Problem Chosen	2020 MCM/ICM Summary Sheet	Team Control Number
C		2012671

Stories that Amazon reviews tell: Marketing Strategies Development and Product Design Improvement

Summary

We first perform preliminary processing on the data. For text-based reviews, we consider the use of sentiment scoring with the sentiment dictionary to score the emotional level of each review between -5 and 5, and then multiply the degree coefficient given by the degree word in. the sentence. After emotionally scoring the review language, we extract the length of each review and use them as two characteristics to describe and examine the reviews further.

After quantifying each factor, we obtain a strong correlation between the factors through the correlation coefficient graph, thus we consider using LASSO to conduct a preliminary screening of variables. Then the filtered variables are used to perform linear regressions, which leads us to the conclusion that the validity of the reviews is mainly affected by discounts, review lengths and star ratings. To be more specific, the validity of reviews on microwave ovens and hair dryers is greatly affected by discounts, while that of pacifiers is mainly determined by the length (number of words) of reviews.

Notice that there might be a lot of hidden factors interacted with the number of helpful votes, it may be more helpful to take it as a two-level categorical factor. We label comments with 0 and 1, indicating the helpfulness of it based on number and ratio of helpful votes, for analysis. For binary dependent variable, we use generalized linear models for regression, and get a relatively more accurate prediction of whether the comment is helpful or not.

The observation of changes of various indicators of the product over time gives us a hint about their similarities in development. After the large fluctuations in the early period, the overall indicators of the product gradually stabilized. In order to reduce the influence of random factors, we perform a time series analysis after differentiating the original data. When performing ARMA analysis on a single factor, we found that its prediction effect was not satisfactory. Considering the possible interaction between various factors, we extend the ARMA model to multivariate analysis. The VARMA model was used to analyze the average sentence length and average star rating of the review indicators as variables. After selecting the appropriate parameters, prediction can be given and compared to the real data, and the result shows that the model fits and predicts well. Finally, we perform sensitivity analysis on the model. From the impulse response function graph, a certain lag effect among the factors can be found, but it tends to stabilize over a rather long period of time.

Finally, we analyze and summarize the advantages and disadvantages of the model, and the model results are combined to propose marketing strategies and recommendations for the company.

Keywords: online costumer review; sentiment analysis; time series analysis; ARMA; VARMA; generalized linear model

Content

1 Introductions

1.1 Background

With the development of technology, electronic retailers have changed our life, allowing customers to buy products without going outside, but generating problems of buying products different from the description at the same time. Thus, reviews have become a significant way to inform users the actual quality of goods, and companies can also get product feedback through online scoring.

In recent years, sentiment analysis has been introduced into product review analysis, to quantify customers' satisfaction. Countless review data from Amazon contains huge information, where the quality level and key characteristics of products can be obtained through product reviews and star ratings, to help companies make better operating decisions.

1.2 Restatement of the Problem

Sunshine Company expects to sell new products including microwave ovens, baby pacifiers and hair dryers, in the online market. In order to help Sunshine understand the product market, we need to analyze the reviews of existing competitors in the Amazon, to find suitable sales strategies, potentially important design features and possible time-based patterns. The issues we need to address in this article include:

- Use mathematical evidence to describe the changing patterns of these products qualitatively or quantitatively, and help evaluate the relationship between star ratings, reviews, and helpfulness rating.
- Identify data measures based on informative ratings and reviews to prepare for these three new products.
- Analyze the change patterns over time for each product, and predict the trend of their reputation.
- Find out common characteristics between potentially successful and failing products, and form combinations of star rating and review text.
- Identify the influence between star ratings and reviews, such as whether previous low-scoring star ratings will affect user reviews.
- Recognize the relevance of the sentiment tendency of the review text and the star rating.

1.3 View of Our Work

For the product review data, we use sentiment analysis to quantify the review text, and then analyze the correlation between sentiment scores, star ratings, and helpfulness ratings. We establish a model to evaluate the processed data, find the changing patterns and influence factors of product reputation, and improve our model to obtain the optimal strategy. Our methods and process are shown as follow:

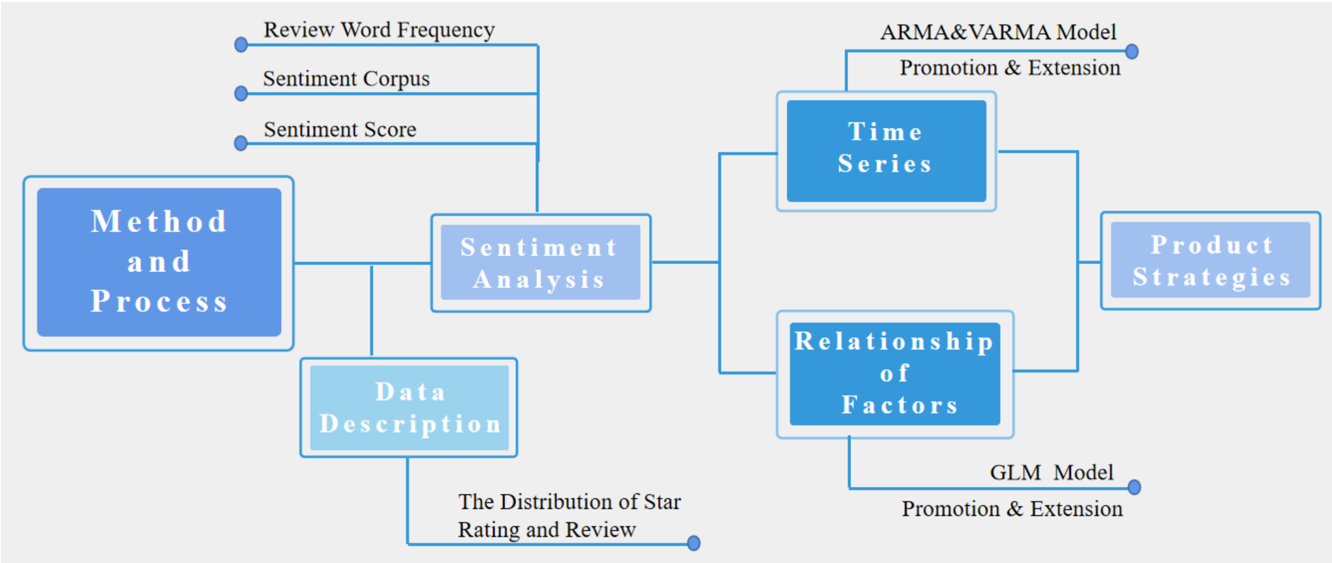


Figure 1. The mind map of our model

2 Data

2.1 Data Description

Checking all the data in the “hair_dryer.tsv”、“microwave.tsv” and“pacifier.tsv”, it is easy to figure out that the review data ranges respectively from 2002-2015、2004-2015、2003-2015, and that the numbers of their reviews, including several competition products, range from a few thousand to nearly 20000. Data cleaning and modification is not a necessity. To explore the pattern of all factors changing over time, we define the first month of review data as *Time 1*, and there are more than 130 months.

After analyzing the data, the distribution characteristics of the three products are as follows:

- In terms of star ratings, the number of five-star reviews is the most for all three products. There are a small number of reviews in the early time, and the overall fluctuation grows around 2011. This can be explained by the fact that online shopping in the early times was not as popular, so there were few sales and few reviews. In recent years, the development of Internet has given people more access to online shopping, so sales and reviews increased rapidly later on. The average star rating of the three products also fluctuated over time. The early fluctuations were large, and the fluctuation became smaller after a period of time. The hair dryer gradually stabilized at 4.2 stars; the baby pacifier gradually stabilized at 4.4 stars; and the microwave oven still showed fluctuations, increasing from 3 stars to 4 stars from 2012~2015. And we will explore more information about these later. (Figure 2)

