International Journal of Knowledge and Language Processing Volume 11, Number 1, **2020**  www.ijklp.org KLP International ©2020 ISSN 2191-2734 pp.39–50

## A Comparison Research of Sentiment Words Based on Chinese Online Product Reviews of Different Domains

Ziyan Cao<sup>1</sup>, Minxuan Feng<sup>1</sup><sup>⊠</sup>, Xuefen Mao<sup>1</sup>, Yang Song<sup>2</sup> and Bin Li<sup>1</sup> <sup>1</sup> School of Chinese Language and Literature Nanjing Normal University N0.122, Ninghai Road, Gulou district, Nanjing, Jiangsu, China caoziyan\_nnu@foxmail.com, fengminxuan@njnu.edu.cn, maoxuefen\_nnu@foxmail.com, libin.njnu@gmail.com

> <sup>2</sup> Department of Computer Science Nanjing University No. 163, Xianlin Avenue, Qixia District, Nanjing, Jiangsu, China songyangnju@hotmail.com

## Selected paper from CLSW2019

**ABSTRACT.** With sentiment analysis attracting considerable attention in the NLP community, sentiment lexicons of various languages have been developed as a basis for linguistic computing. However, sentiment words tend to get different polarities in different domains, which limits the using of sentiment lexicons. This paper extracts the sentiment words from Chinese online product review of four different domains namely the cosmetics, food, clothing and laptops with the improved Pointwise Mutual Information algorithm, comparing the distribution and semantic differences of sentiment words in four domains. The experiment's result shows that sentiment words tend to have various domain adaptability and therefore can be divided into universal sentiment words, specific sentiment words, professional sentiment words, restricted sentiment words, dynamic sentiment ambiguous adjectives, etc. A disambiguation scheme is proposed in order to enhance the domain adaptability of sentiment lexicons, and ultimately helps to improve the accuracy of the sentiment analysis based on sentiment lexicons.

Keywords: Natural language processing, Sentiment analysis, Sentiment words, Domain adaptability

1. Introduction. With the rapid expansion of e-commerce, online reviews on products and services have appeared in large numbers. The researches on sentiment analysis of online product reviews spring up, focusing on the mining of positive and negative opinions and prediction of polarities from natural language. Sentiment words indicate reviewers' opinions and present positive or negative polarities [1], therefore, the sentiment lexicons, which include sentiment words or phrases and their prior polarities, are the most important basic resources of sentiment analysis. However, as some sentiment words' polarities are not completely fixed, and some sentiment words would vary strongly with domains [2], the adaptability of lexicons when used in various domains has always been a difficulty for its construction and application [3]. For example, the phrase "声音大(sound loud)" has a positive orientation when talked in communication device domain, which means the reviewer satisfies with the device as he or she could hear clearly; but when appeared in household appliance domain such as air conditioners, the word "大(loud)" means the machine is noisy and indicates the negative polarity. Therefore, it is unrealistic that a sentiment lexicon can be used through all the domains. As various products have different functions and of different domains, the domain adaptability brings sentiment analysis tasks based on online product reviews greater challenges.

This paper will extract the sentiment words from the Chinese online product reviews of the four domains (cosmetics, food, clothing and laptops) to focus on this topic. Then, a comparative analysis will be conducted to study the varying of semantics and polarities of the sentiment words when used in different domains, and how they are distributed in these four domains.

2. **Problem Statement and Preliminaries.** Many scholars have carried out related researches around sentiment words and sentiment lexicons, mainly focusing on the extraction of sentiment words, polarity classification and basic resource construction. In addition, the adaptability of the sentiment lexicons has also attracted some attention.

