

Integrating Collocation Features in Chinese Word Sense Disambiguation

Wanyin Li

Department of Computing
The Hong Kong Polytechnic
University
Hong Hom, Kowloon, HK
cswyli@comp.polyu.edu.hk

Qin Lu

Department of Computing
The Hong Kong Polytechnic
University
Hong Hom, Kowloon, HK
csqinlu@comp.polyu.edu.hk

Wenjie Li

Department of Computing
The Hong Kong Polytechnic
University
Hong Hom, Kowloon, HK
cswjli@comp.polyu.edu.hk

Abstract

The selection of features is critical in providing discriminative information for classifiers in Word Sense Disambiguation (WSD). Uninformative features will degrade the performance of classifiers. Based on the strong evidence that an ambiguous word expresses a unique sense in a given collocation, this paper reports our experiments on automatic WSD using collocation as local features based on the corpus extracted from People’s Daily News (PDN) as well as the standard SENSEVAL-3 data set. Using the Naïve Bayes classifier as our core algorithm, we have implemented a classifier using a feature set combining both local collocation features and topical features. The average precision on the PDN corpus has 3.2% improvement compared to 81.5% of the baseline system where collocation features are not considered. For the SENSEVAL-3 data, we have reached the precision rate of 37.6% by integrating collocation features into contextual features, to achieve 37% improvement over 26.7% of precision in the baseline system. Our experiments have shown that collocation features can be used to reduce the size of human tagged corpus.

1 Introduction

WSD tries to resolve lexical ambiguity which refers to the fact that a word may have multiple meanings such as the word “walk” in “Walk or Bike to school” and “BBC Education Walk

Through Time”, or the Chinese word “地方” in “地方政府”(“local government”) and “他也有对的地方”(“He is also partly right”). WSD tries to automatically assign an appropriate sense to an occurrence of a word in a given context.

Various approaches have been proposed to deal with the word sense disambiguation problem including rule-based approaches, knowledge or dictionary based approaches, corpus-based approaches, and hybrid approaches. Among these approaches, the supervised corpus-based approach had been applied and discussed by many researches ([2-8]). According to [1], the corpus-based supervised machine learning methods are the most successful approaches to WSD where contextual features have been used mainly to distinguish ambiguous words in these methods. However, word occurrences in the context are too diverse to capture the right pattern, which means that the dimension of contextual words will be very large when all words in the training samples are used for WSD [14]. Certain uninformative features will weaken the discriminative power of a classifier resulting in a lower precision rate. To narrow down the context, we propose to use collocations as contextual information as defined in Section 3.1.2. It is generally understood that the sense of an ambiguous word is unique in a given collocation [19]. For example, “包袱” means “burden” but not “baggage” when it appears in the collocation “思想包袱” (“burden of thought”).

In this paper, we apply a classifier to combine the local features of collocations which contain the target word with other contextual features to discriminate the ambiguous words. The intuition is that when the target context captures a collocation, the influence of other dimensions of

contextual words can be reduced or even ignored. For example, in the expression “恐怖分子焚毁了基因室” (“terrorists burned down the gene laboratory”), the influence of contextual word “基因” (“gene”) should be reduced to work on the target word “分子” because “恐怖分子” is a collocation whereas “分子” and “基因” are not collocations even though they do co-occur. Our intention is not to generally replace contextual information by collocation only. Rather, we would like to use collocation as an additional feature in WSD. We still make use of other contextual features because of the following reasons. Firstly, contextual information is proven to be effective for WSD in the previous research works. Secondly, collocations may be independent on the training corpus and a sentence in consideration may not contain any collocation. Thirdly, to fix the tie case such as “...恐怖分子基因测试...” (“terrorists’ gene checking”), “分子” means “human” when presented in the collocation “恐怖分子”, but “particle” in the collocation “分子基因”. The primary purpose of using collocation in WSD is to improve precision rate without any sacrifices in recall rate. We also want to investigate whether the use of collocation as an additional feature can reduce the size of hand tagged sense corpus.

