Product Reviews Analysis of E-commerce Platform Based on Logistic-ARMA Model

Yingfei Li[1,*], Jingcheng Shi[2], Fan Cao[1], Andi Cui[1]
[1]School of Maritime Economics and Management, Dalian Maritime University, Dalian, China
[2]School of Information Science and Technology, Dalian Maritime University, Dalian, China
*Corresponding author: liyingfei@dlmu.edu.cn

2021 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS) | 978-1-6654-4130-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICPICS52425.2021.9524238

***Abstract*—With the vigorous development of 5G Internet technology, people's demand for purchasing goods through e-commerce platform is daily increasing. E-commerce platforms now mostly launch commodity evaluation function, which will undoubtedly affect other consumers' choice of goods, but it can also provide a reference for platforms to adjust sales strategies at any time. At present, many large transnational e-commerce platforms are committed to scientific commodity evaluation trend prediction, so as to timely understand the popularity of commodities based on the existing data, so as to maximize their own interests. Amazon's platform provides consumers with product ratings and text-based reviews. We collected the recent evaluation data of the three commodities on the platform, and put forward reasonable analysis and prediction. Firstly, the text-based evaluation and commodity rating are processed quantitatively to form a two-dimensional evaluation vector, and the part that is inconsistent with the evaluation emotion is eliminated to realize data cleaning. We use the Logistic Regression Equation Model to complete the function fitting based on time variables for the three commodities, and achieve the statistical analysis of the existing data. In addition, we draw on the principle of filtering out noise outside the band in signal processing, and make innovative use of Trend Moving Average Model (ARMA) to smooth the function to eliminate the impact of various random events on commodity evaluation as much as possible, so as to realize the prediction of the positive Trend of future evaluation. In this paper, an e-commerce platform based on customer's evaluation of products to achieve a high sensitivity, constructive project for future product sales forecast.**

Keywords—Logistic-ARMA Model, Data Quantized, Function Fitting

I. INTRODUCTION

With the popularization of mobile Internet technology to thousands of households, most people have experienced the advantages and convenience of online shopping, but also increase the sales competition among goods[1].

Customer comments in online shopping can provide companies with intuitive experience and timely feedback on products, and help enterprises adjust sales strategies in time[2], so as to improve product quality and meet market demands. However, most online reviews do not provide direct and effective information for the company.

Amazon provides customers with the opportunity to rate and review purchases simultaneously. Customers can express their Level of satisfaction with a product in the module called "Star Ratings", and submit text-based messages in the module called "Reviews", while other purchasers can submit ratings on these Reviews for being helpful or not in the module called "Helpfulness Rating" [3].

We conducted quantitative data processing and analysis on the star rating and text rating of three products on the Amazon platform: laptop, air conditioner and lamp, so as to propose some constructive sales improvement plans.

Since Amazon allows buyers to make star rating and text rating of products, we tend to find out the relationship between rating and text.

We use Logistic Regression Equation Model for function fitting to get the change of evaluation index over time.

Finally, we design a mixed prediction model based on rating and text, which can analyze and obtain the future evaluation trend of the three commodities, and then judge the sales changes of the commodities.

II. MODEL CONSTRUCTION

Firstly, we divide the star ratings into five categories，the rule is shown in Table I.

TABLE I THE RULE TO CLASSIFY THE STAR RATINGS

level of star rating	※	※※	※※※	※※※※	※※※※※
point	**1**	**2**	**3**	**4**	**5**

Then we classify the review-bodies into five groups according to their scores of positivity, the rule is shown in the Table II.

TABLE II THE RULE TO CLASSIFY THE REVIEWS

Positivity score	$[0,0.2)$	$[0.2,0.4)$	$[0.4,0.6)$	$[0.6,0.8)$	$[0.8,1)$
point	**1**	**2**	**3**	**4**	**5**

Some purchasers who are unfamiliar with the review system may give a five-star rating but expressing his or her dissatisfaction in the review, vice versa. Such data would provide little information to the company, so we delete such data from the datasheet.

Then we draw a three-dimensional map to visualize the distribution of the comment, the map is shown in Figure 1. In the map, the x-axis refers to the point of star rating, the y-axis refers to the point of the review positivity, the z-axis refers to the matching rate.

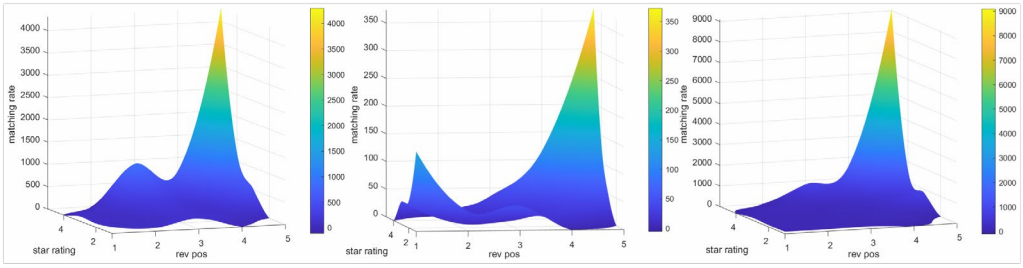

Figure 1 Comment Distribution of the Laptop, the Air Conditioner, and the Lamp

We assume that if a review fits closely with the star rating, it means that the purchaser writes the comment objectively. That is to say, the dots located near the point $(1,1)$ and $(5,5)$ would provide the company with more useful information.

