
A Novel Method to Associate Sensor Data with Domain Ontology

Jin Liu, Yihe Yang, Shengjie Shang

College of Information Engineering, Shanghai Maritime University, Shanghai, China

Email address:

jinliu@shmtu.edu.cn (Jin Liu), yangyihe35@stu.shmtu.edu.cn (Yihe Yang), shangshengjie@stu.shmtu.edu.cn (Shengjie Shang)

To cite this article:

Jin Liu, Yihe Yang, Shengjie Shang. A Novel Method to Associate Sensor Data with Domain Ontology. *International Journal of Data Science and Analysis*. Vol. 5, No. 4, 2019, pp. 52-60. doi: 10.11648/j.ijdsa.20190504.11

Received: July 6, 2019; **Accepted:** July 26, 2019; **Published:** August 16, 2019

Abstract: With the development of the Internet of Things, sensor ontologies have been applied to a variety of fields. Most sensor ontologies are currently built for applications in specific domains, and these ontologies are usually heterogeneous, making it difficult to share or reuse knowledge and concepts. The ontology association methods can be used to construct the semantic mapping between heterogeneous ontologies, so as to effectively determine the similarity between concepts in the ontologies. However, most of the contemporary methods do not make full use of the information that is stored in ontologies and are insufficient for the effective association. This paper proposes a novel association method based on comprehensive similarity. In our proposed method, we first use How-Net to obtain concept representation and calculate the semantic similarity of ontology concepts through sememe Tree and sememe Hierarchy. Then we calculate the structural similarity by the internal structure and the hierarchical relationship between the ontologies and remove the conceptual pairs with low relevance. Finally, we combine the semantic similarity and structural similarity to calculate the similarity matrix between ontology concepts to achieve association. The experimental results on real data show that our method can effectively associate sensor data with domain ontology by combining two different similarity calculation methods.

Keywords: Ontology, Semantic Similarity, Structural Similarity, Sensor Data

1. Introduction

In the past few decades, more IoT (Internet of Things) applications [1], sensor data-based software [2], and social network applications [3] have been generating data. The rapid growth and repeatability of data [4] make it difficult for people to locate desired information efficiently. The presence of the ontology improves this situation. On the basis of the World Wide Web, the Semantic Web was proposed, which adds semantic information that machines can understand for various documents, so that there is a semantic association between Web documents, making it an intelligent network. The construction of the Semantic Web involves several critical technologies, and the ontology is one of them. The concept of ontology first appeared in the field of philosophy, which studied whether a noun represented a real entity or a concept. In the 1970s and 1980s, with the development of artificial intelligence, the ontology was introduced into the computer field. The ontology builds models based on a specific category system to express concepts, entities,

attributes, relationships, and so on. Ontology is a formal representation of a set of important concepts in a particular domain.

The original sensor data lacked semantic information. In 2009, the World Wide Web Consortium established the Semantic Sensor Network (SSN) to develop an ontology that describes the characteristics of sensors and sensor networks, specifying a semantic tagging language for sensor network applications. With the development of the IoT, the construction of sensor ontology is becoming more and more important. The sensor ontology can be defined as a five-tuple:

$$SO = (C, R, M, A, I) \quad (1)$$

Where C represents a set of concepts, R represents a set of relations, M represents a set of methods, A represents a set of axioms, and I represents a set of instances. The concept refers to the abstraction of sensors, sensor data, sensor properties and so on. Relationship refers to the association between concepts and forms a hierarchical structure between

concepts. The method is a further explanation of the concept. The axiom refers to the basic formal statement which can be used to derive other axioms. An instance is a specific object that belongs to a concept.

Ontology has been widely recognized as the basis of a structured description and inference of domain knowledge, providing sharable semantic descriptions for entities in the application domain to better share and reuse knowledge. Ontology is generally applied to the field of semantic-based information processing, especially in the fields of knowledge reasoning, information integration, and knowledge engineering. Although the original intention of establishing ontology is to share knowledge well, with the continuous development of ontology technology, a large number of sensor ontologies are generated, but they have heterogeneity in syntax, pattern, and semantics, which is difficult to share and reuse. At present, there is still great difficulty in establishing a globally universal large ontology. Faced with these heterogeneous ontologies, how to achieve the association among sensor ontologies and realize mutual sharing and reuse to improve the interoperability of sensor ontologies is challenging research topics.

There are many reasons for the emergence of heterogeneous ontology. First, because of the distribution of the ontology itself, the representation of ontology is not completely consistent with its internal logic. Secondly, the existing ontologies are mostly constructed by different ontology development teams, so there is no unified standard to constrain the ontology construction. Different development teams will choose the method that matches the current research background according to their own needs, so that even the ontologies built in the same field may be different. Besides, different team's habits of using terminology, cultural background knowledge, and actual application are different, and their understanding of the objective world and methods are not exactly the same. In heterogeneous ontologies, the same entity may be expressed in different concepts, and the relationships between concepts may be different. The classification of concepts may also be different. The heterogeneity of the ontology will make the application of information communication on the upper layer of the ontology very difficult, and it is contrary to the original intention of sharing knowledge. Therefore, researchers have proposed many solutions to establish a mapping relationship between two heterogeneous ontology concepts.

