Problem Chosen	2020 MCM/ICM Summary Sheet	Team Control Number
C		2006357

Text Mining and Data Analysis with Time Series Based on Amazon Review Data

Summary

As more and more people choose to shop online, online consumer reviews have become an important source of information for sellers, potential customers and researchers. It is significant for sellers to filter out useful items from a large number of product reviews to guide sales.

First, we build a text-emotional analysis model based to caculate sentiment score, quantifying the comment text data after pre-processing, making the value of the quantified evaluation score from large to small to express the positive to negative of the comment, and using PCA to correct the model. And then We use Bayesian average scores to revise the voting ratio for comment usefulness and weight previous scores. Combined with $star\_rating$, we use the entropy weight method to calculate the final comprehensive score of the review(coscore). which can be considered as "reputation" of the products. In this part, we find that there is a correlation between the low rating of the evaluation and the high degree of usefulness.

Next, we use time series analysis to fit and forcast the trend of coscore. According to forcast , hairdryer shows a slightly upward trend, while pacifier shows a slightly downward trend. Microwave is stable in the next several months. And based on forecasts of changes in the reputation of different commodities over time, we can get the best time to enter the relevant market.

Then, we built the TF-IDF model to extract keywords and do cluster analysis, which is based on k-means clustering. We aim to identify the characteristics of different categories of product and recognize the commodity attribute of general concern to buyers. The size and the keypad of a microwave is a big deal, while the balance of power and skin affinity of hair dryers arouse much attention. As for pacifiers, consumers concerned more with the material. We advise Sunshine company to make improvements in these areas. Moreover, we found that different rating stars produce a time.

Finally, what we get is a model of text characteristics, usefulness ratios, rating characteristics and product itself that allows us to analyze the product demand in the corresponding market, and we give Sunshine Company recommendations for product marketing.

Keywords: Review Mining; TF-IDF; Time Series; Text Network

Contents

1 Introduction

1.1 Problem Background

The most intuitive manifestation of a buyer's evaluation of product quality on Amazon's website is a review. Buyers can describe their feelings and opinions about the product while rating the product. These reviews not only provide consumers with feedback on product sellers, but also provide advice and guidance to other consumers. For a certain product, the enterprise can improve product quality and marketing strategy by reading and mining user comments in reviews. Potential consumers can also obtain useful information from other people's reviews of modified products. They are likely to choose to buy, if they are not satisfied with the product information expressed in the corresponding evaluation content, they tend to refuse to buy. The final decision is your purchase intention.

Sunshine Company needs to filter out useful reviews from a large number of product reviews, and extract useful information that can really guide sales and purchases from a lot of redundant information. We have received reviews of purchasers of microwave ovens, baby pacifiers, and hair dryers on Amazon's website, and used these buyers' feedback to develop marketing strategies and product improvement solutions for Sunshine Company's corresponding products.

1.2 Our Goals

Based on our understanding of the problems, we set the following goals:

- Use the given data to analyze the market demand.
- Develop a model system to show the correlations between star ratings, reviews, and helpfulness ratings of each of the three products, and build a portfolio that describes the products' reputation.
- Define the best characteristics of products that fit the customer's needs and establish a system to evaluate the best type of products in the future.
- According to the analysis above, develop a model and set different ratings to predict the impact of ratings on review types and product reputation.
- Based on the established models, decide the timing to market the products and provide advice to the Sunshine Company to make the products on the market successful.

2 Overview

2.1 $star\_rating$ Comparison Among Three Dataset

First, we tried to find the differences of $star\_rating$ between the three sets of data.

Table 1: $star\_rating$ Comparison Among Three dataset

	hair_dryer	**pacifier**	**microwave**
ave_star_rating	3.981	4.299	3.442

It is obvious that pacifier has a higher proportion of 5-star rating, while low-stars rating is rare. Pacifiers currently enjoys a good reputation in the market. In contrast, the microwave has a low 5- star rating and a higher 1-star rating than the other two products, and its current reputation in the market is the lowest among the three. But it's also an opportunity for Sunshine to improve on, and gain reputation in this market.

2.2 Variable

The following graphs are about hair dryer dataset. Pacifier and microwave share the same traits in this part.

2.2.1 Preprocessing

In order to find the relationship between star rating, reviews and helpfulness rating, it's necessary to make preliminary modifications to the original data, and observe the distribution of the three variables respectively. Since the majority of samples have $helpful_voting$ values of 0, and a few of them have large values, $helpful_voting$ is divided into three levels, the frequency and the corresponding relationship with $helpful_voting$ are as follows:

Table 2: The $helpful\_level$

	level 1	**level 2**	**level 3**
helpful_voting	0	(0, 50]	[50, max(helpful_voting)
count	7771	3613	86

Draw a boxplot to observe relationship between $helpful\_level$ and $word\_num$, sen_num.

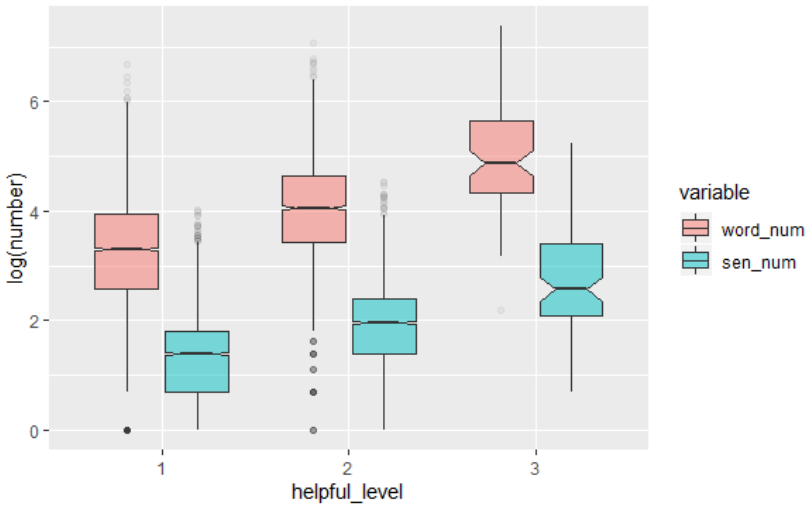

Figure 1: The correlation between $helpful\_level$, $word\_num$ and sen_num

The higher the $helpful\_level$, the more words there are in the comments and the more sentences there are. It is reasonable to assume that because more detailed reviews describr the product informatively to other buyers, thus they receive more $helpful\_votings$.

