User Needs Search Using the Text Mining
- From Commodity Comparison, Understanding
the Difference of Users Awareness

Yukiko Takahashi and Yumi Asahi

3-5-1 Johoku Naka-ku Hamamatsu 432-8561 Japan
`ne210026@gmail.com,`
`tyasahi@ipc.shizuoka.ac.jp`

Abstract. In this paper, the authors establish that "To apply free description message options are good for product development." By using the technique of text mining, we show the awareness of the product user as an example a review site about cosmetic products.

We took up the major review site about cosmetics who can see the review of products freely. By using text mining, we show the awareness of users who are using products and how the images of product formed by elements and this analysis help the product development.

Keywords: text mining, frequently-appearing word, free writing.

1   Introduction

Recently, huge information has been accumulated by the development of the information technology in the Internet. Therefore, the enterprise can obtain the data that can be used for the business. As a result, the enterprise can obtain the data that can be used for the business. Written in the customer data, the questionnaire data, the complaint data, and the electronic bulletin board etc.

Especially, the number of users who used social sites such as "Mixi" and "Twitter" increased at last few years. Knowing the users opinion of the company's product and competitor's product is very important, because it help the development and improvement. It is very important to know competitor's product opinions. If we're making the request for the researching company, we have to spend a lot of money.

So, using text mining for review site is one of the effective measures. If the product is selling, you can get the users opinion about the product easily by using the review sites on the Internet. Review site, which anybody can access, is able to get the user's opinion of competitor's product

To use text mining for business, it will lead to discrimination from other companies. However, the text mining is a language of tongue twisters in Japanese as an actual current state compared with English, and the enterprises that positively analyze it is not so many. There might be information that cannot be obtained from

information on useful choices for the product development and the improvement only to such a free description etc[3]. Not only having a questionnaire companies, but also by using Analyzed opinions on the Internet, It is possible to expand the range of product development strategy.

The purpose of this research is by using review sites about cosmetic we do text mining for 2 items, which supported by difference age and we analyze what elements make difference between ages. We derive a useful strategy for promotion from the result.

2   Target Sites for Analyzing and Items

2.1   Target Sites for Analyzing

In this research, we analyze the ranking data of free description opinion in cosmetic review sites @cosme. This site handled 21000 domestic and foreign brands and cosmetics information about the more than 200 thousand items. Reviews written by user's counts over 10.5million, Page view is 2,400 million per month, unique user is 7.7million. This shows users are interested in this site, so they visit (As of 2013.3).

2.2   Product to the Subject of Analysis

This research is doing text mining for the user's free description opinions on the manicure ranking (June, 2013). Analysis target is LE VERNIIS of CHANEL No.1 of the manicure ranking (June, 2013). These two items is not same age supporter. Under 20's support AT. Under 30's support LE VERNIIS.

Table 1. Analysis target (LE VERNIIS)

Ranking of @cosme	1
Interior content	13ml
Price	¥3,045(About $31)
Data collection period	2012/12/16～ 2013/6/15(half a year)
Number of Sentence	211
Number of character	75852

This research is looking for the elements which acclaimed elements by using text mining of review. We show about two items which target of analyze the information of price or how many reviews there is (Table1-2). The numbers of reviews of LE VERNIIS are 211. AT are 312. As we look the price, there are large differences of 2items. This shows that LE VERNIIS is brand name goods which came out from CHANEL, and AT is sold in 100shops. It is much related with review. The 100yen

shop provides the products all price 100yen. It like ¢99Shop in USA. We should know user's awareness which the difference from the price.

Table 2. Analysis target (AT)

Ranking of @cosme	2
Interior content	10ml
Price	¥105(About $1)
Data collection period	2012/12/16～ 2013/6/15(half a year)
Number of Sentence	312
Number of character	72861

3   About Text Mining

3.1   Summary Introduction of Text Mining

Text-mining-Analysis is a technique that consists of two data mining "natural language processing" and "data mining". Text analysis had been used in automatic translation and automatic summarization of the document originally [2]. And discussions have been made in the field of language model research. Recently, a new phrase (term) "text mining" which means digging up the valuable information from large amounts of text data has been established. Text mining is a general term for technologies to obtain some findings by analyzing document data in large quantities.

3.2   Flow of Text Mining

Text mining is performed as following steps [1].

1. Morphological processing is divided into sentences written in natural language: morphological analysis
2. Dispose the unwanted words such as auxiliary verbs and particles for analysis or create a synonym: unification of synonyms, such as after an unnecessary processing
3. Process of analyzing the dependency relation with words: parsing
4. Analysis: analyze the frequency and number of occurrences, co-occurrence, clustering, and attributes

Is often carried out after the above one, working up to 4-2 is repeated while confirming the results. After step 1, repeat step 2 to 4 while affirming their results.
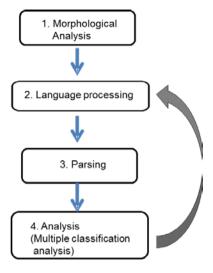

Fig. 1. Flow of text mining

3.3   Analysis Result (LE VERNIIS)

As we look at a result of analysis, frequently-appearing words in LE VERNIIS and AT are color and it is connected with many words. So, we can get to know the importance of color (chart2). We can draw a conclusion that the frequently-appearing words "color" is a strong reason for buying, and pink and beige are popular colors. Lame and Perl like luminescent things is connected with pink, so, we can expect this popularity is supported by young people. Other frequently-appearing words of LE VERNIIS, such as "Coloring" "Pink" "Piece" came at the top. The price in AT is at the top too. So, we checked out the co-occurrence relation and analyzed more. As we did co-occurrence, we got the result that LE VERNIIS is expensive and AT is cheap.