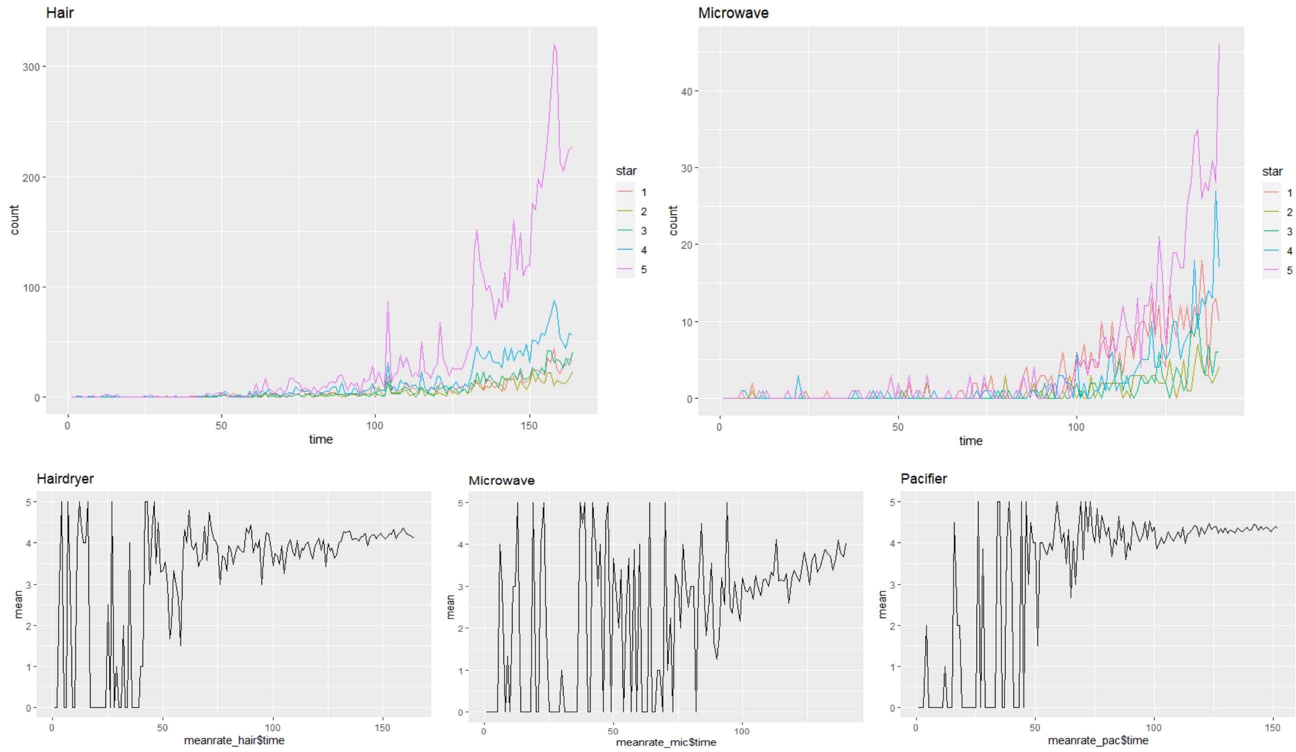
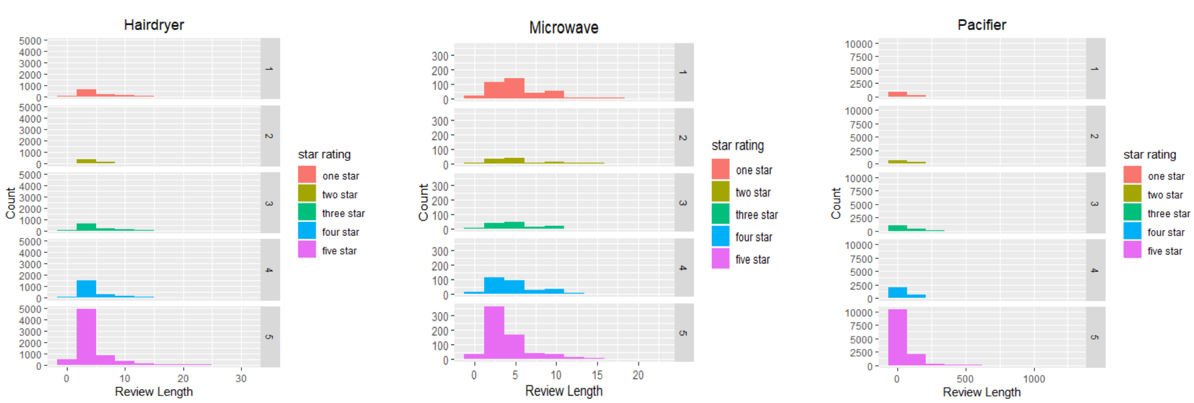


Figure 2. Star rating numbers (hairdryer and microwave oven for example) and the trend of mean star ratings

- In terms of review length, reviews of the microwave oven are mostly 1~6 words; that of the hair dryer include mostly 1~5 words; and that of the pacifier are mostly 1~100 words. The change of the average comment length with time also shows certain characteristics. The average comment length of the three products fluctuates with time. The average review length of the hair dryer increased gradually as a whole, ranging from 3 to 5 words. That of the microwave oven also increased gradually together, with the largest being 12 words in the early months, and later stable range is from 4 to 6 words. For baby pacifiers, with reviews over 300 words in early certain months, the fluctuation range in the later period becomes smaller and less than 100 words, and the overall trend is decreasing. (Figure 3)



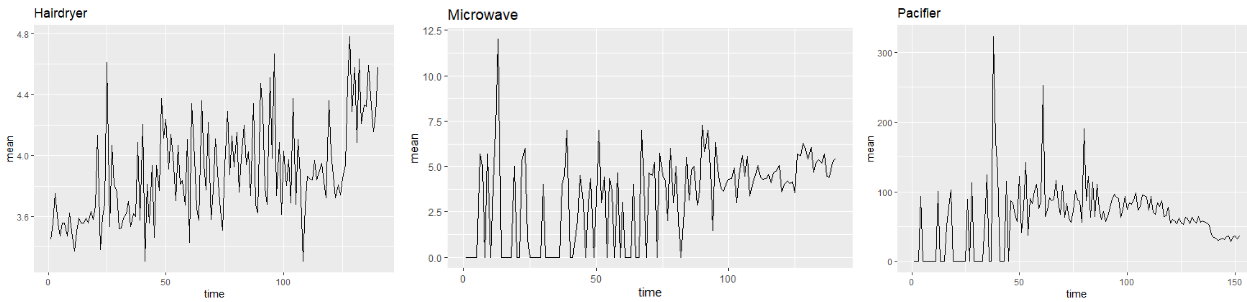


Figure 3. Review length distribution and time pattern

- In terms of helpful ratings, the average helpful votes of 1-2 star reviews for hair dryer are the most; for microwave oven, the most helpful votes in average are 3 and 5 stars; and for baby pacifiers, the average helpful votes of 1-2 star ratings are the most. But for the total number of helpful votes, the most is 5 stars for all the three products. It is shown that although the 5-star rating has a huge advantage in terms of quantity, the low-score reviews are still the most helpful ones for customers. We assume that this may be related to the low-score reviews text containing reasons of low-score and more information concerning products' quality. (Figure 4)

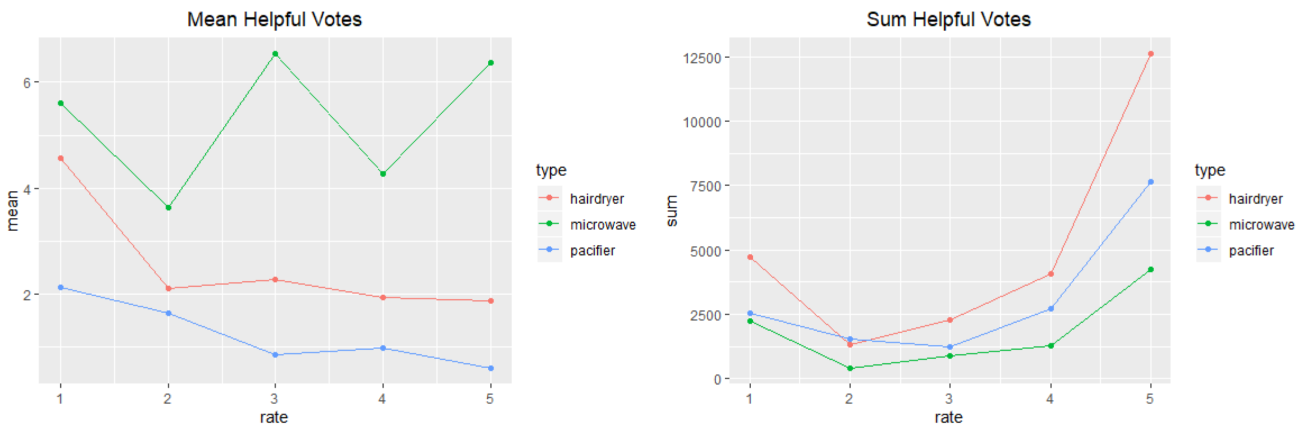


Figure 4. Mean helpful votes and sum helpful votes for each star ratings

2.2 Sentiment Analysis

Reviews from customers can usually be classified into two types, with one type mainly describing the functions and user experience and the other type expressing personal preferences with strong emotional attitudes. Therefore, certain patterns, or more precisely, quality descriptors, are commonly found in the text-based reviews.

While the emotional expressions indicating preference, including verbs like "love" and "hate" and adjectives like "amazing" and "terrible", for various products might be highly similar, the sentence units related to product performance generally differ from one to another as descriptors of certain product characteristics. Below listed are the 10 most frequent words or phrases that appear in the 5-star (Table 1) and 1-star (Table 2) review bodies indicating personal preference and product functions, which will be denoted as EU (emotion unit) and FU (function unit). The frequency each word or unit appears is given in the parenthesis.

