DSSR

Data Science and Service Research Discussion Paper

Interpretable Perceived Topics in Online Customer Reviews for Product Satisfaction and Reader Helpfulness

Mirai Igarashi[*]   Aijing Xing[†]   Nobuhiko Terui[‡]

April 2020

Interpretable Perceived Topics in Online Customer Reviews for Product Satisfaction and Reader Helpfulness

Abstract

Online customer reviews contain useful and important information, particularly, for product development and management, because customers praise or criticize in their reviews certain product attributes. We propose a model that extracts perceived topics from textual reviews using natural language processing under some restrictions created using seed words for improving the topic interpretability. In addition, the proposed model estimates the relationships between the topics and product satisfaction by writers of the review and the perceived helpfulness of reviews by readers, that is, these textual reviews are viewed as current product evaluations by customers who have made purchases and expectations of possible future demand by consumers who have yet to make purchases. The empirical study on e-commerce food reviews shows that our proposed model performs better than the extant alternative models and provides interesting findings such that the "ingredient" topic in the review text decreases the levels of customer satisfaction and the reader's perceived helpfulness. In contrast, the "health" topic increases the levels of both customer satisfaction and the reader's perceived helpfulness. These findings help us understand the product attributes that purchased customers are satisfied with and for which readers of reviews find helpful information.

1 Introduction

Given the increase in electronic commerce, online retailers such as Amazon, Walmart, and Taobao have experienced growth in their number of users. Most online retailers have customer feedback systems to receive information for customer reviews, including satisfaction scores (also called product ratings), textual reviews, and readers helpfulness. In customer reviews, customers argue about their experiences and evaluations of product attributes.

Company and marketer must clarify the perceived quality for products and estimate their preferences for or importance of each product attribute because this information can be utilized to discuss and conduct vast marketing strategies for product development, brand positioning, and advertising. Berger et al. (2020) suggested that companies and marketers can use online reviews to understand whether consumers prefer a product, how consumers feel about a brand, the attributes that are relevant for decision making, or the other brands that fall into the same consideration set.

Customer reviews not only describe customers' evaluations and experiences with a product but also provide information about the product for consumers who read reviews to make purchasing decisions. Chen and Xie (2008) suggested that online reviews can help novice consumers identify products that best match their preferences. These authors concluded that, in the absence of information from reviews, novice consumers may be less likely to buy a product if the only information available is seller-created product attributes. Obviously, customers prefer to read customer reviews before making purchase decisions because such user-generated information is perceived to be trustworthy and in line with their preferences.

A large and growing body of literature has investigated methods for measuring the perceived quality of products, and these methods are related to customer surveys or interviews. However, information is obtained through surveys and interviews at a time lag, tends to be limited, and is costly. Additionally, to create a questionnaire, product attributes need to be selected from a pre-defined set of attributes; that is, such attributes are company-oriented. In contrast, information obtained from online customer reviews is instantaneous

1

and tends to be high in volume and low cost. In addition, these product attributes are selected by customers. In other words, they are customer-oriented. Therefore, for companies and marketers to utilize customer review data and estimate customer preferences or readers' perceived helpfulness is useful.

In the literature, several methods for preference measurement and market research using online customer reviews have been proposed (e.g., Decker and Trusov, 2010). To analyze customer reviews, we first need to extract the product attributes mentioned in the reviews. These studies adopted a rule-based approach that one-to-one translates from words or phrases to product attributes. Through this approach, researchers create lexicons for translation by humans or by using useful tools, such as machine learning, and then map words or phrases to product attributes using the lexicons. Then, they construct a statistical model to explain the relationships between the quantified review text and the dependent variables, such as customer satisfaction.

However, through this approach, which engages in a one-to-one translation of words to attributes, the important information from the contexts of words meanings and sentiments is ignored. For example, "small" could be a positive attribute for a mobile phone and a negative attribute when describing a screen. Some studies proposed a topic modeling approach (e.g., Tirunillai and Tellis, 2014). The topic modeling approach assumes that each word may be assigned to multiple topics (or product attributes) according to the context and can flexibly extract product attributes from review text. In addition, because the approach involves little human intervention, the researchers do not need to know the latent product attribute dimensions in advance, and human error and biases can be minimized as much as possible.

However, the topic modeling approach is problematic because there is no guarantee that we always extract interpretable topics. Basically, the topic modeling approach is based on latent Dirichlet allocation (LDA) as proposed by Blei et al. (2003), and the core of the LDA model is the unsupervised learning method, that is, no word is given any correct topic under supervision and is assigned to the most likely topic in the learning process. Because

the assignment depends entirely on unsupervised learning, a set of words may be obtained with cohesion that is incomprehensible to human. As a result, even if researchers use the topic modeling approach to extract product attributes in customer reviews, they might not achieve preference measurements—the ultimate purpose—if they cannot understand what the attribute actually represents.

In this study, we propose a semi-supervised topic modeling approach to tackle the problem. This approach provides to the topic model symbolic words that are representative of a product attribute, which is quite different from the original topic model that does not offer any prior knowledge. For example, regarding a product's price attribute, "expensive" and "cheap" can be vied as representative words for that attribute. The topic assignments of these words are fixed to corresponding topics. The proposed topic model considers some labeled words whose topic assignments are determined in advance and assigns other non-labeled words to relevant topics based on the labeled words. Therefore, the topics resulting from the analysis contain several pre-assigned representative words, and the proposed approach is expected to improve the interpretability of topics relative to the original unsupervised topic modeling approach. Furthermore, to not require prior knowledge of the analyst is one of the strong points of the topic modeling approach. However, this advantage is wasted if the analyst had considerable knowledge of the domain. In contrast, the proposed semi-supervised approach can naturally incorporate into the model such prior knowledge of the analyst.

The purpose of this research is to propose a topic model that stably extracts interpretable product attributes from customer reviews by providing pre-defined word labels and to answer the following research questions using the proposed model: "What product attributes are mentioned in the online customer reviews and what is the relationship between the product attributes and purchasing customers' satisfaction?" and "What product attributes do readers of customer reviews expect to find as useful information?"

In this empirical study, we use Amazon customer review data to demonstrate the performance of the proposed model and provide a case study that applies our model to a real

dataset. First, we compare the proposed model with existing models from two viewpoints: model selection via information criteria and interpretability of the extracted topics. We use these comparisons to show that differences between our model with the word labeling restriction and the no restriction model are not very large and that our model extracts topics that can be more easily interpreted. Next, we discuss the estimates to obtain interesting findings, such that the "ingredient" topic in the review text decreases the levels of customer satisfaction and the perceived helpfulness of readers, and in contrast, "health" topic increases the levels of both. However, in this study, we use the customer review dataset for only one product, which is a limitation of this work. Because our research is only a case study, generalizing the findings and applying our model to other products and categories is necessary.

