

PAPER • OPEN ACCESS

Research on E-commerce Customer Evaluation System in the Context of Big Data: Taking Amazon as an Example

To cite this article: Xu Yuelin *et al* 2021 *J. Phys.: Conf. Ser.* **1955** 012024

View the [article online](#) for updates and enhancements.

You may also like

- [Contour propagation for lung tumor delineation in 4D-CT using tensor-product surface of uniform and non-uniform closed cubic B-splines](#)
Renchao Jin, Yongchuan Liu, Mi Chen et al.
- [QSSOR and cubic non-polynomial spline method for the solution of two-point boundary value problems](#)
A Sunarto, P Agarwal, J V L Chew et al.
- [Extended analysis of solid cancer incidence among Nuclear Industry Workers in the UK 1955–2011: Comparison of workers first hired in earlier and later periods.](#)
Nezahat Hunter and Richard Haylock

ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Research on E-commerce Customer Evaluation System in the Context of Big Data: Taking Amazon as an Example

Xu Yuelin^{*1}, Xu Danyang², LiHui³

¹Department of Computational Mathematics and Cybernetics, Mathematics and Applied Mathematics, Shenzhen MSU-BIT University, Shenzhen, Guangdong, 518119

²Honours College, Capital Normal University, Beijing, 100048

³School of economics and management, Shenyang Aerospace University, Shenyang, Liaoning, 110000

*duoerxs@163.com

Abstract: Amazon provides the customers an opportunity to assess their purchases. Manufacturers use this data to acquire further insights into the marketing the timing of participation, and potentially successful product designs.

The first step is to preprocess the raw data by detecting outliers and using cubic spline method to complete them and verify the correctness. Then, the data is analyzed in time order, followed by data quantification and visualization. The conclusion is that the comment does affect the star rating.

Finally, as a consultant hired by Sunshine, put forward some suggestions to the company's marketing director: elaborating the establishment of relevant models, explaining the problems solved by the corresponding models, and summarizing the team's analysis and results.

1. Introduction

The Amazon Web Store provides customers with the opportunity to rate the stars of their purchases and submit reviews [1]. In addition, other customers can help evaluate existing evaluations before buying. Product manufacturers can also use these ratings and evaluations to gain a deeper understanding of the consumer market[2]. Now Sunshine Company wants to identify key models and parameters based on historical ratings and reviews provided by customers, so as to better launch its three products[3].

2. Models and Analyses

2.1. The establishment of RPCA model

Step1: Centralize the sample data;

Step2: Find the sample covariance matrix;

Step3: Eigenvalue decomposition of the sample covariance matrix, and mapping through the eigenvectors corresponding to the first k eigenvalues:



$$\partial'_i = \begin{bmatrix} \delta_1^T \partial_i \\ \delta_2^T \partial_i \\ \vdots \\ \delta_d^T \partial_i \end{bmatrix} \quad (1)$$

(1) Maximum variance derivation

Assuming that the original coordinates are $\{v_1, v_2, \dots, v_n\}$ and $\{\partial_1, \partial_2, \dots, \partial_n\} = \{v_1 - \mu, v_2 - \mu, \dots, v_n - \mu\}$ after centralization, the inner product of the vector can be understood as the projection length of the first vector on the second vector, so the projection of ∂_i on δ can be expressed as $(\partial_i, \delta) = \partial_i^T \delta_0$. According to the above, the function we need to optimize is to maximize the projection variance, which is

$$\begin{aligned} W(\partial) &= \frac{1}{n} \sum_{i=1}^n (\partial_i^T \delta)^2 = \frac{1}{n} \sum_{i=1}^n (\partial_i^T \delta)^T (\partial_i^T \delta) \\ &= \frac{1}{n} \sum_{i=1}^n \delta^T \partial_i \partial_i^T \delta \\ &= \omega^T \left(\frac{1}{n} \sum_{i=1}^n \partial_i \partial_i^T \right) \delta \end{aligned}$$

For this reason, we can get the maximum projection direction to be solved is the eigen matrix corresponding to the maximum eigenvalue of the covariance matrix.

$$\begin{cases} \max \{\delta^T \Sigma \delta\} \\ \text{s.t. } \delta^T \delta = 1 \end{cases} \quad (2)$$

(2) RPCA

RPCA is used to solve data with multiple influence factors. The basic assumption is that the data matrix contains structural information (the matrix is low rank) and influence factor matrix (the matrix is sparse), so RPCA hopes to decompose the original matrix into $M = W + Q$ form:

$$\min_{W, Q} \text{rank}(\cdot) + \lambda \|Q\|_0 \quad \text{s.t. } W + Q = M \quad (3)$$

The $\text{rank}(\cdot)$ and L_0 norms here are non-convex and non-smooth, so they need to be scaled down. That is, use the kernel norm instead of $\text{rank}(\cdot)$ and L_1 instead of L_0 :

$$\min_{W, Q} \|W\|_* + \lambda \|Q\|_1 \quad \text{s.t. } W + Q = M \quad (4)$$

2.2. Data training

Taking the hair dryer data given in the title as an example, the RPCA model was trained by MATLAB to obtain training result data [4]. The training result data is imported into Excel for sorting. The sorted data is shown in Table1.

