International Journal of Innovative Computing, Information and Control
Volume **9**, Number **3**, March **2013**

ICIC International ©2013 ISSN 1349-4198
pp. 1189–1200

A NEW APPROACH OF WORD SENSE DISAMBIGUATION AND KNOWLEDGE DISCOVERY OF ENGLISH MODAL VERBS BY FORMAL CONCEPT ANALYSIS

Jianping Yu[1,*], Wenxue Hong[2], Shaoxiong Li[2], Tao Zhang[3] and Jialin Song[2]

[1]College of Foreign Studies
[2]Institute of Electrical Engineering
[3]College of Information Engineering
Yanshan University
No. 438, Hebei Ave., Qinhuangdao 066004, P. R. China
*Corresponding author: yjp@ysu.edu.cn

Received January 2012; revised May 2012

Abstract. *A new approach is proposed in this paper for improving the accuracy of word sense disambiguation (WSD) and knowledge discovery of English modal verbs. The concept hierarchical relation diagrams of some English modal verbs are generated based on the mathematical descriptions of the optimization of formal context. The diagrams can not only be used to disambiguate word senses, but also show clearly the complex relations between the objects (senses) and the attributes (contextual features) and the relations among the attributes of modal verbs. The proposed approach is verified by the experiments with the data of some English modal verbs. The results of WSD by the new approach are compared with those by other existing methods. It is proven that the new approach can not only improve the accuracy of WSD of English modal verbs, but also display the hierarchical relations of the complex data for knowledge discovery. The discovered knowledge provides the basis for feature selection in natural language processing and for the semantic study of English modal verbs.*
Keywords: English modal verb, Formal concept analysis, Word sense disambiguation, Knowledge discovery

1. **Introduction.** It has been a challenging issue to improve the accuracy of word sense disambiguation (WSD) and the quality of feature selection in natural language processing. Since the knowledge behind the target words may provide the basis for feature selection, it is significant to display hierarchical relations of the complex data in WSD. Some methods have been applied to the WSD in natural language processing, such as BP neural network [1-3], adaptive network-based fuzzy inference system [4] and fuzzy c-means clustering [5]. However, there is still room for further improvement in the accuracy of WSD, and the knowledge behind the target words can not be discovered by the existing methods because they are designed just as classifiers, and they can not visualize the hierarchical relations of the complex data, which are necessary for knowledge discovery.

English modal verbs are a complex semantic system. The semantic studies of modal verbs have mainly focused on the sense categorization and the description of semantic and syntactic features and functions. With the advances in the studies of WSD and lexical semantics, more knowledge about the deep structures and relations of the features of modal verbs is needed, such as the relations between word senses and contextual features, the hierarchical orders of the features and the regularity of clusters of the features. Experts in natural language processing and linguistics need to know how the contextual features interact with each other and influence the sense of a modal verb, which features are the extents and intents of a modal verb, what relations exist between the contextual features and which features influence the senses of a modal word greater than others, etc., because this knowledge may provide important and valuable basis for the proper feature selection for WSD and the correct understanding of the senses of the modal verbs. However, the knowledge is unobtainable by the existing methods of WSD, and it is not easy to further improve the accuracy of WSD by them. Therefore, the theory and approach of formal concept analysis (FCA) are used to visualize the hierarchical relationships of the objects (senses) and the attributes (contextual features) of English modal verbs, to improve the accuracy of WSD, to discover the knowledge behind the English modal verbs and to provide basis for the proper feature selection in natural language processing.

The theory of FCA proposed by Wille [6] is based on the philosophical understanding of a concept. The idea is that a concept is composed of extent and intent. The extent is a set of all the objects of a concept, and the intent is a set of common attributes of all the objects. All the concepts and their relationships in generalization and specification may constitute a lattice which may be visualized by a Hasse diagram, which is used for displaying the complex network relationships between the objects and the attributes.

FCA has been studied extensively and used widely in machine learning, software engineering, information searching, medical diagnosis, decision making in management. Baixeries et al. [7] presented a faster algorithm for building the Hasse diagram of a concept lattice. Yang et al. [8] examined the problems of rule acquisition and attribute reduction of real decision formal contexts, and proposed an approach of rule acquisition and attribute reduction in real decision formal contexts. Moriki et al. [9] proposed a document clustering algorithm based on formal concept analysis for text classification and summarization. Andrews et al. [10] presented a way for discovering knowledge through creating formal contexts. Yoshinaga et al. [11] employed formal concept analysis to classify web pages and visualize the lattice structure of web pages and keywords as line diagram. Wang et al. [12] proposed a new algebra system for the formal context in order to derive the mathematical properties of formal concepts. The derived properties can be applied to explore concept hierarchy and ontology merging. Pelaez-Moreno et al. [13] employed FCA to provide a conceptual interpretation of confusion matrices which enables the analysis of the structure of confusions for both human and machine performances, and analyzed phonetic confusions. Poelmans et al. [14] made a survey of studies in applying formal concept analysis to knowledge discovery based on 702 collected related papers published during 2003-2009, and gave an extensive overview of the studies. Martin et al. [15] applied fuzzy FCA to creating simple taxonomies, which can be used to structure data and a novel form of fuzzy association rule for extracting simple knowledge from data organized hierarchically. Liu et al. [16] proposed a visualization approach for differential diagnosis in traditional Chinese medicine based on FCA. Zhang et al. [17] proposed a dynamic optimization algorithm of ontology based on FCA in order to make the relations between the concepts clearer and the critical information more prominent.

