International Journal of Innovative
Computing, Information and Control
Volume 10, Number 1, February 2014

 ISSN 1349-4198
pp. 257–268

TARGET BASED REVIEW CLASSIFICATION FOR FINE-GRAINED SENTIMENT ANALYSIS

Changqin Quan[1] and Fuji Ren[2]

[1]AnHui Province Key Laboratory of Affective Computing and Advanced Intelligent Machine
School of Computer and Information
HeFei University of Technology
No. 193, Tunxi Road, Hefei 230009, P. R. China
quanchqin@gmail.com

[2]Faculty of Engineering
University of Tokushima
2-1 Minami-Josanjima, Tokushima 770-8506, Japan
ren@is.tokushima-u.ac.jp

Received November 2012; revised March 2013

Abstract. *Target based sentiment classification is able to provide more fine grained sentiment analysis. In this paper, we propose a similarity based approach for this problem. Firstly, a new measure of PMI-TFIDF by combining PMI (Pointwise mutual information) and TF-IDF (term frequency-inverse document frequency) is proposed to measure the association of words for extending related features for a given target. Then Polynomial Kernel (PK) method is applied to get the similarities between a review and the related features of different targets. The sentiment orientation of a review is determined by comparing their similarities with the target based opinion words. The comparisons between PMI and PMI-TFIDF showed that the extracted features that measured by PMI-TFIDF have closer association with the targets than the extracted features measured by PMI. And the association values measured by PMI-TFIDF showed better distinction between different features. The experiments also demonstrated the effectiveness and validation of the proposed approach on target based review classification, opinion words extraction, and target based sentiment classification.*
Keywords: Sentiment analysis, Target based, Opinion words extraction, Word similarity

1. **Introduction.** Sentiment analysis aims to determine the positive and negative attitudes of a writer with respect to some topics in consumer generated media. One of the most benefits of a sentiment analysis system is the business values that it delivers. Because so many people use social media, enterprises will understand how about people like (or dislike) their products and service. And people post things that can be analyzed and shown to be indicators of their intents (to buy, to complain, or cancel their service, etc.). This information can be helpful for improving customer service and product quality, can also be used to identify and act on opportunities.

The early research of sentiment analysis mainly focused on text classification of positive or negative attitude [1-3]. Based on these pilot studies, follow-up studies considered rating sentiment on a scale (e.g., from -10 to $+10$) to capture sentiment intensity [4,5]. As simply judging the sentiment polarity of a text that is not sufficient for many customer review analysis, with the in-depth study, researchers begun to commit their efforts for more fine grained sentiment analysis: feature-level [6-8] or target-dependent [9] sentiment classification. Feature-level or target-dependent sentiment analysis predicts the sentiment orientation related to different review features, which are considered as opinion targets. In this problem, identifying implicit review features is a big challenge. Many approaches achieved good performance for identifying explicit review features that appear in reviews. Representative methods include template extraction based method [8] and association rule mining based method [10]. However, a lot of review features are implicit in real customer reviews, so recent studies have focused on mining implicit or hidden sentiment association in reviews. Du and Tan proposed an iterative reinforcement framework based on the improved information bottleneck algorithm to detect the implicit review features and mine the hidden sentiment association in reviews [11]. Jiang et al. proposed to detect target-dependent features for twitter sentiment classification [9].

Although a review in some domains may contain multiple opinions on more than one topics, especially for electronic and car product reviews, this situation is rare in some other domains, such as book reviews. A statistics on a corpus of 4000 Chinese book reviews [12] (containing 510,650 Chinese characters) shows that there are less than 3% book reviews contain multiple opinions on more than one topics. From the enterprises' point of view, they are more concerned about users' overall evaluation on the different targets or topics, such as content, price, service for book reviews. So it is necessary to classify a review according to a certain target, and then identify the sentiment orientation of this review based on the target that it belongs to.

In this paper, we propose a similarity based approach for target based sentiment classification. Firstly, a word similarity measure (PMI-TFIDF) is applied to extend related features for a certain target. Then a polynomial Kernel (PK) in vector-space model is constructed to get the similarities between a review and the related features of different targets. The sentiment orientation of a review is determined by comparing their similarities with the target based opinion words. The experiments have demonstrated the effectiveness and validation of the proposed approach.

The remainder of this paper is organized as follows. Section 2 presents the method of target based review classification. Section 3 describes the method of target based sentiment classification. Section 4 presents the experimental setup and results. Section 5 concludes this paper with closing remarks and future directions.

