Ontology matching by actively propagating user feedbacks through upper ontologies

Juego de ontología que propaga activamente evaluaciones de usuarios a través de ontologías superiores

Raúl Ernesto Menéndez-Mora*

Ryutaro Ichise**

Resumen

Coincidencia ontología es un proceso complejo y en gran parte impulsado por los usuarios para encontrar correspondencias entre entidades de diferentes ontologías. Se han propuesto muchos algoritmos para automatizar la generación de coincidencias. Sin embargo, no pueden ser totalmente automatizados ya que se requiere la intervención del usuario para aceptar, rechazar, o crear nuevas alineaciones o matchings.

En este trabajo se extiende sobre el marco de aprendizaje activo para la adaptación de la ontología, que trata de encontrar las coincidencias de candidatos más informativos para consultar al usuario. En nuestro enfoque de retroalimentación del usuario explota ontologías superiores como puentes semánticos. Estos puentes contribuyen al proceso de correspondencia en general teniendo en cuenta la información supervisada y su propagación en la corrección de concordancias de error. En la experimentación nuestro trabajo superó a la versión anterior en el que se no ha de utilizar ningún elemento de ontología supe-

*   National Institute of Informatics 2-1-2 Hitotsubashi Chiyoda-ku, Tokyo, 101-8430 Japan. E-mail: menendez@nii.ac.jp

**   National Institute of Informatics 2-1-2 Hitotsubashi Chiyoda-ku, Tokyo, 101-8430 Japan. E-mail: ichise@nii.ac.jp

85

rior, mientras que sigue siendo tan competitivo como el estado del sistema de concordancia de la ontología arte.

Palabras clave: Coincidencia de Ontología, Usuario Feedback, Ontología Superior, Aprendizaje Activo.

Abstract

Ontology matching is a complex and largely user-driven process of finding correspondences between entities belonging to different ontologies. Many algorithms have been proposed to automate the matching generation. However, they can't be fully automated since the user input is required to accept, reject, or create new alignments or matchings.

This paper extends on active learning framework for ontology matching, which tries to find the most informative candidate matches to query the user. In our approach the user's feedback exploits upper ontologies as semantic bridges. Such bridges contribute to the overall matching process while considering the supervised information and its propagation in correcting mistake matchings. In the experimentation our work outperformed the previous version where none upper ontology was used, while it remains as competitive as state of the art ontology matching system.

Key words: Ontology Matching, User Feedback, Upper Ontology, Active Learning

1. Introduction

Comprehensive ontology matching is an active field of study for at least a decade but it still requires more automatic methods. The tedious, cumbersome task of manually accepting, rejecting or creating new matchings, remains a bottleneck that severely slows the development of the semantic web.

Human users, especially the domain experts, are capable of discovering complex relationships between candidate's pairs of ontologies' entities. During the ontology matching process, their knowledge is commonly ignored, and human effort is made only for judging whether there is a matching or not.

This paper extends on active learning framework for ontology matching, which tries to find the most informative candidate matches to query the user. Thus our approach allows the user's feedback to exploit upper ontologies as semantic bridges. Such bridges contribute to the overall matching process in correcting mistake matchings. We propose a correct propagation algorithm which consid-

ers the upper ontologies to spread the user relevance feedbacks.

The paper is organized as follows. The next section reviews related work. Insection 3 we describe our framework and the new correct propagation algorithm. Section 4 discusses the experiment and its results, while section 5 summarizes our work, draws some conclusions, and outlines future work.

2. Related Work

- Relevance Feedback for Ontology Matching Systems. Recently some researches have introduced some kind of relevance feedback into the ontology matching process. In [8] it is used for determining threshold values to classify matching pairs from non-matching ones and for detecting, correcting and propagating the error matches. Other works like [1, 9] used it for helping to train classifiers. In the next subsection we introduce one of these approaches.

- Active Learning for Ontology Matching. The key idea behind active learning is that a machine learning algorithm can achieve greater accuracy with fewer training labels if it is allowed to choose the data from which it learns [7]. An active learner may pose queries, usually in the form of unlabeled data instances to be labeled by an oracle. An Active Learning Framework for Ontology Matching was proposed by Shi et al. [8]. These researchers find the most informative candidate correspondences (matches) to query, and propagate the user correction according to the ontology structure to improve the matching accuracy.

