Journal of
Management and Humanity Research

Characteristics of Useful Reviews: A Text-Mining Research Based on Clothing Products

Lan-ling Huo[1] and *Xiao-yu Liu[2*]*

[1]School of Management, Shenzhen Polytechnic, Guangdong 518055, China.
[2]School of Economics and Management
Chongqing University of Posts and Telecommunications
Chongqing 400065, China.
[*]Corresponding Author: Email: 374060528@qq.com

Abstract. This paper uses octopus software to capture 3170 online review data of popular clothing products in Jingdong Mall as the data set of this research. The comment features are quantified by text mining method, and the comment keywords are extracted by TF-IDF feature extraction algorithm. Based on the results of previous studies on the characteristics and usefulness of reviews, this paper proposes five hypotheses and makes an empirical analysis using Poisson regression model. It is concluded that comment pictures and comment heat significantly positively affect the usefulness of online comments on clothing products, while comment timeframe, additional comment and comment length do not have significant influence on the usefulness of comments. Then, according to the results of empirical analysis in this paper, suggestions are put forward for online shopping platforms and merchants to strengthen platform control, increase activity and improve product marketing methods.

Keywords: comment feature, useful comment, Poisson regression, text mining

1. Introduction

In recent years, the development of network technology has brought huge changes to the e-commerce industry, and many well-known shopping platforms have emerged, such as the three e-commerce giants Taobao, Tmall, JD, etc., and now Pinduoduo is also among the forefront. People have brought great convenience to life through these emerging online shopping platforms. At the same time, the improvement of the economic level has caused the income of Chinese residents to continue to increase, and the upgrading of consumption level has constructed China's current online shopping market with a good development trend. The sudden new crown epidemic in 2020 has brought more attention

to the originally well-developed online shopping. The quarantine of the epidemic has restricted people's travel, the convenience of online shopping has attracted a large number of people, and online shopping has gradually become a part of many people's lives. According to data released by iiMedia Consulting in 2020, the sudden new crown epidemic has caused my country's total consumption of consumer goods to drop by 3.9%. However, the online retail sales of physical goods reached 9.76 trillion RMB. According to the information on the characteristics and development trend of my country's online retail market in 2021 released by China Research Network in May 2021, my country's online retail sales have reached 3.8 trillion RMB from January to April.

Although online shopping has brought many benefits, it can save shopping time and cost, and submit commodity transaction rates, but it is difficult for consumers to judge the quality of products through the Internet. Buyers can only understand the product through the pictures displayed by the seller in the store on the platform, the parameters set, and the text description. These data are all set by the seller, and the authenticity cannot be verified, which is likely to cause inconsistency with the actual product. As a result, buyers may buy undesirable goods, increasing the risk of shopping. Compared with online shopping, under the traditional offline transaction mode, consumers can not only see the real goods, but also check the fabrics, materials and workmanship, and even try them on directly to check the upper body effect of the clothing. Therefore, in order to reduce the risk of shopping, consumers can only use other consumers who have purchased the product in order to reduce the risk of shopping, in addition to understanding the product from the appearance of the product according to the introduction of the text and the picture displayed in the window. Some of the reviews, such as purchase volume, number of reviews, favorable rate, negative rate and many other factors to determine whether the product meets the demand and whether it can meet expectations.

By viewing the online reviews of the product, consumers can understand the product more intuitively and comprehensively, improve the authenticity of the product, and reduce the risk of purchase. Therefore, online product reviews have gradually become more trusted and concerned information by consumers. However, the accumulated amount of online review data is already very large, which creates a pain point for buyers and sellers that the amount of information is too large and it is difficult to obtain useful information. Therefore, how to obtain useful information from review data and summarize the characteristics of useful reviews is another topic worthy of research.

