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

ICIC International ©2013 ISSN 1349-4198
pp. 3627–3638

INTERACTIVE RELATIONS BETWEEN SEMANTIC AND SYNTACTIC FEATURES IN WORD SENSE DISAMBIGUATION OF SEMANTICALLY COMPLEX WORDS

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

[1]College of Foreign Studies
[2]College 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 August 2012; revised December 2012

Abstract. *The interactions between the semantic and syntactic features in the word sense disambiguation (WSD) of semantically complex words are studied in this paper. A structural partial-ordered attribute diagram (SPOAD) is generated for the WSD of English secondary modal verb "should", and the accuracy of WSD reaches 98%. Some rules are extracted from the SPOAD for WSD of "should", and the accuracy of WSD by the rules reaches 97%. The discovered interactive relations between semantic and syntactic features in WSD and contributions of the features to WSD provide significant and valuable evidence for the proper feature selection in natural language processing, decision making of modality in information processing and semantic study of English modal verb and other semantically complex words.*

Keywords: Interaction, Semantic and syntactic features, Word sense disambiguation, Formal concept analysis, English modal verb *should*

1. **Introduction.** It has been a significant and challenging issue to discover the interactive relations between semantic and syntactic features in word sense disambiguation (WSD), because the discovery of the knowledge may provide important information for decision making in many fields, such as natural language processing, artificial intelligence, machine translation, intelligent control of robot or spacecraft, management, medicine, forecast of weather or disaster and linguistic study [1-3]. However, few related studies are found. In the aspect of WSD, He and Wang [4] proposed an automatic feature selection algorithm based on maximum entropy model for Chinese WSD. Specia et al. [5] investigated a way of improving the accuracy of WSD of English verbs and nouns by constructing a set of features using semantic, syntactic and lexical information. Huang and Lu [6] put forward a knowledge based WSD method by dependency relation and syntax tree. Orhan and Altan [7] determined effective features for WSD of Turkish. Chung and Lee [8] proposed a method of WSD using co-occurring concept codes. Chen et al. [9] presented an approach of WSD of English verbs using linguistically motivated features and produced systems with high performance. Zhang et al. [10] explored syntactic structured features over parse trees for the extraction of semantic relation between entities. Abbasi et al. [11] proposed a rule-based multivariate text feature selection method for semantic classification. Hristea and Colhon [12] put forward a method of feeding syntactic versus semantic knowledge to a knowledge-lean unsupervised WSD algorithm with an underlying Naïve Bayes model. Mirroshandel et al. [13] proposed an improved algorithm

for exploiting some useful syntactic features in order to improve the accuracy of temporal relation classification. It is known that most of the previous studies have been focused on obtaining, or exploring, or selecting features for WSD. Up to now, the interactive relations between semantic and syntactic features in WSD have not been discovered and studied.

With the advance of natural language processing and linguistic studies, more and more knowledge discovery based on WSD is demanded. Experts in natural language processing and linguists need to know how the semantic and syntactic features interact with each other to determine and influence the sense of a word, what relations exist between the them and which features influence the senses of a word greater than the others, in order to obtain important and valuable evidence for the proper feature selection for WSD and the proper understanding of the senses of words. However, the knowledge is not obtainable by the existing methods of WSD, such as BP neural network, support vector machine, Bayesian model and adaptive network-based fuzzy inference system because they can not visualize the structural relations between the senses and the contextual features of the target word. Wille [14] proposed an approach of formal concept analysis (FCA). Since this approach has a good capacity of classification and can visualize the hierarchical relations of the objects and the attributes of concepts, it has been studied extensively and used widely in machine learning, software engineering, information searching, medical diagnosis, and decision making in management.

For this reason, this paper focuses on investigation and discovery of the interactive relations between semantic and syntactic features in WSD using the approach of FCA. 1) Generate a structural partial-ordered attribute diagram (SPOAD) of secondary English modal verb *should*; 2) Extract rules from the SPOAD for WSD of *should*; 3) Discover the interactive relations between semantic and syntactic features and contributions of the features in WSD of *should*.

2. Word Sense Disambiguation (WSD) of English Modal Verb *Should*.

Perkins [15] categorized English modal verbs into primary modal verbs (*may*, *must*, *can*, *shall* and *will*) and secondary modal verbs (*might*, *ought to*, *could*, *should*, *would*). Since the secondary modal verbs are even more ambiguous in senses than the primary ones, the WSD of the secondary modal verb must be more difficult. Generally, the senses of English modal verb *should* are categorized into 2 categories [16]: root *should* (RT*should*) and epistemic *should* (EP*should*).

$$\textit{should}\begin{cases}\text{RT}\textit{should} \text{ – express a weak sense of moral obligation, duty; offer advice} \\ \qquad\qquad\quad \text{or describe correct procedure} \\ \text{EP}\textit{should} \text{ – express a less confident assumption}\end{cases}$$

For instance,

(1) Every citizen *should* obey the law. (RT*should*)
(2) Several potential limitations *should* be considered. (RT*should*)
(3) The report is based on many investigations, so it *should* be reliable. (EP*should*)
(4) With his talent and experience, he *should* do well for himself. (EP*should*)

Since the ambiguity of the word senses of *should* brings troubles to the natural language understanding and processing, the approach of FCA is used to disambiguate RT*should* from EP*should* and to discover the knowledge of the interactive relations between semantic and syntactic features in the WSD of *should*.

2.1. Data preparation.

A corpus is established for this study. It contains about 1 million words and it is composed of written and spoken data, including news reports, novels, research articles, legal documents, movie lines, interviews and presidency debates.

Table 1. Statistic results of occurrence of *should* in corpus

total *should*	RT*should*	EP*should*
771	649	122

The senses of RT*should* and EP*should* are tagged firstly, and then the occurrences of them are counted by Concordance Tool of Wordsmith 4.0. Table 1 shows the statistic results.

