

Summary

This is the age of information explosion. Numbers, text, video and audio are all information. How to dig out the instructive information we are interested in from the large amount of data is crucial. A rating model based on user comments is constructed.

To understand consumer preferences and basic information, we used TextBlob, a natural language processing library in Python, to perform word frequency statistics and sentiment analysis on the content of product reviews. We have seen an exponential increase in the number of reviews, with more and more consumers choosing to shop online, with beauty and baby products more popular. Specific quality descriptors of text-based reviews such as 'love', 'disappointed', and others, strongly associated with rating levels.

In order to understand product reputation and predict the success of product promotion, we defined product reputation score and product future development score through a rating model based on user reviews. In order to understand whether the product evaluation content has the Matthew effect, we use the moving average algorithm in the time series to analyze the correlation between the number of comments of each product evaluation quantity and the number of comments, and find that the comments of the following commenters are correlated with the comments of the previous commenters.

Finally, we performed sensitivity analysis and extended the model.

Keywords: keyword1; keyword2

Contents

1	Introduction	1
1.1	Background	1
1.2	Restatement of the Problem	1
1.3	Overview of Our Work	1
2	Assumptions and Justifications	1
3	Notation	1
4	Name of Model	1
5	Name of Mode2	1
6	Results and Recommendations	1
6.1	Inform Online Sales Strategy	1
6.2	Potentially Important Design Features	3
6.3	The Product's Reputation	4
6.4	The Future Trend of the Product	6
7	Modifications of our model	7
8	Sensitivity Analysis	7
9	Strengths and Weaknesses	7
9.1	Strengths	7
9.2	Weaknesses	7
	Reference	7

1 Introduction

1.1 Background

.

1.2 Restatement of the Problem

1.3 Overview of Our Work

2 Assumptions and Justifications

•

•

•

•

3 Notation

Table 1

Symbol	Description
--------	-------------

4 Name of Model

5 Name of Mode2

6 Results and Recommendations

6.1 Inform Online Sales Strategy

First, we do some basic analysis of the known raw data to understand the basic situation of online sales. Then we raised a series of questions that the sun company might be interested in.

1. Is online shopping becoming more common in America?
2. Will online sales replace traditional marketing?
3. Whose user experience is better online than traditional sales?
4. What products are suitable for online sales?

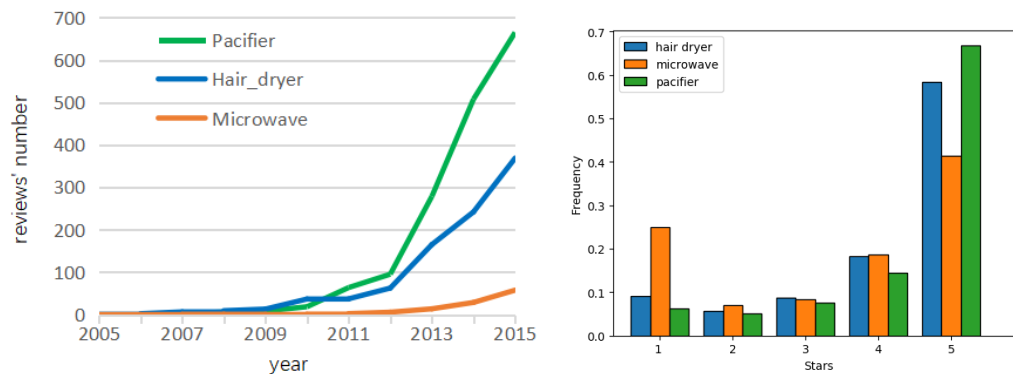


Figure 1: Figure on the Left: the average number of monthly reviews of the three products per year; Figure on the right: star frequency of all data of three types of product

We know that the number of comments is directly proportional to the number of online shoppers, so we can see the figure on the right as the relationship between the number of online shoppers and time. From the left chart, we can find that before 2008, very few people chose to shop online. After 2008, online shopping became more and more popular. To analyze the reasons, the subprime mortgage crisis in the United States in 2008 triggered the global financial crisis, which led to the failure of many banks and the collapse of small and medium-sized enterprises. In order to survive, many small businesses have switched to online sales, because it is known that online sales cost is small, with no rent, sales staff.

We found that the number of online shoppers increased exponentially after 2008, indicating that online shopping is becoming more and more popular with customers. We have reasons to believe that online sales will beat traditional marketing, and online shopping will become the new shopping habit of people in the new era. Here we strongly suggest the sun company to develop an online sales business.

