Problem Chosen	2020 MCM/ICM Summary Sheet	Team Control Number
C		2004951

A K-means-based Model for Product Evaluation

Abstract

Nowadays, the proportion of online shopping in the consumer market is increasing and people generate a large amount of data when shopping online. Important information such as product evaluation can be obtained from the data, which can be used for merchants' research on marketing strategies.

In order to incorporate reviews into our model, we did sentiment analysis to quantify the text, which effectively reduced the difficulty of processing the text. Then we obtained the emotional polarity of the reviews headline and reviews body for analysis. To illustrate the intrinsic relationship among star ratings, reviews, and helpfulness ratings, we first process them with correlation test and determine that there is a certain linear relationship among them. In that case, we decided to use a multiple linear regression model and gain and the relationship among them. Autocorrelation is also adopted for acquiring more accurate results.

After the classification of KNN, we acquire an intuitive result of costumer evaluation of the products. As a commonly used unsupervised clustering, k-means attracted us with its excellent performance. In order to find the best number of clusters, we optimized the k-means by using Silhouette Coefficient as indicators. According to the distance between the cluster center and the category to which the sample belong, we propose a credibility index of the review and a success index of the product.

To anticipate the change of product reputation over time, our team take advantage of the outstanding representative LSTM on the time series model to predict the change in the reputation of the product, and have obtained a credible prediction. We also discover the relationship between specific quality descriptors in reviews and the rating levels.

In general, our model has extensive adaptability and can provide stable and accurate judgments for enterprises. At the same time, our model has broad application prospects for it adopt machine learning method to process the data.

Keywords: KNN, K-means, Linear Regression, LSTM

Contents

1. Introduction

1.1. Problem Background

With the rapid development of Internet and express delivery industry, the proportion of online shopping in the consumer market is increasing. Since online shopping has the characteristics of non-physical contact, the evaluation of consumers who have already purchased products has become the main basis for consumers' online shopping. Due to the large number of reviews, consumers can only refer to a part of them. In that case, product ratings become an important reference for consumers when they shop online. At the same time, the scores also indicate the product is potentially successful or failing.

Product evaluation is an important type of user-generated content, and many scholars have studied the product feature extraction, sentiment analysis, and utility analysis in product evaluation. Researchers have found that consumer evaluation has a causal relationship with consumer buying behavior and has an impact on consumer choice of products[1]. A highly credible product evaluation can bring clear product information to consumers and reduce the uncertainty of the shopping process. In addition, user feedback can provide valuable suggestions for product's improvement, which is beneficial for producer.

The rules exhibited by product scoring, evaluation and other behaviors help merchants to predict consumer purchasing behavior and adjust marketing strategy, which is showed in figure 1. By analyzing these data, we can provide reasonable guidance for e-commerce merchants.

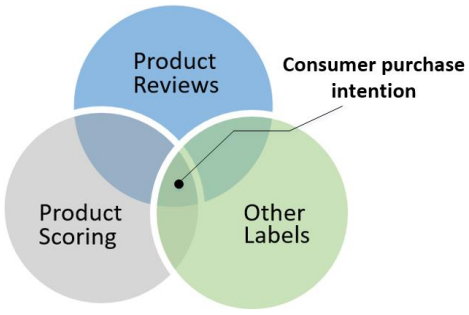


Figure 1. Factors affecting consumer purchase intention

1.2. Our Work

In order to help Sunshine Company better sell their three new products in the online marketplace, we need to analyze key patterns, relationships, measures, and parameters in past customer supplied ratings and reviews associated with other competing products. Based on the analysis, we can draw a conclusion about the important design features that would enhance product desirability and offer a suitable marketing strategy.

To solve those problems, we will proceed as follows:

1) Our team adopt Natural Language Processing (NLP) method to divide the data into positive and negative categories and return the quantitative results between -1 and 1.
2) Using regression analysis to illustrate the intrinsic relationship among star ratings, reviews, and helpfulness ratings.
3) We design an evaluation model to measure the product is successful or failing and obtain a data-based measure for Sunshine Company to track.
4) A time-based model to anticipate the changes in product reputation is established.
5) The connection between specific star ratings and reviews is analyzed.
6) We analyze the relationship between specific quality descriptors in reviews and the rating levels.

To illustrate our model better, we display the process in figure 2.

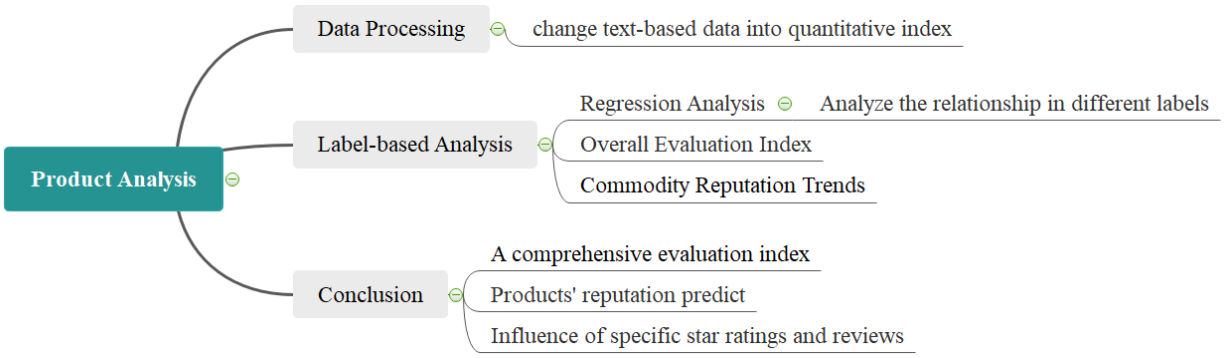


Figure 2. The general frame of our model

2. General Assumptions

Our model makes the following assumptions:

- The results of our sentiment analysis are sufficient enough to reflect consumers' evaluation of the product.
- The ratings of reviews are under the control of time.
- The degree of success of a product is positively related to the overall rating.
- If reviews are close in Euclidean space, they may belong to the same cluster.
- The more reviews deviate from the cluster center, the less credible they are.
- Vine has a great influence on subsequent reviews.

