Industrial Engineering
& Management Systems
Vol 14, No 4, December 2015, pp.335-346
ISSN 1598-7248 │EISSN 2234-6473 │
http://dx.doi.org/10.7232/iems.2015.14.4.335

A New Latent Class Model for Analysis of Purchasing and Browsing Histories on EC Sites

Masayuki Goto*, Kenta Mikawa, Shigeichi Hirasawa
Department of Industrial and Management Systems Engineering, Waseda University, Tokyo, Japan

Manabu Kobayashi
Department of Information Engineering, Shonan Institute of Technology, Yokohama, Japan

Tota Suko
School of Social Sciences, Waseda University, Tokyo, Japan

Shunsuke Horii
Global Education Center, Waseda University, Tokyo, Japan

(Received: June 12, 2015 / Revised: November 7, 2015 / Accepted: November 20, 2015)

ABSTRACT

The electronic commerce site (EC site) has become an important marketing channel where consumers can purchase many kinds of products; their access logs, including purchase records and browsing histories, are saved in the EC sites' databases. These log data can be utilized for the purpose of web marketing. The customers who purchase many product items are good customers, whereas the other customers, who do not purchase many items, must not be good customers even if they browse many items. If the attributes of good customers and those of other customers are clarified, such information is valuable as input for making a new marketing strategy. Regarding the product items, the characteristics of good items that are bought by many users are valuable information. It is necessary to construct a method to efficiently analyze such characteristics. This paper proposes a new latent class model to analyze both purchasing and browsing histories to make latent item and user clusters. By applying the proposal, an example of data analysis on an EC site is demonstrated. Through the clusters obtained by the proposed latent class model and the classification rule by the decision tree model, new findings are extracted from the data of purchasing and browsing histories.

Keywords: Web Marketing, Big Data, Latent Class Model, Aspect Model, Customer Segmentation, Business Analytics

* Corresponding Author, E-mail: goto@it.mgmt.waseda.ac.jp

1. INTRODUCTION

The electronic commerce site (EC site) has become an important marketing channel where consumers can purchase many kinds of products; their access logs, including purchase records and browsing histories, are saved in the EC sites' databases. These log data can be utilized for the purpose of web marketing (Hughes, 2006). For the purpose of marketing, it is valuable to analyze the purchasing and browsing behaviors of customers. The customers who purchase many product items are considered good customers (Curry *et al.*, 2000), whereas the other customers who do not purchase many items must not be good customers, even if they browse many items. If the attributes of good customers and those of other customers (normal and dormant customers) are clarified, such information can be valuable as input for making a new marketing strategy. Regarding the product items, the characteristics of good items that are bought by many users are valuable information. However, the numbers of product items and customers are quite large, and there are many attributes of customers and product items. It is, therefore, necessary to construct a method to efficiently analyze the relation between the customer attributes and purchasing and browsing activities.

The effectiveness of the latent class model, which is a discrete type of the latent variable model, has been described in many studies (Green *et al*., 1976; Swait and Adnmowicz, 2001; Bhatnagar and Ghose, 2004; Train, 2009). The latent class model assumes hidden latent classes for users and items. By introducing a latent class model, the latent user and item clusters can be modeled, and this assumption is consistent with marketing models (Train, 2009). A well-known latent class model is the aspect model, which was proposed to represent the relation between documents and words (Hofmann, 1999). The original aspect model proposed by Hofmann (1999) was a model of the statistical structure between documents and words for the purpose of information retrieval and document analysis. In order to apply the aspect model to a recommender system on EC sites, the relation between users, items, and preference values was modeled by introducing a discrete variable of a latent class for predicting individual choices and preferences (Hofmann and Puzicha, 1999). This model was extended to the model predicting users' ratings by introducing the Gaussian distribution (Hofmann, 2004). The flexible mixture model (FMM) is also an extended model that assumes two latent variables independently for users and items (Si and Jin, 2003; Jin *et al*., 2003). These models would be useful for customer segmentation (Green *et al.*, 1976) for marketing purposes. The clustering of customers is one of the important problems in the field of marketing (Magidson and Vermunt, 2002). However, these conventional models do not use the browsing history data and do not take account of the differences between purchasing and browsing activities.

This paper focuses on the analysis of customers and product items for customizing marketing strategies on an EC site. From the viewpoint of marketing, it is useful to analyze the differences of characteristics between good customers and others. Here, the available data saved in an EC site's database is not only purchasing history but also users' attributes and browsing history data. A customer has to create an account with a password to purchase the product items and register attributes such as address, sex, and age. These user attributes are saved in a database table. Regarding product items, the attributes of the item, such as item category, maker (brand), and shop, are also available. Moreover, it is desirable to utilize the information of users' browsing histories. This is because the number of purchased items is considerably smaller than that of browsed items. Some users may purchase only a small number of items, so that the purchasing history data may not be sufficient for determining a latent class model from the viewpoint of the relation between model structure and data size. Under those circumstances, we propose a new framework to analyze the data saved in the database on an EC site. The framework proposed in this paper includes the following distinctive points.

(1) In order to treat the browsing activities of customers, we propose a new latent class model expressing both purchasing and browsing activities directly in the model. The proposed model enables us to utilize not only the purchasing data but the huge browsing history data for model construction. It is desirable to construct a model expressing endogenously both purchasing and browsing activities, and it is advisable from the viewpoint of statistics.
(2) Additionally, we demonstrate a method for using the proposed model to make clusters of users and items. By using the proposed latent class model, the clusters of the good customers and the good product items are specified. The cluster sizes can be estimated by their probabilities of belonging to latent classes, and the conditional probabilities of purchasing and browsing activities can be useful for comprehending the estimated model and the meaning of each cluster.
(3) For the analysis of characteristics of each cluster, the decision tree model (Breiman *et al*., 1984) is introduced to investigate the relation between clusters given by the proposed latent class model and attributes of items and users. The CART algorithm can automatically construct a decision tree model expressing the relation between attributes and the latent classes of product items. Usually, there are many attributes of items and users. An efficient procedure to analyze the relation between the latent classes and attributes is provided by this analysis.

By the combination of the proposed latent class model and the decision tree, a tool for the efficient analysis of characteristics of item and customer clusters is given in practice. The proposed analysis framework can be useful for clustering product items and customers and characterizing each cluster by using attributes.

As an application of the proposal, an example of data analysis on an apparel EC site in Japan is demonstrated. This EC site provides a virtual shopping mall and apparel shops can open their virtual stores on the site by paying a store opening charge. In this site, the product items have the attributes of shops, brand, area, and item category. By applying the proposed model, the purchasing and browsing activities on the EC site can be modeled. As a result, the characteristics of each latent class of product items and users can be investigated from the viewpoint of the pattern of browsing and purchasing. On the application to the apparel EC site, we show the results obtained by applying the proposed latent class model and the analysis by the decision tree model. Through the proposed model and analysis approach, new findings are extracted from the data of purchasing and browsing histories. We show that effective discussions are possible on the basis of the result of our proposed analysis.

2. ANALYSIS FRAMEWORK

In this section, we show several related works and the overall framework of the proposed analysis.

2.1 Notations

A product item and a user (meaning a customer) are de-noted by $x \in \mathbf{A}$ and $y \in \mathbf{U}$, respectively. Here, $\mathbf{A} = \{a_1, a_2, \cdots, a_A\}$ is the set of product items and $U = \{u_1, u_2, \cdots, u_U\}$ is that of customers. The notations used in this study are listed here.

