

# Sentiment Analysis with Multi-source Product Reviews

Hongwei Jin, Minlie Huang, and Xiaoyan Zhu

State Key Laboratory of Intelligent Technology and Systems  
Tsinghua National Laboratory for Information Science and Technology  
Department of Computer Science and Technology, Tsinghua University  
Beijing, 100084, China  
cs.jin.hongwei@gmail.com, {aihuang, zxy-dcs}@tsinghua.edu.cn

**Abstract.** More and more product reviews emerge on E-commerce sites and microblog systems nowadays. This information is useful for consumers to know the others' opinion on the products before purchasing, or companies who want to learn the public sentiment of their products. In order to effectively utilize this information, this paper has done some sentiment analysis on these multi-source reviews. For one thing, a binary classification framework based on the aspects of product is proposed. Both explicit and implicit aspect is considered and multiple kinds of feature weighing and classifiers are compared in our framework. For another, we use several machine learning algorithms to classify the product reviews in microblog systems into positive, negative and neutral classes, and find OVA-SVMs perform best. Part of our work in this paper has been applied in a Chinese Product Review Mining System.

**Keywords:** product review, sentiment analysis, microblog, SVM.

## 1 Introduction

With the development of Internet, more and more customers get used to purchasing products on E-commerce sites such as 360buy<sup>1</sup> and Newegg<sup>2</sup>. They also write reviews on the products after using them, which produce a large number of reviews on the Internet. In addition, microblog, a system that allows users to post messages of no more than 140 words, and share information instantaneously based on the relationship between users, is under rapid development, such as Twitter<sup>3</sup>, Sina microblog<sup>4</sup> and Tencent microblog<sup>5</sup>. A lot of microblogs contain latest product reviews.

Reviews from the above two large data sources contain much useful information for users and companies. Users can make better purchasing decisions based on these reviews, while companies can also analyze customers' satisfaction according to these reviews, and further improve the quality of their products. Since there is a mass of

---

<sup>1</sup> <http://www.360buy.com>

<sup>2</sup> <http://www.newegg.com.cn>

<sup>3</sup> <http://www.twitter.com>

<sup>4</sup> <http://weibo.com>

<sup>5</sup> <http://t.qq.cn>

product reviews and a single user cannot read all of them, automatically mining the reviews from multiple sources is particularly important.

Most reviews in Chinese E-commerce sites are labeled with advantage or disadvantage, which is naturally suitable for binary classification. The state-of-art research in Chinese sentiment analysis mainly focuses on the whole review classification, while customers often desire a more detailed understanding of products. For example, they want to know others' opinion on the battery of a cell phone. Therefore, we propose a framework of sentiment classification at aspect level to solve this problem. In our framework, not only explicit but also implicit aspects are taken into account. To our knowledge, no implicit aspect discovery work of product review in Chinese language has been reported before. For the reviews of each aspect, the unigram features of words are used as text features. We also compare the performance of three feature weighing strategies, three reduction dimension, and three classification approaches.

The sentiment analysis for reviews of products on microblogs is in its infancy. Besides the microblogs that express opinion on the products, some microblogs only give some statements relative to the products, which contain no sentiment polarity, or are neutral. Therefore, in this paper, we exploit linear regression, multi-class classification, two-stage classification and Mincut model optimization to classify the product related microblogs into three classes, and compare the performance of these methods.

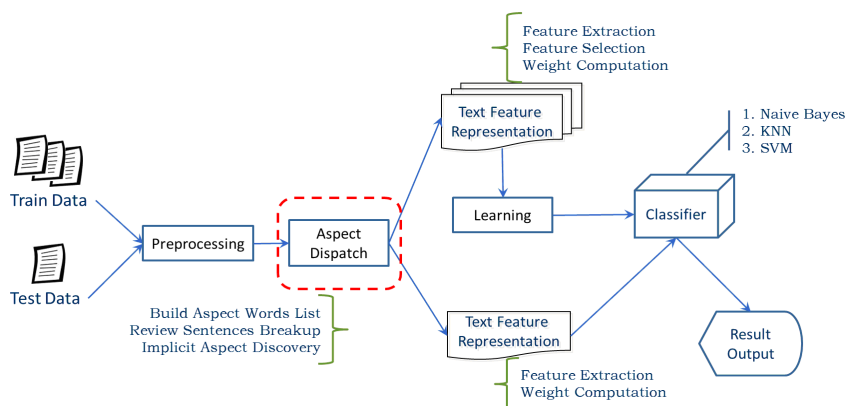
## 2 Related Work

Sentiment classification mainly includes two methods: supervised learning and unsupervised learning. Turney[1] proposed a simple unsupervised method to classify reviews into positive and negative categories. Pang[2] used Naive Bayes (NB), Max entropy (ME) and SVM separately as supervised learning algorithms for binary classification of reviews. Besides these, Pang[3] exploited regression and One-Vs-All SVMs to predict the score for classification, namely the multi-classification, which can also be realized by combining binary classifiers in a two-stage manner. Pang[4] and Su[5] both optimized their multi-classification results using the Mincut model.

