A Similarity Measure in Formal Concept Analysis Containing General Semantic Information and Domain Information

FUGANG WANG\textsuperscript{1,2}, NIANBIN WANG\textsuperscript{1}, SHAOBIN CAI\textsuperscript{3}, AND WULIN ZHANG\textsuperscript{2}

\textsuperscript{1}College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China
\textsuperscript{2}School of Physics and Electronic Engineering, Mudanjiang Normal University, Mudanjiang 157011, China
\textsuperscript{3}School of Electronic Information Engineering, Foshan University, Foshan 528225, China

Corresponding author: Nianbin Wang (wangnianbin@hrbeu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772152, in part by the Basic Research Project under Grant JCKY2017604C010 and Grant JCKY2019604C004, in part by the National Key Research Development Program of China under Grant 2018YFC0806800, in part by the Technical Basic Research Project under Grant JSQB2017206C002, and in part by the Pre-Research Project under Grant 10201050201.

\section*{ABSTRACT} Formal concept analysis (FCA) gets into good graces by increasing big data scientists due to its unique advantages. Concept similarity measurement is the key to the FCA-based application. Most of the previous methods are based on set theory and less concerned with semantic information, whereas those methods focusing on semantic information usually rely on ontologies or knowledge bases to obtain the relevant semantic knowledge. However, it is difficult for knowledge methods to obtain domain knowledge in formal contexts (datasets), which are not suited well for domain text data. To tackle these problems, this paper proposes a novel formal concept similarity measure that synthesizes the Semantic information in knowledge bases and Domain information in the formal context (S&D measure). S&D uses word vectors as the representations of words to obtain the semantic information in general knowledge bases while defining novel semantic relations of intent words to obtain the domain information contained in the data itself. It can measure the similarity relation of concepts more comprehensively and precisely, particularly in a domain textual formal context, and it can be implemented automatically and unsupervisedly without any knowledge base, ontology or external corpus. Compared with other related works, experiments show that this method has a better correlation with human judgment.

\section*{INDEX TERMS} Concept lattice, domain information, formal concept analysis, similarity measure.

\section*{I. INTRODUCTION} Formal Concept Analysis (FCA) was introduced by R. Wille as a mathematical theory \cite{1, 2} and became an effective technique that constructs a concept hierarchy (i.e., the concept lattice) by using the binary relation between objects and attributes in the data set (i.e., the formal context, and the context for short), and then analyzes data by using lattice algebraic theory, ensuring the maximum decomposition of information while retaining all data details. Since its inception, FCA has been applied in software engineering, machine learning, knowledge discovery, information retrieval, and other fields \cite{3, 4, 5}. New research results on the FCA-related algorithms \cite{6, 7, 8} have recently made a wider application of FCA in the large-scale data processing field, among which the application under the text context (the textual formal context, that is, take text documents as objects and take the words in the documents as attributes) has a promising future \cite{9}. Just as the recognition of semantic similar concept becomes a fundamental technological component in many fields, such as cognitive science, artificial intelligence, and semantic web, the similarity measurement of formal concept serves as the basis of almost all FCA-based data processing applications and plays a vital role \cite{10, 11, 12, 13, 14, 15}.

Many studies focus on the similarity measurement of formal concepts. However, in those studies, most formal contexts extract fewer features from objects to serve as attributes, which is different from the text contexts that use most words in the documents (objects) as attributes. These techniques work based on set theory \cite{16, 17, 18, 19, 20, 21}, so their results lack semantic content and they are not suitable for processing...
domain text data. The methods in [22]–[24], and [12] take the semantic relation of attribute words into account, which combines the semantic information from ontology (or WordNet) with the statistical information of words from the external corpus (such as the Brown Corpus in [25]), and then calculate the Information Content (IC) of words by using the two information to evaluate the similarity between the attribute words. The contexts of them are mainly limited to small-scale text data in the general domain, such as geographic names, hotel facilities, etc.. Paper [13] evaluates the semantic relations by using Wikipedia as the source of semantic knowledge, and paper [26] utilizes the linked data of DBpedia and the synsets of WordNet to judge the semantic correlation of words. It can be learned that the existing semantic-related methods obtain semantic information mainly by the knowledge base. However, reliance on some alternative knowledge bases inevitably increases the complexity of the system and simultaneously reduces the reliability of it. Furthermore, as a knowledge structure, knowledge bases undoubtedly have the problem of Vocabulary Deficiency (VD). It will be intractable if some words in the data don’t exist in the knowledge bases. Recently, the performance of distributional methods (such as word embedding) for lexical semantic similarity has far exceeded those of knowledge base methods [27]. So, it is simple and effective to calculate the semantic similarity of words based on the semantic information represented by word vectors. Moreover, it is easy for word vector methods to solve the VD problem. On the other hand, domain data contain abundant domain-related information, so it is obviously imprecise to judge the relations of concepts without such information, which cannot be obtained through a general knowledge base.

To overcome these problems, this paper proposes a novel formal concept similarity measure that synthesizes the Semantic information in knowledge bases and Domain information in the formal context (S&D measure), but it is independent of any knowledge base, ontology, or corpus (except the data set). Besides considering the number of the common members (objects or attributes) from the set theory perspective, this method considers the importance of members from the concept lattice perspective, and the measurement of the importance is put forward. Moreover, the semantic similarity of concepts is also reflected in terms of text context, and the measurement of semantic similarity of intents is given. Therefore, the unsupervised S&D method can automatically measure the similarity relation of concepts more comprehensively and precisely, particularly in a domain text context.

The rest of this paper is organized as follows. Section II introduces the basic theory of FCA. Section III gives a brief overview of previous relevant works. Section IV describes the proposed method for measuring formal concept similarity systematically, and then compares it with benchmarks and evaluates it by the real-world data in experiments. The last section makes a summary and proposes future work directions.

II. BASIC THEORY OF FCA

FCA is a mathematical formalism [15] that provides a conceptual framework for analyzing, visualizing and structuring data to get a better understanding of them. The formal concept of FCA is defined in a formal context [22]. A formal context \( F \) is a triple \((O, A, R)\), in which \( O \) is a set of objects, \( A \) is a set of attributes, and \( R \) is a binary relation defined by the Cartesian product of \( O \) and \( A \), that is \( R \subseteq O \times A \). When \( o \in O \) and \( a \in A \), \((o, a) \in R \) or \( oRa \) denotes that object \( o \) possesses attribute \( a \). Given two sets \( E \subseteq O \) and \( I \subseteq A \), then \( E' = \{a \in A | \forall o \in E, oRa\}, I' = \{o \in O | \forall a \in I, oRa\} \). A formal concept \( c \) of \((O, A, R)\) is defined as a pair \((E, I)\), in which \( E' = I \) and \( I' = E \). Here, \( E \) and \( I \) are called the extent and intent of the formal concept \((E, I)\) respectively.

