

Data Mining of New Snack E-commerce Reviews Based on Text Sentiment Analysis and Latent Dirichlet Allocation Topic Model

Qian Yang^{1,*}

¹College of Tourism, Sichuan Agricultural University, Chengdu, China

*Corresponding author. Email: 13880735835@163.com

ABSTRACT

In the new retail era, in order to promote the development of new snack e-commerce enterprises, increase customer satisfaction with their products and services and the desire to repurchase, this article makes sentiment analysis on the 430,000 online reviews of new snack e-commerce as crawled online from the perspective of customers. Firstly, it makes an overall sentiment judgment on the online review texts of five stores. Secondly, it constructs the Latent Dirichlet Allocation (LDA) topic model, gets ten categories of keywords, analyzes the cluster analysis results and discusses for improving any consumer dissatisfaction point. Empirical analysis shows that in order to improve customer satisfaction and stickiness and attract more potential consumers, new snack e-commerce companies need to pay attention to optimizing brand reputation, product quality, service level, marketing strategy, and review information.

Keywords: e-commerce online reviews, text sentiment analysis, LDA topic model, Python, new snacks

I. INTRODUCTION

With the advent of the new retail era and gradual updating of consumption, consumers have more and more understanding of various commodities, and more and more mature concept of consumption. The world of snacks is also quietly changing, people's requirement for snacks becomes higher and higher, and health becomes an important standard for measuring commodities. "New snacks" are healthy snacks that are processed to the minimum extent without additives or with necessary dosage of additives, and are flavor, pigment, preservatives and the like additives free provided that the color and flavor of the food itself are unchanged. Major snack brands such as Qiaqia, Three Squirrels and BESTORE take nuts and the like new snacks as the new breakthrough point and e-commerce as the new sales channel. Casual snack industry ushers in another wave of growth E-commerce online review refers to consumer reviews on the quality and prices of a commodity and the service of the store after the e-commerce transaction is completed. Text review often carries a strong emotional tendency. It not only reflects the emotional expression of customers after shopping, but also provides a reference for potential customers to make shopping decision. It has become an important information source for new snack e-commerce enterprises to obtain customers' consumption preferences and carry out precise marketing.

II. OVERVIEW OF RELEVANT RESEARCHES

A. Researches on e-commerce online reviews

With the vigorous promotion and development of e-commerce in the Internet era, the status of online review is becoming more and more important. According to survey, nearly 90% of consumers check customer reviews before making shopping decision. Therefore, in recent years, online reviews have received extensive attention from academic circle. In foreign countries, there have been many years of studies on the impact of online reviews on marketing management activities of e-commerce companies. As early as 2004, Villas proposed that in order to obtain more customers and sales, e-commerce companies can use online reviews to enhance product promotion and form a good brand reputation [1]. Deloitte (2007), Hitt (2011), and et al. further emphasized the importance of online reviews for potential consumers to understand the actual quality of goods and make purchase decision [2] [3].

Chinese scholars have conducted researches on the impact of online reviews on consumers' purchase intention and satisfaction, product sales, and brand reputation, and generally recognize that online reviews have an important impact on e-commerce sales. Zheng Xiaoping (2008) explored the impact of online reviews on online customers' purchase intentions and purchase

decisions, and found that the quality and number of online reviews are positively proportional to online consumers' purchase intentions [4]. Later, Zhai Likong (2011), Wa Yu (2014), Du Xuemei (2016), and et al. also made researches in this field. They respectively explored the impact of online reviews on potential consumers' purchase intention based on different perspectives or theories. As a result, they concluded that "the number and strength of online reviews, and the seniority of reviewers positively affect potential consumers' willingness to buy" [5] [6] [7]. At the same time, Zhai Likong introduced the emotional tendency of reviews into research, and proposed that "negative online reviews have a stronger impact on consumers' purchase intention than positive online reviews [5]". Zhang Xiaojuan (2016) even made an empirical analysis on the impact of the interaction between the emotional polarity and emotional intensity of online reviews on consumers' purchase intention and satisfaction [8]. Of course, scholars such as Wang Junjun (2013), Chen Zaifei (2014), and Li Yanfei (2016) also demonstrated the significant impact of online reviews on product sales by comparing different types of products and reviews from different sources [9] [10] [11]. In recent years, with the popularization of big data technology, the research of online reviews has been exploding, and scholars using big data to conduct research has become a development trend. For example, Wang Jianfei (2019) made up the deficiency in research method by researching the dynamic interaction between brands, reviews and sales based on the PVAR model [12]. Zhang Chaohui et al. (2020) used Bayesian and neural network mixed algorithms to study e-commerce credit evaluation methods to meet the actual needs of e-commerce to ensure fairness and authenticity of network transactions [13].

