Problem Chosen	2020	Team Control Number
C	MCM/ICM Summary Sheet	2012797

Exploring hidden information in customer-supplied data

Summary

As more and more e-commerce platforms plunge into the fierce competition of online sales market, it has become essential for companies to dig out hidden valuable information in data gathered. Thus, we build several models to explore the interaction between customer-supplied data.

To begin with, we established a product popularity model which takes into account a product's total number of reviews (TNR) and average star rating (ASR). The weights of the two indexes are determined by entropy method, which are 0.82 and 0.18, respectively. Then the popularity score of a product is obtained by the weighted sum of the two indexes.

Next, we chose TNR as dependent variable, ASR, number of one-star rating reviews, review helpfulness (The ratio of helpful votes and total votes), number of verified purchases, total votes and variance of star ratings as independent variables, and used SPSS to conduct a multiple regression analysis. This model is based on the assumption that TNR can represent a product's sales. The result shows that the total votes a product received and the variance of the star ratings are the most informative indicators that Sunshine company should place much importance on.

Then, we built a reputation model based on time series. First, we calculated the sentiment score of a product's reviews, using the NLTK library and TensorFlow with Python. Then, the reputation score of a product is calculated by the weighted sum of ASR and the sentiment score by entropy method. Next, we selected four variables: review helpfulness, variance of star ratings, number of one-star rating reviews, and total votes, and used SPSS to calculate the Spearman correlation coefficients of them and the reputation score. We found that all of them, except total votes, can indicate a product's future reputation.

What is more, we found a successful way to combine text-based and data-based measures and predict a product's performance.

Besides, interestingly, we also discovered that low star ratings, one-star and two-star ratings especially, would stimulate more buyers to write reviews using regression analysis.

Finally, we obtained the most frequently appeared quality descriptors in each star rating reviews using the text mining and sentiment dictionary. We found that many words are highly connected to rating levels.

Furthermore, our model is proven to be robust by the sensitivity analysis. Only moderate changes to the result were noticed when we altered key parameters. Thus, we firmly believe that our model is valid.

Keywords: reputation model; regression analysis; sentiment score; text mining

Dear Marketing Director:

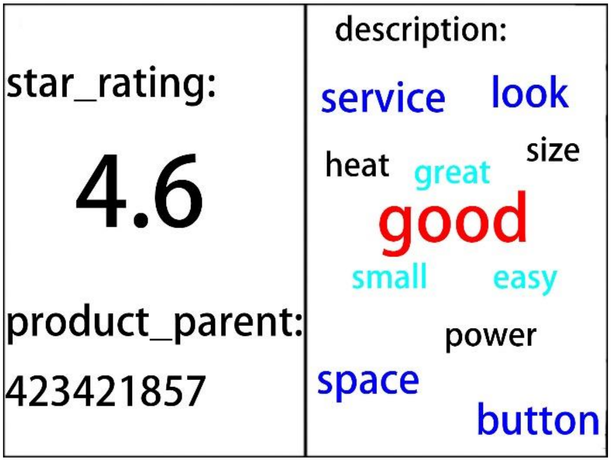
It is my pleasure to write you on behalf of the MCM team #2012797. We understand that you are very concerned about the valuable information hidden in the data of customer-supplied ratings and reviews. And after carefully analyzing the interaction within and between the data you provided for us with mathematical tools, we present you our results and offer you the following suggestions with confidence.

In order to help you better evaluate and predict a product's popularity, we built a product popularity model. Using the MATLAB codes we wrote, you will be able to obtain the popularity score of a product.

After your company's new products come to the online market, we sincerely advise you to pay attention to the total votes a product received and the variance of the star ratings, as we have discovered that they are the most informative indicators of a product's reputation. The increase in total votes and decrease in the variance of the star ratings suggest that a product's reputation is getting better.

Besides, we believe that your company should encourage customers to write more sentimental reviews, as they are more likely to be voted as helpful, and incur more potential customers. It is also a good idea to absorb more vine members, since the helpfulness of their reviews is higher than others.

What is more, we found a successful way to predict a product's performance with combined text-based and data-based measures, which is creating a product identification card like this:



Finally, we obtained the most frequently appeared quality descriptors from one to five star rating reviews using text mining and sentiment dictionary. We found that negative words like " bad ", " disappointed ", and " heavy " usually appeared in one and two-star rating reviews. " Good ", " great ", " love ", and " nice " frequently come in high star rating reviews. We hope that recognizing this will help you improve your products.

