

Marketing Improvement of Chinese Original Picture Books From Dissatisfaction Evaluation - Text Mining Based on LDA Model

Ziyan Tang and Wen Yuan*

Interdisciplinary Center for Culture, Science and Technology, Sichuan University, Chengdu, China

Abstract: Chinese original picture books play an important role in inheriting traditional culture and forming cultural identity, which is very important for children. We analyzes the dissatisfaction evaluation of Chinese original picture books by using the topic model of Latent Dirichlet Allocation (LDA). It is found that the dissatisfaction of consumers mainly focuses on smell, AR function, preaching, quality, picture, content , painting style and so on. In the future, we should take quality as the bottom line, focus on the content creation, and carry out integrated marketing through social media.

1 Introduction

Picture book is a kind of reading material with pictures and little or no words, which is of great significance to children. It is not only to tell stories and learn knowledge, but also to help children build a comprehensive spiritual world and cultivate multiple intelligences [1]. With the increasing attention on the public to early childhood education, the social recognition of the effectiveness of picture book education is also higher and higher, and the sales volume of picture books is increasing year by year. However, in the picture book market, the imported picture books still occupy a dominant position, while the sales of Chinese original picture books are dwarfed. There are about 4000 kinds of picture books published in China every year, among which more than 2000 kinds of domestic original picture books are published [2]. Although they account for half of the total, the market performance is not ideal. In terms of sales revenue, the imported picture books accounts for 80% of the total sales [3]. In terms of sales ranking, the top ten of Jingdong Mall (One of the largest online shopping websites in China) children's picture books are all imported .

This study conducted a preliminary test on mothers of children under 3 years old (n = 30), the results show that all of them would buy picture books, but chinese original picture books only account for less than 20% of they buy.

The imported picture books play an important role in children's aesthetic cultivation, character education, emotional development, etc. But they are created under their specific cultural background and educational philosophy, which is not conducive to the inheritance

*Corresponding author: yuanwen@scu.edu.cn

of Chinese culture [4]. The values conveyed by picture books will be imperceptibly learned or even internalized by children, which will have a profound impact on children's cognitive development and moral consciousness. Therefore, Chinese original picture books take the important role of inheriting traditional culture and forming cultural identity [5]. However, how to improve it and how to make it recognized by the market have been puzzling us and hindering the further development. Previous studies have found that the dissatisfaction expressed by consumers is often the place where the work needs to be improved [6,7]. Based on this, the current study aimed to analyze the dissatisfaction evaluation of Chinese original picture books. It is expected that the results of this study will provide some marketing suggestions for publishers and marketers. This paper also seeks to offer theoretical contributions to the existing literature on customer satisfaction.

2 Literature review

2.1 Dissatisfaction evaluation

Customer satisfaction is a kind of feeling, which comes from the comparison between the performance or output of a product or service and its expectation [8]. Satisfaction has always been the focus in the field of marketing [9-12]. With the rapid development of the Internet, consumers are increasingly inclined to express their emotions through online comments. The text information of these comments can help enterprises grasp consumers' attitude towards products and manage product quality [13]. Many researchers have carried out the analysis of book reviews. For example, Wang uses content analysis and text analysis method to analyze the content of the online comment of popular picture book Peppa Pig, the result shows that the affective factors and derivative services supporting factors should not be neglected [14].

Among many comments, dissatisfied evaluation is particularly important. First of all, most consumers tend to leave with no comments, only 4% of dissatisfied customers choose to express themselves. Although only one complaint is received, it actually means that more than 26 consumers have the same problem. Second, on average, customers who have been satisfied with the complaint will tell five people what they are treated, which will save a lot of marketing costs for the enterprise [15]. Therefore, the simplest and most effective way to understand what consumers really want and don't want is to listen to their dissatisfaction [6]. For example, Li Ming analyzed 924 dissatisfied evaluations of the book "A Dream in Red Mansions" from the online shopping platforms, he found that some dissatisfied evaluations have played a supervisory and restrictive role of the book [7].

2.2 The Latent Dirichlet Allocation Model

Topic models are receiving extensive attention in NLP (natural language processing) [16-20]. It has obvious advantages in topic recognition, semantic mining and so on, can better represent and organize document information [21]. In short, a topic is like a "box", which contains a number of words and is arranged according to the probability of occurrence. These words have a strong correlation with the topic, or it is these words that jointly construct the topic. Topic models not only depend on the frequency of literal word repetition, but also depend on the semantic association behind the text. It is a method to model the hidden topic of text which overcomes the shortcomings of traditional document similarity calculation methods.