2.2. **Feature exaction for WSD of *should*.** Both semantic features and syntactic co-occurrence features are considered as contextual features in order to disambiguate the sense of RT*should* from EP*should*. The mutual information (MI) between *should* and its adjacent word is considered as semantic feature for WSD and knowledge discovery since it has been proven in the previous study [17] that generally MIs have a greater influence to the result of WSD of English modal verb than other features. The 4 MIs considered are:

1) MI (subject, RT*should*); 3) MI (RT*should*, main verb);
2) MI (subject, EP*should*); 4) MI (EP*should*, main verb).

And some co-occurrence syntactic features are selected based on Coates' summarization [16] and the statistic results from the corpus as the potential syntactic features for the WSD and knowledge discovery of *should*:

5) No-strong implication of future time reference; 8) *should* + agentive verb;
6) passive voice; 9) animate subject + *should*;
7) *should* + dynamic verb; 10) Negation.

Some optimal features are determined from among potential syntactic features according to their occurrence and distribution in the 100 samples. The ones which show clear difference in distribution between RT*should* and EP*should* are selected. And those which do not show the clear distinction between the two are deleted. For instance, it is found that the 9$^{\text{th}}$ feature "animate subject + *should*" occurs frequently in both the epistemic sense and the root sense of *should*. It is the same to the 7$^{\text{th}}$ feature "*should* + dynamic verb", they would not contribute much to the WSD of *should*; therefore, they are deleted. And the 10$^{\text{th}}$ feature "negation" also does not contribute much to the WSD of *should* because it occurs very few, and it occurs in both RT*should* and EP*should*; therefore, it is also deleted. Finally, 7 features are determined as follows:

Semantic features:	**Syntactic features:**
1. MI (subject, RT*should*)	5. No-strong implication of future time reference
2. MI (subject, EP*should*)	6. passive voice
3. MI (RT*should*, main verb)	7. *should* + agentive verb
4. MI (EP*should*, main verb)	

Then, 2 data sets are prepared based on the 7 features for the WSD of *should*. One is training set which is composed of 100 sample sentences (objects) evenly taken out from the corpus in which 50 samples contain RT*should* and another 50 samples contain EP*should*. The other data set is testing set which is also composed of 100 sample sentences, 50 for RT*should* and 50 for EP*should*.

2.3. **Vectorization of the data.** After the constitution of the 2 data sets, the 4 MIs are calculated according to Formula (1) [18].

$$MI(w_1, w_2) = \log \frac{P(w_1, w_2)}{P(w_1)P(w_2)} \tag{1}$$

here, w_1 and w_2 are two words, $MI(w_1, w_2)$ embodies the degree of relevance of w_1 and w_2. If $MI(w_1, w_2) > 0$, or $= 0$, or < 0; then w_1 and w_2 are relevant, or independent, or compensate, respectively. After that logical values are given to the syntactic features of the samples in order to vectorize the data for each object. If a feature supports the sense of RT*should*, give it 1; otherwise, give it 0. Finally, the data set for training is obtained; see Table 2.

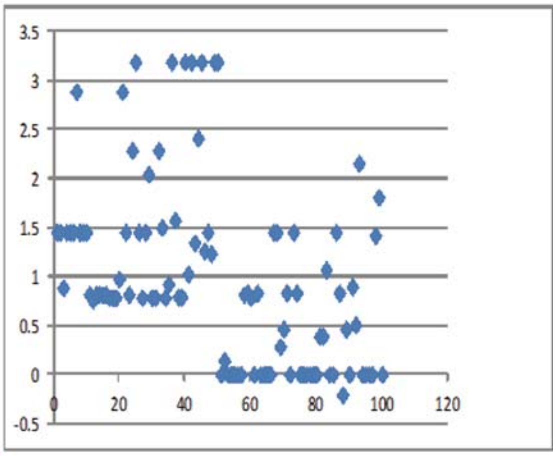
The first 4 data (MIs) in Table 2 need to be symbolized because the formal context can only be expressed by binary values, not continues values. Therefore, a symbolization of the continuous values is carried out.

2.4. **Symbolization of the data set of *should*.** An easy way of symbolizing the continuous value in a data set is to divide the continuous values into ranges, and each range is set to be 0.5. However, this linear equal dividing method is obviously not an optimal one because 0.5 is a random value but not an optimal one. Besides, this method may result in over-refined division of the continuous values, which may lead to great increase in the number of attributes. It may also make the analysis of the model of WSD more complicated. Therefore, an optimal binary value division method is proposed. The determination of the dividing value of the first 4 data (MIs) would directly influence the accuracy of WSD. In order to find a dividing value which can nicely separate the MI for RT*should* from the MI for EP*should*, the MI distribution diagrams are generated; see Figure 1.

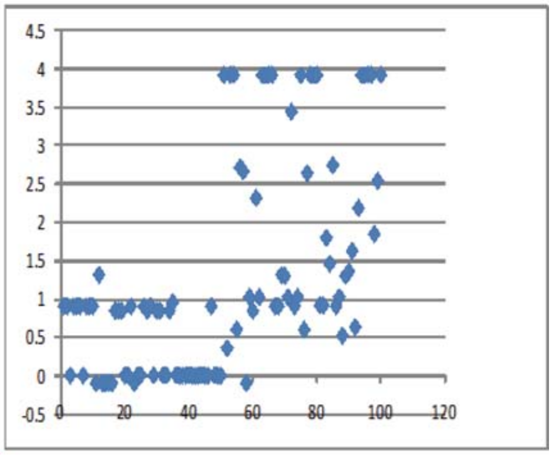
It is known from Figure 1(a), all the MIs(s, RT*should*) for the first 50 samples (containing RT*should*) are distributed at the area of 0.76 or above, and the MIs(s, RT*should*) for most of the second 50 samples (containing EP*should*) are distributed at the area of below 0.76. Therefore, the division for MIs(s, RT*should*) is set to be 0.76. Similarly, the divisions of the other MIs are determined based on Figure 1 as follows:

Table 2. Training set of *should*

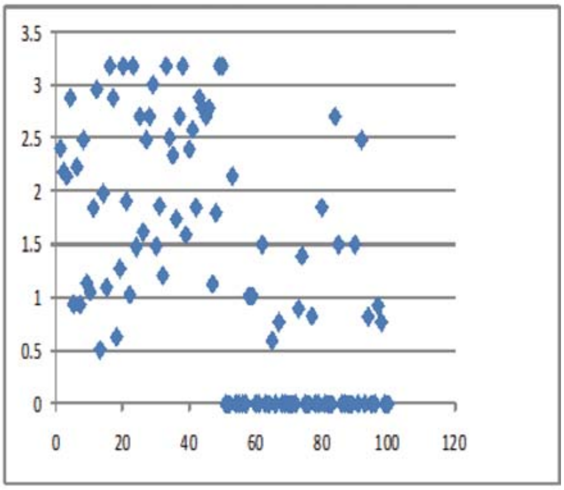
No.	data	No.	data	No.	data	No.	data
1	1.45,0.9,2.41,0,1,1,0	26	1.45,0.9,1.62,0,0,0,1	51	0,3.9,0,3.4,0,0,0	76	0,0.6,0,3.9,0,0,0
2	1.45,0.9,2.19,0,1,0,1	27	0.78,0.84,2.49,0,1,0,1	52	0.14,0,35,0,3.9,0,0,1	77	0,2.6,0.83,0.7,0,0,0
3	0.88,0,2.15,0,1,1,1	28	1.45,0.9,2.71,0,1,0,0	53	0,3.9,2.15,2.87,0,0,0	78	0,3.9,0,3.6,0,0,0
4	1.45,0.9,2.89,0,1,0,1	29	2.04,0,3.01,0,0,1,1	54	0,3.9,0,3.1,0,0,1	79	0,3.9,0,2.2,0,0,0
5	1.45,0.9,0.94,0,1,0,1	30	0.78,0.84,1.49,0,1,0,1	55	0,0.59,0,3.9,1,0,0	80	0,3.9,1.85,2.1,1,0,0
6	1.45,0.9,2.23,0,1,0,1	31	0.78,0.84,1.87,0,0,0,1	56	0,2.7,0,1.95,0,0,0	81	0.39,0.9,0,3.9,1,0,0
7	2.89,0,0.94,1.66,1,0,1	32	2.28,0,1.21,0,1,0,0	57	0,2.66,0,3.6,1,0,0	82	0.39,0.9,0,1,13,0,0,0
8	1.45,0.9,2.49,0,1,0,1	33	1.49,0,3.19,0,1,0,1	58	0.82,−0.1,1,1.35,0,0,1	83	1.07,1.8,0,3.4,0,0,0
9	1.45,0.9,1.14,0,0,0,1	34	0.78,0.84,2.51,0,1,0,1	59	0.83,1,1,1.04,0,0,0	84	0,1.5,2.7,3.4,1,0,0
10	1.45,0.9,1.06,0.83,0,0,0	35	0.92,0.95,2.34,0,1,0,1	60	0.78,0.84,0,2.9,1,0,0	85	0,2.7,1.5,1.8,1,0,0
11	0.82,−0.1,1.85,0,1,0,0	36	3.19,0,1.74,0,0,1,1	61	0,2.3,0,3.3,0,0,0	86	1.45,0.9,0,2.6,0,0,1
12	0.76,1.31,2.97,0,1,1,1	37	1.57,0,2.71,0,1,0,0	62	0.83,1,1.5,1.8,1,0,0	87	0.83,1,0,2.9,0,0,0
13	0.82,−0.1,0.51,0,1,0,0	38	0.79,0,3.19,0,1,1,1	63	0,3.9,0,2.46,0,0,0	88	−0.21,0.5,0,2.3,1,0,0
14	0.82,−0.1,1.98,0,1,0,1	39	0.79,0,1.59,0,1,0,0	64	0,3.9,0,3.3,0,0,0	89	0.46,1.3,0,2.5,0,0,0
15	0.82,−0.1,1.1,0,1,0,1	40	3.19,0,2.4,0,0,1,0	65	0,3.9,0.6,1.32,1,0,0	90	0,1.35,1.5,1.8,1,0,0
16	0.82,−0.1,3.18,0,1,0,1	41	1.02,0,2.59,0,1,1,0	66	00,3.9,0,3.9,0,0,0	91	0.89,1.6,0,1.9,0,0,0
17	0.78,0.84,2.89,0,1,0,0	42	3.19,0,1.85,2.1,1,1,1	67	1.45,0.9,0.8,2.2,0,0,0	92	0.5,0.6,2.5,3.2,0,0,1
18	0.78,0.84,0.63,0,0,0,0	43	1.34,0,2.89,0,1,1,1	68	1.45,0.9,0,3.6,0,0,0	93	2.15,2.2,0,2.8,1,0,1
19	0.78,0.84,1.28,0,1,0,0	44	2.41,0,2.79,0,0,1,1	69	0.28,1.31,0,2.9,0,0,0	94	0,3.9,0.8,1.2,1,0,1
20	0.97,0,3.19,0,1,1,0	45	3.19,0,2.71,0,1,1,1	70	0.46,1.3,0,2.5,0,0,0	95	0,3.9,0,3.2,1,0,0
21	2.89,0,1.91,0,1,1,1	46	1.26,0,2.79,0,1,1,1	71	0.83,1,0,1,0,0,0	96	0,3.9,0,3.3,1,0,0
22	1.45,0.9,1.o3,0,1,0,1	47	1.45,0.9,1.13,0,1,0,1	72	0,3.4,0,3.9,0,0,0	97	0,3.9,0.9,1.7,0,0,1
23	0.82,−0.1,3.19,0,0,0,1	48	1.23,0,1.8,0,0,1,1	73	1.45,0.9,0.9,2.1,0,0,0	98	1.4,1.8,0.8,2.2,0,0,0
24	2.28,0,1.48,0,1,0,0	49	3.19,0,3.19,0,0,1,1	74	0.83,1,1.4,2.12,0,0,0	99	1.8,2.5,0,3.9,0,0,0
25	3.19,0,2.71,0,1,1,0	50	3.19,0,3.19,0,1,1,0	75	0,3.9,0,2.5,0,0,0	100	0,3.9,0,3.4,1,0,0