Ontology mapping refers to the process of mapping the entity of the source ontology to the target ontology entity through a semantic association between the two ontologies. Ontology mapping enables heterogeneous ontologies to achieve a consistent understanding of the same thing and determine how different ontologies are mapped or related [5]. Ontology mapping solves the problem of complex information exchange in the Semantic Web. It can effectively share the knowledge in heterogeneous ontologies [6]. In practice, ontology mapping takes two heterogeneous ontologies as input and then outputs a semantic mapping

relationship between two ontology concepts based on the mapping rules. The mapping relationships between semantics are generally determined by their similarity. The initial mapping process was done manually. At that time, the amount of data was relatively small so that the manual construction method could accomplish the task better. However, as the number of data increases, the information represented by the ontology was more and more. Merely relying on these methods cannot get good results. Moreover, the mapping process constructed by hand is too subjective and difficult to reuse. In addition, in some traditional ontology mapping systems, there are still problems such as large computation, low accuracy, and low mapping efficiency. Therefore, there is an urgent need to find some ways to better complete the association process.

Considering the shortcomings of current research, we propose a new method based on the comprehensive similarity to calculate the similarity between sensor data and domain ontology. The rest of this paper is structured as follows. In Section 2, we introduced the related work of ontology mapping and association method in the literature. In Section 3, we reconstruct the association rules of sensor data and domain ontology and propose a new similarity calculation method. In Section 4, the experimental result of a case study for semantic inference in berth management is introduced. Finally, in Section 5, we summarize the entire paper.

2. Related Work

Sensors are the low-level hardware that plays an essential role in detecting, sensing and collecting environmental information, such as warehouse security management [7], fire detection and warning [8], home automation, and remote patient monitoring [9]. Due to different descriptions of concepts in the sensor field and data processing methods, heterogeneous problems exist in sensor networks, which makes it difficult to share, reuse, and integrate different sensor data. Therefore, scholars have introduced ontology into sensors to construct sensor ontologies. Huang and Javed [10] applied the structure of the Semantic Web to the sensor network so that sensor data could be understood and processed by computers. Ni *et al.* [11] proposed a semantic sensor network that incorporates dynamic semantic information into sensor data, allowing sensor data to be reused. Liu *et al.* [12] proposed a new approach to construct multi-domain ontology for large-scale unstructured text. Kim *et al.* [13] proposed a service-oriented sensor ontology. Atanasov *et al.* [14] proposed a layered approach to model semantic information for IoT applications, transforming IoT data into reusable knowledge, and improving system energy efficiency.

To associate sensor data to domain ontology, we need help from sensor ontology. Therefore, we need to calculate the similarity between the sensor ontology and the domain ontology. The sensor ontology refers to the interrelationship

between concepts in the domain of sensor. The sensor ontology is also a domain ontology. In the field of sensors, there are mainly 12 famous sensor ontologies. [15] However, these sensor ontologies are built for specific applications, and the standards for their construction are not uniform, which leads to the definition of each concept is not the same, so it is difficult to achieve the sharing of the sensor ontology [16]. Therefore, this paper studies the similarity calculation method in the ontology mapping system to associate the sensor data with domain ontology. The system establishes a connection between heterogeneous ontologies to achieve a consistent understanding of the same thing between the two ontologies. The ontology mapping method determines how different ontologies are mapped and associated. In general, the ontology mapping system in a specific domain uses specific domain rules, heuristic learning, or background knowledge to assist association. [17]

The methods in ontology mapping system can be classified from different angles. According to the principle of methods, there are mainly the following three categories: methods based on rules, methods based on statistical and methods based on machine learning. According to the characteristics of the ontology, there are mainly following four categories: methods based on linguistic features, methods based on structural features, methods based on external resources, and methods based on logical reasoning.

2.1. Methods Based on Rules

The rule-based method is to define some heuristic rules in the ontology mapping. These rules are mostly derived from the definition of concepts and their structural information. In some specific application areas, domain rules can be applied to ontology mapping as ontology mapping rules. Fan et al. [18] proposed an improved multi-strategy concept similarity computing approach, which discusses the similarity of ontology concepts from four aspects: ontology concept name, attributes, instances, and hierarchies. Ehrig et al. [19] proposed a similarity calculation method based on heuristic rules. They implement ontology mapping by developing corresponding coding rules by experts in the ontology domain. After careful evaluation, the experimental results show that the method is effective. A novel comprehensive method was proposed by Li et al. [20]. This method uses different mapping rules to compute ontology similarity. This method constructs a structural tree model of association rules and calculates the similarity of conceptual structures according to association rules. Then the method calculates the final result based on Multiple feature similarities. This method achieves better results and reduces time complexity compared to RiMOM [21]. Wang et al. [22] proposed a new ontology mapping method, which can perform a series of filtering processes through the structural relationship between ontology entities, avoiding unnecessary mapping operations between source ontology and target ontology. Based on ontology concepts, Faria et al. [23] proposed AWL, a novel ontology mapping system. In the experiment, AML achieved high accuracy, but the F1 value and recall rate were relatively low.

