Interactive Cross-Lingual Ontology Matching

Article *in* IEEE Access · June 2019

DOI: 10.1109/ACCESS.2019.2920881

Junfeng Chen
Hohai University

Xingsi Xue
Fujian University of Technology

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 5, 2019, accepted May 22, 2019. Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2019.2920881

Interactive Cross-Lingual Ontology Matching

JUNFENG CHEN[1], XINGSI XUE[2,3,4,5], YIKUN HUANG[6], AND XUEWU ZHANG[1]
[1]College of IoT Engineering, Hohai University, Changzhou 213022, China
[2]College of Information Science and Engineering, Fujian University of Technology, Fuzhou 350118, China
[3]Intelligent Information Processing Research Center, Fujian University of Technology, Fuzhou 350118, China
[4]Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, Fuzhou 350118, China
[5]Fujian Key Lab for Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou 350118, China
[6]Concord University College, Fujian Normal University, Fuzhou 350117, China

Corresponding author: Xingsi Xue (jack8375@gmail.com)

This work was supported in part by the National Key R&D Program of China under Grant 2018YFC0407101, in part by the National Natural Science Foundation of China under Grant 61503082 and Grant 61403121, in part by the Natural Science Foundation of Fujian Province under Grant 2016J05145, in part the Fundamental Research Funds for the Central Universities under Grant 2019B22314, in part by the Program for New Century Excellent Talents in Fujian Province University under Grant GY-Z18155, in part by the Program for Outstanding Young Scientific Researcher in Fujian Province University under Grant GY-Z160149, and in part by the Scientific Research Foundation of Fujian University of Technology under Grant GY-Z17162 and Grant GY-Z15007.

ABSTRACT Recently, with the growing number of ontologies defined in different languages, to bridge the semantic gaps between them, it is necessary to identify the correspondences between their heterogeneous entities, so-called cross-lingual ontology matching. Due to the complexity and the intricacy of the cross-lingual ontology matching, it is essential to get an expert involved in the matching process to guarantee the alignment's quality. In this paper, we propose an interactive cross-lingual ontology matching technique that makes the user and automatic matcher work together to create high-quality alignments in a reasonable amount of time. In particular, we present a cross-lingual similarity metric to calculate the similarity value of two cross-lingual entities, construct an optimal model for the cross-lingual ontology matching problem, and propose an interactive compact differential evolution (ICDE) algorithm to effectively match the cross-lingual ontologies. The experiment exploits the ontology alignment evaluation initiative (OAEI) multifarm track to test our proposal's performance. The experimental results show that the ICDE significantly outperforms other EA-based matchers and OAEI's participants, and the interacting mechanism can significantly improve the alignment's quality.

INDEX TERMS Cross-lingual ontology matching, user interaction, interactive compact differential evolution.

I. INTRODUCTION

An ontology provides a formal specification on the concepts and their relationships in a domain of interest, which is utilized by Semantic Web to overcome the heterogeneity issue and achieve the semantic inter-operability among ontology-based intelligent applications. Recently, with the growing number of ontologies defined in different languages, to bridge the semantic gaps between them, it is necessary to identify the correspondences between their heterogeneous entities, so-called cross-lingual ontology matching [1]. In particular, cross-lingual ontology matching can be defined as the process of discovering semantic mappings between two independent ontologies where each one is lexicalized in a different natural language, which has become a major issue in ontology matching field. Indeed, considering cross-lingual information is becoming more and more important, in view particularly of the growing number of content-creating non-English users and the clear demand of cross-language inter-operability leading to the need of bringing multilingual semantic information and knowledge together in an explicit manner [2].

Due to the complexity and intricacy of the cross-lingual ontology matching, with each task having its particularities, dictated by both the domain and the design of the ontologies, there are limits to the performance of automatic ontology matching technique, as adopting more advanced alignment techniques has brought diminishing returns [3]. Therefore, it is essential to get an expert involved into the cross-lingual ontology matching process, and make user validation an

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

essential step to guide the matching algorithm and guarantee the alignment's quality. In this work, we propose an interactive cross-lingual ontology matching technique, which makes users and automatic cross-lingual matcher cooperate with each other to create high quality alignments in a reasonable amount of time. In particular, we present a Compact Differential Evolution (CDE) algorithm to implement the automatic matching process, and introduce a mechanism of user intervention into CDE's evolving process to improve the result's quality. In particular, our contributions made in this paper are as follows:

- A cross-lingual similarity metric is proposed to calculate the similarity value of two cross-lingual entities,
- An optimal model for the cross-lingual ontology matching problem is constructed,
- A problem-specific Interactive CDE (ICDE) is proposed to effectively match the cross-lingual ontologies.

The rest of the paper is organized as follows: Section II describes the related work; Section III presents the cross-lingual ontology matching problem and cross-lingual similarity metric; Section IV presents in details the ICDE-based cross-lingual ontology matching technique; Section V shows the experimental results; and finally, Section VI draws the conclusion.