2.1. **Construction of Sentiment Lexicons.** The sentiment lexicon contains sentiment words and a small number of phrases for sentiment analysis. In addition to manual construction, most studies treat this task as the sentiment word extraction and classification. According to Ref. [2], the main method is to take use of co-occurrence of conjunction, or of seed word. The former is to summarize the rules of the polarities of words before and after conjunctions and use these patterns to extract sentiment words, while the latter uses the Pointwise Mutual Information, SO-LSA, etc. to examine the co-occurrence characteristics of the inspected words and seed words to judge whether they indicate the same polarity. It is also a common approach to solve the classification problem of sentiment words by using traditional machine learning methods and neural network models. With the methods above, most of the sentiment words of a corpus can be obtained, but they cannot be approached exhaustively.

2.2. Resources of Sentiment Lexicons. Some sentiment lexicons have been constructed to

serve for sentiment analysis. As for the English resources, General Inquire<sup>1</sup> is the earliest sentiment lexicon and it labels each word according to the polarity, intensity and category. As for the Chinese resources, sentiment lexicon of National Taiwan University (NTUSD) collects 2,810 positive words and 8,276 negative words [4]. Chinese Emotion Word Ontology provided by Dalian University of Technology contains most Chinese Psychological verbs and adjectives [5]. HowNet also provides a word set for sentiment analysis<sup>2</sup>, including 8 subsets of Chinese and English evaluation words and emotion words, and it collects the subjective words and adverbs of degree in addition. The existing sentiment lexicons and other resources contain most of the words or phrases with sentiment information. However, they don't deal with the sentiment words with unfixed polarity.

2.3. Relevant researches on the domain adaptability of sentiment lexicons. The domain adaptability of sentiment lexicons has been paid attention to by many researchers in the application of sentiment lexicons. Ref. [2] proposed that "the sentiment words closely relate to domain of the product" and pointed out it was essential to take use of domain knowledge. Xi [6] believed that the using of previous sentiment lexicons were limited by adaptability of domains. Tang [7] also emphasized the construction of domain-adaptive sentiment lexicon. Ref. [3] considered the domain adaptability as a difficultly of lexicon construction. Therefore, quite a few studies constructed sentiment lexicons based on one or two specific domains, such as hotels, blogs or shopping [8-10]. Qi [6] also proposed a method of constructing domain-specific sentiment lexicon with pointwise mutual information and contextual constraints. Ref. [11] and Ref. [12] attempted to use different lexicons for sentiment classification in different domains. Ref. [13] used the N-gram model, the co-occurrence of words and pointwise mutual information to construct a cross-domain sentiment dictionary. Although the domain adaptability of sentiment lexicons is a common difficulty, their origins, performance and types lack systematic research and analysis from the perspective of linguistics. This is precisely the important theoretical basis for improving the domain adaptability of sentiment lexicons.

3. The extraction of Sentiment Words. In order to explore the domain adaptability of sentiment lexicons, referring to the domains of existing product review database and the balance of product types, four categories of product are selected: the cosmetics, food, clothing and laptops, and this paper used the improved pointwise mutual information to extract sentiment words and construct the sentiment lexicon.

3.1. The collection and processing of review texts. With the web crawlers, this paper obtained about 4,500 reviews of each domain (cosmetics, food, clothing and laptops) from a popular shopping website. For example:

还不错的,没出现什么大问题,就是外放的喇叭有时候会莫名其妙的沙沙响,重启

<sup>&</sup>lt;sup>1</sup> http://www.wjh.harvard.edu/~inquirer/

<sup>&</sup>lt;sup>2</sup> http://www.keenage.com/html/e\_index.html

了又不会了,很奇怪。(笔记本电脑领域)

Not bad, there is no big problem, but the speaker sometimes noises and when restarted it will not, that's very strange. (In the domain of laptops)

The processing of the text can be divided into four steps. The first stage is the segmentation of clauses and words, part-of-speech tagging, and the remove of duplicates and blank items. The second stage is to manually label the polarity of the reviews (positive, negative). Positive sentiment sentences are labelled with P, and negative ones are labelled with N<sup>3</sup>. The third stage is to manually deal with the sentimental polarity shifting, which is mainly caused by the adversative conjunctions and the negators in Chinese. These sentences are split with polarity shifting, and the opposite polarity is marked of each sentence. The fourth stage is to check and correct. Through the four steps, about 5,000 processed reviews are obtained in each domain (Table 1). Among these reviews, positive reviews are more than the negative ones, which is determined by the phenomenon that the Chinese shopping websites' positive feedback rate is higher than the negative feedback rate.