2. **Target Based Review Classification.** In target based review classification, it is relatively easier to identify by keyword spotting when the target is a specific thing, such as "battery life of a camera". However, when the target is abstract, such as "content of a book", it would be difficult to identify because there are numerous expressions for such a target. Many previous researches restricted target to nouns and noun phrases [7,9,11], but this way will be not applicable for an abstract target. For example, sentences (1)-(3) show different means of expressing negative opinion for "the content of a book".

(1) 题目好，但内容空。(Good title, but the content is empty.)

(2) 很一般啊，并不精彩。(It is just so-so, not wonderful.)

(3) 才读了2页就睡着了。(Only read two pages and then fell asleep.)

Sentences (4)-(6) show the different meanings of expressing negative opinions on "after-sales service" in book reviews.

(4) 送货太慢了，将近一个月时间，而且还不给任何解释。(Delivery is too slow, nearly a month, without any explanation.)

(5) 预定了书，但是没有拿到书，郁闷！(I had made a reservation for the book, but did not get the book, how depressing!)

(6) 接线员态度极差！(Attitude of the operator was extremely bad!)

As shown in the above examples, it is not easy to classify a review to a relatively abstract target. Before we determine the positive and negative attitude to a certain target, we should firstly distinguish reviews relating to a target from others.

From the above examples, we find that although there are many expressions for an abstract target, some keywords still can be found, including nouns, adjectives, verbs, etc. For example, sentences (1) uses nouns of "题目 (title)", "内容 (content)"; sentences (2) uses adjectives of "很一般 (just so-so)", "不精彩 (not wonderful)"; sentences (3) uses verbs of "读 (read)", "睡着 (fell asleep)" to express opinions on "the content of a book". Sentences (4) uses nouns of "送货 (delivery)", "解释 (explanation)"; sentences (5) uses nouns of "预定 (made a reservation)", sentences (6) uses nouns of "接线员 (operator)" to express opinions on "after-sales service" in book reviews. If we can get these feature words, we may know what target a review relates to.

2.1. **Extend related features by measuring the association of words.** PMI (Pointwise mutual information) [13] is a measure of association used in information theory. PMI has been applied in many sentiment analysis methods to measure the association of words, see Equation (1).

$$PMI(w_i, w_j) = \log \frac{\Pr(w_i, w_j)}{\Pr(w_i)\Pr(w_j)} \tag{1}$$

where $\Pr(w_i, w_j)$ is the probability of a sentence containing word w_i and word w_j in corpus, $\Pr(w_i)$ is the probability of a sentence containing word w_i in corpus.

The main problem of PMI for measuring the association of words is that PMI value is sensitive to corpus size. In a small corpus, different words often get the same PMI value that is not helpful to distinguish the degree of associations. So we combine TF-IDF with PMI in our method to evaluate the associations of candidate features and targets.

The TF-IDF weight (term frequency – inverse document frequency) [14] is a numerical statistic which reflects how important a word is to a document in corpus, see Equations (2)-(4).

$$TF_{ij} = \frac{n_{ij}}{\sum\limits_k n_{kj}} \tag{2}$$

where n_{ij} is the frequency of word i in document j, $\sum\limits_k n_{kj}$ is the number of all words in document j.

$$IDF_i = \log \frac{|D|}{|d : w_i \in d|} \tag{3}$$

where $|D|$ is the number of documents in corpus, $|d : w_i \in d|$ is the number of documents containing word w_i.

The importance of word i to document j can be weighted by Equation (4).

$$w_{ij} = TF_{ij} \cdot IDF_i \tag{4}$$

We combine PMI and TF-IDF to measure the association of words for extending related features, see Equation (5).

$$PMI - TFIDF(w_i, w_j) = \log \frac{\Pr(w_i, w_j) \cdot \sum\limits_K TFIDF(w_i, d_k) \cdot \sum\limits_K TFIDF(w_j, d_k)}{\Pr(w_i)\Pr(w_j)} \tag{5}$$

where $\sum\limits_K TFIDF(w_i, d_k)$ is the sum of word i to all documents in corpus. The multipliers of $\sum\limits_K TFIDF(w_i, d_k)$ and $\sum\limits_K TFIDF(w_j, d_k)$ add the importance measure of word i and word j to all documents in corpus. So it can distinguish words with different degrees of importance in corpus.

Given a target w_i, PMI-TFIDF can get its related features with higher association values, and this target can be extended by these related features.

2.2. **Classifying reviews based on targets.** Before we determine the positive and negative attitude to a certain target, we should firstly distinguish reviews relating to targets. Based on the measure of PMI-TFIDF, we can get related words with higher relevance for a target. Then we use Polynomial Kernel (PK) method to get the similarities between a review and the related features of different targets.

