*Research Article*

# Competitive Product Identification and Sales Forecast Based on Consumer Reviews

**Guoquan Zhang** [ID] **and Haibin Qiu**

*School of Management, Jilin University, Changchun, Jilin 130022, China*

Correspondence should be addressed to Guoquan Zhang; gqzhang@jlu.edu.cn

Sellers readily obtain consumer product evaluations from online reviews in order to identify competitive products in detail and predict sales. Firstly, we collect product review data from shopping websites, social media, product communities, and other online platforms to identify product competitors with the help of word-frequency cooccurrence technology. We take mobile phones as an example to mine and analyze product competition information. Then, we calculate the product review quantity, review emotion value, product-network heat, and price statistics and establish the regression model of online product review forecasts. In addition, the neural-network model is established to suggest that the relationships among factors are linear. On the basis of analyzing and discussing the impact of product sales of the competitors, product price, the emotional value of the reviews, and product-network popularity, we construct the sales forecast model. Finally, to verify the validity of the factor analysis affecting the sales and the rationality of the established model, actual sales data are used to further analyze and verify the model, showing that the model is reasonable and effective.

## 1. Introduction

The new innovative enterprises, such as Xiaomi, take the internet as their main venue and rely on high-quality and low-price products to earn a high reputation among consumers. With the rapid spread of online product reputation information, these enterprises can quickly seize market share from traditional enterprises. With traditional enterprise development, the lengthy intermediate links create high enterprise costs, making business transformation difficult. Therefore, a low profit rate cannot support expansion and desired market competitiveness. Internet enterprises, on the other hand, omit many intermediate links, relying on low inventory and high turnover rates of capital to reduce costs. However, with business management, owing to insufficient sales forecasts, there are often shortages that seriously affect the competitive position of enterprises [1]. Thus, sales forecasting is an important task for internet companies [2].

In consumer-behavior research, it is generally believed that only when the utility obtained by the consumer is positive will the consumer consider purchasing a product [3]. Based on utility theory, a consumer's willingness to purchase is related to price satisfaction. In the context of multiple competitive retailers, the market share of the product can be obtained [4]. Additionally, consumer-purchase intent is affected by product prices, seasonal factors, market expansion, repeat-purchase willingness, technology growth, etc. Sales also have an impact on consumers' willingness to purchase, which, in turn, affects enterprise sales forecasts [5]. Within the sales forecasting process, product sales are random across time, and there is no fixed rule for product sales. This requires dynamic forecasting and production planning during the product life cycle [6]. There remains an important issue to be considered during the forecasting process. Prediction error increases over time, and the bullwhip effect intensifies. For products having a large market demand, the bullwhip effect can seriously reduce prediction accuracy [7]. Thus, the actual data in the procurement and sales process need to be analyzed to find the right forecasting method [8].

Compared with traditional enterprises, internet companies have a unique advantage regarding the acquisition of consumer feedback, owing to online product transactions and feedback venues [9]. Consumers obtain products from a shopping website, and they send feedback on the product experience via online reviews, social media, and product communities (i.e., user-generated content (UGC)) [10]. With the help of massive consumer feedback, data mining methods help companies analyze the advantages and disadvantages of products [11] and forecast sales [12].

When studying consumer reviews for this research, it was found that consumers not only describe product features and apply emotional feedback in their comments, they also compare their products with those of similar manufacturers, implying that not only are product sales influenced by the products themselves, they are also influenced by competitors [13]. Competitors can thus seize market share, and their ads can be tailored to grab consumers' attention based on the online feedback phenomena.

The previous research focused on the factors affecting consumer reviews and the relationship between consumer reviews' emotional scores and product sales by empirical research. The research mainly considered the impact of the product review on the product itself and rarely involved the impact of other competitive products. Due to the existence of market competition, ignoring the existence of competitors in sales forecasting will result in decreased forecast accuracy, shortages, or warehouse explosions. Therefore, the influence of competitors must be considered in sales forecasting.

Many scholars have carried out fruitful research in three aspects: sales forecasting, consumer online reviews, and competition analysis, and they have also achieved rich research results. But no research can fully utilize the advantages of the three aspects of research. Therefore, in the study of sales forecast in this article, we consider the influence of competitors on product sales, product prices, the emotional values of the product from massive consumer review data, and product-network popularity to construct the regression models to predict comment volume and sales.

The remainder of this paper is organized as follows. The reviews of competitor identification research, sales forecasting research, and consumer review research are conducted in Section 2. Data processing and statistical analysis are presented in Section 3, including using web crawlers to grab consumer review data from different platforms, deduplicate the collected data, and count the number of product reviews and sales, and calculating the emotional value. In Section 4, identifying the main competitors of the product based on product review is proposed. We use statistical data to establish the regression model to forecast the sales and conduct the neural-network models in Section 5. Through comparing the obtained models, regression model is used with the best-fitting effect for further verification analysis. In Section 6, we obtain the regression model for sales forecast and use experiments to prove the effectiveness of the model. Finally, conclusions and future research directions are discussed in Section 7.

## 2. Literature Review

### 2.1. Review of Competitor Identification Research.
Identifying competitors of an enterprise mainly requires their analysis and assessment using the obtained enterprise information to locate their enterprise positions. Relevant identification methods can thus be identified from the different perspectives, including resource- and internet-based methods.

Within the same market, different companies can provide similar products. If they compete, they will likely have relatively similar products, similar technologies, and similar market strategies. Therefore, competitors can be identified according to a series of indicators related to these factors [14–16]. Of course, the traditional method of investigating competitors requires the use of a questionnaire, because the product is ultimately aimed at consumers who are most strongly aware of the competition. Using questionnaires can obtain direct data from the front-line market consumers. This method is the most direct, but it takes a long time and consumes many resources. Furthermore, information feedback may be incomplete, causing data errors that lead to faulty conclusions [17].

Ruan et al. [18] proposed a new unsupervised method based on a multistrategy learning algorithm for identifying competitors in a prospectus. In their enterprise competition analysis, the positioning of the competitors was crucial. It was found that only when the competitors were clearly positioned could the company develop an effective marketing strategy. Pant and Sheng [19] found that competitor recognition prediction models based on online indicators were much better than those using offline data, such as standard industry classification codes and company market value. Zheng [20] used a quantitative statistical method of clustering-and-matching analysis to identify and classify dynamic competitors of target enterprises. Liu et al. [21] proposed a Dirichlet process model based on sparse biterms and a bipartite graph model with random walk algorithm to analyze asymmetric competition using user favorite data.

### 2.2. Review of Consumer Review Research.
With the facilitation of the internet, social and consumption activities have gradually shifted to online platforms, and social platforms and online-shopping platforms have become important entertainment and consumption channels. Relying on network user data, such as social activity data and user online comment data on online sales platforms, it can serve all social participants through information extraction and mining. For example, the extraction of data information through social platforms can help identify and discover social dynamics and social behavior patterns and help the manager feed back corresponding demands, prevent and adjust strategies in time, and determine the priorities of work. Ibrahim Arpaci et al. [22] sampled 665 volunteers from two state universities to conduct an online survey and then used principal component analysis, exploratory factor analysis, and maximum likelihood analysis to analyze the scale data. The research results show that there is a

significant correlation between individualism, cyber bullying, and internet addiction. By studying the psychological conditions of young social media users, it is helpful to design more effective communication strategies to prevent internet addiction and cyber bullying. Ibrahim Arpaci et al. [23] used evolutionary cluster analysis to study the public's understanding of topics related to the COVID-19 outbreak. Studies have shown that analyzing social media users' attention trends can help organizations make better decisions and prioritize tasks. For example, governments and institutions can timely make strategic preparations according to the analysis of users' word trend data on social media.

Consumer reviews are mainly presented in text form on both shopping and social media sites. The incorporation of consumer commentary information into predictive models requires the translation of textual data into specific variables, which, in turn, requires analysis of the characteristics of consumer reviews [24, 25]. Many scholars have conducted extensive research in the field of consumer commentary research and have achieved great results. Specifically, consumer review research has focused on empirical analysis and practical techniques. In terms of empirical analysis, this has mainly included the characteristics and values of consumer reviews [26], consumer-group characteristics, and product advantages and disadvantages [11]. In terms of technology, the research has focused on consumer review feature extraction [11], review sorting and display [27], review sentiment analysis [28], and more.

