Global Plastic Waste Model based on Linear Programming and Support Vector Regression

Summary

Plastic waste (PW) has been one of intractable environmental dilemma for modern humankind. In order to mitigate this tough problem, we combine production of plastics and management of plastic waste to establish a model, which can evaluate the impacts of plastic waste and find out an achievable minimal level of plastic production that can be reduced with a time line.

Firstly, we develop a plastic waste estimate model based on Product Life Cycle (PLC) theory to calculate the maximal levels of single-use or disposable plastic product waste within the environmental capacity in a certain region or country. Then we independently explore the four possible ends of PW and formulate their impacts on environment by several functions. We use linear programming that includes one particular objective function and four constraints. As a result, we can output the maximal amount of plastic products by analyzing the input digits of any region.

Secondly, to find out to what extent plastic waste can be minimized, we established an model called HSVR which combines happiness-index analysis and SVR method. By analyzing the characteristics of different regions and people's living standards, we find out the relationship between the amount of plastics produced and people's happiness index. Based on this, we can analyze how much plastic production we can reduce without significantly reducing happiness.

Thirdly, we follow the previous happiness index to assess how much human life is changed, and apply the Environmental Assessment of Solid Waste Systems and Technologies (EASE-WASTE) model to evaluating how the the environment is affected by plastic reduction. Moreover, we take computable general equilibrium (CGE) model to quantify the effects of the target plastic production. Then we normalized the outputs and into non-dimensional indices, and calculate the weight for each indicator above by the analytic hierarchy process (AHP). The total impact for achieving the confined level is a linear combination of the three indices.

Furthermore, reduction of plastic production will impact countries unequally because of their different development level. We use dual programming to fine the dual model of the linear programming based model we established first. Then shadow prices of every factor formulated by constraints can be calculated, which indicate the significance of each factors. Then we divided them into environmental and governmental factors. Sensitivity analysis is be used to identify which factors that is most sensitive to a particular country, and the equitable reduction of plastic of the country depends on the categories that its sensitive factors belong to.

Finally, we analyze how to approach the minimum level following the time line by an achievable way. Many unexpected circumstance would definitely appear so we empirically choose most significant possibilities that may delay or accelerate the achievement process. The ultimate result are specified into a memo which will be provided for ICM.

Keywords: Plastic waste; LCA; SVR; EASEWASTE; AHP

Contents

8 Conclusion **16**

1   Introduction

1.1   Background

For e-commerce platforms such as Amazon, there is no shortage of the customer's ratings and reviews. Ratings are usually expressed in five grades of 1 to 5 stars, while reviews are expressed in text form, giving customers the opportunity to show their feedback. But how to make good use of this huge rating and reviewing data is also a big challenge. Taking good advantage of these data can not only help companies to provide targeted customer service, find potential high sales and low sales goods but also help them adjust operational strategies, or enhance products in a timely manner[8].

1.2   Problem statement

Now, Sunshine Company plans to launch three new products to sell online: microwave, pacifier and hair dryer. The company has obtained historical ratings and comments from customers of these three products. They hope to dig out effective information from reviews and ratings to help companies better formulate sales strategies and enhance the effectiveness of products. They hope that the following problems can be reasonably solved:

1. What is the relationship between product star rating, review and help votes? Can we quantify the reviews and explore the mathematical relationship among them?

2. Based on task1 to address the following specific requests from the Sunshine Company Marketing Director:

   (a) How to build a model that can integrate ratings and reviews, and help company get effective information and adjust marketing strategies?

   (b) How to describe the reputation of the product from the perspective of time, so as to see and predict the change of the product's reputation in the online market?

   (c) Based on text evaluation and rating, how to find potential successful or failed products?

   (d) Is there a relationship between a specific star rating and the number of reviews? In other words, what kind of ratings attract customers to comment?

   (e) Are subject words of a review, such as "happy", "disappointed" related to specific ratings?

2   Model Analysis

2.1   Data analysis based on LDA and Spearman correlation coefficient

The product data given contains a lot of information, but the deep-seated information information must be processed through information mining so that it can be used for Sunshine Company.

During the mining process, there must be some invalid data, so we filter it through data cleaning. Secondly, for the non numerical field in the data – review, we need to make qualitative analysis. After the key words in the reviews are classified by LDA model[6], we can get the quantitative results.

Thirdly, we make statistics on the comment itself, and get the characteristic information such as sentence number, average sentence length, etc. And finally, we analyze internal relationship between star ratings and helpful ratings and quantified reviews features by Spearman correlation coefficient[7].

2.2   Rating-review based reputation model

We can infer the reputation of a product from its star rating and reviews. The better the reputation of a product, the higher the potential sales volume of the product.

Here, we can't simply measure the reputation of a product by star rating or evaluation alone, because there are often situations of "five star with negative feedback" or "one star with praise". Therefore, we integrate star ratings and reviews to build a reputation model. By comparing the reputation of products, Sunshine Company can enhance the stock and publicity of products with high reputation.

