Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **1**, January **2018**

An Overview on Evolutionary Algorithm based Ontology Matching

Xingsi Xue and Jeng-Shyang Pan
College of Information Science and Engineering
Fujian Provincial Key Laboratory of Big Data Mining and Applications (Fujian University of Technology)
Fujian University of Technology
No.3 Xueyuan Road, University Town, Minhou, Fuzhou, Fujian, 350118, China
Corresponding Author:jack8375@gmail.com

Received March, 2017; revised July, 2017

Abstract. *Evolutionary Algorithm (EA) based ontology matching technologies are emerging as the most suitable approach for solving ontology matching problem. In this paper, we introduce the basics of EA based ontology matching technology, and discuss the state-of-the-art EA based ontology matching approach. EA based ontology matching is making a measurable progress, though it is slowing down. In order to address this situation, we presented four challenges for EA based ontology matching, accompanied for each of these with an overview of the recent advances in the field and a discussion of the potentially useful ways to approach the challenges under consideration. We believe that addressing the outlined challenges should accelerate the progress of the EA based ontology matching field and direct the corresponding research into the most promising tracks.*

Keywords: Evolutionary Algorithm, Ontology matching, Ontology alignment

---

1. **Introduction.** Ontology is an explicit specification of a conceptualization [1], i.e. the formal specification of the objects, concepts, and other entities that are presumed to exist in some area of interest and the relationships that hold them. Ontology is the main component of Semantic Web [2], which is regarded as the solution to data heterogeneity on the Semantic Web. However, different tasks or different points of view lead ontology designers to produce different conceptualizations of the same domain of interest. The subjectivity of the ontology modeling results in the creation of heterogeneous ontologies, which is characterized by terminological and conceptual discrepancies. Examples of these discrepancies are the use of different words to name the same concept, the use of the same word to name different concepts, the creation of hierarchies for a specific domain region with different levels of detail and so on. The arising so-called semantic heterogeneity problem poses as a barrier to achieve the semantic collaboration on the ontology level among various Semantic Web applications. Ontology matching is a ground solution to the ontology heterogeneity problem, which is able to identify the correspondences between semantically related entities of ontologies [3]. The increasing relevance of performing an ontology alignment process in several domains of application such as knowledge management, information retrieval, medical diagnosis, e-Commerce, knowledge acquisition, search engines, bioinformatics, the emerging Semantic Web and so on, have led to the development of many diverse fully automatic or semi-automatic ontology matching technologies. See [4] for some contributions of the last decades and [5] for recent surveys. However, despite the many matching technologies that have been developed so far, there

is no integrated technology that is a clear success or robust enough to be the basis for future development.

In general, ontology matching technology corresponds to finding an isomorphism between the sub-graphs [6]. Since modeling the two ontologies under alignment is a complex (nonlinear problem with many local optimal solutions) and time-consuming task (large scale problem), particularly when the considered ontologies are characterized by a significant number of entities (resulting in large scale problem), approximate methods are usually used for computing the correspondence. From this point of view, evolutionary optimization methods could represent an efficient approach for addressing this problem. Evolutionary Algorithm (EA) is an evolutionary based stochastic optimization algorithm, which was proposed by John Holland [7]. It's considered to be a flexible and robust technique, which can deal successfully with a wide range of difficult optimization problems and they are generally good at finding acceptably good solutions to problems acceptably quickly. Recently, among the existing ontology matching technologies, those based on EA, are appearing as the most suitable methodology to face the ontology matching problem [8]. In general, EA based ontology matching technology can be classified into two categories. First class solves the ontology meta-matching problem, which dedicates to optimize ontology matching system's parameters, e.g. weights and threshold, for aggregating different ontology entity matchers' matching results. While the second kind deals with the ontology entity matching problem, which tends to directly determine the optimal entity correspondence set with the given ontology entity matcher. In each category, both single-objective and multi-objective EAs based approaches have been designed and utilized to obtain high quality ontology alignments. In this paper, we discuss the main trends in the EA based ontology matching domain, and overview the recent advances in this field. This should direct research into the critical path and accelerate progress of EA based ontology matching field.

The remainder of the paper is organized as follows: Section 2 presents the basic concepts in EA based ontology matching domain; Section 3 and Section 4 respectively overview the EA based technology for ontology meta-matching problem and ontology entity matching problem; Section 5 shows four challenges in EA based ontology matching domain; and finally, Section 6 draws the conclusions.

2. **Preliminaries.** Figure 1 shows an example of ontology matching, where two electronics ontologies O_1 and O_2 that respectively describe the same knowledge in different ways. Here, we need to determine the semantically related candidates in O_1 and O_2, for instance, the elements with labels Price in O_1 and in O_2 are candidates, while the element with label Digital Cameras in O_2 should be subsumed by the element with label Photo and Cameras in O_1.