B. Researches on text sentiment analysis and LDA topic model

In recent years, due to the rapid development of the Internet and online shopping, the number of online reviews has been increasing day by day, but some reviews have complicated content or ambiguous attitudes. It is hard to feed back the actual purchaser information and demand to e-commerce enterprises and further affects the business activities of such enterprises. Therefore, text sentiment analysis came into being, and attracted many Chinese and foreign scholars to carry out a lot of research. Before that, Hatzivassiloglou (1997) and et al defined text sentiment analysis as "the process of analyzing, processing, summarizing, and reasoning subjective texts with sentimentalism flavor [14]", and the earliest sentiment analysis stemmed from people's emotional judgment on sentimental words; for example, "satisfaction" is a word with positive emotion, and "disgust" is a word with negative emotion. With the gradually increasing in

length of subjective text content, scholars have gradually transitioned from simple research and analysis of emotional words to that of more complex emotional phrases, sentences, chapters and even multiple chapters [15], while broadening the research content of text sentiment analysis. For example, Cui Anqi (2013) conducted sentiment analysis on the hot events in Weibo and correctly evaluates the opinions of netizens [16]; Li Hanyu (2017) and et al. automatically extracted the product attributes that customers pay attention to and the customers' evaluation opinions on the corresponding attributes, and further made text sentiment tendency analysis on the review of commodity [17].

Blei (2003) and others proposed a complete Latent Dirichlet Allocation topic model (LDA for short) [18], which was later widely used in text sentiment research, involving in multiple service areas such as tourism and e-commerce. Zheng Xianwei (2014) and others created the LDA topic model of scenic area reviews by extracting keywords or sentences about products or services in the reviews, and achieved the goal of individualizing the setting and querying of Internet tour reviews [19]. Yao Yao (2019) conducted a cluster analysis of the positive and negative reviews of cold-chain agricultural product e-commerce through the LDA topic model, which provided reference significance for improving consumer satisfaction and purchase desire, and promoting the development of e-commerce enterprises [20].

C. Researches on e-commerce online review based on text sentiment analysis

Text sentiment analysis refers to a computational research method that uses natural language processing (NLP), data mining algorithms, etc. to perform sentiment judgment analysis on text content, and then to grasp the attitude and viewpoint of the text [21]. As online reviews play an increasingly important role in e-commerce marketing operations, many scholars have begun to use text sentiment analysis research methods when studying e-commerce online reviews, and have achieved certain results. Dasgupta (2016) and others conducted text mining case studies on online reviews of Samsung mobile phones in order to understand that customers use different words to express the characteristics of mobile phones [22]. Lei Yuntao (2018) uses efficient sentiment analysis technology to make the computer automatically classify sentiment of e-commerce product reviews, provide consumers and e-commerce companies with more accurate sentiment review information, help customers choose products correctly, and provide e-commerce enterprises with suggestions for improving the quality of goods and services [23]. Liu Yulin (2018) conducted empirical research on the online reviews of Tmall Pecans, using Python to deeply explore reviews to make sentiment

judgment, and create an sentiment index to more accurately understand the sentiment change in online reviews of e-commerce, and provide a theoretical basis for e-commerce companies to establish new review methods and grasp the emotional tendency of the entire e-commerce industry [21]. Yao Yao (2019), from the perspective of potential topics, studied customers' concerns about Huainan's "vegetable basket project", a cold-chain agricultural product e-commerce, and extracted the most practical reviews for enterprises and customers to make sentiment analysis and fully explore customers' actual demand [20].

