

OPTIMIZING MULTIPLE ONTOLOGY ALIGNMENTS THROUGH HYBRID EVOLUTIONARY ALGORITHM BASED ON A NEW ELITIST STRATEGY

LI JIANG¹, PEI-WEI TSAI^{2,3} AND XINGSI XUE^{2,3,*}

¹Computer Department
Fuzhou Polytechnic
No. 8, Lianrong Road, University Town, Minhou, Fuzhou 350108, P. R. China
86068374@qq.com

²College of Information Science and Engineering

³Fujian Provincial Key Laboratory of Big Data Mining and Applications
Fujian University of Technology
No. 3, Xueyuan Road, University Town, Minhou, Fuzhou 350118, P. R. China
pwtsai@fjut.edu.cn; *Corresponding author: xxs@fjut.edu.cn

Received November 2015; accepted February 2016

ABSTRACT. *In this paper, we propose a hybrid evolutionary algorithm based on a new elitist strategy to address the problem of optimizing the multiple ontology alignments simultaneously. Comparing with the conventional approach, our method is able to deal with multiple pair of ontologies at a time, and through a new elitist strategy, avoid the bias improvement caused by f-measure. Experimental results show that our approach is effective.*

Keywords: Ontology alignment, Hybrid evolutionary algorithm, Elitist strategy

1. Introduction. Ontologies are regarded as the solution to data heterogeneity on the semantic web. However, because of human subjectivity, the ontologies could themselves introduce heterogeneity: given two ontologies, one entity can be given different names or simply be defined in different ways [5]. Addressing this heterogeneity 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 [1].

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. Since none of the similarity measures could provide the satisfactory result independently, most ontology alignment systems combine a set of different similarity measures together by aggregating their aligning results. How to select the appropriate similarity measures, weights and thresholds in ontology aligning process in order to obtain a satisfactory alignment is called meta-matching which can be viewed as an optimization problem and be addressed by evolutionary approaches like Genetic Algorithm (GA) and hybrid evolutionary algorithm such as Memetic Algorithms (MA). Among those meta-matching systems using evolutionary algorithm, the most notable system is GOAL (Genetics for Ontology Alignments) [6]. Although GOAL does not directly compute the alignment between two ontologies, it determines, through a GA, the optimal weight configuration for a weighted average aggregation of several similarity measures by considering a reference alignment. The same idea of implementing a meta-matching system to combine multiple similarity measures into a single aggregated metric is also developed in papers [7, 8]. Inspired by

GA, Acampora et al. employ MA in the alignment problem [5]. Since the MA has the capability of realizing local search process within the successive generations, it improves the performance of genetic approach in both quality of solutions and computational efficiency.

Nevertheless, there are two main drawbacks of current evolutionary approaches which determine the weights by one reference alignment of two ontologies. One of the drawbacks is that it is difficult to deal with several reference alignments of various pairs of ontologies by these approaches. The other one arises from f-measure which is generally used as the evaluation of the alignment's quality. Although there is an implicit consensus among researchers that the f-measure is the best way of combining evaluation metric pairs, it is controversial to rank the quality of individuals produced by MA according to the f-measure only since an overall improvement in f-measure often derives from an improvement in one of the metrics at the expense of a decrement in the other. To overcome these two defects, in this paper, Unanimous Improvement Ratio (UIR) [2], a measure that complements f-measure, is utilized to compare the qualities of individuals produced by MA. Given a set of reference alignments, the aim of this paper is to utilize an MA using both f-measure and UIR to find the optimal set of parameters, such as weights and thresholds, to combine multiple similarity measures into a single aggregated metric and avoid the bias improvement caused by f-measure.

2. Preliminaries.

2.1. Ontology and ontology alignment. There are many definitions of ontology over years. However, 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 as follows [5].