Consumer reviews typically are reflective of product characteristics. After a review area is examined, subsequent consumer-purchase decisions will be influenced by these reviews [29]. Thus, the characteristics of consumer reviews will always have an impact on product sales [30]. *Pan* et al. [31] used online review information and Yelp review data to find that negative reviews had a greater impact on consumer purchasing decisions than did positive ones. However, positive reviews had a bigger impact than did negative ones when the reviews were from the consumers' friends. These findings are similar to those of Yin et al. [26], who found that not only did the review content reflect whether or not the consumer's attitude toward the product was positive or negative, it also reflected the specific attitude of the consumer. Bollen and Mao [32] used OpinionFinder and Google-Profile of Mood States to further refine consumers' attitudes from their comments. The user text content of social media was analyzed, showing that their comments included not only positive and negative emotions, but also notes about stability, warnings, happiness, peaceful aspects, etc. Furthermore, specific emotional dimensions were known to have an impact on consumer decisions. With the deepening of this research, researchers have found that consumer reviews not only reflect the attitude of consumers, but they also predict the characteristics of sellers from review features. Zhang et al. [33] attempt to identify and extract product innovative ideas from online reviews. Luca and Zervas [34] analyzed Yelp reviews and combined the ratings of various shops to identify fraud and other behaviors. Sun et al. [35] have studied how online reviews affect the return decisions of two competing online sellers selling products of different quality. Li et al. [36] extracted six factors related to comments and two incentive factors for product sales from 28 studies focusing on online comments and sales to study how they affect product sales. Through online reviews, consumers' attributes about products and services can be obtained. Xu [37] studies the relationship among consumer online reviews behavior, satisfaction, and demand in the accommodation-sharing hotel industry. When consumers are at different sharing levels, the factors that affect customer satisfaction and demand are different.

It is known that consumer reviews reflect consumer attitudes toward products. This important indicator was quantified and modeled through textual sentiment calculations by Felbermayr and Nanopoulos [38]. Owing to the development of natural-language processing technologies, textual emotional calculation is relatively mature. However, owing to consumer online-shopping characteristics, the emotional analysis of product reviews has a variety of features. Thus, the emotional calculation that considers shopping characteristics has become an important study item. Thelwall et al. [28] proposed a new algorithm, SentiStrength, to examine informal internet text to help extract emotional intensity. Extant research on the emotional polarity of most texts has only considered positive and negative attributes. However, Bingkun et al. [39] proposed a fuzzy-computing model to identify the emotional temperature of Chinese emotional words. The emotional tendency of online reviews affects consumers' purchase intentions. Emotional prediction technology is usually used to assign digital scores to online reviews to measure the emotional tendency. Basiri et al. [40] used machine learning methods to detect the polarity of reviews and proposed a new score aggregation method based on Dempster–Shafer evidence theory. Xu et al. [41] studied the influence of online consumption evaluation on commodity marketing decision through the method of data mining. They established BP network model to identify "useful consumer reviews", used ARIMA model to realize the time-series prediction of product reputation and then analyzed the radiation effect of the comments. Lin et al. [42] identified and classified innovative users based on online comments in automotive forums. On the basis of summarizing the six typical characteristics of innovative users, three interactive innovation methods are proposed, and then a pyramid model is constructed to show the distribution of innovative users.

*2.3. Review of Sales Forecasting Research.* The application of data mining method can be used as an important prediction technology. Güven and Şimşir [43] had established an artificial neural-network (ANN) and support vector machine (SVM) model that considered color parameters for the retail apparel industry to forecast demand. Tony et al. [44] used three machine learning models (linear regression, decision tree (DT), and random forest (RF)) for sales forecasting. From extant literature, we find that consumer review data play a huge role in sales forecasting [45, 46]. A classic study in this area was that of Bollen and Mao [32], who used the emotional changes of users on Twitter to predict the Dow

Jones Industrial Index. It was found that there is a strong relationship between the two indicators, and the prediction accuracy was much better than the traditional method. Some traditional forecasting models [47] can be combined with consumer reviews. The repeated purchasing index in the Bass diffusion model can [48], for example, be estimated using the emotional tendency of consumer reviews, and historic sales data can be used to fit the model to improve the accuracy of forecasting [49]. Dellarocas et al. [50] added online product review metrics to the benchmark model when using the Bass diffusion model [48] to predict movie box-office returns. Zhang et al. [51] use user-generated Internet information and macroeconomic data to improve the Bass model to achieve more accurate demand forecasting. Due to the lack of historical sales data, it is very difficult to predict the sales of short life cycle products. On the basis of the improved evidence theory, Tang and Dong [52] proposed multiple evidence dynamic weighted combination forecast method. Therefore, from the work of these scholars, we find that the content of consumer reviews can help us make more accurate estimations of prediction model parameters, thus improving prediction accuracy [53].

## 3. Data Processing and Statistical Analysis

Innovative products (viz. mobile phones) have been closely integrated with our daily lives and are now core products of our digital age. Thus, this type of product has a broad consumer market. The characteristics of mobile phones are evident and ubiquitous, and consumers have a variety of mobile phone needs. These needs are conducive to the division of the related consumer markets. Finally, owing to the rise of e-commerce, social media, and other internet resources, the online display of goods, transactions, and reviews has provided a wholesome source of data for acquisition and analysis.

The analysis and mining of mobile phone product review information first require a reliable source. Currently, user commentary is mainly concentrated on social media sites, shopping websites, and product communities. Because this article needs to mine competitive information, and each user's comment may contain competitive information, we leverage the independent database of Weibo, JD, Tmall, and Suning as our source. Because this source was opened on April 19, 2017, the data collected in this article spans April 19, 2017, to May 5, 2018.

*3.1. Text Deduplication.* We found that there are many similar reviews and even duplicates across social media. Unfortunately, the differences are only reflected by text characters. This limitation causes errors when determining the similarity of comments and competitors. Thus, duplicate information must be algorithmically removed.

There are two main algorithms for text deduplication [54]: consistency and similarity testing. Consistency checking uses the MD5 message-digest algorithm to generate a digital certificate to verify the consistency of files or text strings. MD5 (i.e., the Information Digest Algorithm 5th

Ed.) is a hash function widely used in computer security to provide information integrity protection and to encrypt and securely transmit documents. The algorithm mainly converts text information into a short information digest using a specific hash method combining compression, encryption, and hashing. When using the MD5 algorithm to deduplicate, the text content is first encrypted with the algorithm. Then, the MD5 value is compared [55]. In short, for short commentary texts, the MD5 steps are too complicated, causing the time complexity to increase. Another widely used method is the similarity test, which is used to calculate similarity among texts. It uses similar cosine-angle and Euclidean distance calculations [56].

The calculation methods of Euclidean distance and cosine-angle are simple and require a short time. However, the threshold control and the criterion for judging the degree of text repetition are neither intuitive nor accurate, which can easily cause errors.

The deduplication algorithm used in this paper is the simhash algorithm [57]. After the text is converted to a vector, the original vector with more dimensions is mapped to a feature vector with fewer dimensions using a hash algorithm. Then, the Hamming distance [58] of the two vectors is calculated to determine the similarity of the pieces of text. The Hamming distance reflects the number of different characters in the same position of two equal-length strings. Thus, if one changes one string to another, the number of characters will require updating. For example, the Hamming distance between 1100101 and 1001101 is 2. Using this method, the similarity can be obtained by calculating the Hamming distance of multiple texts.

The simhash algorithm is a locally sensitive hash function that is divided into five steps: word segmentation, hashing, weighting, merging, and dimensional reduction. The algorithm has the characteristics of low time complexity, simple steps, and high accuracy. In this paper, a total of 489,058 units of data were obtained, and the simhash algorithm was applied to remove 32,382 similar and duplicate data. Owing to the openness of the network environment and the language habits of different consumers, emoticons appear in the text, and different consumers have different internet access methods, which can cause garbled information in the reviews. If these problems were not processed, emotional value calculation errors would be created. Therefore, it is necessary to clean up the garbled characters and convert emojis after deduplication.

*3.2. Emotional Calculation.* Another important task related to consumer comment analysis includes the calculation of the emotional value of a review. The commentary information expresses the emotional orientation of the consumer, and this tendency must be translated into a specific value in the model. In the emotional analysis process of this paper, the first step is reading the consumer review data and using punctuation to segment the comments. The second step is processing the word segmentation, removing useless stop words, and retaining the emotional degree negative words. The third step finds the emotional words and records their

positions and their positivity and negativity. The fourth step finds degree words in front of emotional words. If there is a degree word, it sets the weight for the degree word and multiplies it by the sentiment value. Simultaneously, it finds the negative word. If the number of negative words is odd, the emotion value is multiplied by $-1$. Otherwise, it is multiplied by 1. Finally, it calculates the emotional value of all clauses in a comment. It then calculates the emotional value of the entire review text.

During the process of sentiment analysis, to ensure the accuracy of emotional value calculation, it is necessary to accurately segment the text. Considering the computational complexity and the maturity of the algorithm, this paper uses the Jieba word segmentation algorithm. The emotional value of each review is then calculated for all review data obtained by the above method, and the daily review emotion value of each product is calculated to form time-series data.

