Print ISSN: 1738-3110 / Online ISSN 2093-7717
JDS website: http://kodisa.jams.or.kr/
http://dx.doi.org/10.15722/jds.18.10.202010.65

Determinants of Online Review Helpfulness for Korean Skincare Products in Online Retailing*

Yun-Kyung OH[1]

Received: August 18, 2020. Revised: September 09, 2020. Accepted: October 05, 2020.

Abstract

Purpose: This study aims to examine how to review contents of experiential and utilitarian products (e.g., skincare products) and how to affect review helpfulness by applying natural language processing techniques. **Research design, data, and methodology:** This study uses 69,633 online reviews generated for the products registered at Amazon.com by 13 Korean cosmetic firms. The authors identify key topics that emerge about consumers' use of skincare products such as skin type and skin trouble, by applying bigram analysis. The review content variables are included in the review helpfulness model, including other important determinants. **Results:** The estimation results support the positive effect of review extremity and content on the helpfulness. In particular, the reviewer's skin type information was recognized as highly useful when presented together as a basis for high-rated reviews. Moreover, the content related to skin issues positively affects review helpfulness. **Conclusions:** The positive relationship between extreme reviews and helpfulness of reviews challenges the findings from prior literature. This result implies that an in-depth study of the effect of product types on review helpfulness is needed. Furthermore, a positive effect of review content on helpfulness suggests that applying big data analytics can provide meaningful customer insights in the online retail industry.

Keywords : Big Data Analytics, Online Consumer Reviews (OCRs), Review Helpfulness, Product Type, Online Retailing

JEL Classification Code : M31, L81, L86

1. Introduction

Online consumer reviews (OCRs) benefit online distribution channels by providing useful information to prospective customers and attracting them (Mudambi & Schuff, 2010; Pavlou & Dimoka, 2006; Siering, Muntermann, & Rajagopalan, 2018). OCRs can be greatly persuasive in that the voices of general consumers, not companies, are reflected (Bickart & Schindler, 2001). Helpfulness of OCR measures the extent that a review can help a consumer's purchase decision. In addition, the usefulness of online information has a positive effect on online information acceptance, attitude, and intent to use it (Cheung & Thadani, 2012; HAN, 2020). However, as the number of OCRs increases, consumers may experience a difficult time identifying useful reviews. To overcome this limitation, many online retailers have adopted the helpfulness voting system. With this function, consumers can easily make purchase decisions easily by sorting OCRs in the order of helpfulness. Previous studies have noted that the review usefulness index plays an important role in filtering opinions efficiently by following consumers (Mudambi & Schuff, 2010).

In the last decade, research on the determinants of the OCRs' helpfulness has been growing, which allows us to broaden our understanding of the factors influencing helpfulness votes. One of the key research topics in this area is the effect of extreme sentiment on the review helpfulness, and controversy exists. Reviews with neutral polarity are perceived to be more helpful than extreme

* This study was supported by the Dongduk Women's University grant.
2 Associate Professor, Department of Business Administration, Dongduk Women's University, Seoul, Korea. Tel:+82-2-940-4471, Email:ykoh1@dongduk.ac.kr

polarity (Salehan & Kim, 2016). The mixed view can be perceived as credible when delivering the pros and cons of a product in a balanced manner (Crowley & Hoyer, 1994). On the other hand, very positive or negative reviews can be perceived as more helpful than those with mixed sentiment. This view is based on the theory that extreme information can help consumers reduce uncertainty by confirming positive aspects of a product or avoiding risks (Forman, Ghose, & Wiesenfeld, 2008).

Prior studies have documented how consumers' perception of review helpfulness varies across product types (Mudambi & Schuff, 2010; Pan & Zhang, 2011). Nelson (1970) argued that consumers value the information for experience goods more than that for search goods because they can easily find the difference when purchasing search goods. Search goods, such as laptop computers, are those whose quality and specifications are easily identified before purchase. By contrast, experience goods such as music, movies, hotels, restaurants, and cosmetics, are those whose quality is difficult to assess before using the product. Therefore, consumers are likely to be dependent on OCRs before making purchase decisions when purchasing experience goods.

Cosmetics have the characteristics of experience products. Considering that skincare products are directly applied to the skin, safety can be an important issue. For color cosmetics, consumers may make purchase decisions to satisfy social needs and self-expression. In this way, cosmetics have a strong characteristic of experience products. Then helpful consumer reviews can play an important role in other consumers' purchase decisions at the online retail channel. This study aims to extend research on online review helpfulness in two ways.

First, this study uses novel text-mining techniques in conducting a content analysis for big data and provides useful insights into customer feedback (Willemsen et al., 2011). The crucial issues after product usage can be identified by applying bigram natural language processing (NLP). In this study, the sample includes OCRs generated at Amazon.com for Korean skincare products. By applying the bigram analysis technique, result shows that the most frequently referred topic is user skin type. Furthermore, this study provides evidence that the variables operationalized based on content analysis can significantly affect the helpfulness of the OCRs.

Second, this study presents additional results on the moderating effect of product types on the influence of review extremity on helpfulness. Mudambi and Schuff (2010) provided empirical evidence that reviews with extreme ratings are perceived as less helpful than those with equivocal ratings. Their study only utilized the experience goods with hedonic characteristics, such as a music CD, an MP3 player, and a video game. However, depending on the product characteristics—hedonic or utilitarian—the effect of review extremity on helpfulness can be different. This study hypothesizes and tests the positive effect of extreme reviews on helpfulness by focusing on skincare products with a utilitarian trait.

The rest of this paper is organized as follows. Section 2 discusses the existing literature on the determinants of the OCR helpfulness. Section 3 shows the research model and hypotheses derived based on prior literature. Then, Section 4 and 5 explain the data, the empirical analysis model, and the analysis results. Last, Section 6 discusses the theoretical and practical implications, limitations, and future research topics.