Starting from the left, table 3 the word of price back and belly, how many words in back and belly, how many words in back and total of back and belly, T level. No matter whether there is a co-occurrence relation or not, as judgment, it is intended that the value T in corpus linguistics statistically. T value, the statistical analysis is an index, which is used for example, when performing a test of the difference between the average values. This is corpus linguistics is widely used as an index to determine the presence or absence of co-occurrence relations. T value formula is the following. [10] Based on formula (1), the value of T is the result after certifying the difference of means.
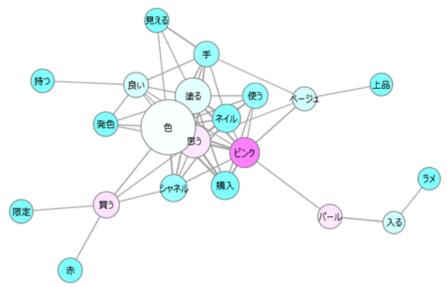
Fig. 2. Co-occurrence network of words (LE VERNIIS)

Value T = (measured value - mean) / Square root of actual
measurement value (1)

"The square root of the measured value" in the denominator represents the
approximate value of the standard deviation about the words co-occur with language
center here.

In corpus linguistics it is considered as equal to or greater than 1.65, the co-
occurrence of two terms is no coincidence. There is a co-occurrence relation if there is
more than 1.65 T value based on corpus linguistics, this time you shall make any
sense.

Also we can see that in LE VERNIIS, brand name CHANEL comes out many
times. But we can't find the same situation in AT. Therefore, LE VERNIIS is trying
to integrate the brand value of "Chanel". We connect these results to the product
development, improvement and advertising slogan.

If we use the result of this analysis, we can promotion as CHANEL and except the
increasing of repeater like expecting coloring, dry. And we can plan the new mini
bottle which is cheap and reduced the amount of content. But when we check out
frequently-appearing word, maker name is higher occurring frequently than name of
article. So, we get to know the name of CHANEL has brand value. Now therefore,
when we plan the new bottle promotion, it is not good to sell doing competition to
keep the prices down. It's good to sell more expensive than competitor. And more it is
good to sell by the price which teenagers can buy and a little more expensive than
competitor.

3.4   Analysis Result (AT)

"Colors" is most frequently-appearing word as same as LE VERNIIS. But number of advent words difference is smaller than LE VERNIIS. But advent word is large. So, we can get to know reviews are not organized. And "word of mouth" "watch" is adverted. So, it affects the review. Brush is larger than nail. It is a special feature.
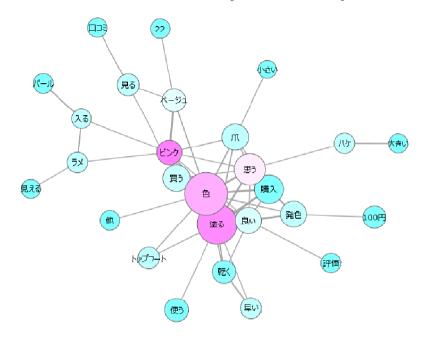

Fig. 3. Co-occurrence network of words (AT)

Table 3. The co- occurrence relation to the price(AT)

Rank	Word	Before	After	Span	T
1	low	1	4	5	2.144629
2	deserved	0	4	4	1.991654
3	and over	0	4	4	1.968704
4	think	0	3	3	1.70555

We did Co-occurrence analysis   for the words "put on" "come off" "dry" " thick" "price". As a result, LE VERNIIS has almost positive opinions, But AT has negative opinions. There are positive and negative opinions. Positive opinions are "Easily put on" "Easy to use" "Easily to dry", negative opinion are "come off easy" "Easy to become irregularity". Furthermore result of the co-occurrence analysis for price,

it became Table 3. The user recognizes that the price of this product is low. So, this product is easy to attract new users.

3.5   Analysis Summary of 2 Products

To summarize the analysis below, what the user feels for the two products are different. And furthermore we were able to visualize it. The quality of LE VERNIIS is good. And there is a brand value. However, since the price is expensive, purchase rate of young users is low. In contrast, AT does not broadcast TV commercial, but name recognition is up. This is because there is an effect of the review and cheap price. But there is room for improvement in quality. In this way, text mining would be helpful in the sales strategy of their products. And it is beneficial in corporate activities.

4   Future Outlook

This research analyzed the difference of user review comments not take the money to browse. Target prediction is essential to product planning. And knowing the opinion of the target's opinion will lead to good product development and improvement. There is definitely a demand for the text mining. But that analysis is difficult.

Then, the future task is to ensure that apply to business with less cost and less effort. Excellent open source software of the morphological analysis "Chasen" and "MeCab" is announced. In addition, it becomes a practicable level and has spread. We become necessary that while using them, we need to summarize user's opinion by using multivariate analysis. Making a text mining is one of the important thing for the company. And the text mining technology large amount of s benefits even in business activities. The text mining technology is expected to have on future research and application.

References

1. Hirohiko, A.: Introduction to questionnaire survey (2011)
2. Yumi, I., Takaaki, H., Isamu, W., Mitsuhiro, S.: Text mining- Case introduction. Japanese Society for Artificial Intelligence Magazine 16(2)
3. Tetsu, K., Seisyou, M., Masaaki, N., Tatsuo, O., Kenzi, Y.: Language and statistics of psychology. Iwanami Shoten Publisher (2003)
4. Barnbrook, G.: Language and computers. Edinburgh University Press (1996)