In Section 2, we discuss related works in the relevant body of literature. In Section 3, we describe the details of dataset used in this study and how to construct "labeled" topics. Then, in Section 4, we propose the supervised-labeled LDA model and derive the inference algorithm. Section 5 presents an application of the proposed model for Amazon customer review data and a discussion of the results. Finally, in Section 6, we provide concluding remarks and directions for future research.

2 Literature Review

2.1 Customer Review Analysis for Customer Satisfaction and Reader Helpfulness

This section provides a brief overview of related studies on customer review analysis. First, we discuss several studies that aimed to capture the feelings that customers expressed in their reviews. In the fields of marketing and management, researchers proposed several approaches to extract the product attributes mentioned in customer reviews and estimate their relationships with customer satisfaction and product sales. Decker and Trusov (2010)

4

and Xiao et al. (2016) estimated the impact of product attributes that were positively and negatively mentioned in ratings using a latent class Poisson regression and heterogeneous multinomial choice model. Archak et al. (2011) proposed a demand estimation model that captures the effects of product attributes on sales considering the heterogeneity of each word that qualifies the attributes.

These studies treated all reviews equally as subjects for the analysis. However, Qi et al. (2016) used human and machine learning methods to select only useful reviews and to analyze the effects of attributes on product ratings. Companies and marketers can use the relationship between product attributes and customer satisfaction and sales to analyze the market structure (Lee and Bradlow, 2011; Netzer et al., 2012) and improve the search algorithm for product pages (Ghose et al., 2012).

Customer reviews contain assessments of product attributes by customers who have already purchased the product and affect the perceived helpfulness of consumers who are potential buyers. Some studies analyzed the relationship between the review attributes and readers' helpfulness. For example, Chen and Xie (2008) explored the interactive effects of seller-created product attributes and buyer-created review information on readers' usage experiences. They clarified that customer reviews help consumers identify products that best match their idiosyncratic usage conditions.

Hence, consumers are observed to read reviews in search of useful information. Yin et al. (2017) analyzed the helpfulness that readers receive from customer reviews by taking into account the variations in the degrees of emotional arousal. Additionally, Mudambi and Schuff (2010) revealed the factors that make reviews helpful to customers during the purchase decision process and described the effect of review ratings, review depth, and product types, search and experience goods. In this study, we build on these previous studies and propose a model that identifies the product attributes that make reviews helpful.

2.2 Extracting Product Attributes from Customer Reviews

The approaches for analyzing customer reviews to reveal the relationship between the product attributes mentioned therein and the dependent variables, such as satisfaction, sales, and reader helpfulness, can be broadly stated as being a two-step process that extracts product attributes from review text and constructs a statistical model to estimate the effect of product attributes on dependent variables. In this study, we focus on improving approaches for extracting product attributes from review text and start with identifying how the proposed methods extract attributes and their limitations.

Most of the studies previously discussed adopted a rule-based approach that maps words and product attributes on a one-to-one basis. After preprocessing the review text, including removing stop words and word stemming, these studies created rules that determined the product attribute that a word represents using either human power (Moon and Kamakura, 2017; Hou et al., 2019) or machine learning techniques, such as clustering (Lee and Bradlow, 2011; Archak et al., 2011). By identifying the correspondence between words and the attributes, unstructured text data can be transformed into quantified variables, and researchers can explore the effects of attributes on the dependent variables through a regression or choice model.

However, these rule-based approaches, which transform words into attributes one-to-one, ignore important information, fluctuations in the meanings and sentiments of words according to the context, similar to the example of "small" in Section 1. To capture these fluctuations, approaches using topic modeling, especially LDA (Blei et al., 2003), have been proposed in the literature. The LDA model reflects the generative process of text, and a word is assumed to be generated from the vocabulary on its latent topic. The characteristic of the LDA is that topics may differ according to the context, even if it is about the same word.

In customer review analysis, the topics in the review text can be treated as product attributes, and some studies used the LDA model and proposed an extension model to

extract product attributes from the review text. For example, Tirunillai and Tellis (2014) proposed an extension of the LDA model by adding a mechanism that considers the latent sentiment of words. They extracted the key latent dimensions of consumer satisfaction to conduct brand positioning analysis. Büschken and Allenby (2016) proposed a topic model that considers sentence-based topics rather than independently assigning words to topics to analyze the importance of each product attribute to customer satisfaction.

As briefly discussed in previous sections, these topic modeling approaches also have limitations in terms of interpretability. Because a topic model is based on unsupervised learning for topic assignments to words, some words with less semantic cohesion may be clustered in a topic. This clustering leads to a meaningless analysis of the relationship between product attributes and satisfaction or helpfulness, which is the ultimate goal of using the topic model.

In this study, we propose a semi-supervised topic model that provides representative words for a product attribute as supervision. The model can also assign topics to each word based on these seed words. Similar approaches using seed words or seed labels have been proposed in the literature. Lin et al. (2012) and Tirunillai and Tellis (2014) used the seed words approach to estimate word sentiments by fixing some word sentiments to either polarity, such as good and great as positive words, and bad and terrible as negative words. Additionally, Ramage et al. (2009) proposed for the labeled LDA model that some tags previously assigned to documents are considered to be seed topics, and the word topics in the documents are determined from the set of seed topics. Therefore, the proposed model can be said to be an application of the approach used in these previous studies to assign topics to words.

In this study, we adopt the supervised topic model (Blei and McAuliffe, 2007) to estimate the relationships among product attributes extracted by the labeled LDA model and customer satisfaction and reader helpfulness. Some studies integrated the text information quantified by the LDA into regression models, such as linear regression and multinomial probit, to treat the topic extraction part and the regression part as separate models (Qi et al.,

2016; Bi et al., 2019), whereas others treated them together as a single model (Büschken and Allenby, 2016; Puranam et al., 2017). We take the latter position but compare both approaches in the section of the empirical analysis.

3 Data

In this section, we explain the details of the dataset used in this study and construct "labeled" topics. We use an Amazon customer review dataset collected by the authors that consists of 1,178 customer reviews on a specific potato chip product on Amazon platform from March 2009 to October 2019. Selecting a single product facilitates the assumption of perceived topics on the point of product features.

