A proposal of review extraction method considering the knowledge amount of consumers and the purchase stage

Koichi Horie, Masataka Onoda, Takeshi Kaneko*

Tokyo City University: 1-28-1 Tamatsukuri Setagaya-ku Tokyo, 158-8557 Japan

*tkaneko@tcu.ac.jp

Abstract:
Today internet shopping sites are widely used. On such sites, consumers cannot take products in their hands, so user reviews of products must be informative for the consumers. However, number of reviews are very large, so consumers cannot read them all. The may miss a very important information which can catch reading all reviews. Purpose of our research is to make a method of indicating only useful reviews for reducing consumer’s burden.

We assume that useful reviews are different by consumers. We found that usefulness of reviews differs from consumers’ purpose of browsing review, and also from consumers’ knowledge about purchase target product group. We can classify consumers into six groups. Next we identified factors of useful reviews for each group. Finally we constructed models for evaluating usefulness of each reviews for each group.

We made dictionaries for each factor to judge which factors are included in each review.

For validating our models, we conducted questionnaires’ survey. Consumers were shown two groups of user reviews, one group contains top 10 useful reviews based on our model, and the other contains original 10 reviews which shopping site shows. We could confirm that our model can show more useful reviews to consumers.

Keywords
Electronic commerce site, Information overload

1. Background

The number of people using electronic commerce site (I will write it as EC site hereafter) is increasing. EC site is convenience that they can purchase products without going to a real store, but on the other hand, they cannot watch the real products and take the actual products in their hands.

Therefore, EC site users judge purchase of goods by referring to the evaluation of the product review posted on EC site. From this, it turns out that the product review is an important source of information by consumers. However, now, there are so many pieces of information on the Internet. Therefore, the amount of selectable information that consumers can obtain on the Internet has increased sharply. As a result, various information can be obtained, but problems arise that consumers cannot consume so many pieces of information, and it takes time to find product reviews that consumers will find useful. Also, it is difficult to read all product reviews, so there is a problem that there is a possibility of missing important information.

2. Precedent Study

Various researches have been made to solve the problem.

Hirayama et al (2011) propose a method to display the overall evaluation by to extract reputation attribute representing the function and feature from the product review, and to aggregate positive and negative evaluations.

Kawai et al (2011) propose a method to visualize reputation using graphs based on information that classified as positive and negative sentences from product reviews.

Hu et al (2005) propose a method to extract and summarize feature of evaluation using natural language analysis from product reviews of EC sites.
Precedent study focuses on summarizing the content of product reviews. However, it is reader who uses product reviews, it is not clear what kind of information they want to know.

So, in this study, we evaluate product review with emphasis on readers of that.

3. Propose

Based on the above points, the purpose of this study is to clarify the useful product reviews for each consumer and to propose a method for mechanically extracting useful reviews. As a result, it is think consumers can shorten the time to search useful product reviews for consumers, and it leads to reducing the burden of product selection.

4. Content of study

4.1 Classification of Consumers

It is difficult to extract suitable reviews for each consumer according to various kinds of preferences and circumstances of consumers. Therefore, we thought that it is possible to classify consumers depending on the difference in useful product reviews, and we conducted interview.

4.1.1 Interview Summary

In order to classify consumers, we conducted interview on EC site users “What purpose they view product reviews for, and what product reviews are useful when they select products”. In addition, we asked them to view the actual product reviews and choose reviews that they thought it was useful.

Also we asked them why the review was judged to be a useful review.

A total of 48 people in their 20s to 40s who routinely use EC sites asked for product search on the premise of purchasing cameras and asked which reviews were useful. When selecting a specific product from among multiple candidate products or excluding a specific product, it was helpful that the information written in the review was helpful. Information based on the actual feeling of use that cannot be obtained only by the product description or specification by the manufacturer or seller is described in the review. The subjects were asked to determine whether useful information was provided to determine whether or not it is a product that matches the purpose of purchase or use of the consumer.

