

Extraction of Harmful Ingredients of Cosmetics using User Comments

Hirotooshi HONMA¹ Yoko NAKAJIMA¹

Abstract: Recent years have witnessed web services drastically becoming popular in our daily lives, and many consumers take user reviews of products into account when planning purchases. The number of cosmetic review sites, users, and products posted have been increasing year by year. However, there are very few functions that prompt attention to purchase or use for products that may not be good for the user’s skin quality. Also, since the number of user reviews is small for new products, it is difficult to identify products that are not good for the user’s skin. In this study, we assume that there is common harmful ingredients causing skin disorder for each user group with similar skin quality and age. Using review information from @cosme and that from *Bihada-Mania* websites, we extract the harmful cosmetic ingredients for each user group.

Key words: Computer simulation, Natural language processing, Review information, Cosmetic ingredient information.

1 Introduction

Recent years have witnessed web services drastically becoming popular in our daily lives, and many consumers take user reviews of products into account when planning purchases. The number of studies on recommendation methods using website review information is increasing [1, 2, 3, 4]. The @cosme website¹ is the largest-scale website in the field of cosmetics in Japan. Many users use it for product selection and information gathering. In basic cosmetics, by mistakenly using a product incompatible with the individual skin of the user, there is a risk of skin disorder such as skin roughness, itching, and pain. However, there are very few functions that prompt attention to purchase or use for products that may not be good for the user’s skin quality. Also, since the number of user reviews is small for new products, it is difficult to identify products that are not good for the user’s skin. In this study, we assume that there is common harmful ingredients causing skin disorder for each user group with similar skin quality and age. Using review information from @cosme and that from *Bihada-Mania* websites², we extract the harmful cosmetic ingredients for each user group.

2 Websites to Use

The @cosme is the largest-scale website in the field of cosmetics in Japan and provides a user review posting and viewing service. Users registers their information such as age and skin quality at the time of registra-

tion. The user can refer to the user’s review of similar skin quality herself when purchasing the product. At @cosme, the user evaluates the recommendation degree for each product. It is evaluated by the user in 8 steps from ‘0’ to ‘7’ for the product. The image of a review of @cosme is shown in Fig. 1

Bihada-Mania has released ingredient information of cosmetics, and as of January 2018 information on 31,336 products has been released. The ingredients are listed in the order of content and a brief explanation of the ingredients is given. We use Bihada-Mania to acquire ingredient information because @cosme does not present them. Figure 2 shows an example of an ingredient list.

3 System Outline

The product categories covered in this research were skin lotion of basic cosmetics, which represent the largest number of products in cosmetics. We manually corrected the notation of brand name and ingredient name for 2,014 skin lotions that exist in both @cosme and Bihada-Mania sites.

All users set one skin quality and one age at the time of user registration. The user attribute in this research is defined by information of the set skin quality and age. For user attributes, we consider 30 classes composed of combinations of six skin types (normal, oily, mixed, dry, sensitive, and atopic skin) and 5 age groups (10s, 20s, 30s, 40s, 50s). In particular, we have experimented for two skin types of “atopic” and “sensitive”.

¹Department of Creative Engineering, National Institute of Technology, Kushiro College

¹@cosme: <http://www.cosme.net>

²Bihada-Mania: <http://www.bihada-mania.jp>



Figure 1: The review on @cosme

スキンコンディショナー(ハトムギ化粧水)の全成分

●成分名をクリックすると、その成分を含む他の化粧品を探すことができます。

成分名	成分分析
水	溶剤
BG	保湿剤、溶剤、粘度低下剤
グリチルリチン酸2K	香料、消炎作用
ハトムギエキス	油剤、保護剤、美白性
(スチレン/アクリルアミド)コポリマー	合成ポリマー、不透明化剤
エタノール	溶剤、可溶化剤、収れん剤
クエン酸	キレート剤、pH調整剤
クエン酸Na	緩衝剤、キレート剤
EDTA-2Na	キレート剤、殺菌防腐剤
メチルパラベン	防腐剤