In similar way, the reputation calculation needs to first carry out LDA processing on the reviews of a product, and then convert them into discrete level values.

In order to consider the credibility of the reviews, we also consider several fields such as vine, verified_purchase, helpful_votes, and total_votes, which can measure the accuracy of the review, because reviews with product experience are more credible, and reviews of high helpful votes are more referential.

By modeling and calculating the pacifier samples with more than 25 comments from July to August 2015, we find that in July to August, the reputation of sample 'wubbanub soft toy and pacifier' is higher and the sales prospect is the largest; the reputation of 'Philips aventbpa free another pacifier, 3 + months, green, 2 count' is the lowest and the sales prospect is the worst.

2.3   Relations between time and reputation

The reputation of products will fluctuate with the change of time, and the latent regulation may have great inspiration for sunshine company. We can use RBF neural network[1] to fit (reputation time) function. After getting the fitting curve, we can directly observe the change trend of reputation. Take the pacifier for example,

- First of all, we select the sample with the most reviews in the pacifier–'philips avent bpa free soothie pacifier, 0-3 months, 2 pack' with 833 reviews

- Then we use the reputation transformation model to get the reputation of all samples

- Then, taking one day as the unit time, RBF neural network is used to fit the reputation and time

- Finally, according to the fitting curve, we find that the reputation of Philips pacifier has a minimum value on May 30, 2015 and a maximum value on May 10, 2015. In the next question, the causes of the maxima and minima are analyzed.

2.4   Indication of potentially successful or failing product

As mentioned above, reputation will fluctuate with time, so the curve we fit will naturally have inflection point. Therefore, we define that if the curve will usher in a maximum value, the

product will become a potential successful product; if the curve will usher in a minimum value, the product will become a potential failure product.

In order to facilitate processing, we only consider a set of maxima and minima of the curve when modeling. We can cluster the stars and reviews near the maxima and minima of the curve to find the topic words and stars with the highest word frequency. When these themes and stars appear, it means that the reputation of the product will probably usher in a turning point.

In our experiment, it is found that when 'baby cant' and 2-3 stars appear, the reputation of the product will decline, so the product will become a potential failure product; when 'baby love' and 4-5 stars appear, the reputation of the product will rise, and the product will become a potential success product.

2.4.1 Relationship between specific star ratings and more reviews

Whether a particular star rating will cause more comments. We can judge this is a statistical problem and one-way ANOVA can be used. We need to count the average star rating in the previous period and the total number of comments in the next period to find their relationship. Take pacifier for example.

- First of all, we selected the most commented samples in pacifier, which is 'philips avent bpa free soothie pacifier, 0-3 months, 2 pack, packaging may vary'(833 samples).

- Then we count the number of comments between different time periods of the product and the average star rating before the time period.

- Because the independent variable only has star rating, we discretize it into high star rating, medium star rating and low star rating. Then we use one-way ANOVA to find the relationship between star ratings and reviews.

- Finally, according to the statistics, we found that The relationship between the average star rating in the current time and the total number of comments in the next time period is not strong, but after the semantic analysis of the comments, we divide the comments into positive comments, medium comments and bad comments. We found that in the face of a period of time when products are low stars, users are more willing to give negative comments; in the face of a period of time when products are high stars, users are more willing to give positive comments.

2.4.2 Relationship between specific quality descriptors of text-based reviews and rating levels

The topic asks us to find out whether the specific words appearing in the comment, such as "disappointed", "enthusiasm", will be related to the rating of the comment. Because we are looking for the relationship between the data, we can determine whether this is a statistical problem. We can still use one-way ANOVA, taking the blower as an example.

- First of all, we make word frequency statistics for the comments in the hair dryer, and select the words with higher frequency.

- Then, we select the comment with key words in the hair dryer. We code the comment 0 or 1 based on whether it contains key words or not.

- Then because the independent variable only contains keywords, we continue to use one-way ANOVA to find the relationship between keywords and star ratings.

- Finally, according to the statistics, we found that comments with the keyword "disappointed" are closely related to star rating of one star, and comments with the keyword "love" or "great" or "good" are closely related to star rating of five stars.

3   Assumptions and justifications

To simplify the problem, we strive for the following assumptions:

- Users can't see the rating stars of other users long ago.

- Higher product reputation means better sales prospects.

- Star ratings in a time period satisfies the normal distribution.

- The head and body of a review are always consistent.

- Poor reviews with high star rating exists.

- Positive reviews with low star rating also exist.

- Reviews are more valuable than ratings.

4   Parameter Table

Important symbols that in this article are listed in Table 1.