2. Theoretical Background

2.1. Determinants of OCR Helpfulness

OCR helpfulness is an indicator of the value of the review and is measured by the number of users who voted useful for that review. Prior studies have addressed the fundamental research question about which factors determine the helpfulness of a review (Mudambi & Schuff, 2010; Salehan & Kim, 2015; Schindler & Bickart, 2012). Several studies have focused on the influence of quantitative metrics, such as star ratings and the text length.

Product ratings usually range from one to five, with 1 being a very negative rating, 5 being a very positive rating, and 3 being a moderate rating. The effect of review polarity on review helpfulness has two-sided arguments. One line of research in advertising has shown that two-sided messages can have higher reliability than an extreme opinion, and thus, the helpfulness of information is high (Eisend, 2006; Hunt & Smith, 1987). By contrast, other studies argued that extreme reviews provide useful information to readers by providing new information (Forman et al., 2008; Pavlou & Dimoka, 2006). For example, Forman et al. (2008) showed that neutral reviews provide less information than extreme ones, so they receive significantly low helpfulness votes. Mudambi and Schuff (2010) suggested that the mixed findings may be caused by not considering the product type.

Findings on the effect of positive and negative reviews on the perception of helpfulness vary. Based on negativity bias theory, several studies have shown that negative reviews are more useful than the positive ones. According to the theory of negativity bias, people have difficulty in reasoning about expected behaviors but tend to reason easily about behaviors that are not expected. Therefore, many people tend to provide positive reviews, so people find negative reviews greatly useful (Salehan & Kim, 2015; Sen & Lerman, 2007). Other studies find no significant difference between the effects of positive reviews and negative ones (e.g., Schindler & Bickart, 2012).

The review length also positively affects the helpfulness (Ghose & Ipeirotis, 2010; Pan & Zhang, 2010; Schindler & Bickart, 2012). Long reviews can convey more detailed information regarding product usage than short reviews. Additional information can help decision-makers make decisions with confidence (Tversky & Kahneman, 1974).

2.2. Effect of Review Text on the Helpfulness

Prior studies on review helpfulness have focused on the role of OCRs in a consumer's purchase decision process based on the economics of information theory. Online reviews serve as useful information to consumers by lowering the perceived risk associated with purchasing (Siering et al. 2018). In recent years, as the text mining techniques developed, studies on variables based on content analysis were also being conducted (Ghose & Ipeirotis, 2010; Salehan & Kim, 2016; Schindler & Bickart, 2012; Siering et al., 2018; Singh et al. 2017).

Ghose and Ipeirotis (2010) showed that the objective or subjective information extracted from the text of online reviews and readability is an important predictor of usefulness. Schindler and Bickart (2012) analyzed the effect of the review content and style. The review content represents the information the review conveys, and the style represents the words used to convey the information. They found that the proportion of product descriptions in the review content and that of reviewer descriptions significantly affect review usefulness. Siering et al. (2018) applied content analysis on Amazon reviews and identified words related to product quality, positivity, negativity, and uncertainty on review helpfulness.

2.3. Review Helpfulness and Product Type

Product type is an important moderating variable of product rating effect on the review helpfulness. Consumers tend to pursue aesthetic, sensory, or emotional pleasure when judging products for which the hedonic aspect is important (Hirschman & Holbrook, 1982; Pan & Zhang, 2011). By contrast, consumers tend to be highly cognitive, instrumental, and goal-oriented when judging which utilitarian value is important (Pan & Zhang, 2011; Strahilevitz & Myers, 1998). Therefore, the decision-making process for a hedonic versus utilitarian product can moderate the positive or negative review's helpfulness.

Sen and Lerman (2007) investigated the existence of negativity effect in eWOM helpfulness for utilitarian and hedonic products. They found that negative reviews are more useful in utilitarian products. As utilitarian consumption aims to maximize one's utility, consumers may seriously consider what negative factors may arise in their purchasing decisions. By contrast, they found that positive reviews are evaluated as more useful in hedonic products than negative ones. This result is because evaluating a hedonic product is related to meeting the expectation of achieving a value (e.g., happy life), which can be subjective. Consumers likely perceive the reviewer's experience as an individual's subjective experience and as less helpful.

Review helpfulness is also considered differently depending on search vs. experience goods. Nelson (1970, 1974) classified the type of products into search and experience goods. For the former, consumers can easily access information on product quality before making a purchase decision, whereas consumers find evaluating product quality for experience goods challenging.

Mudambi and Schuff (2010) found that extreme reviews' helpfulness may differ depending on search or experience goods as different products have different information needs. For the former, reviews with very high or low star ratings are perceived as more helpful than the moderate ones. By contrast, extreme reviews are perceived to be less helpful than the moderate ones for experience goods. The reason is that consumers often show high confidence in their tastes for experience goods and can be skeptical of peer consumers' extreme reviews. Chua and Banerjee (2016) highlighted that the helpfulness for a positive review of experience goods was greater than that of search products. They demonstrated that review valence was negatively related to the perceived usefulness of reviews discussing experience products.

3. Model and Hypotheses

Following prior literature, this study considers the numeric review attribute and the semantic content reflected in the text as major determinants of online review helpfulness. The text mining technique was applied to extract the major topics related to consumer's skincare product usages. Figure 1 summarizes the proposed research model of online review helpfulness for the skincare product category.