This dataset includes five variables, review texts, product rating scores, voted helpfulness counts, and some control variables. Specifically, after buying and accepting goods, customers are allowed to post their feelings and experiences by writing a textual review. At the same time, these customers may grade products by assigning "star" ratings from 1 to 5. The reader of this online review can evaluate whether the review is helpful to his or her product consideration and purchase decision by clicking on the "helpfulness" button. As control variables, we consider four types of status badges: purchase verification, top contributor, top reviewer, and vine voice. Purchase verification indicates that the customer who wrote the review was verified for purchasing on the Amazon platform. The top contributor badge is awarded to customers who frequently share reviews or answers questions from other customers. Top reviewers are identified by Amazon's reviewer rankings, but the ranking system algorithm is not clarified. Vine voice is an invitation program that gives Amazon reviewers advance access to not-yet-released products for the purpose of writing reviews. Therefore, these variables also may have some impact on product ratings and readers' attitudes, as well as perceived topics. In the next section, we construct a statistical model to unveil these relationships.

Before the data analysis, review texts were preprocessed. First, for each document, we split the text into word sets and substituted capital letters with lower-case ones. Next, we excluded numbers, punctuation, and popular stop words (e.g., a, the, I, they). We also transformed inflected words to their word stems to reduce the redundancy of the created corpus. Finally, we excluded frequently used and rare words because they adversely affect the extraction of topics or the interpretation of the topics after extraction. As a result, the created corpus included 716 unique words, and each review consisted of an average of 9.8 words.

The purpose of this study is to extract objective topics from the textual review and explore the connections among the topics, review writers' satisfaction, and readers' perceived helpfulness. This study aims to assist online retailers and marketers in developing and improving their products according to specific discovered relationships between objective product features and satisfaction or helpfulness.

To extract objective topics, we assume a certain number of perceived topics on features of food. Scholars mentioned several features of food, such as culinary quality (Chi et al., 2013), taste and flavor (Andersen and Hyldig, 2015), service and price (Mason and Nassivera, 2013), health/nutrition, and low-fat ingredients (Küster and Vila, 2017). At the same time, to identify the possible features of food in the dataset, we read subsets of customer reviews and observe the frequently used words of the entire vocabulary. Based on observations, the features of food are related to flavor, taste, packing, weight, food ingredients, cooking method, and purchases. According to the previous related literature and the observed product features in the dataset, we combine synonym topics and propose five perceived topics: (1) flavor and taste, (2) packing, (3) healthy, (4) money and buying, and (5) ingredient. In this study, we need to select seed words for each labeled topic, and these words should be symbolic and representative of each product feature. We used the literature and the subset of customer reviews to select the seed words for each labeled topic. Table 1 provides a list of labeled words.

Flavor and Taste:	salt, vinegar, flavor, cheddar, test, bbq, crunchy, texture, tasty, sweet, pepper, salty
Packing:	case, box, product, store, bag, pack, weight
Healthy:	healthy, fat, calorie
Money and Buying:	try, buy, get, bought, find, price
Ingredient:	oil, flour, ingredient, starch, flake

Table 1: List of labeled words for each product feature

4 Model

4.1 Supervised Learning Model with Labeled Topics

To improve the interpretability of the extracted topics in the text analysis, we first employ a labeled topic model. "Labeled" means that some labeled words are fixed to their respective topics to distinguish meanings among latent topics, where topics represent product attributes in customer reviews. The meanings of extracted topics are easily interpreted if the synonyms are assigned in common topics. For example, "hungry," "famished," and "starving" are assigned to the common topic "hunger." Second, to supervise the model, we use other information sources, where "supervise" means that the observed numeric data—including ordered categorical satisfaction score and positive integer helpfulness count—work as dependent variables, and the frequencies of the topics in customer review are covariates. In this article, we propose the supervised-labeled LDA (SL-LDA) model, which is the combination of two types of models, the LDA model with labeled words and the supervised topic model.

First, we demonstrate how to accomplish the labeled perceived topics using labeled words. After removing unnecessary words, we consider the total vocabulary, which includes all words in customer reviews as text content. To make topics interpretable, some meaningful and frequently used *labeled words* are selected from the total vocabulary and fixed in respective perceived topics, as discussed in the previous section. Simultaneously, remaining words with

$$
\begin{array}{c}
\phantom{\text{Topic 1, }\beta_1^*}
\begin{matrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \end{matrix} \\
\begin{matrix}
\text{Topic 1, } \beta_1^* \\
\text{Topic 2, } \beta_2^* \\
\text{Topic 3, } \beta_3^* \\
\text{Topic 4, } \beta_4^* \\
\text{Topic 5, } \beta_5^*
\end{matrix}
\left(
\begin{matrix}
\beta_1 & \times & \times & \times & \times & \beta_6 & \beta_7 \\
\times & \beta_2 & \times & \times & \times & \beta_6 & \beta_7 \\
\times & \times & \beta_3 & \times & \times & \beta_6 & \beta_7 \\
\times & \times & \times & \beta_4 & \times & \beta_6 & \beta_7 \\
\times & \times & \times & \times & \beta_5 & \beta_6 & \beta_7
\end{matrix}
\right)
\end{array}
$$

Figure 1: Labeled and non-labeled words for each topics

no pre-defined labels are called *non-labeled words* and assigned to topics according to a specific distribution.

Figure 1 provides an example of constructing labeled topics. Given five topics on product attributes and the total vocabulary of v_1, v_2, v_3, v_4, v_5, v_6, and v_7, we assume that word v_1 works as the labeled word of topic 1 and, thus, is excluded from the vocabularies of the other four topics. Other labeled words, v_2, v_3, v_4, and v_5, are specified as labeled words for the remaining topics. The non-labeled words, v_6 and v_7, can be assigned to any topic with a certain probability. After establishing the vocabulary for topic k, where $k = 1, 2, \ldots, 5$ according to the LDA model, we expect that customers choose words from the vocabulary of topic k with specific probabilities to express their feelings for the product attribute of topic k.

To develop the vocabulary for topic k from the total vocabulary, we imitate and simplify the methodology from the literature, such as labeled LDA (Ramage et al., 2009) and joint sentiment LDA (Lin et al., 2012; Tirunillai and Tellis, 2014). We first generate the topic's *transformation matrix* $\Lambda^{(k)}(V_k \times V)$ conditional on λ_1, λ_2, $\ldots$, λ_{V_k}, where λ_1, λ_2, $\ldots$, λ_{V_k} is the sequence of the number of labeled words for topic k and non-labeled words. Similar to 1, 6, and 7 for topic 1 in Figure 1, $V_k = 3$ and $V = 7$. For each row $i \in \{1, 2, \ldots, V_k\}$ and

column $j \in \{1, 2, \ldots, V\}$,

$$\Lambda_{i,j}^{(k)} = \begin{cases} 1 & \text{if } \lambda_i = j \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In this model, the topic-vocabulary vector ϕ_k, called word distribution, is assumed to follow the Dirichlet distribution, $\phi_k \sim Dirichlet(\beta_k^*)$, where β_k^* is transformed from the V-dimensional hyper parameter β via the transformation matrix, $\beta_k^* = \Lambda^{(k)}\beta$. Hence, the length of β_k^* is V_k.

