Amazon Platform Shopping Review Data Analysis

Xiaoqi Xue*

Business school, Northwest University, Xi'an, Shanxi, China

*Corresponding author e-mail: xuexiaoqi@stumail.nwu.edu.cn

Abstract. This paper studies the development of three products on Amazon platform. This paper analyzes the text emotion of the comments on Amazon platform, quantifies the emotional score of each comment based on LSTM, selects relevant indicators of the comment data, establishes a comprehensive evaluation model of the comprehensive praise of the comment, and recommends the products with the greatest potential success for businesses based on the comment data.

Keywords: LSTM Neural Network DID TF-IDF Sentiment Analysis.

1. Introduction

1.1 Background Investigations

In the online marketplace it created, Amazon provides customers with an opportunity to rate and review purchases. Sunshine Company is planning to introduce and sell three new products in the online marketplace: a microwave oven, a baby pacifier, and a hair dryer. Sunshine Company are time-based patterns in these data, and whether they interact in ways that will help the company craft successful products.

1.2 Journals reviewed

This paper mainly:

Identify data measures based on ratings and reviews that are most informative for Sunshine Company to track.

Identify and discuss time-based measures and patterns within each data set that might suggest that a products reputation is increasing or decreasing.

Determine combinations of text-based measure and ratings-based measures that best indicate a potentially successful or failing product.

2. Products' Reputation Analysis Model

2.1 Establishment of Emotional Model for Comments

we first slice the comments according to the words and transform the words into word vectors. We establish a neural network emotion training algorithm based on LSTM. Then we use the reputation comprehensive evaluation model of the first question to analyze the situation of the products so as to quantify the emotion degree of each comment.

2.1.1 Extraction of Word Vector

We hope to extract text emotion from comments and reviews for analysis. For the prepro- cessed text, morphemes are extracted to establish a lexicon S, and if there are n words in the lexicon, the word vector in the form of one-hot can be expressed as

$$s = \underbrace{[0,0,..0,1,0,\dots,0]}_{n}$$

In order to avoid dimension explosion caused by too long lexicon, Word2Vec method is adopted here, and one-hot sparse word vector mapping is called an N-dimensional dense vector based on one-layer neural network (CBOW)[1]

Establishing a lexicon matrix and compressing words into a high-dimensional dense matrix：

$$G = w_1 \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{m1} \end{bmatrix} + w_2 \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{m2} \end{bmatrix} + \ldots + w_m \begin{bmatrix} x_{1m} \\ x_{2m} \\ \vdots \\ x_{mm} \end{bmatrix}$$

2.2 Establishment of LSTM Recurrent Neural Network

In natural language processing, recurrent neural network or cyclic neural network is usually used. The input in matrix form is encoded into one-dimensional vectors of lower dimensions while retaining most useful information. RNNs pays attention to the reconstruction of adjacent positions, which shows that RNNs is more persuasive than CNN [2] (language is always com- posed of adjacent words, adjacent words form phrases, and adjacent phrases form sentences, which requires effective integration of information of adjacent positions, or reconstruction). So, we chose LSTM [3] neural network under cyclic neural network to model. And a deep learning emotion scoring model based on Word2Vec transformed word vector and LSTM neural network is established.

The principle of neural network is that recurrent neural network is a kind of cyclic neural network, which expands the information of the previous moment according to the time sequence to be used as the feedback of the next node to process the sequence information.

The neural network is divided into three layers: input layer, hidden layer and output layer. The input layer X is a vector, which is each statement after standardization; S is a vector that represents the value of the hidden layer;U is the weight matrix from the input layer to the hidden layer; O is also a vector that represents the value of the output layer;V is the weight matrix from the hidden layer to the output layer. The weight matrix W is the weight of the previous value of the hidden layer as the input of this time. The output layer is a vector

$$[x_1', x_2', \ldots, x_n']$$

Xn indicates the "memory" of each morpheme in the sentence to favorable comments, that is, the probability that the word is most likely to appear after it. take

$$max\{x_1', x_2', \ldots, x_n'\}$$

As the emotional memory h(t) of the sentence, there are:

$$\begin{cases} h_t \geq 0.5, \text{positive emotion ,command this product} \\ h_t < 0.5, \text{nagtive emotion ,dont command this product} \end{cases}$$

Among them:$h_t \subset [0,1]$

The idea of long-short memory network is based on RNN, but the hidden layer of RNN has only one state, H, which is very sensitive to short-term input. Then, if we add another state, C, to preserve the long-term state.

