Journal of Computers Vol. 33 No. 4, August 2022, pp. 15-27
doi: 10.53106/199115992022083304002

Data Analysis of Amazon Product Based on LSTM and GPR

Zi-Yang Ye[1], Xuan Ji[1], Ming-Zi Ye[2], Yu-Tong Shan[3], Xiang-Rong Shi[4*]

[1] Department of Data Science, Zhejiang University of Finance and Economics, 310018, Hangzhou, China
13194387533@qq.com
[2] Department of Finance, Zhejiang University of Finance and Economics, 310018, Hangzhou, China
[3] Department of Accounting, Zhejiang University of Finance and Economics, 310018, Hangzhou, China
[4] Department of Information Management, Zhejiang University of Finance and Economics, 310018, Hangzhou, China

Received 24 July 2021; Revised 2 February 2022; Accepted 2 March 2022

Abstract. In this paper, we propose a method that combines models such as GPR with PSO optimization to predict the time series data. We use LSTM and TOPSIS with entropy weight method modification to process various types of data from various aspects, taking into account both tabular and textual data, and to mine valuable contents from them. Based on shopping data, we analyze the historical situation and predict the future sales of products. So that we can recommend the most suitable products for customers. At the same time, for merchants, this paper provides directions for product optimization and improvement of advertising and marketing strategies.

Keywords: PSO, GPR, LSTM, Natural Language Process (NPL), TOPSIS, entropy weight method

1 Literature Review

The development of the Internet has led to the progress of online shopping, and the importance of e-commerce platform data has emerged. Yun-yun Zeng argues that the research and analysis for online store operations need to be strengthened in terms of customer data, so as to improve the economic returns of online stores and promote further development of online stores [1]. Yu-hang Li suggested that with the outbreak of NCCP in 2020, offline consumption is greatly impacted and the share of online channels is rapidly increasing, which makes digital consumption increasingly important in the consumption field. In particular, during the Newcastle pneumonia epidemic in 2020, China's consumption level did not decline, but achieved a counter-trend growth [2]. Ze-yang Gao and Jia-Nan Hao argue that because the degree of socio-economic development also varies, the distribution of online shopping outlets and the consumption preferences of residents for online shopping differ, and the natural resources vary from region to region, the population of online shoppers shows significant spatial variability. The consumption characteristics of the population in different regions can be studied by [3].

And the relevant data can be obtained very easily from the back-end of e-commerce platforms, how to analyze is indeed an important aspect. To analyze the Amazon platform shopping data, we used a method that combines PSO, GPR, entropy weight method, LSTM, TOPSIS and other models. Particle swarm optimization (PSO) is a bionic algorithm proposed by Kennedy and Eberhart in 1995, which is a group intelligence model inspired by the activity pattern of flocking birds and based on sociology and psychology [4]. W. Hu, G. G. Yen, X. Zhang proposed Particle Swarm Optimization (PSO) particle swarm optimization algorithm is considered as one of the most promising methods for solving multi-objective optimization problems because of the advantages of simple form, fast convergence and flexible parameter adjustment mechanism, as well as the ability to obtain multiple solutions in one run and the ability to approximate the non-convex or discontinuous Pareto optimal front end [5]. And Zhikun He and Guangbin Liu et al. argue that Gaussian process regression is a new machine learning method developed based on Bayesian theory and statistical learning theory, which is suitable for dealing with complex regression problems such as high dimensional number, small samples and nonlinearity. Based on the principle of this method, the problems of large computation and noise must obey Gaussian distribution were analyzed, and the improvement method was given [6]. Combining PSO with GPR can better construct the model and thus improve the performance of Gaussian process regression, and an important method used to reduce the computational effort is hyperparameter optimization. We introduce PSO combined with GPR for innovation. Wenliang Cao and Lanlan Kang proposed that in order to solve the problems of lack of a priori knowledge in hyperparameter optimization problems, over-dependence on initial values and easy to fall into local optimality, a particle swarm optimization algorithm can be introduced, and an adaptive differential particle swarm-Gaussian process regression optimiza-

* Corresponding Author

tion algorithm is proposed for hyperparameters in GPR by combining the differential velocity update formula and adaptive variation strategy optimization. The algorithm optimizes the hyperparameters and the GPR has high fitting accuracy and generalization ability [7].

The entropy method is an objective assignment method, which is used in information theory to measure the amount of information, i.e., the more orderly a system is, the lower the information entropy; on the contrary, the higher the information entropy. Therefore, information entropy can also be said to be a measure of the degree of disorder in a system. Wang Qingyuan and Xuhai proposed that in the evaluation process, the size of the information obtained is one of the determinants of the evaluation accuracy and reliability, and if the information entropy of the index is smaller, the more information the index provides, the greater the role in the comprehensive evaluation, and the higher the weight [8]. The long short-term memory (LSTM) network is a modification of recurrent neural networks. Since it was proposed by Hochreiter and Schmidhuber in 1997, LSTM has achieved good research results in many fields [9]. According to Sunrich, it is a novel, efficient, gradient-based method that truncates gradients without causing harm. LSTM can learn to bridge the minimum time lag of more than 1000 discrete time steps by forcing a constant error stream through constant error rota within a special cell. The multiplicative gate unit learns to turn on and off access to the constant error stream [10].