We wish our suggestions can be of help for your company's future online sales strategy and genuinely hope you adopt them.

Your sincerely

MCM team #2012797

Contents

1 Introduction

1.1 Problem Background

With the rapid development of electronic commerce, online shopping has gained much popularity among all walks of life. E-commerce platforms have developed and adopted various methods to attract buyers and facilitate their decision-making process. One significant way that Amazon uses is to allow customers to give data-based and text-based feedback called "star ratings" and "reviews", respectively. Besides, other customers can also rate these reviews as helpful, known as "helpfulness ratings", if they find them useful in reducing the information asymmetry as they shop [1]. Along with this, companies can gather a wealth of data that will provide them with valuable information if used properly.

Previous scholars have conducted lots of researches on what reviews are more likely to be voted as helpful. For example, Wu jiang et al discovered that the credibility of the writer, the amount of information the review contains and the extremity of the review have a positive impact on the possibility of a review being voted as helpful [1]. Korfiatis et al [2] found that the readability of a comment is more important than its length.

However, few researches have used combined data of star ratings, reviews, and helpfulness ratings to discover the hidden patterns, relationships, measures, and parameters within and between them.

1.2 Our Work

We are hired by Sunshine company to explore key patterns, relationships, measures, and parameters in their previous customer feedback, so as to help them gain success in the release of their three new products in the online market, which are a microwave oven, a baby pacifier, and a hair dryer.

Firstly, we build a product popularity model in order to study how ratings and reviews of products can show and influence their future popularity. Next, we assumed that the numbers of reviews can represent the sales of products, and built a multiple regression analysis model to find how other variables influence the number of reviews through quantitative analysis. A reputation model based on time series is also established. What is more, we found a successful way to combine text-based and data-based measures and predict a product's performance.

Finally, we obtained the most frequently appeared quality descriptors in each star rating reviews using the text mining and sentiment dictionary.

2 Assumptions

We make the following assumptions to simply our model:

- The review date is consistent with the purchase date, that is, review date can be used to substitute purchase date.

- The number of reviews is positively related to that of purchases. Namely, the more review one product has, the more it has been sold.
- Every customer browse reviews before purchase.
- The influence of product price will not be taken into consideration. It is possible that when customers shop for expensive goods, they may pay more attention to reviews, thus leading to more helpfulness votes on those reviews. But since we were not provided with relevant data, we will not consider this factor.

3 Nomenclature

In this paper, we use the nomenclature in Table 1 to describe our model.

Table 1: Nomenclature

Symbol	Definition / Description
X	The average star rating of a certain product
N	The sum of reviews a certain product gets
S	The popularity score of a product
ρ_{jX}	The proportion of the j th product under index X
ρ_{jN}	The proportion of the j th product under index N
E_X	Information entropy E_X
E_N	Information entropy E_N
W_k	The weight of index k based on information entropy
$help$	The ratio of total helpful votes a product received and its total votes
var	The variance of each product's star ratings
$lowrate$	The number of one-star ratings pertain to a product
vp	The number of reviews given by people who were not verified by Amazon to have purchased the product on its website
tv	Number of total votes reviews received

4 Model I- A product popularity model based on entropy weight method

As we know, a product's sales records demonstrate its popularity from the time it was released to the current time. However, sometimes it is observed that a hit product may suffer from sales reduce due to the decrease of its rating and bad reviews. On the other hand, products that were not sales champions in the past may witness a sales boom with the improvement of its feedback. Thus, we built a product popularity model based on entropy weight method in order to study how ratings and reviews of products can show and influence their future popularity.

4.1 Additional assumptions

（1）The number of reviews pertain to a certain product has a positive impact on users' choice of the product.

（2）The average star rating of a product has a positive influence on the product' popularity.

4.2 Indexes chosen

We classify products that belong to the same product parent but may have different product ids into one type of product in order to reduce the number of products.

4.2.1 The average star rating of a product (ASR)

As we assumed, the average star rating of a product shows its reputation among all purchasers, and the better the reputation is, the more customers will be attracted. Suppose a product has *n* star ratings, and x_i represents the *i*th star rating of the product, then:

$$X = \frac{\sum_{i=o}^{n} x_i}{n} \quad (1)$$

4.2.2 The number of reviews pertain to a certain product N

We presume that a product that sales well tends to have more comments, and there is a positive correlation between the two based on life experience.

4.3 Content of the model

The entropy weight method is adopted to obtain the corresponding weight of X and N in this model, which are W_X and W_N. And then the popularity score of a product is obtained by calculating the weighted sum of its X and N.

