

# User Similarity Calculating Method for Cosmetic Review Recommender System

Yuki Matsunami, Asami Okuda, Mayumi Ueda, and Shinsuke Nakajima

**Abstract**—Most of shopping web sites support a function of user-provided item reviews. It has been shown that reviews have a profound effect on item conversion rates. Cosmetic item reviews carry particular importance in purchasing decisions because of their personal nature, and particularly because of the potential for irritation with unsuitable items. In this paper, we try to develop a user similarity calculating method for a cosmetic review recommender system. To realize such a recommender system, we propose a method for automatic scoring of various aspects of cosmetic item review texts based on an evaluation expression dictionary curated from a corpus of real world online reviews. Furthermore, we consider how to calculate user similarity of cosmetic review sites.

**Index Terms**—explanation, automatic review rating, evaluation expression dictionary, cosmetic review, user similarity

## I. INTRODUCTION

IN recent years, Most of shopping web sites have added support for user-provided reviews. These are very helpful for consumers to decide whether to buy a commercial item, and they have been shown a significant impact on conversion rates. In particular, consumers make careful choices about cosmetics since unsuitable items frequently cause skin irritations. “@cosme”[1] is a cosmetics review site that is very popular among Japanese young women. While the site can be helpful for their decision making, it is not easy to find truly suitable cosmetics because of the lack of explanation and granularity in user’s reviews and scores of cosmetic items. As an example, there is no guarantee that a cosmetics item mentioned as good for dry skin is always suitable for people who have dry skin. Since the compatibilities between skin and cosmetics items differ from one user to another, we believe that it is important to identify users who share common preferences for cosmetic items and to share reviews among those niche communities. In order to realize the proposed approach, 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 item reviews.

To develop the review recommender system, our system adopts the scores by the similar users. We have proposed a basic concept of automatic scoring method of various aspects of cosmetic items in our previous work[2].

Therefore, in this paper, we propose the user similarity calculating method for cosmetic review recommender system.

Manuscript received January 8, 2017; revised February 2, 2017. This work was supported in part by the MEXT Grant-in-Aid for Scientific Research(C) (#26330351, #16K00425).

Y. Matsunami, A. Okuda and S. Nakajima are with Faculty of Computer Science and Engineering, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto-city, Kyoto, 603-8555, Japan, e-mail: ({g1245108, g1444341, nakajima}@cc.kyoto-su.ac.jp).

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).

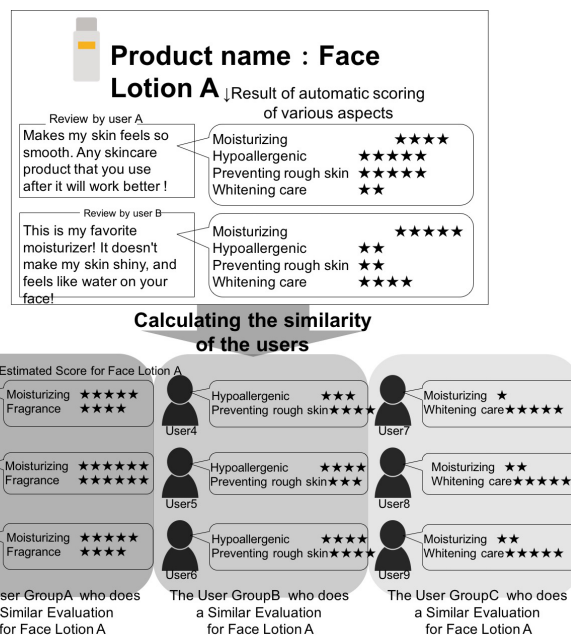


Fig. 1. Example of automatic scoring and user clustering

In particular, we realize the system by using the following method. (see Fig.1)

- 1) Automatic scoring of various aspects of cosmetic item (Section III)
- 2) User similarity calculating method considering various aspects of cosmetic item (Section IV)

The remainder of this paper is organized as follows. The related work is given in Section II. Then Section III describes the method for automatic scoring of various aspects of cosmetic item review texts based on evaluation expression dictionary. In Section IV, we show the User similarity calculating method. We conclude the paper in Section V.