II. RELATED WORK

A. CROSS-LINGUAL ONTOLOGY MATCHING

Currently, most of the cross-lingual ontology matching techniques use the general-purpose machine translators and corpus to reduce the problem to monolingual English-only matching. SimCat [4] splits the labels into words and normalizes them, and then translates the normalized entities into English through the Yandex.[1] CroLOM [5] first applies natural language process technique on each language, and then it also uses the Yandex to translate all entities into English and computes the similarity values through Wordnet.[2] Ngai *et al.* [6] implement a translator by a machine learning method. They first construct a bilingual corpus from the American English Wall Street Journal and Mandarin Chinese People's Daily newspaper, and then compute the similarity values by the Term FrequencyĺCInverse Document Frequency (TF-IDF) and cosine based similarity metrics. Helou *et al.* [7] propose to makes use of Google Translate[3] as the background knowledge base to implement the interpretation. Trojahn *et al.* [8] also proposes a translation-based cross-lingual ontology matching technique, which uses a multi-agent architecture. The mediator agent sends an ontology to the translation agent to translate it into the target language using a dictionary. DSSim [9] uses DBpedia to associate concepts in English and Dutch, and then propose a DSSim tool to solve the monolingual ontology matching problem. Bouma *et al.* [10] uses EuroWordnet[4] to translate English into Dutch, and map Dutch acronym for common Thesaurus for Audiovisual Archives (GTAA) to Wordnet and DBpedia.[5] The quality of the translations used has a major impact on its performance, and the possible misinterpretations could significantly reduce the quality of the alignment. To ensure the alignment's quality, we choose the Babelnet translate, which performs better than other Web translation services in the context of concept mapping [2], to transform the cross-lingual problem into monolingual problem. Moreover, we also make use of the user knowledge to validate those ambiguous mappings, which can further improve the precision of the results.

B. EVOLUTIONARY ALGORITHM BASED AUTOMATIC ONTOLOGY MATCHING TECHNIQUE

Among all automatic ontology matching techniques, EA could present a good methodology for solving ontology matching problem. Genetics for Ontology ALignments (GOAL) [11] firstly utilizes EA to optimizing the aggregating parameters of several similarity measures. Naya *et al.* [12] also use EA to optimize the alignment's quality by aggregating different similarity metrics. Ginsca and Iftene [13] use EA to optimize all the parameters in ontology matching process, which includes the aggregating weights of different similarity measures and a threshold for filtering the final alignment. Later, Acampora *et al.* [14] propose a Memetic Algorithm (MA) to optimize the ontology alignments, which introduces the local search process into classic EA's evolving process to improve its converging speed. Xue and Wang [15] adopt the similar idea but they propose a novel evaluation metric on the alignment's quality, which can overcome three drawbacks of EA-based matchers. All these matching techniques require to calculate several similarity matrices through the similarity measures, which consume huge memory consumption and make the matching process inefficient. Therefore, recent studies mainly focus on the entity matching technique, which use EA to directly determine the optimal ontology entity mapping set instead of tuning the aggregating weights of the similarity measures. GAOM (Genetic Algorithm based Ontology Matching) [16] regards ontology regards two ontologies as two feature sets, and defines the fitness function as a global similarity measure function between two ontologies based on feature sets, and then employ EA to match them. MapPSO [17], instead, addresses the ontology entity matching problem as an discrete optimization problem, whose fitness function depends on the similarity values computed by performing a combination of lexical, linguistic and structural matchers. Since DE is a famous and effective EA for tackling complex optimization problems and being inspired by its success in various application domains [18], [19], in this work, we propose an ICDE to effectively address the cross-lingual ontology matching

[1] https://translate.yandex.com/?lang=es-en&text=administrar&ncrnd=5317
[2] https://wordnet.princeton.edu/
[3] https://translate.google.com/
[4] https://ssli.ee.washington.edu/people/duh/multilingual/eurowordnet.html
[5] https://wiki.dbpedia.org/

problem. Through the introduction of user knowledge into the evolving process, our approach can be more efficient than other compact EA such as the population based incremental learning algorithm.

C. INTERACTIVE ONTOLOGY MATCHING TECHNIQUE

To ensure the ontology alignment's quality, the alignments determined by the automatic tools should be validated by the experts to distinguish which ones are correct and which ones are not [3]. To this end, various interactive ontology matching techniques are developed, which present different strategies on user interaction exploitation. AgreementMakerLight (AML) [20] employs an interactive selection algorithm, which utilized the alignments returned by various ontology matchers to detect suspicious mappings. Above the threshold 70%, AML queries the user for suspicious mappings, otherwise, it rejects all the suspicious mappings. AML ensures that the reasonable workload for the user by setting the query limit as 45% of the determined correspondences for small scale ontology matching tasks, and 15% for the others. ALIN [21] generates an initial set of candidate correspondences, and requires the user to validate them. If the user judges a candidate mapping as correct, it will be moved to the final alignment. Then, ALIN removes all candidate mappings that are not consistent with the approved correspondences. The interactions continue until there are no more candidate correspondences left. LogMap [22] presents problematic mappings to the user for validation, and the validating results are utilized to detect the conflicts with already found mappings. LogMap allows to pause the user interaction and continue the validation work in the future. XMap [23] cooperate with user in the post-matching steps to filter the final alignment. It uses two thresholds to implement this procedure, where the mappings with similarity value higher than the upper threshold are directly added to the final alignment, and those mappings with similarity values lower than the lower threshold are presented to the user for validation. SOCOM++ [24] requires a user to configure properties in the process of selecting label translations, which can adjust the mapping results. Dragoni [25] proposes a suggestion-based cross-lingual ontology matching technique, which introduces an information retrieval-based (IR-based) approach for generating candidate mappings for the user validation. The existing interactive ontology matching techniques exploit user involvement in either pre-matching or post-matching phrase, but our approach gets user involved into DE's evolving process, and make use of a user's validating results in an iterative way, which is a more effective way of utilizing user's knowledge to improve the alignment's quality [3].

III. PRELIMINARIES

A. CROSS-LINGUAL ONTOLOGY MATCHING PROBLEM

Cross-lingual ontology matching dedicates to determine a cross-lingual entity mapping set, so-called cross-lingual ontology alignment. During the matching process, some external resources, such as the electronic dictionaries and knowledge bases, are required in matching process. Since the quality of a cross-lingual ontology alignment can be measured by the number of entity mappings and the mean similarity value of them, in this work, we utilize the following equation to measure an alignment's quality:

$$f(A) = \frac{2 \times |A| \times \sum simValue_i}{(|C_1| \times |C_2|) \times (|A| + \sum simValue_i)} \in [0, 1] \quad (1)$$

where $|C_1|$, $|C_2|$ and $|A|$ are respectively the cardinalities of two entity sets C_1 and C_2, and an alignment A between them, $simValue_i$ is the similarity value of the ith entity mapping.