Microwave Oven		Baby Pacifier		Hair Dryer	
EU	FU	EU	FU	EU	FU
great (272)	works (173)	great (3200)	little (1937)	great (2340)	dry (1232)
well (152)	space (147)	love (3104)	easy (1901)	love (1920)	blow (1171)
love (130)	up (145)	like (1959)	holds (801)	like (1280)	works (1160)
good (123)	easy (142)	well (1493)	clean (711)	good (1142)	heat (818)
like (114)	small (141)	cute (1404)	soft (698)	well (973)	years (755)
nice (74)	fits me (76)	good (1249)	quality (656)	recommend (672)	powerful (606)
recommend (59)	enough (55)	perfect (1056)	size (585)	best (607)	easy (638)
perfectly (55)	large (55)	recommend (900)	works (558)	nice (593)	light (592)
happy (52)	power (48)	best (806)	sleep (489)	perfect (447)	fast (555)
right (46)	clean (46)	nice (726)	fit (457)	happy (396)	small (405)

Table 1. Most Frequent Sentence Units in 5 Star Review Bodies

Microwave Oven		Baby Pacifier		Hair Dryer	
EU	FU	EU	FU	EU	FU
not buy (167)	out of service (200)	not like (266)	hard (91)	not like (202)	in months (266)
problem (97)	no warranty (140)	took back (163)	plastic (82)	money (162)	hot (194)
not like (93)	in months (146)	waste money (136)	return (78)	waste (94)	return (145)
cost (70)	repair (117)	not work (114)	little use (75)	low (73)	buy another (117)
even less (65)	replaced (74)	disappointed (104)	big (63)	disappointed (79)	no warranty (94)
not purchase (51)	sharp (69)	not good (98)	not fit (60)	bad (66)	fire (81)
died (48)	heat (60)	never used (85)	broke (63)	died (65)	broke (68)
not good (48)	stopped working (58)	waste (82)	cheap (63)	cheap (61)	sparks (60)
not recommend (35)	fire (38)	not recommend (64)	not safe (34)	problems (61)	burned (56)
error (30)	broke (37)	bad (34)	useless (34)	junk (41)	loud (55)

Table 2. Most Frequent Sentence Units in 1 Star Review Bodies

It can be seen that the most frequent words in the reviews of the 3 products have a lot in common. Customers who wrote 5-star review were happy and satisfied with the product, considering it as just the right fit and would recommend it to others. Those who wrote 1-star reviews, on the contrary, were disappointed and viewed it as a waste of money and a product full of problems. It's also worth noticing that a common and critical property for satisfying products is whether it is long-lasting and safe. Customers will also tend to pay attention to whether a product has warranty when purchasing electrical appliances in case of a breakdown. However, when it comes to function descriptors, customers' focus for various product start to differ, which may give us a hint about the important design features that would enhance product desirability.

- As for microwave oven, customers are looking for products that generally doesn't take too much "space", but has "large" or "enough" room inside. It should also be "clean" and "heats" up well with great "power" without being caught in "fire", and easy to use as well. Here we have identified key features for a microwave oven: size and power.

- A good baby pacifier has to fulfill multiple requirements for the safety of babies. It should be "small" to fit babies' mouth size, "clean" and "soft" to make sure it will not hurt babies even when they are in "sleep". Other than size, material and quality are other essential features. A good pacifier should be made with materials like soft "rubber" instead of "hard" "plastic", and should not break easily, which ought to be absolutely safe with babies.
- As can be seen from the reviews, highly-rated hair dryers should be "powerful" with strong "blows" and "heat", and "dries" hair "fast". Bonus features for it would be "small", "light" and portable. Undesirable features of a hair dryer include "hotness" and "loudness" during use. In extreme cases, "sparks", "burns" and "smoke" may even occur in use, indicating potential danger and quality problems. Generally speaking, power and safety should be the main concerns for hair driers, and those with smaller size are more desired.

To better describe the reviews quantitatively, we use sentiment analysis to transform the review text to score. First, we define all the reviews as $V_r = \{V_{r1}, V_{r2}, \cdots, V_{rn}\}$. Then we perform word frequency analysis on the review title and review body of the three products, and define the high frequency words as $W_r = \{W_{r1}, W_{r2}, \cdots, W_{rn}\}$. Keywords of the three products can be found out respectively for determining important design features, and are marked with a sentiment score, ranging from -5-5 (-5 for the lowest negative sentiment, 5 for the most positive sentiment). We choose these words and scores to define a sentiment corpus: the words are $S_r = \{S_{r1}, S_{r2}, \cdots, S_{rn}\}$, and the corresponding scores are $s_r = \{s_{r1}, s_{r2}, \cdots, s_{rn}\}$. Part of the sentiment corpus is displayed below. (Table 3)

word	grade	word	grade
bitch, bastard	-5	outstanding, breathtaking	5
shit, fraud	-4	win, amazing	4
bad, fake	-3	good, beautiful	3
annoyed, unprofessional	-2	like, stable	2
limited, waste	-1	huge, accept	1
word	grade	word	grade
absolute, extreme	1.64	just, merely	0.89
above, too	1.33	better, really	1.25
more, relatively	1.22	a bit, quite	1.15

Table 3. Demonstration of the Sentiment Corpus

Find sentiment words in these review headlines and review bodies, sum up the scores by our sentiment corpus, and then we get the scores of the reviews. Analyze the sentiment scores and star ratings of the reviews, and a positive correlation can be found. The higher the sentiment score is, the higher the star rating will be. According to the results of sentiment analysis, we can conclude the following:

- For the hair dryer and the baby pacifier, reviews of 1~2 star mostly have scores of -5 to 0, existing outliers well below the lower edge; reviews of 3 star get scores between 0 and 5, existing both outliers well below the lower edge and higher ones; and reviews of 4~5 stars get scores between 0 and 10, with outliers far above the upper edge.(Figure 5)

- For the microwave oven, reviews of 1~3 star mostly get -5~0 score, existing outliers well below the lower edge; and reviews of 4~5 stars get score between 0 and 10, existing outliers far above the upper edge. (Figure 5)

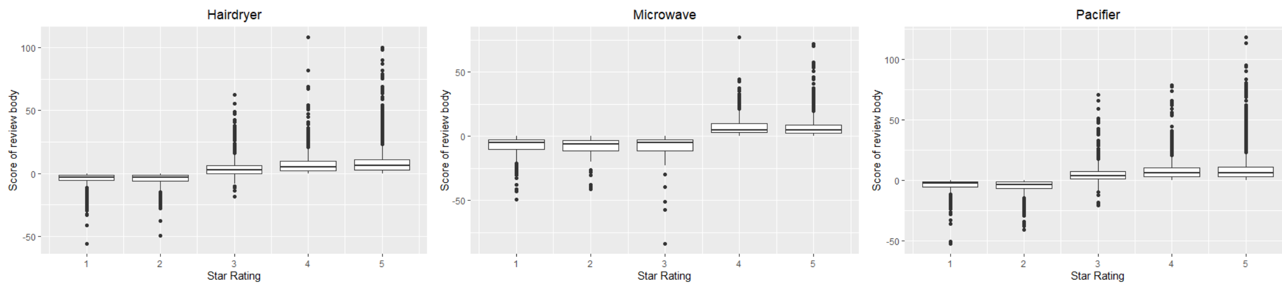


Figure 5. Distribution of the score of review body for each star rating

- In terms of the correlation of these four factors, including helpful ratings, score of review headline, score of review body and review length, we can draw some interesting insights. As follows, *Var i* (*i=1-4*) represent the distribution of each star rating for these four factors; the top right corner shows the correlation coefficient between each two factors; and the left bottom are scatter diagrams. For these three products, we can find the score of review headline and the score of review body shows a strong correlation with correlation coefficient between 0.43 and 0.62. And for the baby pacifier, the correlation coefficient between the score of review body and review length is high as well. It is shown that there is some special connection among review length and the sentiment of review headline and review body. We shall explain the connection and its influence in models afterwards. (Figure 6)

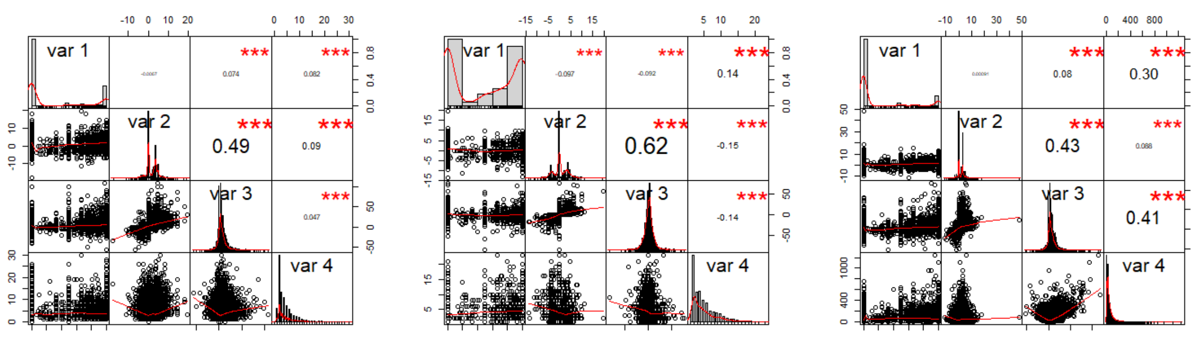


Figure 6. Correlation among helpful ratings, review headline, review body and review length for the hair dryer, the microwave oven and the baby pacifier (from left to right)

3 Assumptions & Nomenclature

- The three types of product are sold via Amazon independently, which means that their sales do not affect each other. But there can be some similar patterns of sales and reviews between them.