Table 1: Notations

Symbols	Definition
R_{vec}	Topic frequency vector of a review
R_{t_i}	The keyword frequency score under the ith topic
R_{t_mid}	Word frequency score of the topic in middle reviews.
R_{t_low}	Word frequency score of the topic in poor reviews.
R_α	Star rating value.
R_β	Qualified value of a review.
W_r	The relationship coefficient of rating and review.
R_h	Helpful votes.
R_u	Unhelpful votes.
β	Hyper-parameter in our reputation model.
R_{s_num}	The number of sentences in each reviews.
R_{l_avg}	The average length of each sentence.
$\overline{s_j}$	The length of the business cycle
$\overline{s_j}$	The average star in the j-th business cycle
C_{j+1}	The total number of comments in the (j+1)-th business cycle

5 Model Construction

5.1 Data analysis model based on LDA and Spearman correlation coefficient

5.1.1 Data analysis model based on LDA

The product data contains thousands of items, so mining the connotation information is very helpful for Sunshine Company.

First of all, we clean the data: remove the data that contains empty fields, random codes and other invalid fields.

Then we use Latent Dirichlet Allocation(LDA)[6] to do analysis, because LDA can identify the hidden subject information in large-scale document collection in an unsupervised way. What's more, it can transform each document sample into a word frequency vector about the topic, thus transforming the text information into intuitive numerical information. Our LDA topic classification modeling is as follows:

1. Do text preprocessing on reviews.

    (a) Unify case and remove symbols and numbers from review text

    (b) Delete manually filtered stop words and calibrate misspelled words.

    (c) Do word normalization based on stem extraction.

2. The key words are extracted and selected by counting the frequency distribution of every word, and then integrated with the high frequency key words selected by the 2-gram model[8], so that each review is transformed into a key words array.

3. Put the reviews data into LDA model in the form of keyword array, and set the number of different topics, then train.

4. According to coherence score, the number of topics T_n that are most suitable is filtered, and the topics are assigned according to the usefulness of keywords in each topic.

5. Each review is converted into topic frequency vector $R_{vec} = [R_{t\_1}, ..., R_{t\_n}]$ (where R_{t_i} is the keyword frequency score under the ith topic) according to its keyword array. And take the topic score of the maximum word frequency as the review score.

The results of LDA subject classification are as follows (take microwave oven as an example):

Figure 1: The Distribution of Three Topics with Coherence Score 0.543

(a) topic 1   (b) topic 2   (c) topic 3

Figure 2: Keyword Cloud for Each Topic in Microwave

From the figure, we can see that

- The distance between the three categories is relatively large, and the overall topics consistency score is high.

- According to the key words in the three topics, we can see that topic1 mainly contains a kind of positive emotional key words such as 'perfect', 'nice', 'love' etc;

- Topic3 contains some feedback on 'problems' and other specific topics of poor functional evaluation;

- The key words of Topic 2 are relatively neutral, so the three topics are scored 3,2,1 respectively (corresponding to good reviews, medium reviews and poor reviews)

And the other two products are also scored in this way which is not show in the paper due to space constraints:

Table 2: Microwave's Topic Score Based on Keyword Attribute

Topic	Key Words	Score
TOPIC1	nice, good, perfect, cook, small, love, need, etc	3
TOPIC2	paint, door, button, back, light, etc	2
TOPIC3	repair, problem, return, not work, etc	1

5.1.2 Spearman Correlation Coefficient

We make characteristic statistics on the reviews, and get the number of sentences in each reviews R_{s_num} and the average length of each sentence R_{l_avg}.

Then we use quantile-based discretization assignment method to grade the continuous value features.

And last, we use spearman correlation coefficient formula as follows to calculate the correlation between the star rate $R\alpha$, and the quantitative feature of reviews text $\{R_\beta, r_h, R_{s\_num}, R_{l\_avg}, R_{vec}\}$:

$$\frac{\sum_i^n (a_i - \bar{a})(b_i - \bar{b})}{[\sum_i^n (a_i - \bar{a})^2 \sum_i^n (b_i - \bar{b})^2]^{\frac{1}{2}}} \qquad (1)$$

where a, b are two characteristic sequences respectively, $\bar{a}$ and $\bar{b}$ are the average values of the two sequences respectively.

Taking the microwave oven as an example, the following results are obtained:

Table 3: Correlation between Star Rating and Review features

Feature - Feature	Correlation value
star ratingR_α - topic score R_β	0.651
star ratingR_α - review sentence num R_{s_num}	0.209
star ratingR_α - review sentence average length R_{l_avg}	-0.121
star ratingR_α - review helpful votes r_h	-0.301

Table 4: Correlation Among Review Features

Feature - Feature	Correlation value
helpful votesR_α - R_{t_mid}	0.493
review helpful votesR_α - review sentence average length R_{l_avg}	0.459
review helpful votesR_α - review sentence num R_{s_num}	0.213
review helpful votesR_α - R_{t_low}	-0.390

In table4, we use R_{t_mid}, R_{t_low} to represent the feature: word frequency score of the topic in middle evaluation and in low evaluation.