The process of the entropy weight method is as followed:

Suppose there are *n* products in each category, two positive indexes, X and N, have been adopted.

(1) Data normalization: $X = \{X_1, X_2, X_3 \dots X_n\}$, $N = \{N_1 N_2 N_3 \dots N_n\}$. X_j, N_j stand for the average star rating and review numbers of the *j* th product in each category. The normalized data are X^*, N^*, then:

$$X^*{}_j = \frac{X_j - min(X)}{max(X) - min(X)} \quad (2)$$

$$N^*{}_j = \frac{N_j - min(N)}{max(N) - min(N)} \quad (3)$$

(2) Calculate the proportion of the j th product under index X and N:

$$\rho_{jX} = \frac{X_j^*}{\sum_{j=1}^{n} X_j^*} \quad (4)$$

$$\rho_{jN} = \frac{N_j^*}{\sum_{j=1}^{n} X_j^*} \quad (5)$$

(3) Calculate the information entropy E_X, E_N. The smaller the E of an index is, the greater its degree of variation is, the more information it provides, the larger its influence becomes, and the greater its weight is. And vice versa:

$$E_X = -\frac{1}{\ln(n)} \sum_{j=1}^{n} \rho_{jX} \ln(\rho_{jX}) \ (6)$$

E_N can be obtained using the same equation.

(4) Calculate the weight of indexes based on information entropy

$$W_k = \frac{1 - E_k}{2 - \sum E_k}, (k = X, N) \ (7)$$

(5) Calculate the popularity score of a product.

$$S = W_X \times X + W_N \times N \quad (8)$$

4.4 Result of the model

After running the data provided for us in MATLAB, we successfully obtained 5 most popular products from each category. The results can be seen as followed:

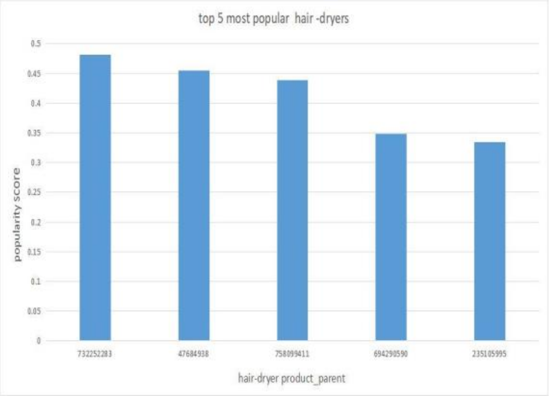


Figure 1 top 5 most popular hairdryers

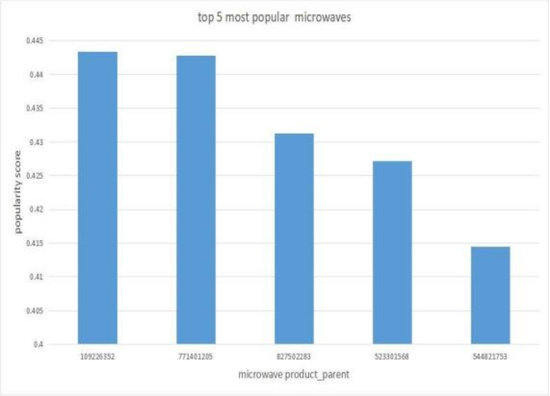


Figure 2 top 5 most popular microwaves

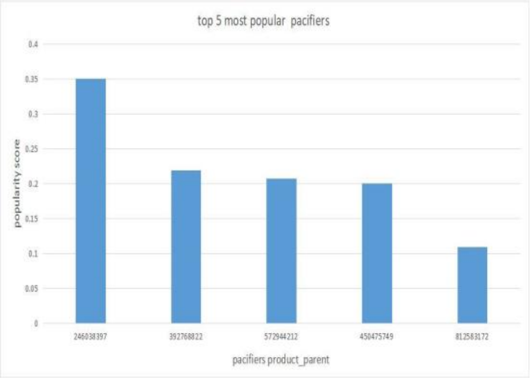


Figure 3 top 5 most popular pacifiers

$$S = 0.82 \times X + 0.18 \times N \quad (9)$$

Based on our model, these products will be more popular in the future than others, in later part of our study, we will study what features these products have, so as to inform Sunshine company their sales strategy.

4.5 Sensitivity analysis

We altered the key parameters in this model, and only moderate changes to the result were noticed. Thus, we believe that this model can fit other data in real life very well.

