Entropy Weight-TOPSIS Method Considered Text Information with an Application in E-Commerce

Ailin Liang, Xueqin Huang, Tianyu Xie, Liangyan Tao, and Yeqing Guan

1 Introduction

As the rapid development of Internet economy, many companies are planning to publish and sell new products in the online market. However, how to design products and sell them online are the two major problems in the business process. As we all know, there are too much data that can be analyzed from Amazon, Taobao, JD.COM and other websites. These data have great commercial value. Through efficient collection, processing and analysis, we can help companies make more reasonable decisions on how to design products and how to sell online.

Some researchers take e-commerce products as the research object, based on the sentiment analysis of reviews (Liu Yulin & Tong Lirong, 2018; Li Huizong et al., 2019), putting forward the selection scheme and sales forecast of foreign trade e-commerce products, (Zhang Yue, 2019) studied online reviews of cold chain agricultural product e-commerce, and there are many other similar studies, such as (Yang Ruixin, 2017) on air conditioning and (Zhao Zhibin et al., 2017) on Chinese products. However, few studies focus on finding successful features from comments combined with other data, and help companies make decisions. Finding these features is very important, because for those companies who want to sell things online, there is a lot of wealth in data. This will help companies to make more correct decisions and gain more profits.

At present, there are many technical processes of generating data and information in real time. New technologies, such as big data and artificial intelligence, are the tools to analyze these data, which provide important value for companies. Sentiment analysis is a good way to analyze people's speech and individual behavior. It can

A. Liang (✉) · X. Huang · T. Xie · L. Tao · Y. Guan
College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China
e-mail: ailinliang@nuaa.edu.cn

H. Yang et al. (eds.), *AI and Analytics for Public Health*, Springer Proceedings in Business and Economics, https://doi.org/10.1007/978-3-030-75166-1_8

be used in many fields, such as predicting political election results (Paul et al., 2017), mental health care (Zhai et al., 2010), review analysis (Ye et al., 2005), and product analysis (Wang et al., 2018). Sentiment analysis mainly studies the emotional tendencies of texts from grammar, semantic rules and other aspects (Yanghui Rao et al., 2014). The texts from social network has the characteristics of few words, irregular grammar and noisy data, which increases the difficulty of sentiment analysis (Jun He et al., 2020). Therefore, a reasonable method is also needed to analyze the text in product reviews. Then, the sentiment analysis model based on LDA (Jun He et al., 2020; Wang Rui et al., 2020) is used to score each comment in the data set, and the success of each product is judged by considering the score. After obtaining the emotional score for each product, the data set needs preprocessing to facilitate further analysis. Actually, there are only nine useful data metrics: product_id, star_rating, helpfulvotes, total_votes, vine, verified_purchase, review_header, review_body and review_date. Based on Excel and Python as tools for this study, we can get other data characteristics through analysis and processing (see the subsequent analysis for details). The LDA-based sentiment analysis is used to score each comment, and then the score is used as an index to judge whether each product is successful, which also requires a reasonable weight matrix. There is a method of establishing weight matrix, namely entropy weight method. According to the basic principles of information theory, information is a measure of the degree of system order, and entropy is a measure of the degree of disorder. The smaller the information entropy of the index, the greater the information provided by the index. The greater the effect, the higher the weight. Therefore, we can use this method to calculate the weight of each index, which provides a basis for comprehensive evaluation of multiple indicators. Therefore, the method of weights determination based on entropy weight is objective and effective enough. Entropy method has been applied in many aspects, such as evaluation, decision-making and selection (Jian-qiang Zhang & Hai-hong Sun, 2019; Farhadinia, 2017; Yanmeng Zhang et al., 2017; Xiangxin Li et al., 2011) In 1981, C. L. David Henry Hwang and k. Yoon first proposed the method of ranking by similarity with ideal solution. TOPSIS is a method to calculate the relative distances between objects and the idealized target, and then sorting these objects based on distances. TOPSIS is a general multiple attribute decision making method. And entropy weight is always combined with TOPSIS method to tackle reality problems, such as supplier selection of home appliance industry supply chain (Yanmeng Zhang et al., 2017) and safety evaluation of coal mines (Xiangxin Li et al., 2011). Entropy method and TOPSIS are good at evaluating and find the best choice. On this basis, we can find ten successful products for people to analyze why these products sell well on the Internet.

