Semi-Supervised WSD in Selectional Preferences

with Semantic Redundancy

Xuri TANG[1,5] , Xiaohe CHEN[1] , Weiguang QU[2,3] and Shiwen YU[4]

1. School of Chinese Language and Literature, Nanjing Normal University
{xrtang,chenxiaohe5209}@126.com
2. Jiangsu Research Center of Information Security & Privacy Technology
3. School of Computer Science, Nanjing Normal University
wgqu_nj@163.com
4. Institute of Computational Linguistics , Peking University
yusw@pku.edu.cn
5. College of Foreign Studies, Wuhan Textile University

Abstract

This paper proposes a semi-supervised approach for WSD in Word-Class based selectional preferences. The approach exploits syntagmatic and paradigmatic semantic redundancy in the semantic system and uses association computation and minimum description length for the task of WSD. Experiments on Predicate-Object collocations and Subject-Predicate collocations with polysemous predicates in Chinese show that the proposed approach achieves a precision which is 8% higher than the semantic-association based baseline. The semi-supervised nature of the approach makes it promising for constructing large scale selectional preference knowledge base.

1 Introduction

This paper addresses word sense disambiguation (WSD) which is required in the construction of selectional preference (SP) knowledge database. In previous literature of SP, four different types of formalization models are explicitly or implicitly employed. Two types are distinguished in Li and Abe(1998):

Word Model: $P(n \mid v, r) = \sigma$ (1)

Class Model: $P(C \mid v, r) = \sigma$ (2)

where v stands for verb, n for noun, C for the semantic class of n, r for the grammatical relation between v and n, and P for the preference strength. Most of the researches(Resnik 1996; Li and Abe 1998; Ciaramita and Johnson 2000; Brockmann and Lapata 2003; Light and Greiff 2002) uses the class model, and a few(Erk 2007) uses the word model. The other two types of model are given as below:

Class-Only Model: $P(C_n \mid C_v, r) = \sigma$ (3)

Word-Class Model: $P(n, C_n \mid v, C_v, r) = \sigma$ (4)

where C_n , C_v are semantic classes for the noun and verb respectively. Class-Only model considers solely the semantic classes, while Word-Class model considers both words and semantic classes. Agirre and Martinez(2001) and Zheng et al(2007) adopted the Class-only Model in research, while in McCarthy and Carroll(2003) and Merlo and Stevenson(2001) the Word-Class Model is employed.

Among the four models, the Word-Class Model is the type which possesses the most granulated knowledge and is the most potential in applications. McCarthy and Carroll(2003) reports that the Word-Class Model performs well in unsupervised WSD. In other NLP tasks such as metaphor recognition, this model may be indispensable. For instance, to distinguish the predicate verb "浮动(float)" in Ex(1a) as

```
Ex. 1
a. 树 叶 浮 动    b. 价 格 浮 动
   leaf floats        price floats
```

literal and Ex(1b) as metaphorical requires different interpretations of the verb.

The present research is concerned with WSD as in the Word-Class model. Particularly, it aims at disambiguating predicates in subject-predicate (Subj-Pred) and predicate-object (Pred-Obj) constructions. The motivations behind the research are two folds. Firstly, semi-supervised and unsupervised WSD in SP are not fully explored. Merlo and Stevenson(Merlo and Stevenson 2001) employs supervised learning from large annotated corpus, which is difficult to obtain. One known unsupervised learning approach for WSD in SP is McCarthy and Carroll(2003) which addresses the issue via conditional probability. The other motivation derives from the fact few research is done on selectional preferences in languages other than English, as is stated in Brockmann and Lapata(2003). For instance, studies on construction of SP knowledge database in Chinese can only be found in Wu et al(2005), Zhen et al(2007), Jia and Yu(2008) and some others.