For sentiment classification in microblogs, Go[6] was the first to classify Twitter data into two classes (positive and negative). Barbosa[7] classified Twitter data into three classes in a two-stage manner. Jiang[8] analyzed topic related sentiment for Twitter text. Based on the two-stage binary classification result, and the forwarding relationship between users, the performance was promoted using the method of graph. But for Chinese microblogs systems, only text can be obtained instead of the relationship between users and microblogs.

## 3 Binary Sentiment Classification Based on Product Aspect

The overall flow chart about polarity classification (positive and negative) based on product aspect is shown in Fig.1.



**Fig. 1.** The polarity classification (positive and negative) based on product aspect framework

**Aspect, Aspect Word.** Product aspect refers to some attribute, component, or function of the product. The word or phrase used to represent the aspect of product is called the aspect word. Aspect is actually the concept at semantic level, and the aspect word is the external presentation form of the aspect. Different categories of products generally have a different set of aspects. Product aspect can be divided into explicit aspect and implicit aspect. Explicit aspect refers to the aspect word which describes the product performance or function can be directly found in the product reviews. Implicit aspect may be found after the sentence semantic understanding because there are no aspect words in review sentence.

### 3.1 Preprocessing

In order to get relatively clean text about product reviews, messy code will be removed in the preprocessing stage according to a user-defined messy code list. Next we use the word segmentation software tools to make the sentence processed. There are some words that appear in almost all of the text which are called stop words. The text content is represented more accurately after deleting stop words. Some words in the traditional stop words list is retained because they have sentiment polarity which is useful information for sentiment classification.

### 3.2 Build Aspect Words List

We use a statistics-based algorithm to automatically build the vocabulary about the product aspects. Here is the algorithm:

1. Count occurrences number of all the nominal words.

We count the occurrences number of all the unigram whose part of speech (POS) tagging is noun in the reviews of same product category.

2. Filter high-frequency words

The final aspect words list is filtered by threshold of high frequency and product specifications.

### 3.3 Review Sentences Breakup

First of all, we break a review into sentences according to various sentence end punctuations, each of which is further cut into short sentences (SS) with comma. Then each SS is assigned to different sentiment categories of different aspects according to the aspect vocabulary made in advance. Finally, if an SS has implicit aspect, its words should also be correctly distributed.

### 3.4 Implicit Aspect Discovery

For example, if there is a word "beautiful" in the sentence and without any explicit aspect word appeared, we also need to know "beautiful" is likely in the description of the product's appearance which is just the implicit aspect of product.

The discovery of implicit aspects needs to balance the following factors:

1. Calculate mutual information (MI) scores between adjectives in the current SS and each aspect. Adjectives with most relevant aspects will be calculated and the aspect of the entire SS is determined by voting.
2. If the above rule doesn't work, the SS is just assigned to the previous SS's aspect.

### 3.5 Text Feature Representation

After the processing mentioned above, we get all the review sentences in each aspect in each category, and vector space model is employed to represent them.

Features of the review text using word-based unigram, but different word have the different ability and importance to represent the text which is called weight. In this paper, BOOL, TF, and TF-IDF are used as feature weights. BOOL weight means that if the count of the term appears in the current document greater than zero, the weight is 1, otherwise 0. TF (term frequency) refers to the frequency of the words occurrence in the file. IDF (inverse document frequency) measure the common importance of the word. In addition, we use chi-square statistics to reduce feature dimension.

### 3.6 Classifier

**Naive Bayes.** Navie Bayes is a simple model which calculates posterior probability of each class based on the priori probability and the likelihood. Class with the largest posterior probability is assigned as the class of the document.

**KNN.** The basic idea of the KNN algorithm is considering the K nearest (the most K similar) texts in the training set of the new given text. The label of the new text is determined by these K texts.

**SVM.** SVM is a popular classification algorithm which compresses the raw data set to support vector set and learn a decision hyperplane in the vector space. This hyperplane best splits the data points into the two classes.

In this paper, NB, KNN and SVM are implemented using data mining tool Weka<sup>6</sup>.