- $\mathbf{A} = \{a_1, a_2, \cdots, a_A\}$: the set of product items
- $U = \{u_1, u_2, \cdots, u_U\}$: the set of customers
- $x \in \mathbf{A}$: a product item
- $y \in \mathbf{U}$: a customer (user)
- $W_1 = \{0, 1, 2, \cdots, R_1\}$: the set of numbers of purchasing actions
- $W_2 = \{0, 1, 2, \cdots, R_2\}$: the set of numbers of browsing actions
- $w_1 = W_1$: the number of purchasing actions
- $w_2 = W_2$: the number of browsing actions
- $w = (w_1, w_2)$: the two-dimensional vector expressing both purchasing and browsing actions

2.2 Overall Framework to Analyze the Data of Purchasing and Browsing Activities

We propose a new framework to analyze the data saved in the database on an EC site, including the users' purchasing and browsing history data and the attribute information of customers and product items. The main procedure of the analysis proposed in this paper is composed of modeling by a new latent class model, soft clustering of product items and customers, and analyzing the characteristics of each cluster on the basis of decision tree model.

(1) **[Modeling by A Latent Class Model]** In order to treat the browsing activities of customers, a new latent class model expressing both purchasing and browsing activities directly in the model is introduced. The proposed model enables us to utilize not only the purchasing data but the immense browsing history data for model construction. It is desirable to construct a model expressing endogenously both purchasing and browsing activities, and it is also desirable from the viewpoint of statistics.

(2) **[Soft Clustering by the Latent Classes Model]** Additionally, we demonstrate a way to use the proposed model to make clusters of users and items. By using the proposed latent class model, the clusters of good customers and good items are specified. The cluster sizes can be analyzed by their probabilities of belonging to latent classes, and the conditional probabilities of purchasing and browsing activities can be useful for comprehending the estimated model and the meaning of each cluster.

(3) **[The Analysis of Characteristics of Each Cluster by Decision Trees]** For the analysis of characteristics of each cluster, the decision tree model (Breiman *et al*., 1984) is introduced to investigate the relation between clusters given by the proposed latent class model and attributes of items and users. The Classification and Regression Tree (CART) algorithm can automatically construct a decision tree model expressing the relation between attributes and the latent classes of product items. Usually, there are many attributes of items and users. An efficient procedure to analyze the relation between the latent classes and attributes is provided by this analysis.

2.3 Related Works with Latent Class Models

In the field of marketing science, the effectiveness of latent class models is well known. This is because market segmentation strategies are generally used to identify the target consumers in practice, and the concept of the discrete latent class is consistent with this fact. The market is usually characterized by an aggregate of many different consumer groups. The latent class model assumes this general situation, and each latent class represents a consumer group with similar preferences. Through experimental studies, many researchers have stated that the assumption of latent classes is reasonable in marketing problems.

A basic latent class model is the aspect model (Hofmann, 1999). Introducing a latent class variable $z \in \mathbf{V}$ to model the statistical relation between an item $x \in \mathbf{A}$ and a customer $y \in \mathbf{U}$, the conditional probability $P(x \mid y)$ of a latent class model is defined as follows:

$$P(x \mid y) = \sum_{z \in \mathbf{V}} P(x \mid z)\, P(z \mid y), \tag{1}$$

where the latent variable z takes a discrete value expressing a latent class on $\mathbf{V} = \{v_1, v_2, \cdots, v_z\}$. In this model, the conditional probabilities $P(x \mid z)$ and $P(z \mid y)$ are the parameters to be estimated by a set of training data. This model represents the purchase probability of each item $x \in \mathbf{A}$ by each user $y \in \mathbf{U}$ and does not contain the information about the probability of user $P(y)$. Rewriting the model by using the Bayes rule, the following model is given:

$$P(x, y) = \sum_{z \in \mathbf{V}} P(x \mid z) P(y \mid z) P(z). \tag{2}$$

This expression gives the form of a joint probability of $x \in \mathbf{A}$ and $y \in \mathbf{U}$, so that the log likelihood function given

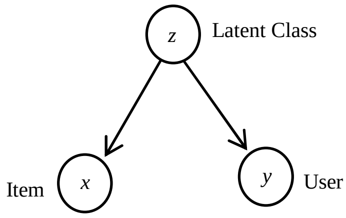

Figure 1. Graphical representation of the aspect model.

the training data $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ can be directly expressed as

$$L = \sum_{i=1}^{n} \log \sum_{z=\mathbf{V}} P(x_i \mid z) P(y_i \mid z)\ P(z). \tag{3}$$

This formulation of a latent class model is the aspect model proposed by Hofmann. The aspect model is also called the probabilistic latent semantic indexing (PLSI) model to represent a document-word statistical structure. The graphical model of the aspect model is shown in Figure 1.

The parameters of the aspect model, $P(x \mid z)$, $P(y \mid z)$, and $P(z)$, can be estimated from the training data set by applying the EM algorithm (Dempster *et al.*, 1977; McLachlan and Krishnan, 2007). The procedure of the Expectation–maximization algorithm (EM algorithm) for the aspect model is shown in Appendix A.

Many types of modified aspect models have been proposed. For the application to recommender systems, several models predicting users' ratings of each item have been proposed. The FMM is one of the most basic latent class models to predict users' ratings (Si and Jin, 2003). In FMM, the two independent latent classes for users and items are introduced. However, the purchasing and browsing history data cannot be treated in FMM because the prediction of users' rating is the purpose. In order to estimate the parameters of FMM with high accuracy, a modified learning algorithm using both purchasing and browsing history data is effective (Oi *et al.*, 2015). Suzuki *et al.* (2014) proposed a recommendation logic considering both predicted purchase probability and predicted ratings based on FMM.

Though many kinds of latent class models have been proposed for the purpose of collaborative filtering (see Hofmann and Puzicha, 1999; Hofmann, 2004; Si and Jin, 2003; Jin *et al.*, 2003; Magidson and Vermunt, 2002; Langseth and Nielsen, 2011; Jin *et al.*, 2006; Sitkrongwong *et al.*, 2013; Suzuki *et al.*, 2014), there is no model to directly represent the purchasing and browsing actions in EC sites for marketing purposes. In order to increase the estimation accuracy of the FMM, the modified learning algorithm using both purchasing and browsing history data was proposed and its effectiveness was clarified by Oi *et al.* (2015). In this case, the browsing history data is used only for learning the parameters of FMM. For the purpose of clustering by a latent class model, both browsing and purchasing activities should be directly represented in the model.

2.4 Proposed Latent Class Model to Represent Both Purchasing and Browsing Activities

The proposed model is a way to make use of the browsing history data in addition to the purchasing history data. There are two main reasons to use the data of both purchasing and browsing activities:

(1) The browsing activity of each customer is important to clarify the characteristics of each customer. Because customers browse product items depending on their own preferences, the browsing history data is valuable for making customer clusters reflecting user preferences. Moreover, more detailed analysis of customer characteristics can be realized by the combination of characteristics of both browsing and purchasing activities.
(2) Compared with the purchasing history data, the size of the browsing activity data is much larger. From the viewpoint of the accuracy caused by the sample size, it is useful to make use of both data of purchasing and browsing activities.