- There may be some customers who are not willing to grade and evaluate the product, but the whole sales are proportional to the number of reviews.
- The reviews and star ratings in the tables are based on real experience, instead of paid Internet trolls.

Symbol	Definition
R	Star rate of the product
S_h	Emotional score of review headline
S_b	Emotional score of review body
L_r	Length of reviews
H_v	Helpful votes of reviews
H_r	Helpful ratings of reviws
t_i	The i_{th} month of time series
v_1	Vine
V_2	Verification
$\overline{R}$	Mean rate of one month
$\overline{L}$	Mean length of one month
$\overline{S}$	Mean sales of one month

Table 4. Nomenclature

4 Our Model of GLM

4.1 Introduction to the Idea of Method

To estimate and predict the future development of three product, we should determine a smaller subset of predictors first. LASSO works well for linear regression. The basic form of LASSO is

$$\hat{\beta}^{lasso} = argmin \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

where parameter λ can be adjusted to screen variables.

4.2 Modeling Process & Results

To figure out the relationship among review factors, we use GLM model to analyze data. Take the microwave oven for example. Calculate the var and other index of review length, star rating, vine, sentiment score of review headline and body. According to these indexes, we choose $df = 2$, follow the order of verify, length, star, vine, body and head. According to the rule that cp should be smaller, we choose 2 or 3 parameters.

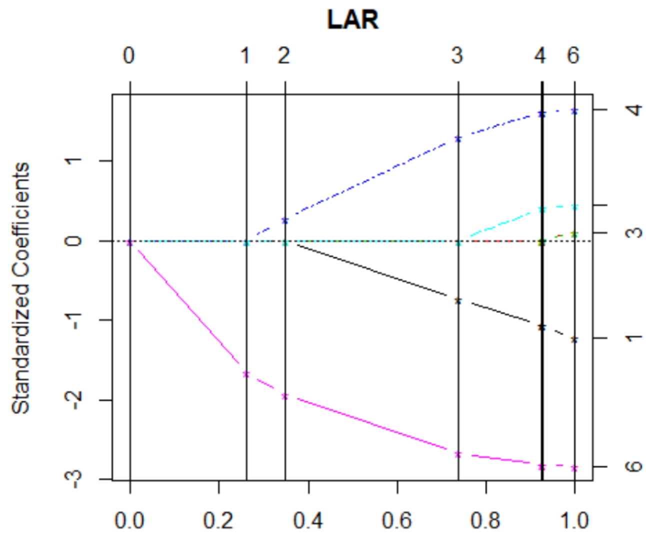


Figure 7. Standardized Coefficients for the microwave oven

The above figure shows that whether there is a discount has the most significant impact on the informative evaluation, followed by sentence length. Besides, the discount and star rating have a negative correlation with informative evaluation.

We can get the regression function:

$$H_r = 0.554 - 0.184V_2 + 0.012L_r$$

According to the regression results, it can be seen that discount has a negative impact on informative evaluation, and the comment length is positive. However, the coefficient of the comment length is much smaller than the discount, which indicates that the fluctuation of the comment length does not greatly change the result.

Considering that there are many 0-1value of helpful rating, we selected samples with helpful rating = 0 or 1 as the dependent variable, and then tried to use the generalized linear model of the original variable for logistic regression.

$$g(\pi) = logit(\pi) = \alpha + \beta x$$

We get

Microwave : $g(\pi_1) = -0.045 - 0.829V_2 + 0.053L_r$

Hairdryer : $g(\pi_2) = 0.140 - 0.563V_2 + 0.022L_r - 0.291R + 0.038S_b$

Pacifier : $g(\pi_3) = 1.281 - 1.070V_2 + 0.005L_r - 0.265R + 0.0008V_2L_r$

All factors show good correlation, and the coefficient of verification is negative, which is consistent with the results of the previous linear regression. And the ROC curve also fits well. (Figure 8)

While for the hairdryer, informative evaluation is mainly affected by discounts, stars and reviews, where discounts and stars are still not negatively related to results. The regression function of the hairdryer is as follows. According to regression, the first four variables are more significant for the model.

$$H_r = 0.642 - 0.175V_2 - 0.069R + 0.005S_b + 0.005L_r + 0.05V_2S_b$$

Unlike the first two products, for pacifiers, the length of the review is the biggest factor affecting the validity of the information. But star ratings and discounts are still negatively related to it. And the regression function of the baby pacifier is as follows. We can figure out that the validity of pacifier reviews is mainly determined by discounts and ratings.

$$H_r = 0.331 - 0.106V_2 - 0.035R + 0.005S_h + 0.001L_r + 0.0004V_2L_r$$

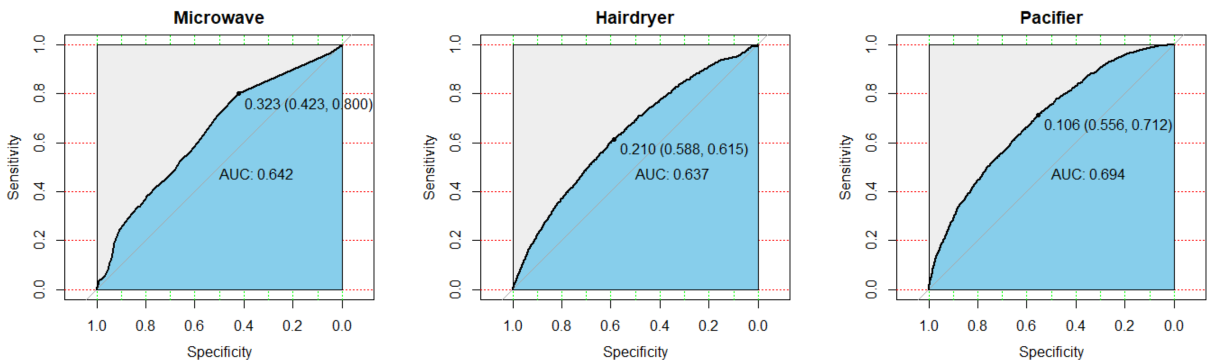


Figure 8. ROC curve for the microwave oven, the hairdryer and the baby pacifier

5 Our Model of Time Series

5.1 ARMA Model

5.1.1 Introduction of Our Method

Based on the analysis above, the factors like star rating and review shows patterns of fluctuating over time. Thus we try to fit an **autoregressive moving average (ARMA) model** to analyze the change of these factors over time.

For an $\boldsymbol{ARMA(p, q)}$ model, p is the order of the autoregressive polynomial, and q is the order of the moving average polynomial.

Then we have

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where φ is the autoregressive model's parameters, θ is the moving average model's parameters, c is a constant and ε is the error term.

5.1.2 Process and Results

For these three products, we want to predict their reputation and sales in the near future, so we use the existed reviews to analyze. Take the baby pacifier for example, according to the time line chart of star ratings, we calculate ACF and PACF respectively for full time period and the period after 50^{th} month with more stable data. Autocorrelation function and partial autocorrelation function are both tailing, thus we apply ARMA model to the period after 50^{th} month. (Figure 7)

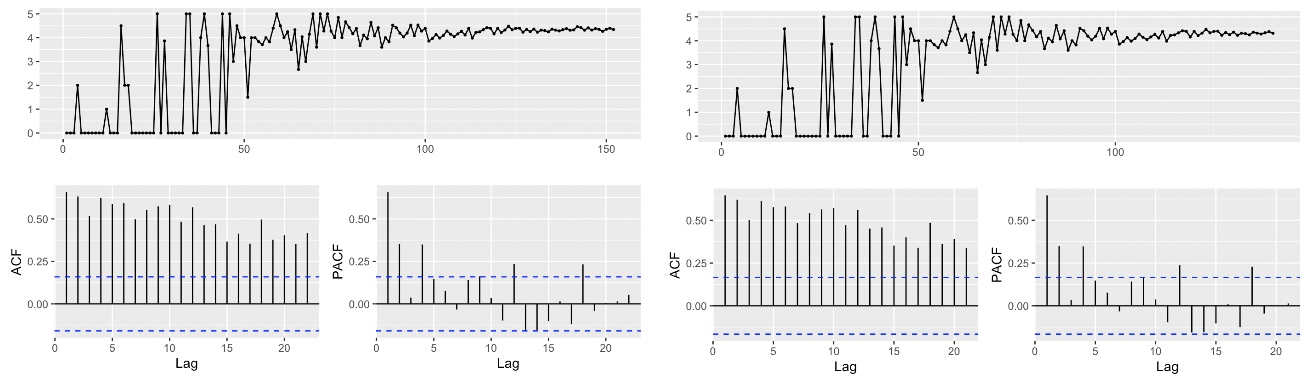


Figure 7. The ACF and PACF of the baby pacifier for full time period and the period after 50th month

Then we choose suitable parameters to find best model for the baby pacifier:

Star rating :

$$X_t^{(1)} = -0.090X_{t-1}^{(1)} + 0.388X_{t-2}^{(1)} - 0.936\varepsilon^{(1)}$$

Sales :

$$X_t^{(2)} = -0.014X_{t-1}^{(2)} + 0.217X_{t-2}^{(2)} - 0.958\varepsilon^{(2)}$$

In this model, the ACF and PACF are more stable and tending to 0. To test our prediction of the model above, we compare our prediction with actual trend of the last 30 months. It is showed that fitted and actual ratings are similar, tending to 4.4 star.