4.1.2 Interview Results

From the results of the interview, we found that usefulness of product reviews differs from consumers’ purpose of browsing product review, and also from consumers’ knowledge about purchase target product group. From that, consumers could be classified into six groups. (Figure-1)

![Figure-1 Classification of consumers](image)

We arranged the three purchase stages below.

1. The product has not been decided at all.
Consumer decided on purchase target product group, and they are checking what kind of products are there.

② The stage of comparing products

After having a look at some products, consumer narrow down the number of products that they want to purchase, and narrow them down to one or a few.

③ The Stage of deciding one product

Consumer narrows down the product to be purchased to one, and are making final confirmation as to whether or not to purchase product.

4.2 Clarification of Useful Product Review

Next, in order to clarify what kind of product review is useful for each of the six group, we conducted a questionnaire for each group.

4.2.1 Summary of Questionnaire

In one review, multiple contents are often written, we divided the product review by content. As a result, we divided 43 product reviews into 111 product review sentences. We asked consumers to sort it into useful product reviews and non-useful product reviews for each group.

By self-declaring in advance, ten college students who clarified which group they belonged were prepared for each group. 111 actual review sentences were classified into 43 types according to the similarity of content and asked whether each review sentence is useful or not useful for product selection.

We thought that the contents of useful product reviews could be different for each product group. So we conducted a questionnaire that limited the target product group to cameras.

4.2.2 Abstraction of Useful Product Review Content

Due to the similarity of contents and similarity in judgment of usefulness in each consumer category, we abstracted the results of the questionnaire to clarify the content of useful product reviews by Affinity Diagram. The results of the questionnaire is more than 25 product reviews that are said to be useful for each group. Abstract review contents are arranged in Table 1. From now on in the study, “product review factors” indicates the result of abstraction.

Table-1 Product Review Factors

| Product Review Factor                        |
|---------------------------------------------|
| Evaluation of function and performance      |
| Convenience of functionality                |
| Usability(Weight, Shape)                    |
| Usability(Function, Performance)            |
| Recommended people(Weight, Shape)           |
| Recommended situation(Weight, Shape)        |
| Explanation for beginners                   |
| Comparison of other products                |
| Evaluation of design(Color, Appearance)     |
| Points to note when choosing                |
| Reason for purchase                         |
| Durability                                  |
| Evaluation of price                         |
| Evaluation of shipping                      |

We arranged useful product review factors and not useful product review factors for each group in Table 2. “○” in Table-2 indicates that more than 7 out of 10 judged useful. “×” in Table-2 indicates that less than 3 out of 10 judged useful. “-” indicates factors not belonging to “○” and “×”. That is, “○” is affixed to a review factor that more than 70% of the consumer category is useful.
4.3 Weighting of Product Review Factors

We thought that there was a difference in the usefulness of useful product review factors for each of the six group. So, we clarified the usefulness of each product review factor for each group. And we created a model for evaluating the usefulness of product reviews for each group. As a method, using Scheffe's paired comparison, we prepared a product review example for each useful product review factors and asked consumer to evaluate the product reviews in 9 grades. We conducted a questionnaire for five consumers per each group, and asked the consumer to repeatedly answer until the answer fulfilled the consistency index. The criteria of consistency index for groups P1, P2, P3, A1 and A2 are 0.1. For group A3, we set the criterion of consistency index to 0.15, because group A3 has as many as 11 factors, it is difficult to make consistency in judgment. We arranged the results of questionnaire in Table-3.

