Multi-granularity opinion comparison on Chinese and English reviews: a case study in IT product domain

Qingqing ZHOU[1, 2], Chengzhi ZHANG[1, 2, *], Rui XIA[3]

1. Department of Information Management, Nanjing University of Science and Technology, Nanjing, China 210094
2. Alibaba Research Center for Complex Sciences, Hangzhou Normal University, Hangzhou, China 311121
3. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China 210094

Abstract. To help manufacturers and customers to obtain the different opinions about a product from various regions or countries via E-commerce websites, it is useful to conduct comparison on reviews written in different languages. This paper focuses on this problem by combining document-level and aspect-level sentiment mining on reviews of the same type of products in two different languages. The main findings show that while reviews in both languages have some common satisfied or unsatisfied aspects, they differ in significant ways. For example, Chinese users are more likely to express positive feelings, while English users have more obvious brand preferences. In addition, Chinese and English users' most satisfied and unsatisfied aspects are different. Thus, different product design and marketing strategies are necessary. It should be further noted that our study methodology is language independent, and it could be applied potentially to cross-language review mining of other languages.

Keywords: review mining, opinion comparison, aspect-level sentiment analysis, cross-language sentiment analysis

1 Introduction

As today's business increasingly operates on a global scale, product design and marketing call for better understanding of customers across national borders. To this end, product reviews can often reveal the attitudes and needs of customers. As an example, two reviews of Apple iPhone 5s are shown in Figure 1. One custom is unhappy about the bad battery, while the other is satisfied for the good package. In fact, these two reviews are typical representatives of English and Chinese reviews, respectively. In general, there are obvious differences between users of different languages. They usually focus on different aspects and have different attitudes to the same products. However, the existing researches pay little attention to differences between reviews in different languages. It is neces-

[*] Corresponding author, E-mail: zhangcz@njust.edu.cn

adfa, p. 1, 2011.

★★☆☆☆ **Not happy,** Battery life is very disappointing. Uses at least 50% charge a day without a lot of use!	★★★★★ **正品，质里好，快递迅速，** 外包装规矩，正品，质量有保证。亚马逊快递迅速，服务好
(a) English review[1]	(b) Chinese review[2]

Fig. 1. Examples of English (a) and Chinese (b) reviews for Apple iPhone 5s

sary to conduct cross-language review analysis at multiple granularities.

In this paper, we conduct multi-granularity opinion comparison on English and Chinese reviews and try to get some sociologically meaningful results. Specifically, we conduct sentiment analysis and opinion comparison at both document and aspect levels on Chinese and English reviews in the IT product domain. We aim at finding preferences of different regions or cultures of the same type of products, so as to contribute potentially to product design, marketing and so on. We get three main interesting results from the perspectives of sentiment polarities (SP), product brands (PB), and frequent aspects (FA), respectively. All of these findings might be valuable information for product designers to enhance the product performance and marketing staff to develop effective marketing strategies.

The reminder of this paper is organized as follows. Section 2 reviews related works. Data collection and annotation are introduced in Section 3. Section 4 presents our framework for multi-granularity opinion comparison. Experimental results are provided in Section 5. Section 6 finally draws the conclusions.

2 Related Works

There are three types of works related to our study: document-level sentiment analysis, aspect-level sentiment analysis and cross-language sentiment analysis.

Document-level sentiment analysis is to predict whether the whole document expresses a positive sentiment or a negative one [1]. It can be conducted by two main types of methods: supervised and unsupervised learning methods. As it is a special text classification problem, many existing supervised learning methods can be applied [2], such as Mullen & Collier [3]. In order to enhance the accuracy of sentiment classification, some techniques specifically for sentiment classification have been proposed, such as Xia [4-7], Li [8] and so on. One of the most classical unsupervised learning approaches to sentiment classification is proposed by Turney [9]. Some other unsupervised learning methods are based on sentiment lexicons, such as Taboada [10], Denecke [11] and so on.