TOPSIS (technique for order preference by similarity to ideal solution) model, proposed by Hwang and Yoon in 1981, is a decision technique for multi-objective decision analysis of finite solutions in systems engineering as distance integrated evaluation method. That is, the method of ranking a finite number of evaluation objects according to their proximity to an idealized goal is an effective method for evaluating the relative merits of existing objects, a ranking method that approximates the ideal solution, and is commonly used in multi-objective decision analysis [11]. In order to improve its objectivity, Dan calculates the weights more scientifically and honestly reduces the dependence of evaluation objectives on human subjective judgment, and defines the importance of attributes through information entropy, and proposes an improved TOPSIS evaluation method based on entropy weight method for the selection of raw material suppliers of enterprises This method provides a reference basis for the selection of raw material suppliers, and then constructs a closeness function between the evaluation objective and the optimal objective [12]. In our paper, we take all these factors and advantages into consideration and propose a shopping data mining framework which combines models mentioned above.

2 Introduction

The innovation of this paper lies in the application of PSO-Gaussian model for simulation and prediction. What's more, we make use of LSTM model to analyze the unstructured data and produce the emotional score based on text data. Then we input the all the data mentioned above into TOPSIS which is improved by entropy weight method so as to obtain the final comprehensive evaluation criteria. Therefore, we can select the best 5 product that can be recommended to customers and predict specific types of products which might be successful in the near future. After that, we analyze the text data of successful products and use LDA to get the key words which help to answer the question why these products are popular among customers. Then, we propose the framework and the specific flow chart is as follows (Fig. 1). This method can help customers to evaluate the quality of specific products and also help the industry to improve their products based on the key word from text data or improve the advertising strategy by adding the attracting key words we find.

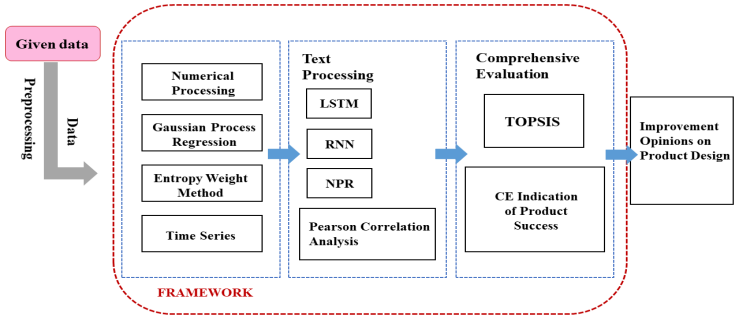


Fig. 1. Flow chart

- Numerical processing: The entropy weight method is used to synthesize the numerical data, and a Gaussian model optimized by PSO is used for fitting and prediction to reflect the changes of product market reputation.

- Text processing: LSTM was used to analyze the textual sentiment to conclude that customer sentiment is closely related to the rating level. Unstructured data were analyzed and it was found that higher star ratings are more likely to trigger positive ratings.

- Determine the comprehensive evaluation criteria: Integration was performed using TOPSIS modified by entropy weight method to obtain the comprehensive evaluation criteria. By ranking the data set, it points to potentially successful products.

- Propose functional improvement and marketing strategies: By analyzing the performance and evaluation of potentially successful products, we obtain functional improvement plans and overall marketing strategies.

3 Assumptions

First, we make some basic assumptions and explain their rationale.

Assumption 3.1. When evaluating products, each customer is absolutely objective in rating their own personal experience of using the product based on the product, and there is no malicious rating of low scores or paying for high scores.

Hypothesis 3.2. Throughout the data study period, their ratings are not influenced by other factors, such as the personal reputation of the endorser, except for the product update iteration.

Assumption 3.3. The data studied in this paper are absolutely realistic and free from any fabrication.

4 Sorting and Classification

Viewing all valid items, we sorted the items in the data file as shown in Table 1 and Fig. 2 to help the reader understand the correlation between the extracted items. Based on the form of user evaluations, we classified all items into four categories: review time, measurement data, numerical evaluation, review text, and composite evaluation, which have different analysis and processing dimensions, and replaced the full names with abbreviations as well. The new indicator composite evaluation in the last row represents the final measure obtained by combining the first four items.

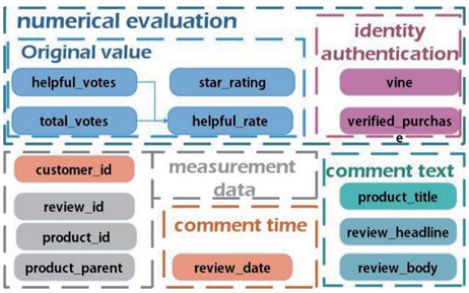


Fig. 2. Data

Table 1. Abbreviation of new categories

Item	Full Name
CM	Comment Time
MD	Measurement Data
NE	Numerical Evaluation
CT	Comment Text
CE	Comprehensive Evaluation

