Optimization of Motorcycle Riders Categorization Based on Emotion Using Decision Tree Analysis

Kodai Obata*　Masashi Sugimoto†　Noriko Nagata‡

Kwansei Gakuin University

ABSTRACT

In the present research, we optimized emotional rider categorization using decision tree analysis. First, we asked participants to evaluate four emotions (" enjoyable/brisk, " " pleasant/comfortable, " " boring/unsatisfying, " and " uneasy/scary ") toward 90 motorcycle pictures. Then, we reduced the amount of evaluation times (from two to seven evaluation). The optimized model succeeded in categorization with 85% accuracy. The optimized model also succeeded in replicating the emotion pattern of the clusters in the original model. The results indicate the high validity of the optimized model.

Index Terms: Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User models; Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies

1 INTRODUCTION

These days, the online advertisement is individuated. The individuated advertisement promotes purchasing based on user's web access history and purchase history. Another way to individuate users is based on sensitivity tendencies. For example, studies that categorized users by preference characteristics and sensitivity tendencies have been attracting attention [2, 3].

Sugimoto [3] conducted experiments to categorize motorcycle riders. Participants were instructed to imagine being on the motorcycle in the picture and rate how much they feel each of the four emotions; " enjoyable/brisk, " " pleasant/comfortable, " " boring/unsatisfying, " " uneasy/scary " . As a result, the participants were categorized into seven rider types. However, the visualization of sensitivity tendencies value required for the user categorization had a problem with involving a lot of time and effort [4].

The present paper aims to construct an optimized rider categorization model that requires much less cost in time, and effort.

2 METHOD

2.1 Participants

The 2,425 participants for this study (2374 males and 51 females, M = 45.28, SD = 10.52) completed an Internet-based survey. All participants are Japanese, and held licensure for regular or heavy motorcycle operations. Each had at least one motorcycle (any displacement).

2.2 Materials

2.2.1 Motorcycle pictures

We used the same motorcycle pictures as the previous study [3]. The pictures contained whole parts of a moving motorcycle and the rider of it, the road it runs on, and the environment it runs in.

2.2.2 Categorization data of riders through emotions evoked in motorcycle riding

We used data from the previous study [3]. The data are from 240 participants and contain four emotional evaluations of the 90 motorcycle pictures. The data also contain the clusters in which the participants are categorized: " standard riders (cluster 1), " " positive riders (cluster 2), " " cool riders (cluster 3), " " super-positive riders (cluster 4), " " own-paced riders (cluster 5), " " active riders (cluster 6), " and " aggressive riders (cluster 7). "

2.3 Procedure

First, participants provided their demographic information (e.g., age, sex and motorcycles they possess). Then they were instructed to imagine they were on the motorcycle in the picture and rate how much they feel each of the four emotions; " enjoyable/brisk, " " pleasant/comfortable, " " boring/unsatisfying, " " uneasy/scary. " They used five-point Likert evaluations (1: " It does not evoke the feeling at all " to 5: " It evokes the feeling a lot. "). They completed 360 emotion evaluations in total.

3 RESULT

3.1 Construction of the optimized model

3.1.1 Analysis target data

In the analysis, we used the data from the previous study (previous dataset) [3], and those we gathered from 2,425 participants (present dataset). Both datasets were composed of evaluation data of four kinds of emotion toward 90 motorcycle pictures. The dataset of the previous study also includes clusters of each participant, but the present dataset does not.

3.1.2 Result of analysis

In the dataset of the previous study, we conducted decision tree analysis whose target variable is seven clusters and whose predictor variable is 360 emotion evaluation data. In the analysis, we used statistic software " R " and its " rpart " package. Considering evaluation times and evaluation accuracy, we adopted a categorization model whose minimum data size in the leaf is three (Table 1). Table 3 shows the evaluation accuracy of the models. In the model, we regarded the category of each leaf as the cluster that contains the most number of participants. Using this analysis, we extracted if-then rules based on the evaluation scores to the picture. Then, we adopted these if-then rules to the present dataset and categorized participants into seven clusters: standard riders (cluster 1'),positive riders (cluster 2'), cool riders (cluster 3'),super-positive riders (cluster 4'),own-paced riders (cluster 5'), active riders (cluster 6'), and aggressive riders (cluster 7'). This rule allows us to categorize all participants at least two-times evaluation and at most seven-times evaluation.In addition, we represent the performance of the decision tree breakdown using sanky diagram, as well as the previous study which used sanky diagram to represent multi-generation transmission of individuals [1]. Every time participants evaluated the picture, they were divided into two subgroups (Fig. 1), and the proportion of the specific type increased (Fig. 2).

*e-mail: kodai-O@kwansei.ac.jp
†e-mail:sugimoto.masashi@kwansei.ac.jp
‡e-mail:nagata@kwansei.ac.jp

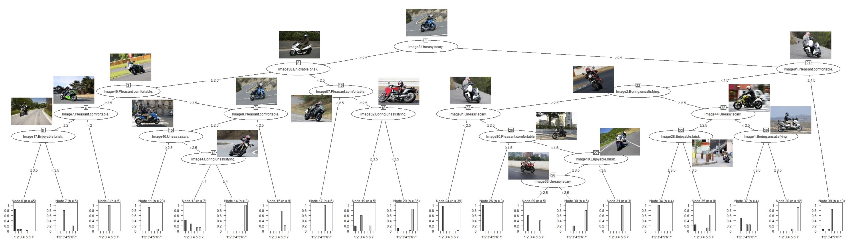

Figure 1: A decision tree in which the minimum data number of each leaf is 3.

