

# Developing Evaluation Expression Dictionaries for the Cosmetic Review Recommendation

Mayumi Ueda, Yuki Matsunami, Panote Siriaraya and Shinsuke Nakajima

**Abstract**—There are many portal sites to support online shopping. These sites provide commercial items and reviews for them. In particular, in the case of purchasing cosmetic items, reviews have an important role in decision making, in order to recommend items that are appropriate to users. Therefore, we are trying to develop a recommender system for cosmetic items and reviews to help consumers. In order to realize such a recommender system, we develop a method for automatic scoring of various aspects of cosmetic items based on evaluation expression dictionary. In this paper, we aim to automatically construct an evaluation expression dictionary for all categories and analyze the tendency of evaluation expressions for each cosmetic category.

**Index Terms**—cosmetic review, review analysis, explanation of review, automatic review rating, evaluation expression dictionary.

## I. INTRODUCTION

THERE are many portal sites to support online shopping. These sites provide many kinds of information, such as item information by the company, cost, user-provided review, ranking information, and so on. Consumers can get effective information by user-provided reviews before they purchase them. In particular, consumers make careful choice about cosmetics since unsuitable items frequently cause skin irritations, therefore, they try to get effective information for themselves, through the online reviews.

@cosme[1] is a portal site that is very popular among Japanese women that provides product information and user reviews. While this site can be helpful for decision making, it is not easy to find truly suitable cosmetics. This is because, most of the existing review sites provide user reviews that contain review text and overall score and users have to read through huge volumes of review texts to find truly effective information for themselves. In this situation, it is difficult to understand the whole context of target cosmetic items intuitively. Therefore, we design and evaluate a collaborative recommender system for cosmetic items, which incorporates opinions of similar-minded users and automatically scores fine grained aspects of product reviews[2][3].

In order to develop such a review recommender system, we have to analyze review text to understand feedback on

TABLE I  
CATEGORIES OF COSMETIC ITEMS

Skin care	Face wash, Makeup remover & Cleansers, Lotion, Emulsion, Serum, Face masks & Packs, Eye & Lip care
Makeup	Brow products, Eyeliner, Mascara, Eye shadow, Lips, Blush, Nails
Base makeup	Foundation, Base foundation, Concealer, Face powder

reviewers' experiences of cosmetic items. While there is a score of each review text on the conventional cosmetic review sites, It mostly consists of overall score, so this makes it difficult to recognize feedback on reviewers' experiences of cosmetic items from it. For example, there are "moisturizing effect", "whitening effect", "exfoliation care effect", "hypoallergenic effect", "aging care effect", etc. as aspects of reviews for "face lotion". Thus, we need a scoring method of such various aspects on cosmetic item review texts to understanding feedback on reviewers' experience of them.

Hence, the purpose of our study is to develop a method for automatic scoring of various aspects of cosmetic item review texts based on evaluation expression dictionary. The method can realize an automatic scoring of various aspects of cosmetic item reviews even if there are no scores for the individual aspects. As a first step, we constructed an evaluation expression dictionary for "face lotion", which has different feedback with each person manually [2]. According to simple experiments carried out, we can get the results of automatic scoring by our method which are quite close to the results of manual scoring from ground truth data. However, there are many categories for cosmetic items, for example, "face lotion", "emulsion", and "serum" as skin care items (see TABLE I). It takes a lot of man-hours to develop evaluation expression dictionary for each categories of cosmetic items manually. Therefore, in this paper, we analyze the tendency of evaluation expression for each categories of cosmetic items for the purpose of developing the evaluation expression dictionaries automatically.

The paper is organized as follows. The related work is given in Section II. Then Section III describes the evaluation expression dictionary for automatic scoring of various aspects of cosmetic items. We analyze the features of evaluation expression for each cosmetic categories in Section IV. We conclude the paper in Section V.

## II. RELATED WORK

There are several websites that provide reviews by consumers. For example, Amazon.com[4] and Priceprice.com[5] are popular shopping sites on the Internet which provide customer reviews of their merchandise. Moreover, "Tabelog" is a popular website in Japan. This website provides information

This work was supported in part by the MEXT Grant-in-Aid for Scientific Research(C) (#16K00425) and Research(B) (#17H01822).

M. Ueda is with Faculty of Economics, University of Marketing and Distribution Sciences, 3-1, Gakuen-nishimachi, Nishi-ku, Kobe, 651-2188, Japan, e-mail: (Mayumi\_Ueda@red.umds.ac.jp).