Rather than gathering isolated opinions about a whole item, users generally prefer to compare specific features of different products, so it is important to con-

[1] Extracted from http://www.amazon.com
[2] Extracted from http://www.amazon.cn

duct fine-grained aspect-level sentiment analysis [12]. It includes two subtasks: aspect extraction and aspect sentiment classification [13, 14]. For the aspect extraction, numerous works have been carried out. Hu & Liu used part-of-speech (POS) tagger to identify nouns and noun phrases as candidate aspects, and chose frequent ones as aspects [15]. Popescu & Etzioni [16] improved the methods with point wise mutual information. Moghaddam and Este [17] presented a set of design guidelines for aspect-based opinion mining by discussing a series of increasingly sophisticated LDA models. For aspect sentiment classification, Ding et al. [14] proposed a simple method based on lexicon and has been demonstrated to perform quite well in a large number of domains. Other works such as Thet [18], Gangemi [19], and Cambria [20] focus on the use of semantics to better infer the conceptual and affective information associated with opinions. In this paper, we use statistical methods to conduct document-level sentiment analysis, and lexical affinity methods for aspect-level sentiment analysis.

There are also some efforts on analyzing sentiment in different languages. Boiy and Moens [21] present machine learning experiments to gain insights into sentiment classification of sentences or statements in blogs, consumer reviews and news forums, written in English, Dutch and French. Wan [22] leveraged available English corpus for Chinese sentiment classification to study the problem of cross-lingual sentiment classification. Brooke et al. [23] has explored the adaptation of English resources and techniques for text sentiment analysis to Spanish.

Generally, researches of cross-language sentiment analysis are mainly conducted at the document level, with few at the aspect level. In this paper, we conduct multi-granularity sentiment analysis and opinion comparison for reviews in two different languages. We try to find the opinion differences through the perspective of sociology, such as brand preferences, concerned aspects, etc., so as to give some suggestions on its design and marketing.

3 Data

3.1 Data collection

We collected English and Chinese reviews from Amazon.com and Amazon.cn under the same type of products respectively. The corpora cover reviews of two kinds of IT products: digital camera (including reviews of Canon and Nikon) and smart phone (including Apple iPhone and Samsung). In total, we have collected 8, 934 Chinese reviews and 14,385 English reviews. In addition, when collecting and annotating the corpus, some obvious redundant examples were deleted. Finally, 8, 801 Chinese reviews and 14, 072 English reviews were left. The details are shown in Table 1.

3.2 Data annotation

In order to construct the training set, we tagged part of reviews manually. For

Table 1. Description of data collection

Products	Brands	#Chinese reviews	#English reviews
Digital camera	Cannon	2,245	2,550
	Nikon	2,172	2,419
Smart phone	Apple	1,805	3,578
	Sam Sung	2,579	5,525

Table 2. Description of reviews annotation

Class labels	#Chinese reviews	#English reviews
#positive	1,200	2,400
#negative	600	1,200
sum	1,800	3,600

Table 3. Cross-validation performance of the reviews annotation

Languages	Macro Recall	Macro Precision	Macro F1 value	Accuracy
Chinese reviews	0.9783	0.9783	0.9783	0.9223
English reviews	0.9983	0.9966	0.9975	0.9531

Chinese reviews, we have tagged 1,800 reviews manually. Among them, 1,200 reviews express a positive feeling towards the entity and 600 reviews express a negative one. For English reviews, we have manually tagged 3,600 reviews, with 2,400 positive ones and 1, 200 negative ones. The annotation details are shown in Table 2.

For the convenience and reliability of the further comparison, we conduct cross validation on the training set to test the performance. Evaluation metrics include Macro Recall, Macro Precision, Macro F1 value and Accuracy:

$$\mathrm{MacroR} = \frac{1}{n}\sum_{j=1}^{n} R_j \tag{1}$$

$$\mathrm{MacroP} = \frac{1}{n}\sum_{j=1}^{n} P_j \tag{2}$$

$$\mathrm{Macro}F_1 = 2 * MacroP * MacroR/(MacroP + MacroR) \tag{3}$$

$$\mathrm{Accuracy} = \#\mathrm{correct}/\mathrm{N} \tag{4}$$

where R_j means recall of class j, P_j means precision of class j, n denotes number of class. *#correct* denotes number of correct classification, and *N* means number of all reviews.