As shown in Fig. 1, this study first collects data from Amazon, and we analyze the data, then delete unnecessary data and keep the data between 2012 and 2014, which are marked as “y” in vine, “n” in verified_purchase, and vice versa, and then sort the six data sets according to product_id. The index system is constructed based on correlation analysis. This system contains 4 important indexes composed of original measurements. Secondly, use LDA-based sentiment analysis with clean data, to quantify every emotional keyword and get the arithmetic mean as the emotional

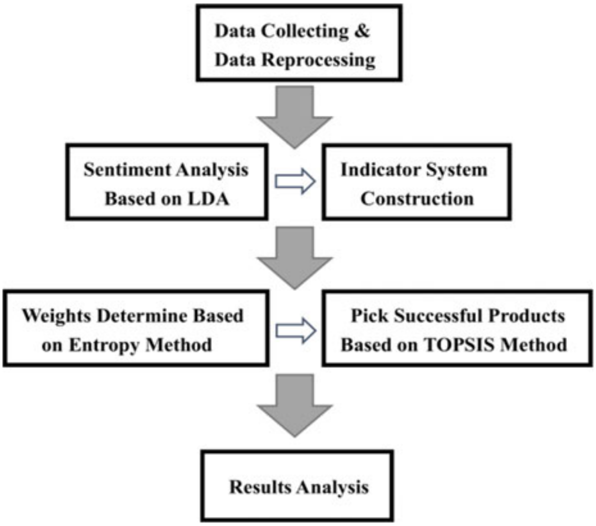
Data Collecting &
Data Reprocessing
Sentiment Analysis Based on LDA
Indicator System Construction
Weights Determine Based on Entropy Method
Pick Successful Products Based on TOPSIS Method
Results Analysis

Fig. 1 Technical route

value of the whole review. Output score for each reviews_body. Then we pick out successful products based on the Entropy Weight-TOPSIS method. Through analyzing these top products, we can extract the successful features.

2 Data

2.1 Raw Data

In the online marketplace it created, Amazon provides customers with an opportunity to rate and review purchases. Individual ratings – called "**star ratings**" – allow purchasers to express their level of satisfaction with a product using a scale of 1 (low rated, low satisfaction) to 5 (highly rated, high satisfaction). Additionally, customers can submit text-based messages – called "**reviews**" – to express further comments and information about the product. Other customers can submit ratings on these reviews as being helpful or not – called a "**helpfulness rating**" – towards assisting their own product purchasing decision. We collect data on three products: a microwave oven, a baby pacifier, and a hair dryer. The three datasets provided include product user ratings and reviews extracted from the Amazon Customer Reviews Dataset through Amazon Simple Storage Service (Amazon S3). A detailed vocabulary of data tag definitions is as follows (Table 1):

Table 1 Data tags define vocabulary

Attributes	Meaning
customer_id (string)	Random identifier that can be used to aggregate reviews written by individual author.
review_id (string)	The unique ID of the review.
product_id (string)	The unique Product ID the review pertains to.
product_category (string)	The major consumer category for the product.
star_rating (int)	The 1–5 stars rating of the review given by people to rate a product with a number of stars.
helpful_votes (int)	Number of helpful votes
total_votes (int)	Number of total votes the review received.
vine (string)	Customers are invited to become Amazon Vine Voices based on the trust that they have earned in the Amazon community for writing accurate and insightful reviews. Amazon provides Amazon Vine members with free copies of products that have been submitted to the program by vendors. Amazon doesn't influence the opinions of Amazon Vine members, nor do they modify or edit reviews.
verified_purchase (string)	A "Y" indicates Amazon verified that the person writing the review purchased the product at Amazon and didn't receive the product at a deep discount.
review_body (string)	The review text.
review_date (bigint)	The date the review was written.