III. The Change of Reputation

A. Logistic-ARMA Model

The sale of a product would be affected by the reviews presented by former purchasers. When a customer browses the online reviews, he or she would first see the latest comments, so his or her consuming behavior would be mostly influenced by nearer reviews. That is to say, reviews have timeliness [4].

Apart from the latest reviews, some particular reviews would also catch a costumer's eyes. For example, reviews with many helpful votes and reviews proposed by "Amazon Vine Voices".

Accordingly, we measure the impact that the time and the value of a review have on a product's reputation and its future sale.

B. Logistic Regression Equation Model

Time-series refer to a sequence of values of the same statistical indicator in the order in which they occur [5].

To find out whether a product is designed successfully, we analyze how the reviews change over time. Here we use the processed data, where the useless information has been deleted.

From the three-dimension-maps above, we find that the percentage of positive reviews changes during the period given, so we use the number of positive reviews to analyze the change. Since the period given is long, we count the number of positive reviews of each quarter.

According to common sense and widely used online shopping methods, we quantify the positive reviews by the following rules:

a) If the star rating and the level of positivity of the review are both higher than 3, then it is a valid positive review, and the symbol of the valid positive review $P_{oi}=1$, otherwise $P_{oi}=0$.

b) If the review is from an "Amazon Vine Voices", then we regard it highly influential, and we assume the impact index $v=2$, otherwise $v=1$.

c) If the review is from a customer who has truly bought the product, then we regard his or her review dependable, and the dependability $b=1$, otherwise $b=0.5$.

Thus we can define the quantified positive review as the following equation:

$$P_{ti}=v\cdot b\cdot P_{oi} \quad (1)$$

Since there are concrete dates in the datasheet, it can be seen as a discrete time-series. We analyze the relationship between the time and the number of positive reviews of each product. The result is shown in Fig 2.

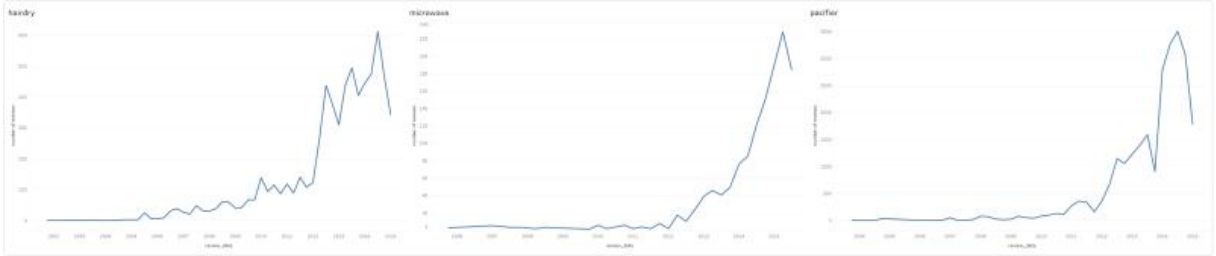

Figure 2 The Relationship Between the Time and the Positive Review

From Fig 2, it can be seen that the number of reviews grew overall. To find a more concrete relationship, we use the Unary Logistic Regression Model to analyze the time and the number of positive reviews each day.

C. Trend Moving Average Model (ARMA)

Based on the simple Moving Average Model, we consider lag bias when there is a straight increase or decrease in the time-series. Therefore, we modify the simple model using Quadratic Moving Average [7]. We use the law of Average Moving Lag Bias to construct a model to predict the line trend.

The average number of motion is

$$M_t^{(1)}=\frac{1}{N}(y_t+y_{t-1}+\ldots+y_{t-N+1}) \quad (2)$$

The formula of the second moving average based on the

former one is as follows:

$$M_t^{(2)} = \frac{1}{N}(M_t^{(1)} + ... + M_{t-N+1}^{(1)}) = M_{t-1}^{(2)} + \frac{1}{N}(M_t^{(1)} - M_{t-N}^{(1)}) \quad (3)$$

Then we discuss how to construct the line trend prediction model.

We assume that time-series y_t has a line trend from the beginning, and the trend would keep till the future. Then we can construct the Line Trend Prediction Model as follows:

$$\hat{y}_{t+T} = a_t + b_t T, T = 1, 2, ... \quad (4)$$

t refers to which quarter it is now; T refers to the number of quarters from t to the time to be predicted; a_t refers to the intercept; b_t refers to the slope.