The height of boxes at different levels is almost the same. The comment characteristics of $helpful\_level$ are relatively stable.

2.2.2 chi-square

In order to test whether there is a correlation between these two discrete variables, the chi-square test should be used. The null hypothesis is that the two discrete variables are independent of each other. The test results are as follows:

Table 3: chi-square

Pearson's Chi-squared test	
df	8
p-value	<2.2e-16

According to this p-value, we reject the null hypothesis. $helpful\_level$ & $star\_rating$ not independent.

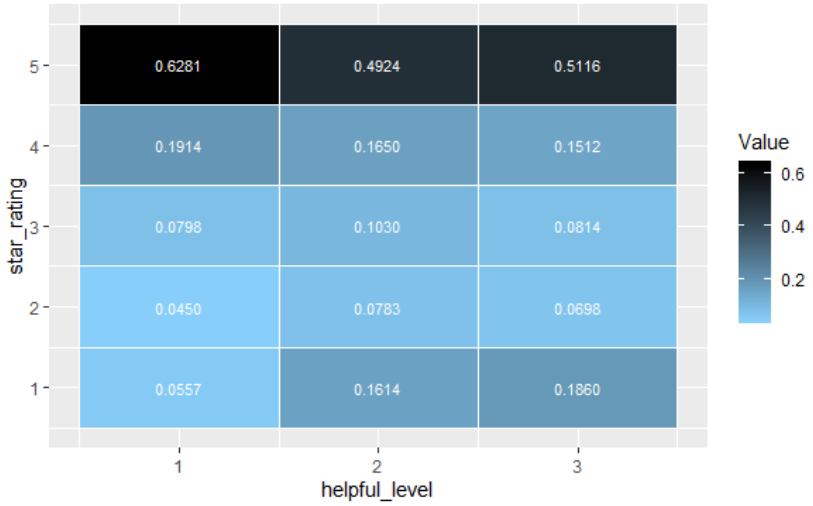

Figure 2: The overview of our model

The data in the figure below indicates that the proportion of comments for star review corresponds to the proportion of $helpful\_level$, with a vertical sum of 1. Firstly, top line of heatedmap are the most dark color, indicating that no matter which $helpful\_level$, the five-star review is always the most. $helpful\_level4$ reviews of 2 or 3 are much higher than those of two stars. This may be due to the low-star evaluation will show some detailed description of the shortcomings of the product, and maybe this information is exactly what consumers need. So its $helpful\_level$ is also relatively high.

2.2.3 Conclusion

As can be seen from the relationship and pattern, the comments for $helpful\_level$ 2 and 3 are more linguistically rich in their own. The number of words and the number of sentences are more, and low-star reviews are more likely to occur in both categories. Sunshine should pay attention to the information $helpful\_level$ the content of the higher comments. It is convenient to improve its own products.

2.3 The Overview of Our Models

We summarize the ideas and methods we use in the flowchart below

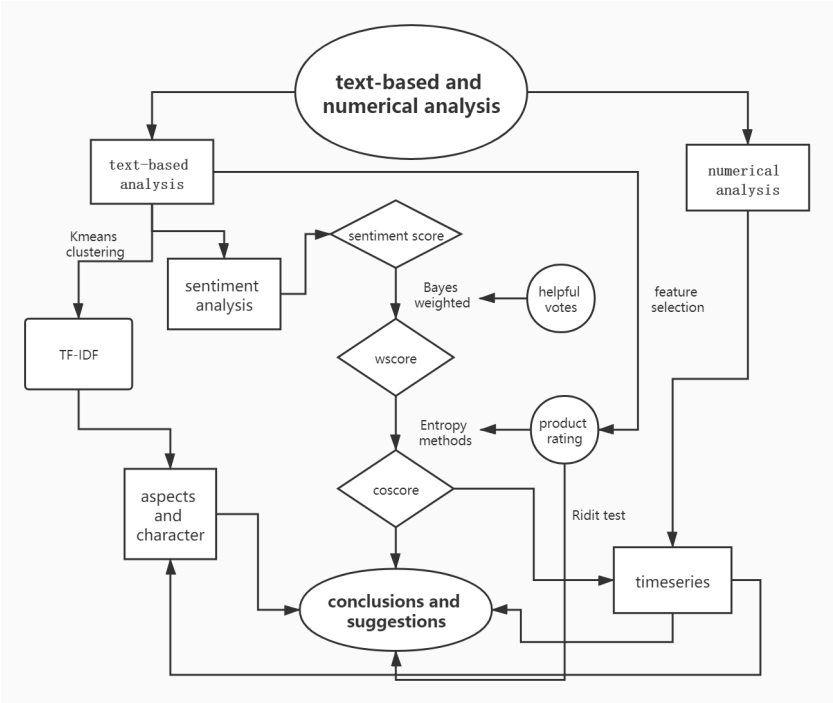

Figure 3: The overview of our models

3 Models

3.1 Model.1 The Sentiwordnet

The sentiment classification mainly uses the semantic lexicon as the basis of judgment. Their main idea is to count the words of various parts of speech and then aggregate them to form a set of words of various parts of speech. The user evaluation information package of the products contains multiple sentences, and each sentence is composed of multiple phrase models. Therefore, the analysis of the emotional tendency of user evaluation information is to analyze the emotional tendency of the phrases contained therein.