Scholars have proposed a variety of topic models, such as TSM (Topic-Sentiment Mixture) [22], DIM (Document Influence Model) [23], BTM (Bitern Topic Model) [24] and

so on. LDA(Latent Dirichlet Allocation)model is one of the most representative topic models, which was proposed by Blei et al. in 2003 [25]. It is a kind of unsupervised machine learning technology. The general steps are as follows: (1) from each document D to the corresponding distribution θ , the topic Z corresponding to each word is extracted from the text; (2) A word W is extracted from the topic Z , and its corresponding multiple distribution is ϕ ; (3) Finally, repeat the above steps until you traverse every word in the document [26].

LDA topic model has been widely used in public opinion analysis, medical treatment, policy making, tourists' attitude, patent analysis, and historical research. Such as, Miller analyzed the national documents about violent crime and rebellion in Qing Dynasty, found out five riot themes, and compared their development trends [27]. Zhao analyzed the discourse changes of early Internet reports in people's daily [28]. An Lu, et al. combined the improved topic model with sentiment analysis, proposed a method to identify opinion leaders in online reviews [29]. Zhang studied the public's feedback on the bicycle sharing draft based on Sina Weibo users' comments [30]. With the continuous development of NLP technology, LDA and its improved model are more and more used in a broader field. LDA topic model has a significant effect on topic extraction and can effectively solve the problem of data sparsity in short text [31]. Thus, we choose it for this study.

3 Data sources and data analysis

3.1 Data sources

This study takes three steps to get consumers' true views on Chinese original picture books. First, we chose the top 100 Picture Books from "JD - Books - children's books - picture books" of JingDong Mall (One of the largest online shopping websites in China). Among the top 100, 12 are Chinese original picture books. Second, we downloaded all the dissatisfied evaluations of the 12 books. It is worthy of note that, all the comments are divided into three grades(Positive, Neutral and Negative Reviews),we define the " Neutral " and " Negative " reviews as dissatisfied evaluation in this study. Third, delete the evaluation with only score, no content and only related to logistics. Finally, 265 effective evaluations were obtained.

3.2 Data preparation

We used R 3.6.0 for text processing and analysis, and use JiebaR package which has good performance in Chinese word segmentation. In the process of data cleaning, considering the particularity of picture book, the segmentation dictionary is adjusted. Such as "picture book", "Make up the bill", "customer service" as fixed words, no longer divided. In the process of stop word lists construction, we combine the stop words list of Harbin Institute of technology, baidu stop words list and Sichuan University Machine Intelligence Laboratory stop words list.²

After word segmentation and elimination of stop words, we continue to build a corpus containing all the words. When using LDA topic model, it is very difficult to get the accurate probability distribution and parameters directly. Therefore, researchers often use approximate inference method. The commonly used parameter estimation methods can be roughly divided into two categories: (1) Variable algorithms, which is a deterministic

† <https://list.jd.com/list.html?cat=1713,3263,4761>

‡ https://gitee.com/chen_kailun/stopwords

method, researchers need to assume some parameter distributions, and compare these ideal distributions with posterior data to find out the closest and main algorithm is the variant expectation maximization algorithm (VEM); (2) Sampling based algorithm, i.e. using randomization method to complete the approximation. For example, Gibbs sampling mainly constructs a Markov chain and randomly extracts samples to estimate the posterior distribution. Gibbs sampling is widely used in the past literature because it is fast and easy to implement. So, we also choose Gibbs sampling in this study.

3.3 Model Fit

Some studys used the subjective judgment method to determine the number of topics K by test many times [32,33]. Considering the characteristics of unsupervised machine learning of LDA topic model, this study aims to minimize the subjective differences caused by different researchers, so maximum likelihood estimation is used to optimize the number of topics. The fitting results are shown in Figure 1. The results show that when the number of topics is 10, the likelihood estimator is the largest. Therefore, the number of topics parameter in LDA topic model is $k = 10$.