Kernel methods (KMs) are state-of-the-art for solving machine learning problems. Kernel-based algorithms exploit the information encoded in the inner-product among all pairs of data items, avoiding explicitly the computation of the feature vector for a given input. KMs approach the problem by mapping the data into a high dimensional feature space, where each co-ordinate corresponds to one feature of the data items, transforming the data into a set of points in a Euclidean space. In that space, a variety of methods can be used to find relations in the data [15]. In the basic vector-space model, documents are represented by a matrix D, whose columns are indexed by the documents and rows are indexed by the terms. The corresponding kernel is given by the inner product between the feature vectors, see Equations (6) and (7).

$$K = D'D \tag{6}$$

$$k(d_1, d_2) =< \phi(d_1), \phi(d_2) >= \sum_{j=1}^{N} tf(t_j, d_1) tf(t_j, d_2) \tag{7}$$

Document d is represented by a row vector, see Equation (8).

$$\phi(d) = (tf(t_1, d), tf(t_j, d), \ldots, tf(t_N, d)) \in R^N \tag{8}$$

where $tf(t_i, d_j)$ is the frequency of term i appeared in document j.

A linear transformation is $\phi(d) = \phi(d) \times S$, where S is an appropriately shaped matrix, can be set by Equation (9).

$$S = RP \tag{9}$$

where R is a term-weight matrix, and is diagonal, whose entire $R(i,i)$ are the weight of the term i, can be defined by the inverse document frequency $idf(t) = \ln(l/df(t))$ [14], l is the total number of documents in the corpus, $df(t)$ is the number of documents that contain the given term. P is a term-document matrix, whose entire $P(i,j)$ are the weight of the term i in document j.

The new kernel K for this feature space is defined by Equation (10).

$$K = D'D = (DS)'DS = (DRP)'DRP \tag{10}$$

For a given kernel $k(d_1, d_2)$, the derived polynomial kernel is defined by Equation (11).

$$k(d_1, d_2) = (k(d_1, d_2) + m)^n \tag{11}$$

where m and n are parameters of the polynomial kernel. K records the similarities between reviews and the related features lists of different targets. By retrieving and comparing the similarity scores, we can get the target that a review is most likely to belong to.

3. **Target Based Sentiment Classification.** After determining the target that a review belongs to, we can further know the emotion that the review may express rather than positive or negative simply. For example, if we know a review is concerned about after-sales service of book, we can roughly determine that the possible emotions of this review may be "happy" or "angry"; if we know a review is concerned about the content of a book, we can roughly determine that the possible emotions may be "love" or "disgust". An emotion annotation on a book review corpus [12] also demonstrates this. So review classification based on target provides a way to look at sentiment in the terms of emotional categories such as "angry", "disgust" or "happy".

Obtaining the opinion words appropriate for a target is important for target based sentiment classification. For example, the positive opinion words for expressing the target of "内容 (content)" of a book may include "丰富 (rich)", "精彩 (wonderful)", the negative opinion words for expressing this target may include "赘述 (redundancy)", "空洞 (empty)"; but the positive opinion words for expressing the target of "服务 (service)" may include "满意 (satisfy)", "耐心 (patient)", the negative opinion words may include "糟糕 (terrible)", "投诉 (complaints)".

In this section, we describe the extraction method of opinion words on different targets. For a certain target, the extended feature words are the words relating to this feature, including the opinion words for expressing this feature. So the problem here is to extract the opinion words from the extended feature words and identify their sentiment orientation.

3.1. **General sentiment lexicons.** As the majority of the literature on sentiment analysis has focused on text written in English, and thus currently, most available sentiment analysis resources are for the English language, such as GI [16], WordNet-Affect [15], NTU Sentiment Dictionary [17], SentiWordnet [18]. Hence, it is currently difficult to analyze opinions written in Chinese. Although Hownet [19] is well used as a resource for the task of Chinese sentiment classification, Quan and Ren showed that, in Hownet sentiment lexicon, most of words occur rarely in real use of language [20]. Since a lot of new words used with high frequency are not included, it is not suitable for sentiment analysis of Internet resources. Therefore, we use a Chinese emotion corpus developed by Ren lab (Ren-CECps) [21] for Chinese sentiment classification.

Ren-CECps consists of 1487 blog articles published at the mainstream blog websites. There are 35,096 sentences, and 878,164 Chinese words contained in this corpus. The emotional words in this corpus are annotated with emotions included in the set of {expect, joy, love, surprise, anxiety, sorrow, angry and hate} and emotion intensities (range from 0.1 to 1.0). Based on this corpus, we can get the general sentiment lexicons. The words with emotions of joy or love are considered as positive opinion words, but the words with emotions of anxiety, sorrow, angry or hate are considered as negative opinion words.

