Consumer evaluation using machine learning for the predictive analysis of consumer purchase indicators

BaoFu Tang
School of Mathematics and Information Engineering, Longyan University
Longyan, Fujian 364012, China
6247787@qq.com

Dong-Meau Chang*
School of Computer Science and Intelligence Education, Lingnan Normal University
Zhanjiang, Guangdong, 524048, China
morganch@me.com

Junjie Yang
School of Computer Science and Intelligence Education, Lingnan Normal University
Zhanjiang, Guangdong, 524048, China
yangjunjie1998@lingnan.edu.cn

Abstract

With the rapid development of the current network platform for online e-commerce, in addition to transparent price competition, buyer feedback also has a reasonable influence on consumers' purchasing decisions. Today, we can see that the feedback behavior of consumers on related websites, including well-known online shopping platforms such as Amazon Shopping, Shopee Shopping and Taobao, has been gradually strengthened in recent years. Whether substantive recommendations from consumer feedback help other superficial consumers read them to improve their shopping habits. In this study, we automatically classify feedback comments using machine learning, and monitor the growth trend of shopping transaction volume, selecting the Shopee shopping platform as an experimental case. The suggestions provided by customers based on reviews are incorporated into the sentiment word management analysis, and words and word scores are weighted. Finally, a shopping engine is built that simulates consumer behavior, filters variable factors using review management, and optimizes metrics for predicting consumer shopping.

1. Introduction

With the maturity of today's Internet technology, it is very convenient to use the Internet to request information about products, logistics management and distribution. The shopping habits of consumers have been changing as a result of the epidemic, and both the online and offline channels have been accelerated. In the past, in the process of industry development promotion, online shopping and online sales were virtual channels for physical stores for the sale of goods [1] [2]. With the development of the epidemic, various brand operators have also invested in online sales models, the consumption behavior of traditional consumers has been gradually changing, and physical retailing is no longer the current consumption model. Consumers are revealing, collecting and analyzing large amounts of data online, changing the way they make decisions and consume. Among today's popular e-commerce platforms, Shopee, Momo, PChome, etc. each have their own unique characteristics. In addition to price exposure and bidding strategies, online shopping platforms also improve customer relationships by transforming e-commerce into value-added and differentiated services. In this study, the Shopee consumer rating was used for analysis, as well as the e-commerce online rating system. Combined with public review information, customer feedback can be used to test the effectiveness of merchants' value-added services and analyses consumer behavior. Fig. 1 Positive feedback from customers after a purchase.

Consumers can rate products, including text and image ratings, within one month of ordering in the Shopee Shopping Centre Rating System Specification. [3] [4] The sys-

Figure 1. Positive customer feedback after making purchases.

Figure 2. Negative emotional assessment and feedback from the consumer after they have made their purchase.

tem will actively disclose to the public platform if there is no return registration in various stores. This reflects consumers' real opinions, and stores can also effectively interact and communicate with customers through measurement. This practice has changed the consumers' doubts about the authenticity and reliability of the reviews on the online platform, the negative emotional evaluation and the post-purchase feedback of the consumers are shown in Fig. 2.

The purpose of this paper is to examine the impact of e-commerce on customer satisfaction. Historically, the status of customer feedback and comments has often been difficult to define, the majority of customer reviews are based on negative emotional states and the reviews are closed, making the results of the analysis less meaningful. This study used the Shopee shopping platform for analysis, mainly because the platform encourages consumers to provide feedback after purchase and offers rewards for feedback, so both positive and negative feedback can be fully expressed. Recent studies have also looked at how to integrate online and offline consumption [5]. At the end of 2021, Google released the "2021 Taiwan cross-border business key report. Based on statistical analysis by Ipsos. Businesses have a significant impact on online marketing. The technical dilemma at this stage is that delivering personalized information to experience and communicate is difficult. Whether it's a stable, mature industry or a traditional one, the main dilemma remains. Therefore, our study focuses on the hotspots of marketing, observing how to improve the effective feedback of marketing to customers, improving consumer orientation to inform consumption and perform key data analysis.

2. Related research

The key to influencing consumer purchase needs is to understand post-purchase consumer feedback information. This can truly reveal opinions, needs and indicators of product satisfaction [1].