| TABLE 1. THE DATASET FOR EXPERIMENT |       |                  |                  |
|-------------------------------------|-------|------------------|------------------|
|                                     | Total | Positive reviews | Negative reviews |
| Cosmetics                           | 5,101 | 4,090            | 1,011            |
| Food                                | 4,983 | 4,553            | 430              |
| Clothing                            | 5,032 | 4,360            | 672              |
| Laptops                             | 4,999 | 3,605            | 1,394            |

3.2. The Extraction of Sentiment Words. In the study of semantic prosody [14-15], the extraction of sentiment words adopts improved Pointwise Mutual Information algorithm, which is based on the co-occurrence of words and positive or negative polarity. Pointwise Mutual Information is often used to measure the correlation between two words. The larger the PMI value is, the more relevant x and y are. As it is shown in formula (1), the PMI of the inspected words in positive or negative polarity, is marked as pPMI or nPMI; and if pPMI-nPMI>0, temporarily put it into the positive sentiment dictionary; instead, add it to the negative sentiment dictionary. In the formula, x represents the polarity of the comment, and y represents the word to be classified. If a word appears only in positive (negative) reviews, it is temporarily classified as a positive or negative sentiment word. Finally, according the category labels, words and phrases including adjectives, nouns, verbs, adverbs, conjunctions, modal particles are retained for further analysis.

$$PMI(x,y) = \log \frac{p(x,y)}{p(x)p(y)}$$
(1)

4. The Analysis of Experimental results and Conclusion. The improved pointwise mutual information algorithm is used to extract sentiment words and construct the lexicons

<sup>&</sup>lt;sup>3</sup> In the extracted sentences, the amount of reviews of neutral polarity is extremely small and will not be considered here.

of positive and negative sentiment words.

4.1. The Extracted Sentiment Words. This experiment temporarily received 2,708 sentiment words (including positive and negative ones). The distribution of sentiment words and the number of them in various domains are shown in Table 2:

| TABLE 2. T | HE AMOUNT OF | SENTIMENT WORDS IN | EACH DOMAIN |
|------------|--------------|--------------------|-------------|
|            | Total        | Positive           | Negative    |
| Cosmetics  | 734          | 439                | 295         |
| Food       | 638          | 495                | 143         |
| Clothing   | 541          | 403                | 138         |
| Laptops    | 795          | 461                | 334         |

The words collected in the experiment include not only adjectives and verbs, but also some adverbs, modal particles, conjunctions and nouns with positive or negative sentiment expectations (such as "problems"), etc. These words can indirectly express the reviewers' attitudes and opinions. The dynamic sentiment ambiguous adjectives (DSAAs) [16] (such as "big", "small", etc.) are also included in this table and are the special type of sentiment words; although the results show these words get positive or negative polarities, they cannot be completely classified by the result of pointwise mutual information algorithm. According to the distribution and polarity in each domain, the sentiment words can be divided into the following types.

4.2. Universal Sentiment Words. The experimental result shows that some sentiment words appear in all the domains of the cosmetics, food, clothing and laptops, and have the consistent polarity. This paper takes such sentiment words as "universal sentiment words", which means they show strong universality of polarity in various domains. Table 3 shows some of the universal sentiment words:

| 值得 worthy 赞(点赞)like 贴心 intimate 惊喜 surprise 感谢 thank 相信 trust     |
|---|
| 支持 support 爱 love (v.) 信赖(信任) 享受 enjoy 期待 expect 满意 satisfied,    |
| 喜欢 like 推荐 recommend 方便 convenient 实惠 affordable 棒 good 好用 useful |