- Specific quality descriptors have a strong guiding effect on the results.

3. Notations

Abbreviation	*Description*
$\boldsymbol{r(x,y)}$	the correlation coefficient between x and y
$\boldsymbol{Cov(x,y)}$	the covariance of x and y
$\boldsymbol{Var}$	the variance
$\boldsymbol{e_i}$	the residual of i
$\boldsymbol{a(i)}$	the average distance from sample i to other samples in the same
$\boldsymbol{b(i)}$	the average distance from sample i to all samples of other clusters
$\boldsymbol{s(i)}$	the Silhouette Coefficient of sample i
$\boldsymbol{P_i}$	the proportion of each cluster
$\boldsymbol{r_i}$	the star rating corresponding to each cluster
$\boldsymbol{R_{PROD}}$	the overall rating of a product
$\boldsymbol{S}$	the degree of success of a product
$\boldsymbol{\mu_i}$	the centroid of cluster
$\boldsymbol{c}$	the credibility of the data
$\boldsymbol{f_t}$	the value of forget gate in LSTM
$\boldsymbol{i_t}$	the value of input gate in LSTM
$\boldsymbol{C'_t}$	the previous cell status
$\boldsymbol{C_t}$	the current cell status
$\boldsymbol{o_t}$	the value of output gate in LSTM
$\boldsymbol{h_t}$	the hidden state used for prediction

4. K-means-based Product Analysis Model

4.1. Reviews Analysis

Before analyzing the reviews, we preprocessed the data to make it more convenient to process. Since user reviews are in the form of text, it is difficult to measure purchaser 's evaluation of the product intuitively. In order to extract the relevant information in reviews and obtain costumers' quantitative evaluation of the product, we need to process the text with a language processing method to obtain a vector or matrix for subsequent analysis. Our team use TextBlob to process the reviews and obtain the quantitative evaluation.

In order to facilitate the subsequent processing of text labels, we first delete the punctuation marks in the review title and review body, then use TextBlob[2] to classify the words and do sentiment analysis. Through analysis we can get the emotional bias of the text and the subjectivity of the results, which are displayed in figure3.

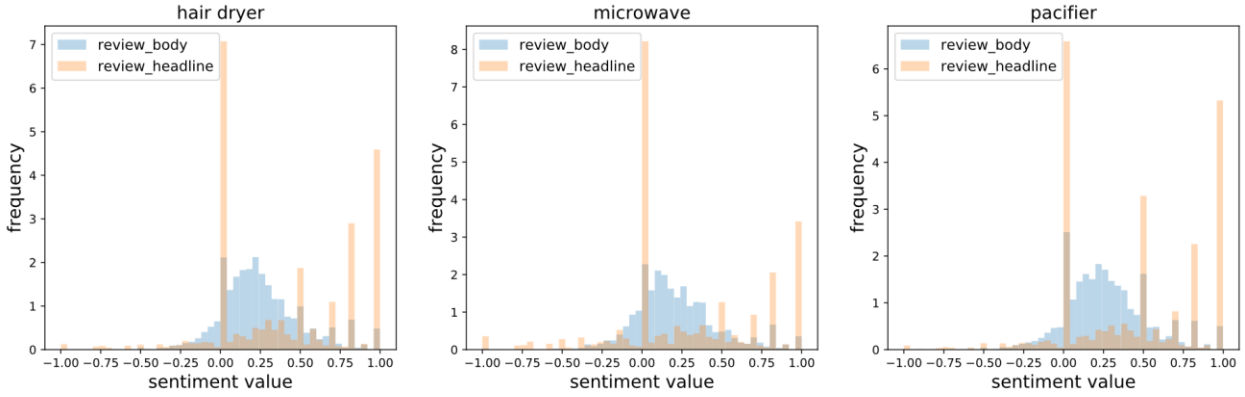


Figure 3. The result of sentiment analysis

TextBlob analyze the emotional polarity of the text and return a number between -1 and 1. The closer that this value is to 1, the more positive the review is, while the value closer to -1 indicate that the review is more negative. Compared to headline, reviews body has clearer emotional tendencies and the result is more accurate. Therefore, the reviews body has more referential value. By quantifying the headline and body of the review, we can analyze its impact on product evaluation more intuitively and accurately.

To illustrate the result of our process better, we list some reviews of the pacifier and their emotional bias and the subjectivity in figure 4.

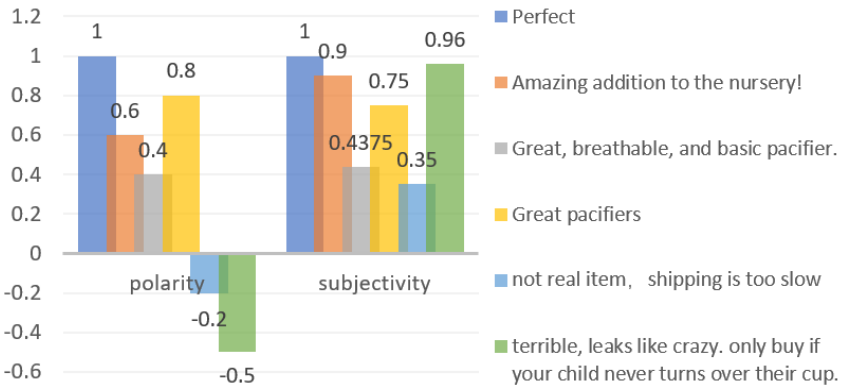


Figure 4. The emotional bias and subjectivity of some reviews

4.2. The Design of the Model

In order to demonstrate our basic model clearly, we divide it into the three following sub models.