## II. RELATED WORK

There are many websites to provide reviews by the consumers. For example, Amazon.com[3] and Priceprice.com[4] are popular shopping sites over the Internet, and these sites provide reviews of their merchandise by the consumers. And “Tabelog” is also popular website in Japan. This website does not sell products, it provides restaurant’s information and reviews. In addition to the algorithmic aspects, researchers have recently focused on the presentation aspects of review data [11]. Furthermore, in recent years, “@cosme” is very popular among Japanese young women. This website is a portal site for beauty and cosmetic items, and it provides various information, such as reviews and shopping information of cosmetic items. According to the report by the istyle

Inc. that is a operation company of this system, at October 2016, the number of monthly page view is 280 million, the number of member is 3.9 million, and the total number of review is 1300 million[5]. From this report, many women exchange information about beauty and cosmetics through the service of @cosme. At the service of @cosme, they provide many information about cosmetic items of various cosmetic brands. Hence, users can compare cosmetic items through the various cosmetic brands. Reviews are composed of review text, scores, tag about effects, etc. Furthermore, the system has profile data that includes information about age and skin type, made by the users when they enroll as a member. Therefore, users who want to browse the reviews can search the reviews according to their own purposes, for example, reviews sorted by the scores or focused on one effect.

Along with the popularization of these review services, several researches about analysis of reviews have been conducted in the past. For example, O'Donovan et al. evaluated their AuctionRules algorithm –a dictionary-based scoring mechanism for eBay reviews of Egyptian antiques. They showed that the approach was scalable and particularly that a small amount of domain knowledge can greatly improve prediction accuracy compared against traditional instance-based learning approaches. In our previous study, we analyze reviews of the cosmetic items[6]. In order to determine if the review is positive review or negative review, we make dictionaries for the Japanese language morphological analysis, which composed of positive expression and negative expression of cosmetic items. This previous research is aimed to develop the system to provide the reviews that take account of the user's profile, then, that system tries to retrieve information from blogs and SNS, and merge the information to the same format. Our final goal of current study is to develop a method for automatic scoring of review texts, according to various aspects of cosmetic items.

Nihongi et al. propose a method for extracting the evaluation expression from the review texts, and they develop the product retrieval system using evaluation expressions[7]. Our research focuses on the analysis of the review for cosmetic items, and we are aimed to find similar users about preferences and feelings in order to recommend truly useful reviews.

Yao et al. conduct a questionnaire to examine the impact of reviews to their purchasing behavior[8]. In order to investigate that what kind of reviews are trusted, they conduct a hypothesis verification. The results obtained from hypothesis verification are contrary to their intention, but the results obtained from the variance analysis show that reviews have impacts to the purchasing behavior.

Titov et al. propose a statistical model for sentiment summarization[9]. This model is a joint model of text and aspect ratings. In order to discover the corresponding topics, this model uses aspect ratings. Therefore, this model is able to extract textual evidence from reviews without the need of annotated data.

Nakatsuji et al. analyze frequency of the description expression to depend on an item of review. And calculate the similarity users in consideration of tendency of the description expression[12]. It is difficult to use this way if calculate the similarity of the users of cosmetic items because

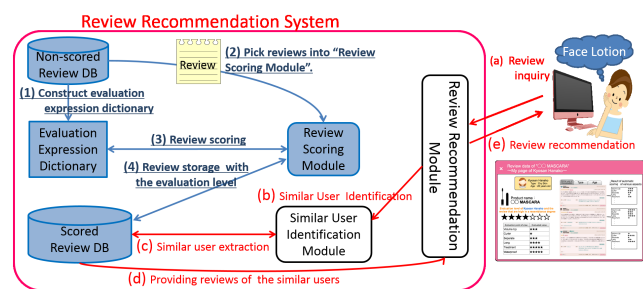


Fig. 2. Conceptual diagram of the cosmetic item review recommender system

Since the compatibilities between skin and cosmetics items differ from one user to another.

