Hierarchical Structure Learning in a Bayesian Network for the Analysis of Purchasing Behavior

Ryota Kawabe1 *, Hiroaki Ito1, Haruka Yamashita2, Masayuki Goto1

1Waseda University, 3-4-1 Okubo, Shinjuku-ku Tokyo, 169-8555, Japan
2Sophia University, 4 Yonbancho, Chiyoda-ku, Tokyo, 102-0081, Japan

* kryota_09@ruri.waseda.jp

Abstract:
In the original Bayesian network model, the hierarchical structure of the variables is not assumed. When modeling the relation between the sales of products in a retail industry, it is better to consider a hierarchical structure of items (e.g., first, second, and third classifications). To apply a Bayesian network to such data, we have to focus on one hierarchy only in order to acquire a Bayesian network model. However, focusing only on the first or second classification provides a high-level view, but makes it difficult to understand customers’ purchasing behavior in detail. On the other hand, focusing on the third classification results in a considerable number of nodes and a complicated network structure. Thus, capturing the overall relationship between products is not straightforward. Therefore, we propose a hierarchical Bayesian network model and a new learning method based on max-min hill-climbing learning algorithm. In the proposed method, we focus on the hierarchical structure of products, which enables us to construct a lower layer that considers the causal relationships in the upper layer. Furthermore, we investigate a case study using actual hierarchical data on consumer purchases, and demonstrate the proposed model using a simulation analysis.

Keywords
Bayesian network, Structure learning, Hierarchical structure, Customer purchasing behavior

1. Introduction

Bayesian networks (BNs), which express the causal relations between variables, have become an increasingly popular area of study (Jensen (1996); Pearl (1985); Ueno (2013)). The structure of BNs is easy to understand, and they make it possible to predict conditional probabilities efficiently, given an event, using an estimated BN from a training data set. A typical analysis using a BN is performed as follows: (i) connect the nodes (i.e., random variables) that have causal relationships using a directed link; (ii) learn a network structure in the form of a directed acyclic graph; and (iii) estimate the posterior probabilities of unobserved nodes using some observation of nodes. A BN model is applicable to a wide range of problems and, thus, has been applied in many data analyses (Zou et al. (2004)). For example, Kamiyama et al. (2009) apply a BN model to a stock price trend analysis, Adachi et al. (2010) provide a solution to a multi-function machine fault diagnosis using a BN model, and Higashi (2014) applies a BN model to quality control. In the field of marketing, the applications of BNs have also expanded (Cui et al. (2006)). For example, Ishigaki et al. (2011) apply a BN model to predict customers’ purchasing behavior. The scope of the application of BNs is still expanding, and its value is increasing in the field of marketing. BNs have also been applied to purchasing data analyses (Hughes (2006); Swait and Adamowicz (2001)) to cluster customers and items simultaneously. In this research, we use a BN model to analyze consumers’ purchasing data and to determine the co-occurrence structure of product purchases.

In general, products in the retail industry are classified hierarchically (e.g., first, second, and third classifications). For example, “lettuce sandwich” is a product’s name, “food” is its first classification, “fabricated
foods” is its second classification, and “bakery” is its third classification. Depending on the retail store, there may also be a category for “sandwiches.” These categories provide an important classification system, used by retail stores to monitor sales and to manage the ordering of products. However, in the original BN model, the hierarchical structure of the variables is not assumed. Thus, in order to model the relation between the sales of products using an original BN model, we must focus on one level of hierarchical classifications. However, focusing on the first or second classification will provide a high-level view of the structure, but make it difficult to understand customers’ purchasing behavior in detail. On the other hand, creating a network that focuses only on the third classification results in a considerable number of nodes and a complicated network structure. No previous studies on BNs have focused on multiple hierarchies of products. Therefore, it is necessary to create a new BN model that represents both global and local viewpoints using the hierarchical structure of products, as well as to construct a learning method for the model.