3.2. **Getting target based opinion words and identifying word sentiment orientation.** In Section 2.1, we have described PMI-TFIDF for measuring the association of words for extending related features. In this section, we use PMI-TFIDF to measure the similarities between each word in the extended feature word list and the general sentiment lexicons for identifying target based opinion words and word sentiment orientation. See Equations (12)-(16).

$$\text{Extended\_feature\_word\_list}_f = \{f_1, f_2, \ldots, f_n\} \tag{12}$$

$$\text{Sentiment\_lexicon}_{pos} = \{p_1, p_2, \ldots, p_m\} \tag{13}$$

$$\text{Sentiment\_lexicon}_{neg} = \{n_1, n_2, \ldots, n_t\} \tag{14}$$

$$\mathrm{sim}_{pos}(f_i, \mathrm{Sentiment\_lexicon}_{pos}) = \frac{1}{freq(f_i)} \sum_{j=1}^{m} Sim\,(f_i, p_j) \tag{15}$$

where Extended_feature_word_list$_f$ in Equation (12) is the extended feature word list for target t; Sentiment_lexicon$_{pos}$ in Equation (13) is the general positive sentiment lexicon extracted from Ren-CECps; Sentiment_lexicon$_{neg}$ in Equation (14) is the general negative sentiment lexicon extracted from Ren-CECps; $\mathrm{sim}_{pos}(f_i, \mathrm{Sentiment\_lexicon}_{pos})$ in Equation (15) is the similarity between feature word f_i and the general positive sentiment lexicon, where $Sim\,(f_i, p_j)$ can be obtained by PMI-TFIDF measure, see Equation (5); $freq(f_i)$ is the frequency of feature word f_i in the corpus. Then Equation (15) can be rewritten by Equation (16).

$$\begin{aligned} &\mathrm{sim}_{pos}\,(f_i, \mathrm{Sentiment\_lexicon}_{pos}) \\ &= \frac{1}{freq(f_i)} \sum_{j=1}^{m} Sim\,(f_i, p_j) \\ &= \frac{1}{freq(f_i)} \sum_{j=1}^{m} PMI - TFIDF\,(f_i, p_j) \\ &= \frac{1}{freq(f_i)} \sum_{j=1}^{m} \log \frac{\Pr\,(f_i, p_j) \times \sum_k TFIDF\,(f_i, d_k) \times \sum_k TFIDF\,(p_j, d_k)}{\Pr(f_i)\Pr(p_j)} \end{aligned} \tag{16}$$

By ranking the similarities between feature word f_i and the general positive sentiment lexicon, we can obtain positive opinion word list for target t. In a similar way, we can also obtain negative opinion word list for feature f_i, see Equation (17).

$$\begin{aligned} &\mathrm{sim}_{neg}\,(f_i, \mathrm{Sentiment\_lexicon}_{neg}) \\ &= \frac{1}{freq(f_i)} \sum_{j=1}^{m} \log \frac{\Pr\,(f_i, n_j) \times \sum_k TFIDF\,(f_i, d_k) \times \sum_k TFIDF\,(n_j, d_k)}{\Pr(f_i)\Pr(n_j)} \end{aligned} \tag{17}$$

Based on the Polynomial Kernel (PK) method, the similarities between reviews and the opinion word lists (including positive and negative opinion word) can be obtained. By retrieving and comparing the similarity scores, we can get the sentiment orientation that a review is most likely to hold.

4. **The Experiments.**

4.1. **Experimental setup.**
Our experiments take book reviews (in Chinese) as data. This corpus is collected by tan [12], and it is composed by 4000 reviews on book, containing 510,650 Chinese characters. Each review has two tags: one is target tag, which includes three types: content, printing and binding quality; the other is sentiment tag, which includes three types: positive, negative and neutral attitudes.

The text preprocessing is Chinese word segmentation and stop words filtration. We use ICTCLAS (www.searchforum.org.cn), a Chinese word segmentation package for this step.