As for the sales of the baby pacifier, it is known that there are products reviews only when people buy them. Thus, we assume that the count of reviews can represent the count of sales, which simply states that the count of sales is at least the count of reviews so we are predicting the minimum sales. And we analyze ACF and PACF of the count of pacifier reviews. Autocorrelation function is tailing, and partial Autocorrelation function shows trend of truncation. We then use the full time period to predict future sales. The same as the prediction of star ratings, fitted and actual sales are similar, tending to 660.

For the microwave oven and the hair dryer, the prediction of star ratings and sales are also suitable for this ARMA model. (Figure 8)

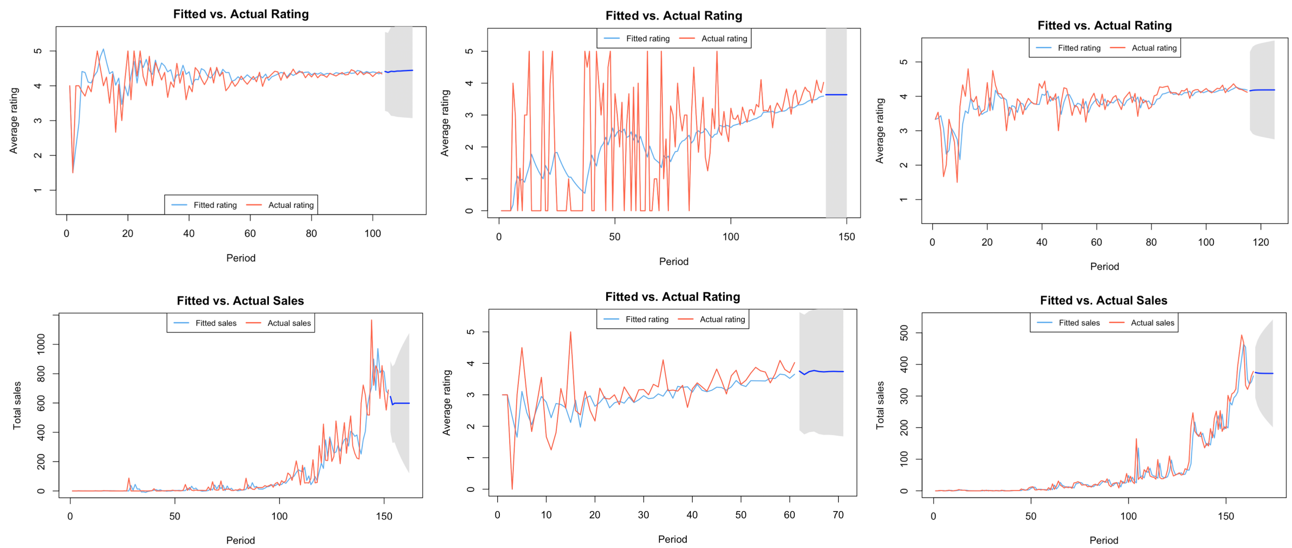


Figure 8. The prediction of star ratings and sales for the baby pacifier, the microwave oven and the hairdryer

5.2 VARMA Model

5.2.1 Introduction of Our Method

Just as what has done to a single factor, we extend the method to a vector which include the factor of text and star rating: mean length of reviews and mean star rate in a month.

Let $z_t = [z_{1t}, \dots, z_{nt}]$ Where n is the dimension of the vector, and t is the time. Assume that z_t is a continuous random vector. We define

$$\Gamma_l = Cov(z_t, z_{t-l})$$

which is the lag l cross-covariance matrix of z_t and z_{t-l}. Then the lag l cross-correlation matrix(CCM) ρ_l is

$$\rho_l = D^{-1}\Gamma_l D^{-1}$$

Set $\hat{a}_t = z_t - \hat{z}_t$ be the forecast error, just as the form of ARMA model, we have

$$z_t = \phi_0 + \sum_{i=1}^{p} \phi_i \, z_{t-1} + a_t - \sum_{i=1}^{q} \theta_i \, a_{t-i}$$

where $p > 0, q > 0$; ϕ_i and θ_i are $k \times k$ constant matrices; $\{a_t\}$ are independently distributed. This is the basic form of **vector autoregressive moving-average (VARMA)** model, which is the extension of **autoregressive moving average (ARMA) model**.

5.2.2 Stationarity

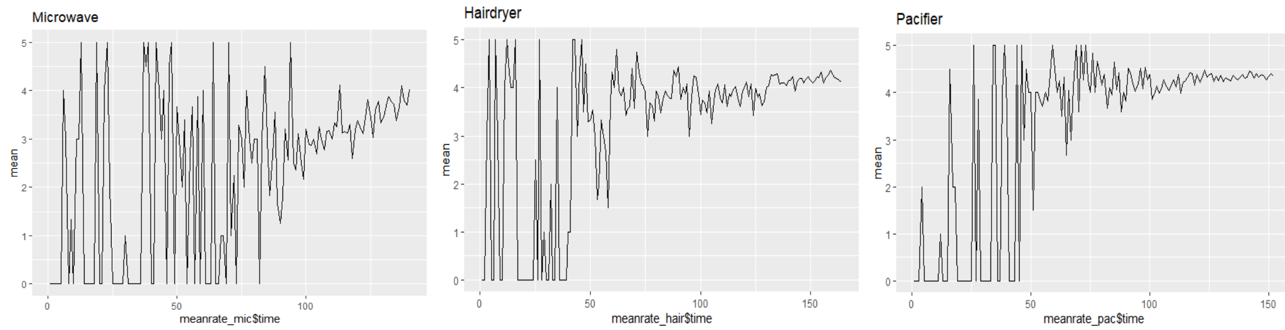


Figure 9. Change of mean rate of reviews over time

By observing how our factors change over time, we can find three product have sharp fluctuations in the early months. For enhancing stationarity of our time series and reduce the impact of seasons to regress better, we make a difference to them at first.

5.2.3 CCM

By combining the text-based factor $\overline{L}$ and rate-based factor $\overline{R}$, we obtain our variable of time series. To study the linear dynamic dependence, we examine $\hat{\rho}_l$ as follows:

$$\hat{\mu} = \frac{1}{T}\sum_{i=1}^{T} z_t \, , \hat{\Gamma}_0 = \frac{1}{T-1}\sum_{t=1}^{T}(z_t - \hat{\rho}_z)(z_t - \hat{\rho}_z)'.$$

$$\hat{\Gamma}_l = \frac{1}{T-1}\sum_{t=l+1}^{T}(z_t - \hat{\rho}_z)(z_{t-l} - \hat{\rho}_z)'.$$

Then the lag l sample CCM is

$$\hat{\rho}_l = \hat{D}^{-1}\hat{\Gamma}_l \hat{D}^{-1}$$

Calculate CCM for hairdryer as an example. For a year, the CCM of first 10 months is all greater than 0.5, which means the Sequences correlate with data from the first 10 months, while the CCMs after 10 months show decreasing trend.

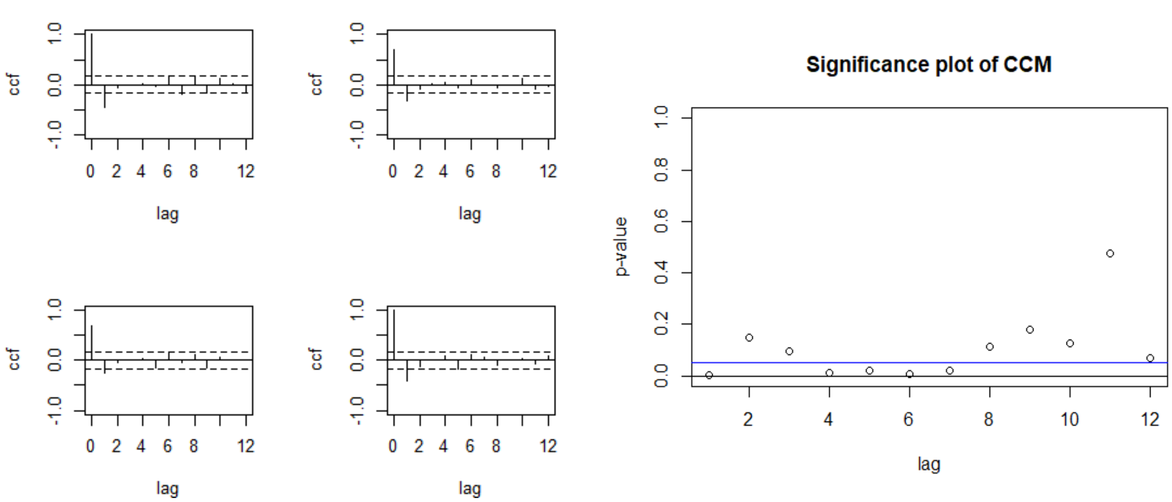


Figure 9. The CCM and p-value of the hairdryer

As is seen above, ccf has truncation properties, and the p-value of CCM is mostly larger than 0.05. Considering that there is a high possibility of zero when lag = 1, we choose Var(2) or Var(3) model.