- **Label-based product analyzing model**: The identification and description of star ratings, reviews, and helpfulness ratings are discussed in this model.
- **Product evaluating model**: In this model, we identify the most informative data measures for company to track and determine the index to judge the success and failure of the product.
- **LSTM-based reputation predicting model**: This model is designed to analyze the relationship between time-based measures and patterns within each data set.

4.2.1. Label-based Product Analyzing Model

In this section, our team establish the Label-based Product Analyzing model to illustrate the intrinsic relationship among star ratings, reviews, and helpfulness ratings. LPA (Label-based Product Analyzing) model analyze the relationship first by doing correlation test to confirm whether there are correlations among these data. If the correlations exist, we can use regression analysis to verify the exact relationship among them. Residual analysis is also needed to verify the reliability of the results of regression analysis. Besides, autocorrelation test is adopted to improve the rationality and accuracy of the model.

1) Using to Pearson's r method[3] to calculate the correlation coefficient r between x and y.

$$r(x,y) = \frac{Cov(x,y)}{\sqrt{Var[x] \cdot Var[Y]}} \tag{1}$$

$Cov(x, y)$ represents the covariance of x and y. $Var[x]$ is the variance of x, and $Var[Y]$ is the variance of y.

2) Using linear regression to obtain the mathematical relationship among the data.

$$h(\omega) = \omega_1 x_1 + \omega_2 x_2 + \cdots \omega_i x_i + b = \omega^T x + b \tag{2}$$

$$\omega^T = \begin{bmatrix} b \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_i \end{bmatrix}, x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_i \end{bmatrix} \tag{3}$$

x_i represents the different labels we input. ω_i represents the weight corresponding to the labels. ω^T is the matrix of ω_i and b. x is the matrix of x_i.

3) Residual is the difference between the observed value and the prediction obtained from the estimated regression equation, which reflects the deviation caused by the estimated regression equation.

$$e_i = y_i - y_i' \tag{4}$$

e_i represents the residual. y_i is the observed value and y_i' is the predictive value.

4) After obtaining the residual of each data, we can use E-VIEWS[4] to test the autocorrelation of residual to determine whether the result require second linear regression. In that case, we can obtain the most suitable parameters to improve the accuracy of our model.

4.2.2. Evaluation Model of Products

In order to observe the customer's evaluation of the products after they launch on the market, we first use k-Nearest Neighbor[5] (KNN) to classify and analyze the data we have. In that case, we can make an intuitive understanding of the customer's evaluation after purchase.

K-means clustering algorithm can divide the data into a specified number of classes we need. In that case, we divide the text-based and ratings-based data of a product into k clusters. Each cluster has five categories according to the star rating, and we can get the proportion of each category. Then we can calculate the overall rating of this product based on the clusters and the percentage of each category.

In order to obtain the clusters number k, we need to calculate the average distance $a(i)$ from sample i to other samples in the same cluster and the average distance $b(i)$ from sample i to all samples of other clusters. Then we can obtain the Silhouette Coefficient $s(i)$.

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{5}$$

$s(i)$ closing to 1 means the sample's cluster is reasonable. Average the Silhouette Coefficient of all points as the total Silhouette Coefficient and choose the highest one as the value k.

$$R_{PROD} = \sum_{i=1}^{5} P_i \cdot r_i \tag{6}$$

R_{PROD} represents the overall rating of the product. P_i is the proportion of each cluster. r_i is the star rating corresponding to each cluster.

Once we obtain the overall rating of product, we can calculate the degree of success of each product as follows.

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad (i = 1,2,3,4,5) \tag{7}$$

C_i represents the category we divide. μ_i is the centroid of C_i.

$$d = \sqrt{(x - \mu_i)^2} \tag{8}$$

$$c = \ln\left(\frac{1}{d}\right) \tag{9}$$

d is the Euclidean Distance between x and μ_i. c is the credibility of the data.

$$S = \frac{1}{n} \frac{\sum_{i=1}^{n} (c_i \cdot R_{PROD})}{\sum_{i=1}^{n} c_i} \tag{10}$$

S represents the degree of success of the product. n is sales volume of the product.

4.2.3. LSTM-based Reputation Predicting Model

In this section, we will analyze the change of each label over time, so as to measure the change of product's reputation. In order to obtain an accurate model of label changes over time, our team use the Long-Short Term Memory[6] (LSTM) algorithm to analyze the problem. The general idea of this predicting model is as follows.