3.1.1 The Foundation of Model

In sentiwordnet (the word network we use) itself has a positive emotional score for each word. We extract the adjectives, verbs, and nouns, and sum the positive emotional scores of these words Measure positive sentiment score for the entire sentence. But considering that score cannot have a strong linear trend similar to the length of the sentence. And the lack of sentiment in the sentence will have some relationship with the length (eg. That's good, that's good. Positive emotions should be better than saying it, but it will not be twice as strong as before), so the process of dividing root by length is used (the length here refers to the sum of adjectives, verbs, and nouns)

So we can get the number of times a comment has a positive word and denote it as

$sum\,(posScores)$, then we get the equation for calculating the sentiment of this comment is

$$S = \frac{sum\,(posScores)}{\sqrt{len\,(posScore)}} \tag{1}$$

where $len\,(posScore)$represents the number of emotional words in the comment

3.2 Improved Model.1 Bayesian Truth Serum

We use Bayesian Truth Serum to improve our model. In the data, more than half of the reviews received only 0 and 1 votes in total. There is a large variance in the total helpful votes and there is a significant long-tail effect. Therefore, the relative level of usefulness of each comment can be more accurately restored by using Bayesian Truth Serum. We use the number of helpfulness votes and the total number of votes for all reviews in the data to calculate the overall average value of the helpfulness ratio which is used as the prior average.[1]

3.2.1 The Foundation of Model

The Bayesian Ratio is the weighted average of the helpfulness vote ratio for that review and the average vote ratio for the full data set.

$$Bayesian\ Ratio_{i,j} = \frac{w\,(a) \times Average\ Ratio + NO.of\ Helpful\_votes_{i,j}}{w\,(a) + NO.of\ Total\_votes_{i,j}} \tag{2}$$

$$Average\ Ratio = \frac{\sum_{i,j} NO.of\ Helpful\_votes_{i,j}}{\sum_{i,j} NO.of\ Total\_votes_{i,j}} \tag{3}$$

Table 4: Symbols of Bayesian Ratio

Symbol	Meaning
i	Product
j	Each review
$\sum_{i,j} NO.of\ Helpful\_votes_{i,j}$	The total number of helpful votes received for all reviews in the set
$\sum_{i,j} NO.of\ Total\_votes_{i,j}$	The total number of votes received for all reviews in the set
$NO.of\ Helpful\_votes_{i,j}$	The number of helpful votes received for a reviews
$NO.of\ Total\_votes_{i,j}$	The total number of votes received for this reviews

We can draw its distribution from the Wscore calculated by Bayesian Ratio, from which we can see that the three commodities have similar distributions, all of which are right-leaning. This information generally indicates that people tend to give more positive comments than negative ones.

[1]Prelec, D.. A Bayesian Truth Serum for Subjective Data. Science, 2004, 306(5695): 462.

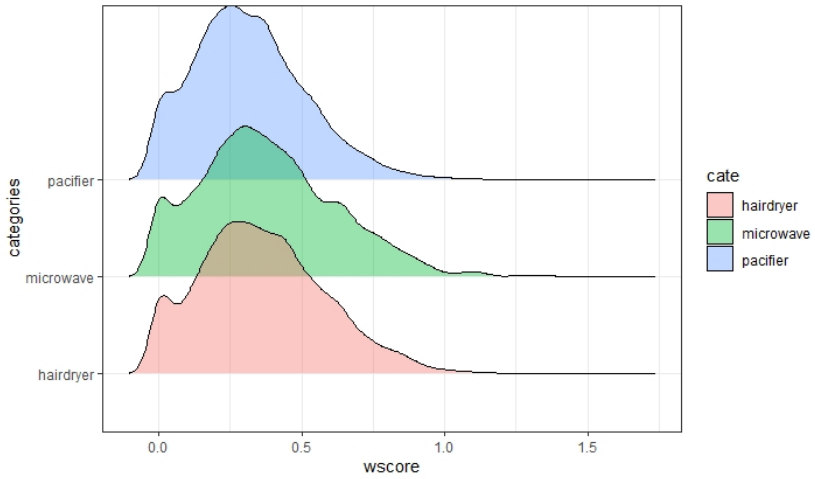

Figure 4: The wscore with Bayesian Ratio

3.2.2 K-means clustering

The comments were clustered according to $wscore$ (weighted sentiment score) and $star\_rating$. Considering the comment homogeneity of products with the same $product\_parent$, we average the values of these two variables, then do clustering. As a result, we cluster 55 kind of microwave product.

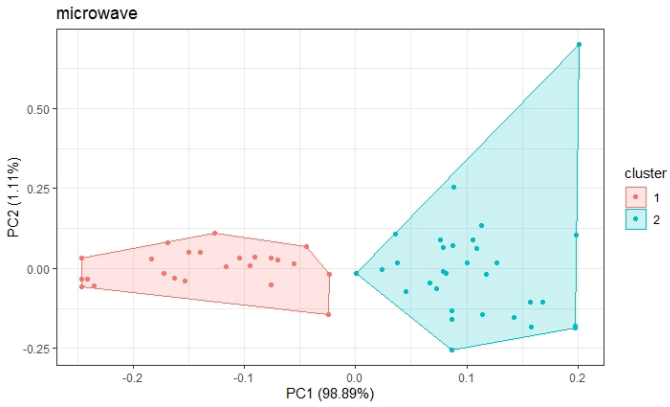

Figure 5: Clusterplot of microwave

The variance contribution rate of the first principal component was 98.89%. The variability of the data is basically explained. Since PC1 is actually calculated by $star\_rating$ and $wscore$, which is weighted by the sentiment score of the review text, and $star\_rating$ is also related to the tendency of emotion, it is reasonable to guess that these two categories may be the division of favorable and unfavorable comments. To test this idea, the next step is to perform a text analysis of the comment text $review\_body$ owned by each of the two categories of products, which is explained in Model 3.

3.3 Model.2 Time Series Analysis

The weighted sentiment score of each observed $star\_rating$ and $BayesianTruthSerum$ is calculated by the entropy method, and the linear weighting constitutes the $coscore$. This index contains information about ratings and reviews. We analyze this indicator in detail.

We now explore the appropriate time to enter the market by constructing a time series model (the change of this indicator over time). And we can filter out a certain category of $product\_parent$ with a higher or lower average share over a longer period of time, and extracting its characteristics. Based on these characteristics, the commonality of people's preferences for a class of goods and the differences and changes in people's attention to product features over time are identified.

Table 5: The Pre-Processing

Data Set	Time	number
hairdryer	2005.04∼2015.08	125
microwave	2011.08∼2015.08	47
pacifier	2006.10∼2015∼0.8	107

3.3.1 The Foundation of Model

First, we draw a figure. According to the figure, we can preliminarily observe that the reputation of this product has a positive trend with the development of time and the fluctuation is smaller and smaller. It indicates that the performance of hair dryer has been continuously improved in recent years and has been recognized by more and more users.

