

Retraction

Retracted: Semisupervised Learning-Based Sensor Ontology Matching

Security and Communication Networks

Received 26 December 2023; Accepted 26 December 2023; Published 29 December 2023

Copyright © 2023 Security and Communication Networks. 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.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] H. Zhu, J. Zhang, and X. Xue, "Semisupervised Learning-Based Sensor Ontology Matching," *Security and Communication Networks*, vol. 2021, Article ID 2002307, 5 pages, 2021.

Research Article

Semisupervised Learning-Based Sensor Ontology Matching

Hai Zhu ¹, Jie Zhang,² and Xingsi Xue ³

¹School of Network Engineering, Zhoukou Normal University, Zhoukou, Henan 466001, China

²School of Computer Science and Engineering, Yulin Normal University, Yulin, Guanxi 537000, China

³School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou, Fujian 350118, China

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

Received 12 June 2021; Revised 30 June 2021; Accepted 10 July 2021; Published 19 July 2021

Academic Editor: Chi-Hua Chen

Copyright © 2021 Hai Zhu 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.

Sensor ontology models the sensor information and knowledge in a machine-understandable way, which aims at addressing the data heterogeneity problem on the Internet of Things (IoT). However, the existing sensor ontologies are maintained independently for different requirements, which might define the same concept with different terms or context, yielding the heterogeneity issue. Since the complex semantic relationship between the sensor concepts and the large-scale entities is to be dealt with, finding the identical entity correspondences is an error-prone task. To effectively determine the sensor entity correspondences, this work proposes a semisupervised learning-based sensor ontology matching technique. First, we borrow the idea of “centrality” from the social network to construct the training examples; then, we present an evolutionary algorithm- (EA-) based metamatching technique to train the model of aggregating different similarity measures; finally, we use the trained model to match the rest entities. The experiment uses the benchmark as well as three real sensor ontologies to test our proposal’s performance. The experimental results show that our approach is able to determine high-quality sensor entity correspondences in all matching tasks.

1. Introduction

Sensor ontology models the sensor information and knowledge on Internet of Things (IoT) in a machine-understandable way [1]. With the help of sensor ontology, different intelligent sensor applications are able to communicate with each other, which is of help to implement their collaboration. Nowadays, more and more sensor ontologies have been developed, which are maintained independently for different requirements. One of the barriers that hampers them from communications is their heterogeneity problem; i.e., one concept could be defined in different ways [2]. Since the complex semantic relationship between the sensor concepts and the large-scale entities is to be dealt with, addressing the sensor ontology heterogeneity problem is an error-prone task. Finding all the identical sensor concept correspondences is called sensor ontology matching, which is regarded as an effective method of addressing the sensor ontology heterogeneity issue [3].

With the quick development of ontology matching domain, more and more matching techniques were

proposed. Most of them need to determine an effective similarity measure to distinguish the heterogeneous entities. However, due to the complex semantic relationships between the entities, there is no single similarity measure which is able to distinguish all the heterogeneous entities, and usually, multiple similarity measures need to work together. A popular aggregating strategy is to first sum the similarity values with linear weighted fashion and filter the results with a proper threshold [4]. However, it is difficult to determine a proper weight set for various matching tasks with different heterogeneous features with completely unsupervised way [5]. Hence, the machine learning-based matching techniques start to attract researchers’ attentions [6–11]. They use a set of correct correspondences to train the regression [12] or classification [13] models, which are then used to determine the final alignment. According to Ontology Alignment Evaluation Initiative (OAEI), the introduction of more learning techniques brings back little improvement on the alignment, and to find more correct correspondences, it is necessary to work with expert’s knowledge. In this work, we propose a semisupervised

learning-based sensor ontology matching technique. In particular, we first require the expert to match a certain number of correct correspondences, which works as the reference alignment in the training phase; then, the Evolutionary Algorithm (EA) is used to train a model of aggregating similarity measures; finally, the obtained model is used to match the rest sensor entities in the training phase.

The rest paper is arranged as follows. Section 2 gives the relevant definitions. Section 3 describes in detail the training examples' construction and EA for addressing the metamatching problem; Section 4 shows the experimental results and makes the corresponding analysis. Finally, Section 5 concludes this work and presents the future work.

