An Analytical Model of Relation between Browsing and Entry Activities on an Internet Portal Site for Job-hunting

Yuuki Sugiyama*1, Takumi Arai1, Tianxiang Yang1, Tairiku Ogihara2, Masayuki Goto1

1. Waseda University : 51-15-03 3-4-1 Okubo, Shinjuku, Tokyo, 169-8555 Japan
2. Recruit Career Co., Ltd. : 1-9-2 Marunouchi, Chiyoda, Tokyo, 100-6640 Japan

* miyus0919yuuki@ruri.waseda.jp

Abstract:
In recent years, many companies conducting recruitment activities and many students looking for a job use internet portal sites for job-hunting in Japan. Companies can post their basic information on individual company pages and recruit applications from students. On the other hand, student users can gauge corporate attractiveness by browsing individual company pages on a job-hunting site and can make entries to companies of interest. Therefore, a large amount of their behavior history data is accumulated on the site. There are several studies on prediction of users’ entries to companies and analysis of preference using user attribute information and entry history data. However, in the conventional researches, browsing activities on individual company pages existing in the background of the user’s entry were not considered, so the relation between browsing and making entries has not been studied. This research proposes a latent class model for analyzing the relation between browsing company pages and making entries to companies. The proposed model enables clarification of target users and consideration of effective promotion activities. Through a demonstrative analysis using actual data on a major job-hunting website in Japan, we show the effectiveness of the proposed model.

Keywords
Big Data Analysis, Analytical Model, Latent Class Model, Business Analytics, Job-Hunting

1. Introduction
In recent years, internet portal sites for job-hunting are used by many companies conducting recruitment activities and many students looking for jobs in Japan. In Japan, the action of applying for a position at a company is called an “entry.” Companies can post their own basic and recruitment information on individual company pages, and recruit entries from students on the job-hunting website. On the other hand, student users can determine corporate attractiveness by browsing individual company pages on the job-hunting site and make entries to interesting companies. Therefore, a large amount of their behavior history data is accumulated on the site. By effectively utilizing the data of users’ behavior history accumulated on the portal site, it is possible to analyze the preferences of student users and propose various actions to the companies.

In previous researches using the data on a job-hunting website, statistical analysis models focusing on the entry tendencies and the users’ company preferences have been proposed. For example, Nagamori et al. (2015) proposed a model for clustering student users by utilizing entry history data, and Yamagami et al. (2015) proposed a prediction model of the finish date of job-hunting based on entry history data and users’ attribute information. These previous works utilized only the entry history data in order to learn a model. However, there are two kinds of user behaviors for making entries to a company: (1) after browsing an individual company page on the portal site, an entry to the company is made; (2) an entry is made only by focusing on the company name or industry type without browsing an individual company page. Previous works have not considered the differences between these entry behaviors. In order to evaluate the effectiveness of individual company pages or analyze the students’ behaviors on the portal site, it is better to construct a model which can distinguish these behavior differences.

This research proposes a new model for analyzing the relation between browsing individual company pages and making entries to companies. Now, we introduce a latent class model (Magidson and Vermunt, 2002; Greene and
Hensher, 2003; Collins and Lanza, 2013; Goto et al., 2014) in order to represent the relation between browsing and entry activities and analyze different trends for each user or company. Through a demonstrative analysis by using actual data on a major job-hunting website in Japan, we show the effectiveness of the proposed model.

2. Preparations

2.1 Job-hunting of Students in Japan

In this section, we introduce the outline of job-hunting activities of Japanese students and recruiting activities of companies. Table 1 shows the basic schedule of students’ job-hunting in Japan.

<table>
<thead>
<tr>
<th></th>
<th>Job-hunting of students who graduated in 2016</th>
<th>Job-hunting of students who graduated in 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start of publicity for recruitment</td>
<td>2015/03/01</td>
<td>2016/03/01</td>
</tr>
<tr>
<td>Start of selection activities such as interview</td>
<td>2015/08/01</td>
<td>2016/06/01</td>
</tr>
<tr>
<td>Presentation of employment offer</td>
<td>2015/10/01</td>
<td>2016/10/01</td>
</tr>
<tr>
<td>Graduation from various schools</td>
<td>2016/03/31</td>
<td>2017/03/31</td>
</tr>
<tr>
<td>Joining a company</td>
<td>2016/04/01</td>
<td>2017/04/01</td>
</tr>
</tbody>
</table>

The style of job-hunting of Japanese students is peculiar compared to the rest of the world. In Japan, most students, such as university students and students of specialized training colleges, look for a job while they are still students. The basic schedule of student job-hunting activities is decided by the Japan Business Federation. Therefore, according to this schedule, many companies publish recruitment information and start hiring activities at the same time every year. Many students use internet portal sites for job-hunting to get to know companies and have the first contact with companies. Especially for university students, job-hunting activities using the website often start from the summer of the third year to the next spring. The flow of job-hunting using the portal sites in Japan is shown in Figure 1.

