Problem Chosen	2020	Team Control Number
C	MCM/ICM Summary Sheet	2007707

Strategies to Online Sales: A Review-tracking System Combined with Deep Learning and Difference Equation Model

With the booming online shopping platforms, online reviews play an increasingly significant role in customers' purchase decisions. How to identify the time-based patterns of reviews and future reputation of newly launched products is a major concern for companies.

In order to address this problem, we make statistics on review indicators of the hair dryers, pacifiers, and microwaves. With a preliminary understanding of data, we set five evaluation indicators based on star-ratings and text-based reviews, namely the average star-rating, favorable rate, number of reviews, average number of review words, and average review sentiment value. We then build a text sentiment analysis model based on deep learning, extract keywords in the reviews by **Term Frequency-Inverse Document Frequency** (**TF-IDF**) algorithm, and calculate the sentiment value of reviews by the **back propagation** (**BP**) neural network. In this way, we explore the specific relationship between review sentiment and star-rating levels and propose design focus for the three products. Later, we build a **difference equation model** to identify the changing patterns of each evaluation indicator as time and other indicators change based on the correlation test results of the five evaluation indicators. Finally, we construct the success index by **principal component analysis** (**PCA**) based on the evaluation indicators, which provides the company with an easy and convenient system to track potential successes or failures of a product.

We find some insightful conclusions based on our results as follows. **(1)** Star-ratings, favorable rate, number of review words, and review sentiment value all decline over time. For hair dryers, high review sentiment value will reduce the increase in star-ratings in the next quarter. Similarly, high star-ratings and favorable rate will also lower the review sentiment value in the next quarter. **(2)** Reviews with low star-ratings cause an increase in negative sentiment, while those with high star-ratings have no significant effect. Also, favorable reviews are strongly associated with high star-ratings, but negative ones are not significantly related to them. **(3)** The emotional words rank the first in the importance of reviews, followed by descriptive adjectives and characteristics words of products. **(4)** The success index effectively reflects the operating conditions and future reputations, which is a valid evaluation indicator.

Based on the above analysis, we provide some strategies on product sales and design for Sunshine Company, which includes **(1)** the design of hair dryers, pacifiers and microwaves should focus on "high-power, light, portable'', "cute, soft, safe", and "easily installed, suitable size, timely repaired", respectively; **(2)** increase the yield to reduce concentrated negative reviews instead of pursuing favorable reviews blindly; and **(3)** ensure that the future product reputation is above the average by employing the success index for timely tracking.

In the end, we make sensitivity analysis, and verify the model's robustness and result adaptability. In a nutshell, our model is **accurate** in sentiment analysis, **consistent** with the reality, as well as **simple, effective** and **practical** in tracking future operating conditions.

Key words: Online Sales Strategies, Text Sentiment Analysis, Deep Learning, Difference Model, Principal Component Analysis

Contents

1 Introduction

Would you refer to online reviews before you purchase a product online? The past decade has witnessed the booming e-commerce. In this context, customers prefer to shop online for convenience, lower cost, and diverse products [3]. Although online shopping makes it easier to reduce cost and match the demand, it also increases the risks of information asymmetry. In order to address the problem, major platforms provide online reviews for potential customers.

Generally, online reviews refer to the positive or negative opinions towards products or service published on websites by consumers [4]. They provide convenience for both customers and companies. For customers, they can rate their purchases and give feedbacks on these platforms, which provides valuable reference for other potential customers. For companies, they can get insightful findings by exploiting the online review data [13]. According to Jupiter Research, online reviews play an increasingly important role in the online market[1].

Online reviews have caught the attention of a multitude of scholars. They mainly focus on content characteristics, such as star-ratings [6], content length [9], and positive or negative reviews [11]. Other scholars try to figure out the impacts of reviewer characteristic, such as reviewer verification [1], and the number of published comments [5]. These literatures study the impacts of online reviews from the aspects of reviews and reviewers.

In this work, we are required to (1) identify relationships among the parameters in text-based and rating-based reviews through time-based data interaction; and (2) provide Sunshine Company with insightful findings on new product launch.

The overview of this work is presented in Figure 1.

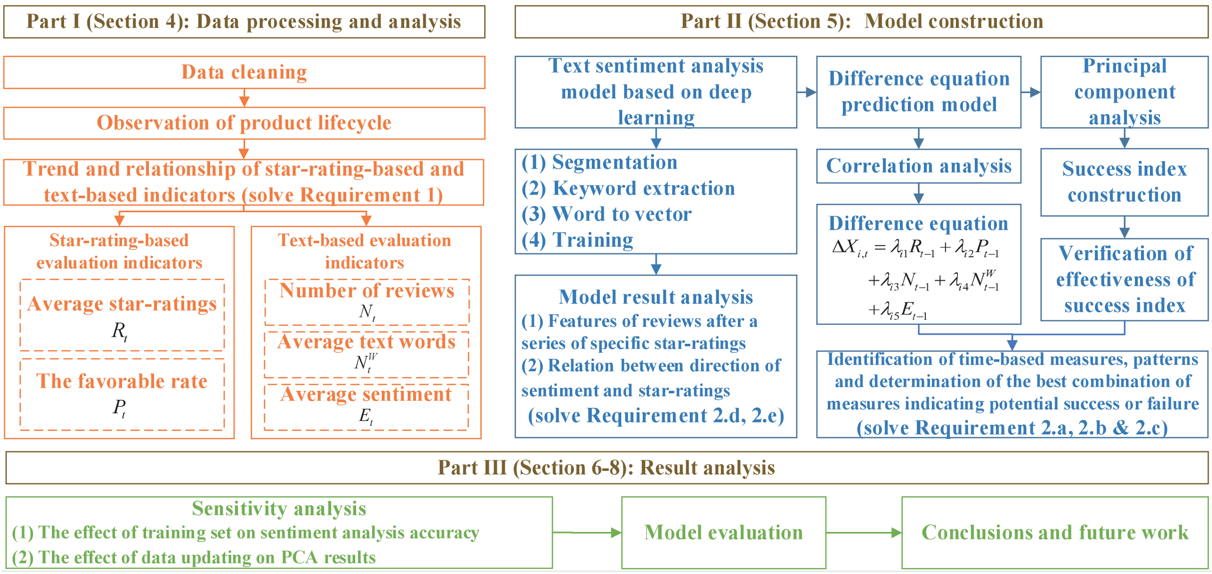

Figure 1: Overview of this work, including data processing and analysis, model construction, and result analysis.

[1]http://web-strategist.com/blog/2006/08/21/jupiter-people-read-user-generated-reviews-before-buys/

2 Assumptions

To simplify the problem, we make the following assumptions.
(1) **The data given in the problem are true and reliable.** The instruction sets the restriction that the data files provided contain the only data we should use for this problem, and our analyses are valid only if these data are true and reliable.
(2) **The impacts of the external environment on product sales and evaluation are not considered**, such as the change of commodity demand. According to the microeconomic theory, the demand of life necessities (i.e., hair dryers, microwaves, and pacifiers) seldom changes sharply [8], so we do not take these factors into account.
(3) **The impacts of Amazon's internal system on product sales and evaluation are not considered**, such as the evaluation and supervision system. Due to lack of these data, we neglect these impacts to simplify the modeling process.

3 Nomenclature

We put the symbols that we use in the model and their explanations in Table 1. They are divided into global variables and local variables in Section 5.1.1 by a solid black line.

Table 1: Symbols and explanations.

Symbol	Explanation
t	The tth quarter in one year*
r_j	The star-ratings of the jth review
P_t	The favorable rate of the tth quarter
S^i	The success index of the ith product
e_j	The sentiment value of the jth review
N_t	The number of reviews of the tth quarter
R_t	The average star-ratings of the tth quarter
v_j	The number of the helpfulness votes of the jth review
N_t^W	The average word number of all reviews of the tth quarter
E_t	The average sentiment value of the reviews of the tth quarter
λ_{mn}	The impact degree of the mth evaluation indicator on the nth indicator
N_C	The number of reviews in C
I_i	The importance of the ith word
C	The collection of all the reviews
TF_i	The term frequency of the ith word
c_i	The number of reviews containing the ith word
IDF_i	The inverse document frequency of the ith word
n_i	The number of times that the ith word appears in C

Note*: t = 1 denotes the first quarter in 2010.

4 Data Processing and Analysis

4.1 Data Cleaning

Before we build the model, we first process and remove the following three types of data: (1) Missing and abnormal data. We find that there are less than 1% data with missing or garbled body and date, which may affect structure uniformity. (2) Product data with only one review. These data make it difficult to analyze the impacts of their star-ratings and text on reputation since they lack continuous reviews in the long term. (3) The transaction data before 2010. The transactions of the three products (hair dryers, microwaves, and pacifiers) before 2010 (excluding 2010) are not continuous, that is, no one purchased such products for a long time (more than 3 months). Since they only account for less than 5%, we delete them for simplified analysis.

4.2 Insights of Data

4.2.1 Observation of Product Lifecycle

We define the product lifecycle as the time between the first and last review of the product. Based on this definition, we calculate the lifecycle of the three products, and the results are shown in Figure 2. Clearly, they show different survival patterns. For hair dryers, as time increases, their quantities also increase, which indicates their lifecycle is very long. Similarly, most microwaves have a lifecycle of 3-4 years, and their technology develops at a low speed. Different from hair dryers and microwaves, most pacifiers have a short lifecycle with commonly 3 months to 1 year, and more than 65% of them are eliminated within 2 years. This implies that pacifiers are updated very quickly with plenty of new products appearing in 1 year.

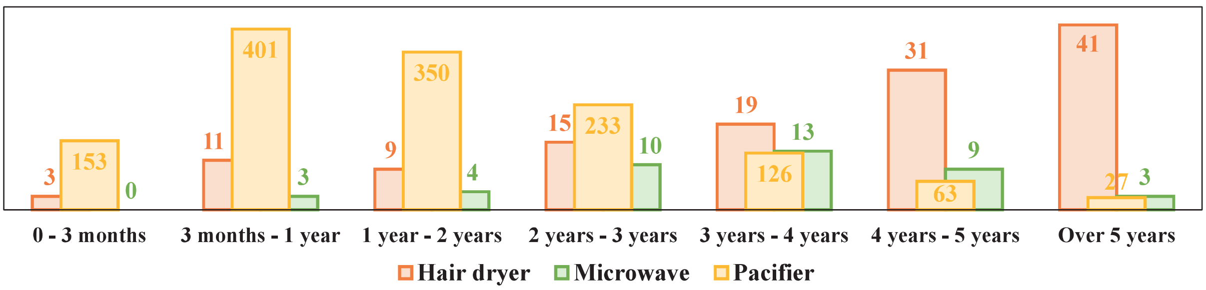

Figure 2: Lifecycle of three products, including hair dryers, microwaves, and pacifiers.