Several studies have been conducted on the determinants of online reviews' helpfulness in the existing information systems and marketing literature. Early research on review helpfulness was based on quantitative information such as review length and rating information (Pan & Zhang, 2010). Review length was one of the significant explanatory variables of review helpfulness (e.g., Bjering, Havro, & Moen, 2015; Mudambi & Schuff, 2010; Schindler & Bickart, 2012). Short reviews reflect the overall evaluation or attitude and are unlikely to include detailed information. On the contrary, long reviews are more likely to contain detailed information. Hence, consumers are more likely to perceive long reviews as helpful than short ones. Thus, a hypothesis was derived as follows:

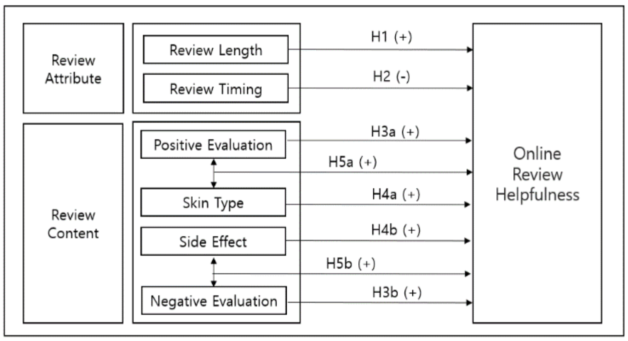

Figure 1: Model of Online Review Helpfulness of Skincare Products

H1: The review length is positively associated with the helpfulness of the review.

In the online review literature, review timing is also considered an important variable in predicting review helpfulness. In predicting hotel review helpfulness, Hu and Chen (2016) found that the number of days since a review is posted is one of the critical determinants of review helpfulness due to review visibility. As the number of reviews increases, Amazon.com provides a review system sorted by the number of helpfulness votes. Hence, the review posted earlier after the registration of a product is likely to have a great chance for many helpfulness votes. Therefore, the date that the review was written can affect the usefulness of the review. Once the review timing is delayed from the registration of a product, the review may have fewer chances to obtain helpfulness votes. For the same reason, Salehan and Kim (2016) showed that older reviews receive more readership and helpfulness votes than the newer ones. Consequently, the review timing is expected to be negatively associated with helpfulness votes.

H2: Reviews written immediately after product registration receive more helpfulness votes than those written later.

A prior study suggests that extreme reviews on experience goods may decrease the perceptions regarding their helpfulness than search goods (Mudambi & Schuff, 2010). They conduct an analysis with experience goods with hedonic nature such as movies, music, and travel. Skincare products belong to experience goods were consumers cannot evaluate products quality before purchase (Chua & Banerjee, 2016). Moreover, these product have a more utilitarian nature rather than hedonic. This is because consumers want to satisfy functional needs, such as skin protection. Therefore, the degree to which consumers perceive the helpfulness of extreme reviews for skincare products may differ from that for movie or music consumption.

In the case of experience goods in which the hedonic needs are important, the meaning of an extreme evaluation may reflect personal tastes and less likely to affect other's purchase decisions. By contrast, the goal of utilitarian consumption is utility maximization, and consumers may seriously consider what the negative issues which may occur after purchase (Sen & Lerman, 2007). Moreover, given that the Korean skincare products have been introduced recently in the U.S. market through the online distribution channel, their brand awareness is relatively low. Consumers may feel a higher perceived risk than relatively well-known cosmetic brands. Therefore, extreme reviews may be considered greatly helpful given its ability to reduce the potential risk (Siering et al., 2018). In sum, review extremity is hypothesized to have positively related to helpfulness votes in the skincare product category. Thus, the following hypotheses are proposed:

H3a: Reviews with very positive ratings receive high helpfulness votes compared with those with a moderate rating.

H3b: Reviews with negative ratings receive high helpfulness votes compared with those with a moderate rating.

When a review content delivers product and reviewer-related descriptions, the reviews can provide detailed information, and consumers are likely to consider it very helpful (Schindler & Bickart, 2012). Willemsen et al. (2011) showed that content characteristics, such as argumentation style may affect the credence of reviews. Beyond the mere sentiment and quantity of reviews, review helpfulness can be associated with the quality of information delivered (Chua & Banerjee, 2016). Accurately presenting a reviewer's experience will be useful for consumers. In the context of skincare products, the suitability of products to their skin type is the most important aspect. Product performance may appear differently depending on the user's skin type. Thus, providing the reviewer's skin type information can be perceived as useful for other consumers.

H4a: Reviews that mention skin types have higher helpfulness votes than those that do not.

H4b: Reviews that mention of skin issues after using the product have higher helpfulness votes than those that do not.

experiences, such as skin type or product use-related trouble are reflected in the reviews, the credibility of extreme reviews can be enhanced. Therefore, the interaction between review polarity and review content variables may influence review helpfulness. Specifically, extremely positive reviews containing the reviewer's skin type or extremely negative reviews reflecting the side effect can boost the credibility of the review. Unlike other experience products for hedonic consumption (e.g., movie, music), consumers may perceive a high level of perceived risk for the skincare product, which does not fit her/his skin type. Therefore, for skincare products, extremely negative reviews on product use or positive reviews mentioning the skin type can be useful for other consumers (Sen & Lerman, 2007). Extremely positive or negative reviews can enhance the helpfulness of reviews because they can provide effectiveness or potential risk of using skincare products. Considering that the review content can provide additional information on the rating, skin type information may enhance the positive effect of favorable reviews on helpfulness. Moreover, skin trouble after usage may boost the positive effect of unfavorable reviews on helpfulness.

H5a: Positive reviews that mention skin types have higher helpfulness votes than those that do not.

H5b: Negative reviews that mention side effects after using the product have higher helpfulness votes than those that do not.

Table 1 shows the prior literature on helpfulness using skincare OCR data. In the context of analyzing OCRs of skincare products, content-based variables such as referring reviewer's skin type or skin trouble experience could be positively associated with review helpfulness.