For some companies, scheduling of briefings and interviews can be done through the portal sites. Moreover, these sites also provide information useful for job-hunting such as summaries of the company/industry, how to write application forms, etc. Therefore, the internet portal sites for job-hunting widely support student users’ in their job-hunting activities.

2.2 Latent Class Model

The preferences of students are very diverse, and groups with hopes of various job placements are mixed. There are also different groups of companies with different characteristics. Latent class models are known as statistical models that are useful for modeling objects with different groups of different features.

Latent class models are models assuming the existence of unobserved latent variables behind the observed data. In real cases, we sometimes encounter a situation in which heterogeneous data groups are mixed, but it is not possible to observe which group an observed datum belongs to. For example, the customers’ purchase records are a set of various customers’ data with different preferences. By applying the latent class model, it becomes possible...
to analyze the latent semantic relation of the data. Latent class models are widely applied from document classification (Rubin et al., 2012) to customer purchasing behavior analysis (Suzuki et al., 2014). Hofmann (1994a; 1994b) proposed an “aspect model,” which is one of the latent class models, assuming that there is a latent topic behind the document and words. That is, this model assumes that words in the document occur according to an unobserved latent topic. With the aspect model, it is possible to make probabilistically clustered documents and words at the same time.

Here, let us define the notations of sets as follows: The set of I documents is $T = \{ t_i; 1 \leq i \leq I \}$, the set of J words is $W = \{ w_j; 1 \leq j \leq J \}$, the set of K latent classes of the companies is $Z = \{ z_k; 1 \leq k \leq K \}$. Then, the probability model is formulated as

$$P(t_i, w_j) = \sum_{k=1}^{K} P(z_k)P(t_i|z_k)P(w_j|z_k)$$

(1)

where $P(t_i, w_j)$ means the joint probability of $t_i$ and $w_j$, $P(z_k)$ is the probability of the latent class $z_k$, and $P(t_i|z_k)$ and $P(w_j|z_k)$ are the conditional probabilities of $t_i$ and $w_j$ conditioned by the latent class $z_k$, respectively.

The graphical model of the aspect model is described in Figure 2.

![Graphical model of the aspect model](image)

Figure 2: Graphical model of the aspect model.

Here, the l-th learning data is expressed as $(x_l, y_l)$, where $x_l \in T$ and $y_l \in W$, and the size of all learning data is L. The parameters of equation (1) in the proposed model are estimated by the EM algorithm (Dempster et al., 1977; McLachlan and Krishnan, 2007). By using the EM algorithm, the parameters $P(z_k)$, $P(t_i|z_k)$, and $P(w_j|z_k)$, locally maximizing the log likelihood can be estimated. In this algorithm, the E-step and M-step of equations (3)—(8) are repeated until the log likelihood function $LL$ of equation (2) converges, and the parameters are updated.

$$LL = \sum_{l=1}^{L} \log P(x_l, y_l)$$

(2)

**E-step**

$$P(z_k|x_l, y_l) = \frac{P(z_k)P(x_l|z_k)P(y_l|z_k)}{\sum_{k=1}^{K} P(z_k)P(x_l|z_k)P(y_l|z_k)}$$

(3)

**M-step**

$$P(z_k) = \frac{\sum_{l=1}^{L} P(z_k|x_l, y_l)}{\sum_{k=1}^{K} \sum_{l=1}^{L} P(z_k|x_l, y_l)}$$

(4)
Analytical Model for Job-hunting Website, Sugiyama + et al.

\[
P(t_i | z_i) = \frac{\sum_{l=1}^{L} P(z_l | x_i, y_i) \delta(x_i = t_i)}{\sum_{l=1}^{L} \sum_{j=1}^{J} P(z_l | x_i, y_i) \delta(x_i = t_i)} \quad (5)
\]

\[
P(w_j | z_k) = \frac{\sum_{l=1}^{L} P(z_k | x_i, y_i) \delta(y_i = w_j)}{\sum_{l=1}^{L} \sum_{j=1}^{J} P(z_k | x_i, y_i) \delta(y_i = w_j)} \quad (6)
\]

Here, \( \delta(k = k') \) is an indicator function that takes 1 when \( k = k' \) and 0 when \( k \neq k' \).