As stated above, there are several studies to analyze reviews. However, there has been no study that tried to develop a method for automatic scoring of review texts, according to various aspects of cosmetic items and calculating the similarity of the users.

### III. AUTOMATIC SCORING OF VARIOUS ASPECTS OF COSMETIC ITEM REVIEW TEXTS

In this section, we describe a method for automatic scoring of various aspects of cosmetic item review texts based on evaluation expression dictionary. At first, we describe the brief overview of our proposed method in section III-A. The method for automatic scoring of various aspects of cosmetic item review texts is given in section III-B. Section III-C describes an experimental evaluation of the automatic scoring method using real review data.

#### A. Overview of proposed method

Our final goal is to develop a cosmetic item review recommender system which can recommend truly useful reviews for a target user. It operates in a similar manner to collaborative filtering, by using a set of similar users who have common both preferences and feedbacks on their experiences of the cosmetic items.

In order to make a significance of our study clear, Fig.2 shows a conceptual diagram of the cosmetic item review recommender system, which is our final goal. In Fig.2, numbers in blue written as (1) - (4) are corresponding to the procedure of cosmetic review automatic scoring process, and Roman alphabets in red written as (a) - (e) are corresponding to the procedure of review recommendation process. More detailed procedures are shown below:

#### Automatic Scoring

- (1) Construct the 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 scored review DB.

#### Review Recommendation Process

- (a) User provides the name of a cosmetic item that she is interested in.
- (b) System Refers to “similar user extraction module” in order to extract similar users to the target user of step (a).
- (c) “Similar user extraction module” obtains the information about reviews and reviewers, and identifies similar users to the target user.
- (d) Provide reviews of the similar users identified in step (c) to “Review recommendation module”.
- (e) System recommends suitable reviews to the target user.

This paper focuses on development of the dictionary-based approach. Developing the live review recommendation method is part of our future work.

### B. Automatic scoring based on evaluation expression dictionary

We describe about automatic scoring based on evaluation expression dictionary in this section. Fig.3 describes a conceptual diagram of constructing the co-occurrence keyword-based dictionary. The procedure of constructing the dictionary is as follows:

- 1) Analyze phrasal evaluation expressions extracted from reviews.
- 2) Divide the phrasal expressions into aspect keywords, feature words and degree words.
- 3) Construct the dictionary by assembling their co-occurrence relations and the evaluation scores.

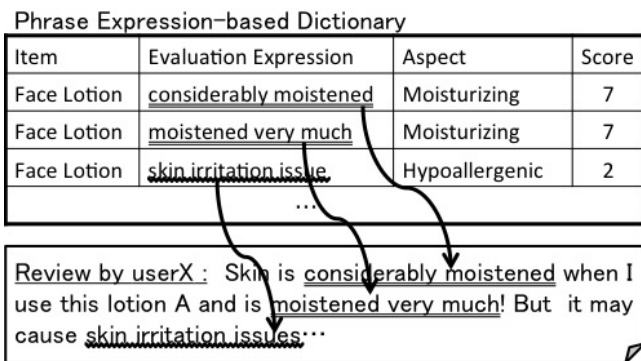


Fig. 3. Review scoring using phrase expression-based dictionary

The procedure of automatic scoring against non-scored reviews based on the evaluation expression dictionary is shown below:

- 1) Acquire a review text data.
- 2) Read a review text from the outside.
- 3) Judge the end of the sentence using punctuation mark when did morphological analysis.
- 4) Extract the sentence of the k unit from one review text.
- 5) Judge the sentence include keyword.
- 6) Pick up the sentence to satisfy co-occurrence condition from sentence including the keyword.
- 7) Survey the evaluation expression and score correspond co-occurrence condition. And calculate score every an evaluation expression to review texts.