In summary, a creative research progress is achieved after applying text sentiment analysis method to the research on online reviews of e-commerce, which to a certain extent improves consumer satisfaction and repurchase rate, and provides reference and theoretical support for e-commerce enterprises to carry out business activities and increase sales. However, there are still some problems in this research field: first, there is little empirical research combined with the LDA topic model in text sentiment analysis of e-commerce online reviews; second, the research on e-commerce online reviews based on text sentiment analysis still stops at the theoretical level and there are few e-commerce companies that empirically apply research results and form evaluation reports.

Therefore, this article collects 450,000 online reviews of three squirrels, Wolong, Qiaqia and other new snack e-commerce companies, uses Python to perform natural language processing on the reviews, judges emotional tendencies based on text sentiment analysis, and constructs LDA topic models to assess the customer's recognition on enterprise, explore customer preferences, explore the operation and management status of various e-commerce enterprises, and provide more targeted suggestions to help e-commerce enterprises improve product quality, improve service levels, and achieve more precise and individualized marketing.

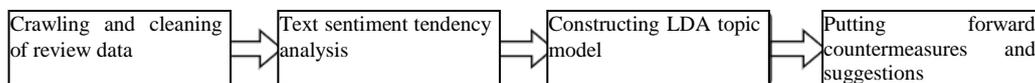


Fig. 1. Flow chart of text sentiment analysis on new snack e-commerce online reviews.

IV. EMPIRICAL ANALYSIS

A. Data source

The author selects five top new snack stores on Tmall including BESTORE, BE&CHEERY, Three Squirrels, Qiaqia, and Wolong, takes the nut products of each store as the research objects, then uses Python to design crawler programs to acquire customers' online reviews after purchasing. The online review texts from January 2018 to January 2020 are selected and made analysis. High-selling products of the store have high

III. RESEARCH METHODS

A. Crawling and cleaning data

Python is used for crawling the selected new snack e-commerce online reviews. Since data from five e-commerce stores at different times are involved, it is needed to preprocess (such as merge and clean) the data. Wherein, the data cleaning includes text deduplication, processing missing data, and extracting target data text.

B. Judging emotional tendency

This method is to introduce the SnowNLP model into Python, define the emotion function, introduce the processed e-commerce online review text data into the text sentiment analysis model for different shops, and run the programming for multiple times to ensure the accuracy of sentiment tendency judgment.

C. Constructing LDA topic model

This method is to load in Jieba library, then introduce the LDA model, define the get_lda function, and introduce the processed online review text data of the e-commerce into LDA topic model taking shops as the dimension, and ten respectively get ten categories of topic keywords of five stores.

D. Putting forward countermeasures and suggestions for e-commerce enterprises

Through emotional tendency judgment and LDA key words, this method deeply explores the emotional expression in online reviews of e-commerce, explores the marketing and operation problems existing in e-commerce enterprises, and gives countermeasures and suggestions.

The specific research process is shown in "Fig. 1". In particular, all the above steps are implemented through Python programming.

amount of online reviews; a total of 435,350 online reviews were obtained through the crawler. In addition, data such as the store name, product ID, review time, URL, and review text of new snack e-commerce online reviews were collected during the crawling.

Word Art is applied to make a word cloud map of online reviews (as shown in "Fig. 2"). By extracting high-frequency vocabularies from the pre-processed reviews and making word cloud map, it is available to more intuitively reflect the focus of the reviews, which is conducive to summarizing customers preferences.

C. Cluster analysis of LDA topic model for e-commerce online reviews

According to the classification results, the LDA topic model was applied to make topic modeling on reviews of the five new snack e-commerce stores, and clustered them according to the maximum necessary

probability of the review words or phrases on all topics [20]. Due to length limitation, this article only lists some of the topics and important words clustered under the topics of the store reviews of Wolong and BESTORE with the highest positive and negative emotions, as shown in "Table III" and "Table IV".