Y. Matsunami is with Graduate School of Frontier Informatics, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto-city, Kyoto, 603-8555, Japan, e-mail: (i1788223@cc.kyoto-su.ac.jp).

P. Siriaraya and S. Nakajima are with Faculty of Information Science and Engineering, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto-city, Kyoto, 603-8555, Japan, e-mail: (spanote@gmail.com, nakajima@cc.kyoto-su.ac.jp).

and reviews of restaurants. In addition to the algorithmic aspects, researchers have recently focused on the presentation aspects of the review data[6].

Furthermore, in recent years, “@cosme” has become a very popular portal site for beauty and cosmetic items, and it provides a variety of information, such as reviews and shopping information regarding cosmetic items. According to the report issued by the company that operates “@cosme”, in June, 2018, the number of monthly page views reached 310 million, the number of members reached 5 million, and the total number of reviews was 14 million[7]. Numerous women exchange information about beauty and cosmetics through the service of @cosme. Hence, users can compare cosmetic items of various cosmetics brands. Reviews consist of a review text, scores and tags regarding effects, etc.

Furthermore, the system stores profile data that include information about age and skin type. Therefore, users who wish to browse the reviews can search the reviews according to their own desired characteristics, for example, reviews can be sorted by scores or filtered for one effect.

Several studies have been conducted on review analysis owing to the spread of web sites that share review information. For example, Hiroshi et al. extracted a remark about product reputation from enormous texts with descriptive sentences such as the questionnaire, and checked the writer’s intention[8].

In our previous study, we analyzed reviews of cosmetic items and constructed an evaluation expression dictionary by analyzing reviews. In addition, we developed a method that can automatically calculate the score of each aspect of a “Face lotion” based on the dictionary[2].

O’Donovan et al. evaluated their AuctionRules algorithm, which is a dictionary-based scoring mechanism for eBay reviews of Egyptian antiques[9]. They showed that the approach was scalable and more specifically, that a small amount of domain knowledge could greatly improve the prediction accuracy compared against traditional instance-based learning approaches.

Titov et al. proposed a statistical model for sentiment summarization[10], which was a joint model of text and aspect ratings. To discover the corresponding topics, this model used aspect ratings; thus, it was able to extract textual evidence from reviews without the need of annotated data.

Pham et al. proposed a new method based on the least-squares model using both known aspect ratings and the overall rating of reviews to identify the overall aspect weights directly from numerous consumer reviews[11]. This method estimated the score of all aspects but there is not content about the evaluation viewpoint within a review.

As previously stated, there are several studies in which reviews have been analyzed. However, no study exists in which efforts have been made to develop a method for the automatic scoring of review texts and for proposing tags for cosmetic items.

### III. EVALUATION EXPRESSION DICTIONARY FOR COSMETIC REVIEW RECOMMENDATION

In this section, we provide an overview of our cosmetic review recommender system and the method used to construct a evaluation expression dictionary.

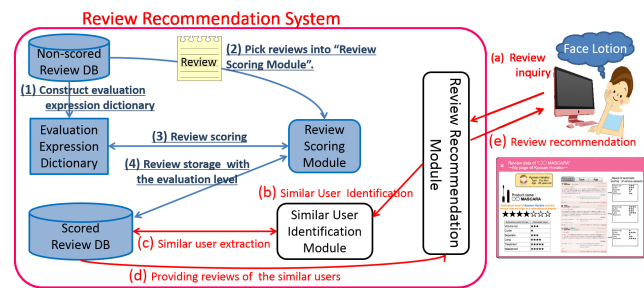


Fig. 1. Conceptual diagram of the cosmetic item review recommender system

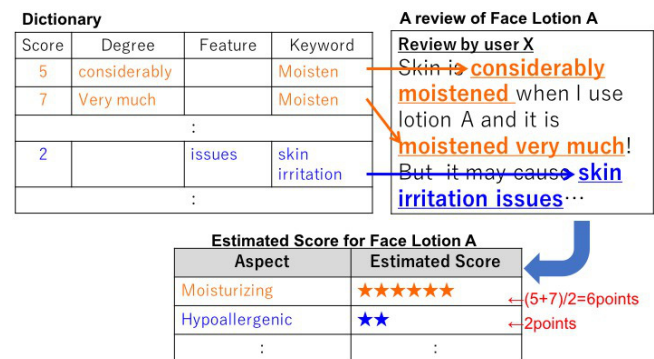


Fig. 2. Review scoring method based on the evaluation expression dictionary

#### A. Overview of the cosmetic review recommender system

The aim of the cosmetic review recommender system is to provide recommendations of truly useful reviews for each target user. Fig.1 shows a conceptual diagram of the cosmetic item and review recommender system, which is the final goal of our research. In Fig.1, the blue numbers from (1) through (4) correspond to the automatic scoring process of the cosmetics review. The red Roman numerals from (i) through (v) correspond to the review recommendation process. More detailed procedure of each process are shown below:

