Detecting Cognitive Workload Using Driving Performance and Eye Movement in a Driving Simulator

Joonwoo Son, Myoungouk Park and Hosang Oh
HumanLAB, Daegu Gyeongbuk Institute of Science & Technology

50-1 Sang-Ri, Hyeonpung-Myeon, Dalseong-Gun
Daegu, 711-813, KOREA
Phone: +82-53-785-4740
Fax: +82-53-785-4759
E-mail: json@dgist.ac.kr

The aim of this study was to develop cognitive workload estimation algorithm using driving performance and eye movement data. The algorithm adopts radial basis probabilistic neural networks (RBPNN) to construct cognitive load estimation models. In order to train and test the models, recordings of driver's gaze and driving performance were captured in a driving simulator during three levels of cognitive demand. As a result, it was found that the proposed RBPNN models were able to differentiate driver's high cognitive workload from the normal driving with high accuracy. The best performance was achieved with a combination of standard deviation of lane position (SDLP) and gaze dispersion of X and Y coordinates over 30 seconds time window. The highest cognitive workload detection accuracy rate in overall model performance was 85.0%.

Topics / Driver-vehicle Systems & Driving Simulator

1. INTRODUCTION

The growing introduction of new technologies inside vehicles generates additional information that drivers have to manage at the same time. Their use can interfere with the driving activity and induce performance decrements. According to car accident statistics, between 13% and 50% of crashes are caused by driver distraction [1]. Thus, identification of a driver's workload and spare capacity is crucial in the design of adaptive automotive user interface [2]. By monitoring driver's workload, the adaptive interface system can provide timely and affordable information when the driver has the spare capacity.

Workload can be measured in a variety of ways including: subjective measures, driving performance based measures, physiological measures, eye movement measures and so on [3-6]. Among those measures, driving performance measures can detect the cognitive workload using easy and less expensive methods through readily available in-vehicle information [7-8]. However, driving performance measures are known to have limitations compared to others due to small changes according to the cognitive workload. That is, the performance measures are able to detect a high cognitive workload condition, but their classification accuracy is not enough to distinguish graded levels of cognitive difficulty. On the other hand, physiological measures have been proposed as useful metrics for assessing workload. Mehler et al. found that a near linear increase in heart rate and skin conductance appeared across the three levels of task difficulty. In the context of vision, the level of cognitive workload has been related to measures such as pupil diameter; blink rates, decreased amplitude of saccadic movements, longer dwell times, and the size of the visual field. Measuring changes in workload by using indices such as blink rates and pupil diameter is difficult in the rapidly changing conditions of an automotive environment. Therefore, measuring changes in the size of the visual field is one of the more easily interpreted alternatives [6].

Thus, this paper suggested a neural network algorithm for estimating driver's cognitive workload using driving performance and eye movement data. The results show that the combination of driving performance and eye movement, or eye movement data can effectively distinguish high cognitive workload condition from the normal driving with the high accuracy rate.

2. MODEL CONSTRUCTION

2.1. Data Source

In order to construct neural network models for estimating the difficulty of cognitive workload, driving experimental data was collected as follows:

2.1.1. Experimental setup

The experiment was conducted in a fixed-based driving simulator, which incorporated STISIM Drive™ software and a fixed car cab (see Figure 1). The virtual roadway was displayed on a 2.5m by 2.5m wall-mounted screen at a resolution of 1024 x 768.

Sensory feedback to the driver was also provided through auditory and kinetic channels. Distance, speed, steering, throttle, and braking inputs were captured at a nominal sampling rate of 30 Hz. Eye behavior data were collected using the FaceLAB® 4.6 eye tracking system (Seeing Machines Ltd., Canberra, Australia), respectively. A display was installed on the screen beside the rear-view mirror to provide information about the elapsed time and the distance remaining in the drive.

Fig. 1 Fixed-based Driving Simulator

2.1.2. Subject

Subjects were required to meet the following criteria: age between 25-35, drive on average more than twice a week, be in self-reported good health and free from major medical conditions, not take medications for psychiatric disorders, score 25 or greater on the mini mental status exam to establish reasonable cognitive capacity and situational awareness, and have not previously participated in a simulated driving study. The sample consisted of 15 males, who are in the 25-35 age range (M=27.9, SD=3.13).