物美价廉 cheap and good 合适 appropriate 超值 overvalued 仔细 careful 严实 tight 美丽 beautiful 细腻 delicate 舒服 comfortable 适合 suitable 适中 moderate

delightful 老顾客 old customer 热情 enthusiasm 耐心 patient 细心 attentive 周 到 thoughtful 良心 conscience 回购(再来/第二次) repo(again/second) 正品

| TABLE 3. THE UNIVERSAL SENTIMENT WORDS OF EACH DOMAIN |
|---|
|---|

| Positive  |  |
|-----------|--|
| Sentiment |  |

完整 complete 干净 clean 完美 perfect 值 valuable 好用 useful 大方 generous 高 Words 端 high-end 高档 high-quality 好看 nice 整齐 neat 开心 happy 高兴 joyful 愉快

real thing 超出 beyond

| Negative           | 郁闷 depressed 累 tired 失望 disappointed 担心 worried 后悔 regret 有点 a little |
|--------------------|---|
| Sentiment<br>Words | 假 false 错 wrong 差 broken 破(破损)broken 垃圾 shit 差别 difference            |
|                    | 出乎意料 unexpectedly 就是 just 超(超级)super 特(特别)special 太 too 相当            |
| Others             | quite 完全 completely 还 also 真心 real(真/真的) 绝对 absolutely 一点(一点          |
|                    | 点)a little(a little bit) 麻烦 trouble 确实 really 稍微(稍)slightly 但是 but    |

For the "others" item in the table above, these words are classified into positive or negative lexicons according to statistical results, but the differences of their pPMI and nPMI are extremely small, and from a semantic perspective, they cannot express the attitudes and emotions of the reviewers strongly and directly, so they cannot be regarded as strictly positive or negative sentiment words, such as the word "有点(a little)". As can be seen from the table, the number of universal sentiment words is small, with 19.07% of cosmetics, 21.94% of food, 25.88% of clothing, and 17.61% of laptops. There are three main types of universal sentiment words:

(1) Absolute commendatory and derogatory evaluation words. Attributes mean the specific aspects of evaluation, such as "logistics", "appearance", "quality", etc. The words which are called evaluation words, are able to describe and modify a certain attribute of the commodity, and can express the attitudes of the reviewers. Some evaluation words in table 3 have fixed sentiment polarity except for some special contexts, such as "clean" and "useful", expressing positive attitudes of the reviewers. Therefore, these words are called absolute commendatory and derogatory evaluation words.

(2) Subjective emotional words (such as some psychological verbs). Subjective emotional words directly express the feelings and emotions of the reviewers, such as "surprise", "happy" and other psychological verbs and adjectives. These words tend to have fixed polarities and can be used in different contexts. They are called subjective emotional words.

(3) Sentiment meaning words. Words like some modal particles, degree adverbs and conjunctions often co-occur with positive or negative words, and also have some positive or negative meanings, which can be called "semantic prosody" in linguistics [14-15], such as the adverb "有点(a little)", always modifying some negative adjectives in Chinese. This pragmatic rule is relatively fixed and has cross-domain adaptability. Such words can be called sentiment meaning words.

4.3. **Specific sentiment words.** Specific sentiment words refer to sentiment words that only appear in a certain domain. The numbers of them in four domains are different, which is caused by the imbalance of the amount of modifiers and product categories, and the reviewers have different characteristics. People have different cognition of different items, and the number of angles that each product can be described is different. In terms of products' sub-categories, the cosmetic includes skin care, make-up, and many other products, while the sub-categories of food are relatively simple. What's more, the

consumers of the cosmetics and laptops are mostly young people, so the product reviews are expressed in a variety of ways, including a large number of new words and network words.