2.1. **Ontology, Ontology Matching Process and Ontology Alignment.** For the convenience of describing the work in EA based ontology matching domain, the ontology can be defined as following:

Definition 1 [9]. An ontology is a 9-tuple $O = (C, P, I, A, \leq_C, \leq_P, \phi_{CP}, \phi_{CI}, \phi_{PI})$, where:

- C is a nonempty set of classes,
- P is a nonempty set of properties,
- I is a set of instances (it can be empty),
- A is a set of axioms which should not be empty,
- $\leq_C$ is a partial order on C, called class hierarchy or taxonomy,
- $\leq_P$ is a partial order on P, called property hierarchy,

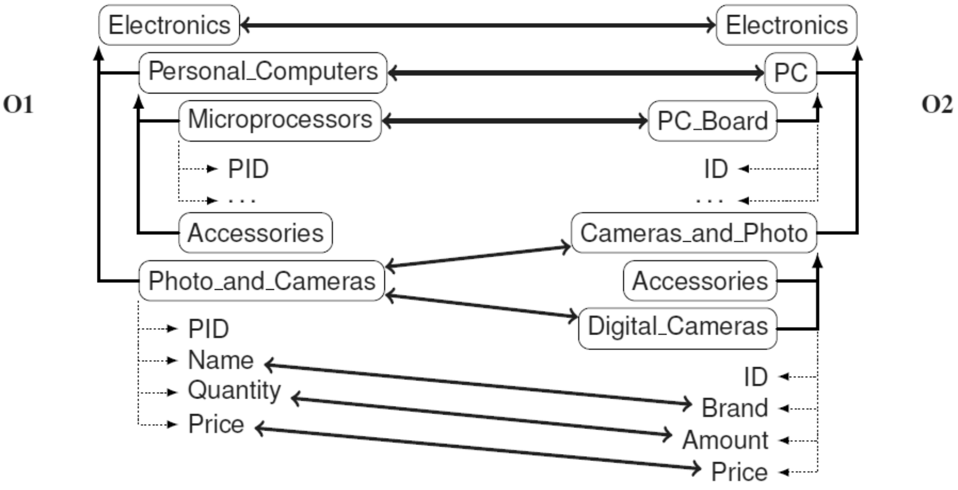


FIGURE 1. An example of ontology matching.

- $\phi_{CP} : P \rightarrow C \times C$ is a function which associates a property $p \in P$ with two linked classes through the property p. We denote the domain by $dom(p) := \pi_1(\phi_{CP}(p))$ and the range by $ran(p) := \pi_2(\phi_{CP}(p))$ where π_1 and π_2 are two functions obtaining the domain class and range class respectively,
- $\phi_{CI} : C \rightarrow \mathcal{P}(I)$ is a function which associates a concept $c \in C$ with a subset of I representing the instances of the concept c,
- $\phi_{PI} : P \rightarrow \mathcal{P}(I^2)$ is a function which associates a property $p \in P$ with a subset of cartesian product $I \times I$ representing the pair of instances related through the property p.

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

To solve the heterogeneity problem between ontologies, a so-called ontology matching process is necessary. Formally, an alignment between two ontologies can be defined as presented by Definition 2.

Definition 2 [10]. An alignment between two ontologies is a set of mapping elements. A mapping element is a 5-tuple (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 ontologies, respectively,
- n is a confidence measure in some mathematical structure (typically in the range $[0, 1]$) holding for the correspondence between entities e and e',
- R is a relation, e.g. equivalence, more general and disjointedness, of the correspondence between entities e and e'.

In principle, all relations between entities in the given ontology language can be used as the correspondence relation, and the interpretation of correspondences and alignments is strongly case-dependent. However, in many cases, a correspondence between ontological entities is always thought of expressing the “equivalent” or at least somewhat “similar” entities. A common assumption is to regard a correspondence as equivalence axiom for two corresponding entities. Furthermore, the ontology matching process can be defined as follows:

Definition 3 [10]. 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, a set of resources r, returns a new alignment A':

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

2.2. **Alignment Evaluation.** The alignment is normally assessed on the basis of two measures commonly known as recall and precision [11]. 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:

$$recall = \frac{|R \bigcap A|}{|R|} \tag{1}$$

$$precision = \frac{|R \bigcap A|}{|A|} \tag{2}$$

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

$$f - measure = \frac{recall \cdot precision}{\alpha \cdot recall + (1-\alpha) \cdot precision} \tag{3}$$

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, f-measure computes their harmonic mean.

3. **Evolutionary Algorithm for Ontology Meta-matching Problem.** Since different ontology matching algorithms, which are also called ontology entity 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 [12]. 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. In addition, among different compositions, the parallel composition of basic matchers, due to its ability of dynamically tuning the basic matchers to obtain the high quality output, becomes the key breakthrough for obtaining first-rate matching performance [13].

3.1. **Ontology Entity Matcher and Aggregation Strategy.** Ontology entity matcher takes as input two ontologies O_1 and O_2 and output a $|O_1| \times |O_2|$ similarity matrix S, whose element s_{ij} is the similarity score between ith entity in $|O_1|$ and jth entity in $|O_2$. In general, the basic ontology matchers can be divided into four categories, i.e. syntactic-based matcher, linguistic-based matcher, structure-based matcher and instance-based similarity matcher.

3.1.1. *Syntactic-based Matcher.* Syntactic-based matcher calculates the edit distance between two ontology entities. SMOA distance [14], which is the most performing measure for the ontology alignment problem. Formally, given the labels of two entities w_1 and w_2, the SMOA distance between them can be defined by the following equation:

$$SMOA(w_1, w_2) = c(w_1, w_2) - d(w_1, w_2) + winklerImprove(w_1, w_2) \tag{4}$$

where $c(w_1, w_2)$ stands for the commonality between w_1 and w_2, $d(w_1, w_2)$ for the difference and $winklerImprove(w_1, w_2)$ for the improvement of the result proposed in [15].