Definition 2.1. *An ontology is a triple $O = (C, P, I)$, where C is the set of classes, P is the set of properties, I is the set of individuals, i.e., the set of objects of the real world, representing the instances of a concept.*

In general, classes, properties and individuals are referred as entities.

To solve the heterogeneity problem between ontologies, a so-called ontology alignment process is necessary. Formally, an alignment between two ontologies can be defined as presented as follows [5].

Definition 2.2. *An alignment between two ontologies is a set of mapping elements. A mapping element is a 5-uple (id, e, e', n, R) , where id is a unique identifier for the mapping, e and e' are the entities of the first and the second ontology respectively, n is a confidence measure in some mathematical structures (typically in the $[0, 1]$ range) holding for the correspondence between the entities e and e' , and R is a relation (typically the equivalence) holding between the entities e and e' .*

The ontology alignment process can be defined as follows [5].

Definition 2.3. *The alignment process can be seen as a function Φ where given a pair of ontologies O and O' , a partial (and optional) input alignment A , a set of parameters p , and a set of resources r , returns a new alignment A' :*

$$A' = \Phi(O, O', A, p, r).$$

The ontology alignment process computes a mapping element by using a similarity measure, which determines the closeness value n (related to a given relation R) between the entities e and e' in the range $[0, 1]$, where 0 stands for complete inequality and 1 for complete equality.

2.2. Aggregation strategy. To combine all the similarity measures mentioned above, an aggregation strategy is needed. In this work, we utilize weighted average aggregation which is defined in the following:

$$\phi(\vec{s}(c), \vec{w}) = \sum_{i=1}^n w_i s_i(c) \quad \text{with} \quad \sum_{i=1}^n w_i = 1 \quad \text{and} \quad w_i \in [0, 1] \quad (1)$$

where $\vec{s}(c)$ is the vector of similarity measure results, \vec{w} is the vector of weights, and n is the number of similarity measures.

Since the quality of resulting alignment, the correctness and completeness of the correspondences found already, need to be assessed, we will introduce some conformance measures which derive from the information retrieval field in the next section.

2.3. F-measure and UIR.

2.3.1. Recall, precision and f-measure. The alignment is normally assessed on the basis of two measures commonly known as recall and precision. Recall (or completeness) measures the fraction of correct alignments found in comparison to the total number of correct existing alignments. Typically, recall is balanced against precision (or correctness), which measures the fraction of found alignments that are actually correct. Given a reference alignment R and some alignment A , recall and precision are given by the following formulas [3]:

$$recall = \frac{|R \cap A|}{|R|} \quad (2)$$

$$precision = \frac{|R \cap A|}{|A|} \quad (3)$$

In most instances, it requires considering both recall and precision to compare alignments' performance. The most common combining function is the f-measure which is defined as follows:

$$f\text{-measure} = \frac{recall \cdot precision}{\alpha \cdot recall + (1 - \alpha) \cdot precision} \quad (4)$$

where α is the relative weight of recall and precision which is in the range $[0, 1]$. When $\alpha = 0$ or 1 , f-measure can be transformed into recall or precision; when $\alpha = 0.5$, both recall and precision have the same relative weight, and f-measure computes their harmonic mean.

2.3.2. UIR. A problem of f-measure is that the relative weight is established intuitively for a given task, but at the same time a slight change in the relative weight may produce substantial changes in the rankings of individuals' quality. If individual A improves another individual B in precision with a loss in recall, f-measure may say that A is better than B , depending on the relative weight α of precision and recall. Therefore, it is controversial to rank the individuals' quality according to the f-measure only. To overcome the shortcoming of f-measure, we employ the UIR which is a measure that allows to compare two individuals using recall and precision without dependency on the relative weight α in f-measure. For two individuals A and B , UIR can be given by the following formula:

$$UIR(A, B) = \frac{(|T_A| - |T_B|)}{|T|} \quad (5)$$

where T_A is the set of cases for which individual A achieves precision and recall that are greater than or equal to that of individual B ; T_B is the set of cases for which individual B achieves precision and recall that are greater than or equal to that of individual A ; T is the set of all cases.