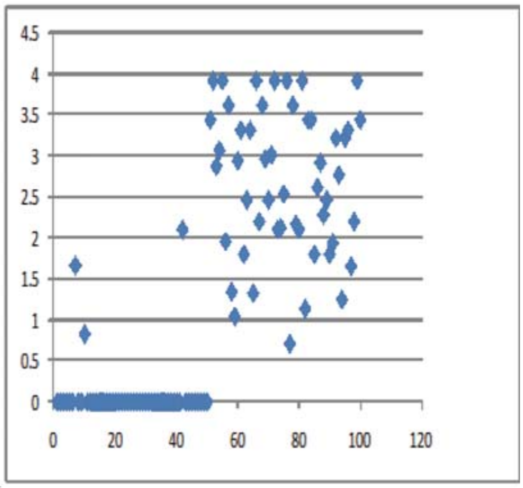
(a) MI(s, RT*should*)



(b) MI(s, EP*should*)



(c) MI(RT*should*, v)



(d) MI(EP*should*, v)

FIGURE 1. Distribution of MIs of *should*

a1 = 0 for MI(s, RT*should*) ≤ 0.76 a2 = 1 for MI(s, RT*should*) > 0.76
a3 = 1 for MI(s, EP*should*) ≤ 1.1 a4 = 0 for MI(s, EP*should*) > 1.1
a5 = 0 for MI(RT*should*, v) ≤ 1.05 a6 = 1 for MI(RT*should*, v) > 1.05
a7 = 1 for MI(EP*should*, v) ≤ 1 a8 = 0 for MI(EP*should*, v) > 1

In this way, the symbolization of MIs is finished. In the symbolization of the MIs, we also follow the principle that if a feature supports the sense of RT*should*, it is given 1; otherwise, it is given 0. The formal context is formed based on the result of symbolization, as shown in Table 3.

2.5. **Generation of structural partial-ordered attribute diagram and WSD of *should*.** In formal concept analysis, WSD is concerning the classification of word senses according to the attributes the objects have. Objects have certain attributes, and objects can be divided into certain categories by attribute characteristics. The objects in formal concept analysis correspond to the samples of *should* and the attributes correspond to the features owned by the samples in this study. Based on the formal context, a structural partial-ordered attribute diagram (SPOAD) is generated by a SPOAD generation tool [19]. Since there are 100 objects in the training data set and the generated diagram is too big to show in this paper, the formal context is clarified by deleting the objects with repeated attribute (feature) patterns, and the clarified formal context is shown in Table 4.

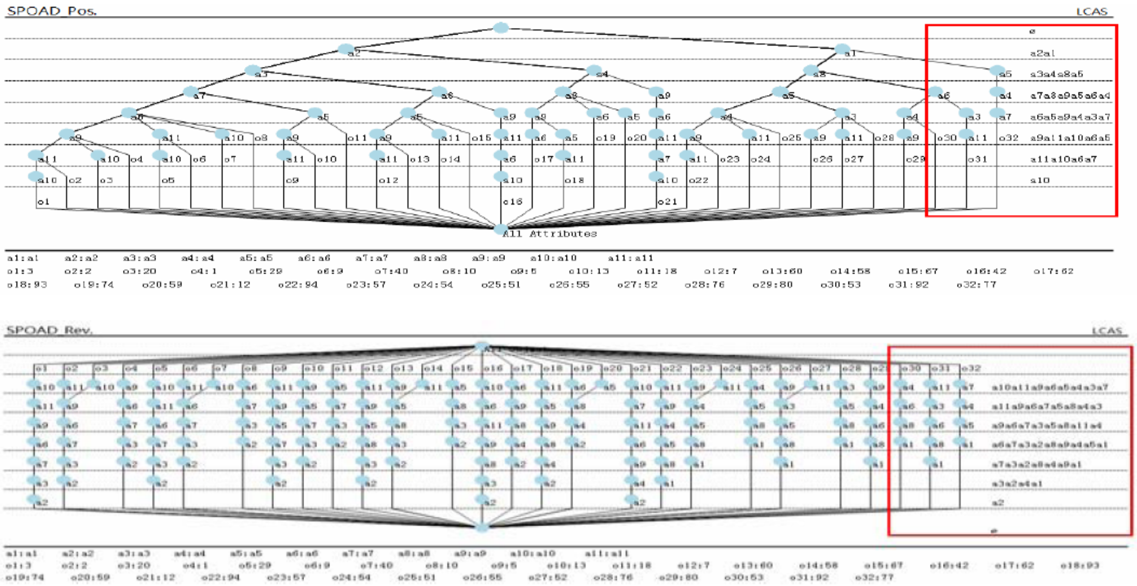


FIGURE 2. Clarified SPOAD Pos. and SPOAD Rev. of *should*

Finally, the corresponding SPOAD Pos. (referring to the positive SPOAD) and SPOAD Rev. (referring to the reversed SPOAD) of *should* are generated, as shown in Figure 2.