Given two concepts \( c_1 = (E_1, I_1) \) and \( c_2 = (E_2, I_2) \) within the context \( F \), we call \( c_2 \) the subconcept of \( c_1 \) and \( c_1 \) the super-concept of \( c_2 \) (expressed as \( c_2 \subseteq c_1 \)), if \( E_2 \supseteq E_1' \) (or, equivalently, \( I_2 \supseteq I_1 \)) and \( E_2 \subseteq E_1 \). The relation \( \subseteq \) is a partial order relation here, which denotes the double inclusion among concept components (extent and intent), that is, the extent of concept \( c \) is contained in the extent of each of its super-concepts, in turn, the intent of \( c \) contains the intent of each of its super-concepts[22]. For concept \( p, c \) and \( s \) of \( F \), we call \( p \) the parent-concept of \( s \), and \( s \) the son-concept of \( p \), if \( p > s \), and \( \forall c \in F \) while \( p > c > s \). The son-concepts of all the parent-concepts of \( c \) and the parent-concepts of all the subconcepts of \( c \) belong to the same level as \( c \) in terms of inheritance, so these concepts are called general sibling-concepts.

All concepts within \( F \) and their partial order relations constitute concept lattice, denoted as \( L(F) \) or \( L(O, A, R) \). Hasse Diagram is the best and most general representation of concept lattice [2], in which nodes denote concepts and arcs denote their partial order relations. Two most special nodes in concept lattice, the top nodes and the bottom nodes (also called the most general and the most specific concepts), contain the most objects and the most attributes respectively. To construct a concept lattice, it is necessary to identify the relevant objects related to the application domain of the data first and their relevant features. From this point of view, the concept in FCA is not only an abstraction, but also a clustering of objects and their common attributes based on careful observation of reality [22].

III. RELATED WORK

Similarity measurement of formal concepts is of great importance to FCA and many works about it have been done. Paper [28] put forward two similarity measures: the local similarity \( S_l \) and global similarity \( S_g \). For any two formal concepts \( c_1 = (E_1, I_1) \) and \( c_2 = (E_2, I_2) \), \( S_l \) and \( S_g \) are defined as follows:

\[
S_l (c_1, c_2) = \frac{1}{2} \left( \frac{|E_1 \cap E_2|}{|E_1 \cup E_2|} + \frac{|I_1 \cap I_2|}{|I_1 \cup I_2|} \right) \tag{1}
\]

\[
S_g (c_1, c_2) = \frac{1}{2} \left( \frac{|E_1 \cap E_2|}{|O|} + \frac{|I_1 \cap I_2|}{|A|} \right) \tag{2}
\]
where $|O|$ and $|A|$ refer to the number of objects and attributes in the formal context respectively. It can be learned that both measures take into account the coincidence degree of the members in the extents and the intents of two concepts. Local similarity takes the two concepts as a universe, whereas global similarity takes the whole concept lattice.

Paper [29] and [30] discussed the distance of concepts. Similar to the terminologies in [28], local distance and global distance were defined respectively, and then they were combined by some way to obtain the distance of two concepts. These two methods, in essence, are an integration or development of [28]. Paper [29] also discussed the relation between the concept distance in terms of set theory (i.e., the similarity of concepts) and that in terms of lattice theory (i.e., the sibling relationships of concepts), and the conclusion was drawn that there is no consequent correlation between the similarity and the physical distance of concepts in the same lattice. Moreover, the experimental results of [30] showed that for acquiring information, there is no guarantee that the concepts with the best similarity will be found in any of the sibling spaces.

Paper [16] introduced rough set theory into a measure model. First, it finds meet irreducible and join irreducible concepts by virtue of superma and infima structures in a lattice, and then, using these two concepts, calculates Tversky similarities [31] between the extents and the intents of two concepts respectively, finally obtains the similarity of the two concepts by combining the extent Tversky similarity with the intent Tversky similarity. This method preserves both feature information and structural information of the lattice, which can be regarded as the development of Tversky model.

The first similarity measure proposed by [17] is called weighted concept similarity, which is also a combination of two similarity values calculated from the extents and the intents of two concepts respectively, and the formula to compute the similarity of the two extents or intents, according to user’s needs, can be Jaccard index, Sorensen coefficient or Symmetric difference. Paper [17] also proposed another measure, which utilizes the truth that formal concept or the lattice. For two concepts, the intuitive and naive similarity relays on the number of common members, as described in equation (1). However, with regard to text data, the importance of each element (term or word) in bag-of-words (a document or paragraph) models may be different. Therefore, for two pairs of documents, which possess the same number of common members but the different names of them, there is a strong possibility that the similarity of one pair and

that the intents are the decisive factors, so are the only things that need to be compared. Also, paper [21] believes that the attributes in the intent should have different weights based on their categories, and the weight of each attribute should be assigned by counting the frequency proportion of this attribute in different categories.

Some of the above methods are the variants of [28] or Tversky, and the others set the weights of common members in different ways. They are based upon set theory, and none of them think about the effect of the text context on the similarity. As a result, the semantic relation between concepts is ignored.

Paper [22] and [23] observed the influence of semantic relation of words (as attributes) on concept similarity. For two concepts, they calculated the similarity of the extents (based on set theory) and that of the intents (based on the semantic relation of words) respectively, and finally integrated them in some proportion. Measuring the semantic relation of words is the atomic operation of measuring that of intents. The difference between methods of [22] and [23] lies right in the difference of their atomic operations. The former is based on human domain knowledge (ontology), whereas the later on the IC [25] synthesized by the structure of WordNet and the statistical data of external corpora.

Focusing on the semantic similarity measures for FCA, besides WordNet, paper [26] utilized Linked Data provided by the network knowledge base DBpedia as the resource of information. Combining with the idea of possibility theory, it holds that the similarity of two attribute words depends on the number of similar links that they share while the dissimilarity on the number of different links that they have respectively. The dominant feature of this approach lies in the similarity degree it gives, which is not a definite value but an interval composed of minimum and maximum of the similarity of two concepts. As the author says, it will be pursued to apply such interval similarity in different fields that FCA is applied in.

Aiming at the domain text context, this paper proposes a novel formal concept similarity measure, which does not need any knowledge base, ontology or other corpora, but synthesizes the semantic information in the knowledge base and domain information in the formal context, so it is called S&D measure for short. The specific implementation process will be discussed in detail in the next section.