3.1.2. *Linguistic-based Matcher.* Linguistic-based matcher utilizes synonymy, hypernymy and other linguistic relations to calculate the similarity score between ontology entities. To this end, a lexicon and thesauri are needed, and the most popular one is WordNet [16] which is an electronic lexical database where various senses of words are put together into sets of synonyms. Given the labels of two entities w_1 and w_2,

$$LinguisticDistance(w_1, w_2) = \begin{cases} 1, & \text{if } w_1 \text{ and } w_2 \text{ are synonymous,} \\ 0.5, & \text{if } w_1 \text{ is the hypernym of } w_2 \text{ or vice versa,} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

3.1.3. *Structure-based Matcher.* Structure-based matcher computes a similarity score between two ontological entities based on their ontology taxonomy hierarchy structure, and the common intuition is that two distinct ontology entities are similar when their adjacent entities are similar. The most popular structure-based matcher works based on the well known Similarity Flooding (SF) algorithm [17], where an iterative fix-point computation (see also the following equation) is utilized to produce an alignment between the elements of two ontologies.

$$\delta^{i+1} = normalize(\delta^i + f(\delta^i)) \tag{6}$$

where the function f increments the similarity of an element pair (δ^{i+1}) based on the similarity of its neighbors, and the previous iterations value (δ^i) changes in each variation.

3.1.4. *Instance-based Matcher.* Instance-based matcher exploits the instances associated to the entities to determine their similarity. One of the outstanding instance-based matcher utilizes a soft TFIDF [18] to measure similarity between the instances, and then propagate the instance similarities to the entities in the schema-level [19]. Particularly, given two concepts c_1 and c_2 and their direct instance set S_1 and S_2, the similarity score of c_1 and c_2 can be calculated by the following formulas:

$$\frac{\sum_{i=1}^{f} \max\limits_{j=1\cdots g}(sim(s_{1i}, s_{2j})) + \sum_{j=1}^{g} \max\limits_{i=1\cdots f}(sim(s_{1i}, s_{2j}))}{f+g} \tag{7}$$

where:

- f and g are the cardinalities of S_1 and S_2,
- s_{1i} is the ith instance of c_1 and s_{2j} is the jth property of c_2,
- sim is the similarity function returns the similarity value of s_{1i} and s_{2j} which is calculated by the soft TFIDF .

3.1.5. *Aggregation Strategy.* Since the application of a single ontology entity matcher is often not enough to produce an acceptable output alignment, the common strategy is to combine different matchers to compute a unique confidence value as an aggregated similarity value. Obviously, the quality of the alignments is strongly dependent on the selection of the appropriate aggregation strategy. However, the determination of the adequate aggregation strategy is not an easy task, which is a complex process that normally is achieved manually by an expert or by means of general approaches (e.g. maximum, minimum, geometric mean, harmonic mean function, etc.), which are not appropriate to provide optimal results for all alignment problems. Therefore, in this work, we choose the weighted average strategy to aggregate different similarity measures into a single similarity metric, and further utilize MA to automatically find the best manner of aggregating different similarity measures into a single similarity metric. In general, the weighted average aggregation is defined in the following:

$$\phi\left(\overrightarrow{s}\,(c)\,,\overrightarrow{w}\right)=\sum_{i=1}^{n}w_i s_i\,(c) \tag{8}$$

where:

- $\sum_{i=1}^{n} w_i = 1$ and $w_i \in [0,1]$,
- $\overrightarrow{s}\,(c)$ is the vector of similarity measure results,
- $\overrightarrow{w}$ is the vector of weights,
- n is the number of similarity measures.

3.2. **Ontology Meta-Matching Problem.** The ontology meta-matching problem is a six-tuple $(O_1, O_2, A_{set}, R, W_{set}, F)$, where:

- O_1 and O_2 are the ontologies to align, A_{set} is the set of various alignments determined by diverse basic ontology matchers beforehand, and R is the reference alignment given by the domain experts;
- W_{set} is the set of all possible weight set which is used for aggregating various alignment;
- $F : W_{set} \to S \in [0,1]$, S is objective function for evaluating the quality of a weight set $W \in W_{set}$:

$$F(W) = f(A), r(A) or p(A), \quad A=\sum_{i=1}^{|A_{set}|} w_i A_i \quad with \quad w_i \in W \quad \text{and} \quad A_i \in A_{set} \tag{9}$$

where $f, r, p : A \to [0,1]$ calculate the the f-measure, recall and precision of A respectively.

According to the above definition, the single-objective and multi-objective optimal model for the ontology meta-matching problem can be respectively defined as follows:

$$\begin{cases} max \quad f(X) \\ s.t. \quad X=(x_1,x_2,\cdots,x_n)^T \\ \qquad x_i \in [0,1], i=1\cdots n \\ \qquad \sum_{i=1}^{n} x_i = 1 \end{cases} \tag{10}$$

$$\begin{cases} max \quad F(X)=(r(X),p(X)) \\ s.t. \quad X=(x_1,x_2,\cdots,x_n)^T \\ \qquad x_i \in [0,1], i=1\cdots n \\ \qquad \sum_{i=1}^{n} x_i = 1 \end{cases} \tag{11}$$

where n is the number of ontology entity matchers, $x_i, i = 1, 2, /cdots, n$ is the i-th ontology entity matcher's aggregating weights, and f, r and p are three functions calculating the obtained alignment's f-measure, recall and precision, respectively.