The basic idea of the approach proposed for WSD in the paper is that the most acceptable interpretation of senses for a given construction is the pair of senses which encodes the most redundant information in the semantic system of the language. Two principles, namely Syntagmatic Redundancy Principle and Paradigmatic Redundancy Principle, are proposed in the paper to capture the intuition. Two corresponding devices are employed to model the two principles: Association for Syntagmatic Redundancy Principle and Minimum Description Length for Paradigmatic Redundancy Principle. Two experiments are conducted in the paper. The first is based on semantic association, achieving a 61.98% precision for predicates in Subj-Preds and 62.54% in Pred-Objs. This experiment is used as baseline as the approach is also used in McCarthy and Carroll(2003) for verb and adjective disambiguation. In the second experiment, both semantic association and MDL are employed, the precision of WSD amounts to 69.88% and 69.09% for predicates in Subj-Preds and Pred-Objs respectively, indicating that a combination of the two devices are fairly effective in disambiguating word senses for SP.

The rest of the paper is organized as below. The second part gives further illustration of the rationale for the approach. The third part describes the procedure and the fourth part discusses the experiment result. The thesis concludes with some speculations in further researches.

2 Rationale

2.1 Task Formalization

Consider a Subj-Pred or Pred-Obj collocation C=$<W_{pred}, W_{arg}>$, where W_{pred} is the word of predicate and W_{arg} is the word of argument. W_{pred} has M senses, denoted by set S_{pred}. W_{arg} has N senses, denoted by S_{arg}. The possible interpretation of C has M*N possibilities, denoted by

$$S_C = S_{pred} \times S_{arg} = \{ \varsigma_j^i \mid \varsigma_j^i = <s_{pred}^i, s_{arg}^j> \},$$

where ς_j^i is called a sense collocation. The task of WSD is to search for a particular sense collocation in S_C and assign it to C as its interpretation. At the initial stage, each sense collocation in S_C is considered to have an even number of frequency, namely $f(\varsigma_j^i) = 1/(N \times M)$. Accordingly, for each $s_{pred}^i \in S_{pred}$, $f(s_{pred}^i) = 1/M$, For each $s_{arg}^i \in S_{arg}$, $f(s_{arg}^i) = 1/N$.

2.2 Syntagmatic Redundancy Principle

Syntagmatic Redundancy Principle (SRP) can be stated as following: among all possible sense collocations for a word collocation, the most appropriate is the one in which senses exhibit the most redundant information between each other.

The syntagmatic redundancy between words has been noticed very early by linguists and has been applied in WSD. Firth(1957) argues that there exists "mutual expectancy" between words in collocations, and the meaning of word is partially encoded in its juxtaposition. Lyons(1977:261) comments that Porzig has noticed in 1934 the "essential meaning relation" between words of collocations like "dog barks" and "tree fells"

and emphasizes that the meanings of collocationally restricted lexemes such as "bark" and "fell" can only be explained by taking into account the collocates they occur with. This notion is also employed in Yarowsky(1995) for WSD, in which the key is the "one-sense-per-collocation" statement. McCarthy and Carroll(2003) also uses this type of redundancy for disambiguation in SP.

SRP can be explained as a statistic correlation between s_{pred} and s_{arg}. The more co-relevant these two senses are, the more likely the pair is to be accepted as the appropriate interpretation. This can be described as below:

$$\overline{\varsigma_j^i} = \arg\max Assoc(s_{pred}^i, s_{arg}^j) \qquad (5)$$

where $Assoc(s_{pred}^i, s_{arg}^j)$ is the function for sense association. Four methods can be considered for association computation: conditional probability (Formula 6 and 7), Lift(Han and Kamber 2006:261) (Formula 8), All-Confidence(Han and Kamber 2006:263) (Formula 9) and cosine (Formula 10). Note that two versions of conditional probability are considered, as are denoted in Formula 6 and 7. The first version, Cond-Prob 1, takes argument sense as condition, while the second version Cond-Prob 2 takes predicate sense as condition.