Further, we construct an optimal model for the cross-lingual ontology matching problem:

$$\begin{cases} max \quad F(X) \\ s.t. \quad X = (x_1, x_2, \cdots, x_{|C_1|})^T \\ \qquad x_i \in \{0, 1, 2, \cdots, |C_2|\}, \quad i = 1, 2, \cdots, |C_1| \end{cases} \quad (2)$$

where $|C_1|$ and $|C_2|$ respectively represent the cardinalities of two entity sets C_1 and C_2, x_i, $i = 1, 2, \cdots, |C_1|$ represents the ith entity mapping, and in particular, $x_i = 0$ means the ith entity in one ontology is mapped to none of the concept in the other ontology. Supposing A is X's corresponding alignment, the objective function $F(X)$ is equal to $f(A)$.

B. CROSS-LINGUAL SIMILARITY METRIC

Cross-lingual similarity metric is the foundation of the cross-lingual ontology matching technique [1]. In this work, we utilize a profile-based similarity metric to measure to what extent two cross-lingual concepts are similar to each other. Given the concept hierarchies of two cross-lingual ontologies, for each concept, we construct a profile for it by collecting its label and the labels of all its direct ascendants and descendants. Then, the similarity value of two entities e_1 and e_2 is calculated through their profiles p^1 and p^2:

$$\frac{\sum_{i=1}^{|p^1|} \max_{j=1\cdots|p^2|} (sim'(p_i^1, p_j^2)) + \sum_{j=1}^{|p^2|} \max_{i=1\cdots|p^1|} (sim'(p_j^2, p_i^1))}{2 \times min(|p^1|, |p^2|)} \quad (3)$$

where:

- $|p^1|$ and $|p^2|$ are the cardinalities of p^1 and p^2, respectively,
- p_i^1 and p_j^2 are respectively the ith element of p^1 and jth element of p^2,
- $sim'()$ computes the similarity value between p_i^1 and p_j^2

Before calculating $sim'(p_i^1, p_j^2)$, we utilize the natural language processing technique and Babelnet Translate[6] which covers 271 different languages and becomes an appropriate machine translation tool in cross-lingual ontology matching domain [2], to process p_i^1 and p_j^2. In particular, this process consists in the following successive steps:

[6]https://babelnet.org/

<table>
<tr><td>Gene</td><td>1</td><td>0</td><td>……</td><td>1</td><td>0</td><td>1</td><td>……</td><td>1</td><td>……</td><td>1</td><td>1</td><td>……</td><td>0</td></tr>
<tr><td>Index</td><td colspan="4">1</td><td colspan="4">2</td><td>……</td><td colspan="4">n</td></tr>
</table>

FIGURE 1. An example of encoding and decoding mechanism, where *n* is the number of source concepts.

- remove the numbers, punctuations and stop-words;
- split the strings into words;
- translate the words into English, and convert them into lower-case;
- lemmatizing and stemming the English words;

Then, $sim'(p_i^1, p_j^2)$ is calculated with soft TF-IDF [26]. In particular, two words are identical when they are the same literally or they are synonymous in the English Wordnet [27]. With the concept alignment, we can further match the object properties through computing their domain and range's similarity values, and determine the identical data object properties by calculating their labels' similarity value.

IV. INTERACTIVE COMPACT DIFFERENTIAL EVOLUTION ALGORITHM

In this work, we propose a problem-specific Interactive Compact Differential Evolution (ICDE) algorithm to address the cross-lingual ontology matching problem. ICDE utilizes a probabilistic representation of the population to save the memory consumption, and introduces the adaptive schemes on control parameters to improve the converging speed. In the following, three kernel components of ICDE are presented, i.e. the encoding mechanism, the mutation operator and the user interaction. Last, the pseudocode of ICDE is presented.

A. ENCODING MECHANISM

In our proposal, a Probability Vector (PV) [28] is utilized to characterize the entire population, and each element inside stands for the probability that holds true for a correspondence. We utilize the gray encoding mechanism to encode each cross-lingual entity mapping. When decoding, the number obtained represents the index of a target entity, and in particular, value 0 means a source entity is not mapped to any target entity. An example of the encoding and decoding mechanism is shown in the Fig. 1, where the index means the source entity index and the corresponding gene values are the target entity index that is encoded through gray encoding mechanism.

B. MUTATION OPERATOR

In ICDE, three solutions, namely ind_r, ind_s, and ind_t, are sampled from the PV, and an offspring ind'_{off} is generated as follows;

$$ind'_{off} = ind_t + F(ind_r - ind_s) \tag{4}$$

where F is a scale factor that determines how far the generated offspring is from ind_t. Generally, ICDE generates ind'_{off} by altering ind_t according to the distance between ind_r and ind_s. For the cross-lingual ontology matching problem, which is a discrete optimization problem, we introduce the edit distance to measure two individual's distance. In the following, we present the equation about the calculation of two individuals ind_r and ind_s's edit distance:

$$editDistance(ind_r, ind_s) = \sum_{i=1}^{|ind_r|} |ind_{r,i} - ind_{s,i}| \tag{5}$$

where $|ind_r|$ is the cardinality of ind_r, $ind_{r,i}$ and $ind_{s,i}$ are respectively the ith element of ind_r and ind_s. Next, the offspring ind'_{off} are generated by partly flipping the elements of ind_t, and the number of altered elements is determined by a random number in [0,1] and $editDistance(ind_r, ind_s)$. For the sake of clarity, the pseudo-code of mutation operator is given in Algorithm 1:

Algorithm 1 Mutation Operator

```
ind'_off = ind_t.copy();
for i = 0; i < |ind_r|; i++ do
    if ind_r[i]! = ind_s[i] then
        append i to an array index;
    end if
end for
totalNum = round(rand(0, 1) × editDistance(ind_r, ind_s));
j = 0;
k = 0;
while j < totalNum do
    if rand(0, 1) < F then
        ind'_off[index[k]] = (ind'_off[index[k]] + 1) mod 2;
        remove the kth element from index;
        j++;
    end if
    k = (k + 1) mod index.length();
end while
```

C. USER INTERACTION

Since the number of user intervention needed highly depends on the performance of its automatic ontology matching process, reducing user involvement requires an efficient automatic matching technology as the basis. In this work, we utilize ICDE to implement the automatic ontology matching process, and when the evolving process of ICDE gets stuck, i.e. the elite solution keeps unchanged for $\epsilon = 20$ generations, the user gets involved to redirect the search direction of it. Moreover, the number of candidate correspondences can be reduced by only selecting the correspondences with similarity values between 0.4 and 0.6, i.e. the problematic correspondences, as the candidate correspondences.

User is not able to validate a whole alignment, rather, automatic matching tools must limit their demand for user intervention. Therefore, through simply using the limited user intervention, it is difficult to effectively guide the matching algorithm's search direction and improve the matching result's quality. Propagating a user's validating results is an effective approach to maximize his work's value [29]. Based on the observation that a correct alignment should not be inconsistent with ontology's concept hierarchies [30], in this work, we propose a context path based propagating method. Particularly, given a user-validated concept correspondence, the shortest path between the source concept (or target concept) and the root is called the context path of that source concept (or target concept). It is obvious that source context path's concepts should be similar to those in the target context path if two paths are alike. On this basis, ICDE implements the propagation by checking all the possible mappings between each source concept's ancestor and target concept's ancestor. When a better concept correspondence is found, the elite's codes and the corresponding PV's elements will be updated.

D. THE PSEUDOCODE OF INTERACTIVE COMPACT DIFFERENTIAL EVOLUTION ALGORITHM

Given the length of a solution (or PV) *length*, the maximum number of generations $maxGen = 3000$, the binomial crossover probability $p_{cr} = 0.6$, the step length for updating PV $st = 0.1$, the pseudo-code of ICDE is given in Algorithm 2.

Algorithm 2 Interactive Compact Differential Evolution Algorithm

```
** Initialization **
generation = 0;
tag = 0;
for i = 0; i < length; i++ do
   PV[i] = 0.5;
end for
generate an individual ind_elite through PV;
while generation < maxGen do
   ** Mutation **
   generate three solutions ind_r, ind_s and ind_t through PV;

   generate an offspring ind'_off through Equation 4;
   ** Crossover **
   ind_off = ind'_off;
   for i = 0; i < length; i++ do
      if rand(0, 1) < p_cr then
         ind_off[i] = ind_elite[i]
      end if
   end for
   ** Elite Update **
   [winner, loser] = compete(ind_elite, ind_new);
   if winner == ind_new then
      ind_elite = ind_new;
      tag = 0;
   else
      tag = tag + 1;
   end if
   ** PV Update **
   for i = 0; i < length; i++ do
      if winner[i] == 1 then
         PV[i] = PV[i] + st;
      else
         PV[i] = PV[i] - st;
      end if
   end for
   ** User Involvement **
   if tag > 20 then
      get user involved and validate the alignment;
      propagate the user validation;
      tag = 0;
   end if
   generation = generation + 1;
end while
output ind_elite;
```

V. EXPERIMENT

A. EXPERIMENTAL SETUP

In order to study the effectiveness of our proposal, we exploit Ontology Alignment Evaluation Initiative (OAEI)'s Multifarm track[7] where the user validation is simulated by using an oracle. In Multifarm track, matchers need to deal with 45 different language pairs, i.e. ar-cn, ar-cz, ar-de, ar-en, ar-es, ar-fr, ar-nl, ar-pt, ar-ru, cn-cz, cn-de, cn-en, cn-es, cn-fr, cn-nl, cn-pt, cn-ru, cz-de, cz-en, cz-es, cz-fr, cz-nl, cz-pt, cz-ru, de-en, de-es, de-fr, de-nl, de-pt, de-ru, en-es, en-fr, en-nl, en-pt, en-ru, es-fr, es-nl, es-pt, es-ru, fr-nl, fr-pt, fr-ru, nl-pt, nl-ru, pt-ru. For instance, en-cn refers to the test cases involving the English and Chinese languages while cn-fr refers to the test cases involving the Chinese and French languages. For more details, please see also OAEI's official web site on Multifarm track.

We carry out the statistical comparison on the alignment's quality among ICDE-based matcher, EA-based matcher [31], MA-based matcher [32], DE-based matcher, CDE-based matcher and OAEI's participants that take part in the Multifarm track, i.e. AgreementMakerLight (AML) [33], KEPLER [34], LogMap [35] and XMap [36]. In order to compare with OAEI's participants, we use f-measure [37] as the performance metric to evaluate the quality of obtained alignments. The results of OAEI's participants are from OAEI's official website,[8] and the configurations of EA and MA based approaches are respectively referred to their corresponding literatures. EA, MA, DE, CDE and ICDE's results shown in the tables are the mean values of thirty independent executions, they use the parameters (see also Section IV-D) which represent a trade-off setting obtained in an empirical way to achieve the highest average alignment quality on all exploited testing cases, the configurations of EA and MA based matcher refer to their own literatures.

[7] http://oaei.ontologymatching.org/2018/multifarm/index.html
[8] http://oaei.ontologymatching.org/2018/index.html

B. STATISTICAL COMPARISON ON ALIGNMENT'S QUALITY

In this work, the statistical comparisons are carried out among various approaches. The comparison is carried out by means of a multiple comparison procedure: firstly, the Friedman's test [38] is used to find out whether the results of various

TABLE 1. 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.