#### Cosmetic Review Automatic Scoring Process

- (1) Construct an evaluation expression dictionary which includes pairs of evaluation expression and its score by analyzing reviews sampled from non-scored DB.
- (2) Pick up reviews from non-scored DB to score them.
- (3) Automatically score reviews picked up in step (2) based on the evaluation expression dictionary constructed in step (1).
- (4) Put reviews scored in step (3) into a scored review DB.

#### Review Recommendation Process

- (a) User gives the name of a cosmetic item that he or she is interested in.
- (b) System Refers to the “similar user identification module” in order to extract similar users to the target user of step (a).
- (c) “Similar user identification module” obtains the information of reviews and reviewers and identify similar users to the target user.
- (d) Provide reviews of similar users identified in step (c) to the “Review recommendation module”.

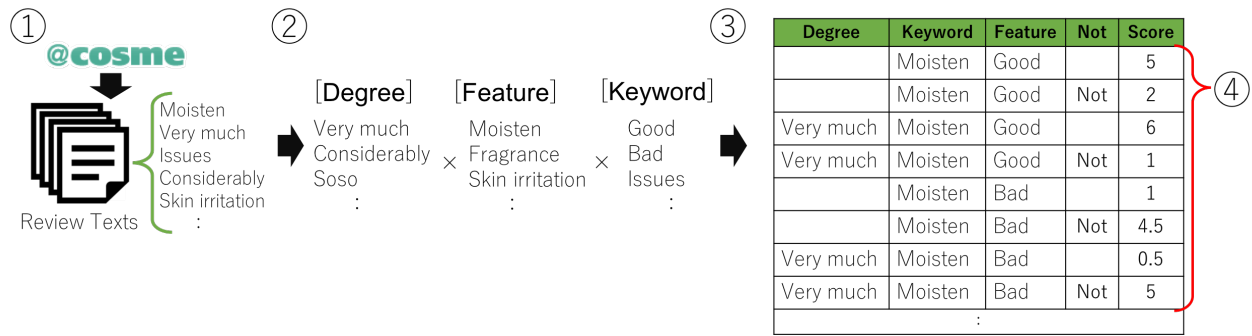


Fig. 3. Constructing the Co-occurrence Keyword-based Evaluation Expression Dictionary

(e) System recommends suitable reviews to the target user.

Our system calculates the scores for various aspects to judge the similarity of users and to recommend cosmetic reviews. Fig.2 shows the review scoring method for non-scored reviews based on the evaluation expression dictionary. The system reads a non-scored review text and identifies evaluation expressions existing in the reviews. Afterwards, the system gives a score to each evaluation expression according to the dictionary. In this step, if there are several expressions for one aspect, the system calculates the average score for the aspect(Fig.2). For example, in Fig.2, two expressions for “moisturizing aspect” exist in the review and our system calculates the average score as the score of “moisturizing aspect”.

#### B. Constructing the evaluation expression dictionary

Fig.3 shows the method used to construct the evaluation expression dictionary. To construct the evaluation expression dictionary, we utilized review data provided by @cosme. First, we analyzed phrasal evaluation expressions extracted from reviews. Second, we divided the phrasal expressions into aspect keywords, feature words and degree words. Finally, we constructed the dictionary by assembling their co-occurrence relations and the evaluation scores. Overall, there are several types of cosmetic items. As a first step, we sought to construct an evaluation expression dictionary of the “Face lotion” aspect because “Face lotion” is used by numerous people. In addition, various evaluations may exist for only one “Face lotion” product owing to differences in the skin types of the users. As shown in Fig.4, the phrases “It is easy to get a smooth skin” and “It can easily make your skin smooth” are semantically nearly identical; however, as a phrase, each of them is different from one another. Hence, it may be possible to detect more evaluation expressions based on the co-occurrence keyword-base dictionary than based on the phrase-based dictionary. Therefore, we use the co-occurrence keyword-base dictionary.