On the other hand, the accuracy of predictions and inferences depend on the structure-learning method used in a BN. As one such method, Tsamardinos et al. (2006) proposed the max-min hill-climbing method (MMHC), which can learn a structure with high accuracy while reducing the computation time. This paper proposes a BN model that considers the hierarchical structure of variables, as well as a learning method for the model based on the MMHC. In this method, we focus on the hierarchical nature of products, and make it possible to construct lower-layer networks that consider the causal relationships in the upper-layer networks. Moreover, we apply the proposed method to consumers’ purchase history data, using a hierarchy provided by Macromill, Inc., and discuss the analysis results. The effectiveness of the proposed BN model is demonstrated by applying it in a data analysis and simulation experiment. Such experiments using the proposed BN model can help to identify efficient promotion policies.

2. Preparation

2.1 The Bayesian network

A BN is a model that expresses a causal relation stochastically using a network structure with directed links. In this study, each node in the BN is considered a product purchase event. The input values to the BN are obtained by categorizing the frequency of the purchases of goods. When applying a BN model to a problem, there are two steps when learning the structure of the BN and when inferring probabilities using the learned structure. First, in the structure-learning step, causal relations between the nodes of the learning data are expressed using directed links, creating a network structure. Here, a directed link is drawn between two nodes that have a causal relation and includes the direction of the conditional probability. Next, for probabilistic inferences (Cooper (1992); Murphy (2013)), values are input to the node from which an observation value is obtained, and the probability of the occurrence of an event of an unobserved node is predicted in a chain.

2.2 The max-min hill-climbing algorithm

Tsamardinos et al. (2006) proposed the MMHC method to enable structural learning in BNs. In this method, independent tests are performed between the nodes, an undirected graph is created between two nodes that are highly likely to be drawn, and then an undirected graph is searched explicitly using a score-based method. In the MMHC method, it is possible to reduce the calculation time by creating directed link candidates and by learning the network structure using a score-based method, based on the constraints.

2.2.1 Creating directed link candidates

In order to create directed link candidates between nodes by MMPC algorithm, the $G^2$ test (a likelihood ratio test) is used to test for independence between the nodes. Let Y be a target node, X be a node that needs to be verified with the target node, Z be the set of nodes that have already been adjudged to be related, a and b be the discrete numeric values taken on node X and Y, respectively, and c be a combination of discrete numeric values on the node set Z. Letting $S_{X,Y,Z}$ be the occurrence frequency of the combination $X,Y,Z$ in the data, the $G^2$ test statistics are given by equation (1):

$$G^2 = 2 \sum_{a,b,c} S_{X=a,Y=b,Z=c} \ln \frac{S_{X=a,Y=b,Z=c}}{S_{X=a,Z=c} S_{Y=b,Z=c}}$$  \hspace{1cm} (1)$$

2.2.2 Directed link structure

The structure-learning methods for BNs are classified into constraint-based methods (Cooper and Glymour (1999)) and score-based methods (Koivisto and Sood (2004)). Here, we focus on the greedy hill-climbing (GHC)
algorithm as a score-based method. In this algorithm, a structure with no links is set as the initial state. Then, the structure of the network is learned so that the Bayesian information criterion (BIC) (Akaike (1996)) is improved by choosing “add,” “invert,” “delete” functions on the directed links. The BIC is used in model selection and considers the model complexity and data size of the likelihood. In the step in which the link structure is created, efficient network structure learning is realized by applying the GHC algorithm to the undirected graph created in 2.2.1.

3. Proposed Method

Because the original BN model does not assume a hierarchical structure in the data, there is a risk that the direction of a causal relation in the network learned in the upper layer may contradict that of a relation in the lower layer. Moreover, when teaching a BN model using the MMHC method, the network becomes excessively complicated, making it difficult at times to interpret the output when there is a large number of nodes (e.g., the third classification data of products.) Therefore, we propose the following new BN structural-learning method: (i) extract only those upper-layer nodes that have co-occurrence relationships in the upper layer network; and (ii) construct a network at the lower level for the extracted pair. When using the MMHC method on the lower layers, the possible links and directions are restricted to those that are consistent with the co-occurrence relationship links and directions in the upper layer of the network structure. This makes it possible to interpret the co-occurrence between nodes, even when the number of nodes is enormous in the lower layer, and to reflect the co-occurrences in the upper layer in the co-occurrence relationships between the nodes. The proposed algorithm is shown below.