## 4. Competition Analysis and Competitor Identification

This paper uses a method of identifying competitors from consumer comment data. Extant methods judge the relationship between enterprises based on the frequency of occurrences on the same page on the same network. This method collects a large number of related web page documents in advance. It then counts the number of times that different enterprises cooccur on the same page. If different enterprises cooccur, it is more likely that they are in the same market. The consumer will thus encounter these two companies simultaneously. Thus, the two companies have a competitive relationship.

The cooccurrence matrix used in this paper is similar to the above method and is used to calculate the frequency at which different products cooccur in the consumer review text. It is generally believed that if two products appear in the same review at the same time, then consumers tend to compare the two products. This indicates that the two products have a relatively close competitive relationship. This paper uses consumer comment data as the text to be analyzed, assuming $c(x, y)$ is the number of times that enterprises $x$ and $y$ cooccur. $c(x)$ is the number of occurrences of enterprise $x$, and $p(y \mid x) = (c(x, y))/(c(x))$ indicates the probability that enterprise $y$ is the competitor of enterprise $x$. Then, the statistics of different $c(x, y)$ can provide the probability of different enterprises being competitors. Thus, the greater the probability is, the more likely it is a competitor of enterprise $x$.

In a product review, the more frequently a user mentions a brand or product, the closer the relationship is and the more competitive the two products are. Therefore, this article identifies and analyzes competitors by counting the number of times competitive brands and products appear in reviews.

The competitors of Xiaomi 6 are identified by using the competition-identification method mentioned earlier. Owing to the large number of mobile phone product-library entries, for the convenience of calculation, the brands of competitors are identified first. We establish a cooccurrence matrix between brands; we then use NetDraw software to visualize the data [59]. We finally obtain a cooccurrence network of different brands. It can be found from Figure 1 that Xiaomi's main competitors are Apple, Samsung, Huawei, and Honor. Additionally, owing to the competitive relationship between products within a brand, a brand can also be competitive to itself.

After positioning the competing brands, it becomes necessary to locate the competing products of Xiaomi 6 under different brands. In this analysis, the branded products are used to construct a cooccurrence matrix with Xiaomi 6, and NetDraw is used for visualization. As shown in Figure 2, Xiaomi 5x has a competitive relationship with Xiaomi 6. The analysis methods of other brands are the same. Finally, the main competitive products of Xiaomi 6 include iPhone 6s Plus, iPhone 7, iPhone 7 Plus, Honor 8, Honor 9, Samsung S7, Samsung S8, and Xiaomi 5X.

## 5. Predictive Model of Consumer Reviews

From our research, we found that many factors influence the sales of enterprise products. In the same market, corporations compete for the market, and the growth of competitor sales is likely accompanied by a decline in sales. Dellarocas et al. [50] found that daily reviews could be used to represent the daily sales of products, because there is a stable proportional relationship between daily review volume and sales volume. Thus the daily review volume of the product used in the model can replace product sales. In addition to market competition, owing to the display of consumer review information, comments having higher emotional value have a higher positive impact on subsequent consumers, and subsequent consumers' purchase intentions will be enhanced and vice versa. Therefore, product sales are affected by comments. With product advertisements and consumer word-of-mouth, product information will spread rapidly on the internet. The wider the spread, the more people will discuss it, the more attention it will receive, and the more likely a purchase will be made. Thus, network popularity will affect product sales. Finally, price is the most important factor and has always affected product sales. Thus, it is included in the forecast model of this paper.

### 5.1. Model Hypothesis

*5.1.1. Impact of Competitive Product Reviews on the Volume of the Product Reviews.* In the same consumer market, because the number of consumers is certain for a given period of time, product-sales volumes are closely related to those of competitive products. Under the premise that the market scale remains stable, the growth of sales of competitive products will inevitably accompany a decline in sales of its own products, and vice versa. Because there is a relatively stable relationship between the number of reviews and sales volume, there is a correlation between product reviews.

We collected all product comments on Xiaomi 6 and those of its competitors from Tmall, JD, and Suning (April
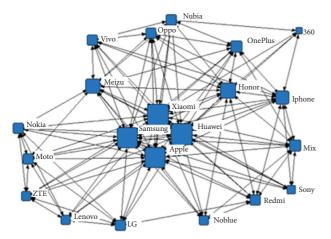
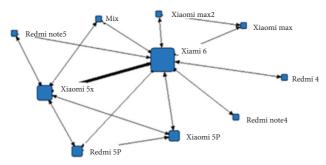Figure 1: Visualization of the main competitors of Xiaomi 6.



Figure 2: Visualization of the competitive products of Xiaomi 6 inside Xiaomi.

19, 2017, to May 5, 2018) for a total of 432,231 data items, including separate daily reviews for different products.

The interval in Table 1 refers to the interval between the dataset associated with the dependent variable and the one associated with the independent variable. The independent dataset is older than the dependent variable dataset, and the data in the 5th–12th columns are the regression coefficients of the corresponding products at different intervals. From this table, it can be found that when the interval is about 20 days, the number of competitive product reviews is most strongly correlated with the number of reviews for this product, and the multiple $R$, $R^2$, and adjusted $R^2$ values are nearly maximized. When the interval is 21 days, the three values are at their maximum. Taking the data from this day as an example, the number of reviews of iPhone6sP, iPhone7, iPhone7P, Samsung S7, and Samsung S8 had a positive impact on the number of reviews of Xiaomi 6, indicating that users of these products had a lower degree of overlap with the users of Xiaomi 6. Owing to the release of new products, the entire product market will promote change, thus driving market expansion. As the main competitor of Xiaomi in the domestic market, the Honor series has a high degree of overlap with the target users of Xiaomi. Thus, the number of product reviews for the Honor series is negatively correlated with the number of reviews of Xiaomi 6. The growth of Glory mobile phone sales will thus lead to a decrease in sales of Xiaomi mobile phones.

*5.1.2. Correlation between the Daily Review Volume and the Comment Emotion Value.* The emotional data of the product reviews represent consumer subjective attitudes toward the product. Thus, subsequent consumers browse existing product reviews to form a subjective attitude toward the product and decide whether or not to make a purchase. Thus, there is a correlation between the daily review sentiment value and the product daily review volume. From these statistics, it is found that the adjusted $R^2$ value in the regression analysis between the product daily review volume and the daily emotion is 0.98, and the standard error is 8.86. Therefore, it can be concluded that there is a strong correlation between the daily review volume of the product and the daily emotion. Because the sum of the daily emotional scores is the sum of the emotional scores of all reviews and the number of daily reviews is greatly affected, it is necessary to analyze the relationship between the daily average emotional score and the daily review amount.

We selected data having an interval of 0–7 days. From Table 2, we can see that the multiple $R$, $R^2$, and the adjusted $R^2$ reach their maximum values when the interval is 2 days. Simultaneously, when the time interval is 2 days, the $P$ value is 0.00087, which is much smaller than 0.001. Thus, the model is significant at $\alpha = 0.001$. That is, the confidence of the model is 99.9%, and the confidence interval is [362.86, 1,252.50]. There is thus a significant correlation between the amount of product reviews and the average daily emotional value of the product.

*5.1.3. Correlation between the Daily Review Volume and the Price.* The price has a strong influence on product sales. The increase in product price will lead to a decrease in sales volume. The most influential factor in the product sales process is the product price.

Figure 3 shows the change of the $R^2$ value between the quantity of product reviews and the price. In the figure, the abscissa is the number of intervals between the price and the dataset used by the number of reviews. The ordinate is the adjusted $R^2$ value between the two datasets. As can be seen from the figure, the number of product reviews is most relevant to prices before 10 to 15 days. This is also in line with the actual situation. Consumers make purchases after observing that the price is at a lower level, and the time interval between purchase and review is about 7–15 days.

In Table 3, we find that all of the regression coefficients of price are less than 0, indicating that the price drop will bring about an increase of product sales, which is consistent with the positioning of Xiaomi 6 being cost-effective. Thus, the impact of promotional activities on product sales is positive. Simultaneously, when the interval is 11 days, the values of multiple $R$, $R^2$, and adjusted $R^2$ reach their maximum levels, and the $P$ value is 0.00003, which is much smaller than 0.0001. Thus, the model is significant at $\alpha = 0.0001$, and the confidence of the model reaches 99.99%, which shows a significant correlation between product prices and the number of product reviews.