Figure 2: The ingredient list on Bihada-Mania

3.1 Identification of Harmful Ingredients

We define representation of skin abnormality such as “Roughness”, “Swelling”, or “Aches” as *harmful expression*. We have manually listed 55 harmful expressions. Also, we define user comments meaning that skin abnormalities occurred by the use of products as *harmful comments*. Many users who cause skin abnormalities tend to evaluate the recommendation degree low. We extracted harmful comments using harmful expressions for user comments whose recommendation degree is 4 or less. In this study, products whose number of harmful comments accounts for 40% or more of the total number of comments are regarded as harmful products. In order to identify the characteristic ingredients (harmful ingredients) contained in harmful products, we utilize the IF-IDF [4] value applying the concept of TF-IDF.

IF (Ingredient Frequency) is the proportion of content of ingredient i included in the highly effective

cosmetics group X . IF is given by the following equation (3). Here, n_p is the number of ingredients included in product p in skin lotion group X , m is the number of products in X , and $\alpha_{p,i}$ is the rank of ingredient i listed in product p . In ingredient with high content ranking, the IF value becomes high.

$$IF(i, X) = \sum_{p=1}^m \frac{n_p - \alpha_{p,i}}{n_p} \quad (1)$$

IPF (Inverse Product Frequency) is the reciprocal of the number of products containing an ingredient. For example, “water” used in many products has a lower IPF value. IPF is given by equation (2). Here, N is the number of skin lotion products and $pf(i)$ is the number of products including ingredient i .

$$IPF(i) = \log \frac{N}{pf(i)} \quad (2)$$

IF-IPF is given by

Table 1: Harmful ingredients for atopic skin users

Age	Ingredient	IF-IPF
10s	Arginine Hydrochloride	1.9571
	Pionin 201	1.8868
	DL-PCA·NA	1.8313
20s	Xylitol	6.4714
	DPG	5.1271
	BG	4.1592
30s	Dipotassium Glycyrrhizinate	6.7842
	1,3-Butylene Glycol	4.4337
	DPG	4.3706
40s	Isononyl Isononate	2.1279
	Sulfated Castor Oil	1.9577
	Isostearyl Glyceryl Ether	1.9151

$$IF \cdot IPF(i, X) = IF(i, X) \times IPF(i) \quad (3)$$

Table 1 shows the results of specifying harmful ingredients by age for atopic skin users.

4 System Evaluation

We validate the user group of “atopic skin 20” with the largest number of reviews. Deprecated products are those that containing all of the IF-IPF top 3 components. Approximately 52.5% of users of atopic skin users who comment on products derived as deprecated products describe negative reviews. So, we think that the validity of this system is very high. In addition, our system also has the function of warning products containing any of IPF-IF top three harmful ingredients as deprecated products. For users with atopic skin 20s, there were 1,585 items of deprecated products.

5 Conclusion

In this study, we identified harmful ingredients of cosmetics and estimated deprecated products for each user group. We evaluated the implemented system using the harmful rate of review to the estimated deprecated products.

Extracting harmful expressions automatically and improving estimation accuracy of deprecated products are left for future research.

Acknowledgment

This work was partially supported by JSPS KAKENHI Grant Number 17K00324.

References

- [1] Y. Matsunami, M. Ueda, and S. Nakajima, “Proposal of a review analysis method aimed at judging the feeling of use and preference of cosmetics items,” DEIM Forum, P3-1, 2015.
- [2] S. Abe and I. Kobayashi, “Development of review recommender system using cosmetic review data,” SIG-DBS 2015-A-22, 2015.
- [3] Y. Matsunami, “Evaluation item review automatic scoring method using cosmetic item evaluation expression dictionary,” DEIM Forum, B1-1, 2016.
- [4] R. Iwabuchi, Y. Nakajima, and H. Honma, “Proposal of recommender system based on user evaluation and cosmetic ingredients,” Advanced Informatics, Concept, Theory, and Applications (ICAICTA), IEEE, 2017.