[Proposed Algorithm]

Step 1: Apply MMHC to the purchase history data, focusing on the upper layers, to build a BN at this level.
Step 2: Create a list of pairs of nodes that are linked in the upper layers of the network, where the co-occurrence relationships were confirmed in Step 1.
Step 3: For each pair extracted in Step 2, all directions of the relationships at the upper layer level are regarded as directed link candidates, and a network at the lower layer level is constructed using the MMHC method.

The proposed algorithm can prevent inconsistencies in the network structure between the upper and lower layers. Moreover, it provides a realistic method of constructing a relational model that covers many products as a whole. In retail shops, the number of goods is often enormous, some of which are goods with low sales. As a result, modeling the relationship between the sales of goods using a network becomes difficult. The proposed algorithm achieves this by constructing a hierarchical network structure.

4. An experiment

4.1 Experimental condition

In this research, we use consumer purchase history data from the Quick Purchase Report (QPR), provided by Macromill, Inc., to examine the usefulness of the proposed model. Here, we use the purchase history data of consumer monitor for 2015. Each site user sends a purchase record using a barcode reader system. In this way, the purchase histories of all site users are accumulated. There are 25 second classifications of products and 275 third classifications. This classification based on JAN code is defined by The Distribution Systems Research Institute. In addition, there are 7,872 users, and 7,827,088 data items. Here, because the BN is a model that uses categorical variables, the variables representing numbers of purchased items need to be discrete. Therefore, the numbers of purchased items are quantized in order to model the co-occurrence structure of consumers’ product purchases. When considering how to quantize the number of purchases of an item, the distribution of the number of purchases is important. The quantization process needs to ensure that the number of data items in each category after quantization is as equalized as possible. The quantization method employed here is based on the following three cases:

Case 1: When the median value of the number of purchased items is 0, the value is binarized as “0” or “1 or more” for the item.
Case 2: When the median value of the number of purchased items is one, the value is binarized as “1 or less” or “2 or more” for the item.
Case 3: In all other cases, the number of purchased items is ternarized so that the number of members in each category becomes equal.

4.2 Analysis focusing on second classification data

In order to show a BN of the co-occurrence structure of consumers’ product purchases, abbreviations are used because the names of the second classifications are too long to be used in the figure as they are. Table 1 shows the second classification names and their abbreviations.

Then, Fig.1 shows the network for the second classification level obtained from Step 1 of the proposed method, based on the abbreviations shown in Table 1.

| Table 1 Second classifications and their abbreviations |
|------------------------------------------------------|
| First | Second | Abbreviation |
|-------|--------|--------------|
| Food  | Fabricated food | Fab f |
|       | Fresh food | Fre f |
|       | Confectionery | Conf |
|       | Beverage | Bev |
| Daily necessities | Daily miscellaneous goods | Dmg |
|       | Drugs | Dru |
|       | Cosmetics | Cosm |
|       | Household goods | Hg |
|       | Carpentry goods | Carp |
|       | Pet Supplies | Pet |
| Cultural goods | Stationery office supplies | Sta |
|       | Toy | Toy |
|       | Books | Books |
|       | Instruments | Ins |
|       | Information equipment | Inf |
| Durable consumer goods | Furniture | Fur |
|       | Vehicle supplies | Vehi |
|       | Watches and glasses | W.G |
|       | Photographic supplies | Photo |
|       | Household appliances | Ha |
| Clothings | Clothing | Clo |
|       | Beddings | Bed |
|       | Personal supplies | Per |
|       | Shoes | Sho |
|       | Sports equipment | Spo |

Fig.1 Part of the network at the second classification level
For example, from Fig. 1, we see a co-occurrence relationship between purchasing “fresh food” and “household goods.” Calculating the conditional probability, the probability of buying “household goods” is 3.8% when not purchasing much “fresh food,” but this increases to 56.4% when buying “fresh food” as well. In the next section, we analyze the results between these two second classifications at the third classification level.

4.3 Analysis focusing on third classification data
Fig. 2 shows an example of the network at the third classification level from Step 3 of the proposed method. Note that the co-occurrence relation in the medium classification data has been maintained. Fig. 2 shows the relationship between “fresh food” and “household goods” as an example.