| Product Review Factor                  | P1  | P2  | P3  | A1  | A2  | A3  |
|---------------------------------------|-----|-----|-----|-----|-----|-----|
| Evaluation of function and performance |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |
| Functionality of convenience          |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |
| Usability(Weight, Shape)              |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |
| Usability(Function, Performance)      |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |〇〇〇〇 |
| Recommended people(Weight, Shape)     |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |
| Recommended situation(Weight, Shape)  |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |
| Explanation for beginners             |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |
| Comparison of other products          |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |
| Evaluation of design(Color, Appearance)|××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |
| Points to note when choosing          |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |
| Reason for purchase                   |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |×××〇〇〇 |
| Durability                            |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |−−−〇〇〇 |
| Evaluation of price                   |××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |
| Evaluation of shipping                |××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |××〇〇〇〇 |
For each subject, two review statements corresponding to each review element are presented, and one pair is compared in nine steps which is more useful. From the completed paired comparison table, find the geometric mean of people belonging to each category. Furthermore, the weights shown below are standardized by the geometric mean of the usefulness for each consumer category. It can be interpreted that the weights of reviews that many of the consumers included in the consumer category judged to be more useful become larger.

4.4 The Method of Extraction Review

4.4.1 Identification of Product Review Factors

In order to extract useful reviews for each group, we need to identify which product review factors the one review contains. So, we created a dictionary to identify each product review factor. Regarding "Evaluation of function and performance", we prepared two dictionaries "function" dictionary and "evaluation" and identified the product review factor. Regarding "Convenience of functionality ", we prepared two dictionaries "function" and "presence" and identified the product review factors. We did not create a dictionary of two review factors "Reason of purchase" and "evaluation of shipping".

Regarding the product review factors of "Reason of purchase "and "evaluation of delivery", it is known that they are not included in useful product review factors from the results of 4.2.2. So we did not create a dictionary of them because we do not need to extract reviews on that factor. Therefore, we created a dictionary to identify 12 product review factors. We arranged the created dictionary for each product review factor in Table-4.

Table-4 Dictionary

| Product Review Factor | Dictionary |
|-----------------------|------------|
| Evaluation of function and performance | Function and performance + Evaluation |
| Convenience of functionality | Function and performance + Presence or absence |
| Usability(Weight, Shape) | Weight and shape + Usability |
| Usability(Function, Performance) | Function and performance + Usability |
| Recommended people(Weight, Shape) | Weight and shape + Person + Recommended |
| Recommended situation(Weight, Shape) | Situation + Recommend |
| Explanation for beginners | Beginners |
| Comparison of other products | Comparison |
| Evaluation of design(Color, Appearance) | Design |
| Points to note when choosing | Points to note when choosing |
| Durability | Durability |
| Evaluation of price | Price |

4.4.2 Score of the review

When extracting product reviews, we want to extract high usefulness product reviews. So it is necessary to evaluate the review. As a method, we adopted the method of using the model created in 4.3 and adding the weight to the product review considered to contain useful product review factors. And, the top 10 product reviews with high scores are extracted.

From the above contents, we arranged extraction method in Figure-2. Decompose the review sentences into words and check those words against the dictionary shown in Table 4. For each word included in each product review factor, weights are added to make the review usefulness. Naturally, if it is a long review sentence, it is assumed that the usefulness will be high, but taking into consideration the large amount of information, we do not standardize by length here.
4.5 Measurement of Extraction Accuracy

We measured how accurately the keywords registered in the dictionary can extract the contents of each dictionary. We also measured how correctly the dictionary combination can extract the product review factor. After that, we measured how much useful reviews were included in the reviews extracted based on 4.4.

For the measurement method, we calculated the precision and the recall. The precision indicates the ratio of how much the correct answer is included in the search result and expresses accuracy. The recall indicates the ratio of how much hit the search from within the correct answer, and indicates coverage. If the measured results is low, we adjust the keywords registered in the dictionary and improved the results.

In this study, the precision and the recall were calculated as follows.
The precision: Number of product reviews extracted correct answers / Number of product reviews extracted
The recall: Number of correct answers extracted product reviews / Number of correct answers of product reviews in all data

4.5.1 Measuring Accuracy of dictionaries

We measured the accuracy for each dictionary. We asked consumers to divide 102 reviews into each dictionary, and measured accuracy by compared that results with the product review as determined by dictionary. As a result of measurement, we adjusted the keywords registered in the dictionary and improved the results, because the result of measurement is low.