5 Processing of Numerical Evaluations

5.1 Performance of the Time Series Numerical Evaluation

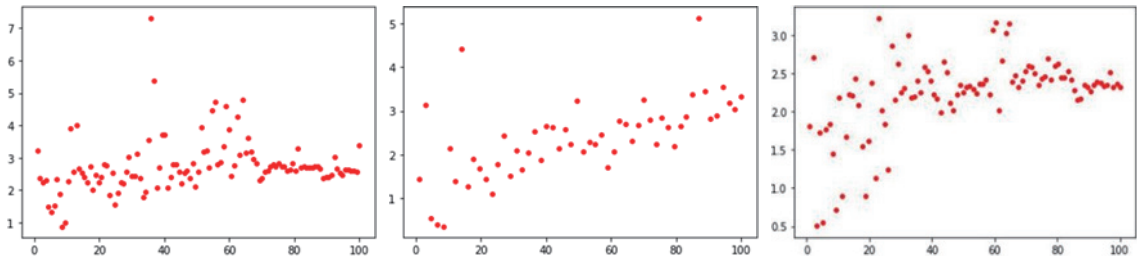
Based on the dataset it is known that the data are sourced from a very wide range of time, with hair dryer reviews spanning even 13 years, from 2002 to 2015. In this chapter, we place the users' numerical evaluation in the time dimension for observation and prediction. This is illustrated in Fig. 3.

From the figure, we can learn that

1. the strong fluctuations in project values require a model to capture the instability;
2. the model needs to accommodate the stochastic noise in the time series.

Gaussian Process Regression (GPR) is a powerful model that can be used to represent the distribution of a function. Currently, a common approach to machine learning is to parameterize the function and then use the generated parameter modeling to avoid distribution representation. However, GPR is different in that it generates non-parametric models for function modeling directly. One of the outstanding advantages is that it can model not only black box functions, but also uncertainty.

For these challenges, we use the nonparametric probabilistic model Gaussian Process Regression (GPR) to fit the time series.



(a) CE of hair dryer (b) CE of microwave (c) CE of pacifier

Fig. 3. Composite evaluation distribution of the three commodities on the time axis

Based on our observations, we summarize the following challenges when describing time series.

Strong fluctuations in item values require a model to capture the instability, and the model needs to accommodate the random noise in the time series.

5.2 PSO Introduction

Particle swarm optimization algorithms are an evolutionary computational technique. Particle swarm algorithms mimic the swarming behavior of insects, herds of animals, flocks of birds, schools of fish, etc. These groups search for food in a cooperative manner, with each member of the group constantly changing its search pattern by learning from its own experience and the experience of other members. The basic idea of the particle swarm optimization algorithm is to find the optimal solution through collaboration and information sharing among individuals in a population. The advantage is that it is simple and easy to implement and does not require many parameters to be adjusted. It has been widely used in function optimization, neural network training, fuzzy system control and other applications of genetic algorithms. In this model, we use the PSO algorithm to optimize the parameters used in the Gaussian process regression model.

5.3 Gaussian Model for Description and Prediction

The input to the function is x and the output of the function is the mean and variance of the Gaussian distribution.

For the data set D: (X, Y), let $F(x_i) = y_i$ and obtain the vector $f = [f(x_1), f(x_2), \ldots, f(x_n)]$. The set of xi to be predicted is defined as X, and the corresponding predicted value is f. According to the Bayesian formula.

$$p(f*|f) = \frac{p((f|f*)}{p(f)} = \frac{p(f,f*)}{p(f)}. \quad (1)$$

Gaussian regression first calculates the joint probability distribution $f \sim N(\mu, K)$ among the samples in the data set, μ is a vector consisting of the means of $f(x_1), f(x_2),\ldots$, a vector consisting of the mean of $f(x_n)$, and K is its covariance matrix, and then calculates the posterior probability distribution of f* based on the previous probability distributions $f^* \sim N(\mu^*, K^*)$ and f^* of $f \sim N(\mu, K)$.

5.4 Calculation of the Covariance Matrix

Defining the function $m(x) = E(f(x))$, $k(x, x^T) = K$, we can obtain from the basic formulas of probability theory.

$$f(x) \sim N(m(x), k(x, x^T)) . \tag{2}$$

$$m(x) = E(f(x)) . \tag{3}$$

$$k(x, x^T) = E\left(f(x) - m(x)(f(x) - m(x))^T\right) . \tag{4}$$

We need to know the following two theorems.
Theorem 4.1. The covariance matrix must be a positive semi-definite matrix;
Theorem 4.2. All kernel analytic quantities are positive semi-definite matrices.
This time, we choose the RBF kernel as follows

$$k(x_i, x_j) = \exp\left(-\frac{1}{2\theta^2}\left\|x_i - x_j\right\|^2\right) . \tag{5}$$

We can add diagonal elements here, which is a regularization method (to avoid over fitting).

$$k(x_i, x_j) = \exp\left(-\frac{1}{2}(x_i, x_j)^T diag(\theta)^{-2}(x, x')\right) . \tag{6}$$

where is a super parameter that can be used to control the width of the kernel. To adjust its parameters, the kernel $k(x, x')$ is evaluated for its advantages and disadvantages in terms of $f(x) \sim N(m(x), k(x, x^T))$ to maximize $p(Y|X)$

To facilitate the derivation, we set the objective function as

$$\log p(Y|X) = \log N(\mu, K_\gamma - \frac{1}{2} tr\left(K_\gamma^{-1}\frac{\partial K_\gamma}{\theta}\right) . \tag{7}$$