2.2. Methods Based on Statistical

The statistical-based method refers to the use of statistical methods in the mapping process. In general, the traditional method assumes that the numerical properties of the concept are entirely independent and does not consider the dependencies therein. In fact, the value of each numeric attribute corresponds to the numerical attribute in the concept. Therefore, the joint distribution of different attribute values can be considered as features for ontology mapping. Prasad et al. [24] proposed a method based on Bayesian, while Doan et al. [25] proposed a method based on joint probability distribution. The method combines a variety of machine learning methods and adopts a hybrid ontology mapping system architecture to further improve the accuracy of ontology concept matching. However, the mapping accuracy of the experimental results on different test sets is 66% to 97%, indicating that the method does not have good stability.

2.3. The Methods Based on Machine Learning

Ontology mapping based on machine learning generally refers to the use of machine learning methods to transform mapping problems into classification problems in the mapping process. These methods typically require a data set of known mapping results as a training set to train a particular machine learning model. Common machine learning methods are support vector machines, random forests, deep learning networks, and so on. Xiang et al. [26] used vocabulary information, linguistic information, and semantic information of How-Net and Chinese Concept Dictionary (CCD) as mapping features to train the machine learning model. Finally, the initial mapping results are adjusted by using the similarity propagation algorithm. This method has achieved high accuracy in the experiment. Liu et al. [29] proposed a gradient descent algorithm based on the stochastic conjugate gradient and optimized the negative sampling process based on part-of-speech tagging to increase the accuracy of the language model. The experimental results show that the method is effective.

2.4. Methods Based on Linguistic Features

The domain ontology contains the concept definitions and language descriptions for specific domains, so the similarity between concepts can be obtained by comparing differences in character sequences. This method generally uses the edit distance to represent the difference between the sequences of characters. If the value of edit distance is large, the similarity between strings is small. There are many ways to calculate the edit distance, such as Levenshtein distance [27], I-sub method [28], and so on. With the development of deep learning, the word vector constructed by the deep learning model has also appeared to calculate the similarity between texts. Liu et al. [29] uses global information and sliding window context to characterize textual information.

2.5. Methods Based on Structural Features

The ontology can usually be represented as a tree or a graph, so ontology mapping can be performed using a tree operation or a graph theory method. These methods usually require constructing an output map between the ontologies as input, and a new mapping can be obtained by using a similarity propagation algorithm. This approach suggests that if the two concepts are similar, it is likely that their child concept or ancestor concept is similar. Sekine *et al.* [30] used WordNet and EDR as references and conducted several experiments based on the parent node, child nodes, and grandchild nodes. The experimental result proved that parent nodes and child nodes are more important.

2.6. Methods Based on External Resources

Domain resources usually have known background knowledge, and the use of dictionaries or third-party ontologies can provide a better reference for ontology mapping. At present, the use of dictionaries for ontology mapping generally based on WordNet. WordNet can classify words according to their meanings. Therefore, when calculating the similarity, the judgment of the similarity can be based on whether the concept of the ontology belongs to the same synonym set in WordNet or whether the distance is close enough. Budanitsky *et al.* [31] calculated semantic similarity based on WordNet ontology and path distance. Jin *et al.* [32] proposed a How-Net based ontology mapping method, which first calculates the conceptual similarity of ontology and then analyses its structural similarity. The method based on the third-party ontologies can usually treat the instance data in the ontologies as an external resource. When the instance sets of the two classes have a large intersection, the two classes are likely to be similar. A typical method is the Jaccard method [33]. This method is used for the similarity between sets, but it is not possible to measure the specific difference value of the measure set, just to get the result like whether the two sets are the same. Therefore, the Jaccard coefficient method only cares about whether the features between the sets are consistent.

2.7. Methods Based on Logical Reasoning

Logical reasoning-based methods are often used in conjunction with other methods because using this method alone does not produce good ontology mapping results. The mapping method based on logical reasoning can effectively utilize the semantic information in the ontology for reasoning and can also verify the mapping results, such as, reasoning or correcting the results calculated by other methods. There are currently not many mapping methods based on logical reasoning. Kumar *et al.* [34] construct a domain-specific ontology mapping based on a method of describing logic.

2.8. Other Methods

In addition to some of the above methods, it is also possible to find semantic relations between similar entities in

heterogeneous ontologies to implement ontology mapping. The association of semantic information is the cornerstone of the semantic sensor network, which plays an important foundation for it. For semantic association problem, Wang *et al.* [35] proposed a data access control model for individual users. This model is mainly based on the semantic integration nature of XML data. Xiong *et al.* [36] proposed a new neural network model. The model improves the traditional algorithm and constructs a distributed vector representation of semantics based on CBOW model. What is more, the model uses the CNN model to train a large amount of text to achieve the semantic association between texts. Experiments show that this neural network model has high semantic relevance and accuracy. Wang *et al.* [37] proposed a mapping algorithm based on semantic similarity. The method constructs the connection between the knowledge ontologies and the Predicate-Argument structure. This mapping algorithm is also capable of querying the semantics of multiple ontologies. Pinkel *et al.* [38] extended the previous ontology mapping benchmark and proposed a new version of RODI. RODI tests the accuracy of the system by using test samples from scientific conferences and various exploration fields. The newly proposed RODI can evaluate all ontology mapping system that associate concepts with ontologies. Wang *et al.* [39] proposed a dynamic mapping strategy, which analyzes similar information and entities, and uses function decision methods to determine the combined weights of different ontology features. The experiment has achieved good results. Bockj *et al.* [40] used the MapPSO algorithm to solve the ontology alignment problem of semantic applications under heterogeneous knowledge base.