---

<sup>6</sup> <http://www.cs.waikato.ac.nz/ml/weka/>

## 4 Sentiment Analysis on Product Reviews in Microblog System

### 4.1 Text Feature Extraction

Like mentioned in Section 3.1, stop words and messy code in the microblog will be removed, and then the Chinese word segmentation and POS tagging will be done. The vector space model is still employed to represent each microblog. The features are unigrams and the weight is BOOL value. However, we can take the advantage of characteristic of microblog to portray it better and reduce the feature dimension.

**Emoticons.** For example, the text form of emoticon 🤔 in microblog text is "[ha ha]". All the positive emoticons will be converted into token "POS", and all negative into token "NEG" according to the manual defined emotions list.

**Usernames.** Forwarding mechanism of microblog systems makes microblog amazing transmitted. It is in the form of @ + username (such as @ryanking1219). So we use string "USERNAME" to instead all words beginning with @.

**Links.** Users like to include the URL which usually begins with "http" when they share videos or news, such as <http://t.cn/aCKddG>. All website links are normalized to the token "URL".

**Topics.** In microblog systems, topic starts with "#" and also ends with "#". We will replace all the string in this form with an equivalence class string "TOPIC".

### 4.2 Two-Stage SVM

#### Step 1: Subjectivity Classifier

Neutral samples in the training data is considered as positive examples, while positive and negative samples as negative cases to train the classifier. Do binary classification on the test data for subjective and objective detection.

#### Step 2: Polarity Classifier

Positive samples in the training data is considered as positive examples, while negative samples as negative cases to train the classifier. Do binary classification on the test data which is correctly divided into subject class in Step 1 for positive and negative detection.

### 4.3 Minimum Cut Model Optimization

Inspired by work of Pang[4] and Su[5], we also use Minimum cut (Mincut) model to optimize the Two-stage SVM result. Binary classification with Mincut in graph is based on the idea that similar items should be split in the same cut. We build an undirected graph  $G$  with vertices  $\{s, t, v_1, v_2, \dots, v_n\}$ ;  $s$  is the source and  $t$  is the sink, and all items in the test data are seen as vertices  $v_i$ . Each vertex  $v$  is connected to  $s$  and  $t$  via a weighted edge. The weight is the estimation of the probability converted from SVM classifier output. Each edges  $(v_i, v_k)$  with weight  $assoc(v_i, v_k)$  expresses the

similarity between  $v_i$  and  $v_k$  and how important they should be in the same class. Then we remove a set of edges to divide graph into two disconnected sub-graphs. The vertices via  $s$  are positive instances and the vertices via  $t$  are negative instances. We penalize when putting highly similar items into different classes, so the best split is one which removes edges with lowest weights. This is exactly the Mincut of graph. Because the capacity of the Mincut equals the max flow value, we can employ Edmonds-Karp algorithm to compute the Mincut.

In experiment, we set

$$assoc(v_i, v_j) = \begin{cases} \alpha * \cos(v_i, v_j), & \text{if } \alpha * \cos(v_i, v_j) > t \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $\cos(v_i, v_j)$  is the text similarity between  $v_i$  and  $v_j$ . Constant  $\alpha$  is the scale factor of association scores. Only scores higher than threshold  $t$  will be taken into consideration.

#### 4.4 Regression

Text classification is also a problem that utilizes text features to predict a category label, and similar entries should have similar category labels. The most classic model is the linear regression function which is actually a linear weighted sum of all the features. The final score is rounded as a category label.

#### 4.5 OVA-SVMs

Although the original SVM is used to solve binary classification problem, we can combine several SVM classifiers to achieve multi-class classification, such as one-versus-all method (for short of OVA-SVMs). When Training each classifier we mark one category samples as positive examples, and all the rest samples as negative examples. So that  $k$  categories of samples can construct  $k$  SVM classifiers in turn. Class label of the test sample assigned to the class which has the largest classification function value.

### 5 Experiments

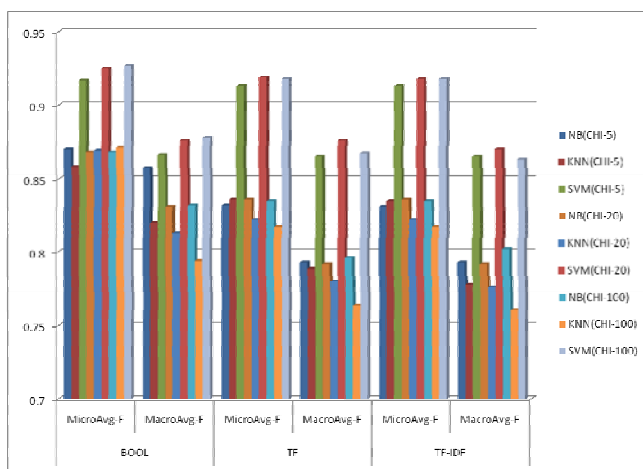
#### 5.1 Sentiment Classification Based on Product Aspect

We downloaded 821 cell phone products data from the 360buy and Newegg, and finally get 663,537 advantages reviews and 314,529 disadvantage reviews.