The results show that there are two types of specific sentiment words: One is the professional word. These words indicate a fixed polarity in one domain and have a specific and single meaning. For example, "轻薄的(light)" in the cosmetics domain can only describe the texture, and "二手机(second-hand machine)", "蓝屏(blue-screen)" are also the professional words in the domain of laptops. On the other hand, some words which appear in one domain in this research may also appear in other domains with the expansion of the corpus, and many of them are synonymies of other universal sentimental words. For our research, the sentiment words with application value are mainly the first category, namely professional sentiment words. The distribution of specific sentiment words and professional sentiment words in four domains are shown in Table 4 below. Table 5 gives some of the professional sentiment words in each domain:

| TABLE 4. THE DISTRIBUTION OF PROFESSION | NAL AND SPECIFIC SENTIMENT WORDS IN EACH |
|---|--|
| DO                                      | MAIN                                     |

|           | Professional sentiment words | Proportion | Specific sentiment words | Proportion | - |
|-----------|------------------------------|------------|--------------------------|------------|---|
| Cosmetics | 77                           | 10.49%     | 310                      | 42.23%     |   |
| Food      | 41                           | 6.43%      | 291                      | 45.61%     |   |
| Clothing  | 58                           | 10.72%     | 188                      | 34.75%     |   |
| Laptops   | 69                           | 8.68%      | 364                      | 45.79%     |   |

| <b>FABLE 5. PROFESSIONAI</b> | SENTIMENT WORDS | S IN EACH DOMAIN |
|------------------------------|-----------------|------------------|
|------------------------------|-----------------|------------------|

|               | Positive     | 持久 lasting 滋润(水润 弹润) moisturizing(moist) 轻盈      |
|---------------|--------------|--|
|               |              | lightweight 光泽 luster (光泽感)柔顺 supple 上镜          |
| Constitu      |              | photogenic 提亮 brighten                           |
| cosmetic<br>s |              | 拔干 pull out 暗沉无光(暗沉)dull 假面 mask 假白 false        |
|               | Negativ<br>e | white 炸毛 fried hair 厚重感 heaviness 干枯 dry 毛躁      |
|               |              | buttercup 刺激(刺激性)stimulate 闷痘 stuffy 飞粉 fly      |
|               |              | powder   |
| D :::         |              | 咬得动 biting 醇香 mellow 爽口(可口)tasty 美味 delicious    |
| Food          | Positive     | 正宗 authentic 香脆 crispy                           |
|               | Negativ<br>e | 腥味 astringency 生味 raw 臭臭的 nasty 骚味 savor 怪味      |
|               |              | weird 受潮 damp 变质 metamorphism 咸 salty 腻 greasy 苦 |
|               |              | bitter 烂掉 rotten                                 |

|          |              | 透气breathable 吸汗 sweat 纯棉(棉 全棉) cotton 合身 fit 修      |
|----------|--------------|---|
| Clothing | Positive     | 身 slim fit 休闲 leisure 保暖 keep warm 贴身 close-fitting |
|          |              | 百搭 wild 性感 sexy 耐穿 anti-wear 耐磨 wear resistant      |
|          |              | 贴肉 sticking meat 粘毛(沾毛 黏毛 粘身) stick 污渍 stain        |
|          | Negativ<br>e | 起皱(褶皱) wrinkle 缩水 shrink 刺刺的 prickly 脱线             |
|          |              | off-line 起毛 fluff 扎身 tied up 磨坏 worn out            |
|          |              | 高清 HD 科技感 technological 果粉 Apple Fans 高配 high       |
|          | Positive     | configuration 全新机 brand new machine 定制机 custom      |
|          |              | machine 定制 custom made 双系统 dual system 顶配 top       |
| Laptops  |              | match 防眩光 anti-glare                                |
|          |              | 杂音 noise 静电 Static electricity 缝隙 gap 掉帧 drop frame |
|          | Negativ<br>e | 波动fluctuation 卡顿freeze up 松动loosen 发热fever 降频       |
|          |              | underclocking 耗电 consume power 发烫 hot 跳屏 skip       |
|          |              | screen  |