The rest of this paper is organized as follows. Section 2 summarizes the existing Word Sense Disambiguation techniques based on annotated corpora. Section 3 describes the classifier and the features in our proposed WSD approach. Section 4 describes the experiments and the analysis of our results. Section 5 is the conclusion.

2 Related Work

Automating word sense disambiguation tasks based on annotated corpora have been proposed. Examples of supervised learning methods for WSD appear in [2-4], [7-8]. The learning algorithms applied including: decision tree, decision-list [15], neural networks [7], naïve Bayesian learning ([5],[11]) and maximum entropy [10]. Among these leaning methods, the most important issue is what features will be used to construct the classifier. It is common in WSD to use contextual information that can be found in the neighborhood of the ambiguous word in training data ([6], [16-18]). It is generally true that when

words are used in the same sense, they have similar context and co-occurrence information [13]. It is also generally true that the nearby context words of an ambiguous word give more effective patterns and features values than those far from it [12]. The existing methods consider features selection for context representation including both local and topical features where local features refer to the information pertained only to the given context and topical features are statistically obtained from a training corpus. Most of the recent works for English corpus including [7] and [8], which combine both local and topical information in order to improve their performance. An interesting study on feature selection for Chinese [10] has considered topical features as well as local collocational, syntactic, and semantic features using the maximum entropy model. In Dang’s [10] work, collocational features refer to the local PoS information and bi-gram co-occurrences of words within 2 positions of the ambiguous word. A useful result from this work based on (about one million words) the tagged People’s Daily News shows that adding more features from richer levels of linguistic information such as PoS tagging yielded no significant improvement (less than 1%) over using only the bi-gram co-occurrences information. Another similar study for Chinese [11] is based on the Naive Bayes classifier model which has taken into consideration PoS with position information and bi-gram templates in the local context. The system has a reported 60.40% in both precision and recall based on the SENSEVAL-3 Chinese training data. Even though in both approaches, statistically significant bi-gram co-occurrence information is used, they are not necessarily true collocations. For example, in the express “掌握/和/监视/本州/各地/新纳粹/分子/的/活动/情况”, the bi-grams in their system are (掌握, 和), (和, 监视), (监视, 本州), (本州, 各地), (各地, 新纳粹), (新纳粹, 的), (的, 活动), (活动, 情况). Some bi-grams such as (活动, 情况) may have higher frequency but may introduce noise when considering it as features in disambiguating the sense “human|人” and “symbol|符号” like in the example case of “水分子活动情况”. In our system, we do not rely on co-occurrence information. Instead, we utilize true collocation information (新纳粹, 分子) which fall in the window size of (-5, +5) as fea-

tures and the sense of “human|人” can be decided clearly using this features. The collocation information is a pre-prepared collocation list obtained from a collocation extraction system and verified with syntactic and semantic methods ([21], [24]).

Yarowsky [9] used the one sense per collocation property as an essential ingredient for an unsupervised Word-Sense Disambiguation algorithm to perform bootstrapping algorithm on a more general high-recall disambiguation. A few recent research works have begun to pay attention to collocation features on WSD. Domminic [19] used three different methods called bilingual method, collocation method and UMLS (Unified Medical Language System) relation based method to disambiguate unsupervised English and German medical documents. As expected, the collocation method achieved a good precision around 79% in English and 82% in German but a very low recall which is 3% in English and 1% in German. The low recall is due to the nature of UMLS where many collocations would almost never occur in natural text. To avoid this problem, we combine the contextual features in the target context with the pre-prepared collocations list to build our classifier.

3 The Classifier With Topical Contextual and Local Collocation Features

3.1 The Feature Set

As stated early, an important issue is what features will be used to construct the classifier in WSD. Early researches have proven that using lexical statistical information, such as bi-gram co-occurrences was sufficient to produce close to the best results [10] for Chinese WSD. Instead of including bi-gram features as part of discrimination features, in our system, we consider both topical contextual features as well as local collocation features. These features are extracted from the 60MB human sense-tagged People’s Daily News with segmentation information.

