

# Optimization analysis of online marketing strategy based on the regression prediction model

JingyaYang<sup>1</sup>, XiwangXiao<sup>1</sup>, XianfengZhang<sup>1</sup>, BinWen<sup>1</sup>, YuehuaZhang<sup>1</sup>, JinQiuGuan<sup>1\*</sup>, LanLi<sup>2\*</sup>

<sup>1</sup>College of Science & Mechanical Engineering, Jiamusi University, Jiamusi City, Heilongjiang Province, 154007, China

<sup>2</sup>College of Big Data and Software Engineering, Wuzhou University, Wuzhou City, Guangxi Province, 543000, China

\*Co-Corresponding author's e-mail: jmsdx8888@sohu.com; lljmsu@163.com

**Abstract.** In response to the three new product sales recommendations of the Amazon online market "Sunshine", based on the data set provided by the enterprise data center, the data in the NLP feature package is preprocessed in Python, and the AHP model is established to measure and establish the data. This experiment according to the quarterly time series ARIMA model to forecast the product sales, and then through the standardization of OLS regression model to predict the trend of the development of products. The results showed that customer text reviews were an important reference for the sunshine company and found that microwave oven sales increased by an average of 6.28%, while sales of baby pacifiers and hair dryers increased by an average of 1.02%. This experiment based on this, puts forward the corresponding marketing strategy. Online marketing strategy can pay more attention to the customer's comments, make updates on the features of the product, the use of to meet customer demand. The study enriches the existing research results, and the sun of the company's sales strategy provides feasibility suggestions, thus has important theoretical significance and practical value.

## 1. Introduction

Amazon online shopping with abundant online evaluation system, can accurately express the product customer satisfaction through star evaluation and text comments, make increase the understanding of the other customers for products to promote online merchants to improve[1]. Experiment can submit a rating to "help" at the same time, the auxiliary for evaluation, thus increasing online evaluation effect on sales strategy [2]. According to Amazon's 2019 hot-selling product data, Sunshine's highest, medium and lowest-selling products are electronics, home kitchen appliances, baby supplies, and select the most representative microwave ovens, hairdryers and baby pacifiers as the subjects of research. "The evaluation dataset gets useful information based on past sales of similar products on Amazon's online market, establishes an online sales strategy and predicts future sales trends." Through the analysis of the PCA model, two parameters representing quantitative quantity are extracted from the original data set, which mainly includes direct evaluation score and indirect evaluation score[3]. Using hierarchical analysis, we can get the data measurement measures of rating and evaluation[4]. The time series in the data set is organized to reflect the quarterly measurement form of more timely information, and the quarterly decomposition of the time series is observed in the time measurement. The SARIMA model is used to fit the time series curve and follow the trend of the data. Finally, using the OLS standardized



regression model [5], a linear combination of the best text evaluation score and rating is obtained, which shows that there are large random perturbations between the overall score and the purchase result[6].

## 2. Data Sources

"Reviews of microwave ovens, baby nipples and hairdryers are in customer reviews, taking annual data from the online market, customer review indicators, and annual sales of products in the Sunshine Company Statistical Yearbook 2012-2019".

## 3. Research method

### 3.1. Hierarchical Analysis (AHP)

Hierarchical Analysis (AHP) refers to the breakdown of elements related to decision-making into targets, standards, and procedures. On this basis, qualitative and quantitative analysis methods are used. The three levels in the hierarchical analysis structure are set according to the parameters in the dataset.

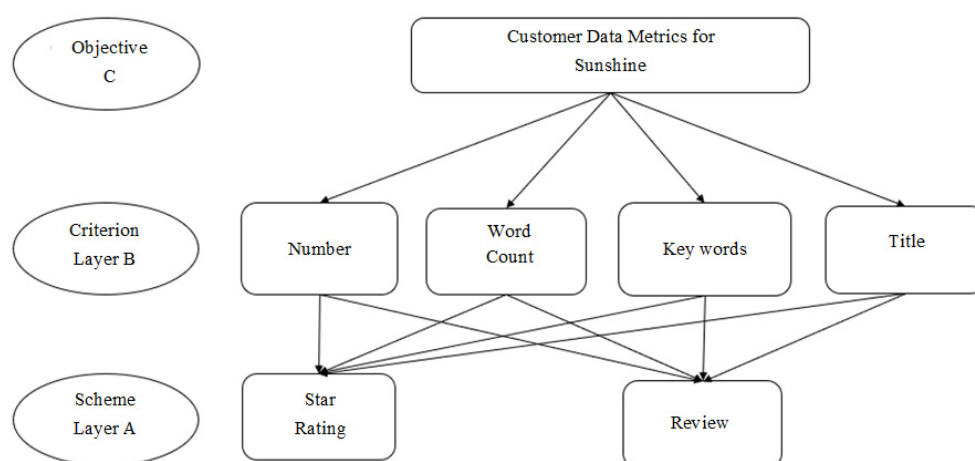


Fig. 1 Customer evaluation hierarchy map

The C-B comparison matrix is constructed, and the four evaluation indicators of the criteria layer are compared, and the C-B comparison matrix is obtained.

