

Research Article

Optimizing Biomedical Ontology Alignment through a Compact Multiobjective Particle Swarm Optimization Algorithm Driven by Knee Solution

Xingsi Xue ^{1,2,3,4}, Xiaojing Wu,⁵ and Junfeng Chen ⁶

¹Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, Fuzhou 350118, China

²Intelligent Information Processing Research Center, Fujian University of Technology, Fuzhou 350118, China

³Fujian Key Lab for Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou 350118, China

⁴Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

⁵College of Information Science and Engineering, Fujian University of Technology, Fuzhou 350118, China

⁶College of IOT Engineering, Hohai University, Changzhou 213022, China

Correspondence should be addressed to Xingsi Xue; jack8375@gmail.com

Received 19 January 2020; Revised 15 March 2020; Accepted 19 March 2020; Published 1 May 2020

Guest Editor: Chi-Hua Chen

Copyright © 2020 Xingsi Xue et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nowadays, most real-world decision problems consist of two or more incommensurable or conflicting objectives to be optimized simultaneously, so-called multiobjective optimization problems (MOPs). Usually, a decision maker (DM) prefers only a single optimum solution in the Pareto front (PF), and the PF's knee solution is logically the one if there are no user-specific or problem-specific preferences. In this context, the biomedical ontology matching problem in the Semantic Web (SW) domain is investigated, which can be of help to integrate the biomedical knowledge and facilitate the translational discoveries. Since biomedical ontologies often own large-scale concepts with rich semantic meanings, it is difficult to find a perfect alignment that could meet all DM's requirements, and usually, the matching process needs to trade-off two conflict objectives, i.e., the alignment's recall and precision. To this end, in this work, the biomedical ontology matching problem is first defined as a MOP, and then a compact multiobjective particle swarm optimization algorithm driven by knee solution (CMPSO-K) is proposed to address it. In particular, a compact evolutionary mechanism is proposed to efficiently optimize the alignment's quality, and a max-min approach is used to determine the PF's knee solution. In the experiment, three biomedical tracks provided by Ontology Alignment Evaluation Initiative (OAEI) are used to test CMPSO-K's performance. The comparisons with OAEI's participants and PSO-based matching technique show that CMPSO-K is both effective and efficient.

1. Introduction

Decision-making requires finding an optimum solution to a decision problem in the process of identifying and evaluating alternatives. Nowadays, most real-world decision problems consist of two or more incommensurable objectives to be optimized simultaneously, so-called multiobjective optimization problems (MOPs) [1]. In general, MOP does not have one optimum solution but a set of solutions, so-called Pareto optimal solutions, which are superior to the others in terms of one or more objectives. In many situations, a decision maker (DM) only prefers one

single optimum solution in the Pareto front (PF) [2], and therefore, the optimization and the decision process are often combined. One of the multiobjective methods is driven by the PF's knee solution which is the one for which an improvement in one of its objectives will result in a deterioration in another. Often, PF's knee solution is logically preferred to the DM if there are no user-specific or problem-specific preferences [3].

In this context, the biomedical ontology matching problem in the Semantic Web (SW) domain is investigated. Although biomedical ontology is extensively used in the biomedical domain to model the biomedical knowledge, it is

developed and maintained by different communities, and the same biomedical knowledge could be described with different terminologies or in different contexts. To bridge the semantic gaps between two biomedical ontologies and support their communications, it is necessary to find their identical concept mapping, which is the so-called biomedical ontology matching. Since biomedical ontologies often own large-scale concepts with rich semantic meanings, it is difficult to find a perfect alignment that could meet all DM's requirements, and usually, the matching process needs to trade-off two conflict objectives, i.e., the alignment's recall and precision. To this end, in this work, the biomedical ontology matching problem is defined as a MOP. Although there are many studies on the single-objective approaches for addressing the ontology matching problem [4–8], the research on multiobjective ontology matching techniques is still in this infancy [9]. Being inspired by the success of the particle swarm optimization algorithm (PSO) in the ontology matching domain [10], large-scale discrete optimization domain [11–13], and biomedical engineering [14], in this work, a compact multiobjective particle swarm optimization algorithm driven by knee solution (CMPPO-K) is further proposed to address this problem. In particular, CMPPO-K uses a compact multi-objective evolutionary mechanism to efficiently optimize the alignment's quality and a max-min approach to determine the PF's knee solution. The contributions made in this paper are as follows: (1) a discrete multiobjective optimal model for the biomedical ontology matching problem is constructed; (2) a hybrid biomedical concept similarity metric is proposed, which can effectively calculate the similarity value of two biomedical ontology concepts; and (3) a CMPPO-K is proposed to optimize the biomedical ontology alignment, which takes into consideration both DM's preference and algorithm's performance.

The rest of the paper is organized as follows: Section 2 overviews the SIA-based ontology matching techniques; Section 3 defines the biomedical ontology matching problem and presents the similarity measure on biomedical concepts; Section 4 presents the CMPPO-K-based biomedical ontology matching technique in detail; Section 5 shows the experimental results; and finally, Section 6 draws the conclusion and presents the future work.