Using the gradient descent method, we find the optimal value. Now we find the gradient function as follows.

$$\frac{\partial \log(Y|X}{\theta} = \frac{1}{2} y^T K_y^{-1}\frac{\partial K_y}{\theta} K_y^{-1} y - \frac{1}{2} tr\left(K_y^{-1}\frac{\partial K_y}{\theta}\right) . \tag{8}$$

Given $f(x) \sim N(\mu, K), f(x^*) \sim N(\mu, K(x, x))$, the prior of its joint probability distribution can be computed as

$$\begin{pmatrix} f \\ f* \end{pmatrix} \sim \left(\begin{pmatrix} \mu \\ \mu* \end{pmatrix}, \begin{pmatrix} K & K* \\ K*^T & K** \end{pmatrix}\right) . \tag{9}$$

K **is the covariance matrix of $f(x^*)$

$$K** = k(X*, X*) . \tag{10}$$

$$K* = k(X*, X*) . \tag{11}$$

Based on the prior distributions of $P(f)$ and $p(f, f^*)$ computed above, the posterior probability of p (f *|f) can be computed according to the Bayesian formula

$$p(f*|\ \mathrm{f}) = \frac{p(f|\ \mathrm{f}*)}{\mathrm{p(f)}} = \frac{p(f, f*)}{p(f)} \quad . \tag{12}$$

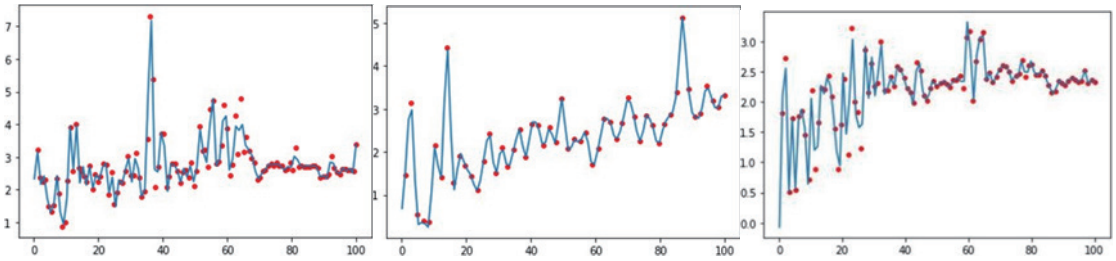
From this, the estimate of f^* , $f^* \sim (u', K')$, is obtained as

$$\mu' = K^T K^{-1} f \ . \tag{13}$$

$$K' = K*^T \ K^{-1} K* + K** . \tag{14}$$

5.5 Gaussian Regression Results

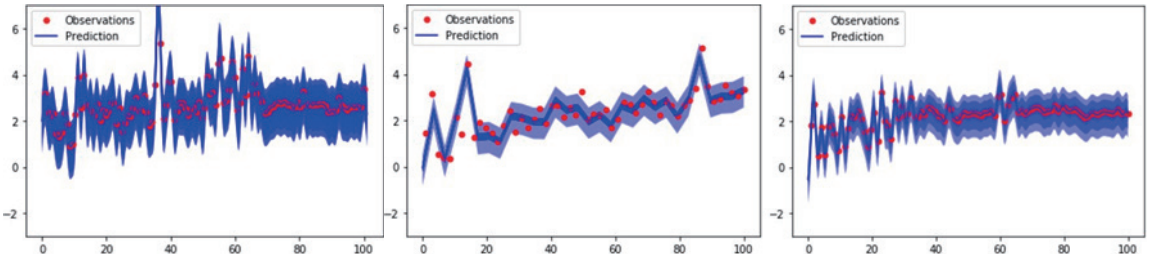
The results of the NE fit of Gaussian regression for the three products are shown in Fig. 4: the total evaluation of the three products shows an increasing trend. There are many data points in the comprehensive evaluation of hair dryer (a), which is the most intensive in 2012 to 2015. The comprehensive evaluation of microwave (b) has fewer data points and the score shows a steady upward trend. The comprehensive evaluation of soothers (c) spans a very large range, with a clear upward trend in the overall evaluation, but the evaluation has been relatively stable over the last three years.



(a) Gauss fit of hair drier (b) Gauss fit of microwave (c) Gauss fit of pacifier

Fig. 4. Distribution of numerical evaluation of three commodities on time axis

The Gaussian model has a higher correlation with temporal proximity values, which is consistent with closer information in product evaluation and prediction. Also, considering random errors, the Gaussian model can predict output probabilities with confidence intervals. This is shown in Fig. 5.