TABLE 1: Regression results of Xiaomi 6 review volume with major competitor's product review volume.

| Interval/d | Multiple R | $R^2$ | Adjusted $R^2$ | iPhone6sP | iPhone7P | iPhone7 | Honor 8 | Honor 9 | S7 | S8 | Xiaomi 5X |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 14 | 0.79 | 0.63 | 0.52 | 0.17 | 0.00 | 0.32 | 1.89 | −0.15 | −3.33 | 0.04 | 0.18 |
| 15 | 0.82 | 0.68 | 0.58 | −0.53 | 0.82 | 1.25 | 2.35 | 0.05 | −2.61 | −0.01 | −0.24 |
| 16 | 0.79 | 0.62 | 0.50 | −0.14 | 1.69 | −0.95 | 1.02 | 0.33 | −5.82 | −0.06 | −0.89 |
| 17 | 0.77 | 0.59 | 0.45 | 0.06 | 1.73 | 0.28 | −0.27 | 0.55 | −3.44 | −0.09 | −1.05 |
| 18 | 0.70 | 0.49 | 0.32 | 0.34 | 1.13 | 0.17 | −0.60 | 0.42 | 1.07 | −0.03 | −0.43 |
| 19 | 0.82 | 0.68 | 0.56 | 0.79 | 1.05 | 0.92 | −1.64 | 0.40 | −3.46 | −0.02 | −0.39 |
| 20 | 0.83 | 0.70 | 0.58 | 0.44 | 1.05 | 1.19 | −1.50 | −0.31 | 1.82 | 0.03 | 0.36 |
| 21 | 0.86 | 0.74 | 0.63 | 0.44 | 0.83 | 1.66 | −1.00 | −0.64 | 1.78 | 0.03 | 0.29 |
| 22 | 0.83 | 0.69 | 0.55 | 0.51 | 0.65 | 1.92 | 0.92 | −0.90 | −1.26 | −0.01 | −0.52 |
| 23 | 0.70 | 0.49 | 0.27 | −0.60 | 1.14 | 1.07 | −0.49 | −0.43 | 0.52 | 0.00 | 0.48 |
| 24 | 0.81 | 0.65 | 0.49 | −0.09 | 0.86 | 0.52 | −1.16 | −0.09 | −1.53 | −0.03 | 0.03 |
| 25 | 0.70 | 0.50 | 0.25 | 0.41 | 0.04 | 0.19 | −0.50 | −0.11 | 0.75 | −0.03 | −0.03 |

TABLE 2: Regression results of product review emotion value with product review volume.

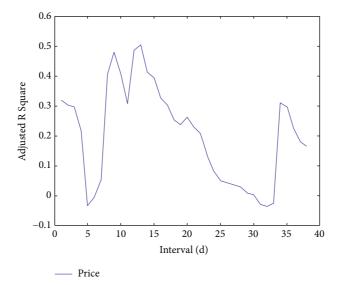| Interval/d | Multiple R | $R^2$ | Adjusted $R^2$ | Coefficients | $t$ Stat | $P$ value | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.29 | 0.08 | 0.05 | 400.73 | 1.69 | 0.10 | −82.74 | 884.20 |
| 1 | 0.48 | 0.23 | 0.20 | 663.73 | 2.99 | 0.01 | 210.14 | 1117.31 |
| 2 | 0.57 | 0.32 | 0.30 | 807.68 | 3.71 | 0.00 | 362.86 | 1252.50 |
| 3 | 0.45 | 0.21 | 0.18 | 649.25 | 2.70 | 0.01 | 157.29 | 1141.21 |
| 4 | 0.33 | 0.11 | 0.07 | 454.75 | 1.79 | 0.09 | −67.69 | 977.20 |
| 5 | 0.50 | 0.25 | 0.22 | 685.62 | 2.97 | 0.01 | 210.62 | 1160.62 |
| 6 | 0.37 | 0.14 | 0.10 | 491.09 | 2.01 | 0.06 | −13.35 | 995.53 |
| 7 | 0.46 | 0.21 | 0.17 | 366.19 | 2.50 | 0.02 | 64.41 | 667.96 |



FIGURE 3: The adjusted $R^2$ of price with product review volume.

TABLE 3: Regression results of price with product review volume.

| Interval/d | Multiple R | $R^2$ | Adjusted $R^2$ | Coefficients | $t$ Stat | $P$ value | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|---|---|
| 7 | 0.42 | 0.17 | 0.15 | −0.06 | −2.54 | 0.02 | −0.11 | −0.01 |
| 8 | 0.56 | 0.31 | 0.29 | −0.08 | −3.69 | 0.00 | −0.13 | −0.04 |
| 9 | 0.59 | 0.35 | 0.32 | −0.09 | −3.92 | 0.00 | −0.13 | −0.04 |
| +10 | 0.64 | 0.41 | 0.39 | −0.09 | −4.39 | 0.00 | −0.14 | −0.05 |
| 11 | 0.69 | 0.48 | 0.46 | −0.10 | −5.00 | 0.00 | −0.14 | −0.06 |
| 12 | 0.68 | 0.47 | 0.45 | −0.10 | −4.78 | 0.00 | −0.14 | −0.05 |
| 13 | 0.58 | 0.34 | 0.31 | −0.08 | −3.59 | 0.00 | −0.12 | −0.03 |
| 14 | 0.62 | 0.38 | 0.36 | −0.05 | −3.85 | 0.00 | −0.07 | −0.02 |
| 15 | 0.50 | 0.25 | 0.22 | −0.03 | −2.76 | 0.01 | −0.06 | −0.01 |

### 5.1.4. Correlation between the Daily Review Volume and the Search Index.
The popularity of the product-network mainly reflects the popularity and attention of the products in the consumer-group. The higher the consumer's attention, the higher the likelihood that the product will be purchased and the greater the number of product reviews. We use the Baidu index as an indicator to reflect the popularity of product-network. This index is a data analysis platform based on network-behavior big data. It is a primary statistical analysis platform and is an important basis for many companies' marketing decisions.

The search index reflects the popularity of the product in the network. A high index indicates a large number of consumers who know the product. Thus, product sales will increase, and the regression coefficient of the search index for the number of product reviews will be greater than 0. In Table 4, when the values of multiple $R$, $R^2$, and adjusted $R^2$ reach their maximum, the $P$ value is 0.00008, which is less than 0.0001. Thus, the model is significant at $\alpha = 0.0001$, and the confidence of the model reaches 99.99%, which shows a significant correlation between the search index and the number of product reviews.

### 5.2. Establishment of Product Reviews Forecasting Model.
From the above analysis, we find that the number of daily product reviews has a strong correlation with those of competitor products, product price, daily emotional value of the product, and product-network popularity. Supposing the daily review volume of a product is $R$, the daily review volume of the competitor product is $x_1$, $x_2$, .... The product daily emotional value is $s$, the product price is $p$, and the product-network popularity is $w$. Thus, the regression model for the number of daily reviews of the product is $R = A \times X_{t+\alpha_1} + b \times s_{t+\alpha_2} + c \times p_{t+\alpha_3} + d \times w_{t+\alpha_4}$, where $X_{t+\alpha_1}$ represents the review quantity vector of competitor products, and A is the coefficient corresponding to the number of competitors' different reviews, $s_{t+\alpha_2}$ is the daily emotional value of the product, $p_{t+\alpha_3}$ is the product price, and $w_{t+\alpha_4}$ is the product-network popularity. Because we find that different intervals will lead to different correlation coefficients, $A$, $b$, $c$, $d$, we must use the corresponding time dataset in the calculation process, and $\alpha$ is the corresponding interval.

Table 5 displays the main parameters of the regression model. It can be seen that the multiple $R$, $R^2$, and adjusted $R^2$ values are all greater than 0.9, indicating that the model-fitting effect is good enough. The significance F value is $9.38E - 09$, indicating that the confidence of the model is above 99.99%. $\alpha_1$, $\alpha_2$, $\alpha_3$, and $\alpha_4$ represent the interval between competitor product reviews, sentiment index, price, and network popularity and the number of reviews for this product. The regression coefficient results are similar to the previous analysis. As a direct competitor, the Honor 8 and Honor 9 reviews and the price increase will bring about a decline in the number of reviews for Xiaomi 6. To analyze the impact of different products and different indicators on the number of reviews on this product, this paper normalizes the data for further analysis.

Table 6 contains the regression coefficient results obtained by normalizing the data. From the information in the table, we can find that the impact of Honor 8 in several different competitor products is greater, because Honor and Xiaomi apply nearly the same ideas and models via internet marketing. Their target group includes users who focus on cost performance, and they mainly engage the online market. At the same time, compared with Honor 9, the prices of Honor 8 and Xiaomi 6 are closer. Thus, the impact of Honor 8 on Xiaomi 6 is even greater. Thus, target users pay attention to cost performance, and the emotional tendency of product reviews is particularly important for product sales. Thus, consumer reputation and product popularity are present on the network.