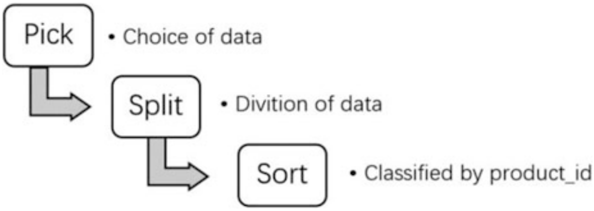

Fig. 2 Data preprocessing flow chart

2.2 Data Preprocessing

1. Import the three data set: hair_dryer.tsv, microwave.tsv and pacifier.tsv, into Excel sheet. Then observe them and analyze them by common sense. Choose the data of 2012, 2013 and 2014, because those deleted data (before 2012 and after 2015) are totally unreliable and redundant.
2. Split three datasets into six datasets. (Extract data measures "vine" and "verified_purchase", if "vine = Y", it means that the reviews are more accurate and insightful. Similarly, "verified_purchase = Y" indicates that the reviews are more objective. Therefore, the original data is divided into two parts according to "vine=Y & verified_purchase= N" or "vine=N & verified_purchase= Y".)
3. Sort all selected data by product_id & time(month) on python 3.7 (Fig. 2).

3 Modeling Preparation

3.1 Indicator System Construction

First, according to the previous analysis, we find that after mathematical processing, these flow data measures can be used as candidate indicators of the indicator system, and these four indicators are sorted by product _ id:

① $\overline{star_{rating}}$: The average of "star_rating" for each product. It's an important factor giving people reference, and the average can be more useful and meaningful, therefore we pick it into our indicator system.

② $\overline{\frac{\text{helpful}_{\text{votes}}}{\text{total}_{\text{votes}}}}$: The average of "helpful_votes/total_votes" for each product. It's an important factor to prove the authenticity and reliability of views, which displays the reputation of a product. (If total_votes = 0, $\frac{\text{helpful}_{\text{votes}}}{\text{total}_{\text{votes}}} = 0$)

③ $\overline{score}$: The Number which translated from "review_headline & review_body" based on sentiment analysis, shows the average product quality and consumer satisfaction for each product.

④ Sales: Sales volumes of each product according to product_id (Table 2).

3.2 LDA-Based Sentiment Analysis

1. Assumptions

① Consumers' emotions is closely related to product quality. ② Data is valid for completing emotion analysis.

2. The Foundation of LDA

LDA (Linear Discriminant Analysis) is a topic model, which can display the topic of each text in the form of probability distribution, and reverse the distribution of the topic based on a given document. After analyzing some texts to extract their topics (distribution), these texts and topics can be classified into clusters. At the same time, it is a typical bag-of-words model, that is, a text is composed of a group of words, and there is no sequential relationship between words. Establishing variables for LDA (Table 3):

Table 2 Indicators' symbols

Contents	Symbols
$\overline{star_{rating}}$	x_1
$\overline{\frac{\text{helpful}_{\text{votes}}}{\text{total}_{\text{votes}}}}$	x_2
$\overline{score}$	x_3
Sales	x_4

Table 3 Variables' define for LDA

Contents	Symbols
Keyword	K
Number of keywords in comments	n
Number of negative words in comments	m
Single keyword sentiment score	$e_i, i \in \{1, 2, \cdots, n\}$
Emotion score	E
Degree word sentiment value	level1(0.2 ~ 0.5)level2(0.4 ~ 0.7)level3(0.6 ~ 0.9)
	level4(1.2 ~ 1.6)level5(1.6 ~ 1.9)level6(2.0 ~ 2.5)
Negatives	D

The initial emotional value of positive word is 1, and the initial emotional value of negative word is − 1. We use this model and this idea to implement the text sentiment analysis of the views. The realization principle is as follows:

Step1: Loading emotional dictionary (divided into positive word/negative word/degree word/negative word) and word segmentation.