TABLE III. EXAMPLE OF DISTRIBUTION OF LDA TOPICS FOR ONLINE REVIEWS OF WOLONG STORE

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Good review	Rush to purchase	Nuts	Fresh	Price	Express delivery	Taste	Children	Purchase	Activity
Logistics	Buy one and get one free	Every day	Date	Good	Service	Commodity cost	Eat	For the New Year	Delicious
Satisfied	Nutritious	Brand	Package	Special purchases for the Spring Festival	Happy	performance	Quick	Give a gifts	NO
Favorable price	Good	Activity	Quality	Taste	Substantial	high	love	Less	This time
Fast logistics	Like it	One bag	Delivery	Cost-efficient	Bag	Review	Cheap	Two boxes	Discounts

TABLE IV. EXAMPLE OF DISTRIBUTION OF LDA TOPICS FOR ONLINE REVIEWS OF BESTORE

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Good	Activity	Taste	Thing	Flavor	Logistics	Review	Less	Purchase	Fresh
Good product	Cost-efficient	Super	Customer service	Package	Very fast	Fill	Feeling	Dry	Date
Store	Double tents	love	Service attitude	cost performance	Delivery	User	NO	Special	Eat
Repurchase	Snacks	Children	As always	Commodity	Express delivery	Delicious	Thing	Sweet	This time
Purchase	Cheap	Like it	Good	Satisfied	Speed	Biscuits	Amount	It's ok	Quality

Based on the keywords extracted from the LDA topic model, the following conclusions can be drawn: the keywords of the two stores are mostly positive, and the most frequent ones are delicious, good, buy, cost-effective, special purchases for the Spring Festival, etc., which indicates that new snack companies' factors such as product taste, flavor, quality, price, service, logistics, and promotion all affect consumers' purchase intention and satisfaction. By making comparison on the keywords of Wolong and BESTORE, it is shown that the quality and taste of the new snacks in the two stores are both good, the packaging is generous and beautiful, the store has a good service attitude, and the logistics and transportation are fast. They are customized products targeted to children. Hence, the customer satisfaction and willingness to repurchase are high, but there are both problems of small amount and feeling not enough to eat. In addition, Wolongsheng wins in its lower price, affordable nature and high cost performance of new snacks, and is good at taking advantage of various holidays and festivals such as New Year's shopping festival, Double Eleven, and Double Twelve to organize promotional activities. And

BESTORE wins in its good brand image and word of mouth. The term "BEST" has become a synonym for its product quality and has penetrated into the hearts of consumers, attracting many potential consumers to buy.

V. CONCLUSION

The online reviews of new snack e-commerce contain a lot of emotional information. This article analyzes the sentiment tendency of online reviews and extracts LDA topic words to achieve the purpose of understanding customers' overall emotional tendency towards new snack e-commerce and their satisfaction and dissatisfaction with their products or services. Result of empirical research taking 430,000 online reviews of Three Squirrels, Wolong, Qiaqia and other new snack e-commerce companies as an example shows that the sentiment tendency judgments made in text sentiment analysis method and LDA cluster analysis method can be well applied to new snack e-commerce online review. Through the interpretation of the data analysis results, the author has extracted the problems of brand reputation, product quality, service

level, marketing strategy, review information and other aspects existing in new snack e-commerce companies, and based on the problems, put forward the following four targeted countermeasures and suggestions in order to provide a reference for the business management activities of the new snack e-commerce industry, improve customer satisfaction and stickiness, and thereby increase the sales and brand reputation of the new snack e-commerce enterprises.

A. Shaping a high-quality brand image and word of mouth, and improving consumer satisfaction and willingness to repurchase

In order to further expand the consumer market and improve consumer satisfaction and repurchase willingness, new snack companies must pay attention to the shaping and management of new snack brand, especially to the differentiation shown between the brand culture and value output. Brands need to be synonymous with a way of life, attitudes, and concepts of life, not just varieties. [24] According to the data analysis results, consumers are very fond of branded products, and are more willing to buy old brand of product that already have a good quality image and word of mouth than new brands. Therefore, they especially prefer Three Squirrels and BESTORE, two old-fashioned snack companies. This indicates that their good brand images have penetrated into the minds of consumers. Wolong and Qiaqia still need to spend a lot of time and money in Internet brand marketing, and strive to improve the brand reputation and image, improve customer satisfaction and willingness to repurchase, and strive to attract old customers to bring new customers.