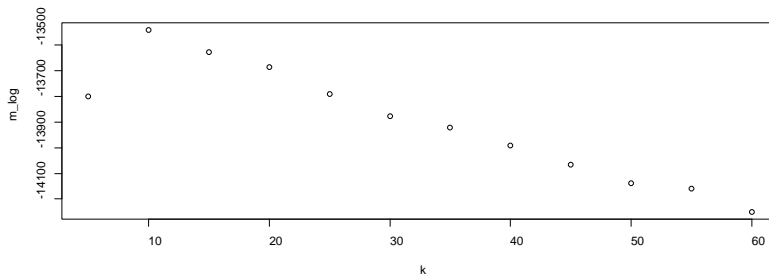


Fig. 1. Maximum likelihood estimation

The most contributing words of LDA topic model are determined by the text features it analyzes. This study collected small number comments and most of them are shorts texts. So, we tested the method of selecting 3-8 most contributing words, and finally found that it is more appropriate to select top 5 words per topic. The results of topic analysis is shown in Table 1 (translated the result into English by the author).

Table 1. Results of Topic analysis

Topic 1	Bad, smell, feel, a copy, disappointment
Topic 2	It's not worth it, AR, too, compare, bad
Topic 3	content, recommend, somewhat, Preaching, very
Topic 4	children, quality, A set of books, thin, printing
Topic 5	frame, especially, commonly, very, inside
Topic 6	see, story, sweep, feel, garbage
Topic 7	Few pages, thin, book, especially, quite
Topic 8	Not good, Don't buy it, at all, proposal, bad
Topic 9	JD,99 yuan, Cost effective, adult, book
Topic 10	A set of books, Painting style, somewhat, content, say

3.4 Topic interpretation

As we can see, the differences between these topics are not at the level of words, but rather in the distribution of groups of words. Let us give a general impression based on the proportion words within each topic one by one.

The most contributing words in topic 1 include "bad, smell, disappointment", which mainly reflect consumers' concern that the smell of picture books may hurt children. For example, some comments mentioned that "the smell is too smelly. I haven't read the book yet, I put it on the balcony to blow and remove the formaldehyde".

In topic 2, the words "unworthiness and AR" reflect the buyers' serious disappointment with the so-called AR (Augmented Reality) function added to picture books. For example, some consumers commented that "the gimmick of AR is just like this in fact", "but the AR that we are looking forward to is very disappointing, and it is not generally difficult to use." The words such as "content and preaching" in topic 3 express consumers' dissatisfaction with the strong preaching of domestic picture books. For example, some consumers think that "very bad stories, which are not interesting and have confused logic, are completely forced to preach".

Topic 4 is still about book quality, Topic 5 focuses on the screen. Consumers feel that the pictures are very general and nothing special. Topic 6 is still the evaluation of AR function. After scanning the code of book, consumers find that the operation is very inconvenient, which cannot meet the expected. In topic 7, most contributing words express that the reader thinks the book is too thin.

Topic 8 reflects consumers' recommendation attitude after purchase. Topic 9 mainly focuses on the evaluation of "Choose 10 books for 99 yuan on JD Mall" and Topic 10 reflects consumers' attitude towards picture book style.

In general, without considering the influence of JD's promotional activities, consumers' dissatisfaction mainly comes from seven aspects: smell, AR function, preaching, quality, picture, content richness and painting style.

4 Conclusions

From the text analysis of LDA topic model to consumer dissatisfaction evaluation, Chinese original picture book should take quality as the bottom line, focus on the content creation, and carry out integrated marketing through social media.

First, quality is of first importance. The paper and material of picture books are the carrier of content. Small gram paper and the usual right angle design make the paper sharp and easy to cut hands. Ordinary ink printing may lead to heavy smell and high formaldehyde content of picture books. Although it seems to be able to reduce manufacturing costs, it is actually at the expense of consumer recognition. Moreover, it can no longer meet the current consumers' pursuit of "the Good life" and "high quality". In the production of Chinese original picture books, enterprises need to strictly control the quality and keep the bottom line of safety and harmlessness.