3.1.1 Topical Contextual Features

Niu [11] proved in his experiments that Naïve Bayes classifier achieved best disambiguation accuracy with small topical context window size (< 10 words). We follow their method and set

the contextual window size as 10 in our system. Each of the Chinese words except the stop words inside the window range will be considered as one topical feature. Their frequencies are calculated over the entire corpus with respect to each sense of an ambiguous word w . The sense definitions are obtained from HowNet.

3.1.2 Local Collocation Features

We chose collocations as the local features. A collocation is a recurrent and conventional fixed expression of words which holds syntactic and semantic relations [21]. Collocations can be classified as fully fixed collocations, fixed collocations, strong collocations and loose collocations. Fixed collocations means the appearance of one word implies the co-occurrence of another one such as “历史包袱” (“burden of history”), while strong collocations allows very limited substitution of the components, for example, “地方院校” (“local college”), or “地方大学” (“local university”). The sense of ambiguous words can be uniquely determined in these two types of collocations, therefore are the collocations applied in our system. The sources of the collocations will be explained in Section 4.1.

In both Niu [11] and Dang’s [10] work, topical features as well as the so called collocational features were used. However, as discussed in Section 2, they both used bi-gram co-occurrences as the additional local features. However, bi-gram co-occurrences only indicate statistical significance which may not actually satisfy the conceptual definition of collocations. Thus instead of using co-occurrences of bi-grams, we take the true bi-gram collocations extracted from our system and use this data to compare with bi-gram co-occurrences to test the usefulness of collocation for WSD. The local features in our system make use of the collocations using the template (w_i, w) within a window size of ten (where $i = \pm 5$). For example, “政府部门/ 和/ 地方/ 政府/ 认为” (“Government departments and local government commanded that”) fits the bi-gram collocation template (w, w_j) with the value of (地方, 政府). During the training and the testing processes, the counting of frequency value of the collocation feature will be increased by 1 if a collocation containing the ambiguous word occurs in a sentence. To have a good analysis on collocation features, we have also developed an algorithm using lonely adjacent bi-gram as locals features(named Sys-

adjacent bi-gram as locals features(named System A) and another using collocation as local features(named System B).

3.2 The Collocation Classifier

We consider all the features in the features set $F = F_t \cup F_l = \{f_1, f_2, \dots, f_m\}$ as independent, where F_t stands for the topical contextual features set, and F_l stands for the local collocation features set. For an ambiguous word w with n senses, let $S_w = \{w_{s1}, w_{s2}, \dots, w_{sn}\}$ be the sense set. For the contextual features, we directly apply the Naïve Bayes algorithm using Add-Lambda Smoothing to handle unknown words:

$$score_1(w_{si}) = \log p(w_{si}) + \sum_{f_j \in F_t} \log p(f_j | w_{si}) \quad (1)$$

For each sense w_{si} of an ambiguous word w :

$$p(w_{si}) = \frac{freq(w_{si})}{freq(w)} \quad (2)$$

For each contextual feature f_j respects to each sense w_{si} of w :

$$p(f_j | w_{si}) = \frac{freq(f_j, w_{si})}{\sum_{f_t \in F_t} freq(f_t, w_{si})} \quad (3)$$

To integrate the local collocation feature $f_j \in F_l$ with respect to each sense w_{si} of w , we use the follows formula:

$$score(w_{si}) = score_1(w_{si}) + \alpha \cdot score_2(w_{si}) \quad (4)$$

where α is tuned from experiments (Section 4.5), $score_1(w_{si})$ refers the score of the topical contextual features based on formula (1) and $score_2(w_{si})$ refers the score of collocation features with respect to the sense w_{sj} of w defined below.

$$score_2(w_{si}) = \sum_{f_j \in F_l} \delta(f_j | w_{sj}) \quad (5)$$

where $\delta(f_j | w_{sj}) = 1$ for $f_j \in F_l$ if the collocation occurs in the local context. Otherwise this term is set as 0.