(a) Confidence interval of hair drier (b) Confidence interval of microwave (c) Confidence interval of pacifier

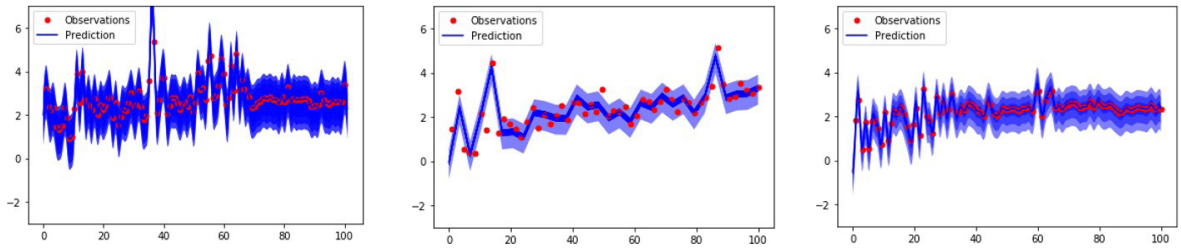
Fig. 5. 95% confidence interval of three commodities on time axis

Since the Gaussian model does not output deterministic values, we show the 95% configuration interval for

each output. The gap surrounded by the upper and lower confidence lines is called the confidence interval. We can see that the NE for microwave ovens shows a continuous increase with a fairly small confidence interval, which indicates a high reliability of the prediction. In contrast, the confidence interval is larger for the NE of the hair dryer, which fluctuates the most, which indicates that the customers' evaluation of the hair dryer is highly random, less predictable, and sensitive to external factors. The NE of the pacifier showed a smooth fluctuation trend with a small confidence interval, indicating that the prediction results were more reliable.

5.6 Prediction Using Gaussian Model

Forecasting using Gaussian is shown in Fig. 6. We can see that the reputation of the three commodities in the market is increasing. Among them, microwave (b) is increasing the most significantly. While hair dryers (a) and soothers (c) are increasing, the overall trend is relatively flat with little to no increase.



(a) Confidence interval of hair drier (b) Confidence interval of microwave (c) Confidence interval of pacifier

Fig. 6. 95% confidence interval of three commodities on time axis

6 Processing of Text Comments

In the evaluation data given, each customer's evaluation not only scores the product, but also comments based on the customer's subjective opinions, which are different from cold scores, they are more vivid and reflect the image of the product in the customer's mind in the computer field also known as text sentiment analysis. As a common task in natural language processing (NLP), text sentiment analysis is of high practical value. In this chapter, we will use an LSTM model to train a classifier that can identify three emotions: positive, neutral and negative emotions and assign a corresponding score to each comment. In the previous section, we know that NE is the result of GPR regression after consumer scoring, while CT is the result of text evaluation after classification and recognition by the LSTM model.

6.1 Text Sentiment Analysis Based on LSTM

RNN (Recurrent Neural Network). This model is good at natural language modeling and it converts sentences of any length into a floating point vector of a specific dimension. The model solves the problem of short-term memory modeling of natural language sentence vectorization. LSTM is a special structural type of RNN model that adds three control units (cells) (Fig. 7). When information enters the model, the cells in the LSTM will judge the information and the information that matches the rules will be left behind and the information that does not match the rules will be forgotten.

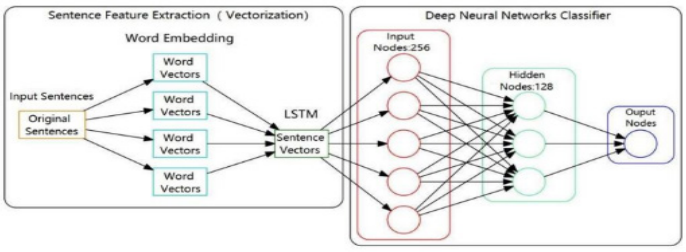


Fig. 7. RNN principle

In NLP, x is considered as a word in a sentence and y is the context word of the word, so f is a frequent language model in NLP. The goal is to determine whether the sample (x, y) conforms to the laws of natural language. word2vec comes from this idea, which only cares about the model parameters after the model is trained, uses high-dimensional vectors (word embeddings) to represent words, and places words with similar meanings in similar positions.

We divide and convert words into high-dimensional vectors, so that sentences correspond to sets of word vectors, i.e., matrices.

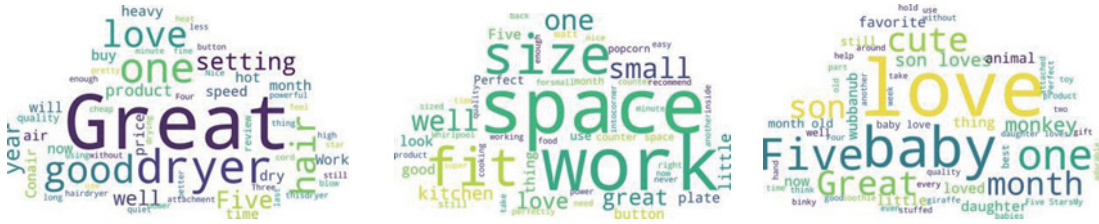
LSTM triple classification model

In addition to the two emotions, positive and negative, there is a natural emotion. We choose three categories here because the first two category models are difficult to classify neutral evaluations because we added neural emotions to make the model more practical. For the three feelings after classification, please use -1, 0, and 1 instead.

In the initial stage, we trained the model with the dataset we accumulated, and then used the model for sentiment analysis. Please note that the final criterion is CT.

(a) Emotional keywords of hair (b) Emotional keywords of microwave (c) Emotional keywords of pacifier

Fig. 8. When CT>0, emotional keywords of three products



(a) Ek2 of hair drier (b) Ek2 of microwave (c) Ek2 of pacifier

Fig. 9. When CT<0, emotional keywords of three products

We separated the three keywords of the items. Fig. 8 shows the proportion of valid keywords when the reviews are dominated by positive sentiment (CT > 0), and Fig. 9 shows the proportion of valid keywords when the reviews are dominated by negative sentiment (CT < 0).