5 Model II- identifying factors influencing product sales via multiple regression analysis

5.1 Building process

We built a multiple regression analysis model based on the assumption that the numbers of reviews can substitute the number of products sold. Our purpose is to find how other variables influence the number of reviews through quantitative analysis, and find the data-measure that has the biggest impact on sales. Then, we can advise Sunshine company to pay more attention to this measure.

5.1.1 Dependent variable-*Y*, the sum of reviews

Based on previous assumptions, we use the sum of reviews Y to represent the sales of a product, and see Y as the dependent variable in our model.

5.1.2 Independent variables

- X -the average star rating.

X can directly give customers an impression of one product’s performance, thus, we assume X can have an impact on the sales of products.

- *help*- helpfulness ratio. Namely, the ratio of total helpful votes a product received and its total votes:

$$help = \frac{helpful\ votes}{total\ votes}\ (10)$$

- *var*- the variance of each product's star ratings:

$$var = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\ (11)$$

A big variance suggests that big disparity exists in customers' opinions on a product, and this may be caused by brands hiring people to promote products. So we hope to find out its influence on the sales, in order to alter our online sales strategy.

- *lowrate* -the number of one-star ratings pertain to a product.

Some researchers have found out that customers are more likely to be influenced by bad reviews [3], thus, we adopt lowrate as one of the independent variables.

- *vp*- the number of reviews given by people who were not verified by Amazon to have purchased the product on its website.

Those reviews may be given by people who shopped on other platforms and received a deep discount, and they come to leave a comment on Amazon because they have a strong opinion that they want to share.[1]

- *tv*-number of total votes reviews received.

It is predicted that the more helpfulness votes a product's reviews receive, the more helpful these reviews are in assisting other customers' decision-making process, thus, the more likely this product will be sold.

5.1.3 Content of the model

$$Y = \beta_1 X + \beta_2 help + \beta_3 var + \beta_4 lowrate + \beta_5 vp + \beta_6 tv + \varepsilon\ \ (12)$$

5.2 Data selection

We employ SPSS to process data, and we only select data from products that have more than ten reviews.

5.3 Conclusion

When data are selected from products with more than ten reviews, the result we get with the help of SPSS is:

$$Y = 0.09X + 0.028help - 0.116var + 0.234lowrate - 0.50vp + 0.886tv\ \ (13)$$

	R	R^2	Adjusted R^2	Std.error	Durbin-Watson
1	.887[a]	.787	.760	26.93434	2.167

Table 1

From Table 1, we can see that the adjusted R^2 is 0.760, which means that the independent variables we chose can account for 76% of the dependent variable. DW test is also passed. From this aspect, our model is very convincing.

Anova[a]

	quadratic sum	df	Mean square	F	Sig.
Regression	128444.718	6	21407.453	29.509	.000[b]
Residual error	34822.010	48	725.459		
Sum	163266.727	54			

Table 2

From Table 2, we can see that our model also passed F test with a Sig less than 0.001.

		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		β	standard error	Standardized β		
1	ε	-13.022	14.497		-.898	.374
	X	4.059	3.837	.090	3.058	.000
	help	5.145	13.886	.028	3.370	.000
	var	-6.293	4.350	-.116	-4.447	.001
	lowrate	1.560	1.483	.234	5.052	.000
	vp	-.247	1.110	-.050	-.222	.825
	tv	.149	.016	.886	9.440	.000

Table 3

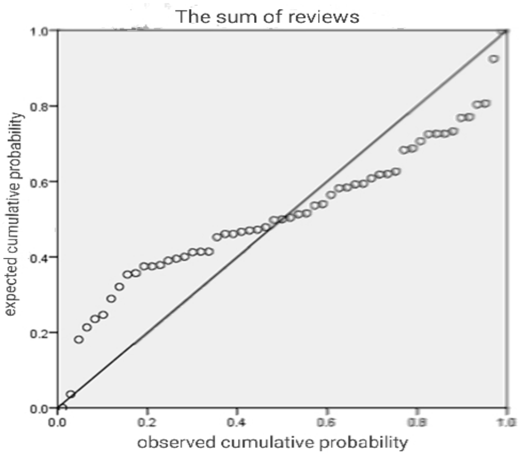


Figure 4 The sum of reviews

From Table 3, we can learn that ε should be removed from the model because its sig is much higher than 0.05. And all variables except *vp* have passed T test, which means that *vp* has no prominent influence on *Y* and should be eliminated from the function. This means that whether a review writer is verified to have bought the product on Amazon does not make much difference. A possible explanation is that these people may have purchased the product elsewhere, some even received a big discount, but they have such a deep impression or opinion on it which they want to share on Amazon. So these comments will also be accepted by other viewers.