In summary, the contemporary ontology mapping systems lack effective measures for big data. Besides, since the number of ontologies increases rapidly and the ontologies lack a relatively uniform framework, the ontologies established in the same domain is not exactly the same, and there is no suitable general method for associating the sensor ontology with the domain ontology. In some ontology mapping systems, the calculation of similarity between ontologies is complex, and some methods only calculate similarities in certain aspects, such as concepts, attributes, and instances. However, not all concepts, attributes, and instances have similarities. There are also some methods with a single algorithm, resulting in lower computational accuracy, making mapping inefficient. In addition, there are also many methods to use only names as features of similarity calculations. Therefore, how to implement automatic mapping and reduce semantic conflict and human intervention are important research directions. For the problems mentioned above in the similarity calculation in ontology mapping, we propose a novel association method based on comprehensive similarity.

3. Method

We propose a method for correlating sensor ontology and

domain ontology based on compound similarity. The compound similarity is built on semantic similarity and structural similarity. In order to implement ontology mapping, we need to calculate the similarity between ontology concepts. The higher the similarity between the ontologies, the more likely it is to represent the same ontology concept. In the same domain, the ontologies constructed for different applications have certain similarities in its structure. Besides, the semantics of different concepts between ontologies also have certain similarities. Therefore, this paper proposes a new method to calculate similarity based on the similarity of conceptual semantics and the similarity of ontology structure.

3.1. Semantic Similarity Calculation

First, we calculate the similarity of the conceptual semantics of the ontologies. We take advantage of How-Net, which can express the relationship between the commonality and personality of the concept. We calculate the semantic similarity of sememe through sememe Tree and sememe Hierarchy. The method of this paper combines the depth of sememe and the density of sememe.

In general, the deeper the depth of sememe, the smaller the distance of sememe; the greater the density of sememe, the smaller the distance of sememe. The sememe depth is the distance from the sememe to the root node in the sememe tree. When the other conditions are the same, the similarity of the sememe pair in the upper layer of the sememe tree is lower than that of the sememe pair in the lower layer. The sememe density refers to the proportion of the number of brothers in the sememe layer. When other conditions are the same, the similarity of a sememe pair with more siblings will be higher than that of a sememe pair with a small number of siblings.

$$\text{Sim}_{sem}(S_1, S_2) = \frac{\alpha P(S_1, S_2)}{D(S_1, S_2) + \alpha P(S_1, S_2)} \quad (2)$$

$$D(S_1, S_2) = \frac{Q(S_1, S_2)}{W(S_1) + W(S_2)} \quad (3)$$

$$W(S) = \beta \text{Dep}(S) + (1 - \beta) \text{Den}(S) \quad (4)$$

$$\text{Den}(S) = \frac{\log f(S)}{\log \max} \quad (5)$$

Where S refers to sememe, S_1, S_2 refers to two different sememes, and $P(S_1, S_2)$ is the distance from the smallest common parent node of S_1, S_2 to the root node. $D(S_1, S_2)$ refers to the distance metric between two different sememes, and $Q(S_1, S_2)$ refers to the sum of the distances of two different sememes to the smallest common parent node, $W(S)$ means the distance weight of the sememe, which is determined by the depth of sememe and the density of sememe. $\text{Dep}(S)$ represents the depth of the sememe, $\text{Den}(S)$ represents the density of the sememe, and $f(S)$ represent the number of siblings of the sememe. \max represents the number of nodes of the sememe tree, and α, β are adjustment weights.

3.2. Structure Similarity Calculation

Then we calculate the structural similarity of the

ontologies. First, we need to turn the ontology structure chart into ontology structure tree. Then we calculate the structural similarity between the conceptual nodes in the heterogeneous ontology trees according to the corresponding rules.

In the two heterogeneous ontology trees, we perform similar searches by setting different search radius so that we can get a set of concepts for the same search scope. We set the search radius $r = 3, 5, 7$ and use $N(S)$ to represent these two related concept sets. Before calculating the structural similarity, we need to make certain assumptions about the structure of the ontology: in the constructed ontology tree, if we want to determine that the two concepts are similar, then their uncle nodes of the parent nodes need to be similar. Second, if the two concept nodes are similar, their respective sibling nodes and their children nodes are similar. Based on the above hypothetical rules, we choose the Jaccard coefficient method to compare the similarities between the two concepts. The structural similarity we propose needs to be calculated as follows:

$$\text{Sim}_{str}(S_1, S_2) = \frac{\sigma \text{sim}_u(S_1, S_2) + \tau \text{sim}_s(S_1, S_2) + \nu \text{sim}_b(S_1, S_2)}{\sigma + \tau + \nu} \quad (6)$$