Then we determine the smoothing factor according to the moving average. According to the equation above, we can find that

$$M_t^{(1)} = \frac{1}{N}(y_t + y_{t-1} + ... + y_{t-N+1}) = \frac{1}{N}\{y_t + (y_t - b_t) + ... + (y_t - [(N-1)b_t]\}$$

$$= \frac{Ny_t - [1 + 2 + ... + (N-1)]b_t}{N} = y_t - \frac{N-1}{2}b_t \quad (5)$$

So

$$y_t - M_t^{(1)} = \frac{N-1}{2}b_t \quad (6)$$

We do a derivation similar to the equation **(3),** finding out that

$$M_t^{(1)} - M_t^{(2)} = \frac{N-1}{2}b_t \quad (7)$$

Finally, we get the following formula to calculate the smoothing factor:

$$\begin{cases} a_t = 2M_t^{(1)} - M_t^{(2)} \\ b_t = \frac{2}{N-1}(M_t^{(1)} - M_t^{(2)}) \end{cases} \quad (8)$$

IV. Problem Solving

Based on the model above, we analyze the data of the three products and predict the numbers of positive reviews in the next four quarters, thus predicting their reputation. We set $N = 8$, and then we get the linear trend prediction model of the laptop, the air conditioner, and the lamp respectively and thus predicting the numbers of positive reviews in the next four quarters, the result is shown in Fig 3.

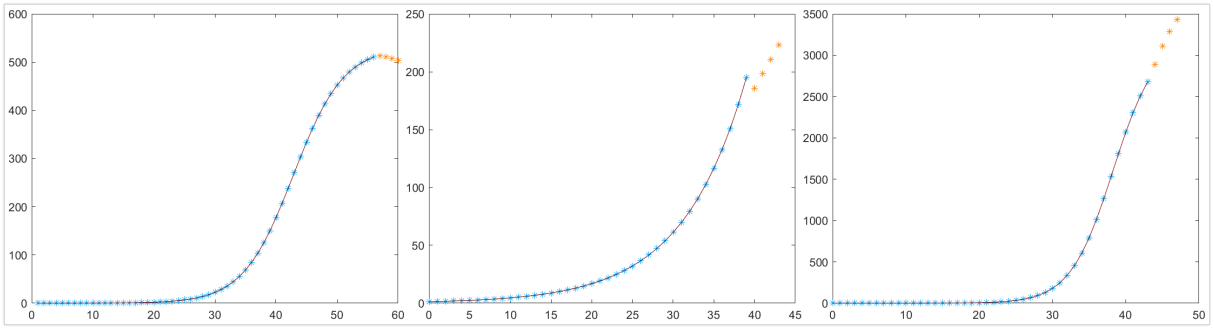

Figure 3 Moving Smooth Prediction Results

From Fig 3, we predict that the reputation of the air conditioner and the lamp is expected to rise, while the reputation of the laptop is anticipated to go down.

V. Conclusions

For the text evaluation and star evaluation analysis of e-commerce platform, we use the Star Rating and Text Rating of three products on Amazon platform, carry out quantitative fitting processing for them, and put forward an innovative data fitting processing model based on time series. Firstly, we normalized and quantified the Star Rating and Text Rating, and based on this, put forward the Logistic Regression Equation Model to generate the fitting function. However, if we set a threshold, it would be influenced by outliers. That is to say, we should take the outliers and uncommon changes into account. According to the Logistic Regression Equation Analysis, we find that the relationship between the time of the reviews and the number of valid positive reviews. In addition, we establish a Logistic-ARMA model, analyze the change characteristics of the comment orientation of three products over time, then analyze and predict the development and change of the comment orientation in the future, and finally determine the potential successful products and failed products.

In general, the error of the model based on the actual situation is small, which can provide a certain reference for the product sales strategy of the e-commerce platform from the level of commodity evaluation.

REFERENCES

[1] Wenji Wei. Study on inventions of fresh food in commercial aspects using e-commerce over internet[J]. Acta Agriculturae Scandinavica, Section B — Soil & Plant Science, 2021, 71(4) : 303-310.

[2] Zhenyang Hu et al. Study on the Impact of Rural E-commerce Development on Farmers' Income Increase[J]. Journal of Innovation and Social Science Research, 2021, 8(4)

[3] Wu P F . Motivation Crowding in Online Product Reviewing: A Qualitative Study of Amazon Reviewers[J]. Information & Management, 2019.

[4] Anna Nikishova,Giovanni E. Comi,Alfons G. Hoekstra. Sensitivity analysis based dimension reduction of multiscale models[J]. Mathematics and Computers in Simulation,2020,170.

[5] Xialu Liu,Rong Chen. Threshold factor models for high-dimensional time series[J]. Journal of Econometrics,2020.

[6] Opinion-lexicon-English attained from http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

[7] Kai Zhou. Real estate price based on ARMA-ANFIS Research on prediction model[C]. International Informatization and Engineering Associations. Proceedings of 2019 International Conference on Educational Reform,Management Science and Sociology(ERMSS 2019). International Informatization and Engineering Associations: Computer Science and Electronic Technology International Society, 2019:342-349.