In order to treat both purchasing and browsing activities, we define the two-dimensional vector $\boldsymbol{w} = (w_1, w_2)$. Here, $w_1 \in W_1 = \{0, 1, 2, \cdots, R_1\}$ is the number of purchasing actions, and $w_2 \in W_2 = \{0, 1, 2, \cdots, R_2\}$ is that of browsing actions.

On the basis of the original idea of the aspect model, the following model is a simple expansion:

$$P(x, y, \boldsymbol{w}, z) = P(z)P(x \mid z)P(y \mid z)P(\boldsymbol{w} \mid z). \tag{4}$$

However, the interpretation of the latent class is difficult for the purpose of marketing analysis, though both latent clusters should be analyzed for marketing purposes. Therefore, the latent class of item x and that of user y are introduced independently in the same way as with FMM. Letting the latent class of item x be z_x and that of user y be z_y, the following latent class model is proposed:

$$P(x, y, \boldsymbol{w}) = \sum_{z_x \in \mathbf{V}_x} \sum_{z_y \in \mathbf{V}_y} P(z_x)P(z_y)P(y \mid z_y)P(x \mid z_x)P(w \mid z_x, z_y), \tag{5}$$

where $\mathbf{V}_x = \{v_1^x, v_2^x, \cdots, v_{z_x}^x\}$ and $\mathbf{V}_y = \{v_1^y, v_2^y, \cdots, v_{z_y}^y\}$ are the sets of the latent classes of items and users, respectively. The probability model of the complete data including the latent classes is given by

$$P(x, y, w, z_x, z_y) = P(z_x)P(z_y)P(x \mid z_x)P(y \mid z_y)P(w \mid z_x, z_y), \tag{6}$$

where $x \in \mathbf{A}$, $y \in \mathbf{U}$, $\boldsymbol{w} \in W_1 \times W_2$, $z_x \in \mathbf{V}_x$, and $z_y \in \mathbf{V}_y$. The components of Eq. (11) are defined as follows:

(1) $P(z_x)$ is the multinomial distribution on $\mathbf{V}_x$.
(2) $P(z_y)$ is the multinomial distribution on $\mathbf{V}_y$.
(3) $P(x \mid z_x)$ is the multinomial distribution on $\mathbf{A}$.
(4) $P(y \mid z_y)$ is the multinomial distribution on $\mathbf{U}$.
(5) $P(\boldsymbol{w} \mid z_x, z_y)$ is the multinomial distribution on $W = W_1 \times W_2$.

That is, this model assumes the two latent classes for users and product items, and $\boldsymbol{w}$ follows the conditional multinomial distribution $P(\boldsymbol{w} \mid z_x, z_y)$.

It is not a novelty point to assume two different latent classes for items and users. The FMM model proposed by Si and Jin (2003) assumes the two latent class variables of items and users independently. This assumption is reasonable in the field of marketing science. However, the FMM is a model proposed for the prediction of users' rating values. Therefore, the proposed model is different from the probability model of the FMM.

The graphical models of FMM and the proposed model are shown in Figure 2 and Figure 3. The structure of the graphical model of our proposal is similar to that of FMM because both latent classes of users and items are supposed.

Because the probability model of FMM is given by

$$P(x, y, r, z_x, z_y) = P(z_x)P(z_y)P(x \mid z_x)P(y \mid z_y)P(x \mid z_x, z_y), \tag{7}$$

where r is a rating value of a user to an item, the equations of the probability model are also similar. However, the meanings of $P(x \mid z_x)$ and $P(y \mid z_y)$ are different from each other. For FMM, the meanings of $P(x \mid z_x)$ and $P(y \mid z_y)$ are interpreted as follows:

(1) $P(x \mid z_x)$ is the probability that the item x belonging to the latent class z_x is given a rating value (by some user).
(2) $P(y \mid z_y)$ is the probability that the user y belonging to the latent class z_y gives a rating (to some item).

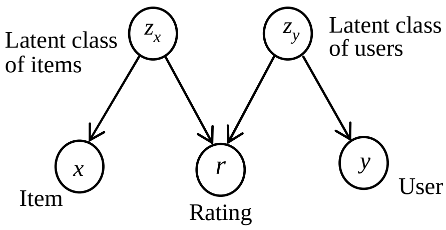

Figure 2. Graphical representation of FMM (Si and Jin, 2003).

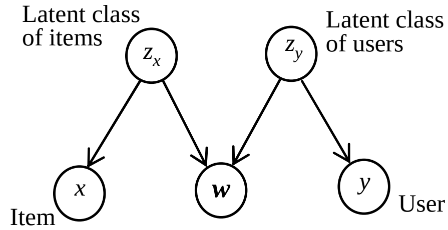

Figure 3. Graphical representation of the proposed model.

On the other hand, the meanings of $P(x \mid z_x)$ and $P(y \mid z_y)$ of the proposed model are interpreted as follows:
(1) $P(x \mid z_x)$ is the probability that the item x belonging to the latent class z_x is purchased or browsed (by some user).
(2) $P(y \mid z_y)$ is the probability that the user y belonging to the latent class z_y purchases or browses (some item).

On the condition that a user purchases or browses an item, the probability that the user y purchases the item x is given by

$$P(x, y, w_1 \neq 0) = \sum_{z_x} \sum_{z_y} \sum_{\boldsymbol{w} \mid w_1 \neq 0} p(x, y, \boldsymbol{w}, z_x, zy),$$

and the probability that the user y browses the item x is given by

$$P(x, y, w_2 \neq 0) = \sum_{z_x} \sum_{z_y} \sum_{\boldsymbol{w} \mid w_2 \neq 0} p(x, y, \boldsymbol{w}, z_x, zy),$$

As shown in the preceding discussion, there are some differences in the interpretations of FMM and the proposed model. The FMM was introduced with the motivation to model the multivalued rating for each combination (x, y). The proposed model can represent the case where there are different meanings of co-occurrence of x and y. That is, there are two meanings, purchasing activity and browsing activity. Obviously, our proposed model can be generalized to represent more actions, such as "browsing," "searching," "purchasing," and "bookmarking." If the users' activities on an EC site are these four actions, we may construct a four-dimensional vector $\boldsymbol{w}$ whose elements are the numbers of these actions.

The probability of the i-th complete data $(x_i, y_i, \boldsymbol{w}_i, z_i^x, z_i^y)$ is given by

$$P(x_i, y_i, \boldsymbol{w}_i, z_i^x, z_i^y) = P(z_i^x)P(z_i^y)P(x_i \mid z_i^x)P(y_i \mid z_i^y)P(\boldsymbol{w}_i \mid z_i^x, z_i^y), \tag{8}$$

where $x_i \in \mathbf{A}$, $yi \in \mathbf{U}$, $w_i \in \mathbf{W}$, $z_i^x \in \mathbf{V}_x$, and $z_i^y \in \mathbf{V}_y$. Because the latent classes z_i^x and z_i^y cannot be observed, the parameters $P(z_x)$, $P(z_y)$, $P(x \mid z_x)$, $P(y \mid z_y)$, and $P(w \mid z_x, z_y)$

should be estimated by the EM algorithm. Here, the EM algorithm is formulated with the observed data $(x_i, y_i, \boldsymbol{w}_i)$.

