



Ecology and Biodiversity Ontology Alignment for Smart Environment *via* Adaptive Compact Evolutionary Algorithm

Xingsi Xue^{1,2*} and Pei-Wei Tsai³

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

² Intelligent Information Processing Research Center, Fujian University of Technology, Fuzhou, China, ³ Department of Computer Science and Software Engineering, Swinburne University of Technology, Hawthorn, VIC, Australia

OPEN ACCESS

Edited by:

Yu Xue,
Nanjing University of Information
Science and Technology, China

Reviewed by:

Ying Huo,
Nanjing Institute of Technology
(NJIT), China
Xuejian Zhao,
Nanjing University of Posts and
Telecommunications, China
Li Chaorong,
Yibin University, China

*Correspondence:

Xingsi Xue
jack8375@gmail.com

Specialty section:

This article was submitted to
Sustainable and Intelligent
Phytoprotection,
a section of the journal
Frontiers in Plant Science

Received: 16 February 2022

Accepted: 14 March 2022

Published: 14 April 2022

Citation:

Xue X and Tsai P-W (2022) Ecology
and Biodiversity Ontology Alignment
for Smart Environment *via* Adaptive
Compact Evolutionary Algorithm.
Front. Plant Sci. 13:877120.
doi: 10.3389/fpls.2022.877120

Smart Environment (SE) focuses on the initiatives for healthy living, where ecological issues and biodiversity play a vital role in the environment and sustainability. To manage the knowledge on ecology and biodiversity and preserve the ecosystem and biodiversity simultaneously, it is necessary to align the data entities in different ecology and biodiversity ontologies. Since the problem of Ecology and Biodiversity Ontology Alignment (EBOA) is a large-scale optimization problem with sparse solutions, finding high-quality EBOA is an open challenge. Evolutionary Algorithm (EA) is a state-of-the-art technique in the ontology aligning domain, and this study further proposes an Adaptive Compact EA (ACEA) to address the problem of EBOA, which uses semantic reasoning to reduce searching space and adaptively guides searching direction to improve the algorithm's performance. In addition, we formally model the problem of EBOA as a discrete optimization problem, which maximizes the alignment's completeness and correctness through determining an optimal entity corresponding set. After that, a hybrid entity similarity measure is presented to distinguish the heterogeneous data entities, and an ACEA-based aligning technique is proposed. The experiment uses the famous Biodiversity and Ecology track to test ACEA's performance, and the experimental results show that ACEA-based aligning technique statistically outperforms other EA-based and state-of-the-art aligning techniques.

Keywords: ecology ontology, biodiversity ontology, ontology alignment, adaptive compact evolutionary algorithm, semantic reasoning

1. INTRODUCTION

Smart Environment (SE) is a knowledge-based system that focuses on initiatives for healthy living with an emphasis on environment and sustainability, where ecological issues and biodiversity play a vital role in urban citizens' life (Kumar, 2020). In particular, ecology research focuses on the ecosystems, habitat restoration practices, and communities of interest, which is of help to further our understanding of the environment and opportunities to affect change; biodiversity investigates the species' variability as well as their relationship to the environment. Since ecology and biodiversity are the most complex entities on this planet, the corresponding knowledge is usually modeled with the ontology (Madin et al., 2008), which is a powerful domain knowledge

modeling technique (Berners-Lee et al., 2001). Currently, more and more ecology and biodiversity ontologies have been developed, such as Environment Ontology (ENVO) and Plant Trait Ontology (PTO). However, since they are developed and maintained independently, a concept might be defined with different contexts, granularities, and terminologies, yielding the ontology heterogeneity problem (Karam et al., 2020). Examples of heterogeneous ecology and biodiversity ontologies are shown in **Table 1**.

Therefore, to preserve the ecosystem and biodiversity simultaneously and manage the knowledge on ecology and biodiversity, it is necessary to link the data entities in different ecology and biodiversity ontologies, which is the so-called Ecology and Biodiversity Ontology Alignment (EBOA).

Aligning ecology and biodiversity ontologies aims at finding a 0–1 Aligning Matrix (AM), whose element denotes whether two corresponding entities (the source ontology's entities in row and the target ontology's entities in column) are mapped by 1 or not by 0. Since the scale of the ecology and biodiversity ontologies are usually large, and the constraint of single cardinality on the aligning result, the problem of EBOA needs to find a large-scale AM (the number of its row and column is large) with sparse solutions (most of its element values are 0). Due to the large search space and richness of semantic meaning on different data entities, it is a complex task of aligning ecology and biodiversity ontologies. In recent years, Evolutionary Algorithm (EA) (Mirjalili, 2019) has become a popular technique in the ontology aligning domain (Acampora et al., 2013; Xue et al., 2018). Due to the population-based evolving paradigm, the classic EA's searching performance is low in terms of memory consumption and run time. To improve the efficiency, a new category of EA with the name Compact EA (CEA) is presented, which uses compact encoding mechanism to describe the whole population with probability estimation. CEA mimics EA's searching process by simplifying the evolving operators, but it is easy to get stuck in the local optima especially when two ontologies' scale is large. To overcome this drawback, this study further proposes an Adaptive CEA (ACEA), which uses the semantic reasoning to filter the negative correspondences, and adaptively alters the algorithm's searching direction to explore the unknown region. In the following, we list the contributions of this study:

- The optimization model of the problem of EBOA is presented;
- A hybrid entity similarity measure is proposed to distinguish the heterogeneous ecology and biodiversity data entities;
- An ACEA-based aligning technique is proposed, which uses semantic reasoning to reduce searching space, and adaptively guides the searching direction to efficiently align the ecology and biodiversity ontologies.

The introduction process of this study is as follows: before defining the problem of EBOA and entity similarity measure (Section 3), the EA-based aligning techniques are overviewed (Section 2); after that, the problem-specific ACEA is presented (Section 4), followed by the experimental results (Section 5); and finally, we draw the conclusion on this article's study (Section 6).

2. EVOLUTIONARY ONTOLOGY ALIGNING TECHNIQUE

With the rapid development of ontology engineering, the scale of an ontology has grown from hundreds of entities to tens of thousands of entities, and the semantic representation of the entities also become more and more complex, which makes the determination of a high-quality ontology alignment become an open challenge (Shvaiko and Euzenat, 2011). Essentially, the ontology aligning problem can be regarded as an optimizing issue that aims at maximizing the quality of final alignment, and EA-based aligning techniques have become a popular methodology to address this problem.

The first EA-based ontology aligning technique is proposed by Martinez-Gil et al. (2008) which tries to find an optimal way of combining different similarity measures for determining the final alignment. Later on, researchers have done a lot to improve this category of EA-based aligning techniques. Based on this study, Ginsca and Iftene (2010) further optimize the threshold for filtering final alignment. Acampora et al. (2012) propose a Hybrid EA (HEA) to improve the efficiency of classic EA's performance. Alves et al. (2012) further use the instance-level information in an ontology to construct the similarity measure and then use HEA to combine it with others. Currently, it is necessary to enhance the performance of population-based EA in terms of running time and memory so as to address the large-scale aligning task, such as addressing the problem of EBOA where the ontology contains tens of thousands of entities. To this end, an efficiency improvement strategy should be introduced to improve classic EA's performance. The most popular way is the utilization of a compact encoding based evolving paradigm, which describes the population with a probability distribution, and on this basis, it approximates the classic EA's evolving process. The first CEA-based aligning technique is proposed in Xue et al. (2015), which executes the evolving process by one Probability Vector (PV). According to the experimental results, CEA is able to significantly reduce EA's running time and memory consumption without sacrificing the alignment's quality. Later on, a Hybrid CEA (HCEA) (Xue and Wang, 2015a) and a CEA with a Re-sample Inheritance Mechanism (RIM) (Xue and Liu, 2022) are respectively proposed to further enhance CEA's performance. To address the large-scale aligning task, a divide-and-conquer

TABLE 1 | The examples of heterogeneous ecology and biodiversity ontologies.

