

An Evolutionary Algorithm based Ontology Alignment Extracting Technology

Xingsi Xue

school of computer science
University of Birmingham
Selly Oak, Birmingham B15 2TT, UK
jack8375@gmail.com

Aihong Ren

Department of Mathematics
Baoji University of Arts and Sciences
No.1 Hi-Tech Avenue, Baoji, Shaanxi, 721013, China
raih2003@hotmail.com

Received February 2017; revised May 2017

ABSTRACT. *Ontology matching is able to identify correspondences between heterogeneous ontology entities. During ontology matching process, since different ontology matchers do not necessarily find the same correct correspondences, usually several competing matchers are applied to the same pair of entities in order to increase evidence towards a potential match or mismatch. How to select, combine and tune various ontology matchers to obtain the high quality ontology alignment is one of the main challenges in ontology matching domain. Recently, Evolutionary Algorithms (EAs) are appearing as another suitable methodology to determine the optimal aggregating weights for the matchers. However, existing EA based ontology matching approaches regard various ontology matchers as the black boxes, and try to determine the optimal weights to aggregate their output. Ignoring the effects brought about by different entity mapping's preference on different matchers could significantly reduce the quality of ontology alignment. Moreover, weights tuned in this way could be problem specific, which means they might not be reused in other ontology matching scenarios. In this paper, we present an EA based ontology alignment extracting technology, which can directly extract the ontology alignment from different matchers' alignments without tuning their aggregating weights. The experiment is carried out on the bibliographic track of OAEI 2016, and the statistical comparisons with three EA based ontology matching approaches show that our approach is effective to match various heterogeneous ontologies.*

Keywords: ontology alignment extracting technology, Evolutionary Algorithm, OAEI

1. **Introduction.** Ontology is regarded as the kernel technology to solve the data heterogeneous problem in semantic web. However, different ontology engineers might use different ways to describe the same domain concepts, causing ontology heterogeneous problem, which raises the heterogeneous problem to a higher level. Ontology matching is a ground solution to face the ontology heterogeneous problem, which can identify correspondences between heterogeneous ontology entities. During ontology matching process, since different ontology matchers do not necessarily find the same correct correspondences, usually several competing matchers are applied to the same pair of entities in order to increase evidence towards a potential match or mismatch [1]. How to select, combine

and tune various ontology matchers to obtain the high quality ontology alignment is one of the main challenges in ontology matching domain [2]. Recently, Evolutionary Algorithms (EAs) are appearing as another suitable methodology to determine the optimal aggregating weights for the matchers. GOAL [3] is the first ontology matching system that utilizes EA to determine the weight configuration for a weighted average aggregation of several matchers by considering a reference alignment. Similar idea of combining multiple ontology matchers is also developed in [4] and [5]. More recently, Xue et al. present an EA based ontology matching framework [6], a segment-based approach for EA based large scale ontology matching [18], and the Multi-Objective EA (MOEA) based approaches to determine the optimal weights being assigned to different matchers [8, 9]. All these methods dedicate to tune the aggregating weights of different matchers, and ignores the effects brought about by different entity mappings' preferences on different matchers, which could decrease the quality of ontology alignment.

However, existing EA based ontology matching approaches regard various ontology matchers as the black boxes, and try to determine the optimal weights to aggregate their output. Ignoring the effects brought about by different entity mapping's preference on different matchers could significantly reduce the quality of ontology alignment. Moreover, weights tuned in this way could be problem specific, which means they might not be reused in other ontology matching scenarios. In this paper, we present an EA based ontology alignment extracting technology, which can directly extract the ontology alignment from different matchers' alignments without tuning their aggregating weights. In particular, we propose a framework of EA based ontology alignment extracting technology, construct a novel optimal model for ontology matching problem, propose a new similarity measure and design a problem specific EA for automatically extract the optimal ontology alignment from various alignments obtained by ontology matchers.

The rest of the paper is organized as follows: Section 2 describes the ontology matching problem; Section 3 presents the EA based ontology extracting technology; Section 4 presents the experimental studies and analysis; finally, Section 5 draws the conclusions.

2. Ontology and Ontology Alignment Extraction. In this work, an ontology O is defined as $O = (C, P, I, \Lambda, \Gamma)$ [10], where C, P, I, Λ, Γ are respectively referred to class set, property set, instance set, axiom set and annotation set. In addition, an ontology alignment A between two ontologies is a correspondence set, and each correspondence is a 4-tuples $(e_1, e_2, confidence, relation)$, where e_1 and e_2 are respectively the entities of two ontology, $confidence \in [0, 1]$ is a confidence value for the correspondence between e_1 and e_2 and $relation$ is the relation of equivalence.