The remaining Labeled LDA part of the proposed model is similar to the LDA model. That is, the topic assignment for the n-th word in review d is assumed to follow the categorical distribution, $z_{dn} \sim Categorical(\theta_d)$, where θ_d is a topic distribution for review d and follows the Dirichlet prior distribution, $\theta_d \sim Dirichlet(\alpha)$. We assume that a word assigned to topic k is generated from the corresponding word distribution, $w_{dn} \mid z_{dn} = k \sim Categorical(\phi_k)$.

Next, we develop the "supervised" part of our model. In this model, we use two dependent variables: satisfaction score and helpfulness count. The satisfaction score reflects the current evaluation by customers who already purchased the product based on their past experiences, and we assume that helpfulness counts imply the interest and expectations of the product by other consumers who may purchase it in the future. Additionally, as discussed in the previous section, the labeled topics in customer reviews can directly serve as covariates for the variations in the satisfaction scores and as helpful product expectation references. For example, if one product feature is particularly satisfying to customer needs, we expect that the topic related to this feature co-occurs with a high satisfaction score in online customer reviews. In contrast, topics related to dissatisfying features should occur with low satisfaction scores. Regarding the connection between topic assignment and review helpfulness, a textual review is likely to be regarded as helpful by readers if it contains the topics in which they are interested.

Given the word-topic assignment, z_{dn}, $N_{dl} = \sum_{n=1}^{N_d} \mathbb{I}\{z_{dn} = l\}$ is the number of words assigned to topic l and works as explanatory variables after a logarithm transformation. Note that the number of explanatory variables for these topic assignments is the same as the number of labeled topics, L (set to five in this study). However, the number of total topics can be different from the number of labeled topics. Let the number of total topics be K, and the rest of the topics, $K - L$, are non-labeled and do not include any labeled words. Because non-labeled topics are extracted from the review text but not considered to be explanatory variables, they work as topics for collecting outlier words or topics in the text and do not affect the following regression analysis even if their cohesion and meanings are difficult for humans to interpret.

The number of total topics, K, is determined by comparing models with varying numbers using deviance information criterion (DIC) and widely applicable information criterion (WAIC), as is demonstrated in Section 5. In addition to the topic assignment variables, some control variables also work as explanatory variables. In this study, we consider the following four status variables: purchase verification, top contributor, top reviewer, and vine voice. The status badges are displayed next to the user icon if the user qualifies for the status. Word count variables, which represent the number of words included in the customer review, are also considered.

We assume that satisfaction scores measured using five point scales follow the ordered probit model, and helpfulness counts that are positive integers follow the Poisson regression model. First, let the satisfaction score of review d be $y_{s,d}$, which follows the ordered probit model:

$$y_{s,d} = r, \quad \text{if } \tau_{r-1} \leq y_{s,d}^* < \tau_r \tag{2}$$

$$y_{s,d}^* = \sum_{l=1}^{L} \gamma_{s,l} \cdot \log\left(N_{dl} + 1\right) + \sum_{m=1}^{5} \delta_{s,m} \cdot x_{s,dm} + \epsilon_d, \tag{3}$$

where the thresholds $\{\tau_r\}$ work for realizing discrete satisfaction scores through the latent

continuous variable, $y_{s,d}^*$, and thresholds τ_0 and τ_R ($R = 5$ in Amazon data set) are set to $-\infty$ and ∞, respectively. $x_{s,d}$ is a vector of the control variables—status variables and word counts. Additionally, to identify the SL-LDA model, the error term ϵ_d is assumed to follow the standard normal distribution and does not include the intercept term. Therefore, we can freely estimate the remaining $R - 1$ thresholds. Next, let the helpfulness count for review d be $y_{h,d}$, which follows the Poisson regression model:

$$y_{h,d} \sim Poisson(y_{h,d}^*) \tag{4}$$

$$y_{h,d}^* = \sum_{l=1}^{L} \gamma_{h,l} \cdot \log\left(N_{dl} + 1\right) + \sum_{m=1}^{6} \delta_{h,m} \cdot x_{h,dm}, \tag{5}$$

where $x_{h,d}$ is a vector of control variables—status variables, word counts, and satisfaction score. The explanatory variables for the helpfulness regression model have much common with the variables of the satisfaction probit model, except that the effect of the satisfaction scores on helpfulness is considered, which is based on the findings in the literature (e.g. Ho-Dac et al., 2013; Mauri and Minazzi, 2013; Ludwig et al., 2013). These studies showed that positive and negative online customer reviews affect readers' purchase intentions and expectations. We also explore the effect of the valence of customer reviews on readers' helpfulness which respect to the review.

Therefore, the full joint likelihood of SL-LDA model is as follows:

$$
\begin{aligned}
&p(W, Y_s, Y_h, X_s, X_h, Z, \theta, \phi, \tau, \alpha, \beta, \Lambda, \gamma_s, \gamma_h, \delta_s, \delta_h) \\
&= \prod_{d=1}^{D} \left\{ p(y_{s,d} \mid z_d, x_{s,d}, \gamma_s, \delta_s, \tau) p(y_{h,d} \mid z_d, x_{h,d}, \gamma_h, \delta_h) \prod_{n=1}^{N_d} p(w_{dn} \mid z_{dn}, \phi) p(z_{dn} \mid \theta_d) \right\} \\
&\quad \left\{ \prod_{d=1}^{D} p(\theta_d \mid \alpha) \right\} \left\{ \prod_{k=1}^{K} p(\phi_k \mid \Lambda^{(k)}, \beta) \right\} p(\gamma_s, \gamma_h, \delta_s, \delta_h), \tag{6}
\end{aligned}
$$

where $p(\gamma_s, \gamma_h, \delta_s, \delta_h)$ is the prior distribution for the regression coefficients, and the settings of the distribution and hyper parameters are provided in the Appendix.