We employed linear SVM as the classification model. Specifically, we used the LibSVM with the default parameters to conduct experiments with 5-fold cross validation and present evaluation results in Table 3. From Table 3, we can find that the performances of both Chinese reviews and English reviews annotation are very good. Therefore, it is trustable to use it as training data to conduct sentiment analysis on the whole corpus.

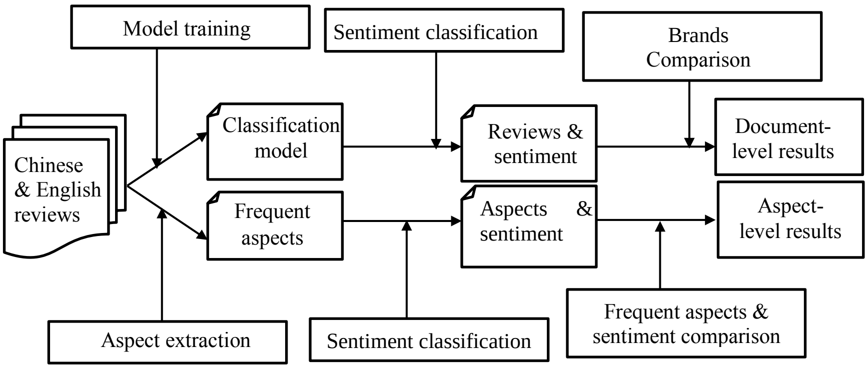

Fig. 2. Framework of opinion comparison of Chinese and English reviews

4 Methodology

4.1 Framework

We conducted document-level and aspect-level sentiment classification on English and Chinese reviews respectively. In the process of document-level sentiment classification, difference of opinions about the product between English and Chinese reviews can be analyzed. In the process of aspect-level sentiment classification, frequent aspects were selected and compared between reviews of the two languages. The framework of opinion comparison of Chinese and English reviews is shown in Figure 2.

4.2 Document-level Sentiment Classification

Most existing document-level classification use supervised learning methods. In this paper, we used linear SVM as the classification model to conduct document-level sentiment classification. Like other supervised learning applications, one of the key issues of sentiment classification is engineering of features selection. It is important to select features and compute their weight value efficiently. In this paper, we chose CHI as feature selection method and TF-IDF as feature weighting method. CHI is calculated with formula (5).

$$X^2(t,c) = \frac{N*(AD-BC)^2}{(A+C)(B+D)(A+B)(C+D)} \tag{5}$$

where A denotes the number of co-occurrence times of feature t and category c, B means the number of times t occurs without c, C denotes the number of times c occurs without t, D means c or t occurs, and $N=A+B+C+D$.

4.3 Aspect-level Sentiment Classification

Aspect-level sentiment analysis includes two important parts: aspect extraction and aspect sentiment classification. When people comment on different aspects of an entity, the vocabulary that they use usually converges. Thus, nouns that are frequently talked about are usually genuine and important aspects [1]. Therefore, we extract aspects of products by frequent nouns and nouns phrases. We use POS tagger[34] to find nouns or noun phrases as candidate aspects, then count their occurrence frequencies and choose frequent ones as aspects. We use IDF value to calculate frequency, if the IDF value is lower, it would be more frequent.

For every review, we can identify sentiment words via sentiment lexicons. As a review may express sentiments on multiple aspects, and it is hard to determine which aspect a sentiment word describes. In general, if the distance between an aspect word and a sentiment word is shorter, the sentiment word is more likely to describe the aspect. So we can compute the sentiment polarity of each aspect in a review by measuring the distance between an aspect word and a sentiment word. Sentiment polarity of aspect A in a review can be calculated via formula (6) [14].

$$\mathrm{Score(A)} = \sum_{w_i : w_i \in s \cap w_i \in V} \frac{w_{i.}SO}{dis(w_i, A)} \tag{6}$$

where w_i denotes a sentiment word, V means the set of all sentiment words and s is the review that contains the aspect A, and $dis(w_i, A)$ denotes the distance between aspect A and sentiment word w_i in the review s. $w_{i.}SO$ is the sentiment score of the word w_i. If word w_i is a positive word, $w_{i.}SO$ equals to +1, else it equals to -1. If $Score(A) > 0$, the sentiment polarity of aspect A in the review s is positive, else it is negative.