Testing Case	AML	KEPLER	LogMap	XMap	EA	MA	DE	CDE	ICDE
ar-cn	0.26 (3)	0.14 (8)	0.19 (6)	0.00 (9)	0.16 (7)	0.23 (5)	0.28 (2)	0.25 (4)	0.32 (1)
ar-cz	0.40 (2.5)	0.23 (8)	0.40 (2.5)	0.00 (9)	0.28 (7)	0.32 (6)	0.37 (4)	0.33 (5)	0.42 (1)
ar-de	0.37 (2.5)	0.25 (8)	0.37 (2.5)	0.00 (9)	0.27 (7)	0.30 (6)	0.35 (4.5)	0.35 (4.5)	0.45 (1)
ar-en	0.39 (3)	0.26 (8)	0.41 (2)	0.00 (9)	0.30 (7)	0.34 (6)	0.36 (4)	0.35 (5)	0.42 (1)
ar-es	0.44 (2)	0.22 (8)	0.36 (5)	0.00 (9)	0.25 (7)	0.34 (6)	0.40 (3)	0.38 (4)	0.47 (1)
ar-fr	0.37 (2.5)	0.20 (8)	0.29 (5)	0.00 (9)	0.21 (7)	0.27 (6)	0.36 (4)	0.37 (2.5)	0.42 (1)
ar-nl	0.39 (3)	0.23 (8)	0.41 (2)	0.00 (9)	0.25 (7)	0.29 (6)	0.35 (4.5)	0.35 (4.5)	0.44 (1)
ar-pt	0.48 (1.5)	0.28 (7)	0.38 (3.5)	0.00 (9)	0.27 (8)	0.31 (6)	0.38 (3.5)	0.36 (5)	0.48 (1.5)
ar-ru	0.29 (5)	0.23 (7)	0.41 (2)	0.00 (9)	0.21 (8)	0.28 (6)	0.33 (4)	0.35 (3)	0.44 (1)
cn-cz	0.32 (3)	0.17 (8)	0.27 (5)	0.00 (9)	0.20 (7)	0.24 (6)	0.31 (4)	0.33 (2)	0.35 (1)
cn-de	0.35 (3)	0.23 (7.5)	0.23 (7.5)	0.00 (9)	0.24 (6)	0.34 (4.5)	0.36 (2)	0.34 (4.5)	0.40 (1)
cn-en	0.32 (2)	0.27 (5)	0.22 (8)	0.00 (9)	0.23 (7)	0.24 (6)	0.30 (3)	0.28 (4)	0.36 (1)
cn-es	0.41 (2)	0.19 (8)	0.25 (6)	0.00 (9)	0.24 (7)	0.30 (5)	0.37 (3)	0.35 (4)	0.46 (1)
cn-fr	0.40 (2)	0.20 (8)	0.23 (6)	0.00 (9)	0.21 (7)	0.28 (5)	0.35 (3)	0.32 (4)	0.48 (1)
cn-nl	0.34 (4)	0.19 (8)	0.21 (7)	0.00 (9)	0.25 (6)	0.28 (5)	0.41 (2)	0.38 (3)	0.46 (1)
cn-pt	0.41 (2)	0.21 (8)	0.25 (6)	0.00 (9)	0.24 (7)	0.26 (5)	0.36 (3)	0.30 (4)	0.46 (1)
cn-ru	0.39 (2)	0.21 (8)	0.31 (3)	0.00 (9)	0.22 (7)	0.28 (4)	0.27 (5)	0.25 (6)	0.43 (1)
cz-de	0.47 (2)	0.34 (6)	0.39 (5)	0.08 (9)	0.28 (8)	0.30 (7)	0.42 (3.5)	0.42 (3.5)	0.52 (1)
cz-en	0.48 (3)	0.31 (7)	0.50 (2)	0.14 (9)	0.27 (8)	0.32 (6)	0.44 (5)	0.46 (4)	0.56 (1)
cz-es	0.57 (2)	0.31 (8)	0.39 (5)	0.05 (9)	0.33 (7)	0.35 (6)	0.42 (4)	0.48 (3)	0.63 (1)
cz-fr	0.54 (2)	0.27 (7)	0.39 (5)	0.01 (9)	0.25 (8)	0.31 (6)	0.45 (3)	0.41 (4)	0.58 (1)
cz-nl	0.56 (2)	0.30 (8)	0.45 (3.5)	0.08 (9)	0.31 (7)	0.35 (6)	0.45 (3.5)	0.43 (5)	0.60 (1)
cz-pt	0.55 (2)	0.41 (7)	0.44 (5)	0.10 (9)	0.39 (8)	0.43 (6)	0.52 (3)	0.48 (4)	0.60 (1)
cz-ru	0.52 (2)	0.35 (7)	0.46 (3.5)	0.00 (9)	0.28 (8)	0.36 (6)	0.44 (5)	0.46 (3.5)	0.58 (1)
de-en	0.47 (2)	0.39 (6)	0.44 (3)	0.10 (9)	0.34 (8)	0.36 (7)	0.42 (4.5)	0.42 (4.5)	0.50 (1)
de-es	0.48 (2)	0.33 (8)	0.39 (5.5)	0.02 (9)	0.36 (7)	0.39 (5.5)	0.46 (3)	0.44 (4)	0.50 (1)
de-fr	0.50 (2)	0.29 (8)	0.43 (5)	0.03 (9)	0.38 (7)	0.41 (6)	0.46 (3.5)	0.46 (3.5)	0.52 (1)
de-nl	0.48 (2)	0.34 (6)	0.45 (3)	0.08 (9)	0.31 (8)	0.33 (7)	0.43 (4)	0.37 (5)	0.52 (1)
de-pt	0.49 (2)	0.41 (6)	0.38 (8)	0.07 (9)	0.39 (7)	0.43 (5)	0.45 (3.5)	0.45 (3.5)	0.52 (1)
de-ru	0.41 (4)	0.28 (8)	0.44 (2)	0.00 (9)	0.31 (7)	0.35 (6)	0.42 (3)	0.40 (5)	0.48 (1)
en-es	0.45 (2.5)	0.30 (8)	0.45 (2.5)	0.08 (9)	0.35 (7)	0.41 (5)	0.44 (4)	0.40 (6)	0.46 (1)
en-fr	0.45 (2)	0.27 (8)	0.43 (3)	0.11 (9)	0.30 (7)	0.34 (6)	0.42 (4.5)	0.42 (4.5)	0.49 (1)
en-nl	0.48 (3)	0.32 (6)	0.54 (1.5)	0.12 (9)	0.26 (8)	0.30 (7)	0.42 (4)	0.40 (5)	0.54 (1.5)
en-pt	0.49 (3)	0.39 (6)	0.52 (2)	0.12 (9)	0.32 (8)	0.36 (7)	0.45 (5)	0.48 (4)	0.58 (1)
en-ru	0.38 (5)	0.30 (6.5)	0.48 (2)	0.00 (9)	0.28 (8)	0.30 (6.5)	0.45 (3)	0.40 (4)	0.51 (1)
es-fr	0.55 (2)	0.29 (8)	0.40 (5)	0.02 (9)	0.34 (7)	0.38 (6)	0.52 (3)	0.48 (4)	0.57 (1)
es-nl	0.58 (1.5)	0.33 (8)	0.40 (5)	0.00 (9)	0.34 (7)	0.35 (6)	0.48 (3)	0.45 (4)	0.58 (1.5)
es-pt	0.58 (1.5)	0.39 (6)	0.45 (5)	0.08 (9)	0.32 (8)	0.35 (7)	0.52 (3)	0.50 (4)	0.58 (1.5)
es-ru	0.51 (2)	0.39 (7)	0.41 (5)	0.00 (9)	0.36 (8)	0.40 (6)	0.50 (3)	0.48 (4)	0.56 (1)
fr-nl	0.55 (2)	0.27 (8)	0.42 (5)	0.10 (9)	0.30 (7)	0.36 (6)	0.46 (3)	0.44 (4)	0.57 (1)
fr-pt	0.55 (2)	0.33 (8)	0.39 (6)	0.02 (9)	0.35 (7)	0.42 (5)	0.49 (3)	0.46 (4)	0.56 (1)
fr-ru	0.49 (2)	0.30 (8)	0.36 (6)	0.00 (9)	0.32 (7)	0.38 (5)	0.45 (3.5)	0.45 (3.5)	0.52 (1)
nl-pt	0.59 (2)	0.36 (8)	0.45 (5)	0.07 (9)	0.37 (7)	0.42 (6)	0.52 (3.5)	0.52 (3.5)	0.61 (1)
nl-ru	0.51 (2)	0.29 (8)	0.46 (4)	0.00 (9)	0.32 (7)	0.35 (6)	0.48 (3)	0.44 (5)	0.55 (1)
pt-ru	0.49 (2)	0.37 (7)	0.47 (3.5)	0.00 (9)	0.34 (8)	0.38 (6)	0.47 (3.5)	0.42 (5)	0.53 (1)
Average	0.45 (2.41)	0.28 (7.40)	0.38 (4.33)	0.03 (9.00)	0.29 (7.28)	0.33 (5.85)	0.41 (3.53)	0.40 (4.13)	0.50 (1.04)