Finally, we choose the right w_{sk} so that

$$s = \arg \max_{s_k} score(w_{sk}) \quad (6)$$

4 Experimental Results

We have designed a set of experiments to compare the classifier with and without the collocation features. In system **A**, the classifier is built with local bi-gram features and topical contextual features. The classifier in system **B** is constructed from combining the local collocation features with topical features.

4.1 Preparation the Data Set

We have selected 20 ambiguous words from nouns and verbs with the sense number as 4 in average. The sense definition is taken from HowNet [22]. To show the effect of the algorithm, we try to choose words with high degree of ambiguity, high frequency of use [23], and high frequency of constructing collocations. The selection of these 20 words is not completely random although within each criterion class we do try to pick word randomly.

Based on the 20 words, we extracted 28,000 sentences from the 60 MB People’s Daily News with segmentation information as our training/test set which is then manually sense-tagged. The collocation list is constructed from a combination of a digital collocation dictionary, a return result from a collocation automatic extraction system [21], and a hand collection from the People’s Daily News. As we stated early, the sense of ambiguous words in the fixed collocations and strong collocations can be decided uniquely although they are not unique in loose collocations. For example, the ambiguous word “面目” in the collocation “崭新的面目” may have both the sense of “appearance|外观” or “reputation|名声”. Therefore, when labeling the sense of collocations, we filter out the ones which cannot uniquely determine the sense of ambiguous words inside. However, this does not mean that loose collocations have no contribution in WSD classification. We simply reduce its weight when combining it with the contextual features compared with the fixed and strong collocations. The sense and collocation distribution over the 20 words on the training examples can be found in **Table 1**.

Table 1. Sense and Collocation Distribution of the 20 target words in the training corpus

Am.	T#	S1	S2	S3	S4	S5	S6
W		co#	co#	co#	co#	co#	co#

保管	31	1	30	NA				
		1	10					
保险	499	479	18	0	0	NA		
		324	0					
纪念	944	908	1	17	18	0	NA	
		129	1	10	0			
程序	409	3	389	17	NA			
		2	171	0				
包袱	110	3	101	6	NA			
		0	36	9				
结晶	41	3	37	1	NA			
		0	6	0				
关系	4885	26	34	72	4492	261	NA	
		0	0	0	1356	1		
精神	3508	7	48	3194	259	NA		
		0	4	1448	194			
接触	348	312	22	14	NA			
		117	11	4				
活动	4438	3983	33	123	153	102	44	
		721	10	37	123	23	5	
报告	1987	1712	274	NA				
		723	10					
攻击	83	36	47	00	NA			
		14	4					
程度	995	168	827	NA				
		108	513					
面目	31	11	20	NA				
		3	11					
基本	2725	227	498	102	1898	NA		
		1772	49	424	201			
把握	592	1	208	367	16	1	NA	
		0	63	124				
贷款	1155	756	399	NA				
		571	135					
决定	2792	691	1765	336	0	NA		
		98	113	29				
地方	2460	82	36	1231	877	NA		
		63	11	474	103			
意思	125	11	64	15	32	3	NA	
		0	0	3	4	0		

T#: total number of sentences contain the ambiguous word
s₁- s₆: sense no; co#: number of collocations in each sense

4.2 The Effect of Collocation Features

We recorded 6 trials with average precision over six-fold validation for each word. Their average precision for the six trials in the system A, and B can be found in **Table 2** and **Table 3**. From **Table 3**, regarding to precision, there are 16 words have improved and 4 words remained the same in the system B. The results from the both system confirmed that collocation features do improve the precision. Note that 4 words have the same precision in the two systems, which fall into two cases. In the first case, it can be seen that these words already have very high precision in the system A (over 93%) which means that one sense dominates all other senses. In this case, the additional collation information is not necessary. In fact, when we checked the intermediate outputs, the score of the candidate senses of the ambiguous words contained in the collocations get improved. Even though, it

would not change the result. Secondly, no collocation appeared in the sentences which are tagged incorrectly in the system A. This is confirmed when we check the error files. For example, the word “关系” with the sense as “亲疏” (“closeness”) appeared in 4492 examples over the total 4885 examples (91.9%). In the mean time, 99% of collocation in its collocation list has the same sense of “亲疏” (“closeness”). Only one collocation “关系户” has the sense of “势力” (“power”). Therefore, the collocation features improved the score of sense “亲疏” which is already the highest one based on the contextual features.