Table 1: Previous Studies on Review Helpfulness of Skincare Products

Study	Data and Product (#of Reviews)	Findings
Willemsen et al. (2011)	Amazon.com Sunscreen (n=465)	Negative reviews lead to higher perceived helpfulness than positive ones due to negativity bias.
Cheung, Xiao, and Liu (2014)	Online Review Forum in Asia (n=39,897)	Peer consumer purchases will positively influence the consumer purchase decision.
Bjering et al. (2015)	Amazon.com Perfume (n=21,724)	The association between helpfulness and the length of the review is stronger for search goods than for experience ones.
Chua and Banerjee (2016)	Amazon.com Skincare Products (n=346)	Among experience products, the information quality for review helpfulness is the greatest for negative reviews.
The present study (2020)	Amazon.com Korean Skincare Products (n=69,699)	Review content (e.g., skin compatibility) moderates the review extremity effect on helpfulness.

4. Empirical Applications

4.1. Data

For this study, we collected Amazon consumer review data for the Korean skincare product category. Exports of Korean cosmetics have increased mainly in Asia due to the increasing interest in Hallyu and improving product quality (Park, 2015; Yoon et al., 2020). Cosmetics companies in South Korea have actively diversified their targets through a global e-commerce platform Amazon.com. In the U.S. market, Korean cosmetics are newly introduced, and information about the brand is limited. Hence, U.S. consumers are likely to make purchase decisions based on the opinions of prior consumers. Given that beauty products have a character of experience goods, online reviews can serve as an essential source of information for consumers.

The sample includes the 69,633 OCRs generated from March 2013 to August 2019 for 467 products manufactured by the 13 Korean cosmetic brands. Figure 2 shows the change in the number of products registered on Amazon.com for the last seven years in these 13 brands. Most brands increase the number of products listed on Amazon.com. Table 2 shows the brands, the number of registered items, average item price, average star rating, and the average number of reviews. The reviews with 0 helpfulness vote are excluded in the sample to investigate the determinants of review helpfulness.

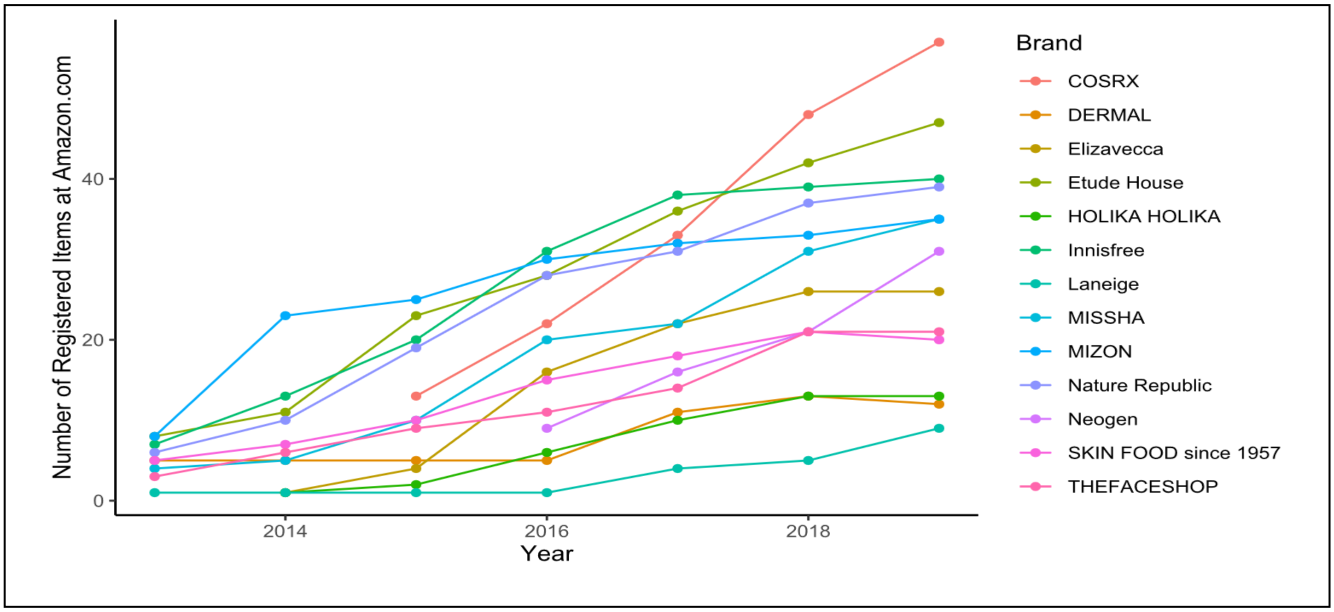

Figure 2: The Number of Registered Items of Korean Skincare Products at Amazon.com (Source: own)

Table 2: Summary of Korean Skincare Products at Amazon.com

Brand	# of Items	Avg. Price($)	Avg. Rating	Avg.# of Reviews
COSRX	58	16.52	4.07	149.71
DERMAL	13	12.14	4.27	464.38
Elizavecca	27	12.25	4.07	437.30
Etude House	49	12.67	4.09	158.02
HOLIKA HOLIKA	13	12.70	4.15	79.23
Innisfree	42	14.57	4.08	96.60
Laneige	9	23.96	4.33	109.67
MISSHA	36	19.17	4.06	69.89
MIZON	36	16.04	4.06	377.61
Nature Republic	41	14.74	4.16	101.20
Neogen	31	15.67	4.29	73.39
SKIN FOOD	21	15.65	4.21	205.19
THE FACE SHOP	21	18.75	4.24	219.48

To identify the frequent expressions reflected in online consumer reviews, the NLP(Natural Language Processing) procedure was applied to text data. First, the data cleansing process was conducted to remove stop words (a, the, and, in, on, at, etc.). Then, the bigram NLP analysis, which has been widely adopted in other academic domains (Computer Science- Bhakkad & Kulkarni, 2013; Medical – Wakamiya et al., 2019), is applied. Bigram NLP analysis is useful in the sense that the researchers can detect the most frequently appearing two-word combinations. As a result, it helps us to understand the context in which specific words are used.

