Sentiment Analysis of Consumers Purchasing Internet Red Food Based on Data Mining

Chao Zheng[1], Yunting Miao[2*], Liqing Guo[3]
162243153@qq.com[1], 771022025@qq.com[2], 1374078837@qq.com[3]

Artificial Intelligence college, Jiangxi University of Applied Science[1],

Innovation and Entrepreneurship College, Jiangxi University of Applied Science[2],

Intelligent Manufacturing College, Jiangxi University of Applied Science[3]

Abstract. In recent years, the rapid development of Internet Red Food, set off a wave of Internet Red Food consumption boom on the Internet, but about the Internet Red Food safety incidents frequently, unlicensed sales, three no products and other topics also rushed to the list. This study is based on this paper based on the Python crawl "JD" shop evaluation and bilibili video comments and to the data cleaning, by using TF-IDF algorithm to classify comments, and make co-word analysis, the Gephi software is used to generate word co-occurrence network graph, and conduct emotion analysis based on Hownet emotion dictionary and SnowNLP dictionary, the mining and analysis of comment data are realized. The results show that: from the TF-IDF algorithm, consumers are more concerned about the taste, nutrition and freshness of Internet Red Food. After emotional analysis, it is found that about 86% of consumers have a good impression on Internet Red Food, and negative emotions account for about 14%, which can reflect consumers' love for Internet Red Food.

Keywords: Internet Red Food; Python; emotional analysis; word co-occurrence analysis

1 Introduction

In the era of "fast consumption" background, the rise of the webcast let many characteristics of web celebrity food into the public eye, web celebrity food has gradually become the main consumer shopping choice, people for web celebrity food evaluation heat also become higher, but consumers in the process of buying web celebrity food appeared a large number of high-tech food, unlicensed sales, comment of the three no products this phenomenon may be the market for the essential purchase behavior and its internal mechanism is not perfect. Existing research in the process of purchasing behavior on the benefits and risks of consumers, electronic commerce in the rapid development of emerging things (web celebrity food) research level, only stay in the rise of web celebrity food external factors and market environment, and the research on web celebrity food related quantity is less, the lack of empirical validation research, difficult to verify the market rule, guide the development of the market. Therefore, this paper will establish an emotional analysis model based on data mining, explore the main factors that consumers pay close attention to when buying web celebrity food, and provide marketing ideas and suggestions for web celebrity food.

MSIEID 2022, December 09-11, Chongqing, People's Republic of China

DOI 10.4108/eai.9-12-2022.2327591

2 Literature Review

2.1 Internet celebrity Characteristics

Internet Red Food is food that has become extremely popular in a short period of time as a result of online publicity. It is sold on WeChat Moments and other e-commerce platforms through special marketing methods because of its eye-catching packaging and fashionable and novel taste. Some scholars have found that the food quality and price of Internet celebrities have a significant impact on consumers' purchase intentions. (MOjing Chen etc,2016)[1].Some scholars conclude that the good social image of Internet celebrity food will enhance consumers' purchase intention and that the authenticity, integrity, and reliability of Internet celebrity food information will affect consumers' consumption decisions.(TaiPing Li etc, 2018)[2].

2.2 Consumer purchase behavior

Consumer behavior is a series of processes in which an individual or organization studies from product selection to purchase, and finally meets their own needs through actual use. Consumers 'purchasing behavior will be affected by various factors to different degrees. Research shows that consumers' purchasing attitude and identity will enhance consumers 'trust in the payment security of the platform and also promote consumers' purchasing behavior (Lee EJ, Overby JW, 2004) [2].

2.3 Text emotion analysis

Text sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text mining, and computer linguistics to identify and extract the subjective information in the original material (Yuqin Xia, Xuewei Shan, 2018) [4]. Some scholars have pointed out that the information on the Internet is increasing day by day, containing a huge amount of information with emotional color. The e-commerce platform for buying Internet Red Food contains a large number of comments, which analyzes a large number of post-purchase comments, which is of great significance for merchants to understand consumer opinions (Tianyun Yu, Peijin Li, 2021) [5].