As can be seen from **Table 3**, the collocation features work well for the sparse data. For example, the word “保管” in the training corpus has only one example with the sense “人” (“human”), the other 30 examples all have the sense “管理” (“management”). Under this situation, the topical contextual features failed to identify the right sense for the only appearance of the sense “人” (“human”) in the training instance “有相应的粮食检验、保管专业人员...”. However, it can be correctly identified in the system B because the appearance of the collocation “保管专业人员”.

To well show the effect of collocations on the accuracy of classifier for the task of WSD, we also tested both systems on SENSEVAL-3 data set, and the result is recorded in the **Table 4**. From the difference in the relative improvement of both data sets, we can see that collocation features work well when the statistical model is not sufficiently built up such as from a small corpus like SENSEVAL-3. Actually, in this case, the training examples appear in the corpus only once or twice so that the parameters for such sparse training examples may not be accurate to forecast the test examples, which convinces us that collocation features are effective on handling sparse training data even for unknown words. **Fig. 1** shows the precision comparison in the system A, and B on SENVESAL-3.

Table 2. Average Precision (5/6 training, 1/6 test) of system A on People’s Daily News

Amb. W	T1	T2	T3	T4	T5	T6	Ave. Prec.
保管	1.00	1.00	1.00	1.00	1.00	.83	.972
保险	.90	.97	1.00	1.00	.97	.98	.972
纪念	.97	.96	.96	.92	.98	.96	.958

程序	.94	.94	.97	.92	.97	.97	.951
包袱	1.00	1.00	.77	.94	.88	1.00	.932
结晶	.83	1.00	1.00	1.00	.83	.90	.927
关系	.93	.95	.91	.92	.92	.92	.925
精神	.93	.94	.89	.91	.89	.90	.91
接触	.94	.93	.86	.93	.89	.87	.903
活动	.83	.94	.89	.90	.88	.94	.897
报告	.86	.88	.92	.84	.82	.87	.865
攻击	.92	.84	.92	.76	.84	.72	.833
程度	.84	.83	.88	.82	.88	.71	.827
面目	.80	.60	.80	.20	1.00	1.00	.733
基本	.68	.72	.67	.77	.70	.68	.703
地方	.51	.67	.47	.60	.68	.59	.586
决定	.70	.63	.66	.64	.64	.64	.652
把握	.57	.74	.55	.64	.72	.67	.648
贷款	.65	.58	.66	.64	.54	.47	.58
意思	.55	.50	.45	.45	.45	.64	.507
Total Average Precision							0.815

Table 3. Average Precision (5/6 training, 1/6 test) of system B on People’s Daily News