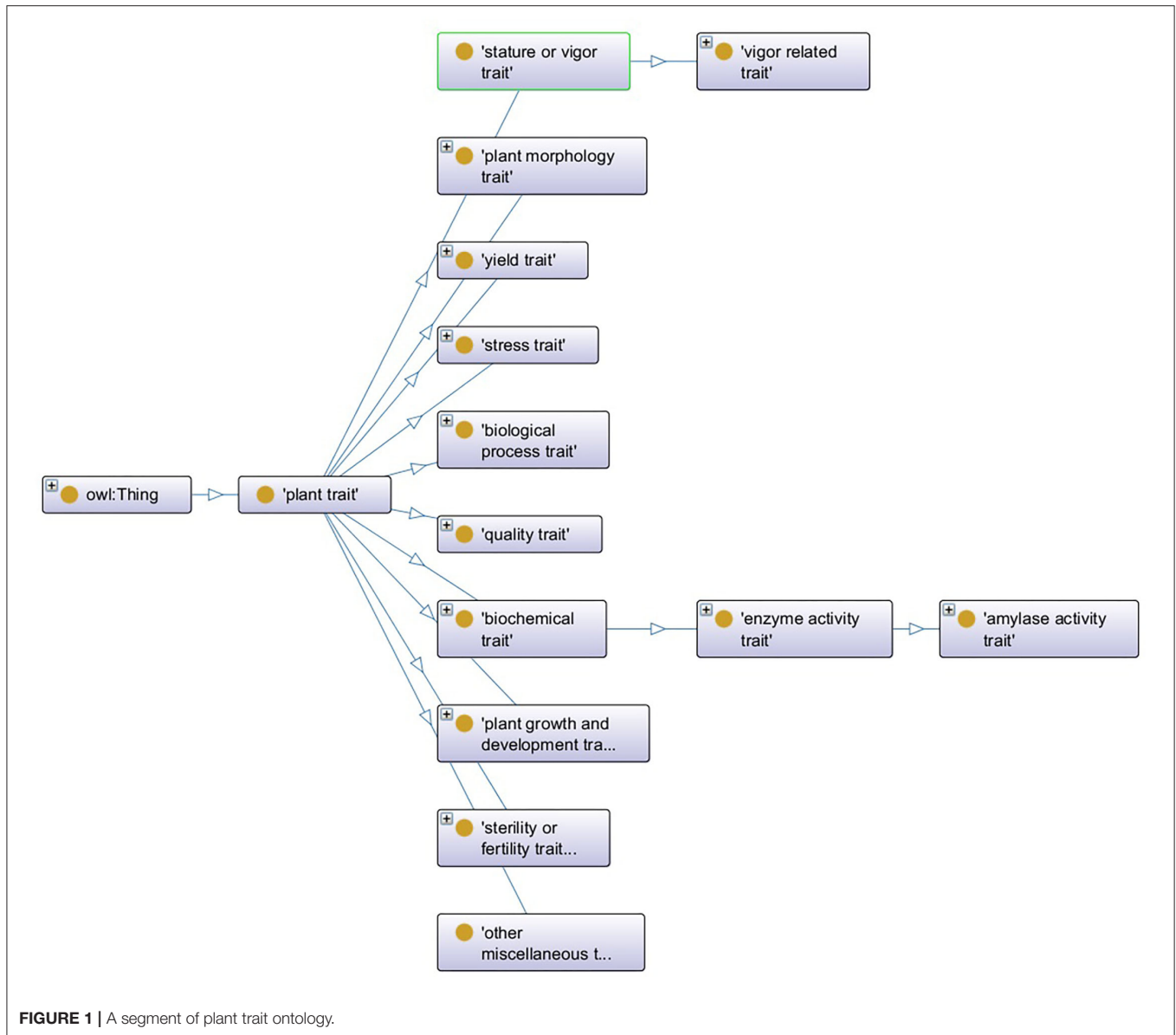
ENVO ontology^a	SWEET ontology^b
Divergent tectonic movement	Plate divergence
Tectonic movement	Continental drift
FLOTO ontology^c	PTO ontology^d
Inflorescence absent	Inflorescenceless
Leaf alternate placement	Phyllotaxy

^a<http://agroportal.lirmm.fr/ontologies/ENVO>.

^b<https://bioportal.bioontology.org/ontologies/SWEET>.

^c<http://agroportal.lirmm.fr/ontologies/FLOPO>.

^d<http://agroportal.lirmm.fr/ontologies/TO>.



method is also presented, which is of help to reduce HCEA's searching space (Xue and Wang, 2015b; Xue and Zhang, 2021).

Existing EA-based aligning approaches need to maintain each similarity measure's corresponding AM, and on this basis, the optimization on the alignment can be executed, which greatly raises the computational complexity. In this study, we try to directly find a set of correspondences with the given similarity measure, which only needs to save several entity pairs' similarity value instead of maintaining all the similarity measures' corresponding entity pairs' similarity values. In addition, classic CEA only uses one PV to execute the optimizing process, which makes it easy to get stuck in the local optima when facing a complex optimization problem. To overcome this drawback, our approach proposes to adaptively maintain several Probability Matrices (PMs) to guide the algorithm's searching direction. Finally, since the problem

of EBOA is a large-scale issue with sparse solutions, we propose semantic reasoning based initialization to reduce the algorithm's searching space and evenly distribute the AM's element values.

3. ECOLOGY AND BIODIVERSITY ONTOLOGY ALIGNMENT

3.1. The Problem of Ecology and Biodiversity Ontology Alignment

An ontology consists of the concepts, the datatype properties, and the object properties, which are referred to as entities (Xue et al., 2021). **Figure 1** shows a segment of PTO, where the oval symbol describes the concept's name, e.g., "plant trait," the arrow line is the object property or relationship between two concepts,

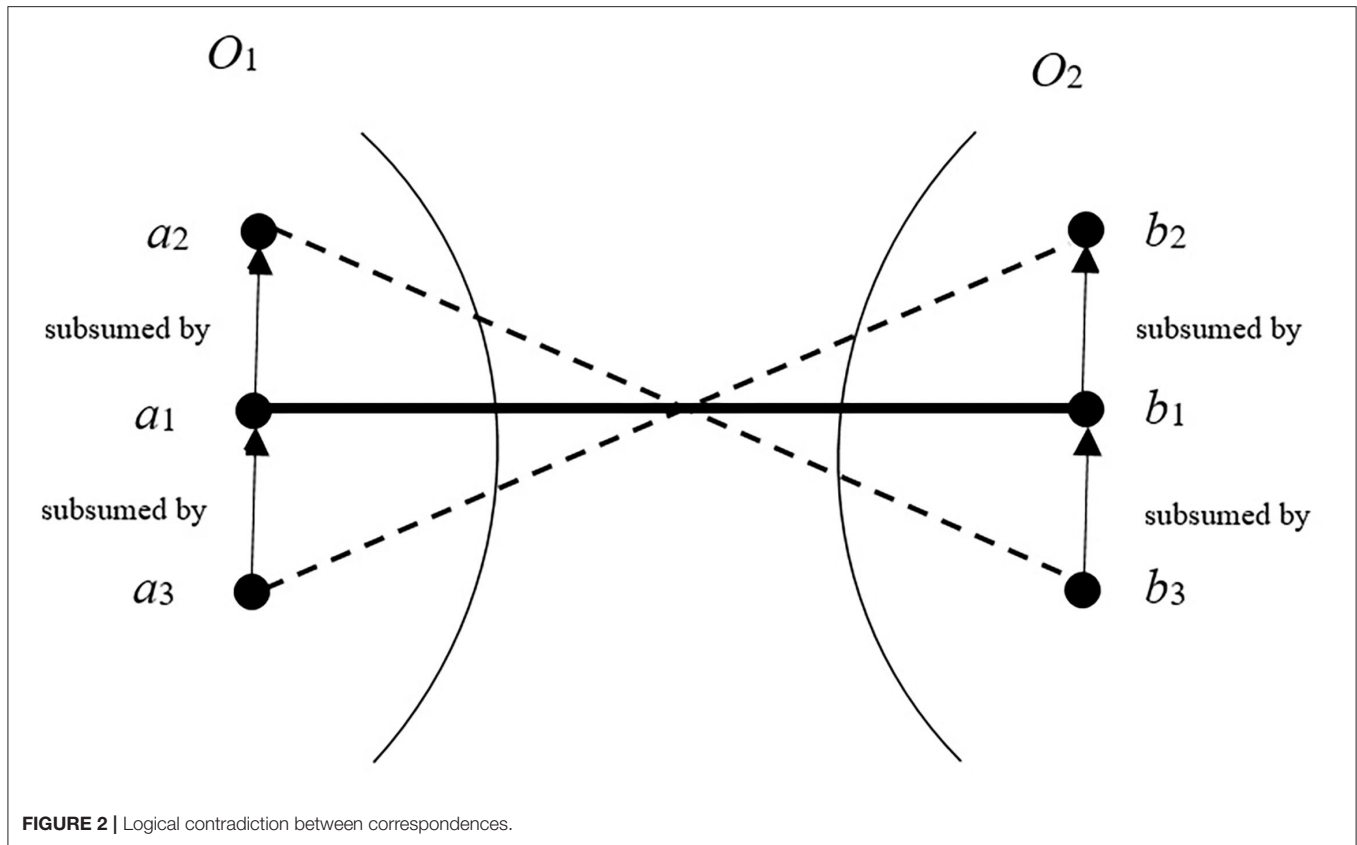


FIGURE 2 | Logical contradiction between correspondences.

e.g., the concept “quantity trait” is subsumed by “plant trait,” and each concept has several datatype properties to describe its feature, e.g., the concept “plant trait” has the datatype property “definition” whose value is “A plant trait (TO:0000387) that is the commercial and /or economical value of the plant product, or its overall improvement.”

An entity correspondence consists of 4 elements, i.e., $e_1, e_2, rel, conf$, where e_1 and e_2 are respectively two ontologies’ entities, rel is the type of their relationship (typically the equivalence \equiv), and $conf$ denotes the confidence level that the correspondence holds, which is often measured by e_1 and e_2 ’s similarity value. For example, in Table 1, $(ENVO : Tectonicmovement, SWEET : Continentaldrift, \equiv, 0.9)$ denote a correspondence between the concept “Tectonic movement” from ENVO and the concept “Continental drift” from SWEET, their relationship is the equivalence \equiv , and this correspondence’s confidence value is 0.9. The ontology alignment is a correspondence set, whose quality is typically evaluated with recall, precision, and f-measure (Rijsberge, 1975). Since reference alignment is often not available in the practical aligning tasks, this study proposes three new metrics to approximate them. According to Wang et al. (2006), $recall(A)$ can be estimated by the number of correspondences found in A , i.e., $recall(A) = norm(|A|)$, $precision(A)$ can be evaluated by the average similarity values of all the correspondences in A , i.e., $precision(A) = \frac{\sum sim(corr_i)}{|A|}$ where $corr_i$ is i -th correspondence