2. Sensor Ontology Metamatching Problem

2.1. Concept Similarity Measure. A sensor ontology is a 3-tuple (C, P_d, R) , where C , P_d , and R are, respectively, the concepts' set, data property set, and concept relationship set in the sensor domain [14, 15]. A concept similarity measure (CSM) is a function that maps two sensor concepts to a real number in $[0,1]$, where 1 means two concepts are the same and 0 means they are totally different. Different CSMs measure the similarity value with different ontology information, and in general, they can be divided into three categories, i.e., name-based CSM, dictionary-based CSM, and datatype property-based CSM.

A name-based CSM calculates the edit distance between two concepts' names. In this work, we use the Levenshtein distance [16]. Given two concepts' names s_1 and s_2 , the Levenshtein distance is defined as follows:

$$LN(c_1, c_2) = \frac{\max(0, \min(|s_1|, |s_2|) - d(s_1, s_2))}{\min(|s_1|, |s_2|)}, \quad (1)$$

where s_1 and s_2 are, respectively, the names of two concepts c_1 and c_2 , $|s_1|$ and $|s_2|$ are, respectively, their number of characters, and $d(s_1, s_2)$ is the number of operations that convert s_1 into s_2 .

Dictionary-based CSM makes use of electronic dictionaries, such as WordNet [17], to calculate two concepts' similarity value on their name, which is defined as follows:

$$WN(c_1, c_2) = \frac{2 \times |m(s_1) \cap m(s_2)|}{|m(s_1)| + |m(s_2)|}, \quad (2)$$

where s_1 and s_2 are, respectively, the names of two concepts c_1 and c_2 , and $m(s_1)$ and $m(s_2)$ are, respectively, their meaning sets.

Datatype property-based CSM [18] makes use of two concepts' datatype properties to calculate their similarity value, which is defined as follows:

$$DT(c_1, c_2) = \frac{\sum_{i=1}^{|c_1|} \sum_{j=1}^{|c_2|} \text{sim}(c_1^i, c_2^j) + \sum_{i=1}^{|c_2|} \sum_{j=1}^{|c_1|} \text{sim}(c_1^i, c_2^j)}{|c_1| + |c_2|}, \quad (3)$$

where c_1^i and c_2^j are, respectively, the i th and j th datatype property names of concepts c_1 and c_2 ,

$$\text{sim}(c_1^i, c_2^j) = \max\{\max(0, \min(|c_1^i|, |c_2^j|)) - d(c_1^i, c_2^j) / \min(|c_1^i|, |c_2^j|), 2 \times |m(c_1^i) \cap m(c_2^j)| / (|m(c_1^i)| + |m(c_2^j)|)\}.$$

3. The Optimization Model of Metamatching Problem

Taken two sensor ontologies and a concept similarity measure CMS_i as input, we can determine a similarity matrix M_{csm_i} , whose row and column are, respectively, the concepts from two sensor ontologies and the elements inside are the similarity value determined by CMS_i and two corresponding concepts. On this basis, we can convert the SSM aggregating problem into their corresponding similarity matrix aggregating issue, which can be defined as follows:

$$\text{aggregate}(CMS_1, CMS_2, \dots) = \sum_I w_i \times M_{csm_i}, \quad (4)$$

where $w_i \in [0, 1]$ and $\sum_I w_i = 1$.

The sensor ontology metamatching problem is defined as follows:

$$\begin{cases} \max & f(W, t) \\ & W = \{w_1, w_2, \dots\} \\ \text{s.t.} & \sum_i w_i = 1 \\ & w_i, t \in [0, 1], \end{cases} \quad (5)$$

where $W = \{w_1, w_2, \dots\}$ is the aggregating weight set and $w_i \in W$ is the i th similarity measure's weight, t is the threshold for filtering the final alignment, and the objective function $f(W, t)$ is to evaluate the alignment's quality determined by W and t . Assuming A is the alignment determined by W and t , R is the reference alignment provided by the expert and $f(W, t)$ is equal to A 's f -measure [19], which is defined as follows:

$$\begin{aligned} \text{recall} &= \frac{|R \cap A|}{|R|}, \\ \text{precision} &= \frac{|R \cap A|}{|A|}, \end{aligned} \quad (6)$$

$$f\text{-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{recall} + \text{precision}},$$

where $|A|$, $|R|$, and $|R \cap A|$ are, respectively, the number of correspondence in A and R and their intersection. In particular, recall measures the ratio of found correct correspondences in the reference alignment and precision calculates the ratio of correct correspondences in all the found correspondences.

4. Semisupervised Learning-Based Sensor Ontology Matching

Given a partial reference ontology alignment that are determined by the expert, our approach first uses EA to address the ontology metamatching problem, which trains the model of aggregating similarity measures, and then, the obtained

model are used to match the rest entities. In the next, we first introduce the training example construction and then present the EA for training the model of aggregating similarity measures.