IV. S&D SIMILARITY IN FCA
A concept in FCA consists of extent and intent, in which an object or attribute can be called a member of the concept or the lattice. For two concepts, the intuitive and naive similarity relays on the number of common members, as described in equation (1). However, with regard to text data, the importance of each element (term or word) in bag-of-words (a document or paragraph) models may be different. Hence, for two pairs of documents, which possess the same number of common members but the different names of them, there is a strong possibility that the similarity of one pair and
that of another are different. For instance, there are three sets, \( S_1 = \{a/0.7, b/0.3\} \) (set \( S_1 \) includes elements \( a \) and \( b \), their importance are 0.7 and 0.3 respectively), \( S_2 = \{a/0.6, c/0.4\} \) and \( S_3 = \{a/0.3, d/0.7\} \). Comparing the similarity between \( S_1 \) and \( S_2 \) and that between \( S_1 \) and \( S_3 \), although the two pairs of sets share the same number of the elements, obviously, the two similarity values are different and the former is higher than the latter factually. Therefore, to accurately measure the similarity between two concepts, especially in a text context, the importance degree, rather than the number only, of common members should be taken into account. Also, in regard to the text context, particularly the domain text context, the semantic similarity of the two concepts’ intents, which depends on the number of common words and the semantic relationship between non-common words, should be taken into account.

A. MEMBER IMPORTANCE
Inspired by the idea of the well-known TF-IDF, we believe that if a member is contained by many concepts, the member is relatively ordinary; conversely, if it is contained by few concepts, its peculiarity is stronger. So, if two concepts share a certain number of peculiar members, the similarity of them is higher than that of those share the same number of ordinary members. In view of this, we propose the terminology of Inverse Concept Frequency (ICF). If we refer to the proportion of the concepts including a certain member as Concept Frequency (CF) in a lattice, then ICF measures the reciprocal of CF, which shows the peculiarity degree of a member in a lattice (rather than in a concept).

Definition 1: Given a concept lattice \( L \), the ICF of a member \( e_i \) denoted as \( ICF_i \) can be calculated as follows:

\[
ICF_i = \log \frac{N}{n_i}
\]

where \( N \) is the number of all concepts in the lattice \( L \) and \( n_i \) denotes the number of the concepts including the member \( e_i \). It can be learned from the above statements that a member in a lattice belongs to at least one concept, so the value of \( n_i \) will not be zero. Note that the right-hand side of (3) was used in [19] and [21], here we just give it a name.

For a member (no matter an object or attribute), the higher its ICF is, the stronger its peculiarity is. And, for a pair of concepts, the stronger their common members’ peculiarities are and the bigger these members’ number is, the larger the similarity value of the two concepts is. Different from an object appearing once in a concept, an attribute appears frequently since it is shared by all objects of the concept. Moreover, the times of the attribute appearing in each object vary. Viewing the sum of occurrence times of an attribute in each object of one concept as the total occurrence times of the attribute appearing in the concept, we can conclude that the importance of an attribute with different total occurrence times should be different. Considering the different lengths of objects, we define Attribute Frequency (AF) to measure the importance of an attribute in a certain concept.

\[
AF_{i,j} = \frac{af_{i,j}}{l_j}
\]

where \( af_{i,j} \) denotes the number of the times that \( a_i \) occurs in \( o_j \) and \( l_j \) denotes the length of \( o_j \). It is obvious that \( AF_{i,j} \), in essence, is the mean value of occurrence times of \( a_i \) in each object in \( c_k \). Such a definition avoids that the AF of some attribute becomes oversized, which is caused by a large extent, and expresses the importance degree of every attribute in a certain concept fairly.

In a text context, the intents of two concepts are two different sets of terms (words), between which the similarity depends on, in textual view, not only the number of their common terms but also their own semantic meaning. In the next section, we will discuss the semantic relations of intents.

B. SEMANTIC RELATIONS OF INTENTS
The intent of a concept is a set of words, so the semantic similarity of two words is an element of the semantic similarity of two intents. The measurement of lexical semantic similarity is a lasting topic. At present, computational linguists adopt the methods based on either general knowledge base (such as WordNet, Wikipedia, etc.) [27] or distributional word representation (like word embedding) [32], and, of course, the combination of these two kinds of methods [33]–[35]. In recent years, as using the neural networks trained on very large corpora, the word embedding methods have outperformed knowledge-based methods for predicting human judgments of lexical similarity. Compared with the previous WordNet-based methods, Word2set, a recently published WordNet-based method [27], has improved significantly in performance, which is, however, only close to that of word embedding methods. In terms of the ideal of sense in WordNet, each word always has multiple senses, each of which is usually bound to a particular semantic context, and a context-free word (that is, a word without a context and can be referred to as a dictionary word) is a synthesis of all its senses in WordNet. When a word is in a context, it becomes the specific expression of one of its senses. So it can be said that the semantic relation between words is embodied in the relationship between their senses. The word vector obtained by word embedding in an ideal condition (trained on an ideal large-scale corpus) is an integrated expression of all the senses of the word, so it is not perfect for the meaning of words in a specific context. The model of [36] overcomes the problem, and its word vectors are context-sensitive, which can be employed for tasks such as Word in Context (WiC) similarity [37], [35]. Although the FCA application with a
text context needs to consider the semantics of words in a specific document, it is more important to consider the semantics of them in the entire dataset (the formal context). So it is not a good choice to use the BERT vector in such FCA programs.

The set of words as an intent derives from a set of documents (extent) and is the common words of them. In terms of a group of words, typical differences exist in the expressions of senses between the words shared in multiple documents and those in one document. Multiple terms in a document belong to the same semantic environment, and probabilistically, it is more possible for them to show the single meaning (assuming that the document has a single semantic environment). However, for a group of documents, because the semantic environment of each document is probably to be different, each common term in multiple documents may have a different sense in each document. Therefore, when the document set is large enough, the semantic meanings of common words correspond to those of the word vectors trained on the set. In other words, word vector is essentially the expression of the meaning of a word as a common word in the documents of the large-scale training corpus. Because of this, the word vector [32], [38]–[42] trained on a large general corpus is a proper representation of the words in intent. As mentioned in the introduction, any word representation model will face the challenge of the VD. There is usually a lack of phrase word (such as “New York”), instead of individual word (namely single word, such as “New” or “York” in “New York”; in various models. One advantage of the word representation of vector is that you can combine the word vectors of individual words into the vector of a phrase word, which easily solves the problem of VD.

One obvious problem is that a formal context cannot be infinite or as large as the general training corpus normally, that is, an intent word is not exactly a dictionary word in the ideal (large enough) corpus. Furthermore, their senses will tend to the overall semantics of the formal context due to the influence of its domain properties. In summary, to measure the relation of a pair of intent words, we have to consider not only the comprehensive semantic meanings of them as dictionary words but also the domain-dependent semantic meanings as the members of a context.