$$\text{Sim}_u(S_1, S_2) = \text{Sim}_{jac}(N_u(S_1), N_u(S_2)) \quad (7)$$

$$\text{Sim}_s(S_1, S_2) = \text{Sim}_{jac}(N_s(S_1), N_s(S_2)) \quad (8)$$

$$\text{Sim}_b(S_1, S_2) = \text{Sim}_{jac}(N_b(S_1), N_b(S_2)) \quad (9)$$

$$\text{Sim}_{jac}(N(S_1), N(S_2)) = \frac{J(N(S_1), N(S_2))}{J(\overline{N(S_1)}, N(S_2)) + J(N(S_1), \overline{N(S_2)}) + J(N(S_1), N(S_2))} \quad (10)$$

$$J(\overline{N(S_1)}, N(S_2)) = \frac{|U_1^{\overline{N(S_1)}, N(S_2)}| + |U_2^{\overline{N(S_1)}, N(S_2)}|}{|U_1| + |U_2|} \quad (11)$$

$$J(N(S_1), \overline{N(S_2)}) = \frac{|U_1^{N(S_1), \overline{N(S_2)}}| + |U_2^{N(S_1), \overline{N(S_2)}}|}{|U_1| + |U_2|} \quad (12)$$

$$J(N(S_1), N(S_2)) = \frac{|U_1^{N(S_1), N(S_2)}| + |U_2^{N(S_1), N(S_2)}|}{|U_1| + |U_2|} \quad (13)$$

S_1, S_2 represent the concepts in two ontologies. $\text{Sim}_u, \text{Sim}_s, \text{Sim}_b$ respectively represent the similarity between the sets of uncle nodes, child nodes, and sibling nodes of parent nodes of S_1, S_2 . $N(S)$ represents the set of related nodes of the concept S . N_u, N_s, N_b represent the set of uncle nodes, children nodes, and siblings of the parent nodes. Sim_{jac} is the calculated Jaccard similarity. $U_1^{N(S_1), N(S_2)}, U_2^{N(S_1), N(S_2)}$ respectively represent a set in the ontology O_1 and ontology O_2 that belongs to both the set $N(S_1)$ and the set $N(S_2)$. $|U_1|, |U_2|$ respectively represent the number of concept sets in ontology O_1 and ontology O_2 . $U_1^{\overline{N(S_1)}, N(S_2)}, U_2^{\overline{N(S_1)}, N(S_2)}$ respectively represent a set in the ontology O_1 and ontology O_2 that belongs only to the set $N(S_2)$. Similarly, $U_1^{N(S_1), \overline{N(S_2)}}, U_2^{N(S_1), \overline{N(S_2)}}$ respectively represent a set in the ontology O_1 and ontology O_2 that belongs only to the set $N(S_1)$. In the structure tree, the structure of uncle nodes, child nodes, and sibling nodes are difficult to have the same effect on the similarity between concepts. Therefore, in order to indicate the contributions of different node structures, we set the contribution coefficients σ, τ, ν for

uncle nodes, the children nodes, and sibling nodes, respectively. The larger the value of the coefficient, the more important it is for the node to calculate the conceptual similarity. In addition, we set $\sigma + \tau + \nu = 1$, $\sigma \geq \tau \geq \nu$.

In addition to the different sets of nodes affecting the calculation of similarity, the different search radius we initially set will also have different effects on it. Therefore, we set the influence coefficients σ_r, τ_r, ν_r for different search radius. The calculation formula is as follows:

$$\text{Sim}_{str,r}(S_1, S_2) = \frac{\sigma_r \text{sim}_u(S_1, S_2) + \tau_r \text{sim}_s(S_1, S_2) + \nu_r \text{sim}_b(S_1, S_2)}{\sigma_r + \tau_r + \nu_r} \quad (14)$$

$$\text{Sim}_{str}(S_1, S_2) = \frac{\sum_r \text{sim}_{str,r}(S_1, S_2)}{\text{num}(r)} \quad (15)$$

Where $\text{num}(r)$ represents the set number of searches.

3.3. Improvement Strategy and Comprehensive Similarity

In the process of ontology mapping, we can unify the mapping results among ontologies into a Cartesian product of the number of concepts among mapping ontologies. The output of the ontology mapping system is a $m \times n$ matrix, where m, n represent the number of concepts in the ontology O_1, O_2 . In the matrix, $\text{Mat}[i][j]$ represents the similarity between the i th concept in the ontology O_1 and the j th concept in the ontology O_2 . Where $0 \leq i \leq m - 1, 0 \leq j \leq n - 1$.

According to the similarity calculation strategy in Sections 3.1 and 3.2, we calculate the similarity for each concept pair. We do not directly calculate the sum of the two similarities as the final criterion. Since the calculation of semantic similarity and structural similarity has different computational focus, we have performed some optimization operations on the calculation process. Firstly, we can calculate an initial matrix Mat based on semantic similarity. The results of semantic similarity calculations will have some conceptual pairs with low similarity. We exclude these low-degree concept pairs from consideration to reduce the number of calculations. We set a threshold parameter T_{sem} here to preserve the concept pairs with higher similarity. Next, we calculate the structural similarity. We only need to select the conceptual similarity of the concept pairs we are interested in in the matrix. We will still set a threshold parameter T_{str} here for filtering.