Tab. 1 M-C Comparison matrix

C	B1	B2	B3	B4
B1	1	1/2	1/4	1/3
B2	2	1	1/2	1/3
B3	4	2	1	2
B4	3	3	1/2	1

Conformity testing is carried out according to the following formula:

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (1)$$

$$CR = \frac{CI}{RI} \quad (2)$$

Note: *CI*: Consistency index; *CR*: Consistency ratio; *RI*: Consistency index;  $\lambda_{\max}$ : Maximum eigenvalue.

### 3.2. ARIMA Time series model

The original time series in the data set are discrete. In order to establish and analyze the model based on time measurement, the original time measurement is converted into quarters, which is conducive to a

clearer observation of the data changes in a year. By using the first comprehensive score to represent the product's reputation index, the data can be more analysis. Through the seasonal decomposition of the data, the time series of reputation trend is decomposed into: long-term variation trend, seasonal variation rule, periodic variation rule and irregular variation, which makes the time series diagram easier to highlight the rule and achieve the purpose of intuitive analysis. The equation of the product model of seasonal decomposition of time series is given by the following equation.

$$Y = T \times S \times C \times I \quad (3)$$

Note:  $Y$ : The final change of the indicator value;  $T$ : Long-term trend value;  $S$ : Seasonal cycle trend value;  $C$ : Cyclically changing value;  $I$ : Irregular value.

According to the results of seasonal decomposition, the comprehensive time series diagram of the components of each time series was obtained by SPSS-24.

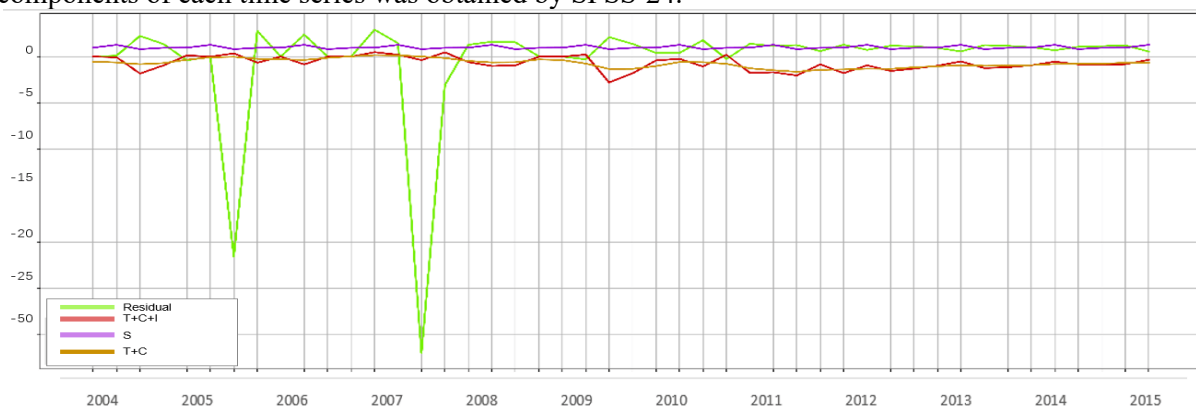


Fig. 2 Seasonal decomposition time series

For the sake of express the time series changes more clearly,  $f(x)$  of the seasonal decomposition time series curve can be obtained by fitting. Using Mat lab to fit the raw data linearly, Fig 2 shows that the reputation index of the three products sold online by Sunshine Company depends on the time change, the overall stab trend is stab, is a steady time series, there is a clear seasonal change trend. So it is considering using a time series prediction model for prediction.

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) \left(1 - \sum_{i=1}^p \Phi_i L^{mi}\right) (1-L)^d (1-L^m)^D y_t = \alpha_0 + \left(1 + \sum_{i=1}^q \theta_i L^i\right) \left(1 + \sum_{i=1}^q \theta_i L^{mi}\right) \varepsilon_t \quad (4)$$

Note:  $\Phi$  is the autoregressive coefficient, but  $\theta$  is the moving average coefficient, and  $p$ ,  $d$ ,  $q$  are the regression order, division number, and moving average of the law.

### 3.3. OLS Standardized regression model

Using Mat lab for data set if purchases, the first principal component data, and data processing, the second principal components analysis, the function relationship between shows sun online sales of three kinds of products of the company whether to buy depends on changes in the attributes of different components, identify and describe the quantitative and qualitative model, and its different components, the relationship between variables OLS regression model is analyzed.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu \quad (5)$$

Note:  $\mu$ : interference term;  $\beta_i$ : regression coefficient;  $y$ : customer purchase volume;  $x_1$ : first principal component;  $x_2$ : second principal component.