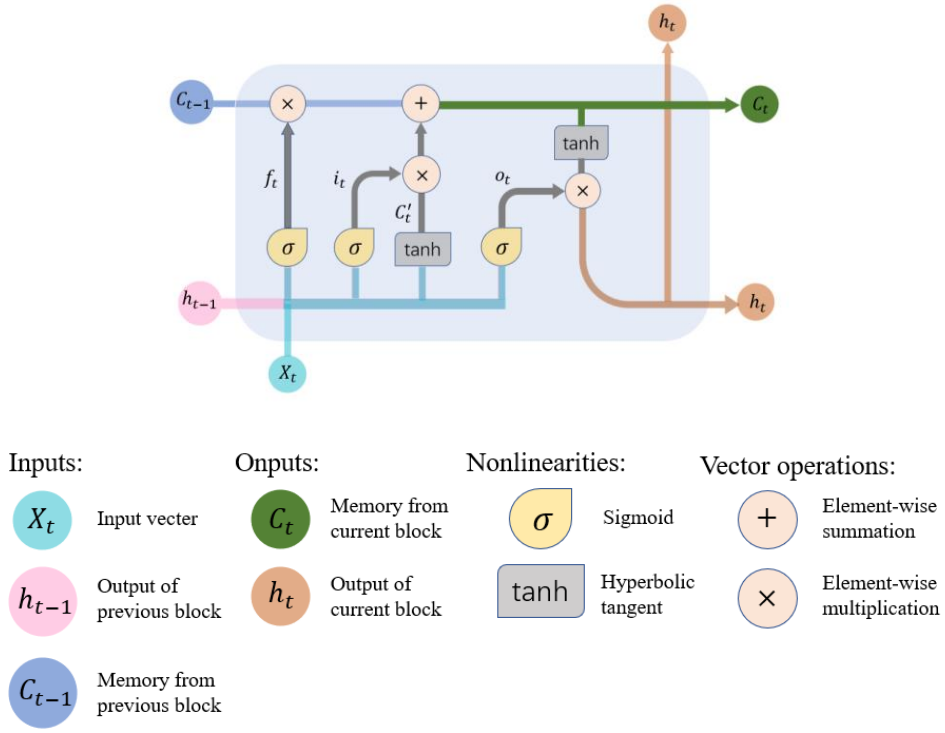


Figure 5. The process diagram of LSTM

1) Input the feature vector of three different products and set the forget gate. The parament f_t represents the value of forget gate.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \ + \ b_f\right) \tag{11}$$

2) Then the input gate is used to update the unit status.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] \ + \ b_i) \tag{12}$$

$$C'_t = tanh \cdot \left(W_C \cdot [h_{t-1}, x_t] \ + \ b_f\right) \tag{13}$$

$$C_t = f_t * C_{t-1} + i_t * C'_t \tag{14}$$

The parament i_t represents the value of input gate.

C_t represents the current cell status and C'_t represents the previous cell status.

3) The gate that controls the output of C_t is called an output gate, which is represented by o_t. The parameter h_t represents the hidden state, which contains information about the previous input. More importantly, the hidden state can be used for predicting the trend of each label. By analyzing the trends, we can draw a conclusion about the reputation of the product.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (15)$$

$$h_t = o_t * tanh \cdot (C_t) \quad (16)$$

5. Implementation and Results

We use Matlab and python to realize the model programmatically and get a set of virtue data, then we process the data to see the relationship among the labels.

5.1. Regression Analysis and Data Testing

In this section, we illustrate the result of our model from regression analysis, data testing and accuracy test. The first step is to calculate the correlation coefficients $r(x, y)$ in equation (1). $r(x, y) > 0$ represents positive correlation and $r(x, y) < 0$ represents negative correlation. A higher general correlation coefficient indicates a closer relationship between the two variables.

The result of calculating is displayed in figure 6. Through analyzing we discover that the helpful vote data has a strong correlation with total vote, and the correlation coefficients among star ratings, reviews, and helpfulness ratings are above 0.4. Therefore, we can draw a conclusion that these three labels have positive correlations.

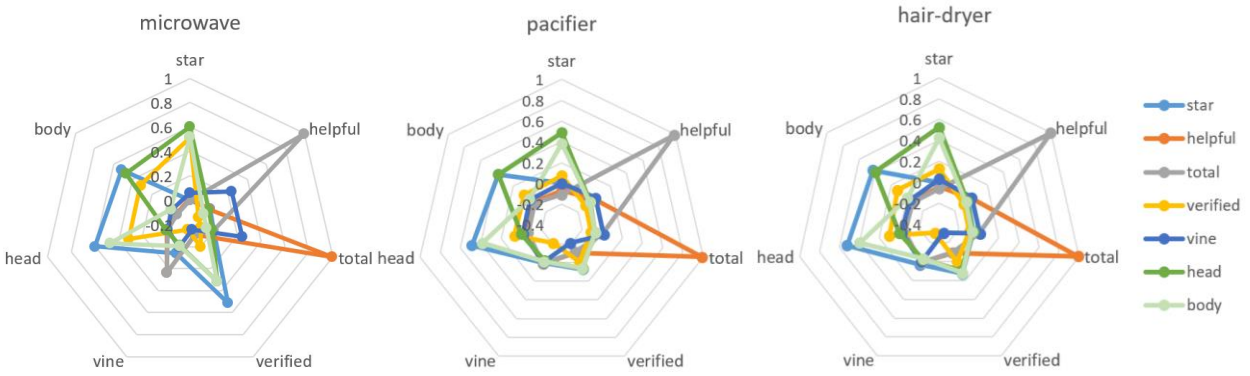


Figure 6. The Correlation Coefficient Among Different Labels

Then we use linear regression to obtain the mathematical relationship among the data. Through processing we acquire the linear relationship among the data and their mathematical expressions. The following picture is the result of hair-dryer after linear regression.

$$y = k_1 x_1 + k_2 x_2 + b \quad (17)$$

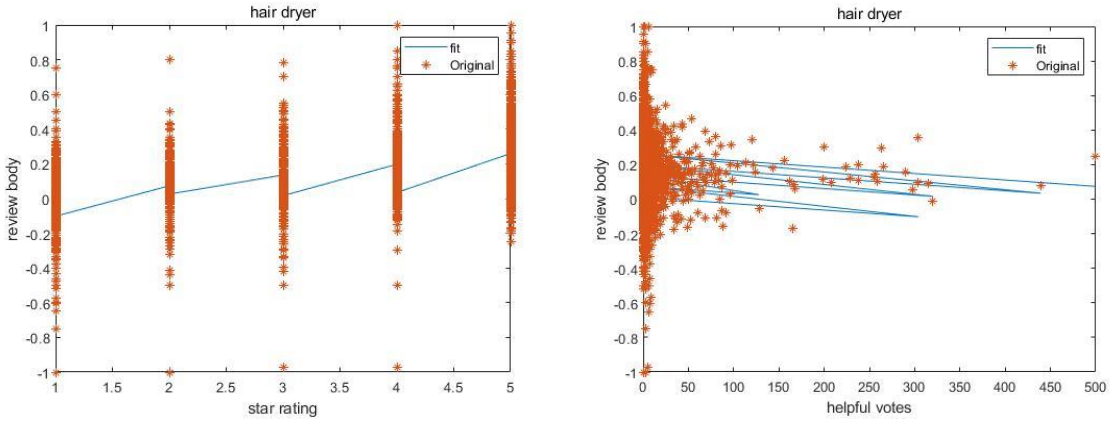


Figure 7. First Linear Regression of Hair-dryer

The linear regression is based on the assumption that the mean of b is zero and it is a random variable with equal variance and obeying normal distribution $N(0, \sigma_2)$. To analyze whether the assumption is correct or not, the residual analysis is needed. If the assumption is correct, all points in the residual plot fall in the middle of a horizontal band. The following picture is the result of hair-dryer after residual analysis.

