https://doi.org/10.20965/jaciii.2013.p0221

Paper:

Bayesian Network Model that Infers Purchase Probability in an Online Shopping Site

Yutaka Matsushita* **and Syunsuke Maeda****

*Department of Psychological Informatics, Kanazawa Institute of Technology
3-1 Yatsukaho, Hakusan, Ishikawa 924-0838, Japan
E-mail: yutaka@neptune.kanazawa-it.ac.jp
**Asobimo Incorporated
5-26-19 Nishiikebukuro, Toshima-ku, Tokyo 171-0021, Japan
E-mail: s_maeda@asobimo.com

[Received November 16, 2012; accepted January 29, 2013]

In order to understand the properties of online shopping that contribute to visitors' purchasing habits, we have developed a Bayesian network model that infers the probability of purchase from eye movement and web log data. The results obtained from this model imply that a short visit time on catalog pages and a high frequency of fixation on all pages are related to increased purchase probability. Furthermore, it is shown that websites conforming to Internet Usability Guidelines (IUG) make visitors feel little stress regardless of browsing patterns, and that websites not conforming to IUG require a very short visit time on catalog pages if low stress is to be maintained.

Keywords: Bayesian network, inference, shopping site, product purchase, eye movements

1. Introduction

In general, websites (e.g., online shopping or information provision sites) are designed to help visitors achieve their objectives. Many studies on web usability have been carried out already. Jakob Nielsen [1] established Internet Usability Guidelines (IUG), a set of guidelines concerning the contrast colors of characters vs. backgrounds and the desirable number of pages. Thus, it can be assumed that sufficient research has been done on web usability.

A recent trend in website research is the shift toward studies involving inference of visitors' behavior. For example, Khosravi et al. [2], using naive Bayes classification, inferred the next browsing page from the construction of pages and web log data. Adachi et al. [3] predicted visitors' target pages from the word-frequency of anchor texts and their web log data. However, these studies did not address whether visitors could achieve their objectives, despite the fact that their doing so is a prime concern for website designers. Furthermore, these studies did not use eye movement data as explanatory variables. Here, eye movement data appear to be useful in evaluating websites. Indeed, Toda et al. [4] investigated the relationship between duration of fixation and the behavior patterns of subjects through a search experiment with key words, and concluded that if the duration of fixation is long, subjects are making a judgment on the compatibility of their found information with wanted information; if short, then they are exploring the location of each key word.

Considering these factors, this paper focuses on an online shopping site to develop a probabilistic model that infers whether a correlation exists between the purchase of a product and the visitor's eye movement and web log data. With this approach, a Bayesian network proves to be a useful analytical method. The reason for this is that, given eye movement and web log data, what we want to know is the possibility of purchase, which corresponds well to the concept of a posterior probability. Another reason is that many existing research examples [5–7] demonstrate that feelings and behavior are sufficiently inferred by the use of Bayesian network models.

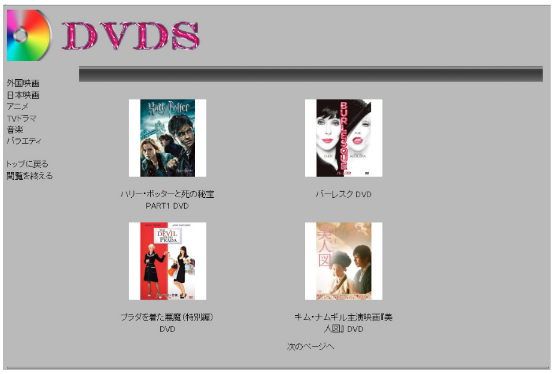
2. Experiment

The aim of this experiment was to track behavioral properties in a situation where the subjects could freely browse an online shopping site for DVD movies and to examine whether these properties contribute to the purchase of a product. Henceforth, if subjects purchased certain products, we say the objective has been achieved; otherwise, the objective has not been achieved.