It is known from the previous studies that FCA as a mathematic device of describing concept and conceptual hierarchy can be used for the conceptual unfolding of the data in contexts. Its applications in different fields imply the feasibility of applying it to the WSD and the knowledge discovery of English modal verbs. However, no such studies have been found from the previous literature. Therefore, this paper focuses on the generation of the hierarchical relation diagram for improving the accuracy of WSD and the knowledge discovery of English modal verbs.

2. **Approach of Generation of Hierarchical Relation Diagram.** A new approach is proposed for the generation of the hierarchical relation diagram. The formal context is optimized in order to obtain a clearer hierarchy of attributes and reduce the unnecessary attributes. A theorem and two definitions for the optimization of the formal context are explained as follows:

Theorem 2.1. *$K = (O, A, R)$ is a formal context. O is a set of objects, A is a set of attributes, R is the binary relation between O and A. The objects in O are in the order of $\{o_1, o_2, o_3, \ldots, o_i, \ldots, o_m\}$, and the attributes in A are in the order of $\{a_1, a_2, a_3, \ldots, a_j, \ldots, a_n\}$, a_{ij} is the value of an attribute. If an object o_i has the attribute a_j, then $a_{ij} = 1$, if not, $a_{ij} = 0$. The sufficient and necessary condition for transforming K into a hierarchical formal context $K_{y0} = (O, A, R)$ is that O is in the order of $\{o_{1'}, o_{2'}, o_{3'}, \ldots, o_{i'}, \ldots, o_{m'}\}$ and A is in the order $\{a_{1'}, a_{2'}, a_{3'}, \ldots, a_{j'}, \ldots, a_{n'}\}$. The new orders of objects O and attributes A can be determined by the following method:*

1) *Solve $\mathrm{MS}a_{mj} = \mathrm{Max}(\sum a_{i1'}, \sum a_{i2'}, \sum a_{i3'}, \ldots, \sum a_{in'})$ $(i = 1, 2, 3, \ldots, m)$. $\mathrm{MS}a_{mj}$ means that for object o_m, the jth column has the maximum sum of attributes.*
2) *Exchange the first column with the jth column, a new order of attributes $\{a_1, a_2, a_3, \ldots, a_j, \ldots, a_n\}$ is obtained.*
3) *Exchange the rows by ranking the values of attributes $a_{jl} = 1$ from a_{11} successively, a new order of objects $\{o_{1'}, o_{2'}, o_{3'}, \ldots, o_{i'}, \ldots, o_{m'}\}$ is obtained.*

Definition 2.1. *$K = (O, A, R)$ is a formal context, and $H \in O$, $N \in A$, $(H, N, R \cap H \times N)$ is defined as a subcontext of (O, A, R).*

Definition 2.2. *Given $K_1 = (O_1, A_1, R_1)$ and $K_2 = (O_2, A_2, R_2)$ are two sub-contexts, the disjoint union of K_1 and K_2 is $K = (O_1 \cup O_2, A_1 \cup A_2, R_1 \cup R_2)$.*

$K_y = (O_y, A_y, R_y)$ can be divided into two sub-contexts: $K_1 = (O_1, A_1, R_1)$ and $K_2 = (O_2, A_2, R_2)$ according to the above two definitions. K_1 is the part of context corresponding to $a_{i1} = 1$ of $a_{1'}$. K_2 is the part of context corresponding to $a_{i1} = 0$ of $a_{1'}$. The two sub-contexts are transformed as follows:

For K_2, transform the column with the maximum sum of attributes to the second column and rank $a_{i2} = 1$ from a'_{12} successively (a'_{12} is the value of attribute of the first row and the second column of K_2) according to the above Theorem 2.1. Note that when transforming a column, the corresponding column in K_1 transforms accordingly. Then, transform the rows in K_1 by ranking the values of attributes $a_{jl} = 1$ from a_{11} successively. Now, a new formal context K_{y1} is obtained with $\{o_{1''}, o_{2''}, o_{3''}, \ldots, o_{i''}, \ldots, o_{m''}\}$ as its object order and $\{a_{1''}, a_{2''}, a_{3''}, \ldots, a_{j''}, \ldots, a_{n''}\}$ as its attribute order. Note that the first attributes of $K_{y0} = (O, A, R)$ and $K_{y1} = (O, A, R)$ are the same.

The transformation will be repeated until when the sum of attributes reaches the minimum.

From the viewpoint of mathematics, the hierarchy process is essentially to divide A into subsets according the generality of the attribute set. The purpose is to include the extent of the subsets with low generality in the union of the extent of the subset with high generality. A mathematical description of the process is as following: Given an attribute set $A = \{a_1, a_2, \ldots, a_m\}$ of a formal context, a_i denotes the ith attribute. The degree of attribute a_i is defined as Degree $(a_i) = \|a'_i\|_0$, which denotes the degree of generality of the current attribute. The greater the Degree (a_i) is, the greater generality the attribute has in the current formal context. On the contrary, the less the Degree (a_i) is, the more distinctive the attribute is. Since Degree (a_i) is calculated without the consideration of the inclusive relation between the sets, some modification is needed.