4.4. **Restricted sentiment words.** Restricted sentiment words have fixed polarities but their scopes of domain are restricted, or their polarities are limited to attributes, domains or entities. The details are shown in Table 6:

|                 | Restriction | Name  | Examples   |
|-----------------|-------------|---|--|
| Type 1          | Domain      | Domain-restricted sentiment words                     | 有型 Cool(only for laptops and clothing, not for food)                       |
| Type 2 Polarity | Dolority    | Relative commendatory and derogatory evaluation words | Fragrant, expensive: Lipstick is too fragrant; I like expensive computers; |
|                 | roiarity    | Dynamic sentiment<br>ambiguous adjectives             | Big, small, etc.   |

TABLE 6. THE CLASSIFICATION OF LIMITED SENTIMENT WORDS

The first type is the fixed-polarity but domain-restricted sentiment words. Such restricted sentiment words can appear in at least two domains. The second type is the sentiment words which polarities are restricted, which can be divided into two categories, relative commendatory and derogatory evaluation words and dynamic sentiment ambiguous adjectives (DSAAs). The relative commendatory and derogatory evaluation words mean that they seem to indicate positive or negative polarities, but are easily changed in different contexts; the DSAAs are polysemous and neutral, but they are important to evaluate products, and need to be combined with the texts to analyze their polarity. Here are some examples of DSAAs selected:

| Domain<br>DSAA      | Cosmetics                                     | Food                               | Clothing                          | Laptop  |
|---------------------|---|------------------------------------|-----------------------------------|---|
| 软 soft              | Pos:毛刷/粉扑的材质<br>material of brush and puff    | N-p:4口感 taste                      | Pos:材质 material                   | Neg:材质 material                                       |
| 硬 hard              | Neg:毛刷/粉扑的材质<br>material of brush and puff    | N-p:口感 taste                       | Pos:质量 quality<br>Neg:材质 material | Pos:质量 quality<br>N-p:材质 material                     |
| 大 Big               | N-p:外形 appearance                             | Pos:规格<br>specification            | Pos:弹性 elasticity<br>Neg:尺码 size  | N-p:外形 appearance<br>Neg:问题 problem                   |
| 小 Small             | N-p:外形 appearance<br>Pos:副作用 associate work   | N-p:外形<br>appearance/包装<br>package | Neg:尺码 size                       | Pos:声音 sound<br>N-p:外形 appearance                     |
| 高 High              | Pos:饱和度 saturation<br>Neg:价格 price            | Neg:价格 price                       | Neg:价格 price/外形<br>appearance     | Pos:性价比 cost-effective<br>Neg:价格 price                |
| 低 Low               | Pos:价格 price<br>Neg:体验感 experience<br>feeling | Pos:价格 price/脂<br>肪含量 fat content  | Pos:价格 price<br>Neg:设计 design     | Pos:价格 price<br>Neg:配置 configuration                  |
| 快 Fast              | Pos: 物流 logistic                              | Pos:物流 logistic                    | Pos:物流 logistic                   | Pos:运行 run/物流 logistic<br>Neg:耗电 power<br>consumption |
| 慢 Slow <sup>5</sup> | Neg:物流 logistic                               | Neg:物流 logistic                    | Neg:物流 logistic                   | Neg:运行run/物流logistic                                  |
| 薄 Thick             | Pos:质地 texture<br>Neg:包装 package              | /                                  | N-p:材质 material                   | Pos:外形 appearance/尺<br>寸 size                         |
| 厚 Thin              | Neg:质地 texture                                | Neg:包装 package                     | Neg:材质 material                   | Neg:外形 appearance/尺<br>寸 size                         |
| 轻 light             | Pos:质地 texture                                | N-p:口味 taste                       | Pos:材质 material                   | Pos: 重量 weight  |
| 重 heavy             | Neg:气味 smell/妆感<br>make-up                    | Neg:重量 weight                      | Neg:气味 smell                      | Neg: 重量 weight  |
| 长 long              | Pos: 时间 time<br>N-p: 外形 appearance            | Neg:物流 logistics                   | Neg:尺寸 size                       | Pos: 待机时间 standby<br>time/ 更新时间 update<br>time        |
| 短 short             | Neg:外形 appearance                             | /                                  | Neg:长度 length                     | Neg:续航时间 battery<br>duration/长度 length                |

TABLE 7. THE ANALYSIS OF DYNAMIC SENTIMENT AMBIGUOUS ADJECTIVES

<sup>&</sup>lt;sup>4</sup>"Non-polarity" means that when a word is used to evaluate an attribute of the field, it only represents the attribute value without emotional orientation itself.