### 5.3. Neural-Network Model.
During the analysis process, the regression model is linear, indicating that the impact of different variables is mainly linear. However, according to extant research, the influence between different variables is more nonlinear [60, 61], such as the problem of collinearity between prices and competitors' reviews. However, different variables have different kinds of interference factors, owing to their different consumer characteristics. Thus, the prediction accuracy of linear models is not as good as those of nonlinear models. In this paper, TensorFlow is used to establish a neural-network model to verify this conjecture.

This paper uses the back-propagation (BP) neural-network model, which continuously corrects network weights and thresholds during data training to reduce the error function along the negative gradient. This neural network is widely and commonly used to perform time-series prediction. The BP neural-network generally consists of an input layer, a hidden layer, and an output layer. The hidden layer may have one or more layers. Through the error function, $E = (\sum_i (t_i - e_i)^2 / 2)$ ($t_i$ is the calculation result; $e_i$ is the test data), the weight is continuously adjusted to minimize $E$.

According to the above principle, this paper establishes a neural-network model for sales forecasting. The number of neurons in the input layer is 11, the hidden layer is two, and the number of output neurons is one. The rectified linear-unit function is selected as the excitation function of the hidden layer.

The predicted mean-square error calculated by the neural-network model is 76.096644, which is much larger than the mean-square error of 73.60849 predicted by the regression model, which indicates that the regression model is more suitable for the data.

## 6. Sales Forecast Model and Experimental Analysis

The above process was analyzed experimentally using automotive supplies, and we selected a total of 73,040 automotive gasoline additive products from August 5, 2018, to September 4, 2018, corresponding to a total of 287,490 data reviews.

From the analysis of the gasoline additive market, Chevron products were selected as the analysis object, and

TABLE 4: Regression results of search index with product review volume.

| Interval/d | Multiple R | $R^2$ | Adjusted R2 | Coefficients | t Stat | P value | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|---|---|
| 7 | 0.41 | 0.17 | 0.14 | 0.01 | 2.50 | 0.02 | 0.00 | 0.02 |
| 8 | 0.53 | 0.28 | 0.26 | 0.01 | 3.44 | 0.00 | 0.01 | 0.02 |
| 9 | 0.53 | 0.28 | 0.26 | 0.02 | 3.37 | 0.00 | 0.01 | 0.03 |
| 10 | 0.66 | 0.43 | 0.41 | 0.02 | 4.63 | 0.00 | 0.01 | 0.03 |
| 11 | 0.63 | 0.40 | 0.37 | 0.02 | 4.22 | 0.00 | 0.01 | 0.04 |
| 12 | 0.64 | 0.41 | 0.39 | 0.02 | 4.28 | 0.00 | 0.01 | 0.04 |
| 13 | 0.61 | 0.37 | 0.35 | 0.02 | 3.85 | 0.00 | 0.01 | 0.04 |
| 14 | 0.62 | 0.38 | 0.36 | 0.01 | 3.85 | 0.00 | 0.01 | 0.02 |
| 15 | 0.48 | 0.23 | 0.19 | 0.01 | 2.59 | 0.02 | 0.00 | 0.02 |
| 16 | 0.33 | 0.11 | 0.07 | 0.01 | 1.65 | 0.11 | 0.00 | 0.02 |

TABLE 5: Regression result of the product reviews forecasting model.

| Parameter | Value | Variable | Regression coefficients |
|---|---|---|---|
| Multiple R | 0.96 | iPhone6sP | 0.632 |
| R square | 0.92 | iPhone7P | 0.435 |
| Adjusted $R$ square | 0.90 | iPhone7 | 1.375 |
| F | 49.95 | Honor 8 | −0.520 |
| Significance F | 0.00 | Honor 9 | −0.534 |
| Standard error | 33.89 | Samsung S7 | 0.082 |
| $\alpha_1$ | 21 | Samsung S8 | 0.032 |
| $\alpha_2$ | 9 | Xiaomi 5X | 0.017 |
| $\alpha_3$ | 11 | Emotion | 0.050 |
| $\alpha_4$ | 42 | Price | −0.004 |
| Intercept | 0.617 | Search index | 0.008 |

TABLE 6: Normalized regression coefficient of the product reviews forecasting model.

| Variable | Regression coefficients |
|---|---|
| iPhone6sP | 0.04 |
| iPhone7P | 0.08 |
| iPhone7 | 0.01 |
| Honor 8 | −0.18 |
| Honor 9 | −0.03 |
| Samsung S7 | 0.02 |
| Samsung S8 | 0.00 |
| Xiaomi 5X | 0.09 |
| Emotion | 0.92 |
| Price | −0.03 |
| Search index | 0.15 |

the main brands in the market were determined to be BASF, WHIZ, Hummer, Red Line, Goodway, Chief, BAFU, Prince Edward, WD, and 3M. We counted the number of times each brand product appeared in the Chevron product review and located the competitive products. The main competitive brands were found to be Red Line, Hummer, BAFU, 3M, and BASF.

The data used in this section included the daily order volume of Chevron products and their competitors from August 5, 2018, to September 4, 2018, as well as the price changes of Chevron. According to the model established in this paper, regression analysis was performed on the above experimental data to establish a regression model. It can be

seen that the model's adjusted $R^2$ value is 0.933154662, indicating that the model fits very well. Product sales will thus be affected by its price, the sales of competitors' products, the popularity of its own products, and the emotional sentiment of its product reviews. For each variable's impact on product sales, we can analyze the adjusted $R^2$ value of the regression result of each variable with the sales volume.

### 6.1. Model Analysis

*6.1.1. Impact of the Sales Volume of Competitive Products on the Sales.* Figure 4 reflects the relationship between the sales of competitor products and those of enterprise products. In the figure, the abscissa is the number of intervals between the competitor's sales and the dataset used by the company's product sales, and the ordinate is the adjusted $R^2$ between the two datasets. As can be seen from the figure, there is a strong correlation between competitor and product sales, indicating that, in the same market, enterprise products and competitors' products share the market. Because consumers face multiple choices, the product value, sales strategy, and other factors will affect the sales of the company, but the sales of competitors will also affect them. As can be seen from the figure, because the competition occurred during the same period, the sales volume of the competitors had the greatest impact on the sales of the products during the period of product sales. Additionally, from the previous analysis, we know that competitor reviews also affect product sales, and the time of impact of this effect was about 16 days. Thus, in the figure, we can find that, at intervals of 15 days, there is a maximum effect on product sales.

The data in columns 5–11 of Table 7 is the regression coefficient of the product at different intervals. When the interval is 16 days, the values of the multiple $R$, $R^2$, and adjusted $R^2$ reach the maximum value at the same time. When the competitor product sales volume and the product sales interval are 16 days, the correlation is the strongest. Simultaneously, it can be known from the data in the table that 3M 7481848 had the greatest impact on the sales volume of this product. For each additional unit of 3M 7481848 sales, the sales volume of this product increased by 66 units. This is because the price difference between 3M products and this product is large, and the target customers are different. The other product of 3M 717303 has a similar market position
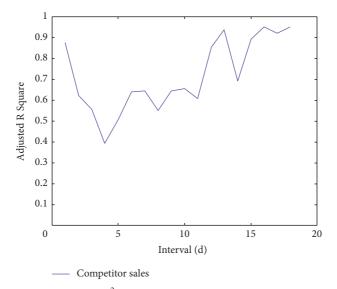
Figure 4: Adjusted $R^2$ of competitor product sales with own sales.

Table 7: Regression results of competitor's product sales with this product's sales.

| Interval/d | Multiple R | $R^2$ | Adjusted $R^2$ | Hummer | Redline 682371 | Redline 5841309 | BAFU | 3M 717303 | 3M 906035 | 3M 7481848 |
|---|---|---|---|---|---|---|---|---|---|---|
| 8 | 0.55 | 0.30 | −0.14 | −5.84 | −12.02 | 3.57 | 1.56 | 10.46 | 0.56 | −1.90 |
| 9 | 0.64 | 0.42 | 0.01 | −6.93 | −7.56 | 1.42 | 4.44 | 6.83 | 1.31 | −34.70 |
| 10 | 0.66 | 0.43 | −0.01 | −9.19 | 2.03 | −0.06 | 5.07 | 22.40 | 2.43 | −34.87 |
| 11 | 0.61 | 0.37 | −0.19 | −2.42 | 4.26 | −2.16 | 9.94 | 3.21 | 1.04 | −28.92 |
| 12 | 0.85 | 0.72 | 0.45 | 3.58 | 14.70 | −7.70 | 9.14 | −23.35 | 0.21 | 23.95 |
| 13 | 0.94 | 0.88 | 0.74 | 14.81 | −0.85 | −6.63 | 16.21 | −23.12 | −3.50 | −4.30 |
| 14 | 0.69 | 0.48 | −0.26 | 23.44 | 7.18 | −3.04 | −10.54 | −48.73 | 2.02 | 22.99 |
| 15 | 0.89 | 0.79 | 0.43 | 22.51 | −9.71 | 7.57 | −5.86 | −48.34 | 0.70 | 22.25 |
| 16 | 0.95 | 0.90 | 0.68 | 3.75 | −22.72 | 10.96 | −6.75 | −34.55 | 3.87 | 66.67 |
| 17 | 0.92 | 0.85 | 0.31 | −0.85 | 0.91 | 9.24 | −23.73 | 8.81 | −3.30 | 44.48 |
| 18 | 0.95 | 0.91 | 0.26 | −2.87 | 0.39 | 7.41 | −27.55 | 64.93 | −7.48 | −3.97 |

and is in a competitive relationship because of the low price of this product.