Given a set of Degree (a_i), $D = \{\|a_i'\|_0,\ i = 1, 2, \ldots, m\} = \{0, 1, 2, \ldots, d | d \in N\}$, according to set theory, there must be $d \leq$ Degree (A), and $d \leq$ Degree (O). There exists a set of objects with the jth attribute, $D_j = \{a_i' | \|a_i'\|_0 = j,\ i = 1, 2, \ldots, m\}$ and a set of attributes with the jth attribute, $MD_j = \{a_i | \|a_i'\|_0 = j,\ i = 1, 2, \ldots, m\}$. Compare the degrees of two adjacent sets D_j and D_{j-1} $(j > 0)$, do this for all $a_i \subseteq MD_{j-1}$. If $a_i'|_{A_i \subseteq MD_j} \subseteq (\cup a_i')$, it means that, for the current $a_i \subseteq MD_{j-1}$, the union of all the elements in D_j includes the union of all the elements in D_{j-1}, and no modification is needed. Otherwise, let $\|a_i'\|_0 = j$, renew D_j, MD_j, and repeat the modification process until $\|a_i'\|_0 = d$.

The theorem and definitions are used in the optimization of the formal context.

3. **Experiments of WSD of English Modal Verb *May* by the Approach.** In this study, *may* is chosen as the target word for WSD because it is one of the most frequently used modal verbs. The meanings of *may* are categorized as follows [18]:

$$may \begin{cases} \text{epistemic meaning} - \text{epistemic possibility} \\ \text{root meaning} - \text{root permission or root possibility} \end{cases}$$

For instance,

(1) It *may* rain today. (epistemic *may*)

(2) I *may* be a few minutes late, but I don't know. (epistemic *may*)

(3) *May* I use your phone? (root *may*)

(4) Visitors *may* use the swimming pool between 5:00 and 7:00 pm. (root *may*)

In order to disambiguate epistemic *may* from root *may* by formal concept analysis and make a comparison of the result with the one by BP neural network, the data in [2] were used in this study, as shown in Table 1 and Table 2.

The first 4 data in Table 1 and Table 2 are the values of MI (mutual information [19]) and they are all continuous values. Since only binary data are allowed in the formal context, the continuous value data need to be transformed into binary data. First, the 4 data were titled Attributes A, B, C, D and E, respectively, and then a linear equal

TABLE 1. Data for training set

No.	Vectors	No.	Vectors	No.	Vectors
1	1.74, 1.89, 0.3, 1.57, 0	18	1.32, 1.52, 1.47, 0.38, 0	35	0, 0.67, 1.19, 1.72, 0
2	1.17, 1.68, 0, 0, 0	19	1.21, 1.67, 0, 0, 0	36	0.57, 0, 1.02, 2.13, 0
3	1.74, 1.92, 0.3, 0, 0	20	0.81, 1.92, 0, 0, 0	37	0, 2.29, 1.96, 2.54, 1
4	2.22, 1.64, 0, 0, 0	21	0.8, 2, 0, 0, 0	38	0.34, 0, 0.49, 1.24, 1
5	0.29, 1.1, 0, 0, 0	22	1.07, 1.16, 1.04, 0, 0	39	2, 0, 2.45, 1.83, 0
6	0.97, 0.91, 0, 0, 1	23	1.08, 1.78, 0, 1.74, 0	40	0.58, 0, 0.78, 1.25, 0
7	1.75, 0.92, 0, 0.9, 0	24	1.05, 1.07, 1.02, 1.30, 0	41	0, 0, 1.35, 1.97, 1
8	0.79, 2.22, 0, 0, 1	25	1.23, 1.98, 0, 0, 0	42	0, 0.4, 1.06, 0.85, 1
9	0.53, 1.42, 0, 0, 0	26	0, 0, 2.27, 1.96, 1	43	1.32, 0, 1.47, 1.3, 1
10	1.78, 1.11, 0, 0, 0	27	0.46, 0.93, 1.2, 1.38, 1	44	0.69, 0, 1.14, 1.46, 0
11	1.74, 1.45, 0, 0, 0	28	0.38, 0, 0.83, 1.26, 1	45	0, 0, 1.14, 2.02, 1
12	0.8, 0.76, 0, 0, 0	29	0, 0, 1.67, 1.46, 0	46	0, 0, 1.24, 1.29, 1
13	1.28, 1.16, 0, 0, 0	30	0, 0, 1.57, 1.36, 1	47	0, 0, 2.3, 1.24, 1
14	0.63, 1.14, 0.77, 0, 0	31	0, 0, 1.36, 2.1, 0	48	0.43, 0, 1.36, 1.77, 1
15	0.9, 1.52, 0.99, 0, 0	32	1.2, 0, 1.3, 0.85, 1	49	0, 0, 1.69, 2.72, 1
16	2, 1.83, 0, 0, 0	33	0, 0, 2.6, 1.39, 1	50	0, 0, 2.02, 1.82, 1
17	0.36, 0.8, 0, 0, 0	34	0.51, 0, 1.04, 1.87, 1		