From above tables, we can draw the following conclusions:

1. The correlation coefficient between star rating and topic score R_α is the highest, and it is positive correlation, which shows that some complimentary high-frequency keywords in the positive topic are closely related to the high score ratio of customers.

2. Helpful reviews tend to be rated at low and medium star levels.

3. The highest correlation coefficient between helpful votes and the score of the subject in the middle evaluation indicates that the helpful reviews of these commodities is relatively more pertinent, more objective, and will not praise excessively or turn to the bad side.

4. The positive correlation between helpful votes and review sentence average length and review sentence number indicates that helpful reviews are relatively long.

5. According to the above information, Sunshine Company should pay more attention to the longer reviews with medium star rating and more keywords, so as to find out customers' concerns and ideas and improve products.

5.2   Rating-Review Based Reputation Model

By modeling ratings and reviews, we can get product reputation, and higher reputation means higher potential sales volume, which can help Sunshine Company make better decisions. The modeling process is as follows:

Assume the star rating of the product is R_α, the review is quantified as R_β, the helpful votes of the review is R_h, so that unhelpful votes(R_u) is total votes minus helpful votes, net helpful votes is($R_h - R_u$), We use net helpful votes plus 1 as a factor and multiply the quantified review to judge the credibility of the evaluation.

We use the logarithmic form to make the intermediate calculation smoother, then average the reputation during specific period, and finally refer to the article[4] we magnified the reputation differences between samples in the form of index, and the reputation model is initially established as follows:

$$F_{repu} = e^{\frac{\sum_{i=1}^{n}(log_2(R_\alpha+1)+\beta \times log_2(R_\beta \times (R_h-R_u+1)+1))}{n}} \tag{2}$$

where n is the number of evaluation in a certain period. For the convenience of processing, when the net helpful votes are negative in the model, it will be set to -1, that is, this kind of comment has no contribution to reputation.

Furthermore, on the one hand, the super parameter β should be able to reflect the weight relationship between rating and comment, on the other hand, it should be able to represent the credibility of the review, and the review with the vine or verified_purchas field equals 'y' is a review that has experienced the product, which is more referential. So the expression of β is as follows:

$$\beta = \begin{cases} 1 + W_r & if \quad vine \quad or \quad verified\_pruchase =' Y' \\ W_r, & else \end{cases} \tag{3}$$

where W_r is the relationship coefficient of rating and review mentioned above. Set 0.651 here.

In order to test our model, we first counted more than 25 pacifier samples from July to August 2015, which are:

- 'philips avent bpa free soothie pacifier, 0-3 months, 2 pack, packaging may vary', marked as Sample1

- 'wubbanub soft toy and pacifier', marked as Sample2

- 'wubbanub brown monkey pacifier', marked as Sample3

- 'wubbanub infant pacifier - giraffe', marked as Sample4

- 'philips aventbpa free soothie pacifier, 3+ months, green, 2 count', marked as Sample5

- 'wubbanub brown puppy pacifier', marked as Sample6

Then we calculated the average reputation value of six samples in these two months, and the results are as follows:
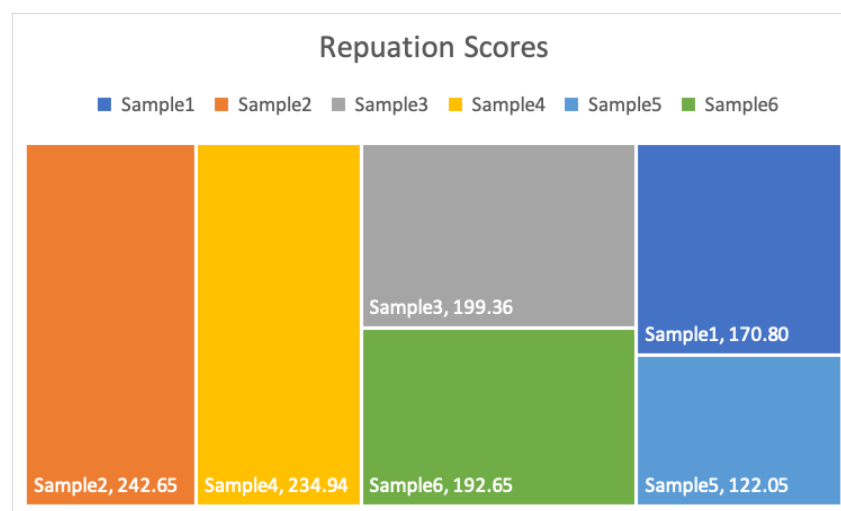

Figure 3: The reputations of six samples calculated from July to August 2015

As shown in the figure:

1. Sample2 and Sample4 belong to the same class, with high product reputation value, which means better sales prospects;

2. Sample3 and Sample6 belong to the same class; Sample1 and Sample5 belong to the same class, with low reputation value, which means that the sales prospects are not clear.

3. In fact, we found that Sample5's rating is indeed full of 1-star and 2-star, and there are many bad comments in the reviews.