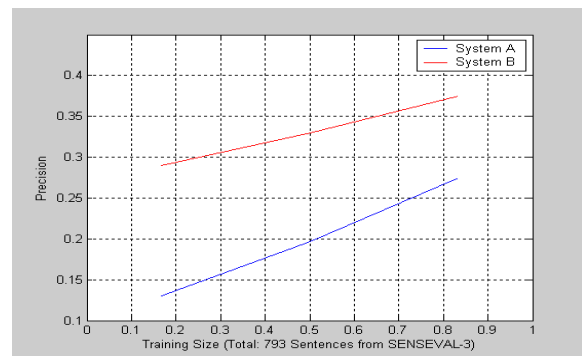
Amb. W	T1	T2	T3	T4	T5	T6	Ave. Prec.
保管	1.00	1.00	1.00	1.00	1.00	1.00	1.00
保险	.90	.97	1.00	1.00	.97	.98	.970
纪念	.96	.98	.97	.96	.98	.96	.968
程序	.94	.94	.97	.94	.97	.98	.957
包袱	1.00	1.00	.77	.94	.88	1.00	.931
结晶	.83	1.00	1.00	1.00	.83	.90	.927
关系	.93	.95	.91	.92	.92	.92	.925
精神	.92	.95	.92	.92	.91	.91	.922
接触	.94	.94	.86	.93	.91	.87	.908
活动	.80	.95	.89	.93	.89	.94	.902
报告	.87	.88	.92	.84	.83	.91	.875
攻击	.84	1.00	.92	.76	.84	.77	.855
程度	.88	.86	.89	.84	.90	.74	.852
面目	1.00	.80	.80	.20	1.00	1.00	.800
基本	.69	.72	.68	.79	.75	.72	.725
把握	.69	.76	.73	.74	.82	.79	.755
贷款	.58	.59	.70	.67	.64	.59	.628
决定	.68	.67	.66	.63	.65	.63	.653
地方	.65	.68	.71	.61	.70	.69	.673
意思	.60	.55	.54	.54	.54	.64	.568
Total Average Precision							0.840

Table 4. Average Precision of System A & B on SENSEVAL-3 Data Set

Amb. Word	Total S	Ave. Prec. in Sys A	Ave. Prec. in Sys B
日子	48	.207	.290
材料	20	.742	.742
走	49	.165	.325
坐	25	.325	.325
活动	36	.260	.373
研究	30	.167	.267
没有	30	.192	.392

分子	36	.635	.635
路	57	.238	.275
地方	36	.327	.385
把握	31	.100	.322
钱	40	.358	.442
起来	40	.308	.308
包	76	.110	.123
穿	28	.308	.475
突出	30	.500	.667
少	42	.165	.260
老	57	.037	.422
冲击	28	.833	.103
Total Ave. Precision			.376

Fig. 1. The precision comparison in system A, and B based on SENSEVAL-3

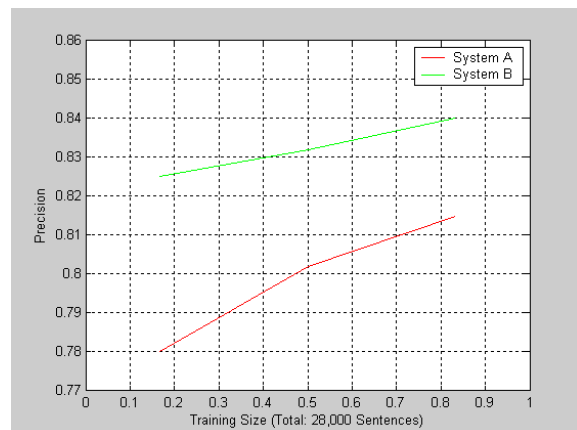


4.3 The Effect of Collocations on the Size of Training Corpus Needed

Hwee [21] stated that a large-scale, human sense-tagged corpus is critical for a supervised learning approach to achieve broad coverage and high accuracy WSD. He conducted a thorough study on the effect of training examples on the accuracy of supervised corpus based WSD. As the result showed, WSD accuracy continues to climb as the number of training examples increases. Similarly, we have tested the system A, and B with the different size of training corpus based on the PDN corpus we prepared. Our experiment results shown in **Fig 2** follow the same fact. The purpose we did the testing is that we hope to disclose the effect of collocations on the size of training corpus needed. From **Fig 2**, we can see by using the collocation features, the precision of the system B has increased slower along with the growth of training examples than the precision of the system A. The result is reasonable because with collocation feature, the statistical contextual information over the entire corpus becomes side effect. Actually, as can be seen from **Fig. 2**, after using collocation features

in the system B, even we use 1/6 corpus as training, the precision is still higher than we use 5/6 train corpus in the system A.