2. Swarm Intelligence Algorithm-Based Ontology Matching Technique

The first generation of swarm intelligence algorithm- (SIA-) based matcher dedicates to optimize the ways of aggregating various ontology matchers' corresponding alignments. The very first matching system is Genetics for Ontology ALignments (GOAL) [7], which uses evolutionary algorithm (EA) to optimize the aggregating weight set of different ontology matchers. Later, Alexandru-Lucian and Iftene [15] further used EA to optimize one more parameter to filter the unauthentic concept mappings to final alignment. Acampora et al. [4] introduced a local search process into EA's evolving process to improve the algorithm's performance. Xue and Wang [16] used a new metric as the fitness function to guide the algorithm's search direction. Their approach can

address the holistic matching problem and determine a universal weight configuration for matching several pairs of ontologies at a time. He et al. [6] proposed to utilize artificial bee colony algorithm (ABC) to optimize all the parameters in the matching process, whose results are better than the EA-based matchers. More recently, Xue et al. [17] proposed a new approach that uses NSGA-III [18] to combine various similarity measures without tuning the aggregating parameters. However, when the scale of the similarity measures becomes huge, e.g., more than 50 similarity measures, this approach could be inefficient.

The above matchers need to calculate all the matchers' alignments and store them in the main memory before aggregating them, which requires huge memory consumption. Recently, the second generation of the SIA-based matcher tries to directly find an entity correspondence set that is close to the golden alignment. GAOM (genetic algorithm-based ontology matching) [8] regards two ontologies as two discrete concept sets and employs EA to determine the optimal entity mapping set. Alves et al. utilized a memetic algorithm (MA), which combines EA with a local search strategy, to execute the instance-based ontology matching process [5]. They first matched the instances and then propagated the instance pair's similarity value to the corresponding concepts. More recently, MapPSO [10] uses PSO to determine the optimal entity correspondence set. In particular, MapPSO introduces a new quality measure on the ontology alignment, which depends on the statistical results on the alignment. More recently, Chu et al. [19] first modeled the ontologies in the vector space so that two entities' similarity value can be calculated through the cosine function, and then EA was used to determine the optimal alignments. Xue [20] proposed a new similarity metric for measuring the biomedical concepts' similarity value and then used the firefly algorithm (FA) to optimize the biomedical ontology alignment. However, their proposal suffers from the premature convergence when matching large-scale ontologies. Our proposal also belongs to this category, but different from the existing work, in this work, the biomedical ontology problem is regarded as a discrete MOP, and a multiobjective SIA is presented to address it, which takes into consideration the DM's preference.

3. Preliminaries

3.1. Biomedical Ontology Matching Problem. Ontology matching aims at determining the identical entity mappings, which is the so-called ontology alignment. In the past, the ontology alignment's quality was often measured by f-measure [21], which is defined as follows:

$$\begin{aligned} recall &= \frac{|R \cap A|}{|R|}, \\ precision &= \frac{|R \cap A|}{|A|}, \\ f\text{-measure} &= \frac{recall \times precision}{\alpha \times recall + (1 - \alpha) \times precision}, \end{aligned} \quad (1)$$

where R and A are, respectively, the alignments determined by the domain expert and ontology matcher. Recall and precision, respectively, measure A 's completeness and soundness, f -measure is their harmony mean, and $\alpha \in [0, 1]$ is the weight to trade-off recall and precision. However, f -measure requires domain experts to provide R , which limits its application in real practice. Supposing a golden biomedical ontology alignment's cardinality is 1:1, since the larger cardinality of the concept mapping set is and the higher the found correspondences' mean similarity value is, the better the alignment quality is, three approximate measures, i.e., recall', precision', and f -measure', are used to, respectively, approximate the original recall, precision, and f -measure [22]:

$$\begin{aligned} \text{recall}' &= \frac{|A|}{\max\{|C_1|, |C_2|\}}, \\ \text{precision}' &= \frac{\sum_{i=1}^{|A|} \text{simValue}_i}{|A|}, \\ f\text{-measure}' &= \frac{\text{recall}' \times \text{precision}'}{\alpha \times \text{recall}' + (1 - \alpha) \times \text{precision}'}, \end{aligned} \quad (2)$$

where $|C_1|$ and $|C_2|$ are two ontologies' concept scale, $|A|$ is A 's concept mapping number, and simValue_i is the i -th concept mapping's similarity value.

Finally, biomedical ontology matching problem's discrete multiobjective optimal model is defined as follows:

$$\begin{cases} \max f(x) = (r(x), p(x)), \\ \text{s.t. } X = (x_1, x_2, \dots, x_{|C_1|})^T, \\ x_i \in \{1, 2, \dots, |C_2|\}, i = 1, 2, \dots, |C_1|, \end{cases} \quad (3)$$

where $|C_1|$ and $|C_2|$ are, respectively, the concept number of two biomedical ontologies, x_i means the i -th pair of concept correspondence, i.e., i -th source concept is mapped to target x_i -th concept, and two objective functions r and p are to, respectively, maximize recall' and precision' of X 's corresponding alignment.

3.2. Similarity Measure on Biomedical Concept. In this work, a hybrid similarity measure is used to enhance the confidence of the calculated similarity value, which takes into consideration the concept's syntax and linguistic and context information. First, for each biomedical concept, the information (the label, comment, and property name) from itself and its context concept are put into its separated profile. Then, two biomedical concepts' similarity value is calculated according to the following equation:

$$\text{sim}(p_1, p_2) = \frac{\sum_{j=1}^f \max_{i=1 \dots g}(\text{sim}(p_{1i}, p_{2j})) + \sum_{j=1}^g \max_{i=1 \dots f}(\text{sim}(p_{1i}, p_{2j}))}{f + g}, \quad (4)$$

where p_1 and p_2 are, respectively, two biomedical concept's corresponding profiles and $\text{sim}(p_{1i}, p_{2j})$ is calculated by

integrating the SMOA [23] based on syntax measure and Unified Medical Language System (UMLS) [24] based on linguistic measure.

$$\text{sim}(p_{1i}, p_{2j}) = \begin{cases} 1, & \text{if two elements are synonymous in UMLS} \\ \text{SMOA}(p_{1i}, p_{2j}), & \text{otherwise} \end{cases}. \quad (5)$$

4. Compact Multiobjective Swarm Optimization Algorithm Driven by Knee Solution

This work proposes a CMPSO-K for solving the biomedical ontology matching problem, which approximates the population-based PSO's evolving process through a probability vector (PV) [25]. In the next, the objective decomposition approach, the encoding mechanism, and knee solution determination are presented, respectively, and finally, the details of CMPSO-K are shown through the pseudo-code.