5.3 Reputation-Time model Based on RBF Regression

It is helpful for Sunshine Company to adjust its strategy and increase its profit by exploring the changing law of product reputation over time. Inspired by paper[3], we can use neural network to fit and predict reputation.

In this paper, we use RBF neural network as the regression fitting of time based reputation, because RBF has a strong nonlinear fitting ability, can map any complex nonlinear relationship,

and the learning rules are simple, which is convenient for computer implementation. The establishment process of our RBF based reputation time model is as follows:

1. We first filter the data and select the most representative samples of each type of product:

   - **Pacifier:** 'philips avent bpa free soothie pacifier, 0-3 months, 2 pack, packaging may vary' with 833 reviews on 557 dates.
   - **Hair Dryer:** 'remington ac2015 t|studio salon collection pearl ceramic hair dryer, deep purple', with 587 reviews on 436 dates.
   - **Microwave:** 'danby 0.7 cu.ft. countertop microwave' with 394 reviews on 224 dates.

2. By substituting the selected samples into the reputation model, we get the discrete sequence of sample reputation changing with time.

3. Building RBF network:

   (a) Take the first half of the sample as the training set and the second half as the test set, then normalize them.

   (b) Set the number of neurons in the network to increase gradually, the most is the number of training samples, and set the spread speed to 0.1, start training.

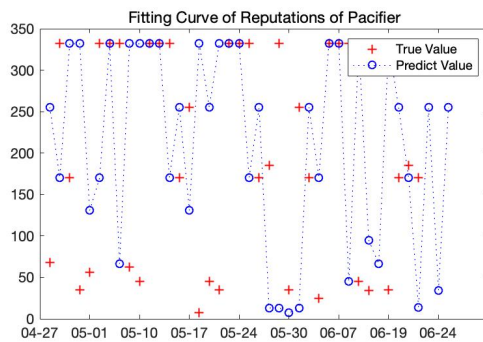   (c) The RBF network was tested and the fitting result was obtained.

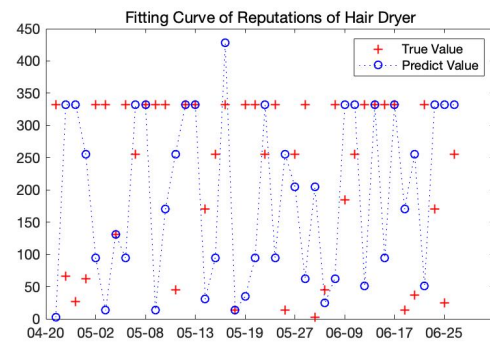Figure 4: Fitting Curve of Reputations of Pacifier

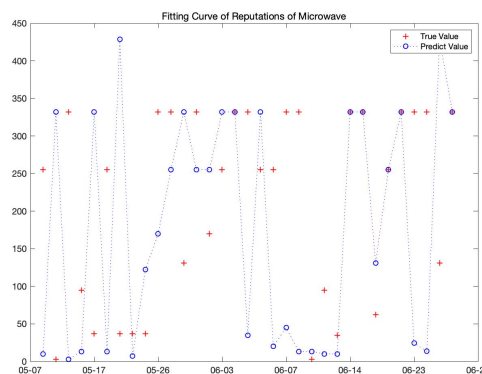Figure 5: Fitting Curve of Reputations of Hair Dryer

Figure 6: Fitting Curve of Reputations of Microwave

From the figure, we can observe the following phenomena:

- For pacifier, the minimum reputation appears around May 30, 2015, for hair dryer the minimum appears around May 19, 2015 and microwave's minimal reputation was around June 7. The causes for the minimum will be explained in the next question.

- For pacifier, the maximum reputation value appears around May 10, 2015, for hair dryer the maximum appears around June 9, 2015 and microwave's maximal reputation was around May 26. Again, we will discuss the causes for the maximum in the next question.

5.4 Model to Indicate Potentially Successful or Failing product

In our modeling, the theme words and ratings near the maximum value of fitting curve are the signs of potential successful products, while the theme words and ratings near the minimum value of fitting curve are the signs of potential failed products.

Take pacifier for example, according to the Fig.4, The maximum value of fitting curve appears on May 10, 2015 (marked as t_1), The minimum value of fitting curve occurs on May 30, 2015 (marked as t_2), we extract stars and reviews at t_1 and t_2's 7 nearby time points respectively, We conduct LDA text analysis and clustering, and the results are as follows:

Figure 7: World Cloud for Reviews around May 10, 2015

Figure 8: World Cloud for Reviews around May 30, 2015

From the clustering results, we can conclude that:

- If the reviews are full of words like 'baby love', 'easy to hold', 'onto love' with 4-5 star ratings, it means that the reputation of the product will usher in a maximum value and it is a potential successful product. So it's recommended for Sunshine Company to grow the portability of pacifiers.