In order to explore the time patterns of evaluation indicators, it is necessary to specify the observation interval. Since most of the three products can survive over one quarter, we set the interval as one quarter. Accordingly, the time unit in the model is also set as one quarter.

4.2.2 Trend of Star-rating-based and Text-based Indicators

In the dataset, the star-ratings and text-based reviews are two main indicators of products, so we calculate the star-ratings and evaluation changes of the three products over time. In addition, we find that reviews with higher helpfulness ratings will be preferentially seen by buyers[2], which indicates that they have great impacts on the sales. Therefore, we count the star-ratings and reviews with helpful votes greater than 0 separately. The results are shown in Figure 3.

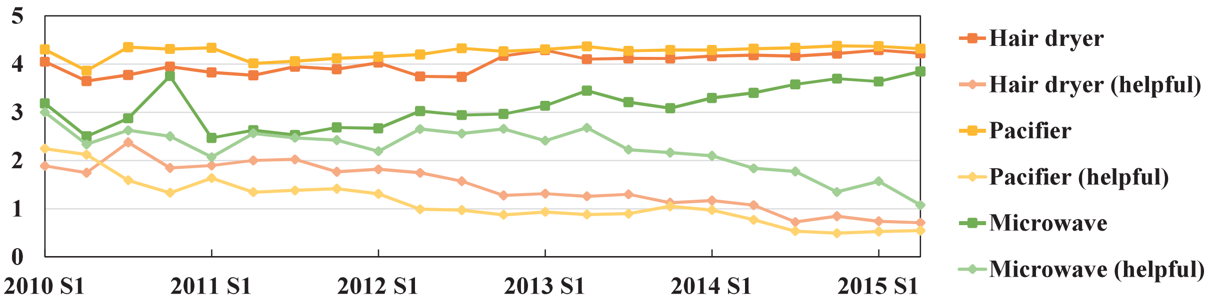

Figure 3: Changes of star-ratings over time, including hair dryers, microwaves, and pacifiers.

When we calculate the star-ratings of all reviews, the average star-ratings of hair dryers and pacifiers are in a relatively stable trend without dramatic changes, while those of microwaves are in an upward trend. However, when we only calculate those with helpful votes, the average star-ratings of three products have a downward trend to varying degrees. Interestingly, for products with high average star-ratings in all reviews, the average star-ratings displayed by helpful reviews are low. Based on this, if we want to predict their future business conditions, we need tofocus on those reviews with more helpfulness ratings, because the higher the product's average star-ratings of a helpful review, the more attractive it is to potential customers.

We define reviews with 4 to 5 star-ratings as favorable reviews, and we calculate the favorable rate of all quarters. As is shown in Figure 4, the favorable rate and average star-ratings show aucorrelation. In other words, the previous reviews will take effect for a long time.

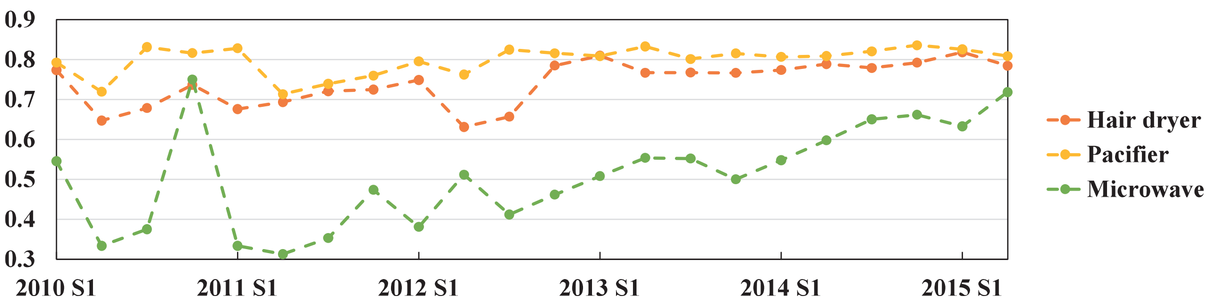

Figure 4: Changes of the favorable rate over time, including hair dryers, microwaves, and pacifiers.

As for text-based reviews, we count the number of reviews and the average text word number of each quarter. As shown in Figure 5, the number of reviews of the three products has an

[2]https://www.amazon.com/

upward trend, while the average text word number has a downward trend. Similarly, there is evident autocorrelation in the number of reviews and the average text word number.

The **sentiment tendency** of reviews is also an important indicator of product sales. Since

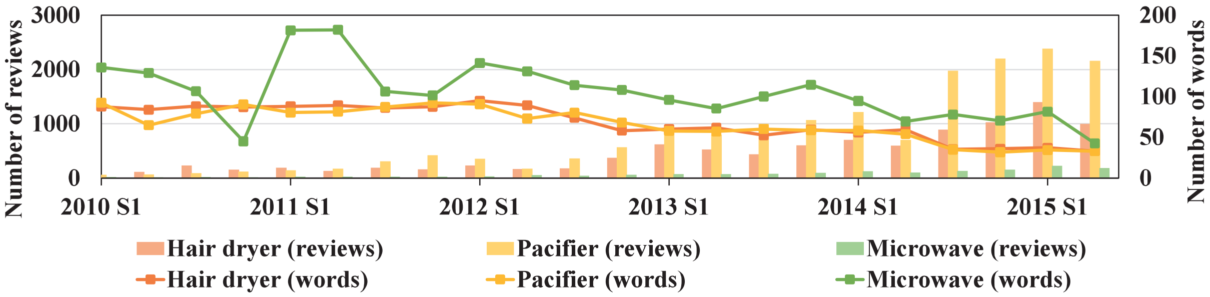

Figure 5: Changes of the number of reviews and average text word number over time, including hair dryers, microwaves, and pacifiers.

we have not built complete analysis model for text sentiment in this section, we define reviews that meet the following two requirements as favorable reviews: (1) words containing positive sentiment; and (2) words without negative sentiment or negators (see Appendix A).

In this way, we can get the percentage of favorable reviews of three products in each quarter, as shown in Figure 6. Since favorable reviews are generally paired with high star-ratings, the percentage of favorable reviews and average star-ratings show similar patterns. The difference lies in that the favorable percentage of helpful reviews for microwaves has remained the same trend as the favorable percentage of all reviews until the third quarter of 2012, but showed the opposite trend subsequently. According to Figure 6, we can draw a similar conclusion that the favorable review percentage has strong autocorrelation.

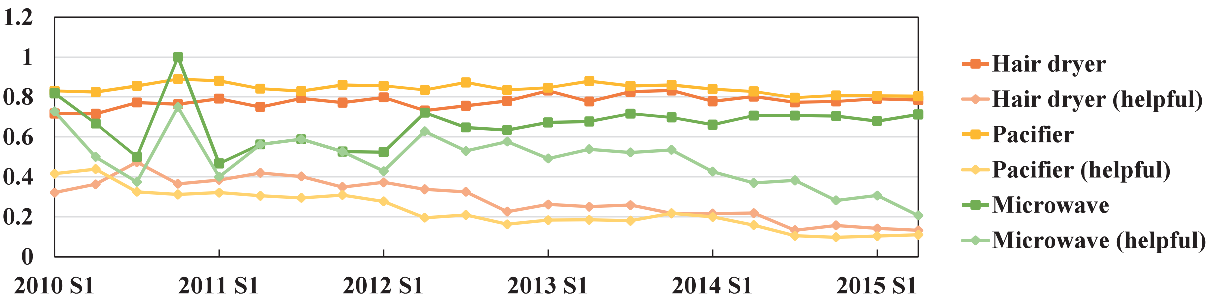

Figure 6: Changes of the percentage of favorable reviews over time, including hair dryers, microwaves, and pacifiers.

4.2.3 Relationship Between Star-rating-based and Text-based Indicators

Next, we observe the relationship between star-rating indicators and text indicators. We visualize the star-ratings and percentage of favorable reviews, as well as the number of reviews and average text word number, as shown in Figure 7 and Figure 8, respectively. Obviously, there is a positive correlation between star-ratings and percentage of favorable reviews—the higher the star-ratings, the higher the percentage. Moreover, the number of reviews has no evident impacts on the percentage of favorable reviews. In addition, the average text word number does not have fixed effects on the percentage of favorable reviews—for microwaves, the average text word number has a significantly negative correlation with it, but this is not obvious for hair dryers and pacifiers.

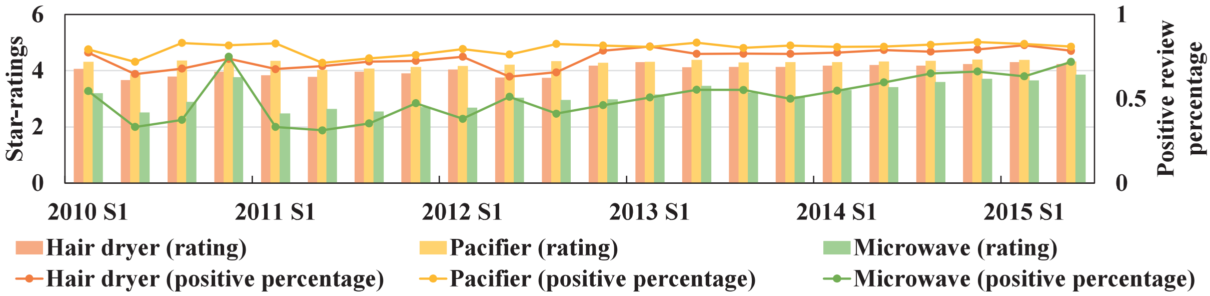

Figure 7: The relationship between star-ratings and percentage of favorable reviews, including hair dryers, microwaves, and pacifiers.

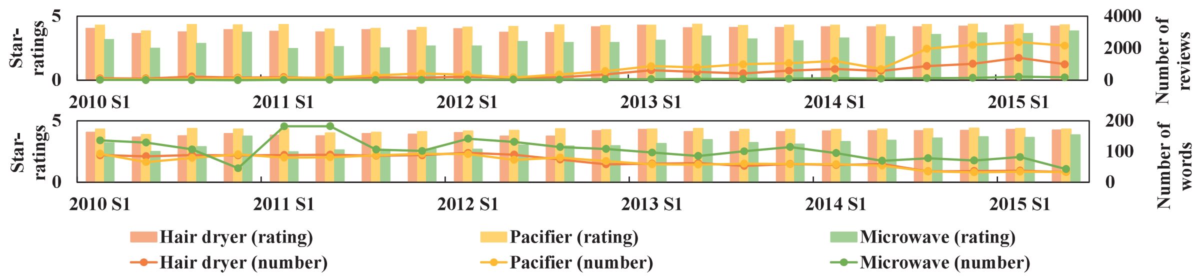

Figure 8: The relationship among star-ratings, number of reviews, and average text word number, including hair dryers, microwaves, and pacifiers.