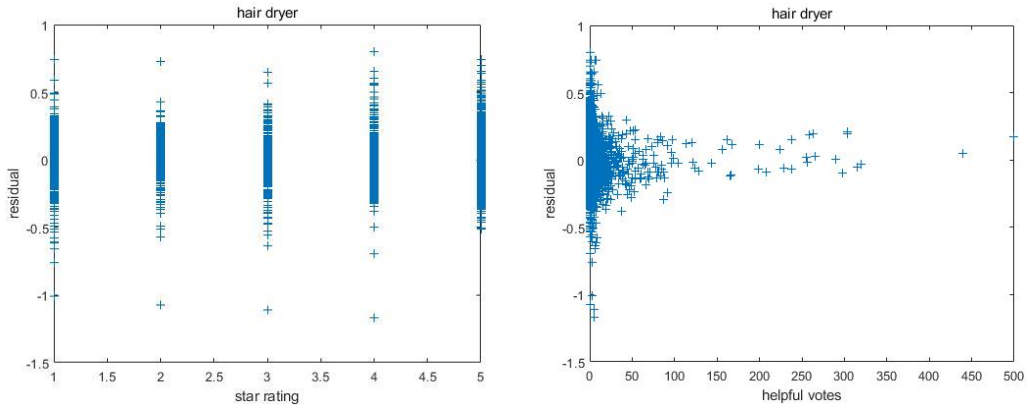


Figure 8. The Result of Residual Analysis

After testing, the autocorrelation exists in the result so we do the second linear regression to gain the accurate mathematical expression of different labels. The coefficients of each label in the mathematical expression are given in table 1.

$$\dot{y} = k_1 \dot{x_1} + k_2 \dot{x_2} + b \tag{18}$$

$$\dot{y_t} = y_t - p * y_{t-1} \tag{19}$$

$$\dot{x_{1t}} = x_{1t} - p * x_{1,t-1} \tag{20}$$

$$\dot{x_{2t}} = x_{2t} - p * x_{2,t-1} \tag{21}$$

	b	*Confidence bound*	k_1	*Confidence bound*	k_2	*Confidence bound*
Microwave	-0.014	(-0.035,0.006)	0.064	(0.051,0.077)	-2.472	(-6.511,1.567)
Hair-dryer	-0.009	(-0.016,0.002)	0.061	(0.048,0.072)	-4.329	(-8.311,-2.673)
Pacifier	-0.004	(-0.012,0.004)	0.058	(0.044,0.072)	5.279	(-5.982,7.038)

Table 1. The Coefficients of Different labels in Mathematical Expression

5.2. KNN Classification and K-means Clustering

By analyzing the text-based data, ratings-based data and other labels, we can acquire costumers' satisfaction with the product. In order to make the results more intuitive, we use star ratings as the standard to quantify costumers' satisfaction and obtain a chart as follows. Analysis shows that costumers have the worst satisfaction with microwave, which indicates that company should make appropriate adjustments to the product. By comparing the results with the star ratings of products, we can get the accuracy of our evaluation in table 2. The results show that the accuracy of our evaluation reach 70%, and the accuracy of pacifier's evaluation is nearly 80%.

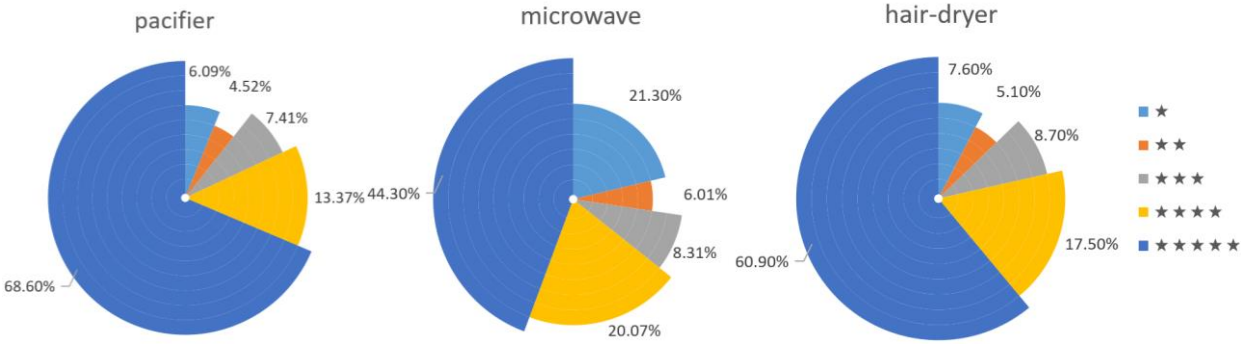


Figure 9. The proportion of different star ratings in three products

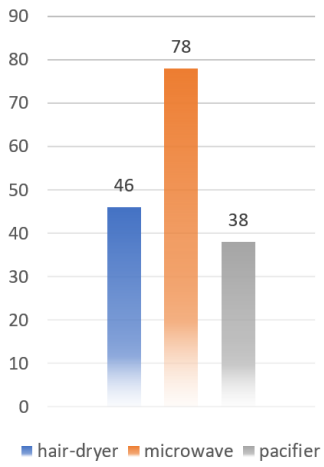
	pacifier	*microwave*	*hair-dryer*
Accuracy	0.7917	0.6929	0.7010

Table 2. The Accuracy of KNN Classification

Based on the result of LPA model and KNN classification, we can obtain a preliminary judgment on the customers' satisfaction of the product. Our task is to measure whether the product is successful or not and obtain a data-based measure for company to track, so we adopt k-means clustering algorithm to achieve it.