## 4. Model parameter selection and testing

### 4.1. Parameter selection and test of the analytic hierarchy process model

In the APH model, the consistency ratio CR was all less than 0.1, and the consistency test was all passed, and the calculated result data were summarized.

According to the quantitative value of each factor, two schemes of three products were calculated with a score of 0.2814 for rating and 0.7187 for evaluation. According to the above results, the maximum value of the evaluation results is obtained, and the evaluation scheme is preferred.

Tab. 2 Customer review weight summary

	Index Weight	Rating	Evaluation
digital	0.0941	0.6667	0.3333
Word count	0.1627	0.3333	0.6667
Keywords	0.4276	0.2	0.8
title	0.3156	0.25	0.75

4.2. ARIMA Parameter selection and testing of time series model

SARIMA (p, d, q) (p, d, q) model was calculated by SPSS-24 software, and the results are shown in the following Tab:

Tab. 3 SarimaModel parameters

	Estimate	SE	t	Sig
Microwave oven	-.303	.095	-3.182	.003
Baby pacifier	.088	.032	2.736	.009
hairdryer	-.157	.093	-1.692	.098

As can be seen from the ACF and PACF graphs above, some of the data in the residual PACF have a lag order that is significantly different from 0, indicating that the remaining is not white noise. In addition, it can be seen from the above Tab that the P value is 0.036 after the Q test of the residual error, and the null hypothesis cannot be rejected at the 99% confidence level. The model goodness of fit R<sub>2</sub> is 0.531, which is more than 50%. Although the goodness of fit of the model is not very high, the use of the SARIMA model aims to explore the trend changes of the remaining scores, so that the model can meet the expected requirements.

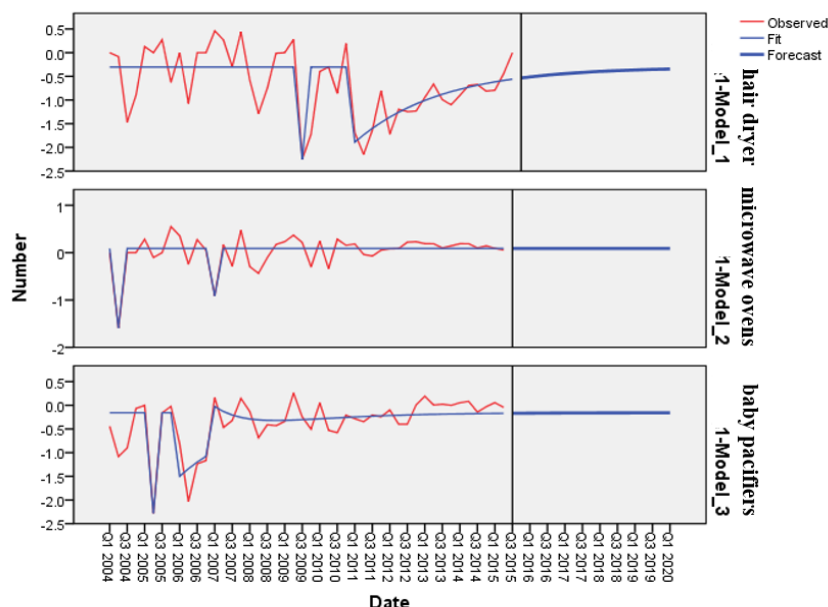


Fig. 3 Raw data fitting curve

It can be seen from the comparison between the final fitting diagram and the results that although the model does not fit the original data, it can also reflect the changing trend of the original data. Use the time measurement model to predict reputation scoring data to the first quarter of 2020. It can be seen

from the three sales surplus scores, microwave oven scoring in the early fluctuations, indicating that in previous years of similar products, not popular with the public, but after 2012 gradually showed an upward trend. From the forecast results, the upward trend is still extending. For baby products, it can be seen from the data trend in previous years that, compared with other two types of products, the fluctuation range is the smallest, and the fluctuation gradually converges and becomes stab after 2012. In hair dryer products, the product has two minimum situations of reputation evaluation during the early period. In the subsequent time, there are still fluctuations, but gradually converge, from the forecast of the results of the trend, also showed a stab state.

