

Comparative Analysis of Two Chinese Major Online Review Platforms in Customer Concern Identification

Jing Wang^{1,a,*} and Xiangbin Yan^{1,2,b}

¹ Department of Management Science and Engineering, Harbin Institute of Technology, 92 West Dazhi Street, Nan Gang District, Harbin, China

² Department of Management Science and Engineering, University of Science & Technology Beijing, 30 Xueyuan Road, Haidian District, Beijing, China

^a jingwangxr@gmail.com, ^b xbyan@hit.edu.cn

*corresponding author

Keywords: customer concern, online platform, online reviews, LDA.

Abstract: Analysing online reviews is a good method to identify customers' needs and concerns. We conducted a comparative analysis of two Chinese major online review platforms (JD vs. ZOL) in customer concern identification using the LDA model. We find that there is a big difference between customer concerns of JD users and that of ZOL users. JD platform users pay more attention to sound quality, exterior, standby time, and performance, while ZOL platform users focus on function, configuration, brand, and tutorial. The former concentrates on specific attributes while the latter takes note of attributes in general and brand information. The results of this study are beneficial for retails to find target customer and for manufacturers to improve customer satisfaction.

1. Introduction

Analysing online reviews is a good method to identify customers' needs and concerns. Compared with the traditional methods, online reviews are published voluntarily by consumers. The data generated from the reviews have three features. First, publishing reviews online can be anonymous. Therefore customers have less social pressure to express their true feelings as compared to social investigations. As a result, the data have the validity in nature. Second, online reviews often provide more details while surveys are often based on single, isolated, or even symbolized description words, like star rating. Manufacturers can hardly obtain detailed information out of these rating-kind or single adjective based reviews. Last but not least, collecting data from websites presumably is easier with the right methods. It also costs less funds and time compared to using questionnaires.

So far, much research has been done to identify customers' needs and concerns using text mining on online reviews^[1; 2]. However, little attention has been paid to platform differences. Since users of different platforms have different motivation in posting online reviews, their concerns may appear different accordingly. For instance, the users of an online retail may express about product attribute while the users of a third-party platform may express which attributes they are looking for. Meanwhile, understanding customer concerns clearly is beneficial for retails to find target customer and for manufacturers to improve customer satisfaction. Therefore, it is necessary to include platform difference in customer concern analysis.

In this paper, we make a comparative analysis of two Chinese major online platforms in customer concern identification. In the following sections, we first review related literatures about the application and methods of customer concern analysis as well as Latent Dirichlet Allocation (LDA). Then we describe the proposed research framework of this paper. Next, empirical experiments are conducted and we present the results. Finally, we conclude our paper with future research direction.

2. Literature Review

2.1. Customer Concern Analysis Application

Some researchers are interested in solving business problems with customer concern identification methods. Motivated by augmenting traditional methods for concept generation, Lee applied adaptive text extraction methods to learn customer needs and product attributes^[3]. Lee and BradLow proposed a method to support the analysis and visualization of market structure by automatically uncovering product attributes and brand's relative positions from online customer reviews^[4]. Abrahams et al. defined a new class of vehicle 'smoke' words and proposed an approach to discover vehicle safety defects based on the words^[5]. Zhang et al. also described a system to find manufacturers' product weakness from Chinese reviews by using aspects based sentiment analysis^[6]. Marrese-Taylor et al. extended Bing Liu's aspect-based opinion mining technique to identify customer preferences about tourism products^[7].

2.2. Customer Concern Analysis Methods

Many studies mainly aimed to develop methods for analyzing online reviews to find customer concerns. Moghaddam and Ester introduced Interdependent Latent Dirichlet Allocation (ILDA) model^[1]. According to the characteristics of Chinese reviews, Pan and Wang proposed a method based on pattern matching to mine features and opinions^[8]. To generate an aspect hierarchy, Yu et al. developed a domain-assisted approach by integrating domain knowledge and consumer reviews^[2]. In order to discover latent aspects automatically, Ma and Qu adopted Local LDA^[9]. They also used Latent Rating Regression (LRR) model to infer the latent rating and weight for each of the discovered aspects. These methods are the deformation of LDA that is a classical and one of the most popular method for analysing customer concerns.

2.3. Latent Dirichlet Allocation (LDA)

Latent Dirichlet allocation (LDA) is a generative probabilistic model for topic modelling^[10] and it is an unsupervised learning model. After conducting the LDA model, documents will be represented a set of topics with probabilities and each topic has a collection of features associated with different probabilities. As a classical model, LDA has been applied in many domains, such as cellular phone products, tourism, mobile device applications, etc. Santosh et. al. used this model to extract feature and opinions mixtures from product reviews^[11]. Guo et. al. identified important dimensions of customer service in customer satisfaction analysis using LDA^[12]. Hao et. al. applied LDA to extract new features of different versions of mobile device applications^[13].