[E-Step]

$$\hat{P}(z_x, z_y \mid x_i, y_i, \boldsymbol{w}_i) = \frac{P(x_i, y_i, w_i, z_x, z_y)}{\sum_{z_x'} \sum_{z_y'} P(x_i, y_i, w_i, z_x', z_y')}, \quad (9)$$

$$\hat{P}(z_x \mid x_i, y_i, \boldsymbol{w}_i) = \sum_{x_y} \hat{P}(z_x, z_y \mid x_i, y_i, \boldsymbol{w}_i), \quad (10)$$

$$\hat{P}(z_y \mid x_i, y_i, \boldsymbol{w}_i) == \sum_{x_y} \hat{P}(z_x, z_y \mid x_i, y_i, \boldsymbol{w}_i). \quad (11)$$

[M-Step]

$$P\left(z_x, z_y\right) = \frac{1}{n} \sum_{i=1}^{n} \hat{P}(z_x, z_y \mid x_i, y_i, \boldsymbol{w}_i), \quad (12)$$

$$P(z_x) = \sum_{z_y \in \mathbf{V}_y} P(z_x, z_y), \quad (13)$$

$$P(z_y) = \sum_{z_x \in \mathbf{V}_x} P(z_x, z_y), \quad (14)$$

$$P\left(x \middle| z_x\right) = \frac{1}{nP(z_x)} \sum_{i=1}^{n} \delta(x_i = x) \hat{P}(z_x \mid x_i, y_i, \boldsymbol{w}_i), \quad (15)$$

$$P\left(y \mid z_y\right) = \frac{1}{nP(z_y)} \sum_{i=1}^{n} \delta(y_i = y) \hat{P}(z_y \mid x_i, y_i, \boldsymbol{w}_i), \quad (16)$$

$$P\left(\boldsymbol{w} \mid z_x, z_y\right) = \frac{1}{nP(z_x, z_y)} \sum_{i=1}^{n} \delta(\boldsymbol{w}_i = \boldsymbol{w}) \hat{P}(z_x, z_y \mid x_i, y_i, \boldsymbol{w}_i), \quad (17)$$

For the derivation of the EM algorithm, see Appendix B.

3. CASE STUDY

3.1 Data and Analysis Policy

In this study, we demonstrate the analysis by applying the proposed model for the data given by the 2013 data analysis competition of the Joint Association Study Group of Management Science in Japan (2014). These data are composed of the purchasing and browsing histories on an apparel EC site. In this case, the number of users is 99,924, and that of product items is 1,150,443. The purchasing and browsing data from September 1, 2011, to March 31, 2013, are used for this analysis. Usually, the size of the purchasing history is relatively small for the purpose of estimation of users' preferences to the immense number of items, but the size of the browsing data is relatively large. In this case, the number of purchasing actions of all users in this period was 707,857, and the number of browsing actions was 37,278,907. In the analysis, the browsing history is also valuable log data expressing a tendency of users' actions and preferences. Because the browsing activity is conducted by users on the basis of their interests and preferences, and the size of the browsing history is much larger than that of the purchasing history, the browsing history data is more valuable for building the model. Therefore, we consider the clustering of items and users by using both purchasing and browsing history data. Such clustering becomes possible by applying the proposed latent class model in Section 2.4.

After the clustering, it is effective analysis to clarify the characteristics of the representative attributes of each given cluster. However, the numbers of attributes of users and items are large and it is difficult to find distinctive features by repeating stratifications by hand. If the relation between users' attributes and given clusters can be analyzed by a machine learning technique, such a method provides a reasonable and easy way.

On the basis of the preceding discussion, we analyze the data by the following two steps in order to analyze the relation between users' attributes and their tendency of purchasing and browsing.

(1) The items and users are clustered respectively through both purchasing and browsing history data by applying the proposed latent class model.
(2) The relation between the given clusters and the attributes of items and users are analyzed by applying the decision tree classifier.

3.2 Learning of Proposed Latent Class Model

The browsing and purchasing histories, and the attributes of product items and users, are saved in tables separately in the database. Therefore, we should create a data matrix from the raw data for learning the latent class model. The image of the data structure for training the model is shown in Table 1.

In the learning phase, the data whose frequencies of purchasing and browsing are both 0 are not used. The multinomial distribution is assumed in the proposed latent class model, so that the frequencies of purchasing and browsing are quantized. Regarding the frequency of purchasing data, all frequencies over 4 are transformed to 4. The frequencies of browsing are quantized to 14 levels. That is, $R_1 = 4$ and $R_2 = 14$.

In order to analyze the characteristics of good customers and good product items, the numbers of the latent class are set as 2, such that $z_x \in \{0, 1\}$ and $z_y \in \{0, 1\}$. We can make four clusters by the combination $\left(z_x, z_y\right)$: (0, 0), (0, 1), (1, 0), and (1, 1). After the learning phase, the characteristics of each cluster are clarified by the conditional probability $P(\boldsymbol{w} \mid z_x, z_y)$. Figure 4 shows an example of the cluster $\left(z_x = 0, z_y = 0\right)$.

Table 1. Data structure

	x	y	w	
No.	Item ID	User ID	Purchasing frequency	Browsing frequency
1	100,251	25	0	3
2	255,613	25	1	2
3	369,154	3,566	0	1
4	556,976	5,789	1	4
⋮	⋮	⋮	⋮	⋮

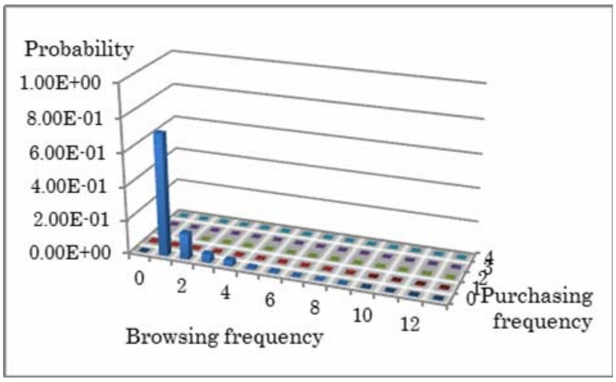

Figure 4. Probabilities $P(\boldsymbol{w} \mid z_x = 0,\ z_y = 0)$.

From Figure 4 showing the probabilities $P(\boldsymbol{w} | z_x = 0,$ $z_y = 0)$ of the cluster (0, 0), we find out that the users belonging to the latent class $z_y = 0$ tend to browse the item in the latent class $z_x = 0$ a few times, but they do not tend to purchase it.

As shown in Figure 4, the proposed model can treat both browsing and purchasing events. Therefore, the clusters can be characterized by the patterns of browsing and purchasing frequencies. This is a desirable characteristic of the proposed model in practice.

Figure 5 shows a comparison of the same figures between four clusters: $(z_x = 0, z_y = 0)$, $(z_x = 1, z_y = 0)$, $(z_x = 0, z_y = 1)$, and $(z_x = 1, z_y = 1)$. The probability of each cluster $P(z_x, z_y)$ and the purchase probability $P(w_1 \geq 1)$ are shown in Table 2 and Table 3.

Table 2. Belonging probabilities $P(z_x, z_y)$

	$z_x = 0$	$z_x = 1$
$z_y = 0$	65.87%	0.03%
$z_y = 1$	0.06%	34.04%

Table 3. Purchasing probabilities $P(w_1 \geq 1)$

	$z_x = 0$	$z_x = 1$
$z_y = 0$	1.96%	99.95%
$z_y = 1$	1.23%	2.29%

From these results, we learn the following facts.

a) When the user latent class z_y is fixed, the purchase probability of item latent class $z_x = 1$ is larger than that of $z_x = 0$.
b) When the item latent class z_x is fixed, the purchase probability of the latent item class $z_y = 0$ is larger than that of $z_y = 1$.