2. Literature review

With the development of social networks, online communication has become an indispensable way for people to conduct social activities. At the same time, the

emergence of products such as communities and forums has facilitated people's communication, triggered online evaluations by the whole people, and attracted many scholars to start researching online reviews. Online comments refer to people using online platforms to subjectively or objectively express their views and opinions on a certain thing or event. At present, most social network platforms have added the comment area module. The platform has the function of online comments. Users can post their views, view other people's comments, understand other people's ideas, discuss and like them in the comment area. As far as the platform is concerned, with social functions, users can interact, which can enhance the user experience and increase the activity of the platform, so that the product can get more attention and attract more users. Since online reviews are based on a network platform, the time cost is low, and information can be exchanged quickly and flexibly, and there is no restriction on quantity and quality. Therefore, the number of comments will continue to accumulate, forming a huge amount of data, which will make it difficult to easily obtain valuable information from the comment area. Therefore, many scholars have initiated research on the usefulness of online reviews.

At present, the definition of useful comment that is most recognized by the majority of scholars is to start from the perspective of perceived value and judge whether the comment is valuable to other comment readers. The more helpful the comment is to others, the more useful it proves to be. But because usefulness is subjective, it cannot be accurately measured. Therefore, some scholars perceive the usefulness of online comments by consumers based on the number of useful votes obtained from online comments on websites in their research. In studying the usefulness of online comments in the context of user communities, Li Yumeng uses the number of comments collected as a dependent variable to measure the usefulness of comments [1]. Zhang Yanhui uses Taobao's online comment data to measure the usefulness of comments by the number of useful votes obtained by the comments [2]. High-quality reviews can effectively affect consumers' desire to buy and shopping decisions, which in turn affect product sales, platform activity, and user trust.

Most of the previous studies on online reviews have adopted the form of model verification and questionnaires. On the one hand, they include from the perspectives of merchants, review readers, review authors, etc., to find out the influencing factors of purchasing decisions based on reviews. Infer the consumer needs of buyers, and then make improvements based on these feedbacks and formulate relevant marketing plans. The other is to find out the comments that are considered useful and to analyze the characteristics of these useful comments.

According to the research of many scholars, the useful comment characteristics obtained mainly include the text length of the product review, the favorable rate, the

negative review rate, the rating level, the popularity of the review, the number of votes for the "useful" review, and the sales volume of the product. The 2015 Phocuswright survey on Trip Advisor's user evaluation service shows that users gradually value the content of reviews. By combining the text and digital features of online reviews, Zhou Yusheng conducted research and analysis on the content and title of the review, and finally concluded that the features that affect the usefulness of the review include the title and content of the review, the text feature of the review, and the digital feature. These features will also Affected by the type of comment [3]. In studying the influence of online reviews on the sales of clothing products, Zhu Shujiang analyzed the characteristics of the number of comments with pictures, the number of reviews, the number of comments, and product prices [4]. Dong Shuang and others believe that on different shopping platforms, consumers pay more attention to product features and emotional expression features in reviews, while paying less attention to service features contained in reviews [5]. Yu Mingnan, Wu Fang and others believe that the timeliness of online reviews positively affects consumers' purchase intentions [6]. Zhang Li and others believe that the quality of online reviews, graphic reviews, and additional reviews will have a positive impact on consumers' purchase intentions [7]. Song Sujuan and others found that the number of pictures, follow-up reviews, videos, and the number of comment responses significantly affect the usefulness of comments [8]. Research by Cai Limei et al. found that comment length has a more obvious impact on the usefulness of experiential product reviews [9]. Susan M. Mudambi found that the extreme degree of reviews, the depth of reviews and product types will affect the perceived usefulness of reviews [10].

Combining previous studies, this article will focus on the timeliness characteristics of reviews, the length of the text of the reviews, the number of pictures in the reviews, the number of high-frequency words that appear in the reviews, that is, the popularity of the reviews, and the number of follow-up reviews. These characteristics are used to explore the characteristics of the reviews Usefulness.