Figure 3 shows the results of applying bigram NLP with OCRs in the cosmetics category. The mention of the skin type of consumers showed the highest percentage in the bigram frequency. For example, "sensitive skin," "oily skin," and "dry skin," "acne prone," and "combination skin" are frequently mentioned other than product category-related bigrams (e.g., "skin care," "eye cream," and "sheet masks"). These results suggest that in purchasing cosmetics, consumers mostly consider the suitability of products to their skin type. Given that the performance of the product may appear differently depending on the user's skin type, providing a reviewer's skin type information can be perceived as useful to other consumers. Therefore, this study aims to analyze not only quantitative information on reviews—positive/neutral/negative ratings but also the effect of information reflected in the text on review helpfulness. A dummy variable labeled "skin type" was constructed to evaluate whether the inclusion of skin type information affects the review helpfulness.

Furthermore, when applying NLP analysis to OCRs with a low rating of 3 or less, word combinations related to skin problems appeared frequently (e.g., "reaction," "bump," "red," and "greasy"). Skin troubles effects that may occur when using cosmetics can play a role of provider of valuable information for consumers who are in the process of considering purchasing skincare products in the future. Therefore, a dummy variable was constructed to indicate whether the review contains skin trouble related words.

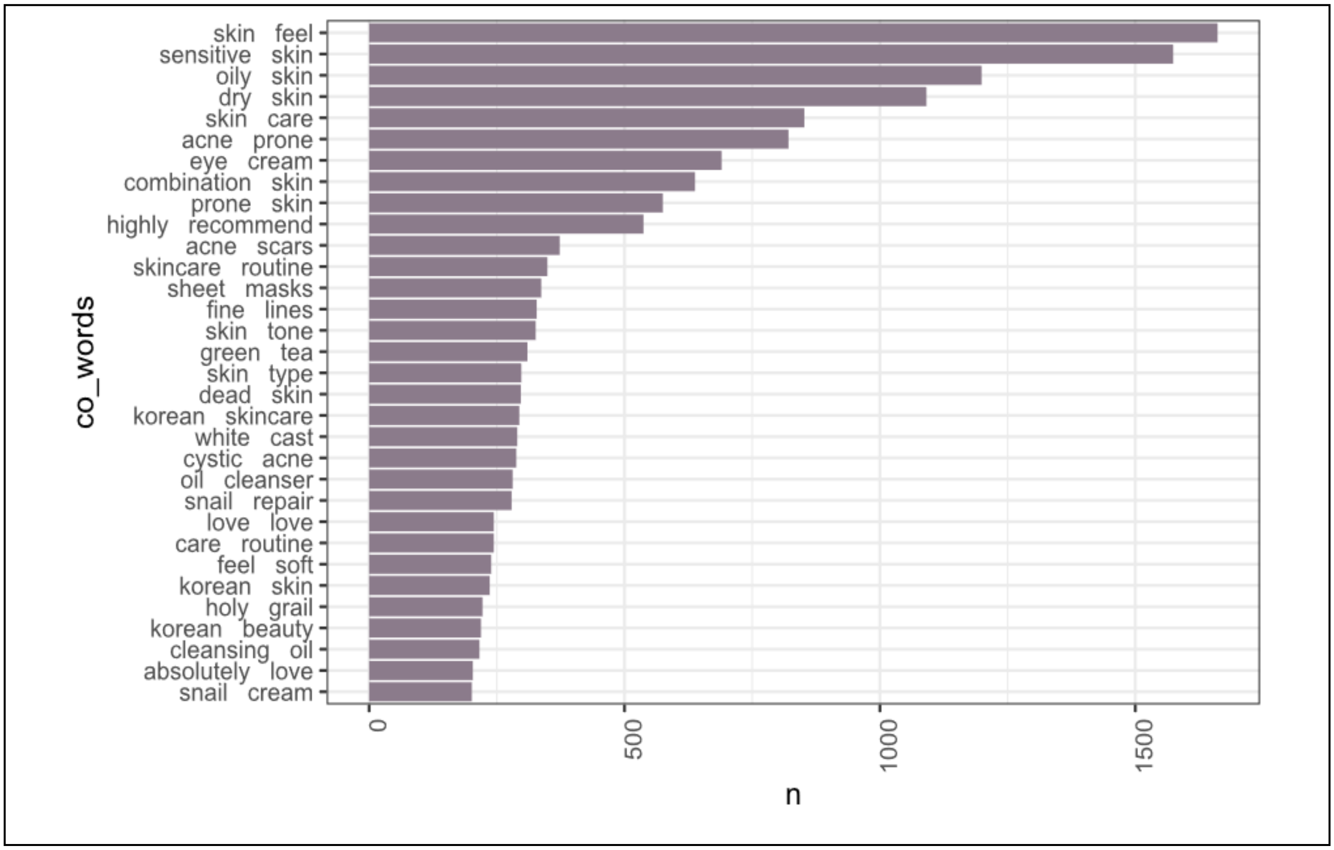

Figure 3: Bigram Analysis Results of K-Beauty Skincare Products (Source: own)

Table 3 shows the measurement of variables, and Table 4 shows the descriptive statistics of variables used in the model. The sample includes 20,550 reviews with at least one helpfulness vote, with an average value of 5 votes and a maximum value of 985 votes. The average number of characters is 404, and the review date was up to 2,355 days since the product was registered at Amazon.com. $highscore_{ik}$ measures whether a review k for the product i represents a very positive evaluation, which is a dummy variable with a value of 1 when the rating is 5 points. More than 56% of reviews with a rating of 5 points in the sample, which can be interpreted as the products included in the sample, are products that received a high evaluation of above 4 points on an average. $lowscore_{ik}$ measures whether the review k represents a very negative valence for the product i, with a value of 1 when the overall rating is 1 or 2 points (20% of the sample). Thus, this study considers 3 or 4 points out of 5 rating systems, which stands for a neutral point of view (24% of the sample). For the review of content-related variables, 21% of reviews reveal the reviewer's skin type ($skintype_{ik}$), and 27% of reviews includes skin trouble ($trouble_{ik}$) after product usage. $price_i$ captures the price effect on the helpfulness.