14

4.2　Model Estimation

In this section, we derive the estimation procedure of the SL-LDA model from the joint distribution (6). First, we apply the collapsed Gibbs method (Griffiths and Steyvers, 2004) for sampling the topic assignment variable z by integrating out topic distribution θ and word distribution ϕ. The conditional probability density of topic assignment $z_{dn} = k$ is given as:

$$
p(z_{dn} = k \mid w_{dn} = v_k, W_{\backslash dn}, y_{s,d}^*, y_{h,d}, x_{s,d}, x_{h,d}, Z_{\backslash dn}, \alpha, \beta^*, \gamma_s, \gamma_h, \delta_s, \delta_h, \tau)
$$

$$
\propto \left(N_{dk \backslash dn} + \alpha \right) \frac{N_{kv_k \backslash dn} + \beta_{kv_k}^*}{\sum_{v=1}^{V_k} N_{kv \backslash dn} + \beta_{kv}^*} p(y_{s,d}^* \mid z_{dn} = k, x_{s,d}, \gamma_s, \delta_s, \tau) p(y_{h,d} \mid z_{dn} = k, x_{h,d}, \gamma_h, \delta_h),
$$

$$
\tag{7}
$$

where N_{kv} represents the counts of word v that are assigned to topic k, and the symbol $\backslash$ represents the exclusion of the word from the counts. $p(y_{s,d}^* \mid \cdot)$ and $p(y_{h,d} \mid \cdot)$ are the probability density function of the normal distribution and the Poisson distribution, respectively.

Next, for the ordered probit model of satisfaction scores, we apply Gibbs sampling with data augmentation (Albert and Chib, 1993). Using the results from the literature, the conditional densities of the regression coefficients γ_s and δ_s, the augmented continuous satisfaction $y_{h,d}^*$, and the threshold parameters τ are multivariate normal, truncated normal, and uniform distributions, respectively. The details of the posterior density and Markov chain Monte Carlo (MCMC) procedure are provided in the Appendix.

Finally, we employ the random walk Metropolis-Hastings algorithm to estimate the helpfulness-Poisson regression parameters—the regression coefficients γ_h and δ_h. The joint conditional density of γ_h and δ_h is given by the product of the Poisson density for Y_h and the normal density for the prior distribution. Because the constant term of this posterior density is unknown and obtaining samples from the posterior is not easy, we employ the Metropolis-Hastings algorithm for sampling from the posterior. The proposal density is a normal distribution with a mean vector of previous samples.

Therefore, the MCMC process by Gibbs and Metropolis-Hastings samplings is as follows.

First, we initialize all of the parameter values and then execute the sampling from the posterior according to the previously described densities. After repeated sampling until all of the parameters converge, we calculate the estimates and the highest posterior density for each parameter using the samples after the burn-in period. Additionally, to estimate the expectation of the parameters integrated out in (7), θ and ϕ, the results of the topic assignment are used: $\theta_{dk} = \frac{N_{dk}+\alpha}{N_d+\alpha \cdot K}$, $\phi_{kv} = \frac{N_{kv}+\beta^*_{kv_k}}{\sum_v N_{kv}+\beta^*_{kv}}$.

5   Empirical Analysis

5.1   Comparison Results

In this section, we compare the proposed model with the extant models to clarify the performance of the proposed model. The differences between these models are compared from two viewpoints. The first is model selection through the information criterion, and the second is the interpretability of the extracted topics. Through these two comparisons, we examine how well the proposed model can improve the interpretability of the topics against the existing models while improving the fit and prediction of the data. The Appendix contains the details of the estimation settings.

We consider two comparison models—the separate model and the supervised LDA model. In the separate model, two processes—extraction of product attributes from the text using the LDA model and the regression model for satisfaction and helpfulness using the Poisson regression and the ordered probit model—are separately conducted as divided another models. The supervised LDA model has no constraints related to the labeled words; that is, the model is equivalent to the proposed model that treats all words as non-labeled. For these models, the parameters are estimated using the Gibbs and Metropolis-Hastings sampling and the same setting as in the proposed model, as described in the previous section.

First, we discuss the results of the model comparison using the information criteria. To measure the model comparison, we use the DIC (Spiegelhalter et al., 2002) and the WAIC

(Watanabe, 2010). The two criteria compare models from different perspectives: the DIC considers the model's goodness of fit and complexity, and the WAIC assesses the model's generalization error. Figure 2 illustrates the values of the DIC and WAIC as the summation of those for two dependent variables of product satisfaction and readers helpfulness in the range of the number of topics from 5 to 15 for three models. The comparison of the separate and supervised models show that the separate model is inferior to the others for all numbers of topics and both criteria. Whereas both the separate and supervised models have been used in previous studies (e.g., Moon and Kamakura, 2017; Büschken and Allenby, 2016, respectively), the supervised model is superior in terms of model comparison based on the information criterion.

A comparison of the proposed model with the supervised LDA model shows that the supervised LDA model is better overall, even though the proposed model has smaller DIC and WAIC for certain numbers of topics. One possible explanation for this findings is that the proposed model has restrictions of labeled words, and the parameters with respect to these words are fixed in certain values and might deviate from values that achieve a better fit, whereas the supervised LDA model does not have such restrictions. Interestingly, the difference between both models is not large, especially when compared with the WAIC, for which both models are almost the same. This findings indicates that the impact of the labeled word constraints on generalization is relatively small, and the proposed model is worthwhile if the benefits from the improvements in topic interpretability by labeling seed words are greater than the impact of the restriction.

Next, we compare the proposed model with the supervised LDA model from the perspective of topic interpretability. Tables 2 and 3 provide the top 15 words distributed for each topic, in descending order. For a simple comparison, the number of topics is five for both models, which is the same as the number of expected perceived topics determined before the analysis. Table 2 shows that, for the proposed model, five interpretable topics extracted are attributable to the labeled words. In addition, the proposed model can assign seemingly
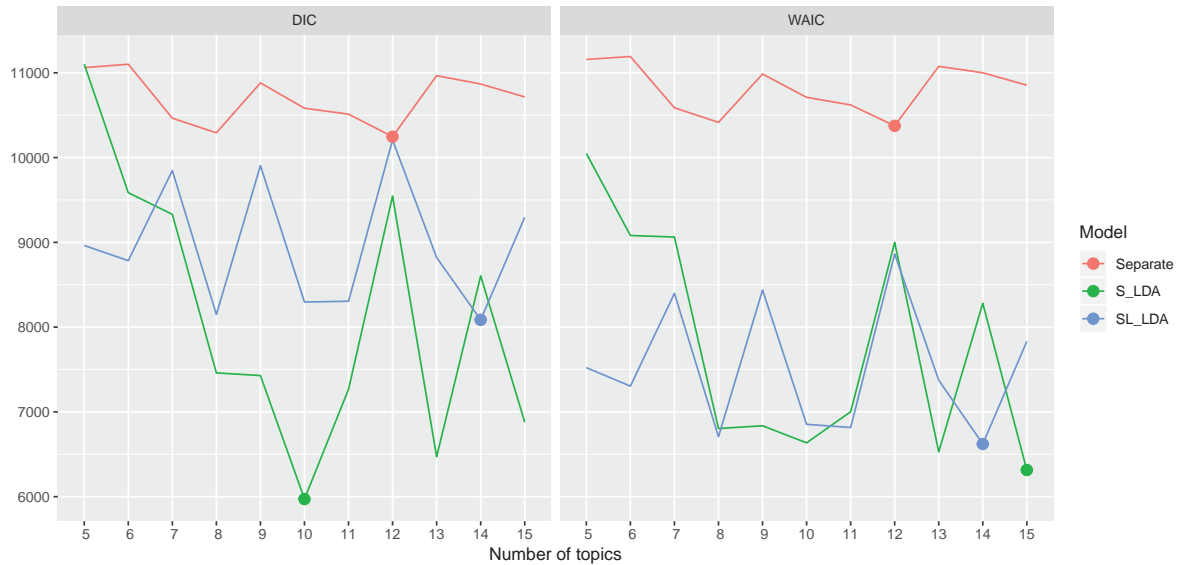
17