The following two experiments have been conducted:

(1) Review classification based on targets: the reviews are classified based on its target (content, printing and binding quality). At first, we use the measure of PMI-TFIDF to extend related feature words by measuring the associations between feature words and other words to get the related feature word list for the three targets. After that, Polynomial Kernel method is applied to compute the similarities between a review and

the related feature word lists of different targets. Then we can get the target that a review is classified

(2) Review sentiment classification based on targets: the reviews are classified based on its sentiment orientation (positive, negative, or neutral). We first get the general opinion word lexicons (including positive and negative) from Ren-CECps. The words are selected with higher intensity (>=0.7) (Table 4 shows examples of emotion words with different emotion intensities extracted from Ren-CECps). After that, target based opinion words are extracted from the extended feature words by comparing their similarities with the sentiment word lists from a Chinese emotion corpus (Ren-CECps). The sentiment orientation of a review is determined by Polynomial Kernel (PK) method. By retrieving and comparing the similarity scores, we can get the sentiment orientation that a review is most likely to hold.

4.2. **The experimental results.**

4.2.1. *Experimental results of extending related feature words by measuring the association of words.* We first compare the performance of PMI measure and the proposed PMI-TFIDF measure for extending related words. Given a target word "内容 (content)", Table 1 compares its related words measured by PMI and PMI-TFIDF.

The comparisons between PMI and PMI-TFIDF in Table 1 shows that the extracted features that measured by PMI-TFIDF have closer association with the targets than the extracted features measured by PMI. In addition, the association values measured by PMI-TFIDF showed a good distinction between different features. In contrast, the association values measured by PMI showed a poor distinction. The reason for this can be seen from the definition of PMI and PMI-TFIDF (see Equations (1) and (5)). In a small corpus, most words appear only once or twice, that means the probability $\Pr(t, w_i)$ of a sentence containing target t and the word w_i in this corpus very may be equal to the probability $\Pr(w_i)$ of a sentence containing target t in this corpus. In such situation, the value of PMI would be only determined by $\Pr(t)$, so many words will have the same PMI value.

TABLE 1. Related words of target "内容 (content)" measured by PMI and PMI-TFIDF

PMI measure		PMI-TFIDF measure	
Related words	PMI values	Related words	PMI-TFIDF values
本月(this month)	5.355	内容(content)	13.256
胎教(prenatal education)	5.132	一般(general)	12.363
中小学生 (Primary and secondary students)	5.132	质量(quality)	11.056
大好(excellent)	4.844	页(pages)	11.046
金属(metal)	4.844	不错(not bad)	10.678
赶集(go to market)	4.844	感觉(feel)	10.655
机床(machine tool)	4.844	买(buy)	10.622
摄影师(photographer)	4.844	没什么用(no use)	10.577
关子(climax of a drama or novel)	4.844	实用(practical)	10.555
要素(elements)	4.844	重复(repeat)	10.543
赘述(redundancy)	4.844	写(write)	10.487
眼皮(eyelid)	4.844	介绍(introduce)	10.432
亲笔(write by oneself)	4.844	作者(writer)	10.363
地点(location)	4.844	本书(this book)	10.357
乐于(be happy)	4.844	适合(suitable)	10.356
累积(accumulation)	4.844	简单(simple)	10.355

Table 2. Related words of “内容 (content)” , “服务 (service)” and “装订 (binding)” with higher PMI-TFIDF value

内容(content)		服务(service)		装订(binding)	
Related words	PMI-TFIDF values	Related words	PMI-TFIDF values	Related words	PMI-TFIDF values
内容(content)	13.256	服务(service)	12.346	装订(binding)	12.040
一般(general)	12.363	货(goods)	10.945	中间(middle)	10.371
质量(quality)	11.056	快递(express delivery)	10.859	差(lack)	10.191
页(pages)	11.046	差(poor)	10.709	郁闷(boring)	10.166
不错(not bad)	10.678	收到(receive)	10.628	锯齿(sawtooth)	10.131
感觉(feel)	10.655	显示(show)	10.298	页(pages)	10.076
买(buy)	10.622	速度(speed)	9.915	精美(fine)	9.485
没什么用(no use)	10.577	付款(pay)	9.887	钉(nail)	9.417
实用(practical)	10.555	送(send)	9.822	损坏(damage)	9.102
重复(repeat)	10.543	态度(attitude)	9.801	整齐(neat)	8.982
写(write)	10.487	抱歉(apologize)	9.643	错误(error)	8.945
介绍(introduce)	10.432	退货(return)	9.621	正版(authorized edition)	8.933
作者(writer)	10.363	糟糕(terrible)	9.430	缝(sew)	8.924
本书(this book)	10.357	打电话(make a phone)	9.397	皮(cover)	8.866
适合(suitable)	10.356	询问(ask)	9.227	纸面(paper)	8.768
简单(simple)	10.355	投诉(complaints)	9.153	割伤(cut)	8.724
丰富(rich)	10.18	满意(satisfy)	9.090	封底(back cover)	8.707
专业(professional)	9.946	顾客(customer)	9.077	印刷(print)	8.429