5.2.4 Choose order

To choose the parameter of model, we calculate AIC, BIC and HQ of the regression and select the suitable order based on them.

p	AIC	BIC	HQ
0	0.344	0.344	0.007
1	-0.219	-0.143	-0.188
2	-0.319	-0.168	-0.258
3	-0.565	-0.337	-0.472
4	-0.588	-0.285	-0.465
5	-0.546	-0.166	-0.391
6	-0.559	-0.103	-0.374
7	-0.614	-0.082	-0.398
8	-0.625	-0.018	-0.387

Table 5. The information criterion of VARMA model for hair dryer

For a regression model, the criterion AIC and BIC should be as small as possible. As shown in the table, for hairdryer product, $p = \{2,3\}$ can be suitable.

Similarly, we choose $p = \{1,2,3\}$ for microwave and $p = \{2,3,4\}$ for pacifier.

For another parameter q, we use the extended cross-correlation matrices to choose.

$$\omega_{1,t}^{(j)} = a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$

0	1	2	3	4	5	6
0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0003
0.0000	0.0000	0.0008	0.0043	0.0040	0.0022	0.0001
0.0000	0.0042	0.0125	0.0065	0.0004	0.0004	0.0000
0.0154	0.0208	0.0619	0.0092	0.0002	0.0000	0.1151
0.1399	0.0281	0.0301	0.0116	0.0000	0.6017	0.4990
0.1188	0.0954	0.0251	0.0323	0.1618	0.4842	0.7713

Table 6. The p-value table for extended cross-correlation matrices of hairdryer.

As is shown in the table, for hairdryer, $p = 3$ and $q = 1$ can be a good choice. Similarly, we select $VARMA(1,1)$ for microwave and $VARMA(2,1)$ for pacifier.

5.2.5 Estimation

As we know, p is the order of z and q is the order of a. We substitute the data into the respective model and get the estimation as the following:

Microwave:

$$z_t = \begin{bmatrix} 0.017 \\ 0.030 \end{bmatrix} + \begin{bmatrix} 0.120 & -0.006 \\ 0.127 & -0.152 \end{bmatrix} z_{t-1} + \hat{a}_t + \begin{bmatrix} 1.120 & -0.080 \\ 0.051 & 0.971 \end{bmatrix} a_{t-1}$$

Hairdryer:

$$z_t = \begin{bmatrix} 0.013 \\ 0.010 \end{bmatrix} + \begin{bmatrix} -0.489 & -0.232 \\ 0.330 & -0.585 \end{bmatrix} z_{t-1} + \begin{bmatrix} -0.400 & -0.122 \\ 0.293 & -0.414 \end{bmatrix} z_{t-1} + \hat{a}_t + \begin{bmatrix} -0.198 & -0.007 \\ 0.266 & -0.355 \end{bmatrix} a_{t-1}$$

Pacifier:

$$z_t = \begin{bmatrix} 0.047 \\ 1.802 \end{bmatrix} + \begin{bmatrix} -0.114 & -0.003 \\ -9.362 & -0.090 \end{bmatrix} z_{t-1} + \begin{bmatrix} -0.074 & -0.002 \\ -8.491 & 0.078 \end{bmatrix} z_{t-2}$$
$$+ \hat{a}_t + \begin{bmatrix} 0.734 & -0.0004 \\ -15.494 & 1.026 \end{bmatrix} a_{t-1}$$

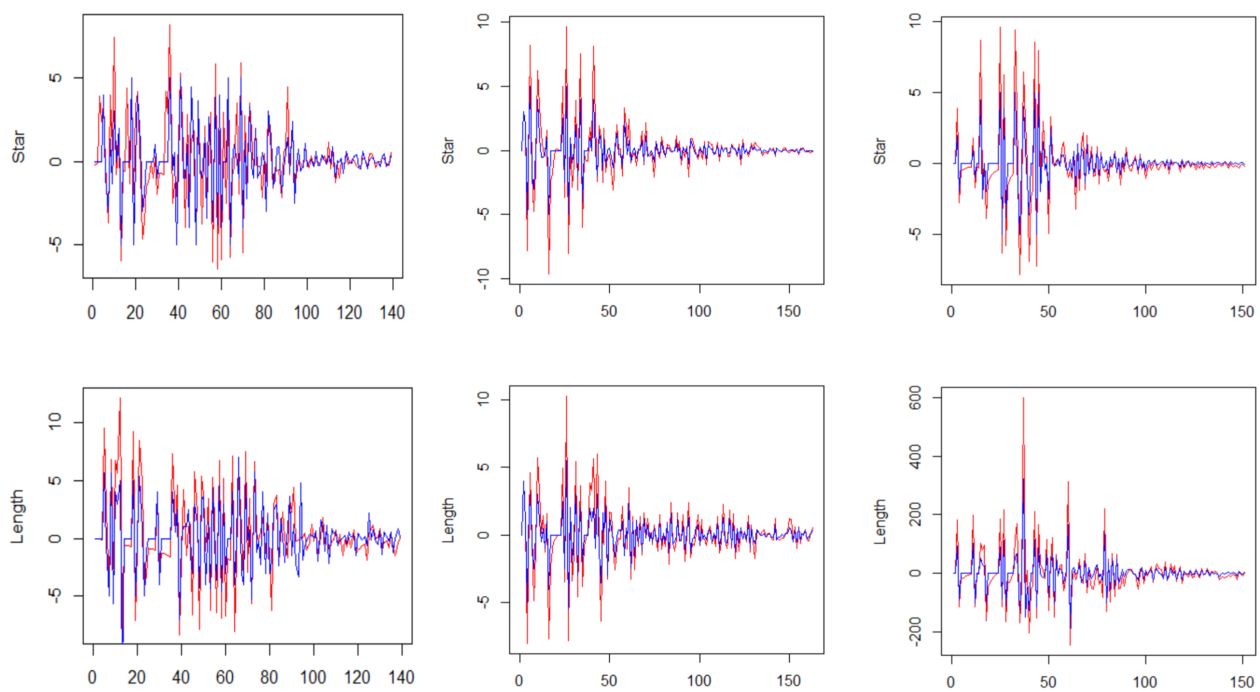


Figure 9. The estimation of models v.s. the original data.

The figures above shows the fitting result of our model. As we can see, the VARMA model fits well.

5.2.6 Prediction

The final purpose of our time series model is to predict the development of the product. Thus, we combine every 50 months as a group to predict the situation of next month. The simulation results are as follows.

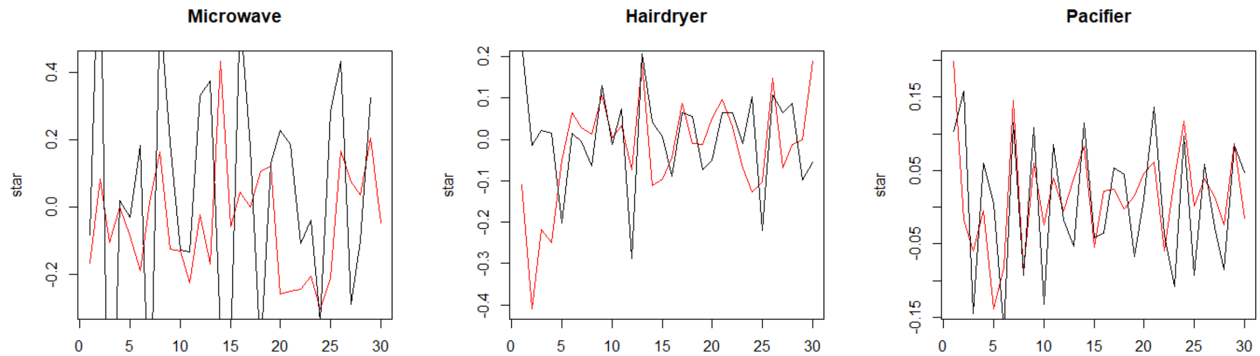


Figure 9. The prediction of star ratings for three product v.s. the original data.

We can see from the picture that the whole trend of our predictions is consistent with the original data, which means that our models perform well in prediction.

For microwaves, the star ratings go up and down around 0+. Since our data has been differentiated, this figure shows that star ratings will increase with a decreasing speed and then tends to be stable.