$$p(s_{pred}^i \mid s_{arg}^j) = \frac{p(s_{pred}^j, s_{arg}^i)}{p(s_{arg}^j)} \qquad (6)$$

$$p(s_{arg}^i \mid s_{pred}^j) = \frac{p(s_{pred}^j, s_{arg}^i)}{p(s_{pred}^i)} \qquad (7)$$

$$lift(s_{pred}^i, s_{arg}^j) = \frac{p(s_{pred}^i, s_{arg}^j)}{p(s_{pred}^i) * p(s_{arg}^j)} \qquad (8)$$

$$all\_conf(s_{pred}^i, s_{arg}^j) = \frac{f(s_{pred}^i, s_{arg}^j)}{\max(f(s_{arg}^i), f(c_{pred}^j))} \qquad (9)$$

$$cosine(s_{pred}^i, s_{arg}^j) = \frac{p(s_{pred}^i, s_{arg}^j)}{\sqrt{p(s_{pred}^i) * p(s_{arg}^j)}} \qquad (10)$$

2.3 Paradigmatic Redundancy Principle

Paradigmatic Redundancy Principle (PRP) can be stated as following: among all possible sense collocations for a word collocation, the most appropriate is the one which is also implicitly or explicitly expressed by other synonymous, metonymic or metaphorical word collocations.

Ex(2) illustrates the explicit redundancy in synonymous and metaphorical ways, in which the sense collocation "[Price|价格] [QuantityChange|量变]" is expressed by five word collocations, each with a different predicate：变化(change), 浮动(float), 调整 (adjust), 起伏(go up and down), 变动(alter).

Ex 2.
a. 价格变化 price changes　b. 价格浮动 price floats　c. 价格调整 price adjusts
d. 价格起落 price goes up and down　e. 价格变动 price alters

Ex(3) reveals the implicit redundancy in metonymic way, in which the meaning "人 (human) 安心 (is eased)" is implicitly expressed in all the six collocations, established by semantic relatedness among the arguments "马拉多纳(Maradona)", "学生 (student)", "工作(work)", "劳动(labour)", "行车(driving)", and "生活(life)".

Ex 3.
a. 马拉多纳 安心 了 Maradona is eased　b. 学生 安心 了 Student is eased
c. 工作 安心 了 work is eased　d. 劳动 安心 了 labour is eased
e. 行车 安心 了 driving is eased　f. 生活 安心 了 Life is eased

To apply PRP, WSD in SP is casted as an issue of model selection. Given a set of word collocations Θ, the process of WSD is to assign to each word collocation one sense collocation from a number of possibilities. Those assigned sense collocations form a set, or a model for Θ. The goal of WSD in SP is to select from all those models the one which best interprets Θ. For this purpose, Miminum Description Length(Barron et al. 1998; Michell 2003; MacKay 2003) can be used. MDL selects models by relying on induction bias based on Occam's Razor, which stipulates that the simplest solution is usually the correct one. One way to interpret MDL in Bays' analysis is as below(Michell 2003:124):

$$m' = \arg\min L_M(m) + L_D(D \mid m) \qquad (11)$$

In (11)，$L_M(m)$ is the model description length when model m is considered, $L_D(D \mid m)$ is the data description length when model m is used for description. The model with minimum length is the best model.

For model description length, we have adopted the method used in (Li and Abe 1998) which considers only the size of the model:

$$L_M(m) = \frac{size(m) - 1}{2} \log(N) \qquad (12)$$

where size(m) is the number of sense collocation contained in model m, and N is the number of word collocation in consideration. In this study, the set of word collocation with the same predicate word, denoted by Θ, is used as the unit for model description length calculation instead of the whole corpus, so as to reduce computation complexity. Accordingly, each word collocation in Θ can be assigned one and only one sense collocation in the model m, out of all the potential sense collocations as is explained in section 2.1.

Data description length is calculated on model m and Θ, as is denoted in formulas (13), (14) and (15) below. The calculation is

$$L(\Theta \mid m) = -\sum \log(p(\varsigma_j^i)) = -\sum \log\left(\frac{\overline{f}(\varsigma_j^i)}{Num_\Theta^2}\right) \qquad (13)$$

$$\overline{f}(\varsigma_j^i) = f(\varsigma_j^i) + f(\varsigma_j^i) * \sum_{\varsigma_j^i, \varsigma_l^i \in m} w(\varsigma_j^i, \varsigma_l^k) \qquad (14)$$