TABLE 2. Data for testing set

No.	Vectors	No.	Vectors	No.	Vectors
1	0.43, 1.76, 0, 0, 1	18	0.74, 1.85, 0.83, 1.94, 1	35	0, 0, 3.27, 2.07, 0
2	0, 0, 0.98, 2.06, 1	19	0.4, 0.74, 0, 0, 1	36	1.59, 1.83, 0, 0, 0
3	0.77, 0, 1.82, 2.57, 1	20	0, 0, 1.02, 1.67, 1	37	1.51, 1.37, 0, 1.54, 0
4	0.13, 0, 0.87, 2.18, 1	21	3, 1.79, 0, 0, 0	38	2.05, 1.55, 0, 0, 0
5	0.77, 1.6, 1.82, 0, 0	22	1.11, 1.97, 0, 0, 0	39	0.95, 2.19, 1.52, 0, 0
6	1.21, 1.71, 0.99, 0, 0	23	0.56, 1.84, 0, 0, 0	40	1.83, 0.72, 0, 0.99, 0
7	1.35, 2.49, 0, 0, 0	24	1.11, 2.05, 0, 0, 0	41	0.94, 1.62, 0, 0, 0
8	0.12, 1.53, 0, 0, 0	25	0.81, 1.79, 0, 0, 0	42	0.87, 0.98, 0, 0, 0
9	1.1, 1.06, 0, 0, 0	26	0, 0, 1.04, 1.86, 1	43	2.10, 0, 2.37, 2.12, 0
10	0.43, 0.91, 1, 1.01, 0	27	0, 0.67, 2.69, 0.76, 0	44	2.32, 1, 0, 1.10, 1
11	0.89, 2.23, 0, 0, 0	28	0, 0.67, 2.69, 0.76, 0	45	1.01, 1.51, 0, 0, 0
12	1.17, 1.88, 0, 2.08, 0	29	0, 0.39, 1.18, 0.96, 1	46	2.30, 0.93, 0, 0, 1
13	1.48, 1.39, 0, 0, 0	30	1.01, 1.72, 0, 0, 0	47	1.14, 0.81, 0.76, 0, 1
14	0.57, 2.53, 1.14, 0, 0	31	1.08, 1.62, 0, 0, 0	48	0.88, 0.99, 0, 0, 0
15	0, 0, 2.34, 1.91, 1	32	1.52, 1.45, 0, 0, 0	49	0.74, 2.7, 0.83, 0, 0
16	1.54, 1.55, 0, 1.82, 0	33	0.88, 1.88, 0, 0, 0	50	1.05, 1.12, 0, 1.69, 0
17	0.77, 1.72, 0, 0, 0	34	2.23, 1.64, 0, 0, 0		

TABLE 3. Division and symbols of the attributes

Attribute A		Attribute B		Attribute C		Attribute D		Attribute E	
a_1	$0 \leq A \leq 0.5$	b_1	$0 \leq B \leq 0.5$	c_1	$0 \leq C \leq 0.5$	d_1	$0 \leq D \leq 0.5$	e_1	$E = 1$
a_2	$0.5 < A \leq 1$	b_2	$0.5 < B \leq 1$	c_2	$0.5 < C \leq 1$	d_2	$0.5 < D \leq 1$	e_0	$E = 0$
a_3	$1 < A \leq 1.5$	b_3	$1 < B \leq 1.5$	c_3	$1 < C \leq 1.5$	d_3	$1 < D \leq 1.5$		
a_4	$1.5 < A \leq 2$	b_4	$1.5 < B \leq 2$	c_4	$1.5 < C \leq 2$	d_4	$1.5 < D \leq 2$		
a_5	$2 < A \leq 2.5$	b_5	$2 < B \leq 2.5$	c_5	$2 < C \leq 2.5$	d_5	$2 < D \leq 2.5$		
a_6	$2.5 < A \leq 3$	b_6	$2.5 < B \leq 3$	c_6	$2.5 < C \leq 3$	d_6	$2.5 < D \leq 3$		
				c_7	$3 < C \leq 3.5$				

dividing method was used to divide the continuous values of the data into ranges and then transform them into binary values, i.e., the symbolization of the data. Each range was set to be 0.5. Finally, the divisions and symbols of the data were obtained, as shown in Table 3.

The symbolized data together with other binary data formed a formal context, as shown in Table 4. After that the formal context was optimized according to the theorem and the definitions in Section 2, and an optimized formal context of *may* was obtained, as shown in Table 5.

Finally, the corresponding concept hierarchical relation diagram was generated by a SPOAD (structural partial-ordered attribute diagram) tool, see Figure 1. The subscripts of the attributes were changed into regular size numbers in Figure 1 and Figure 4 considering the size of the generated diagram and the visibility of the symbols for the attributes.