4.1. Objective Decomposition. In this work, the weighted sum approach is used to transform a MOP into a set of subproblems and solves them simultaneously.

$$\begin{cases} \max g^{ws}(x) = \sum_{i=1}^N w_i \cdot f_i(x), \\ \text{s.t. } x \in \Omega \subset R^n \end{cases}, \quad (6)$$

where N is the number of decomposed problems and $w_i \in [0, 1]$ $i = 1, \dots, N$, $\sum_{i=1}^N w_i = 1$. Two objectives of our problem are to maximize recall' and precision', and the i th subproblem's objective is defined as follows:

$$\max \frac{\text{recall}' \times \text{precision}'}{\alpha_i \times \text{recall}' + (1 - \alpha_i) \times \text{precision}'}, \quad (7)$$

where $\alpha_1 = 0$ and $\alpha_i = i - 1/N - 1, \dots, \infty, \alpha_N = 1$.

4.2. Encoding Mechanism. In this work, the Gray coding, a binary encoding mechanism, is used to encode an alignment. As can be seen from Figure 1, an example of the encoding mechanism is shown, in which the source concept "uterine gland" with index 8 is mapped to target concept "Uterine Gland" with index 6 whose Gray code is 110. In particular, Gray code 000 means a source concept is not mapped to any target concept.

Moreover, a PV is utilized to characterize a population for solving one decomposed subproblem, whose element number is equal to that of an individual, and each element in PV represents the probability of being 0. Therefore, various individuals can be generated with the binary code through a PV. Figure 2 shows an example of generating an individual through PV. Through comparing each of PV's element with the random number, the new individual's element value can be determined. It is obvious that when all of PV's elements are close to 1 or 0, the algorithm tends to converge. At the end of each generation, PV is updated according to the best individual by increasing (or decreasing) its element value if the corresponding element of that individual is 1 (or 0).

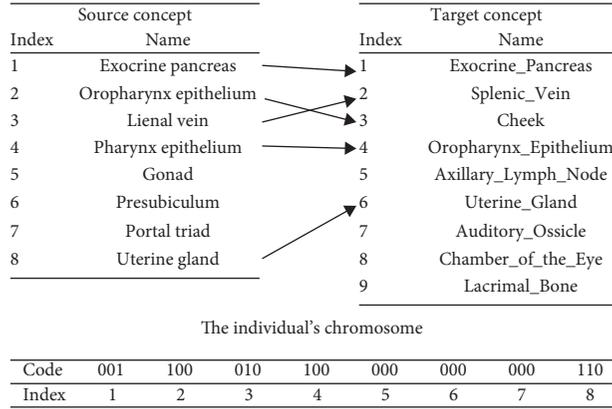


FIGURE 1: An example of the encoding mechanism.

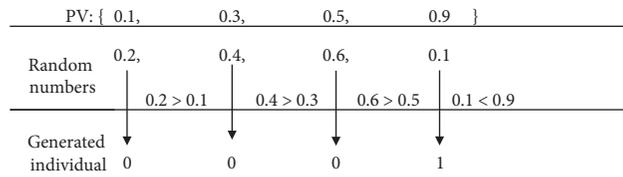


FIGURE 2: An example of generating an individual through PV.

Input: two individuals ind_1 and ind_2 , crossover probability p_{cr}
 Output: a new individual ind_{new}

- (1) $ind_{new} = ind_1$;
- (2) **for** $ind\ i = 1; i \leq ind_1.length, i++$ **do**
- (3) **if** $rand(0, 1) < p_{cr}$ **then**
- (4) $ind_{new,i} = ind_{2,i}$;
- (5) **end if**
- (6) **end for**
- (7) **return** ind_{new} ;

ALGORITHM 1: Uniform crossover.

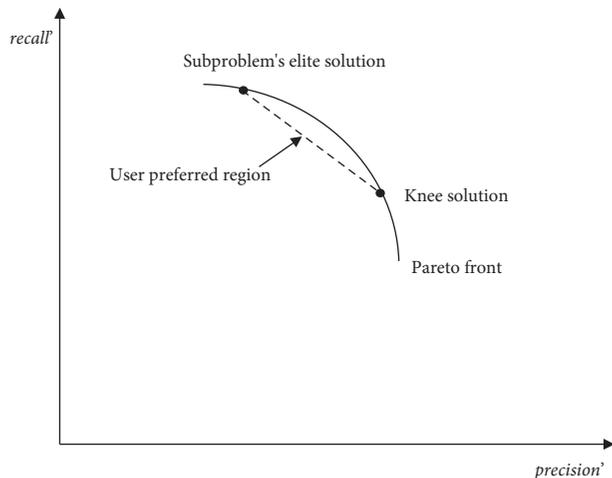


FIGURE 3: An example of knee solution and user preferred region.

```

Input: PF solution set  $IND_{pf} = \{ind_1, ind_2, \dots\}$ 
Output: a knee solution  $ind_{elite}$ 
(1)  $max = 0;$ 
(2) for  $ind\ i = 1; i \leq IND_{pf}.length, i++$  do
(3)    $min_i = \min\{ind_i.recall', ind_i.precision'\}$ 
(4)   if  $min_i > max$  then
(5)      $max = min_i;$ 
(6)      $ind_{elite} = ind_i;$ 
(7)   end if
(8) end for
(9) return  $ind_{elite}$ 

```

ALGORITHM 2: Determination of knee solution.