Secondly, content is the king. Chinese traditional culture is broad and profound, which provides materials for original picture books. Many elements in excellent masterpieces, such as "Shan Hai Jing" or "Journey to the west", can be used for reference for picture books. However, some traditional stories may not be understood by young children. We should not simply copy the whole books, but innovate on the basis of respecting history according to readers' cognitive level. In addition, Chinese traditional stories lay particular stress on conception and preaching, asking whether it is meaningful and what kind of educational role it has when they are conceived and narrated, but ignoring the decoding link of the audience, which reduces the narrative energy of picture books [34]. In order to solve

this problem, picture book designers should adopt the children based narrative mode, focusing on story and interesting. In addition, the style of picture books, the sense of detail of pictures are also very important, this will be a long-term direction for improvement. As for the additional functions such as AR, if they can't achieve the expected effect, they can be abandoned to avoid putting the cart before the horse.

Thirdly, build high-quality products of public praise. As mentioned above, many publisher have launched a variety of Chinese original picture books. The number is large, but the word of mouth is not good. This consumes a lot of enterprise resources, which eventually leads to the high-quality products are not outstanding. So publisher need to concentrate resources to create classical picture books, occupy the minds of consumers with the image of market segmentation first.

Finally, seize the fission effect of social media. With the rapid development of mobile Internet, consumers' time and attention are the limited resources. After creating a blockbuster, publishers should make full use of various tools for word-of-mouth communication, attract a group of key people at the center of the social network first, so as to drive more users to follow. In fact, ordinary mothers are not very clear about which picture books they should buy. They buy books according to the recommended list recommended list lists, and they don't deliberately distinguish between imported picture books and Chinese original picture books. Some mothers even said they would like to buy Chinese original picture books for their children's traditional culture education, but they don't know what to buy. The big V(verified weibo and wechat users who have more followers) of mother-to-infant has gathered a large number of fans through sharing the knowledge of child care, and naturally become the KOL (key opinion leader). The Press can cooperate with them base on various forms such as book introduction, recommendation and group purchase, as to form fission effects.

5 Limitations and future research

This study mainly uses LDA topic model to analyze the dissatisfaction evaluation of Chinese original picture books and puts forward relevant suggestions. Limited by time and funds, We only collected data from one online shopping platform. Future research could expand the source of comments, get more dissatisfaction evaluation through multiple platforms. Additionally, For LDA topic model, the number of topics is an important factor affecting the quality of text mining , we simply apply the MLE to determine the topic numbers. In the future, it would be interesting to use DIC(Bayesian Deviance Information Criterion) and perplexity for determination. Finally, choosing top 5 words per topic seems a little confused to interpretation of the topics, Although many experiments have proved that this is the best explanation in this study. In addition to empirical interpretation of the topics, automatic coherence measures could have been used such as UMASS and UCI.

This research was supported by Fundamental Research Funds for the Central Universities (Grant number 2019whkjzy04); Innovation Spark Project of Sichuan University (Grant number 2018hhf-58) ;Ideological and Political Education Research Center Project (Grant number CSZ11007) ;and Fundamental Research Funds for the Central Universities (Grant number 2021whkjzy05).

References

1. Q.Fang. Picture books: a bridge for preschool children to read with multiple intelligences, *Journal of the Chinese Society of Education*, s1:32-33 (2010)