Table 3. The experimental results of review classification based on targets by Polynomial Kernel (PK) method

PMI-TFIDF value	Num. of Related words			prec(%)
	内容 (content)	服务 (service)	装订 (binding)	
>=0.0	1425	326	503	78.8
>=6.0	1090	272	90	80.3
>=7.0	**589**	**180**	**68**	**82.5**
>=8.0	180	52	26	82.2

In larger corpus, the effectiveness of PMI would gradually emerge. However, acquiring a large domain review corpus is a very high demand for some practical applications. So it is not suitable for the feature extraction task based on a small corpus. We also find that the related words to the target “内容 (content)” include nouns, adjectives, verbs, etc., that extends the scope of related words to a target.

Table 2 gives the related words of target “内容 (content)”, “服务 (service)”, “装订 (binding)” with higher PMI-TFIDF value.

4.2.2. *Experimental results of classifying reviews based on targets.* Table 3 shows the experimental results of review classification based on targets by Polynomial Kernel (PK) method. The empirical parameters of Polynomial Kernel are set by: $c = 0$, $d = 0.5$ (see Equation (11)). The performances are evaluated by precision measure.

PMI-TFIDF value is used to control the number of related words of a target. As shown in Table 3, the highest precision is obtained when the PMI-TFIDF value is above or equal 7.0, which means the results is sensitive to PMI-TFIDF value, and the number of related words of a target is not the much the better.

4.2.3. *Experimental results of target based sentiment classification.* To compare the performances of using general sentiment lexicons and using target based opinion word lists for target based sentiment classification, we first experiment the use of general sentiment lexicons for this task. Table 4 gives some examples of emotional words with different emotion intensities extracted from Ren-CECps.

TABLE 4. Examples of emotion words with different intensities

Emotion intensity=0.7		**Emotion intensity=0.8**		**Emotion intensity=0.9**		**Emotion intensity=1.0**	
POS	**NEG**	POS	**NEG**	POS	**NEG**	POS	**NEG**
满足 (meet)	反思 (reflection)	值得 (deserve)	陈词滥调 (cliche)	有意思 (interesting)	欺骗 (deceive)	畅销 (sell well)	愤恨 (resentment)
适应 (adapt)	偏见 (bias)	安慰 (comfort)	盗版 (piracy)	感动 (moving)	自私 (selfish)	开心 (happy)	大吵 (kicked up)
教育 (educate)	不足 (inadequate)	便利 (convenient)	借口 (excuse)	创意 (creative)	生气 (angry)	欢呼 (cheer)	让人抓狂 (makes you go crazy)
理解 (understand)	忽视 (neglect)	心曲 (heart songs)	抄袭 (plagiarism)	实用 (practical)	讨厌 (hate)	难忘 (memorable)	骂声滔天 (a myriad of condemning)
引导 (guide)	世故 (sophisticated)	特色 (characteristics)	不满 (dissatisfied)	丰富(rich)	讽刺 (satire)	喜欢 (like)	不可饶恕 (unforgivable)

TABLE 5. The experimental results of review sentiment classification by Polynomial Kernel (PK) method on general sentiment lexicons

Emotion intensity	**Num. of words in sentiment lexicons**		**Prec(%) on different targets**		
	POS	**NEG**	**内容(content)**	**服务(service)**	**装订(binding)**
>=0.7	4710	4406	46.5	60.84	58.65
>=0.8	2739	2513	58.45	77.67	67.67
>=0.9	852	777	61.44	78.96	72.9
=1.0	276	289	**70.65**	**89.64**	**86.47**

TABLE 6. Target based sentiment words of "内容 (content)" , "服务 (service)" and "装订 (binding)"

内容(content)		**服务(service)**		**装订(binding)**	
Positive words	**Negative words**	**Positive words**	**Negative words**	**Positive words**	**Negative words**
丰富(rich)	空洞(empty)	满意(satisfy)	糟糕(terrible)	整齐(neat)	差(lack)
精彩(wonderful)	赘述 (redundancy)	卓越 (excellent)	无耻 (shameless)	舒服 (comfortable)	破(break)
实用(practical)	没什么用(no use)	信任(trust)	投诉 (complaints)	漂亮 (beautiful)	错误(error)
新鲜(fresh)	不通(barrier)	好听(nice)	差(poor)	精美(fine)	危险(danger)
好看 (good-looking)	流水账(running account)	及时(timely)	敷衍了事 (half-heartedly)	清楚(clear)	割伤(cuts)
喜爱(favorite)	吹捧(flatter)	耐心(patient)	小人(villain)	鲜艳(bright)	损坏(damage)
严谨(rigorous)	罗嗦(wordy)	精细(fine)	简直(at all)	精致(fine)	倒置(invert)
激发(excite)	废话(nonsense)	快(quick)	推卸(shirk)	清晰(clear)	粗糙(coarse)
幽默(humorous)	名不符实 (fabled)	赞(praise)	骗人(deceive)	结实(solid)	裂(crack)