### 6.1.2. Impact of Product Prices on the Sales.

Figure 5 reflects the relationship between product prices and sales. In the figure, the abscissa is the number of intervals between the product price and the dataset used by the company's product sales, and the ordinate is the adjusted $R^2$ between the two data. In the market, one of the most influential and most timely factors for product sales is the product price. When the company adjusts the price, the consumer perceives the price change and will make the purchase decision in any case. The consumer's purchase decision corresponds to the price of the same period. Product promotion usually occurs in a short time, and the consumer will make a quick decision, which is reflected in the figure: the impact of the price has the biggest impact on the product. Then, over time, the impact of price on product sales continues to diminish and eventually remains stable.

It can be seen from Table 8 that as the interval increases, the correlation between price and sales volume gradually decreases. It shows that the impact of price on product sales is time-limited. People will decide whether or not to purchase according to the price of the product at the right time. Therefore, the price elasticity of this type of product is very strong, and promotion is very necessary. When the interval is 0 days, the product price has the greatest impact on product sales. The regression coefficient is −14.79, and the product sales are reduced by 14 units for each additional unit price. Simultaneously, the $P$ value is 0.003, which is less than 0.01. Thus, the model is significant at $\alpha = 0.01$. That is, the confidence of the model reaches 99%, indicating that there is a significant correlation between price and product sales.

### 6.1.3. Impact of Product-Network Popularity on the Sales.

Figure 6 reflects the relationship between the information index (news index) and the sales of the company's products. In the figure, the abscissa is the number of intervals between the information index and the dataset used by the company's product sales, and the ordinate is the adjusted $R^2$ between the two datasets. The information index reflects the degree of consumer response to related products on the network. The higher the information index, the more consumers praise
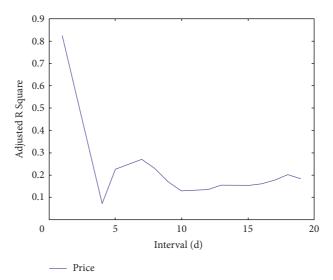
FIGURE 5: Adjusted $R^2$ of price with product sales.

TABLE 8: Regression results of price with product sales.

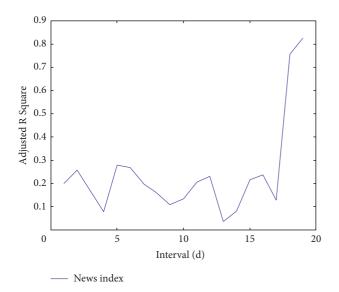| Interval/d | Multiple R | $R^2$ | Adjusted $R^2$ | Coefficients | $t$ Stat | $P$ value | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.82 | 0.68 | 0.67 | −14.79 | −7.14 | 0.00 | −19.07 | −10.52 |
| 1 | 0.58 | 0.34 | 0.31 | −10.34 | −3.41 | 0.00 | −16.61 | −4.07 |
| 2 | 0.32 | 0.11 | 0.06 | −5.75 | −1.61 | 0.12 | −13.16 | 1.67 |
| 3 | 0.07 | 0.01 | −0.04 | 1.26 | 0.33 | 0.75 | −6.73 | 9.24 |
| 4 | 0.23 | 0.05 | 0.00 | 3.79 | 1.04 | 0.31 | −3.78 | 11.37 |
| 5 | 0.25 | 0.06 | 0.01 | 4.12 | 1.12 | 0.28 | −3.56 | 11.81 |



FIGURE 6: Adjusted $R^2$ of News Index with product sales.

and share the behavior. From the previous analysis, it can be found that the information index takes about 24 days to have an effect. This product has a similar effect. The information index has less influence on the product sales during the same period. The spread of product information and familiarity with consumers takes about 24 days.

*6.1.4. Impact of Review Emotion Value on the Sales.* Figure 7 reflects the relationship between product review sentiment value and product sales. In the figure, the abscissa is the number of days between the product review sentiment value and the dataset used by the company's product sales, and the ordinate is the adjusted $R^2$ between the two datasets. Emotion reflects the evaluation of the product by the consumer and the degree of consumer acceptance of the product. The higher the emotional value, the better the consumer's evaluation of the product over a period of time. Thus, more consumers will buy the product. It can be seen from the figure that the emotional value has the greatest impact at intervals of about 15 days, which is related to the product review display algorithm. Observing the product review date, we find that most of the reviews that consumers can see are reviews from 15 days prior. Thus, the impact of emotional value on product sales will be maximized at intervals of about 15 days.

*6.2. Establishment of the Sales Forecasting Model.* From the analysis in the previous section, the regression model of product sales is $Y = B \times C_{t+\alpha_1} + b \times s_{t+\alpha_2} + c \times p_{t+\alpha_3} + d \times w_{t+\alpha_4}$, where $C_{t+\alpha_1}$ represents the sales quantity vector of competitor products, $s_{t+\alpha_2}$ is the daily emotional value of the product, $p_{t+\alpha_3}$ is the product price, $w_{t+\alpha_4}$ is the product-network popularity, and $A, b, c, d$ represent the different correlation coefficients. Because we found that different
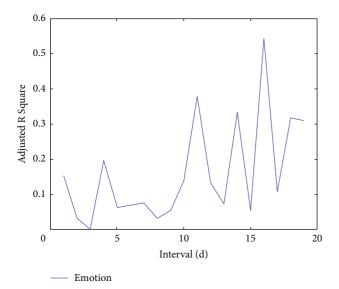
FIGURE 7: Adjusted $R^2$ of review emotion score with product sales.

TABLE 9: Regression results of the sales forecasting model.

| Parameter | Value | Variable | Regression coefficients |
|---|---|---|---|
| Multiple R | 0.98 | Hummer | 1.21 |
| R square | 0.97 | Redline 682371 | −3.30 |
| Adjusted $R$ square | 0.95 | Redline 5841309 | 2.50 |
| F | 40.37 | BAFU | −2.34 |
| Significance F | 0.00 | 3M 717303 | −6.11 |
| $\alpha_1$ | 15 | 3M 906035 | 0.68 |
| $\alpha_2$ | 0 | 3M 7481848 | 2.59 |
| $\alpha_3$ | 19 | Price | −12.36 |
| $\alpha_4$ | 16 | Search index | 0.62 |
| | | Emotion | 15.62 |

variables have different intervals during the analysis, then $C_{t+\alpha_1}$, $s_{t+\alpha_2}$, $p_{t+\alpha_3}$, $w_{t+\alpha_4}$ is the dataset corresponding to the time in the calculation process. $\alpha_1$, $\alpha_2$, $\alpha_3$, and $\alpha_4$ represent the time interval among competitor product sales, sentiment index, price, network popularity, and product reviews, respectively. Table 9 shows the main parameters of the regression model. It can be seen that multiple $R$, $R^2$, and adjusted $R^2$ are all greater than 0.9, indicating that the model-fitting effect is very good. The significance F value is $1.14E-08$, indicating that the confidence of the model has reached 99.99% or more. Similar to the previous analysis, the sales growth of Hummer and 3M 7481848, which are higher-end products, will drive the sales growth of this product, and the sales growth of similar products will lead to a decline in sales of this product.

The data are also normalized to analyze the impact of different indicators on the sales of this product. As can be seen from Table 10, the price has the greatest impact on the sales volume of this product. Therefore, it is very effective to reduce the price to increase the sales volume of the product. Additionally, because the target user group of the product is relatively simple, the influence of the network popularity on the product sales is low. Because the Redline at 5841309 is

TABLE 10: Regression coefficient after normalization.

| Variable | Coefficients |
|---|---|
| Hummer | 0.08 |
| Redline 682371 | −0.27 |
| Redline 5841309 | 0.54 |
| BAFU | −0.17 |
| 3M 717303 | −0.21 |
| 3M 906035 | 0.09 |
| 3M 7481848 | −0.22 |
| Price | −0.58 |
| Search index | 0.10 |
| Emotion | 0.10 |

positioned at a higher level than this product, the competition with this product is weak. The growth of Redline means increased market demand, which naturally leads to the growth of sales of this product.