2.1.3. Cognitive workload

An auditory delayed digit recall task was used to create periods of cognitive demand at three distinct levels. This form of n-back task requires participants to say out loud the nth stimulus back in a sequence that is presented via audio recording [9]. The lowest level n-back task is the 0-back where the participant is to immediately repeat out loud the last item presented. At the moderate level (1-back), the next-to-last stimuli is to be repeated. At the most difficult level (2-back), the second-to-the-last stimulus is to be repeated. The n-back was administered as a series of 30 second trials consisting of 10 single digit numbers (0-9) presented in a randomized order at an inter-stimulus interval of 2.1 seconds. Each task period consisted of a set of four trials at a defined level of difficulty resulting in demand periods that were each two minutes long.

2.1.4. Procedure

As shown in Fig. 2, Following informed consent and completion of a pre-experimental questionnaire, participants received 10 minutes of driving experience and adaptation time in the simulator. The simulation was then stopped and participants were trained in the n-back task while remaining seated in the vehicle. N-back training continued until participants met minimum performance criteria. Performance on the n-back was subsequently assessed at each of the three demand levels with 2 minute breaks between each level. When the simulation was resumed, participants drove in good weather through 37km of straight highway. Minutes 5 through 7 were used as a single task driving reference (baseline). Thirty seconds later, 18 seconds of

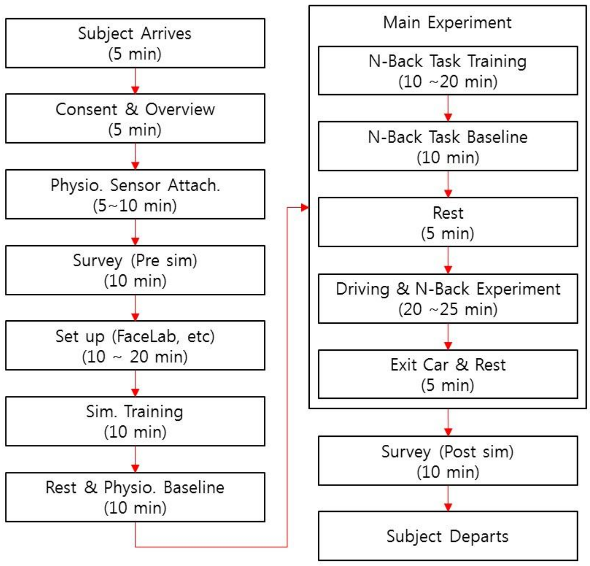

Fig. 2 Experimental Procedure

instructions introduced the task (0, 1 or 2-back). Each n-back period was 2 minutes in duration (four 30 second trials). Two-minute rest/recovery periods were provided before presenting instructions for the next task. Presentation order of the three levels of task difficulty was randomized across participants.

2.2. Model Characteristics and Training

2.2.1. Definition of cognitive workload

Although the cognitive workload was applied with three levels of complexity, the present study classified cognitive workload into two categories, i.e. normal driving and high cognitive workload condition. The normal driving means the situation of driving without cognitive workload and the high cognitive workload was defined as the durations of performing the highest difficult level of the cognitive task, so called 2-back task

2.2.2. Input features

Two driving performance measures, the standard deviation of lane position (SDLP) and steering wheel reversal rate (SRR), and two eye movement data, the standard deviation of horizontal gaze (Gaze X) and the standard deviation of vertical gaze (Gaze Y), were considered as input features to detect high cognitive workload demand in the RBPNN models.

SDLP was calculated from 0.1 Hz high pass filtered lateral position data with removing lane changes using the AIDE project guidelines. SRR was calculated by counting the number of steering wheel reversal from the 2Hz low pass filtered steering wheel angle data per minute. For cognitive workload, the reversal angles, which have more than 0.1 degree of the gap size, were counted.

Before calculating eye-movement measures, raw gaze data were filtered with the following criteria that was suggested by earlier study [10]: 1) the FaceLAB's automated gaze quality index for the left and right eyes was categorized as optimal, 2) the x-axis position was between -1.5m and +1.5m, the y-axis position was between -1.0m and +1.0m, and 3) the data point was contained within a set of six valid measurements (approximately 100ms). With the filtered data, the S.D. of vertical and horizontal gaze were calculated.

2.2.3. Summarizing parameters

In this paper, window size was considered as the summarizing parameter for the inputs. Window size denotes the period over which performance and eye movement data were averaged. The comparisons of window size could identify the appropriate length of data that can be summarized to reduce the noise of the input data without losing useful information. This paper considered three window sizes: 10, 20 and 30 seconds.