### C. Experimental evaluation of automatic scoring using real review data

We examine an experimental evaluation of the automatic scoring method using real review data in order to verify the effectiveness of our proposed method. As a first step, we analyze 5,000 reviews randomly extracted from review data for “face lotion” posted at @cosme, for understanding characteristics of the data.

a) Procedure of experimental evaluation: In this experiment, we use 10 review data for “face lotion” randomly selected from 5,000 reviews as described above, and then compare results by the following methods:

- Manual scoring method without the Dictionary (as ground truth data).
- Automatic scoring method based on the evaluation expression dictionary (proposed method).

In the case of the manual scoring, evaluators actually read review texts and score them between 0 to 7 stars for 10 aspects of “face lotion” set in advance. The evaluators are 30 people. They are 20’s to 50’s females.

In the case of automatic scoring, the method scores review texts between 0 to 7 stars for the 10 aspects based on co-occurrence keyword-based dictionary.

The 10 aspects for “face lotion” set for the experiment in advance are Cost performance, Moisturizing, Whitening care, exfoliation & Pore care/Cleansing effect, Refreshing feeling/Preventing sebum shine, Refreshing↔Thickening, Hypoallergenic, Preventing rough skin, Aging care and Fragrance.

b) Result of the Experiment: Fig.4 describes results of review scoring against 10 reviews based on co-occurrence keyword-based dictionary and manual scoring by evaluators.

The contents of 10 reviews are different from each other, so that the detected aspects are different. The average score (# of stars) of all aspects by manual scoring is 4.76, and the average score by our proposed method is 4.32. The mean absolute error (MAE) is 1.20.

Scores of the manual scoring tend to a little higher than scores of proposed scoring method. However, the range of the score is from 0 to 7 and MAE is 1.20, so that we may say that the results of automatic scoring by proposed method are quite close to the results of manual scoring as ground truth data. Total number of detected aspects are 31 aspects by manual scoring (ground truth) and 22 aspects by automatic scoring (proposed method). Therefore, the achievement rate of our proposed method against manual scoring is about 71%.

There are several “N/A” in Fig.4 by automatic scoring method. However, there is room for improving the result of aspect detection for reviews by updating the dictionary. In future work we will try to analyze larger number of reviews, and then improve and tune the dictionary. Moreover, we will develop a review recommender system for cosmetic items to further evaluate our novel scoring method.

Next, we focus on MAE of automatic review scoring. Fig.5 show result of review scoring for each evaluation aspect. The range of the score is from 0 to 7. Most MAE scores achieve results under 1.30. But MAE of “Hypoallergenic” achieve results more than 2.00. It is caused by scoring expression about problem of the quality of user’s skin like “My skin easily feels stimulation because I have a sensitive skin,