Basically, the precision and the recall are a trade-off relationship. Regarding the precision and the recall of the dictionary, we judged both accuracy and coverage to be important, so we adjusted the results of them to be 80% or more.

We arranged the final measurement results in Table 5.

Table-5 Accuracy of dictionary

| Function and performance | Function and performance | Usability | Situation | Beginners |
|--------------------------|--------------------------|----------|-----------|-----------|
| precision                | 86%                      | 91%      | 91%       | 96%       | 91%       |
| recall                   | 88%                      | 98%      | 87%       | 79%       | 91%       |
| Comparison               | Design                   | Points to note when choosing | Durability | Price |
| precision                | 67%                      | 89%      | 75%       | 80%       | 93%       |
| recall                   | 100%                     | 89%      | 67%       | 100%      | 86%       |
| Evaluation               | Presence or absence      | Person   | Recommend | Terminology |
| precision                | 92%                      | 60%      | 94%       | 100%      | 88%       |
| recall                   | 93%                      | 67%      | 84%       | 74%       | 88%       |
4.5.2 Measuring accuracy of product review factor extraction

We also measured the accuracy of the product review factor extraction method as well as the dictionary. We asked consumers to divide 102 reviews for each review factor, and measured accuracy from the results and results divided by the dictionary.

Regarding the precision and the recall of the product review factor extraction method, we judged both accuracy and coverage to be important, so we adjusted the results of them to be 80% or more.

We arranged the final measurement results in Table 6.

Table-6 Accuracy of product review factor extraction

| Evaluation of function and performance | Convenience of functionality | Usability(Weight, Shape) | Usability (Function, Performance) | Recommended people (Weight, Shape) |
|---------------------------------------|-----------------------------|--------------------------|----------------------------------|-----------------------------------|
| precision 95%                         | 100%                        | 97%                      | 87%                              | 100%                              |
| recall 81%                            | 78%                         | 71%                      | 87%                              | 56%                               |
| precision 100%                        | 83%                         | 58%                      | 89%                              | 83%                               |
| recall 27%                            | 100%                        | 93%                      | 89%                              | 63%                               |
| precision 80%                         | 81%                         |                          |                                  |                                    |
| recall 100%                           |                             |                          |                                  |                                    |

4.5.3 Measuring accuracy of extracted reviews

It is impossible to reduce consumer’s burden if there are too many reviews to show to consumers. Also, the purpose of this study is to reduce the consumer’s burden. So we set the number of reviews to extract to 10.

In the same way, we also measured accuracy of the review extraction method. We arranged the result in Table-7.

Table-7 Accuracy of extracted reviews

| P1 | P2 | P3 |
|----|----|----|
| precision | 70% | precision | 80% | precision | 90% |
| recall | 39% | recall | 47% | recall | 50% |

4.6 Confirmation of information coverage

From the result of 4.5.3, the recall is low, since the number of reviews to extract is limited to 10. However, there are many reviews of similar contents on the EC site. Therefore, even though the recall is low, we thought that we could cover useful information unless new important information was included in the not extracted reviews. So we checked whether the product reviews which we did not extract not contains important information.

As for the checking method, at first, we plotted product reviews which we did not extract on a graph composed of two axes of usefulness and information volume. The usefulness in the graph is the score given to the product review, and the information volume is the number of product review factors included in the product review.
Next we extracted product reviews which is usefulness and high information volume content on the graph. Finally, we had consumers answer whether the product review we extracted for each group is a useful product review. We arranged the result in Table-7.

We judged that reviews not contain important information, if 80 percent of consumers said it is not useful. From the above results, we found that useful information can be covered in 10 extracted reviews.

The reason for paying attention to product reviews with low usefulness and large information volume is that since the information volume of the review is fixed, the review is extracted based on only the usefulness of the review. So we only extracted reviews with low usefulness and high information content to determine if the scale of usefulness is appropriate.