Finally, we calculate the similarity results by combining semantic similarity and structural similarity:

$$\text{sim}(S_1, S_2) = w \text{Sim}_{sem}(S_1, S_2) + (1 - w) \text{Sim}_{str}(S_1, S_2) \quad (16)$$

Where w and $1 - w$ are the weights of the two similarities. In this process, we filter the similarity between semantic similarity and structural similarity to get the mapping matrix. When all the similarity calculations are completed, we determine that the i th concept in the ontology O_1 is related to the j th concept in the ontology O_2 by looking up the i th row and the j th column of the maximum value of the matrix row by row. After calculating the mapping matrix Mat , the mapping rules of the concept pairs in the heterogeneous

ontologies can be obtained.

4. Experimental Results

In the previous chapter, we have introduced our similarity calculation method in detail. In order to verify that our proposed method of combining semantic similarity and structural similarity can be effectively applied to ontology mapping tasks, we have experimented with semantic reasoning cases in berth management.

We used the depth conditions and climatic conditions of the port berths from the sensors registered by 52North [41]. In addition, SOS services [42] also store data collected by different sensors. In the experiment, we will calculate the similarity between the seven concepts in the domain ontology and the nine concepts in the sensor ontology to test whether our method can accurately associate sensor data with domain ontology.

The 16 concepts in the experiment are Light Intensity, Discharge of Water, Rainfall, Wind Velocity, Depth of Water, Temperature, Pressure Value, Air Water Content, PH Value, Air Temperature, Atmospheric Pressure, Wind Power, Humidity, Silt Amount, Water Quality, Geology respectively.

Table 1. Similarity calculation results for Light Intensity, Discharge of Water and Rainfall.

Concepts	Light Intensity	Discharge of Water	Rainfall
Air Temperature	0.200	0.317	0.187
Atmospheric Pressure	0.213	0.113	0.101
Wind Power	0.076	0.075	0.044
Humidity	0.220	0.060	0.263
Silt Amount	0.034	0.008	0.041
Water Quality	0.172	0.121	0.101
Geology respectively	0.085	0.306	0.263

Table 2. Similarity calculation results for Wind Velocity, Depth of Water and Temperature.

Concepts	Wind Velocity	Depth of Water	Temperature
Air Temperature	0.195	0.199	0.274
Atmospheric Pressure	0.249	0.160	0.122
Wind Power	0.049	0.046	0.193
Humidity	0.116	0.012	0.086
Silt Amount	0.039	0.223	0.127
Water Quality	0.177	0.133	0.183
Geology respectively	0.175	0.228	0.015

Table 3. Similarity calculation results for Pressure Value, Air Water Content, and PH Value.

Concepts	Pressure Value	Air Water Content	PH Value
Air Temperature	0.012	0.119	0.245
Atmospheric Pressure	0.314	0.211	0.120
Wind Power	0.079	0.005	0.189
Humidity	0.050	0.166	0.031
Silt Amount	0.207	0.213	0.077
Water Quality	0.132	0.164	0.194
Geology respectively	0.207	0.122	0.145

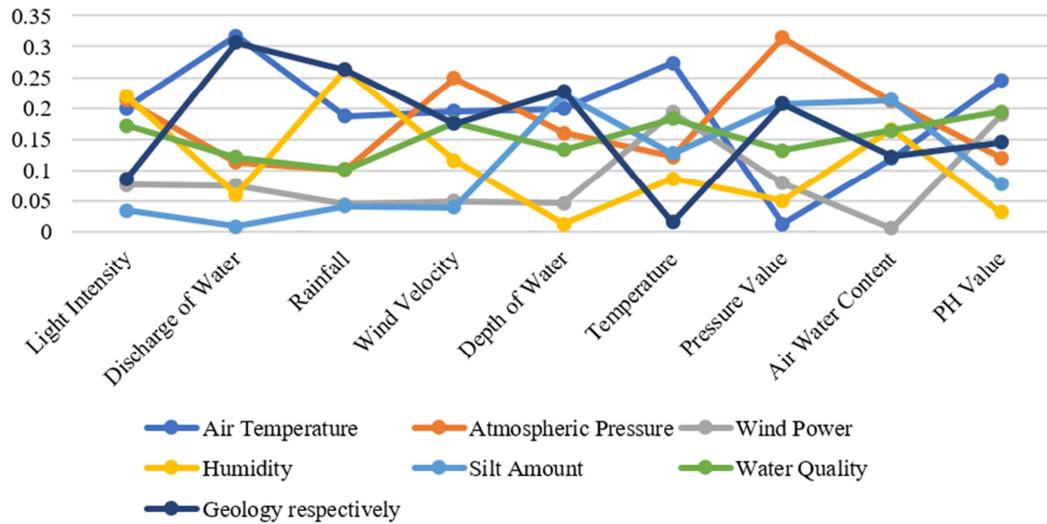


Figure 1. Line chart of similarity calculation results.

We calculate the similarity between concepts by comprehensively evaluating the semantic similarity and structural similarity of the concepts in the sensor ontology and domain ontology to formulate association rules. The similarity calculation results are shown in Tables 1, 2, 3. It can be seen from Tables 1, 2, 3, and Figure 1 that the comprehensive similarity calculated by combining semantic similarity and structural similarity can well associate sensor ontology and domain ontology concepts.