From these results, the clusters from the viewpoints of "good" or "others" are obtained. Here, "good" means that the purchase probability is relatively high. Conver-

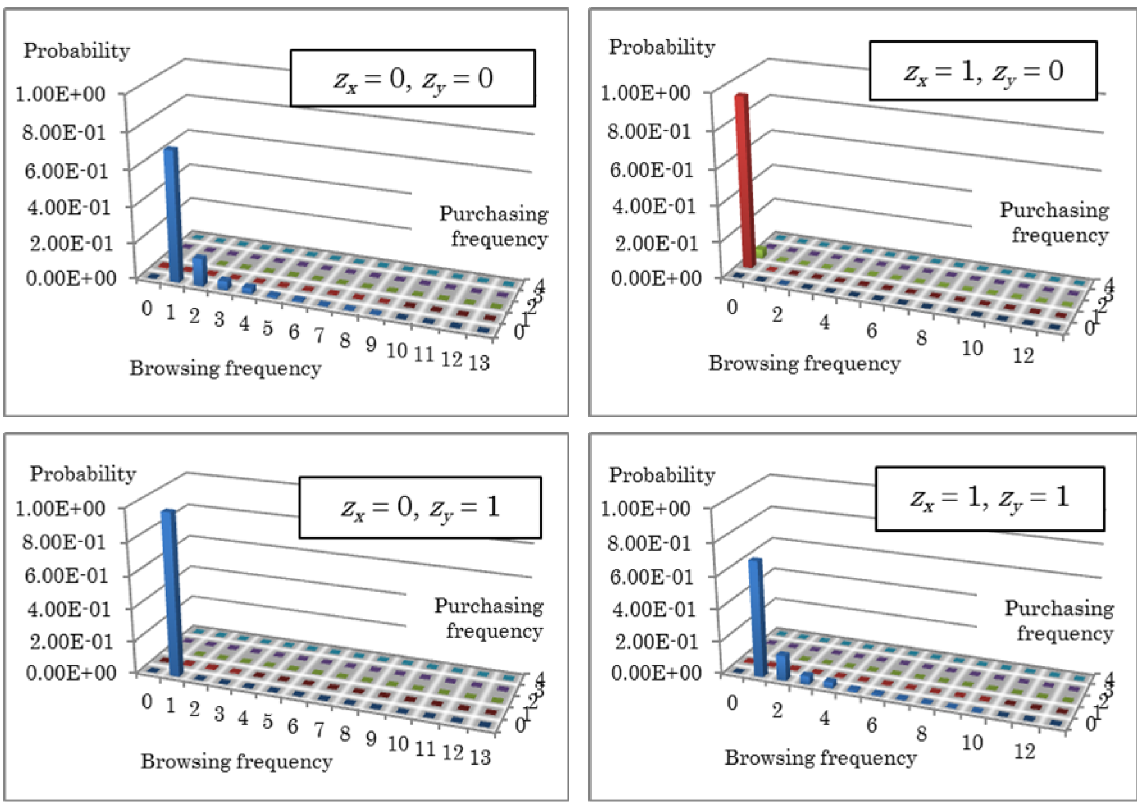

Figure 5. Result of clustering by applying the proposed latent class model.

Table 4. Cluster matrix

	"Other" item cluster $z_x = 0$	"Good" item cluster $z_x = 1$
"Good" user cluster $z_y = 0$	"Other" items "Good" customers	"Good" items "Good" customers
"Other" user cluster $z_y = 1$	"Other" items "Other" customers	"Good" items "Other customers

sely, "others" means that the purchase probability is relatively lower than that of "good." Therefore, the matrix of clusters by the "good" and the "other" customers, and the "good" and the "other" items, is shown in Table 4.

Through the proposed latent class model, the "good" and the "other" clusters can be extracted. The attributes of product items and users are not used for the clustering. If the factors to distinguish "good" or "others" are easily clarified, it is useful for marketing research.

3.3 Analysis by Decision Tree Model

In this section, the factors to distinguish "good" from "other" are analyzed. In this study, we formulated a classification problem of the latent class based on the attributes of items and users. If the classification rule is learned from the training data, then the relation between the latent class and the attributes is modeled. We apply the decision tree model with the CART learning algorithm (Breiman *et al*. 1984). The decision tree model has an advantage in that an analyst can easily interpret the model structure.

3.3.1 Analysis for Product Items

First, the results of the analysis for item clusters are shown. For the learning of the decision tree, the following attributes of product items are used. The number of broad categories is 26, that of subcategories is 216, that of shops is 535, that of brand is 5,456, and that of zone is 35. The code numbers are used for the attributes of items in Table 5. Using the item table, the category name can be known in practice. Examples of the broad item categories are "Bags," "TOPS," "Shoes," "Pants," and "Skirt." Examples of subcategories for "TOPS" are "Tank top," "Shirt Blouse," "T-Shirt," "Parka (Hooded Sweatshirt)," and "Cardigan." Table 5 shows an example of the item data structure.

Table 5. Data of item cluster analysis

z_x	x Item ID	Product Item Attributes: Broad category	Sub-category	Shop	Brand	Zone
1	63516	139	2065	26	1140	67
0	215435	111	2035	55	767	1
0	1015538	113	2246	55	767	1
1	1118094	101	2021	149	3671	6
⋮	⋮	⋮	⋮	⋮	⋮	⋮

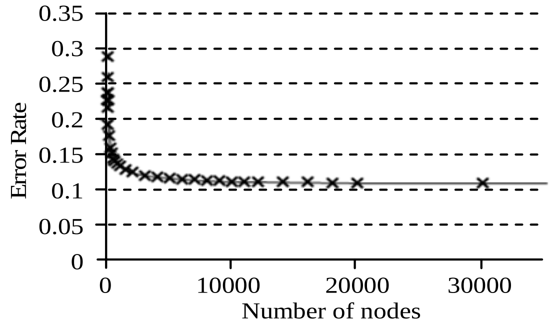

Figure 6. Error rate of item class for test data.

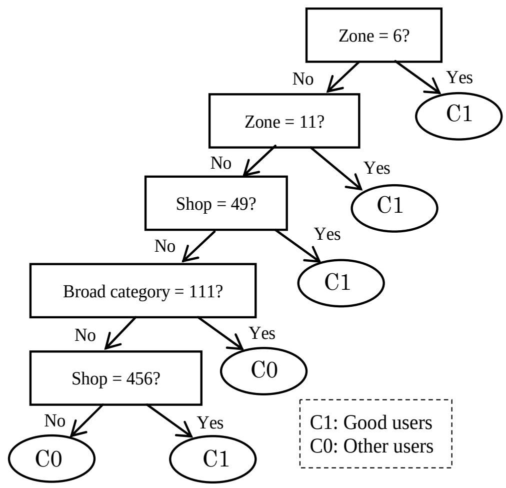

Figure 7. CART for item when the number of nodes is 11

In order to clarify the performance of CART, the classification tree is learned by the set of 800,000 training samples, and the error rate for the test data with 350,000 samples is evaluated. The error rate of CART for different numbers of nodes is shown in Figure 6.