3. Research model
3.1. Research hypothesis
Online reviews refer to the evaluation information published on the shopping platform by buyers who have previously purchased a product, describing the characteristics of the product, shopping experience, logistics, and after-sales service. Because sellers deliberately brush good reviews and malicious bad reviews from bad competitors among merchants, the accuracy and usefulness of online reviews are reduced.

Based on previous research results, combined with the particularity of the clothing product (wearable products) that this article focuses on, this article mainly studies the usefulness of online comments from five aspects: comment length, comment popularity,

comment timeliness, additional comments, and comment pictures.

The text length of online reviews largely represents the amount of information carried by the product experience. So many scholars study the relevance of comment length and useful comments. Hao Yuanyuan believes that the more text in the comments, the more useful the comments will be. Friske believes that the longer the text length of the review, the lower the usefulness of the review for buyers and sellers, that is, the length of the review is negatively correlated with the usefulness of the review [11]. This article uses the word count of the comment text to represent the length of the comment.

H1: Comment length positively affects the usefulness of online product reviews.

The high-frequency words appearing in the content of online reviews can reflect the buyer's focus on goods and services, that is, the popularity of reviews. Generally, the higher the frequency of the appearance of the words describing the characteristics of the product or the representative of the service appear in the buyer's reviews, the more the buyer pays attention to the product attributes or service characteristics of the product. Dong Shuang believes that the more information about the product is expressed in the review content, the more useful this comment is [5]. This article uses the frequency of high-frequency words that appear in comments to measure the popularity of comments.

H2: The popularity of reviews positively affects the usefulness of online product reviews.

After the buyer receives the goods and publishes product reviews through the online platform, the platform will display the time of the review. To a certain extent, the comment time can show the changes in the user's shopping experience. There are contradictory conclusions in the current research on the correlation between the timeliness of reviews and the usefulness of reviews. Ipeirotis et al. believe that the longer the comment interval, the higher the usefulness of the comment. On the contrary, Yin Guopeng's research believes that the comment interval is negatively correlated with the usefulness of comments [12]. In order to explore the correlation between the two under clothing products, this article uses the time interval between comment release and comment collection to measure the timeliness of comments.

H3: Comment timeliness positively affects the usefulness of online product reviews.

The number of comments with pictures refers to the number of physical pictures of the product released when the product is commented online after receiving the goods. Since the physical images released by buyers and the effect images of the try-on are closer to reality, they have more reference value and are worthy of the trust of other buyers. Because the product information contained in the picture is more direct, the buyer's emotions when posting a comment with the picture are more real. So when doing research, you also need to pay attention to the number of pictures uploaded in the comments. This article uses the number of pictures included in the comment to represent

the characteristics of the commented picture.

H4: Comment pictures positively affect the usefulness of online product reviews.

The number of additional reviews refers to the number of additional reviews left by consumers after product reviews. Consumers can reply to previous reviews in the review area of the shopping platform to show their supplements to the reviews and describe the shopping experience more truthfully. The characteristics of additional comments in this paper are measured by the number of follow-up comments.

H5: Additional reviews positively affect the usefulness of online product reviews.

3.2. Variable measurement

In previous studies, most scholars have modeled product sales or the number of useful votes for reviews as dependent variables, but product sales do not directly reflect the usefulness of a single review, and the number of useful votes for comments on Taobao is also rarely, thus the research value is not high. Therefore, this article uses the number of likes of comments on Jingdong as a measure of the usefulness of the dependent variables of this article, and selects the length of the comments, the popularity of the comments, the timeliness of the comments, the additional comments, and the comment pictures as the explanatory variables (that is, the potential characteristics of the usefulness of the comments). Verification to discover the main characteristics of usefulness reviews. The measurement indicators and methods of variables are shown in Table 1 below.