The aspect model was initially proposed as a document model and applied to collaborative filtering. Several works related with the probability model with latent topics have been developed, for example, Si and Jin (2003) proposed the flexible mixture model (FMM) and Blei and Jordan (2003) proposed the Latent Dirichlet Allocation (LDA).

Goto et al. (2015) proposed a latent class model representing customers’ browsing and purchasing behaviors toward items on an EC site. In this model, the behavior of customers’ purchasing and viewing of items is represented by a two-dimensional vector \( w = (w_1, w_2) \), which is composed of the number of purchases \( w_1 \) and the number of views \( w_2 \). For the purpose of detailed marketing analysis, the latent class of items and that of users are introduced independently. Depending on the combination of two latent classes, customers’ browsing and purchasing behavior is expressed. By this model, it is possible to analyze customers’ tendencies, such as customers who tend to purchase items and compare many items but rarely purchase, and so on.

3. Proposed Model

3.1 Overview

In this study, we apply the idea of Goto et al. to model user behavior on a job-hunting website, and propose a latent class model that expresses the relation between user’s browsing and entry behaviors toward companies.

In the proposed model, the two kinds of user’s behaviors, that is, browsing of individual company pages and entries to companies, are represented by a two-dimensional vector composed of binary values. For example, if a user browsed an individual company page but did not make an entry to this company, it is represented as \( w = (1, 0) \). Also, like the model of Goto et al., we model the relation between browsing and entries by users for companies, based on the assumption of latent classes for users and companies independently.

Here, let us define that the set of \( P \) companies is \( C = \{ c_p : 1 \leq p \leq P \} \), the set of \( Q \) users is \( U = \{ u_q : 1 \leq q \leq Q \} \), the set of \( I \) latent classes of the companies is \( V_c = \{ s_i : 1 \leq i \leq I \} \), the set of \( J \) latent classes of the company is \( V_t = \{ t_j : 1 \leq j \leq J \} \). Then, the probability model is formulated as

\[
P(c_p, u_q, w) = \sum_{s_i} \sum_{t_j} P(s_i) P(t_j) P(c_p | s_i) P(u_q | t_j) P(w | s_i, t_j) \quad (7)
\]

The graphical model of the proposed model is described in Figure 3.

Figure 3: Graphical model of the proposed model.
3.2 Training Procedure

Here, let \( L \) be the number of all training data, the user in the \( l \)-th data be \( y_l \in U \), the company in the \( l \)-th data be \( x_l \in C \), and let the vector of browsing and entry activities by the user \( x_l \) for the company \( y_l \) be \( w = (w_{1l}, w_{2l}) \). The log-likelihood function for parameter estimation is given by the following equation (8):

\[
LL = \sum_{l=1}^{L} \log P(x_l, y_l, w_l)
\]

The parameters of equation (1) in the proposed model which locally maximizes equation (8) can be estimated by the EM algorithm. By using this algorithm, the parameters \( P(s_j), P(t_j), P(c_p|s_i), P(u_q|t_j) \) and \( P(w|s_i, t_j) \) maximizing the log likelihood can be estimated. In this algorithm, the E-step and M-step of equations (9)—(14) are repeated until the log likelihood function \( LL \) of equation (8) converges, and the parameters are updated.

**[E-step]**

\[
P(s_i, t_j|x_l, y_l, w_l) = \frac{P(s_i)P(t_j)P(x_l|s_i)P(y_l|t_j)P(w_l|s_i, t_j)}{\sum_{i=1}^{L} \sum_{j=1}^{J} P(s_i)P(t_j)P(x_l|s_i)P(y_l|t_j)P(w_l|s_i, t_j)}
\]