In this work, we set $\alpha = 0.5$ to favor neither precision nor recall. To estimate which individuals' performance differences in f-measure where $\alpha = 0.5$ are robust against different α values, we employ the rule of thumb in [2]: if $UIR(A, B) > 0.25$, then an observed improvement of individual A over B in f-measure where $\alpha = 0.5$ is robust.

The main advantage of UIR is that no metric weighting is necessary. However, there remain two main limitations of UIR. First, as well as the unanimous improvement, it is not transitive [2]. Therefore, it is not possible to define a linear individual ranking based on UIR. In addition, there is some information lost when comparing systems given that the ranges in evaluation results are not considered. Thus, we propose an MA utilizing both the f-measure and UIR for optimizing multiple reference ontologies' alignments.

3. MA Based on a New Elitist Strategy.

3.1. Encoding mechanism. We incorporate in an individual both the weights associated with the similarity measures and the threshold to decide whether a pair of entities is an alignment or not. Therefore, one individual can be divided into two parts where one stands for several weights and the other for threshold. Concerning the characteristics of the weights which are mentioned in Section 2.2, our encoding mechanism indirectly represents them by defining the cut or separation point in the interval $[0, 1]$ that limits the value of the weights. If p is the number of weights required, the set of cuts can be represented as $c' = \{c'_1, c'_2, \dots, c'_{p-1}\}$. The individual decoding is carried out by queuing the elements of c' in ascending order, then we get $c = \{c_1, c_2, \dots, c_{p-1}\}$, and calculate the weights as follows:

$$w_k = \begin{cases} c_1, & k = 1 \\ c_k - c_{k-1}, & 1 < k < p \\ 1 - c_{p-1}, & k = p \end{cases} \quad (6)$$

3.2. Fitness functions. Fitness functions are objective functions that evaluate the quality of the alignments obtained by using the weights and the threshold encoded in the individual. In our work, the fitness function should take into account the quality of all the alignments. This is done by means of the average of all the f-measures for each reference alignment as follows:

$$fitness = \frac{f-measure_1 + f-measure_2 + \dots + f-measure_n}{n} \quad (7)$$

where n is the number of reference alignments.

3.3. A new elitist strategy. Elitist strategy puts the best individual (elite) of the current population unaltered in the next population. This assures the survival of the elite that has been obtained up to the moment. In our work, we utilize f-measure and UIR to obtain the elite of current generation. First, we calculate each individual's reference individual which is defined as follows: given an individual a , its reference individual $I_{ref}(a)$ is the one that improves a with maximal UIR:

$$I_{ref}(a) = Arg \max(UIR(I, a)) \quad (8)$$

In other words, $I_{ref}(a)$ represents the individual with which a should be replaced in order to robustly improve across f-measures with different α values. And the one with the maximal value of being the reference of other individuals has better quality than the others. However, there may be several individuals with the same maximal number of being the reference for other individuals. Then a baseline individual, whose weights are all 0.25 and the threshold is 0.80, is taken. Comparing with baseline individual, the individual with larger UIR is regarded as the one that has the better quality. Finally, if there still exist some individuals whose quality cannot be winnowed via UIR, the one with largest f-measure will be taken as the best individual of the current population.

3.4. Local search process. In general, the local search strategies perform iterative search for optimum solution in the neighborhood of a candidate. In order to trade off between the local search and the global search, the local search process in our work is designed according to the following rules: (1) the local search is applied within each evolutionary cycle; (2) the local search is executed after crossover and mutation; (3) the local search is applied to the best individual of population; (4) the local search method is the hill climbing algorithm.

In particular, the hill climbing algorithm is a local search iterative method. During iterations, the algorithm attempts to find a better individual by randomly mutating the current one. If a mutation improves the current individual, then the new individual replaces the current one. The search is repeated until no further improvement can be found or after a maximum number of iterations.