By using this procedure, we developed evaluation expression dictionary for “face lotion” manually[2]. This dictionary stores 1,332 evaluation expressions for “face lotion”. Then, in order to verify the effectiveness of our proposed method, we evaluated the results of the automatic scoring based on the dictionary. As the results, our system scores with around 81% accuracy. However, as mentioned in Section I, it is not a good idea to develop the dictionaries manually because of it takes lot of man-hours.

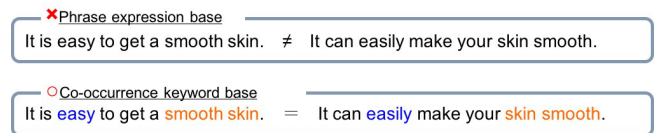


Fig. 4. Differences in detecting the evaluation expression between the Phrase expression base and the Co-occurrence keyword base

#### IV. FEATURE ANALYSIS OF EVALUATION EXPRESSION FOR EACH CATEGORY OF COSMETIC ITEMS

In this section, we describe the features analysis of evaluation expressions for each cosmetic category. In order to construct an evaluation expression dictionary efficiently, we analyze the relationship between the purpose of use and the evaluation expressions.

##### A. Target datasets for the analysis

We already constructed an evaluation expression dictionary for “face lotion”. Aiming to construct new dictionaries efficiently by using already constructed dictionary, we analyzed the evaluation expressions of “emulsion” which is the same intended use of “face lotion” like “moisturizing”. Furthermore, we analyze the “cleansing and face wash”. Intended use of “cleansing and face wash” is different from “face lotion”, but the large cosmetic category is the same (skin care). Thus, we consider the possibility of expansion of the dictionary for “face lotion” to the “skin care” category. In order to expand the dictionaries to all the cosmetic categories, we analyze the “Lips” and “Blush” which are coloring purpose items. According to these analysis, we will discuss an automatic construction policy of evaluation expression dictionary for all cosmetic categories.

For this analysis, we use review texts of over 50 characters for each cosmetic categories. We gather these 241 review texts, then extract characteristic expressions for each category by focusing on the frequent appearance of words of verb, adjective, noun, and adverb. Detailed information of target review data are shown in TABLE II.

##### B. Results of the analysis

In order to analyze the features of each cosmetic category, we use two cases (CASE 1, CASE 2) to calculate the similarity of frequently appearing words. For calculating the similarity, we adopt the cosine similarity metric.

###### CASE 1



TABLE II  
THE NUMBER OF REVIEWS AND FREQUENT APPEARING WORDS FOR EACH COSMETIC CATEGORY

Category	Cleansing	Face wash	Face lotion	Emulsion	Lips	Blush
number of reviews	42	40	41	40	38	40
number of frequent appearing words	45	51	43	47	43	42

TABLE III  
PART OF APPEARANCE COUNT FOR FREQUENTLY OCCURRENCE WORDS FOR EACH COSMETIC CATEGORY

	Cleansing	Face wash	Face lotion	Emulsion	Lips	Blush
Freshen	3	7	4	4	0	0
Pearl	0	0	0	0	1	5
Dry	6	7	4	9	3	5
Mat	0	0	0	0	1	2
:	:	:	:	:	:	:

TABLE IV  
CASE 1 : THE RESULTS OF COSINE SIMILARITY BETWEEN DIFFERENT COSMETIC CATEGORIES

	Cleansing	Face wash	Face lotion	Emulsion	Lips	Blush
Cleansing	-	0.51	0.18	0.23	0.25	0.19
Face wash	0.51	-	0.28	0.31	0.08	0.08
Face lotion	0.18	0.28	-	0.84	0.09	0.12
Emulsion	0.23	0.31	0.84	-	0.11	0.14
Lips	0.25	0.08	0.09	0.11	-	0.86
Blush	0.19	0.08	0.12	0.14	0.86	-

TABLE V  
PART OF THE PRESENCE OR ABSENCE(0 OR 1) OF FREQUENTLY OCCURRING WORDS FOR EACH COSMETIC CATEGORY

	Cleansing	Face wash	Face lotion	Emulsion	Lips	Blush
Freshen	1	1	1	1	0	0
Pearl	0	0	0	0	1	1
Dry	1	1	1	1	1	1
Mat	0	0	0	0	1	1
:	:	:	:	:	:	:

First, we create 129 dimensional feature vectors based on the frequency of appearing words extracted from all review texts. Then, we calculate the similarities between each cosmetic category. TABLE III shows a part of the appearance count for frequently occurrence words of each cosmetic category. And TABLE IV shows the results of cosine similarity between different cosmetic categories.