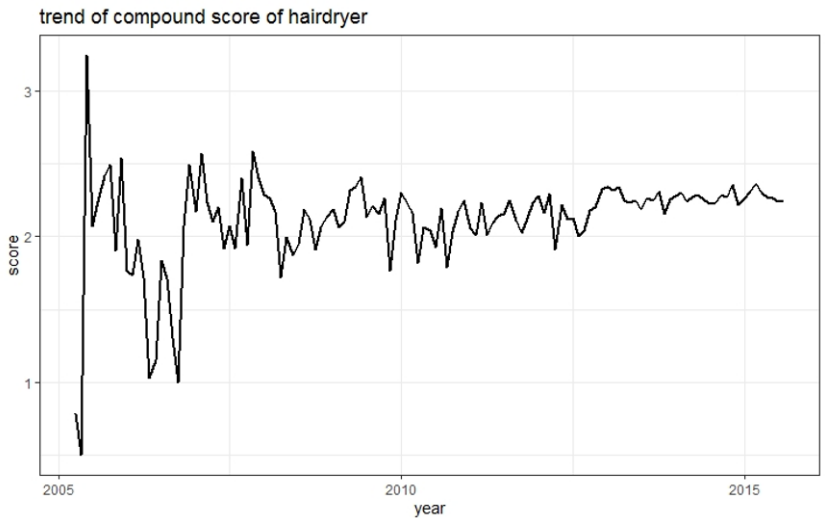

Figure 6: The time series of hairdryer

We draw the autocorrelation and partial correlation figure of the time series. It can be seen from the figure that both acf and pacf show when lag > 2. We preliminary judged that we can use the ARIMA (p, d, q) model to fit.

According to the performance of hairdryer data, we found that the sequence after doing the second-order differential is stable (the acf and pacf graph sits below). According to the illustration, we find that the self-correlation coefficient after lag one into a trailing shape and the partial coefficient after lag s 3 also into a trailing pattern. Thus we can fit ARIMA (1,2,3), which takes into account AIC and Log-likelihood; we fit ARIMA (1,2,1).

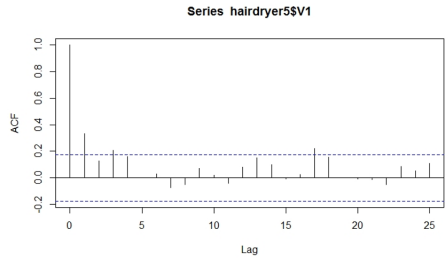

Figure 7: The autocorrelation

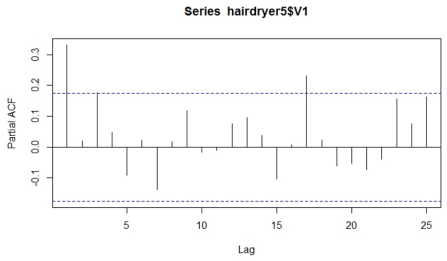

Figure 8: The partial correlation

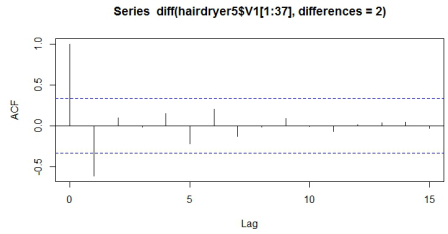

Figure 9: The acf

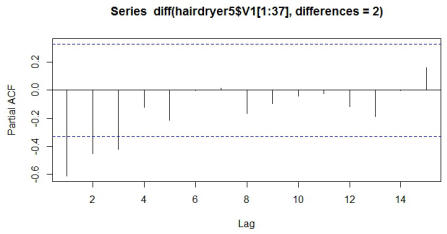

Figure 10: The pacf

The model is calculated by R software, and the form is as follows:

$$\nabla^2 X_t = -0.403\nabla^2 X_t + \epsilon_t - \epsilon_{t-1} \tag{4}$$

where $\nabla^2 X_t = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2}$

Table 6: Model Summary

	Hairdryer	**Microwave**	**Pacifier**
differnce	2	NA(white noise)	1
Ar1	-0.403		-0.103
Ma1	-1.000		-0.980
Ar2	NA		0.252
Ma2	NA		NA
Ar3	NA		0.180
AIC	92.66		19.98
Log likelihood	-43.33		-4.99
Sigma^2	0.1131	0.0264	0.0652
Mean	NA	2.669	NA

After the model parameters are estimated, we have to carry out white noise test of the residuals of the model. If the residuals are white noise, the parameters of the model have excellent estimation properties. We examined the residual white noise test of lag = 1, 6, 12 separately; we could not reject the null hypothesis at 90% confidence, that is, we considered the residual is to be a white noise sequence. The tests for the other two datasets are shown in the following table.

Table 7: White noise test of sequence residuals

Data	lag	X-squared	df	pvalue
hairdryer	1	0.12	1	0.973
	6	8.56	6	0.201
	12	15.91	12	0.196
pacifier	1	0.10	1	0.749
	6	9.21	6	0.162
	12	16.28	12	0.179

3.3.2 Solution and Result

Based on the time series obtained, it is possible to predict trends in future comprehensive scores, thus determining the timing of market entry. In the following chart , the years after 2015 (not too long), the sequence has a tendency to rise over time, so that hair dryer sellers can choose to enter the market.

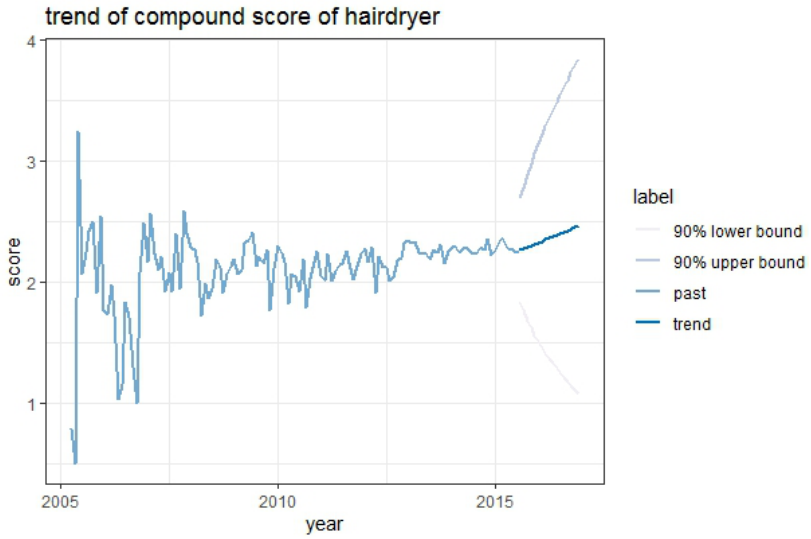

Figure 11: The trend of score of hairdryer

3.3.3 Supplement the Microwave and Pacifier Conclusion

We draw a line plot of microwave and pacifier data based on the time period intercepted above.