From this result, the error rate converges to approximately 11% when increasing the number of nodes. The given classifier can classify the item to "good" or "others" with approximately an 11% error rate. When a new product item is added, we can predict the latent class of this item by using its attributes.

An example for the case where the number of nodes is 11 is shown in Figure 7.

The error rate for test data in this case was 22.8%. The code number "111" of the broad category is "One-piece dress." The customers who tend to purchase one-piece dresses may not become good customers.

By the analysis based on CART, we found the result that the "Zone" and "Shop" are selected on the early stage of branch making. This fact means that the "Zone" and "Shop" are more important than "Broad category" or "Subcategory." The items of popular shops are being purchased with high frequencies. There is a possibility that the users are narrowing down an item in their favorite shops.

Table 6. Data of user cluster analysis

z_y	y User ID	User Attributes		
		Sex	Birth Year	Habitat
0	25	1	1974	27
1	107	1	1977	28
1	3658	2	1983	12
0	16850	2	1978	14
⋮	⋮	⋮	⋮	⋮

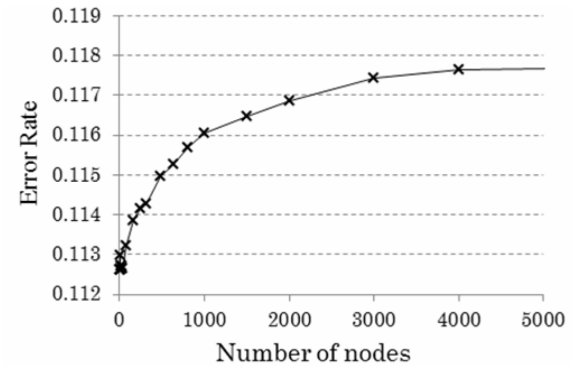

Figure 8. Error rate of user class for test data.

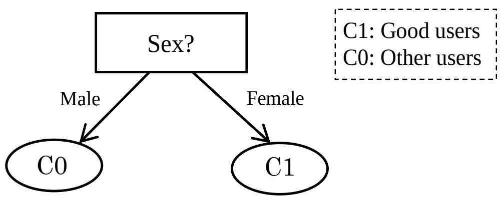

Figure 9. CART for user when the number of nodes is 3.

Table 7. The rate of sex in each latent user class

	"Good" user cluster $z_y = 0$	"Not good" user cluster $z_y = 1$	Sum
Male	17.6 %	82.4 %	100%
Female	93.2 %	6.8 %	100%

3.3.2 Analysis for Users

Next, the results of the analysis for user clusters are shown. For the learning of decision tree, the attributes of users in Table 6 are used.

In order to clarify the performance of CART, the classification tree is learned by the set of 70,000 training samples, and the error rate for the test data with 27,000 samples is evaluated. The error rate of CART for different numbers of nodes is shown in Figure 8.

From this result, the error rate is not improved by increasing the node number. The tree when the number of nodes is 3 is shown in Figure 9.

The error rate for test data in this case was 11.3%. That is, the user latent class can be explained by the attribute "Sex." The rate of male and female in each latent user class is shown in Table 7.

Females tend to become good customers, but most male users are not good customers. The reason for this result may be the target of the web design. Because this EC site is focused on fashions, the web design may be more attractive to women customers.

3.4 Discussion

From the analysis based on CART, we can discuss the necessary strategies of the EC site.

Regarding the product items, the "Zone" and "Shop" were more important than "Broad category" or "Subcategory" of items. There are several ways for a user to search an item in this EC site. The users can search an item through the links of "Broad category" and "Subcategory" of items. However, if a user clicks the link of "One-piece dress," then a large number of one-piece dresses will be listed. It may be impossible for users to find the best one from a list with a large number of items. It is necessary to build a strategy to raise the sales of the ordinary shops (i.e., other than the popular shops).

For the user latent cluster, females tend to become good customers, but most male users are not good customers. The construction and design of web pages that are attractive to male customers may be different from those to females. If the present web design is modified to be attractive to males, such a design may not be preferred by females. It may be better to build two different top pages, for males and females. The recommended product items and information can be customized separately for males and females. In order to increase the product items for males, a different strategy may be effective for the different target. This EC site already has good female customers who are purchasing many items. It is an effective way to promote gift items for a boyfriend or a husband and to expand the web page to have women purchase products for men as gifts. For example, the function to choose a lapping may be important for gift items.

4. CONCLUSION

This paper proposes a new latent model to analyze both purchasing and browsing histories efficiently. By using the proposed latent class model, customer clustering was conducted and the clusters of good customers and good product items were extracted. Moreover, the relations between these clusters and attributes of items and users were analyzed by the decision tree model. An example of data analysis on an EC site is demonstrated. Through the proposed model and approach of our analysis, new findings are extracted from the data of purchasing and browsing histories.

Although the latent class model was constructed for the purpose of clustering and analyzing the characteristics of good items and good customers, the other types of latent class models expressing both purchasing and browsing actions may be useful for other purposes. For example, a model for recommendation is proposed by

Fujiwara *et al.* (2014). The relation with such other extensions should be discussed as a future work. Moreover, the idea of Bayesian statistics is sometimes effective in the field of machine learning and data analysis, so that a combination with the Bayesian model can realize a more exquisite model. The Bayesian estimation of the parameters can be introduced for our proposed model. These improvements are also planned for future works.

ACKNOWLEDGMENTS

The authors would like to express our thanks to the Joint Association Study Group of Management Science in Japan for giving us the opportunity to analyze the real data of an EC site. The authors also thank the area editor and reviewers for their constructive feedback. A part of this study was supported by JSPS KAKENHI Grant Numbers 26282090 and 26560167.

REFERENCES

Bhatnagar, A. and Ghose, S. (2004), A Latent Class Segmentation Analysis of E-Shoppers, *Journal of Business Research*, **57**(7), 758-767.

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer.

Breiman, L., Friedman, J. H. Olshen, R. A., and Sone, C. J. (1984), *Classification and Regression Trees*, Wadsworth.

Curry, J. and Curry, A. (2000), *The Customer Marketing Method*, Free Press.

Dempster, A., Laird, N., and Rubin, D. (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. Royal Statist. Soc*., *Series B*, **39**(1), 1-38.

Fujiwara, N., Mikawa, K., and Goto, M. (2014), A New Estimation Method of Latent Class Model with High Accuracy by Using Both Browsing and Purchase Histories, *The 15th Asia Pacific Industrial Engineering and Management Systems Conference*, APIEMS.

Goto, M., Minetoma, K., Mikawa, K., Kobayashi, M., and Hirasawa, S. (2014), A Modified Aspect Model for Simulation Analysis, *IEEE International Conference on Systems, Man, and Cybernetics*.

Green, P. E., Carmone, F. J., and Wachspress, D. P. (1976), Consumer Segmentation Via Latent Class Analysis, *Journal of Consumer Research*, **3**(3), 170-174.

Hofmann, T. (1999), Probabilistic Latent Semantic Indexing, *The 22nd Annual International SIGIR Conference on Research and Development in Information Retrieval*.

Hofmann, T. and Puzicha, J. (1999), Latent Class Models for Collaborative Filtering, *Proc. 16th International Joint Conference on Artificial Intelligence*, 688-693.

Hofmann, T. (2004), Gaussian Latent Semantic Models for Collaborative Filtering, *Proc. the 26th Annual International ACM SIGIR Conference*, **22**(1), 259-266.