Step2: The emotional value of positive and negative words is accumulated by keyword matching. The emotional default value of positive and negative words are 1/−1.

$$K = \begin{cases} 1, word\ is\ positive \\ -1, word\ is\ negative \end{cases} \tag{1}$$

Step3: Retrieving and verifying the semantics of negatives D in text, and confirming the positive and negative again by multiple checks.

$$K' = -1^m K \tag{2}$$

Step4: Controlling the degree of repeated emphasis by a specific interval attenuation function. Determining the level of emphasis and multiplying K' with L (Degree word sentiment value) which is determined by the upper and lower limits of the interval and the attenuation coefficient.

$$e_i = K' \times L \tag{3}$$

Step5: Get an overall emotional value E of the text which is the arithmetic average of each single keyword's emotional value e_i, and the results of piecewise analysis.

$$E = \frac{1}{n}\sum_{i=1}^{n} e_i \tag{4}$$

3. Results (Run this model in python 3.7)

Output scores of each reviews_body. There are samples of results (Table 4):

Table 4 Sentiment scores on reviews for pacifiers

Reviews_body	Scores (E)
Sample1: Easy to clean, no weird smell, feels secure. These are pretty similar to Nuk pacifiers, but I like the material better. Nice pacifier.	2.67
Sample2: this is not worth it if you have any wind in your neighborhood. it does not have any weight to it and will simply blow down the street with any degree of wind. not worth it unless you get a 5 lb weight to hold it down.	−2.00
Sample3: very disappointed on this one. we need it for my husband to hold onto to enter the bath and it doesn't stick. it will lose suction at random times and you'll hear it crashing down into the tub. it's on smooth ceramic tiles as well. it should stay put. not happy with this purchase.	−2.89

4 Decision Making Based on Entropy Weight-TOPSIS Method

4.1 The Foundation of Model

To observe and confirm the characteristics of successful products by using multi-attribute decision-making method, it is necessary to establish a reliable index system. Then we choose entropy weight method to gain the weight matrix for indicators and TOPSIS method to find successful products. The modelling steps are as follows:

Step1: According to the indicator system and definitions, a decision matrix is obtained $X = (x_{ij})_{m \times n}$.

n: number of indicators, m: number of products, x_{ij} represents the j-th indicator value of the i-th evaluation object.

Step2: Normalize the raw matrix $Z = (z_{ij})_{m \times n}$ by column.

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \, (1 \le i \le m, 1 \le j \le n) \tag{5}$$

Step3: Entropy can be expressed as the following formula by definition:

$$S\left(z_j\right) = \sum_{i=1}^{m} z_{ij} \ln z_{ij} \tag{6}$$

Step4: Calculate the entropy of each indicator.

$$S_j = -k \cdot s\left(z_j\right) \tag{7}$$

k is related to the number of samples m, and $k = \frac{1}{\ln m}$. The supplementary definition: If $z_{ij} = 0$, let $z_{ij} \, ln \, z_{ij} = 0$.

Step5: The Degree of difference of indicators.

$$G_j = 1 - S_j \tag{8}$$

Step6: Calculate the weight coefficient of each indicator.

$$c_j = \frac{G_j}{\sum_{i=1}^{n} G_i} \tag{9}$$

Step7: Because the importance of various indicators is different, the entropy weight of each indicator should be considered, and the normalized data is weighted to obtain the following weighted normalization matrix.

$$V = \left(v_{ij}\right)_{m\times n} = \begin{bmatrix} c_1 z_{11} & c_2 z_{12} & \cdots & c_n z_{1n} \\ c_1 z_{21} & c_2 z_{22} & \cdots & c_2 z_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ c_1 z_{m1} & c_2 z_{m2} & \cdots & c_n z_{mn} \end{bmatrix} \tag{10}$$