- If the reviews are full of words like 'baby cant', 'baby not', 'without soothing' with 2-3 star ratings, it means that the reputation of the product will usher in a minimum value and it is a potential failing product. And we recommend Sunshine Company to fix the soothing problem with the specific pacifier.

5.5 Model to the relationship between specific star ratings and more reviews

Because there are only specific star ratings can cause more comments from users. Obviously, there is only one independent variable, so we use one-way ANOVA [?]. Our Model is built as follows.

1. We first filter the data and select the most representative samples of each type of product as we mentioned in sec.5.3

2. Further, we screen samples to remove comments whose star ratings are inconsistent with comments.

3. Then we split the time into d days (generally 7 days for business) and assume that the comments in the next period are related to the average star rating in the previous period. Next, the average star rating $\overline{s_j}$ of the i product in the j-th time period and the number of comments in the next time period C_{j+1}. We need to mine the relationship between $\overline{s_j}$ and C_{j+1}:

   The average star level can be any value between $[1, 5]$. For the convenience of seeking the relationship, we can discretize it by dividing $[1, 5]$ into four sections, each section is corresponding to one class, and then $\overline{s_j}$ is divided into different classes according to its value. So, the question is translated into, whether a specific star rating range corresponds to C_{j+1}. Because we only consider the influence of star rating, we use one-way ANOVA.

One-way analysis of variance requires that all variables follow normal distribution, and the variance is uniform

Table 5: Test of Homogeneity of Variances

Levene Statistic	df1	df2	Sig.
2.054	2	812	0.129

From the graph, we find that the variance is homogeneous, and we can use one-way ANOVA:

Table 6: ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12.31	2	6.15	16.599	0.001
Within Groups	13402.85	812	16.51		

We can see from the picture that:

- The relationship between the average stars in the current time period and the total number of comments in the next time period is not strong

Later, we made a semantic analysis of the comments.

1. According to the results of semantic analysis, we divide comments into positive comments, medium comments and negative comments.

2. Then we calculated the relationship between the proportion of positive comments, medium comments and negative comments and A_p

The final statistical results are as follows:

Table 7: Descriptive of the percentage of negative comment

rating	N	Mean	Std.Deviation	Std.Error	F	Sig.
2	2	56.25	8.83	6.25	16.599	0.001
3	18	36.64	42.52	10.05		
4	175	8.68	19.47	1.47		

Table 8: Descriptive of the percentage of medium comment

rating	N	Mean	Std.Deviation	Std.Error	F	Sig.
2	2	37.50	17.68	12.50	0.40	0.671
3	18	20.40	27.16	6.40		
4	175	22.32	25.52	1.93		

Table 9: Descriptive of the percentage of medium comment

rating	N	Mean	Std.Deviation	Std.Error	F	Sig.
2	2	6.25	8.83	6.25	9.828	0.001
3	18	42.95	40.88	9.64		
4	175	68.99	29.36	2.21		

From the figure we can draw the following conclusions:

-   In the face of a period of time when products are low star rating, users are more willing to give negative comments.

-   In the face of a period of time when products are high star rating, users are more willing to give positive comments.

5.6 Model to the relationship between specific quality descriptors of text-based reviews and rating levels

To determine whether there is a relationship between keywords and star ratings in the comments, it is obvious that there is only one independent variable, so we still use one-way ANOVA. Our model to relationship between specific quality descriptors of text-based reviews and rating levels based on one way ANOVA is established as follows:

1. We first filter the data and find the specific quality descriptors with high frequency, such as "disappointed", "love", "great".

2. Further, we also filter samples, because some people will add "not" before "disappointed" to make the meaning opposite. So we need to remove these samples.

3. Based on whether the comments contain keywords, we divided the samples. And based on this, one-way ANOVA is carried out.

The final statistical results are as follows:

Table 10: "disappointed"

	N	Mean	Std.Deviation	Std.Error	F	Sig.
0	11250	4.16	1.27	0.012	745.936	0.000
1	220	1.82	0.84	0.06		

Table 11: "love"

	N	Mean	Std.Deviation	Std.Error	F	Sig.
0	9042	3.97	1.36	0.01	553.465	0.000
1	2428	4.65	0.85	0.02		

Table 12: "great"

	N	Mean	Std.Deviation	Std.Error	F	Sig.
0	8700	3.99	1.37	0.02	346.118	0.000
1	2770	4.51	0.93	0.02		

These show that the distribution of star ratings are significantly affected by the key words. Now we use histogram to show the specific distribution:
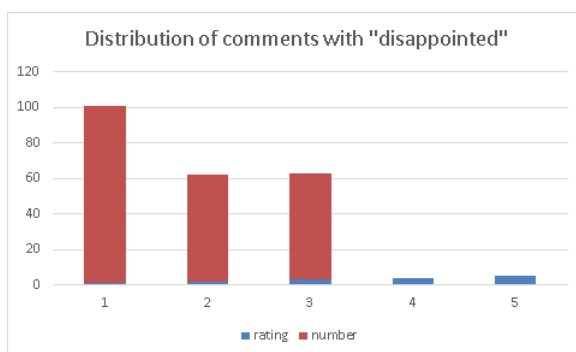