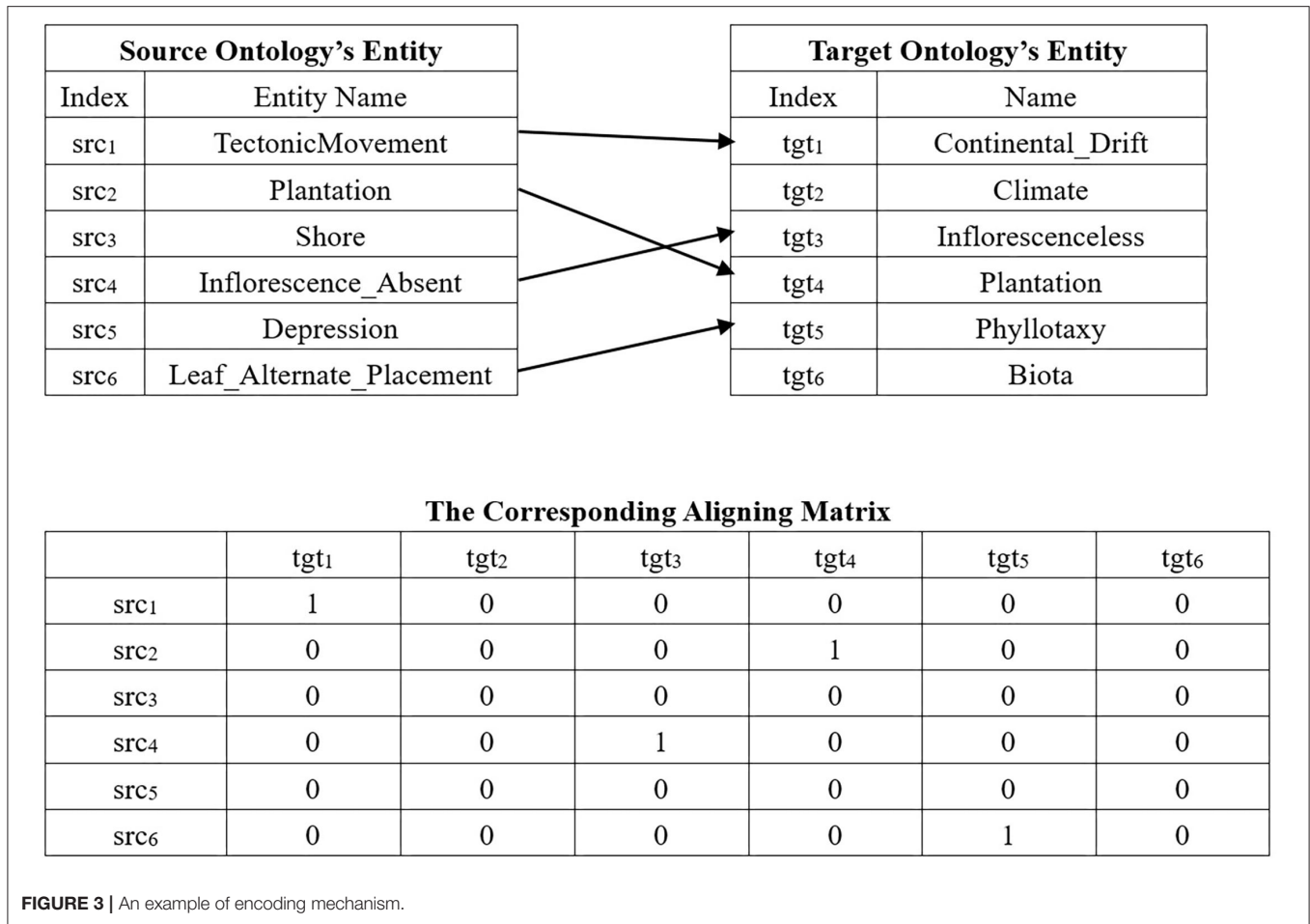
in A . To evaluate an alignment’s quality in terms of both completeness and correctness, a comprehensive metric $f - measure(A) = \frac{2 \times recall(A) \times precision(A)}{recall(A) + precision(A)}$ is presented, which calculate an alignment’s harmony mean of its $recall(A)$ and $precision(A)$. Given two ontologies O_1 and O_2 , a 0–1 matrix M and its corresponding alignment A , the problem of EBOA is defined as follows:

$$\begin{cases} \max & f(M) \\ \text{s.t.} & M = [m_{i,j}]_{|O_1| \times |O_2|} \\ & m_{i,j} \in \{0, 1\}, i = 1, 2, \dots, |O_1| \end{cases} \quad (1)$$

where $|O_1|$ and O_2 are respectively O_1 and O_2 ’s entity numbers, and $f(M)$ is equal to $f - measure(A)$, and the model of EBOA aims at finding an optimal matrix by maximizing its corresponding alignment’s $f - measure$. In particular, the decision variable is a 0–1 matrix whose row and column are respectively two ontologies’ entities, and its element value 1 means two corresponding entities are mapped, and 0 means not.

3.2. Entity Similarity Measure

The entity similarity measure calculates two entities’ similarity value $conf$, which is a real number in $[0,1]$. $conf = 1$ means two entities are the same, and $conf = 0$ means they are totally different. To improve the result’s confidence, usually, it is necessary to comprehensively consider three categories



of similarity measures, i.e., string-based, linguistic-based, and structure-based similarity measures (Xue and Huang, 2022). To this end, this study proposes a hybrid entity similarity measure to comprehensively calculate the similarity value: (1) given two entities e_1 and e_2 , before calculating their similarity value, the numbers, punctuations, and stop-words in their names are first removed; (2) the strings are split into the words, which are further lemmatized and stemmed; (3) in each word set, the word will be removed if it is the same literally or synonymous to the other one in Wordnet (Miller, 1995), and we obtain two word sets s_1 and s_2 ; finally, e_1 and e_2 's similarity value is the same as the similarity value of two string s_1 and s_2 :

$$sim(s_1, s_2) = \frac{\sum_{i=1}^{|W_1|} \max_{j=1 \dots |W_2|} (sim(w_{1,i}, w_{2,j})) + \sum_{j=1}^{|W_2|} \max_{i=1 \dots |W_1|} (sim(w_{1,i}, w_{2,j}))}{|W_1| + |W_2|} \tag{2}$$

where $|W_1|$ and $|W_2|$ are respectively the numbers of words in W_1 and W_2 , and $w_{1,i}$ and $w_{2,j}$ are respectively the i th and j th words in W_1 and W_2 ; and $sim(w_{1,i}, w_{2,j})$ is calculated with

Wordnet and N-gram measure (Kondrak, 2005):

$$sim(w_{1,i}, w_{2,j}) = \begin{cases} 1, & \text{two words are synonyms} \\ & \text{in Wordnet} \\ N\text{-gram}(w_{1,i}, w_{2,j}), & \text{otherwise} \end{cases} \tag{3}$$

4. ADAPTIVE COMPACT EVOLUTIONARY ALGORITHM

Adaptive compact evolutionary algorithm adaptively maintains PMs according to the current generation's population information, which is able to help the algorithm effectively exploit the unexplored domains. In addition, ACEA uses the anchor-based semantic reasoning strategy to initialize the individual and refine the new individuals, which can effectively reduce the algorithm's searching domain. The framework of ACEA is presented in **Algorithm 1**, which takes as input two ontologies to be aligned, and the output the alignment with best fitness value.

In the next, we successively present the Semantic Reasoning Based Initialization and adaptive PM maintenance.

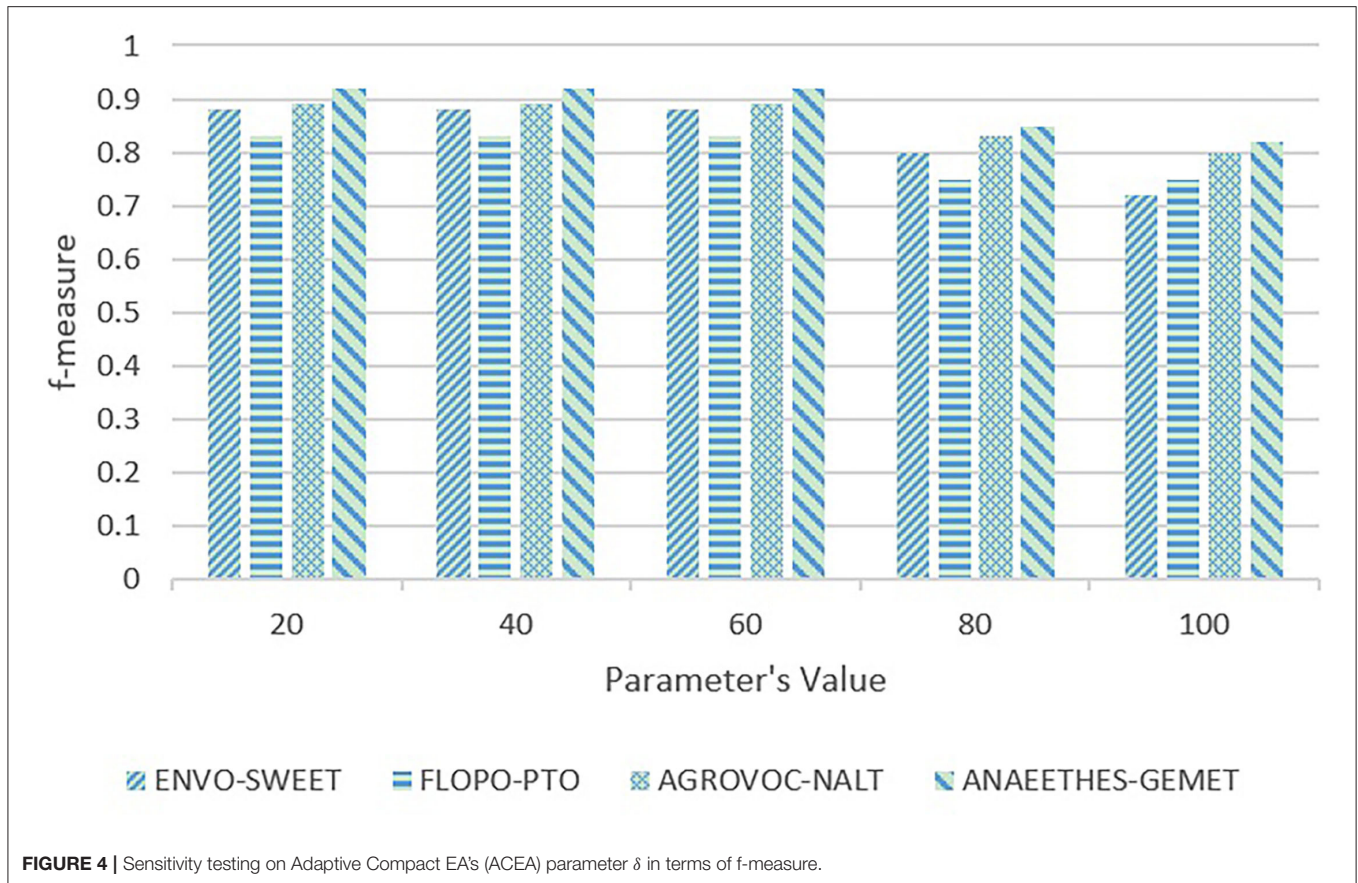


FIGURE 4 | Sensitivity testing on Adaptive Compact EA's (ACEA) parameter δ in terms of f-measure.