Pacifiers are baby products, hairdryers are beauty products, and microwave ovens are large appliances. On the left, we found that compared to hairdryers and pacifiers, the number of microwave reviews is very small, that is, the number of people who buy microwave ovens online is very small. Combined with the picture on the right, the one-star microwave oven can be found to be very large, which indicates that consumers do not like to buy large appliances online. In summary, beauty and baby products are more suitable for online sales.

The figure on the right shows the star distribution map of online reviews of three kinds of products. The frequency of five-star pacifiers and hairdryer is more than 50%, which indicates that most people have a good online shopping experience for these

5 stars	4 stars	3 stars	2 stars	1 star
love 2431	pacifier 545	pacifier 309	pacifier 194	product 178
loves 2381	love 394	son 187	bought 133	pacifier 159
pacifier 2257	easy 355	pacifiers 172	son 126	bought 142
easy 1678	son 355	daughter 149	product 103	time 117
daughter 1500	daughter 312	bought 142	pacifiers 101	buy 113
son 1468	pacifiers 293	time 124	time 86	son 111
pacifiers 1333	bought 267	product 119	buy 86	pacifiers 100
bought 1166	product 263	months 118	months 80	month 91
product 791	loves 263	love 116	daughter 76	daughter 89
months 786	time 227	cute 107	bag 65	months 86
cute 752	nice 219	mouth 103	nipple 64	received 80
recommend 748	seat 216	easy 102	cute 63	hard 75
hold 747	bit 210	seat 98	seat 63	child 71
mouth 724	months 193	month 95	monitor 62	waste 71
time 724	month 175	hard 89	month 61	plastic 71
perfect 720	car 169	bit 88	hard 61	item 70
loved 702	mouth 154	fit 87	love 57	return 69
month 609	bag 153	pretty 78	received 56	monitor 69
buy 596	cute 152	hold 75	mouth 54	money 64
nice 571	size 151	nipple 75	size 52	disappointed 64

Figure 3: Top20:The frequency of words corresponding to different stars of the baby pacifier

Bad reviews are what we're really focused on, where we can identify areas where consumers are dissatisfied, so we adjust our product or online sales strategy to improve our competitiveness. As can be seen from the figure 3 :

1. When evaluating products, consumers prefer to describe their subjective feelings rather than product attributes.
2. Specific quality descriptors of text-based reviews such as 'love', 'disagree', and others, strongly associated with rating levels.
3. In the low-star rating, "hard" and "plastic" reflect that the potentially important design functions of the baby nipple are its hardness and material safety; Potentially important design functions of the hairdryer are drying effect and service life. Potentially important design features of microwave ovens are functional diversity and after-sales service.

6.3 The Product's Reputation

Consumers' real evaluation score only considers the content and stars of the review, not the importance of the review. The four most important factors when considering product reputation are star rating, reviews, vine, and helpful votes. ^[1]

Each commenter contributes p_I to the product's reputation:

$$p_i = \sum_i^{n_i} [s_i + k_i(r_{i1} + r_{i2})] \quad (1)$$

The reputation of a product takes into account the reputation of the product rated by each reviewer at a certain point in time, and the following indicators are obtained. This indicator can describe the reputation of the product in a certain period, whether it goes down or up, where n_k is the number of reviews of j product in the k month. K month reputation score of j product rp_{jk} :

$$rp_{jk} = \frac{\sum_i^{n_k} [s_{ji} + k_{ji}(r_{ji1} + r_{ji2})]}{n_i} \quad (2)$$

In drawing the presentation, we did some manipulation on the data: each reviewer's reputation for the product was divided into rp_i by converting it to a number on $[-1, 1]$. The value on $(0, 1]$ indicates that each reviewer's contribution to the product's reputation is positive; The value on $[-1, 0)$ indicates that each reviewer has a negative reputation contribution to the product; 0 indicates that each reviewer has no reputation contribution to the product; since the maximum number of baby pacifiers is 18,939, the data we selected is pacifier data.