Table 7. The experimental results of review sentiment classification by Polynomial Kernel (PK) method on target based sentiment lexicons

Target	Num. of words in sentiment lexicons		Prec(%)
	POS	NEG	
内容(content)	89	92	75.65
服务(service)	46	39	92.12
装订(binding)	23	26	89.56

As shown in Table 5, the higher emotion intensity of opinion words, the higher precision is. The results demonstrate that the performance of sentiment classification is sensitive to sentiment lexicon. The factors of emotion intensity of opinion words, the number of words in sentiment lexicons, and the proportion of positive and negative words can affect the performance.

Based on these results, the emotion words with intensity value of 1.0 are used as general Sentiment_lexicon in Equation (15) to get target based sentiment words. Table 6 gives some examples of the positive and negative opinion words of target "内容 (content)" , "服务 (service)", "装订 (binding)" with higher similarities with general sentiment lexicons.

After obtaining the positive and negative opinion word list for target t, we can get the sentiment orientation that a review is most likely to hold, based on the Polynomial Kernel (PK) method. Table 7 shows the experimental results of sentiment classification by using target based sentiment lexicons.

As shown in Table 7, the experimental results of using target based sentiment lexicons is much higher than using general sentiment lexicons. The precision scores of the targets on"内容 (content)", "服务 (service)" and "装订 (binding)", gained 5.0%, 2.5% and 3.0% increase respectively, which demonstrates the effectiveness of using target based sentiment lexicons.

We also find that the precision for the target of "内容 (content)" is lower than "服务 (service)" and "装订 (binding)". This is because "内容 (content)" is more abstract than "服务 (service)" and "装订 (binding)", that is, there are more expressions for such a target. In contrast, the expressions for "服务 (service)" and "装订 (binding)" are relatively less, so they are easier to be identified.

Conducting an error analysis, we find that the reviews containing negative words, such as "不 (not)", "没法 (cannot)" "并非 (be not)" are difficult to be classified. Negative expressions in reviews are still an open problem for sentiment analysis because the use of negation is so flexible in natural language. As an example of sentence (7):

我不喜欢这本书，也不讨厌。(I do not like this book, and not hate.)

There are two negative words "不 (not)" and two opinion words "喜欢 (like)" and " 讨厌 (hate)", but we still cannot understand what the reviewer's opinion without more contexts.

Some other errors are due to the unclear opinions by reviewers, for example of sentence (8):

(8) 越看心情越沉重，推荐大家购买。(The more I read, the heavier I feel, recommend a purchase.)

The first sentence of this review expresses a negative opinion, but the second sentence expresses a positive opinion. The overall opinion of this review may be positive because we usually think that a review expresses the overall opinion at the end of this review. In this case, a strategy of adding weight on each sentence of a review (for example, adding higher weight on the last sentence) may help to recognize its opinion.

5. **Conclusions and Future Work.** Internet has dramatically changed the way that people express their opinions, and has made it possible for a company to find consumer opinions about its products and those of its competitors by collecting and analyzing the user-generated content on the Web. Numerous companies have a lot of demands on sentiment analysis and have been working on it.

Sentiment analysis has traditionally been performed using technology that evaluates an article by judging its sentiment polarity, which is not sufficient for many customer review analysis. Target-based sentiment analysis provides more fine grained sentiment analysis. In this paper, we proposed a new measure of PMI-TFIDF by combining PMI and TF-IDF to measure the association of words for extending related features for a target. The comparisons between PMI and PMI-TFIDF showed that the extracted features that measured by PMI-TFIDF had closer association with the targets than the extracted features measured by PMI. In addition, the association values measured by PMI-TFIDF showed better distinction between different features. With the extended feature words, Polynomial Kernel (PK) method was applied to classify reviews based on targets.

After determining reviews relating to targets, targets based opinion words were extracted from the extended feature words by comparing their similarities with the opinion word lists from a Chinese emotion corpus (Ren-CECps). The sentiment orientation of a review was determined by comparing their similarities with the target based opinion words. The experimental results demonstrated that the performance of sentiment classification is sensitive to sentiment lexicons. So the quality of sentiment lexicons can affect the system performance. The experiments also showed the effectiveness of using target based sentiment lexicons for this task.