3.3. **Single-objective Evolutionary Algorithm based Ontology Meta-matching.** EA usually represents the problem in the form of a bit vector, and then for each chromosome, it evaluates the fitness using an appropriate fitness function suitable for the problem. Based on this, the best chromosomes are selected into the mating pool, where they undergo crossover and mutation thus giving a new set of solutions (offspring). The first ontology matching system utilizes EA to solve Ontology Meta-matching problem is GOAL (Genetics for Ontology ALignments) [20]. GOAL determines, through EA, the optimal weight configuration for a weighted average aggregation of several basic ontology matchers by optimizing one of these conformance measures: precision, recall and f-measure. The same idea of combining multiple basic ontology matcher is also developed by Naya et al. [21]. Alexandru et al. try to optimize the combination of similarity measures by means of a genetic algorithm but, different from the previous works, they focus on optimizing the whole ontology alignment process as a single unit, including the threshold value in the chromosome [22].

The slow convergence and premature convergence are two main shortcomings of the classical EA for ontology meta-matching problem [23]. It makes EA incapable of effectively searching the optimal solution for large scale and complex problems. To overcome this problem, a newly emergent class of EA, named Memetic Algorithm (MA), is introduced to efficiently face the problem of ontology meta-matching. MA is also a population-based search method which combines EA (global search) and local refinements (local search). This marriage between global search and local search allows keeping high population diversity via strong mutation (thus, reducing the possibility of the premature convergence) and increasing the convergence speed via the local search (in fact, local search can greatly improve the solution quality and thus make the solution approaches to optimal solution more quickly). Acampora et al. first define an ontology alignment process based on MA able to efficiently aggregate similarity measures without using a priori knowledge about ontologies under alignment [23]. Through the statistical multiple comparison procedure, MA based approach yields high performance in terms of alignment quality with respect to top-performers of well-known Ontology Alignment Evaluation Initiative (OAEI) campaigns. However, their proposal requires the human expert to provide the reference alignment, which is difficult to obtain especially when the ontology scale is large, to evaluate various solutions. To overcome this drawback, Xue et al. propose to use a Partial Reference Alignment (PRA), i.e. a part of golden alignment given by the domain expert, to evaluate the alignment's quality [24]. On the basis of the PRA based quality measures, a problem-specific MA are designed to tune the parameters of the meta-matching system. Further, they dedicate to overcome three drawbacks of MA based ontology matching technology: (1) is it is difficult to simultaneously deal with several pairs of ontologies, i.e. finding a universal weight configuration that can be used for different ontology pairs without adjustment; (2) a reference alignment between two ontologies to be aligned should be given in advance which could be very expensive to obtain especially when the scale of ontologies is considerably large; (3) simply using f-measure to measure the ontology alignments quality may cause the bias improvement of the solution. In particular, they propose to use both MatchFmeasure, a rough evaluation metric on no reference alignment to approximate f-measure, and Unanimous Improvement Ratio (UIR), a measure that complements MatchFmeasure, in the process of optimizing the ontology alignments by MA. The details of encoding mechanism, fitness function, evolutionary operators and local search strategy please see also papers [24, 9].

3.4. **Multi-objective Evolutionary Algorithm based Ontology Meta-matching.** Since a suitable computation of parameters could be better performed by evaluating the right compromise among different objectives involved in the matching process, approaches based on multi-objective EAs are emerging as an innovative and efficient methodology to face the meta-matching problem [25]. Acampora et al. first propose to use one of the most popular multi-objective "a posteriori" approach, the Non-dominated Sorting Genetic Algorithm II(NSGA-II) [26], in order to perform an ontology meta-matching process by tuning the appropriate values for ontology matching system's parameters. The application of NSGA-II allows to improve the semantic interoperability by finding high quality solutions. Xue et al. also propose to utilize NSGA-II to determine various non-dominated ontology matching system's parameters in terms of recall and precision [27]. Further, they propose a new ontology alignment quality measures which do not require the experts to provide reference alignment [28], and on this basis, a novel optimal model is constructed for ontology meta-matching. In addition, Xue et al. try to use the Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D) [29] to improve the performance of NSGA-II based ontology meta-matching technology [30]. They present the decomposition approach of the objective, the encoding mechanism, the problem-specific evolutionary operators and the principle of selecting the representative solutions for different decision makers. More recently, Acampora et al. compare six multi-objective EAs' performance [25], i.e. NSGA-II, SPEA2 [31], PESA-II [32], OMPPSO [33], DENSEA [34], when solving the ontology meta-matching problem. They also construct a novel optimal model using different ontology alignment measures, and evaluate the considered algorithms by using three metrics: (1) hypervolume [35], which takes into consideration the size of the dominated volume in the objective space; (2) Δ index [26] (distribution and spread PI), which is based on distance and includes information about both spread and distribution; (3) coverage of two sets [35] (binary cardinality-based PI), which is a binary one because it is computed by considering two fronts to be compared one against the other.

4. **Evolutionary Algorithm for Ontology Entity Matching Problem.**

4.1. **Ontology Entity Matching Problem.** The ontology entity matching problem is a five-tuple (O_1, O_2, A, R, F), where:

- O_1 and O_2 are the ontologies to align, A is an ontology alignment determined by given ontology matcher, and R is the reference alignment given by the domain experts;
- $F : A \to [0, 1]$, where F calculate the the f-measure, recall or precision of A respectively.