5 Experiments

5.1 Comparison on document-level sentiment analysis

5.1.1 Comparison on sentiment polarities of Chinese and English reviews

Sentiment classification was conducted with the SVM classifier, and the classification results are shown in Table 4. From Table 4, we can find that, for digital cameras, there are 4,417 Chinese reviews with 4,054 positive reviews and 363 negative ones. The number of English reviews is 4,969, among them, 4,697 reviews express positive feelings and 272 express negative feelings. The former one's ratio of positive reviews is 92%, while the latter one's is 94%, which mean that both Chinese users and English users are satisfied with the camera, while English ones' satisfaction tendency is a little higher. For smart phones, there are 4,384 Chinese reviews (include 3,607 positive reviews and 777 negative reviews)

[3] https://github.com/fxsjy/jieba
[4] http://www.nltk.org/

Table 4. Descriptive statistics of review sentiment polarities

Languages	Products	#Review	#Pos.	#Neg.	Ratio of Pos.
Chinese reviews	Camera	4,417	4,054	363	0.9178
	Smart phone	4,384	3,607	777	0.8227
English reviews	Camera	4,969	4,697	272	0.9452
	Smart phone	9,103	6,460	2,643	0.7096

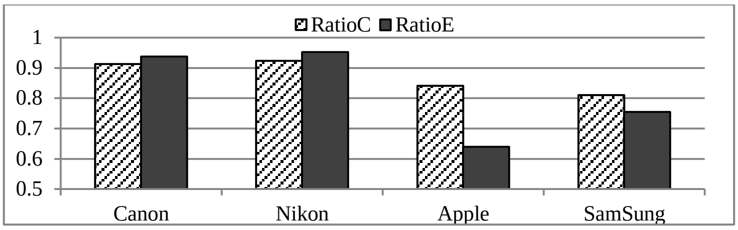

Fig. 3. Ratio of positive reviews of different brands (RatioC means ratios of Chinese positive reviews, while RatioE denotes English ones (the same below))

and 9,103 English reviews (include 6,460 positive ones and 2,643 negative ones). The former one's ratio of positive reviews is 82%, while the latter one's is 70%, which means that both Chinese users and English users are generally satisfied with the smart phone, while English users' ratio is significantly lower than Chinese users.

From the analysis above, we can draw the SP (sentiment polarity) conclusion: the amounts of positive reviews are much larger than negative ones both in Chinese and English reviews. In different fields, the possibilities of giving positive reviews are different, and Chinese users are more likely to express positive feelings.

5.1.2 Comparison on Chinese and English reviews of different brands

Figure 3 shows the ratios of positive reviews of different brands. From Figure 3, we can find that, for digital camera reviews, ratios of English positive reviews are higher than Chinese ones', while for smart phone reviews, ratios of English positive reviews are lower. In domain of digital camera, both Chinese and English users prefer Nikon to Canon. However, in field of smart phone, Chinese users prefer Apple, while English users more like Samsung.

In general, we can draw the PB (product brands) conclusion: English users have more obvious brand preferences than Chinese users.

From the above analysis, we can find that in the domain of smart phone, Samsung and Apple are both high-end phone brands, Chinese users do not care much about the specific aspects, so they have no obvious brand preference, while English users may prefer some aspects of Samsung. It may be related to differences in cultures or users' experiences in different countries or language areas. From our experiments results, we can find that English users pay more attention to details than Chinese ones.