4.3. Crossover Operator. The crossover operator generates one child individual by mixing the information of two parent individuals, which is carried out according to the crossover probability. In this work, the uniform crossover operator is used, whose pseudo-code is shown in Algorithm 1. Given two parent solutions ind_1 and ind_2 , each gene bit value of their offspring ind_{new} is the same with the corresponding gene bit value of ind_1 (or ind_2) when a random number $rand(0, 1)$ is larger than (or smaller than) the crossover rate p_{cr} .

4.4. Determination of Knee Solution. Figure 3 shows an example of knee solution and user preferred region. As can be seen from the figure, the knee solution, a subproblem's elite solution, and the PF form a user preferred region for that subproblem. In each generation, the newly generated solution tends to move toward that region to meet a DM's requirement. Thus, the knee solution's determination is of utmost importance for the algorithm's search performance. In this work, for those solutions in the PF, a max-min approach is used to determine the knee solution. In particular, suppose ind_1 and ind_2 are two solutions in the PF, and $recall'_i$ and $precision'_i$, $i = 1, 2$, are, respectively, their recall' and precision', and they can be compared according to the following formula:

$$x_i = \arg \max \{ \min \{ recall'_i, precision'_i \} \}. \quad (8)$$

On this basis, a best solution can be selected from PF solutions, and this procedure is shown in Algorithm 2.

4.5. Pseudo-Code of Compact Multiobjective Particle Swarm Optimization Algorithm Driven by Knee Solution. The pseudo-code of CMPSO-K is presented in Algorithm 3. CMPSO-K first divides the problem into three subproblems that, respectively, maximize $f_{\alpha=0}$, $f_{\alpha=0.5}$, and $f_{\alpha=1}$. First, three PVs and local best individuals are, respectively, initialized for three subproblems, and then the knee solution (or global best individual) is initialized by first using the nondominated sorting algorithm [26] on the population and then the knee solution in the PF is determined. In each generation, CMPSO-K tries to solve each subproblem by approximating PSO's position updating strategy, i.e., crossover an individual with the local best individual and

global best individual to obtain a new individual, and then use the new one to update the local best individual and PV. After solving each subproblem, a population in the current generation is obtained by 3 local best individuals and 6 individuals in total generated through $PV_{\alpha=0}$, $PV_{\alpha=0.5}$, and $PV_{\alpha=1}$, respectively, and then the nondominated sorting algorithm is used to determine its PF and the current generation's knee solution ind'_{elite} . Through comparing ind'_{elite} with historical knee solution ind_{elite} through the max-min approach, the latter can be updated. Finally, when the generation approaches the maximum generation $maxGen$, the algorithm terminates and returns ind_{elite} .

5. Experiment

To test CMPSO-K's performance, the experiment exploits three biomedical tracks in Ontology Alignment Evaluation Initiative (OAEI), i.e., anatomy track, large biomed track, and disease and phenotype track. The testing cases in these tracks are all practical ontologies that are widely used in the biomedical domain and open to achieve, which have lots of overlapping information with different representations. OAEI provides the reference alignments for each track to test a matcher's performance, and Table 1 briefly describes the ontologies in these tracks.

In terms of the alignment's quality, the EA-based matcher [8], ABC-based matcher [6], PSO-based matcher [10], and OAEI's participants are compared in Tables 2 and 3, and also, in terms of the memory consumption and converging speed, CMPSO-K-based matcher and other SIA-based matchers are compared in Figures 4 and 5. All the SIAs' results are the mean values of 30 independent executions. EA, ABC, and PSO's configurations are referred to their literatures, and CMPSO-K uses the following configuration: number of decomposed problems: $N = 3$; maximum generation: $maxGen = 3000$; crossover probability: $p_{cr} = 0.6$; and step length for updating PV: $step = 0.1$. This configuration is determined in an empirical way, which represents a trade-off setting in the experiment to obtain the highest average results on all testing cases.

First, Friedman's test [32] is used to figure out whether all the matchers present any difference and then determines whether one matcher statistically outperforms others through Holm's test [33]. In Friedman's test, the null

Input: maximum generation $maxGen$, crossover rate p_{cr} , step length for updating PV $step$