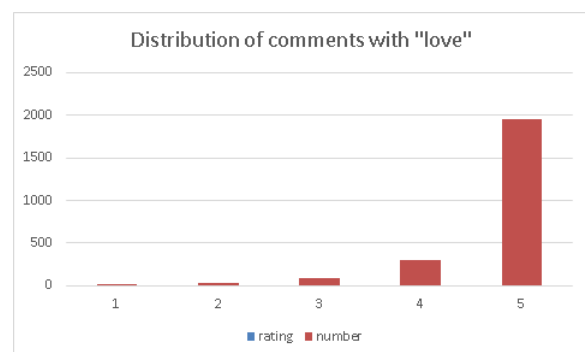Figure 9: Fitting Curve of Reputations of Pacifier

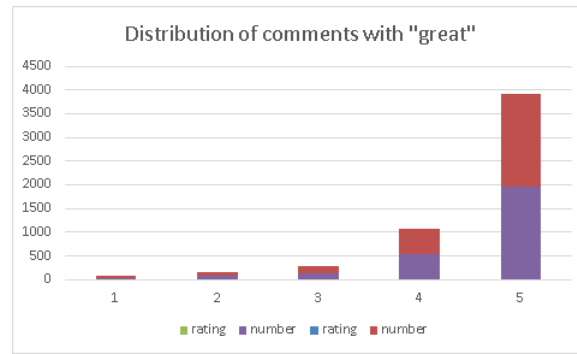Figure 10: Fitting Curve of Reputations of Hair Dryer

Figure 11: Fitting Curve of Reputations of Microwave

From the figure we can draw the following conclusions:

- Comments with the keyword "disappointed" are closely related to star rating of one star;

- Comments with the keyword "love" or "great" or "good" are closely related to star rating of five stars.

6   Sensitivity Analysis

6.1   Hyper-parameter β

In our reputation model, the super parameter β is set to be related to the relationship coefficient W_r. And we think e-commerce product reviews are more important than ratings[8], especially the reviews written after experiencing the product, which can better reflect the customer experience. So β in our model increased the weight of reviews.

In fact, we find that our model is robust to β, i.e. slight fluctuations in β, will not affect the final result of reputation:

We changed W_r from 0.651 to 0.2 and 1.2 in Eq.3 respectively, and recalculated the samples in section.5.2. The results are as follows:
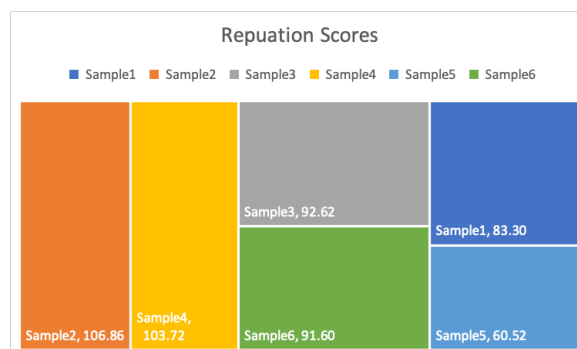
Figure 12: The reputations of six samples calculated from July to August 2015 with lower β
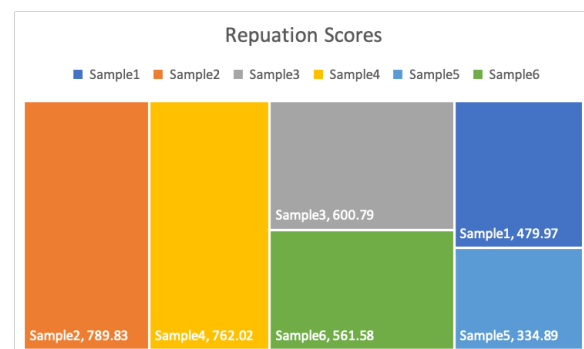
Figure 13: The reputations of six samples calculated from July to August 2015 with higher β

We can see that although the absolute value of reputation has changed, the relative value

remained the same. Sample2 and Sample4 are still the first reputable products, and Sample5 is still the product with the worst reputation among the samples.

7   Strength and weakness

7.1   Strengths

Model rationality. The models we use for reference are all models that have been proved to have strong generalization ability, and our own reputation model also shows certain robustness to hyper-parameters.

Reasonable results. We analyzed the results and found that our results are reasonable and scientific.

Make good use of given data. We eliminate the redundant information and interference information in the given data, and fully mine the potential data wealth in the remaining information.

7.2   Weaknesses

Lack of sales in given data. If the sales volume corresponding to the product can be provided in the given data, then our model will be better tested in practice. **Limited data.** The sample size is not enough, which leads to that our classification of star rating is not fine and the lack of obvious relationship between medium star rating and comments.