Hofmann, T. (2004), Latent Semantic Models for Collaborative Filtering, *ACM Trans. Information Systems*, **22**(1), 89-115.

Hughes, A. M. (2006), *Strategic Database Marketing*, McGraw-Hill.

Jin, R. Si, L., and Zhai, C. X. (2003), Preference-based Graphic Models for Collaborative Filtering, *UAI Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*, 329-336.

Jin, R., Si, L., and Zhai, C. (2006), A Study of Mixture Models for Collaborative Filtering," *Journal of Information Retrieval*, **9**(3), 357-382, DOI 10.1007/s10791-006-4651-1.

Joint Association Study Group of Management Science in Japan (2014), http://jasmac-j.jimdo.com/.

Langseth, H. and Nielsen, T. D. (2011), A Latent Model for Collaborative Filtering, *preprint submitted to Elsevier*.

Magidson, J. and Vermunt, J. K. (2002), Latent Class Models for Clustering: A Comparison with K-means, *Canadian Journal of Marketing Research*, **20** (1), 37-44.

McLachlan, G. and Krishnan, T. (2007), *The EM Algorithm and Extensions*, Wiley-Interscience.

Oi, T., Mikawa, K., and Goto, M. (2013), A Study of Recommender Systems Based on the Latent Class Model Estimated by Combining Both Evaluation and Purchase Histories, *The 14th Asia Pacific Industrial Engineering and Management Systems Conference*, APIEMS.

Si, L. and Jin, R. (2003), Flexible Mixture Model for Collaborative Filtering, *Proc. 20th International Conference on Machine Learning*, **2**, 704-711.

Sitkrongwong, P., Maneeroj, S., and Takasu, A. (2013), Latent Probabilistic Model for Context-Aware Recommendations, *IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT)*, DOI 10.1109/WI-IAT.2013.14.

Suzuki, T., Kumoi, G., Mikawa, K., and Goto, M. (2014), A Design of Recommendation Based on Flexible Mixture Model Considering Purchasing Interest and Post-Purchase Satisfaction, *Journal of Japan Industrial Management Association*, **64**(4E), 570-578.

Swait, J. and Adnmowicz, W. (2001), Consumer Choice: A Latent Class Model of Decision Strategy Switching, *Journal of Consumer Research*, **28**(1), 135-148.

Train, K. E. (2009), *Discrete Choice Methods with Simulation-Second edition*, Cambridge University Press.

APPENDIX A

Here, the EM algorithm for the aspect model is shown. Letting $n(x, y)$ be the number of occurrences of (x, y) in n training data, $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, the log likelihood of Eq. (3) can be rewritten as

$$L = \sum_{x \in \mathbf{A}} \sum_{y \in \mathbf{U}} \log n(x, y) \sum_{z \in \mathbf{V}} P(x|z) P(y|z) P(z).$$

The estimator to maximize the log likelihood of Eq. (3) is given by the following EM algorithm.

[E-Step]

$$P(z \mid x, y) = \frac{P(z)P(x \mid z)P(y \mid z)}{\sum_{z'} P(z')P(x \mid z')P(y \mid z')},$$

[M-Step]

$$P(x|z) = \frac{\sum_{y'} n(x, y')P(z \mid x, y')}{\sum_{x'}\sum_{y'} n(x', y')P(z \mid x', y')},$$

$$P(y|z) = \frac{\sum_{x'} n(x, y')P(z \mid x, y')}{\sum_{x'}\sum_{y'} n(x', y')P(z \mid x', y')},$$

$$P(z) = \frac{\sum_{x'}\sum_{y'} n(x', y')P(z \mid x', y')}{\sum_{x'}\sum_{y'} n(x', y')}.$$

APPENDIX B

Here, the derivation of the EM algorithm for the proposed latent class model is described. At first, the following notations are introduced.

$$\boldsymbol{X} = (x_1, x_2, \cdots, x_n)$$
$$\boldsymbol{Y} = (y_1, y_2, \cdots, y_n)$$
$$\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_n)$$
$$\boldsymbol{Z_x} = \left(z_1^x, z_2^x, \cdots, z_n^x\right)$$
$$\boldsymbol{Z_y} = \left(z_1^y, z_2^y, \cdots, z_n^y\right)$$

The probability of the complete data set $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{Z_x}, \boldsymbol{Z_y})$ is given by

$$P(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{Z_x}, \boldsymbol{Z_y}) = \prod_{i=1}^{n} P\left(x_i, y_i, \boldsymbol{w}_i, z_i^x, z_i^y\right)$$
$$= \prod_{i=1}^{n} P\left(z_i^x\right) P\left(z_i^y\right) P\left(x_i \middle| z_i^x\right) P\left(y_i \middle| z_i^y\right) P\left(\boldsymbol{w}_i \middle| z_i^x, z_i^y\right),$$

where $x_i \in \mathbf{A}, y_i \in \mathbf{U}, \boldsymbol{w}_i \in \mathbf{W}, z_i^x \in \mathbf{V}_x, z_i^y \in \mathbf{V}_y$.

E-Step:
The data of latent classes $\boldsymbol{Z_x}, \boldsymbol{Z_y}$ cannot be observed; $\tilde{P}(\boldsymbol{Z_x}, \boldsymbol{Z_y} | \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W})$ is prepared to calculate the expectation of the log likelihood.

$$\tilde{P}(\boldsymbol{Z_x}, \boldsymbol{Z_y} | \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}) = \frac{P(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{Z_x}, \boldsymbol{Z_y})}{\sum_{Z_x}\sum_{Z_y} P(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{Z_x}, \boldsymbol{Z_y})}$$
$$= \frac{\prod_{i=1}^{n} P\left(x_i, y_i, \boldsymbol{w}_i, z_i^x, z_i^y\right)}{\prod_{i=1}^{n}\sum_{Z_x}\sum_{Z_y} P\left(x_i, y_i, \boldsymbol{w}_i, z_i^x, z_i^y\right)}$$
$$= \prod_{i=1}^{n} \frac{P\left(x_i, y_i, \boldsymbol{w}_i, z_i^x, z_i^y\right)}{\sum_{Z_x}\sum_{Z_y} P\left(x_i, y_i, \boldsymbol{w}_i, z_i^x, z_i^y\right)}$$
$$= \prod_{i=1}^{n} \tilde{P}\left(z_i^x, z_i^y \middle| x_i, y_i, \boldsymbol{w}_i\right)$$

Using this probability, the Q-function can be formulated as follows:

$$Q = \sum_{Z_x}\sum_{Z_y} \tilde{P}\left(\boldsymbol{Z_x}, \boldsymbol{Z_y} \middle| \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}\right)$$
$$\times \log\left\{\prod_{i=1}^{n} P\left(z_i^x\right) P\left(z_i^y\right) P\left(x_i \middle| z_i^x\right) P\left(y_i \middle| z_i^y\right) P\left(\boldsymbol{w}_i \middle| z_i^x, z_i^y\right)\right\}$$
$$= \sum_{i=1}^{n}\left\{\sum_{z_x} \tilde{P}\left(z_x \middle| x_i, y_i, \boldsymbol{w}_i\right) \log P\left(z_x\right)\right.$$
$$+ \sum_{z_y} \tilde{P}\left(z_y \middle| x_i, y_i, \boldsymbol{w}_i\right) \log P\left(z_x\right)$$
$$+ \sum_{z_x} \tilde{P}\left(z_x \middle| x_i, y_i, \boldsymbol{w}_i\right) \log P\left(x_i \middle| z_x\right)$$
$$+ \sum_{z_y} \tilde{P}\left(z_y \middle| x_i, y_i, \boldsymbol{w}_i\right) \log P\left(y_i \middle| z_y\right)$$
$$\left. + \sum_{z_x}\sum_{z_y} \tilde{P}\left(z_x, z_y \middle| x_i, y_i, \boldsymbol{w}_i\right) \log P\left(\boldsymbol{w}_i \middle| z_i^x, z_i^y\right)\right\}$$