Output: knee solution ind_{elite}

```

(1) **Initialization**
(2) initialize generation  $t = 0$ ;
(3) initialize  $PV_{\alpha=0}$ ,  $PV_{\alpha=0.5}$  and  $PV_{\alpha=1}$  by setting all the elements inside as 0.5;
(4) initialize three local best individuals  $ind_{\alpha=0,elite}$ ,  $ind_{\alpha=0.5,elite}$  and  $ind_{\alpha=1,elite}$ ;
(5) generate 5 individuals through  $PV_{\alpha=0}$ ,  $PV_{\alpha=0.5}$  and  $PV_{\alpha=1}$ ;
(6) NonDominatedSort ();
(7) initialize the knee solution (or global best individual)  $ind_{elite}$ ;
(8) **Evolving Process**
(9) while  $t < maxGen$  do
(10) **Updating  $ind_{\alpha=0,elite}$ **
(11) generate an individual  $ind_{\alpha=0,new}$  through  $PV_{\alpha=0}$ ;
(12)  $ind_{\alpha=0,new} = crossover(ind_{\alpha=0,new}, ind_{\alpha=0,elite})$ ;
(13)  $ind_{\alpha=0,new} = crossover(ind_{\alpha=0,new}, ind_{elite})$ ;
(14)  $[winner, loser] = compete(ind_{\alpha=0,new}, ind_{\alpha=0,elite})$ ;
(15) updatePV ( $winner, loser$ );
(16) if  $winner == ind_{\alpha=0,new}$  then
(17)  $ind_{\alpha=0,elite} = ind_{\alpha=0,new}$ ;
(18) end if
(19) **Updating  $ind_{\alpha=0.5,elite}$ **
(20) generate an individual  $ind_{\alpha=0.5,new}$  through  $PV_{\alpha=0.5}$ ;
(21)  $ind_{\alpha=0.5,new} = crossover(ind_{\alpha=0.5,new}, ind_{\alpha=0.5,elite})$ ;
(22)  $ind_{\alpha=0.5,new} = crossover(ind_{\alpha=0.5,new}, ind_{elite})$ ;
(23)  $[winner, loser] = compete(ind_{\alpha=0.5,new}, ind_{\alpha=0.5,elite})$ ;
(24) updatePV ( $winner, loser$ );
(25) if  $winner == ind_{\alpha=0.5,new}$  then
(26)  $ind_{\alpha=0.5,elite} = ind_{\alpha=0.5,new}$ ;
(27) end if
(28) **Updating  $ind_{\alpha=1,elite}$ **
(29) generate an individual  $ind_{\alpha=1,new}$  through  $PV_{\alpha=1}$ ;
(30)  $ind_{\alpha=1,new} = crossover(ind_{\alpha=1,new}, ind_{\alpha=1,elite})$ ;
(31)  $ind_{\alpha=1,new} = crossover(ind_{\alpha=1,new}, ind_{elite})$ ;
(32)  $[winner, loser] = compete(ind_{\alpha=1,new}, ind_{\alpha=1,elite})$ ;
(33) updatePV ( $winner, loser$ );
(34) if  $winner == ind_{\alpha=1,new}$  then
(35)  $ind_{\alpha=1,elite} = ind_{\alpha=1,new}$ ;
(36) end if
(37) **Updating  $ind_{elite}$ **
(38) generate 2 individuals through  $PV_{\alpha=0}$ ,  $PV_{\alpha=0.5}$  and  $PV_{\alpha=1}$ , respectively;
(39) NonDominatedSort;
(40) determine current generation's knee solution  $ind'_{elite}$ ;
(41)  $[winner, loser] = compete(ind_{elite}, ind'_{elite})$ ;
(42) if  $winner == ind'_{elite}$  then
(43)  $ind_{elite} = ind'_{elite}$ ;
(44) end if
(45)  $t = t + 1$ ;
(46) end while
(47) return  $ind_{elite}$ ;

```

ALGORITHM 3: Compact multiobjective particle swarm optimization algorithm driven by knee solution.

TABLE 1: Description on the OAEI's tracks.

Track	Subtask	Ontology	Scale
Anatomy	MA-HA	Adult Mouse Anatomy (MA)	2,744 classes
		Human Anatomy (HA)	3,304 classes
Large biomed	FMA-SNOMED SNOMED-NCI	Foundation Model of Anatomy (FMA)	78,989 classes
		Systemized Nomenclature of Medicine (SNOMED)	122,464 classes
		National Cancer Institute Thesaurus (NCI)	66,724 classes
Disease and phenotype	HP-MP DOID-ORDO	Human Phenotype Ontology (HP)	33,205 classes
		Mammalian Phenotype Ontology (MP)	32,298 classes
		Human Disease Ontology (DOID)	24,034 classes
		Orphanet Rare Disease Ontology (ORDO)	68,009 classes

TABLE 2: Friedman’s test on the alignment’s quality. Each value represents the f-measure, and the number in round parentheses is the corresponding computed rank.

	Anatomy	FMA-NCI	FMA-SNOMED	NCI-SNOMED	HP-MP	DOID-ORDO	Average
AML [27]	0.94 (1.5)	0.93 (1.5)	0.83 (2)	0.80 (2)	0.84 (3)	0.64 (7)	0.83 (2.83)
LogMap [28]	0.89 (4)	0.92 (3)	0.79 (3)	0.77 (3)	0.85 (2)	0.84 (2)	0.84 (2.83)
XMap [29]	0.89 (4)	0.86 (5.5)	0.77 (4)	0.69 (4)	0.47 (8.5)	0.70 (6)	0.73 (2.33)
DOME [30]	0.76 (6)	0.86 (5.5)	0.33 (9)	0.64 (6)	0.47 (8.5)	0.60 (9)	0.61 (7.33)
POMAP++ [31]	0.89 (4)	0.88 (4)	0.40 (8)	0.68 (5)	0.68 (6)	0.83 (3)	0.73 (5.00)
EA [8]	0.68 (9)	0.73 (8)	0.64 (6)	0.53 (8)	0.70 (5)	0.61 (8)	0.65 (7.33)
ABC [6]	0.72 (7)	0.75 (7)	0.64 (6)	0.62 (7)	0.62 (7)	0.72 (5)	0.68 (6.50)
PSO [10]	0.70 (8)	0.72 (9)	0.64 (6)	0.51 (9)	0.72 (4)	0.75 (4)	0.67 (6.67)
CMPSO-K	0.94 (1.5)	0.93 (1.5)	0.85 (1)	0.82 (1)	0.86 (1)	0.86 (1)	0.88 (1.17)