$$
\begin{aligned}
w(\varsigma_j^i, \varsigma_l^k) &= rel(< s_{pred}^i, s_{arg}^j >, < s_{pred}^k, s_{arg}^l >) \\
&= \begin{cases} rel(s_{arg}^j, s_{arg}^l) & \text{if } s_{pred}^i = s_{pred}^k \\ 0 & \text{if } s_{pred}^i \neq s_{pred}^k \end{cases}
\end{aligned} \qquad (15)
$$

based on the probability of sense collocation $\varsigma_j^i = < s_{pred}^i, s_{arg}^j >$, which in turn is calculated on a modified frequency of the collocation $\overline{f}(\varsigma_j^i)$. The frequency is modified by counting the explicit occurrence of the sense collocation itself and the implicit occurrence expressed by other sense collocations in Θ. This idea is equivalent to enlarge the corpus by 1 fold, thus the overall collocation number is the two times of the original number.

The modified frequency is a sum of two parts, denoted in formula (14). The first part is $f(\varsigma_j^i)$, the frequency of ς_j^i. The second part is the weighted frequency of ς_j^i. The weight is determined by the relatedness of the sense collocation ς_j^i and all the other sense collocation ς_l^k in the model m. According to this formula, if the sense collocation is found to be more similar to other sense collocations, it should obtain a higher modified frequency,

and thus more likely to be the correct one for the word collocation.

The way to calculate the weight is given in formula (15). If two sense collocations have identical predicate sense, namely $s_{pred}^i = s_{pred}^k$, then the weight between the two sense collocations is measured by $rel(s_{arg}^j, s_{arg}^l)$, the semantic relatedness between the argument sense s_{arg}^j and s_{arg}^l. Otherwise, 0 is returned. There are different ways to measure sense relatedness. The present study has used semantic similarity based on HowNet(Liu and Li 2002) to calculate the semantic relatedness.

3   Procedure

Figure 1 maps out the procedure for WSD in SP in the present study. The procedure is divided into two phases: data collection and disambiguation. The collocation data are collected from three sources: Sketch Engine, Collocation Dictionary and HowNet Examples. Two types of collocation data are collected: subject-predicate collocations (Subj-Pred) and predicate-object collocations (Pred-Obj) from Sketch Engine and Collocation Dictionary. Collocation Retriever reduces HowNet examples into Subj-Preds and Pred-Objs using simple heuristic methods. As a result, about 70,000 subject-predicate collocations and 106,000 predicate-object collocations are obtained.
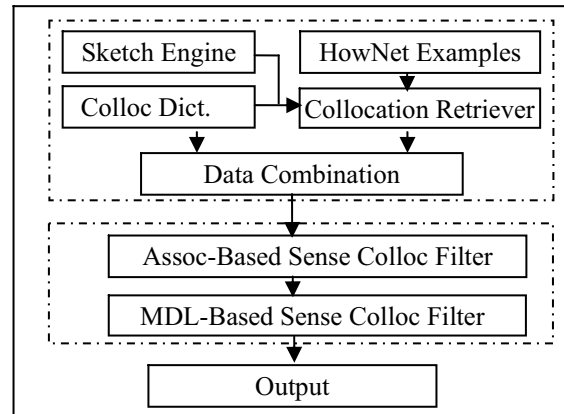

Figure 1. WSD Procedure

In disambiguation phase, two devices are employed to filter out unlikely sense collocations: Association-Based Sense Collocation Filter, following SRP, and MDL-Based Sense Collocation Filter, following PRP.

In this phase, Subj-Preds and Pred-Objs are processed independently but following the same route.

Each phase alone can perform WSD independently. Accordingly, two experiments are conducted to evaluate the method proposed in this paper. The first experiment uses association-based filter for word sense disambiguation, which is also used as the baseline. The approach is also used in (McCarthy and Carroll 2003) to disambiguate verbs and adjectives in collocations. To be particular, the method used by McCarthy and Carroll(2003) is formula (6). The second experiment is based on the result of the first one so as to observe the improvement obtained by MDL-Based approach. In the second experiment, unsupervised and semi-supervised WSD are also investigated by including some annotated collocations in the evaluation data.