Algorithm 1 The Framework of Adaptive Evolutionary Algorithm

```

 $PM_{num} = 2$ ; //Initialize the number of PM
for  $i = 0$ ;  $i < PM_{num}$ ;  $i++$  do
   $AM_{elite}^i = initializeAM(PM^i)$ ; //Initialize the elite AM
end for
 $gen = 0$ ;
while  $gen < maxGen$  do
  for  $i = 0$ ;  $i < PM_{num}$ ;  $i++$  do
     $Update(PM^i)$ ; //Execute the evolutionary operator and update PM
  end for
   $PM_{num} = adaptivePMMaintenance()$ ; //Adaptively maintain the PMs
   $gen = gen + 1$ ;
end while

```

4.1. Semantic Reasoning Based Initialization

Typically, the correspondence with a high confidence value is referred to as Positive Anchor (PA), and the one with a low confidence value is called the Negative Anchor (NA) (Wang, 2010). The concepts in an ontology are modeled with the

hierarchy graph (Chu et al., 2020), where the node denotes the concept and the edge represents the relationships between two concepts. **Figure 2** shows an example of correspondences' logical contradiction. As shown in the figure, the entities a_1 , a_2 , and a_3 are three entities of ontology O_1 , and the entities b_1 , b_2 , and b_3 belong to ontology O_2 . In O_1 (O_2), a_3 (b_3) is subsumed by a_1 (b_1), and a_1 (b_1) is subsumed by a_2 (b_2). Assuming the correspondence (a_1, b_1) is a PA, the correspondences (a_2, b_3) and (a_3, b_2) logically contradict with (a_1, b_1). It is obvious that the correspondences that contradict with some PA will not hold, and the confidence of correspondences that are in line with some NA should be reduced. According to this reasoning rule, the searching space of the algorithm can be reduced.

Each individual of ACEA is represented by a 0–1 matrix, i.e., the so-called AM. **Figure 3** shows an example of an encoding mechanism, wherein the top of the figure is a real alignment, and its corresponding AM is given below it. ACEA uses Probability Matrix (PM) to approximately describe the population, which has the same size as AM. PM's elements are the real number in $[0,1]$, which denotes the probability of being 1 with respect to the corresponding gene bit. Therefore, we can use PM to generate AM by comparing its elements with a random number in $[0,1]$.

Since the problem of EBOA is a large-scale issue with sparse solutions, it is necessary to evenly distribute the gene value

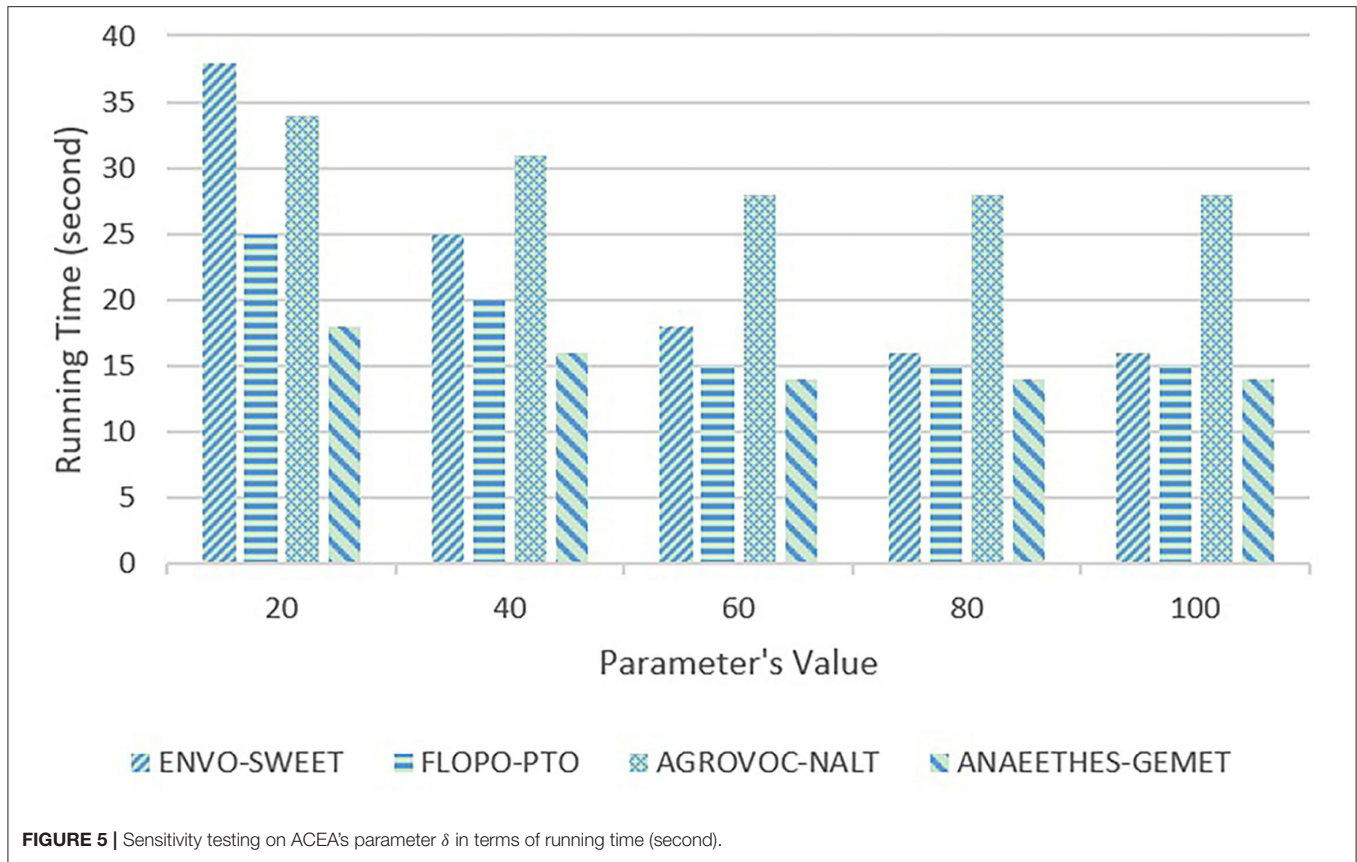


FIGURE 5 | Sensitivity testing on ACEA's parameter δ in terms of running time (second).

TABLE 2 | Comparisons among Adaptive Compact EA (ACEA), Compact EA (CEA), and Hybrid CEA (HCEA) in terms of mean f-measure and standard deviation.

Testing case	CEA f-measure (stdDev)	HCEA f-measure (stdDev)	ACEA f-measure (stdDev)
ENVO-SWEET	0.79 (0.03)	0.83 (0.01)	0.88 (0.01)
FLOPO-PTO	0.76 (0.01)	0.76 (0.02)	0.83 (0.01)
AGROVOC-NALT	0.78 (0.03)	0.80 (0.01)	0.89 (0.02)
ANAEETHES-GEMET	0.74 (0.02)	0.82 (0.01)	0.92 (0.01)

TABLE 3 | T-test on alignment's quality.

Testing case	(CEA, ACEA) t-value (p-value)	(HCEA, ACEA) t-value (p-value)
ENVO-SWEET	-5.69 (0.014)	-7.07 (0.009)
FLOPO-PTO	-9.89 (0.005)	-6.26 (0.012)
AGROVOC-NALT	-6.10 (0.012)	-8.04 (0.007)
ANAEETHES-GEMET	-16.09 (0.001)	-4.14 (0.026)

when initializing AM. Algorithm 2 shows the pseudo-code of initialization.

We first initialize AM by setting all its elements as 0 and determine the positive anchor set PAS with the similarity

Algorithm 2 Initialization

```

Initialize Aligning Matrix AM by setting all its elements as 0;
if Probability Matrix PM is not given then
    initialize all the elements in PM as 0.5;
end if
initialize the Positive Anchor Set PAS;
for i = 0; i < PAS.length; i++ do
    for j = 0; j < AM.row; j++ do
        for k = 0; k < AM.column; k++ do
            if (entityi, entityj) is contradicted with PAS; then
                PMi,j = 0;
            else
                if random(0, 1) < PMi,j then
                    AMi,j = 1;
                end if
            end if
        end for
    end for
end for
end for
    
```

measure. If probability matrix PM is not given, all its elements will be initialized as 0.5. Then, we compare all the correspondences in AM with PAS. If the correspondence is logically conflicted with PAS's correspondence, its AM and PM's values will be set as 0, otherwise, its value is decided by

comparing its corresponding PM 's value with a random number in $[0,1]$. Through semantic reasoning, the searching space can be significantly reduced, and initializing PM 's elements as 0.5 is also of help to ensure the even distribution of the gene values.