Our online shopping site consisted of a top page, catalog pages (CPs) for DVD movies, and product pages (PPs) with a summary of each movie. Links were provided between the top page and each CP, and between each CP and each PP (the page of a product contained in the CP). Movies were divided into 6 genres, and 25 products were provided per genre, contributing to a total of 150 products. Each PP was equipped with a purchase button, and all pages had an exit button. Subjects could finish the experiment by clicking either the purchase button on an arbitrary PP or the exit button on every page.

Usability is considered an important factor in purchase

(a) Conforming stimulus



(b) Non-conforming stimulus

Fig. 1. A catalog page in stimuli of each type.

Table 1. Factors and levels for design of stimuli.

Factor	Level	
IUG	Conforming	Non-conforming
Product	A	B

A total of 50 students in the Kanazawa Institute of Technology participated in this experiment. After we had explained the methods of purchase and escape, each subject was left alone in a laboratory, and asked to visit our shopping site as usual. While subjects were exposed to the stimuli, web log data and eye movement data were recorded via Web Tracer (SRA) and EMR-9 (NAC Image Technology), respectively. There were 50 samples for stimuli of each type. including those conforming and not conforming to IUG, and the total number of all samples was 100. Subjects spent on average 155.16 sec and 138.95 sec visiting the conforming and non-conforming websites, respectively. In addition, after finishing the first and second experiments, subjects were required to fill out questionnaires, rating the degree of several aspects – including usability, interest in DVD movies, and stress – on a scale of one to five. It should be noted that "interest" here refers to the degree of interest in the products for everyday life (i.e., an existing interest), not satisfaction with the products specifically listed in this experiment.

behavior. The good or bad usability of a stimulus was decided according to whether it conformed to IUG. More precisely, a stimulus conforming to IUG was defined as a website in which the contrasting colors of characters vs. backgrounds were black vs. white, and the number of CPs was 3. On the other hand, a stimulus not conforming to IUG was defined as a website in which the contrasting colors of characters vs. backgrounds were black vs. gray and the number of CPs was 7. **Figs. 1(a)** and **(b)** show respective samples of the CPs contained in conforming and non-conforming stimuli. CPs involved the biggest difference between the two types of stimuli because they listed different numbers of products and were equipped with different link menus according to their conforming status. Moreover, the top page was similar except for the contrasting colors of characters vs. backgrounds, as were PPs, although they also supplied the different link menus according to their conforming status.

The aspect of usability was to be evaluated from the perspective of achieving the purpose. Since every subject was asked to visit stimuli both conforming and not conforming to IUG, it was important to keep his or her interest at a similar level. For this reason, two different sets of products were provided, which were alternately presented in the first and second experiments. Consequently, the stimuli were designed by two factors, each of which contained two levels (**Table 1**), resulting in a total of four stimuli. Note that a combination of the levels of these two factors and the presenting order of stimuli were counterbalanced by the subjects.

3. Results of the Experiment

Items from the web log and eye movement data are listed as follows. The first three items belong to web log data, and the last three items correspond to eye movement data.

- The (total) number of visited pages
- Average visit time in PPs (PP time)
- Average visit time in CPs (CP time)
- The total number of fixations (the number of fixations)
- Average duration of fixation (duration of fixation)
- Average eye movement velocity (movement velocity)

The average visit time in PPs or CPs refers to the time that a visitor spent on average on one page of PPs or CPs. The average duration of fixation is the sum of the duration divided by the number of fixations. The average eye movement velocity means the average of eye movement velocities across all visited pages. Although these data were in themselves independently measured, in view of the problem of multiple comparisons, we will henceforth use Šidák's method [8] to test the statistical significance of the difference between these averages.

Table 2 shows the rates of objective achievement and the averages of usability, PP time, and CP time for stimuli both conforming and not conforming to IUG. The difference between usability evaluations of these two types

Table 2. Comparison between conforming and non-conforming stimuli.

	Conforming	Non-conforming
Achievement rate	0.78	0.68
Usability	3.76	2.50
PP time (sec)	11.44	11.95
CP time (sec)	5.55	3.69

Table 3. Averages of web log and eye movement data.