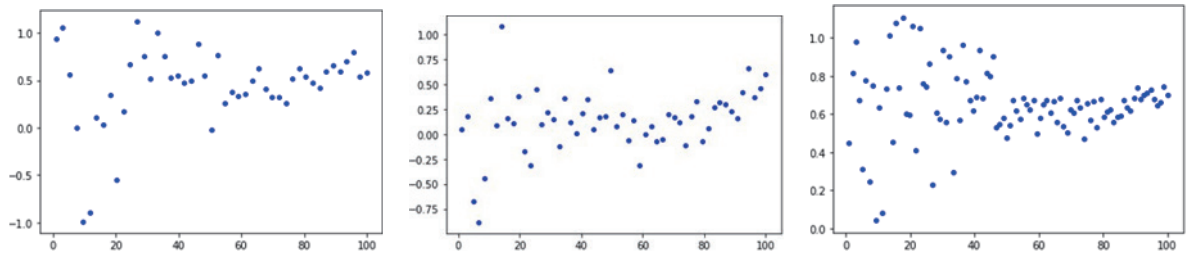
According to the word cloud, when CT>0, some words with emotional and evaluation nature appear very frequently, such as “Like, great”, which are all praising the products from the subjective perspective. However, when CT<0, consumers do not mention many words that directly reflect negative emotions in their evaluations, but rather evaluate the product from the point of view that its functions do not satisfy their needs, such as “space, size”, thus, we can obtain the hypothesis.

Theorem 8. A clever and cruel hypothesis. The name of this hypothesis comes from a study by amabile, which supports the argument that “negative evaluators are perceived as smarter, more competent, and more professional than positive evaluators.”

Since CT <0 reviews are more focused on discussions of product design and features, it provides more professional information to others who are considering whether to purchase the product, which provides them with more evidence of consumer support.

6.2 Findings from the LSTM Model

The distribution of CT ratings for each product over time is shown in Fig. 10.



(a) CT of hair drier (b) CT of microwave (c) CT of pacifier

Fig. 10. Distribution of CT scores of each commodity over time

7 Comprehensive Evaluation

7.1 The Principle of Entropy Weight Method

Let's suppose we give k indexes $X_1, X_2, ..., X_k$, where $X_i = \{x_i, x_2, ..., x_n\}$. Assuming that the standardized value of each indicator data is $X_1, X_2, ..., X_k$, then

$$Y_{ij} = \frac{X_{ij} - \min(X_i)}{\max(X_i) - \min(X_i)}. \tag{15}$$

Here we mainly allocate weights to VP and HR. The number of stars scored by users is scoreSR, the total amount of data is n.

Find the Information Entropy of Each Index. According to the definition of information entropy in information theory, the information entropy of a group of data is

$$E_j = -In(n)^{-1} \sum_{i=1}^{n} p_{ij} Inp_{ij}. \tag{16}$$

where

$$p_{ij} = Y_{ij} / \sum_{i=1}^{n} Y_{ij}. \tag{17}$$

If $p_{ij} = 0$, defines

$$\lim_{pij \to 0} pijInpij = 0. \tag{18}$$

Determine the Weight of Each Index. According to the formula of information entropy, the information entropy of each index is calculated as $E_1, E_2, ..., E_k$. Calculate the weight of each index through information entropy:

$$W_I = \frac{1 - E_i}{k - \sum E_i} (i = 1, 2, ..., k). \tag{19}$$

$$Z_t = \sum_{i=1}^{n} X_{ti} W_i. \tag{20}$$

Results of Entropy Weight Method. Because VI has a very high reference value, we fixed the weight of VI as 3, and used entropy weight method to allocate the weight of VP and HR:

$$score_{NE} = (Z_t + 3X_{VI}) score_{SR}. \tag{21}$$

The results are shown in Table 2.

Table 2. Weight distribution of VP, HR by entropy weight method

Product	VP	HR
hair dryer	0.403492	0.5811362
microwave	0.7333429	0.2666571
pacifier	0.4188638	0.5811362

7.2 Integrating Data Using a Comprehensive TOPSIS Evaluation Method Modified by the Entropy Weight Method

Algorithm 1. TOPSIS Algorithm Flow [10]

Input: Original data set $X = \{x_1, x_2, ..., x_n\}$; Weight of each index $w = (w_1, w_2, ..., w_m)$

Output: Evaluation results of TOPSIS of each data sample

1: The index attribute of the original data set is isotropic X'

2: Constructing normalized matrix $Z = \{z_1, z_2, ..., z_n\}$ after vector normalization

3: **for** Z_i: each column of Z **do**

4: The i-th dimension of the worst scheme Z^- The minimum value of the Z_i element

5: The i-th dimension of the best scheme Z^+ The maximum value of the Z_i element

6: **end for**

7: **for** z_i Z **do**

8: The approach degree between Zi and the optimal scheme D^+ Formula (6.5)

9: The approach degree between Zi and the worst scheme D_i^- Formula (6.6)

10:The close degree between Zi and the optimal scheme D_i Formula (6.7)

11: **end for**

12: Sort by C_i size

A Comprehensive TOPSIS modified by entropy weight method uses the distance scale to measure the sample gap. Using the distance scale, we need to process the index attributes in the same direction.