For instance, two intent words may have different senses in each document in the context. The closer the semantic relation between the two sets of senses is, which are comprised of all senses of the two words respectively, the more similar the two words are. Similarly, for two words in different concepts, the bigger the intersection of their respective senses is, the more similar the two concepts are. The size of the intersection of senses does not depend on the sizes of senses sets of the two words, but on the formal context (that is, depends on the global semantic properties of the dataset). This implies that the intersection of the two senses sets is a subset of the intersection of all senses of the two words as dictionary words. Therefore, we perform semantic clustering on the context data. Two words in the same semantic category indicate that their senses have a strong correlation, so their semantic relation is closer. Since one word has multiple senses, each word after clustering should be allowed to belong to different categories. Thereupon, we adopt topic clustering, to a certain degree, a clustering of each word according to its senses in the formal context. Each topic is a semantic category, and each word may belong to different categories, which depends completely on the semantic properties of the corpus itself.

Based on it, we define the semantic similarity of two terms in an attribute set.

Definition 3: For concept lattice \((O, A, R)\), given \(t_1, t_2 \in A\), the semantic similarity of term \(t_1\) and \(t_2\) is defined as follows:

\[
Tse(t_1, t_2) = \begin{cases} 
1 & : t_1 = t_2 \\
\frac{V_1 \cdot V_2}{\|V_1\| \cdot \|V_2\|} & : \text{otherwise}
\end{cases}
\]

where \(V_i\) is the word vector of term \(t_i\), \(\|V_i\|\) represents the 2-norm of the vector \(V_i\), and \(n\) is the times that \(t_1\) and \(t_2\) belong to the same topic category. Obviously, when the topics that they belong to have no intersection, the semantic similarity of \(t_1\) and \(t_2\) is the cosine similarity of their respective semantic meaning in the dictionary; and as the number of their common topics increases, they become more similar.

Following the thought of the maximum weighted matching problem in bipartite graphs [43], the semantic similarity between two intents (term groups) is the maximum similarity value matched between them. A non-formal definition is given here.

Definition 4: Sample two sets of terms (that is, two intents) without replacement respectively to form the set of term-pairs, and many such sets will be obtained. For one of the pair sets, calculate and cumulate the semantic similarity value of each pair within the set, and then take the maximum value as the semantic similarity of the two intents.

For example, there are two intents \(I_1 = \{a_1, a_2, \ldots, a_m\}\) and \(I_2 = \{b_1, b_2, \ldots, b_n\}\) (let \(m < n\), and the attributes in the two intents exist in form of the term). Without replacement, take terms out one by one from \(I_1\) and \(I_2\) respectively to form \(m\) term-pairs, which are called one sample. Repeat the sampling process until all combinations are finished. Each of these combinations (i.e., a sample) is a subset of \(I_1 \times I_2\), and in each sampling, any two term-pairs do not share the same term. Find the sum of the similarity values of the \(m\) pairs within each sampling (denoted as \(\sum Tse(a_i, b_j)\) and \(m\) values will be obtained, each of which corresponds to a set of term-pairs. The maximum of these \(m\) values is the semantic similarity between intents \(I_1\) and \(I_2\), expressed as \(STse(I_1, I_2)\).

C. S&D SIMILARITY

Based on the discussion above, the importance of each object or attribute in a lattice varies with its frequency appearing in the lattice, whereas, for a certain attribute, its importance varies with the different concept in which it occurs, because the frequency that it occurs in different concept is different
usually. Therefore, the semantic similarity of the two concepts is related closely with the importance of their members, especially those they share. In addition, for the text context, since the intent is represented as a word set, there must be a semantic relationship between two intents, which will inevitably affect the similarity of the two concepts that they exist in respectively. So, the proposed S&D similarity, which measures the overall similarity of two formal concepts within a text context, is defined as follows.

Definition 5: Given concept lattice \( L \), the semantic similarity of any two concepts \( c_1 = (E_1, I_1) \) and \( c_2 = (E_2, I_2) \) in \( L \) is defined as:

\[
FCS(c_1, c_2) = w_1 \cdot \frac{\sum_{o_k \in E_1 \cap E_2} (ICF_{o_k})^2}{\left( \sum_{o_k \in E_1} (ICF_{o_k}) \cdot \sum_{o_k \in E_2} (ICF_{o_k}) \right)^{1/2}} + w_2 \cdot \frac{\sum_{a_m \in I_1 \cap I_2} \left( (AF_{a_m}^1 + AF_{a_m}^2) \cdot ICF_{a_m} \right)}{2 \left( \sum_{a_m \in I_1} (AF_{a_m}^1 \cdot ICF_{a_m}) \cdot \sum_{a_m \in I_2} (AF_{a_m}^2 \cdot ICF_{a_m}) \right)^{1/2}} + (1 - w_1 - w_2) \cdot \frac{STse(I_1, I_2)}{|I_1| + |I_2|} \tag{7}
\]

where, \( o_k, a_m \) and \( o_k \) are objects, \( a_l, a_{m} \) and \( a_n \) are attributes, \( |I_1| \) and \( |I_2| \) represent the numbers of attributes in \( I_1 \) and \( I_2 \) respectively, and \( 0 < w_1, w_2 < 1 \) are the weight factors.

The first two terms of (7) express set-theory-based similarity between two concepts, which is obtained by using our AF-ICF, and the third term does semantic similarity of the two concepts when they are treated as two sets of words. Generally, set-theory-based concept similarity measures consider that the extents and intents of two concepts make equal contributions to the similarity of the concepts, that is \( w_1 \) and \( w_2 \) should usually be equal. So \( w_1 = w_2 = 1/3 \) becomes an ideal default for (7), where the semantic component (the third term) of the equation is considered to be ideal, which can fully and accurately measure the semantic correlation between the two word-sets (i.e. concepts). Semantic information is contained in the human language, so statistical computation on ideal (large-scale) corpus is a typical way to obtain the exact semantics of words. Obviously, the size of the corpus affects the quality of semantic information obtained. Many other factors affect this quality, which is beyond the scope of this paper and will not be discussed in detail. Here, we just want to show that, in general, the available corpus is difficult to reach the ideal, so we suggest that the weight of the third term of (7) should not exceed 0.3.

When two terms (words) are the same, obviously, they have the largest similarity value of 1. Hence, when two concepts have the same attributes, the semantic similarity between two intents can be simply calculated by:

\[
STse(I_1, I_2) = |I_1 \cap I_2| + STse(I_1 - I_2, I_2 - I_1) \tag{8}
\]

Therefore, another expression of equation (7) for simplified calculation is as follows:

\[
FCS(c_1, c_2) = w_1 \cdot \frac{\sum_{o_k \in E_1 \cap E_2} (ICF_{o_k})^2}{\left( \sum_{o_k \in E_1} (ICF_{o_k}) \cdot \sum_{o_k \in E_2} (ICF_{o_k}) \right)^{1/2}} + w_2 \cdot \frac{\sum_{a_m \in I_1 \cap I_2} \left( (AF_{a_m}^1 + AF_{a_m}^2) \cdot ICF_{a_m} \right)}{2 \left( \sum_{a_m \in I_1} (AF_{a_m}^1 \cdot ICF_{a_m}) \cdot \sum_{a_m \in I_2} (AF_{a_m}^2 \cdot ICF_{a_m}) \right)^{1/2}} + (1 - w_1 - w_2) \cdot \frac{STse(I_1, I_2)}{|I_1| + |I_2|} \tag{9}
\]

Proposed S&D similarity considers the importance of members in both views of the lattice and the concept, which is based on our AF-ICF. Also, it considers the semantic relation between intents as a collection of terms and the semantic relation reflects the domain-related information contained in the dataset, which is implemented by topic clustering. Thus, the S&D measure can work on the similarity relationship between two formal concepts in the domain text context more precisely.