TABLE 3: Holm’s test on the alignment’s quality.

i	Approach	z value	Unadjusted p value	$\alpha/k - i, \alpha = 0.05$
8	XMap	2.33	0.045	0.050
7	AML	2.83	0.004	0.025
6	LogMap	2.83	0.004	0.016
5	POMAP++	5.00	1.28×10^{-7}	0.012
4	ABC	6.50	9.82×10^{-8}	0.010
3	PSO	6.67	3.79×10^{-8}	0.008
2	EA	7.33	5.38×10^{-11}	0.007
1	DOME	7.33	5.38×10^{-11}	0.006

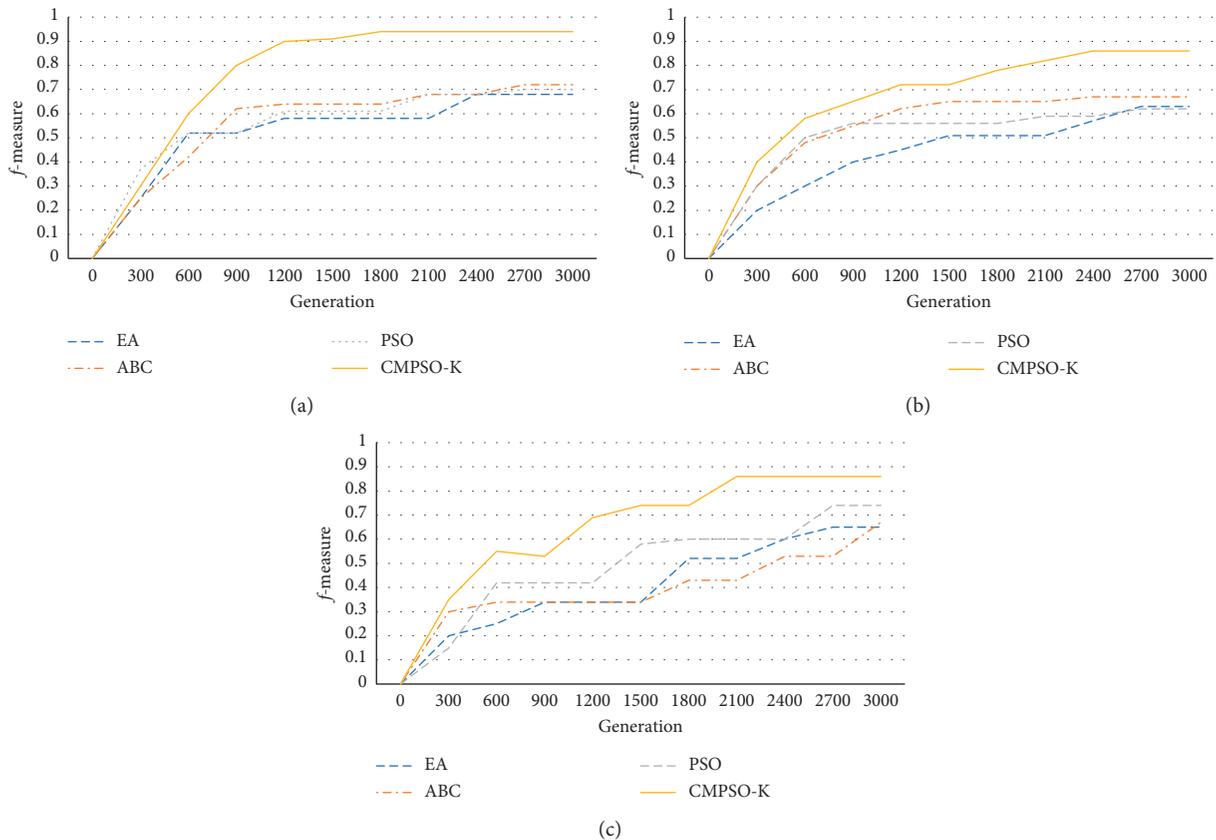


FIGURE 4: Comparison on the converging speed, (a) comparison on the anatomy track, (b) comparison on the large biomed track, and (c) comparison on the disease and phenotype track.

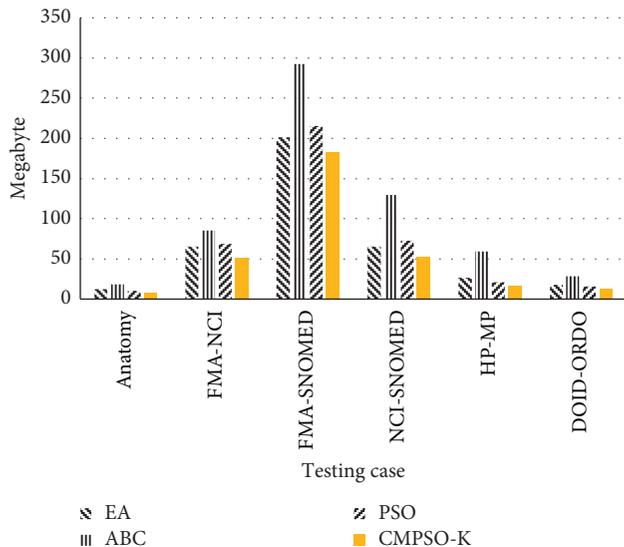


FIGURE 5: Comparison on the memory consumption.

hypothesis is that all the matchers are equivalent, and if the computed value χ_r^2 must be equal to or greater than the tabled critical chi-square value at the specified level of significance [34], this hypothesis can be rejected. In this work, a level of significance = 0.05 is chosen, and the critical value for 8 degrees of freedom (since there are 9 matchers), i.e., $\chi_{0.05}^2 = 15.507$, needs to be considered.