Moving to standardized β, we can see that the β of *tv* is the largest among all, the second largest is that of *lowrate*, and then *var, X*, and *help*. We can reason that when a product's reviews get more votes, whether those votes are for helpful or not, it means that the product has received more attention, indicating that the product is more popular and therefore easier to sell.

One-star ratings are seen as extreme feedback, the text reviews of them are usually filled with more strong emotions. Scholars have pointed out that buyers are more likely to be influenced by extreme reviews [3]. However, as our results indicate, *lowrate* has a positive contribution to the sales of product, which clearly disagrees with our life experience. This leads to our later study on whether certain star rating can incur more reviews.

Var suggests that big disparity in different customers' reviews will exert negative influence on the product's sales. We think that this can happen when a company hires people to give positive reviews, whereas real buyers leave opposite reviews, and when fake reviews are being recognized, other purchases would feel beguiled and decide to not buy the product.

X shows that high average star ratings promote product sales. Last but not least, *help* indicates that the helpfulness of a product's reviews (the ratio of total helpful votes a product received and its total votes) is positively related to its sales. This can be interpreted as that when a product has more helpful reviews, those reviews will help customers make quicker decisions.

6 Model III- A time-based reputation model

6.1 Executive process of the model

To begin with, we gave a thorough consideration on what can represent a product's reputation.

There is no doubt that the ASR of a product can directly show the word of mouth (WOM) of it among customers. However, we decided to introduce another index, called sentiment score to achieve a better assessment of a product's reputation together with ASR, considering that products with the same ASR may have reviews that contain different amount of sentiment, which will also make a difference on these products' reputation.

Sentiment score can be calculated by using the NLTK library and TensorFlow with Python. Then, we use the entropy weight method we elaborated in the first model to obtain the weight of ASR and sentiment score. Finally, we calculate the reputation score of each product by the weighted sum of the two indexes.

Next, using the SPSS in quarterly time units, we obtained the curve of the reputation score of the products over time. Four indicators that might affect the reputation of a product were then selected: *help*, *var*, *lowrate,* and *tv* whose meanings have been introduced in model II. Then, we obtained their curves over time as well. If there is little data (under 5 reviews) about a product in a certain quarter, we will add these data to the next quarter.

Finally, we calculate the Spearman correlation coefficient of the four indicators and the reputation score to arrive at how relevant these indicators relate to a product's reputation score. The greater the correlation, the more these indicators are able to predict the change in the reputation of the product in the future. The math process is done with SPSS.

6.2 Results

Considering the difference characteristics of different products, we ran the model for each category, and obtained three figures (Figure 8-10) to illustrate our results:

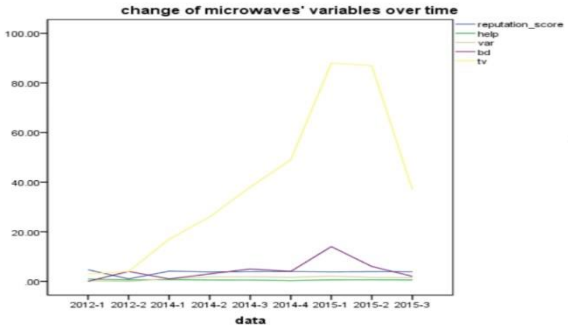


Figure 5 change of microwaves' variables over time

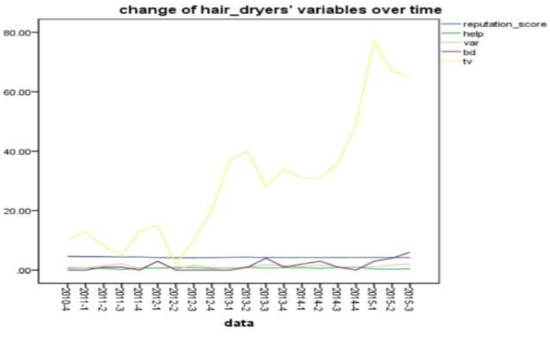
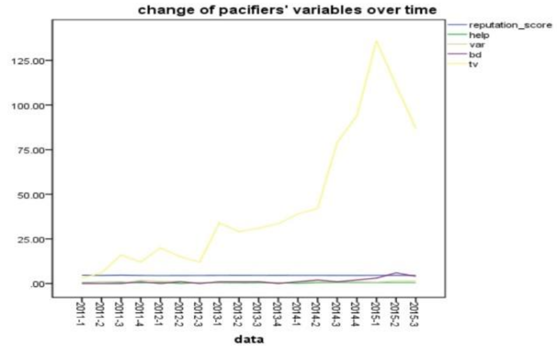