B. Guaranteeing the quantity and quality of products and improving the quality and level of service

New snacks must be of high quality and quantity in order to win customers' love. They not only must be fresh, taste good, and nutritious, but also should meet the quantitative needs of most consumers. This is the so-called high cost performance. It can be seen from the extracted keywords that customers attach great importance to the taste, flavor, quality, freshness of snacks and the efficiency, quality, and level of service. Many consumers choose to repurchase because of the friendly customer service and consistent service attitude. At the same time, in the field of Internet e-commerce, delivery time and logistics speed are the key concerns and considerations of consumers and directly affect consumers' purchase decisions. Therefore, new snack e-commerce companies should strengthen the strict supervision on the production line, strictly check and screen new snacks in accordance with the standards of freshness, taste, flavor, quality, etc., to ensure product safety and quality, fully mobilize the production initiative, enthusiasm and creativity of the

employees of the enterprise, increase the importance they attach to food quality in their production activities, and take measures to improve the service quality of the entire enterprise. E-commerce services include consulting services, logistics services, after-sales service, and so on. Enterprises should cultivate the service awareness of all employees, implement service ideas, and improve service levels through training, evaluation and other means [25].

C. Adjusting the price strategy and seizing the opportunity to carry out promotional activities

The target market of new snacks can be subdivided according to the potential consumers' gender, age, income, occupation, dietary preferences and other variables, and each price level of snacks can be developed in a targeted way based on consumers various lifestyles, dietary habits and nutritional needs. At the same time, it is available to formulate different price strategies for products of different grades, and adjust the prices of products appropriately according to factors such as seasonal changes, consumer psychology, and feedback information. Judging from the statistical results of sentiment tendency classification and the cluster analysis results of LDA topic model, the two new brands, Wolong and Qiaqia, fully grasped the competitive advantage of the low price strategy and attracted many potential consumers to buy, which shows that customers are very concerned about the price and promotion activities of new snacks, especially New Year's shopping festival, Double Eleven, Double Twelve and other large promotional events. Therefore, new snack e-commerce companies should seize the occasions of various holidays and festivals including double eleven to carry out promotional activities, stimulate consumers' desire to buy, attract more customer groups, and enhance consumer satisfaction and repurchase rate through the sale of big gift packages, and holding "buy one and get one free" or Internet promotional activities.

D. Paying attention to online review information and using the Internet for precise marketing

The arrival of 5G era has facilitated the marketing of new snack companies. New snack companies can crawl customer review information to understand customers' emotional attitudes towards products and services and existing problems, timely carry out after-sales service, improve its product and service in response to the feedback, and at the same time create a good review atmosphere and environment to attract more customers. Individualized marketing strategies that are widely used nowadays include individualized keyword marketing, precise marketing recommended by association rules, and precise marketing based on content recommendation [26]. First of all, new snack e-commerce companies can learn more about the market

characteristics, predict consumer behavior and further more precisely recommend and guide consumption by mining a large amount of data and information about consumers, such as consumers' basic personal characteristics, shopping behavior characteristics and psychological characteristics. Secondly, new snack e-commerce companies can keep track of their overall emotional tendencies, degree and changes through dynamic monitoring and analysis of customers' online reviews, analyze the key factors they pay attention to when making purchase decisions, grasp customers satisfaction with their products and services, and based on which improve the improper part of business management and formulate more precise and personalized marketing strategies and programs.

Above all, this research can provide a reference and theoretical support for vast consumers and new snack e-commerce companies. However, this research method is not limited to the new snack e-commerce industry, but is also applicable to other commercial fields with a large amount of review text data.