Table 1: Research variables

Variable type	variable	Measurement methods
Independent variable	Comment Heat	Number of high-frequency words
	Comment length	Comment text word count
	Comment timeliness	The number of months between the release time of the comment and the collection time
	Additional comments	Review number
	Comment image	Number of comments
Dependent variable	Comment usefulness	Comment number of votes

3.3. Model construction

The measurement characteristics of the dependent variable will affect the choice of analysis model. In reality, the value of some variables can only be measured by the

number, such as the number of urban patents and published papers, the number of births and deaths in the country, and the number of crimes in a city. These data have one thing in common. That is, they cannot be negative numbers and decimals, and they are countable and finite. The analysis of this type of data is not suitable for the use of traditional multiple linear regression models (requiring the dependent variable to be a continuous value) or logarithmic model for analysis (requiring only a small number of discrete values for the dependent variable. Because this article uses the number of likes as the measurement comment Useful proxy variable. This is a non-negative, discrete, and low-incidence count variable. This dependent variable is suitable for analysis using a Poisson regression model. The Poisson regression model constructed in this paper is as follows:

$$\log \mu = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + x_5\beta_5$$

where β_i is the parameter vector, xi represents the characteristic variable, and μ represents the expected mean value of votes.

4. Collection of comment data

4.1. Data selection

Currently, the main domestic online shopping platforms are Taobao and JD. For Taobao.com, its daily website visits have reached more than 60 million, and an average of at least 48,000 products can be sold per minute. The accumulation of online review data for these products is also extremely large. However, because the number of useful votes on Taobao is very small, using them as representative data of the usefulness of comments will lead to lower research effects, and the number of useful votes on Taobao cannot be crawled by tools now, which is not conducive to the research of this article.

As a self-operated e-commerce platform in my country, JD.com adopts a typical B2C model to directly communicate with buyers face-to-face through platform merchants. Jingdong Mall ranks first in the number of visitors to electronic products, and is currently the mainstream online shopping platform in China like Taobao. According to JD's latest performance report for the first quarter of 2021, JD.com currently has nearly 500 million active purchasers each year. Its net income in the first quarter of this year has reached 203.2 billion yuan, and its market share has reached 30%. The quality and reputation of JD.com products are recognized by consumers. In the product evaluation on JD.com, the number of people who like the evaluation is generally higher than that on Taobao, and the number of likes and the number of follow-up comments can be collected through crawler tools. After comparison, it is found that the data of Jingdong Mall meets the research needs. The hot-selling products will be more valued by a wide range of consumers, since a large amount of evaluation information will be generated, and women will pay more attention to evaluation information than men. Therefore, this paper selects the comment

data of dresses and T-sleeves, which are popular clothing products on JD.com, for research.

4.2. Data crawling

The research data in this article comes from Jingdong Mall, so it is necessary to crawl the product evaluation information of the webpage to collect the data. Since the main research variables are comment popularity, comment pictures, comment length, additional comments and comment timeliness, these five comment characteristics and comment usefulness. And it is measured by the number of high-frequency words in the comment content, the number of pictures in the comment, the number of text words in the comment, the number of follow-up comments, the month between comment release and collection, and the number of comment likes. This information can be collected by the crawler tool Octopus, No need to write code.

Because the octopus can set the data collection process and create collection tasks according to its own needs. After entering Octopus, click on the custom task and enter the URL that needs to collect comment information. After saving the settings, Octopus will open the URL. Since the webpage of Jingdong Mall is a dynamic webpage, it is associated with a database, and the content of the page can be changed according to the parameter changes in the link address. And the product evaluation information needs to log on to the website, enter the account number and password to view it. Therefore, when using octopus to collect data, you need to set up Cookies to maintain the login status and avoid repeated input of account and password to log in during the crawling process.

Since the product review data is displayed by pagination, it is necessary to crawl the data of each page in a loop, and then select the corresponding field of each required review data in each page and put it into the list, and set the list loop. The comment data in each page can be extracted cyclically. When the data on this page is crawled, click to turn the page, you can also crawl the data on the next page. Embed the cycle list and page turning settings into the outer cycle page turning, and then click the save process to automatically perform local collection.