The test of the model was based on similarity which is described as following:

Object O_1 can be described by the features of $X_1 = \{x_1, x_2, x_3, \ldots, x_i\}$, and object O_2 has and can be described by the features of $X_2 = \{x_1, x_2, x_3, \ldots, x_m\}$. Suppose X_1 is a sample in the training set and it represents a model, the similarity of X_1 and X_2 is represented by S:

$$S = \frac{|X_1 \cap X_2|}{\max(|X_1|, |X_2|)} \tag{1}$$

If S is 1, then X_2 and X_1 have the same description and are in the same class. If $S \neq 1$, and S is the maximum of a group of S_i, then X_2 is most approximate to X_1, and they are classified into one class. If S of X_2 is the same as S_i of several other samples,

Table 4. Original formal context of *may*

	e_1	a_1	a_2	a_3	a_4	a_5	b_1	b_2	b_3	b_4	b_5	c_1	c_2	c_3	c_4	c_5	c_6	d_1	d_2	d_3	d_4	d_5	d_6
1					1					1		1									1		
2				1						1		1						1					
3					1					1		1						1					
4						1				1		1						1					
5		1							1			1						1					
6	1		1					1				1						1					
7					1			1				1							1				
8	1		1								1	1						1					
9			1						1			1						1					
10					1				1			1						1					
11					1				1			1						1					
12			1					1				1						1					
13				1					1			1						1					
14			1						1				1					1					
15			1							1			1					1					
16					1					1		1						1					
17		1						1				1						1					
18				1						1				1				1					
19				1						1		1						1					
20			1							1		1						1					
21			1							1		1						1					
22				1					1					1				1					
23				1						1		1									1		
24				1					1					1						1			
25				1						1		1						1					
26	1	1					1									1					1		
27	1	1						1						1						1			
28	1	1					1						1							1			
29		1					1								1					1			
30	1	1					1								1					1			
31		1					1							1								1	
32	1			1			1							1					1				
33	1	1					1										1			1			
34	1		1				1							1							1		
35		1						1						1							1		
36			1				1							1								1	
37	1	1									1				1								1
38	1	1					1					1								1			
39					1		1									1					1		
40			1				1						1							1			
41	1	1					1							1							1		
42	1	1					1							1					1				
43	1			1			1							1						1			
44			1				1							1						1			
45	1	1					1							1								1	
46	1	1					1							1						1			
47	1	1					1									1				1			
48	1	1					1							1							1		
49	1	1					1								1								1
50	1	1					1									1					1		

Table 5. Optimized formal context of *may*

	b_6	b_5	c_5	e_1	a_1	c_1	a_2	b_3	c_7	a_3	d_1	b_2	a_4	b_1	b_6	c_2	c_3	c_6	d_2	b_4	d_3	a_5	c_4
1	1	1		1	1				1														
2	1			1	1	1			1														
3	1			1	1	1					1												
4	1			1	1	1					1												
5	1			1	1	1												1					
6	1			1	1	1													1				
7	1			1	1				1						1								
8	1			1	1				1							1							
9	1			1	1				1								1						
10	1			1	1				1														1
11	1			1	1						1						1						
12	1			1	1						1						1						
13	1			1	1											1					1		
14	1			1		1	1				1												
15	1			1		1			1	1													
16	1			1		1				1								1					
17	1				1	1													1				
18	1				1				1							1							
19	1					1	1		1														
20	1					1	1												1				
21	1						1		1						1								
22	1										1		1				1						
23		1	1	1			1							1									
24		1	1	1			1													1			
25		1	1		1							1											
26		1	1		1									1									
27		1	1				1	1															
28		1	1				1	1															
29		1	1				1					1											
30		1	1				1							1									
31		1	1					1		1													
32		1	1					1		1													
33		1	1					1		1													
34		1	1					1					1										
35		1	1					1					1										
36		1	1					1														1	
37		1	1							1		1											
38		1	1									1	1										
39		1	1									1	1										
40		1						1		1	1												
41		1						1			1		1										
42		1											1	1				1					
43			1			1		1		1													
44			1			1				1		1											
45			1				1	1							1								
46			1				1					1			1								
47				1	1	1			1					1									
48				1	1											1				1	1		
49					1	1					1			1									
50						1			1	1		1											

then X_2 is classified into the class which has the maximum members of samples having co-occurred features. In this way, the model can be tested. The whole process of WSD of *may* is shown in Figure 2. By comparing the similarities of the model of testing data with the generated model, the accuracy of WSD of *may* reached 88%. This accuracy is higher than that in [2] (78%), which implies 10% increase in accuracy of WSD of *may* by the new approach.