Table 5. Top aspects of digital camera and smart phone reviews sort by numbers of reviews

Products	Chinese Reviews			English Reviews		
	Aspects	#Review	Pos. ratio	Aspects	#Review	Pos. ratio
Digital camera	package	704	0.8906	lens	2413	0.9428
	price	475	0.8484	quality	1598	0.9874
	lens	392	0.8392	flashlight	945	0.9714
	…	…	…	…	…	…
	aperture	17	0.8235	viewfinder	97	0.7804
	charger	14	0.5714	Memory	85	0.6434
	flashlight	14	0.2142	service	82	0.8352
Smart phone	logistics	502	0.8924	apps	3587	0.4502
	screen	473	0.8562	screen	3099	0.8144
	service	374	0.6417	battery	2405	0.8336
	…	….	…	…	…	…
	charger	45	0.7333	button	411	0.7761
	headphone	39	0.7692	network	281	0.6761
	camera	32	0.8125	resolution	176	0.8238

Table 6. Top aspects of digital camera and smartphone reviews sort by ratios of positive reviews

Products	Chinese Reviews			English Reviews		
	Aspects	#Review	Pos. ratio	Aspects	#Review	Pos. ratio
Digital camera	logistics	312	0.890625	quality	1598	0.9874
	package	704	0.848421	flashlight	945	0.9714
	button	20	0.839286	lens	2413	0.9428
	…	…	…	…	…	…
	charger	14	0.823529	viewfinder	97	0.7804
	Memory	25	0.571426	surface	155	0.7246
	flashlight	14	0.214286	memory	85	0.6434
Smart phone	logistics	502	0.8924	quality	1137	0.9683
	screen	473	0.8562	camera	1556	0.8650
	surface	220	0.8454	surface	1921	0.8625
	…	….	…	…	…	…
	price	343	0.7317	price	1474	0.7625
	service	374	0.6417	network	281	0.6761
	button	84	0.5357	apps	3587	0.4502

5.2 Comparison on aspect-level sentiment analysis

We selected top 12 aspects of smart phone reviews and top 15 aspects of digital camera reviews according to IDF values. Table 5 shows parts of frequent aspects of digital camera and smart phone reviews sort by numbers of reviews. For digital camera reviews, the most frequent aspect in Chinese reviews is package, followed by price and lens, while English users are most concerned about the lens, followed by quality and flashlight. For smart phone reviews, the most frequent aspect in Chinese reviews is logistics, followed by screen and service, while English users are most concerned about the apps, followed by screen and battery.

Table 6 shows parts of frequent aspects of digital camera and smart phone reviews sort by the ratios of positive reviews. For digital camera reviews, the most popular aspect of Chinese users is logistics, followed by package and button, while English users' popular aspects are quality, flashlight and lens. Both Chinese and English users are dissatisfied with memory. For smart phone reviews, Chinese users' most popular aspect is logistics, followed by screen and surface, while English users' popular aspects are quality, camera and surface. Both Chinese and English users are dissatisfied with price, while Chinese users' most unsatisfied aspect is button and English users are most dissatisfied with apps.

Generally, the FA (frequent aspects) conclusion can be drawn from above analysis: aspects that Chinese and English users concerned most are different, neither are their most satisfied and unsatisfied aspects.

From the analysis above, we can find that Chinese users concern surrounding conditions of products, such as logistics, packaging and so on, while English users care about the performances of products themselves, such as lens and quality etc. From the sociological point of view, Chinese users concern about external conditions, however, English users pay more attention to the internal conditions.

6 Conclusions

In this paper, we conducted multi-granularity sentiment analysis and opinion comparison for English and Chinese reviews in the IT product domain. Three main conclusions can be drawn according to our above mentioned comparison and analysis:

(1) **SP (sentiment polarity) conclusion**: Chinese users are more likely to pose positive reviews than English users do;

(2) **PB (product brands) conclusion**: English users have more obvious brand preferences than Chinese users in both digital camera and smart phone fields;

(3) **FA (frequent aspects) conclusion**: Chinese and English users' most satisfied and unsatisfied aspects are different. Neither are their most concerned aspects.

According to the three conclusions, we may infer some differences between Chinese and English consuming habits in the perspective of sociology. For example, the Chinese customers incline to focus on external conditions, while the English customers tend to care about the internal conditions and pay careful attention to details. Product designers and marketing personnel are thus required to pay attention to the habits and preferences of users from different regions. For example, for those aspects that are dissatisfied by both Chinese and English users, improvements are needed in the process of product design and production. For those aspects that are dissatisfied by only Chinese or English users, personalized marketing strategies are needed.