4.2.4 Determination of Evaluation Indicators Based on Star-ratings and Text

Based on the above analysis, we determine the evaluation indicators based on star-ratings and text. Above all, the average star-ratings of a product in one quarter is an important indicator, but according to Figure 3, if we simply average the star-ratings of all the reviews, it may not be able to find the future impacts of helpful reviews. Therefore, we introduce the number of helpful votes in the average star-ratings as a weight, and consider the impacts of verified purchase and vine, and then we determine the average star-ratings R_t of a product in the tth quarter as

$$R_t = \frac{\sum_{j=1}^{N_t} r_j \min\left\{1 + \frac{v_j}{20}, 5\right\} \left(\frac{1}{2} + \frac{1}{2}\mathbf{I}_{\text{Verified}}\right) \left(1 + \frac{1}{2}\mathbf{I}_{\text{Vine}}\right)}{N_t}, \tag{1}$$

where N_t is the product's number of received reviews in the tth quarter, r_j is the star-ratings of the jth review, v_j is the helpfulness votes of the jth review, and $\mathbf{I}_{\text{Verified}}$ as well as $\mathbf{I}_{\text{Vine}}$ are indicator functions.

We have defined the favorable rate P_t in Section 4.2.2—the percentage of the reviews with star-ratings of more than 4 in all reviews, and its calculation formula is as follows,

$$P_t = \frac{\sum_{j=1}^{N_t} \mathbf{I}_{r_j \geq 4}}{N_t}. \tag{2}$$

As shown in Eq.(2), when $r_j \geq 4$, $\mathbf{I}_{\text{Verified}} = 1$; otherwise, it is 0.

With rating-based indicators determined, we further determine the text-based indicators. To begin with, the number of reviews N_t reflects the popularity of a product. Furthermore, according to Figure 8, the average text word number N_t^W has possible impacts on average star-ratings. Additionally, text sentiment is also significant. Although favorable review percentage can be regarded as a kind of sentiment expression, it cannot reflect all the sentiment—the sentiment intensity of different favorable reviews is not the same, i.e., "*Perfect! I love it!*" is obviously more intense than "*Good.*", which has different impacts on the potential customers. Therefore, we need to get a more specific indicator to express reviews' sentiment intensity. We build this indicator by a text sentiment analysis model. Assume that the review's sentiment value obtained by the subsequent model is e_j, the average sentiment value E_t of the product in the t quarter is

$$E_t = \frac{\sum_{j=1}^{N_t} e_j \min\left\{1 + \frac{v_j}{20}, 5\right\}\left(\frac{1}{2} + \frac{1}{2}\mathbf{I}_{\text{Verified}}\right)\left(1 + \frac{1}{2}\mathbf{I}_{\text{Vine}}\right)}{N_t}. \tag{3}$$

In Eq.(3), the average sentiment value E_t is basically the same as the average star-ratings. Next, we start to build a text sentiment analysis model to calculate the sentiment value of the reviews.

5 Model Construction

5.1 A Text Sentiment Analysis Model Based on Deep Learning

5.1.1 Steps of the Text Sentiment Analysis Model

After collecting text data, we judge the sentiment of the text and its degree as follows.

(1) Segmentation.

Usually, it starts with judging the keywords in the sentence. The sentiment contained in the keywords reflects most information in the whole sentence [7], so we need to segment the sentence into several words. First, we use spaces and punctuation as separators to segment a sentence. Then, we delete the stop words that appear very frequently with no practical meanings, i.e., "a, is", as well as "and"—because they increase the storage space but help little with sentiment judgements. Here, we delete 891 stop words in total, and the stop word list is attached in Appendix B.

(2) Extract keywords by TF-IDF.

Next, we extract the keywords in each sentence. Here, keywords refer to those words appearing frequently in the text with practical meanings, such as "great", "bad", "love", and "hate". After that, we make sentiment judgement on these keywords, and get the sentiment orientation (positive, neutral, and negative) and their degrees.

In this work, we use the TF-IDF algorithm to extract keywords. The main idea of this algorithm is that the importance of a word is directly proportional to its word frequency in the entire text database C, but inversely proportional to the number of reviews containing the word. Therefore, for the ith word, there are two indicators—Term Frequency (TF) and Inverse Document Frequency (IDF), and their calculation formulas are as follows,

$$TF_i = \frac{n_i}{\sum n_i}, \tag{4}$$

$$IDF_i = \ln \frac{N_C}{c_i + 1}, \tag{5}$$

, where n_i denotes the word frequency, N_c denotes the number of all reviews, and c_i denotes the number of reviews containing the word. In Eq.(4), TF_i reflects the proportion of the word frequency in all words—the bigger the TF_i, the more important the word. In Eq.(5), IDF_i reflects the word frequency in all the reviews (we add 1 to the denominator to prevent it from being 0)—the more frequent the word, the more likely that it is a common word in general sentences, i.e., the article (the, a), conjunction (and, so), and preposition (in, about). Based on these two indicators, the importance of the ith word is defined as

$$I_i = TF_i \times IDF_i. \tag{6}$$

As Eq.(6) indicates, the more important the word, the more frequently it appears in the reviews, and the less number of reviews containing it.

With stop words removed, the average word number in a review is about 75. According to [2], the useful information accounts for 20% in a sentence. Therefore, we extract 20% useful information, or rather, the top 15 most important words for each review (if the review contains less than 15 words, all words are used as keywords).

(3) Turn all keywords into word vectors.

There are two ways to judge text sentiment. One is to refer to the sentiment dictionary: since the sentiment value of all keywords is determined, we can get the sentiment value of the keywords in the text by summarizing them. We use this simple and fast method originally to analyze the text sentiment, but it is at a disadvantage in less comprehensive consideration. When sentence expressions are changed without sentiment changes, there are great changes in the sentiment value, resulting in inaccuracy of sentiment value in some sentences, as shown in Table 2.

The other is the deep learning model. With the development of science and technology, deep learning models in text sentiment analysis are gaining increasing popularity. Analysis based on deep learning makes for the lack of context in the sentiment dictionary, and improves the judgement accuracy effectively. It mainly converts keywords into mathematical vectors, inputs them into the neural network, and trains the network.

Here, we adopt deep learning to judge the text sentiment. We convert the extracted keywords into mathematical vectors. First of all, we count the number of keywords and encode them with one-hot encoding: the vector corresponding to each keyword is a column vector, and the number of rows is the number of keywords—the value of the row that the keyword is in equals 1, and the others' are all 0. In this way, all the keywords have their own vectors, as shown in Figure 9.

Table 2: Some problems in judging sentiment value based on a sentiment dictionary.

Original Review	Sentiment Value*	Review	Actual Sentiment Value**
Good	6	Correct, because “good” is a positive word and should have positive value	6
Works good	8.4	Wrong, because “works” cannot add any positive sentiment here	6
Good as expected	0	Wrong, “as expected” does not change any sentiment here so the review should have same value with “Good”	6
Good experience will order again	6	Wrong, “order again” increases the positive sentiment and should have higher value	10

*Based on sentiment directory

**Based on subjective judgment.

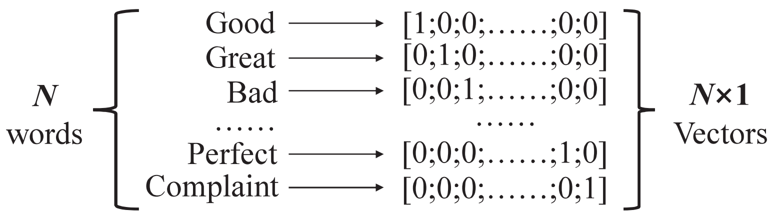

Figure 9: A schematic diagram of one-hot encoding.

However, if there are too many keywords, the length of the vectors will be too long. If we directly put them into the neural network, it will lead to dimensionality curse and long-time training. To prevent this, we employ the Continuous Bag-of-Word (CBOW) model to reduce the dimensions. It puts the one-hot vectors of the neighboring keywords of the ith keyword in the text as independent variables into the neural network (the number of selected neighboring keywords is also called the window). There is only one neuron in the one hidden layer, which is the final converted word vector, and the output layer is the one-hot vector of the ith keyword.

Figure 10 is an example of CBOW. We set the window to 2 after extracting the keywords, that is, the one-hot vectors of the two neighboring keywords of each keyword are put into the neural network as independent variables. If we use “so” as the keyword, the input is “surprisedly”, “good”, and “powerful”, respectively. After that, we train the weight matrix $W_{V \times N}$ to minimize the difference between the result vector of the output layer and the original one-hot vectors of the keyword “so”. Then, we can get a word vector with reduced dimensions by multiplying this matrix with the original one-hot vectors.

(4) Turn word vectors into sentence vectors and train them by the neural network.

After we convert all the keywords into word vectors, we add all these word vectors, and take the average as the vector of this review. Then, we input this review vector into a new neural network model, where the output layer is the final text sentiment value.

Notice that the neural network needs a part of reviews with sentiment values as the training

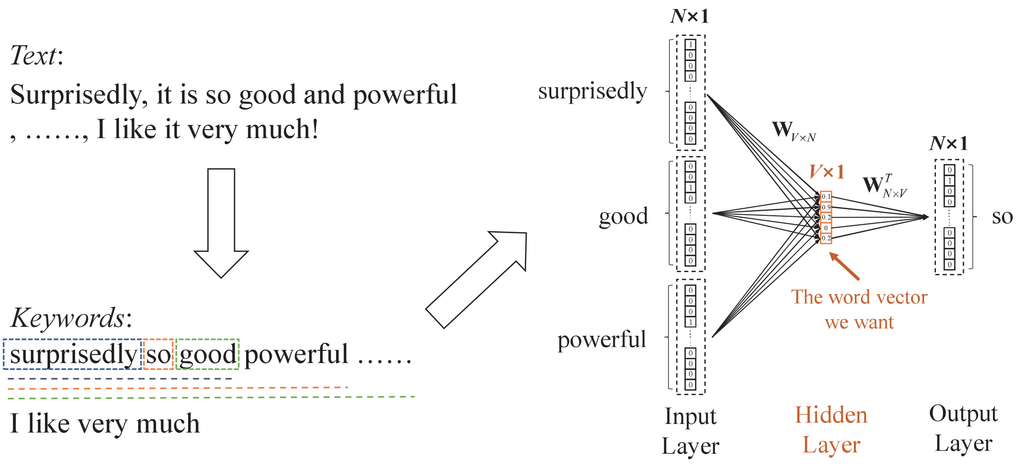

Figure 10: A schematic diagram of CBOW.

samples. Therefore, we first use the sentiment dictionary (SentiWordNet[3]) to calculate sentiment value of the top 200 reviews of the three products, and then manually adjust the problematic ones. After that, we put these reviews and sentiment words into the neural network for training. Figure 11 is the whole framework of the text sentiment analysis model based on deep learning.