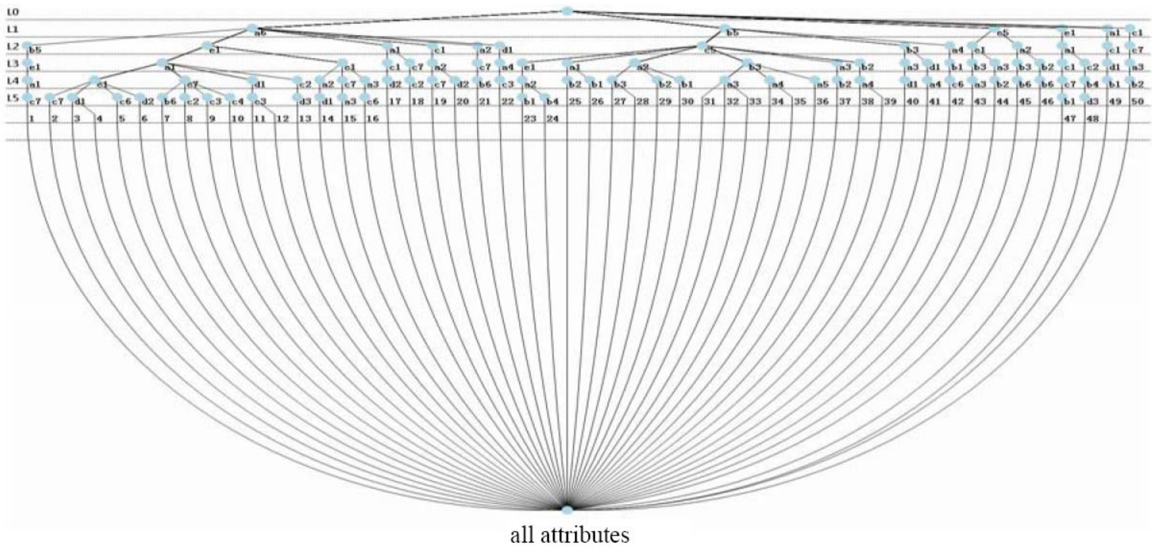

FIGURE 1. Generated concept lattice of *may*

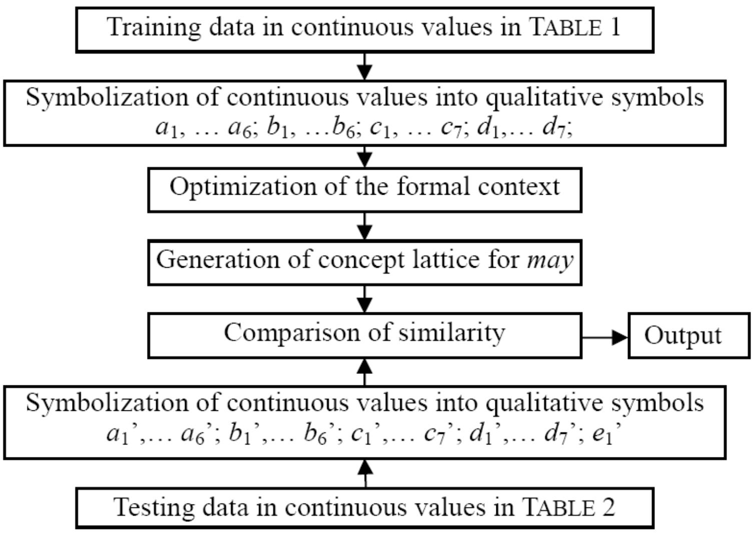

FIGURE 2. Process of WSD of *may*

TABLE 6. Comparison of accuracies of WSD of 4 English modal verbs

modal verbs	*may*	*must*	*can*	*will*
approach: rate	BPNN: 78%	BPNN: 84%	ANFIS: 76%	FCM: 90%
approach: rate	FCA: 88%	FCA: 96%	FCA: 80%	FCA: 99%
improvement	10%	12%	4%	9%

BPNN – BP neural network approach; ANFIS – Adaptive network-based fuzzy inference system approach; FCM – Fuzzy *c*-means clustering approach and FCA – formal concept analysis approach

The similar experiments were also carried out on *must*, *can* and *will* using the new approach in order to verify the effectiveness of the approach in WSD of the English modal verbs. The experimental data were from [3-5], and the experimental results were compared with those from [3-5]. The comparative results are listed in Table 6.

It is known from Table 6 that the rates of correct disambiguation of different modal verbs by FCA approach are commonly higher than that by other approaches, which proves the effectiveness of FCA approach for the WSD of English modal verbs.

It should be noted that the linear equal dividing approach is used to divide the continuous values of the attributes in this study. The linear equal dividing approach is an easy approach but not the optimum one. If the division of the attributes can be optimized, the accuracy of disambiguation of the modal verbs can be further improved.

4. **Knowledge Discovery.** Besides improving the accuracy of disambiguation, the new approach can also be used for the discovery of knowledge behind the modal verbs, such as the relations between the objects and the attributes, and the relations between the attributes.

As shown in Figure 1, there are 6 attributes at L1: a_6, b_5, c_5, e_1, a_1 and c_1, among which a_6 and b_5 cover the most of members of the objects (44% and 40%, respectively). Attribute a_6 mainly covers the objects of epistemic *may*; and attribute b_5 mainly covers the objects of root *may*. This implies that a_6 and b_5 are two attributes with great extents. Since a_6 and b_5 are both the high values of MI between epistemic *may* and the subject and between epistemic *may* and the main verb, respectively, and high values of mutual information embody high degree of relevance of the two words, it may be inferred that the attributes of high semantic relevance of epistemic *may* and the adjacent words have greater extent than the other attributes. This also implies that the MI of epistemic *may* and the adjacent word have the function of generalization of the concept of *may*.

On the contrary, c_k $(k = 1, 2, \ldots, 7)$ and d_m $(m = 1, 2, \ldots, 6)$ are the intents of epistemic *may*, and a_i $(i = 1, 2, \ldots, 6)$ and b_j $(j = 1, 2, \ldots, 6)$ are the intent for root *may*. Here, c_k and d_m represent the MI of root *may* and the subject and the MI of root *may* and the main verb, respectively; a_i and b_j represent the MI of epistemic *may* and the subject and the MI of epistemic *may* and the main verb, respectively. Therefore, in the WSD of epistemic *may* from root *may*, attribute a_6 is the extent of epistemic *may*, and c_k and d_m are the intents of epistemic *may*; attribute b_5 is the extent of root *may*, and attributes a_i and b_j are the intents of root *may*, as shown in Figure 3(a).