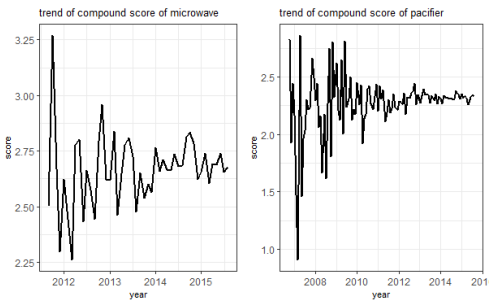

Figure 12: The trend of microwave and pacifier

Microwave oven data itself is white noise sequence after the white noise test. It is found that the lag 1 to 6 test did not reject the original hypothesis, so can not refuse the null hypothesis. That is, microwave ovens fluctuated up and down with a fixed mean (2.67) between 2011 and 2015, so there was no difference in the average for sellers to enter the market after 2015.

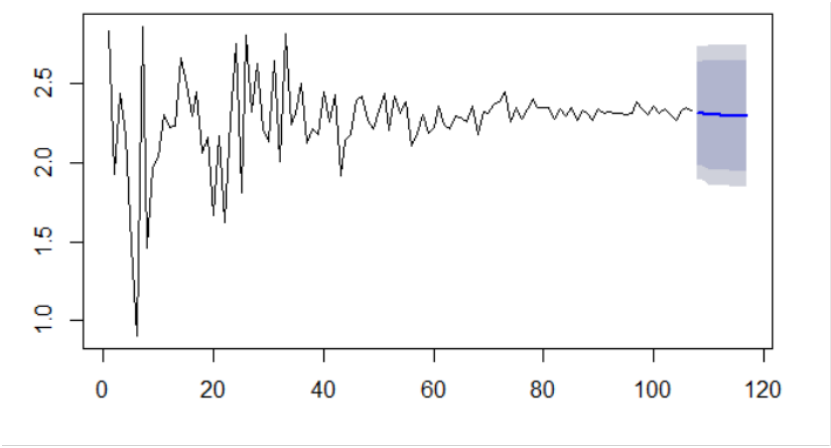

Figure 13: Forecast from ARIMA(3,1,1)

The result of the pacifier sequence fitting is ARIMA (3,1,1), and the residual sequence is also white noise, indicating that the estimated nature is good. The forecast is shown below. The forecast results show that over the next 1 to 2 years, the pacifier 's reputation is in a slight decline in volatility. So the next 1 to 2 years is not very suitable for merchants to enter the nipple market.

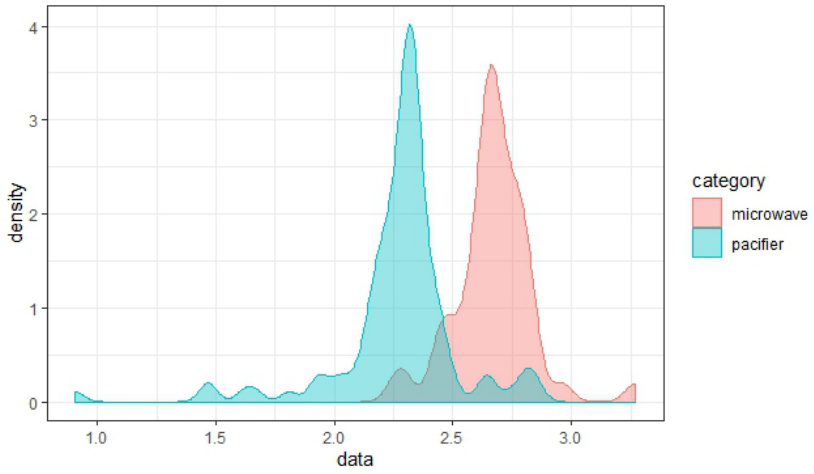

Figure 14: Horizontal comparison

Comparing horizontally, the reputation of pacifier is obviously not as high as the microwave oven. The possible reason is that the pacifiers' audience is toddler's parents, and the products are related with their children's health. The product requirements are higher and the score is more stringent.

3.4 Model.3 TF-IPF

In order to combine numerical data and textual data to get a feature that can predict the potential success and failure of products, we can use the clustering model generated previously

in Model 1 to get numerical index. After clustering, the texts are analyzed to identify the characteristics of the clustered category. According to the word frequency outcome, we found it just corresponds to the classification of favorable reviews and unfavorable reviews. We marked the favorable review category as category 1 and the unfavorable category as category 2. Then, observe the useful feature nouns, and then go back to the text to extract the reviews that mentioned this specific useful feature noun, and put forward some suggestions for improving the characteristics of commodity attributes. To quantify the difference between the two categories of text, we introduce TF-IDF model.

The purpose of TF-IDF algorithm is to evaluate the importance of a word to a text. If a word or phrase appears frequently in an article and less frequently in a document set, it is considered that the word or phrase has a good ability to distinguish categories.