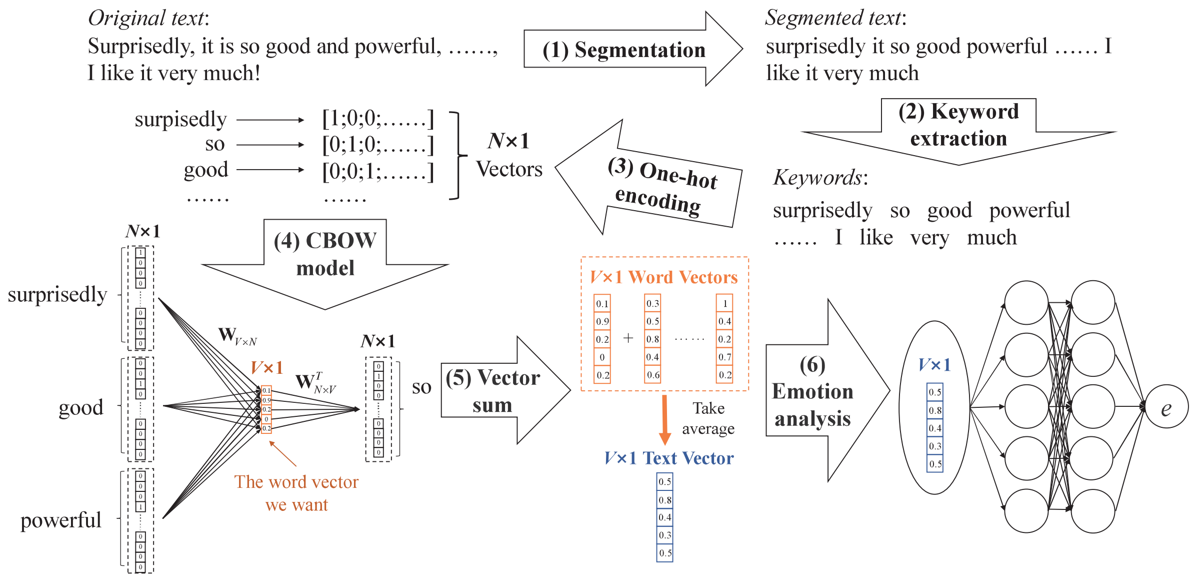

Figure 11: The framework of the text sentiment analysis model based on deep learning.

5.1.2 Analysis of Model Results

The trained deep learning model works well, and the mean absolute percentage error (MAPE) after fitting on the training set is 6.34%. In this way, we get all the sentiment value of the text after applying the deep learning model to analyze the text of the non-training set.

[3]https://github.com/aesuli/sentiwordnet

Next, we focus on the relationship between review sentiment and star-ratings. Figure 12 shows the average sentiment value, favorable rate, and negative rate (the percentage of the reviews with star-ratings less than 2 in all the reviews in the quarter) of the three products in each quarter. To further explore the impacts of star-ratings on sentiment value, we also examine the average sentiment value of positive and negative reviews. Please note that the sentiment of positive reviews are greater than 0; otherwise, it is less than 0.

According to Figure 12, during 2010 and 2012, the average sentiment value of all the reviews and the average sentiment value of positive reviews were generally positively related to the favorable rate. Nevertheless, after 2012, the relationship between the favorable rate and the average sentiment value was not obvious.

In terms of the negative rate, the relationship between the negative rate and the average sentiment value of negative reviews are obvious in the three products—the higher the negative rate, the smaller the average negative sentiment value, and this pattern persisted during 2010 and 2015. It indicates that when customers see a series of low star-rating reviews, they are more likely to make more negative reviews, which is likely to persist for a long time in the future.

We also observe the number of high and low star-ratings in positive and negative reviews,

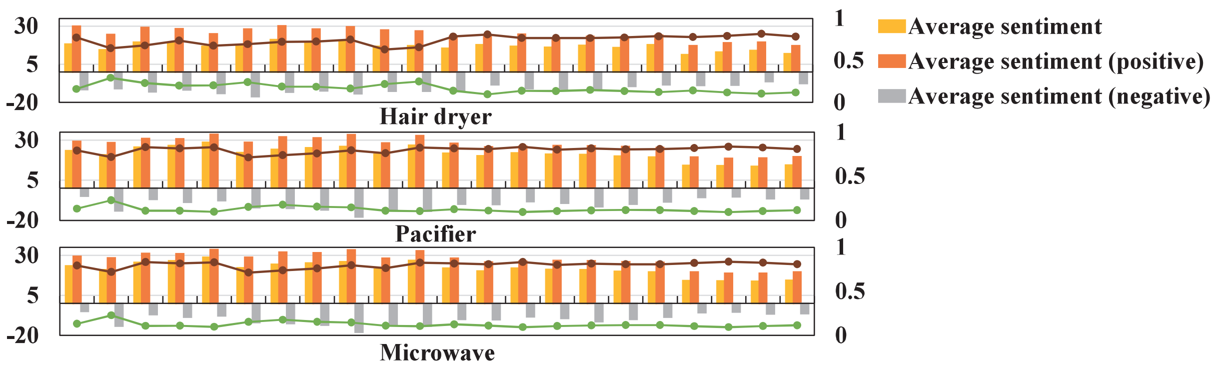

Figure 12: The relationship among the average sentiment value, favorable rate, and negative rate, including hair dryers, microwaves, and pacifiers.

as shown in Figure 13 and 14, respectively. For all the three products, high star-ratings account for the majority of positive reviews.

Nevertheless, it is hard to relate negative reviews with specific star-ratings. Since the third quarter of 2010, there have been high star-ratings in negative reviews of hair dryers and pacifiers, and the proportion of high and low star-ratings in each quarter was roughly the same. For microwaves, although low star-ratings still accounted for the majority, there was an upward trend for the high star-rating reviews.

Last but not least, we observe the keywords extracted by TF-IDF, with the aim to provide suggestions on product design for Sunshine Company. We select the top 200 notional keywords (i.e., noun, verb, adjective, etc.) in the reviews of hair dryers, pacifiers, and microwaves. As shown in Figure 15, the bigger the word, the more important it is to the product.

In Figure 15, the most important keywords are emotional words (i.e., great, love, well, and good), followed by descriptive adjectives (i.e., hot, cool, new, clean, and small). Coming last

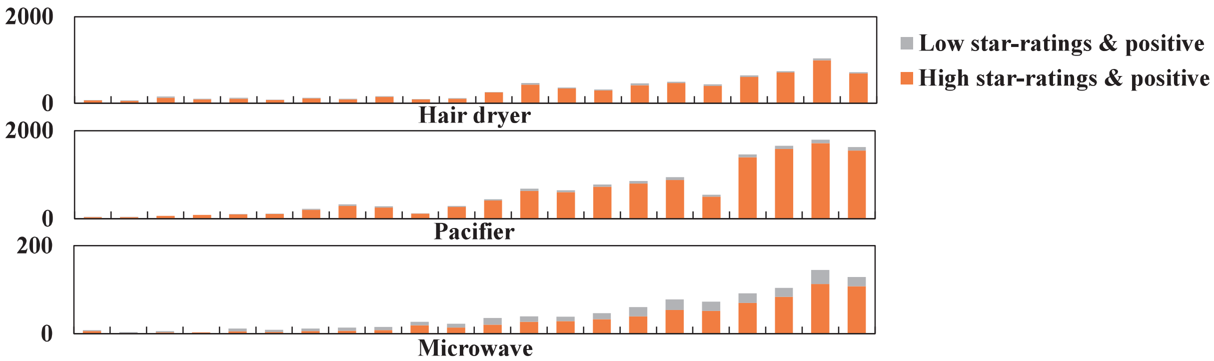

Figure 13: The number of high and low star-ratings in positive reviews, including hair dryers, microwaves, and pacifiers.

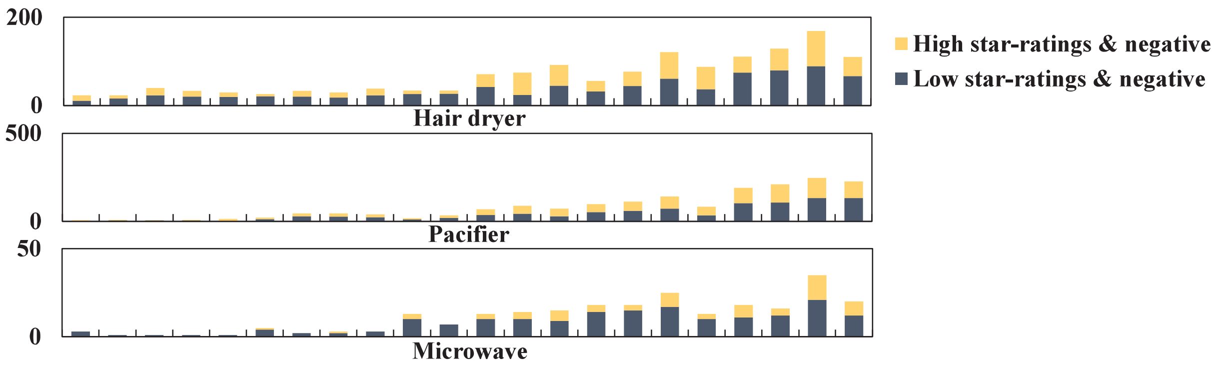

Figure 14: The number of high and low star-ratings in negative reviews, including hair dryers, pacifiers, and microwaves.

are feature words of products (i.e., heat, cord, nipple, suck, cooking, and fix). The sequence of importance indicates that most people express their feelings after using the product, followed by the product description. This provides design ideas for Sunshine Company. For hair dryers, words like hot, cool, heat, light, small, heavy, cord, settings, and button are more important, so the company should design high-power, light, and portable products. For pacifiers, cute, soft, such, small, new, cat, size, clean, and hard are frequent, so it should focus on designing products with cute apparency, soft texture, and clean material. As for microwaves, the most important

Figure 15: Keywords in reviews, including hair dryers, pacifiers, and microwaves in sequence.

keywords are *small, cooking, fix, big, works, repair, install, size, and fix*, so the design should be focused on easy installation, medium size, and easy repair.

5.2 A Difference Equation Prediction Model

5.2.1 Correlation Analysis of Evaluation Indicators

In Section 4.2.4, we define five evaluation indicators based on star-ratings and text, namely the average star-ratings R_t, favorable rate P_t, number of reviews N_t, average text word number N_t^W, and average review sentiment value E_t. To explore their changing patterns, it is necessary to obtain their influencing factors—whether the change of one indicator will be affected by itself or the others. To address this problem, we make the correlation analysis.