4. Experimental Results and Analysis. In the experiments, the well-known benchmarks provided by the Ontology Alignment Evaluation Initiative (OAEI) 2011 [4] are used. Each benchmark in the OAEI data set is composed of two ontologies to be aligned and a reference alignment to evaluate the quality of alignment. Moreover, according to OAEI policies, the benchmark reference alignments take into account only the matching between ontology classes and properties. Table 1 shows a brief description about the benchmarks of OAEI 2011.

TABLE 1. Brief description of benchmarks

ID	Brief description
101-104	The ontologies under alignment are the same or the first one is the OWL Lite restriction of the second one
201-210	The ontologies under alignment have the same structure, but different lexical and linguistic features
221-231	The ontologies under alignment have the same lexical and linguistic features, but different structures
301-304	The ontologies under alignment are real world cases

4.1. Experiments configuration. In our experiment, the MA uses the following parameters: (1) Search space for each parameter is the continuous interval $[0, 1]$; (2) Numerical accuracy = 0.01; (3) Population size = 20 individuals; (4) Crossover probability = 0.6; (5) Mutation probability = 0.01; (6) Max generation = 5. After ten independent executions, we noticed that the genetic algorithm does not improve the results beyond the fifth generation, so we have set a limit of five generations.

4.2. Results and analysis. Table 2 and Table 3 show the recall, precision and f-measure values of solutions obtained by the baseline proposal, MA using f-measure only and MA using f-measure and UIR by one reference alignment and multiple reference alignments, respectively. In Table 2 and Table 3, R , P and F refer to recall, precision and f-measure values and the combinations of benchmarks in Table 3 are selected randomly. Table 4 and Table 5 present each solution's UIR values and average f-measures, which are calculated through the values in Table 2 and Table 3, where S_A and S_B refer to the solutions obtained by MA using f-measure only and those by MA using f-measure and UIR, respectively.

As can be seen from Table 4, S_B is better than S_A in benchmarks 204, 205, 223, 301 and 302. While among the other benchmarks, the qualities of solutions are alike. From the second column, the values of last two rows show the apparent improvement of S_B over S_A . Although it is unable to distinguish the quality of solutions obtained by two approaches from the second column among the rest benchmarks, the values in the third

TABLE 2. Comparison of the results obtained by one reference alignment

ID	Baseline (R, P, F)	MA using f-measure only (R, P, F)	MA using f-measure and UIR (R, P, F)
101	(0.00, 1.00, 0.00)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)
103	(0.96, 1.00, 0.98)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)
104	(0.99, 1.00, 0.99)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)
203	(0.80, 1.00, 0.89)	(0.98, 1.00, 0.99)	(0.98, 1.00, 0.99)
204	(0.69, 1.00, 0.82)	(0.98, 0.99, 0.98)	(0.91, 1.00, 0.95)
205	(0.21, 1.00, 0.34)	(0.89, 0.99, 0.93)	(0.37, 1.00, 0.54)
221	(0.98, 1.00, 0.99)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)
222	(0.98, 1.00, 0.99)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)
223	(0.96, 1.00, 0.98)	(0.98, 1.00, 0.98)	(0.98, 1.00, 0.99)
224	(0.00, 1.00, 0.00)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)
225	(0.97, 1.00, 0.98)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)
228	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)
301	(0.29, 1.00, 0.45)	(0.76, 0.74, 0.75)	(0.34, 1.00, 0.51)
302	(0.21, 1.00, 0.34)	(0.63, 0.83, 0.71)	(0.40, 1.00, 0.57)