Acknowledgments

This work is supported by the Major Projects of National Social Science Fund (No. 13&ZD174), the National Social Science Fund Project (No. 14BTQ033), the National Natural Science Foundation of China (No. 61305090) and the Opening Foundation of Alibaba Research Center for Complex Sciences, Hangzhou Normal University (No. PD12001003002003).

References

1. Liu B. Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1): 1-167, 2012.
2. Pang B, Lee L, Vaithyanathan S. Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing, pp. 79-86, 2002.
3. Mullen T, Collier N. Sentiment Analysis using Support Vector Machines with Diverse Information Sources. In: Proceedings of EMNLP, pp. 412-418, 2004.
4. Xia R, Zong C, Li S. Ensemble of feature sets and classification algorithms for sentiment classification. Information Sciences. 181(6): 1138-1152, 2011.
5. Xia R, Zong C. Exploring the use of word relation features for sentiment classification. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pp. 1336-1344, 2010.
6. Xia R, Wang T, Hu X, Li S, Zong C. Dual Training and Dual Prediction for Polarity Classification. In: Proceedings of 51st Annual Meeting of the Association for Computational Linguistics, pp. 521-525, 2013.
7. Xia R, Zong C, Hu X, Cambria E. Feature Ensemble plus Sample Selection: Domain Adaptation for Sentiment Classification. IEEE Intelligent Systems, 28(3), pp. 10-18, 2013.
8. Li S, Lee SYM, Chen Y, Huang C-R, Zhou G. Sentiment classification and polarity shifting. In: Proceedings of the 23rd International Conference on Computational Linguistics, pp. 635-643, 2010.
9. Turney PD. 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, pp. 417-424, 2002.
10. Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. Computational linguistics. 37(2): 267-307, 2011.
11. Denecke K. Using sentiwordnet for multilingual sentiment analysis. In: Proceedings of 24th International Conference on Data Engineering Workshop, pp. 507-512, 2008.
12. Cambria E, Hussain A, Havasi C, Eckl C, Munro J. Towards crowd validation of the UK national health service. In: Proceedings of WebSci, 2010.
13. Yu J, Zha ZJ, Wang M, Chua TS. Aspect ranking: identifying important product aspects from online consumer reviews. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 1496-1505, 2011.
14. Ding X, Liu B, Yu PS. A holistic lexicon-based approach to opinion mining. In: Proceedings of the International Conference on Web Search and Data Mining, pp. 231-240, 2008.

15. Hu M, Liu B. Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining; pp. 168-177, 2004.
16. Popescu AM, Etzioni O. Extracting product features and opinions from reviews. Natural language processing and text mining, Springer London; 9-28, 2007.
17. Moghaddam S, Ester M. On the design of LDA models for aspect-based opinion mining. In: Proceedings of the 21st ACM international conference on Information and knowledge management, pp. 803-812, 2012.
18. Thet TT, Na JC, Khoo CS. Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of Information Science. 36(6): 823-848, 2010.
19. Gangemi A, Presutti V, Reforgiato Recupero D. Frame-Based Detection of Opinion Holders and Topics: A Model and a Tool. Computational Intelligence Magazine, IEEE. 9(1): 20-30, 2014.
20. Cambria E, Hussain A, Havasi C, and Eckl C. SenticSpace: Visualizing opinions and sentiments in a multi-dimensional vector space. In: LNAI, vol. 6279, pp. 385–393, 2010.
21. Boiy E, Moens MF. A machine learning approach to sentiment analysis in multilingual Web texts. Information retrieval. 12(5): 526-558, 2009.
22. Wan X. Co-training for cross-lingual sentiment classification. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, pp. 235-243, 2009.
23. Brooke J, Tofiloski M, Taboada M. Cross-linguistic sentiment analysis: From english to spanish. In: Proceedings of the International Conference RANLP, pp. 50-54, 2009.