2. M.G.Pan. On the development path of Chinese original picture books for children, China Publishing Journal, 9:32-35 (2016)
3. A.X.Yang. *Chinese Original Picture Books:the Ideal and the Road*, Journal of Minzu University of China(Philosophy and Social Sciences Edition), 4: 156-160 (2012)
4. W.W.Yuan. How to integrate traditional cultural elements into children's Picture Book Creation -- from the perspective of children's psychology, China Publishing Journal 10: 32-34 (2016)
5. X.Y.Liu &Z.Wang. Cold thinking of picture book publishing fever from the perspective of cultural identity, Publishing Research,09:69-72 (2016)
6. Z.Q.Sun. *Dissatisfaction survey: an effective strategy tool*, Journal of Commercial Economics, 23:16-17 (2004)
7. M.L. Quality problems and Countermeasures of paper books from the poor evaluation of online stores, Editorial Friend, 03:34-37 (2013)
8. C.H.Bai &Y.Liao. *Research on customer satisfaction based on customer perceived value*, Nankai Journal(Philosophy,Literature and Social Science Edition) ,06: 14-20 (2001)
9. J.L.Han &F.X.Wei. Research on the interactive relationship between customer satisfaction and customer loyalty, Nankai Business Review, 06: 8-10+29 (2001)
10. C.H.Bai &Z.Liu. Research on customer loyalty and its determinants of service enterprises, Nankai Business Review ,06: 64-69 (2002)
11. A.M.Deng,B.Tao &Y.Y.Ma. *Empirical Study of Influential Elements of E-loyalty*, Chinese Journal of Management Science,22(06):94-102 (2014)
12. W.Z.Zhong ,L.L.Xi &R.J.Wu. Empirical Study on Influencing Factors of online shopping satisfaction based on ACSI model, Soft Science, 28(02):100-105 (2014)
13. D.H.Yang, et al. A Comparative Study on Positive, Neutral and Negative Reviews of Online Products Based on Jingdong Mall Data , Information Science, 37(02):125-132(2019)
14. W.Wang. The Influence Factors to Customer Satisfaction of Online Children's Picture Books :Taking "Dangdang" 's Best-selling Picture Book as an Example, Publishing Journal, 26(04):74-78 (2018)
15. Gorden, W. I. *Service America: Doing business in the new economy*, Employee Responsibilities and Rights Journal 1.1:89-89(1988)
16. G.X.&H.F.Wang . The development of topic models in natural language processing,Chinese journal of computers,34(8) :1423-1436(2011)
17. H.Xu,M.l.Zhang,D.C.Sun.The research topic mining and evolutionary trends of opinion mining and sentiment analysis in china based on LDA, Journal of Henan University of Technology(Social Science),37(2):38-46(2021)
18. X.Feng,K.Tang. *Research on Topic Mining Technology Based on Network Short Article*, Computer & Digital Engineering,49(5):952-956+992(2021)
19. J.J.Huang, et al.*Review of Deep Learning-Based topic model*,Chinese journal of computers,43(5) :827-855(2021)
20. L.P.Li,X.B.Zhao. A *Research Summary of Topic Discovery Methods Based on Topic Model*, Journal of MUC(Natural Sciences Edition),30(2):59-66(2021)
21. X.Yang, et al. Automatic multi-document summarization based on the latent Dirichlet topic allocation model, CAAI Transactions on Intelligent Systems 5.02:169-176(2010)

22. Mei, Q. , et al. *Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs*, Proceedings of the 16th International Conference on World Wide Web(2007)
23. Gerrish S, Blei D M. A language-based approach to measuring scholarly impact, Proceedings of the ICML(2010)
24. Yan X H, et al. *A bitern topic model for short texts*, International World Wide Web Conference (2013)
25. Blei, et al. *Latent Dirichlet Allocation*, Journal of Machine Learning Research ,3: 993 – 1022 (2003)
26. X.Z. Yang. Research on the Shui Literature Topic Mining Based on LDA Model and Text Clustering. *Modern Computer* ,05:13– 17 (2019)
27. L. Song & P.J. Zhang. *System design of micro-blog public opinion based on LDA topic modeling method*, Network Security Technology & Application ,04:5– 6 (2014)
28. Miller, I. M. . Rebellion, crime and violence in Qing China, 1722–1911: A topic modeling approach, *Poetics*,41.6:626–649(2013)
29. L.L. Zhao & D.L. The discourse changes of early Internet reports in people's daily (1978-2000): an analysis based on theme model and annotation package, *Publishing Research*, 01:57-63 (2019)
30. L. An, J. Hu & G. Li. A Method of Identifying Comment Opinion Leaders Based on Topic Consistency and Emotional Support, *Journal of Management Science*, 32(01):3-13 (2019)
31. Y. Yang, et al. The Public Sectors Adoption of Civil Response on Sharing Cycle Policy Based on LDA—the Comparative Analysis of Policy Revision and Data Mining, *Information Science*, 37(01):86-93 (2019)
32. Y. Zhang, Y. Yang & W. Deng . *China's Cyber Space Governance from the Perspective of Policy and Departments—Big Data Analysis based on LDA and SNA*, Journal of Beijing Institute of Technology(Social Sciences Edition), 21(02):127-136 (2019)
33. S. Dong & Q. Wang . LDA-based Tourist Perception Dimension Recognition: Research Framework and Empirical Research—Taking the National Mine Park as an Example, *Journal of Beijing Union University(Humanities and Social Sciences)* ,17(02):42-49(2019)
34. H.F. Cao. Audience needs and cultural identity -- on the development bottleneck and coping strategies of domestic picture books, *Chinese Editors Journal* ,11:21-25(2017)