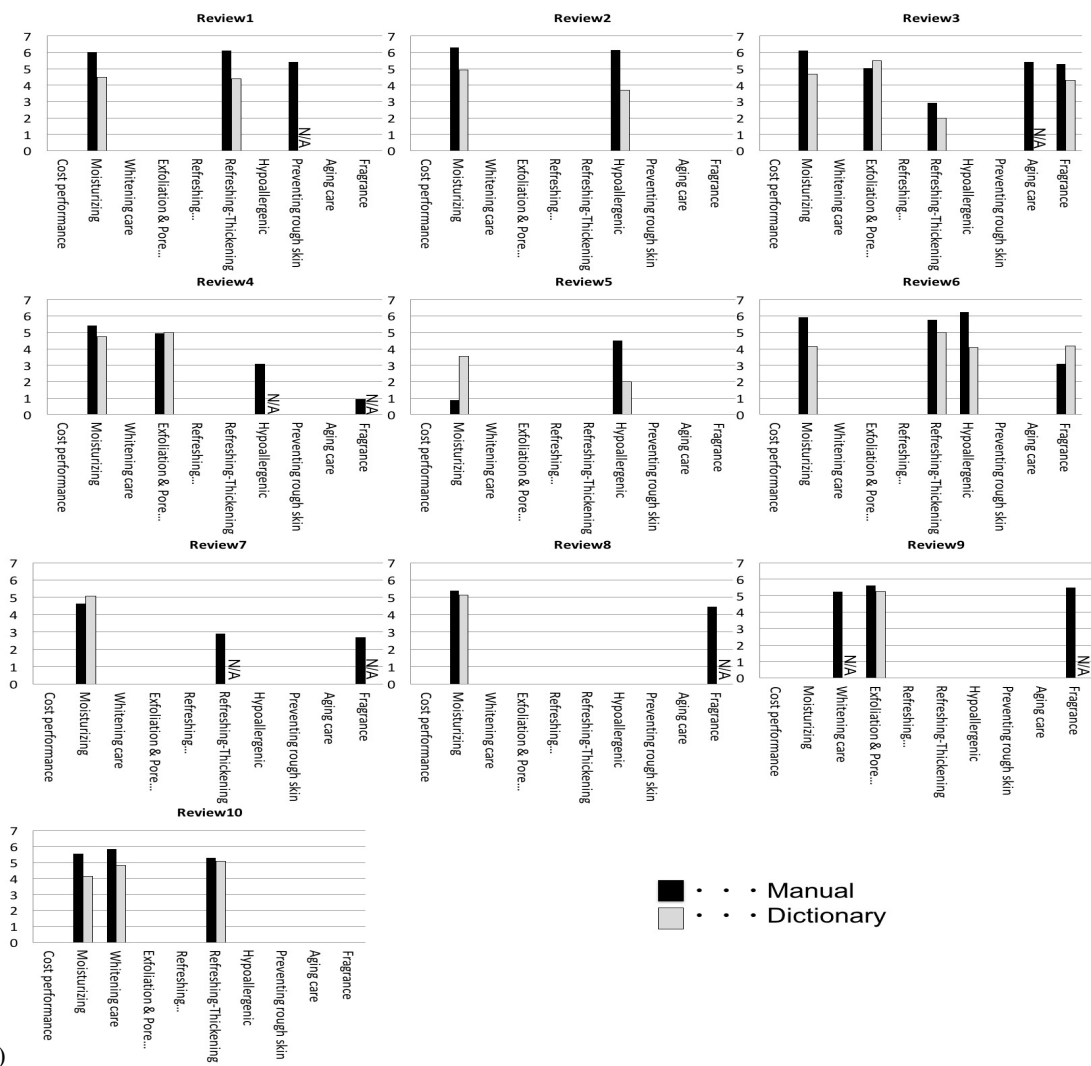


Fig. 4. Result of review scoring based on co-occurrence keyword-based dictionary

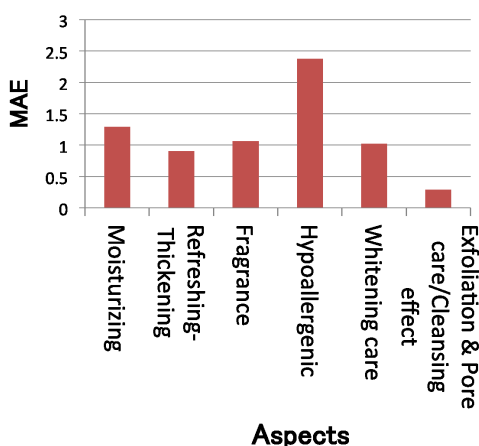


Fig. 5. MAE of review scoring for each aspect (except N/A)

but this lotion is good for me!”. We have to remove the noise evaluation expression using a correlative conjunction. In future work, we will try to analyze larger number of reviews, and also try to improve the expression dictionary. Moreover, we will develop a review recommender system for cosmetic items to further evaluate our novel scoring method.