3. Research Framework

Fig. 1 presents the proposed research framework that consists of three phases: data preprocessing phase, customer concerns analysis phases, and comparative analysis phases.

3.1.1. Data Preprocessing Phase

Because there are no spaces between words in Chinese, we use NLPIR (Natural Language Processing and Information Retrieval Sharing Platform)^[14] to segment all the online reviews into words. NLPIR is one of the most popular tool for Chinese word segmentation^[15]. Then Chinese stop words were removed because they cannot provide useful information^[16].

3.1.2. Customer Concern Analysis Phase

Commonly the customer concerns are reflected by nouns^[17]. Therefore, we select nouns from online reviews to analyse customer concerns. We use nouns as features and apply Latent Dirichlet Allocation (LDA) for clustering online reviews into different topics^[18]. The customer concerns will be inferred from the representative features of the topics.

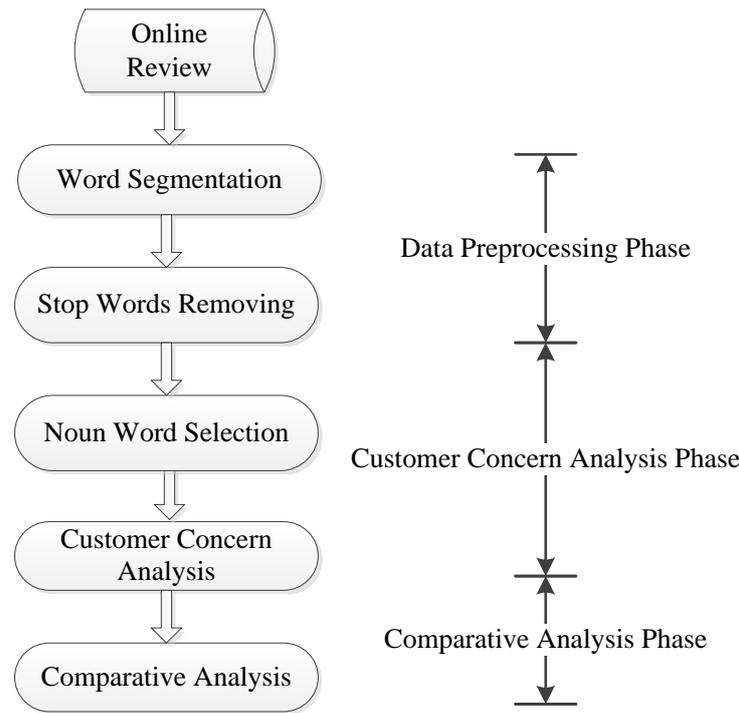


Figure 1 Research framework.

3.1.3. Comparative Analysis Phase

Jingdong (JD, <https://www.jd.com/>) is one of the most popular online retailers in China and Zhongguancun online (ZOL, <http://www.zol.com.cn/>) is a professional IT business website in China that provides reviews, news, ranking, etc. The difference between these two websites is that the reviews on JD are written by people who purchased the term there while there is no limitation for commenting on ZOL. In other words, the motivation for writing reviews between JD users and ZOL users is different. Accordingly, we infer that the concerns of their users are different. Hence, we conduct a comparative analysis in customer concerns between JD and ZOL.

4. Empirical Results

Table 1 Comparative analysis results of JD and ZOL in customer concern identification.

JD platform	ZOL platform
Sound quality: 声音(voice), 音质(sound quality), 铃声(ringtones)	Function: 功能(function), 模式(mode), 视频(video)
Exterior: 屏幕(screen), 大气(gorgeous), 款(style)	Configuration: 系列(series), 处理器(processor), 配置(configuration)
Standby time: 电池(battery), 时间(time), 充电器(charger)	Brand: 苹果(Apple), 品牌(brand), 销量(sales)
Performance: 卡(stuck), 速度(speed), 内存(memory)	Tutorial: 教程(tutorial), 故障(fault), 原因(reason)

We collect online reviews about cell phone from JD and ZOL. We got 10280 JD online reviews and 115486 ZOL online reviews. The time ranges are from 16th January 2016 to 3rd August 2016. Because there is a big difference between the number of JD online reviews and that of ZOL online reviews, we randomly select the same number (i.e. 10280) of ZOL online review as the empirical data. Then we segmented all the reviews into words and removed stop words. We got 1418 nouns for JD online review and 5805 nouns for ZOL online reviews, which were used as features in the LDA model. We conducted the LDA model with four clusters^[19]. The comparative analysis results

of JD and ZOL in customer concern identification are shown in Table 1. We summarized the topics based on their features and listed three representative features to explain.