V. EXPERIMENTS AND EVALUATION

Many works have contributed to the measurement of formal concept similarity, but the experimentation and evaluation of it remain a challenge. Paper [23] and [26] explained the reason that each study has its own different inherent assumptions, and, of course, individual advantages. For example, some methods require the support of domain ontology [22], some have the advantage of simplicity [16], some need a specific classification structure [21], and some require the performance of the downstream application to reflect its effect [19], [13]. More importantly, no public data set can be used, on which the performances of various approaches can be compared with each other. Therefore, it is difficult to implement the previous approaches to form an effective benchmark method and further generate comparable results [12], [26]. At present, it is an advisable choice to compare the computerized results of various concept similarity measures with human judgment [45], [13], [26], which provides objective evaluation for these different works.

Paper [26] gives a simple concept lattice and the human judgment scores of the similarity of 8 formal concepts from the lattice. Therefore, this paper selects 10 documents to be the object set \( O = \{o_1, o_2, \ldots, o_{10}\} \) according to the requirement of this paper and defines the same set \( A = \{a_1, a_2, \ldots, a_9\} \), an attribute set containing 9 attributes, in which \( a_1 \sim a_9 \) refer to nine noun phrases respectively: Data mining, Formal Concept Analysis, Artificial Intelligence, Ontology Engineering, Description Logic, Fuzzy Logic, Ontology Learning, Information Retrieval, and Pattern Recognition. Table 1 shows the formal context \( F = (O, A, R) \) corresponding to the above two sets \( O \) and \( A \), the concept lattice \( L(O, A, R) \) is shown in Table 2, and 8 typical concepts, which are same as [28], are extracted from Table 2 to form Table 3. For the convenience of observation and comparison, we do not
TABLE 1. Formal context $F = (O, A, R)$.

|    | a1 | a2 | a3 | a4 | a5 | a6 | a7 | a8 | a9 |
|----|----|----|----|----|----|----|----|----|----|
| o1 | x  | x  |    |    | x  |    |    |    | x  |
| o2 | x  |    | x  | x  |    |    |    | x  |    |
| o3 | x  | x  |    |    | x  |    |    | x  |    |
| o4 |    | x  |    |    | x  |    |    |    |    |
| o5 | x  | x  |    |    | x  | x  |    |    |    |
| o6 |    |    | x  | x  |    |    | x  |    |    |
| o7 | x  | x  | x  |    | x  |    |    |    |    |
| o8 | x  |    | x  |    | x  |    |    |    |    |
| o9 |    |    | x  | x  |    | x  |    |    |    |
| o10| x  |    |    | x  |    |    | x  |    | x  |

TABLE 2. Lattice $L(F)$ corresponding to $F$ of Table 1.

| Concept | Extent | Intent |
|---------|--------|--------|
| top     | o1, o2, ..., o10 | $\emptyset$ |
| c1      | o3, o6, o8, o10 | a9     |
| c2      | o2, o3, o5, o8 | a2     |
| c3      | o1, o3, o5, o6, o10 | a3     |
| c4      | o1, o4, o5, o8, o10 | a1     |
| c5      | o1, o2, o5, o8 | a5     |
| c6      | o2, o4, o5, o7, o8, o9 | a7     |
| c7      | o3, o6, o10 | a3, a9 |
| c8      | o1, o3, o6, o7 | a6     |
| c9      | o1, o6, o7, o8 | a8     |
| c10     | o1, o5, o10 | a1, a3 |
| c11     | o2, o5, o8 | a2, a7 |
| c12     | o4, o5, o8 | a1, a7 |
| c13     | o2, o5, o9 | a5, a7 |
| c14     | o2, o4, o9, o10 | a4  |
| c15     | o3, o6, o10 | a3, a6, a9 |
| c16     | o6, o8 | a8, a9 |
| c17     | o3, o8 | a2, a9 |
| c18     | o6, o10 | a1, a9 |
| c19     | o6, o7 | a6, a8 |
| c20     | o1, o6 | a3, a8 |
| c21     | o3, o5 | a2, a3 |
| c22     | o1, o8 | a1, a8 |
| c23     | o7, o8 | a7, a8 |
| c24     | o5, o8 | a1, a2, a7 |
| c25     | o7, o9 | a6, a7 |
| c26     | o1, o5 | a1, a3, a5 |
| c27     | o2, o5 | a2, a5, a7 |
| c28     | o4, o10 | a1, a4 |
| c29     | o2, o4, o7 | a4, a7 |
| c30     | o3 | a2, a3, a6, a9 |
| c31     | o6 | a3, a6, a8, a9 |
| c32     | o8 | a1, a2, a7, a8, a9 |
| c33     | o10 | a1, a3, a4, a9 |
| c34     | o1 | a1, a3, a5, a8 |
| c35     | o7 | a6, a7, a8 |
| c36     | o5 | a1, a2, a3, a5, a7 |
| c37     | o9 | a4, a6, a7 |
| c38     | o2 | a2, a4, a5, a7 |
| c39     | o4 | a1, a4, a7 |

present the lattice in the form of a Hasse Diagram, but in a table. We choose the local similarity in [28] (denoted as Sl), the matrix method in [17] (denoted as Sz), the method in [19] (denoted as Sf) and the method in [26] as the benchmarks. Paper [26] proposes several methods, among which two best performing methods are selected: CassJac and CassSor (denoted as CASJac and CASSor in Table 1 respectively). Also, we have improved the approach in [21] by replacing the attribute weights in the original literature with the AF in this paper, and the improved approach (denoted as Sw+) is also a benchmark. The technique in [23] is not selected due to the weakness of knowledge-based methods, namely the VD problem. We evaluate the performances of all compared methods by computing the correlation scores between the results of them and those of human judgment. Pearson Correlation Coefficient and Spearman’s Rank Correlation Coefficient are usually used to measure the correlation of two sets of data from a different perspective. The former assesses the linear correlation between them, and the latter does the monotonic one. Therefore, to reflect the correlation of the two sets comprehensively, we take theAverage of the two Coefficients (denoted as ASP) as another evaluation metric.