TABLE 3. Comparison of the results obtained by multiple reference alignments

Case No.	ID	Baseline (R, P, F)	MA using f-measure only (R, P, F)	MA using f-measure and UIR (R, P, F)
1	101	(0.00, 1.00, 0.00)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)
	103	(0.96, 1.00, 0.98)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)
2	104	(0.99, 1.00, 0.99)	(1.00, 0.98, 0.99)	(1.00, 1.00, 1.00)
	203	(0.80, 1.00, 0.89)	(0.97, 1.00, 0.99)	(0.86, 1.00, 0.92)
3	204	(0.69, 1.00, 0.81)	(0.98, 0.99, 0.98)	(0.94, 0.99, 0.96)
	205	(0.21, 1.00, 0.34)	(0.89, 0.99, 0.93)	(0.45, 1.00, 0.62)
4	204	(0.69, 1.00, 0.81)	(1.00, 1.00, 1.00)	(1.00, 1.00, 1.00)
	221	(0.23, 1.00, 0.37)	(0.49, 0.87, 0.63)	(0.67, 0.73, 0.70)
5	228	(1.00, 1.00, 1.00)	(1.00, 0.89, 0.94)	(1.00, 1.00, 1.00)
	301	(0.29, 1.00, 0.45)	(0.66, 0.75, 0.70)	(0.37, 1.00, 0.54)
6	301	(0.29, 1.00, 0.45)	(0.80, 0.70, 0.75)	(0.41, 0.96, 0.57)
	302	(0.21, 1.00, 0.34)	(0.63, 0.75, 0.68)	(0.40, 1.00, 0.57)
7	101	(0.00, 1.00, 0.00)	(1.00, 0.94, 0.97)	(1.00, 0.98, 0.99)
	221	(0.98, 1.00, 0.99)	(1.00, 0.94, 0.97)	(1.00, 1.00, 1.00)
	301	(0.29, 1.00, 0.45)	(0.64, 0.79, 0.71)	(0.44, 0.96, 0.60)
8	103	(0.96, 1.00, 0.98)	(1.00, 0.98, 0.99)	(1.00, 1.00, 1.00)
	223	(0.96, 1.00, 0.98)	(0.99, 0.94, 0.96)	(0.98, 1.00, 0.99)
	302	(0.21, 1.00, 0.34)	(0.54, 0.90, 0.68)	(0.40, 1.00, 0.57)
9	104	(0.99, 1.00, 0.99)	(1.00, 0.98, 0.99)	(1.00, 1.00, 1.00)
	225	(0.97, 1.00, 0.98)	(1.00, 0.92, 0.96)	(1.00, 1.00, 1.00)
	301	(0.29, 1.00, 0.45)	(0.63, 0.77, 0.69)	(0.24, 1.00, 0.38)

and the fourth columns show that S_B is better than S_A in benchmarks 204 and 205. In the fifth and the sixth columns, the values of benchmark 223 show the priority of S_B . To conclude, through the comparison of the results obtained by one reference alignment by two approaches, MA using both f-measure and UIR is effective.

As Table 5 shows, except case number 1, S_B is better than S_A in the rest cases. From the third column, S_B is apparently better than S_A in case number 5, 7, 8 and 9. While

TABLE 4. Comparison of UIR values of the results obtained by one reference alignment

ID	$UIR(S_B, S_A)$	$UIR(S_A, Baseline)$	$UIR(S_B, Baseline)$	fitness of S_A	fitness of S_B
101	0.00	1.00	1.00	1.00	1.00
103	0.00	1.00	1.00	1.00	1.00
104	0.00	1.00	1.00	1.00	1.00
203	0.00	1.00	1.00	0.99	0.99
204	0.00	0.00	1.00	0.98	0.95
205	0.00	0.00	1.00	0.93	0.54
221	0.00	1.00	1.00	1.00	1.00
222	0.00	1.00	1.00	1.00	1.00
223	0.00	1.00	1.00	0.98	0.99
224	0.00	1.00	1.00	0.99	0.99
225	0.00	1.00	1.00	1.00	1.00
228	0.00	0.00	0.00	1.00	1.00
301	1.00	0.00	1.00	0.75	0.51
302	1.00	0.00	1.00	0.71	0.57