Figure 2: Values of DIC (left) and WAIC (right): Lines indicate comparison models, separate model (red), supervised LDA model (green), and proposed model (blue). Dots represent the smallest values among the variations in the number of topics for each model.

correct topics to words that are related to the labeled words and, hence, are appropriate to make up the same topic. For example, lime and chili are in the flavor and taste topic, size is in the packing topic, and diet is in the healthy topic; these words are non-labeled but appropriate for constituting the same topic along with the labeled words. In contrast, most topics in Table 3 for the supervised LDA model consist of meaningless sets of words, and the same words are related to multiple topics. For example, different topic words, such as weight, price, and calorie, are in the same topic 3, and the same words, such as bag and flavor, are related to multiple topics. Therefore, humans face difficulties interpreting the meanings of topics using the supervised LDA model.

In summary, these results show that the proposed model is reliable in both explaining the variation in product satisfaction and readers' helpfulness and interpretability of topics by labeling certain seed words as representative words for the topics. In the next section, we discuss the estimation results of the regression parameters.

Table 2: Top 15 words in descending order of word distribution for each topic in supervised labeled LDA model (asterisks next to words indicate labeled words)

Flavor & Taste	Packing	Healthy	Money & Buying	Ingredient
flavor*	bag*	calorie*	try*	ingredient*
taste*	box*	fat*	get*	oil*
salt*	case*	healthy*	buy*	rice
sweet*	product*	fry	find*	flour*
bbq*	pack*	delicious	bought*	definitely
vinegar*	store*	always	price*	corn
salty*	weight*	crave	order	kid
pepper*	size	per	origin	back
texture*	order	diet	know	thought
lime	whole	star	since	final
crunch	need	ever	bad	disappoint
crunchy*	watcher	recommend	healthier	kind
chili	variety	sodium	far	purchase
cheddar*	keep	lunch	greasy	say
tasty*	addict	look	addict	list

Table 3: Top 15 words in descending order of word distribution for each topic in supervised LDA model

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
sweet	bag	bag	flavor	flavor
fat	taste	find	taste	salt
taste	try	weight	try	vinegar
calorie	buy	buy	lime	taste
flavor	get	price	chili	try
rice	package	calorie	texture	bbq
healthy	box	store	expect	pepper
bag	order	healthy	though	origin
ingredient	pack	get	get	calorie
salty	bought	bbq	spicy	garlic
oil	purchase	size	product	cheddar
bad	never	order	need	order
flour	look	every	salt	prefer
fry	said	day	calorie	fat
sodium	people	keep	say	sour

5.2 Discussion of Estimation Results

Table 4 shows the estimation results of the proposed model when the number of total topics is 14, which is the smallest value of both criteria in Figure 2. Table 4 provides that the estimated posterior mean of the thresholds parameters and the regression coefficients of labeled topics and control variables. The asterisks next to the posterior mean indicate the parameter significance calculated by the 95% highest posterior density.

First, the threshold parameters (τ_r) indicate that an approximate 0.50 increase in the latent continuous rating is associated with a one-point increase in the observed discrete rating. Because the explanatory variables for the topics represent the log-transformed number of words assigned to the topic, the regression coefficients on the labeled topics (γ_l) can be interpreted as substantial changes in the ratings for a 1% increase in the number of words for that topic. For example, if the number of words associated with the topic "flavor and taste" increases by 1%, the expected change in the latent rating is -0.58, translating to an almost one-point decline in customer satisfaction.

Table 4 also provides some interesting findings related to the regression coefficients. The coefficients of topics 1, 2, and 5 for the satisfaction score are negatively estimated, and only the coefficients of topic 3 are positively estimated. This finding indicates that dissatisfied customers are more likely to talk about "flavor and taste," "packing," and "ingredient" topics, and satisfied customers are more likely to talk about the "health" topic. Additionally, significant coefficients of the helpfulness regression are related to the health and ingredient topics. This result suggests that reviews including "health" topic words are regarded as helpful, and reviews including "ingredients" topic words are less helpful.

Remarkable findings are also found in the coefficients for the control variables (δ_m). Because the coefficients for the purchase verification are positively significant for both objective variables, customers who purchased the product can be viewed as being satisfied relative to customers who do not make a purchase, and readers find that reviews by customers who made purchase are more helpful. Similarly, more satisfied customers write longer reviews,

Table 4: Estimation results of proposed model

| Parameters | Posterior Mean (95 % HPD interval) | |
	Satisfaction Score	Helpfulness Count
τ_1	-2.095^* $(-2.340, -1.843)$	—
τ_2	-1.635^* $(-1.808, -1.425)$	—
τ_3	-1.102^* $(-1.272, -0.937)$	—
τ_4	-0.587^* $(-0.761, -0.392)$	—
γ_1 (Flavor & Taste topic)	-0.584^* $(-0.773, -0.404)$	0.081 $(-0.018, 0.182)$
γ_2 (Packing topic)	-0.382^* $(-0.605, -0.160)$	0.009 $(-0.123, 0.139)$
γ_3 (Health topic)	1.046^* $(0.797, 1.412)$	0.238^* $(0.050, 0.439)$
γ_4 (Money & Buying topic)	0.075 $(-0.190, 0.338)$	-0.114 $(-0.274, 0.044)$
γ_5 (Ingredient topic)	-1.612^* $(-2.147, -1.115)$	-0.274^* $(-0.535, -0.019)$
δ_1 (Purchase verification)	0.294^* $(0.073, 0.502)$	0.234^* $(0.059, 0.410)$
δ_2 (Top contributor)	0.678 $(-0.852, 2.161)$	1.004^* $(0.234, 1.773)$
δ_3 (Top reviewer)	0.748 $(-1.494, 3.128)$	-0.489 $(-1.682, 0.707)$
δ_4 (Vine voice)	0.593 $(-0.924, 2.043)$	0.076 $(-0.742, 0.931)$
δ_5 (Word counts)	0.012^* $(0.006, 0.018)$	0.004^* $(0.001, 0.007)$
δ_6 (Satisfaction score)	—	-0.152^* $(-0.190, -0.113)$

and such reviews are regarded as helpful for readers. The last finding is the negative effect of the satisfaction score on readers' helpfulness. Readers tend to find critical reviews with low satisfaction scores more helpful than positive and satisfied reviews.