- Ontology Matching via Upper Ontologies. Two concrete attempts to exploit upper ontologies for solving the ontology matching problem have been reported in the literature. LOM, a Lexicon-based Ontology Mapping Tool [3] and Automatic Ontology Matching via Upper Ontologies [4]. Some other works have been reported where background ontologies were used for ontology matching.

3. A Framework for Actively Propagating User Feedback through Upper Ontologies

3.1. The Framework

In this section we combine all these previously introduced knowledge to extend on active learning framework for ontology matching [8]. Specifically we propose a new correct propagation algorithm which includes the predisposition of being a good match. Our framework for ontology matching, actively propagates user feedbacks through upper ontologies, using those ontologies as semantic bridges. Such bridges contribute to the overall matching process while considering the supervised information and its propagation in correcting error matches[2].

The framework requires two ontologies to be aligned (O_S and O_T), a set of ontology matching algorithms $\{M_k\}$, a number of user feedbacks (N) about the matches (correspondences) and a set of upper ontologies $\{U_l\}$. During the initialization phase each matching algorithm M_l is used to obtain a map between the source O_S and the target ontology O_T, and to create the bridges from each entity of the source and target ontology to entities in the upper ontologies.

---

2 A match is an error match if its correctness is different after the user confirmation

Figure 1. Example of a similarity propagation graph with match predisposition created from the source and target ontologies
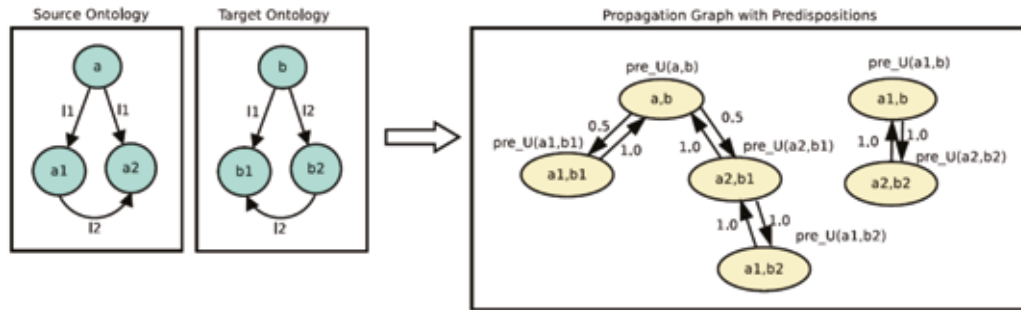
Reference: investigation.

Each match pair entity (e_S, e_T) in the propagation graph, where $e_S \in O_S$ and $e_T \in O_T$, will use the bridge with the shortest distance through the upper ontology U_l as it's predisposition. For improving the matching result the correct propagation algorithm have to solve the following problems:

1. how to select the most informative candidate match to query, and
2. how to improve the matching result with the confirmed matches.

The first problem is solved in [8] where three measures to select informative matches to query were proposed. However, we will use just the two of them whose combination achieved the best results: Confidence metric and Similarity disctance metric. The second problem received the focus of our attention because now a new correct propagation algorithm is needed.

3.2. The New Correct Propagation Algorithm

The new correct propagation algorithm is based on the similarity propagation graph whose construction follows the rules described in [5]. However, each node (match pair) in the propagation graph will have a real value associated to it, as shown in Figure 1. This value represents the predisposition of being a good match after exploiting the semantic bridges created through the upper ontologies.

Match predisposition. Our algorithm try to find the entities in the upper ontologies U_l which stands for the highest value of similarity with the entities e_S and e_T in the source and the target ontology respectively. Knowing those entities (u_l and u'_l) from U_l then the distance (other metrics can also be used) between them is computed and the smallest value from the possible combinations is returned as the predisposition between the entities e_S and e_T. For example, if $u_l = u'_l$ then $dist(u_l, u'_l) = 0$ and consequently $pre(e_S, e_T) = 0$, which means there is none predisposition for the match e_S, e_T.