At time t, LSTM has three inputs: the input value xt of the network at the current time, the output value $h_{t-1}$of LSTM at the previous time, and the cell state C_{t-1}at the previous time.

There are two outputs of LSTM: LSTM output valueht at the current time and cell state ct at the current time.

Here, three control switches are used to control the memory level of the neural network. The first switch, the forgetting gate, is responsible for controlling the continuous storage of the long-term state C; The second switch, the input gate, is responsible for controlling the input of the immediate state to the long-term state C; The third switch, the output gate, is responsible for controlling whether the long-term state C is taken as the output of the current LSTM.

2.2.1Forward Calculation

In order to control the input and output information of neurons, we use activation function to control the three gates. Sig function can map the input information from the interval to the interval of [0,1], and the function is relatively smooth without abrupt changes.

$$\sigma(z) = y = \frac{1}{1 + e^{-z}}$$

The sig function is used as the activation function for the control of the forgetting gate, that is, the output ft of the forgetting gate can be expressed as:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

Wf is the weight matrix of the forgetting gate, h_{t-1} is the output of the previous moment, x_t is the input of the neural network.bf is an offset item to divide the soft boundary of classification. Since the dimensions of the weight matrix are consistent, the above formula can be further rewritten as:

$$f_t = \sigma(W_{fh}h_{t-1} + x_t W_{fx} + b_f)$$

The input gate determines how much input xt of the network is saved to the cell state ct at the current time. The sig function is also used to control the input gate, then the outputit of the input gate at time t can be written as

$$i_t = \sigma(W_f[h_{t-1}, x_t] + b_i)$$

The cell state of the current input is calculated according to the previous output and the current input.

$$\widetilde{c_t} = tanh(W_c[h_{t-1}, x_t] + b_c)$$

The cell state ct at time t consists of the cell state of the previous time and the output cell state of the previous time. The specific method is to multiply the forgetting gate by the element, multiply the input gate by the element by the current input cell state, and add up the two products.

$$c_t = f_t c_{t-1} + i_t \widetilde{c_t}$$

The output gate controls the influence of long-term memory on the output at the current moment, and uses sig function to control, namely:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_0)$$
$$h_t = o_t\, tanh(c_t)$$

The final output htof LSTM is jointly determined by the output gate and the cell state. The final output ht can be expressed as:

So far, the network construction and forward calculation have been completed.

2.2.2 LSTM Training Algorithm

LSTM training algorithm uses the algorithm of error reverse transfer for training. Assuming that the error between the output value and the original value of the neural network is e, we use the square loss function, that is

$$E = (y - h_t)^2$$

We define the error term at time t δt is

$$\delta_t = \frac{\partial E}{\partial h_t}$$

Then four weight inputs to LSTM ft, it, ct, ot ,

the error is respectively defined as

$$\delta_{t,f} = \frac{\partial E}{\partial net_{f,t}}$$
$$\delta_{t,i} = \frac{\partial E}{\partial net_{i,t}}$$
$$\delta_{t,\tilde{c}} = \frac{\partial E}{\partial net_{\tilde{c},t}}$$
$$\delta_{t,o} = \frac{\partial E^{,t}}{\partial net_{o,t}}$$

Net is the input of each weighted term at time t . The above four errors can befurther reduced to:

$$\delta_{t,o} = \delta_t\, tanh(c_t) o_t(1 - o_t)$$
$$\delta_{f,t} = \delta_t o_t(1 - tanh(c_t)^2) c_{t-1} f_t(1 - f_t)$$
$$\delta_{i,t} = \delta_t o_t(1 - tanh(c_t)^2) \widetilde{c_t} i_t(1 - i_t)$$
$$\delta_{\widetilde{c_t},t} = \delta_t o_t(1 - tanh(c_t)^2) i_t(1 - \widetilde{c_t}^2)$$