According to the above definition, the single-objective and multi-objective optimal model for the ontology entity matching problem can be respectively defined as follows:

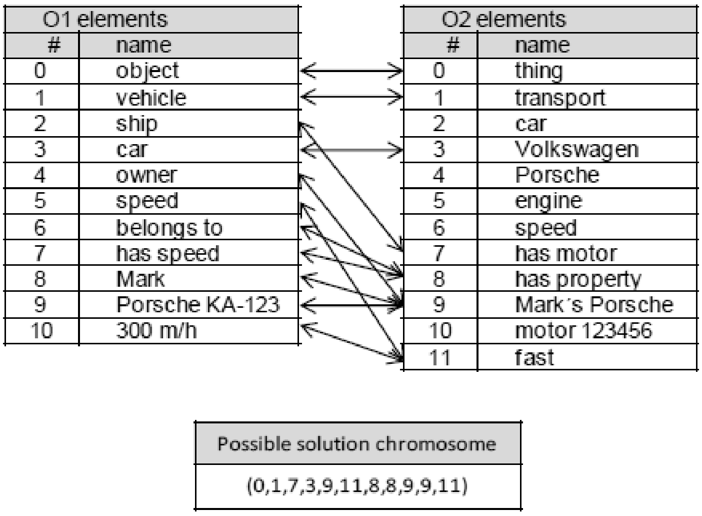
$$\begin{cases} max \quad f(X) \\ s.t. \quad X = (x_1, x_2, \cdots, x_{|O_1|})^T \\ \qquad x_i \in \{1, 2, \cdots, |O_2|\}, i = 1, 2, \cdots, |O_1| \end{cases} \tag{12}$$

$$\begin{cases} max \quad F(X) = (r(X), p(X)) \\ s.t. \quad X = (x_1, x_2, \cdots, x_{|O_1|})^T \\ \qquad x_i \in \{1, 2, \cdots, |O_2|\}, i = 1, 2, \cdots, |O_1| \end{cases} \tag{13}$$

where $|O_1|$ and $|O_2|$ represent the cardinalities entity sets in two ontologies O_1 and O_2, respectively, $x_i, i = 1, 2, \cdots, |O_1|$ represents the ith pair of correspondence, and functions f, r and p calculate the obtained alignment's f-measure, recall and precision, respectively.

Figure 2 gives an example of a fictitious alignment between the elements from two ontologies in a car domain (where each double-ended arrow connects a pair of corresponding elements) and the chromosome representing this alignment. On this basis, various EAs are designed to solve the ontology entity matching problem.

4.2. **Single-objective Evolutionary Algorithm based Ontology Entity Matching.** The first ontology matching system utilizes EA to solve the ontology entity matching problem is GAOM (Genetic Algorithm based Ontology Matching) [36]. GOAM presents the EA based optimization procedure for ontology matching problem as a feature-matching process. First, from a global view, GOAM regards two ontologies as two feature sets, and employ EA to match them. Given a certain ontology alignment, GOAM defines the fitness function as a global similarity measure function between two ontologies based on feature sets. MapPSO [37], instead, addresses the ontology entity matching problem as an optimization problem to be minimized through a computational intelligence technique,



O1 elements	
#	name
0	object
1	vehicle
2	ship
3	car
4	owner
5	speed
6	belongs to
7	has speed
8	Mark
9	Porsche KA-123
10	300 m/h

O2 elements	
#	name
0	thing
1	transport
2	car
3	Volkswagen
4	Porsche
5	engine
6	speed
7	has motor
8	has property
9	Mark´s Porsche
10	motor 123456
11	fast

Possible solution chromosome
(0,1,7,3,9,11,8,8,9,9,11)

FIGURE 2. An example of ontology entity matching alignment between ontologies O_1 and O_2 and its chromosome representation.

i.e., the discrete Particle Swarm Optimization (PSO) [38]. In detail, MapPSO exploits a fitness function depending on the similarity values computed by performing a combination of lexical, linguistic and structural matchers. Also MapPSO employs aggregation techniques (minimum, weighted average aggregation, ordered weighted average aggregation) whose weights are manually set.

In order to overcome the slow convergence and premature convergence of EA for ontology entity matching problem, Acampora et al. [39] first propose a MA to perform an efficient matching process capable of computing a suboptimal alignment between two ontologies. Their experimental results show that MA based approach is more suitable for ontology entity matching problem than a classical EA based approaches. Alves et al. argue that in scenarios where ontologies contain instances, the knowledge embedded in these instances can be useful to improve the alignments. Therefore, the ontology elements that may be considered for the alignment comprise its concepts, relations, or instances. The matching approach proposed by Acampora et al. considered only the first two elements, and they extend it by also considering instances [40]. More recently, Xue et al. [6] first design a MA based approach to solve ontology instance matching problem in Linked Open Data (LOD) [41] which is a cornerstone in the realization of the Semantic Web vision. Their approach works in a sequential stage, i.e. ontology instance matching is carried out with the result of ontology concept matching. First, they respectively propose a profile similarity measure and the rough evaluation metrics with the assumption that the golden alignment for both ontology concept alignment and instance alignment are one to one, i.e. one source ontlogy entity is mapped with one target ontology entity and vice versa. Then, they construct new optimization models for ontology concept matching problem and ontology instance matching problem, respectively, and design a problem-specific MA to solve them. Furthermore, we give the details of the MA. They compare their proposal with the state of the art ontology matching systems on OAEI benchmark and real-world datasets, and the experimental results show that when dealing with ontology entity matching problem, MA based ontology matching technology is more efficient than other state of the art ontology matching systems.