Two corpora are constructed for evaluation. One corpus is a set of 1034 subject-predicate constructions. The other is a set of 1841 predicate-object constructions. Both are manually annotated by the authors with sense definitions defined in HowNet(Dong 2006). All together there are 52 highly ambiguous predicates involved in the study.

4 Experiments and Discussion

4.1 Collocation Retriever

The major task in data collocation is in Collocation Retriever, which retrieves collocations from HowNet examples. Ex(4) gives a partial entry structure in HowNet,

```
Ex 4.
W_C=浮动
E_C=人心~，开始~，工作指标不停
地~
DEF=[change|变]
```

in which W_C stands for Chinese Word, DEF for definition, E_C for Examples of Chinese, and the wave "~" for the word in question. From E_C, possible Subj-Preds such as "人心 (public opinion) 浮动(floats)", "指标(index) 浮动(floats)" can be retrieved, in which the sense of "浮动(float)" is annotated with DEF. But there are also noises. A simple heuristic method is applied to automatically filter out unwanted collocations. The heuristic method

checks whether the collocation retrieved from HowNet share possible sense collocations with collocations in Collocation Dictionary. If yes, it is accepted as a collocation of the type, otherwise, it is rejected. Procedures are given below:

(a) Use Subj-Pred collocations and Pred-Obj collocations in Collocation Dictionary to build sense collocation set $\Gamma_{Subj-Pred}$ and $\Gamma_{Pred-Obj}$;

(b) For each example sentence in E_C, segment it using ICTCLAS[1] to obtain an array of words. Words before "~" forms potential Subj-Pred collocations $A_{Subj-Pred}$ and Words after form potential Pred-Obj collocations $B_{Pred-Obj}$.

(c) For each $a \in A_{Subj-Pred}$ or $b \in B_{Subj-Pred}$, construct possible sense collocation set Γ_a or Γ_b, if $\Gamma_a \cap \Gamma_{Subj-Pred} \neq \phi$ or $\Gamma_b \cap \Gamma_{Pred-Obj} \neq \phi$, add it as a Subj-Pred collocation or Pred-Obj collocation.

Evaluation on partial retrieved collocations shows that about 70% of obtained collocations are valid collocations, while about 30% are errors. Thus manual edition has been applied to rid those invalid collocations.

4.2 Association-Based Filter

Association-Based Sense Collocation Filter filters out those sense collocations that are very unlikely to be the right interpretation for a word collocation. Table 1 gives association computation result for the six senses related to the predicate "粗 (rough)" in Subj-Pred collocation "性格 (personality) 粗 (rough)". The 2nd, 3rd, 4th, and 6th are very unlikely interpretations and should be filtered, while the 5th seems to be the most appropriate.

Table 1. Association-Based Filter Example

No.	Pred Sense	Arg Sense	Assoc. Dgr
1	[Behavior|举止]	[careless|粗心]	0.0019
2	[Behavior|举止]	[coarse|糙]	0.0002
3	[Behavior|举止]	[hoarse|沙哑]	0.0004
4	[Behavior|举止]	[roughly|大概]	0.0002
5	[Behavior|举止]	[vulgar|俗]	0.0071
6	[Behavior|举止]	[widediameter|粗]	0.0002

Following the procedure in Figure 1, to filter out those unlikely sense collocations, average

association value is used as the filter and those below the average are dropped and those above are chosen for MDL-Based Filter. In Table 1, the average is 0.0017, and the 1rd and 5th are chosen.

However, in order to obtain a baseline and to decide which association computation model to use, we have followed the definition in Formula 5 and perform WSD test by choosing the sense collocation with highest association as the correct sense tags. for used this step solely for WSD, as is defined in Formula 4. Table 2 gives the experiment results for Subj-Pred and Pred-Obj collocations with all the association computation models denoted in Formula 6-10.