Item	Achievers	Non-achievers
The number of visited pages	26.81	21.56
PP time (sec)	11.85	11.28
CP time (sec)	4.38	5.27
The number of fixations	251.89	183.33
Duration of fixation (msec)	525.37	632.29
Movement velocity (mm/sec)	1847.53	1666.47

of stimuli was significant at a level of 0.1%; hence, conforming stimuli were considered superior in usability to non-conforming. However, since the difference between the achievement rates was not significant, it was revealed that an improvement in usability did not always increase the achievement of the objective. In addition, the CP time was significant at a level of 0.1%, whereas the PP time was not. This implies that visitors tended to finish browsing CPs faster in non-conforming stimuli, and that they spent much the same time on PPs in both stimuli. In addition, the difference in the achievement rates was not significant ($\chi^2 = 2.49$, n.s.) between the sets (A and B) of products. Therefore, the provision of the two sets was considered rational.

Subjects were divided into objective achievement and non-achievement groups. The number of samples belonging to the objective achievement and non-achievement groups was 73 and 27, respectively. **Table 3** shows the average of each item across subjects belonging to each group. The number of fixations and the duration of fixation were significant at a level of 5% and the CP time, at a level of 15%. **Table 3** illustrates that in the achievement group, the CP time and the average duration of fixation was shorter, but the number of fixations was larger. According to the conclusion of Toda et al. [4], it is considered that subjects who were able to achieve the objective had the tendency to search information rather than to judge it.

Table 4 reveals averages of feeling evaluations for the achievement and non-achievement groups. The interest evaluation was significant at a level of 0.1% and the stress evaluation at a level of 5%, but usability was not significant. Therefore, it is suggested that subjects who achieved the objective had more interest in DVD movies and were under less stress while browsing pages, but did not necessarily perceive better usability.

Table 4. Averages of feeling evaluations.

Item	Achievers	Non-achievers
Interest	4.18	3.22
Stress	2.70	3.26
Usability	3.22	2.89

4. Inference by Bayesian Network Models

Since we have learned that the difference between the achievement rates is not significant between conforming and non-conforming stimuli, our study proceeds with the following approach: an optimal graph structure of the inference model is chosen for the two types of stimuli as a whole; prior and posterior probabilities are then evaluated for each.

4.1. Choice of Explanatory Variables

All items included in the web log and eye movement data from **Table 3** are candidates for explanatory variables. The reason is that by also including non-significant items, we can maintain the possibility of increasing the accuracy of inference. Since these data were quantitative, they had to be discretized to evaluate probability values in a discrete form. On the assumption that the distribution of each data set was normal, each of the data sets was divided into three categories: S (Small), M (Medium), and L (Large), such that the frequencies were similar. Using web log and eye movement data (transformed into qualitative data) as explanatory variables, and achievement or non-achievement (two states) as objective variables, we chose a set of optical explanatory variables through the use of the analytical software Bayonet.

As for the graph structure, we were required to create links between nodes having strong correlations with each other. Considering the correlation coefficients of any two explanatory variables (**Table 5**), the following links were initially generated:

- From the number of visited pages to the number of fixations.
- From each CP time and PP time to duration of fixation.
- From movement velocity to CP time, PP time, duration of fixation, and the number of fixations separately.

An optimal graph structure was selected on the basis of AIC and the accuracy rate. Note here that we deemed a link invalid whenever at least one of the linked nodes was eliminated in the choice process. AIC is expressed as

$$\begin{aligned}\mathrm{AIC} = &-2 \times \mathrm{MLL}\\ &+2 \times (\text{the number of probability parameters}),\end{aligned}$$

where MLL refers to a maximum log-likelihood and the number of probability parameters denotes the total number of probability values running freely through the infer-

Table 5. Correlation coefficients between variables.