Step8: Determine positive ideal solution and negative ideal solution.

$$V^+ = \left\{\left(\max_i v_{ij} | j \in J_1\right), \left(\min_i v_{ij} | j \in J_2\right)\right\} (i = 1, 2, \ldots, m) \tag{11}$$

$$V^- = \left\{\left(\min_i v_{ij} | j \in J_1\right), \left(\max_i v_{ij} | j \in J_2\right)\right\} (i = 1, 2, \ldots, m) \tag{12}$$

Among them J_1 is the benefit-type indicator set and J_2 is the cost-type indicator set.
Step9: Calculate distance.
The distances between the evaluated indicators and the positive ideal solution:

$$d_i^+ = \left[\sum_{j=1}^{n}\left(v_{ij} - v_j^+\right)^2\right]^{\frac{1}{2}} (i = 1, 2, \ldots, m) \tag{13}$$

The distance from the evaluated indicators to the negative ideal solution:

$$d_i^- = \left[\sum_{j=1}^{n}\left(v_{ij} - v_j^-\right)^2\right]^{\frac{1}{2}} (i = 1, 2, \ldots, m) \tag{14}$$

Step10: Calculate relative proximity. The relative proximity of each indicator to the ideal solution:

$$C_i = \frac{d_i^-}{d_i^+ + d_i^-} (i = 1, 2, \ldots, m) \tag{15}$$

Step11: Ranking relative proximity C_i.

According to the relative proximity values obtained above. The greater the relative closeness of the product, the more successful the product will be.

4.2 Results and Analysis (Run This Model in Python 3.7)

Six datasets have six weight matrixes, as shown in Table 5.

There is a sample result (picking the top 10 of pacifiers) in Table 6.

Firstly, $\boldsymbol{x_3}$ is the number which is translated from "review_headline & review_body" based on sentiment analysis. By sentiment analysis, the evaluation and qualitative analysis of products by consumers are transformed into quantitative indicators. Therefore, the index size of $\boldsymbol{x_3}$ represents the satisfaction degree of customers. The higher the score, the greater the customer's satisfaction and the more successful the product is. Therefore, the index of $\boldsymbol{x_3}$ can reflect the superiority of some characteristics of the product itself. For example, high-quality raw materials, aesthetic appearance, good customer experience, etc., greatly affect or determine the success of online sales of products.

In addition, $\boldsymbol{x_2}$ is the average value of "helpful_votes/total_votes" for each product. $\boldsymbol{x_2}$ represents the proportion of the number of consumers who voted that the

Table 5 The weights of datasets

Datasets	x_1	x_2	x_3	x_4
pacifier_vine	0.00857046	0.31428073	0.53113709	0.14601172
pacifier_verified_purchase	0.00920789	0.46741284	0.46477582	0.05860344
microwave_vine	0.00641744	0.01582893	0.41178463	0.56596899
microwave_verified_purchase	0.03262573	0.18562624	0.55308866	0.22865938
hairdryer_verified_purchase	0.01570945	0.27572281	0.61800476	0.09056299
hairdryer_vine	0.00728758	0.6814491	0.28472485	0.02653847

Table 6 Decision making on products

product_id	d_i^+	d_i^-	C_i	order
B0009XH6TG	0.136476	0.302574	0.689156	1
B00132ZG3U	0.136454	0.301603	0.688501	2
B003V264WW	0.137605	0.286589	0.675609	3
B00005O0MZ	0.147816	0.264795	0.641755	4
B000R80ZTQ	0.160629	0.224238	0.582637	5
B001QTW2FK	0.200724	0.158145	0.440677	6
B001UE7D2I	0.204837	0.155294	0.431216	7
B000A3I2X4	0.208026	0.153092	0.423938	8
B00APV7OWG	0.230470056	0.120650182	0.343615003	9
B0009XH6WI	0.238325699	0.114978564	0.325437806	10

product is useful to the total number of consumers who voted. The larger the index is, the greater the proportion of people who think the product is useful. Therefore, this index reflects the reputation and popularity of the product among consumers who purchase the product. This index reflects whether the product is welcomed by consumers more truly and reliably. Thus, this index can also reflect some important features of the product itself, which have the magic of attracting consumers and being favored by more consumers.