We calculate the autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) of the indicators by

$$ACF = \frac{E\left[(X_t - EX_t)(X_{t-k} - EX_t)\right]}{E\left(X_t - EX_t\right)^2}, \tag{7}$$

$$PACF = \frac{E\left[(X_t - EX_t)(X_{t-k} - EX_{t-k})\right]}{E\left(X_{t-k} - EX_{t-k}\right)^2}, \tag{8}$$

where k denotes the lags, X_t denotes the indicator value of the tth quarter. Table 3 shows the ACF and PACF results of the five indicators of hair dryers. Due to space limitations, the results of microwaves and pacifiers are shown in the Appendix C.

Table 3: The ACF and PACF of Hair dryers' indicators.

	R_t		P_t		N_t		N_t^W		E_t	
Lags	ACF	PACF	ACF	PACF	ACF	PACF	ACF	PACF	ACF	PACF
1	**0.325**	**0.325**	**0.499**	**0.499**	**0.816**	**0.816**	**0.838**	**0.838**	**0.481**	**0.481**
3	0.091	0.117	0.264	0.331	0.442	0.151	0.539	-0.127	0.342	0.070
5	0.058	0.076	0.251	0.218	0.259	-0.063	0.309	0.116	0.159	0.042
7	-0.247	-0.296	-0.002	0.073	0.091	0.076	0.145	0.025	0.092	0.044
9	-0.122	-0.271	-0.068	-0.367	-0.096	-0.235	-0.134	-0.178	-0.023	-0.010

If the first-order ACF of the indicator is large and trailing, and the first-order PACF is truncated, it indicates that the indicator has a strong first-order autocorrelation. The results indicate that N_t shows obvious autocorrelation in all the three products, P_t and N_t^W show autocorrelation in pacifiers and microwaves, but R_t and E_t show autocorrelation only in hair dryers. Apart from the average star-ratings and average sentiment value, the conclusion mentioned in Section 4.2.2 that the other three indicators have autocorrelation has been verified. After setting different weights for helpfulness ratings, verified purchase, and vine, the average star-ratings and sentiment value remain basically stable in each quarter. This indicates that these two indicators may be mainly affected by other indicators, rather than themselves.

After that, we observe the correlation between the variables. We make the correlation analysis on the indicators of the $(t-1)$th and tth quarter. We calculate the Pearson correlation coefficient for the five indicators in pairs by

$$\rho_{XY} = \frac{E\left[(X_t - EX_t)(Y_{t-1} - EY_{t-1})\right]}{\sigma_X \sigma_Y}, \tag{9}$$

where σ_X and σ_Y denote the standard deviations of the indicator X and Y, respectively.

The specific values of the Pearson correlation coefficient of hair dryer indicators are shown in Table 4 (see Appendix D for those of microwaves and pacifiers). According to the principle of correlation judgment, we identify the combinations of variables with correlations, including (R_{t-1}, E_t), (P_{t-1}, R_t), (N_{t-1}, R_t), (N_{t-1}, E_t), (N^W_{t-1}, R_t), and (E_{t-1}, R_t).

Table 4: Pearson correlation coefficient of the hair dryer indicators.

	R_t	P_t	N_t	N^W_t	E_t
R_{t-1}	-	-0.221	0.046	-0.198	**-0.461***
P_{t-1}	**-0.770****	-	-0.044	-0.121	-0.204
N_{t-1}	**-0.437***	-0.056	-	-0.150	**-0.418***
N^W_{t-1}	**0.455***	0.132	0.001	-	-0.012
E_{t-1}	**-0.469***	-0.173	-0.054	-0.062	-

$**$. Significant at $\alpha = 0.01$

$*$. Significant at $\alpha = 0.05$

5.2.2 Construction of Difference Equation Model

After determining each indicator's autocorrelation and the correlation with other variables, we establish a difference equation as follows,

$$\begin{cases} \Delta R_t = \alpha_1 + \lambda_{11} R_{t-1} + \lambda_{12} P_{t-1} + \lambda_{13} N_{t-1} + \lambda_{14} N^W_{t-1} + \lambda_{15} E_{t-1} \\ \Delta P_t = \alpha_2 + \lambda_{21} R_{t-1} + \lambda_{22} P_{t-1} + \lambda_{23} N_{t-1} + \lambda_{24} N^W_{t-1} + \lambda_{25} E_{t-1} \\ \Delta N_t = \alpha_3 + \lambda_{31} R_{t-1} + \lambda_{32} P_{t-1} + \lambda_{33} N_{t-1} + \lambda_{34} N^W_{t-1} + \lambda_{35} E_{t-1} \\ \Delta N^W_t = \alpha_4 + \lambda_{41} R_{t-1} + \lambda_{42} P_{t-1} + \lambda_{43} N_{t-1} + \lambda_{44} N^W_{t-1} + \lambda_{45} E_{t-1} \\ \Delta E_t = \alpha_5 + \lambda_{51} R_{t-1} + \lambda_{52} P_{t-1} + \lambda_{53} N_{t-1} + \lambda_{54} N^W_{t-1} + \lambda_{55} E_{t-1} \end{cases}. \tag{10}$$

According to the correlation analysis, if there is no autocorrelation in the mth indicator, then $\lambda_{mm} = 0$ in the corresponding difference equation; and if there is no correlation between the mth variable and the nth variable, then $\lambda_{mn} = \lambda_{nm} = 0$.

5.2.3 Results and Analysis of Parameter Fitting

We put the indicators of the three products in each quarter into the model for fitting, and get the parameters in Eq.(10). We take hair dryers for an example, and the results are shown in Table 5. Please see the Appendix E for the parameter fitting results of pacifier and microwave.

By analyzing the parameter fitting results, we find that the increase in star-ratings is negatively correlated to the star-ratings and sentiment value in the previous quarter. Similarly, the increase in the favorable rate, number of reviews, the sentiment value are also negatively correlated to the indicators of the previous quarter. This shows that the large number of high

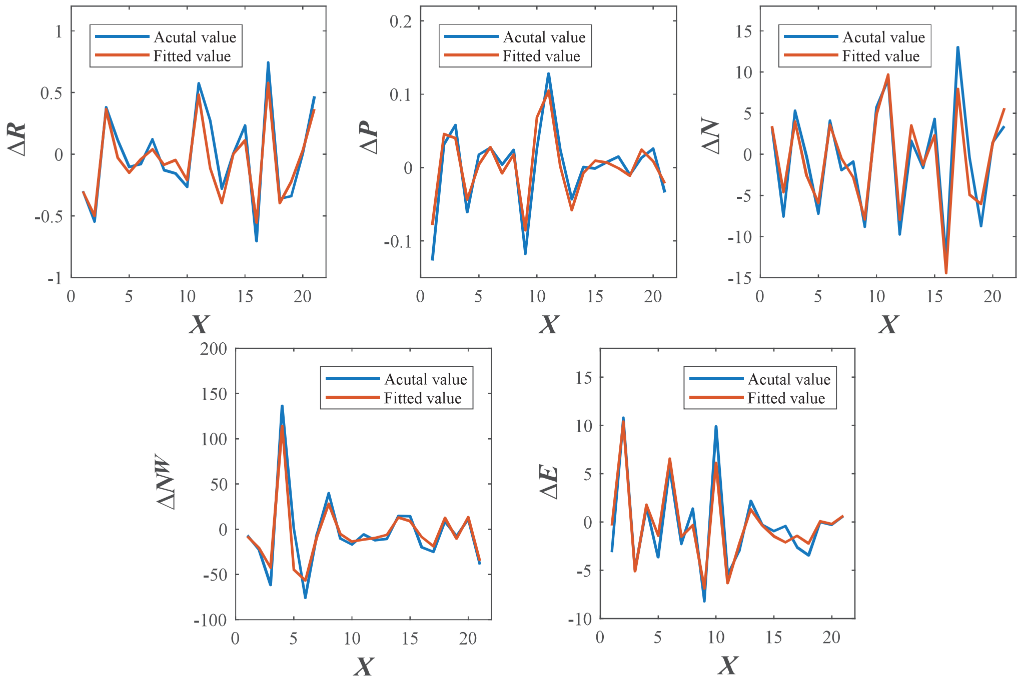

Figure 16: Results of parameter fitting of hair dryers, including ΔR, ΔP, ΔN, ΔN^W, and ΔE.

Table 5: Parameter fitting results of the difference equation model for hair dryers.

m	α_m	λ_{mn}				
		n				
		1	2	3	4	5
1	3.38	-1.05	0	0	0	-0.02
2	1.06	-0.29	-0.16	0	0	0
3	958.81	-75.21	0	-0.29	0	-38.99
4	-50.94	15.85	0	0	-0.03	0
5	17.8	-4.66	0	0	0	-0.22

star-ratings or positive reviews in one quarter cannot increase the performance of indicators in the next quarter, and may even bring negative growth. To explore its reasons, we find that there have been related reports[4] reflecting the phenomenon of Amazon's click farming[5]. The eye-catching favorable reviews highlighted on the shopping platforms might instead give customers the feeling of "intentionally clicking farming for sales", which reduces the customers' desire to buy or affect subsequent product reviews [12]. This is consistent with the obseravtion in Section 4.2.2—the average star-ratings are always high without considering the helpfulness, but they decrease after taking the helpfulness into account. Consistently favorable reviews bring negative reviews in the future, until they reach a level that can actually reflect the product conditions.

[4]https://www.wsj.com/articles/how-sellers-trick-amazon-to-boost-sales-1532750493

[5]Click farm is an organized group of low-paid workers employed to click on particular parts of web pages, especially approval buttons in social media as a way of making businesses seem popular.

5.3 A Principal Component Analysis Model

5.3.1 Principle Introduction

Principal component analysis (PCA) is a multivariate statistical method often used for reducing dimensionality. It transforms a group of variables with possible correlation into linearly unrelated ones through orthogonal transformation, and the converted variables are called principal components [10]. It believes that the amount of information contained in a variable is usually measured by the variance or the sum of squared deviations, and the number of principal components is selected according to the variance contribution rate.