5. Conclusion

With the development of the Internet of Things, sensor data will become more complex. At the same time, more sensor ontologies are applied to sensor data-based applications in various fields. By using an ontology association method, the semantic mapping between heterogeneous ontologies can be constructed, so as to effectively determine the correlation between ontologies. However, most association methods only use part of the information in the ontology so that the heterogeneous ontology cannot be well correlated. In this paper, we propose a novel method by building a compound similarity with semantic similarity and structural similarity to associate sensor data with domain ontology. We first make use of HowNet to obtain concept representation, calculate semantic similarity through sememe Tree and sememe Hierarchy, and remove conceptual pairs with low relevance. Then, we calculate the structural similarity of the ontologies, and also remove the conceptual pairs with low relevance. Finally, we combine two different similarity calculations to obtain the similarity matrix between the two ontology concepts for the association. The experimental results show that the proposed method can well determine the similarity of different concepts between the sensor ontology and the domain ontology through the combination of two similarities. However, some parts of the model use predefined weights, which are subjective to a certain extent. In our future work,

the similarity calculation strategy will continue to be optimized. We will extract more information from the sensor data to achieve a more accurate and efficient sensor ontologies association method.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (61872231, 61701297).

References

- [1] Wang, J., Zhang, Z., Li, B., Lee, S., & Sherratt, R. S. An enhanced fall detection system for elderly person monitoring using consumer home networks. *IEEE transactions on consumer electronics*. Vol. 60, No. 1, 2014, pp. 23-29.
- [2] Liu, J., Li, Y., Tian, X., Sangaiah, A. K., & Wang, J. Towards Semantic Sensor Data: An Ontology Approach. *Sensors*. Vol. 19, No. 5, 2019, pp. 1193.
- [3] Liu, W., Chen, X., Jeon, B., Chen, L., & Chen, B. Influence maximization on signed networks under independent cascade model. *Applied Intelligence*. 2018, pp. 1-17.
- [4] Liu, J., Song, J. J., Kong, L., Kim, J. U., & Wang, J. A Novel Parallel Method for Denoising and Deduplicating Mass Web Documents. *Journal of Internet Technology*. Vol. 17, No. 5, 2016, pp. 889-896.
- [5] Hooi, Y. K., Hassan, M. F., & Shariff, A. M. A survey on ontology mapping techniques. *Lecture Notes in Electrical Engineering*. Vol. 279, No. 1, 2014, pp. 15-33.
- [6] Ma, Z., Zhang, F., Yan, L., & Cheng, J. Fuzzy semantic web ontology mapping. *Studies in Fuzziness & Soft Computing*. Vol. 306, 2014, pp. 157-180.
- [7] Trab, S., Bajic, E., Zouinkhi, A., Abdelkrim, M. N., & Chekir, H. RFID IoT-enabled warehouse for safety management using product class-based storage and potential fields methods. *International Journal of Embedded Systems*. Vol. 10, No. 1, 2018, pp. 71-88.