The weight ratio of $\boldsymbol{x_2}$ and $\boldsymbol{x_3}$ far exceeds the weight ratio of $\boldsymbol{x_1}$ and $\boldsymbol{x_4}$, which is up to 76.6326303%. This shows that consumers' preference information about products can be transformed into the data of these two indicators. Therefore, it is of more important theoretical and practical significance to analyze the characteristics of products based on $\boldsymbol{x_2}$ and $\boldsymbol{x_3}$ index data. Thus, for a company planning to design and sell online products, it should pay more attention to the information of consumers' favorite degree reflected by $\boldsymbol{x_2}$ and $\boldsymbol{x_3}$ indicators, as well as some characteristic information of the products themselves reflected based on this. According to the different characteristics of different products, it is easier to grasp the key features and contradictions of online sales products, design or optimize the corresponding products, and achieve the success of online sales. Moreover, It is also suggested that online sellers should pay attention to customer feedback information, use various measures to win consumers' favor, and establish good reputation, so as to increase the possibility of repurchase by old customers and attract new customers.

By analyzing the characteristics of the top ten products, we can provide some useful proposals for the company to making better decisions. Taking pacifiers as an example, this paper finds out the IDs of ten pacifiers by this method, and analyzes some characteristics of them as examples (Table 7).

For example, in terms of price, excluding the highest price and the lowest price, and calculating the average price of the remaining eight products, it can be found that if the price of the products is set at about 6.315$, the products are more likely to succeed in online sales. In terms of product materials, it can be found that 9/10 products are made from silica gel. Therefore, products made of this raw material are

Table 7 Product_id and some typical characteristics

product_id	Price	Material	Whether have free gift	...
B0009XH6TG	4.39$	Wacker silica gel	Yes	...
B00132ZG3U	5.99$	LSR food grade silica gel	No	...
B003V264WW	5.29$	Wacker silica gel	Yes	...
B00005O0MZ	4.39$	Nano silver liquid silica gel	Yes	...
B000R80ZTQ	9.99$	Food grade silica gel	Yes	...
B001QTW2FK	3.49$	Silastic	No	...
B001UE7D2I	16.39$	Food grade silica gel	Yes	...
B000A3I2X4	8.99$	Food grade silica gel	Yes	...
B00APV7OWG	5.99$	Nano silver liquid silica gel	Yes	...
B0009XH6WI	5.49$	Wacker silica gel	No	...

easier to sell successfully on the internet. As to whether there are free gifts or not, 7/10 products provide consumers with free gifts, so it is one of the successful ways to obtain online sales. There are more than three factors influencing the success of online sales. Therefore, when a company intends to open up the Internet market of a certain product, it can comprehensively evaluate the characteristics and analyze the advantages and disadvantages of the product, so as to make more scientific and effective decisions.

5 Conclusions

In our research, we use LDA sentiment analysis to convert online products reviews from text to numbers. Then, an index system with 4 indexes is constructed. Based on the entropy weight method, each index is given appropriate weight, and TOPSIS is used for comprehensive evaluation according to the characteristics of products, which provides quantitative basis for enterprises to make timely and reasonable product market decisions.

Through the above work, we find that x_2 and x_3 are very important and meaningful. Companies should pay more attention to x_2 and x_3 on the Internet to gain more trust from consumers. Different data has different values for different products. Managers should focus on different products, pay attention to different points, grasp the main contradictions, and analyze them, which can often grasp consumers' hearts.