Table1 Hair dryer processed data

Star rating		Effective rate	Vine		Verified purchase	
Grade	Number		Y	N	Y	N
1	1025	77.38%	0	1025	736	289
2	637	76.55%	3	634	508	129
3	992	78.25%	20	972	857	135
4	2090	86.60%	51	2039	1803	287
5	6673	88.83%	105	6568	5891	782

Among them, Effective rate represents the ratio of the number of helpful votes to the number of total votes. In order to obtain the relationship between the data of each component in Table 1, the data is visualized, as shown in Figure1 and Figure2.

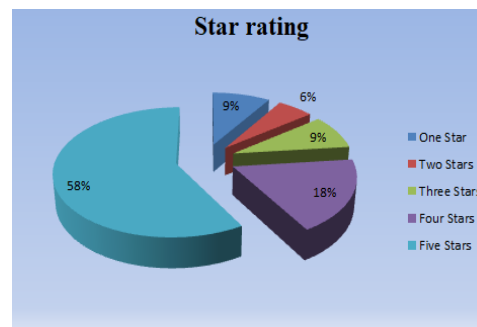


Figure1 Proportion of star rating

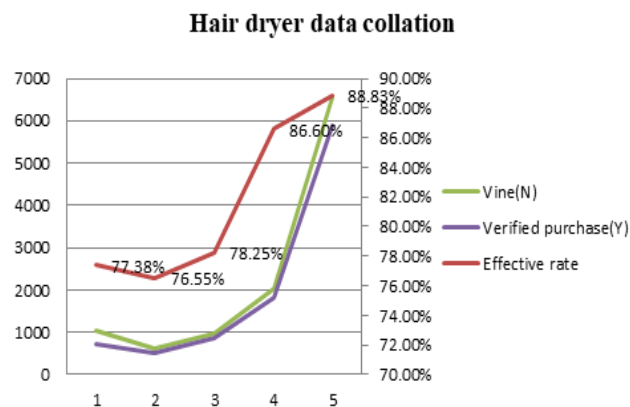


Figure2 Relationship between component data and star rating

It can be seen from Figure1 that the five-star rating accounts for the largest proportion and the two-star rating accounts for the smallest proportion. Can get most customers tending to choose products with high star rating [5]. It can be seen from Figure2 that the Effective rate is the lowest among the two-star rated products, indicating that the evaluation of the product is the worst among most of the evaluations. Effective rate is the highest among the five-star rated products, indicating that the product's evaluation is the best in most reviews [6]. When Vine(N) and Verified purchase are the lowest, they are also in the two-star rating, and the highest is in the five-star rating. It shows that there is a certain quantitative relationship between the components, which together affect the star rating and star evaluation.

According to the above method, the data of the microwave oven and the pacifier are processed to obtain the following data tables and figures, respectively.

Table2 Microwave processed data

Star rating		Effective rate	Vine		Verified purchase	
Grade	Number		Y	N	Y	N
1	399	79.82%	0	399	115	284
2	110	69.66%	0	110	56	54
3	131	81.10%	2	129	90	41
4	296	82.15%	8	288	248	48
5	662	90.24%	9	653	580	82

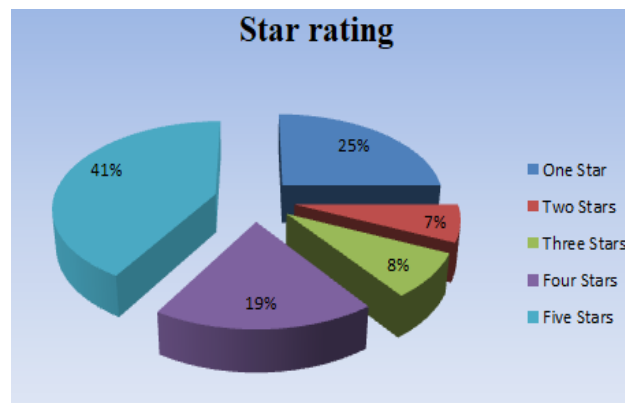


Figure3 Proportion of star rating

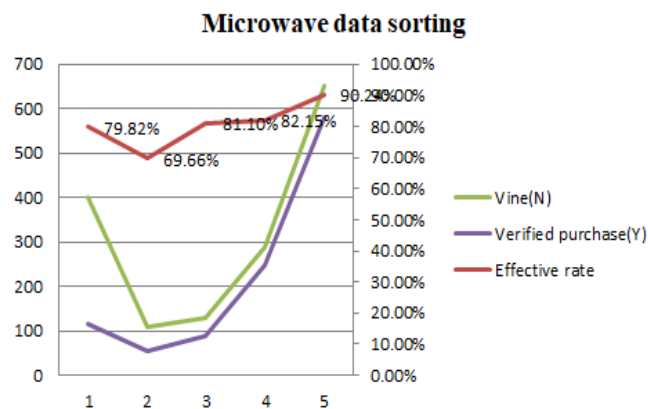


Figure4 Relationship between component data and star rating

Table3 Pacifier Processed data

Star rating		Effective rate	Vine		Verified purchase	
Grade	Number		Y	N	Y	N
1	1169	59.49%	3	1166	887	282
2	933	66.15%	2	931	781	152
3	1416	63.34%	14	1402	1230	186
4	2692	77.43%	43	2649	2328	364
5	12554	79.45%	71	12483	10961	1593