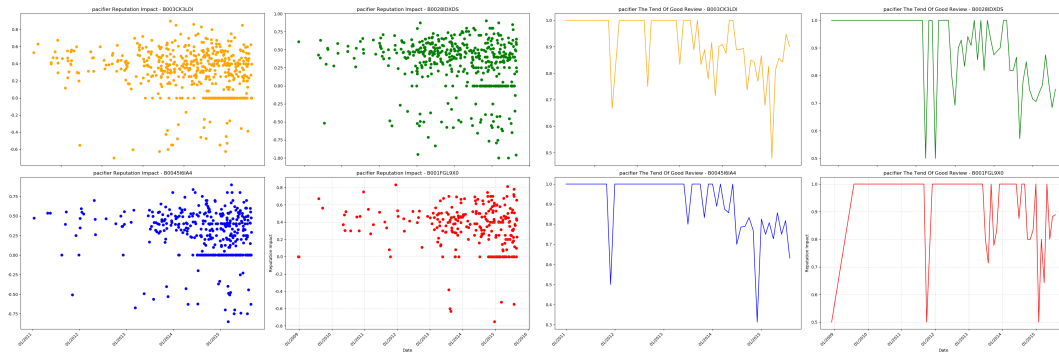


Figure 4: Left: reputation score of each reviewer of the top four products reviewed in the baby pacifier data; Right: monthly reputation scores for the top four products reviewed in the baby pacifier data

Figure 4 shows that the reputation scores of the top four products with the number of comments are mostly concentrated above 0, which indicates that most people have positive comments on the products (the same conclusion as the figure 1 on the right). Here we are looking at the top four products in terms of the number of reviews, so these products also have high sales.

Figure 4 the right graph studies the proportion of the number of five-star comments to the total number of comments in a month. Reputation scores are stable at 0.5 or above, and the occasional five-star review for the month is good for the sales company. Around 2015, reputation took a turn for the worse and then stabilized. As online shopping becomes more and more common and the review data keeps increasing, it is not possible to display every comment on the shopping platform, so

the network companies make corresponding adjustments to the display of comments on the shopping platform.

To sum up, sales and reputation are closely related, we strongly suggest that the sunshine company should pay attention to the reputation of products.

6.4 The Future Trend of the Product

How do we define success after a product has been sold in an online marketplace? If the product has been bought by many people, the comments after purchase are mostly positive, and the product has a good reputation on the Internet, then the future trend of the product must be good.

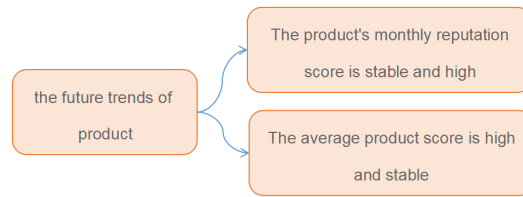


Figure 5: the combination to indicate a potentially successful or failing product.

Reputation score and product score are important factors affecting the future development of a product. According to the formula ??, it can be known that the product score f_{3i} is a comprehensive indicator to measure the product quality. According to the formula 2, we can know that the reputation score will affect the future sales of the product. These two indicators, one to measure the quality of product attributes, the other reflects the social value, together they can have a comprehensive forecast of the future trend of the product. Product future development score ψ :

$$\psi_j = \text{avg} \sum (f_{2jk} + rp_{jk}) \quad (3)$$

Table 2: Comparison of three indicators for different products

product id	review number n_j	product score f_{2j}	product reputation rp_j	total ψ_j
B0009XH6TG	555	53.76	13.30	67.06
B00VRN7SB8	15	3.37	16.00	19.37
B005GSZB9Q	78	13.53	15.42	28.95
B00NXRHIO8	13	4.57	17.62	22.18
B003CK3LDI	515	54.13	13.04	67.17
B00UH2XBOS	14	2.65	13.31	15.96

From the table 2: the more reviews the product has, the higher the future development score. Because the more reviews a product sells, the more reviews there are, the more reviews there are, so the number of reviews indirectly represents the number of sales. If we can prove that the number of comments is positively correlated with

the score of future development, then it can be shown that the higher the score of ψ_j in future development, the more sales will be generated and the more successful the product is likely to be.

7 Modifications of our model

8 Sensitivity Analysis

9 Strengths and Weaknesses

9.1 Strengths

-
-
-
-

9.2 Weaknesses

-
-
-
-

References

- [1] W. U. Wann-Yih, PO-CHING HUANG, and CHEN-SU FU. The influence of an online auction's product price and e-retailer reputation on consumers' perception, attitude, and behavioral intention. *Scandinavian Journal of Psychology*, 52(3):290–302.