In Table 2, computed $\chi_r^2 = 54.43$, which is greater than 15.507, and therefore, the null hypothesis can be rejected, and Holm's test can be further carried out. As shown in Table 2, since the CMPSO-K-based matcher ranks with the lowest value, it is set as a control matcher that will be compared with others.

In Holm's test, z value is the testing statistic for comparing the i -th and j -th matchers, which is used for finding the p value that is the corresponding probability from the table of the normal distribution. p value is then compared with $\alpha = 0.05$, which is an appropriate level of significance. According to Table 3, it is possible to state that our approach statistically outperforms other biomedical ontology matchers on f -measure at 5% significance level. Since the multiobjective evolving mechanism can better trade-off two objectives and the knee solution can effectively guide the algorithm's search direction, CMPSO-K's solutions are much better than other SIAs.

Figures 4 and 5, respectively, compare CMPSO-K with other SIAs on the memory consumption and converging speed. As can be seen from the figures, CMPSO-K can significantly improve the converging speed and reduce the memory consumption, which shows the effectiveness of the compact encoding mechanism and the compact evolutionary operators. To sum up, CMPSO-K-based ontology matching technique can efficiently optimize the biomedical ontology alignments.

6. Conclusion and Future Work

To optimize biomedical ontology alignment's quality, in this paper, a discrete multiobjective optimal model is constructed, and a hybrid similarity metric to distinguish the

heterogeneous biomedical concepts, a CMPSO-K-based ontology matching technique is then proposed for addressing it. Compared with the most existing SIA-based ontology matching techniques, CMPSO-K takes into consideration both the algorithm's performance and the DM's preference. Accordingly, three methods are proposed to achieve this goal: (1) the compact encoding mechanism for saving the memory consumption and runtime; (2) multi-objective decomposition and evolutionary mechanism for trading-off different objectives; and (3) the max-min strategy for determining the knee solution and guiding the algorithm's searching direction. Our work presents a novel compact multiobjective evolutionary framework that can improve the efficiency of the current SIA-based ontology matching technique.

In the experiment, the f -measure values obtained by CMPSO-K outperform all the other competitors, which shows that CMPSO-K can effectively optimize the ontology alignments. In particular, the quality of alignment of CMPSO-K is better than EA ABC and PSO, which shows that CMPSO-K's multiobjective evolutionary mechanism driven by the knee solution can effectively trade-off two optimal objectives and find the better solution. Moreover, since the compact encoding mechanism uses PVs to represent the swarms and the compact evolutionary operators can simplify the population-based PSO's evolving process, CMPSO-K can significantly reduce the memory consumption and runtime. Since none of the similarity measures can effectively distinguish all the heterogeneous concepts in any situations, it is necessary to aggregate several similarity measures to improve the result's precision. We utilize a hybrid similarity measure which combines three kinds of similarity measures to calculate the entity similarity value, and therefore, CMPSO-K's results are significantly higher than other systems that only take into consideration one or two categories of similarity measure, such as DOME, POMAP++, and LogMap. However, AML applies too many similarity measures that lead to the conflicting results, which decrease its recall value. Thus, how many similarity measures should be selected and combined to ensure the quality of the alignment will be one of our future work. Last but not the least, CMPSO-K takes into consideration of DM's preference and utilizes the knee solution to guide the algorithm's search direction, which can effectively trade-off two optimal objectives and find better solutions.

To further improve the matching process's efficiency, in the future, we will be interested in developing a biomedical ontology partitioning technique to split two ontologies into disjoint ontology segments so that the large-scale problem can be converted into several small-scale segment matching problems. After that, the parallel computation can also be further utilized to match the similar segments and improve the matching process's performance. With respect to the similarity measure, we would like to develop an adaptive biomedical concept similarity measure framework, which can instantiate an effective similarity metric according to two biomedical ontologies' heterogeneous characteristics.

Data Availability

The data used to support this study can be found in <http://oaei.ontologymatching.org>.

Conflicts of Interest

The authors declare that they have no conflicts of interest in the work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 61503082), the Natural Science Foundation of Fujian Province (no. 2016J05145), the Program for New Century Excellent Talents in Fujian Province University (no. GY-Z18155), the Program for Outstanding Young Scientific Researcher in Fujian Province University (no. GY-Z160149), and the Scientific Research Foundation of Fujian University of Technology (nos. GY-Z17162 and GY-Z15007).