On the one hand, this research is a way to enable managers to discover key points in big data faster and more accurately. On the other hand, it's a new and extensive method for TOPSIS. It enables TOPSIS to tackle indicator systems which contain text attributes.

Acknowledgements This research was supported by a project of the National Natural Science Foundation of China (72071111). At the same time, the authors would like to acknowledge the partial support of the project of Intelligence Introduction Base of the Ministry of Science and Technology (G20190010178).

References

Farhadinia, B. (2017). A multiple criteria decision making model with entropy weight in an interval-transformed hesitant fuzzy environment [J]. *Cognitive Computation, 9*(4).

Jian-Qiang Zhang & Hai-Hong Sun. (2019). Study on the evaluation index of cadet's physical training based on the entropy weight [C]. In *Proceedings of 2019 2nd International Conference on Informatics, Control and Automation (ICA 2019)* (pp. 88–92). Advanced Science and Industry Research Center: Science and Engineering Research Center.

Jun He, Hongyan Liu, Yiqing Zheng, Shu Tang, Wei He, & Xiaoyong Du. (2020). Bi-Labeled LDA: Inferring interest tags for non-famous users in social network [J]. *Data Science and Engineering, 5*(1).

Li Huizong, Yao Yao, Wang Xiangqian, et al. (2019). Sentiment analysis of online reviews of cold chain agricultural product e-commerce based on LDA [J]. *Journal of Nanyang Institute of Technology, 11*(2), 25–30.

Liu Yulin, & Tong Lirong. (2018). E-commerce online review data mining based on text sentiment analysis [J]. *Statistics and Information Forum, 33*(12), 119–124.

Paul, D., Li, F., Teja, M. K., Yu, X., & Frost, R. (2017). Compass: Spatio temporal sentiment analysis of US Election what Twitter says! In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, Canada* (pp. 1585–1594).

Wang, F. F., Zhang, S. T., Zhang, J. L., et al. (2018). Research on the majority decision algorithm based on WeChat sentiment classification [J]. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 35*(3), 2975–2984.

Wang Rui, Long Hua, Shao Yubin, & Du Qingzhi. (2020). Text feature extraction method based on labeled-LDA model [J]. *Electronic Measurement Technology, 43*(01), 141–146.

Xiangxin Li, Kongsen Wang, Liwen Liu, Jing Xin, Hongrui Yang, & Chengyao Gao. (2011). Application of the entropy weight and TOPSIS method in safety evaluation of coal mines [J]. *Procedia Engineering, 26*.

Yang Ruixin. (2017). *Emotion analysis of review data of e-commerce air conditioning products [D]*. Shanxi University.

Yanghui Rao, Jingsheng Lei, Liu Wenyin, et al. (2014). Building emotional dictionary for sentiment analysis of online news [J]. *World Wide Web, 17*(4), 723–742.

Yanmeng Zhang, Shengshi Zhou, & Di Wu. (2017). Research on supplier selection of home appliance industry supply chain based on entropy weight and TOPSIS method [C]. Research Institute of Management Science and Industrial Engineering. In *Proceedings of 2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017)* (pp. 672–676). Research Institute of Management Science and Industrial Engineering.

Ye, Q., Li, Y., & Zhang, Y. (2005). Semantic-oriented sentiment classification for Chinese product reviews: An experimental study of book and cell phone reviews. *Tsinghua Science and Technology, 10*(S1), 797–802.

Zhai, Z., Xu, H., & Jia, P. (2010). An empirical study of unsupervised sentiment classification of Chinese reviews. *Tsinghua Science and Technology, 15*(6), 702–708.

Zhang Yue. (2019). *Research on selection of foreign trade e-commerce based on comment sentiment analysis and sales forecast* [D]. Beijing Jiaotong University.

Zhao Zhibin, Liu Huan, Yao Lan, et al. (2017). Dimension mining and sentiment analysis of Chinese product reviews * [J]. *Computer Science and Exploration*.