Table 2. WSD Result by Association

	Subj-Pred(%)	Pred-Obj(%)
Cond-Prob 1	61.98	62.54
Cond-Prob 2	55.15	42.4
Lift	63.09	40.84
All_Conf	56.16	48.54
Cosine	58.83	55.72

One interesting phenomenon about all the five models is null-invariance. In selecting models for association computation, null-invariance is an important feature to be considered(Han and Kamber 2006). A model with null-invariance is not influenced by additional irrelevant data and thus is more stable. In the experiment, the model Lift is the only one not featured with null-invariance. The experiments show that Lift is not stable in different collocation types, achieving high precision in Subj-Pred but low precision in Pred_Obj.

A second interesting phenomenon is collocation directionality exposed by the experiments, which can be observed in the two models of conditional probability: Cond-Prob 1, with argument as condition, and Cond-Prob 2, with predicate as condition. Directionality in collocation has been noticed earlier in some researches, for example Qu(2008). Our experiment shows that when using Cond-Prob 1, we are able to get a precision of 61.98% and 62.54% for Subj-Pred and Pred-Obj respectively, while Cond-Prob 2 gets a much lower precision. This fact can be interpreted that arguments tend to have a stronger selectional preference strength, and the possible selection range is comparatively narrower, while predicates have weaker selectional preference strength and a wider selectional range.

4.3   MDL-Based Filter

MDL-Based Filter takes as input result from Association-Based Filter using Cond-Prob 1 for association computation and average association as filter. Table 3 and 4 give the final experiment outcome for Pred-Obj and Subj-Pred constructions and individual predicates.

It can be seen in Table 3 that MDL-Based Filter Several inferences can be made from the experiments. Firstly, comparison between Association-Based WSD (Table 2) and MDL WSD (Table 3) shows that MDL can improve overall performance up to 8%. As is mentioned earlier, Association-Based WSD is used as baseline in the present study. Given the fact that the average number of senses for word in question is fairly high, the improvement is considered as significant.

Table 3. General WSD Results[2]

	Ave. N.O.S.	Assoc. WSD (%)	MDL WSD (%)
Subj-Pred	4.16	61.98	69.09
Pred-Obj	5.03	62.54	69.88

Analysis on the individual predicates in Table 4 gives a clearer picture of WDL-based WSD. Firstly, it can be seen that MDL is especially effective when the demarcation of word senses is clear-cut. Predicate words such as "安静 (quiet)", "肮脏 (dirty)", "困难 (difficult)" in Subj-Preds and "锤炼(beat)", "触动(touch)" and "打断(break)" in Pred-Objs are successfully disambiguated in Table 4. These words generally have 2 or 3 senses, and the senses generally differ in terms of abstractness and concreteness, as is indicated in table 5. This is due to the fact that the arguments in these collocations are clearly delimitated in HowNet and this delimitation is well captured by the modified frequency calculation defined in formula (14). Via the formula, the concrete sense collocations can

---

Table 4. Detailed WSD Experiment Results

Results for Pred-Obj.				Results for Subj-Pred.			
Pred.	N.O.S	Assoc. WSD (%)	MDL WSD (%)	Pred.	N.O.S.	Assoc. WSD (%)	MDL WSD (%)
追(v)	5	69.23	80.77	困难(a)	2	61.14	92.00
上(v)	14	70.59	70.59	垮(v)	2	72.73	86.36
交(v)	6	56.25	90.62	细腻(a)	2	47.83	58.7
介绍(v)	3	72.22	88.89	错(a/v)	5	52.17	78.26
倒(v)	9	50	60.53	肮脏(a)	3	56.76	81.08
停(v)	8	86.67	93.33	浓(a)	5	40	40
套(v)	5	68.75	62.5	破裂(v)	2	55.17	41.38
展开(v)	3	73.91	81.16	安静(a)	3	75.76	93.94
开(v)	17	55.93	44.07	沉(a)	4	96.3	66.67
把握(v)	3	80.36	78.57	沉闷(a)	3	47.37	42.11
打断(v)	2	66.67	92.31	淡(a)	6	88.24	88.24
洗刷(v)	2	57.14	80.95	低(a)	6	46	60
挨(v)	6	76.27	79.66	开发(v)	3	44.44	44.44
锤炼(v)	3	83.33	100	开阔(a)	2	38.46	65.38
磨(v)	8	63.64	63.64	麻木(a)	2	93.33	53.33
排(v)	3	77.14	80	漂浮(v)	3	85.19	88.89
触动(v)	2	88.24	100	轻(a)	10	50	50
散布(v)	2	83.87	80.65	陷落(v)	2	60.53	63.16
看(v)	9	61.84	68.42	长(a/v)	9	39.66	53.45
陷入(v)	3	40.28	51.39	粗(a)	6	59.46	51.35
带(v)	4	48.08	53.85	掉(v)	6	48.72	74.36
破坏(v)	3	73.49	73.49	动摇(v)	3	48.15	44.44
恢复(v)	2	15.32	40	黑暗(a)	2	88.57	57.14
糟蹋(v)	2	84.91	83.02	厚(a)	6	68.18	40.91
找(v)	3	86.54	85.58	坏(v)	8	52.03	65.04
爱(v)	4	72.51	72.99	结实(a)	2	95.35	95.35