Using k-means clustering algorithm we divide the hair-dryer, microwave and pacifier into 46, 78 and 38 clusters, which is displayed in figure 10. Each cluster has a specific value, which represents the corresponding overall rating of the product. After clustering the categories of different products, we obtain their overall rating that can be used as the most informative measure for Sunshine Company to track.

To demonstrate our result intuitively, we sort the product categories based on the number of reviews, and list the top four categories of the three products separately in table 3. In the overall rating, the full mark is 5 and lowest mark is 1.



Hair-dryer				
Product_parent	732252283	955015830	531479992	127431946
Ratings	5	4.7281	1.0025	4
microwave				
Product_parent	423421857	544821753	109226352	295520151
Ratings	3.8521	4.2542	4.4414	2.2568
pacifier				
Product_parent	246038397	392768822	450475749	337629353
Ratings	5	4.8957	3.6695	1.5262

Figure 10. The number of clusters

Table 3. The overall ratings of four categories

From table 3, we discover that the ratings of microwave's top four categories generally lower than the other two products, which match the result of KNN Classification. Besides, the category with product_parent 531479992 in hair-dryer and the category with product_parent 337629353 in pacifier have a bad reputation. Generally speaking, the overall ratings can clearly reflect the quality of products.

In order to indicate a potentially successful or failing product, we construct the index S in equation (10). Considering the different clusters in different product, the success indicators of the three products are also different.

Based on the clustering and reviews analysis, we set different standards for three products. In hair-dryer, S above 0.8 means the product is successful, the more the better. In microwave, S

above 6.0 means the product is successful. In pacifier, the standard is 0.9. To demonstrate our result intuitively, we also choose four categories from the three products separately in table 4. Besides, we also draw a graph (figure 11) of different categories in order to contrast the success indicator of one category with others. Obviously, products A, E and F are more successful than other displayed products.

hair-dryer				
Product_parent	732252283 (A)	47684938 (B)	758099411 (C)	197856712 (D)
Degree of success	0.9990	0.8858	0.8087	0.7258
microwave				
Product_parent	423421857 (E)	544821753 (F)	109226352 (G)	295520151 (H)
Degree of success	6.5841	6.5873	6.4697	5.5898
pacifier				
Product_parent	246038397 (I)	392768822 (J)	572944212 (K)	911821018 (L)
Degree of success	1.1720	1.0042	1.3080	0.9985

Table 4. The success indicator of four categories

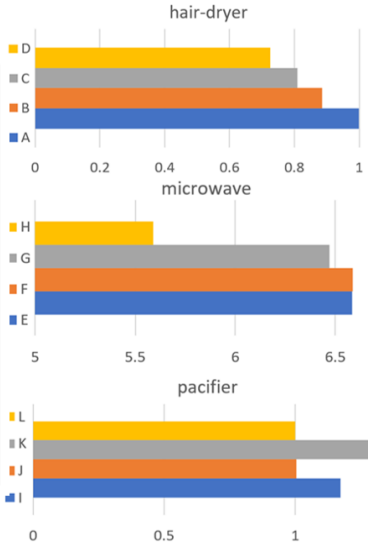


Figure 11. Indicators of success

5.3. LSTM-based Reputation Predicting

In this section, LSTM is adopted to anticipate the changes in product reputation. Since the data of products is a time-dependent sequence, we can take advantage of the characteristics of the data in the past to predict the change of the data in the future. To analyze the patterns within each data, we use the overall rating obtained in section 5.2 as the parameter to predict. The result of hair-dryer is showed in figure 12.

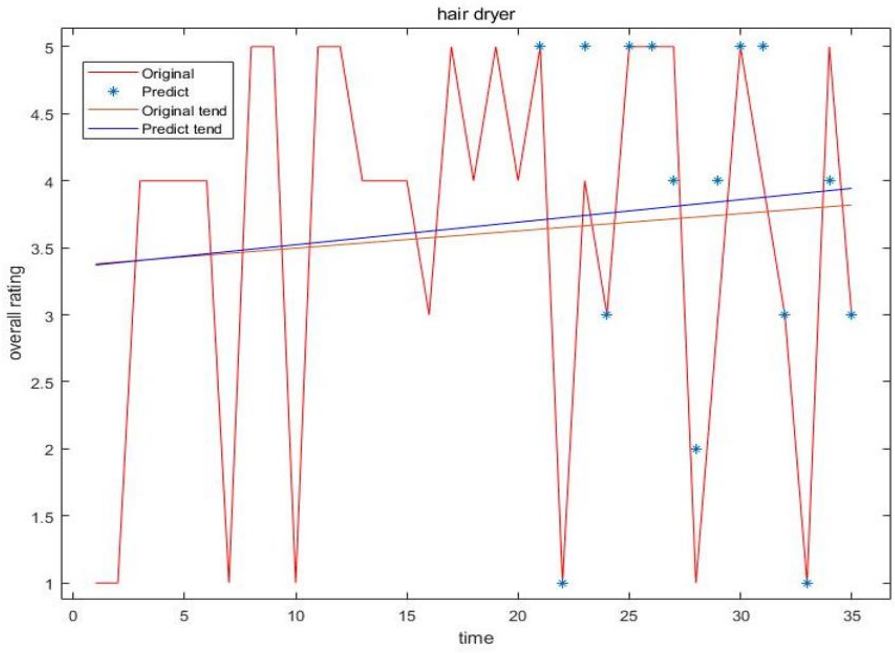


Figure 12. The result of LSTM prediction and its linear regression

To verify the accuracy of the results, the SSE(The sum of squares due to error) and MSE(Mean squared error) of three product are given in figure 13.