**[M-step]**

\[
P(s_i) = \frac{\sum_{j=1}^{J} \sum_{t_j \in Y_i} P(s_i, t_j|x_l, y_l, w_l)}{L}
\]

\[
P(t_j) = \frac{\sum_{i=1}^{I} \sum_{s_i \in Y_i} P(s_i, t_j|x_l, y_l, w_l)}{L}
\]

\[
P(c_p|s_i) = \frac{\sum_{j=1}^{J} \sum_{t_j \in Y_i} P(s_i, t_j|x_l, y_l, w_l) \delta(x_l = c_p)}{L \times P(s_i)}
\]

\[
P(u_q|t_j) = \frac{\sum_{i=1}^{I} \sum_{s_i \in Y_i} P(s_i, t_j|x_l, y_l, w_l) \delta(y_l = u_q)}{L \times P(t_j)}
\]

\[
P(w|s_i, t_j) = \frac{\sum_{j=1}^{J} \delta(w_l = w)P(s_i, t_j|x_l, y_l, w_l)}{\sum_{j=1}^{J} P(s_i, t_j|x_l, y_l, w_l)}
\]

By using \( P(w|s_i, t_j) \) calculated by equation (14), we can analyze the relation between browsing individual company pages and entries to companies by combining latent classes \( s_i \) and \( t_j \).
4. Analysis

4.1 Data Set

In order to show the effectiveness of the proposed model, we demonstrate an analysis by using actual data on a major job-hunting website, “Site-A,” in Japan. In this analysis, we use the browsing and entry data from March 1, 2015, to March 31, 2015. In addition, we target users who have made entries to more than 10 companies, and target the companies that have been browsed by and received entries at least once from target users. Based on the pre-analysis, the number of latent classes of companies is set to $I = 4$, and the number of latent classes of users is set to $J = 3$.

4.2 Result

As a result of parameter estimation by the EM algorithm, the log likelihood converged as follows.

![Figure 4: Convergence process of log likelihood](image)

Tables 2, 3, and 4 show the conditional probabilities of browsing and application actions for each combination of latent classes.

Table 2: $P(w_1 = 1(\text{browse}), w_2 = 0(\text{not entry})|s, t)$

<table>
<thead>
<tr>
<th>User \ Company</th>
<th>$s_1$</th>
<th>$s_2$</th>
<th>$s_3$</th>
<th>$s_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>0.989</td>
<td>1.000</td>
<td>1.000</td>
<td>0.718</td>
</tr>
<tr>
<td>$t_2$</td>
<td>0.000</td>
<td>0.727</td>
<td>0.056</td>
<td>0.000</td>
</tr>
<tr>
<td>$t_3$</td>
<td>0.017</td>
<td>0.996</td>
<td>0.540</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Table 3: $P(w_1 = 0(\text{not browse}), w_2 = 1(\text{entry})|s, t)$

<table>
<thead>
<tr>
<th>User \ Company</th>
<th>$s_1$</th>
<th>$s_2$</th>
<th>$s_3$</th>
<th>$s_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$t_2$</td>
<td>0.840</td>
<td>0.057</td>
<td>0.000</td>
<td>0.270</td>
</tr>
<tr>
<td>$t_3$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table 4: $P(w_1 = 1(\text{browse}), w_2 = 1(\text{entry})|s, t)$

<table>
<thead>
<tr>
<th>User \ Company</th>
<th>$s_{1}$</th>
<th>$s_{2}$</th>
<th>$s_{3}$</th>
<th>$s_{4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{1}$</td>
<td>0.011</td>
<td>0.000</td>
<td>0.000</td>
<td>0.282</td>
</tr>
<tr>
<td>$t_{2}$</td>
<td>0.160</td>
<td>0.216</td>
<td>0.944</td>
<td>0.730</td>
</tr>
<tr>
<td>$t_{3}$</td>
<td>0.983</td>
<td>0.004</td>
<td>0.460</td>
<td>0.996</td>
</tr>
</tbody>
</table>

By comparing the cells of the same latent class combination in these three tables, it is possible to grasp the user behavior toward each company latent class. Focusing on the latent class $s$ on the company side, from Table 2, companies belonging to the latent class $s_{2}$ are viewed by student users but tend to have difficulty in receiving many entries. On the other hand, from Table 4, it can be confirmed that companies belonging to the latent class $s_{4}$ tend to be both browsed and receive entries. Next, focusing on the latent class $t$ on the user side, from Table 2, users belonging to the latent class $t_{1}$ do not tend to make entries by only viewing. On the other hand, from Table 3, it can be confirmed that users belonging to the latent class $t_{2}$ tend to make only the entry action without browsing the individual company page. From these results, it can be said that the relation between browsing and entry actions is represented by the combination of latent classes of users and companies by applying the proposed model.