The similarity values of the 8 typical concepts from various techniques and human judgment (HJ) are listed in Table 4, in which HJ and the methods in [26] all directly adopt the results given in [26]. In section III.C, we briefly discussed the setting of weight factors of S&D. Considering the data size and the formal context of this experiment, we roughly selected the weight of the semantic component of S&D to be 0.2. Therefore, in this experiment, all variants of our method ran with the weight factors $w_1 = w_2 = 0.4$. The above three correlation coefficient values of the different methods are shown in Table 5, in which CAJac-a (or CASor-a) represents the average of the upper limit and the lower limit of CASJac (or CASSor) method. For showing the relationship between the correlations of the methods clearly, the values in Table 5 are reshown as Fig. 1 using line charts.

As can be seen from Fig. 1-(a), Sz has the lowest correlation with HJ, and its result is significantly different from those of the others. Fig. 1-(b) is a partial enlargement of Fig. 1-(a), in which the gaps between the correlation coefficients of the methods other than Sz can be shown clearly. In this figure, both S&D-f (the first two terms of (7) or (9)) and Sf measure the similarity of concepts based on set theory while considering the importance of common members. And the former has better performance, which indicates that the AF-ICF-based method measuring the importance of members is superior to the methods merely considering the proportion...
TABLE 3. Typical concepts from Table 2.

| Concept | Name | Extent | Intent |
|---------|------|--------|--------|
| c4      | FC1  | o1,o4,o5,o8,o10 | a1     |
| c2      | FC2  | o2,o3,o5,o8     | a2     |
| c25     | FC3  | o1,o5           | a1,a3,a5|
| c21     | FC4  | o3,o5           | a2,a3  |
| c27     | FC5  | o2,o5           | a2,a5,a7|
| c36     | FC6  | o5              | a1,a2,a3,a5,a7|
| c7      | FC7  | o3,o6,o10       | a3,a9  |
| c10     | FC8  | o1,o5,o10       | a1,a3  |

TABLE 4. Between eight typical formal concepts by different methods.

| Concept pairs | HI   | SI   | Sz  | Sf  | Sw+  | CASJac | CASOr | S&D  |
|---------------|------|------|-----|-----|------|--------|-------|------|
| (FC1,FC2)     | 0.38 | 0.14 | 0.64| 0.13| 0.00 | 0.29,64| 0.37,72| 0.13 |
| (FC1,FC3)     | 0.80 | 0.37 | 0.67| 0.52| 0.66 | 0.70,70| 0.79,79| 0.56 |
| (FC1,FC4)     | 0.30 | 0.08 | 0.67| 0.10| 0.00 | 0.23,59| 0.29,65| 0.12 |
| (FC1,FC5)     | 0.20 | 0.08 | 0.62| 0.10| 0.00 | 0.23,59| 0.29,65| 0.09 |
| (FC1,FC6)     | 0.60 | 0.20 | 0.60| 0.36| 0.50 | 0.60,60| 0.67,67| 0.39 |
| (FC1,FC7)     | 0.40 | 0.07 | 0.67| 0.15| 0.00 | 0.28,57| 0.33,62| 0.16 |
| (FC1,FC8)     | 0.90 | 0.55 | 0.80| 0.72| 0.82 | 0.80,80| 0.88,88| 0.70 |
| (FC2,FC3)     | 0.20 | 0.10 | 0.65| 0.12| 0.00 | 0.25,60| 0.32,60| 0.17 |
| (FC2,FC4)     | 0.90 | 0.50 | 0.75| 0.74| 0.91 | 0.75,75| 0.83,83| 0.71 |
| (FC2,FC5)     | 0.85 | 0.42 | 0.75| 0.69| 0.79 | 0.75,75| 0.83,83| 0.64 |
| (FC2,FC6)     | 0.75 | 0.23 | 0.65| 0.44| 0.65 | 0.63,63| 0.71,71| 0.45 |
| (FC2,FC7)     | 0.28 | 0.08 | 0.67| 0.07| 0.00 | 0.28,59| 0.34,65| 0.09 |
| (FC2,FC8)     | 0.30 | 0.08 | 0.67| 0.09| 0.00 | 0.23,59| 0.29,65| 0.10 |
| (FC3,FC4)     | 0.50 | 0.29 | 0.75| 0.34| 0.22 | 0.49,67| 0.57,75| 0.30 |
| (FC3,FC5)     | 0.55 | 0.27 | 0.73| 0.48| 0.28 | 0.42,64| 0.50,72| 0.28 |
| (FC3,FC6)     | 0.73 | 0.55 | 0.80| 0.71| 0.77 | 0.75,75| 0.83,83| 0.62 |
| (FC3,FC7)     | 0.40 | 0.12 | 0.65| 0.16| 0.43 | 0.36,50| 0.36,50| 0.25 |
| (FC3,FC8)     | 0.95 | 0.67 | 0.89| 0.72| 0.82 | 0.84,84| 0.91,91| 0.73 |
| (FC4,FC5)     | 0.75 | 0.29 | 0.75| 0.42| 0.72 | 0.48,67| 0.56,75| 0.45 |
| (FC4,FC6)     | 0.80 | 0.45 | 0.70| 0.60| 0.70 | 0.75,75| 0.83,83| 0.54 |
| (FC4,FC7)     | 0.45 | 0.29 | 0.75| 0.31| 0.39 | 0.47,63| 0.55,71| 0.32 |
| (FC4,FC8)     | 0.55 | 0.29 | 0.75| 0.37| 0.27 | 0.45,63| 0.53,71| 0.28 |
| (FC5,FC6)     | 0.85 | 0.55 | 0.80| 0.64| 0.84 | 0.75,75| 0.83,83| 0.62 |
| (FC5,FC7)     | 0.15 | 0.00 | 0.56| 0.00| 0.00 | 0.16,50| 0.16,50| 0.03 |
| (FC5,FC8)     | 0.30 | 0.12 | 0.65| 0.13| 0.00 | 0.26,63| 0.34,71| 0.13 |
| (FC6,FC7)     | 0.35 | 0.08 | 0.54| 0.13| 0.26 | 0.36,50| 0.36,50| 0.16 |
| (FC6,FC8)     | 0.70 | 0.37 | 0.67| 0.51| 0.60 | 0.67,67| 0.75,75| 0.50 |
| (FC7,FC8)     | 0.45 | 0.27 | 0.73| 0.43| 0.52 | 0.46,60| 0.47,67| 0.44 |