Fig. 2. The precision variation respect to the size of training corpus in system A, and B based on PDN corpus



4.4 Investigation of Sense Distribution on the Effect of Collocation Features

To investigate the sense distribution on the effect of collocation features, we selected the ambiguous words with the number of sense varied from 2 to 6. In each level of the sense number, the words are selected randomly. **Table 5** shows the effect of sense distribution on the effect of collocation features. From the table, we can see that the collocation features work well when the sense distribution is even for a particular ambiguous word under which case the classifier may get confused.

Table 5. The Effect of Sense Distribution on the Effect of collocation Features

Amb. word	Prec. Without coll	Prec. With coll	Sense #	Sense Distri.
保管	.972	1	2	97% *
保险	.97	.97	4	96% *
纪念	.957	.968	5	96% *
程序	.951	.957	3	95% *
包袱	.931	.931	3	92% *
结晶	.927	.927	3	90% *
关系	.925	.925	5	92% *
精神	.915	.922	4	91% *
接触	.903	.908	3	90% *
活动	.902	.902	6	90% *
报告	.865	.875	2	86% o
攻击	.833	.855	3	^
程度	.823	.852	2	83% o
面目	.733	.8	2	^

基本	.706	.725	4	^
把握	.65	.653	4	^
贷款	.618	.755	4	^
决定	.582	.628	2	^
地方	.563	.673	4	^
意思	.507	.568	5	^

*: over 90% samples fall in one dominate sense

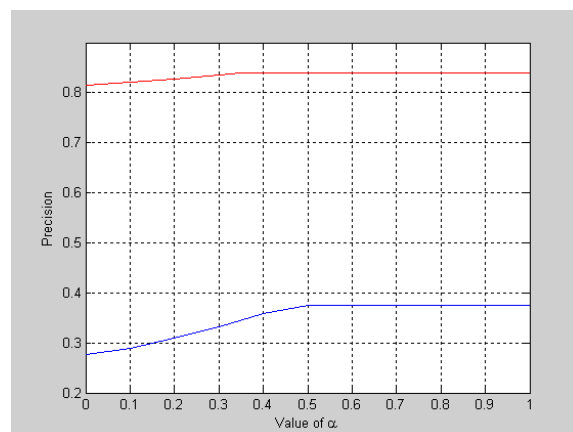
^: Even distribution over all senses

o: 83% to 86% samples fall in one dominate sense

4.5 The Test of α

We have conducted a set of experiments based on both the PDN corpus and SENSEVLA-3 data to set the best value of α for the formula (4) described in Section 3.2. The best start value of α is tested based on the precision rate which is shown in **Fig. 3**. It is shown from the experiment that α takes the start value of 0.5 for both corpora.

Fig. 3. The best value of α vs the precision rate



5 Conclusion and the Future Work

This paper reports a corpus-based Word Sense Disambiguation approach for Chinese word using local collocation features and topical contextual features. Compared with the base-line systems in which a Naïve Bayes classifier is constructed by combining the contextual features with the bi-gram features, the new system achieves 3% precision improvement in average in Peoples' Daily News corpus and 10% improvement in SENSEVAL-3 data set. Actually, it works very well when disambiguating the sense with sparse distribution over the entire corpus under which the statistic calculation prone to identify it incorrectly. In the same time, because disambiguating using collocation fea-

tures does not need statistical calculation, it makes contribution to reduce the size of human tagged corpus needed which is critical and time consuming in corpus based approach.

Because different types of collocations may play different roles in classifying the sense of an ambiguous word, we hope to extend this work by integrating collocations with different weight based on their types in the future, which may need a pre-processing job to categorize the collocations automatically.

6 Acknowledgements

We would like to present our thanks to the IR Laboratory in HIT University of China for sharing their sense number definition automatically extracted from HowNet with us.

