
Construction of Models for Predicting Arousal Level in Advance based on Features of Face Images

Yuki Mekata and Miwa Nakanishi

Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, Japan

ABSTRACT

Sleepiness is a major factor in accidents or errors. Recently, a method for maintaining arousal level has been developed by conversing with the user using artificial intelligence agents. As a result, we expect to realize a system that predicts a decrease in arousal levels and works on it proactively. In order to realize such a system, we attempt to construct a model that can predict arousal levels in advance. We obtained data on face images and sleepiness assessment during the experiment, and we constructed models using machine learning methods. The discrepancy between the predicted and measured values is less than one in about 60% of the test data based on the result of the prediction model created by extracting features using deep learning and creating the model using a neural network. We think the model can help us figure out when interactions happen in a proactive system.

Keywords: Arousal level, Human-systems interaction, Artificial intelligence agents, Autonomous system

INTRODUCTION

Sleepiness is a major factor in accidents or errors. Even in these days of promoting system automation, users should monitor the state of the autonomous system and take over in the event of an emergency. Thus, it is important for systems to understand the user's arousal level and manage the user's arousal level appropriately. In recent years, a method for maintaining arousal levels has been proposed by using artificial intelligent agents to converse with the user (Mekata et al. 2019). The interaction method used by such a system is considered to be a more natural way to maintain arousal level than the conventional method of using stimulation such as a beep sound (Mekata et al. 2020). As a result, we anticipate realizing a system that not only detects sleepiness and responds to it reactively, but also predicts and responds to a decreasing arousal level in advance. In this study, we attempt to construct a model that predicts decreasing arousal levels in advance with the goal of achieving such a system. We think this will lead to the optimization of system interaction.

Sleepiness detection has already been studied, particularly for automobile drivers. In addition, the sleepiness prediction was tested in a few studies. Shikii et al. focused on the ambient temperature of drivers, and they discovered a

correlation between the drowsiness level after 10–15 minutes and the ambient temperature of drivers (Sikii et al. 2018). Using eyelid closure, gaze and head movement, and driving time, Jacobé de Naurois et al. studied predict when a driver's state may become impaired (Jacobé de Naurois et al. 2019). In the study, the model could predict when the driver's state would become impaired to within five minutes. Liang et al. studied night shift workers' electroencephalograms and electrooculography. They were able to predict sleepiness events within 10 minutes (Liang et al. 2019). In these studies, the focus is mainly on predicting whether a driver is in a dangerous state or not, and predicting the stages of arousal level remains a challenge. We would be able to create a more comfortable system if we could adapt the frequency and content of system interactions based on arousal level. On the other hand, in the assessment of arousal level, there is a method to assess a user's arousal level in five levels based on facial expression by trained raters (Wierwille & Ellsworth 1994; Kitajima et al. 1997). As a result, we assumed to construct a model predicting the five stages of arousal level in the method based on features of face images.

METHOD

In this study, we conducted experiments to obtain data on face images and subjective arousal level assessments, and then used this information to construct a model for predicting arousal level.

Experiment to Acquire Data for Model Construction

In the experiments to acquire data for model construction, autonomous driving was assumed as an example of an autonomous system. For an hour, participants monitored autonomous driving. Participants could only change lanes by pressing a button, and all other operations were handled by the autonomous system. During the task, participants responded to the arousal level in five levels by pressing a button. The participants were advised of the timing of answering the arousal level by a synthesized voice every 30 seconds, requesting that they answer the current sleepiness. Figure 1 shows an example of an experimental screen, whereas Figure 2 shows the experimental equipment.

During the monitoring tasks, we hypothesized that the variation in scenery affects the ease of changing the arousal level. Thus, in order to change the visual image, we varied the driving scenes depending on four categories: road, places, weather, and traffic. In the road category, the shape of the road (straight or curved), the number of lanes (two or three lanes), and the appearance of tunnels were set. In the places category, the surrounding landscape was changed to set up a road surrounded by mountains or buildings. In the weather category, in addition to the normal sunny weather, rainy and foggy weather were set. In the traffic category, the number and behavior of other vehicles in the vicinity were changed to generate the merging of other vehicles from the side lines and the occurrence of traffic jams. We are set up to generate a variety of driving settings by separately altering each of these four categories. Figure 3 shows examples of experimental screens in which the elements of each category are changed in different ways.



Figure 1: An example of experimental screen.