4.2. Updating Probability Matrix

Adaptive compact evolutionary algorithm combines the mechanisms of a classic EA with a competitive learning mechanism, which is effective to lead the algorithm to determine the optimal solution. To be specific, ACEA first uses its fitness function to evaluate its solution's fitness value by calculating its corresponding alignment's $f - measure'$, and through competitions between the individuals, the algorithm updates PM by moving it toward the elite individual. The process of updating PM is presented in **Algorithm 3**.

Algorithm 3 Updating Probability Matrix

```

 $AM^{new} = PM.generateAM();$ 
 $AM' = PM.generateAM();$ 
for  $i = 0; i < AM^{new}.row; i++$  do
  for  $j = 0; j < AM^{new}.column; j++$  do
    if  $random(0, 1) < 0.5$  then
       $AM_{ij}^{new} = AM'_{ij};$ 
    end if
  end for
end for
 $compete(AM^{new}, AM^{elite});$ 
if  $winner == AM^{new}$  then
   $AM^{elite} = AM^{new};$ 
end if
for  $i = 0; i < PM.row; i++$  do
  for  $j = 0; j < PM.column; j++$  do
    if  $AM_{ij}^{elite} == 1$  then
       $PM_{ij}^{elite} + = 0.01;$ 
    else
       $PM_{ij}^{elite} - = 0.01;$ 
    end if
  end for
end for
for  $i = 0; i < PM^{elite}.row; i++$  do
  for  $j = 0; j < PM^{elite}.column; j++$  do
    if  $corr(e_i, e_j).conf < 0.2$  then
       $PM_{ij}^{elite} = 0;$ 
      if  $corr(e_m, e_n)$  is  $corr(e_i, e_j)$ 's neighbor correspondence then
        if  $corr(e_m, e_n)$  does not logically contradict with  $corr(e_i, e_j)$  then
           $PM_{m,n}^{elite} - = 0.01;$ 
        end if
      end if
    end if
  end for
end for
end for

```

In **Algorithm 3**, we first generate two AM s and use them to obtain a new AM^{new} with the uniform crossover operator. Then, AM^{new} is compared with the elite AM^{elite} , and the winner will become the elite AM . After that, we use AM^{elite} to update its corresponding PM : if $AM_{ij}^{elite} == 1$, then $PM_{ij}^{elite} + = 0.01$; otherwise, $PM_{ij}^{elite} - = 0.01$. We update PM so that the newly generated AM will be closer to the elite AM . Finally, we find the NA $corr(e_i, e_j)$ from AM^{elite} , and their neighbor correspondence $corr(e_m, e_n)$ where the shortest path between e_m (or e_n) and e_i (or e_j) in the ontology hierarchy graph is less than 2, we pick up those do logically contradict with $corr(e_i, e_j)$ and reduce their corresponding PM elements' values by 0.01. In particular, the step length of updating PM determines the algorithm's learning rate. If the step length is too large, the algorithm converges quickly, i.e., the value of PM 's elements are close to 1 or 0; and if it is too small, the algorithm consumes a long running time. Here, we empirically set the step length as 0.01, which is able to ensure the highest average quality of alignments on all testing cases.

4.3. Adaptive Probability Matrix Maintenance

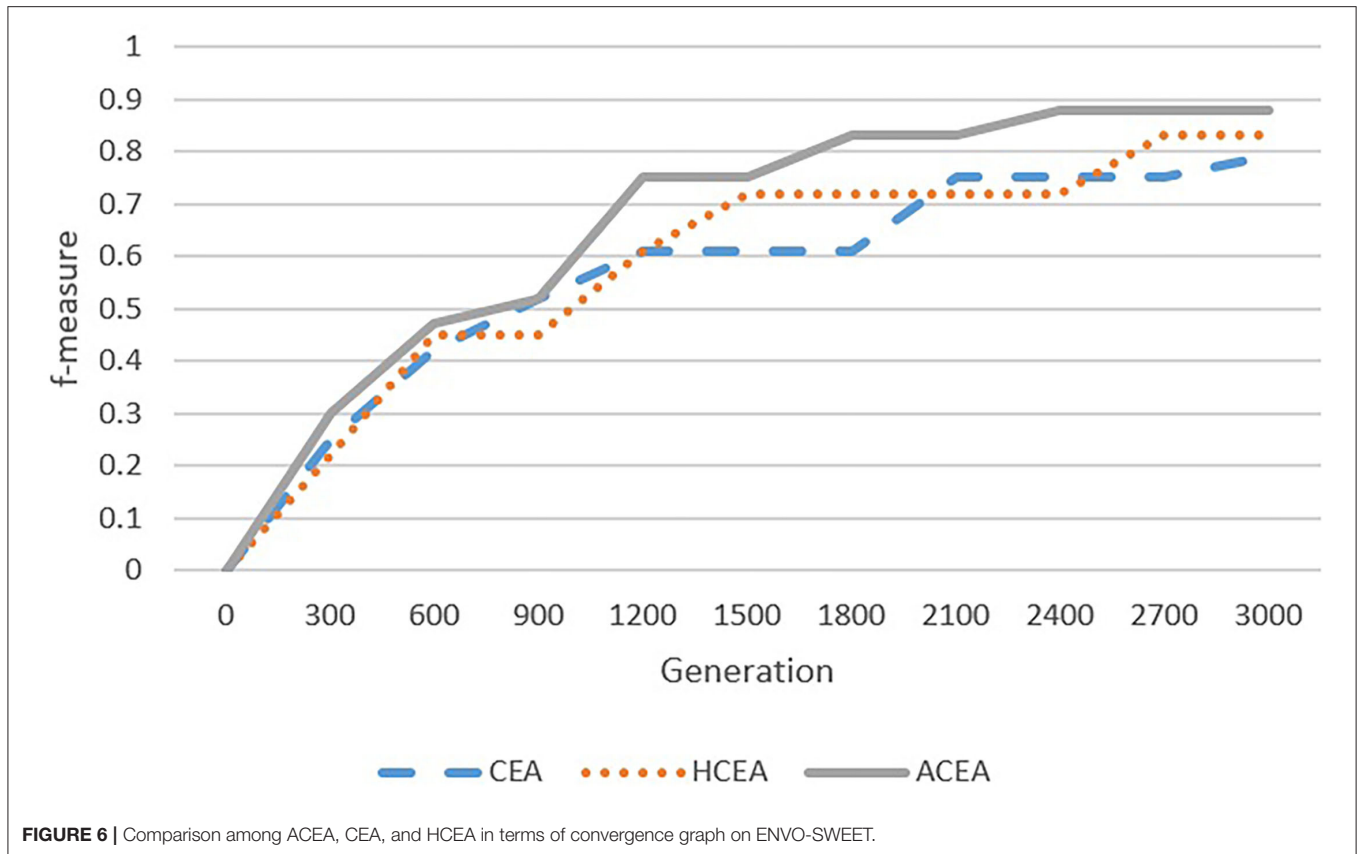
At the end of the generation, adaptive PM maintenance is executed to adjust the algorithm's

Algorithm 4 Adaptive Probability Matrix Maintenance

```

if  $\sum(|PM_{ij}^a - PM_{ij}^b|) < 0.5$  then
  if  $PM_{elite}^a$  is better than  $PM_{elite}^b$  then
    remove  $PM^b$ ;
  else
    remove  $PM^a$ ;
  end if
end if
if All elite  $AM$ s keep unchanged for  $\delta$  generations then
  for  $i = 0; i < PM^{new}.row; i++$  do
    for  $j = 0; j < PM^{new}.column; j++$  do
      if  $PM_{ij}^{max} = PM_{ij}^{min} = 1$  or 0 then
         $PM_{ij}^{new} = 1$  or 0;
      end if
      if  $PM_{ij}^{max} < 0.5$  then
         $PM_{ij}^{new} = PM_{ij}^{max} + rand(0, 1)(1 - PM_{ij}^{max});$ 
      end if
      if  $PM_{ij}^{min} > 0.5$  then
         $PM_{ij}^{new} = (1 - rand(0, 1))PM_{ij}^{min};$ 
      end if
      if  $PM_{ij}^{max} > 0.5$  and  $PM_{ij}^{min} < 0.5$  then
         $PM_{ij}^{new} = 0.5;$ 
      end if
    end for
  end for
   $AM_{elite}^{new} = initializeAM(PM^{new});$ 
end if

```



searching direction. The pseudo-code of adaptive population maintenance is shown in **Algorithm 4**.

In **Algorithm 4**, we first calculate the distance between the existing PMs. The smaller distance indicates a larger overlap between their searching directions, and therefore, the one with worse elite AM should be deleted. When all elite AMs keep unchanged for δ generations, i.e., the algorithm gets stuck in the local optima, we add a new PM PM^{new} with a different searching direction. To determine PM^{new} 's elements, we need to analyze the existing PMs' elements distribution through the maximum probability and minimum probability. To be specific, for each of PM^{new} 's element PM_{ij}^{new} , we find its corresponding maximum probability PM_{ij}^{max} and minimum probability PM_{ij}^{min} from the existing PMs. If they are all equal to 1 or 0, we will set PM_{ij}^{new} as 1 or 0. If $PM_{ij}^{max} < 0.5$, PM_{ij}^{new} will be put in the left of all existing probabilities, i.e., $PM_{ij}^{new} = PM_{ij}^{max} + rand(0,1)(1 - PM_{ij}^{max})$. If $PM_{ij}^{min} > 0.5$, PM_{ij}^{new} will be put in the right of all existing probabilities, i.e., $PM_{ij}^{new} = (1 - rand(0,1))PM_{ij}^{min}$. If $PM_{ij}^{max} > 0.5$ and $PM_{ij}^{min} < 0.5$, PM_{ij}^{new} will be put in the middle, i.e., $PM_{ij}^{new} = 0.5$. Finally, we initialize the elite AM AM_{elite}^{new} for PM^{new} .