We also find that it is more difficult to classify reviews for abstract targets and the reviews containing negative words. Therefore, in our future work, we will consider to improve the current system performance by negation analysis in opinions and weighting each sentence of a review.

In the future, we plan to apply this method for the problem of feature-level sentiment analysis. After that, we plan to further extend this method for more applications such as product ontology construction and product attribute analysis.

Acknowledgment. This research has been partially supported by National Natural Science Foundation of China under Grant No. 61203312, National High-Tech Research and Development Program of China 863 Program under Grant No. 2012AA011103, the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, Key Science and Technology Program of Anhui Province, under Grant No. 1206c0805039, and the Ministry of Education, Science, Sports and Culture of Japan under Grant-in-Aid for Scientific Research (A) No. 22240021.

REFERENCES

[1] V. Hatzivassiloglou and K. McKeown, Predicting the semantic orientation of adjectives, *Proc. of the 40th Annual Meeting on Association for Computational Linguistic*, pp.174-181, 1997.

[2] B. Pang, L. Lee and S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, *Proc. of the 2002 Conference on Empirical Methods in Natural Language Processing*, 2002.

[3] P. Turney and M. Littman, Measuring praise and criticism: Inference of semantic orientation from association, *ACM Transactions on Information Systems*, vol.21, no.4, pp.315-346, 2003.

[4] B. Pang and L. Lee, Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, *Proc. of the 49th Annual Meeting on Association for Computational Linguistic*, 2005.

[5] C. Quan, F. Ren and T. He, Sentiment classification based on kernel methods, *International Journal of Innovative Computing, Information and Control*, vol.6, no.6, pp.2681-2690, 2010.
[6] X. Ding and B. Liu, The utility of linguistic rules in opinion mining, *Proc. of SIGIR-2007 (Poster Paper)*, pp.23-27, 2007.
[7] M. Hu and B. Liu, Mining and summarizing customer reviews, *Proc. of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2004.
[8] A. Popescu and O. Etzioni, Extracting product features and opinions from reviews, *Proc. of the HLT-EMNLP*, 2005.
[9] L. Jiang, M. Yu, M. Zhou, X. Liu and T. Zhao, Target-dependent twitter sentiment classification, *Proc. of the 55th Annual Meeting on Association for Computational Linguistic*, pp.151-160, 2011.
[10] M. Hu and B. Liu, Mining opinion features in customer reviews, *Proc. of AAAI*, 2004.
[11] W. Du and S. Tan, An iterative reinforcement approach for fine-grained opinion mining, *Proc. of the 53th Annual Meeting on Association for Computational Linguistic*, pp.486-493, 2009.
[12] *http://www.searchforum.org.cn/tansongbo/corpus-senti.htm.*
[13] G. Bouma, Normalized (pointwise) mutual information in collocation extraction, from form to meaning: Processing texts automatically, *Proc. of the Biennial GSCL Conference*, pp.31-40, 2009.
[14] G. Salton and C. Buckley, Term-weighting approaches in automatic text retrieval, *Information Processing & Managemen*, vol.24, no.5, pp.513-523, 1998.
[15] C. Strapparava and A. Valitutti, Wordnet-affect: An affective extension of wordnet, *Proc. of the 4th International Conference on Language Resources and Evaluation*, pp.1083-1086, 2004.
[16] P. Stone, D. Dunphy, M. S. Smith and D. M. Ogilvie, *The General Inquirer: A Computer Approach to Content Analysis*, MIT Press, 1966.
[17] L. Ku, Y. Liang and H. Chen, Tagging heterogeneous evaluation corpora for opinionated tasks, *Proc. of the 6th International Conference on Language Resources and Evaluation*, pp.667-670, 2006.
[18] A. Esuli and F. Sebastiani, SentiWordNet: A publicly available lexical resource for opinion mining, *Proc. of the 6th International Conference on Language Resources and Evaluation*, pp.417-422, 2006.
[19] Z. Dong and Q. Dong, *HowNet*, http://www.keenage.com, 1999.
[20] C. Quan and F. Ren, Recognizing sentence emotions based on polynomial kernel method using Ren-CECps, *Proc. of the IEEE NLP-KE2009*, pp.478-484, 2009.
[21] C. Quan and F. Ren, A blog emotion corpus for emotional expression analysis in Chinese, *Computer Speech and Language*, vol.24, no.4, pp.726-749, 2010.