CHI dimension reduction selects 5%, 20% and 100% words most relevant to each aspect of each product category, represented as CHI-5, CHI-20, and CHI-100. Of course, from the whole view of each product category, the most relevant words of each aspect will be overlapped with each other.

Performance evaluation methods commonly apply PRECISION, RECALL, and F-values, but these only represent local significance (i.e., the performance of each aspect in each product category). If you want to evaluate the overall classification performance, all the aspects of product category need to be taken into account to calculate the F-value of the MICRO AVERAGING and MACRO AVERAGING.

Micro-average gives the same weight to each document in each product category while Macro-average gives the same weight to each aspect in each product category. Results are illustrated in Fig.2.



**Fig. 2.** Cell phone experiment result (17 aspects)

The following conclusions could be drawn from the experimental result:

First, the best combination is BOOL-SVM. Second, the consequent of TF is very close to the TF-IDF, but is slightly lower than BOOL. In addition, the SVM classification performance is better than NB and KNN. Last but not least, CHI-5 result is similar with CHI-20 and CHI-100, even sometimes CHI-5 perform best, so dimension reduction is proved effective.

## 5.2 Sentiment Analysis on Product Reviews in Microblog System

We select 5 popular items including "Nokia", "iphone 4s", "E72i", "Lenovo" and "Canon". For all of these items, we downloaded 2,100 microblogs through the API provided by Sina microblog and Tencent microblog from October 2011 to November 2011. We manually labeled all the microblogs and finally obtain 729 positive, 345 negative, and 1,026 neutral microblogs respectively. The results are listed in Table 1.

**Table 1.** Result for microblog classification

	Regression	OVA-SVMs	Two-stage SVM	Two-stage SVM + MinCut
Accuracy	51.6%	<b>64.5%</b>	62.6%	64.0%

As shown in Table 1, the OVA-SVMs classifier is best. Two-stage SVM classifier after optimization with the minimum cut model (when parameter  $\alpha = 0.2$ ,  $t = 0.5$ ) has 1.4% increase in performance, and reach almost the same effect of OVA-SVMs.

## 6 Conclusion

This paper makes sentiment analysis on product reviews from multiple data sources. Firstly, a binary sentiment classification framework based on the aspects of the product is proposed. On the granularity of aspect, we use unigram as review feature, and adopt BOOL weight and SVM classifier to get the best results. Secondly, we classify our product review in microblog systems into three classes. OVA-SVMs method offers the optimal result. Finally, Work in Section 3 has been partially applied in a Chinese Product Review Mining System<sup>7</sup>, and Section 4 will be applied soon.

## References

1. Turney, P.D.: Thumbs up or Thumbs down?: Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL 2002, pp. 417–424. Association for Computational Linguistics, Stroudsburg (2002)
2. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: Sentiment Classification Using Machine Learning Techniques. In: Proceedings of the ACL 2002 Conference on Empirical Methods in Natural Language Processing, EMNLP 2002, vol. 10, pp. 79–86. Association for Computational Linguistics, Stroudsburg (2002)
3. Pang, B., Lee, L.: Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL 2005, pp. 115–124. Association for Computational Linguistics, Stroudsburg (2005)
4. Pang, B., Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, ACL 2004. Association for Computational Linguistics, Stroudsburg (2004)
5. Su, F., Markert, K.: Subjectivity Recognition on Word Senses via Semi-supervised Mincut. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, NAACL 2009, pp. 1–9. Association for Computational Linguistics, Stroudsburg (2009)
6. Go, A., Bhayani, R., Huang, L.: Twitter Sentiment Classification Using Distant Supervision. Technical report, Stanford Digital Library Technologies Project (2009)
7. Barbosa, L., Feng, J.: Robust Sentiment Detection on Twitter from Biased and Noisy Data. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING 2010, pp. 36–44. Association for Computational Linguistics, Stroudsburg (2010)
8. Jiang, L., Yu, M., Zhou, M., Liu, X., Zhao, T.: Target-dependent Twitter Sentiment Classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, HLT 2011, vol. 1, pp. 151–160. Association for Computational Linguistics, Stroudsburg (2011)

---

<sup>7</sup> <http://166.111.138.18/cReviewMiner/>