![Part of the network at the third classification level](image)

Table 2 Third classifications and their abbreviations (for three second classifications)

| Second                  | Thrid         | Abbreviation |
|-------------------------|---------------|--------------|
| Household goods         |               |              |
| Food rapping            | Fr            | Fr           |
| Cleaning supplies       | Cs            | Cs           |
| Laundry supplies        | Ls            | Ls           |
| Kitchenware             | Kit           | Kit          |
| Bathroom supplies       | Bath          | Bath         |
| Tablware                | Tabl          | Tabl         |
| Cooking equipment       | Cook          | Cook         |
| Living equipment        | Liv           | Liv          |
| Other household goods   | Other h       | Other h      |
| Fresh food              |               |              |
| Fisheries               | Fish          | Fish         |
| Animal husbandry        | Anim          | Anim         |
| Agricultural products   | Agri          | Agri         |
| Other fresh food        | Other f       | Other f      |
| Carpentry goods         |               |              |
| Painting tools          | Pt tl         | Pt tl        |
| Painting materials      | Pt mt         | Pt mt        |
| Gas water goods         | Gas           | Gas          |
| Gardening equipment     | Gard          | Gard         |
| Other carpentry goods   | Other c       | Other c      |

For example, from Fig. 2, we find a co-occurrence relation between purchasing “fishery” and “tableware.”
Similarly, to Section 4.2, the conditional probability of buying “tableware” is 42.4% when not purchasing much “fishery,” but this improves to 72.4% when also buying “fishery.”

From Fig. 4 there is a co-occurrence relation between three third classifications. By connecting the graphs in the same way as those in Fig. 2 and Fig. 3, the overall network is completed. The completed graph shows a change
in the conditional probabilities at both ends between the other nodes in Fig. 4. Next, we focus on the co-occurrence relation between “fishery” and “gas water goods.” The conditional probability of buying “fishery” is 10.55% when not purchasing “gas water goods,” but 24.16% when buying “gas water goods,” confirming the conditional probability changes.

In the proposed method, possible links in the network at the third classification level are limited to nodes that have a co-occurrence relationship at the second classification level. That is, the structures of the directed links in the network at the third and second classification levels are consistent. Fig. 2 shows the detailed local structures at the third classification levels, and the second classification level (see Fig. 1) shows the overall structure of the co-occurrence of consumers’ purchases. By hierarchically visualizing the overall structure and the details of local structures, the model is that much easier for a user to understand.

4.4 Simulation analysis

In this section, we describe a simulation analysis to demonstrate how to plan improvements related to customers’ purchases using the estimated model as an example. In this simulation, we analyze the data using the proposed model and show which products are strongly linked to purchases of other products. In this way, we identify the product categories that maximize the overall purchase probability. We perform a probabilistic inference using the learned network of each item hierarchy in Section 4.1. Specifically, we focus on the category with the largest number of purchase items at nodes other than the target node in order to obtain the average variation ratio when the occurrence probability of the category with the largest number of items of the target node is changed to 1. This approach helps us to show the impact on purchase increases throughout the network. That is, the average variation ratio $R_l \ (0 \leq R_l \leq 1)$ is a value indicating the ease of connection between a purchased product and some other product.

In order to define the average variation ratio $R_l$, the category with the most purchase amount of the target node $n_l$ and that of another node $n_k$ are denoted by $c_l$ and $c_k$ respectively. Here, $P(X_k = c_k)$ is a prior probability of the category with the most purchase amount of the note $n_k$, and $P(X_k = c_k | X_l = c_l)$ is the conditional probability given by $X_l = c_l$. If the probability $P(X_k = c_k | X_l = c_l)$ on the node $n_k$ considerably increases when the purchase amount of the note $n_k$ is changed to $X_l = c_l$, then the influence of the note $n_l$ for the node $n_k$ is relatively strong. Therefore, we can regard $P(X_k = c_k | X_l = c_l) - P(X_k = c_k)$ as the degree of the influence of the node $n_l$ to the node $n_k$. Letting $D$ be the total number of nodes, we can define the average variation ratio $R_l$ for all nodes by the following equation (2):

$$R_l = \frac{\sum_{k \neq l} P(X_k = c_k | X_l = c_l) - P(X_k = c_k)}{D-1}$$ (2)

When $R_l$ in (2) is determined for the second classification of all $D$ items, the overall purchase probability increased for “household goods,” where $R_l$ is equal to 11.9%. On the other hand, the average of the average variation rates of the other nodes is 5.4%. Detailed variation rates are shown in Table.3.