References

1. Hwee Tou Ng, Bin Wang, Yee Seng Chan. Exploiting Parallel Texts for Word Sense Disambiguation. ACL-03 (2003)
2. Black E.: An experiment in computational discrimination of English word senses. IBM Journal of Research and Development, v.32, n.2, (1988) 185-194
3. Gale, W. A., Church, K. W. and Yarowsky, D.: A method for disambiguating word senses in a large corpus. Computers and the Humanities, v.26, (1993) 415-439
4. Leacock, C., Towell, G. and Voorhees, E. M.: Corpus-based statistical sense resolution. In Proceedings of the ARPA Human Languages Technology Workshop (1993)
5. Leacock, C., Chodorow, M., & Miller G. A. Using Corpus Statistics and WordNet Relations for Sense Identification. Computational Linguistics, 24:1, (1998) 147-165
6. Schütze, H.: Automatic word sense discrimination. Computational Linguistics, v.24, n.1, (1998) 97-124
7. Towell, G. and Voorhees, E. M.: Disambiguating highly ambiguous words. Computational Linguistics, v.24, n.1, (1998) 125-146
8. Yarowsky, D.: Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French. In Proceedings of the Annual Meeting of the Association for Computational Linguistics, (1994) 88-95
9. Yarowsky, D.: Unsupervised word sense disambiguation rivaling supervised methods. In Proceedings of the Annual Meeting of the Association for Computational Linguistics, (1995)189-196
10. Dang, H. T., Chia, C. Y., Palmer M., & Chiou, F.D., Simple Features for Chinese Word Sense Disambiguation. In Proc. of COLING (2002)
11. Zheng-Yu Niu, Dong-Hong Ji, Chew Lim Tan, Optimizing Feature Set for Chinese Word Sense Disambiguation. To appear in Proceedings of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (SENSEVAL-3). Barcelona, Spain (2004)
12. Chen, Jen Nan and Jason S. Chang, A Concept-based Adaptive Approach to Word Sense Disambiguation, Proceedings of 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational linguistics. COLING/ACL-98 (1998) 237-243
13. Rigau, G., J. Atserias and E. Agirre, Combining Unsupervised Lexical Knowledge Methods for Word Sense Disambiguation, Proceedings of joint 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL/EACL'97), Madrid, Spain (1997)
14. Jong-Hoon Oh, and Key-Sun Choi, C02-1098.: Word Sense Disambiguation using Static and Dynamic Sense Vectors. COLING (2002)
15. Yarowsky, D., Hierarchical Decision Lists for Word Sense Disambiguation. Computers and the Humanities, 34(1-2), (2000) 179-186
16. Agirre, E. and G. Rigau (1996) Word Sense Disambiguation using Conceptual Density, Proceedings of 16th International Conference on Computational Linguistics. Copenhagen, Denmark, COLING (1996)
17. Escudero, G., L. Márquez and G. Rigau, Boosting Applied to Word Sense Disambiguation. Proceedings of the 11th European Conference on Machine Learning (ECML 2000) Barcelona, Spain. 2000. Lecture Notes in Artificial Intelligence 1810. R. L. de Mántaras and E. Plaza (Eds.). Springer Verlag (2000)
18. Gruber, T. R., Subject-Dependent Co-occurrence and Word Sense Disambiguation. Proceedings of 29th Annual Meeting of the Association for Computational Linguistics (1991)
19. Dominic Widdows, Stanley Peters, Scott Cederberg, Chiu-Ki Chan, Diana Steffen, Paul Buitelaar, Unsupervised Monolingual and Bilingual Word-Sense Disambiguation of Medical Documents using UMLS. Appeared in Natural Language Processing in Biomedicine., ACL 2003 Workshop, Sapporo, Japan (2003) 9-16
20. Hwee Tou Ng., Getting serious about word sense disambiguation. In Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How? (1997) 1-7
21. Ruifeng Xu , Qin Lu, and Yin Li, An automatic Chinese Collocation Extraction Algorithm Based On Lexical Statistics. In Proceedings of the NLPKE Workshop (2003)
22. D. Dong and Q. Dong, HowNet. <http://www.keenage.com>, (1991)
23. Chih-Hao Tsai, <http://technology.chtsai.org/wordlist/>, (1995-2004)
24. Q. Lu, Y. Li, and R. F. Xu, Improving Xtract for Chinese Collocation Extraction. Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering, Beijing (2003)