4.1. Training Example Construction. To ensure the training result's quality, it is necessary to construct a training example set through determining the most representative entities in the sensor ontologies. Here, we borrow the definition of centrality from the social network [20]; i.e., the representative entities should be the central ones in the ontology hierarchy graph, which denotes the entities as nodes and their relationships as edges. To be specific, we measure a concept c 's centrality as follows:

$$\text{centrality}(c) = |\text{sub}(c)| + |\text{super}(c)|, \quad (7)$$

where $\text{sub}(c)$ and $\text{super}(C)$ are, respectively, c 's direct descendant classes and all its ascendant classes, $|\text{sub}(c)|$ and $|\text{super}(c)|$ are, respectively, the cardinality of $\text{sub}(c)$ and $\text{super}(c)$, and c 's centrality is the number of all its direct descendant classes and ascendant classes. We sort all the concepts in the descending order and select first 30% concepts as the representative ones. After that, we require the expert to manually match two representative concept sets from two ontologies, and we can obtain the partial reference alignment. With this reference alignment, we can train the model of aggregating similarity measures by finding the optimal aggregating weight set and the threshold with EA.

4.2. Evolutionary Algorithm. The real number encoding mechanism is applied in this work to improve the algorithm's efficiency. To be specific, there are $n + 1$ gene bits in a chromosome, where n is the number of similarity measures and the last gene bit represents the threshold's information. When decoding, the i th aggregating weight is $\text{gene}_i / \sum_{j=1}^n \text{gene}_j$. With aggregating weights and the threshold, each solution corresponds to a particular alignment, and its fitness value is equal to the alignment's f -measure.

We use the selection operator based on roulette wheel strategy and the single-point crossover. During the mutation, we first generate a random number ranNum in $[0,1]$ if it is large than 0.5, the new gene value $\text{gene}' = \text{gene} + \text{ranNum} \times (1 - \text{gene})$; otherwise, the updated gene value $\text{gene}' = \text{gene} - \text{ranNum} \times \text{gene}$. To improve the converging speed, we also introduce the elite strategy, which replaces the solution with worse fitness value with the elite solution (the best solution found so far) at the end of each generation.

5. Experiment and Results

5.1. Experimental Configuration. In the experiment, we use the benchmark (<http://oaei.ontologymatching.org/2016/results/benchmarks/index.html>) and three sensor ontologies, i.e., CSIRO sensor ontology (CSIRO) (<https://www.w3.org/2005/Incubator/ssn/wiki/SensorOntology2009>), Sensor Network ontology (SSN) ([\[ssn\]\(https://www.w3.org/TR/vocab-ssn\)\), and MMI Device ontology \(MMI\) \(<https://marinetmetadata.org>\). The selected sensor ontologies are all widely used in SSW, which shares lots of overlapping information. Table 1 shows the detailed descriptions on all the ontologies in the testing cases.](https://www.w3.org/TR/vocab-</p>
</div>
<div data-bbox=)

Table 2 shows the configuration of EA, which is a trade-off configuration that ensures our approach's performance in all testing cases.

We compare our method with SOBOM [21], CODI [22], ASMOV [23], and FuzzyAlign [24], which are four state-of-the-art techniques in the sensor ontology matching domain. The results of our method are the mean value of thirty independent runs.

5.2. Experimental Results and Analysis. We carried out the sensitive experiment to show the effectiveness of EA's configuration, and Table 3 compares in terms of f -measure among our approach and all competitors on all testing cases. The results shown in Tables 3 and 4 are the mean value on all testing cases.

The population size and maximum generation depend on the complexity of the problem, and their recommended scopes are, respectively, $[20,100]$ and $[500,5000]$. The larger value it takes, the longer runtime it needs. Our problem is a 3-dimension problem, which is not a very complicated optimization problem. As shown in Table 3, the population size and maximum generation are set as 40 and 2000. Crossover probability and mutation probability, respectively, affect EA's exploitation and exploration. If the crossover probability is too large, EA would easily suffer from the premature converge; if it is too small, the algorithm would be difficult to converge. On the contrary, if the mutation probability is too large, EA would become a stochastic algorithm; if it is too small, the algorithm tends to fall in the local optima. From the experimental results in Table 3, we can see that when crossover probability and mutation probability are, respectively, 0.6 and 0.02, the results are the best.