The partial derivative of the timet error with respect to the weight can be obtained to obtain the weight gradient of the four weighted inputs

$$\frac{\partial E}{\partial W_{o,t}} = \frac{\partial E}{\partial net_{0,t}} \frac{\partial net_{o,t}}{\partial W_{o,t}} = \delta_{o,t} h_{t-1}$$
$$\frac{\partial E}{\partial W_{f,t}} = \frac{\partial E}{\partial net_{f,t}} \frac{\partial net_{f,t}}{\partial W_{f,t}} = \delta_{f,t} h_{t-1}$$
$$\frac{\partial E}{\partial W_{i,t}} = \frac{\partial E}{\partial net_{i,t}} \frac{\partial net_{i,t}}{\partial W_{i,t}} = \delta_{i,t} h_{t-1}$$
$$\frac{\partial E}{\partial W_{c,t}} = \frac{\partial E}{\partial net_{\tilde{c},t}} \frac{\partial net_{\tilde{c},t}}{\partial W_{c,t}} = \delta_{\tilde{c},t} h_{t-1}$$

Add up the gradients at each moment and get Wc,x, Wi,x, Wf,x, ∂W0,x. The total gradient from the start of training to time t is further simplified

$W_{c,x}, W_{i,x}, W_{f,x}, \partial W_{0,x}$The calculation formula is:

$$\frac{\partial E}{\partial W_{0,x}} = \frac{\partial E}{\partial net_{0,t}} \frac{\partial net_{o,t}}{\partial W_{x}} = \delta_{o,t} x_t$$
$$\frac{\partial E}{\partial W_{f,x}} = \frac{\partial E}{\partial net_{f,t}} \frac{\partial net_{f,t}}{\partial W_{f,x}} = \delta_{f,t} x_t$$
$$\frac{\partial E}{\partial W_{i,x}} = \frac{\partial E}{\partial net_{i,t}} \frac{\partial net_{i,t}}{\partial W_{ix}} = \delta_{i,t} x_t$$
$$\frac{\partial E}{\partial W_{c,x}} = \frac{\partial E}{\partial net_{\tilde{c},t}} \frac{\partial net_{\tilde{c},t}}{\partial W_{c,x}} = \delta_{\tilde{c},t} x_t$$

The training stops when the given precision is finally reached through continuous iteration. The final output of the neural network is

h(t) = WX

In the output set of vectors, the elements at each position represent the memory of the words for favorable comments, and the value with the highest probability represents the emotion h(t) represented by the comment, i.e., favorable comments.

2.2.3Construction of Training Set

In general, the closer the comment score is to a 5-star comment, the higher the positive degree of emotion, and the closer the comment score is to a 1-star comment, the higher the negative degree of emotion, i.e., the positive degree of emotion is positively correlated with the comment score. Based on this, we randomly selected about 5,000 comments with a score of 1 star and a score of 5 stars from the three data sets given, set 5 stars as number 1 and 1 star as number 0, and manually marked out the data with too large a difference between the score and language emotion to construct the training set. The input of LSTM neural network is the word vector converted from training set comments through Word2Vec algorithm, and the output layer is the converted score, i.e. 0 or 1, so that the final training result of the model is an emotional value between 0 and 1.