3 Mining and analysis based on the data of JD and bilibili

This paper uses Python to write related programs to text mine Internet Red Food related fields on Jingdong and other e-commerce platforms, and run the program to crawl the famous Internet Red Food review data. The word frequency statistics was conducted by NLP word segmentation (Yafei Zhang, Jianhua Fan, 1998) [6], the Hownet dictionary and SnowNLP dictionary were used for sentiment analysis (Yanyan Zhao etc, 2010) [7] and the word co-occurrence analysis (Qin Zhang, Feicheng Ma, 2007) [8] by statistic was used to realize the mining and analysis of comment data. The data collection work in this paper was completed by using Python. After writing the data collection logic, the program was run, and 10,870 comments on the emergence of web celebrity food were successfully captured in Jingdong and B station and 3,060, respectively. Through the preliminary algorithm screening,4,000 invalid comments were removed, and the effective rate was 78%.

3.1 Text data feature extraction

Online comments are a collection of document data in natural language. Each document is composed of several words with certain semantic logic. According to the frequency of the terms in the text and the specific topics expressed, the vector space model (Vector Space Model, VSM) (Yaohong Kang,1989) [9] is used to establish a text feature model for the massive text, so that the text is transformed into quantifiable structured data for feature mining. The basic principle of VSM is that the comment document Di can be expressed as Di=D (T1,Wi1;T2,Wi2;...;Tm, Wim) for the document collection, Where (T1,T2,...,Tm) is a set of m-dimensional reciprocal terms, (Wi1, Wi 2,..., Wim) is the weight of the corresponding m-dimensional terms in the document, That is, the degree of importance in the document Di, Wij is generally defined as in the document Di (i=1,2,3,...,n) in the term Tj (j=1,2,3,..., m) occurrence frequency (Term Frequency, TF) of the function Tf (dij), The constructed text feature models are shown in **Table 1**.

Table1. Text feature model

document	mouth feel	tasty	...	lexical item
document1 (D1)	W_{12}	W_{21}	...	W_{1m}
document2 (D2)	W_{13}	W_{22}	...	W_{2m}
⋮	⋮	⋮	⋮	⋮
documentn (Dn)	W_{1n}	W_{2n}	...	W_{nm}

Word frequency-inverse document frequency (term frequency-nverse Document Frequency, word frequency-inverse file frequency) (Yuntao Zhang etc, 2005) [10] algorithm is used to extract text features for DTM. Where TF indicates the frequency of the entry(keyword) in the text; IDF is the reverse document frequency, which refers to the reverse value of the document number in the total document to represent the ability of the word to distinguish the document. For the TF-IDF calculation formula, see equation (1).

$$\text{TF - IDF} = TF \times IDF = t_{fi} \times \log\left(\frac{N}{df_i} + 1\right) \tag{1}$$

Where tfi is the frequency of word i in the document: dfi is the number of documents with word i; N is the total number of documents. The results are shown in **Table2**.

Table2.TOP5 key words

Keyword	word frequency
tasty	534
nutrition	356
fresh	330
recipe	276
unusual	250

3.2 Analysis of word co-occurrence

In order to clarify the relationship between keywords, 17 high-frequency keywords in the data were analyzed by python to generate a 17*17 co-occurrence matrix, and a word co-occurrence network graph was generated by using Gephi software. The shorter the connection of each

keyword in the word co-occurrence network graph is, and the closer the connection of the corresponding keywords is. The Fig.1 is the keyword co-occurrence network diagram:

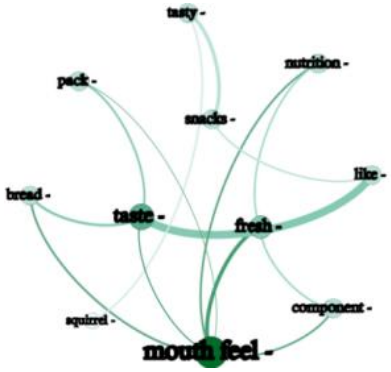

Fig. 1. Network diagram of the co-occurrence of 17 keywords

3.3 Sentiment Analysis

With the help of the emotion dictionary, python is used to compare the words of each comment with the words in the emotional dictionary to obtain emotional scores. In order to more scientific data visualization analysis, we introduce widely recognized and easy rate concept, it refers to the analysis of product reviews comprehensive calculated reference value, this through with big data technology for network public opinion data, is quite consistent with the current actual situation, and in the early stage of the data processing, classification of rigorous method for rough evaluation of web celebrity food. The calculation formula is shown in the equation (2).

$$FC = \frac{NP}{(NP + NN)} * 100\% \tag{2}$$

where FC is the positive rate, NP is the number of positive comments, and NN is the number of negative comments.

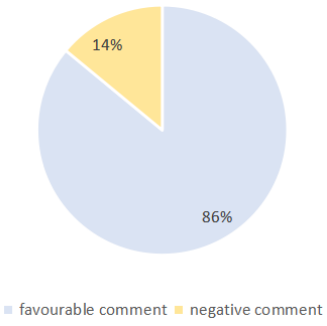

Fig. 2. Sentiment Analysis

The analysis of the review data points out that about 86% of consumers have a good impression on web celebrity food for web celebrity food, and negative emotions account for

about 14%. It can be seen that web celebrity food has great potential for sustainable development.

4 Conclusion

This paper constructs the emotional attitude analysis model of consumer buying celebrity food based on data mining, and the following suggestions are put forward based on the results of word frequency statistics, word co-occurrence analysis and emotion analysis:

First of all, consumers can use big data technology to effectively analyze existing reviews, which can help them understand the real situation of merchants and goods. Furthermore, in the era of big data, data mining technology can dig out valuable information behind the massive data. Businesses can use the technical means of data mining to establish the database of web celebrity food for quantitative analysis and scientific prediction, and improve their market competitiveness.

The research model in this paper is insufficient, and the effective data obtained in this study is not comprehensive enough. The future study can expand the mining scope to improve the number of samples, and make the universality of the research results more general. In this study, only the default corpus is used. In the future research, an accurate corpus of web celebrity products can be constructed to make the accuracy higher.

References

[1] Mojing Chen, Guifang Chao, Xinxin Ding.: Impact of Online Comments on Purchase Intention of College Student Consumers under Online Shopping. Asian Agricultural Research. pp. 29-34 (2013)

[2] Taiping Li, Jiarui Li.: Study on consumers' purchase intention of traceable agricultural products and its influencing factors. Rural South. pp. 14-19 (2018)

[3] Eun-Ju Lee, Jeffrey W.: Creating Value for Online Shoppers: Implications for Satisfaction and Loyalty. The Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior. pp. 54-67 (2015)

[4] Yuqin Xia, Xuewei Shan.: Research on the online shopping Evaluation system based on Emotion Analysis. China Journal of Commerce. pp. 67-70 (2021)

[5] Tianyun Yu, Peijin Li.: Simple text sentiment analysis based on the Python. Yinshan Academic Journal. pp. 58-62 (2018)

[6] Yafei Zhang, Jianhua Fan.: Natural language processing: Methods and techniques. Computer Applications and Software. pp. 1-15 (1998)

[7] Yanyan Zhao, Bin Qin, Tin Liu.: Text emotion analysis. Journal of Software. pp. 1834-1848 (2010)

[8] Qin Zhang, Feicheng Ma.: Foreign knowledge management research paradigm——takes common word analysis as the method. Journal of Management Sciences in China. pp. 65-75 (2007)

[9] Yaohong Kang.: Vector space model based on word dependency. Journal of the China Society for Scientific and Technical Information. pp. 177-185 (1989)

[10] Yuntao Zhang, Ling Gong, Yongcheng Wang.: An improved TF-IDF approach for text classification. Journal of Zhejiang University Science Al. pp. 50-56 (2005)