	The number of visited pages	PP time	CP time	The number of fixations	Duration of fixation	Movement velocity
The number of visited pages	1					
PP time	0.12	1				
CP time	-0.35	0.05	1			
The number of fixations	0.78	0.34	-0.09	1		
Duration of fixation	0.06	0.20	0.23	-0.12	1	
Movement velocity	-0.18	0.36	0.45	0.37	-0.29	1

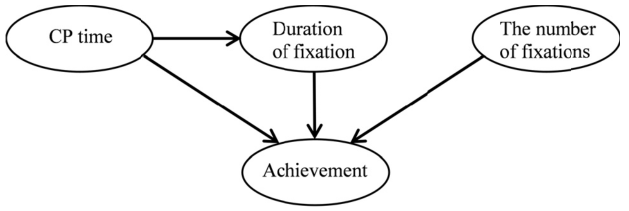


Fig. 2. Graph structure for achievement probability.

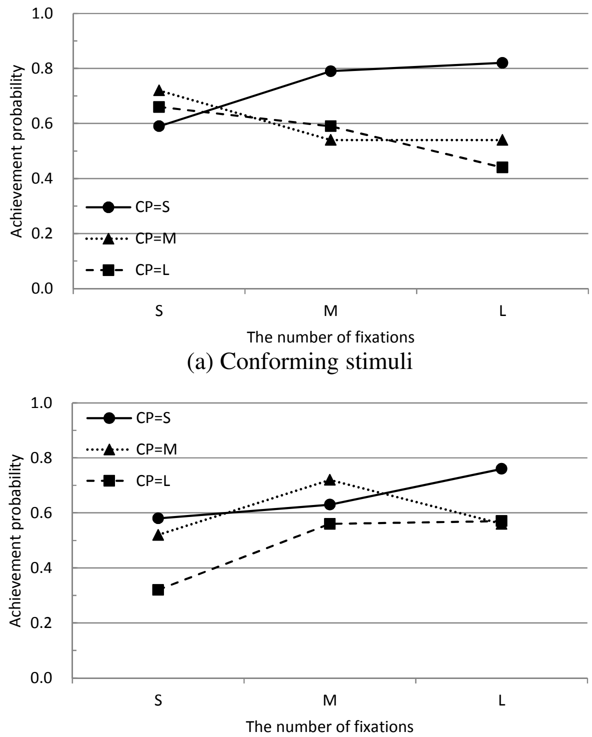
Table 6. Accuracy rates of all states and the states of achievement and non-achievement.

	All	Conforming	Non-conforming
All states	0.83	0.90	0.86
Achievement	0.89	0.97	0.85
Non-achievement	0.67	0.64	0.88

ence model (for example, see Chapter 3 of [9] for the evaluation of these values). An inferred result is deemed a correct answer if it proves compatible with the observation. The accuracy rate is defined as the division of the number of correct answers by that of all samples. Since, in this experiment, the rate of achievement for all stimuli rested at the high level of 0.73, it was possible to obtain a high accuracy rate (0.73) as long as one inferred that each sample achieved the objective. Hence, given the requirement that the accuracy rate of all states be greater than or equal to 0.73 and that of the state of non-achievement be greater than 0.66, an optical model was structured as a graph satisfying the requirement and possessing the smallest AIC. **Fig. 2** shows the optical model, in which the number of parameters was 37. The accuracy rates of all states and the achievement and non-achievement states are shown in the first row of **Table 6**. This table reveals that the CP time, duration of fixation, and number of fixations are important factors in inferring achievement because all accuracy rates are high.

4.2. Inference of Achievement Probabilities for Two Types of Stimuli

Under the graph structure of **Fig. 2**, we calculated conditional probabilities separately in the cases of conforming and non-conforming stimuli. The second and third rows in **Table 6** show the accuracy rates of conforming and non-conforming stimuli, respectively. Moreover,



(a) Conforming stimuli

(b) Non-conforming stimuli

Fig. 3. The relationship between achievement probability and the number of fixations.