M-Step:
On the following constraints, the Q function is maximized with respect to the parameters.

$$\sum_{z_x \in \mathbf{V}_x} P\left(z_x\right) = 1,$$
$$\sum_{z_y \in \mathbf{V}_y} P\left(z_y\right) = 1,$$
$$\sum_{x \in \mathbf{A}} P\left(x \middle| z_x\right) = 1,$$
$$\sum_{y \in \mathbf{U}} P\left(y \middle| z_y\right) = 1,$$
$$\sum_{z_x \in \mathbf{V}_x}\sum_{z_y \in \mathbf{V}_y} P\left(\boldsymbol{w} \middle| z_x, z_y\right) = 1.$$

The method of the Lagrange multiplier is applied to solve the optimization problem.

$$\mathrm{L} = \mathrm{Q} - \alpha_x\left(\sum_{z_x \in \mathbf{V}_x} P(z_x) - 1\right) - \alpha_y\left(\sum_{z_y \in \mathbf{V}_y} P(z_y) - 1\right)$$
$$- \sum_{z_x} \beta_{z_x}\left(\sum_{x \in \mathbf{A}} P(x|z_x) - 1\right) - \sum_{z_y} \gamma_{z_y}\left(\sum_{y \in \mathbf{U}} P(y|z_y) - 1\right)$$
$$- \sum_{z_x}\sum_{z_y} \lambda_{z_y z_y}\left(\sum_{w \in \mathbf{W}} P(\boldsymbol{w}|z_x, z_y) - 1\right)$$

(1) Optimization for $P(z_x)$ and $P(z_y)$:

From

$$\frac{\partial L}{\partial P(zx)} = \sum_{i=1}^{n} \tilde{P}\left(z_x \middle| x_i, y_i, \boldsymbol{w}_i\right) \frac{1}{P(z_x)} - \alpha_x = 0,$$

we have

$$P(z_x) = \frac{1}{\alpha_x} \sum_{i=1}^{n} \tilde{P}\left(z_x \middle| x_i, y_i, \boldsymbol{w}_i\right).$$

Because $\sum_{z_x \in \mathbf{V}_x} P(z_x) = 1,$ the equation

$$\sum_{z_x \in \mathbf{V}_x} P(z_x) = \sum_{z_x \in \mathbf{V}_x} \frac{1}{\alpha_x} \sum_{i=1}^{n} \tilde{P}\left(z_x \middle| x_i, y_i, \boldsymbol{w}_i\right) = 1,$$

is satisfied. Therefore, we have

$$\alpha_x = \sum_{z_x \in \mathbf{V}_x} \sum_{i=1}^{n} \tilde{P}\left(z_x \middle| x_i, y_i, \boldsymbol{w}_i\right) = n,$$

so that

$$P\left(z_x\right) = \frac{1}{n} \sum_{i=1}^{n} \tilde{P}\left(z_x \middle| x_i, y_i, \boldsymbol{w}_i\right),$$

is given. Similarly, we have

$$P(z_y) = \frac{1}{n} \sum_{i=1}^{n} \tilde{P}\left(z_y \middle| x_i, y_i, \boldsymbol{w}_i\right).$$

(2) Optimization for $P(x|z_x)$ and $P(y \mid z_y)$:

From

$$\frac{\partial L}{\partial P(x \mid z_x)} = \sum_{i=1}^{n} \delta(x_i = x) \tilde{P}(z_x \mid x_i, y_i, \boldsymbol{w}_i) \frac{1}{P(x \mid z_x)} - \beta_{z_x} = 0,$$

we have

$$P(x \mid z_x) = \frac{1}{\beta_{z_x}} \sum_{i=1}^{n} \delta(x_i = x) \tilde{P}(z_x \mid x_i, y_i, \boldsymbol{w}_i) = 1.$$

Because $\sum_{x \in \mathbf{A}} P(x \mid z_x) = 1,$ the equation

$$\frac{1}{\beta_{z_x}} \sum_{x \in A} \sum_{i=1}^{n} \delta(x_i = x) \tilde{P}(z_x \mid x_i, y_i, \boldsymbol{w}_i) = 1,$$

is satisfied. Therefore, we have

$$\beta_{z_x} = \sum_{x \in \mathbf{A}} \sum_{i=1}^{n} \delta(x_i = x) \tilde{P}(z_x \mid x_i, y_i, \boldsymbol{w}_i) = nP(z_x),$$

so that

$$P\left(x \middle| z_x\right) = \frac{1}{nP(z_x)} \sum_{i=1}^{n} \delta(x_i = x) \tilde{P}\left(z_x \middle| x_i, y_i, \boldsymbol{w}_i\right),$$

is acquired. Similarly, we have

$$P\left(y \middle| z_y\right) = \frac{1}{nP(z_y)} \sum_{i=1}^{n} \delta(y_i = y) \tilde{P}\left(z_y \middle| x_i, y_i, \boldsymbol{w}_i\right).$$

(3) Optimization for $P\left(w \middle| z_x, z_y\right)$:

From

$$\frac{\partial L}{\partial P\left(\boldsymbol{w} \middle| z_x, z_y)\right)} = \sum_{i=1}^{n} \delta(\boldsymbol{w}_i = \boldsymbol{w}) P\left(z_x, z_y \middle| x_i, y_i, \boldsymbol{w}_i\right) \frac{1}{P\left(\boldsymbol{w} \middle| z_x, z_y\right)} - \lambda_{z_y z_y} = 0,$$

we have

$$P\left(\boldsymbol{w} \middle| z_x, z_y\right) = \frac{1}{\lambda_{z_y z_y}} \sum_{i=1}^{n} \delta(\boldsymbol{w}_i = \boldsymbol{w}) \tilde{P}\left(z_x, z_y \middle| x_i, y_i, \boldsymbol{w}_i\right) = 1.$$

Because $\sum_{z_x \in \mathbf{V}_x} \sum_{z_y \in \mathbf{V}_y} P\left(w \middle| z_x, z_y\right) = 1,$ the equation

$$\sum_{z_x \in \mathbf{V}_x} \sum_{z_y \in \mathbf{V}_y} \frac{1}{\lambda_{z_y z_y}} \sum_{i=1}^{n} \delta(\boldsymbol{w}_i = \boldsymbol{w}) \tilde{P}\left(z_x, z_y \middle| x_i, y_i, \boldsymbol{w}_i\right) = 1,$$

is satisfied. Therefore, we have

$$\lambda_{z_y z_y} = nP(z_x, z_y),$$

so that

$$P\left(\boldsymbol{w} \middle| z_x, z_y\right) = \frac{1}{nP\left(z_x, z_y\right)} \sum_{i=1}^{n} \delta(\boldsymbol{w}_i = \boldsymbol{w}) P\left(z_x, z_y \middle| x_i, y_i, \boldsymbol{w}_i\right),$$

is acquired.