4.3. **Multi-objective Evolutionary Algorithm based Ontology Entity Matching.** Since the perfect ontology alignment is difficult to determine, various decision makers have different preferences on the sub-optimal ontology alignment's quality. For example, someone prefers the alignment with higher precision, while others prefer the one with better recall. Although there are urgent needs of providing different non-dominated alignments, the study on multi-objective EA based technology for solving the ontology entity matching problem is still in its infancy. The only open published literature on this topic is that in 2015, Xue et al. propose a problem specific NSGA-II for multi-objective ontology entity matching problem [19]. Particularly, they first determine the optimal instance correspondences, and then through a similarity propagation algorithm which makes use of various semantic relations, instance similarity values are propagated to other entities of ontologies. The process of determining the ontology instance alignment and the propagation is integrated inside the NSGA-II. The experimental results show the effectiveness of their approach with respect to the high precision values.

5. **Challenges.**

5.1. **Large Scale Ontology Matching.**

For EA based ontology matching technology, how to efficiently solve the large scale ontology matching problem is one of the major challenges since the correct and complete identification of semantic correspondences are difficult to achieve in the huge searching space. Most EA based technologies focus on the improvement of ontology alignment quality, whereas the scalability problem has attracted interest only recently. Although the existing EA based approaches are efficient to solve small scale ontology matching problem, with the number of entities in ontology increases, the search space of the algorithm becomes huge and their performance drop dramatically. Therefore, for large scale ontology matching problem, how to reduce the search space becomes critical to EA based ontology matching technology. Xue et al. [42] propose a segmentation-based ontology matching approach to divide the large scale ontology into small ontology segments, where the term "segment" is referred to a fragment of an ontology which stands alone on its right [43]. They dedicate to use the ontology partition algorithm to transform the large scale ontology matching problem into several ontology segment matching problems, which can be solved in parallel by EA. They further applied this idea to solve the large scale ontology instance matching problem in LOD [6]. They utilize the ontology concept alignment to partition the large scale instance set into small similar instance subsets, and EA based matching process is executed on them in parallel. In addition, for solving the scalability issue, MapPSO [37] propose to use cloud infrastructure's ability to harness parallel computation resources, and on the basis of it, the PSO based matching process is executed. Their deployment is done with a focus on parallel efficiency, taking into account both communication latency and computational inhomogeneity among parallel execution units. To conclude, the efficient ontology partition algorithm and parallel computation mechanism could be two critical technologies to develop the EA based large scale ontology matching.

5.2. **Matching Efficiency.**

However, for dynamic applications, it is necessary to perform the similarity measures combination and system self-tuning at run time, and thus, beside quality (correctness and completeness) of the aligning results, the efficiency (execution time and main memory) of the aligning process is of prime importance especially when a user cannot wait too long for the system to respond or when memory is limited. In order to improve the efficiency of EA based approach, Xue et al. [27] argue that during the process of optimizing the ontology alignment through EA, a large number of evaluations are needed, and the function evaluations for the problem of optimizing ontology alignment are time and memory consuming. In order to reduce the number of time and memory consuming evaluations, they propose to introduce the metamodel, which could be understood as surrogate evaluation models that are built using existing information [44], to approximate the objective function value using solutions that have already been evaluated. As the core technology, metamodeling approach is helpful to considerably improve the efficiency of solving process by using a large number of precise evaluations. Their proposal is able to highly reduce the time and main memory consumption of the tuning process while at the same time ensures the correctness and completeness of the alignments. Further, since according to [45], if properly designed, a population-based algorithm with a very small population size can efficiently solve large scale problems, Xue et al. propose a compact EA to improve the efficiency of EA based ontology matching process. Compact EA employs the search logic of population-based algorithms but do not store and process an entire population, but on the contrary make use of a probabilistic representation of the population in order to perform the optimization process [46]. Thus, their compact EA based ontology matching technology can highly improve the efficiency of the population-based EA approach. More recently, they propose a parallel version of compact EA to further improve the efficiency of compact EA based ontology matching technology [47]. In summary, the effective parallel and compact mechanisms are two main technologies to improve the EA based ontology matching efficiency.

5.3. **Ontology Alignment Evaluation.** The ontology alignment evaluation direct the evolutionary direction of EA, which is critical to the design of EA based ontology matching technology. Although the classic alignment evaluation metrics, recall, precision and f-measure can reflect the quality of the resulting alignment, they require that the perfect matching result, i.e. the reference alignment, should be given in advance. However, this perfect match result is generally unknown to difficult real-life match problems, especially for large heterogeneous ontologies. To overcome this drawback, based on the observations that the quality of an alignment is calculating by taking into account these two reasonable observations: (1)the higher the average of the confidence values of the correspondences and the better the alignment quality; (2)the higher the number of correspondences and the better the alignment quality, Acampora et al. utilize two new alignment evaluation metrics which are totally independent of reference alignment [23]. Xue et al. also propose two alignment evaluation metrics, i.e. MatchCoverage and Frequency, to respectively approximate recall and precision, without requiring the expert to provide the reference alignment [28]. Further, they present two partial reference alignments based alignment evaluation metrics, which require expert to provide a part of representative reference alignment, to help the algorithm evaluate the solutions [24]. However, the above alignment evaluation metrics works well on the condition that the reference alignment is one to one, i.e. the entity of one ontology can correspond to only one entity of the other and every entity of an ontology participates in a correspondence. Currently, there are no such evaluation metrics are able to effectively approximate the quality of alignment in the matching scenario that reference alignment is many to many.