## 7. Conclusion

Online sales platforms provide a convenient means. Consumers can obtain product-related attributes based on online reviews and word-of-mouth learning, so as to make purchasing decisions, while sellers can improve products and make more scientific sales and pricing decisions based on consumer reviews. As the scale of the internet continues to expand, every participant will become a content producer. User-generated content reflects consumers' most intuitive perceptions of products and can be used as an important reference for enterprises to make product decisions. It not only will provide ideas and directions for product design and improvement, but also will become an important means for enterprises to participate in market competition. As shown in the research of this paper, we can use consumer comments to identify competitors, and we can use their attitudes to evaluate competing products and sales to predict market changes. When companies predict product sales, they adjust the supply chain accordingly, determine the appropriate production scale and inventory scale, reduce unnecessary waste, and improve business efficiency.

In terms of competitor identification, the most direct market feedback and the symbiotic matrix method combined with NetDraw software were used to quickly identify competitors of enterprise product markets. Traditional methods require industry analysis, questionnaires, and other methods, which require a great deal of manpower and resources. Thus, it may not provide the most direct market feedback. Owing to the universality and dispersion of consumers who have experienced market competition, their feedback information cannot be collected effectively by these old methods. This paper used a web crawler to obtain all relevant consumer content from consumers' main feedback platforms. This is a wide-ranging practice at present, and the cost is low. Combined with other methods, relevant research can thus be quickly carried out.

The emotional content of consumer feedback is also important. Using natural-language-processing tools, we can

find subtle changes in consumer attitudes toward a product. Although these changes may be imperceptible to other consumers, they directly affect purchasing decisions. Thus, emotional analysis must be considered in product sales forecasting, which is inseparable from mathematical models. The existing forecasting models are mainly linear regression models, Bass diffusion models, etc. Sales volume is affected by many factors, and these influencing factors are actually nonlinear. Using the TensorFlow tool, we built a neural-network model and compared it with a linear regression model, finding that the nonlinear relationship between competitors' sales/comment volume, product price, network heat, and commentary sentiment was weak, but the effect of linear regression prediction model was better. Additionally, because it is impossible for consumers to distinguish imitators from innovators, and the potential market size is difficult to determine, various problems arise with the Bass diffusion model.

In addition to the above methods and models, the research in this paper provides a new perspective for user-generated content-related research and predictive analysis. In the prediction model, competitive factors are considered. With the application of data mining, predictive models of competition are no longer comprised of simple variables; they are real and effective data.

The dissemination of product information by enterprises is influenced by the comments of consumers. Mobile internet capabilities have amplified the influence of consumer-created media. Thus, the product-sales process can be affected by network popularity. With relevant data statistics, this paper drew upon Baidu index data, similar to Google Trends, which helped us understand the depth and breadth of a product's spread on the internet and was an important part of the research process.

The research in this paper focused on mobile phone and automotive products as examples. However, there are countless commodities in the market. The influence factors of different products are different, as are the impacts. Therefore, in the future research, different product types should be examined to find a common forecasting model. Furthermore, in the data mining process, owing to the complexity of internet information, figuring out how to eliminate useless information is a major concern, not to mention the elimination of errors generated by data noise.

Online sales make information transmission and dissemination faster. Social events will also be quickly and accurately reflected in the online sales, such as consumer enthusiasm mentioned in this article. The COVID-19 epidemic has changed the normal production and living order, as well as online sales activities. The survey results of the project "COVID-19 and e-commerce" (which was organized by the United Nations Conference on Trade and Development (UNCTAD)) show that the epidemic has accelerated the world's digitalization process and has a permanent impact on people's shopping behavior. More than half of the interviewees said that, under the epidemic, online consumption has increased and they rely more on the internet for news, health-related information, and entertainment [62]. Mukhisa Kituyi, Secretary-General of UNCTAD, said

that the epidemic has accelerated the transition to a digital world.

Under the background of COVID-19, information diffusion and dissemination have an important impact on online sales of enterprises. Based on the changes of COVID-19, how do online reviews and public opinion propagation affect online sales? How are online emergency supplies sold and matched in the context of COVID-19? Further research on these issues will help to expand the existing research and make the application of the model more extensive and more realistic.

## Data Availability

Most of the data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

## References

[1] P. C. Verhoef and L. M. Sloot, *Out-of-Stock: Reactions, Antecedents, Management Solutions, and a Future Perspective*, Springer, Berlin Heidelberg, Germany, 2010.

[2] D. F. Davis, J. T. Mentzer, T. M. Mccarthy, and S. L. Golicic, "The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices," *Journal of Forecasting*, vol. 25, pp. 303–324, 2010.

[3] C. Forman, A. Ghose, and B. Wiesenfeld, "Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets," *Information Systems Research*, vol. 19, no. 3, pp. 291–313, 2008.

[4] F. Caro and V. Martínez-de-Albéniz, "Product and price competition with satiation effects," *Informs Management Science*, vol. 58, no. 7, pp. 1357–1373, 2012.

[5] C.-F. Chien, Y.-J. Chen, and J.-T. Peng, "Manufacturing intelligence for semiconductor demand forecast based on technology diffusion and product life cycle," *International Journal of Production Economics*, vol. 128, no. 2, pp. 496–509, 2010.

[6] R. Qin and D. A. Nembhard, "Demand modeling of stochastic product diffusion over the life cycle," *International Journal of Production Economics*, vol. 137, no. 2, pp. 201–210, 2012.

[7] F. Chen, Z. Drezner, J. K. Ryan, and D. Simchi-Levi, "Quantifying the bullwhip effect in a simple supply chain: the impact of forecasting, lead times, and information," *Management Science*, vol. 46, pp. 436–443, 1999.

[8] S. Thomassey, "Sales forecasts in clothing industry: the key success factor of the supply chain management," *International Journal of Production Economics*, vol. 128, no. 2, pp. 470–483, 2010.

[9] S. C. Nettelhorst, L. A. Brannon, and W. Trey Hill, "Examining the impact of consumer feedback on internet product evaluation: comparing base-rate and case history information," *Computers in Human Behavior*, vol. 29, no. 4, pp. 1290–1294, 2013.

[10] C. Hildebrand, G. Häubl, A. Herrmann, and J. R. Landwehr, "When social media can be bad for you: community feedback stifles consumer creativity and reduces satisfaction with self-designed products," *Information Systems Research*, vol. 24, no. 1, pp. 14–29, 2013.

[11] C.-P. Wei, Y.-M. Chen, C.-S. Yang, and C. C. Yang, "Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews," *Information Systems and E-Business Management*, vol. 8, no. 2, pp. 149–167, 2010.

[12] M. J. Schneider and S. Gupta, "Forecasting sales of new and existing products using consumer reviews: a random projections approach," *International Journal of Forecasting*, vol. 32, no. 2, pp. 243–256, 2016.

[13] T. Huang, R. Fildes, and D. Soopramanien, "The value of competitive information in forecasting FMCG retail product sales and the variable selection problem," *European Journal of Operational Research*, vol. 237, no. 2, pp. 738–748, 2014.

[14] M. Bergen and M. A. Peteraf, "Competitor identification and competitor analysis: a broad-based managerial approach," *Managerial and Decision Economics*, vol. 23, no. 4-5, pp. 157–169, 2002.

[15] C. D. B. Montgomery, "Managerial identification of competitors," *Journal of Marketing*, vol. 63, pp. 67–83, 1999.

[16] M. A. Peteraf and M. E. Bergen, "Scanning dynamic competitive landscapes: a market-based and resource-based framework," *Strategic Management Journal*, vol. 24, no. 10, pp. 1027–1041, 2003.

[17] R. M. Morgan and S. Hunt, "Relationship-based competitive advantage," *Journal of Business Research*, vol. 46, no. 3, pp. 281–290, 1999.

[18] T. Ruan, Y. Lin, H. Wang, and J. Z. Pan, "A multi-strategy learning approach to competitor identification," *Joint International Semantic Technology Conference*, vol. 89, pp. 197–212, 2014.

[19] G. Pant and O. R. L. Sheng, "Web footprints of firms: using online isomorphism for competitor identification," *Information Systems Research*, vol. 26, no. 1, pp. 188–209, 2015.

[20] Z. Zheng, "The identification of enterprise dynamic competitor based on cluster analysis," in *Proceedings of the International Conference on Information Management*, IEEE, Bangalore, India, December 2011.

[21] Y. Liu, Y. Qian, Y. Jiang, and J. Shang, "Using favorite data to analyze asymmetric competition: machine learning models," *European Journal of Operational Research*, vol. 287, no. 2, pp. 600–615, 2020.