Buyers can evaluate the order in the order information, post comments, and express their views on the purchase after purchasing goods, express delivery, and confirmation of receipt in Jingdong Mall. After the text information entered by the buyer on the evaluation page is published, it will be displayed in the product review as the text part of the review. After text processing, the text content that will be crawled in this article can be used as a feature of comment popularity. If there are uploaded videos and pictures, they will be displayed below the text comment. What Octopus collects is the link corresponding to each picture in the comment, so this article uses the number of picture links in each comment as the comment picture feature. The platform will display the

specific release time of the comment below the picture. The timeliness of the comment is reflected by the month difference between the comment release and the collection date, so this article intercepts the date part of the collected specific release time. Remove the specific time point, and then calculate the month between the date and the date of collecting comments in this study as the quantitative value of the comment timeliness characteristics.

After the comment is published, other buyers obtain information by reading the comment, and can reply to the comment, which will generate a follow-up comment. Jingdong Mall will count the number of follow-up comments for each comment and display it below the comment. Therefore, the additional comment feature can be obtained directly by crawling the number of follow-up comments below each comment.

After other buyers read the review, they will subjectively perceive whether the review contains useful information. If the reader finds it useful, they will like the review information, and for a single review, each reader only I can like it once. In other words, how many likes are, how many people think the comment is useful. Therefore, the number of likes for each comment displayed in the comment area can be used to quantify the usefulness of the comment. The comment length feature needs to be obtained by calculating the text length of the comment content. According to the research needs, the useful fields captured by the octopus collector in this article include the text content of the comment, the link of each picture contained in each comment, the number of likes, the number of follow-up comments, and the specific comment release time. After selecting the content that needs to be captured, modify the field name and save the settings to start automatic data collection.

By removing the remaining 3170 comments automatically by the system, deleting the default value and the data whose comment text length is less than 10, and then selecting the comment data whose number of likes is not 0 as useful comment data, totaling 2,234.

4.3. Data mining

The research object of this paper is online comment data, which belongs to text information, which needs to be processed by text mining to quantify its characteristics. Since the data is crawled directly through the crawler tool, the types and formats of the fields need to be unified. We unify the comment content into text type in Excel, and comment pictures, additional comments, and comment usefulness are obtained directly through crawling in the data set studied in this article, so there is no need to calculate them, just unify them Processed as numerical data. The length of the comment can be calculated by using the len() function in Excel to calculate the length of the comment text, and the result is also numeric data. The comment release time is cut out by the column

function, and then the comment release date and the comment collection time are unified into a short date type, and then the interval month between the two dates is calculated through the datedif() function, and the value obtained is used as Comment timeliness. Comment popularity is measured by the number of high-frequency words in the comment content, so the next step is to conduct text mining on the text content.

Words in English documents are separated by words, while Chinese characters do not have obvious separators to divide words. After word segmentation, the computer can recognize the meaning of the sentence more effectively. Therefore, when doing text mining, the Chinese text data must be segmented. At present, the mainstream word segmentation algorithms are completed on the basis of word frequency statistics, knowledge understanding, and dictionary and thesaurus matching. There are also many common word segmentation tools. Since the use of jieba word segmentation is relatively simple and widely used, this article uses jieba word segmentation for processing. Python can directly import the jieba package and call the cut() method in the package to automatically segment the text. Therefore, this article directly calls the jieba package for word segmentation. The word segmentation mode can be set when calling the package. The first input parameter of the cut() method is the text data that needs word segmentation, and the second parameter cut_all can specify the word segmentation mode. Since the full mode cannot handle the problem of Chinese ambiguity, and the search engine mode is mainly for search engines, this article uses precision mode. That is, the value of cut_all is False.

The next step is to remove the stop words in the text. Because there will be many words with no practical meaning in Chinese, such as the conjunctions "and" and "or", the modal particles "ah" and "oh", and the prepositions. The proportion of these words in the text is very high, which will affect the results of the text analysis, so they need to be removed. As clothing product reviews often appear as interpersonal appellations such as "pro" or "girlfriend", these words have no meaning for research, so they also need to be added to the stop word file for processing. This article implements the removal of stop words by writing Python code.