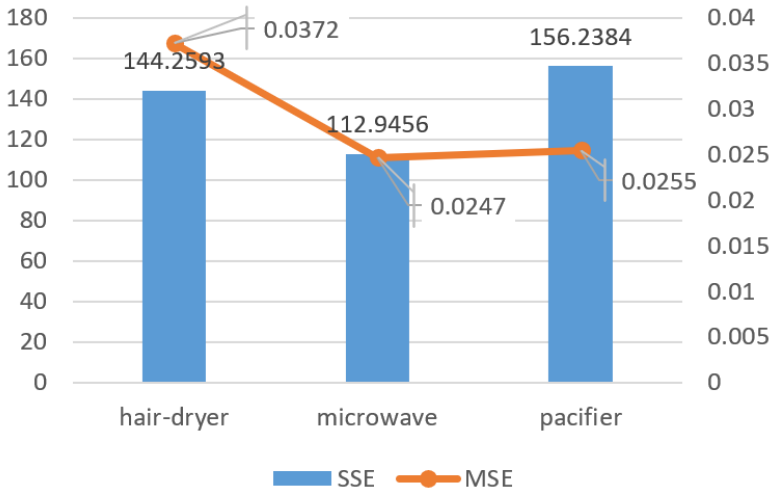


Figure 13. The SSE and MSE of three products

Given the amount of data we have, the value of SSE and MSE is reasonable, which indicates that our Reputation Predicting model (RP) is accurate enough. Though the overall rating seems random and discord, we can still analyze its trend by using linear regression. The result of linear regression is displayed in figure 12, and we can draw a conclusion that the reputation of hair-dryer is increasing.

5.4. The Influence of Specific Labels

5.4.1. The Connection Between Specific Star Ratings and Reviews

To explore the connection between specific star ratings and reviews, we observe the change of costumer evaluation before and after a vine comment the product. Figure 14 shows the change of hair-dryer.

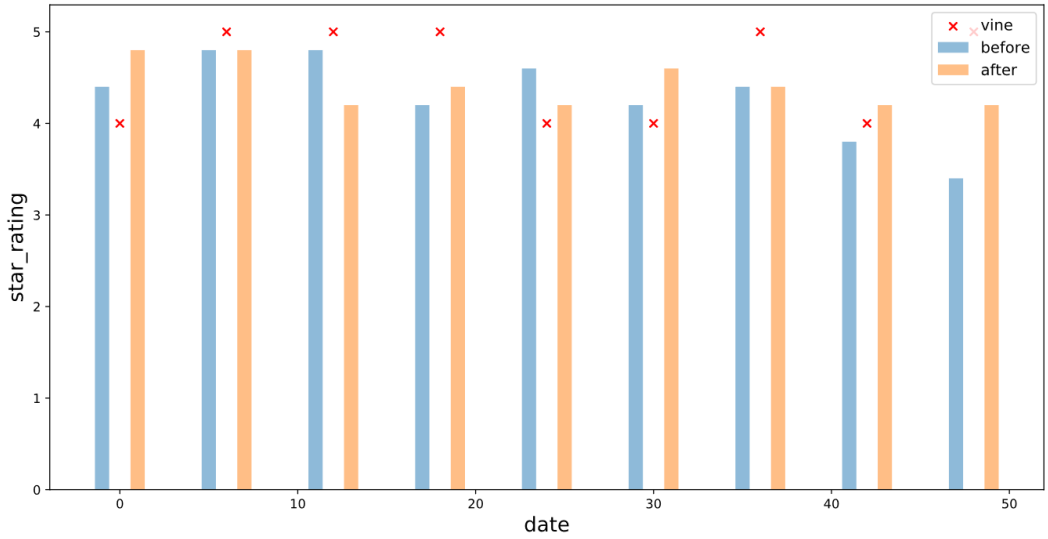


Figure 14. The star ratings before and after a vine comment in hair-dryer

Through analyzing the chart, we obtain that in most case, when the vine gives a low score, the costumer ratings will also decline; when the vine gives a high score, the costumer ratings also rise. While in some case, costumer ratings show opposite changes from vine. Besides, the ratings also decline after lots of negative reviews appear, and the ratings will rise after a large number of positive reviews appear.

In general, specific star ratings does incite more reviews and it also influence costumer ratings showed by making corresponding changes.

5.4.2. The Relationship Between Specific Descriptors and Rating levels

In this section, we will discuss the relationship between specific quality descriptors of text-based reviews and rating levels.

To analyze the relationship accurately, our team pick out ten specific quality descriptors based on the result of sentiment analysis, of which the sentiment value ranges from 1 to -1. Then we describe them with their overall ratings in figure 15.