According to the results of CASE 1, we can find the cosmetic categories that have the most similar feature vectors of frequently appearing words. Thus, we can estimate that categories, such as "Cleansing and Face wash", "Face lotion and Emulsion", and "Lips and Blush", are similar cosmetic categories that contain similar evaluation expressions. By using the analysis method in CASE 1, we can construct the dictionaries for similar purpose categories.

#### CASE 2

In CASE 2, we create dimensional feature vectors based on the presence or absence(0 or 1) of appearing words extracted from all review texts. Then, we calculate the similarities between each cosmetic category, in order to find similar relationships other than "Cleansing and Face wash", "Face lotion and Emulsion", and "Lips and Blush". TABLE V shows a part of the presence or absence(0 or 1) of frequently occurring words for each cosmetic category. TABLE VI shows the results of cosine similarity between different cosmetic categories in CASE 2. In addition, TABLE VII shows Concordance rates of frequent words between different cosmetic categories.

The results of CASE 2 indicated that there are many common words in reviews for "Cleansing", "Face wash",

"Face lotion" and "Emulsion". Thus, we may say that our analysis method in CASE 2 can estimate similarities between different cosmetic categories.

#### C. Construction Policy of Evaluation Expression Dictionaries

TABLE VIII shows the frequent occurrence words in the high similarity categories of the four skin care categories, such as "Cleansing and Face wash", and "Face lotion and Emulsion" category. We discuss the construction policy of evaluation expression dictionaries based on this result. We aim to share the dictionary between categories as much expressions as possible.

The expressions, "moisture", "freshen", "hypoallergenic", "rough skin", "aging care", and "cost-performance" occurred commonly with four "Skin care" categories. The expressions that related to Foam" and "Tautness" occurred commonly with "Cleansing and Face wash" categories. The expressions that related to Refreshing or Thickening" occurred commonly to "Face lotion and Emulsion" categories. According to above results, we can find the evaluation expressions that can be shared among the categories(Fig.5). We can also find this fact from the concordance rate of frequency words between categories shown in TABLE VII.

TABLE IX shows the frequently occurring words in the high similarity categories between "Lips", "Blush", and "Skincare" categories. The expressions that related to "Moisturizing", "Cost-performance" occurred in common to all six categories(Lips, Blush, and Skin care). The expressions related to "coloring", "Color holding", "lame/pearl", and

TABLE VI  
CASE 2 : THE RESULTS OF COSINE SIMILARITY BETWEEN DIFFERENT COSMETIC CATEGORIES

	Cleansing	Face wash	Face lotion	Emulsion	Lips	Blush
Cleansing	-	0.58	0.41	0.48	0.34	0.32
Face wash	0.58	-	0.51	0.47	0.28	0.28
Face lotion	0.41	0.51	-	0.65	0.35	0.28
Emulsion	0.48	0.47	0.65	-	0.38	0.34
Lips	0.34	0.28	0.35	0.38	-	0.49
Blush	0.32	0.28	0.28	0.34	0.49	-

TABLE VII  
CONCORDANCE RATES OF FREQUENT WORDS BETWEEN DIFFERENT COSMETIC CATEGORIES

	Cleansing	Face wash	Face lotion	Emulsion	Lips	Blush
Cleansing	-	55%	42%	47%	35%	33%
Face wash	62%	-	56%	49%	30%	31%
Face lotion	40%	47%	-	62%	35%	29%
Emulsion	49%	45%	67%	-	40%	36%
Lips	33%	25%	35%	36%	-	50%
Blush	31%	25%	28%	32%	49%	-

TABLE VIII  
FREQUENT WORDS OF “SKINCARE”

Category	Cleansing, Face wash, Face lotion, Emulsion	Cleansing, Face wash	Face lotion, Emulsion
Frequent words	Moisturizing, Refresh, Allergies, Aging, Cheap, Value, Smooth, Dry, Expensive, Stimulation, Moisture, Price, Rough skin	Clear, Smooth, Dirty, Wash, Moderate, Washing, Tautness, Transparency, Buying, Foam, Removing	Refreshing, Thickening, Plump, Rough, High price, Penetration, Sensitive, Wrinkle

TABLE IX  
FREQUENT WORDS OF “LIPS”, “BLISH” AND “SKINCARE”