Among them, jieba word is the text data that needs to remove the stop words, and stopword is the stop word list. After reading the text data in a loop and comparing the stop words list, the data after removing the stop words is stored in the clean words file. The purer comment data obtained after word segmentation and removal of stop words.

5. Data analysis of online comments
5.1. Descriptive statistical analysis
This article uses Python to quantify the comment data of the hot-selling clothing products on JD.com, and then analyzes the comment data's comment length, comment timeliness,

comment usefulness, additional comments, comment popularity, and comment pictures. Descriptive statistics, the descriptive statistical results of each variable are shown in Table 2 below.

Table 2: Descriptive analysis results

Variable name	Sample size	Minimum	Max	average value	Standard deviation	Median
Comment usefulness	2234	1.000	12.000	5.000	3.366	4.000
Comment Heat	2234	1.000	17.000	6.500	4.994	4.500
Comment image	2234	0.000	5.000	1.600	1.350	2.500
Additional comments	2234	0.000	5.000	1.000	0.494	0.000
Comment length	2234	28.000	238.000	76.800	64.677	49.500
Comment timeliness	2234	5.000	60.000	25.200	17.169	20.000

5.2. Regression results

According to the following verification results in Table 3, it can be seen that P<0.05 indicates the original hypothesis: whether there are five independent variables of comment pictures, comment length, comment timeliness, additional comments, and comment popularity has no effect on the usefulness of the model Not true, which means that when constructing the analysis model in this article, putting these five independent variables will have an impact on the quality of the Poisson regression model, which means that the Poisson regression model constructed in this article is effective.

Table 3: Verification of regression analysis

model	-2 times log likelihood	Chi-square value	df	P	AIC value	BIC value
Intercept only	451.895					
Final model	213.913	237.982	5	0.021	225.913	237.264

According to the results of Poisson model regression analysis, the following conclusions can be drawn. Poisson regression takes the comment pictures, comment popularity, comment timeliness, additional comments, and comment length as independent variables, and the usefulness of online reviews of clothing products as dependent variables analysis. The regression analysis results are shown in Table 4.

Table 4: Regression results

Independent variable	coefficient	Standard error	z value	p value	OR value	OR value 95% CI
Comment Heat	2.965	1.377	2.154	0.031	19.402	1.306 ~ 288.328
Comment image	0.250	0.118	2.113	0.035	1.284	1.018 ~ 1.619
Additional comments	-0.029	0.089	-0.322	0.747	0.972	0.817 ~ 1.157
Comment length	-0.013	0.016	-0.797	0.425	0.987	0.957 ~ 1.019
Comment timeliness	-0.000	0.043	-0.007	0.994	1.000	0.918 ~ 1.088
intercept	1.358	1.837	0.739	0.460	3.886	0.106 ~ 142.214
R square	0.527					

Dependent variable: Comment usefulness (number of votes)

Based on the regression results in Table 4, we can draw the following research conclusions.

First, for the factor of comment popularity, the regression coefficient is 2.965 and it is statistically significant ($p=0.031<0.05$), which indicates that comment popularity has a significant positive impact on the usefulness of online clothing reviews. In other words, comments Popularity is one of the obvious characteristics of usefulness reviews. Therefore, Hypothesis 2 holds.

Second, for the factor of comment pictures, the regression coefficient is 0.250, and it is statistically significant ($p=0.035<0.05$), which indicates that comment pictures will have a significant positive impact on the usefulness of clothing online reviews, which means Commenting pictures are one of the obvious features of useful comments on clothing products. Therefore, Hypothesis 4 is true.

Third, for the element of additional reviews, the regression coefficient value is -0.029 but it is not statistically significant ($p=0.747>0.05$), which means that additional reviews will not have a significant impact on the usefulness of clothing online reviews, so suppose 5 is not true.