5.3.2 Model Results and Construction of the Success Index

Here, we employ PCA to construct the success index S^i. Firstly, we standardize the data by

$$X_1 = \frac{R - \mu_R}{\sigma_R}, X_2 = \frac{P - \mu_P}{\sigma_P}, X_3 = \frac{N - \mu_N}{\sigma_N}, X_4 = \frac{N^w - \mu_{N^W}}{\sigma_{N^W}}, X_5 = \frac{E - \mu_E}{\sigma_E}, \tag{11}$$

where μ is the mean, σ is the standard deviation, and X is the standardized data. More specifically, X_1 is the average star-ratings, X_2 is the favorable rate, X_3 is the sentiment value, X_4 is the word number of the review, and X_5 is the number of reviews.

$$\begin{cases} Z_1 = c_{11}X_1 + c_{12}X_2 + \ldots + c_{1p}X_P \\ Z_2 = c_{21}X_1 + c_{22}X_2 + \ldots + c_{2p}X_P \\ \ldots \\ Z_P = c_{p1}X_1 + c_{p2}X_2 + \ldots + c_{pp}X_P \end{cases} \tag{12}$$

In Eq.(12), for each i, (1)$c_{i1}^2 + c_{i2}^2 + \ldots + c_{ip}^2 = 1$, and $[c_{11}, c_{12}, \ldots, c_{1p}]$ maximizes the variance of Z_1; (2) $[c_{21}, c_{22}, \ldots, c_{2p}]$ is orthogonal to $[c_{11}, c_{12}, \ldots, c_{1p}]$, and it maximizes the variance of Z_2; and (3) $[c_{31}, c_{32}, \ldots, c_{3p}]$ is orthogonal to $[c_{11}, c_{12}, \ldots, c_{1p}]$ and $[c_{21}, c_{22}, \ldots, c_{2p}]$, and it maximizes the variance of Z_3. Similarly, we can get all the p and principal components.

Generally, we select the cumulative variance contribution rate of the first few principal components as the analyzed principal components. In other words, most information can be described by these comprehensive evaluation indicators. Here, we select the principal components whose cumulative variance contribution reaches 90%. The PCA results of hair dryers are shown in Table 6.

Table 6: Results of Principal component analysis of hair dryers.

No.	Characteristic root	Variance contribution rate	Cumulative variance contribution rate
1	2.05	40.92%	40.92%
2	1.20	23.93%	64.85%
3	0.93	18.51%	83.35%
4	0.51	10.21%	93.56%
5	0.32	6.44%	100%

$$\begin{cases} Z_1 = 0.6251X_1 + 0.5088X_2 + 0.5584X_3 + 0.0731X_4 + 0.1826X_5 \\ Z_2 = -0.0678X_1 - 0.3594X_2 + 0.2317X_3 - 0.5229X_4 + 0.7343X_5 \\ Z_3 = -0.0118X_1 - 0.3352X_2 + 0.0733X_3 + 0.8438X_4 + 0.4126X_5 \\ Z_4 = 0.0472X_1 + 0.5208X_2 - 0.6928X_3 + 0.0253X_4 + 0.4959X_5 \end{cases} . \tag{13}$$

We can calculate the success index of hair dryers S^1 by

$$S^1 = 0.4092Z_1 + 0.2393Z_2 + 0.1851Z_3 + 0.1021Z_4, \tag{14}$$

and then we put the value of Z into Eq.(14) and get

$$S^1 = 0.2422X_1 + 0.1133X_2 + 0.2267X_3 + 0.0635X_4 + 0.3774X_5. \tag{15}$$

Then we destandardize X and get

$$S^1 = 0.0962R + 0.3032P + 0.0062E + 0.0085N + 0.0064N^W - 1.3199. \tag{16}$$

Similarly, we can obtain the success index of microwaves and pacifiers as follows,

$$S^2 = 0.0727R + 0.3158P + 0.0042E + 0.0244N + 0.0260N^W - 0.9226, \tag{17}$$

$$S^3 = 0.0962R + 0.1197P + 0.0069E + 0.0548N + 0.0061N^W - 1.1190. \tag{18}$$

Here, Eq.(16)(17)(18) represents the influence of main information on the success index. The specific PCA results of hair dryers and pacifiers are attached in Appendix F.

5.3.3 Effectiveness and Application of the Success Index

To verify the effectiveness of the success index, we calculate the average success index of hair dryers with the top 50% and bottom 50% of the number of reviews in each quarter since 2014, as shown in Figure 17. We set 2014 as the starting point because these representative products sold in 2014 continued to be sold afterwards without being eliminated.

As Figure 17 indicates, the success index is a good indicator of the difference between

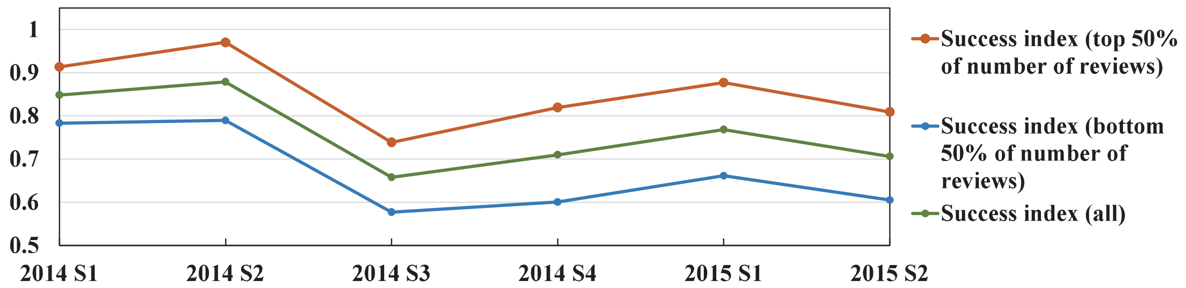

Figure 17: Comparison of the success index of top 50% and bottom 50% of the number of reviews in each quarter during 2014 and 2015.

successful and unsuccessful products. The sales volume with the top 50% number of reviews ranks the top, and its success index is always above the average during 2014 and 2015. Instead, the success index of products with less sales volume is generally smaller than the average.

In brief, the success index does effectively reflect a product's operating conditions, future reputation and success possibility. For Sunshine Company, it can calculate the expected success index of its products for this quarter through the difference equation model. If the index is greater than the average, the company develops well with its current strategies. Otherwise, it should change the sales plan to improve reviews of this quarter, such as lowering the price and increasing the yield rate, to raise the actual success index of this quarter above the average.

6 Sensitivity Analysis

In the text sentiment analysis, we employ a text sentiment analysis model based on deep learning. We obtain the training set of the model by manually adjusting the sentiment value of the top 200 reviews of various products based on the sentiment dictionary. To verify the stability of the deep learning algorithm, we randomly select 5% and 10% text of the training set without manual correction, and observe how this affects the final overall sentiment value, respectively.

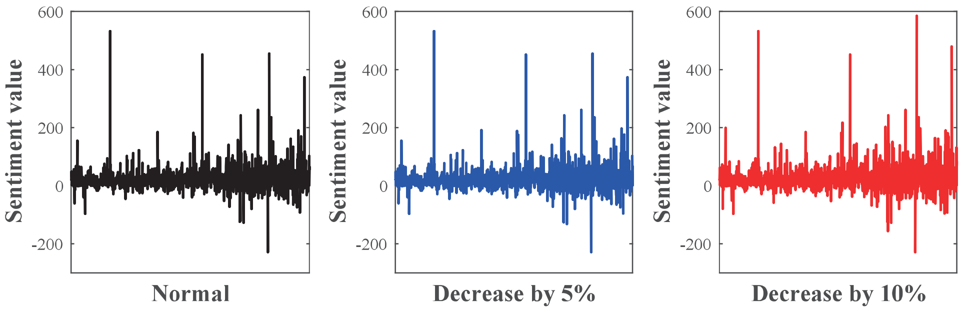

Figure 18: The sentiment value of pacifiers, including normal text, reduced manually-corrected text by 5% and 10%, respectively.

Table 7: Changes of sentiment value for three products with 5% and 10% reduced text*.

	Hair dryer	Pacifier	Microwave
Reduced manually-corrected text by 5%	0.72%	0.80%	1.10%
Reduced manually-corrected text by 10%	1.15%	1.29%	1.33%

*The percentage change in sentiment score.

According to the results shown in Figure 18 and Table 7, reducing manually-corrected text by 5% and 10% only exert little effect. This verifies the robustness and adaptability of our deep learning model, and the generated sentiment value is accurate and reliable.

In addition, we analyze the effect of the number of data on each coefficient of the product's success index. We use data for 22 quarters to calculate the success index, so we reduce the data by 2 and 4 quarters to explore the impacts of the number of data on the coefficient of the index. We take hair dryers for an example, and as Table 18 indicates, the number of data have no significant impacts on the PCA model since the coefficients are not obviously changed.

Table 8: Changes of coefficient, including data for 22 quarters, 20 quarters, and 18 quarters.

	R*	P*	N*	N_w*	E*	Constant
Data for 22 quarters	0.0962	0.3032	0.0085	0.0064	0.0062	-1.3199
Data for 20 quarters	0.0964	0.3001	0.0086	0.0062	0.0061	-1.2999
Data for 18 quarters	0.0960	0.3049	0.0086	0.0063	0.0063	-1.3156
Average change	0.21%	1.58%	1.18%	1.56%	1.61%	1.22%

*The coefficient of the parameter.

7 Model Evaluation

7.1 Strengths

(1) **Improve sentiment analysis accuracy by advanced deep learning model**. In text sentiment analysis, compared with traditional sentiment dictionary models, we employ a more advanced deep learning model to overcome the shortcoming of less comprehensive consideration and get more accurate sentiment analysis results (see Section 5.1).

(2) **Explore the time-based measures and patterns in the review data which are consistent with the real data**. In our difference equation model, the results obtained in the correlation analysis and parameter fitting are consistent with the statistical analysis of reviews, which indicates that our model successfully explores the decisive measures contained in the reviews and their time-based changing patterns (see Section 5.2).

(3) **Provide a simple, effective, and practical tracking system for companies.** By PCA, we get the equation for calculating the success index of the three products. For Sunshine Company, after calculating the review sentiment value by the trained neural network, it can calculate the success index of its products and compare them with other companies' products to make accurate business plans in time without tedious calculations or procedures (see Section 5.3).

7.2 Weaknesses

(1) The training of neural network requires longer time and more data sets. The training of neural networks in text sentiment analysis model takes a long time. In addition, to improve the accuracy, we need to manually fix the training set. However, once the network is trained, it can be used for a long time, so the time cost is relatively small.

(2) Our discussions are confined to 2015, but the evaluation indicator model may change after that. We give the conclusions based on data for a given time period. However, the patterns might change in the future, so it will be better to use the latest data.

8 Conclusions and Future Work

In this work, we build a simple and convenient business tracking system and provide insights on sales strategies on new product launch for Sunshine Company.

To begin with, we perform statistical analysis on the lifecycle and review indicators of hair dryers, pacifiers, and microwaves, and set five star-rating-based and text-based evaluation indicators, namely the average star-ratings, favorable rate, number of reviews, average word number of reviews, and average review sentiment value. Here, we set different weights for helpfulness, verified purchase, and vine in these indicators, respectively. After that, we build a text sentiment analysis model based on deep learning, extract review keywords by TF-IDF, reduce word vector dimension by CBOW, train the word vectors by BP neural network, and get the sentiment value of reviews. Later, we explore the interaction between review sentiment and star-ratings, and provide suggestions on product design based on the ranking of keyword importance. Then, we verify the autocorrelation and correlation between indicators by ACF, PACF and Pearson correlation coefficient, select the variables with strong correlations to establish a difference equation model, and discuss the changing patterns of the indicator increment with itself and other indicators. Finally, we design a success index equation by PCA, and verify its effectiveness, which provides a practical tracking system for success or failure.