In conclusion, these results answer our research questions: "What is the relationship between product attributes and the satisfaction of customers who make purchases?" and "What attributes do readers of the customer review expect to find as useful information?" Additionally, these findings are obtained by labeling seed words and extracting interpretable perceived topics, and paying attention to interpretability when using topic models in customer review analysis may be necessary.

6 Conclusion

We introduced a supervised labeled latent Dirichlet allocation (SL-LDA) model that combines word labeling to extract interpretable perceived topics with the supervised LDA to explore the structure of the satisfaction of experienced customers and the expectation of review readers. To obtain stably interpretable latent topics in online customer reviews, priori labeled words related to product attributes are assigned respective topics as semi-supervised learning. Accordingly, we connect product attributes with customer satisfaction as feedback from past customers and consumer interests for products as perceived helpfulness of future customers through supervised learning.

The model comparison in Section 5.1 shows that the proposed model is not only reliable in explaining the variations in customer satisfaction and readers' helpfulness but also can interpret topics relative to the extant model. When comparing the proposed model with the supervised LDA model, the difference between both models is not large in terms of model selection through the information criterion. Furthermore, a comparison of the word sets of extracted topics shows that the words obtained in the proposed model are easy to interpret because they consist of labeled words and other words assigned to the topic according to the labeled words. In contrast, because the supervised LDA model has no such restriction, we may obtain a set of words whose cohesion is incomprehensible to humans. Therefore, the proposed model is worthwhile if the benefits from improving the topic interpretability using word labeling are greater than the impact of the restriction.

In this empirical study, we find that the "flavor and taste," "packing," and "ingredient" topics are mentioned by dissatisfied customers, and reviews including the "ingredient" topic are likely recognized as unhelpful by readers of the reviews. In contrast, "health" topic has a positive effect on both customer satisfaction and perceived helpfulness. Companies and marketers can use the proposed model to understand the relationship between product attributes and the satisfaction of customers making purchases and the attributes that readers of customer reviews expect to find as useful information.

Future research is needed in two aspects related to marketing improvement. One aspect is improving the product attribute analysis through a consideration of consumer heterogeneity (Xiao et al., 2016) and word sentiments (Decker and Trusov, 2010; Archak et al., 2011) to develop a more sophisticated model. Another aspect is using the topic model to predict more "supervised" objectives such as detecting fake or deceptive reviews (Qi et al., 2016), which are meaningless reviews and high score ratings after false purchase behavior.

References

Albert, J. H. and Chib, S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.

Andersen, B. V. and Hyldig, G. Food satisfaction: Integrating feelings before, during and after food intake. *Food Quality and Preference*, 43:126–134, 2015.

Archak, N., Ghose, A., and Ipeirotis, P. G. Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 57(8):1485–1509, 2011.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., and Schweidel, D. A. Uniting the Tribes: Using Text for Marketing Insight. *Journal of Marketing*, 84(1):1–25, 2020.

Bi, J.-W., Liu, Y., Fan, Z.-P., and Zhang, J. Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management*, 70:460–478, 2019.

Blei, D. M. and McAuliffe, J. D. Supervised topic models. In *Advances in neural information processing systems*, pages 121–128, 2007.

Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(4-5):993–1022, 2003.

Büschken, J. and Allenby, G. M. Sentence-based text analysis for customer reviews. *Marketing Science*, 35(6):953–975, 2016.

Chen, Y. and Xie, J. Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix. *Management Science*, 54(3):477–491, 2008.

Chi, C. G.-Q., Chua, B. L., Othman, M., and Karim, S. A. Investigating the Structural Relationships Between Food Image, Food Satisfaction, Culinary Quality, and Behavioral Intentions: The Case of Malaysia. *International Journal of Hospitality & Tourism Administration*, 14(2):99–120, 2013.

Decker, R. and Trusov, M. Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4):293–307, 2010.

Ghose, A., Ipeirotis, P. G., and Li, B. Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science*, 31(3): 493–520, 2012.

Griffiths, T. L. and Steyvers, M. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Supplement 1):5228–5235, 2004.

Ho-Dac, N. N., Carson, S. J., and Moore, W. L. The Effects of Positive and Negative Online Customer Reviews: Do Brand Strength and Category Maturity Matter? *Journal of Marketing*, 77(6):37–53, 2013.

Hou, T., Yannou, B., Leroy, Y., and Poirson, E. Mining Changes in User Expectation Over Time From Online Reviews. *Journal of Mechanical Design*, 141(9), 2019.

Küster, I. and Vila, N. Health/Nutrition food claims and low-fat food purchase: Projected personality influence in young consumers. *Journal of Functional Foods*, 38:66–76, 2017.

Lee, T. Y. and Bradlow, E. T. Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48(5):881–894, 2011.

Lin, C., He, Y., Everson, R., and Rüger, S. Weakly supervised joint sentiment-topic detection from text. *IEEE Transactions on Knowledge and Data Engineering*, 24(6):1134–1145, 2012.

Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., and Pfann, G. More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing*, 77(1):87–103, 2013.

Mason, M. C. and Nassivera, F. A Conceptualization of the Relationships Between Quality, Satisfaction, Behavioral Intention, and Awareness of a Festival. *Journal of Hospitality Marketing & Management*, 22(2):162–182, 2013.

Mauri, A. G. and Minazzi, R. Web reviews influence on expectations and purchasing intentions of hotel potential customers. *International Journal of Hospitality Management*, 34: 99–107, 2013.

Moon, S. and Kamakura, W. A. A picture is worth a thousand words: Translating product reviews into a product positioning map. *International Journal of Research in Marketing*, 34(1):265–285, 2017.

Mudambi, S. M. and Schuff, D. What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly: Management Information Systems*, 34(1):185–200, 2010.

Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science*, 31(3):521–543, 2012.

Puranam, D., Narayan, V., and Kadiyali, V. The effect of calorie posting regulation on consumer opinion: A flexible latent dirichlet allocation model with informative priors. *Marketing Science*, 36(5):726–746, 2017.

Qi, J., Zhang, Z., Jeon, S., and Zhou, Y. Mining customer requirements from online reviews: A product improvement perspective. *Information and Management*, 53(8):951–963, 2016.