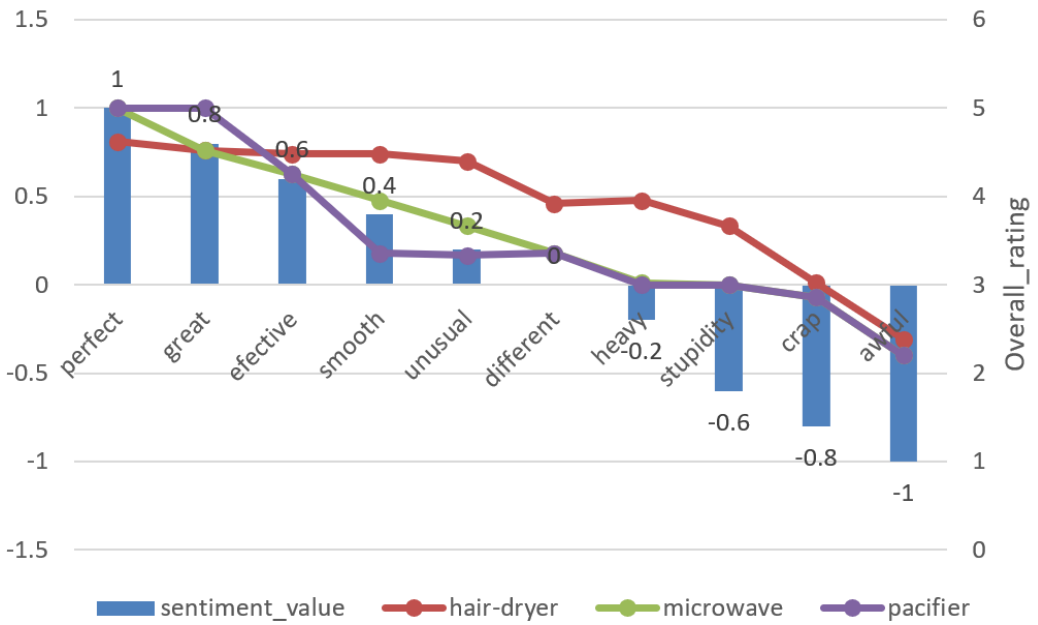


Figure 15. The diagram of sentiment value and overall rating

After analyzing, we discover that with the decline of the sentiment value, the overall ratings all show a downward trend. Besides, the difference in evaluation level of the three products is small when the sentiment value is close to 1 or -1, but when the sentiment value is close to 0, the difference is large. Besides, the overall ratings of hair dryer is higher than microwave and pacifier.

In conclusion, there is a correlation between the evaluation level of a product and the sentiment value of quality descriptors.

6. Conclusion

6.1. Strengths and Weaknesses

6.1.1. Strengths

a) We design a comprehensive and intuitive evaluation index for company to track their products.

b) Our models are adaptable to different conditions. The three sub models can still provide stable and accurate judgement for company though there may be some impacts.

c) Our models have broad application prospect. Based on machine learning, the evaluation model can provide more accurate results if there are more labels and data. The RP model can also predict the changes in sales.

6.1.2. Weaknesses

a) We adopt TextBlob to do sentiment analysis, but the results cannot reflect the sentimental bias of the text accurately. The analysis of some reviews is vague and unreliable.

b) In the LPA model, we process the data by using linear regression to obtain an intuitive result, but the mathematical expression of them is just our speculation, which may be a little different from reality.

c) In the RP model, we use LSTM to predict the change of reputation. However, considering the length of time-based data and the certain randomness of costumer comments, the accuracy of our forecast may reduce.

6.2. Future Work

Since we still have some weaknesses in our current work, future efforts are needed to better work out this problem.

Firstly, more accurate Natural Language Processing method should be used to analyze the sentiment in reviews headline and reviews body.

Secondly, more factors should be considered in LPA in order to increase the accuracy of the model. Besides, more reasonable formula of regression should be undertaken.

Finally, we will try to generalize our model and approach in this problem to more complex application scenarios after more punishments.

Letter

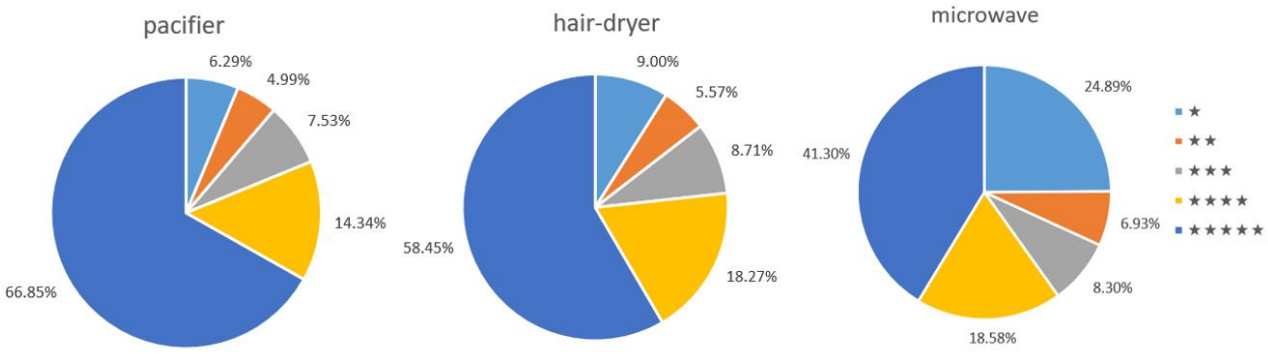
Dear Marketing Director:

It is my pleasure to give recommendation to Sunshine Company. We have analyzed the data offered by company and here are the results and our recommendations.

To begin with, our team analyze the reviews of each product and discover that there are some important design features that would enhance product desirability.

- For hair dryer, the weight of the product and the temperature of the wind have a great impact on the consumers' experience. To attract more consumers, the weight of the hair dryer should be reduced, and the temperature of the blowing wind should be more suitable for use.
- For microwave, a large number of costumers unsatisfied with the microwaves on the market because of their quality. If the quality of our company's microwave is better than the ones on sales, the product should be very popular.
- For pacifier, lots of costumers complain that the pacifiers on sales are too big for baby to use, offer different size of pacifier can be a beneficial measure for the product sales.

Besides, we adopt the methods of Machine Learning to establish a comprehensive index for company to track the products after they are placed on sale in the online marketplace. Furthermore, we analyze the reputation of three products on sales and the result is displayed in the chart.



As is shown on the chart, the microwave has the lowest reputation so it is a good chance to grab market share. Compared to microwave, pacifier and hair-dryer have already occupied large market share so that it may be a long way for company to go.

Thank you again for taking the time to read our suggestions. We sincerely hope that our recommendations will be helpful!

Best Regards,

Sincerely

Reference

1 WANG Q, WANG L, ZHANG X, et a1. The impact research of online reviews' sentiment polarity presentation on consumer purchase decision[J]. Information Technology & People, 2017, 30(3):522—541.

2 https://textblob.readthedocs.io/en/dev/

3 https://blog.csdn.net/u013129109/article/details/79896910

4 https://bbs.pinggu.org/thread-5144056-1-1.html

5 https://www.cnblogs.com/jyroy/p/9427977.html

6 Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory[J]. Neural Computation, 1997,9(8): p.1735-1780.