Table 5. Word Sense Distinction

Pred	Concrete Sense	Abstract Sense(s)
安静	[quiet|静]	[calm|镇静], [peaceful|宁]
肮脏	[dirty|龌]	[despicable|卑劣], [immoral|不道德]
困难	[difficult|难]	[poor|穷]
锤炼	[beat|打]	[MakeBetter|优化], [cultivate|培养]
触动	[touch|触]	[excite|感动]
打断	[break|折断]	[obstruct|阻止]

increase the modified frequency of concrete sense collocations, and the abstract sense collocation can increase the modified frequency of abstract sense collocations, thus

leading to the clear demarcation of abstract senses and concrete senses.

The role of semantic relevance can also be clearly noticed in the predicates which have a decreased precision in MDL in Table 4. Via Paradigmatic Redundancy Principle, the information encoded in one collocation are diffused to other collocations. Consequently, errors can be diffused. This explains why the precisions of some predicates such as "沉 (sink)", "麻木(dumb)", "黑暗(dark)" in Subj-Pred and "开(open)", "套(harness)" in Pred-Objs decrease after MDL. Further analysis shows that this is because MDL has diffused the errors produced by Association Filter. For instance, at Association Filter phase, the collocation "箱子(box) 沉(sink)" is assigned with the only sense collocation "[tool|用具] [very| 很]" and all other potential sense collocations are filtered. When MDL is applied, other collocations such as "机器(machine) 沉 (heavy)", "镐头(pick) 沉(heavy)", "镢头(chaw) 沉(heavy)", "篮子(basket) 沉(heavy)", "盒子 (box) 沉(heavy)", "家具(furniture) 沉(heavy)", in which the arguments are tightly correlated with that of "箱子(box) 沉(sink)"，all takes the sense "[very| 很]", thus leading to the decrease of precision.

The diffusion of senses can also best seen in the comparison between those predicates whose WSD are semi-supervised and those whose WSD are not supervised. Some predicates have collocations successfully retrieved from HowNet examples in which the word sense is already identified. These collocations are diffused in MDL filtering and play important roles in improving precision, while some other predicates do not have such resource. In Table 4, those unsupervised predicates are "陷落(fall)", "垮(collapse)", "细腻(exquisite)", "麻木(dumb)", "开阔(wide)", "开发(develop)" in Subj-Preds and "展开(spread)", "洗刷(brush)", "陷入(get into)", "带(bring)", and "糟蹋(mar)" in Pred-Objs. The other predicates are semi-supervised. As can be seen in Table 4, most of these unsupervised predicates generally have a precision of 40%-60%, while those semi-supervised predicates enjoy are much higher precision between 50%-100%. The explanation

for the result is straight forward. When one sense collocation of one word collocation is correctly identified, by way of Paradigmatic Redundancy Principle, the sense collocation which is similar to the correctly identified will have a higher modified frequency and is thus singled out as the best choice. This feature of MDL has great significance in the process of annotating large scale collocation data. With only a small number of annotated collocations for each predicate, a fairly high precision can be achieved for all the rest of the data through MDL.