Figs. 3(a) and **(b)** show the conditional probabilities of achievement given the number of fixations and CP time for both stimuli. The vertical axis indicates the inferred probability of achievement, and the horizontal axis denotes the number of fixations. Three types of lines correspond to the three category values of the CP time. These figures imply that in the case of conforming stimuli, the probability of achievement is high if the CP time is S and the number of fixations is M or L (with a total of 16 samples), and that in the case of non-conforming stimuli, the probability is also high if the CP time is S and the number of fixations is L (with a total of 6 samples). With this information, it is considered that in both cases of stimuli, if a visitor stayed on each CP briefly and looked at the information on all the pages (the number of fixations is large), then the possibility that they would achieve their objective increased. In other words, we may say that a prompt judgment of preference and thorough reading of contents brings an increased achievement probability. However, conforming stimuli have a greater effect on this increase

Table 7. Evaluation scores of the model for stimuli of each type.

The number of fixations	All	L	M	S
Conforming	0.62	0.72	0.57	0.60
Non-conforming	0.57	0.66	0.60	0.49

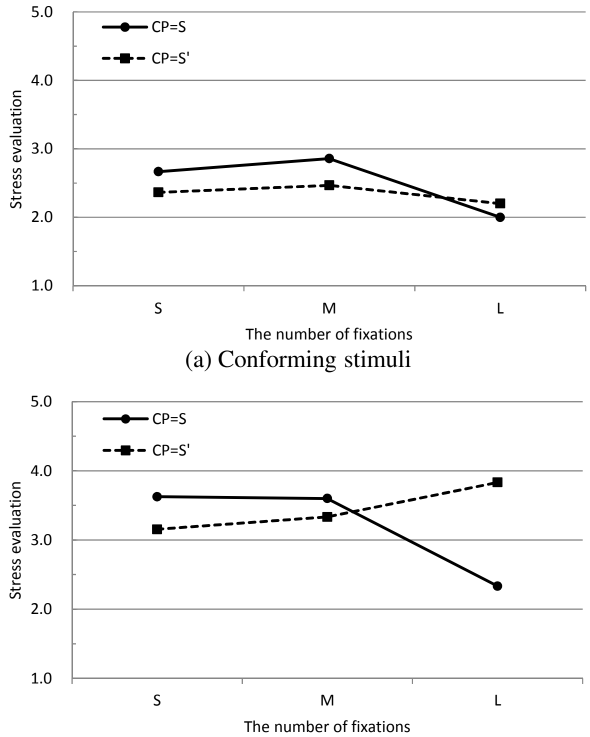
and enlarge the sufficient range of the number of fixations more than non-conforming stimuli do.

In a manner similar to the approach of Nomori et al. [7], we examined by *leave-one-out* cross validation [10] the predictive capability of the model for each of conforming and non-conforming stimuli. That is, the *leave-one-out* method was performed as follows: using a single subject's sample as validation data and the remaining samples as training data, the probability of the subject's actual state (achievement or non-achievement) was inferred, which was defined as an individual score for the validation data. The inference was repeated such that each subject's sample was used once as the validation data, and the average of all individual scores was calculated to obtain an "evaluation score." Thus, the evaluation score is interpreted as a probability value at which the model can infer the correct state arising from a subject's experiment. **Table 7** shows the evaluation scores for conforming and non-conforming stimuli. In addition, the table lists the average of individual scores across the samples classified under each category (S, M, L) of the number of fixations. Unfortunately, **Table 7** suggests that these models are not entirely accurate in terms of conforming and non-conforming stimuli, but comparatively capable of inference in the category of L. In the next subsection, we will consider reasons why the browsing pattern, as stated in the first paragraph, gives rise to a high achievement probability from a psychological viewpoint.

4.3. Effect of Stress and Interest on High Achievement Probabilities

In this subsection, the effects of the CP time and the number of fixations on stress and interest were tested by analysis of variance. In order to make simple effect analyses possible, the levels M and L of the CP time were integrated into a single level, denoted as S′. Thus, two factors were provided in which one has two and the other has three levels.