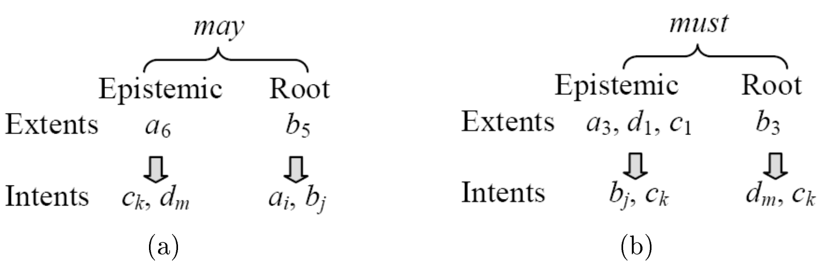

FIGURE 3. General extents and intents of *may* (a) and *must* (b)

In the same way, the concept lattice of *must* is generated with the data in [3], as shown in Figure 4. In the case of *must*, a_i $(i = 1, 2, \ldots, 6)$ are the MI of root *must* and the subject; b_j $(j = 1, 2, \ldots, 6)$ are the MI of epistemic *must* and the subject; c_k $(k = 1, 2, \ldots, 5)$ are the MI of root *must* and the main verb; d_m $(m = 1, 2, \ldots, 7)$ are the MI of epistemic *must* and the main verb; e_1 denotes that the sample sentence is in active voice and f_1 means that the sample sentence is negative. In the training set, the first 25 samples are for root *must* and the other 25 samples are for epistemic *must*.

It can be seen from Figure 4 that b_3 at Level 1 covers all the members of root *must*, and d_m and c_k take the prominence as the end nodes; a_3, d_1 and c_1 at Level 1 covers

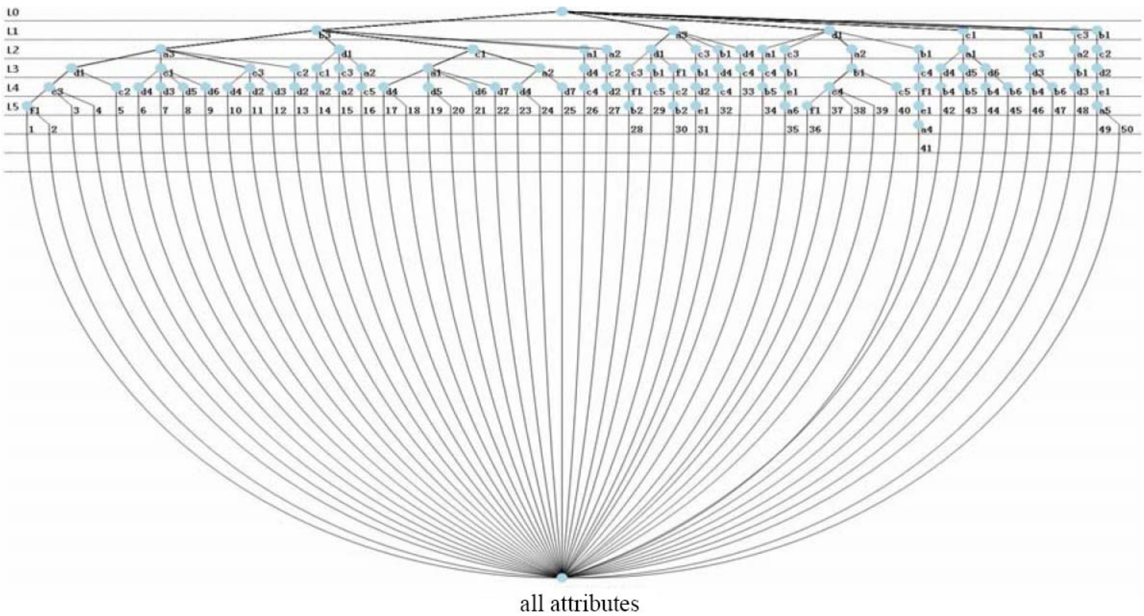

Figure 4. Generated concept lattice of *must*

most of the members of epistemic *must*, and b_j and c_k take the prominence as the end nodes. Therefore, we may generalize that in the WSD of root *must* from epistemic *must*, attributes b_3 is the extent of root *must*, and d_m and c_k are the intents; attribute a_3, d_1, c_1 are the extents of epistemic *must*, and attributes b_j and c_k are the intents, as shown in Figure 3(b).

It is noted for the case of *must* that the extent of root *must* is centralized at b_3, and the extents of epistemic *must* are relatively decentralized at different attributes, such as a_3, d_1 and c_1, which means that epistemic *must* is sensitive to different attributes.

In the same way, some knowledge behind other modal verbs can also be discovered. The knowledge discovered by the proposed approach includes the extents and intents of a concept (a sense of a modal verb), the hierarchical relations between the attributes and the relations between the objects and the attributes. This knowledge provides an important basis for the proper feature selection for WSD in natural language processing, and it brings new light for the semantics of English modal verbs.