2.1. Perceived Value:

Zeithaml (1988) [6] has already suggested that perception should be the independent choice of each individual and that it should be a process of redefining, adjusting and shifting product specifications and meanings. Perceived value [7] refers to the consumer's evaluation of a product's usefulness based on the product's income and payment integrity. [8] [9] [10] Therefore, if the results of the analysis of perception comments can be adjusted by valuation, the perceived value of a product or service can be thought of as the consumer's assessment of the usefulness of the product or service. In the effort/reward trade-off, perceived value has been seen as a mediator of the relationship between price and information, unrelated to purchase intention.

2.2. Convolutional neuronal network CNN model: [11]

Convolutional neural networks use a variety of modelling blocks. These include convolutional layers, clustering layers and fully connected layers. They learn the spatial hierarchy of the features automatically and in an adaptive way by means of retrocopulation via the connection of the data. [12] Input layer: divided into static (static word vector), non-static (non-static word vector), multichannel, CNN-rand (randomly initialized). Convoluted layer: The first layer key is extracted and the operation will be in accordance with the different characteristics of the number of words in each article.

$$ci = f(w \cdot xi : i + h - 1 + b). \qquad (1)$$

(1) Pooling layer: The pooling layer is the same as the convoluted layer. Its main function is to extract functionality from the fully wired layer and to reduce the number of settings in the fully wired layer. Fully wired layer: The information or the information is passed through the convolutive layer, the pooling layer and the fully connected layer in order to obtain the values of the more valuable attributes. [13] [14] In research, after sampling the test data, the data is analyzed and classified to determine the analytical significance of related words. At the same time, after

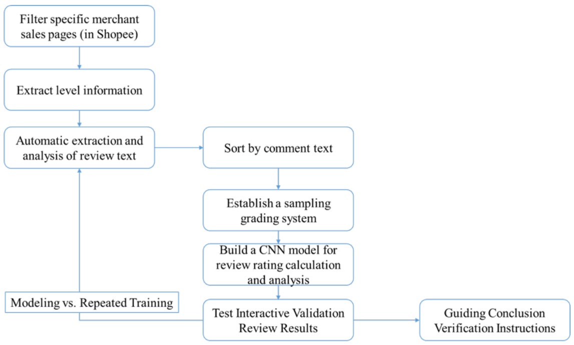

Figure 3. Design of the experimental research.

word analysis, semantic database correlation comparison is performed to improve the match of word analysis results, the transaction amount results are compared to ensure relevance.

2.3. Web Crawler: [15]

The auto crawler program simulates the operation of the user on the computer, simulates the operation of the user to search the specific data of the website through the programming language, and searches and stores the code of the website. In this study, data from the web pages were first captured, and relevant data from the relevant parameters were analyzed, for example print(rss.entries[0]['link']) captures line 0 of the matrix and stores it in a field, and stores pertinent information through the field store. Conduct classification and storage and conduct subsequent word segmentation benchmarking.

3. The Research Framework

This research selects one of the merchant's products and magazines for sampling, analysis and verification of data extraction in an experimental case study based on the Shopee shopping platform. Crawl is used to comment on the results of the verification, import the CNN model for comparison and verification of word segmentation, and monitor each consumption. When consumers are shopping, they are likely to look at reviews from other consumers to help them make a decision. However, not everyone knows the meaning of words, so word frequency classification is used to determine which words consumers use often. The CNN model has been tested on these words and a conclusion has been reached.

4. Experimental description

The study found that most sellers on the Shopee shopping platform are in the 4-4.9 range overall, with a focus on special offers and hot items. When it comes to commenting, consumers are unable to quantify and understand product

序号	字词	频次	频率 %
75	好看	9	0.2839
84	很快	8	0.2524
105	糟糕	7	0.2208
117	完整	6	0.1893
120	不好	5	0.1577
122	超快	5	0.1577
125	非常	5	0.1577
136	快速	5	0.1577
190	下次	4	0.1262
207	不能	3	0.0946
208	不是	3	0.0946
209	不太	3	0.0946
220	壞	3	0.0946