2.2.4. Model construction

Radial basis probabilistic neural networks (RBPNN) were used to construct the driver's cognitive workload estimation models. For training and testing RBPNN models, data of four task periods, which consist of a single task (driving only condition) and three dual tasks (n-back task condition), were used. A task was divided into multiple segments based on window size. In each task, half of the segments were used for training and the other segments were used for testing. Model performance was evaluated with testing accuracy, which is the ratio of the number of instances correctly identified by the model to the total number of instances in the testing set.

2.2.5. Model training and testing

Radial basis probabilistic neural networks (RBPNN) were used to construct the driver's cognitive workload estimation models. In this paper, the models were trained using the NEWPNN function in MATLAB. For training and testing RBPNN models, data of two task periods among four task levels, which consist of a single task (driving only condition) and the most difficult tasks (2-back task condition), were used. A task was divided into multiple segments based on window size. For example, if the model uses 30s window, one task period divided into four segments as shown in Figure 3. In the same manner, 20s window set has six segments and 10s window set has twelve. In each task, half of the segments, i.e. two segments per subject in 30s window, were used for training and the other segments were used for testing. Thus, each neural net was trained and tested using different sets of measurements, i.e. 15 (subjects) x 2 (levels) x 2 (segments), 15x3x2 and 15x6 examples for 30s, 20s and 10s window, respectively. Since the estimator is always evaluated on the data disjoint from the training data, the performance evaluated through the cross validation scheme correctly reflects the actual generalization capability of the derived estimator [7]. Model performance was evaluated with testing accuracy, which is the ratio of the number of instances correctly identified by the model to the total number of instances in the testing set.

3. RESULTS AND DISCUSSION

The performance of the RBPNN models varies from the combined input features and window sizes. Among different combinations of inputs, i.e. SDLP, SRR, Gaze X and Gaze Y, the performance using a multiple domain combination of driving performance, i.e., SDLP and eye behavior, i.e., Gaze X & Y, and eye movement domain only, i.e. Gaze X & Y outperformed as shown in Table 1.

Due to the fact that eye movement had higher concentration of gaze dispersion with the levels of cognitive load complexity and lower SDLP was observed under cognitive workload condition, the best performance appeared when the models have SDPL, Gaze X and Gaze Y as an input feature. The best performing model, which uses SDLP, Gaze X and Gaze Y data over a 30s-window, could detect the highest level of cognitive workload with an average accuracy of 85.0%. With this model, the estimation accuracy rate of driving only criteria, i.e. no cognitive workload condition, was 73.3%, and the accuracy of the most difficult cognitive load estimation was 96.7%. It should be noted that the proposed model outperforms detecting the highest cognitive workload that must be detected correctly with very high accuracy.

The results demonstrated that the model using SDLP and Gaze X & Y was outperforming than the other combinations among performance and eye behavior measures. The main contributor of the high

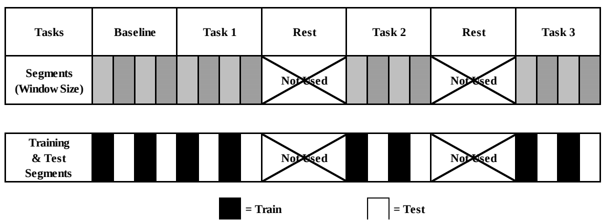

Fig. 3 Allocation of Segments to Training and Testing Sets

accuracy rate in this model was Gaze X, which provides clear changes associated with difficult level of cognitive workload, but relatively lower threshold to distinguish higher mental workload. According to Reimer et al., an