As shown in Table 4, our approach outperforms other competitors on all testing cases, which shows that it is able to effectively determine high-quality sensor ontology alignments in different matching tasks.

6. Conclusion and Future Work

To implement the collaborations among intelligent applications on IoT, it is necessary to distinguish the heterogeneous sensor entities. To find all the sensor entity correspondences, this work proposes a semisupervised learning-based sensor ontology matching technique. In the training process, the training example set is constructed by extracting the most important concepts from two ontologies, which are matched by the expert. Then, an EA-based metamatching technique is proposed to train the model of aggregating different similarity measures. Finally, the weight set and threshold in the model are used to determine the rest correspondences in the testing phase. The experimental results show the effectiveness of our approach.

TABLE 1: Descriptions on testing cases.

Testing case	Description
OAEI's benchmark	1XX Two identical ontologies
	2XX Two ontologies with different heterogeneous characteristics
	3XX Two ontologies from real application
Real sensor ontology	CSIRO CSIRO focuses on the description on the sensor grounding and operation specification
	SSN SSN focuses on modelling the sensor network's information and knowledge
	MMI MMI focus on modelling the sensor knowledge in the marine domain

TABLE 2: Evolutionary algorithm's configuration.

Parameter	Value
Population size	40
Maximum generation	2000
Crossover rate	0.6
Mutation rate	0.02

TABLE 3: Compare with sensor ontology matching techniques in terms of f -measure.

Testing case	SOBOM	CODI	ASMOV	FuzzyAlign	Our method
<i>OAEI's benchmark</i>					
101	1.00	1.00	1.00	1.00	1.00
103	1.00	1.00	1.00	1.00	1.00
201	1.00	1.00	0.94	1.00	1.00
203	0.86	0.81	0.89	1.00	1.00
204	0.88	0.76	0.91	1.00	1.00
221	0.94	0.84	1.00	0.84	1.00
223	1.00	0.97	1.00	1.00	1.00
228	0.91	0.85	0.96	1.00	1.00
230	0.87	1.00	1.00	0.95	1.00
231	1.00	0.88	1.00	1.00	1.00
301	0.72	0.68	0.80	0.80	0.89
302	0.81	0.72	0.83	0.71	0.85
<i>Real sensor ontology</i>					
MMI vs SSN	0.77	0.80	0.73	0.88	0.94
CSIRO vs SSN	0.78	0.79	0.75	0.88	0.95
MMI vs CSIRO	0.72	0.78	0.75	0.87	0.91

In the future, we will be interested in adaptively selecting the similarity measures according to the heterogeneous characteristics on two sensor ontologies to be aligned. Moreover, when the scale of the sensor ontology becomes huge, an efficiency-improving strategy should be introduced, such as the divide-and-conquer method [25].

Data Availability

The data used to support the findings of the study are available from the corresponding author upon request.

Conflicts of Interest

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

TABLE 4: Sensitive experiment on evolutionary algorithm's configuration.

Parameter name	Parameter value	The alignments' mean f -measure
Population size	20	0.90
	40	0.96
	60	0.96
	100	0.96
Maximum generation	500	0.90
	1000	0.92
	2000	0.96
	5000	0.96
Crossover probability	0.2	0.61
	0.4	0.74
	0.6	0.96
	0.8	0.92
Mutation probability	0.01	0.82
	0.02	0.96
	0.05	0.85
	0.10	0.71

Acknowledgments

This work was supported by the Natural Science Foundation of Fujian Province (no. 2020J01875) and the National Natural Science Foundation of China (nos. 61103143 and 61801527).

References

- [1] X. Wang, X. Zhang, and M. Li, "A survey on semantic sensor web: sensor ontology, mapping and query," *International Journal of u-and e-Service, Science and Technology*, vol. 8, no. 10, pp. 325–342, 2015.
- [2] X. Xue and J. Chen, "Using compact evolutionary Tabu search algorithm for matching sensor ontologies," *Swarm and Evolutionary Computation*, vol. 48, pp. 25–30, 2019.
- [3] P. Shvaiko and J. Euzenat, "Ontology matching: state of the art and future challenges," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 158–176, 2011.
- [4] X. Xue and Y. Wang, "Optimizing ontology alignments through a memetic algorithm using both MatchFmeasure and unanimous improvement ratio," *Artificial Intelligence*, vol. 223, p. 6581, 2015.
- [5] Z. Wang, "A semi-supervised learning approach for ontology matching," in *Proceedings of the 8th Chinese Semantic Web and Web Science Conference*, pp. 17–28, Wuhan, China, August 2014.