5.4. **User Involvement.** Due to the complexity of the ontology matching process, ontology alignments generated by the EA based matching technologies should be validated by the users to ensure their qualities [3]. The technology makes users and automatic tools cooperate with each other to create high quality matchings in a reasonable amount of time are referred to as semi-automatic ontology matching [48]. Since user is a rare and valuable resource, we can not expect them to validate the whole alignment especially in the context of large scale ontology matching [49]. Rather, how to minimize user involvement, but at the same time, maximize user involvement's value is one of the main challenges in semi-automatic ontology matching domain [5]. To this end, a feasible way of further improving the matching performance of EA based matching technology is to get user involved into its evolutionary process and utilize user validation to improve the solution's quality in an iterative way. Particularly, EA could adaptively determine the timing of getting user involved, determine the problematic entity correspondences for user's validation and propagate the validation to maximize the user involvement's value. In terms of this, an interactive EA could be taken into consideration to solve the ontology matching problem. Although semi-automatic ontology matching is a hot research topic in ontology matching domain, currently, there is only one semi-automatic EA based ontology matching approach which is proposed by Xue et al. in 2014 [24].

6. **Conclusion.** EA based ontology matching technologies are emerging as the most suitable approach for solving ontology matching problem. In this paper, we introduce the basics of EA based ontology matching technology, and discuss the state-of-the-art EA based ontology matching approach. EA based ontology matching is making a measurable progress, though it is slowing down. In order to address this situation, we presented four challenges for EA based ontology matching, accompanied for each of these with an overview of the recent advances in the field and a discussion of the potentially useful ways to approach the challenges under consideration. We believe that addressing the outlined challenges should accelerate the progress of the EA based ontology matching field.

Acknowledgment. This work is 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 Key Project of Fujian Education Department Funds (No. JA15323).

REFERENCES

[1] T. R. Gruber, et al., A translation approach to portable ontology specifications, *Knowledge acquisition* vol.5, no.2, pp.199-220, 1993.

[2] T. Berners-Lee, J. Hendler, O. Lassila, et al., The semantic web, *Scientific American*, vol.284, no.5, pp.28-37, 2001.

[3] P. Shvaiko, 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.

[4] P. Shvaiko, J. Euzenat, Ten challenges for ontology matching, *OTM Confederated International Conferences*, Springer, pp.1164-1182, 2008.

[5] L. Otero-Cerdeira, F. J. Rodriguez-Martinez, A. Gmez-Rodriguez, Ontology matching: A literature review, *Expert Systems with Applications*, vol.42, no.2, pp.949-971, 2015.

[6] X. Xue, Y. Wang, Using memetic algorithm for instance coreference resolution, *IEEE Transactions on Knowledge and Data Engineering*, vol.28, no.2, pp.580-591, 2016.

[7] J. H. Holland, Genetic algorithms, *Scientific American*, vol.267, no.1, pp.66-72, 1992.

[8] J. Martinez-Gil, J. F. A. Montes, Evaluation of two heuristic approaches to solve the ontology meta-matching problem, *Knowledge and Information Systems*, vol.26, no.2, pp.225-247, 2011.

[9] X. Xue, Y. Wang, Optimizing ontology alignments through a memetic algorithm using both matchfmeasure and unanimous improvement ratio, *Artificial Intelligence*, vol.223, pp.65-81, 2015.

[10] J. Euzenat, P. Shvaiko, et al., Ontology matching, Springer, vol.18, 2007.

[11] C. J. Van Rijsbergen, Foundation of evaluation, *Journal of Documentation*, vol.30, no.4, pp.365-373, 1974.

[12] T. T. A. Nguyen, S. Conrad, Ontology matching using multiple similarity measures, *7th International Joint Conference on Knowledge Discovery*, Knowledge Engineering and Knowledge Management, vol.1, pp.603-611, 2015.

[13] M. Gulic, B. Vrdoljak, M. Banek, Cromatcher: An ontology matching system based on automated weighted aggregation and iterative final alignment, *Web Semantics: Science, Services and Agents on the World Wide Web*, vol.41, pp.50-71, 2016.

[14] G. Stoilos, G. Stamou, S. Kollias, A string metric for ontology alignment, *International Semantic Web Conference*, Springer, pp.624-637, 2005.

[15] W. E. Winkler, The state of record linkage and current research problems, Statistical Research Division, US Census Bureau, Citeseer, 1999.

[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, E. Rahm, Similarity flooding: A versatile graph matching algorithm and its application to schema matching, *18th International Conference on Data Engineering*, 2002, pp.117-128.

[18] M. Cheatham, P. Hitzler, String similarity metrics for ontology alignment, *International Semantic Web Conference*, Springer, pp.294-309, 2013.

[19] X. Xue, Y. Wang, Ontology alignment based on instance using nsga-ii, *Journal of Information Science*, vol.41, no.1, pp.58-70, 2015.

[20] J. Martinez-Gil, E. Alba, 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, 2008.

[21] J. V. Naya, M. M. Romero, J. P. Loureiro, C. R. Munteanu, A. P. Sierra, Improving ontology alignment through genetic algorithms, *Soft computing methods for practical environment solutions: Techniques and studies*, pp.240-259, 2010.

[22] G. Alexandru-Lucian, A. Iftene, Using a genetic algorithm for optimizing the similarity aggregation step in the process of ontology alignment, *Roedunet International Conference (RoEduNet)*, pp.118-122, 2010.