3.4.1 The Foundation of Model

Assuming that the frequency of feature words i in the review text$d$ is$tf_i\left(d\right)$, and n_i is the number of texts containing feature words, the function of TF-IDF is

$$TFIDF_i\left(d\right) = tf_i\left(d\right) \times \ln\left(\frac{N}{n_i}\right) \tag{5}$$

l refers to the total number of characteristic words. In order to weaken the suppression effect of very few high frequency words on low frequency words, it is necessary to normalize the above TF-IDF values.

$$V_d = \sum_{i=1}^{l} \sqrt{\left(TFIDF_i\left(d\right)\right)^2} \tag{6}$$

$$TFIDF_i\left(d\right) = \frac{TFIDF_i\left(d\right)}{V_d} \tag{7}$$

3.4.2 Steps

- Combining the two types of text as a whole, a dictionary is constructed. Text vectorization is conducted according to this dictionary. Each comment corresponds to a vector, which contains the TF-IDF score calculated as above.
- At this point, seperate the vectors according to the previous clustering, and the words with the largest difference in tf-idf between the two categories can be found, and the ranking is conducted from large to small. In particular, since the number of reviews in each of the catogory is basically the same, it's reasonable to use TF-IDF value difference to measure the differences directly. By subtracting the score of each word in one review category from the score of the word in the review of the other category, words with larger difference are the characteristic word of that category.
- According to the characteristic words extracted from the above methods, the first three characteristic words related to commodity attributes were extracted in order of the degree of specificity (that is, the degree of difference between the two reviews on this word), and we got the table.

3.4.3 Analysis

Now we take the first feature of class 1, space, and revert to its original review, looking foradvice on microwave space usage.

Some people love its size because of some special uses, while some others are actually saying something positive while teasing a little about how the microwave is too big and taking up too much space. Similarly, we now extract the first feature word of the second category: keypad. Expectedly, we can find improvements on keypad.

As expected, microwave keypad section didn't do well, and many people commented negatively on it. For example, the plastic film on the keypad, shown in the chart above, bubbled after a few months. It didn't do well in typing numbers as well.

According to these two key words, we draw a conclusion that users with more positive evaluation think the space occupied by microwave oven is too large and unreasonable, while users with more negative evaluation think the keypad is not good. In other words, the defects caused by the space problem will not strongly affect the user's evaluation, while the defects of the keypad will ruin the customer's view of the microwave. If Sunshine is going to enter the online market of microwave ovens, the first thing it should improve is the keypad. Then the improvement of space factor is considered.

Table 8: The feature words

microwave1	**microwave2**	**hairdryer1**	**hairdryer2**	**pacifier1**	**pacifier2**
space	error	blow	skin	easy	straps
size	keypad	cord	nozzle	gift	bags
price	plastic	power	iron	soft	fabric
...	...	...	...	...	...

- For the other two data sets, the same analysis can be carried out to provide advice to sunshine company in the online market of these three commodities
- Microwave: Improve the keypad. As for the size of the space, the designer can make both big and small to meet different needs. Hair dryer: Keep the hair dryer high power, but improve the design of the nozzle as well as the cord material. Ensure high power and fast blow-drying while also considering how to avoid damage to the skin.
- Pacifier: The pacifier is intended for children, so its characteristic of softness should bemaintained. Some people also use it as a gift to give to the parents who need it. The accessories of the existing market for baby pacifiers are not as good as they should be. Both strap and bag have been criticized for being unusable and inconvenient. In addition, better materials should be used in making baby pacifiers.

4 Applications of Our Models

According to the $star\_rating$ in the first thirty days, the six characteristics of the reviews in the last seven days were respectively regressed, that is, users could see whether the weighted average of $star\_rating$ in the previous months had an impact on their review characteristics. To reduce the unstable factors in the relationship, we only selected samples with the number of comments higher than two, and the number of comments greater than two in the corresponding time interval for regression analysis.

Table 9: Comprehensive features of reviews

FUNCTION	PARAMETERS	DESCRIPTION
CV	{CF, PF}	Character of the comments based on different factors
CF	length	Review length based on number of words
	sen_number	Number of sentences based on the punctuation marks
	ave_length	Mean sentence length
PF	adj_number	Number of words that contain emotion
	num_number	Number of features included
	v_number	Number of verbs

4.1 Description of The Correlation of Variables

To reduce the unstable factors in the relationship, we only selected samples with the number of comments higher than two, and the number of comments greater than two in the corresponding time interval for regression analysis.

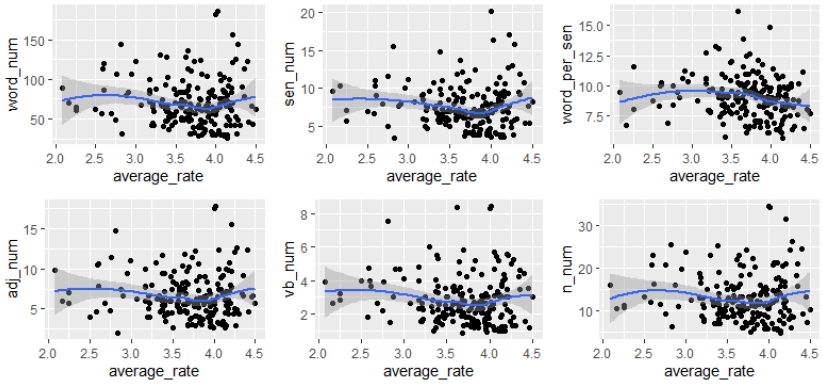

Figure 15: The sensitivity analysis

According to the figure, if the user views one month before the ratings are higher or lower, so the length of the user reviews, sentence number, number of commodity characteristics, characteristics of emotional word number, behavior word quantity are more. On the contrary, if the users see one month before the ratings of the center, such as most of the three-star and four-star evaluation, then the user's comment on the above five indexes will be lower. The rule of average sentence length is different from the other five characteristics. If the user sees a series of low comments or a series of high comments, the average sentence length of his comments will be shorter; on the contrary, it will be higher.

4.2 Analysis of Regression

Assume that the above relation, if it exists, is linear. Using the method of stepwise regression ($verified\_rate$ were not by t test, so be ruled out, finally only constant term, $common\_rate$ and $vine\_rate$). To verify the numerical correlation of the above characteristics, we use the following three independent variables to carry out regression on the dependent variables to observe whether the obtained linear model is significant. If it is significant, it is considered to be significantly correlated; otherwise, it is not considered to be correlated.