417	不便	1	0.0315
418	不大	1	0.0315
419	不符	1	0.0315
420	不行	1	0.0315
421	不合	1	0.0315
422	不久	1	0.0315
423	不怕	1	0.0315
424	不同	1	0.0315
425	不需	1	0.0315
426	不要	1	0.0315
427	不易	1	0.0315
428	不予	1	0.0315
429	不知所云	1	0.0315

Figure 4. Word frequency analysis ratio after sampling.

feedback suggestions when relying solely on star ratings. In the experiments, quantified star ratings were compared to non-quantified semantic ratings and used to help consumers adjust the corresponding comments that are collected after rating. In this study, a textual analysis of the frequency of words in the comments is carried out on comments on the shopping site Shopee. Using word frequency analysis, you can see which words occur frequently in consumer reviews. For example, set it to start staring: 5 stars for a good look, 4 stars for a good look, etc. Place the classification in the convolution layer of the CNN model and simulate to confirm which calculation method is more accurate. Then we write the formula in code, check that the results of the solution on both sides are consistent and draw a conclusion. For experimental, system comments are used as experimental data. For the experience, the product is selected from the products sold on the Shopee website [9], and since the reviews on the Shopee shopping website are rated by 5 to 10 reviews, it is randomly selected between one and five stars and the word frequency analysis system is applied. Frequency analysis ratio in Fig. 4. After searching and sampling, the frequency score classification is calculated, the results are entered for testing and the CNN model convolutional layer. We have used 2x2, 3x3 and 4x4 computations and it is useful to know which method can test the results with greater accuracy.

4.1. Calculate the frequency of words

The corpus was used for sentence classification in this study. In addition to general keywords and contextual queries, due to the complexity of Chinese dictionary classification. In the Chinese lexical function sketching system, the rearrangement and combination of specific words produce different combination analysis results. Therefore, in search, the corpus is used for corpus search and analysis. This paper uses the word frequency statistic for corpus analysis and processing, just add words to the text content and

文字内容（最长100000字）： 共 171 字符

無可挑剔吧超級好看的個人覺得還蠻好看的無可挑剔吧超級好看的個人覺得還蠻好看的無可挑剔吧超級好看的個人覺得還蠻好看的無可挑剔吧超級好看的個人覺得還蠻好看的無可挑剔吧超級好看的個人覺得還蠻好看的質感真的不錯質感真的不錯質感真的不錯質感真的不錯質感真的不錯質感真的不錯質感真的不錯質感真的不錯商品的品質很好商品的品質很好商品的品質很好商品的品質很好

Figure 5. Review sample data.

序号	字词	频次	频率 %
1	的	14	9.7222
2	質	12	8.3333
3	好看	10	6.9444
4	不	8	5.5556
5	錯	8	5.5556
6	感	8	5.5556
7	真的	8	5.5556
8	吧	5	3.4722
9	超	5	3.4722
10	得	5	3.4722
11	個	5	3.4722
12	還	5	3.4722
13	級	5	3.4722
14	覺	5	3.4722
15	可	5	3.4722
16	蠻	5	3.4722
17	人	5	3.4722
18	挑剔	5	3.4722

Figure 6. Word frequency classification.

75	好看	9	0.2839	5
223	良好	3	0.0946	4
281	不佳	2	0.0631	3
120	不好	5	0.1577	2
105	糟糕	7	0.2208	1

Figure 7. Weights of the stars corresponding to the words extracted from the corpus.

choose them, the word frequency statistic will be calculated automatically. Based on the number of words entered by the user, the corpus analyses the frequency of the words. For example In Fig. 5, 171 words were entered. A total of 23 words were analysed, and the total word frequency value is 144. The initial value of the word is 10%, regardless of the number of words entered and the number of words obtained. The frequency value is 0.69444 whenever the word frequency is 1. For example, the word "Nice" appears 10 times. 0.96444×10=6.9444 is used. (Fig. 6)