A Comprehensive TOPSIS Calculation Result Modified by Entropy Weight Method. According to the order of C_i size and the evaluation results, we intercept part of the data and show it in Table 3.

Table 3. A comprehensive TOPSIS calculation result modified by entropy weight method

Item	D_i^+	D_i^-	C_i	Ranking results
i=1	0.569	0.415	0.421	41
i=2	0.423	0.555	0.568	9
i=3	0.839	0.142	0.145	49
...	...	...	...	...

So far, we have combined NE and CT, identified the most informative ratings and reviews, identified a data measure Ci, and recorded its name as Comprehensive Evaluation (CE).

7.3 CE Means Product Success

Let's take the dataset of hair dryers as an example. After analyzing the entire dataset, we found that some products were purchased multiple times, leading in sales field. Whether the product can be successful in the market or not, the comprehensive evaluation of the product by consumers is only one of the factors. As an enterprise, Sunshine Company takes profit as its core, and sales volume is the most important indicator.

In order to study and predict whether the product will be recognized by the market or not, we decided to combine the customer's comprehensive evaluation (CE) with the product sales volume for analysis.

First, we draw a dendrogram with the product ID as the size and CE as the measurement, as shown in Fig. 11. We summarize CE, which is to unify CE and sales. Fig. 11 gives us an overview of the market acceptance of all hair dryer styles, and it also highlights five hair dryer styles: B0009XH6TG, B00132ZG3U, B003V264WW, B0000500MZZ and B000R80ZTQ.

To study the market's reaction to the product in more detail, we used filled bubbles to visualize the data and select exactly the five most recognized products, as shown in Fig. 12.

To further study the change of market recognition of the five commodities, we added the time dimension, as shown in Fig. 13.

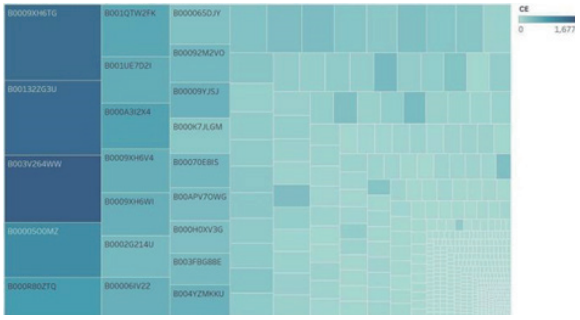
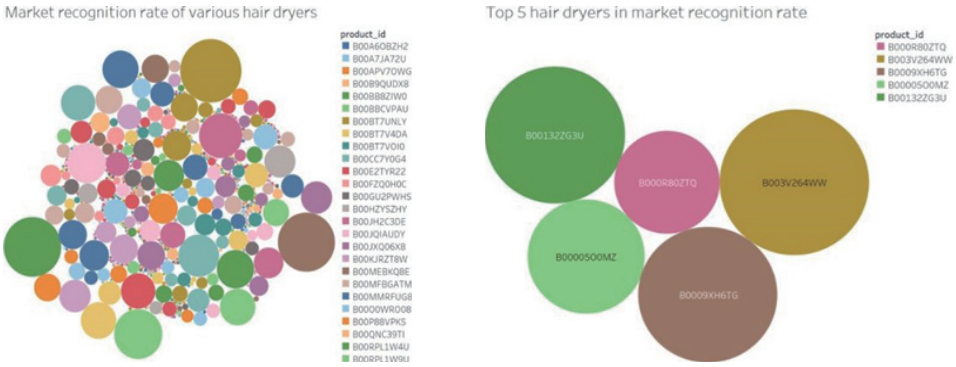


Fig. 11. The dendrogram obtained by summing CE



(a) Market recognition rate of various hair dryers (b) Top 5 hair dryers in market recognition rate

Fig. 12. Packed bubbles visualization

(a) Hair dryers top5 (b) Microwave top6 (c) Pacifier top6

Fig. 13. Top 5 market recognition rates over time

By analyzing the above data and charts, we can conclude that the potential successful products are as shown in Table 4.

Table. 4. Potential successful products

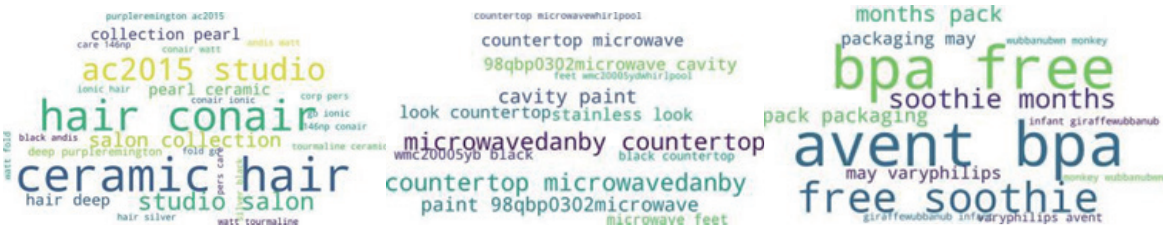
Item	Product ID	Shorthand
Hair Dryer	B000R80ZTQ, B003V264WW, B00005O0MZ, B0009XH6TG, B00132ZG3U	H1, H2, H3, H4, H5
Microwave	B004NXUJ60, B005GSZB3M, B005GSZB7I, B0052G14E8, B0055UBB4O, B005GSZB9Q	M1, M2, M3, M4, M5, M6
Pacifier	B001FGL9X0, B003CK3LDI, B003PCYMP4, B0028IDXDS, B0045I6IA4, B0045I6IAY	P1, P2, P3, P4, P5, P6