5. Verification

5.1 The Purpose of Verification
In order to verify the usefulness of this study, we evaluated and compared the review on existing EC site and the review extracted based on this research. And, we conducted statistical test to clarify whether there are differences between the two methods.

5.2 The Method of Verification
We had consumers in each group evaluate in seven revels whether the top ten product reviews on target products posted on the existing EC site "Amazon" are useful. In the same way, we had consumers in each group evaluate ten product reviews extracted based on this study.

We conducted 2-sample t-test that result level of significance of 5percent. We set the average of the evaluation of the existing site as μ1 and the average of the evaluation of the product review extracted based on this study set as μ2. And we set null hypothesis as μ1=μ2, and alternative hypothesis as μ1<μ2.

The number of questionnaires is 13 for groups P1, P2, P3, and 12 for A1, A2, and A3.

5.3 Result of Verification
As a result of the statistical test, alternative hypotheses were adopted in all groups. We arranged the result of statistical test in Table-8. From this fact, we found that 10 product reviews extracted based on this study are more useful than the top 10 product reviews of existing EC site, when selecting products. And we were able to verify the usefulness of this study.

| Group | Statistic | t (12,0.1) | Group | Statistic | t (11,0.1) |
|-------|-----------|------------|-------|-----------|------------|
| P1    | -2.554    | -1.711     | A1    | -2.64     | -1.1717    |
| P2    | -2.479    | -1.711     | A2    | -2.419    | -1.1717    |
| P3    | -2.14     | -1.711     | A3    | -2.625    | -1.1717    |

6. Conclusion and Task

6.1 Conclusion
In this research, we propose a method to adequately present consumer reviews of product reviews that are considered useful for product selection. According to the interview survey, it was found by consumers how greatly different product reviews are considered useful when considering product selection. Consumers were categorized into two stages of product knowledge concerning the product and three levels of strength of purchase intention of the product, totaling six. The information described in the review was largely classified into 14 categories and clarified which consumer categories think that information is useful for product selection for each category. In order to measure the usefulness of actual review sentences, we proposed a method of creating a dictionary and multiplying the number of words contained in the dictionary by the usefulness weights. Considering the labor for consumers to read the review, we adjusted the dictionary to increase precision rather than recall. When we...
presented the reviews that were determined to be highly useful in the proposed method to their respective consumer categories, it was judged that they could read the useful review for product selection rather than the order displayed on the regular EC site.

With the product description and product specification by the manufacturer or seller, information is not sufficient for consumer's product selection, and the feeling of use by the actual purchaser or user is extremely important. Furthermore, it was also revealed that the information on reviews considered useful is different depending on the attributes of consumers. These findings are extremely important not only in product development but also in marketing. In particular, in order to enhance conversion at the EC site, it is beneficial to make a review presenting these considerations.

6.2 Task
As a future task, in this study we created a dictionary of function only for camera, so it is considered difficult to extract useful reviews in other product groups. Therefore, we think that it is possible to increase versatility by creating a dictionary for each product, or by creating a dictionary for all products.

References:
Takuo Hirayama, Takayuki Yumoto, Manabu Arai, Yutaka Takahashi (2011), “Extraction and presentation of product evaluation based on attribute evaluation model”, DEIM Forum 2011, F2-5, 2011
Yasushi Kawai, Tomriro Yoshikawa, Takeshi Furuhashi (2011), “Study on reputation information analysis of user review by visualization”, HAL Symposium 2011, III - 2 B - 2, 2011
Minqing Hu, Bing Liu (2005), “Mining and Summarizing Customer Reviews”, KDD-2004, 22-25, 2005

Author’s biographical notes
Koichi Horie was graduated from Tokyo City University in 2017, and is enrolling in the Tokyo City University graduate school.
Masataka Onoda was graduated from Tokyo City University in 2017.
Takeshi Kaneko holds a Doctor of Engineering by Tokyo University and is an assistant professor of Tokyo City University, Faculty of Information Engineering. His research interests focus on Quality Management of Software Products, and Information mining from huge text data.