#### 4.3. Parameter selection and testing of OLS standardized regression models

Set the original hypothesis  $H: \beta_1 = \beta_2 = 0$ , and solve it by Stata as follows:

Tab. 4 OLS Model Anova results

Source	SS	DF	MS	Number of obs=26,558 F(2, 26555)=381.89
Model	103.994	2	51.997	Prob > F=0
Residual	3615.646	26,55	0136	R-squared=0.0280
Total	3719.640	26,557	0.401	AdjR-uared=0.0279 Root MSE=.36899

It can be seen from the solution that the P-value of the joint significance test is  $p = 0$  and  $p < 0.05$ , so the null hypothesis is rejected at a 95% confidence level, that is,  $\beta_1, \beta_2$  are not all 0, and  $SSR = 103.994, SSE = 3615.646, SST = 3719.640; FindR^2 = 0.028, R^2_{adjusted} = 0.028$ , the adjusted goodness of fit is the same as the original goodness of fit, so it is believed that there is no effect of multicollinearity, and the model is an explanatory regression. Not concerned.

#### 4.4. Model-checking

Construct a Spearman correlation coefficient model between the quantified text comment score and the rating level, and observe whether there is a strong correlation between the three. Spearman correlation coefficient model:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (6)$$

Hypothesis test on Spearman correlation coefficient results to construct statistics,  $r_s \sqrt{n-1} \sim N(0,1)$ . The original false  $H_0: r_s=0$ , and the  $P$  values of the calculated results were all less than 0.05. Therefore, the null hypothesis was rejected at the 95% confidence level, and the obtained spearman correlation coefficients were all significantly.

## 5. Results and discussion

### 5.1. Result

Based on the model and the results, some suggestions for the improvement of the three kinds of products are given. The baby pacifier should be more novel and adjust the price of the product, the hair dryer should be endowed with more functions, and the microwave oven should mainly improve the quality of the product and the size of the internal use space. In the case of sales, it is recommended that the label of the product should reflect more information about the product and that richer product distribution services should be added.

### 5.2. Discuss

At the present stage, the value of Internet technology applied in marketing is very significant, and the marketing concept and method have been greatly impacted and challenged. With the gradual expansion of consumer groups, consumer demand is more diversified, marketing methods are more diversified, and the efficient convergence of network technology and marketing, but also promote the scope of consumption to expand and extend to a larger field [7]. Therefore, it is urgent to actively explore a more effective marketing model in the era of network economy, so as to keep it fresh and dynamic all the time and it will eventually be eliminated by the market and society. The modern diversified marketing market needs to expand the scope of marketing so as to accurately locate the restrictive problems existing in the traditional marketing market. Under the background of the era of network economy, marketing strategy innovation is not only helpful for enterprises to improve their competitive strength, but also to expand their interests. To the network utilization as the leading expansion of marketing methods, the full use of network technology, through video and images and other different display methods, vibrant product image, can form a strong visual impact on consumers, and thus stimulating consumers' purchase desire. For enterprise marketing, this is not only an opportunity but also a challenge. Therefore, on the basis of catering to the changing times, it is necessary to grasp consumers' psychology and fundamentally improve the competitive strength of enterprises.

### Acknowledgments

Heilongjiang Provincial Department of Education Basic Research Project, China, (2018-KYYWF-0957); China Scientific and technological innovation projects in Jiamusi University, Heilongjiang Province, College students: 201910222075.

### References

- [1] XiaodongJiang. (2015) What kind of product reviews are most useful- A study of the impact of quantitative and textual features of online reviews on their usefulness. *J. Sci. Foreign Economy and Management*, 7: 42-56.
- [2] Chakravarty, A., Liu, Y., & Mazumdar, T. (2010) The differential effects of online word-of-mouth and critics reviews on pre-release movie evaluation. *J. Sci. Social Science Electronic Publishing*, 24: 185-197.
- [3] Yari Saeed Khanloo, Bahman. (2002) Combining simple trackers using structural SVMs for offline single object tracking. *J. Sci. International journal of mineral processing*, 148: 23-31.
- [4] Brusco N K, Taylor N F, Watts J J, et al. Economic Evaluation of Adult Rehabilitation: A Systematic Review and Meta-Analysis of Randomized Controlled Trials in a Variety of Settings. *J. Sci. Archives of Physical Medicine & Rehabilitation*, 95: 94-116.
- [5] PingHan, PengxinWang, MiaoTian, et al. (2007) Application of the ARIMA Models in Drought Forecasting Using the Standardized Precipitation Index. *J. Sci. Applied Mathematics & Computation*, 392: 352-358.
- [6] Gómez-Valent, Adrià, Karimkhani E, Solà, Joan. (2015) Background history and cosmic perturbations for a general system of self-conserved dynamical dark energy and matter. *J. Sci. Journal of Cosmology & Astroparticle Physics*, 25: 48-57.
- [7] Sayadi S. (2016) The multifunctionality of agriculture: challenges and opportunities for new market demands. In: *International Symposium on Organic Greenhouse Horticulture Workshop: Market & Quality*. Fujian, Xiamen. pp: 165-178.

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.