Figure 6 change of hair-dryers' variables over time



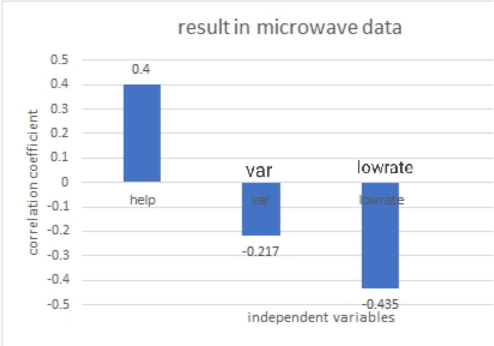
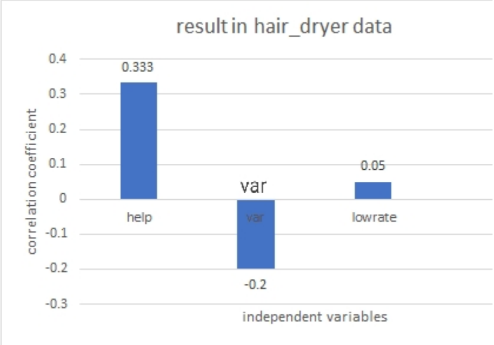


Figure 7 change of pacifiers' variables over time

Figure 8 result in microwave data



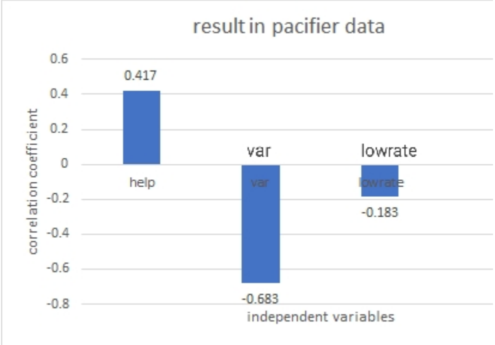


Figure 9 result in hair-dryer data

Figure 10 result in pacifier data

We can see from the curve that *tv* does not fit that of reputation score very well. A possible explanation is that *tv* may contain votes for both positive and negative reviews, thus, we do not think itself alone will serve as a good predictor. Besides, for all three products, *var* and lowrate has a negative impact on products' reputation, whereas *help* has a positive impact.

6.3 Conclusions

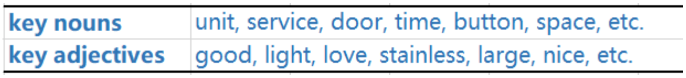
For companies, they should pay more attention to the variance of products' star ratings, the number of one- star ratings and the helpfulness ratio. If *var* and *lowrate* increase, that means the product's reputation is worsening, and they should take actions to meddle if necessary. For example, they can give customers certain credit if they raise the ratings of reviews. Besides, they should also try to encourage customers to write more helpful reviews. In order to achieve that, they can guide customers to write comments that are more sentimental and readable.

7 An indicator consists of text-based and ratings-based measures

7.1 The establishment of the model

- ✧ To begin with, we established a keyword dictionary.

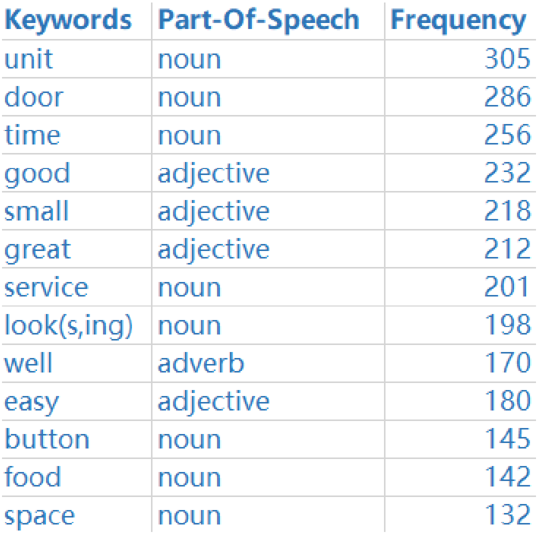
  Like this:



key nouns	unit, service, door, time, button, space, etc.
key adjectives	good, light, love, stainless, large, nice, etc.