The result of WSD of *should* is checked by the leave-one-out testing method and the similarity between the feature pattern of the out left object and the one found in the SPOAD. If o1' is the out left object and its feature pattern can be described by: $X_1' = \{x_1', x_2', x_3', \ldots, x_i'\}$, and o1 is an object in the SPOAD which has similar feature pattern to o1' and can be described by: $X_1 = \{x_1, x_2, x_3, \ldots, x_m\}$, then the similarity of X_1' and X_1 is represented by S:

$$S = |X_1' \cap X_1| / \max\left(|X_1'|, |X_1|\right) \tag{2}$$

If $S = 1$, then X_1' 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_1' is most approximate to X_1, and they are classified into one class. If S of X_1' is the same as S_i of several other objects, then X_1' is classified into the class which has the maximum members of objects having co-occurred features. If the similarity between X_1' and X_1 is greater than 2/3, then the two objects are in the same class. In this way, the model can be checked and the accuracy reaches 100%. After that the model is tested by the testing data set, and the accuracy reaches 98%.

3. **Extraction of Rules for WSD of *Should*.** The rules for WSD of *should* are extracted based on the SPOAD in Figure 2 and the following theoretical descriptions of FCA [20]:

Definition 3.1. *A formal context $K = (G, M, I)$ consists of two sets G and M and a relation I between G and M. The elements of G are called the objects and the elements of M are called the attributes of the context. In order to express that an object g is in a relation I with an attribute m, write gIm or $(g, m) \in I$ and read it as the object g has attribute m.*

Definition 3.2. *Let $K = (G, M, I)$ be a formal context, for a set $A \subseteq G$, $f(A) = \{m \in M | (g, m) \in I, \forall g \in A\}$. Correspondingly, for a set $B \subseteq M$, define $g(B) = \{g \in G | (g, m) \in I, \forall m \in B\}$. A formal concept is a pair (A, B) with $A \subseteq G$, $B \subseteq M$,*

Table 3. Formal context of *should*

No.	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	No.	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11
1		1	1			1	1		1			51	1			1	1			1			
2		1	1			1	1		1		1	52	1		1		1			1			1
3		1	1			1	1		1	1	1	53	1			1		1		1			
4		1	1			1	1		1		1	54	1			1	1			1			1
5		1	1		1		1		1		1	55	1		1		1			1	1		
6		1	1			1	1		1		1	56	1			1	1			1			
7		1	1		1			1	1		1	57	1			1	1			1	1		
8		1	1			1	1		1		1	58		1	1		1			1			1
9		1	1			1	1				1	59		1		1	1			1			
10		1	1			1	1					60		1	1		1			1	1		
11		1	1			1	1		1			61	1			1	1			1			
12		1		1		1	1		1	1	1	62		1		1		1		1	1		
13		1	1		1		1		1			63	1			1	1			1			
14		1	1			1	1		1		1	64	1			1	1			1			
15		1	1			1	1		1		1	65	1			1	1			1	1		
16		1	1			1	1		1		1	66	1			1	1			1			
17		1	1			1	1		1			67		1	1		1			1			
18		1	1		1		1					68		1	1		1			1			
19		1	1			1	1		1			69	1			1	1			1			
20		1	1			1	1		1	1		70	1			1	1			1			
21		1	1			1	1		1	1	1	71		1		1	1			1			
22		1	1		1		1		1		1	72	1			1	1			1			
23		1	1			1	1				1	73		1	1		1			1			
24		1	1			1	1		1			74		1		1		1		1			
25		1	1			1	1		1	1		75	1			1	1			1			
26		1	1			1	1				1	76	1		1		1			1			
27		1	1			1	1		1		1	77	1			1	1		1				
28		1	1			1	1		1			78	1			1	1			1			
29		1	1			1	1			1	1	79	1			1	1			1			
30		1	1			1	1		1		1	80	1			1		1		1	1		
31		1	1			1	1				1	81	1		1		1			1	1		
32		1	1			1	1		1			82	1		1		1			1			
33		1	1			1	1		1		1	83		1		1	1			1			
34		1	1			1	1		1		1	84	1			1		1		1	1		
35		1	1			1	1		1		1	85	1			1		1		1	1		
36		1	1			1	1			1	1	86		1	1		1			1			1
37		1	1			1	1		1			87		1		1	1			1			
38		1	1			1	1		1	1	1	88	1		1		1			1	1		
39		1	1			1	1		1			89	1			1	1			1			
40		1	1			1	1			1		90	1			1		1		1	1		
41		1	1			1	1		1	1		91		1		1	1			1			
42		1	1			1		1	1	1	1	92	1		1			1		1			1
43		1	1			1	1		1	1	1	93		1		1	1			1	1		1
44		1	1			1	1			1	1	94	1			1	1			1	1		1
45		1	1			1	1		1	1	1	95	1			1	1			1	1		
46		1	1			1	1		1	1	1	96	1			1	1			1	1		
47		1	1			1	1		1		1	97	1			1	1			1			1
48		1	1			1	1			1	1	98		1		1	1			1			
49		1	1			1	1			1	1	99		1		1	1			1			
50		1	1			1	1		1	1		100	1			1	1			1	1		

TABLE 4. Clarified formal context of *should*

No.	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	No.	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11
1		1	1			1	1		1			17	1			1		1		1			
2		1	1			1	1		1		1	18	1			1	1			1			1
3		1	1			1	1		1	1	1	19	1		1		1			1	1		
4		1	1		1		1		1		1	20	1			1	1			1	1		
5		1	1		1			1	1		1	21		1	1		1			1			1
6		1	1			1	1				1	22		1		1	1			1			
7		1	1			1	1				1	23		1	1		1			1	1		
8		1		1		1	1		1	1	1	24		1		1		1		1	1		
9		1	1		1		1		1			25		1	1		1			1			
10		1	1		1		1					26		1		1		1		1			
11		1	1			1	1		1	1		27	1		1		1			1			
12		1	1			1	1			1	1	28	1			1	1		1				
13		1	1			1	1			1		29	1			1		1		1	1		
14		1	1			1		1	1	1	1	30	1		1			1		1			
15	1			1	1			1				31		1		1	1			1	1		1
16	1		1		1			1			1	32	1			1	1			1	1		1

$f(A) = B$ and $g(B) = A$. A is called the extent of the concept and B is called the intent of the concept.