<sup>&</sup>lt;sup>5</sup>Although "slow" here shows the same polarity, we still consider it as a dynamic sentiment ambiguous adjective rather than a universal sentiment word. When "slow" modifies "power consumption", "time" or other attributes which showing positive semantic features, we are accustomed to use "snacks can eat for a long time", rather than "snacks eat slowly". Considering the symmetry of DSAA, it is considered as a DSAA together with "fast".

As the table shows, DSAAs present a variety of polarities, and are frequently used in various fields, which mainly related to the following elements:

(1) The product domains. Take "硬(hard)" as an example, as the modifier in terms of hardness, it is used to modify the keyboard and indicates positive polarity in the domain of laptops, which means the keyboard is good to use; while when used to describe food, it means the food is difficult to chew, indicating a negative polarity.

(2) The attribute. For the word "(快)fast", "物流很快(the distribution was fast)" is positive, while "很快就吃完了(It was eaten up quickly)" indicates that the amount of the product is small and it is unsatisfactory.

(3) The object of the review. If "(多) many" means the number/amount is large, it is positive when the object is "赠品(gift)", such as "送了很多赠品(so many gifts to send)"; and negative when the object is "问题/瑕疵(problem)".

(4) Subjective preferences will of the reviewers. For example, in the domain of cosmetic, "很满意包装袋够大(very satisfied with the bag as it is big enough)" and "包装袋太大了 (the bag is too big)", although the above elements are consistent, they show the opposite polarities, which depends entirely on the subjective preferences of the reviewers.

4.5. The Disambiguation of Dynamic Sentiment Ambiguous Adjectives (DSAAs). The DSAAs have no fixed polarities, so they have little contribution to the construction of sentiment lexicons. The judgment of sentimental polarity needs to be based on prior knowledge and feature extraction. Combined with the research results of the conclusion of [16], the following disambiguation scheme is proposed:

(1) Refer to subjective sentiment words and sentiment meaning words. For example, "屏 幕够大我喜欢(The screen is big and I like it)" and "屏幕太大了不适合女生(The screen is too big and not suitable for girls)", in these sentences, "big" describes the size of the computers, only representing the attribute value without sentimental polarity, and what is the decisive factor in the sentiment analysis is the consumers' subjective preferences. It is necessary to use subjective sentiment words and sentiment meaning words to judge the polarities of the word and the whole sentence.

(2) Refer to the positive or negative sentiment expectations of the object. The sentimental polarity of the evaluation phrase can be obtained by multiplying the polarity of the sentiment expectation and the polarity of the DSAAs. For example, the sentiment expectation of the words like "瑕疵(flaw)" / "问题(problem)" are negative (-), and words like "大 (big)", "多 (many)", etc. represent enhancement, are positive (+). Therefore, the phrase "瑕疵多(many flaws)" as a whole is negative.

(3) Refer to the function of the product. For example, "材质柔软(the material is soft)" is positive in the domain of clothing, because the product needs to be soft for comfort. In the domain of laptops, it shows negative polarity when describing the softness of the keyboard

shell material. This often involves the application of the product ontology knowledge.

5. Conclusion and Future Work. In this paper, we extract sentiment words from the review texts of four different domains around the domain adaptability of sentiment lexicons. We proposed three categories of sentiment words according to the distribution and semantic differences, including universal, specific and restricted sentiment words. The universal sentiment words exist in all four domains, while specific sentiment words appear in certain domain. In comparison, the professional sentiment lexicons. As a typical type of restricted sentiment words, dynamic sentiment ambiguous adjectives (DSAAs) are difficult to judge their polarities. And subjective sentiment words, sentiment meaning words, sentiment expectations of objects and the ontology knowledge can be referred as the disambiguation schemes for DSAAs. These researches improve the adaptability of existing sentiment lexicons to be put into use in different domains.