5. **Conclusions.** A new approach is proposed based on the formal concept analysis. The senses of English modal verbs can be disambiguated and some valuable knowledge can be discovered using this approach. The approach is advantageous in the higher accuracy of word sense disambiguation (WSD), the hierarchical representation of the complex data, the visualization of the data structure and the feasibility of knowledge discovery. From the above systemic and comparative study, the following conclusions are drawn:

(1) The new approach is a powerful tool for WSD; it can improve the accuracy of WSD of English modal verbs. The accuracies are increased by 10%, 12%, 4%, 9% for *may*, *must*, *can* and *will*, respectively.
(2) The new approach not only can be used as a classifier, but also can visualize the concept hierarchical relations between the objects and the attributes and the horizontal relations between the attributes. Therefore, the complex knowledge behind the word senses can be discovered using this approach, especially the complex hierarchical relations between the objects and the attributes of English modal verbs.

(3) The knowledge discovered by the new approach provides important basis for the feature selection for WSD in natural language processing. It also provides references for the inference of the senses of English modal verbs and broadens the view of semantic study of modal verbs.
(4) It is feasible to apply the new approach to the WSD and knowledge discovery of other modal verbs and other semantically complex words, such as secondary modal auxiliaries, prepositions, and other words which often occur in ambiguous senses.

Acknowledgments. This work is supported by the National Social Science Foundation of China under Grant No.12BYY121 and the Humanities and Social Sciences Foundation of the Ministry of Education of China under Grant No.12YJA740096. It is also partially supported by National Natural Science Foundation of China under Grant No.61074130 and the doctoral scientific research foundation of Yanshan University. The authors gratefully acknowledge the supports.

REFERENCES

[1] T. He and F. Xie, Chinese word sense disambiguation by BP neural network, *Journal of Central China Normal University*, vol.39, no.4, pp 470-474, 2005.
[2] J. Yu and J. Zhang, Word sense disambiguation of the English modal verb *may* by back propagation neural network, *Proc. of IEEE NLP-KE*, Beijing, China, pp.188-193, 2008.
[3] J. Yu, L. An and J. Fu, Word sense disambiguation of English modal verb *must* by neural network, *ICIC Express Letter*, vol.4, no.1, pp.83-88, 2010.
[4] J. Yu, S. Zhao, D. Mei and W. Hong, Sense inference of English modal verb *must* by adaptive network-based fuzzy inference system, *ICIC Express Letter*, vol.5, no.8, pp.2409-2414, 2011.
[5] J. Yu, L. Huang, D. Mei, W. Hong and H. Liu, Determination of the senses of English modal verb *will* by fuzzy c-means cluster, *ICIC Express Letters, Part B: Applications*, vol.2, no.1, pp.177-182, 2011.
[6] R. Wille, *Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts*, Ivan Rival (ed.), Reidel, Dordrecht-Boston, 1982.
[7] J. Baixeries, L. Szathmary, P. Valtichey and R. Godin, Yet a faster algorithm for building the hasse diagram of a concept lattice, *Lecture Notes in Computer Science*, vol.5548, pp.162-177, 2010.
[8] H. Yang, L. Yee and M. Shao, Rule acquisition and attribute reduction in real decision formal contexts, *Soft Computing*, vol.14, no.1, pp.1-14, 2010.
[9] A. Moriki and S. Yoshida, Order-based clustering using formal concept analysis, *Proc. of World Automation Congress*, Kobe, Japan, pp.376-388, 2010.
[10] S. Andrews and C. Orphanidesy, Knowledge discovery through creating formal contexts, *Proc. of the 2nd International Conference on Intelligent Networking and Collaborative Systems*, Thessaloniki, Greece, pp.455-460, 2010.
[11] N. Yashinaga and H. Nobuhara, Formal concept analysis based web pages classification/visualization and their application to information retrieval, *Proc. of the 10th International Symposium on Communications and Information Technologies*, Tokyo, Japan, pp.153-157, 2010.
[12] L. Wang, X. Liu and J. Cao, A new algebraic structure for formal concept analysis, *Information Science*, vol.180, no.24, pp.4865-4876, 2010.
[13] C. Pelaez-Moreno, A. I. Garcia-Moral and F. J. Valverde-Albacete, Analyzing phonetic confusions using formal concept analysis, *Journal of Acoustical Society of America*, vol.128, no.3, pp.1377-1390, 2010.
[14] J. Poelmans, P. Elzinga, S. Viaene and G. Dedene, Formally analyzing the concepts of domestic violence, *Expert Systems with Applications*, vol.38, no.4, pp.3116-3130, 2011.
[15] T. Martin, Y. Shen and A. Majidian, Discovery of time-varying relations using fuzzy formal concept analysis and associations, *International Journal of Intelligent Systems*, vol.25, no.12, pp.1217-1248, 2010.
[16] X. Liu, W. Hong, T. Zhang, F. Fan and J. Song, A visualization method for differential diagnosis in TCM based on formal concept analysis, *Journal of Yanshan University*, vol.34, no.2, pp.162-168, 2010.

[17] B. Zhang, D. Yu and Z. Liu, An optimization ontology algorithm based on FCA, *ICIC Express Letters*, vol.5, no.6, pp.2029-2034, 2011.
[18] J. Coates, *The Semantics of the Modal Auxiliaries*, Routledge Press, London, 1983.
[19] Q. Zong, *Statistic Natural Language Processing*, Qinghua University Press, Beijing, 2008.