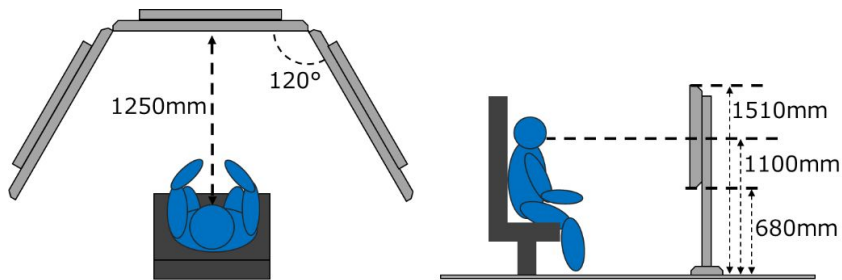


Figure 2: Experimental setting.





<p>Road</p> <ul style="list-style-type: none"> • Curved road • Appearance of tunnel • Number of lanes 	<p>e.g.) Curved road</p> 
<p>Places</p> <ul style="list-style-type: none"> • Surrounded by mountains • Surrounded by buildings 	<p>e.g.) Surrounded by mountains</p> 
<p>Weather</p> <ul style="list-style-type: none"> • Foggy weather • Rainy weather 	<p>e.g.) Foggy weather</p> 
<p>Traffic</p> <ul style="list-style-type: none"> • Occurrence of traffic jam • Merging from the side lane 	<p>e.g.) Occurrence of traffic jam</p> 

Figure 3: Examples of experimental screens when the elements of each category are changed.

The experiment included three male and three female participants, and each participant completed the task three times. Each participant worked on the task only once a day, and the repetition was done on different dates. We measured the participant’s face images at 60 Hz using EMR-ACTUS (nac image technology, Inc.), an eye tracking device placed in front of the participant. In addition, to account for the possibility of model construction based

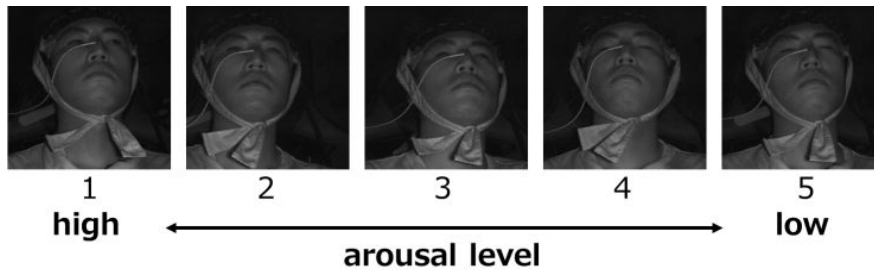


Figure 4: Examples of face images at each arousal level.

not only on face images but also on other features, the physiological indices used to evaluate arousal level were also recorded in this experiment, and the devices were affixed to the participants' faces. Figure 4 shows examples of face images at the time of answering each arousal level. This research was conducted with the permission of the Keio University Faculty of Science and Technology's Research Ethics Review Committee.

Features Extracted from Face Images

For model construction, we used features from a texture distribution generated using Local Binary Patterns Histogram (LBPH) (Ojala et al. 1994; Ojala et al. 2002) and features obtained by embedding using FaceNet (Schroff et al. 2015) for face images. LBPH is a method of describing an image's brightness information in ten dimensions by creating a histogram of how much of the image has patterns extracted by comparing a pixel with its neighbors. FaceNet is a deep learning architecture-based method for representing images in 128 dimensions by embedding them in Euclidean distance space. We constructed models by learning the features of the images obtained by these methods.

Model Construction

As well-known major machine learning methods, we used the neural network, random forest, and support vector machine to construct models. We constructed a model that uses the features extracted from the face images as input and outputs the values of the five-step arousal level assessment answered during the task. The target time to be predicted was changed by changing the value of arousal level used as output based on the number of seconds ahead of input face image. The target time was varied at 30 second intervals, and examples ranging from 30 to 180 seconds were examined. During the learning phase, the parameters of machine learning methods were optimized for each model using grid search method. Data from two trials was used as training data, and data from the remaining trial was used as test data for each participant. We constructed models for three patterns for each participant's three trials, with each trial serving as test data. Furthermore, to reduce the effect of data bias in the arousal levels, we resampled the data for each arousal level to match the training data's lowest number of levels.