7.4 Product Design Improvement Suggestions

Based on the potential successful products, we extract key words from the corresponding product title (PD), analyze the reasons why they are recognized by the market, and put forward suggestions for improvement of the product (Fig. 14).

We found that customers have a unified buying habit, that is, to identify the brand; Other key words are different according to the product function and for different people. Choose hair dryer to pay attention to anion, calcium carbide and other.

Functions: choose microwave to pay attention to size, noise and other design; choose pacifier to pay attention to BPA Free, animal and other materials and appearance, which needs to be more friendly to children.



(a) Function keywords of hair dryer (b) Function keywords of microwave (c) Function keywords of pacifier

Fig. 14. Function keywords of three products

8 Strengths and Weaknesses

The model used in this article has some strengths and weaknesses, which are analyzed as follows:

Strengths. The three-category LSTM model we use is better complemented by neutral reviews and is more realistic, while better analyzing unstructured data.

Our PSO-optimized Gaussian regression can quantify forecast uncertainty in a principled manner.

Weaknesses. The computational complexity of the Gaussian process regression we use here increases with the amount of data, which is about O(n3), and it is only suitable for problems with small and medium samples.

LSTM models require massive amounts of data to train, and if there is not enough data, the effect will be even more worrying.

9 Conclusion

The purpose of this paper is to conduct large-scale analysis with a large amount of different types of data, including text data and rating data, to truly and objectively recommend the most cost-effective products for customers; At the same time, for merchants, our comments on consumers analyzed and extracted keywords, and put forward suggestions for future product optimization direction and marketing strategy designation. Merchants can carry out more in-depth optimization based on the high-frequency words obtained from the model analysis, and conduct targeted marketing in this direction to achieve better sales. Finally, we briefly review the running process of the model:

First, we use PSO-optimized Gaussian Process Regression to fit its time-series forecast and infer the future market performance of different product models for the same commodity from the forecast results.

In addition, we examine user feedback on the product from two perspectives. On the one hand, we use LSTMs to classify and annotate text data, and analyze unstructured data. On the other hand, we compare the keywords on the labels of mainstream products in the market to the average of similar products.

Finally, we use entropy-weighted modified TOPSIS to synthesize text-based and rating-based metrics and derive the final evaluation model. Based on the analysis of its time series performance, the product outlook is predicted.

Regarding how to better improve the method proposed in this paper, in further research in the future, we can replace the LSTM model with the transformer framework to achieve better data processing results in this paper.

10 Acknowledgement

This work is supported by the National Social Science Foundation of China under Grant 17BGL047.

References

[1] Y.-Y. Zeng, Analysis and application of customer data in online store operation, China Storage and Transport (11)(2021) 105-106.
[2] Y.-H. Li, An Empirical study on digital consumer behavior of Chinese residents--based on micro survey data, Journal of Commercial Economics (12)(2021) 43-46.
[3] Z.-Y. Gao, J.-N. Hao, Regional consumption of online shopping population based on e-commerce platform data Characteristic statistical analysis, Journal of ShangQiu Normal University (12)(2020) 49-52.
[4] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proc. ICNN'95-International Conference on Neural Networks, 1995.
[5] W. Hu, G.-G. Yen, X. Zhang, Multiobjective particle swarm optimization based on Pareto entropy, Ruan Jian Xue Bao/ Journal of Software 25(5)(2014) 1025-1050.
[6] Z.-K. He, G.-B. Liu, X.-J. Zhao, M.-H. Wang, Overview of Gaussian process regression, Control and Decision 28(8)(2013) 1121-1129.
[7] W.-L. Cao, L.-L. Kang, Particle swarm optimization for adaptive hyper-parameters acquisition of Gaussian process regression, Journal of Hefei University of Technology (Natural Science Edition) (11)(2019) 1479-1484.
[8] Q.-Y. Wang, H. Xun, Entropy method for major hazards emergency rescue, Journal of Nanjing University of Technology (Natural Science Edition) 33(3)(2011) 87-92.
[9] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9(8)(1997) 1735-1780.
[10]R.-Q. Sun, Research on the price trend prediction model of US stock index based on LSTM neural network, [dissertation] Beijing: Capital University of Economics and Business, 2016.
[11]C.-L. Hwang, K. Yoon, Methods for multiple attribute decision making, in: multiple attribute decision making, Springer, Berlin, Heidelberg, 1981 (58-191).
[12]T. Zhang, H.-C. Yan, Optimization and application of multi-attribute decision algorithm based on entropy weight method, Journal of North China University of Science and Technology (Natural Science Edition) (1)(2022) 82-88.