4.2. Assessment level

As a result of the large number of analyses, there has been a review and adjustment of the relevant data in the study. The analyzed word frequencies are first suppressed for inapplicable and incompatible data, the remaining data are then divided according to the number of word frequencies. For example, words with a frequency of 9 - 6 will be divided into one range. Similarly, the data is divided into five parts. Once the classification is complete, the comment level, defined as 1 to 5, is obtained from the classification. (Fig. 7)

$$x = \begin{bmatrix} x5\ x3\ x3\ x0 \\ x0\ x2\ x1\ x1 \\ x4\ x5\ x2\ x3 \end{bmatrix} \quad y = \begin{bmatrix} y10\ y9 \\ y5\ y11 \\ y10\ y7 \end{bmatrix} \quad h = \begin{bmatrix} h174 \\ h142 \\ h87 \end{bmatrix}$$

$$\begin{array}{|c|c|c|c|} \hline 5 & 3 & 3 & 0 \\ \hline 0 & 2 & 1 & 1 \\ \hline 4 & 5 & 2 & 3 \\ \hline \end{array} \ \times \ \begin{array}{|c|c|} \hline 10 & 9 \\ \hline 5 & 11 \\ \hline 10 & 9 \\ \hline \end{array} \ = \ \begin{array}{|c|c|c|} \hline 174 & 142 & 87 \\ \hline \end{array}$$

Figure 8. 2x2 text converting quantitative weighting calculating analyzing.

5. Experimental evidence

In our research, we will sample the information collected and analyse the results using a CNN model. In this document, scores are arranged by CNN pattern. This paper calculates 2 x 2, 3 x 3, 4 x 4, where x represents 1 to 5. The first mathematics formula is calculated with 2 x 2, where the y value is the sum of the addition of x5x3x0x2, and so on to get y, if the x values in the oil are added to get 6 y values, so calculate the h value, both x and y need to take out 6 values for calculation. The first block of the input layer in a 2 x 2 calculation is 5 + 3 + 0 + 2 + 4 + 5 = 19, 3 + 3 + 2 + 1 + 5 + 2 = 16, 3 + 0 + 1 + 1 + 2 + 3 = 10, based on the added value of the first block is higher than the value of the second block and the third block, so we obtain 174, also because the value of the third block plus The resulting total value operation is also the lowest.(Fig. 8) When calculating 3×3, x5×y25+x3×y20+x3×y18, as the result, in 3×3 has more values than 2×2, to make the value correct, it is calculated again, z is obtained from y, for example: add y25 y20 y19 y22 to get z86, get all the z values, then use the y×z values to get the h value. (Fig.9) When changing the matrix to 4X4, increase the calculation analysis vector, improve the calculation outcome and increase an r value, in Fig. 10. Convolutional layers in a CNN template are used in this research. The experiments were 2 x 2, 3 x 3 and 4 x 4. The test results show that the corresponding opinions and star ratings are too far away from each other, and that the comparison value with the 3x3 method is relatively significant. From the results of the investigation, a matrix computation on consumption words and library ratings has been performed, and sources of interfering factors have been deduced. The computation and analysis results from the 3 x 3 matrix can better satisfy the correlation between planned reviews and consumer behavior.

6. Conclusion

This study uses the characteristics of the CNN convolutional network to analyse the word frequency, the weight classification and the perception of the word frequency to achieve the purpose of the calculation. For search, we built our own search engine to search specific data from the pages

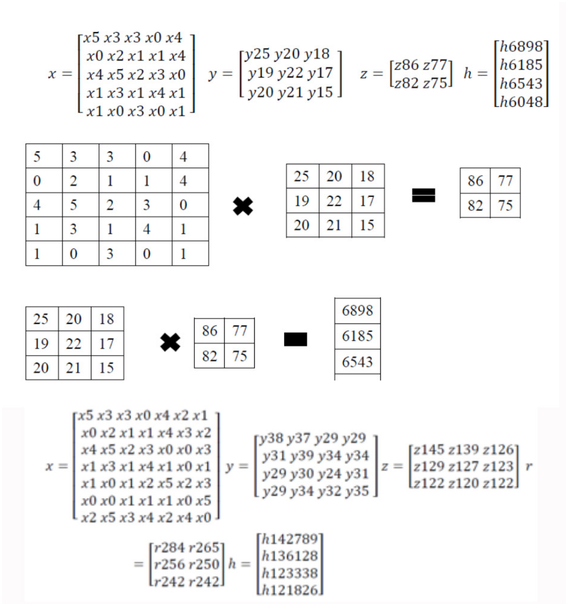

Figure 9. Compare conversion results using 3x3 text evaluation.