Table 10: Regression coefficient and p-value

Y	X	Coefficients	Signif	p-value
word_num	(Intercept)	68.295	0.001	0.02636
	common_rate	-8.904	0.01	
	vine_rate	7.849	0.1	
sen_num	(Intercept)	7.4663	0.001	0.04815
	common_rate	-0.7487	0.05	
	vine_rate	0.6619	0.1	
word_per_num	(Intercept)	9.7434	0.001	0.001279
	common_rate	-0.7072	0.001	
	vine_rate	0.4174	0.1	
adj_num	(Intercept)	6.0678	0.001	0.05538
	common_rate	-0.8181	0.01	
	vine_rate	0.8366	0.05	
vb_num	(Intercept)	2.8939	0.001	0.1072
	common_rate	-0.3417	0.05	
	vine_rate	0.2743	1	
n_num	(Intercept)	11.8746	0.001	0.02692
	common_rate	-1.6209	0.05	
	vine_rate	1.6494	0.05	

By observing the p-value in the above table, we can be confident that these three independent variables are related to the four variables of sentence length, number of sentences, the average number of words per sentence, and the number of behavior words. However, they are not directly related to the two variables of characteristic commodity quantity and characteristic emotion word quantity in the confidence interval. Significance test results of observation variable and constant term highly significant, $common\_rate$ is more significant. The $vine\_rate$ don't show the $verified\_purchase$ the model excludes, shows the user's evaluation although associated with the rating of a month before. But that does not produce too much volatility. When a user reading rating, Whether or not the person who wrote the rating is vine makes only a small difference. And write the ratings are $verified\_purchase$ almost does not affect.

4.3 The Association Between Reviews and Ratings

Take the pacifier as an example to show the data. The other two data sets have a similar pattern. Usually, adjectives are words with strong feelings, We can start by looking at the relationship between the number of adjectives in the comments and $star\_rating$.

Intuitively, We can see that with the increase of $star\_rating$, the number of adjectives tends to be higher, which is consistent with our hypothesis: most of the emotional words are adjectives, so we can preliminarily considered that the words with emotional color and rating are strongly associated. Then, univariate regression was conducted for the above relations, and the regression equation p-value < 2.2e-16 was obtained. Now, we are more confident about this correlation.

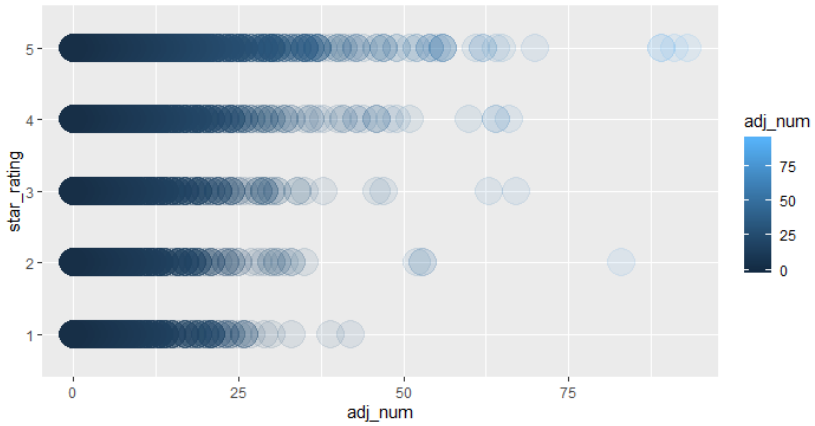

Figure 16: The association between reviews and ratings

Further, to determine whether the ratings were significantly different if words with specific strong emotions were included in the comments. We used word frequency statistics to find words with higher frequency. Negative: "disappoint", "bad"; Positive: "comfortable", "cute". Filter the data that contains these words, the negative is 312 and the positive is 2172 (as shown in the following table).

Because the rating mechanism is five grades, the amount of data may be inconsistent with the objective reality. We use the Ridit test in the nonparametric method. His basic principle is to go to a group with a large number of samples or summarize several sets of data into a reference group. According to the sample structure of the reference group, the original response number of each group is changed into a reference score – Ridit score, and we used the transformed Ridit score to get a fair comparison of the strengths and weaknesses of each treatment.

Table 11: Word frequency table of strongly emotional words

	one star	**two star**	**three star**	**four star**	**five star**
Disappoint & bad	54	40	51	64	103
Comfortable & cute	51	81	185	322	1533

To compare the strength differences between the two treatments of "negative word" and "positive word," we use the ridit.test () function in R to test. The following table shows the test results. According to p-value=4.101e-49 < 0.001 and Mean Ridit(Comfortable & cute) > Mean Ridit(pleasant &cute), reject the null hypothesis, indicating that the score of negative words such as Disappoint & bad in the comment will be lower than that of positive words such as Comfortable & cute. Furthermore, it shows that the customer's experience directly influences the rating of the goods.

Table 12: The Ridit.test

Comfortable & cute	**Disappoint & bad**	**Chi-squared**	**df**	**p-value**
0.5272	0.3103	216.99	1	4.101e-49

Visualization of this relationship is as below.

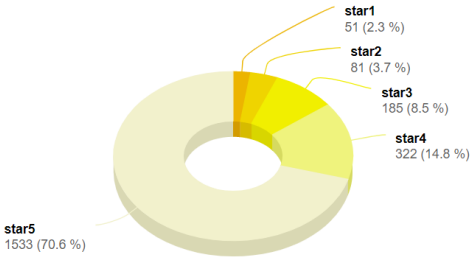

Figure 17: Have bad & disappointed

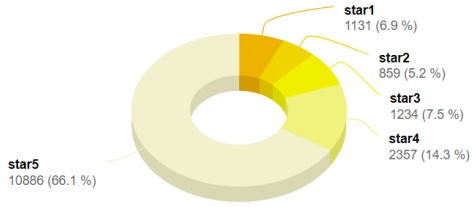

Figure 18: No bad & disappointed

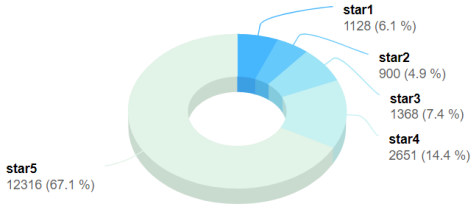

Figure 19: No cute & comfortable

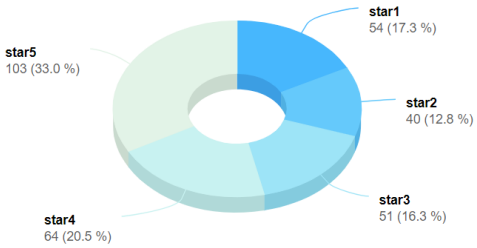

Figure 20: Have cute & comfortable

5 Sensitivity Analysis

In the Bayesian weighting factor calculation, the prior value of the factor is calculated to be 0.8501. We assume that the parameter w0 = 100, and adjust the weighting factor for each piece of data according to the formula to give different weights. The mean of the weights in the sample = 0.8496 and the standard deviation = 0.0088. We set w0 to different values, and we can calculate such a set of mean and variance, as shown in the figure. It can be observed that when w0> 100, both the mean and variance tend to stabilize. When w0 <50, the mean and variance change too quickly. Therefore, it is considered that it is reasonable to take w0 as a number between 50 and 100, which can play a role of weighting without expanding this role excessively.