Given an source ontology entity e_{src} , Table 1 shows its potential mapping e_{tgt} , threshold for mapping (e_{src}, e_{tgt}) and the corresponding similarity values under different ontology matchers. In this work, measuring the similarity between the value of $e_i, i = 1, 2, \dots, e_{|O_{tgt}|}$ under $Matcher_j, j = 1, 2, \dots, n$, has to take into consideration the degree of satisfaction of the threshold and the difficulty of achieving its value. Thus, in this work, we use a factor that considers whether the threshold is surpassed or not, and a penalty factor which penalizes those thresholds that are easier to be satisfied. The formula of it is defined as follows:

$$sim(e_i, Matcher_j) = simFactor(e_i, Matcher_j) \times penaltyFactor(threshold_i) \in [0, 1] \quad (1)$$

The calculation of similarity factor is based on the distance between sim_{ij} , which is the similarity evaluation of mapping e_{src} with $e_i, i = 1, 2, \dots, m$ under $Matcher_j, j = 1, 2, \dots, n$, and $threshold_i$. The more the value of $Matcher_j$ in evaluation e_i overcomes the corresponding $threshold_i$, the greater the similarity value shall be. Specifically, if the

TABLE 1. Similarity value on a source ontology entity e_{src} 's potential mappings under different matchers.

target ontology entity	threshold	$Matcher_1$	$Matcher_2$	\dots	$Matcher_n$
e_1	0.3	0.3	0.2	\dots	0.0
e_2	0.0	0.0	0.1	\dots	0.4
\dots	\dots	\dots	\dots	\dots	\dots
e_m	0.5	0.5	0.5	\dots	0.4

TABLE 2. Similarity factor when mapping an source ontology entity e_{src} with potential target ontology entity under different matchers.

target ontology entity	threshold	$Matcher_1$	$Matcher_2$	\dots	$Matcher_n$
e_1	0.3	1/4	-1/3	\dots	-1
e_2	0.0	1/6	2/6	\dots	5/6
\dots	\dots	\dots	\dots	\dots	\dots
e_m	0.5	1	1	\dots	-1/5

difference $diff_{ij} = (sim_{ij} - threshold_i)$ is equal or greater than 0, we assign a positive similarity in $(0,1]$ that depends on the maximum difference $maxDiff_i = (maxSim_i - threshold_i)$ we can achieve with the given threshold; and if the difference $diff$ is lower than 0, we give a negative similarity in $[-1,0)$, punishing the distance of the value with the threshold. To sum up, the normalized similarity value of mapping (e_{src}, e_i) under $Matcher_i$ can be calculated as follows:

$$simFactor(e_i, Matcher_j) = \begin{cases} (1 + diff_{ij}) / (1 + maxDiff_i), & \text{if } diff \geq 0 \\ (diff_{ij}) / (threshold_i), & \text{otherwise} \end{cases} \quad (2)$$

Table 2 shows the similarity factor corresponding to Table 1.

The penalty factor reflects how difficult it is to overcome the given thresholds. The more difficult to surpass a threshold, the lower the penalty value shall be, and its formula is given as follows: $penaltyFactor(threshold_i) = \frac{threshold_i}{maxSim_i}$. Finally, the similarity value of mapping (e_{src}, e_i) is equal to $\frac{1}{n} \sum_{j=1}^n sim(e_i, Matcher_j)$.

On this basis, supposing the golden alignment is one to one, i.e. one entity in source ontology is matched with only one entity in target ontology and vice versa, the optimal model of ontology matching problem can be defined as follows:

$$\begin{cases} \max & f(X) \\ s.t. & X = (x_1, x_2, \dots, x_{|O_{src}| \times |O_{tgt}|})^T \\ & x_i \in [0, 1], i = 1, 2, \dots, |O_{src}| \times |O_{tgt}| \end{cases} \quad (3)$$

where X is the threshold set for each potential entity mappings, function $f()$ calculates the MatchFmeasure [13] of the ontology alignment that is determined by X , $|O_{src}|$ and $|O_{tgt}|$ respectively represent the cardinalities of source ontology O_{src} and target ontology O_{tgt} .

3. Evolutionary Algorithm based Ontology Alignment Extraction. In this work, we propose a framework of EA based ontology alignment extracting technology to automatically determine the optimal ontology alignment, which is shown in Figure 1. As can be seen from the figure, it consists of three main phases:

TABLE 3. Brief description on benchmark. 1XX, 2XX and 3XX stands for the test case whose ID beginning with the prefix digit 1, 2 and 3, respectively.