#### IV. USER SIMILARITY CALCULATING METHOD FOR COSMETIC REVIEW RECOMMENDER SYSTEM

In this section, we describe the method for calculating the similarities of the users. At first, we introduce the overview of the proposed method in section IV-A. Secondly, we describe the method for clustering cosmetic items based on their rating similarity, in section IV-B. Finally, we show the method for user similarity calculating method in section IV-C.

##### A. The overview of the method

The purpose of our research is to develop cosmetic review recommender system based on providing the effective reviews for each user. In order to realize such recommender system, our system has to estimate the similar user to extract reliable ratings for the items. In this section, we describe the method to estimate the similar user using the rating for similar item clusters.

By using existing review providing services, users can get the reviews by the users who are similar skin type or similar ages. However, these reviews are not truly effective reviews for each users.

For example, we suppose that “Item B” has high esteems by user1, user2 and user3(Fig. 6). User1 and user2 post high score for the “Moisture” aspect of Item B, and low



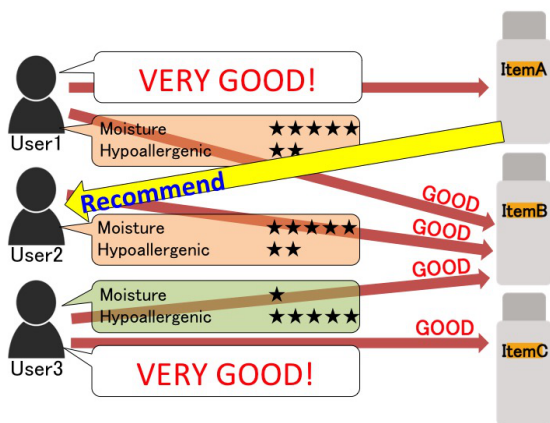


Fig. 6. Conceptual diagram of our recommendation method

score for the “Hypoallergenic” aspect. And User3 post high score for the “Hypoallergenic” aspect, and low score for the “Moisture” aspect. Because of the similarity of the focused aspects, we consider that the User2 would like to know the recommended item, it is deemed desirable to recommend Item A for User2, which post high score by the similar user.

### B. Clustering cosmetic items based on their rating similarities

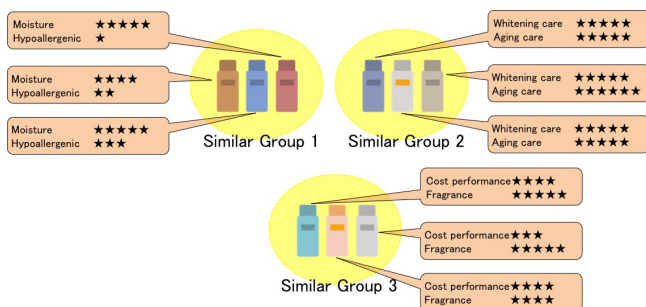


Fig. 7. Clustering cosmetic items based on their rating similarities

We consider that similar user is the user who post similar ratings for the same items. However, because there are huge number of the cosmetic items, it is difficult to find the user who post similar ratings for the same items. Therefore, we make clusters of similar items, then estimate the users who post similar ratings for the items inside the cluster as similar users. The clustering method is as follow(Fig.7).

- 1) Give scores of various aspects for the reviews.
- 2) Calculate the average scores of each item for each aspect.
- 3) Make clusters of cosmetic items by the similarity of the scores.

In order to cluster cosmetic items, we adopt the k-means clustering method.

### C. User similarity calculating method

In this section, we describe a method for calculating user-similarities. In most cases of recommender system using collaborative filtering, the ratings matrices are utilized to understand relationships between users and items. When the number of items and users corresponds to  $m$  and  $n$ , the size

of the ratings matrix become  $m \times n$ . So, it is possible to calculate user similarities based on the Pearson correlation coefficient using the matrix.