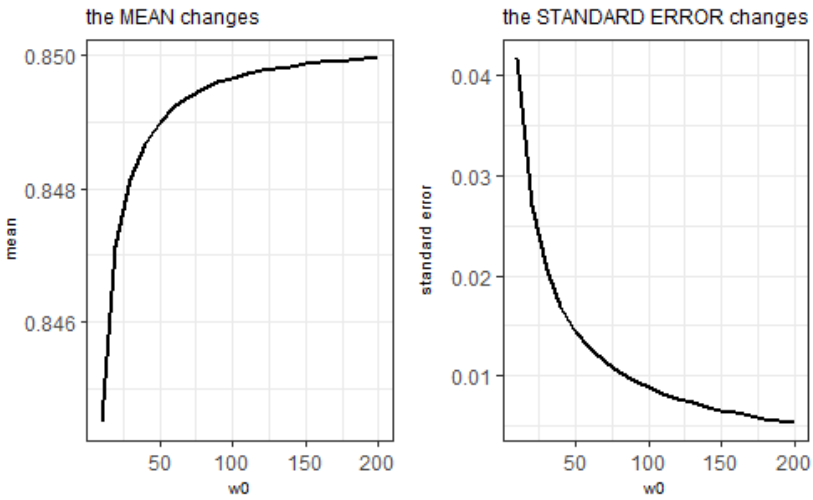

Figure 21: The sensitivity analysis

6 Letter

Dear Marketing Director of Sunshine Company:

It is my pleasure to give a market recommendation to Sunshine Company. We analyzed customer feedback data online. Based on the analysis, we propose three models to discover the customer demand, market trends, and available business opportunities contained in the review data from different dimensions.

As we know, the key to selling a product successfully is to analyze its market competitiveness, predict its reputation trends, and find the direction for improvement. So our model is based on these three factors.

First of all, we segmented the comment text and performed the correlation analysis from the length of the most intuitive comment text, the number of adjectives, the product rating, and the usefulness of the comment. We found that the longer the review text, and the more identifiers or product feature in the review, the more useful the review is. The greater the information a review contains, the more likely it is for potential consumers to remove the uncertainty associated with shopping. Therefore, we suggest that after you company put the products into the online shopping platform, the company can offer some discounts to customers who have already bought the products, and encourage them to write more comments on the website. Potential users are more inclined to read the comments and understand the products, so they are more likely to choose to buy the products.

Then, we dig deeper into the comment text data. Each comment was given an sentiwordnet using a text-based emotional score. According to the score, we can divide the text into positive sentiment evaluation and negative sentiment evaluation. Combining product rating, user characteristics, and evaluation usefulness indicators, we can get an overall score on the assessment, which is a manifestation of commodity reputation. Using the comprehensive scoring model of comments, we can quickly and comprehensively measure the usefulness and emotional tendency of comments. The distribution of the comprehensive score of product reviews can help us to screen the appropriate products to achieve our goals.

Therefore, we added the factor of time and fitted the comprehensive score of goods to get the trend line of time. Through this trend line, we can see the trend of the reputation of different commodities changing over time. We found that the hairdryer has a relatively stable and rising trend over time, so we have better expectations for this product. Immediately investing in a product can help you earn dividends from arising product reputation.

Furthermore, we also introduced the TF-IDF model to analyze the traits of words with positive and negative emotional evaluation. We could obtain the customer's assessment of the characteristics of the goods. We found that the frequency of the word's space was very high in the words of a positive evaluation of microwaves, while keypad frequently appeared in the set of negative assessments. Therefore, our suggestion for your company is to improve the control keypad, and on the basis of this step, consider the improvement of space. Similarly, we suggest to maintain the power of the hairdryer and consider skin protection. The existing pacifiers on the market have feature of soft which we continue to maintain. Although it is mainly used by babies, it is often given to parents as a gift. The accessories of the existing market for baby pacifiers are not as good as expected. Both strap and bag have been criticized not easy to use. So we suggest the company could decorate it and improve its accessories to attract people who want to buy it as a gift. In addition, use better materials when making baby pacifiers.

Thank you for taking the time to read our suggestions. We sincerely hope that our model and advice can be helpful to you!

Sincerely,

MCM Team Members

7 References

[1]Brinton, C. G., Chiang, M.. The Power of Networks: Six Principles that Connect Our Lives. Princeton University Press, 2016.

[2]Bartholomew, David J. "Time series analysis forecasting and control." Journal of the Operational Research Society 22.2 (1971): 199-201.

[3]Haque, Tanjim Ul, Nudrat Nawal Saber, and Faisal Muhammad Shah. "Sentiment analysis on large scale Amazon product reviews." 2018 IEEE International Conference on Innovative Research and Development (ICIRD). IEEE, 2018.

[4] Hogg T.. Inferring Preference Correlations from Social Networks. Electronic Commerce Research & Applications,2010, 9(1): 29-37

[5] Hu, N., Zhang, J., Pavlou, P. A.. Overcoming the J-shaped Distribution of Product Reviews. Communications of the ACM, 2009, 52(10): 144-147.

[6]Lim, Ee-Peng, et al. "Detecting product review spammers using rating behaviors." Proceedings of the 19th ACM international conference on Information and knowledge management. 2010.

[7]Mehrotra, Rishabh, et al. "Improving lda topic models for microblogs via tweet pooling and automatic labeling." Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. 2013.