8   Conclusion

To mitigate the plastic waste problem, we establish a linear programming base model to estimate the environmental impacts of plastic waste, and get the maximal plastic waste amount without further damage to environment. It is proved that the most extent plastic can be reduced to could be confirmed by HSVR model. A minimal target amount of plastic usage is achieved by expending HSVR model. The equity issue can also be solved by dual programming and sensitivity analysis of factors. Finally, we reckon a timeline to approach this realistic minimal target of plastic using amount.

References

[1] M. M. Li and B. Verma, "Nonlinear curve fitting to stopping power data using rbf neural networks," *Expert Systems with Applications*, vol. 45, pp. 161–171, 2016.

[2] M. R. Stoline, "The status of multiple comparisons: simultaneous estimation of all pairwise comparisons in one-way anova designs," *The American Statistician*, vol. 35, no. 3, pp. 134–141, 1981.

[3] D. Tang, B. Qin, T. Liu, and Y. Yang, "User modeling with neural network for review rating prediction," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[4] T. Subkhankulova and F. J. Livesey, "Comparative evaluation of linear and exponential amplification techniques for expression profiling at the single-cell level," *Genome biology*, vol. 7, no. 3, p. R18, 2006.

[5] X. Zhang and C. Dellarocas, "The lord of the ratings: is a movie's fate is influenced by reviews?" *ICIS 2006 proceedings*, p. 117, 2006.

[6] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022.

[7] L. Myers and M. J. Sirois, "Spearman correlation coefficients, differences between," *Encyclopedia of statistical sciences*, vol. 12, 2004.

[8] X. Zhang and C. Dellarocas, "The lord of the ratings: is a movie's fate is influenced by reviews?" *ICIS 2006 proceedings*, p. 117, 2006.

LETTER TO SUNSHINE COMPANY

To: Marketing Director of Sunshine Company

From: Team 2013573

Subject: A Wealth of Data

Date: March 9, 2020

Dear Director:

Thank you very much for your trust and hire our team to solve this important problem for you. As we all know, there is a lot of valuable information in the product ratings and reviews. We hope that the information our team digs from the data of pacifier, microwave and hair dryer can help your company to make reasonable sales strategy, and enhance effectiveness of products timely. We hope our suggestions can be helpful to your company.

We clean the data your company provided and take text processing. We introduce LDA model to cluster the topic of the text, and quantify the text into a number of numerical quantities through statistics. After the correlation analysis of these numerical features with the help votes and star rating, we verify the positive correlation between the high star rating and the rating.

As a result of text processing and correlation analysis, we suggest that your company should pay more attention to the high-frequency keywords under our various topics, which represent customers' preferences or functional concerns or some common defects of products. On the other hand, we also found some interesting information after correlation analysis. For example, helpful reviews tends to be on the low and medium star rating, and more pertinent, more objective, because it is closely related to the subject keywords of the middle evaluation. Therefore, we suggest that your company should pay more attention to longer reviews with medium star rating and more keywords, so as to get more information.

As for how to integrate product ratings and reviews to provide effective information for your company, our team first quantified the text reviews effectively, and then made full use of helpful votes, vine to trace the credibility of the reviews. As the result, a robust model is established. We believe that a good reputation can potentially bring a better sales volume to a product, so we suggest that your company can evaluate the reputation of pacifier, microwave, hair dryer and other products, and then increase the stock of products with high reputation or amplify their publicity. As mentioned in our paper, 'wubbanub infant pacifier - giraffe' is the first choice for online sales of pacifier products from July to August 2015.

Our fitting model uses a powerful RBF neural network to fit the relationship between reputation and time successfully, which can help us better summarize and predict the change of reputation of products. Furthermore we can excavate the potential factors of product success or failure from the maximum and minimum values. As we mentioned in the article, when words like 'baby cant' or 'without soothing' with 2-3 stars appear, we should be extra careful. This is the product the reputation of which will begin to decline. We recommend your company to improve your products in time according to such signs. When word like 'baby love' or 'easy

to hold' with 4-5 stars appears, your company can safely increase the inventory of the product, because the product is about to sell well.

Our statistical model uses one-way ANOVA. After analyzing the average star rating of products in the past period of time and customers' comments of products, we find that when customers see products with lower star rating, they are more willing to publish negative comments; When they see products with higher star rating, they are more willing to publish positive comments. Therefore, we suggest that your company should try its best to improve the products before launching the products, so as to avoid more bad comments caused by low star rating, which will lead to a vicious circle.

We also finds that some words in the comments are related to specific star ratings. For example, the reviews with "disappointed" are closely related to the star rating of one star, reviews with "love" and "great" are closely related to the star rating of five stars. Therefore, we suggest that companies should focus on extracting reviews with emotional words, research and analysis, and optimize products in a timely manner.

We would be very pleased if our suggestion could be of benefit to your company. Wish the given advice above would contribute to your work!

Your sincerely
Team 2013573