Our results indicate that there is incomplete correlation between star-ratings and review sentiment—low star-ratings often lead to negative reviews, but high star-ratings have little effect on review sentiment; and favorable reviews are usually with high star-ratings, but negative reviews are not apparently correlated with star-ratings. Moreover, emotional words are most common in reviews, followed by descriptive adjectives, and feature words of products. Additionally, star-ratings, favorable rate, word number, and sentiment value all decline over time to varying degrees with negative autocorrelation, and the degree of interaction between them varies. Taking the hair dryer for an example, we find that high review sentiment value will reduce the star-rating improvement in the next quarter, and higher star-ratings as well as favorable rate show similar patterns in reducing the review sentiment value in the next quarter. On the contrary, the increment of number of words in reviews is positively related to star-ratings. In brief, the construction of the success index effectively reflects the operating conditions and future reputation of products, which can provide guidance for companies on sales strategies.

In a nutshell, our model is accurate in sentiment analysis, consistent with the reality, as well as effective in tracking future operating conditions. But due to time limitation, there still exists some imperfection in our model. In the future, we can do the following jobs for improvement.

(1) Collect the latest data. In order the overcome the second weakness mentioned in Section 7.2, we need to collect updated data, such as review data during 2016 and 2019, to re-apply our analysis and model to get the latest time-based measures and patterns.

(2) Collect more representative reviews and make more accurate judgments about the review sentiment value. We obtain the training set by manually adjusting the top 200 reviews of various products based on the sentiment dictionary. To the model's sentiment analysis ability, we can put newly-collected reviews with higher helpfulness into the training set, and make more accurate judgments on the sentiment values through questionnaires, experts, and so on.

(3) Introduce more review features to find more convincing indicators. The provided data contain most of the review features, but we can introduce more review features, such as the number of pictures in reviews, and the relevance between these pictures to the product. Also, we can consider reviewers' number of published reviews and received helpfulness votes.

9 A Letter to the Marketing Director of Sunshine Company

To: The marketing director of Sunshine Company
From: Team # 2007707 of 2020 MCM
Date: March 9, 2020
Subject: Findings on Sales and Design Strategies of New Products

---

Dear Sir or Madam:

It is our great honor to be employed as your sales consultants to provide sales and design strategies for your new products (the microwave oven, baby pacifier, and hair dryer). Based on the star-ratings and reviews in the provided data files, we build mathematical models to identify their changing patterns over time, and devise a simply, convenient, and practical business tracking system for you. Our approaches, findings, and suggestions are as follows.

First, we conduct statistical analysis of the review data, including the star-ratings, favorable rate, number of reviews, number of words in reviews, and number of favorable reviews of the three products in each quarter. To accurately analyze the sentiment degree expressed in the text-based reviews, we use advanced models to extract their keywords, and calculate the sentiment value. Also, we build a model to figure out the changing patterns of an indicator while changing time and other indicators. Finally, we put forward the success index to track the future product operating conditions, and verify its effectiveness.

Here are some findings based on our results.

- During the past few years, the number of reviews and their contained number of words were constantly increasing and decreasing, respectively. As for star-ratings, favorable rate, and positive sentiment value, they ascended initially and reached the plateau afterwards. However, if we only focus on reviews with helpfulness votes, the overall reputation of the three products were continually declining.

- The interaction degree between star-ratings and reviews is limited. Customers will make more negative reviews after they read a series of reviews with low star-ratings, but high star-ratings cannot increase their positive sentiment. Also, favorable reviews usually correspond to high star-ratings, but negative reviews are not associated with a specific star-rating level.

- Among the review keywords of the three products, emotional words rank the first, followed by descriptive adjectives and features words of products. In other words, buyers prefer to express their feelings after using the product in the reviews, and then describe the product.

- Among the keywords of hair dryers, words related to power, weight, and portability are most important. As for pacifiers, customers focus on their shape, material and safety. In terms of microwaves, customers pay more attention to their installation, size, and after-sales service.

- If the star-ratings, favorable rate, or review sentiment value in a quarter are too high, they are very likely to decrease sharply in the next quarter. Also, the higher they are in this quarter, the more sharply they will decrease in the next quarter. Nevertheless, the higher the star-ratings, the less the number of words in reviews will reduce in the next quarter. Moreover, the number of reviews is not related to other indicators.

- The success index we define effectively distinguishes between the top 50% and the bottom 50% products. It is an valid business evaluation indicator for companies.

Based on these findings, we put forward some sales and design strategies as follows.

- Do not pursue 100% high star-ratings blindly, because this will not help with future reputation. Instead, they might leave a bad impression of "deliberately clicking farming" on customers, or give them too high product expectations. You only need to ensure that the average star-ratings of the product maintain at a normal level, and the reputation of the product will rank the top among other competing products in the future.
- Ensure that the yield rate of each quarter is basically the same without a reduction in the product quality in a certain quarter, because a series of low star-ratings will result in increase in negative sentiment and decrease in purchase desire.
- Since people pay more attention to the reviews with more helpful votes published by reviewers with vine, you had better increase their star-ratings and positiveness. You can keep an eye on these trustworthy and helpful reviews of other competing products and improve them on the issues mentioned in those reviews.
- As for pacifiers, you can focus on the cuteness of the shape, i.e., some animals such as cats and elephants are popular. Moreover, you should ensure that the pacifiers are made of soft materials, and their production process is hygienic and safe, because the buyers are usually parents, and they pay more attention to whether the pacifier is suitable and safe for baby teething.
- In terms of microwaves, since they update at a low speed, you can concentrate on other aspects, i.e., installation convenience, and suitable size. Additionally, customers particularly want microwaves to be repaired in time, so it would be a good way for you to invest enough energy in microwave after-sales service.
- Finally, you can track the success index of your products and other competing products according to the equation provided in Sectioin 5.3.2 to ensure that the success index of your company's product is above the average. Once the success index is below the average, it is necessary for you to adjust the sales strategies in time, such as the low-price strategy.

We hope our suggestions are helpful. If you have any question, please feel free to contact us.

Sincerely,
Team # 2007707 of 2020 MCM

Appendices

Appendix A Judgement Words for Positive Reviews

(1) Included: great, good, excellent, perfect, okay, nice, well, ok, awesome, love, powerful, fantastic, like, best, quality, surprised, safe, enjoy, worth, and strong.

(2) Not included: no, not, bad, terrible, and hate.

Appendix B The Stop Word List

’d	’ll	’m	’re	’s	’t
’ve	ZT	ZZ	a	a’s	able
about	above	abst	accordance	according	accordingly
across	act	actually	added	adj	adopted
affected	affecting	affects	after	afterwards	again
against	ah	ain’t	all	allow	allows
almost	alone	along	already	also	although
always	am	among	amongst	an	and
announce	another	any	anybody	anyhow	anymore
anyone	anything	anyway	anyways	anywhere	apart
apparently	appear	appreciate	appropriate	approximately	are
area	areas	aren	aren’t	arent	arise
around	as	aside	ask	asked	asking
asks	associated	at	auth	available	away
awfully	b	back	backed	backing	backs
be	became	because	become	becomes	becoming
been	before	beforehand	began	begin	beginning
beginnings	begins	behind	being	beings	believe
below	beside	besides	best	better	between
beyond	big	biol	both	brief	briefly
but	by	c	c’mon	c’s	ca
came	can	can’t	cannot	cant	case
cases	cause	causes	certain	certainly	changes
clear	clearly	co	com	come	comes
concerning	consequently	consider	considering	contain	containing
contains	corresponding	could	couldn’t	couldnt	course
currently	d	date	definitely	describe	described
despite	did	didn’t	differ	different	differently
discuss	do	does	doesn’t	doing	don’t
done	down	downed	downing	downs	downwards
due	during	e	each	early	ed

edu	effect	eg	eight	eighty	either
else	elsewhere	end	ended	ending	ends
enough	entirely	especially	et	et-al	etc
even	evenly	ever	every	everybody	everyone
everything	everywhere	ex	exactly	example	except
f	face	faces	fact	facts	far
felt	few	ff	fifth	find	finds
first	five	fix	followed	following	follows
for	former	formerly	forth	found	four
from	full	fully	further	furthered	furthering
furthermore	furthers	g	gave	general	generally
get	gets	getting	give	given	gives
giving	go	goes	going	gone	good
goods	got	gotten	great	greater	greatest
greetings	group	grouped	grouping	groups	h
had	hadn't	happens	hardly	has	hasn't
have	haven't	having	he	he's	hed
hello	help	hence	her	here	here's
hereafter	hereby	herein	heres	hereupon	hers
herself	hes	hi	hid	high	higher
highest	him	himself	his	hither	home
hopefully	how	howbeit	however	hundred	i
i'd	i'll	i'm	i've	id	ie
if	ignored	im	immediate	immediately	importance
important	in	inasmuch	inc	include	indeed
index	indicate	indicated	indicates	information	inner
insofar	instead	interest	interested	interesting	interests
into	invention	inward	is	isn't	it
it'd	it'll	it's	itd	its	itself
j	just	k	keep	keeps	kept
keys	kg	kind	km	knew	know
known	knows	l	large	largely	last
lately	later	latest	latter	latterly	least
less	lest	let	let's	lets	like
liked	likely	line	little	long	longer
longest	look	looking	looks	ltd	m
made	mainly	make	makes	making	man
many	may	maybe	me	mean	means
meantime	meanwhile	member	members	men	merely
mg	might	million	miss	ml	more
moreover	most	mostly	mr	mrs	much
mug	must	my	myself	n	n't

na	name	namely	nay	nd	near
nearly	necessarily	necessary	need	needed	needing
needs	neither	never	nevertheless	new	newer
newest	next	nine	ninety	no	nobody
non	none	nonetheless	noone	nor	normally
nos	not	noted	nothing	novel	now
nowhere	number	numbers	o	obtain	obtained
obviously	of	off	often	oh	ok
okay	old	older	oldest	omitted	on
once	one	ones	only	onto	open
opened	opening	opens	or	ord	order
ordered	ordering	orders	other	others	otherwise
ought	our	ours	ourselves	out	outside
over	overall	owing	own	p	page
pages	part	parted	particular	particularly	parting
parts	past	per	perhaps	place	placed
places	please	plus	point	pointed	pointing
points	poorly	possible	possibly	potentially	pp
predominantly	present	presented	presenting	presents	presumably
previously	primarily	probably	problem	problems	promptly
proud	provides	put	puts	q	que
quickly	quite	qv	r	ran	rather
rd	re	readily	really	reasonably	recent
recently	ref	refs	regarding	regardless	regards
related	relatively	research	respectively	resulted	resulting
results	right	room	rooms	run	s
said	same	saw	say	saying	says
sec	second	secondly	seconds	section	see
seeing	seem	seemed	seeming	seems	seen
sees	self	selves	sensible	sent	serious
seriously	seven	several	shall	she	she'll
shed	shes	should	shouldn't	show	showed
showing	shown	showns	shows	side	sides
significant	significantly	similar	similarly	since	six
slightly	small	smaller	smallest	so	some
somebody	somehow	someone	somethan	something	sometime
sometimes	somewhat	somewhere	soon	sorry	specifically
specified	specify	specifying	state	states	still
stop	strongly	sub	substantially	successfully	such
sufficiently	suggest	sup	sure	t	t's
take	taken	taking	tell	tends	th
than	thank	thanks	thanx	that	that'll