5. EXPERIMENT

5.1. Experimental Configuration

We use the track of Biodiversity and Ecology in Ontology Alignment Evaluation Initiative (OAEI)¹ to test ACEA's performance. Biodiversity and Ecology track consists of four pairs of ontologies in the biodiversity and ecology domain: (1) ENVO²-SWEET³, (2) PTO⁴-FLOPO⁵, (3) AGROVOC⁶-NALT⁷, (4) GEMET⁸-ANAEETHES⁹. All of these ontologies are widely used in various projects and researches on biodiversity and ecology, which are developed in parallel and are significantly overlapping.

In the experiment, we compare ACEA with CEA (Xue et al., 2015), HCEA (Xue and Chen, 2019), and OAEI's participants. In particular, CEA's configuration is as follows:

¹<http://oaei.ontologymatching.org>

²<http://agroportal.lirmm.fr/ontologies/ENVO>

³<https://bioportal.bioontology.org/ontologies/SWEET>

⁴<http://agroportal.lirmm.fr/ontologies/TO>

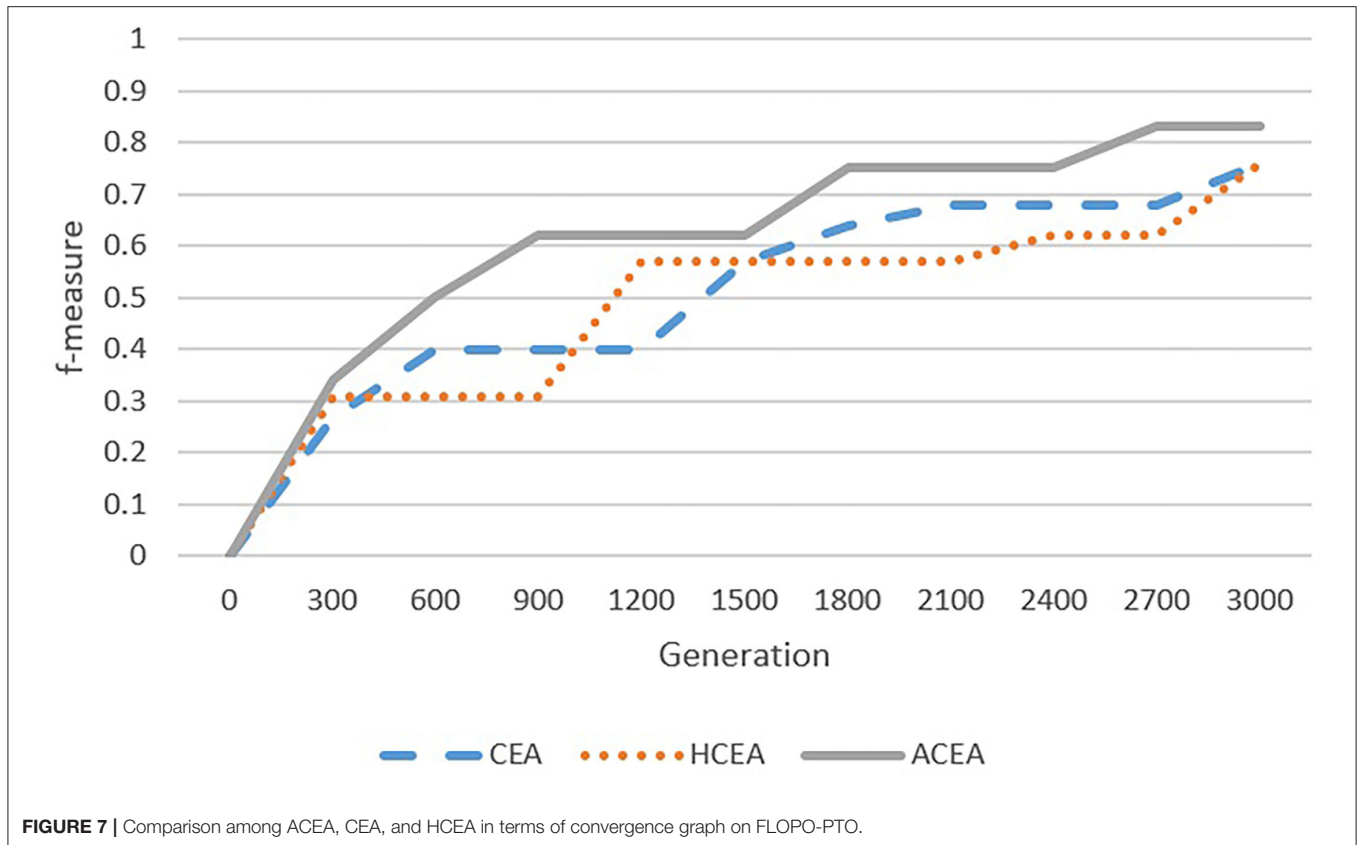
⁵<http://agroportal.lirmm.fr/ontologies/FLOPO>

⁶<http://agroportal.lirmm.fr/ontologies/AGROVOC>

⁷<http://agroportal.lirmm.fr/ontologies/NALT>

⁸<http://agroportal.lirmm.fr/ontologies/GEMET>

⁹<http://agroportal.lirmm.fr/ontologies/ANAEETHES>



- The maximum generation = 3,000;
- The step length for updating PV = 0.01.

The configuration of HCEA's is as following:

- The maximum generation = 3,000;
- The step length for updating PV = 0.01;
- The crossover probability = 0.6;
- The mutation probability = 0.03;
- The mutation shift = 0.05.

Additionally, ACEA's configuration is given as follows:

- The maximum generation = 3,000;
- The threshold for activate PM Maintenance = 60;
- The step length for updating probability matrix = 0.01.

Ontology Alignment Evaluation Initiative's participants' results are from OAEI's official website¹⁰. We first show the sensitivity testing on ACEA's parameter, then ACEA is compared with CEA and HCEA in terms of f-measure and convergence graph, and finally, ACEA is compared with the state-of-the-art ontology aligning techniques. ACEA, CEA, and HCEA's results are the mean value of 30 independent runs.

¹⁰<http://oaei.ontologymatching.org/2021/biodiv/index.html>

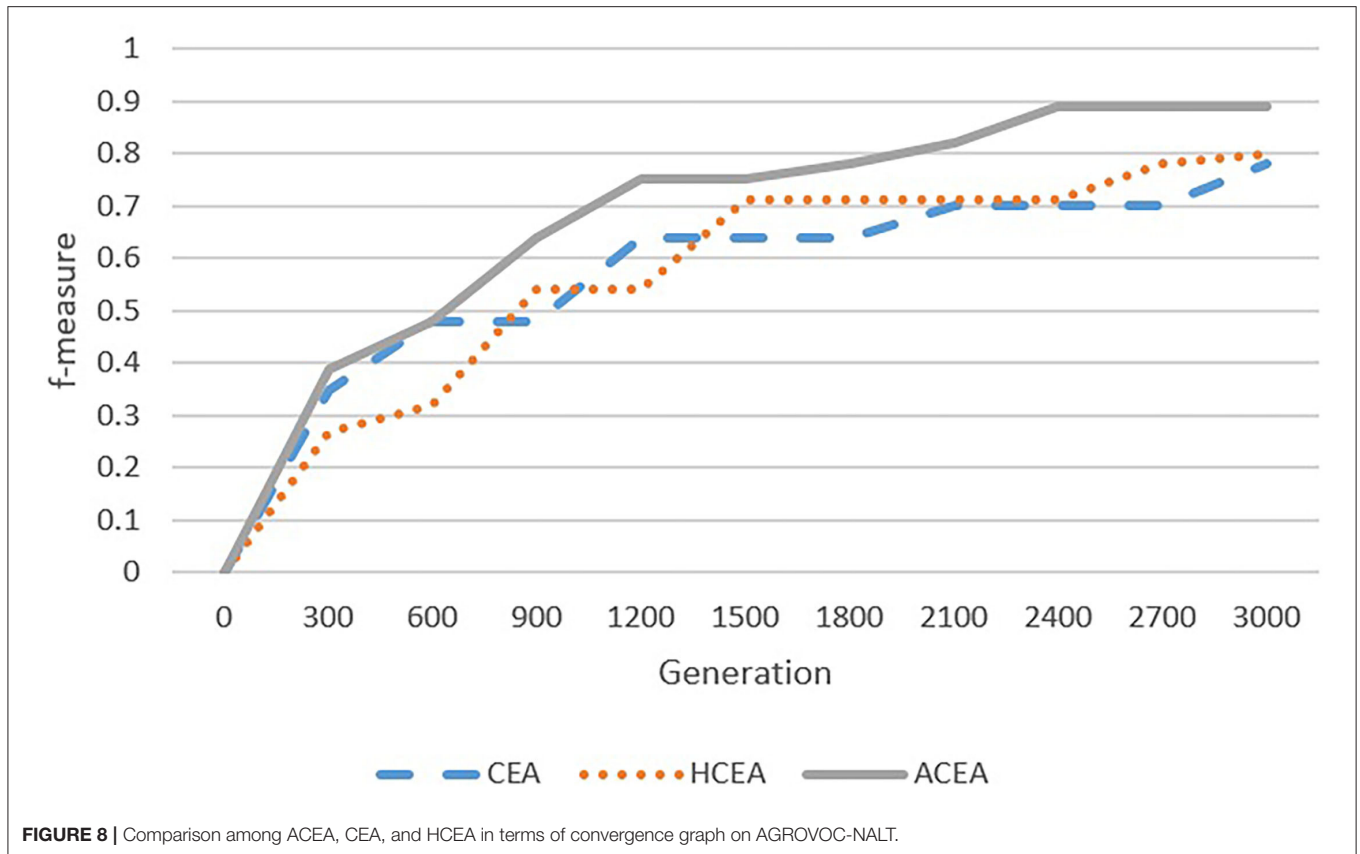
5.2. Experimental Results

First, the sensitivity testings are carried out on ACEA's parameter δ that determines the timing of executing the PM maintenance. If δ is too large, ACEA would get stuck in the local optima for a long time, which would hamper the algorithm from converging on the global optima, and if δ is too small, there would be too many PVs, which increases the computational complexity. We empirically take five representative values, i.e., 20, 40, 60, 80, and 100, to execute the sensitivity testing on δ , whose results are shown in **Figures 4, 5**.