approaches present any difference; secondly, when an difference is found in the first step, the Holm's test [39] is utilized to determine whether one approach statistically outperforms others.

Friedman's test aims at detecting whether a significant difference exists among the algorithms. In particular, under the null-hypothesis, it states that all algorithms are equivalent, hence, a rejection of this hypothesis implies the existence of differences among the performance of all studied algorithms [40]. In order to reject the null hypothesis, the computed value $\mathcal{X}_r^2$ must be equal to or greater than the tabled critical chisquare value at the specified level of significance [41]. In this work, a level of significance $\alpha = 0.05$ is chosen. Since we are comparing 9 approaches, our analysis has to consider the critical value $\mathcal{X}_{0.05}^2$ for 8 degrees of freedom that is equal to 15.507.

In Table 1, by performing the Friedmans test, the computed $\mathcal{X}_r^2$ value is 317.27, which is greater than its associated critical value $\mathcal{X}_{0.05}^2 = 15.507$, the null hypothesis is rejected and it is possible to assess that there is a significant difference between these proposals. According to this result, a post-hoc statistical analysis is needed to conduct pairwise comparisons in order to detect concrete differences among compared algorithms. Holm's procedure works by setting a control algorithm and comparing it with the remaining ones. Normally, the algorithm which obtains the lowest value of ranking in the Friedman's test is chosen as control algorithm. In our experimentation, as shown in Table 1, ICDE is characterized by the lowest value of ranking.

TABLE 2. Holm's test on the alignment's quality. ϵ denotes a number that is very close to 0.

i	approach	z value	unadjusted $p-$value	$\frac{\alpha}{k-i}, \alpha = 0.05$
8	AML	2.37	0.017	0.050
7	DE	4.31	1.63×10^{-5}	0.025
6	CDE	5.35	8.79×10^{-8}	0.016
5	LogMap	5.69	1.27×10^{-8}	0.012
4	MA	8.33	ϵ	0.010
3	EA	10.80	ϵ	0.008
2	KEPLER	11.01	ϵ	0.007
1	XMap	13.78	ϵ	0.006

Holm's test works on a family of hypotheses where each one is related to a comparison between the control algorithm and one of the remaining algorithms. In details, the test statistic for comparing the ith and jth algorithms named z value is used for finding the corresponding probability from the table of the normal distribution (the so-called p-value), which is then compared with an appropriate level of significance α. In our experimentation $\alpha = 0.05$ and the results of Holm's test are shown in Table 2. According to the data in Table 2, it is possible to state that ICDE statistically outperforms other approaches on f-measure at 5% significance level. In particular, CDE's results are quite closed to that of DE in terms of f-measure, which shows that the compact evolutionary mechanism is effective. With the introduction of the user intervention, ICDE's results are significantly better than those of DE and CDE.