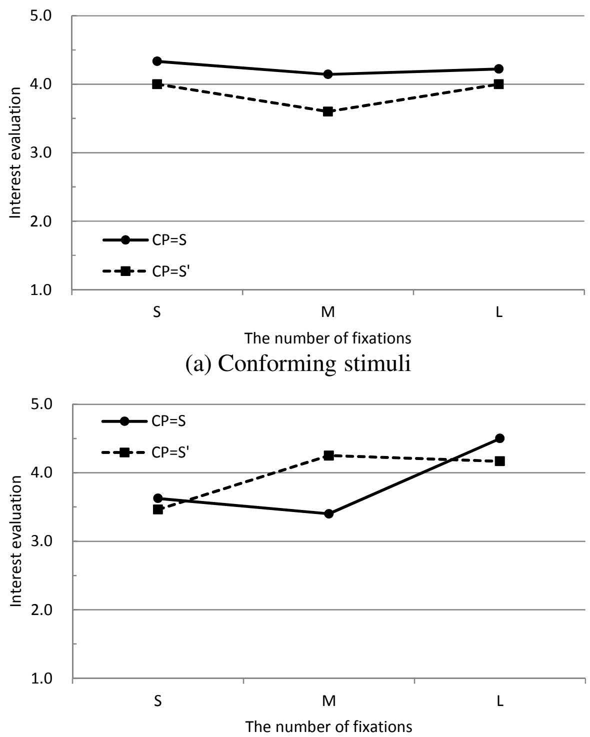
Figures 4(a) and **(b)** demonstrate stress evaluations given the number of fixations and CP time for conforming and non-conforming stimuli. **Figs. 5(a)** and **(b)** show the similar graphs of interest evaluations for both stimuli. In these figures, the vertical axis indicates the evaluation values, and the horizontal axis denotes the number of fixations. Two types of lines correspond to two levels, S and S′, of the CP time. From **Figs. 4(a)** and **5(a)**, it is seen that in conforming stimuli, the stress (also interest) evaluations with CP = S and CP = S′ were nearly the same over all levels of the number of fixations. Moreover, the stress evaluations were kept at a low level, whereas the interest



(a) Conforming stimuli

(b) Non-conforming stimuli

Fig. 4. The relation between stress evaluation and the number of fixations.



(a) Conforming stimuli

(b) Non-conforming stimuli

Fig. 5. The relation between interest evaluation and the number of fixations.

evaluations were kept at a high level. On the other hand, in non-conforming stimuli, as shown in **Figs. 4(b)** and **5(b)**, a couple of lines for CP = S and CP = S′ were different in the stress and interest evaluations. In particular, the stress evaluation of CP = S was obviously different from that of

$\mathrm{CP} = \mathrm{S}'$ when the number of fixations was L. An analysis of variance sustained these features entirely. Neither the factor effects nor interaction was significant in any of the cases, except the case of stress in non-conforming stimuli, in which case interaction was significant at a level of 5%. Also in non-conforming stimuli, when the number of fixations was L, the stress evaluations of $\mathrm{CP} = \mathrm{S}$ and $\mathrm{CP} = \mathrm{S}'$ were significantly different ($p = 0.021$). It should be emphasized that only if the CP time was S and the number of fixations was L, both the stress and interest evaluations for non-conforming stimuli were at the same level as conforming stimuli.

The following conclusions were obtainable from the information provided above. In the case of conforming stimuli, since stress generally remained at a low level and interest at a high level, subjects needed only to browse websites in the manner stated in the first paragraph of the previous subsection in order to achieve their objective. However, in the case of non-conforming stimuli, subjects had to follow this browsing method under the constraint that stress would be kept at a low level. This rendered the browsing method restrictive for non-conforming websites. Hence, it is considered that websites should be created in conformity with IUG.

5. Conclusion

This paper, focusing on an online shopping site, developed a Bayesian network model that infers the probability of purchase from web log and eye movement data. As a result, the following knowledge was obtained.