Table 3: Measurement of Variables

Variable	Measure
$helpfulness_{ik}$	Cumulative number of helpfulness votes for review k for product i
$reviewlength_{ik}$	Number of characters in review k for product i
$reviewtiming_{ik}$	Days passed since review k is written for product i
$highscore_{ik}$	If the review k has the rating of 5, it has a value of 1 (otherwise 0)
$lowscore_{ik}$	If the review k has the rating less than 3, it has a value of 1 (otherwise 0)
$skintype_{ik}$	If the review k has the skin type related words (i.e., sensitive skin/dry skin/prone skin/normal skin/combination skin/oily skin), it has a value of 1 (otherwise 0)
$trouble_{ik}$	If the review k has the skin trouble related words (i.e., bump/ red/ reaction/ greasy), it has a value of 1 (otherwise 0)
$price_{ik}$	Price of product i

Table 4: Descriptive Statistics of Variables (n=20,550)

Variable	Mean	Std. Dev.	Min	Max
helpfulness	5.331	24.944	1	985
reviewlength	404	539	13	20,193
reviewtiming	844	532	1	2,355
highscore(star=5)	0.567	0.496	0	1
lowscore(star<3)	0.204	0.403	0	1
skintype	0.206	0.404	0	1
trouble	0.268	0.443	0	1
price($)	14.016	6.652	2	79

4.2. Model

Considering N products, each of which is characterized by a brand name and price. In equation (1), helpfulness of review k (k=1,...,K) for product i (i=1,....,N) is modeled as a function of review length, review timing, high score, low score, skin type, skin trouble, and product price. Continuous variables are log-transformed to be robust in modeling. $\varepsilon_{ik}$is assumed to be distributed i.i.d. normal with a mean of 0 and a standard deviation of σ.

$$\begin{aligned}\log(helpfulness_{ik}) &= \beta_0 \\ &+ \beta_1 \log(reviewlength_{ik}) \\ &+ \beta_2 \log(reviewtiming_{ik}) \\ &+ \beta_3 highscore_{ik} + \beta_4 lowscore_{ik} \\ &+ \beta_5 skintype_{ik} + \beta_6 trouble_{ik} \\ &+ \beta_7 \log(price_i) + \varepsilon_{ik}\end{aligned} \tag{1}$$

$$\begin{aligned}\log(helpfulness_{ik}) &= \beta_0 \\ &+ \beta_1 \log(reviewlength_{ik}) \\ &+ \beta_2 \log(reviewtiming_{ik}) \\ &+ \beta_3 highscore_{ik} + \beta_4 lowscore_{ik} \\ &+ \beta_5 skintype_{ik} + \beta_6 trouble_{ik} \\ &+ \beta_7 \log(price_i) + \beta_8 highscore_{ik} \\ &\times skintype_{ik} + \beta_9 lowscore_{ik} \\ &\times trouble_{ik} + \sum_{j=1}^{m} \beta_{10(j)} BI_i + \varepsilon_{ik}\end{aligned} \tag{2}$$

5. Results

The models are estimated with the maximum likelihood estimation, and Table 5 shows the results of the analysis. Model 1 shows parameter estimation results of equation (1). Then, Model 2 shows the results when review content variables are added, and Model 3 shows the results with interaction terms between review extremity and review content variables. Moreover, Model 4 represents the full model results, including brand fixed effects shown in equation (2). In all models, review length has a positive, significant impact on helpfulness (H1 supported). As shown in previous studies, review depth (Mudambi & Schuff, 2010) plays an essential role in the perception of helpfulness. The results also provide strong support for H2, hypothesizing that reviews posted later are less likely to have helpfulness votes than ones posted earlier. The coefficients of review timing show negative and significant signs in all models.

For the effect of review extremity, extremely high and low star ratings improve the review helpfulness as expected in H3a and H3b. The positive coefficient for $highscore_{ik}(\beta_3 = 0.101, p < 0.01$ in Model 2) indicates that reviews with extremely high star ratings earn more helpfulness votes than moderate rating reviews. For the reviews with extremely low star ratings, the positive effect of $lowscore_{ik}$ on helpfulness($\beta_4 = 0.268, p < 0.01$ in Model 2) is high, indicating that negative reviews may have an informational value for other consumers for making their purchase decisions. These results are consistent with those of Willemsen et al. (2011), who postulated that negatively valenced reviews induce higher perceived helpfulness than positively valenced reviews ones due to negativity bias.

The empirical results on the effect of review extremity on helpfulness for experience products are different from those of Mudambi and Schuff (2010). The authors considered three types of experience goods—MP3 player, music CD, and P.C. video games. Different from those product categories, skincare products may lead to skin problems once the product does not fit the user's skin type. Hence, negative or positive reviews may provide detailed information to select the most suitable product for personal use. As the role of extremity has been a controversial topic in the review helpfulness literature, further investigation is necessary. This study provides evidence that extreme reviews can be valued more than ones as experience goods are consumed for utilitarian purposes.