VI. CONCLUSION

In this paper, a problem-specific ICDE is proposed to implement the interactive cross-lingual ontology matching, which makes use of CDE to implement the automatic matching process and gets user involved into the automatic matching process to guide the algorithm's search direction and improve the alignment's results. In addition, we construct an optimal model for the cross-lingual ontology matching problem, and present a cross-lingual ontology similarity metric to distinguish the heterogeneous cross-lingual entities. The experimental results show that ICDE significantly outperform other EA-based matchers and OAEI's participants, and the introduction of interacting mechanism can significantly improve the alignment's quality.

REFERENCES

[1] T. Ivanova, "Cross-lingual and multilingual ontology mapping-survey," in *Proc. 19th Int. Conf. Comput. Syst. Technol.* New York, NY, USA: ACM, 2018, pp. 50–57.

[2] A. N. Tigrine, Z. Bellahsene, and K. Todorov, "Light-weight crosslingual ontology matching with LYAM++," in *Proc. OTM Confederated Int. Conf. Move Meaningful Internet Syst.* Rhodos, Greece: Springer, 2015, pp. 527–544.

[3] Z. Dragisic, V. Ivanova, P. Lambrix, D. Faria, E. Jiménez-Ruiz, and C. Pesquita, "User validation in ontology alignment," in *Proc. Int. Semantic Web Conf.* Kobe, Japan: Springer, 2016, pp. 200–217.

[4] A. Khiat, E. A. Ouhiba, M. A. Belfedhal, and C. E. Zoua, "SimCat results for OAEI 2016," in *Proc. OM@ ISWC*, 2016, pp. 217–221.

[5] E. Jiménez-Ruiz, B. C. Grau, and V. Cross, "LogMap family participation in the OAEI 2017," in *Proc. CEUR Workshop*, 2017, pp. 153–157.

[6] G. Ngai, M. Carpuat, and P. Fung, "Identifying concepts across languages: A first step towards a corpus-based approach to automatic ontology alignment," in *Proc. COLING 19th Int. Conf. Comput. Linguistics*, 2002, pp. 1–7.

[7] M. A. Helou, M. Palmonari, and M. Jarrar, "Effectiveness of automatic translations for cross-lingual ontology mapping," *J. Artif. Intell. Res.*, vol. 55, pp. 165–208, Jan. 2016.

[8] C. Trojahn, P. Quaresma, and R. Vieira, "A framework for multilingual ontology mapping," in *Proc. 6th Int. Lang. Resour. Eval.* New York, NY, USA: ACM, 2008, pp. 1–4.

[9] M. Nagy, M. Vargas-Vera, and P. Stolarski, "DSSim results for OAEI 2008," in *Proc. 3rd Int. Conf. Ontol. Matching*, 2008, pp. 147–159.

[10] G. Bouma, S. Duarte, and Z. Islam, "Cross-lingual alignment and completion of Wikipedia templates," in *Proc. 3rd Int. Workshop Cross Lingual Inf. Access, Addressing Inf. Need Multilingual Societies*. Stroudsburg, PA, USA: Assoc. Comput. Linguistics, 2009, pp. 21–29.

[11] J. Martinez-Gil and J. F. Aldana-Montes, "Evaluation of two heuristic approaches to solve the ontology meta-matching problem," *Knowl. Inf. Syst.*, vol. 26, no. 2, pp. 225–247, 2011.

[12] J. M. V. Naya, M. M. Romero, and J. P. Loureiro, *Improving Ontology Alignment Through Genetic Algorithms*. Hershey, NY, USA: Information Science Reference, 2010, pp. 240–259.

[13] G. Alexandru-Lucian, and A. Iftene, "Using a genetic algorithm for optimizing the similarity aggregation step in the process of ontology alignment," in *Proc. 9th Roedunet Int. Conf.*, Sibiu, Romania, 2010, pp. 118–122.

[14] G. Acampora, V. Loia, S. Salerno, and A. Vitiello, "A hybrid evolutionary approach for solving the ontology alignment problem," *Int. J. Intell. Syst.*, vol. 27, no. 3, pp. 189–216, 2012.

[15] X. Xue and Y. Wang, "Optimizing ontology alignments through a memetic algorithm using both matchfmeasure and unanimous improvement ratio," *Artif. Intell.*, vol. 223, pp. 65–81, Jun. 2015.

[16] J. Wang, Z. Ding, and C. Jiang, "GAOM: Genetic algorithm based ontology matching," in *Proc. IEEE Asia–Pacific Conf. Services Comput.*, GuangZhou, China, Dec. 2006, pp. 617–620.

[17] J. Bock and J. Hettenhausen, "Discrete particle swarm optimisation for ontology alignment," *Inf. Sci.*, vol. 192, pp. 152–173, Jun. 2012.

[18] Y. Wang, M. Zhou, X. Song, M. Gu, and J. Sun, "Constructing cost-aware functional test-suites using nested differential evolution algorithm," *IEEE Trans. Evol. Comput.*, vol. 22, no. 3, pp. 334–346, Jun. 2018.

[19] S. Sudha, S. Baskar, and S. Krishnaswamy, "Protein docking using constrained self-adaptive differential evolution algorithm," in *Soft Computing*, no. 1. Berlin, Germany: Springer, 2019, pp. 1–19.

[20] D. Faria, C. Pesquita, B. S. Balasubramani, C. Martins, J. Cardoso, H. Curado, F. M. Couto, and I. F. Cruz, "OAEI 2016 results of AML," *Ontol. Matching*, vol. 1766, pp. 138–145, Jan. 2016.