that’s	that’ve	thats	the	their	theirs
them	themselves	then	thence	there	there’ll
there’s	there’ve	thereafter	thereby	thered	therefore
therein	thereof	therere	theres	thereto	thereupon
these	they	they’d	they’ll	they’re	they’ve
theyd	theyre	thing	things	think	thinks
third	this	thorough	thoroughly	those	thou
though	thoughh	thought	thoughts	thousand	three
throug	through	throughout	thru	thus	til
tip	to	today	together	too	took
toward	towards	tried	tries	truly	try
trying	ts	turn	turned	turning	turns
twice	two	u	un	under	unfortunately
unless	unlike	unlikely	until	unto	up
upon	ups	us	use	used	useful
usefully	usefulness	uses	using	usually	uucp
v	value	various	very	via	viz
vol	vols	vs	w	want	wanted
wanting	wants	was	wasn’t	way	ways
we	we’d	we’ll	we’re	we’ve	wed
welcome	well	wells	went	were	weren’t
what	what’ll	what’s	whatever	whats	when
whence	whenever	where	where’s	whereafter	whereas
whereby	wherein	wheres	whereupon	wherever	whether
which	while	whim	whither	who	who’ll
who’s	whod	whoever	whole	whom	whomever
whos	whose	why	widely	will	willing
wish	with	within	without	won’t	wonder
words	work	worked	working	works	world
would	wouldn’t	www	x	y	year
years	yes	yet	you	you’d	you’ll
you’re	you’ve	youd	young	younger	youngest
your	youre	yours	yourself	yourselves	z
zero	zt	zz			

Appendix C ACF and PACF Results of Microwaves and Pacifiers

Table 10: The ACF and PACF results of microwaves.

	R_t		P_t		N_t		N_t^W		E_t	
Lags	ACF	PACF	ACF	PACF	ACF	PACF	ACF	PACF	ACF	PACF
1	-0.26	-0.26	**0.302**	**0.302**	**0.821**	**0.821**	0.285	0.285	-0.138	-0.138
2	0.232	0.177	0.179	0.096	0.604	-0.217	-0.072	-0.167	-0.033	-0.053
3	0.082	0.197	0.325	0.274	0.47	0.142	0.188	0.291	-0.226	-0.243
4	0.115	0.157	0.308	0.169	0.379	-0.014	0.416	0.296	-0.228	-0.325
5	0.127	0.154	0.094	-0.082	0.287	-0.054	0.131	-0.038	0.272	0.167
6	0.058	0.068	0.113	-0.004	0.15	-0.186	-0.01	0.035	0.044	0.037
7	-0.067	-0.154	-0.007	-0.19	0.067	0.113	0.02	-0.126	0.003	-0.109
8	0.054	-0.112	-0.102	-0.161	-0.015	-0.191	-0.033	-0.208	-0.055	-0.026
9	0.046	0.019	-0.072	-0.047	-0.093	-0.003	-0.076	-0.076	-0.037	0.108
10	-0.166	-0.181	-0.113	-0.066	-0.171	-0.12	-0.011	0.009	-0.083	-0.164

Table 11: The ACF and PACF results of pacifiers.

	R_t		P_t		N_t		N_t^W		E_t	
Lags	ACF	PACF	ACF	PACF	ACF	PACF	ACF	PACF	ACF	PACF
1	-0.03	-0.03	0.245	0.245	**0.804**	**0.804**	**0.763**	**0.763**	**0.59**	**0.59**
2	0.065	0.065	0.125	0.069	0.612	-0.096	0.644	0.147	0.556	0.32
3	-0.232	-0.229	-0.14	-0.198	0.421	-0.118	0.531	-0.002	0.32	-0.157
4	0.37	0.378	0.189	0.286	0.287	0.034	0.344	-0.235	0.27	0.01
5	-0.076	-0.085	0.037	-0.054	0.252	0.177	0.248	0.023	0.072	-0.152
6	0.016	-0.077	0.057	-0.035	0.173	-0.182	0.174	0.036	0.119	0.124
7	-0.157	0.031	-0.039	0.058	0.092	-0.084	0.108	0.018	-0.016	-0.075
8	0.172	0.013	0.075	0.022	0.011	-0.005	0.006	-0.187	-0.028	-0.086
9	0.049	0.101	-0.034	-0.075	-0.092	-0.122	-0.162	-0.332	-0.087	0.004
10	-0.038	-0.095	-0.121	-0.136	-0.202	-0.205	-0.252	-0.048	-0.215	-0.274

Appendix D Pearson correlation coefficient of microwave and pacifier indicators.

Table 12: Pearson correlation coefficient of the microwave indicators.

	R_t	P_t	N_t	N_t^W	E_t
R_{t-1}	-	**-0.426***	0.032	0.13	**-0.441***
P_{t-1}	**-0.443***	-	0.014	0.207	**-0.421***
N_{t-1}	-0.101	**0.484***	-	**-0.411***	-0.104
N_{t-1}^W	0.206	**0.569****	-0.163	-	**-0.474***
E_{t-1}	**-0.537****	-0.157	-0.091	**0.451***	-

$**$. Significant at $\alpha = 0.01$

$*$. Significant at $\alpha = 0.05$

Table 13: Pearson correlation coefficient of the pacifier indicators.

	R_t	P_t	N_t	N_t^W	E_t
R_{t-1}	-	**-0.495***	-0.052	-0.235	**-0.443***
P_{t-1}	**-0.445***	-	-0.044	-0.118	-0.264
N_{t-1}	-0.322	0.057	-	-0.157	**-0.554****
N_{t-1}^W	**-0.401***	**-0.410***	0.043	-	-0.069
E_{t-1}	**-0.740****	**-0.598****	-0.062	-0.144	-

$**$. Significant at $\alpha = 0.01$

$*$. Significant at $\alpha = 0.05$

Appendix E Parameter Fitting Results of the Difference Equation Models for Microwaves and Pacifiers.

Table 14: Parameter fitting results of the difference equation model for microwaves.

m	α_m	λ_{mn}				
		n				
		1	2	3	4	5
1	4.71	0	-6.85	0	0	-0.01
2	0.42	-0.05	-0.47	0	0	0.003
3	-46.46	0	14.02	1.26	6.1	0
4	-0.33	0	0.18	0	0	-0.01
5	-189.47	33.78	0	0	7.02	0

Table 15: Parameter fitting results of the difference equation model for pacifiers.

m	α_m	λ_{mn}				
		n				
		1	2	3	4	5
1	4.86	-0.87	-1.71	0	0	-0.02
2	0.617	-0.03	-0.66	0	0	0
3	822.09	0	0	-0.17	0	-30.62
4	116.57	-10.5	-85.79	0	-0.19	0
5	74.8	-7.99	-47.94	0	0	-0.4

Appendix F PCA Results of Microwaves and Pacifiers

The PCA results of microwaves and pacifiers are shown in Table 16 and 17.

Table 16: The PCA results of microwaves.

No.	Characteristic root	Variance contribution rate	Cumulative variance contribution rate
1	2.03	40.51%	40.51%
2	1.21	24.19%	64.70%
3	0.9	17.89%	82.59%
4	0.54	10.84%	93.43%
5	0.33	6.57%	100%

Table 17: The PCA results of pacifiers.

No.	Characteristic root	Variance contribution rate	Cumulative variance contribution rate
1	2.03	40.64%	40.64%
2	1.28	25.65%	66.29%
3	0.98	19.67%	85.96%
4	0.41	8.25%	94.21%
5	0.29	5.79%	100%

Appendix G Sensitivity Analysis Results of Number of Data on the Success Index

Table 18: The change percentage of the success index with changed number of data, including hair dryers, pacifiers, and microwaves.

	Hair dryer	Pacifier	Microwave
Average coefficient change	1.47%	0.98%	1.34%

References

[1] Aakash, A., Jaiswal, A.: Segmentation and ranking of online reviewer community: The role of reviewers' frequency, helpfulness, and recency. International Journal of E-Adoption (IJEA) **12**(1), 63–83 (2020)

[2] Chen, K.Y., Luesukprasert, L., Seng-cho, T.C.: Hot topic extraction based on timeline analysis and multidimensional sentence modeling. IEEE transactions on knowledge and data engineering **19**(8), 1016–1025 (2007)

[3] Chiang, K.P., Dholakia, R.R.: Factors driving consumer intention to shop online: an empirical investigation. Journal of Consumer psychology **13**(1), 177–183 (2003)

[4] Duan, W., Gu, B., Whinston, A.B.: Do online reviews matter? an empirical investigation of panel data. Decision support systems **45**(4), 1007–1016 (2008)

[5] Jensen, M.L., Averbeck, J.M., Zhang, Z., Wright, K.B.: Credibility of anonymous online product reviews: A language expectancy perspective. Journal of Management Information Systems **30**(1), 293–324 (2013)

[6] Korfiatis, N., GarcíA-Bariocanal, E., SáNchez-Alonso, S.: Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. Electronic Commerce Research and Applications **11**(3), 205–217 (2012)

[7] Lin, C., He, Y.: Joint sentiment/topic model for sentiment analysis. In: Proceedings of the 18th ACM conference on Information and knowledge management. pp. 375–384 (2009)

[8] Mankiw, N.G.: Principles of economics. Cengage Learning (2020)

[9] Mudambi, S.M., Schuff, D.: Research note: What makes a helpful online review? a study of customer reviews on amazon. com. MIS Quarterly **34**, 185–200 (2010)

[10] Pearson, K.: Liii. on lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **2**(11), 559–572 (1901)

[11] Schlosser, A.E.: Can including pros and cons increase the helpfulness and persuasiveness of online reviews? the interactive effects of ratings and arguments. Journal of Consumer Psychology **21**(3), 226–239 (2011)

[12] Tu, R., Ge, J., Feng, W.: The moderating effect of reference group on positive word-of-mouth effect. Enterprise Economy (10), 17 (2018)

[13] Yang, B., Liu, Y., Liang, Y., Tang, M.: Exploiting user experience from online customer reviews for product design. International Journal of Information Management **46**, 173–186 (2019)