In **Figures 4, 5**, with the increasing values, the quality of alignments begins to deteriorate when $\delta > 60$, and the running time start to decrease, and when $\delta = 60$, it reaches the bottom. Therefore, the parameter $\delta = 60$ is able to better trade-off the quality of alignments and the algorithm's running time.

In **Table 2**, we compare ACEA with CEA and HCEA with mean f-measure *f-measure* and the standard deviation *stdDev*. In **Table 2**, the statistical *t*-test (Schmetterer and Lehmann, 1962) is executed on the data presented in **Table 3**.

In **Table 3**, the T-test's degree of freedom of is 2, and the significance level is 0.05. On all testing cases, the *p*-values are all smaller than 0.05, and thus, we can draw the conclusion that ACEA statistically outperforms CEA and HCEA based aligning techniques at the significance level of 5%. It is obvious that through adaptively maintaining PMs, ACEA



can effectively explore the whole feasible region and find high-quality alignments.

As depicted in **Figures 6–9**, with the introduction of semantic reasoning, the searching space of ACEA can be significantly reduced, which makes it able to more efficiently converge to the global optimal solution.

Finally, we compare ACEA with OAEI's participants on Biodiversity and Ecology track through f-measure. In **Table 4**, we can see that ACEA's mean f-measure outperforms all the state-of-the-art ontology aligning systems on all testing cases. ACEA makes use of the evolutionary paradigm to iteratively refine the alignment's quality, which is a more effective way of improving the alignment's quality than other machine learning based aligning approaches (such as ALOD2Vec, POMap, and DOME), logical reasoning based aligning methods (such as Lily, LogMap Family, and XMap) and Word-based aligning techniques (such as AML, Wiktionary, FCAMapKG, ATBox, and KGMatcher).

5.3. Discussions on Experimental Results

Compact evolutionary algorithm combines the mechanisms of a classic EA with competitive learning, which is effective to lead the algorithm to determine the optimal solution. In addition, the simplicity of CEA, which does not require all the

mechanisms of an EA, rather the few steps in the algorithm are small and simple. HCEA further introduces local refinements on the elite solution, which allows increasing the convergence speed *via* the local search. Compared with CEA and HCEA, ACEA works based on the probabilistic modeling of promising solutions, which makes it easier to predict the movements of the populations in the search space. When confronted with complex optimization issues, ACEA is able to jump out of the local optima through adaptively PM maintenance, which guides the algorithm to explore the potential search space and learn a more complex probabilistic model. Therefore, ACEA outperforms CEA and HCEA in terms of both qualities of alignments and computational efficiency.

In addition, ACEA comprehensively aggregates three broad categories of entity similarity measure, i.e., syntactic-based similarity measure, linguistic-based similarity measure, and structure-based similarity measure, which lead to better alignments than the ones that only take into consideration one or two of them, such as AML, LogMap Family, POMap, XMap, DOME, FCAMapKG. This is because when facing a different heterogeneous situation, none of the similarity measures could be effective in all matching tasks, and taking into consideration more similarity measures could be of help to find the correct correspondences.

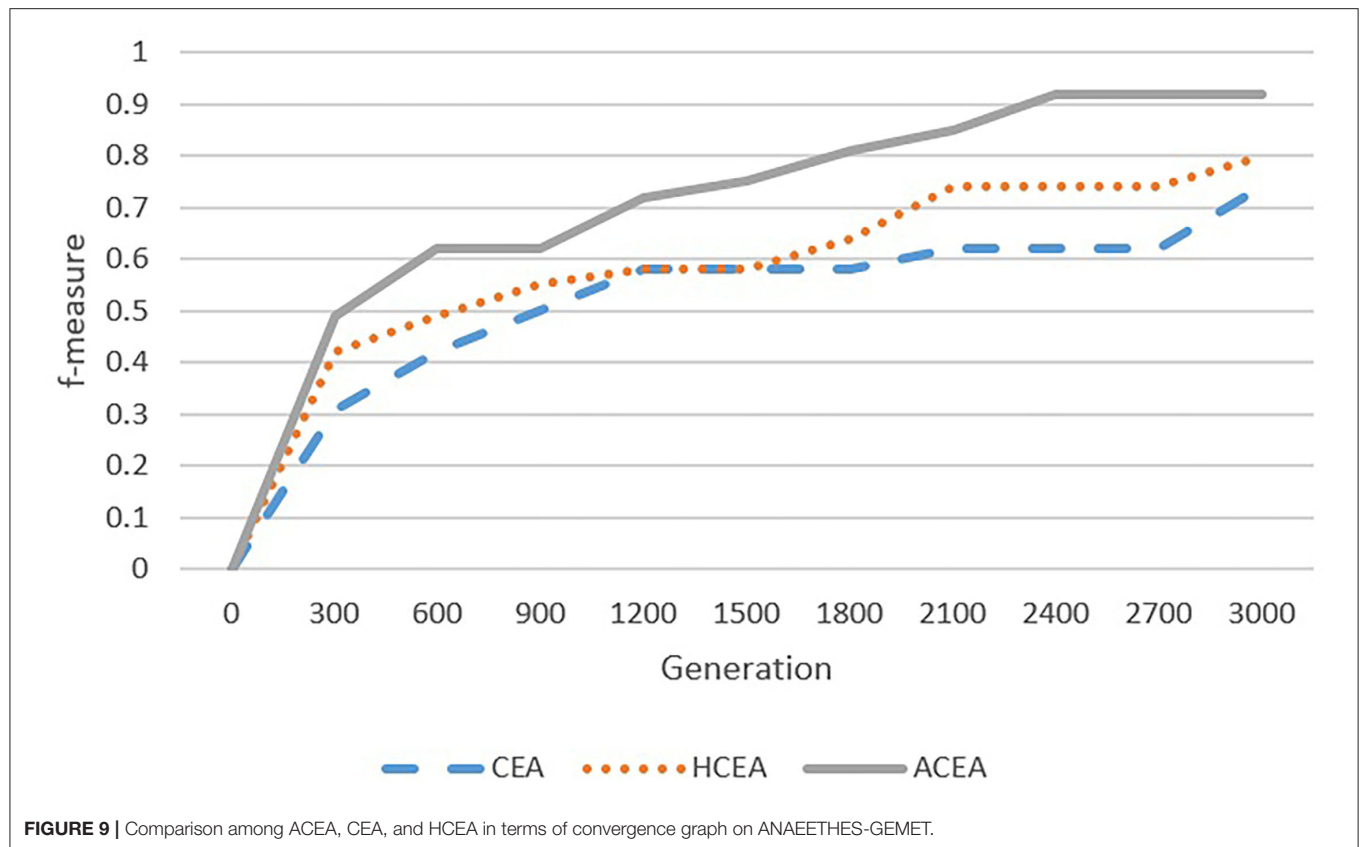


FIGURE 9 | Comparison among ACEA, CEA, and HCEA in terms of convergence graph on ANAEETHES-GEMET.

TABLE 4 | Comparisons among ACEA and state-of-the-art ontology aligning techniques in terms of f-measure.

Aligning system	ENVO-SWEET	FLOPO-PTO	AGROVOC-NALT	ANAEETHES-GEMET
AML (Lima et al., 2020)	0.84	0.86	0.87	0.85
Lily (Wu et al., 2019)	0.73	0.68	-	-
LogMap (Jiménez-Ruiz, 2020)	0.78	0.80	-	0.89
LogMapBio (Jiménez-Ruiz, 2020)	0.77	0.79	-	0.89
LogMapLite (Jiménez-Ruiz, 2020)	0.77	0.75	-	0.49
POMap (Laadhar et al., 2018)	0.78	0.68	-	-
XMap (Djeddi et al., 2015)	0.78	0.76	-	-
DOME (Hertling and Paulheim, 2018)	-	0.73	-	-
FCAMapKG (Algergawy et al., 2019)	0.63	0.69	-	-
POMap (Laadhar et al., 2018)	0.69	0.68	-	-
ATBox (Hertling and Paulheim, 2020)	0.69	0.71	-	-
Wiktionary (Portisch and Paulheim, 2020)	-	0.002	-	-
ALOD2Vec (Portisch et al., 2020)	-	0.002	-	0.10
KGMatcher (Fallatah et al., 2021)	0.005	-	-	0.063
ACEA	0.88	0.83	0.89	0.92

The symbol “-” means that the corresponding matching technique is not able to determine an alignment.

6. CONCLUSION

To manage knowledge on ecology and biodiversity and preserve the ecosystem and biodiversity simultaneously, it is necessary to link the data entities in different ecology

and biodiversity ontologies. To this end, this study proposes an ACEA-based ecology and biodiversity ontology aligning technique. In particular, the problem of EBOA is modeled as a large-scale discrete optimization problem with a sparse solution. Then, a hybrid entity similarity measure is

presented to calculate the ecology and biodiversity entities' similarity. Finally, a problem-specific ACEA is proposed, which introduces semantic reasoning and adaptive PM maintenance to efficiently solve the problem of EBOA. The experimental results show that the evolutionary paradigm is able to find a better alignment than other artificial techniques and the proposed semantic reasoning and adaptive PM maintenance are able to further improve the algorithm's efficiency.

Although ACEA based aligning technique shows its superiority in the experiment, it is not able to detect the m:n correspondence, i.e., multiple source entities are mapped with multiple target entities, which is a common complex correspondence pattern. In addition, ACEA is also not able to find other semantic relationships among the entities, such as the subsumption. Finally, the divide-and-conquer approach has been proved to be a viable method that can facilitate the effectiveness of aligning process (Hu et al., 2008), and we are also interested in utilizing the ontology partitioning technique to pre-process two ontologies.