Ramage, D., Hall, D., Nallapati, R., and Manning, C. D. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. *Empirical Methods in Natural Language Processing*, (August):248–256, 2009.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical*

Methodology), 64(4):583–639, 2002.

Tirunillai, S. and Tellis, G. J. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, 51(4):463–479, 2014.

Watanabe, S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11: 3571–3594, 2010.

Xiao, S., Wei, C. P., and Dong, M. Crowd intelligence: Analyzing online product reviews for preference measurement. *Information and Management*, 53(2):169–182, 2016.

Yin, D., Bond, S. D., and Zhang, H. Keep Your Cool or Let it Out: Nonlinear Effects of Expressed Arousal on Perceptions of Consumer Reviews. *Journal of Marketing Research*, 54(3):447–463, 2017.

Appendix

A Posterior Distributions and Estimation Procedure

In this section, we describe the details of posterior distributions for the proposed model and the MCMC algorithm. Equations for sampling from posterior distributions of the satisfaction regression parameters are as follows:

$$p(\gamma_s \mid Y_s^*, Z, X_s, \delta_s, g_{s,0}) \sim N(\mu_{\gamma_s}, \Sigma_{\gamma_s}),$$

$$\Sigma_{\gamma_s} = \left(\sum_{d=1}^{D} \log(N_{d\cdot} + 1) \log(N_{d\cdot} + 1)^{\top} + g_{s,0}^{-1} \cdot I \right)^{-1},$$

$$\mu_{\gamma_s} = \Sigma_{\gamma_s} \left(\sum_{d=1}^{D} \log(N_{d\cdot} + 1) \left(y_{s,d}^* - \sum_{m=1}^{5} \delta_{s,m} x_{s,dm} \right) \right) \tag{8}$$

$$p(\delta_s \mid Y_s^*, Z, X_s, \gamma_s, d_{s,0}) \sim N(\mu_{\delta_s}, \Sigma_{\delta_s}),$$

$$\Sigma_{\delta_s} = \left(\sum_{d=1}^{D} x_{s,d} x_{s,d}^{\top} + d_{s,0}^{-1} \cdot I \right)^{-1},$$

$$\mu_{\delta_s} = \Sigma_{\delta_s} \left(\sum_{d=1}^{D} x_{s,d} \left(y_{s,d}^* - \sum_{l=1}^{L} \gamma_{s,l} \log(N_{dl} + 1) \right) \right) \tag{9}$$

$$p(y_{s,d}^* \mid y_{s,d}, z_d, x_d, \gamma_s, \delta_s, \tau) \sim N \left(\sum_{l=1}^{L} \gamma_{s,l} \log(N_{dl} + 1) + \sum_{m=1}^{5} \delta_{s,m} x_{s,dm}, 1 \right),$$

$$\text{truncated to } (\tau_{r-1}, \tau_r] \text{ if } y_{s,d} = r \tag{10}$$

$$p(\tau_r \mid Y_s, Y_s^*, \tau_q) \sim U[\tau_{\text{lhs}}^*, \tau_{\text{rhs}}^*], \quad r = 1, \ldots, R-1, q \neq r$$

$$\tau_{\text{lhs}}^* = \max \left(\max\{y_{s,d}^*; y_{s,d} = r\}, \tau_{r-1} \right)$$

$$\tau_{\text{rhs}}^* = \min \left(\min\{y_{s,d}^*; y_{s,d} = r+1\}, \tau_r \right) \tag{11}$$

The parameters for helpfulness regression are sampled Metropolis-Hastings method, and the proposal density is the normal distribution with mean of own values in the previous iteration, $\gamma_h^{(t)} \sim N(\gamma_h^{(t-1)}, \sigma_{\gamma_h}^2 \cdot I)$. $\sigma_{\gamma_h}^2$ is a step size parameter whose value is adjusted in the MCMC procedure so that the acceptance rate falls into the range from 30% to 50%. Because

the proportion of proposal densities can be canceled, the acceptance ratio in sampling $\gamma_h^{(t)}$ consists of the proportion of posterior distributions:

$$\alpha_{\gamma_h} = \min\left\{1, \frac{p(Y_h \mid Z, X_h, \gamma_h^{(t)}, \delta_h)p(\gamma_h^{(t)} \mid g_{h,0})}{p(Y_h \mid Z, X_h, \gamma_h^{(t-1)}, \delta_h)p(\gamma_h^{(t-1)} \mid g_{h,0})}\right\},$$

$$p(Y_h \mid Z, X_h, \gamma_h^{(t)}, \delta_h) = \prod_{d=1}^{D} Po(y_{h,d}; \sum_{l=1}^{L} \gamma_{h,l}^{(t)} \log(N_{dl} + 1) + \sum_{m=1}^{6} \delta_{h,m} x_{h,dm})$$

$$p(\gamma_h^{(t)} \mid g_{h,0}) = N(\gamma_h^{(t)}; 0, g_{h,0}^{-1} \cdot I) \tag{12}$$

δ_h is also sampled in the same way.

Therefore, the algorithm of MCMC procedure is as follows:

1. initialize $Z, \gamma_s, \delta_s, Y_s^*, \tau, \gamma_h, \delta_h, \sigma_{\gamma_h}, \sigma_{\delta_h}$

2. iterate sampling until all parameters converged

   (a) sample Z according to equation (7)

   (b) sample $\gamma_s, \delta_s, Y_s^*, \tau$ according to equations (8) to (11)

   (c) update γ_h and δ_h with the acceptance ratio (12)

   (d) adjust σ_{γ_h} and σ_{δ_h} if the cumulative number of the acceptance falls outside the desired percentage

3. calculate the expectations of θ and ϕ using the last samples of Z

In Section 5, we repeat the above MCMC process 50,000 times, and then we use 25,000 samples excluding the burn-in samples to calculate the posterior means and the intervals of

27

highest posterior density. The settings of prior distribution used in Section 5 are as follows:

$$\theta_d \sim Dirichlet(\alpha), \quad \alpha_k = 1.0 \ \forall k$$

$$\phi_k \sim Dirichlet(\Lambda^{(k)}\beta), \quad \beta_v = 1.0 \ \forall v$$

$$\gamma_s \sim N(0, g_{s,0}^{-1} \cdot I), \quad g_{s,0} = 0.1$$

$$\delta_s \sim N(0, d_{s,0}^{-1} \cdot I), \quad d_{s,0} = 0.1$$

$$\gamma_h \sim N(0, g_{h,0}^{-1} \cdot I), \quad g_{h,0} = 0.1$$

$$\delta_h \sim N(0, d_{h,0}^{-1} \cdot I), \quad d_{h,0} = 0.1$$