Since several matching methods M_k (each of them with a different similarity function f_m) can be used to compute the mappings between the source ontology and the upper ontology $(O_S \square U_l)$ and between the target and the upper ontology $(O_T \square U_l)$, the predisposition can be finally calculated as shown in Formula 1:

$$pre(e_S, e_T) = \begin{cases} \min\{dist(u_l, u'_l)\} \mid \forall u_i, u'_i \in \{U_1, ..., U_h\}, \exists u_l, u'_l \in \{U_1, ..., U_h\}, \\ u_l \neq u_i, u'_l \neq u'_i, f_m(e_S, u_l) > f_m(e_S, u_i) \\ \text{and } f_m(e_T, u'_l) > f_m(e_T, u'_i) \end{cases} \quad (1)$$

- Correct propagation. When the correction or confirmation of the selected matches is provided by users (feedback), the correct propagation updates all the matches according to the following update rules:

$$sim(a_i, b_i) = sim(a_i, b_i) + \alpha * w((x, y), (a_i, b_i)) * (1 - sim(x, y))$$
$$* (1 - er(a_i, b_i)) * (1 - pre(a_i, b_i)); \quad (2)$$
$$(x, p, a_i) \in O_S, (y, p, b_i) \in O_T$$

$$sim(a_i, b_i) = sim(a_i, b_i) - \alpha * w((x, y), (a_i, b_i)) * sim(x, y) * er(a_i, b_i) * pre(a_i, b_i);$$
$$(x, p, a_i) \in O_S, (y, p, b_i) \in O_T$$
$$(3)$$

In Formula 2 and 3, (x, y) is the selected error match, and $sim(x, y)$ is the similarity degree. (a_i, b_i) is one of the matches related to the match (x, y), and $w((x, y), (a_i, b_i))$ is the weight of their relation, and $er(a_i, b_i)$ stands for the error rate of the match (a_i, b_i), and α is an effect factor which is used to control the rate of the propagation. If the match (x, y) is correct (user feedback), the update function uses Formula 2, else it uses Formula 3. Now, our match selection is according to the calculation of error rate, propagation rate and enhanced with the predisposition.

The correct propagation runs in an iterative process. In each iteration, it selects the match for user feedback with the error rate and the propagation rate, and then let users to confirm the selected match. After the confirmation, it updates the similarity degree, error rate and the propagation rate of related matches. Then it repeats this process until convergence (e.g., no any change) or the number of query times reaches a predefined threshold.

4. Experiments and Results

4.1. Experimental Settings, Data and Evaluation Methodology

- Experimental settings. In our enviroment we used the Align API v4.1 [2] for computing the alignments (mappings) between the ontologies and for calculating the performance metrics, the SUMO upper ontology [6] as the semantic bridge and WordNet 3.0 as the lexical database backing up the semantic similarities.

- Data sets. For our experiments we used the OAEI[3] 2010: 101, 301, 302 and 304 benchmark datasets[4]. The reason why just a small subset of the OAEI Benchmark test dataset have been used lies not only on the availability of the ground true (reference alignment) for each of them but because experiments 301 - 304 are real-life ontologies, so using real upper ontologies and their inherent knowledge can be effectively considered.

- Evaluation Methodology. For evaluating our algorithm we computed the precision, recall and F1-measure after matching (aligning) the ontologies in each dataset. Two main type of experiments were conducted depending on the match selection metric (Confidence metric and Similarity Distance metric) considered.

---

3   Ontology alignment evaluation initiative, http://oaei.ontologymatching.org/

4   Dataset 303 was excluded because the Align API [2] could not treat it properly.

89

Table 1. Results for our framework while using 1 and 5 feedbacks and two different match selection metrics

Algorithm	CO_FB1			SD_FB1			CO_FB5			SD_FB5		
test	Prec.	Rec.	FMeas.	Prec.	Rec.	FMeas.	Prec.	Rec.	FMeas.	Prec.	Rec.	FMeas.
101	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.92	0.92	0.93	0.93	0.93
301	0.81	0.79	0.80	0.81	0.79	0.80	0.81	0.79	0.80	0.81	0.79	0.80
302	0.52	0.58	0.55	0.52	0.58	0.55	0.52	0.58	0.55	0.52	0.58	0.55
304	0.77	0.95	0.85	0.77	0.95	0.85	0.78	0.96	0.86	0.77	0.95	0.85
H-mean	0.78	0.84	0.81	0.78	0.84	0.81	0.79	0.84	0.81	0.78	0.84	0.81

Reference: investigation

Table 2. Comparison with the best results of Shi et. al. [8]

	Best FMeas.
Shi et. Al.	0.76
Our approach	0.85

Reference: investigation.