Definition 3.3. *Let $K = (G, M, I)$ be a formal context, if for any objects g_1, $g_2 \in G$ from $f(g_1) = f(g_2)$, it always follows that $g_1 = g_2$ and correspondingly, $g(m_1) = g(m_2)$ implies $m_1 = m_2$ for all m_1, $m_2 \in M$, the context $K = (G, M, I)$ is called clarified context.*

Definition 3.4. *$K = (G, M, I)$ is a formal context, m_i is any attribute in the attribute set. If there is at least one object among the set of objects sharing m_i does not belong to the set of objects sharing the other attributes, then, m_i is called an exclusive attribute.*

Definition 3.5. *$K = (U, M, I)$ is a formal context, $m \in M$. If attribute m satisfies $\{g(m)|m \in M\} = U$, then, m is called maximum common attribute.*

The extraction of the rules for WSD of *should* goes through the following steps:

1) Find an attribute m_i in the SPOAD, go up the line to the top node; the attribute combination constitutes m in the concept pair (g, m); go down the line or lines to the end, and the object set constitutes g in the concept pair (g, m).
2) The calculations are carried out to every two pairs in the pair set by the following algorithm. If $C_1 = (A_1, B_1)$ is a pair of a concept in the SPOAD, and $C_2 = (A_2, B_2)$ is another pair of the concept, then $C_3 = f\{C_1 \oplus C_2\} = \{A_3, B_3\} = \{A_1 \cup A_2, B_1 \cap B_2\}$. In this way, the new pair set is obtained. If any newly generated pair is the same as any pair in the original pair set, the new one is deleted.
3) If the intent of any pair in the new pair set is the same as that of any pair in the original pair set, then the original one is deleted, i.e., the pair that has the larger extent is retained.
4) If there is only one element in the new pair set, the calculation about this attribute stops; if not, go back to step 2) and the calculation will continue.

The flowchart of rule extraction is shown in Figure 3.

The following rules are extracted from Figure 2 for WSD of *should*. Rules for class 1 (RT*should*) are: (1) {1, a7a2}; (2) {1, a10a2}; (3) {1, a11a9a5a8a3a2} and the rules for class 2 (EP*should*) are: (1) {2, a9a5a8a3a2}; (2) {2, a11a5a8a3a2}; (3) {2, a8a4a2}; (4)

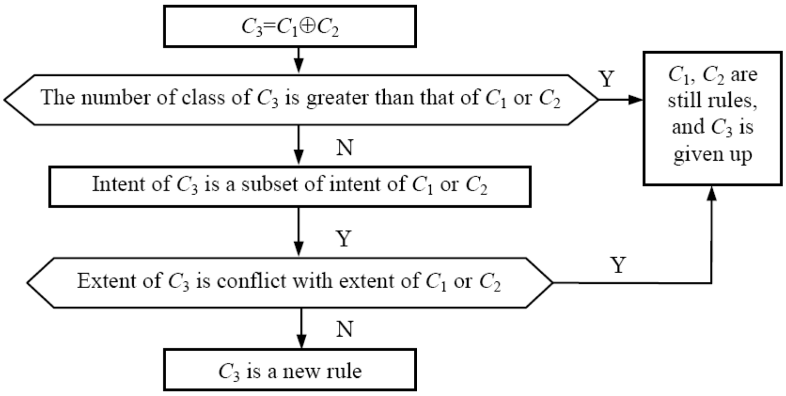


Figure 3. Flowchart of rule extraction

{2, a1}. The extracted rules are used for the WSD of RT*should* from EP*should*, and the accuracy reaches 97%.

4. Knowledge Discovery of *Should.*

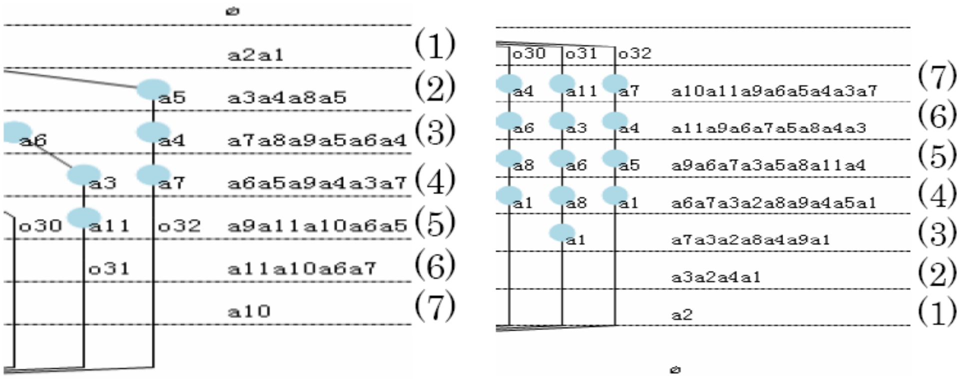
4.1. Interactive relations between the semantic features and syntactic features. It can be seen from Figure 2 that most feature patterns of *should* are a combination of semantic and syntactic features. Some feature patterns are clusters of only semantic features. In the rule extraction, it is found that the semantic feature cluster {a5a8a3a2} some times judges a *should* belonging in class 1 and some times judges a *should* belonging in class 2, which indicates that this intent feature cluster cannot be used as a rule. However, when some syntactic features are added into the intent, some rules are formed, such as the rule (3) for class 1 and the rule (1) and rule (2) for class 2. This implies that in many cases, semantic and syntactic features work together to determine the senses of a word. Syntactic features have the functions of specification and they are playing the role of classification.