Table 4 keyword dictionary

✧ Then, we count the frequency of each keyword, and created a frequency chart.



Keywords	Part-Of-Speech	Frequency
unit	noun	305
door	noun	286
time	noun	256
good	adjective	232
small	adjective	218
great	adjective	212
service	noun	201
look(s,ing)	noun	198
well	adverb	170
easy	adjective	180
button	noun	145
food	noun	142
space	noun	132

Table 5 frequency chart

✧ Next, we build a sentiment dictionary [4].

Keywords	Positive reviews	Total reviews	Positive ratio
Unit	245	305	80.3%
Door	205	286	71.7%
Time	158	256	61.7%
Service	124	201	61.7%
Look	148	198	74.7%
Button	92	145	63.4%
Food	85	142	59.9%
Space	95	132	72%
Heat	105	121	86.8%
Power	91	111	82%
Size	88	116	75.9%

Table 6 sentiment dictionary

✧ Then we conduct a keyword evaluation algorithm.

For each review that contains keywords, we:

a. Locate the keyword

b. Search for sentimental modifiers near the keywords. If a positive modifier is spotted, then the number of positive features of the product is increased by 1. If a negative modifier is captured, the number of negative features is increased by 1.

c. Loop a and b until the product's reviews have been traversed.

d. Count the number of positive and negative features of the product

7.2 Results

We create an identification card for each product to present its text-based and data- based characteristics. One example is:

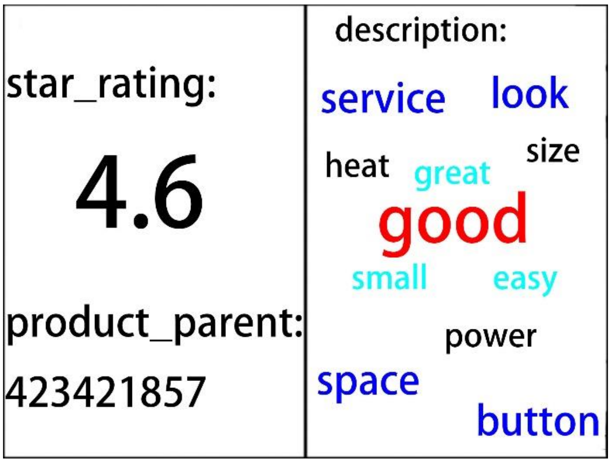


Figure 11: product identification card

8 Modelling the change of review numbers with regression analysis

8.1 Dependent variable- Increased number of product reviews during the quarter, Y_i

Y_i is the newly increased number of the product's reviews in the i th quarter.

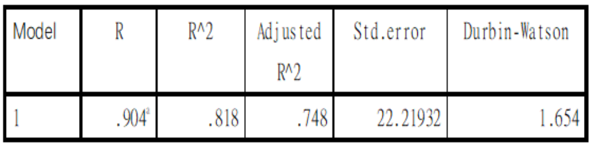
8.2 Independent variable X_{ij}

X_{ij} is the number of *j*-star reviews that the product have at the i th quarter. (*j*=1,2,3,4,5,)

8.3 Regression function

$$Y_i = \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{4i} + \alpha_5 X_{5i} + \varepsilon \quad (14)$$

8.4 Results



Model	R	R^2	Adjusted R^2	Std.error	Durbin-Watson
1	.904ᵃ	.818	.748	22.21932	1.654