Category	cleansing, face wash, face lotion, lips, blush	lips, blush
Frequent words	Moisturizing, Dry, Cheap, Value	Lustrous, Mat, Pearl, Lane, Blood, Color, Coloring, Coloring, Shades, Petit Price



Fig. 6. Common words in the “Lips”, “Blush”, and “Skin care” categories

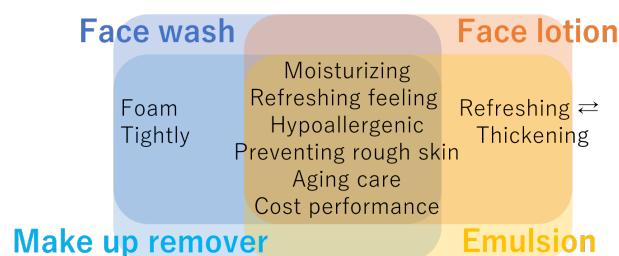


Fig. 5. Common words in the “Skin care” categories

“shiny or matte” occurred commonly to the “Lips and Blush” categories. According to the common words described above, we can find the evaluation expressions that can be shared among the categories(Fig.6).

Consequently, the results and the considerations given above indicate that there are both similar evaluation expressions and dissimilar evaluation expressions between reviews for skincare items and reviews for makeup items. Therefore, we found that we can utilize common evaluation expressions and aspects among different cosmetic categories in order to construct each evaluation expression dictionary.

## V. CONCLUSION

The aim of our research is to develop a recommender system for cosmetic items and reviews to help consumers. In order to realize such a recommender system, we proposed and discussed how to develop a method for automatic construction of evaluation expression dictionary for all categories by analyzing the tendency of both similar evaluation expressions and dissimilar evaluation expressions between reviews for skincare items and reviews for makeup items.

As the result, we recognized that evaluation expressions differ depending on their cosmetic item categories. Moreover, it was found that we can utilize common evaluation expressions and aspects among different cosmetic categories in order to construct each evaluation expression dictionary.

A further direction of this study will be to develop evaluation expression dictionaries for all cosmetic item categories and to work through developing a recommender system for cosmetic items and reviews.

## ACKNOWLEDGMENT

This work was supported in part by istyle Inc. who provided review data for cosmetic items.

## REFERENCES

- [1] @cosme, <http://www.cosme.net/>, (Accessed on 8th January 2019)
- [2] Yuki Matsunami, Mayumi Ueda, Shinsuke Nakajima, Takeru Hashikami, John O'Donovan, and Byungkyu Kang, "Mining Attribute-Specific Ratings from Reviews of Cosmetic Products", *Transactions on Engineering Technologies (International MultiConference of Engineers and Computer Scientists 2016)*, pp.101-114, Springer, 2017.
- [3] Asami Okuda, Yuki Matsunami, Mayumi Ueda, and Shinsuke Nakajima, "Finding similar users based on their preferences against cosmetic item clusters", *Proceedings of the 19th International Conference on Information Integration and Web-based Applications & Services*, pp.156-160, 2017.
- [4] Amazon.com, <http://www.amazon.com/>, (Accessed on 8th January 2019)
- [5] Priceprice.com, <http://ph.priceprice.com/>, (Accessed on 8th January 2019)
- [6] Byungkyu Kang, et al., "Inspection Mechanisms for Community-based Content Discovery in Microblogs" *IntRS'15 Joint Workshop on Interfaces and Human Decision Making for Recommender Systems at ACM Recommender Systems*.
- [7] The site data of @cosme (May.2017), istyle Inc., <http://www.istyle.co.jp/business/uploads/sitedata.pdf> (in Japanese), (Accessed on 8th January 2019)
- [8] Hiroshi Kanayama, et al., "Textual demand analysis: detection of users' wants and needs from opinions." *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2008.
- [9] John O'Donovan, et al., "Extracting and Visualizing Trust Relationships from Online Auction Feedback Comments.," *International Joint Conference on Artificial Intelligence (IJCAI'07)*, 2007.
- [10] Ivan Titov et al., "A Joint Model of Text and Aspect Ratings for Sentiment Summarization," *46th Meeting of Association for Computational Linguistics (ACL-08)*, Columbus, USA, pp.308-316, 2008.
- [11] Duc-Hong Pham, et al., "A least square based model for rating aspects and identifying important aspects on review text data," *the 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science*, pp.265-270, 2015.