4.2. Senses of *should* and its special features. It may be found from the extracted rules that feature a2 occurs in almost all the rules for both class 1 (RT*should*) and class 2 (EP*should*), it occurs in all the rules for class 1, which implies that semantic feature a2 is a maximum common feature of class 1 (RT*should*), while feature a1 only occurs in one rule for class 2, which implies that semantic feature a1 is an exclusive feature of class 2.

4.3. Intent features and extent features of *should.* Extent attributes represent the attributes commonly owned by the objects and they occur at higher layers of the SPOAD representing the generality of the objects. On the contrary, intent attributes show the specific character of the objects – the distinction of one object from the other objects. They occur at the lower layers of the SPOAD. The red square parts in Figure 2 are shown in Figure 4 in order to discover the intent and extent features.

Figure 4(a) shows the feature distributions in the 7 layers of the SPOAD. There is a gradient of concept in the 7 layers from generality to specification. The upper the layer lies at, the more general the features in the layer are; on the contrary, the lower the layer lays, the more specific the features in the layer are. It is observed from Figure 4(a):

1) Features a2 and a1 occur only at layer (1), which implies that they are extent features of *should.*
2) The semantic feathers ai ($i = 3, 4, \ldots, 8$) mostly occur at first half of the 7 layers (layers (2), (3) and (4)), which implies that they tend to be extent features. They have greater extent than intent of *should.*



(a) Positive distribution of features (b) Reversed distribution of features

FIGURE 4. Hierarchical distributions of the features of *should*

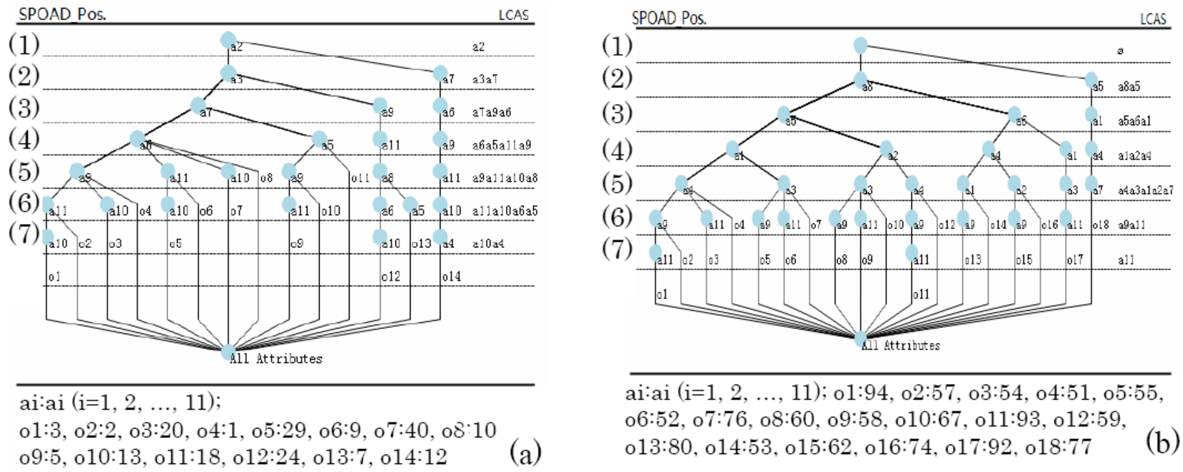


FIGURE 5. Generated SPOAD of RT*should* (a) and EP*should* (b)

3) The syntactic features a9, a10 and a11 mainly occur at the second half of the 7 layers (layer (4), (5), (6) and (7)), which implies that they tend to be extent features.

Figure 4(b) shows the reversed feature distribution of *should*. It shows a clear picture of the intent feature of each object. It can be seen that the three syntactic features (a9, a10 and a11) take the prominent at layer (7) (70% of the objects), which means that 70 objects are finally ended and classified by syntactic features. It can be summarized from the hierarchical distribution of the features of *should* that generally the semantic features tend to be extent of *should* and syntactic features tend to be intent of *should*.

In order to discover the knowledge of the relations between the semantic and syntactic features within one class, the clarified SPOADs of RT*should* and EP*should* are generated, respectively, as shown in Figure 5. Comparing the feature distributions of RT*should* with that of EP*should* in Figure 5, it is known that the features in each corresponding layer of the two classes tend to be mutually exclusive. For instance, at the layer (2), the features for RT*should* are a3 and a7, while the features for EP*should* are a8 and a5. This implies that each sense of *should* has its exclusive features.

It is found from Figure 5: (1) Semantic feature a2 covers every object of RT*should*, and a8 covers almost all the objects of EP*should*. Feature a2 represents MI(s, RT*should*) > 0.76 and a8 represents MI(EP*should*, v) > 1. This implies that the feature MI(s, RT*should*) > 0.76 is the maximum common and extent feature of RT*should* and the

TABLE 5. Experimental results of the WSD

type of features	deleted features	correct disambiguation	accuracy
semantic features	a1-a8	92	92%
	a1-a4	97	97%
	a5-a8	94	94%
syntactic features	a9-a11	98	98%

feature MI(EP*should*, v) > 1 is a common and extent feature of EP*should*. (2) Syntactic features a9, a10 and a11 occur at lower layers. This implies that syntactic features tend to be the intents of both RT*should* and EP*should*.

4.4. **Contributions of contextual features to WSD of *should*.** Another experiment is carried out to investigate the contributions of semantic and syntactic features to the results of WSD. 1) Delete some features from the training formal context; 2) Generate the SPOAD from the rest formal context; 3) Check the accuracy of the WSD. If the accuracy of the WSD changes greatly, the features have a great contribution to the WSD; otherwise, the features have little contribution to the WSD of *should*.