For hairdryer, the volatility of our prediction consists better with the real data than microwave, further verifying our model. It may be partly due to the influence of time for $VARMA(3,1)$ predicts based on the previous 3 months.

For pacifier, the model also predicts very well. It shows more trend of increase, meaning that the ratings may grow in a small speed.

5.2.7 Sensitivity analysis

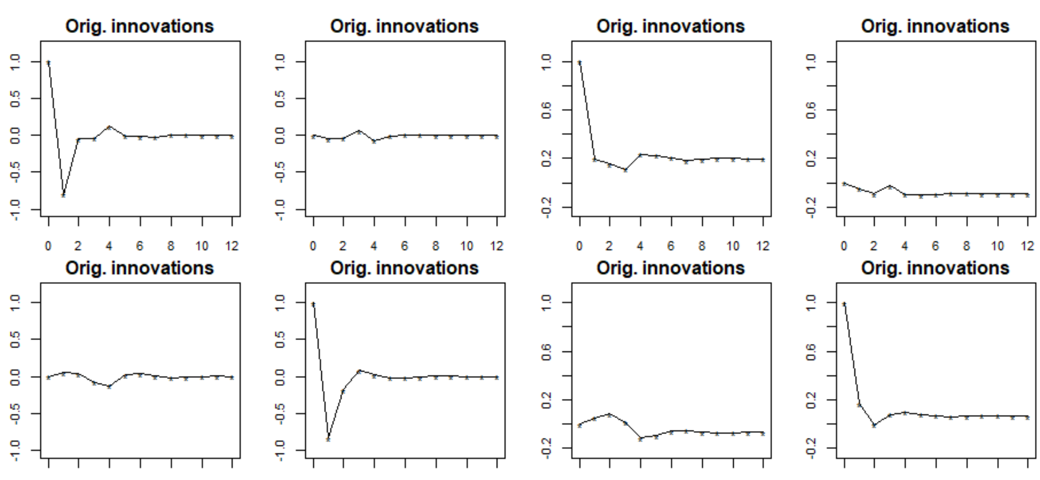


Figure 9. The impulse response functions and accumulated impulse response functions of VARMA(1,1) for microwave

The slight wave of the plots represents delayed effects of one factor to another. Given our two measures, it means that the previous star ratings and reviews can still have influence to the current product to some extent. Like the upper-right plot, customers may tend to write shorter reviews after seeing many high levels of star ratings. But the whole development tends to be smooth after several small wave for the impulse response functions decay to 0 quickly.

6 Strengths and Weaknesses

6.1 Strength

- Effectively quantify reviews through sentiment analysis and play an important role in later regression.
- Use LASSO to select model variables and screen out the most influential factors.
- Linear regression model has strong interpretability.
- Dimension expansion based on the general time series model, making more use of effective factors.
- The prediction results of the VARMA model are good and have guiding significance for future sales.

6.2 Weakness

- Factors such as helpful rating in the table are excessively discrete. The fitting effect and prediction effect using the linear regression model are poor, and the AUC value is only about 0.7.
- The matrix calculation of time series is more complicated. When there are many parameters, the accuracy of the computer will cause the error of matrix inversion, which will affect the accuracy of the final result.
- In this paper, three products are modeled separately, and the results have certain differences. The applicability of each model is narrow and not universal.
- The front part of the data set is relatively sparse, and the latter tends to be dense and stable. The model may have a large error in the fitted value at the beginning of the sale.
- Modeling the same type of products together without considering the brand effects of the products and the effects of competition, which need to be adjusted appropriately when applied to the actual situation.
- The default reviews are all true and effective, without considering the effects of possible false reviews and some casual attitude reviews, which may lead to a gap between the results and the actual situation of the product;

7 Conclusions

First, by analyzing high-frequency vocabulary and using sentiment analysis, we determine the extent to which customer like the product for each evaluation. Reviews with strong emotional attitudes are assigned higher scores in absolute values, and naturally have high or low ratings as well. After splitting high-frequency words into emotion units and function units, we can obtain the desired features for different products from the function units.

Then, we use LASSO to remove some unimportant factors and get the main factors that affect the helpful rating of each product review. The key factors of microwave oven reviews are: discount, review length and star rating, the main factors of hair dryer are discount, star rating and emotional score, and the main factors of pacifier are review length and star rating.

Then we performed a linear regression on it and found that discounts and star ratings have a negative impact on helpful ratings, that is, the higher the discount or star rating, the lower the quality of the review.

Considering the excessive discreteness of the helpful rating and the poor linearity of the ordinary linear regression, we take samples equal to 0 and 1, which are also the two types of samples with the largest proportion for logistic regression. Regression models have good predictions of results.

Next, we use time series models to analyze the time effect of product sales. Through half of the single-factor ARMA model and the multi-factor VARMA model, we obtained the prediction function for the future development of the product. The fitting and prediction results of the ARMA model are not ideal, which may be due to the excessive volatility in the initial period of the data set. The multi-factor model further revised the prediction results to obtain more consistent predicted values. The predicted results of the three types of products all tend to reach 0+, indicating that the overall sales of the products have gradually stabilized, and there is a constant trend. Finally, we perform sensitivity analysis. The impulse response functions will all be 0 quickly, indicating that the product has a certain resistance to the impact of random factors, and the sequence gradually stabilizes.

Our model only refers to the three data given by the title, which can better explain, fit and predict the existing data, but its practicality needs to be further explored.

Memo to the Marketing Director of Sunshine Company

Data: March 9, 2019

To: Marketing Director, Sunshine Company

From: MCM Team #2012671

Subject: Online Marketplace Review Analysis and Product Strategy

Dear Administrator:

By studying the sales and reviews of the three products on the Amazon website, we have initially obtained the basic characteristics and rules of the past sales of each product, and use this to predict future development.

For the textual data of a large number of reviews on the website, we quantify it with the help of sentiment analysis to get the sentiment score. And it has a strong correlation with star ratings, which is in line with expectations. We also calculated the length of comments as another characteristic of comments in addition to emotion. In the preliminary analysis, we found that although the 5-star reviews of the 3 products accounted for the majority, the effective votes focused on low-star reviews. This shows that customers have a common mentality when shopping, and they are more inclined to see and trust the negative reviews of products. Therefore, in the future sales, the company needs to pay special attention to the negative evaluation of the product, and respond in time and remedial measures.

After carefully examining the high-frequency words in the reviews, we have extracted key features that make a product desirable from the function descriptors, which you should carefully design to promote sales. In general, customers expect good microwave ovens not to take too much space but have enough space for food. It should heat up nicely with great power, easy to use, but will never catch on fire. For baby pacifiers, it's better to pay attention to the size and material of pacifiers related to safety issues. It should be soft, clean and small enough to fit babies' mouths, and should not break easily. Hair dryers are expected to be powerful with heat and strong blows to dry hair quickly, and never be too hot or loud. Customers will find a hair dryer more desirable when it is small, light-weighted and portable. One thing to keep in mind is that customers expect every product they purchase to be safe and long-lasting, and electrical appliance should be warranty guaranteed.

Observing the time series of products, we find that all types of products have slower growth and greater volatility in the early stages of sales, and have gradually stabilized after entering the rising period. Therefore, there is no rush in the sales process, and occasional small increases and decreases are common. Products need to accumulate word of mouth slowly. After reaching a certain sales volume and good evaluation, the product sales can maintain a high level. At this time, the random factors have a low impact on them.

In order to analyze the correlation between various factors, in order to select the most refined model, we use LASSO to screen the factors. We found that the most important influencing factors of microwave ovens and hair dryers are discounts, which shows that price, or value for money is still the main reference factor for citizens. Therefore, in the initial pricing and subsequent promotional activities, the company can properly choose between profit and sales volume. Lowering the sales price may bring more profit and better reputation.

Finally, we observe the changes in comments and scoring along the time axis, and use the VARMA model to fit and predict. The model has a good fit and prediction effect. On the one hand, it indicates that the future sales of the product are greatly affected by short-term evaluation. On the other hand, it also indicates the diversity of factors affected by sales. It should be fully considered when formulating the final strategy. fully prepare.

The model parameters finally determined for the three products are small, and the p of the microwave oven is only 1, indicating that the sales situation within one month can predict the future development to a certain extent. In the sensitivity analysis of the model, we found that there is a lag effect between the two factors, further illustrating that the user's psychology is extremely vulnerable to the short-term evaluation of the product. Therefore, the company's reviews and after-sales treatment of products need to pay more attention and be updated in a timely manner in order to maintain product reputation and expand sales.

References

[1] Tsay, Ruey S. Multivariate Time Series Analysis: With R and Financial Applications. 2014. Wiley Ser. in Probability and Statistics. Web.

[2] Muhammad Asif,Atiab Ishtiaq,Haseeb Ahmad,Hanan Aljuaid,Jalal Shah. Sentiment analysis of extremism in social media from textual information[J]. Telematics and Informatics,2020,48.

[3] Goldberg A B . Seeing stars when there aren't many stars : graph-based semi-supervised learning for sentiment categorization[C]// Workshop on Graph Based Methods for Natural Language Processing. Association for Computational Linguistics, 2006.