REFERENCES

- Acampora, G., Loia, V., Salerno, S., and Vitiello, A. (2012). A hybrid evolutionary approach for solving the ontology alignment problem. *Int. J. Intell. Syst.* 27, 189–216. doi: 10.1002/int.20517
- Acampora, G., Loia, V., and Vitiello, A. (2013). Enhancing ontology alignment through a memetic aggregation of similarity measures. *Inf. Sci.* 250, 1–20. doi: 10.1016/j.ins.2013.06.052
- Algergawy, A., Faria, D., Ferrara, A., Fundulaki, I., Harrow, I., Hertling, S., et al. (2019). "Results of the ontology alignment evaluation initiative 2019," in *The Fourteenth International Workshop on Ontology Matching collocated with the 18th International Semantic Web Conference (ISWC 2019)*, Vol. 2536 (Auckland), 46–85.
- Alves, A., Revoredo, K., and Bai ao, F. (2012). "Ontology alignment based on instances using hybrid genetic algorithm," in *Proceedings of the 7th International Conference on Ontology Matching-Vol. 946 (CEUR-WS. Org)* (Boston), 242–243.
- Berners-Lee, T., Hendler, J., Lassila, O. (2001). The semantic web. *Sci. Am.* 284, 28–37. doi: 10.1038/scientificamerican0501-34
- Chu, S.-C., Xue, X., Pan, J.-S., and Wu, X. (2020). Optimizing ontology alignment in vector space. *J. Internet Technol.* 21, 15–22. doi: 10.3966/160792642020012101002
- Djeddi, W. E., Khadir, M. T., and Yahia, S. B. (2015). "Xmap: results for oaei 2015," in *CEUR Workshop Proceedings (Bethlehem)*, 216–221.
- Fallatah, O., Zhang, Z., and Hopfgartner, F. (2021). "Kgmatcher results for oaei 2021," in *Proceedings of the 16th International Workshop on Ontology Matching co-located with the 20th International Semantic Web Conference (ISWC 2021)* (Virtual Conference), 160–166.
- Ginsca, A.-L., and Iftene, A. (2010). "Using a genetic algorithm for optimizing the similarity aggregation step in the process of ontology alignment," in *9th Roedunet International Conference (Sibiu)*, 118–122.
- Hertling, S., and Paulheim, H. (2018). "Dome results for oaei 2018," in *The Thirteenth International Workshop on Ontology Matching collocated with the 17th International Semantic Web Conference (ISWC 2018)* (Monterey), Vol. 2288, 144–151.
- Hertling, S., and Paulheim, H. (2020). "Atbox results for oaei 2020," in *The Fifteenth International Workshop on Ontology Matching Collocated With the 19th International Semantic Web Conference (ISWC 2020)*, Vol. 2788 (Virtual Conference: RWTH), 168–175.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

XX and P-WT proposed the idea. XX performed the literature review and comparative analyses and wrote the manuscript draft. P-WT performed the experiments and made the figures. All authors read the manuscript draft, commented on it, and confirmed it before submission.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 62172095), the Natural Science Foundation of Fujian Province (No. 2020J01875), and the Scientific Research Foundation of the Fujian University of Technology (No. GY-Z17162).

- Hu, W., Qu, Y., and Cheng, G. (2008). Matching large ontologies: a divide-and-conquer approach. *Data Knowl. Eng.* 67, 140–160. doi: 10.1016/j.datak.2008.06.003
- Jiménez-Ruiz, E. (2020). "Logmap family participation in the oaei 2020," in *The Fifteenth International Workshop on Ontology Matching collocated With the 19th International Semantic Web Conference (ISWC 2020)*, Vol. 2788 (CEUR-WS), 201–203.
- Karam, N., Khiat, A., Algergawy, A., Sattler, M., Weiland, C., and Schmidt, M. (2020). Matching biodiversity and ecology ontologies: challenges and evaluation results. *Knowl. Eng. Rev.* 35, 132. doi: 10.1017/S0269888920000132
- Kondrak, G. (2005). "N-gram similarity and distance," in *International Symposium on String Processing and Information Retrieval* (Lille: Springer), 115–126.
- Kumar, V. (2020). "Smart environment for smart cities," in *Smart Environment for Smart Cities* (Singapore: Springer), 1–53.
- Laadhar, A., Ghozzi, F., Bousarsar, I. M., Ravat, F., Teste, O., and Gargouri, F. (2018). "Oaei 2018 results of pomap++," in *The Thirteenth International Workshop on Ontology Matching Collocated With the 17th International Semantic Web Conference (ISWC 2018)* (Monterey), 192–196.
- Lima, B., Faria, D., Couto, F. M., Cruz, I. F., and Pesquita, C. (2020). "Oaei 2020 results for aml and amlc," in *The Fifteenth International Workshop on Ontology Matching collocated With the 19th International Semantic Web Conference (ISWC 2020)* (Virtual Conference), 154–160.
- Madin, J.-S., Bowers, S., Schildhauer, M.-P., and Jones, M. B. (2008). Advancing ecological research with ontologies. *Trends Ecol. Evol.* 23, 159–168. doi: 10.1016/j.tree.2007.11.007
- Martinez-Gil, J., Alba, E., and Aldana-Montes, J. F. (2008). "Optimizing ontology alignments by using genetic algorithms," in *Proceedings of the Workshop on Nature Based Reasoning for the Semantic Web* (Karlsruhe), 1–15.
- Miller, G. A. (1995). Wordnet: A lexical database for english. *Commun. ACM* 38, 39–41. doi: 10.1145/219717.219748
- Mirjalili, S. (2019). "Genetic algorithm," in *Evolutionary Algorithms and Neural Networks* (Cham: Springer), 43–55.
- Portisch, J., Hladik, M., and Paulheim, H. (2020). "Alod2vec matcher results for oaei 2020," in *The Fifteenth International Workshop on Ontology Matching collocated With the 19th International Semantic Web Conference (ISWC 2020)*, Vol. 2788 (Virtual Conference: RWTH), 147–153.
- Portisch, J., and Paulheim, H. (2020). "Wiktionary matcher results for oaei 2020," in *The Fifteenth International Workshop on Ontology Matching collocated with the*

- 19th International Semantic Web Conference (ISWC 2020), Vol. 2788 (Virtual Conference: RWTH), 225–232.
- Rijsberge, C. J. V. (1975). *Information Retrieval*. London: University of Glasgow, Butterworth.
- Schmetterer, L., and Lehmann, E. L. (1962). Testing statistical hypotheses. *Econometrica* 30, 462–465. doi: 10.2307/1909917
- Shvaiko, P., and Euzenat, J. (2011). Ontology matching: state of the art and future challenges. *IEEE Trans. Knowl. Data Eng.* 25, 158–176. doi: 10.1109/TKDE.2011.253
- Wang, J., Ding, Z., and Jiang, C. (2006). “Gaom: genetic algorithm based ontology matching,” in *Proceedings of IEEE Asia8CPacific Conference on Services Computing* (Guangzhou: IEEE), 61–620.
- Wang, P. (2010). “Lily-lom: an efficient system for matching large ontologies with non-partitioned method,” in *Proceedings of CEUR Workshop, Vol. 658* (Shanghai: Citeseer), 69–72.
- Wu, J., Pan, Z., Zhang, C., and Wang, P. (2019). “Lily results for oaei 2019,” in *The Fourteenth International Workshop on Ontology Matching Collocated With the 18th International Semantic Web Conference (ISWC 2019)* (Auckland), 153–159.
- Xue, X., and Chen, J. (2019). Optimizing ontology alignment through hybrid population-based incremental learning algorithm. *Mem. Comput.* 11, 209–217. doi: 10.1007/s12293-018-0255-8
- Xue, X., Chen, J., and Yao, X. (2018). Efficient user involvement in semiautomatic ontology matching. *IEEE Trans. Emerg. Top. Comput. Intell.* 5, 214–224. doi: 10.1109/TETCI.2018.2883109
- Xue, X., and Huang, Q. (2022). Generative adversarial learning for optimizing ontology alignment. *Expert Syst.* doi: 10.1111/exsy.12936. [Epub ahead of print].
- Xue, X., Liu, J., Tsai, P.-W., Zhan, X., and Ren, A. (2015). “Optimizing ontology alignment by using compact genetic algorithm,” in *2015 11th International Conference on Computational Intelligence and Security (CIS)* (Shenzhen: IEEE), 231–234.
- Xue, X., and Liu, W. (2022). Integrating heterogeneous ontologies in asian languages through compact genetic algorithm with annealing re-sample inheritance mechanism. *ACM Trans. Asian Low-Res. Lang. Inform. Process.* 1–21. doi: 10.1145/3519298
- Xue, X., and Wang, Y. (2015a). Optimizing ontology alignments through a memetic algorithm using both matchfmeasure and unanimous improvement ratio. *Artif. Intell.* 223, 65–81. doi: 10.1016/j.artint.2015.03.001
- Xue, X., and Wang, Y. (2015b). Using memetic algorithm for instance coreference resolution. *IEEE Trans. Knowl. Data Eng.* 28, 580–591. doi: 10.1109/TKDE.2015.2475755
- Xue, X., Wu, X., Jiang, C., Mao, G., and Zhu, H. (2021). Integrating sensor ontologies with global and local alignment extractions. *Wireless Commun. Mobile Comput.* 2021, 1–10. doi: 10.1155/2021/6625184
- Xue, X., and Zhang, J. (2021). Matching large-scale biomedical ontologies with central concept based partitioning algorithm and adaptive compact evolutionary algorithm. *Appl. Soft. Comput.* 106, 1–11. doi: 10.1016/j.asoc.2021.107343

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xue and Tsai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.