The estimation results for review content variables ($skintype_{ik}$ and $trouble_{ik}$) demonstrate a differing effect on review helpfulness. The influence of including the reviewer's skin type information on helpfulness turns out insignificant in Model 2. In Model 3 and Model 4, the interaction effect of$skintype_{ik}$and $highscore_{ik}$ becomes significant, although the main effect of $skintype_{ik}$ is negative. The results provide strong support that reviews with a high rating can earn additional helpfulness votes once the reviewer writes a review providing her/his skin type. On the contrary, the main effect of skin trouble ($trouble_{ik}$) on helpfulness is positive and significant in Model 2, but the interaction effect of $trouble_{ik}$ and $lowscore_{ik}$ is insignificant. The results indicate that the skin trouble issue has informative value. However, an extremely low rating review about skin problems does not boost the helpfulness of reviews. Table 6 summarizes the hypothesis testing results.

Table 5: Parameter Estimation Results

Variable	Model 1		Model 2		Model 3		Model 4	
	P.E.	S.E.	P.E.	S.E.	P.E.	S.E.	P.E.	S.E.
log(review_length)	0.390***	(0.007)	0.374***	(0.008)	0.373***	(0.008)	0.376***	(0.008)
log(review_timing)	−0.145***	(0.005)	−0.145***	(0.005)	−0.145***	(0.005)	−0.153***	(0.006)
high_score(star=5)	0.105***	(0.016)	0.101***	(0.016)	0.077***	(0.017)	0.077***	(0.017)
low_score(star<3)	0.278***	(0.019)	0.268***	(0.020)	0.266***	(0.022)	0.264***	(0.022)
skin_type			−0.010	(0.016)	−0.083***	(0.026)	−0.085***	(0.026)
trouble			0.100***	(0.016)	0.100***	(0.017)	0.094***	(0.017)
high_score X skin_type					0.119***	(0.032)	0.121***	(0.032)
low_score X trouble					−0.001	(0.035)	0.011	(0.035)
log(price)	0.049***	(0.016)	0.044***	(0.016)	0.044***	(0.016)	−0.002	(0.020)
Brand Fixed Effects							Controlled	
Constant	−0.768***	(0.064)	−0.683***	(0.066)	−0.668***	(0.066)	−0.491***	(0.078)
Log Likelihood	−27,211(df = 7)		−27,190(df = 9)		−27,183(df = 11)		−27,141(df = 24)	
AIC	54,437		54,399		54,399		54,331	
Deviance	17,001		16,967		16,955		16,886	

Note: *p < 0.1; **p < 0.05; ***p < 0.01

Table 6: Summary of Findings

Description		Variables	P.E. (Model 4)	Result
H1	The review length is positively associated with the helpfulness of the review.	*log(review_length)*	0.376***	Supported
H2	Reviews written immediately after product registration receive more helpfulness votes than those written later.	*log(review_timing)*	−0.153***	Supported
H3a	Reviews with very positive ratings receive high helpfulness votes compared with those with a moderate rating.	*highscore(star=5)*	0.077***	Supported
H3b	Reviews with negative ratings receive high helpfulness votes compared with those with a moderate rating.	*lowscore(star<3)*	0.264***	Supported
H4a	Reviews that mention skin types have higher helpfulness votes than those that do not.	*skintype*	−0.085***	Not supported
H4b	Reviews that mention side effects after using the product have higher helpfulness votes than those that do not.	*trouble*	0.094***	Supported
H5a	Positive reviews that mention skin types have higher helpfulness votes than those that do not.	*highscore X skintype*	0.121***	Supported
H5b	Negative reviews that mention side effects after using the product have higher helpfulness votes than those that do not.	*lowscore X trouble*	0.011	Not supported

6. Conclusions

6.1. Summary

This study examines the effect of OCR contents on the helpfulness by applying text-mining techniques. For this purpose, 69,633 Amazon online reviews generated from March 2013 to August 2019 for 467 skincare products of 13 Korean cosmetic companies are collected. By applying bigram NLP, major consumer experience issues related to cosmetics are identified, namely, skin type and skin trouble. The estimation results support the hypothesis that review content has an exploratory power in the helpfulness model. Specifically, the high-rating review with the reviewer's skin is the type of information which was perceived to be more helpful than that without skin type information. Skin trouble issues are also positively associated with the helpfulness votes. In summary, the positive effect of the review content on helpfulness suggests that applying big data analytics can provide meaningful customer insights for the marketing managers in online retailing.

6.2. Theoretical and Managerial Implications

This study makes useful contributions both, theoretically and practically. In theoretical perspectives, this study furthers the literature on the determinants of review helpfulness in two ways. Firstly, this study challenges the prior literature on the moderating role of the type of product (search/experience good) in the relationship between review extremity and helpfulness. Contrary to prior study (Mudambi & Schuff, 2010), our empirical results demonstrate that the review extremity could be positively related to review helpfulness in the experience goods such as cosmetics. Therefore, whether or why product type moderates the perception of review helpfulness needs to be addressed in future research. Secondly, this study sheds light on the use of big data analytics to gain in-depth insights into post-purchase behavior and their effects on prospective consumers. Major issues regarding product usage can be identified by applying the bigram NLP analysis. In the context of the skincare product category, the product suitability to the user's skin type appears to be one of the critical issues. Thus, whether the review provides the reviewer's skin type can moderate the review helpfulness, specifically in an extremely high rating product. Moreover, the reviews mentioning related potential skin trouble can increase the review helpfulness.