words, in other words, the concept will have greater intent. Therefore, the semantic similarity between concepts must be unignorable, and maybe have a greater effect on the similarity of concepts. To a certain extent, the mini context is also the reason why S&D-t2 and Sw+ perform poorly. Both S&D-t2 and Sw+ work only on the attributes of two concepts, considering the importance of the attributes, and the larger the text context, the better the effect on the importance of attributes reflected by AF-ICF. In summary, the combination of members in a lattice. Fig. 1 also shows that the Pearson Coefficients of Sw+, S&D-t2 (the second term of (7) or (9)), CASor, CAJac, and S&D are higher than the Spearman’s Coefficients of them, indicating that these methods have a better linear correlation with HI than the monotonic correlation, whereas Sz and Sf are on the contrary. By comparing the performance of S&D-t2 and S&D, we can learn that the other two modules of S&D: the object module (the first term of (7) or (9)) and the intent semantic module (the third term of (7) or (9)), promote the overall performance of S&D, especially in terms of the monotonic correlation. This can also explain why the gap between the two coefficients of S&D is lower than the other methods except SI and Sf. On the other hand, the ASP value of all the benchmarks including SI and Sf is lower than that of S&D. All of these illustrate the effectiveness of concept similarity measure based on both the AF-ICF importance of members and the semantic relations between intents containing domain information. One problem is that our method works better when using only the first two terms of (7) or (9) (i.e., S&D-f), that is, the proposed intent semantic module has a negative effect on this experiment. So, is there the necessity to use the semantic similarity of intents to measure the similarity of concepts? It should be noted that although this experiment is based on the real-world data, a mini formal context with only 9 attribute words is used for the comparison with the benchmarks. However, most text contexts in the real-world are larger than that we used, and in this case, the concept will contain more attribute words, in other words, the concept will have greater intent.
of the AF-ICF-based member importance and the domain-information-contained semantic similarity of intents could reflect the difference and relation between two concepts in the specific text context well. For a larger domain text context, we believe and expect that, the presented method will find similar formal concepts more accurately than for a mini one.

VI. CONCLUSION AND FUTURE WORK

This paper proposes a method for measuring the similarity of formal concepts, which aims to measure the similar relation of the FCA concepts in text contexts of domain data, full-automatically, without any direct use of external corpora, ontologies or knowledge bases, without supervision. The method can be used as a component of any system using FCA for knowledge mining and knowledge discovery. The similarity relation of concepts obtained by this technique contains not only syntactic information in terms of text composition but also semantic information in terms of text meaning; which can reflect the semantic knowledge contained in the general knowledge base or other large-scale corpora and embody the domain-related semantic knowledge contained in the data itself. However, as many methods rely on certain knowledge bases (such as WordNet, Wikipedia or DBpedia), the precision of our method depends on that of the word embedding technique to some extent. As mentioned above, in the performance of lexical similarity, the word embedding techniques have been evidently better than other techniques. More optimistically, the word vector used in this paper, to some extent, performs better than other word vectors [40] (http://conceptnet.io/). Besides, this paper compares the proposed method with other related works in a relatively overall way. Experimental results prove the advantages of this method. The future work is to employ the method for knowledge mining in large-scale domain text data.

REFERENCES

[1] R. Wille, “Restructuring lattice theory: An approach based on hierarchies of concepts,” in Proc. ICFLCA. Berlin, Germany: Springer, 2009, pp. 314–339.
[2] B. Ganter and R. Wille, “Concept lattices of contexts,” in Formal Concept Analysis: Mathematical Foundations. Berlin, Germany: Springer, 1999, pp. 17–57.
[3] D. I. Ignatov, “Introduction to formal concept analysis and its applications in information retrieval and related fields,” in Russian Summer School in Information Retrieval, Cham, Switzerland: Springer, 2014, pp. 42–141.
[4] P. K. Singh, C. Aswani Kumar, and A. Gani, “A comprehensive survey on formal concept analysis, its research trends and applications,” Int. J. Appl. Math. Comput. Sci., vol. 26, no. 2, pp. 495–516, Jun. 2016.
[5] J. Poelmans, D. I. Ignatov, S. O. Kuznetsov, and G. Dedeine, “Formal concept analysis in knowledge processing: A survey on applications,” Expert Syst. Appl., vol. 40, no. 16, pp. 6538–6560, Nov. 2013.
[6] S. Andrews, “A ‘Best-of-Breed’ approach for designing a fast algorithm for computing fixpoints of Galois Connections,” Inf. Sci., vol. 295, pp. 633–649, Feb. 2015.
[7] S. M. Neto, L. E. Zárate, and M. A. J. Song, “Handling high dimensionality contexts in formal concept analysis via binary decision diagrams,” Inf. Sci., vol. 429, pp. 361–376, Mar. 2018.
[8] Y. Wan and L. Zou, “An efficient algorithm for decreasing the granularity levels of attributes in formal concept analysis,” IEEE Access, vol. 7, pp. 11029–11040, 2019.
[9] S. Andrews, H. Gibson, K. Domonkosz, and A. Akhtar, “Creating corroborated crisis reports from social media data through formal concept analysis,” J. Intell. Inf. Syst., vol. 47, no. 2, pp. 287–312, Oct. 2016.
[10] F. Fkhri and M. N. Omri, “IRAFC: An O(n) information retrieval algorithm based on formal concept analysis,” Knowl. Inf. Syst., vol. 48, no. 2, pp. 465–491, Aug. 2016.
[11] E. Negm, S. Abdellrahman, and R. Bahgat, “PREPCA: A portal retrieval engine based on formal concept analysis,” Inf. Process. Manage., vol. 53, no. 1, pp. 203–222, Jan. 2017.
[12] A. Formica, “Similarity reasoning in formal concept analysis: From one-to many-valued contexts,” Knowl. Inf. Syst., vol. 60, no. 2, pp. 715–739, Aug. 2019.
[13] Y. Jiang and M. Yang, “Semantic search exploiting formal concept analysis on rough sets and Wikipedia,” Int. J. Semantic Web Inf. Syst., vol. 14, no. 3, pp. 99–119, Jul. 2018.
[14] P. Hutka, T. Low, M. Kotzyba, S. Haun, and A. Nürnberger, “Exploration of Web search results based on the formal concept analysis,” in Semantic Keyword-Based Search on Structured Data Sources. Cham, Switzerland: Springer, 2017, pp. 123–135.
[15] A. Hristoskova, V. Boeva, and E. Tsiporkova, “A formal concept analysis approach to consensus clustering of multi-experiment expression data,” BMC Bioinf., vol. 15, no. 1, p. 151, Dec. 2014.
[16] L. Wang and X. Liu, “A new model of evaluating concept similarity,” Knowl.-Based Syst., vol. 21, no. 8, pp. 842–846, Dec. 2008.
[17] F. Alqadah and R. Batnagar, “Similarity measures in formal concept analysis,” Ann. Math. Artif. Intell., vol. 61, no. 3, pp. 245–256, Mar. 2011.
[18] Y. Du and Y. Hai, “Semantic ranking of Web pages based on formal concept analysis,” J. Softw., vol. 86, no. 1, pp. 187–197, Jan. 2013.
[19] I. Tadrat, V. Boonjing, and P. Pattaraintakorn, “A new similarity measure in formal concept analysis for case-based reasoning,” Expert Syst. Appl., vol. 39, no. 1, pp. 967–972, Jan. 2012.
[20] J. Muangprathub, V. Boonjing, and P. Pattaraintakorn, “A new case-based classification using incremental concept lattice knowledge,” Data Knowl. Eng., vol. 83, pp. 39–53, Jan. 2013.
[21] A. Hristoskova, M. Soljačić, and Somayeh Alizadeh, “A new similarity measure by combining formal concept analysis and clustering for case-based reasoning,” in Proc. ICAIA/AIE. Cham, Switzerland: Springer, 2015, pp. 503–513.
[22] A. Formica, “Ontology-based concept similarity in formal concept analysis,” Inf. Sci., vol. 176, no. 18, pp. 2624–2641, Sep. 2006.
[23] A. Formica, “Concept similarity in formal concept analysis: An information content approach,” Knowl.-Based Syst., vol. 21, no. 1, pp. 80–87, Feb. 2008.
[24] A. Formica, “Semantic Web search based on rough sets and fuzzy formal concept analysis,” Knowl.-Based Syst., vol. 26, pp. 40–47, Feb. 2012.
[25] D. Lin, “An information-theoretic definition of similarity,” in ICML, Madison, WI, USA, 1998, pp. 296–304.
[26] Y. Jiang, M. Yang, and R. Qu, “Semantic similarity measures for formal concept analysis using linked data and WordNet,” Multimedia Tools Appl., vol. 78, no. 14, pp. 19807–19837, Jul. 2019.
[27] S. Jimenez, F. A. Gonzalez, A. Gelbukh, and G. Duenas, “word2vec: WordNet-based word representation rivaling neural word embedding for lexical similarity and sentiment analysis,” IEEE Comput. Intell. Mag., vol. 14, no. 2, pp. 41–53, May 2019.
[28] K. Lengnink, “Ähnlichkeit als Distanz in Begriffsverbänden,” in Begriffliche Wissensverarbeitung. Berlin, Germany: Springer, 2000, pp. 57–71.
[29] D. Dau, Frithjof, Jon Ducrou, and Peter Ekland, “Concept similarity and related categories in searchbubble,” in Proc. IJCCS. Berlin, Germany: Springer, 2008, pp. 255–259.
[30] P. Eklund, J. Ducrou, and F. Dau, “Concept similarity and related categories in information retrieval using formal concept analysis,” Int. J. Gen. Syst., vol. 41, no. 8, pp. 826–846, Nov. 2012.
[31] A. Tversky, “Features of similarity,” Psychol. Rev., vol. 84, no. 4, pp. 327–352, 1977.
[32] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” 2013, arXiv:1301.3781. [Online]. Available: http://arxiv.org/abs/1301.3781
[33] Y. Y. Lee, H. Ke, T. Y. Yen, H. H. Huang, and H.-H. Chen, “Combining and learning word embedding with WordNet for semantic relatedness and similarity measurement,” J. Assoc. Inf. Sci. Technol., to be published, doi: 10.1002/asi.24289.
[34] J. Camacho-Collados and M. T. Pilehvar, “From word to sense embeddings: A survey on vector representations of meaning,” J. Artif. Intell. Res., vol. 65, pp. 743–788, Dec. 2018.
F. Wang et al.: Similarity Measure in FCA Containing General Semantic Information and Domain Information