The results indicate that the customer concerns of JD platform users and that of ZOL platform users are different. JD platform users pay more attention to sound quality, exterior, standby time, and performance, while ZOL platform users focus on function, configuration, brand, and tutorial. The former concentrates on specific attributes, such as how they exactly feel when they use phones. The latter takes note of attributes in general and brand information as well as manuals.

5. Conclusion

We conducted a comparative analysis of two Chinese major online review platforms (JD vs. ZOL) in customer concern identification using the LDA model. We find that there is a big difference between customer concerns of JD users and that of ZOL users. JD users pay more attention to specific some attributes while ZOL users focus on general attributes and brand information. The results of this study are beneficial for retails to find target customer and for manufacturers to improve customer satisfaction. In the future study, we will involve more platforms and products in our model to analyse their differences in customer concerns.

Acknowledgements

This work is partly supported by the National Natural Science Foundation of PRC (No. 71531013, 71490720, and 71401047).

References

- [1] Moghaddam S, Ester M. (2011) ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, 665-674.
- [2] Yu J, Zha Z-J, Wang M, et al. (2011) Domain-assisted product aspect hierarchy generation: towards hierarchical organization of unstructured consumer reviews. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 140-150.
- [3] Lee T Y. (2009) Adaptive text extraction for new product development. ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 769-778.
- [4] Lee T Y, Bradlow E T. (2011) Automated marketing research using online customer reviews. Journal of Marketing Research, 48(5): 881-894.
- [5] Abrahams A S, Jiao J, Wang G A, et al. (2012) Vehicle defect discovery from social media. Decision Support Systems, 54(1): 87-97.
- [6] Zhang W, Xu H, Wan W. (2012) Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. Expert Systems with Applications, 39(11): 10283-10291.
- [7] Marrese-Taylor E, Velásquez J D, Bravo-Marquez F, et al. (2013) Identifying customer preferences about tourism products using an aspect-based opinion mining approach. Procedia Computer Science, 22: 182-191.
- [8] Pan Y, Wang Y. (2011) Mining product features and opinions based on pattern matching. Computer Science and Network Technology (ICCSNT), 2011 International Conference on, 1901-1905.
- [9] Ma G, Qu Y. (2012) A local LDA based method for Latent Aspect Rating Analysis on reviews. Signal Processing (ICSP), 2012 IEEE 11th International Conference on, 2240-2245.

- [10] Blei D M, Ng A Y, Jordan M I. (2003) Latent dirichlet allocation. *Journal of machine Learning research*, 993-1022.
- [11] Santosh D T, Babu K S, Prasad S, et al. (2016) Opinion Mining of Online Product Reviews from Traditional LDA Topic Clusters using Feature Ontology Tree and Sentiwordnet. *International Journal of Education and Management Engineering (IJEME)*, 6(6): 34.
- [12] Guo Y, Barnes S J, Jia Q. (2017) Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59: 467-483.
- [13] Hao Y, Wang Z, Xu X. (2016) Empirical Study on the Interface and Feature Evolutions of Mobile Apps. *International Conference on Service-Oriented Computing*, 657-665.
- [14] Zhou L, Zhang D. (2003) NLPiR: A theoretical framework for applying natural language processing to information retrieval. *Journal of the Association for Information Science and Technology*, 54(2): 115-123.
- [15] Feng L, Liu B, Peng Z. (2016) A patent hotspot discovery method. *Wuhan University Journal of Natural Sciences*, 21(5): 369-375.
- [16] Liu C L, Hsiao W H, Lee C H, et al. (2012) Movie Rating and Review Summarization in Mobile Environment. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(3): 397-407.
- [17] Lark J, Saldarriaga S P, Morin E, et al. (2014) Consumer Concern Extraction in Social Web Reviews. *International Conference on Digital Intelligence 2014*.
- [18] Pham T N, Nguyen L M, Ha Q T. (2012) Named Entity Recognition for Vietnamese Documents Using Semi-supervised Learning Method of CRFs with Generalized Expectation Criteria. *2012 International Conference on Asian Language Processing*, 85-88.
- [19] Lee H, Kihm J, Choo J, et al. (2012) iVisClustering: An interactive visual document clustering via topic modeling. *Computer Graphics Forum*, 1155-1164.