Fourth, for the factor of comment length, the regression coefficient value is -0.013, but it does not show significance ($p=0.425>0.05$), which shows that the comment length does not have a significant impact on the usefulness of online comments, namely Comment length cannot be an obvious feature of usefulness comments, which means that Hypothesis 1 is not true.

Fifth, for the factor of comment timeliness, its regression coefficient value is -0.000 and does not show statistical significance ($p=0.994>0.05$), which shows that comment

timeliness does not have a significant impact on the usefulness of product online reviews , Nor can it become the main feature of usefulness reviews. Therefore, Hypothesis 3 is not true.

Based on the above regression analysis results, we can generate the following conclusions. Review popularity and review pictures have a significant positive influence on the usefulness of online reviews of clothing products, and are important features of useful reviews. However, comment timeliness, additional comments, and comment length have not become obvious features of the usefulness of apparel product reviews.

6. Conclusion and enlightenment
6.1. Conclusion

This paper uses Octopus software to capture the online comment data of popular clothing products on JD.com, such as hot-selling dresses, T-shirts, etc., as the text data needed for this research. Using the literature summary method again, based on previous research results on the characteristics and usefulness of reviews, five research hypotheses were proposed, and a research model was designed for empirical analysis. The research conclusions obtained are as follows:

First, the popularity of reviews significantly positively affects the usefulness of online product reviews, which is one of the main characteristics of clothing product usefulness reviews. This shows that the more features such as product quality, service, and logistics are described in reviews, the more buyers' attention will be attracted.

Second, review pictures significantly positively affect the usefulness of online product reviews, and are one of the main features of clothing product usefulness reviews. This shows that the comments with pictures can better reflect the authenticity of the comments, and at the same time, the comment recipients can estimate the gap between the physical product and the physical image displayed by the merchant through the pictures uploaded by the buyer. Therefore, the information contained in the comments with pictures is more attractive to buyers.

Third, comment timeliness, additional comments, and comment length do not have a significant impact on the usefulness of online product reviews, and have not become the main features of the usefulness reviews of clothing products. This may be because the data selects reviews of popular clothing products, and the continuous accumulation of review data has resulted in a large amount of review data for such products, and the review information displayed on the platform is displayed according to the latest release time. With limited time and energy, consumers are less likely to pay attention to whether there are follow-up and comment release time and text length.

6.2. Enlightenment

Through empirical analysis, this paper finds from a theoretical perspective that the two characteristics of review popularity (the number of high-frequency words in the content) and the number of pictures in the review have a significant positive correlation with the usefulness of online reviews. The research findings in this article provide the following suggestions for online shopping platforms, merchants, and commodity manufacturers:

Merchants can encourage consumers to simultaneously provide physical pictures when they post reviews, so as to enhance the usefulness of online reviews to other consumers, and provide more useful information for consumers' product selection and decision-making. Through the comments with pictures, buyers can clearly know whether the product is in line with the actual product, and it is convenient for buyers to consider buying, in case there are quality problems in the online shopping process. At the same time, it can also let sellers know how much buyers like the items they are shopping, which will help the communication between sellers and buyers. Evaluation with pictures can also allow sellers to improve their goods and provide buyers with better services. In addition, comments with pictures can more truly reflect the buyer's shopping experience, and it is more conducive to the platform to judge the authenticity of the comments, strengthen the management and control of the platform, and thereby increase the trust of buyers and sellers on the platform. Therefore, merchants can encourage buyers to make comments with pictures by means of rebate on comments with pictures, upgrade the level through comments with pictures on the platform, or send gold coins (deductible when shopping again).