The experimental results are shown in Table 5, in which a1-a8 are semantic features among which a1-a4 are the MI(s, *should*) and a5-a8 are the MI(*should*, v); a9-a11 are syntactic features. It is found that when all the semantic features are deleted from the formal context, the accuracy of WSD drops to the lowest (92%). This implies that semantic features have a great influence on the accuracy of WSD. It is also found that among the semantic features, the MI(*should*, v) has a greater influence than the MI(s, *should*), and syntactic features have less influence to the accuracy of WSD of *should*.

5. **Conclusions.** The word sense disambiguation (WSD) of English secondary modal verb *should* is realized using the approach of formal concept analysis with accuracy of 98%.

Generally, semantic features and syntactic features work together to determine the sense of *should*; in some cases, the senses of some objects of *should* are determined only by semantic features.

Semantic features tend to be the extent of *should* with the function of semantic generalization. Syntactic features tend to be the intent of *should* with the function of sense classification. The semantic features contribute more to the WSD of *should* than the syntactic features.

Among the semantic features, the feature of the mutual information of subject and RT*should* being greater than 0.76 is a maximum common feature of RT*should*, while the feature of the mutual information of subject and RT*should* being equal or less than 0.76 is an exclusive feature of EP*should*; and the feature of the mutual information of EP*should* and main verb being greater than 1 is a common feature of EP*should*.

The above discovered knowledge based on the WSD of English secondary modal verb provides very important evidence for the feature selection in natural language processing and for the semantic study of secondary English modal verbs. The established model for WSD of *should* can be used in the automatic natural language processing systems and other applications. Moreover, the proposed approach can also be used in the WSD and knowledge discovery of other secondary modal verbs.

The further study will focus on the mechanism of word sense formation of secondary English modal verbs and other semantically complex words.

Acknowledgments. This work is supported by the National Social Science Foundation of China under Grant No. 12BYY121 and by the Humanities and Social Sciences Foundation of the Ministry of Education of China under Grant No. 12YJA740096. It is also supported by National Science Foundation of China under Grant No. 61273019 and the doctoral scientific research foundation of Yanshan University. The authors gratefully acknowledge the supports.

REFERENCES

[1] F. Mandreoli, Knowledge-based sense disambiguation for all structures, *Information Systems*, vol.36, no.2, pp.406-430, 2010.
[2] M. Hwang et al., Automatic enrichment of semantic relation network and its application to word sense disambiguation, *IEEE Transactions on Knowledge and Data Engineering*, vol.23, no.6, pp.845-858, 2011.
[3] Z. Lu, T. Liu and S. Li, The research of statistical word sense disambiguation, *Acta Electronica Sinica*, vol.34, no.2, pp.333-343, 2006.
[4] J. He and H. Wang, Chinese word sense disambiguation based on maximum entropy model with feature selection, *Journal of Software*, vol.21, no.6, pp.1287-1295, 2010.
[5] L. Specia, A. Srinivasan, S. Joshi, G. Ramakrishnan and M. Das Gracas Vople Nunes, An investigation into feature construction to assist word sense disambiguation, *Machine Learning*, vol.76, no.1, pp.109-136, 2009.
[6] H. Huang and W. Lu, Knowledge-based word sense disambiguation with feature words based on dependency relation and syntax tree, *International Journal of Advancements in Computing Technology*, vol.3, no.8, pp.73-81, 2011.
[7] Z. Orhan and Z. Altan, Determining effective feature for word sense disambiguation in Turkish, *Journal of Electrical and Electronics Engineering*, vol.5, no.2, pp.1341-1352, 2005.
[8] Y. Chung and J. Lee, Practical word sense disambiguation using co-occurring concept codes, *Machine Translation*, vol.19, no.1, pp.59-82, 2005.
[9] J. Chen, D. Dligach and M. Palmer, Towards large-scale high-performance English verb sense disambiguation by using linguistically motivated features, *ICSC 2007 International Conference on Semantic Computing*, pp.378-385, 2007.
[10] M. Zhang, G. Zhou and A. Aw, Exploring syntactic structured features over parse trees for relation extraction using kernel methods, *Information Processing and Management*, vol.44, no.2, pp.687-701, 2008.
[11] A. Abbasi, S. France, Z. Zhang and H. Chen, Selecting attributes for sentiment classification using feature relation networks, *IEEE Transactions on Knowledge and Data Engineering*, vol.23, no.3, pp.447-462, 2011.
[12] F. Hristea and M. Colhon, Feeding syntactic versus semantic knowledge to a knowledge-lean unsupervised word sense disambiguation algorithm with an underlying Naïve Bayes model, *Fundamenta Informaticae*, vol.119, no.1, pp.61-86, 2012.
[13] S. A. Mirroshandel, G. Ghassem-Sani and M. Khayyamian, Using syntactic-based kernels for classifying temporal relations, *Journal of Computer Science and Technology*, vol.26, no.1, pp.68-80, 2010.
[14] R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, in *Ordered Sets*, I. Rival (ed.), Reidel, Dordrecht-Boston, 1982.
[15] M. R. Perkins, *The Expression of Modality in English*, The Polytechnic of Wales, C. N. A. A, 1980.
[16] J. Coates, *The Semantics of the Modal Auxiliaries*, Routledge Press, London, 1983.
[17] J. Yu, H. Dong, J. Fu and T. Bai, An investigation of contributions of different linguistic features to the WSD of English modal verb MAY by BP neural network, *ICIC Express Letters*, vol.3, no.3(A), pp.391-396, 2009.
[18] Q. Zong, *Statistic Natural Language Processing*, Qinghua University Press, Beijing, 2008.
[19] W. Hong, S. Li, J. Yu and J. Song, A new approach of generation of structural partial-ordered attribute diagram, *ICIC Express Letters, Part B: Applications*, vol.3, no.4. pp.823-830, 2012.
[20] G. Ganter and R. Wille, *Formal Concept Analysis*, Springer-Verlag, Berlin, 1999.