2.3 2.3Display of Network Training Parts Results

We use tensorflow to build a neural network on python, and do text sentiment analysis on the titles and contents of reviews of the three products to calculate the favorable comments of each product. Some calculated results are as follows:

Table 1. Table of Comment Emotional Level

review body	The points of emotion	Emotional evaluation
Works great!	0.930513	positive
This dries my hair faster	0.950125	positive
Love this dryer!	0.901746	positive
styling hair in style	0.917792	positive
Excellent dryer.	0.954028	positive
Gets extremely hot	0.239887	negative
I found everything goes well except	0.17608	negative

2.4Solution to Comprehensive Evaluation Model of Product reputation

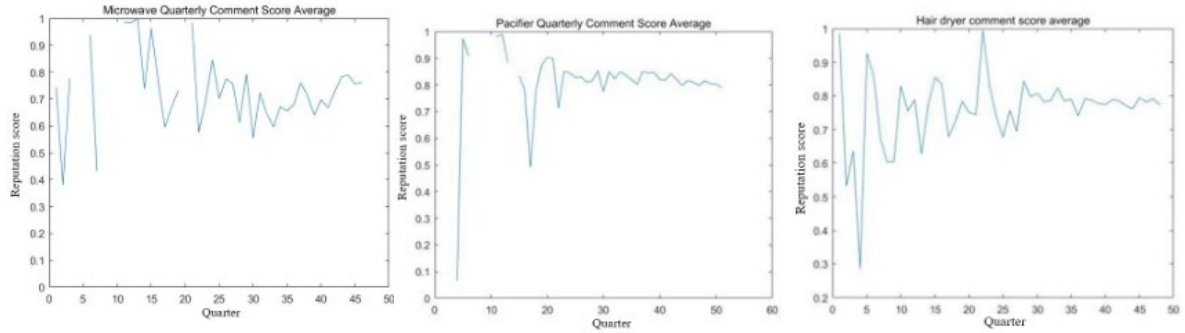
Due to the different number of comments hellpful votes, we believe that the objective degree of comments is different, so all comments hellpful votes are standardized [0,1] and used as weights to re-weight the emotion and stars of each comment.

Take the average emotional score of all comments in the quarter as the score of the product in the quarter. The normalized value is substituted into the reputation comprehensive evaluation model of 1 to calculate the recommendation degree of the three products. The line chart is shown in the following figure: The comprehensive evaluation model of product recommendation can be written as follows:

$$f = \sum_{i=1}^{5} w_i x_i$$

The company can complete product tracking through the user's comment information. If we know the user's attitude through emotional analysis of comments, for example, the emotional quantification of the comment "I love travel blow dryers because they are easy to lift" is 95.6%, which indicates that the user is satisfied. The company can also track the product status through the product's reptilization comprehensive score.

The reputation comprehensive scores of the three products are obtained as follows:



(a) Microwave (b) Pacifier (c) Hair Dryer

Figure 1. The Reputation Comprehensive Scores of the Three Products

The recommendatory degree of microwave oven is low at the beginning of sales, about 0.5. In recent years, the recommendatory degree has increased, about 0.7.

The recommendatory degree of pacifier is relatively high at the beginning of sales, generally fluctuating at about 0.9. In recent years, the recommendatory degree has a downward trend, maintaining at about 0.7.

Hairdyer's reptilization recommendation is high at the beginning of sales, but fluctuates greatly. In recent years, the recommendation tends to be stable, maintaining at about 0.7.

3. The Establishment of Time Series Model

3.1 The Establishment of Time Series Model

The three products have a trend of increasing or decreasing with time. They belong to non-stationary time series, and are analyzed by moving average method. Exponential smoothing method is an improved method of moving average method, which gives different weights to different historical data. However, the predicted value of the first exponential smoothing method usually lags behind the actual value. In order to eliminate this effect, the third exponential smoothing method is used for time series analysis.

For a series of report time series, if is the smoothing parameter of the -th index and is the influence of the observation value of the -th period on the future trend, then the observation value of each period can be weighted in time order as the prediction value, then the recurrence of the fourth index smoothing can be expressed as

$$s_t^{(1)} = \alpha y_t + (1-\alpha)s_{t-1}^{(1)}$$
$$s_t^{(2)} = \alpha {s_t}^{(1)} + (1-\alpha){s_{t-1}}^{(2)}$$
$$s_t{}^{(3)} = \alpha {s_t}^{(2)} + (1-\alpha){s_{t-1}}^{(3)}$$
$$s_t{}^{(4)} = \alpha {s_t}^{(3)} + (1-\alpha)s_{t-1}^{(4)}$$