Merchants/commodity manufacturers should pay full attention to online reviews deemed useful by consumers, especially the high-frequency words involved in the content of useful reviews. These high-frequency words reflect consumers' concerns and needs for products, and can become manufacturers to develop products. The starting point and important reference for innovation or service innovation, and then provide consumers with products with higher satisfaction and maintain their brand loyalty. However, due to the limited time and professionalism of buyers, the content of the comments may not be able to describe the shopping experience and product information more effectively. Therefore, the platform can recommend keyword tags, such as product features, service features, logistics, etc., when buyers make comments. Buyers can choose the corresponding label to describe. This not only facilitates buyers to write reviews, saves review time, but also improves the high-frequency word information described in the reviews, increases the popularity of the reviews, and thereby improves the usefulness of the review information.

REFERENCES

1. Y.M.Li., Research on the influencing factors of the usefulness of online comments in the context of user communities. *Xi'an University of technology*, (2020)

2. Y.H.Zhang and Z.W.Li., Research on the influencing factors of the usefulness of online reviews: the moderating effect based on product type, *Management review,* 28(10) (2016) 123-132.

3. Y.S.Zhou, Research on the effectiveness mechanism of online comments based on text mining, *Zhejiang University of Finance and Economics*, (2019).

4. S.J. Zhu, The influence of online comments on the sales of clothing products, *Yunnan University* (2018).

5. S Dong, X.H Wang and Z.H Ge, Analysis of online comment content characteristics of B2C shopping websites based on text mining, Library theory and practice (2017) 54-58.

6. M.N.Yu, F.Wu and M.M.Zhang, Research on the influence of online reviews on customer purchase intention, *Market Weekly* (Theoretical Research) (2014) 58-60+83.

7. L.F Zhang, Haimeng and Chen Xue, Research on the impact of online reviews on consumers' willingness to purchase, *International Journal of Modelling in Operations Management,* 8(1) (2020).

8. S.J.Song, S.F.Li, Z.Y.Ji, Y.Y.Zeng and W.Peng, Research on the influencing factors of the usefulness of online reviews - based on the empirical research of Jingdong Mall, *Science and Technology Entrepreneurship Monthly,* 34(02) (2021) 72-79.

9. L.M.Cai, M.Qiu, Z.L.Yang, L.Z.Siyang, X.Y.Song and W.Peng, Research on the influencing factors of the usefulness of online comments, *Business and Management*, (2020) 49-55.

10. Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com, *MIS Quarterly*, 34(1) (2010).

11. P.Racherla and W.Friske, Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categorie, *Electronic Commerce Research and Applications*, 11(6) (2012).

12. J.Y.Yan, L.Zhang and L.Zhang, An empirical study on the influence of online comment content on the usefulness of comments in E-commerce, *Information Science*, 30(05) (2012) 713-716.

13. G.P.Yin, What online reviews do consumers think are more useful?——The influence of social factors, *Management World*, (2012) 115-124.

14. X.L.Liu and K.Zhao, Research on the identification of key factors affecting the usefulness of online reviews, *Modern Intelligence*, 37(1) (2017) 94-99.

15. Q.F.Min, L.Qin and K.L.Zhang, Research on the factors affecting the usefulness of online comments, *Management Review,* 29(10) (2017) 95-107.

16. Q.Gu, Research on the principle and application of web crawler technology, *Information and Computer* (theoretical Edition), 33(4) (2021) 174-176.

17. J.Q.Chang and W.Shen, Research on Chinese word segmentation algorithm based on string matching, *Industrial Control Computer,* 29(2) (2016) 115-116.

18. B.Cao, Y.D.Su and Q.Deng, Automatic recognition of Chinese names based on maximum entropy model, *Computer Engineering and Applications* 45(4) (2009) 227-228.

19. S.Y.Zhao, Research and design of false network information filtering platform for public opinion monitoring, *Inner Mongolia University of technology* (2018).

20. Z.S.Zhao, X.A.Fu, X.S.Jin and Y.Liu, Spam SMS recognition based on random forest feature selection, *Computer and Information Technology,* 26(6) (2018) 24-26.

21. X.T.Zhang, Research on Weibo retrieval model integrating user interest and hybrid estimation, *Hebei University*, 45(4) (2019) 227-228.