Next, Table 5 and 6 show the percentages of companies and users belonging to each latent class when companies and users are assigned to each latent class that have the maximum affiliation probabilities $P(s_{i}|c_{p})$ and $P(t_{j}|u_{a})$ calculated by Bayes’ theorem.

Table 5: Percentage of companies belonging to each latent class

<table>
<thead>
<tr>
<th>Latent class</th>
<th>$s_{1}$</th>
<th>$s_{2}$</th>
<th>$s_{3}$</th>
<th>$s_{4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>13.6%</td>
<td>33.1%</td>
<td>35.6%</td>
<td>17.6%</td>
</tr>
</tbody>
</table>

Table 6: Percentage of users belonging to each latent class

<table>
<thead>
<tr>
<th>Latent class</th>
<th>$t_{1}$</th>
<th>$t_{2}$</th>
<th>$t_{3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>26.2%</td>
<td>16.6%</td>
<td>57.1%</td>
</tr>
</tbody>
</table>

From Tables 5 and 6, it is can be said that the percentage of companies belonging to the latent class $s_{2}$, which has difficulty receiving entries with viewing only, and that of users belonging to the latent class $t_{3}$, showing different behavior tendencies depending on the latent class of companies, are high. Therefore, many users and companies are likely to be “with only browsing, but with no entry.” In addition, the user of the latent class $t_{3}$, to which the largest number of users belongs, tends to take two actions, “with browsing, but with no entry” and “with both browsing and entry,” for companies of latent class $s_{3}$, to which the largest number of companies belong with about the same probability.

Next, for each company, the value of $P(w|c_{p})$ is calculated by the following formula.

$$P(w|c_{p}) = P(w_{1}|s_{1})P(s_{1}|c_{p})$$

(15)

Table 7 shows the percentages of patterns of vector $w$ with the highest value of $P(w|c_{p})$ for each company.

Table 7: Percentage of action patterns that companies tend to receive

<table>
<thead>
<tr>
<th>Patterns of vector $w$</th>
<th>$w_{1}$ = 1(\text{browse}), $w_{2} = 0$(\text{not entry})</th>
<th>$w_{1}$ = 0(\text{not browse}), $w_{2} = 1$(\text{entry})</th>
<th>$w_{1}$ = 1(\text{browse}), $w_{2} = 1$(\text{entry})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>72.34%</td>
<td>0.00%</td>
<td>27.66%</td>
</tr>
</tbody>
</table>
companies have a high probability of receiving entries.

### 4.3 Discussion

First, we check the difference of characteristics between latent classes of users. The ratios of the average number of users’ entries and individual company page views for each latent class to the lowest number of entries and page views among latent classes are shown in Table 8 below.

<table>
<thead>
<tr>
<th>Latent class</th>
<th>The ratio of average number of entries</th>
<th>The ratio of average number of page views</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>1.000</td>
<td>3.109</td>
</tr>
<tr>
<td>$t_2$</td>
<td>1.412</td>
<td>1.000</td>
</tr>
<tr>
<td>$t_3$</td>
<td>1.337</td>
<td>1.616</td>
</tr>
</tbody>
</table>

Examining the average numbers of entries and views, there are significant differences in the average numbers of entries and page views between $t_1$ and $t_3$ in Welch's t-test with the significance level of 5%. This means that the behavioral tendencies of the users belonging to the latent class $t_1$ in the average number of entries and page views are different from those of the users belonging to the latent class $t_3$. Here, let us consider the characteristics of each latent class obtained by the results from Tables 2 to 4 together. Compared to other clusters' ratios, the users belonging to the latent class $t_1$ tend to be less inclined to make entries with browsing only because $P(w_1 = 1, w_2 = 0 | s, t)$ is very high in Table 2 and other conditional probabilities are very low in Table 3 and 4. On the other hand, the users belonging to the latent class $t_3$ are more inclined to both browse and make entries because $P(w_1 = 1, w_2 = 1 | s, t)$ is relatively higher than the latent class $t_1$. These behavioral characteristics may cause the significant difference in the average numbers of entries and page views shown in Table 8.

Second, we examine the characteristics of each latent class of companies. The ratio of the number of companies belonging to each latent class for each company size is shown in Table 9 below.

