matchers, and indicate which ones are correct and which are not [4]. To this end, a number of interactive ontology matching technologies are developed to get users involved in the ontology matching process, and various strategies for exploiting user validations are proposed.

2.1. Reduce user's workload

Presenting all correspondences to the user at once is too overwhelming and in fact would annoy the user as they become frustrated sifting through all the correspondences [7]. Based on this consideration, different strategies are proposed to reduce the user's workload. Shi et al. [8] propose to use a threshold, which is determined by an interactive algorithm, as well as a similarity propagation graph to select the most informative and problematic mappings as the candidate correspondences. Jiménez-Ruiz et al. [9] identify three general principles on the consistency, locality and conservativity, and utilize them to filter the candidate correspondences. Beisswanger et al. [10] propose some quality checks for measuring ontology alignment's validity and reusability, which are also used to determine the candidate correspondences. Cruz et al. [11] select problematic mappings for user, which are disagreed by different ontology matchers. SAMBO [12] proposes a conceptual ontology alignment framework that involves user in different interruptible sessions for selecting, combining and tuning different ontology matchers. SAMBO reduces user's workload in each validating process by allowing user to validate a sub-set of the candidate mapping, i.e. user can adjust his workload by interrupting the validation and reduce the erroneous validations by reviewing his historical decisions and those candidate mappings still to validate. In general, SAMBO aims at providing user with more flexible operations in each validating process instead of reducing his total workload through, e.g., decreasing the total request number. GOMMA [13] utilizes a composition-based adaptation and a diff-based adaptation algorithm to reuse unaffected correspondences and adapt only the affected mapping part. PROMPT [14] considers the area of the latest user intervention and uses it to compute a new portion of suggestions to maintain the users focus.

Different from these approaches, in order to reduce the users' workload, our approach first uses the ontology partition algorithm to transform the original large scale ontology matching problem into several small scale ontology segment matching problems, which can reduce the number of candidate correspondences presented for user validation. Then, Evolutionary Algorithm (EA) is utilized to adaptively determine the time of getting users involved, present the most problematic correspondences for users, and help users to automatic validate multiple conflict mappings. Since users don't need to get involved in each iterative matching step and validate all the candidate mappings, our approach can significantly reduce the user's workload and improve the validating process's efficiency.

2.2. Increase user involvement's value

When the scale of the ontology is large, only validating the selected correspondences is far from sufficient. Therefore, it is desirable that an algorithm can increase user involvement's value by propagating the user validation to determine other potential matchings between two ontologies. Propagation algorithm is always carried out during the matching process, where the mapping confidence from some validated mapping is propagated to those in their neighborhood. The neighborhood can be defined according to the structure of the ontologies or the pattern of similarity scores from the various alignment algorithms [4]. Shi et al. [8] propose an active learning framework for ontology matching, which tries to find the most informative candidate matches to query, and propagate the user correction according to the ontology structure to improve the matching accuracy. AgreementMaker [11] utilizes the signature vector to propagate user's feedback to validate other mappings without presenting them to the user. Particularly, it rewards (or penalizes) the similarity value associated with every mapping in the same cluster by a linear function that increases (or decreases) the similarity values, and the similarity of mappings that have been validated (or refuted) is set to 1 (or 0) and not updated anymore.

Since the validation process is a difficult cognitive task, it's risky to take user's input for granted, which might lead to the propagation of the errors. Being inspired by the success of the collaborative recommendation in the applications such as e-government services in the business domain [15], web-page recommendation in the intelligent web domain [16] and movie recommendation [17], our proposal gets multiple users, instead of only one user, to validate the alignment collaboratively, which can effectively reduce the risky of introducing the erroneous user validations. In addition, our approach utilizes the propagation gain measure to adaptively calculate the propagation gain and propagate those validated mappings with high reliability, which can effectively improve the align-

Please cite this article as: X. Xue, J. Liu, Collaborative ontology matching based on compact interactive evolutionary algorithm, Knowledge-Based Systems (2017), https://doi.org/10.1016/j.knosys.2017.09.017