Table 3. Comparison with two of the top state-of-the-art ontology matching systems from OAEI 2010

Experiment	ASMOV	RiMOM	Our Approach
30x	0.82	0.82	0.76

Reference: investigation.

for choosing the most informative matches to query. Later on the number of feedbacks fed into the framework was also changed for each of the previously selected metric.

4.2. Results and Discussion

Table 1 compiles the results of our four experiments. In CO FB1 and CO FB5 the metric used for selecting the most informative match to query was Confidence metric where 1 and 5 matches were supplied to the system as relevance feedback. Experiments SD FB1 and SD FB5 were similar to the previous ones but using Similarity Distance metric.

The results in Table 1 showed none improvement when the number of feed-back was increased from 1 to 5. A similar behavior was observed when using 10 and 15 feedback as well. In [8] a similar behavior was observerd but after 10 queries. An important point to consider is our approach, as well as in [8] can only correct the errors. Thus, if there are no error matches below the threshold, the approach cannot improves the results.

We also analyzed the propagation graph generated and we believe another reason for the no improvement when increasing the number of feedbacks is be-cause of the poor structure of the ontologies being matched. We strongly believe a different improvement would be obtained if the ontologies had an stronger hi-erarchical structure, in which case the propagation of the feedback would achieve higher impact in the overall performance.

However, Shi et al. reported a value of 0.76 for the F1-measure as their best result among the OAEI 2008 30x benchmarks they tested [8]. So, our experiment with 0.85 represent

an improvement of 9% and 4% when compared with the best result reported in [8], see Table 2. Just the result for the benchmark experiment 302 reported lower results than the best reported result in [8].

Table 3 shows a comparison of our approach with two of the top state-of-the-art ontology matching systems from OAEI 2010. In Table 3 we can check our approach, although a bit behind, it still remains as competitive as the top state-of-the-art ontology matching systems performing over the same datasets.

5. Conclusion and Future Work

We presented a framework that actively propagates user feedback through upper ontologies for ontology matching. The proposed feedback propagation algorithm exploits upper ontologies as semantic bridges and the structural properties of the ontologies been matched by computing and propagating the predisposition of pairs in been a good match. In the experimentation our framework outperformed the previous version where none upper ontology was used, while it remains as competitive as state of the art ontology matching system.

As future work for this ongoing reserach, we plan to use different match selection metrics as well as to apply our framework to bigger ontologies. The main limitation related with it is the lack of a reference alignment for further computation of the performance metrics. Another task for the future is to consider other upper ontologies like OpenCyc and DOLCE and to introduce different types of user feedback.

6. References

[1]. S. Duan, A. Fokoue, and K. Srinvias. One size does not fit all: Customizing ontology alignment using user feedback. In Proc. 9th Int. Semantic Web Conf., pages page 177–192, 2010.

[2]. J. Euzenat. An api for ontology alignment. In Proc. 3rd Int. Semantic Web Conf., pages 698–712, 2004.

[3]. J. Li. Lom: A lexicon-based ontology mapping tool. In Proc. The Performance Metrics for Intelligent Systems (PerMIS), 2004.

[4]. V. Mascardi, A. Locoro, and P. Rosso. Automatic ontology matching via upper ontologies: A systematic evaluation. IEEE Transactions on Knowledge and Data Engineering, 22:609–623, 2010.

[5]. S. Melnik, H. Garcia-Molina, and E. Rahm. Similarity flooding: A versatile graph matching algorithm. In Proc. 18th Int. Conf. of Data Engineering (ICDE), pages 117–128, 2002.

[6]. I. Niles and A. Pease. Towards a standard upper ontology. In Proc. Int. Conf. on Formal Ontology in Information Systems, FOIS '01, pages 2–9, New York, NY, USA, 2001. ACM.

[7]. B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[8]. F. Shi, J. Li, J. Tang, G. Xie, and H. Li. Actively learning ontology matching via user interaction. In Proc. 8th Int. Semantic Web Conf., pages 585–600, 2009.

[9]. H.-V. To, R. Ichise, and H.-B. Le. An adaptive machine learning framework with user interaction for ontology matching. In Proc. Workshop on Information Integration on the Web, pages 35–40, 2009

92