Table.3 Top 10 of average variation ratios and the average for the second classification level

| Second classification       | $R_l$ |
|----------------------------|-------|
| Household goods            | 0.119 |
| Daily miscellaneous goods  | 0.098 |
| Fabricated food            | 0.087 |
| Carpentry goods            | 0.083 |
| Fresh food                 | 0.079 |
| Cosmetics                  | 0.078 |
| Watch and glasses          | 0.067 |
| Personal supplies          | 0.061 |
| Confectionery              | 0.060 |
| Stationery office supplies | 0.059 |
| Average                    | 0.054 |

Therefore, in order to increase overall purchases, it is most efficient to improve the purchases of “household goods”. For example, after improving the purchases of “household goods,” purchases of “processed foods” increase by 23%.

Next, to carry out a more detailed analysis, we apply the same simulation analysis to the third classification level. Here, we focus on those nodes that change the occurrence probability of “household goods,” which had the highest average variation ratio in the second classification data (see Table.4).
The results show that the three most influential categories in terms of purchase probabilities are “other household goods,” “kitchen utensils,” and “living goods,” with average fluctuation ratios of 1.7%, 1.4%, and 1.3%, respectively. The average variation rate of all nodes is 1.1%.

As described above, this analysis shows how to increase customers’ overall purchases by increasing the purchases of certain third classification items. This result can be used to plan marketing measures.

### 4.5 Consideration

In the proposed method, because we divide the nodes into several groups, there is an advantage of a reduced computational cost when the number of nodes becomes large. The relationship between the computational time and the number of nodes is shown in Table 5.

Table 5 The number of nodes and calculation time

| Number of nodes | Method | Time (sec) |
|-----------------|--------|------------|
| 25              | MMHC   | 2.12       |
| 50              | MMHC   | 4.60       |
| 275             | MMHC   | 227725.73  |
| 26              | Proposed | 3.30     |
| 45              | Proposed | 8.93     |

The maximum number of nodes between two second classes is 45, with a calculation time of 8.93 seconds. Because the number of links at the second level is 38, the calculation in the proposed method ends after at most 340 seconds. This is considerably faster than calculating 275 nodes simultaneously using the MMHC method. These results clarify the computational advantages of the proposed method.

Table 6 compares the numbers of nodes that are not connected to other nodes at the third classification level when using the proposed method and the MMHC method.

Table 6. The number of unconnected nodes

| Method        | Link | Connected nodes | Isolated nodes |
|---------------|------|-----------------|----------------|
| MMHC          | 231  | 196             | 79             |
| Proposed model | 260  | 147             | 128            |

Although the BN model of the proposed method has many links, there are many isolated nodes. This is because a strong cause-and-effect relationship within the same classification is ignored. This is affected on how to set the classification levels and purchase levels. Classification levels are already defined in dataset. On the other hand, classification of purchase levels is arbitrary; therefore, it is necessary to decide classifying method in consideration of this effect. For isolated nodes, we cannot expect an increase in sales based on their co-occurrence relations and purchases. Thus, it is necessary to promote from another viewpoint.

In general, the number of the data items associated with a node higher in the hierarchy (e.g., the second classification) is larger than that of nodes lower down in the hierarchy (e.g., the third classification). Therefore, setting the direction of the links between the higher hierarchy nodes enables us to construct a highly reliable model in terms of stability, with few exceptions in the lower layer of the hierarchy.
As a practical example, when discount coupons are issued, the results of this study suggest customizing the coupons for individual consumers. In addition, because the influence of a promotion is indicated by a numerical value, a marketing plan can determine an acceptable discount for each item.

5. Conclusions and future work

In this research, we proposed a new model and its learning method using the hierarchical relationship in a BN model, which conventionally does not assume a hierarchical relationship of variables. The proposed method creates a network structure with a lower layer that maintains the direction of co-occurrences in the relevant relations in the upper layers. This method eliminates the difficulty of interpretation when a large number of nodes exists, which we showed by applying it to actual data. In particular, when analyzing the relationships between the sales of products in retail shops, there are usually a great many items. As a result, the proposed method is effective in this context. Because the proposed method uses a hierarchical structure of variables, it can be applied to data with the same structure.