1. The visit time in CPs, number of fixations, and duration of fixation are necessary but not sufficient information for inferring the achievement probability.
2. In stimuli that both conform and do not conform to IUG, if the visit time in CPs is shortened and if the frequency of searching information is high, the achievement probability increases.
3. Conforming stimuli to IUG can keep visitors' stress at a low level throughout any values of the CP time and the frequency of searching information. On the other hand, stimuli not conforming to IUG cannot, preventing visitors from achieving their objectives.

Several topics remain for future research. First, we have to derive other items from web log and eye movement data to make the Bayesian network model more expressive. Furthermore, this experiment was conducted with subjects who had an existing interest in the products. For those visitors with little existing interest, it is important to gain an understanding of the properties of websites that will help them achieve their objectives.

Acknowledgements

We are grateful to the anonymous reviewers for invaluable comments and suggestions.

References:

[1] J. Nielsen, "Usability Engineering," Morgan Kaufmann, 1994.

[2] M. Khosravi and M. J. Tarokh, "Dynamic mining of users interest navigation patterns using naive Bayesian method," IEEE Int. Conf. on Intelligent Computer Communication and Processing, pp. 119-122, 2010.

[3] H. Adachi, I. Kuramoto, Y. Shibuya, and Y. Tsujino, "Finding users' target pages based on web access logs," SIG Technical Report, 2004-HI-112, No.9, pp. 35-42, 2005 (in Japanese).

[4] K. Toda et al., "An information exploration model based on eye movements during browsing web pages," SIG Technical Report, 2005-HI-113(6), No.52, pp. 35-42, 2005 (in Japanese).

[5] Q. Ji, P. Lan, and C. Looney, "A probabilistic framework for modeling and real-time monitoring human fatigue," IEEE Trans. Syst., Man, Cybern. Part A, Vol.36, No.5, pp. 862-875, 2006.

[6] M. Yuasa, Y. Yasumura, and K. Nitta, "A tool for animated agents in network-based negotiation," Proc. of 12th IEEE Int. Workshop on Robot and Human Interactive Communication. pp. 259-264, 2003.

[7] K. Nomori et al., "Developing a computational model to predict infant behavior elicited by environmental factors," The Japanese J. of Ergonomics, Vol.46, No.2, pp. 166-171, 2010 (in Japanese).

[8] T. Furuhashi, "Statistics, multivariate analysis, and soft computing: toward analysis of systems with ultra many degrees of freedom," Kyoritsu Publication, 2012 (in Japanese).

[9] K. Shigemasu, M. Ueno, and Y. Motomura, "Survey of Bayesian Networks," Baifu-kan, 2006 (in Japanese).

[10] M. Stone, "Cross-validation and multinomial prediction," Biometrica, Vol.61, No.3, pp. 509-515, 1974.

Name:
Yutaka Matsushita

Affiliation:
Department of Psychological Informatics, Kanazawa Institute of Technology

Address:
3-1 Yatsukaho, Hakusan, Ishikawa 924-0838, Japan

Brief Biographical History:

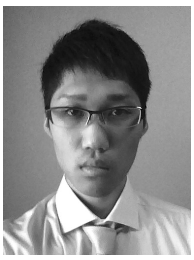
1985- Ohsaki Research Institute, Shimizu Corporation
2000- Chief Researcher, Institute of Technology, Shimizu Corporation
2005- Associate Professor, Kanazawa Institute of Technology

Main Works:

• "An additive representation on the product of complete, continuous extensive structures," Theory and Decision, Vol.69, pp. 1-16, 2010.

Membership in Academic Societies:

• Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
• Architectural Institute of Japan (AIJ)
• The Mathematical Society of Japan (MSJ)
• The Japanese Psychological Association



Name:
Syunsuke Maeda

Affiliation:
Asobimo Incorporated

Address:
5-26-19 Nishiikebukuro, Toshima-ku, Tokyo 171-0021, Japan

Brief Biographical History:

2010- Graduate Program in System Design Engineering, Kanazawa Institute of Technology
2012- Social Game Division, Asobimo Incorporated