ID	Description
1XX	The ontologies under alignment are the same or the first one is the OWL Lite restriction of the second one
2XX	The ontologies under alignment have different lexical, linguistic or structure features
3XX	The ontologies under alignment are real world cases

- Similarity matrices generation: it's the pre-matching phases, which use ontology matcher to determine corresponding similarity matrix whose row and column respectively represent the source ontology and target ontology's entities, and element inside is the similarity evaluation on that entity pair under this matcher,
- EA based ontology alignment extraction: it's the kernel phases, where a problem specific EA is utilized to determine the optimal threshold set, and automatically extract the ontology alignment from various similarity matrices,
- Ontology alignment evaluation: it's the post-matching phases, where we evaluate the optimal alignment through f-measure [12].

In the following, we present three main components of EA, i.e. chromosome encoding mechanism, evolutionary operator and elitist strategy.

3.0.1. *Chromosome Encoding Mechanism.* In this work, we use the binary encoding mechanism. The information encoding inside an individual consists of $|O_{src}| \times |O_{tgt}|$ thresholds. When decoding an individual, we first use the thresholds to calculate the similarity value for $|O_{src}| \times |O_{tgt}|$ pairs of mapping, and then for each entity in source ontology, we select a target entity with the largest similarity value to obtain the ontology alignment.

3.1. **Evolutionary Operator.** In this paper, in order to balance the diversity of the population and the convergence of the algorithm, the selection operator first sorts the chromosomes of population in descending order according to their fitness value. Then we select half of the chromosomes in the front of the population and randomly copy one each time until forming a new population. We use the common one point crossover operator. For each bit in the chromosome we check if the mutation could be applied according to the mutation probability and if it is, the value of that bit is then flipped.

3.2. **Elitist Strategy.** Elitist strategy copies the best chromosome (elite) of the current population unaltered to the next population. In our work, we regard the individual with the highest fitness value as the elite of current generation. When the algorithm terminates, the elite will be recommended to the user as the optimal solution.

4. **Experimental Studies and Analysis.** In the experiment, we utilize the bibliographic track of OAEI 2016 [14] to test our approach's performance. Table 3 shows a brief description on the benchmark of OAEI 2016.

4.1. **Experimental Setup.** Ontology matcher takes as input two ontologies O_{src} and O_{tgt} and output a $|O_{src}| \times |O_{tgt}|$ similarity matrix S , whose element s_{ij} is the similarity score between i th entity in $|O_{src}|$ and j th entity in $|O_{tgt}|$. In our work, we utilize the following four basic similarity matchers:

- Syntactic-based Matcher [15],
- Linguistic-based Matcher [16],

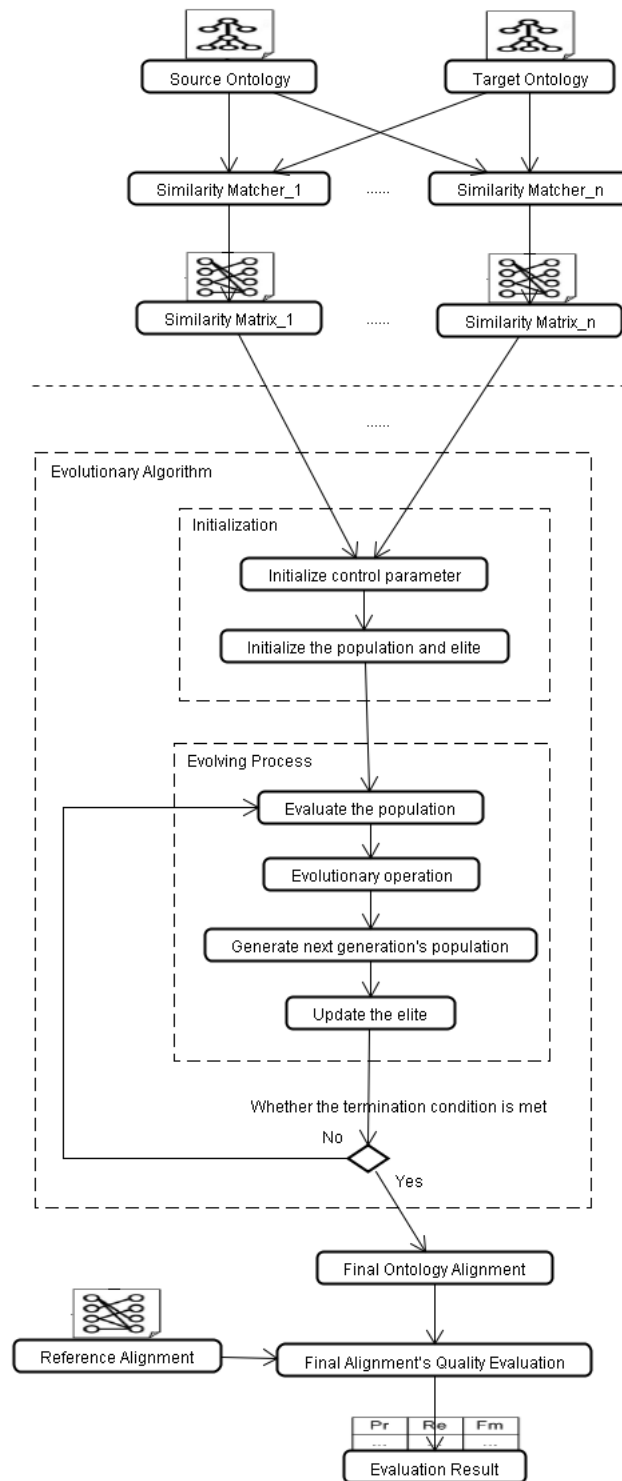


FIGURE 1. The framework of Evolutionary Algorithm based ontology alignment extracting technology

TABLE 4. Friedman’s test on the alignment’s quality obtained by three EA based ontology matching approaches and our approach.

ID	GA	MA	PSO	Our approach
101	1.00 (2.5)	1.00 (2.5)	1.00 (2.5)	1.00 (2.5)
103	0.99 (4)	1.00 (2)	1.00 (2)	1.00 (2)
104	0.99 (4)	1.00 (2)	1.00 (2)	1.00 (2)
201	0.50 (3)	0.62 (2)	0.42 (4)	0.85 (1)
203	0.97 (3)	0.96 (4)	1.00 (1)	0.98 (2)
204	0.94 (4)	0.97 (3)	0.98 (1.5)	0.98 (1.5)
205	0.83 (2)	0.79 (3)	0.73 (4)	0.85 (1)
206	0.84 (4)	0.88 (2)	0.85 (3)	0.89 (1)
221	0.99 (3.5)	0.99 (3.5)	1.00 (1.5)	1.00 (1.5)
222	0.99 (3)	0.99 (3)	0.99 (3)	1.00 (1)
223	0.99 (2.5)	0.99 (2.5)	0.99 (2.5)	0.99 (2.5)
224	1.00 (2.5)	1.00 (2.5)	1.00 (2.5)	1.00 (2.5)
225	0.99 (4)	1.00 (2)	1.00 (2)	1.00 (2)
228	0.99 (3)	0.99 (3)	0.99 (3)	1.00 (1)
230	0.93 (3.5)	0.93 (3.5)	0.98 (2)	1.00 (1)
231	0.99 (3)	0.99 (3)	0.99 (3)	1.00 (1)
301	0.70 (2.5)	0.70 (2.5)	0.64 (4)	0.73 (1)
302	0.61 (3)	0.63 (2)	0.04 (4)	0.71 (1)
304	0.83 (3)	0.87 (2)	0.72 (4)	0.88 (1)
Average	0.90 (3.16)	0.91 (2.63)	0.86 (2.71)	0.94 (1.5)

- Structure-based Matcher [17],
- Instance-based Matcher [18].

The EA uses the following control parameters:

- Numerical accuracy = 0.01,
- Population scale = 20,
- Crossover probability = 0.6,
- Mutation probability = 0.01,
- Maximum generation = 300.

4.2. Results and Analysis. All the experimental results in the tables are the average values over ten independent runs. Specifically, Tables 4 and 5 show the statistical comparison among three EA based ontology matching approaches and our approach. we carry out the statistical comparison on the alignment’s quality in terms of f-measure among Genetic Algorithm (GA) based [3], Memetic Algorithm (MA) based [10], Particle Swarm Optimization (PSO) based [11] ontology matching approaches and our approach, and these compared approaches’ configurations are referred to their corresponding references.

In Table 4, each value represents the f-measure, and the number in round parentheses is the corresponding computed rank. For example, in test case 302, four approaches’ rank values are decided by their f-measure values (the higher the better); in test case 101, since four approaches’ f-measure values are the same, their ranks are calculated by $(1+2+3+4)/4=2.5$.

As can be seen from Table 4, in the Friedmans test, the computed \mathcal{X}_r^2 value is 17.06, which is greater than $\mathcal{X}_{0.05}^2 = 7.81$. In the Holm’s test, as shown in Table 5, our approach statistically outperforms other EA based ontology matching approaches on the alignment’s quality at 0.05 significance level.