- [6] N. Alboukaey and A. Joukhadar, "Ontology matching as regression problem," *Journal of Digital Information Management*, vol. 16, no. 1, pp. 34–42, 2018.
- [7] M. A. Khouidja, M. Fareh, and H. Bouarfa, "Ontology matching using neural networks: survey and analysis," in *Proceedings of the 2018 International Conference on Applied Smart Systems (ICASS)*, pp. 1–6, Medea, Algeria, November 2018.
- [8] M. Tounsi Dhouib, C. Faron Zucker, and A. G. B. Tettamanzi, "An ontology alignment approach combining word embedding and the radius measure," in *Proceedings of the 2019 International Conference on Semantic Systems*, pp. 191–197, Karlsruhe, Germany, September 2019.
- [9] A. Ali, M. Hamid, K. Ahmad, and D. Wajdi, "Context aware instance matching through graph embedding in lexical semantic space," *Knowledge-Based Systems*, vol. 186, pp. 422–433, 2019.
- [10] F. Ali, K.-S. Kwak, and Y.-G. Kim, "Opinion mining based on fuzzy domain ontology and support vector machine: a proposal to automate online review classification," *Applied Soft Computing*, vol. 47, pp. 235–250, 2016.
- [11] A. Siham, M. Sihem, and F. Muhammad, "Decision trees in automatic ontology matching," *International Journal of Metadata Semantics and Ontologies*, vol. 11, no. 3, pp. 180–190, 2016.
- [12] S. Duan, A. Fokoue, and K. Srinivas, "One size does not fit all: customizing ontology alignment using user feedback," in *Proceedings of the 9th International Semantic Web Conference*, pp. 197–192, Bonn, Germany, November 2010.
- [13] M. Ehrig, S. Staab, and Y. Sure, "Bootstrapping ontology alignment methods with APFEL," in *Proceedings of the 4th International Semantic Web Conference*, pp. 186–200, Galway, Ireland, November 2005.
- [14] X. Xue, X. Wu, J. Zhang, L. Zhang, H. Zhu, and G. Mao, "Aggregating heterogeneous sensor ontologies with fuzzy debate mechanism," *Security and Communication Networks*, vol. 2021, Article ID 2878684, 12 pages, 2021.
- [15] X. Xue, X. Wu, C. Jiang, G. Mao, and H. Zhu, "Integrating sensor ontologies with global and local alignment extractions," *Wireless Communications & Mobile Computing*, vol. 2021, Article ID 6625184, 10 pages, 2021.
- [16] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics*, vol. 10, pp. 707–710, 1966.
- [17] G. A. Miller, "WordNet," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [18] X. Xue and Y. Wang, "Using memetic algorithm for instance coreference resolution," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 580–591, 2016.
- [19] E. Jerome, "Semantic precision and recall for ontology alignment evaluation," in *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 348–353, San Francisco, CA, USA, January 2007.
- [20] X. Xue and J. Zhang, "Matching large-scale biomedical ontologies with central concept based partitioning algorithm and adaptive compact evolutionary algorithm," *Applied Soft Computing*, vol. 106, pp. 1–11, 2021.
- [21] P. Xu, Y. Wang, L. Cheng, and Z. Tianyi, "Alignment results of SOBOM for OAEI 2010," *Ontology Matching*, vol. 203, pp. 7–11, 2010.
- [22] J. Noessner, M. Niepert, C. Meilicke, and H. Stuckenschmidt, "Leveraging terminological structure for object reconciliation," in *Proceedings of the Extended Semantic Web Conference*, pp. 334–348, Heraklion, Greece, June 2010.
- [23] Y. R. Jean-Mary, E. P. Shironoshita, and M. R. Kabuka, "Ontology matching with semantic verification," *Journal of Web Semantics*, vol. 7, no. 3, pp. 235–251, 2009.
- [24] S. Fernandez, I. Marsa-Maestre, J. Velasco, and B. Alarcos, "Ontology alignment architecture for semantic sensor web integration," *Sensors*, vol. 13, no. 9, pp. 12581–12604, 2013.
- [25] E. Jimenez-Ruiz, A. Agibetov, M. Samwald, and V. Cross, "We divide, you conquer: from large-scale ontology alignment to manageable subtasks with a lexical index and neural embeddings," in *Proceedings of the International Conference on Digital Culture & AudioVisual Challenges: Interdisciplinary Creativity In Arts And Technology*, pp. 13–24, Corfu, Greece, June 2018.