References

- [1] E. Zitzler and 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.
- [2] S. Sudeng and N. Wattanapongsakorn, "A knee-based multi-objective evolutionary algorithm: an extension to network system optimization design problem," *Cluster Computing*, vol. 19, no. 1, pp. 411–425, 2016.
- [3] S. Sudeng and N. Wattanapongsakorn, "A decomposition-based approach for knee solution approximation in multi-objective optimization," in *Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC)*, pp. 3710–3717, Vancouver, Canada, July 2016.
- [4] 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.
- [5] A. Alves, K. Revoredio, and F. Baião, "Ontology alignment based on instances using hybrid genetic algorithm," in *Proceedings of the 7th International Conference on Ontology Matching*, pp. 242–243, Rio de Janeiro, Brazil, November 2012.
- [6] Y. He, X. Xue, and S. Zhang, "Using artificial bee colony algorithm for optimizing ontology alignment," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 4, pp. 766–773, 2017.
- [7] J. Martinez-Gil and J. F. Aldana-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.
- [8] J. Wang, Z. Ding, and C. Jiang, "Gaom: Genetic algorithm based ontology matching," in *Proceedings of the IEEE Asia-Pacific Conference on Services Computing (APSCC'06)*, pp. 617–620, Guangzhou, China, December 2006.
- [9] X. Xue and Y. Wang, "Improving the efficiency of NSGA-II based ontology aligning technology," *Data & Knowledge Engineering*, vol. 108, pp. 1–14, 2017.
- [10] J. Bock and J. Hettenhausen, "Discrete particle swarm optimisation for ontology alignment," *Information Sciences*, vol. 192, pp. 152–173, 2012.
- [11] W. Z. Guo, J. Y. Chen, G. L. Chen, and H. F. Zheng, "Trust dynamic task allocation algorithm with Nash equilibrium for heterogeneous wireless sensor network," *Security and Communication Networks*, vol. 8, no. 10, pp. 1865–1877, 2015.
- [12] G. Liu, W. Guo, R. Li, Y. Niu, and G. Chen, "XGRouter: high-quality global router in X-architecture with particle swarm optimization," *Frontiers of Computer Science*, vol. 9, no. 4, pp. 576–594, 2015.
- [13] W. Guo, J. Li, and G. Chen, "A PSO-optimized real-time fault-tolerant task allocation algorithm in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 12, pp. 3236–3249, 2014.
- [14] H. Guo, B. Liu, D. Cai, and T. Lu, "Predicting protein-protein interaction sites using modified support vector machine," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 3, pp. 393–398, 2018.
- [15] G. Alexandru-Lucian and A. Iftene, "Using a genetic algorithm for optimizing the similarity aggregation step in the process of ontology alignment," in *Proceedings of the 9th Roedunet International Conference*, pp. 118–122, Sibiu, Romania, June 2010.
- [16] 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.
- [17] X. Xue, J. Lu, and J. Chen, "Using NSGA-III for optimising biomedical ontology alignment," *CAAI Transactions on Intelligence Technology*, vol. 4, no. 3, pp. 135–141, 2019.
- [18] H. Seada and K. Deb, "U-NSGA-III: a unified evolutionary optimization procedure for single, multiple, and many objectives: proof-of-principle results," in *Proceedings of the International conference on evolutionary multi-criterion optimization*, pp. 34–49, Guimarães, Portugal, March 2015.
- [19] S. C. Chu, X. Xue, J. S. Pan, and X. Wu, "Optimizing ontology alignment in vector space," *Journal of Internet Technology*, vol. 21, no. 1, pp. 15–23, 2020.
- [20] X. Xue, "Compact firefly algorithm for matching biomedical ontologies," *Knowledge and Information Systems*, 2020.
- [21] C. J. Van Rijsbergen, "Foundation of evaluation," *Journal of Documentation*, vol. 30, no. 4, pp. 365–373, 1974.
- [22] X. Xue, J. Chen, and X. Yao, "Efficient user involvement in semi-automatic ontology matching," *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1–11, 2019.
- [23] G. Stoilos, G. Stamou, and S. Kollias, "A string metric for ontology alignment," in *4th International Semantic Web Conference (ISWC 2005)*, pp. 623–637, Springer, Berlin, Germany, 2005.
- [24] O. Bodenreider, "The unified medical language system (umls): integrating biomedical terminology," *Nucleic Acids Research*, vol. 32pp. D267–D270, suppl_1, 2004.
- [25] G. R. Harik, F. G. Lobo, and D. E. Goldberg, "The compact genetic algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 287–297, 1999.
- [26] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.
- [27] D. Faria, C. Pesquita, T. Tervo, F. M. Couto, and I. F. Cruz, "AML and AMLC results for OAEI 2019," in *Proceedings of the 18th International Semantic Web Conference (ISWC 2019)*, pp. 101–106, Auckland, New Zealand, October 2019.
- [28] E. Jimenez-Ruiz, "LogMap family participation in the OAEI 2019," in *Proceedings of the 18th International Semantic Web*

- Conference (ISWC 2019)*, pp. 160–163, London, UK, January 2019.
- [29] W. E. Djeddi and M. T. Khadir, “XMAP: a novel structural approach for alignment of OWL-full ontologies,” in *Proceedings of the 2010 International Conference on Machine and Web Intelligence*, pp. 368–373, Algiers, Algeria, October 2010.
 - [30] S. Hertling and H. Paulheim, “DOME results for OAEI 2019,” in *Proceedings of the 18th International Semantic Web Conference (ISWC 2019)*, pp. 123–130, Auckland, New Zealand, October 2019.
 - [31] A. Laadhar, F. Ghazzi, I. Megdiche, F. Ravat, O. Teste, and F. Gargouri, “Pomap++ results for OAEI 2019: fully automated machine learning approach for ontology matching,” in *Proceedings of the 18th International Semantic Web Conference (ISWC 2019)*, pp. 169–174, Auckland, New Zealand, October 2019.
 - [32] M. Friedman, “The use of ranks to avoid the assumption of normality implicit in the analysis of variance,” *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, 1937.
 - [33] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.
 - [34] M. S. Nash, “Handbook of parametric and nonparametric statistical procedures,” *Technometrics*, vol. 43, no. 3, p. 374, 2001.