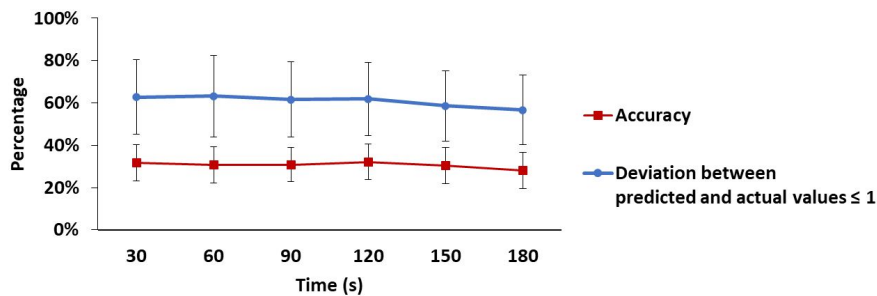


Figure 5: The mean values of the accuracy and the rate of the case where the deviation between predicted and actual values is less than one by a model using a neural network based on features of the embedding by FaceNet.

RESULTS

For model evaluation, the accuracy of the model was calculated by averaging the results of the three models for each participant. In the cases of using random forest and support vector machine, only particular labels were output in many circumstances. On the other hand, the bias of the output value was less when using a neural network than when using other methods. For this reason, we use a neural network to investigate the model's accuracy. In addition, although the overall accuracy was about the same for features based on LBPH and features based on embedding by FaceNet, there was a higher bias in the output value in the case of LBPH than in the case of embedding by FaceNet. Figure 5 shows the results of predicting each participant's model using a neural network based on FaceNet embedding features. For learning individual data, the accuracy of the prediction arousal level after 30 seconds is around 35%. As the prediction target time gets further away from the present time, the accuracy of the prediction target time decreases.

In the five-level prediction, the deviation between the predicted and actual values is less than one in around 60% of the test data, indicating that substantial deviations can be suppressed. The rate of small deviations decreases as the target time for prediction moves further away from the present time, just as it does with accuracy. Although the accuracy gradually decreases, the results up to 120 seconds showed that the accuracy was greater than 30% and the rate of small deviation was greater than 60%. From this result, we think that the model we constructed has some success in setting the timing of interactions in a proactive system, since a two-minute period is considered to be sufficient for conducting dialogue and other approaches.

CONCLUSION

In this study, we constructed a model based on face images to predict an arousal level in advance, with the goal of optimizing interactions with systems. We obtained a model that predicts the arousal level up to 120 seconds later to be less than 1 deviation from the actual value with roughly 60% of the test data by extracting features using deep learning and creating the model using a neural network. While more accuracy is still desired, the model that was

constructed is expected to be used in a proactive system that predicts and responds to decreases in arousal levels in advance.

ACKNOWLEDGMENT

This study is granted by The Keio University Doctorate Student Grant-in-Aid Program from Ushioda Memorial Fund.

REFERENCES

- Jacobé de Naurois, C., Bourdin, C., et al. (2019), "Detection and prediction of driver drowsiness using artificial neural network models." *Accident Analysis and Prevention*, 126, 95–104.
- Kitajima, H., Numata, N., et al. (1997), "Prediction of automobile driver sleepiness: 1st report, rating of sleepiness based on facial expression and examination of effective predictor indexes of sleepiness." *Transactions of the Japan Society of Mechanical Engineers Series C*, 63(613), 3059–3066.
- Liang, Y., Horrey, W. J., et al. (2019), "Prediction of drowsiness events in night shift workers during morning driving." *Accident Analysis and Prevention*, 126, 105–114.
- Mekata, Y., Takeuchi, S., et al. (2019), "Proposal of a method for maintaining arousal by inducing intrinsic motivation and verification of the effect." *Transactions of Society of Automotive Engineers of Japan*, 50(4), 1138–1144.
- Mekata, Y., Hirose, F., et al. (2020), "Comparison of the effects of maintaining arousal level by inducing intrinsic motivation and existing sleepiness countermeasures." *2020 JSAE Annual Congress (Spring) Proceedings*, 26–20.
- Ojala, T., Pertikainen, M., et al. (1994), "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions." *Proceedings of 12th International Conference on Pattern Recognition*, 582–585.
- Ojala, T., Pietikäinen, M., et al. (2002), "Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971–987.
- Schroff, F., Kalenichenko, D., Philbin, J. (2015), "FaceNet: A unified embedding for face recognition and Clustering." arXiv preprint arXiv:1503.03832.
- Sikii, S., Sunagawa, M., et al. (2018), "Driver monitoring system based on drowsiness detection/prediction technology." *Panasonic Technical Journal*, 64(2), 69–74.
- Wierwille, W. W., Ellsworth, L. A. (1994), "Evaluation of driver drowsiness by trained raters." *Accident Analysis and Prevention*, 26(5), 571–581.