[35] R.Navigli and F. Martelli, “An overview of word and sense similarity,” Natural Lang. Eng., vol. 25, no. 6, pp. 693–714, Nov. 2019.

[36] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” 2018, arXiv:1810.04805. [Online]. Available: http://arxiv.org/abs/1810.04805

[37] M. Taher Pilehvar and J. Camacho-Collados, “WiC: The Word-in-Context dataset for evaluating context-sensitive meaning representations,” 2018, arXiv:1808.09121. [Online]. Available: http://arxiv.org/abs/1808.09121

[38] J. Pennington, R. Socher, and C. Manning, “Glove: Global vectors for word representation,” in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), Doha, Qatar, 2014, pp. 1532–1543.

[39] R. Speer, J. Chin, and C. Havasi, “Conceptnet 5.5: An open multilingual graph of general knowledge,” in Proc. AAAI, San Francisco, CA, USA, 2017, pp. 503–513.

[40] R. Speer and J. Lowry-Duda, “ConceptNet at SemEval-2017 task 2: Extending word embeddings with multilingual relational knowledge,” 2017, arXiv:1704.03560. [Online]. Available: http://arxiv.org/abs/1704.03560

[41] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of tricks for efficient text classification,” 2016, arXiv:1607.01759. [Online]. Available: http://arxiv.org/abs/1607.01759

[42] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, “Advances in pre-training distributed word representations,” 2017, arXiv:1712.09405. [Online]. Available: http://arxiv.org/abs/1712.09405

[43] Z. Galil, “Efficient algorithms for finding maximum matching in graphs,” ACM Comput. Surveys (CSUR), vol. 18, no. 1, pp. 23–38, Mar. 1986.

[44] A. Formica, M. Missikoff, E. Pourabbas, and F. Taglino, “Semantic search for matching user requests with profiled enterprises,” Comput. Ind., vol. 64, no. 3, pp. 191–202, Apr. 2013.

FUGANG WANG received the M.E. degree in measuring and testing technologies and instruments from the Chengdu University of Technology, in 2007. He is currently pursuing the Ph.D. degree in computer science and technology with Harbin Engineering University. He is also an Associate Professor with the Department of Electronic Information Science and Technology, Mudanjiang Normal University. His research interests include data mining, data analysis, and machine learning.

NIANBIN WANG received the Ph.D. degree in computer science and technology from the Harbin Institute of Technology, Harbin, China, in 2001. He is a member of CCF. He has been a Professor with the Department of Computer Science and Technology, Harbin Engineering University. His research interests include dataspaces, deep learning, and data integration.

SHAOBIN CAI received the Ph.D. degree in computer system architecture from the Harbin Institute of Technology, in 2005. He is currently a Minjiang Scholar of Fujian Province. He is also a Professor of computer science and technology with the Foshan University. He is a member of the Wireless Sensor Network Committee of the Chinese Computer Society. His research interests include ad hoc networks, wireless sensor networks, underwater acoustic sensor networks, and the blockchain technology.

WULIN ZHANG received the B.E. and M.E. degrees from the Harbin University of Science and Technology, in 2006 and 2009, respectively, and the Ph.D. degree in signal and information processing from Harbin Engineering University, in 2015. He is currently a Research Scholar with Mudanjiang Normal University. His research interests include computer vision, image processing, and deep learning.

* * *