Then the prediction model oft + m period can be expressed as

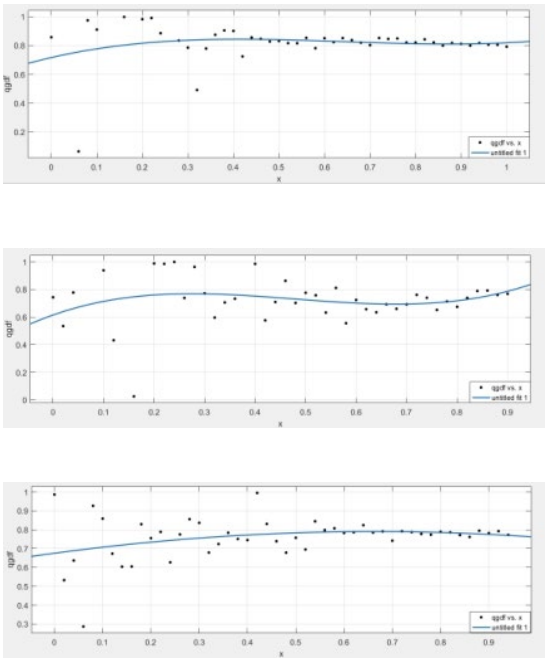
$$y_{t+m}^{'} = a_t + b_t x + c_t x^2 + d_t x^3 + e_t x^4$$

At this time there are

$$a_t = 3s_t^{(1)} - 3s_t^{(2)} + s_t^{(3)}$$
$$b_t = \frac{\alpha}{2(1-\alpha)^2}\left[(6-5\alpha)s_t^{(1)} - 2(5-4\alpha)s_t^{(2)} + (4-3\alpha)s_t^{(3)}\right]$$
$$c_t = \frac{\alpha}{2(1-\alpha)^2}\left[s_t^{(1)} - 2s_t^{(2)} + s_t^{(3)}\right]$$

3.2 Analysis of Model Results

Time Series Model results of three products are as follows:



(a) Microwave (b) Pacifier (c) Hair Dryer

Figure 2. Time Series Model Results

The time-varying law of pacifier reputation changing is

$$f(x) = 0.674x^3 - 1.301x^2 + 0.7296x + 0.7156$$

Pacifier's reputation has been rising gradually in recent years, but it has been relatively stable and fluctuated slightly, indicating that the reputation is relatively good.

The time-varying law of the microwave reputation is

$$f(x) = 2.298x^3 - 3.29x^2 + 1.29x + 0.6144$$

In recent years, the reputation of microwave has increased, but the overall trend fluctuates greatly.

The time-varying law of the hairdryer reputation is as follows:

$$f(x) = -0.00211x^3 - 0.2586x^2 + 0.3456x + 0.6745$$

The fluctuation of hair dryer's reputation in the previous years was relatively large, but was relatively slow in recent years. Generally speaking, the reputation was relatively stable.

3.3 Test of Time Series Model Results

We use R^2 significance to test the goodness of fit of regression equation. R^2 significance measures the fitting degree of the whole regression equation and expresses the overall rela- tionship between dependent variables and all independent variables.

The significance test is to decompose the fitting error into the sum of residual squared sum and regression squared sum. R^2 is equal to the ratio of the sum of regression squared sum to the total squared sum, that is, the percentage of variance of dependent variable that can be explained by regression equation.

Among the total errors between the actual value and the average value, the regression error and the residual error are the relationship between the two. Regression error measures the goodness of fit of linear model from the front, and residual error determines the goodness of fit of linear model from the back. The value of R^2 indicates the goodness of fit.

The goodness of fit of the above models is:

Table 2. Table of Model fit Goodness

Product	R^2 of reputationă
pacifier	0.55
hair dryer	0.61
microwave oven	0.5

It can be seen that the goodness of fit of the simulation is more than 50%, indicating that the fitting effect is better.