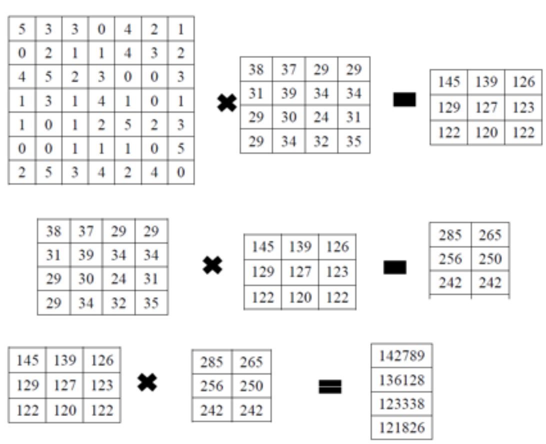

Figure 10. Matrix structure transformation comparative analysis.

of the Shopee website and calculate and convert feedback rating. In our research, we use convolutional network features to combine calculation and estimation to obtain inference calculation results, when positive and negative emotions cross or parallel, whether it affects consumers' product evaluation can be discussed in subsequent derivative works.

References

[1] Saarijärvi H. Yrjölä M., Rintamäki T. and Joensuu J. Consumer-to-consumer e-commerce: outcomes and implications. *The International Review of Retail, Distribution and Consumer Research*, 27(3):300–315, 2017. 1, 2

[2] Wu L. Hong W., Zheng C. and Pu X. Analyzing the relationship between consumer satisfaction and fresh e-commerce logistics service using text mining techniques. *Sustainability*, 11(13):3570, 2019. 1

[3] Ferdows R. Purba M. R., Akter M. and Ahmed F. A hybrid convolutional long short-term memory (cnn-lstm) based natural language processing (nlp) model for sentiment analysis of customer product reviews in bangla. *Journal of Discrete Mathematical Sciences and Cryptography*, 25(7):2111–2120, 2022. 1

[4] Poulin M. Tariq M. U., Babar M. and Khattak A. S. Distributed model for customer churn prediction using convolutional neural network. *Journal of Modelling in Management*, 17(3):853–863, 2022. 1

[5] Verhagen T. Goldman S. P., van Herk H. and Weltevreden J. W. Strategic orientations and digital marketing tactics in cross-border e-commerce: Comparing developed and emerging markets. *International small business journal*, 39(4):350–371, 2021. 2

[6] Zeithaml V. A. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *Journal of marketing*, 52(3):2–22, 1988. 2

[7] Dandawate Y. Kaunchi P., Jadhav T. and Marathe P. Future sales prediction for indian products using convolutional neural network-long short term memory. *In 2021 2nd Global Conference for Advancement in Technology (GCAT)*, pages 1–5, 2021. 2

[8] Li Z. and Hou A. C. Online purchase preference and individual characteristics: A moderation approach. *International Journal of Electronic Commerce Studies*, 10(1):1–21, 2019. 2

[9] Zhu H. A deep learning based hybrid model for sales prediction of e-commerce with sentiment analysis. *In 2021 2nd International Conference on Computing and Data Science (CDS)*, pages 493–497, 2021. 2

[10] Rasoolimanesh S. M. Wong E. and Pahlevan Sharif S. Using online travel agent platforms to determine factors influencing hotel guest satisfaction. *Journal of Hospitality and Tourism Technology*, 11(3):425–445, 2020. 2

[11] Kim H. and Jeong Y. S. Sentiment classification using convolutional neural networks. *Applied Sciences*, 9(11):2347, 2019. 2

[12] Yechuri P. K. and Ramadass S. Semantic web mining for analyzing retail environment using word2vec and cnn-fk. *Ingénierie des Systèmes d'Information*, 26(3):10, 2021. 2

[13] Widiastuti N. I. Convolution neural network for text mining and natural language processing. *In IOP Conference Series: Materials Science and Engineering*, 662(5):52–100, 2019. 2

[14] Guofeng Y. Zikang H., Yong Y. and Xinyu Z. Sentiment analysis of agricultural product ecommerce review data based on deep learning. *In 2020 International Conference on Internet of Things and Intelligent Applications (ITIA)*, pages 1–7, 2020. 2

[15] Khder M. A. Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and Its Applications*, 13(3):1–10, 2021. 3