References

- [1] VILLAS-BOAS, J. Miguel. Consumer learning, brand loyalty, and competition. *Marketing Science*, 2004, 23.1: 134-145.
- [2] DELOITTE, Touche. Most consumers read and rely on online reviews; companies must adjust. *Tech. rep*, 2008.
- [3] LI, Xinxin; HITT, Lorin M.; ZHANG, Z. John. Product reviews and competition in markets for repeat purchase products. *Journal of Management Information Systems*, 2011, 27.4: 9-42.
- [4] Zheng Xiaoping. An empirical study of the impact of online reviews on online consumers' purchase decision [D]. *Renmin University of China*, 2008. (in Chinese)
- [5] Zhai Likong. A Study on the Impact of Online Shop Reviews on Consumers' Purchase Intention [D]. *Dongbei University of Finance & Economics*, 2011. (in Chinese)
- [6] Wa Yu. Study of the impact of online reviews on consumers' purchase intention [D]. *Zhejiang University*, 2014. (in Chinese)
- [7] Du Xuemei, Ding Jingyu, Xie Zhihong, Lei Lifang. An Empirical Study on the Impact of Online Reviews on Consumers' Purchasing Intention [J]. *Management Review*, 2016, 28 (03): 173-183. (in Chinese)
- [8] Zhang Xiaojuan. A Study on the Impact of Online Reviews' Emotional Tendency on Customers' Purchase Intention [D]. *Shanghai University of Engineering Science*, 2016. (in Chinese)
- [9] Wang Junjun, Yan Qiang. An empirical study on the impact of online reviews of search-style products of different popularity on sales [J]. *Chinese Journal of Management Science*, 2013 (S2). (in Chinese)
- [10] Chen Zaifei. Research on the impact of online reviews from different sources on the sales of online retailers [D]. *Nanjing University*, 2014. (in Chinese)
- [11] Li Yanfei. Research on the impact of online reviews on consumer satisfaction and product sales [D]. *Guangdong University of Technology*, 2016. (in Chinese)
- [12] Wang Jianfei. Research on the dynamic relationship between reviews, sales and brand based on PVAR model [D]. *Anhui University*, 2019. (in Chinese)
- [13] Zhang Chaohui, Liu Jiajia, Ran Hui. Study of e-commerce credit evaluation method based on Bayesian and neural network mixed algorithm [J]. *Information Science*, 2020,38 (02): 81-87. (in Chinese)
- [14] Hatzivassiloglou V, McKeown KR. Predicting the semantic orientation of adjectives. In: *Proc. Of the EAACL'97*. Morristown: ACL, 1997.
- [15] Huang XJ, Zhao J. Sentiment analysis for Chinese text. *Communications of CCF*, 2008, 4 (2).
- [16] Cui Anqi. Research on public sentiment analysis of Weibo hot events [D]. *Tsinghua University*, 2013. (in Chinese)
- [17] Li Hanyu, Qian Li, Zhou Pengfei. Sentiment Analysis and Mining of Product Reviews [J]. *Information Science*, 2017,35 (01): 51-55 + 61. (in Chinese)
- [18] Blei D, Ng A, Jordan M. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 2003, 3: 993-1022.
- [19] Zheng Xianwei, Zhang He. Application of LDA Subject Extraction Model in Internet Travel Reviews [J]. *Industrial Control Computer*, 2014,27 (09): 92-94. (in Chinese)
- [20] Yao Yao. Research on sentiment analysis of online reviews of cold chain agricultural products e-commerce platform based on text mining [D]. *Anhui University of Science & Technology*, 2019. (in Chinese)
- [21] Liu Yulin, Jian Lirong. Data Mining of E-commerce Online Reviews Based on Text Sentiment Analysis [J]. *Statistics & Information Forum*, 2018,33 (12): 119-124. (in Chinese)
- [22] Dasgupta Subhasis, Sengupta Kalyan. Analyzing Consumer Reviews with Text Mining Approach: A Case: Study on Samsung Galaxy S3 [J]. *Paradigm*, 2016, 20 (1).
- [23] Lei Yuntao. Research on sentiment analysis based on e-commerce product review text [D]. *South China University of Technology*, 2018. (in Chinese)
- [24] Cheng Xin. Research on Marketing Strategy of Casual Snacks [J]. *Marketing World*, 2019 (47): 34-35. (in Chinese)
- [25] Cui Yu. Analysis of online marketing strategies for casual snack companies in the new retail era [J]. *Modern Marketing (Information Edition)*, 2019 (10): 247-248. (in Chinese)
- [26] Zeng Fue, Zou Zhou, Tao Ran. Does Personalization Marketing Trigger Privacy Concern All the Time: Based on the Moderation of Anthro-pomorphic Communication [J]. *Nankai Business Review*, 2018, 21 (05). (in Chinese)