[22] I. Arpaci, T. Abdeljawad, M. Baloğlu, Ş. Kesici, and I. Mahariq, "Mediating effect of internet addiction on the relationship between individualism and cyberbullying: cross-sectional questionnaire study," *Journal of Medical Internet Research*, vol. 22, Article ID e16210, 2020.

[23] I. Arpaci, S. Alshehabi, M. Al-Emran et al., "Analysis of twitter data using evolutionary clustering during the COVID-19 pandemic," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 193–204, 2020.

[24] A. Kangale, S. K. Kumar, M. A. Naeem, M. Williams, and M. K. Tiwari, "Mining consumer reviews to generate ratings of different product attributes while producing feature-based review-summary," *International Journal of Systems Science*, vol. 47, pp. 1–15, 2015.

[25] N. Archak, A. Ghose, and P. G. Ipeirotis, "Deriving the pricing power of product features by mining consumer reviews," in *Proceedings of the Acm Sigkdd International Conference on Knowledge Discovery & Data Mining*, ACM, Halifax, NS, Canada, August 2007.

[26] D. Yin, S. D. Bond, and H. Zhang, "Keep your cool or let it out: nonlinear effects of expressed arousal on perceptions of consumer reviews," *Journal of Marketing Research*, vol. 54, no. 3, pp. 447–463, 2017.

[27] A. Ghose and P. G. Ipeirotis, "Designing novel review ranking systems: predicting usefulness and impact of reviews," in *Proceedings of the Ninth International Conference on Electronic Commerce*, ACM, Minneapolis, MN, USA, August 2007.

[28] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," *Journal of the Association for Information Science & Technology*, vol. 61, pp. 2544–2558, 2014.

[29] Y.-H. Cheng and H.-Y. Ho, "Social influence's impact on reader perceptions of online reviews," *Journal of Business Research*, vol. 68, no. 4, pp. 883–887, 2015.

[30] F. Zhu and X. Zhang, "Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics," *Journal of Marketing*, vol. 74, no. 2, pp. 133–148, 2010.

[31] X. Pan, L. Hou, K. Liu, and H. Niu, "Do reviews from friends and the crowd affect online consumer posting behaviour differently?" *Electronic Commerce Research and Applications*, vol. 29, pp. 102–112, 2018.

[32] J. Bollen and H. Mao, "Twitter Twitter Mood as a stock market predictor," *Computer*, vol. 44, no. 10, pp. 91–94, 2011.

[33] M. Zhang, B. Fang, N. Zhang, W. Wang, and W. Fan, "Mining product innovation ideas from online reviews," *Information Processing & Management*, vol. 58, pp. 1–12, 2021.

[34] M. Luca and G. Zervas, "Fake it till you make it: reputation, competition, and Yelp review fraud," *Management Science*, vol. 62, no. 12, pp. 3412–3427, 2016.

[35] M. Sun, J. Chen, Y. Tian, and Y. Yan, "The impact of online reviews in the presence of customer returns," *International Journal of Production Economics*, vol. 232, Article ID 107929, 2021.

[36] K. Li, Y. Chen, and L. Zhang, "Exploring the influence of online reviews and motivating factors on sales: a meta-analytic study and the moderating role of product category," *Journal of Retailing and Consumer Services*, vol. 55, Article ID 102107, 2020.

[37] X. Xu, "How do consumers in the sharing economy value sharing? Evidence from online reviews," *Decision Support Systems*, vol. 128, Article ID 113162, 2020.

[38] A. Felbermayr and A. Nanopoulos, "The role of emotions for the perceived usefulness in online customer reviews," *Journal of Interactive Marketing*, vol. 36, pp. 60–76, 2016.

[39] W. Bingkun, H. Yongfeng, W. Xian, and X. Li, "A fuzzy computing model for identifying polarity of Chinese sentiment words," *Computational Intelligence and Neuroscience*, vol. 2015, Article ID 525437, 13 pages, 2015.

[40] M. E. Basiri, A. R. Naghsh-Nilchi, and N. Ghasem-Aghaee, "Sentiment prediction based on dempster-shafer theory of evidence," *Mathematical Problems in Engineering*, vol. 2014, Article ID 361201, 13 pages, 2014.

[41] Y.-H. Xu, L.-F. Huang, R.-R. Guo, X.-Y. Zhang, and J.-M. Zhu, "Metrological analysis of online consumption evaluation influence commodity marketing decision based on data mining," *Mathematical Problems in Engineering*, vol. 2020, Article ID 9345901, 10 pages, 2020.

[42] H. Lin, Y. Chen, and Y. Yang, "Cluster Analysis of automobile innovative users based on interactive innovation value," *Mathematical Problems in Engineering*, vol. 2018, Article ID 5320645, 10 pages, 2018.

[43] İ. Güven and F. Şimşir, "Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods," *Computers & Industrial Engineering*, vol. 147, Article ID 106678, 2020.

[44] A. Tony, P. Kumar, and S. Rohith Jefferson, "A study of demand and sales forecasting model using machine learning algorithm," *Psychology and Education Journal*, vol. 58, pp. 10182–10194, 2021.

[45] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, IEEE, Toronto, ON, Canada, August 2010.

[46] M. Sun, "How does the variance of product ratings matter?" *Management Science*, vol. 58, no. 4, pp. 696–707, 2012.

[47] D. J. Dalrymple, "Sales forecasting methods and accuracy," *Business Horizons*, vol. 18, pp. 69–73, 2006.

[48] F. M. Bass, "A new product growth for model consumer durables," *Management Science*, vol. 15, no. 5, pp. 215–227, 1969.

[49] Z.-P. Fan, Y.-J. Che, and Z.-Y. Chen, "Product sales forecasting using online reviews and historical sales data: a method combining the Bass model and sentiment analysis," *Journal of Business Research*, vol. 74, pp. 90–100, 2017.

[50] C. Dellarocas, X. Zhang, and N. F. Awad, "Exploring the value of online product reviews in forecasting sales: the case of motion pictures," *Journal of Interactive Marketing*, vol. 21, no. 4, pp. 23–45, 2007.

[51] C. Zhang, Y.-X. Tian, and L.-W. Fan, "Improving the Bass model's predictive power through online reviews, search traffic and macroeconomic data," *Annals of Operations Research*, vol. 295, no. 2, pp. 881–922, 2020.

[52] Z. Tang and S. Dong, "A total sales forecasting method for a new short life-cycle product in the pre-market period based on an improved evidence theory: application to the film industry," *International Journal of Production Research*, vol. 2020, Article ID 1825861, 15 pages, 2020.

[53] A. Davydenko and R. Fildes, "Measuring forecasting accuracy: the case of judgmental adjustments to SKU-level demand forecasts," *International Journal of Forecasting*, vol. 29, no. 3, pp. 510–522, 2013.

[54] W. Xia, H. Jiang, D. Feng et al., "A comprehensive study of the past, present, and future of data deduplication," *Proceedings of the IEEE*, vol. 104, pp. 1–30, 2016.

[55] R. Tan, Z. R. Dai, and Y. Xue, "Experiment research on duplicated web pages of Chinese elimination algorithm based on improved TextTiling," in *Proceedings of the Advances in Computer Science, Environment, Ecoinformatics, and Education - International Conference, CSEE 2011*, Wuhan, China, August 2011.

[56] W. H. Gomaa and A. A. Fahmy, "A survey of text similarity approaches," *International Journal of Computer Application*, vol. 68, no. 13, pp. 13–18, 2013.

[57] W. Yao and P. Ye, *Simdedup: A New Deduplication Scheme Based on Simhash*, Web-Age Information Management, Nanchang, China, 2013.

[58] R. D. S. Villaca, L. B. D. Paula, R. Pasquini, and M. F. Magalhaes, "A similarity search system based on the hamming distance of social profiles," in *Proceedings of the IEEE Seventh International Conference on Semantic Computing, IEEE Computer Society*, Anaheim, CA, USA, February 2013.

[59] S. P. Borgatti, M. G. Everett, and L. C. Freeman, *Ucinet for Windows: Software for Social Network Analysis*, Analytic Technologies, Harvard, MA, USA, 2002.

[60] I. Alon, M. Qi, and R. J. Sadowski, "Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods," *Journal of Retailing and Consumer Services*, vol. 8, p. 156, 2001.

[61] L. Aburto and R. Weber, "Improved supply chain management based on hybrid demand forecasts," *Applied Soft Computing*, vol. 7, no. 1, pp. 136–144, 2007.

[62] United Nations Conference on Trade and Development, *COVID-19 and E-Commerce*, United Nations Conference on Trade and Development, Geneva, Switzerland, 2020, https://unctad.org/system/files/official-document/dtlstictinf2020d1_en.pdf.