5 Conclusion

The present paper believes that the Word-Class Model gives the fullest description for selectional preference and thus makes efforts to disambiguate predicates in selectional preferences. From the perspective of semantic system, two principles of semantic redundancy, namely the Syntagmatic Redundancy Principle and Paradigmatic Redundancy Principle, are proposed in the paper and are applied in WSD in SP via Association Computation and Minimum Description Length. The experiments show that the approach proposed is fairly encouraging in disambiguation of polysemous predicates, especially under semi-supervised conditions when a small portion of data is annotated. With such a tool, we are able to build large scale selectional preference knowledge database based on Word-Class Models, which can be applied in various tasks, of which metaphor recognition is the particular one we bear in mind.

Acknowledgement

References

Agirre, E., and D. Martínez. 2001. Learning class-to-class selectional preferences. Paper read at Proceedings of the Conference on Natural Language Learning, at Toulouse, France.

Barron, A. R., J. Rissanen, and B. Yu. 1998. The Minimum Description Length Principle in coding and modeling. IEEE Transactions on Information Theory 44 (6):2743-2760.

Brockmann, C., and M. Lapata. 2003. Evaluating and combining approaches to selectional preference acquisition. Paper read at Proceedings of the European Association for Computational Linguistics, at Budapest, Hungary.

Ciaramita, M., and M. Johnson. 2000. Explaining away ambiguity: Learning verb selectional preference with Bayesian networks. In Proceedingsofthe18thInternationalConferenceon ComputationalLinguistics (COLING 2000), 187-193.

Dong, Z. 2006. HowNet and the Computation of Meaning. River Edge, NJ: World Scientific.

Erk, K. 2007. A Simple, Similarity-based Model for Selectional Preferences. Paper read at Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, at Prague, Czech Republic.

Firth, J. R. 1957. A Synopsis of Linguistic Theory, 1930-1955. In Studies in Linguistic Analysis. Oxford: Blackwell, 1-32.

Han, J., and M. Kamber. 2006. Data Ming: Concepts and Techniques. Singapore: Elsevier.

Jia Yuxiang and Yu Shiwen. 2008. Automatic Acquisition of Selectional Preference and Its Application to Metaphor Processing. Paper read at the Fourth National Student Conference on Computationl Linguistics, at Taiyuan, Shangxi, China.

Li, H., and N. Abe. 1998. Generalizing Case Frames Using a Thesaurus and the MDL Principle. Computational Linguistics 24 (2):217-244.

Light, M., and W. Greiff. 2002. Statistical models for the induction and use of selectional preferences. Cognitive Science 87:1-13.

Liu, Qun and Li Sujian. 2002. Word Similarity Computation Based on HowNet. In Proceedings of the 3rd Chinese Lexical Semantics. Taibei, China.

Lyons, J. 1977. Semantics. Cambridge: Cambridge University Press.

MacKay, D. J. C. 2003. Information Theory, Inference, and Learning Algorithms. Cambridge: Cambridge University Press.

McCarthy, D., and J. Carroll. 2003. Disambiguating Nouns, Verbs, and Adjectives Using Automatically Acquired Selectional Preferences. Computational Linguistics 29 (4):639-654.

Merlo, P., and S. Stevenson. 2001. Automatic Verb Classification Based on Statistical Distributions of Argument Structure. Computational Linguistics 27 (3):374-408.

Michell, Tom M.. Machine Learning. Translated by Zen Huajun and Zhang Yinkui. Beijing: China Machine Press.

Resnik, P. 1996. Selectional constraints: an information-theoretic model and its computational realization. Cognition 61:127-159.

Qu, Weiguang. 2008. Lexical Sense Disambiguation in Modern Chinese. Beijing: Science Press.

Wu, Yunfang, Duan Huiming and Yu Shiwen. 2005. Verb's Selectional Preference on Object. Spoken and Written Language in Practice 2005(2):121-128.

Yarowsky, D. 1995. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. Paper read at Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics, at Cambridge, MA.

Zheng, Xuling, Zhou Changle, Li Tangqiu and Chen Yidong. 2007. Automatic Acquisition of Chinese Semantic Collocation Rules Based on Association Rule Mining Technique. Journal of Xiamen University (Natural Science) 46(3):331-336.