[4] Medhat W , Hassan A , Korashy H . Sentiment analysis algorithms and applications: A survey[J]. Ain Shams Engineering Journal, 2014, 5(4):1093-1113.

[5] Yuan-Yuan H , Qiang Y E . Research on online impact factors of customer reviews usefulness based on movie reviews data[J]. Journal of Management Sciences in China, 2010.

[6] Chevalier J A , Mayzlin D . The Effect of Word of Mouth on Sales: Online Book Reviews[J]. Journal of marketing research, 2006, XLIII(3):p.345-354.

[7] Jin L , Hu B , He Y . The Recent versus The Out-Dated: An Experimental Examination of the Time-Variant Effects of Online Consumer Reviews[J]. Journal of Retailing, 2014, 90(4):552-566.

[8] Yanfeng Z , He L , Lihui P , et al. Empirical Study on Factors Affecting Timeliness of Online User Reviews Behavior[J]. Journal of Modern Information, 2019.

[9] Seung-joon Lee,Hyunjeong Kang. Effects of Characteristics of Online Customer Review on Review Helpfulness[J]. International Journal of Digital Content Technology and its Applications,2013,7(16).

Appendix

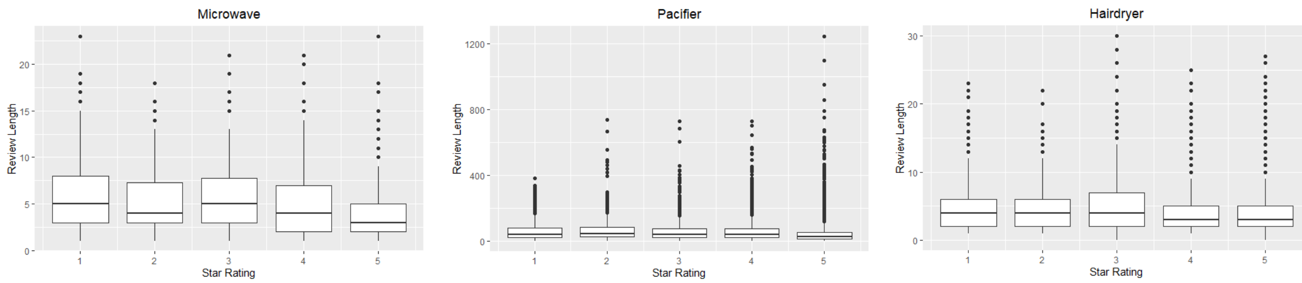


Figure 10. Boxplot of review length at 5 star ratings.

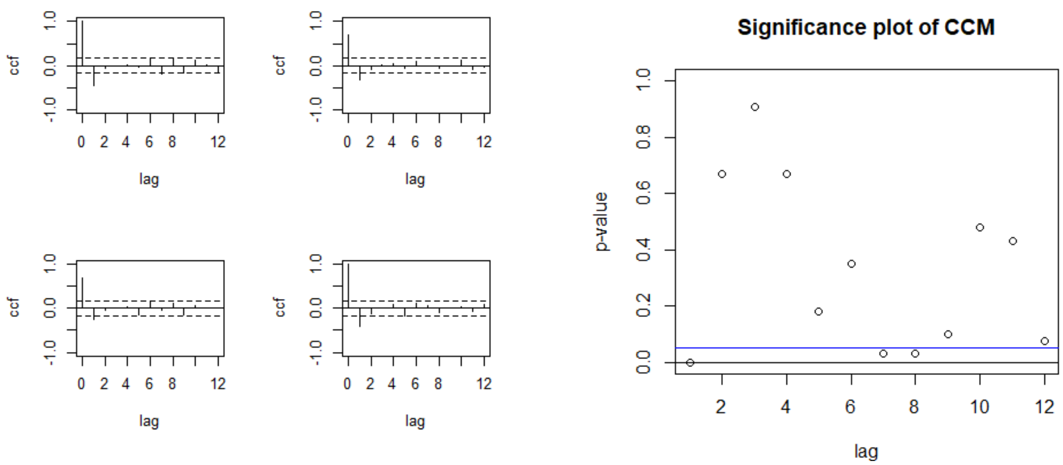


Figure 11. The CCM and p-value of the microwave

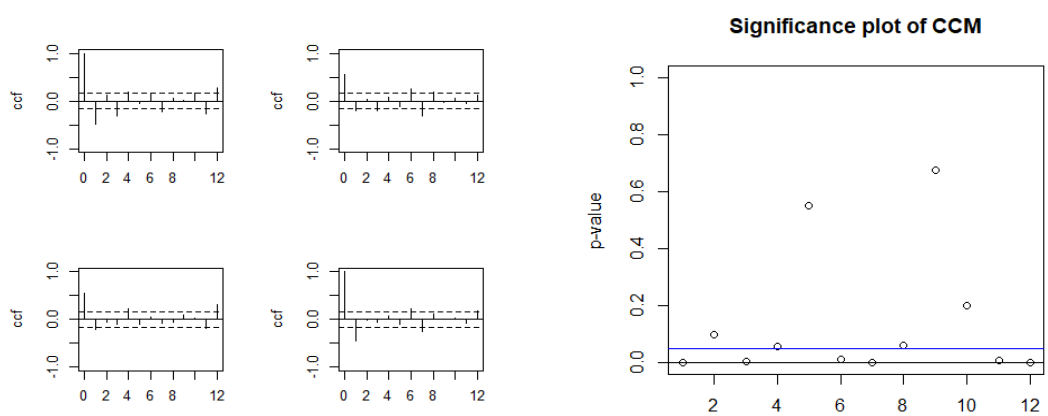


Figure 12. The CCM and p-value of the pacifier

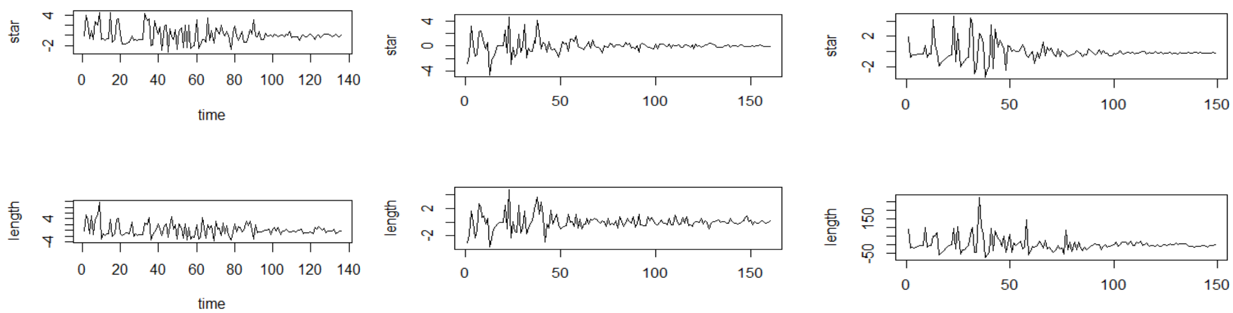


Figure 13. The residuals of VARMA model

	0	1	2	3	4	5	6	7	8
0	0.000	0.025	0.015	0.006	0.003	0.004	0.003	0.013	0.054
1	0.000	0.014	0.010	0.044	0.007	0.005	0.064	0.295	0.579
2	0.007	0.017	0.044	0.004	0.002	0.017	0.767	0.989	0.973
3	0.177	0.049	0.018	0.006	0.999	0.040	0.998	0.965	0.522
4	0.720	0.095	0.035	0.077	0.122	0.008	0.997	0.750	0.298
5	0.863	0.425	0.386	0.958	1.000	0.998	0.997	0.916	0.739
6	0.950	0.879	0.806	1.000	0.990	0.914	0.964	0.849	0.908
7	0.980	0.913	0.999	0.995	0.995	0.939	0.954	0.977	0.800
8	0.986	0.679	0.825	0.999	0.994	0.910	0.365	0.925	0.795

Table 7. The p-value table for extended cross-correlation matrices of microwave

	0	1	2	3	4	5	6	7	8
0	0.000	0.000	0.000	0.000	0.001	0.000	0.002	0.174	0.670
1	0.000	0.000	0.000	0.007	0.409	0.009	0.014	0.937	0.768
2	0.000	0.001	0.000	0.005	0.121	0.386	0.045	0.763	0.931
3	0.001	0.019	0.022	0.062	0.662	0.007	0.003	0.513	0.913
4	0.014	0.176	0.306	0.611	0.386	0.771	0.587	0.386	0.693
5	0.328	0.384	0.624	0.399	0.002	0.495	0.986	0.439	0.750
6	0.387	0.377	0.521	0.297	0.511	0.558	0.854	0.730	0.741
7	0.591	0.947	0.901	0.134	0.021	0.572	0.346	0.599	0.836
8	0.991	0.970	0.962	0.705	0.353	0.168	0.415	0.910	1.000

Table 8. The p-value table for extended cross-correlation matrices of pacifier