TABLE 5. Holm's test on the alignment's quality obtained by three EA based ontology matching approaches and our approach.

i	approach	z value	unadjusted p -value	$\frac{\alpha}{k-i}, \alpha = 0.05$
3	MA	2.69	0.0070	0.05
2	PSO	2.88	0.0039	0.02
1	GA	3.96	0.000073	0.01

5. Conclusion. One of the challenge problems in ontology matching domain is how to select, combine and tune different ontology matchers to obtain the high quality ontology alignment. To face this challenge, in this paper, we present an novel EA based ontology extracting technology, which can automatically extract ontology alignment without tuning the aggregating weights. The experimental results show that our approach can significantly improve the quality of existing EA based ontology matching approaches.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (No. 61503082), Natural Science Foundation of Fujian Province (No. 2016J05145), Scientific Research Startup Foundation of Fujian University of Technology (No. GY-Z15007), Fujian Province Outstanding Young Scientific Researcher Training Project (No. GY-Z160149) and China Scholarship Council.

REFERENCES

- [1] T. T. A. Nguyen and S. Conrad, Ontology Matching using multiple similarity measures, *Knowledge Discovery, 7th International Joint Conference on Knowledge Engineering and Knowledge Management*, vol.1, pp.603–611, Lisbon, Portugal, 2015.
- [2] P. Shvaiko and J. Euzenat, Ontology matching: state of the art and future challenges, *IEEE Transactions on knowledge and data engineering*, vol.25, no.1, pp.158–176, 2013.
- [3] J. Martinez-Gil, E. Alba and J. F. A. Montes, Optimizing ontology alignments by using genetic algorithms, *Proceedings of the First International Conference on Nature Inspired Reasoning for the Semantic Web*, vol.419, pp.1–15, Berlin, Germany, 2008.
- [4] J. Naya, M. M. Romero, J. P. Loureiro, C. R. Munteanu and A. P. Sierr, Improving ontology alignment through genetic algorithms, *Soft computing methods for practical environment solutions: Techniques and studies*, pp.240–259, 2010.
- [5] G. Alexandru-Lucian and A. Iftene, Using a genetic algorithm for optimizing the similarity aggregation step in the process of ontology alignment, *9th Roedunet International Conference*, pp.118–122, Sibiu, Romania, 2010.
- [6] X. Xue and Z. Y. Tang, An Evolutionary Algorithm based Ontology Matching System, *Journal of Information Hiding and Multimedia Signal Processing*, vol.8, no.14, pp.551–556, 2017.
- [7] X. Xue and J. S. Pan, A segment-based approach for large-scale ontology matching, *Knowledge and Information Systems*, vol.52, no.2, pp.467–484, 2017.
- [8] X. Xue and J. Liu, Optimizing Ontology Alignment Through Compact MOEA/D, *International Journal of Pattern Recognition and Artificial Intelligence*, vol.31, no.4, pp.1–19, 2017.
- [9] X. Xue and Y. Wang, Improving the efficiency of NSGA-II based ontology aligning technology, *Data and Knowledge Engineering*, vol.108, pp.1–14, 2017.
- [10] G. Acampora, V. Loia and A. Vitiello, Enhancing ontology alignment through a memetic aggregation of similarity measures, *Information Sciences*, vol.250, pp.1–20, 2013.
- [11] J. Bock and J. Hettenhausen, Discrete particle swarm optimisation for ontology alignment, *Information Sciences*, vol.192, pp.152–173, 2012.
- [12] C. J. Van Rijsberge, *Information Retrieval*, University of Glasgow, Butterworth, London, 1975.
- [13] X. Xue and Y. Wang, Optimizing Ontology Alignments through a Memetic Algorithm Using both MatchFmeasure and Unanimous Improvement Ratio, *Artificial Intelligence*, vol.223, pp.65–81, 2015.
- [14] Ontology Alignment Evaluation Initiative (OAEI), Available at <http://www.oaei.ontologymatching.org>, Accessed on April 15, 2016.
- [15] G. Stoilos, G. Stamou and S. Kollias, A string metric for ontology alignment, *International Semantic Web Conference*, pp.624–637, Galway, Ireland, 2005.

- [16] G. A. Miller, WordNet: a lexical database for English, *Communications of the ACM*, vol.38, no.11, pp.39–41, 1995.
- [17] S. Melnik, H. Garcia-Molina and E. Rahm, Similarity flooding: A versatile graph matching algorithm and its application to schema matching, *Proceedings of 18th International Conference on Data Engineering*, pp.117–128, Washington DC, USA, 2002.
- [18] X. Xue and Y. Wang, Ontology alignment based on instance using NSGA-II, *Journal of Information Science*, vol.41, no.1, pp.58–70, 2015.