- [8] Vijayalakshmi, S. R., & Muruganand, S. Different soft computing algorithms used in fire sensor node of wireless sensor network integrated with IoT. *International Journal of Embedded Systems*. Vol. 9, No. 4, 2017, pp. 310-320.
- [9] Schrickte, L. F., Montez, C. B., Oliveira, R. S. D., & Pinto, A. S. R. Design and implementation of a 6LoWPAN gateway for wireless sensor networks integration with the internet of things. *International Journal of Embedded Systems*. Vol. 8, No. 5-6, 2016, pp. 380-390.
- [10] Huang, V., & Javed, M. K. Semantic sensor information description and processing. In *2008 Second International Conference on Sensor Technologies and Applications*. 2008, pp. 456-461.
- [11] Ni, L. M., Zhu, Y., Ma, J., Li, M., Luo, Q., Liu, Y.,... & Yang, Q. Semantic sensor net: An extensible framework. In *International Conference on Networking and Mobile Computing*. 2005, pp. 1144-1153.
- [12] Liu, J., Zhou, M., Lin, L., Kim, H. J., & Wang, J. Rank web documents based on multi-domain ontology. *Journal of Ambient Intelligence and Humanized Computing*. 2017, pp. 1-10.
- [13] Kim, J. H., Kwon, H., Kim, D. H., Kwak, H. Y., & Lee, S. J. Building a service-oriented ontology for wireless sensor networks. In *Seventh IEEE/ACIS International Conference on Computer and Information Science*. 2008, pp. 649-654.
- [14] Atanasov, I., Nikolov, A., & Pencheva, E. An approach to transform Internet of Things data into knowledge. *International Journal of Embedded Systems*. Vol. 9, No. 5, 2017, pp. 401-412.
- [15] Compton, M., Henson, C. A., Neuhaus, H., Lefort, L., & Sheth, A. P. A Survey of the Semantic Specification of Sensors. *International Conference on Semantic Sensor Networks, CEUR-WS.org*. 2009, pp. 17-32.
- [16] Zong-Hao, L. I., Chun-He, X., & Cheng, Z. Research of the sensor network information sharing based on ontology. *Science & Technology Vision*. No. 29, 2013, pp. 5-6.
- [17] Chen, H., Ning-Jing, H. U., & Song, Y. A research on ontology mapping frame. *Journal of Hunan University of Science & Engineering*. Vol. 30, No. 12, 2009, pp. 58-62.
- [18] Fan, W., & Jian, C. Research on multi-strategy ontology mapping based on conceptual similarity computing. *Computer Technology and Development*. No. 4, 2015, pp. 38-42.
- [19] Ehrig, M., & Sure, Y. Ontology mapping—an integrated approach. In *European Semantic Web Symposium, Springer, Berlin, Heidelberg*. Vol. 3053, 2004, pp. 76-91.
- [20] Li, H., & Su, L. Comprehensive method of computing ontology similarity based on association rules. *Jisuanji Yingyong/ Journal of Computer Applications*. Vol. 32, No. 9, 2012, pp. 2472-2475.
- [21] Li, J., Tang, J., Li, Y., & Luo, Q. RiMOM: A dynamic multistrategy ontology alignment framework. *IEEE Transactions on Knowledge and data Engineering*. Vol. 21, No. 8, 2009, pp. 1218-1232.
- [22] Wang, P., Zhou, Y., & Xu, B. Matching large ontologies based on reduction anchors. *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*. 2011, pp. 2343-2348.
- [23] Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I. F., & Couto, F. M. The agreementmakerlight ontology matching system. In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*. 2013, pp. 527-541.
- [24] Prasad, S., Peng, Y., & Finin, T. A tool for mapping between two ontologies using explicit information. In *AAMAS 2002 Workshop on Ontologies and Agent Systems, Bologna, Italy*. 2002.
- [25] Doan, A., Madhavan, J., Domingos, P., & Halevy, A. Learning to map between ontologies on the semantic web. In *Proceedings of the 11th international conference on World Wide Web, ACM*, 2002, pp. 662-673.
- [26] Chuncheng, X., Zhifang, S., Weidong, Z., & University, P. On mapping between hownet and ccd. *Journal of Chinese Information Processing*. Vol. 29, No. 3, 2015, pp. 44-51.
- [27] Levenshtein, V. I. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*. Vol. 10, No. 8, 1966, pp. 707-710.
- [28] Stoilos, G., Stamou, G., & Kollias, S. A string metric for ontology alignment. In *International Semantic Web Conference, Springer, Berlin, Heidelberg*. 2005, pp. 624-637.
- [29] Liu, J., Lin, L., Ren, H., Gu, M., Wang, J., Youn, G., & Kim, J. U. Building neural network language model with POS-based negative sampling and stochastic conjugate gradient descent. *Soft Computing*. Vol. 22, 2018, pp. 6705-6717.
- [30] Sekine, S., Sudo, K., & Ogino, T. Statistical matching of two ontologies. *SIGLEX99: Standardizing Lexical Resources*. 1999, pp. 69-73.
- [31] Budanitsky, A., & Hirst, G. Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures. In *Workshop on WordNet and other lexical resources*. Vol. 2, 2001, pp. 2-2.
- [32] Hai-Tao, J., & Lin, Z. Integrated similarity calculation method based on domain ontology mapping. *Modern Computer*. No. 14, 2017, pp. 34-39.
- [33] Isaac, A., Van Der Meij, L., Schlobach, S., & Wang, S. An empirical study of instance-based ontology matching. In *The Semantic Web, Springer, Berlin, Heidelberg*, 2007, pp. 253-266.
- [34] Kumar, S. K., & Harding, J. A. Ontology mapping using description logic and bridging axioms. *Computers in Industry*. Vol. 64, No. 1, 2013, pp. 19-28.
- [35] Wang, M., Wang, J., Guo, L., & Harn, L. Inverted XML Access Control Model Based on Ontology Semantic Dependency. *CMC: Computers, Materials & Continua*. Vol. 55, No. 3, 2018, pp. 465-482.
- [36] Xiong, Z., Shen, Q., Wang, Y., & Zhu, C. Paragraph Vector Representation Based on Word to Vector and CNN Learning. *CMC: Computers, Materials & Continua*. Vol. 55, No. 2, 2018, pp. 213-227.
- [37] Wang, S., Zhang, L., Zhang, Y., Sun, J., Pang, C., Tian, G., & Cao, N. Natural Language Semantic Construction Based on Cloud Database. *CMC: Computers, Materials & Continua*. Vol. 57, No. 3, 2018, pp. 603-619.
- [38] Pinkel, C., Binnig, C., Jiménez-Ruiz, E., Kharlamov, E., May, W., Nikolov, A.,... & Heupel, C. RODI: Benchmarking relational-to-ontology mapping generation quality. *Semantic Web*. Vol. 9, No. 1, 2018, pp. 25-52.

- [39] Wang, R., Wang, L., Liu, L., Chen, G., & Wang, Q. Combination of the Improved Method for Ontology Mapping. *Physics Procedia*. Vol. 25, 2012, pp. 2167-2172.
- [40] Bock, J., & Hettenhausen, J. Discrete particle swarm optimisation for ontology alignment. *Information Sciences*. Vol. 192, 2012, pp. 152-173.
- [41] Zhang, L., Yin, C. Y., & Chen, J. Chinese word similarity computing based on semantic tree. *Journal of Chinese Information Processing*. Vol. 24, No. 6, 2010, pp. 23-30.
- [42] Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S.,... & Huang, V. The SSN ontology of the W3C semantic sensor network incubator group. *Web semantics: science, services and agents on the World Wide Web*. Vol. 17, No. 4, 2012, pp. 25-32.