In future work, we will devise an optimal method of categorizing observed values. Because a BN models the relationship between discrete random variables, discretization is necessary when modeling continuous random variables. Thus, there is room for further research on the discretization method.

References:

Adachi, K., Yamada, N., Uwatoko, K., and Yasukawa, K. (2010), “Fault Diagnostic Technology for Duplicator Machine by Bayesian Network,” Fuji Xerox Technical Report, No.19, pp.78–87. (in Japanese)
Akaike, H. (1996), “AIC, MDL and BIC,” Operation Research, 7, pp.375–378.
Cooper, G., and Herskovits, E. (1992), “A Bayesian Method for the Induction of Probabilistic Networks from Data.,” Machine Learning, 9, pp.309–347.
Cooper, G., and Glymour, C. (1999), “Computation, Causation, and Discovery,” MIT Press, Cambridge, MA.
Cui, G., Wong, M.L., and Lui, H.K. (2006), “Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming,” Management Science, 52(4), pp.597–612.
Higashi, R. (2014), “Product Development Completeness Diagnosis Using Bayesian Statistics,” Konica Minolta Technology Report, 11, pp.31–36. (in Japanese)
Hofmann, T. (2001), “Unsupervised Learning by Probabilistic Latent Semantic Analysis,” Machine Learning, 42(1–2), pp.177–196.
Hughes, A. (2006), “Strategic Database Marketing,” McGraw-Hill Education.
Ishigaki, T., Takenaka T., and Motomura Y. (2011), “Customer Behavior Prediction System by Large Scale Data Fusion in a Retail Service,” The Japanese Society for Artificial Intelligence, 26(6), pp.670–681. (in Japanese)
Jensen, F. (1996), “An Introduction to Bayesian Networks,” CRC Press: 1st edition.
Kamiyama, K., Yi, Z., Kamijima, Y., and Kita, E. (2009), “Analysis of Stock Price Fluctuation by Using Bayesian Network,” IPSISIG Technical Report, 2009-MPS-76-16, pp.1–5. (in Japanese)
Koivisto, M., and Sood, K. (2004), “Exact Bayesian Structure Discovery in Bayesian Networks,” Journal of Machine Learning Research, 5, pp.549–573.
Macromill, Inc.: https://www.macromill.com/ (accessed 2017-12-14).
Murphy, K.P., Weiss, Y., and Jordan, M.I. (2013), “Loopy Belief Propagation for Approximate Inference, An Empirical Study,” Proceedings of Uncertainty in AI, pp.467–475.
Zou, M., and Conzen, S.D. (2004), “A New Dynamic Bayesian Network (DBN) Approach for Identifying Gene Regulatory Networks from Time Course Microarray Data,” Bioinformatics, 21(1), pp.71–79.
Pearl, J. (1985), “Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning,” in Proceedings of the 7th Conference of the Cognitive Science Society, 1985, pp.329–334.
Swait, J., and Adamowicz, W. (2001), “The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching,” Journal of Consumer Research, 28(1), pp.135–148.
Tsamardinos, I., Brown, L.E., and Aliferis, C.F. (2006), “The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm,” Machine Learning, 65(1), pp.31–78.
Ueno, M. (2013), “Bayesian Network,” Corona Publishing.
Acknowledgements:

We would like to express our gratitude to Macromill, Inc., for providing us with the consumer purchase history data QPR. We would also like to thank Dr. Kenta Mikawa, Shonan Institute of Technology, and all members of the Goto laboratory, Waseda University, for their helpful suggestions related to this research. A part of this study was supported by JSPS KAKENHI Grant Numbers 26282090 and 26560167.

Author’s biographical notes

**Ryota Kawabe** 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 sparse regularization and graph mining.

**Hiroaki Ito** 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 business data analysis based on graph mining.

**Haruka Yamashita** is an assistant professor at the School of Science and Technology, Sophia University, Japan. She was born in Tokyo, Japan on September 30, 1987. She received the Bachelor’s degree from Tokyo University of Science and Technology in 2010. Moreover, she received the Master’s degree in 2012 and her Ph.D. degree from Keio University in 2015. Her research interests include applied statistics, multivariate analysis, and machine learning. She is a member of the Japanese Society of Applied Statistics, Japanese Industrial Management Association, and Japanese Society of Quality Control.

**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.