In practice, this study can provide useful implications for cosmetic companies that sell products through online distribution channels. A majority of prior literature has shown that online reviews posted on the e-commerce website can play a crucial role in the purchasing decisions of prospective consumers (Chevalier & Mayzlin, 2006; Clemons, Gao, & Hitt, 2006; Mudambi & Schuff, 2010; Oh, 2017). For the experience goods such as cosmetics, online reviews can provide valuable information for prospective consumers and help to boost product sales. Review helpfulness voting systems allow a consumer to effectively manage information. Particularly for the cosmetics category, OCRs can provide how the product works for various skin types, which is hardly provided by sellers. Manufacturers in this product category may use this information to improve their products or find the target market and proper product positioning statement. Online retailers selling beauty products may consider applying text mining techniques in their online review database so that the prospective consumers can easily find how other consumers with similar skin types evaluate the product. In this way, online retailers can utilize their online reviews as an essential asset to increase sales of products.

6.3. Limitations and Future Research

As with other studies, this study has limitations too. This study employs the sample from the Amazon.com website, mainly limited to the skincare products by Korean cosmetic companies. This study has the advantage of being able to extract major consumer experience issues reflected in the text content by focusing on a specific category. However, the validity of the results and the effects of new moderators can be examined by collecting OCRs from other product categories. Such a limitation can be addressed in future studies. First, future research could examine whether the positive effect of the extreme reviews on helpfulness is generalizable by analyzing OCRs for the established brands (e.g., L'Oreal, OLAY).

Furthermore firstly, investigating the moderating effect of brand awareness on the perception of online review helpfulness will be a meaningful research agenda. Secondly, this study focuses on the review helpfulness products of utilitarian value among experience goods such as skincare products. Future research may compare how hedonic/utilitarian products show a differential effect on the perception of review helpfulness by collecting OCRs from color cosmetics with a more hedonic value of consumption.

References

Bickart, B., & Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, *15*(3), 31-40.

Bjering, E., Havro, L. J., & Moen, Ø. (2015). An empirical investigation of self-selection bias and factors influencing review helpfulness. *International Journal of Business and Management*, *10*(7), 16-30.

Bhakkad, A., Dharamadhikari, S. C., & Kulkarni, P. (2013). Efficient approach to find bigram frequency in text document using E-VSM. *International Journal of Computer Applications*, *68*(19), 9-11.

Cheung, C. M., & Lee, M. K. (2012). What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision Support Systems*, *53*(1), 218-225.

Cheung, C. M., Xiao, B. S., & Liu, I. L. (2014). Do actions speak louder than voices? The signaling role of social information cues in influencing consumer purchase decisions. *Decision Support Systems*, *65*, 50-58.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345-354.

Clemons, E. K., Gao, G. G., & Hitt, L. M. (2006). When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, *23*(2), 149-171.

Crowley, A. E., & Hoyer, W. D. (1994). An integrative framework for understanding two-sided persuasion. *Journal of Consumer Research*, *20*(4), 561-574.

Chua, A. Y., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type, and information quality. *Computers in Human Behavior*, *54*, 547-554.

Eisend, M. (2006). Two-sided advertising: A meta-analysis. *International Journal of Research in Marketing*, *23*(2), 187-198.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291-313.

Ghose, A., & Ipeirotis, P. G. (2010). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498-1512.

HAN, S. S. (2020). A factors effecting online social decisions in online consumer behavior. *The Journal of Distribution Science*, *18*(3), 67-76.

Hirschman, E. C., & Holbrook, M. B. (1982). Hedonic consumption: emerging concepts, methods and propositions. *Journal of Marketing*, *46*(3), 92-101.

Hu, Y. H., & Chen, K. (2016). Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. *International Journal of Information Management*, *36*(6), 929-944.

Hunt, J. M., & Smith, M. F. (1987). The persuasive impact of two-sided selling appeals for an unknown brand name. *Journal of the Academy of Marketing Science*, *15*(1), 11-18.

Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS Quarterly*, *34*(1),185-200.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, *78*(2), 311-329.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy*, *82*(4), 729-754.

Oh, Y. K. (2017). The impact of initial eWOM growth on the sales in movie distribution. *The Journal of Distribution Science*, *15*(9), 85-93.

Pan, Y., & Zhang, J. Q. (2011). Born unequal: A study of the helpfulness of user-generated product reviews. *Journal of Retailing*, *87*(4), 598-612.

Park, Y. S. (2015). Does the Rise of the Korean Wave Lead to Cosmetics Export? *Journal of Asian Finance, Economics and Business*, *2*(4), 13-20.

Pavlou, P. A., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, *17*(4), 392-414.

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, *81*, 30-40.

Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing*, *21*(4), 76-94.

Schindler, R. M., & Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behavior*, *11*(3), 234-243.

Siering, M., Muntermann, J., & Rajagopalan, B. (2018). Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. *Decision Support Systems*, *108*, 1-12.

Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Roy, P. K. (2017). Predicting the “helpfulness” of online consumer reviews. *Journal of Business Research*, *70*, 346-355.

Strahilevitz, M., & Myers, J. G. (1998). Donations to charity as purchase incentives: How well they work may depend on what you are trying to sell. *Journal of Consumer Research*, *24*(4), 434-446.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124-1131.

YOON, S. H., SONG, S. Y., & KANG, M. S. (2020). The Growth and Change of Korean Cosmetics Market in Distribution Structure. *The Journal of Distribution Science*, *18*(1), 5-13.

Wakamiya, S., Morita, M., Kano, Y., Ohkuma, T., & Aramaki, E. (2019). Tweet classification toward Twitter-based disease surveillance: new data, methods, and evaluations. *Journal of Medical Internet Research*, *21*(2), e12783.

Willemsen, L. M., Neijens, P. C., Bronner, F., & De Ridder, J. A. (2011). “Highly recommended!” The content characteristics and perceived usefulness of online consumer reviews. *Journal of Computer-Mediated Communication*, *17*(1), 19-38.