TABLE 5. Comparison of UIR values of the results obtained by multiple reference alignments

Case No.	ID	$UIR(S_B, S_A)$	$UIR(S_A, Baseline)$	$UIR(S_B, Baseline)$	fitness of S_A	fitness of S_B
1	101,103	0.00	1.00	1.00	1.00	1.00
2	104,203	0.00	0.50	1.00	0.99	0.96
3	204,205	0.00	0.00	0.50	0.96	0.79
4	204,221	0.00	0.50	0.50	0.82	0.85
5	228,301	0.50	-0.50	0.50	0.82	0.77
6	301,302	0.00	0.00	0.50	0.71	0.57
7	101,221,301	0.67	0.00	0.33	0.88	0.86
8	103,223,302	0.33	0.00	1.00	0.88	0.85
9	104,225,301	0.67	0.00	0.33	0.88	0.79

according to the values in the fourth and fifth columns, S_B is apparently better than S_A in case number 2, 3 and 6. Finally, by comparing the fitness of S_B with that of S_A , S_B is better than S_A in case number 4. To sum up, through the comparison of the results obtained by multiple reference alignments using two approaches, MA using both f-measure and UIR is much better than MA using f-measure only.

5. Conclusions. Ontology alignment is an important step in ontology engineering. In this paper, a novel approach based on MA using both f-measure and UIR has been proposed to aggregate different similarity measures into a single metric, and optimize the quality of multiple reference ontologies' alignments. The experimental results have shown that the MA based on a new elitist strategy is effective to automatically configure the parameters of similarity aggregation process and our approach could deal with multiple reference ontologies' alignments and avoid the bias improvement caused by f-measure.

In continuation of our research, the work is now being done on embedding MA using both f-measure and UIR into a real ontology alignment system. We are also interested in developing an Expert Decision Support System to help the ontology alignment system to automatically decide the parameters and even which similarity measures should be utilized.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (Nos. 61503082 and 61402108) and Natural Science Foundation of Fujian Province (No. 2016J05145).

REFERENCES

- [1] P. Shvaiko and J. Euzenat, Ontology matching: State of the art and future challenges, *IEEE Trans. Knowledge and Data Engineering*, vol.25, no.1, pp.158-176, 2013.
- [2] E. Amigo, J. Gonzalo, J. Artilles and M. F. Verdejo, Combining evaluation metrics via the unanimous improvement ratio and its application to clustering tasks, *Journal of Artificial Intelligence Research*, vol.42, pp.689-718, 2011.
- [3] C. J. Van Rijsbergen, Foundation of evaluation, *Journal of Documentation*, vol.34, pp.365-373, 1974.
- [4] *Ontology Alignment Evaluation Initiative (OAEI)*, <http://oaei.ontologymatching.org/2011/>, 2015.
- [5] G. Acampora, V. Loia, S. Salerno and A. Vitiello, A hybrid evolutionary approach for solving the ontology alignment problem, *International Journal of Intelligent Systems*, vol.3, no.27, pp.189-216, 2013.
- [6] J. Martinez-Gil, E. Alba and J. F. Aldana-Montes, Optimizing ontology alignments by using genetic algorithms, *Nature Inspired Reasoning for the Semantic Web (NatuReS2008)*, vol.419, pp.31-45, 2008.
- [7] J. M. V. Naya, M. M. Romero and J. P. Loureiro, Improving ontology alignment through genetic algorithms, *Soft Computing Methods for Practical Environment Solutions: Techniques and Studies*, pp.1-16, 2010.
- [8] A.-L. Ginsca and A. Iftene, Using a genetic algorithm for optimizing the similarity aggregation step in the process of ontology alignment, *The 9th Roedunet Int Conf. (RoEduNet)*, pp.118-122, 2010.