[23] G. Acampora, V. Loia, A. Vitiello, Enhancing ontology alignment through a memetic aggregation of similarity measures, *Information Sciences*, vol.250, pp.1-20, 2013.

[24] X. Xue, Y. Wang, A. Ren, Optimizing ontology alignment through memetic algorithm based on partial reference alignment, *Expert Systems with Applications*, vol.41, no.7, pp.3213-3222, 2014.

[25] G. Acampora, H. Ishibuchi, A. Vitiello, A comparison of multi-objective evolutionary algorithms for the ontology meta-matching problem, *2014 IEEE Congress on Evolutionary Computation (CEC)*, pp. 413-420, 2014.

[26] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multi-objective genetic algorithm: Nsga-ii, *IEEE transactions on evolutionary computation*, vol.6, no.2, pp.182-197, 2002.

[27] X. Xue, Y. Wang, W. Hao, Optimizing ontology alignments by using nsga-ii., *International Arab Journal of Information Technology*, vol.12, no.2, pp.175-181, 2015.

[28] X. Xingsi, W. Yuping, H. Weichen, H. Juan, Optimizing ontology alignments through nsga-ii without using reference alignment. *Computing and Informatics*, vol.33, no.4, pp.857-876, 2015.

[29] Q. Zhang, H. Li, Moea/d: A multiobjective evolutionary algorithm based on decomposition, *IEEE Transactions on evolutionary computation*, vol.11, no.6, pp.712-731, 2007.

[30] X. Xue, Y. Wang, W. Hao, Using moea/d for optimizing ontology alignments, *Soft Computing*, vol.18, no.8, pp.1589-1601, 2014.

[31] E. Zitzler, M. Laumanns, L. Thiele, et al., Spea2: Improving the strength pareto evolutionary algorithm, TIK-Report, pp.1-21, 2001.

[32] D. W. Corne, N. R. Jerram, J. D. Knowles, M. J. Oates, Pesa-ii: Region based selection in evolutionary multiobjective optimization, *Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation*, Morgan Kaufmann Publishers Inc., 2001, pp.283-290.

[33] M. R. Sierra, C. A. C. Coello, Improving pso-based multi-objective optimization using crowding, mutation and-dominance, *International Conference on Evolutionary Multi-Criterion Optimization*, Springer, pp.505-519, 2005.

[34] D. Greiner, G. Winter, J. M. Emperador, Enhancing the multiobjective optimum design of structural trusses with evolutionary algorithms using densea, *44th AIAA Aerospace Sciences Meeting and Exhibit*, pp.1474, 2006.

[35] E. Zitzler, L. Thiele, Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach, *IEEE transactions on Evolutionary Computation*, vol.3, no.4, pp.257-271, 1999.

[36] J. Wang, Z. Ding, C. Jiang, Gaom: Genetic algorithm based ontology matching, *IEEE Asia-Pacific Conference on Services Computing*, pp.617-620, 2006.

[37] J. Bock, J. Hettenhausen, Discrete particle swarm optimisation for ontology alignment, *Information Sciences*, vol.192, pp.152-173, 2012.

[38] R. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, *Proceedings of the Sixth International Symposium on Micro Machine and Human Science* pp.39-43, 1995.

[39] G. Acampora, V. Loia, S. Salerno, A. Vitiello, A hybrid evolutionary approach for solving the ontology alignment problem, *International Journal of Intelligent Systems*, vol.27, no.3, pp.189-216, 2012.

[40] A. Alves, K. Revoredo, F. Baiao, Ontology alignment based on instances using hybrid genetic algorithm, *Proceedings of the 7th International Conference on Ontology Matching-Volume*, pp.242-243, 2012.

[41] C. Bizer, T. Heath, T. Berners-Lee, Linked data-the story so far, *Semantic services, interoperability and web applications: emerging concepts*, pp.205-227, 2009.

[42] X. Xue, J.-S. Pan, A segment-based approach for large-scale ontology matching, *Knowledge and Information Systems*, pp.1-18, 2017.

[43] J. Seidenberg, A. Rector, Web ontology segmentation: analysis classification and use, *Proceedings of the 15th International Conference on World Wide Web*, pp.13-22, 2006.

[44] M. Emmerich, A. Giotis, M. Ozdemir, T. Back, K. Giannakoglou, Meta-model assisted evolution strategies, *International Conference on parallel problem solving from nature*, Springer, pp.361-370, 2002.

[45] K. E. Parsopoulos, Cooperative micro-differential evolution for high dimensional problems, *Proceedings of the conference on Genetic and evolutionary computation*, pp.531-538, 2009.

[46] F. Neri, G. Iacca, E. Mininno, Compact optimization, vol.38, pp.337-364, 2013.

[47] X. Xue, P.-W. Tsai, L.-L. Zhang, Using parallel compact evolutionary algorithm for optimizing ontology alignment, *International Conference on Genetic and Evolutionary Computing*, Springer, pp.157-165, 2016.

[48] M. S. Falconer, F. N. Noy, *Interactive techniques to support ontology matching, in: Schema Matching and Mapping*, Springer, pp.29-51, 2011.

[49] Z. Dragisic, V. Ivanova, P. Lambrix, D. Faria, E. Jimnez-Ruiz, C. Pesquita, User validation in ontology alignment, *International Semantic Web Conference*, Springer, pp.200-217, 2016.