In the future, we will pay more attention to the basic resource construction of sentiment analysis, such as constructing the lexicons specially for the sentiment meaning words, building domain sentiment lexicons of professional sentiment words, or a more comprehensive and complete product ontology knowledge base for DSAA disambiguation, in order to better solve the adaptive problem of the sentiment lexicons and contribute to the sentiment analysis based on sentiment lexicons.

Acknowledgment. We are grateful for the comments of the reviewers. This work is the staged achievement of the projects supported by National Language Committee project of China(YB135-61), National Social Science Foundation of China (18BYY127) and Jiangsu Higher Institutions' Excellent Innovative Team for Philosophy and Social Sciences project (2017STD006).

## REFERENCES

- Y. Y. Zhao, B. Qin and T. Liu, Sentiment analysis, Journal of Software, vol.21, no.8, pp. 1834-1848, 2010.
- [2] L. Li. Mei, H. Y. Huang, X. Y. Zhou, et al., A Survey on Sentiment Lexicon Construction, Journal of Chinese Information Processing, vol.30, no.5, pp.19-27, 2016.
- [3] K. Wang and R. Xia, A Survey on Automatical Construction Methods of Sentiment Lexicons, Acta Automatica Sinica, vol. 42, no.4, pp.495-510,2016.
- [4] L. W. Ku, Y. S. Lo and H. H. Chen, Using polarity scores of words for sentence-level opinion extraction, Proceedings of the NTCIR-6 Workshop Meeting, pp.316-322,2007.
- [5] J. M. Chen, The Construction and Application of Chinese Emotion Word Ontology, Dalian University of Technology, Liaoning, 2008.
- [6] Y. H. Xi, Construction of Domain-specific Sentiment Lexicon in Product Reviews, Journal of

Chinese Information Processing, vol.30, no.5, pp.136-144, 2016.

- [7] D. Y. Tang, Research on Domain Adaptive Chinese Sentiment Lexicon Construction, Harbin Institute of Technology, Heilongjiang, 2012.
- [8] X. Shi, Research on Sentiment Classification Based on Semantic Lexicon of Hotel Field, Hebei University, Hebei, 2014.
- [9] B. Gui, X. P. Yang, Z. X. Zhang, et al., Research on Building Lexicon for Sentiment Analysis Based on the Chinese Microblogging Smiley, Transactions Beijing Institute of Technology, vol.34, no.5, pp. 537-541, 2014.
- [10] X. Wu and L. Wang, Investigation on Sentiment of Reviews with Shopping Field Dictionary Construction, Computer Technology and Development, vol.27, nol.7, pp. 194-199,2007.
- [11] Y. Choi, C. Cardie, Adapting a Polarity Lexicon Using Integer Linear Programming for Domain-specific Sentiment Classification[C].In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, pp. 590-598, 2009.
- [12] S. Huang, Z. D. Niu and C. Y. Shi, Automatic Construction of Domain-specific Sentiment Lexicon Based on Constrained Label Propagation[J]. Knowledge-Based Systems, vol.56, pp. 191-200, 2014.
- [13] D. Bollegala, D. Weir and J. Carroll, Cross-domain Sentiment Classification Using a Sentiment Sensitive Thesaurus[J]. IEEE Transactions on Knowledge and Data Engineering, vol.25, no.8, pp. 1719-1731, 2013.
- [14] M. Stubbs, Text and Corpus Analysis [M]. Blackwell Publishers Ltd, 1996.
- [15] Y. H. Ji and J. P. Wu, Semantic Prosody Research: Object, Method and Application. Journal of Xiamen University (Arts & Social Sciences), vol.143, no.3, pp.63-68, 2000.
- [16] Y. F. Wu and J. Peng, Disambiguating Dynamic Sentiment Ambiguous Adjectives[C]. In: Proceedings of the 23rd International Conference on Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, pp.1191-1199, 2010.