Table 1 Model performance with different window size

		Gaze			**Driving & Gaze (Combination)**								
					SDLP & Gaze			**SRR & Gaze**			**SDLP & SRR & Gaze**		
		X	**Y**	**X & Y**	**SDLP Gaze X**	**SDLP Gaze Y**	**SDLP X & Y**	**SRR Gaze X**	**SRR Gaze Y**	**SRR X & Y**	**SDLP SRR Gaze X**	**SDLP SRR Gaze Y**	**All**
10s	Normal	43.3	78.9	66.7	43.3	78.9	61.1	66.7	81.1	71.1	66.7	82.2	72.2
	High Load	94.4	56.7	73.3	95.6	57.8	78.9	87.8	72.2	84.4	87.8	71.1	86.7
	Average	68.9	67.8	70.0	69.4	68.3	70.0	77.2	76.7	77.8	77.2	76.7	79.4
20s	Normal	68.9	80.0	80.0	66.7	80.0	80.0	68.9	68.9	68.9	68.9	66.7	68.9
	High Load	86.7	57.8	77.8	86.7	57.8	84.4	73.3	66.7	71.1	73.3	73.3	75.6
	Average	77.8	68.9	78.9	76.7	68.9	82.2	71.1	67.8	70.0	71.7	70.0	72.2
30s	Normal	56.7	83.3	73.3	56.7	83.3	73.3	76.7	73.3	73.3	76.7	76.7	73.3
	High Load	93.3	70.0	93.3	93.3	70.0	96.7	70.0	53.3	60.0	70.0	53.3	60.0
	Average	75.0	76.7	83.3	75.0	76.7	85.0	73.3	63.3	66.7	73.3	65.0	66.7

effect of task performance on Gaze X concentration such that better task performance was associated with less gaze constriction [10]. Thus, SDLP and Gaze X&Y based model provides better performance to detect higher levels of mental demand.

4. CONCLUSION

In this paper, we proposed an algorithm for estimating driver's cognitive workload using driving performance and physiological data. Especially, SDLP and SRR, and SD of Gaze X and SD of Gaze Y were considered as cognitive load indices for the driving performance and eye behavior, respectively. In order to collect driving data, participants drove through highway in a driving simulator and were asked to complete three different levels of auditory recall tasks. The driver's cognitive workload estimation algorithm was developed using RBPNN models that were implemented by MATLAB NEWPNN function.

The results show that the proposed Gaze X&Y and SDLP-based RBPNN models were able to detect the most difficult cognitive workload with high accuracy. The model performance was assessed with the cross-validation scheme, which is widely adopted by the machine learning community. As a result, the highest workload estimation accuracy rate in overall model performance was 85.0%. And it is also expected that the accuracy can be improved by applying more sophisticated algorithms.

REFERENCES

[1] Stutts, J. C. and Hunter, W. W. "Driver inattention, driver distraction and traffic crashes", ITE Journal, Vol. 73, No. 7, 2003, pp. 34-45.

[2] Son, J., Lee, Y., and Kim, M. "Impact of traffic environment and cognitive workload on older drivers' behavior in simulated driving", Int. J. Precis. Eng. Manuf., Vol. 12, No. 1, 2011, pp. 135-141.

[3] Angell, L., Auflick, J., Auustria, P. A., Kochhar, D., Tijerina, L., Biever, W., Diptiman, T., Hogsett, J., and Kiger, S. "Driver Workload Metrics Task 2 Final Report", Nat. Highway Traffic Safety Admin., U.S. Dept. Transp., 2006, DOT HS 810 635.

[4] Mehler, B., Reimer, B., Coughlin, J. F., and Dusek, J. A. "Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers", TRR Journal, Vol. 2138, No. 1, 2009, pp. 6-12.

[5] Son, J., Mehler, B., Lee, T., Park, Y., Coughlim, J., and Reimer, B. "Impact of cognitive workload on physiological arousal and performance in younger and older drivers", Proc. of the DA 2011 Conference, 2011, pp. 87-94.

[6] Reimer, B. "Impact of Cognitive Task Complexity on Drivers' Visual Tunneling", Journal of the TRB of the National Academies., Vol. 2138, 2009, pp. 13-19.

[7] Zhang, Y., Owechko, Y., and Zhang, J. "Driver cognitive workload estimation: A data-driven perspective", Proc. of the ITSC2004, 2004, pp. 642-647.

[8] Son, J. and Park, S. W. "Cognitive Workload Estimation through Lateral Driving Performance", Proc. of the APAC16, 2011, SAE2011-28-0039.

[9] Son, J., Reimer, B., Mehler, B., Pohlmeyer, A. E., Godfrey, K.M., Orszulak, J., Long, J., Kim, M. H., Lee, Y. T., and Coughlin, J. F. "Age and cross-cultural comparison of drivers' cognitive workload and performance in simulated urban driving", International Journal of Automotive Technology, Vol. 11, No. 4, pp 533-539.

[10] Reimer, B., Mehler, B., Wang, Y., and Coughlin, J. F. "A Field Study on The Impact of Variations in Short Term Memory Demands on Drivers' Visual Attention and Driving Performance Across Three Age Groups", Human Factors, Vol. 54, No. 3, 2012, pp. 454-468.