4. Establishment of Optimal Combination Product Model

The question requires us to research the best combination of potential success ratings and reviews based on the review text and star rating, hoping that we can determine the most successful or failed model under a certain product. Based on the reputation comprehensive evaluation model of question 1, we established a multi-objective optimization model with comments and ratings of different products as independent variables and reputation evaluation value F as optimization objective

4.1 Data Preprocessing

We further refined the data set by dividing different types of products into quarters according to the evaluation time in order to better reveal the time characteristics.

4.2 Analysis of Constraint Conditions

Since all the products selected are products with good development trend, the potential success degree m is introduced. M can be expressed as

$$m = f'$$

Where f represents the reptilization score of the product. When a product's score shows an upward trend in four quarters and always maintains a high reptilization, we consider the product to be potentially successful. According to this, the data set is constrained under the following conditions:

$$\begin{cases} f > 0.5 \\ f' > 0 \end{cases}$$

4.3 Establishment of Multi-objective Optimization Model

Assuming that the score is x1, the text emotion is x2, and the corresponding model weights of 1 comprehensive evaluation are w1 and w2 respectively. since they are calculated on the determined feasible region, when a group (x1 and x2) is given, the calculated values of the other three indicators are given as c, then the product representation comprehensive score of this type is:

$$w_1 x_1 + w_2 x_2 + c$$

Then the optimization model with the score x1 and the text emotion x2 as independent variables and the maximum comprehensive score of product reputation as the objective can be written as follows:

$$\max\left(w_1 x_1 + w_2 x_2 + c\right)$$

$$\text{s.t.} \begin{cases} x_1 > 0 \\ x_2 > 0 \\ \left(x_1, x_2\right) \in Q_1 \\ f > 0.5 \end{cases}$$

Traversing the feasible region, the best combination of scoring comments for each product is obtained, and the combination is potentially successful. For potential failure products, a similar method is used to find the potential failure setQ2, and an optimization model is established to minimize the comprehensive score of reputation.

4.4 Model Calculation Results

The data sets of the three products are solved respectively, and the model results are obtained by python programming.

Table 3. Comparison of Good and Bad Commends in Pacifier

Success or Failure	Product Parent	Star Rating
Success	386489526	5
Failure	658527424	1
Success or Failure	Product Parent	Star Rating

Table 4. Comparison of Good and Bad Commends in Hairdryer

Success or Failure	Product Parent	Star Rating
Success	258419037	5
Failure	888313825	1
Success or Failure	Product Parent	Star Rating

Table 5. Comparison of Good and Bad Commends in Microwave

Success or Failure	Product Parent	Star Rating
Success	692404913	5
Failure	771401205	1

5. Conclusion

We study specific issues such as product development trend, optimal product mix, and comment on emotion. We first establish a training set corresponding to comments and comment stars, extract word vectors by word2vec method, and build LSTM neural network to quantify comment emotion into a value in [0,1]. And we divide the comprehensive score into three levels. And the result of clustering of the evaluation results of the products shows that the classification is reasonable. The company can track user's satisfaction through the comprehensive scores and comments of each product.

We also set up a time series analysis model for the quarterly reputation score of prod- ucts,using the four-time exponential smoothing method to fit and obtain the seasonal variation law of products. It can be found that the reproduction of the three products fluctuated greatly in the early stage of sale. In recent years, pacifiers and hair dryers have tended to stabilize, while microwave have tended to rise, with values above 0.5. The goodness of fit of the model is above 0.5, and the model fitting is more accurate.

Further, we select the products with higher potential success as the constraint conditions, and use the comment text emotion and average star rating as the independent variables to establish a single-objective optimization model for reputation comprehensive score. Finally, the best comment and average star rating combination of each product are solved.

References

[1] Jiaqing Liu, xi hai Guo, and yu Sun. Optical power prediction method based on double depth neural network. The Optical power prediction method is based on a double depth neural network, pages 1–9.

[2] Bingjian Lu. Short-term wind speed prediction model based on eemd and lstm. (03):43–48,2020.

[3] YueChen and Qingye He. Influencing factors of buyer evaluation decision: an empirical study based on taobao.com. 8(02):66–87, 2014.