<table>
<thead>
<tr>
<th>Company size \ Latent class</th>
<th>$s_1$</th>
<th>$s_2$</th>
<th>$s_3$</th>
<th>$s_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 50 employees</td>
<td>6.2%</td>
<td>59.1%</td>
<td>30.7%</td>
<td>4.0%</td>
</tr>
<tr>
<td>50~100 employees</td>
<td>8.3%</td>
<td>46.4%</td>
<td>36.6%</td>
<td>8.6%</td>
</tr>
<tr>
<td>100~300 employees</td>
<td>10.0%</td>
<td>37.2%</td>
<td>40.7%</td>
<td>12.2%</td>
</tr>
<tr>
<td>300~500 employees</td>
<td>12.4%</td>
<td>27.8%</td>
<td>40.8%</td>
<td>19.0%</td>
</tr>
<tr>
<td>500~1000 employees</td>
<td>15.4%</td>
<td>21.6%</td>
<td>38.6%</td>
<td>24.4%</td>
</tr>
<tr>
<td>1000~3000 employees</td>
<td>19.8%</td>
<td>17.9%</td>
<td>30.1%</td>
<td>32.3%</td>
</tr>
<tr>
<td>3000~5000 employees</td>
<td>27.2%</td>
<td>12.1%</td>
<td>22.4%</td>
<td>38.3%</td>
</tr>
<tr>
<td>More than 5000 employees</td>
<td>38.3%</td>
<td>13.3%</td>
<td>18.6%</td>
<td>29.6%</td>
</tr>
</tbody>
</table>

In Table 9, the smaller the company, the higher the proportion of companies belonging to the latent class $s_2$ whose companies have difficulty receiving entries with browsing only. On the other hand, it can be said that the proportion of companies belonging to the latent class $s_4$, which is more likely to receive both browsing and entries, is higher for the larger companies.

According to the analysis result, it can be said that companies belonging to the latent class $s_2$ can expect to increase the number of entries by student users who have viewed their individual company pages, for example, by improving individual company pages and improving corporate impression. If the number of page views by student users for the company latent class $s_2$ increases, that of the entries may also increase. This is one possible way for the individual company page to become more attractive to student users. Meanwhile, because companies belonging to the latent class $s_1$ have different behavior tendencies toward each user latent class, the companies should take suitable customized actions for each class of users.

Due to the confidential information of the data donor, the real values cannot be published.
5. Conclusion and Future Works

This research proposed a latent class model to analyze users' browsing and entry behaviors at the same time on the job-hunting website, and showed the results of the analysis by the proposed model. The results of the analysis, we could acquire interesting findings about the relation between users' browsing and entry behaviors. This analysis gives useful information to support student's job-hunting activities.

Future works include modeling the time sequences of individual company page browsing and entry-making, and examining models incorporating other behavior histories. Moreover, it is necessary to analyze the contents of the individual company pages for each latent class of the company. By analyzing the contents of the page, it may be possible to understand the cause of users' behavioral differences.

References:


Yuuki Sugiyama is a graduate student of Department of Industrial and Management Systems Engineering, Graduate School of Creative Science and Engineering, Waseda University, Tokyo, Japan. He is now studying in the field of data mining.
Takumi Arai graduated from Department of Industrial and Management Systems Engineering, Graduate School of Creative Science and Engineering, Waseda University, Tokyo, Japan. He studied in the field of Bayesian prediction.

Tianxiang Yang is now a student of the doctoral program and a research associate in Department of Industrial and Management Systems Engineering at Waseda University, Tokyo, Japan. He received his M.E. degree in Global Information and Telecommunication Studies at Waseda University, Tokyo, Japan in 2015. His research interests include purchase history data analysis based on machine learning.

Tairiku Ogihara graduated from Department of Industrial and Management Systems Engineering, Graduate School of Creative Science and Engineering, Waseda University, Tokyo, Japan. He joined Recruit Career Co., Ltd. in 2015. He is engaged in the development of services using machine learning as a planning position of an internet portal site for job-hunting.

Masayuki Goto is a professor of Department of Industrial and Management Systems Engineering, School of Creative Science and Engineering, Waseda University, Tokyo, Japan. He received Dr.E. degree in Industrial Engineering from Waseda University in 2000. He is now studying in the field of data science, especially in machine learning and business analytics.