[21] J. da Silva, "Alin results for OAEI 2016," in *Proc. 11th Int. Workshop Ontol. Matching*. Kobe, Japan: Springer, 2016, pp. 130–137.

[22] E. Jiménez-Ruiz, B. C. Grau, and V. Cross, "LogMap family participation in the OAEI 2016," in *Proc. 11th Int. Workshop Ontol. Matching*. Kobe, Japan: Springer, 2016, pp. 185–189.

[23] D. E. Warith, M. T. Khadir, and S. Ben Yahia, "XMap: Results for OAEI 2016," in *Proc. 11th Int. Workshop Ontol. Matching*. Kobe, Japan: Springer, 2016, pp. 222–226.

[24] B. Fu, R. Brennan, and D. O'Sullivan, "A configurable translation-based cross-lingual ontology mapping system to adjust mapping outcomes," *J. Web Semantics*, vol. 15, pp. 15–36, Sep. 2012.

[25] M. Dragoni, "Multilingual ontology mapping in practice: A support system for domain experts," in *Proc. Int. Semantic Web Conf.* Bethlehem, PA, USA: Springer, 2015, pp. 169–185.

[26] D. Ngo, Z. Bellahsene, and K. Todorov, "Extended tversky similarity for resolving terminological heterogeneities across ontologies," in *Proc. OTM Confederated Int. Conf. Move Meaningful Internet Syst.* Graz, Austria: Springer, 2013, pp. 711–718.

[27] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[28] X. Xue and J.-S. Pan, "A Compact Co-Evolutionary Algorithm for sensor ontology meta-matching," *Knowl. Inf. Syst.*, vol. 56, no. 2, pp. 335–353, 2018.

[29] X. Xue and X. Yao, "Interactive ontology matching based on partial reference alignment," *Appl. Soft Comput.*, vol. 72, pp. 355–370, Nov. 2018.

[30] P. Wang, "Lily-LOM: An efficient system for matching large ontologies with non-partitioned method," in *Proc. CEUR Workshop*, vol. 658, 2010, pp. 69–72.

[31] J. Martinez-Gil, E. Alba, and J. F. A. Montes, "Optimizing ontology alignments by using genetic algorithms," in *Proc. 1st Int. Conf. Nature Inspired Reasoning Semantic Web*, vol. 419, 2008, pp. 1–15.

[32] G. Acampora, V. Loia, and A. Vitiello, "Enhancing ontology alignment through a memetic aggregation of similarity measures," *Inf. Sci.*, vol. 250, pp. 1–20, Nov. 2013.

[33] D. Faria, C. Pesquita, B. S. Balasubramani, T. Tervo, D. Carriço, R. Garrilha, F. M. Couto, and I. F. Cruz, "Results of AML participation in OAEI 2018," in *Proc. Ontol. Matching, OM-ISWC Workshop*, 2018, pp. 125–131.

[34] M. Kachroudi, G. Diallo, and S. B. Yahia, "KEPLER at OAEI 2018," in *Proc. Ontol. Matching, OM-ISWC Workshop*, 2018, pp. 173–178.

[35] J.-R. Ernesto, G. Bernardo, Cuenca, and V. Cross, "LogMap family participation in the OAEI 2018," in *Proc. 12th Int. Conf. Ubiquitous Inf. Manage. Commun.* New York, NY, USA: ACM, 2018, pp. 187–191.

[36] W. E. Djeddi, S. B. Yahia, and M. T. Khadir, "XMap: Results for OAEI 2018," in *Proc. 13th ISWC Workshop Ontol. Matching (OM)*, 2018, pp. 210–227.

[37] C. J. Van Rijsbergen, "A non-classical logic for information retrieval," *Comput. J.*, vol. 29, no. 6, pp. 481–485, 1986.

[38] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Statist. Assoc.*, vol. 32, no. 200, pp. 675–701, Dec. 1937.

[39] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandin. J. Statist.*, vol. 6, no. 2, pp. 65–70, 1979.

[40] S. García, D. Molina, M. Lozano, and F. Herrera, "A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: A case study on the CEC'2005 Special Session on Real Parameter Optimization," *J. Heuristics*, vol. 15, no. 6, pp. 617–644, 2009.

[41] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, FL, USA: CRC Press, 2003.

JUNFENG CHEN received the Ph.D. degree from the College of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2011. She is currently an Associate Professor with the College of IoT Engineering, Hohai University, Changzhou, China. Her research interests include swarm intelligence, artificial intelligence with uncertainty, cooperative coevolution, data mining techniques, and ontology matching technique. She is also a member of the ACM. She received ICSI 2016 Excellent Paper Award and the Sponsorship of the Jiangsu Overseas Research and Training Program for the University Prominent Young and Middle-aged Teachers and Presidents, China.

XINGSI XUE received the B.S. degree in software engineering from Fuzhou University, China, in 2004, the M.S. degree in computer application technology from the Renmin University of China, China, in 2009, and the Ph.D. degree in computer application technology from Xidian University, China, in 2014. He is currently a Professor with the College of Information Science and Engineering, Fujian University of Technology. He is also the Kernel Member with the Intelligent Information Processing Research Center, Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian Key Lab for Automotive Electronics and Electric Drive, Fujian University of Technology. His research interests include intelligent computation, data mining, and large-scale ontology matching technology. He is a member of the ACM. He received the IIH-MSP 2016 Excellent Paper Award and the 2017 ACM Xi'an Rising Star Award.

YIKUN HUANG received the M.S. degree from the College of Mathematics and Computer Science, Fuzhou University. He is currently an Associate Professor with the Concord University College, Fujian Normal University. His current research interests include information visualization, intelligent computation, and ontology matching technique.

XUEWU ZHANG received the M.S. and Ph.D. degrees from the College of Computer and Information, Hohai University. He is currently a Professor with the College of IoT Engineering, Hohai University. His research interests include computer vision, FMRI image analysis, and underwater detection and imaging.