Table 7

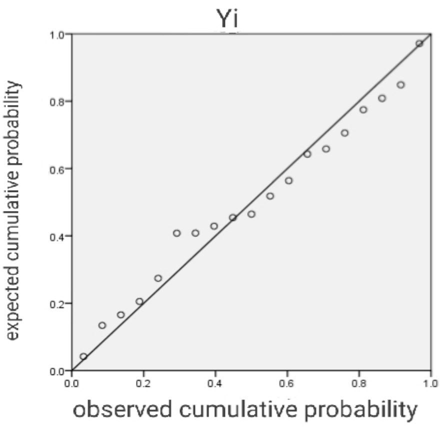


Figure 12

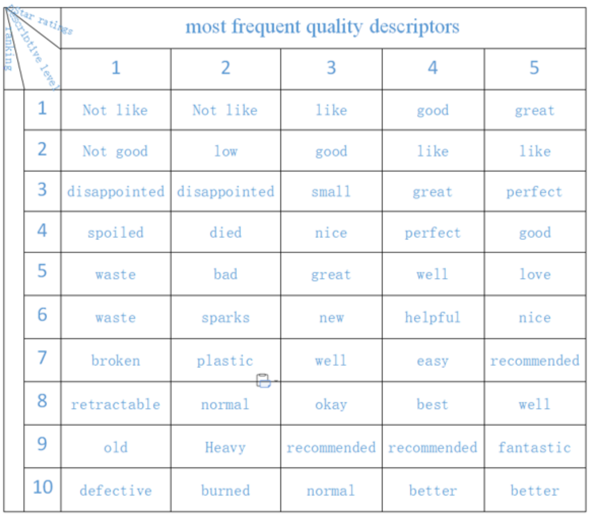
$$Y_i = 4.623X_{1i} + 2.821X_{2i} - 0.072X_{3i} + 0.286X_{4i} - 1.908X_{5i} \quad (15)$$

From the results of our regression model, we can see that the independent variables we chose can account for 74.8% of the dependent model, thus our model is proved effective.

What is more, we can also see that one-star ratings and two-star ratings can incite more reviews.

9 Association between quality descriptors and rating levels – a text mining method

Using the text mining method and sentiment dictionary mentioned in 7.2, we obtained the most frequently appeared quality descriptors in i-star rating reviews:



star ratings / descriptive level / ranking	most frequent quality descriptors				
	1	2	3	4	5
1	Not like	Not like	like	good	great
2	Not good	low	good	like	like
3	disappointed	disappointed	small	great	perfect
4	spoiled	died	nice	perfect	good
5	waste	bad	great	well	love
6	waste	sparks	new	helpful	nice
7	broken	plastic	well	easy	recommended
8	retractable	normal	okay	best	well
9	old	Heavy	recommended	recommended	fantastic
10	defective	burned	normal	better	better

Table 8 most frequent quality descriptors

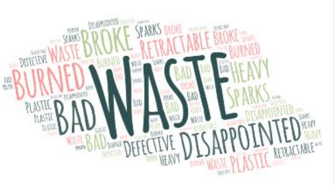


Figure 13 1-star reviews

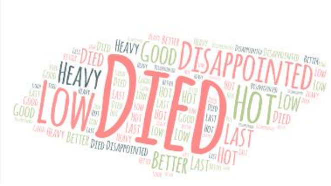


Figure 14 2-star reviews

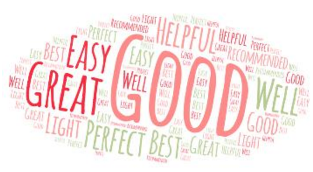


Figure 15 3-star reviews

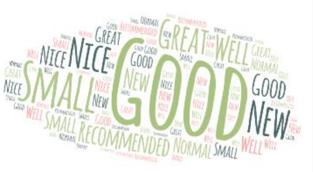


Figure 16 4-star reviews

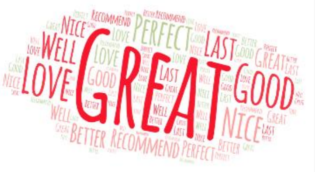


Figure 17 5-star reviews

We noticed that there are quite a few specific quality descriptors that strongly associated with rating levels, for example, " bad ", " disappointed ", and " heavy " usually appears in one and two-star ratings. " Good ", " great ", " love ", and " nice " are frequency noticed in four and five-star rating reviews.

10 Conclusions and Further Discussion

Due to limited time and data provided, there is still much space for improvement in our models. We listed the strength and weakness of our models below:

10.1 Strength

1. We considered a wild range of explanatory variables, which makes our model more convincing.
2. Our regression model fitted especially well.
3. The suggestions we give are highly informative and helpful.

10.2Weaknesses

1. We used the number of reviews to represent the number of sales. But that may not be exactly true in real life.
2. We did not consider the reviews the texts of which are not consistent with the ratings.
3. There is little data in certain period of time when we build the time measure, more data will be needed.

References

[1] Wu Jiang, Liu Wanwan. Identifying Reviews with More Positive Votes——Case Study of Amazon.cn[J]. *Data Analysis and Knowledge Discovery* ,2017,1(09):16-27.

[2] Nikolaos Korfiatis, Elena García-Bariocanal, Salvador Sánchez-Alonso. Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content[J]. *Electronic Commerce Research and Applications*, 2012, 11(3): 205-217.

[3] Cao Q, Duan W, Gan Q. Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach[J]. *Decision Support Systems*, 2010,50(2).

[4] Wei hui ling. Application research on text sentiment analysis in product reviews[D]. 2014, Beijing Jiaotong University.