Figure 2: The number of the participants in each node, and its breakdown. The broad path indicates larger number of the participants, and the thicker blue indicates the higher proportion of a specific type at the node.

Table 1: Categorization accuracy for minimum number of data in a leaf and its evaluation times.

Minimum number of Data in leaf	Categorization Accuracy								Evaluation times		
	Micro			Overall	Macro			Average			
	Precision	Recall	F-measure	Accuracy	Precision	Recall	F-measure	Accuracy	Max	Min	Average
7	0.75	0.75	0.75	0.75	0.76	0.74	0.75	0.74	5	2	3.67
6	0.76	0.76	0.76	0.76	0.77	0.75	0.76	0.75	5	2	3.67
5	0.79	0.79	0.79	0.79	0.80	0.79	0.79	0.79	5	2	4.14
4	0.81	0.81	0.81	0.81	0.82	0.81	0.81	0.81	5	2	4.26
3	0.85	0.85	0.85	0.85	0.87	0.86	0.86	0.86	7	2	4.53

3.2 Confirmation of the validation

3.2.1 Analysis target data

We targeted the dataset of the present research.

3.2.2 Results of Analysis

We categorized the dataset based on the if-then rules we made. Table 2 shows the average emotion evaluation scores in each cluster of the previous and the present study. Table 3 shows Euclidean distance between the scores of the present study (1'-7') and those of the previous study (1-7).

Table 2: Average emotional evaluation scores of each cluster.

Present					Previous				
	Enjoyable /brisk	Pleasant /comfortable	Boring /unsatisfying	Uneasy /scary		Enjoyable /brisk	Pleasant /comfortable	Boring /unsatisfying	Uneasy /scary
1'	3.95	3.92	2.39	2.95	1	3.98	3.95	2.27	2.95
2'	3.70	3.67	2.25	2.03	2	3.88	3.86	2.13	1.94
3'	3.42	3.37	2.61	2.87	3	3.35	3.33	3.09	3.17
4'	4.22	4.22	2.00	1.71	4	4.63	4.33	1.49	1.73
5'	2.89	2.88	2.88	2.94	5	2.74	2.69	2.37	3.07
6'	3.41	3.38	3.38	2.72	6	3.36	3.28	2.73	2.91
7'	3.39	3.37	2.74	1.96	7	3.23	3.24	3.05	1.77

4 Discussion

We constructed the optimized rider categorization model, which reduced 98-99.4% original evaluation times with 85% accuracy using decision tree analysis. This optimized model succeeded in categorizing participants into clusters whose emotion evaluation patterns are are similar with that of the original model. We calculated

Table 3: Euclidean distance between the scores of the clusters in the present model and those in the previous model.

clusters of present model	clusters of previous model						
	1	2	3	4	5	6	7
1'	**0.13**	1.05	1.11	1.71	1.72	0.94	1.67
2'	1.00	**0.29**	1.50	1.39	1.73	1.14	1.06
3'	0.88	1.25	0.57	2.22	1.01	**0.17**	1.21
4'	1.32	**0.56**	2.20	0.66	2.55	1.90	1.74
5'	1.64	1.87	0.72	2.92	**0.57**	0.64	1.28
6'	1.39	1.62	**0.55**	2.63	1.44	0.69	1.03
7'	1.38	0.93	1.27	2.02	1.51	0.96	**0.41**

Note: Bold=the closest, Underline=the second closest.

the Euclidean distances between clusters derived from the original model and clusters from the optimized model, based on their average emotional evaluation scores. They are the closest in standard riders (cluster 1 and 1'), positive riders (cluster 2 and 2'), own-paced riders (cluster 5 and 5'), and aggressive riders (cluster 7 and 7'). The distances were the second closest in cool riders (cluster 3 and 3'), super-positive riders (cluster 4 and 4'), and active riders (cluster 6 and 6'). Although there were minor discrepancies in these clusters, they are acceptable because the clusters were adjacent pairs in the previous data.

In the previous model [3], participants needed 360 evaluation times to be categorized. However, the present model could categorize them with 2 to 7 evaluation times with 85% accuracy. This suggests the high efficiency of the present model.

5 Conclusion

In the present research, we constructed a rider categorization model using decision tree analysis. We succeeded in reducing 98-99.4% of evaluation times and constructed a model requiring much less cost in time, and effort with 85% accuracy. In addition, the clusters in the optimized model showed the same emotion patterns with those in the original model. This indicates the high validity of our model.

References

[1] S. Fu, H. Dong, W. Cui, J. Zhao, and H. Qu. How do ancestral traits shape family trees over generations? *IEEE Transactions on Visualization and Computer Graphics*, 24(1):205–214, Jan 2018. doi: 10.1109/TVCG.2017.2744080

[2] G. Schuitema, J. Anable, S. Skippon, and N. Kinnear. The role of instrumental, hedonic and symbolic attributes in the intention to adopt electric vehicles. *Transportation Research Part A: Policy and Practice*, 48:39–49, 2013.

[3] M. Sugimoto, S. Imai, K. Katahira, Y. Yamazaki, N. Nagata, A. Masuda, K. Iwata, and H. Uchiyama. Indexing of riders emotion on motorcycle based on core-affect model -classification of riders through emotional reaction evaluation toward pictures-. vol. 117, pp. 123–126. The Institute of Electronics, Information and Communication Engineers, may 2017（in japanese with abstract）.

[4] A. Yamada, S. Hashimoto, and N. Nagata. A text mining approach for automatic modeling of kansei evaluation from review texts. In *KEER2018 INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH in press*.