However, in our method, each item has not only one score but multiple scores against aspects of the items. We proposed 10 aspects for face lotion. Thus, when the number of scores and users corresponds to  $p$  and  $n$ , the size of the ratings matrix become  $p \times n$ . In this case,  $p$  corresponds to ten times of  $m$ . By using the ratings matrix  $p \times n$ , our method can consider a correlation coefficient based on not a comprehensive evaluation score but several evaluation scores against each item, in order to calculate user similarity. As a result, we believe that our method can achieve higher recommendation efficiency than conventional method.

## V. CONCLUSIONS

In this paper, we propose a system that can recommend the truly effective reviews of cosmetic items for each user. In order to realize such a recommender system, at first, we proposed a method for automatic scoring of various aspects of cosmetic item review texts based on an evaluation expression dictionary. To develop the automatic scoring method, we constructed a dictionary using co-occurrence keyword-based evaluation expressions. Secondly, we proposed a User Similarity calculating method for cosmetic review recommender system using the rating of the similar cosmetic item clusters.

## ACKNOWLEDGMENT

This work was supported in part by istyle Inc. for providing review data against cosmetic items.

## REFERENCES

- [1] @cosme, <http://www.cosme.net/>, (Accessed 8 January 2017)
- [2] Yuuki Matsunami, Mayumi Ueda, Shinsuke Nakajima, Takeru Hashikami, Sunao Iwasaki, John O'Donovan, and Byungkyu Kang, Explaining Item Ratings in Cosmetic Product Reviews, International MultiConference of Engineers and Computer Scientists 2016, ICI-CWS2016 pp392-397, 2016.3.
- [3] Amazon.com, <http://www.amazon.com/>, (Accessed 8 January 2017)
- [4] Priceprice.com, <http://ph.priceprice.com/>, (Accessed 8 January 2017)
- [5] The site data of @cosme (Oct.2016), istyle Inc., <http://www.istyle.co.jp/business/uploads/sitedata.pdf> (in Japanese), (Accessed 8 January 2017)
- [6] Yumi Hamaoka, Mayumi Ueda and Shinsuke Nakajima, “Extraction of Evaluation Aspects for each Cosmetics Item to Develop the Reputation Portal Site,” *IEICE W12-2012-15*, pp.45-46, 2012.2. (in Japanese)
- [7] Tomohiro Nihongi and Kazuo Sumita, “Analysis and retrieval of the word-of-mouth estimation by structurizing sentences,” *Proceeding of the Interaction 2002*, pp.175-176, 2002.3. (in Japanese)
- [8] Jia Yao, Hiroki Idota and Akira Harada, “The affect of the Internet’s word of mouth on buying behavior: From the questionnaire survey pf the cosmetics purchase of the women students,” *Proceeding of National Conference of JASMIN Spring 2014*, Kanagawa, Japan, pp.231-232, 2014.5. (in Japanese)
- [9] Ivan Titov and Ryan McDonald, “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.
- [10] John O'Donovan, Vesile Evrim, Paddy Nixon and Barry Smyth, “Extracting and Visualizing Trust Relationships from Online Auction Feedback Comments.,” *International Joint Conference on Artificial Intelligence (IJCAI'07)*, Hyderabad, India, January 2007.
- [11] Byunkyu Kang, Nava Tintarev and John O'Donovan, “Inspection Mechanisms for Community-based Content Discovery in Microblogs” *IntRS'15 Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (http://recex.ist.tugraz.at/intRS2015/)* at *ACM Recommender Systems 2015*. Vienna, Austria. September 2015.
- [12] Makoto Nakatsuji, Mitsumasa Kondo, Akimichi Tanaka, Tadasu Uchiyama, “Measuring similarity of users using sentimental tags for items”, *The 24th Annual Conference of the Japanese Society for Artificial Intelligence*, 2010.