Figure5 Proportion of star rating

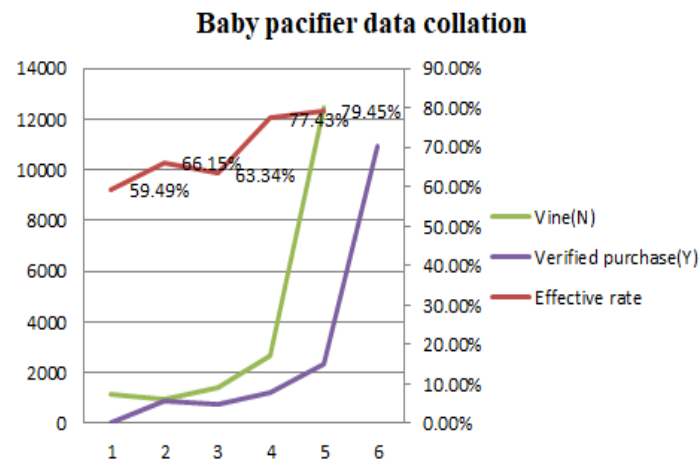


Figure6 Relationship between component data and star rating

From the data and graphs compiled by the microwave oven and pacifier, it can be seen that the proportion of star ratings, the relationship between the data of each component and the star rating, and the results obtained by the hair dryer are similar [7]. From this, it can be concluded that the data of each component affect each other and together affect the star rating and star evaluation. From the data point of view, the impact of each component data on star ratings and star ratings is directly proportional.

3. Conclusion

Our team is proud to be a consultant to Sunshine Company. We built relevant models based on data analysis. These models facilitate the sales of your company's three products.

It is understood that the Amazon online mall provides customers with the opportunity to rate the "star rating" of purchased products and submit reviews. In addition, other customers can help evaluate existing evaluations before buying. Product manufacturers can also use these ratings and evaluations to gain a deeper understanding of the consumer market. First, we used mathematical evidence to analyze based on the data provided, and found out the main factors that affect ratings and reviews--Vine and Verified purchase, thereby establishing the RPCA model. Based on the collated data, pie charts and line charts are drawn, and the following relevant conclusions are drawn: the data of each component affect each other and jointly affect the star rating and star evaluation. From the data point of view, the impact of each component data on star ratings and star ratings is directly proportional.

For the existence of a large amount of text data, your company wants to obtain some information from ratings and reviews, which cannot be obtained by directly analyzing the data. Combining the sample data sorted out in question 1, our team established a dictionary word quantization relationship (QCR). Correlation evaluation between text sentences and error probability index were used for correlation evaluation.

The customer is God. In order for the customer to have a good buying experience, our team recommends a personalized method of emotional matching products. This method analyzes the opinions of different users, uses the principle of equal evaluation weight, and finally uses the SentiByTerm algorithm to calculate the customer's emotions on the product Inclination values to learn more about what customers think.

References

- [1] Lu Hongwei, Wang Shitong. Prediction method of high-dimensional data subspace clustering based on RPCA [J]. Computer Engineering and Science (3).
- [2] Zhong Maosheng. Chinese text segmentation method based on dictionary word quantization relation [J]. CEA, 2008, 44 (21): 25-29.

- [3] Hu Yi, Lu Ruzhan, Chen Yuquan, et al. Research on Chinese Text Clustering Based on Quantified Relations of Words in Dictionaries [J]. High Technology Letters, 17 (8): 778-782.
- [4] Liang Yan, Xia Letian. Application of ARMA Model of Time Series [J]. Journal of Chongqing University of Technology (Natural Science), 2012, 26 (8): 106-109.
- [5] Li Yanshuang, Zeng Zhenxiang. Application of Principal Component Analysis to Multi-index Comprehensive Evaluation Method [J]. Journal of Hebei University of Technology, 1999 (1): 94-97.
- [6] Li Xiaosheng, Chen Zhenzhen. How to Apply SPSS Software for Principal Component Analysis [J]. Statistical Research, 2010, 27(8): 105-108.
- [7] Hou Yinxiu, Li Weiqing, Wang Weijun, et al. Research on personalized recommendation of books based on user preference and emotional matching of product attributes [J]. Modern Library and Information Technology, 2017, 1(8): 9-17.