2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

EOG-based Drowsiness Detection Using Convolutional Neural Networks

Xuemin Zhu, Wei-Long Zheng, Bao-Liang Lu*
Department of Computer Science and Engineering
Key Laboratory of Shanghai Education Commission
for Intelligent Interaction and Cognitive Engineering
Shanghai Jiao Tong University, Shanghai, China

Xiaoping Chen*, Shanguang Chen, and Chunhui Wang
National Key Laboratory of Human Factors Engineering
China Astronaut Research and Training Center
Beijing, China

***Abstract*—This study provides a new application of convolutional neural networks for drowsiness detection based on electrooculography (EOG) signals. Drowsiness is charged to be one of the major causes of traffic accidents. Such application is helpful to reduce losses of casualty and property. Most attempts at drowsiness detection based on EOG involve a feature extraction step, which is accounted as time-consuming task, and it is difficult to extract effective features. In this paper, an unsupervised learning is proposed to estimate driver fatigue based on EOG. A convolutional neural network with a linear regression layer is applied to EOG signals in order to avoid using of manual features. With a postprocessing step of linear dynamic system (LDS), we are able to capture the physiological status shifting. The performance of the proposed model is evaluated by the correlation coefficients between the final outputs and the local error rates of the subjects. Compared with the results of a manual ad-hoc feature extraction approach, our method is proven to be effective for drowsiness detection.**

I. Introduction

Drowsiness refers to the state of near-sleep, a strong desire for sleep, or sleeping for unusually long periods. It is understood that a person in special situations, such as driving a car, operating a machine, needs to remain alert. Or if not, serious casualty may occur. According to the resent study performed in 2006 by the National Highway Traffic Safety Administration (NHTSA) and Virginia Tech Transportation Institute (VTTI) [1], nearly 80 percent of crashes and 65 percent of near-crashes involved some form of driver inattention. This study recognizes driving fatigue as one of the major cases of traffic accidents in the US. Thus, considering the various application scenarios, effective drowsiness detection model is in urgent need and has broad application prospects as well.

Various methods have been proposed to detect drowsiness, which can be divided into video based, multi-sensor based and physiological signal based. Among the video based methods, Dinges *et al.* proposed a method using the percentage of eyelid closure over the pupil over time (PERCLOS) [2], which turns out to be a valid psychophysiological measure of alertness. However, video based methods are sensitive to illumination changes and consequently fail to capture the driver's eyes. Friedrichs *et al.* utilized steering and lane data to monitor drowsiness [3][4]. While both video based methods and multi-sensor based approaches fail to predict driver fatigue in advance, this implicitly signifies physiological signals based methods would be a better choice.

There are various researches on sleep based on physiological signals. In 1968, Rechtschaffen and Kales divided sleep into five stages (S1, S2, S3, S4, REM), plus the state of wakefulness (W), according to the features extracted from electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG) [5]. In our previous work, Shi *et al.* proposed various drowsiness detection models based on EEG signals [6][7][8][9]. Their results have shown that EEG based methods can correctly discriminate between wakefulness and sleepiness. Therefore, EEG based methods are recognized as golden standard for drowsiness detection. Despite the high accuracy of EEG based models for characterizing the drowsiness, EEG signal is regarded to be liable to be affected by noise and difficult to collect. Thus, by reasons of EOG signal's easiness to collect and immunization to slight noise, EOG based methods are considered as a compromise between accuracy and facility. Ma *et al.* used EOG features, mainly slow eye movements (SEM), to estimate vigilance changes during a monotonous task [10]. Feature extraction and feature selection are the key stage in such process. Following their work, we enhance EOG-based method by incorporating recent advances in machine learning with deep learning approaches.

Convolutional neural network (CNN) has been widely used in computer vision. Krizhevsky *et al.* trained a large deep convolutional neural network to classify images in the ImageNet LSVRC-2010 contest into 1000 different classes and achieved fantastic results [11]. Martinez *et al.* successfully built affect models based on skin conductance and blood volume pulse [12]. And Cecotti *et al.* showed that CNN is applicable to processing EEG signals [13]. In this paper, we followed their ideas and applied the deep learning approach to drowsiness detection.

This paper focuses on developing regression models of drowsiness detection by incorporating CNN based on electrooculography. We emphasize on building an unsupervised feature learning model for drowsiness detection as opposed to manual ad-hoc feature extraction process. We employ the experiments described in [10], collect EOG signals of 22 participants, and preprocess the signals with noise removing and band pass filter. Then the raw EOG signals are fed up to the convolutional neural network. The weight matrices in each layer of convolutional neural network are trained in a layer-wise greedy fashion with stacked convolutional auto-encoder

*Corresponding authors: B. L. Lu (bllu@sjtu.edu.cn) and X. P. Chen (xpchen2009@163.com)

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

[14], using no prior knowledge. Such unsupervised training method has been proved to reach awesome results in areas such as object recognition [15] and speech recognition [16]. Once the pre-training is finished, the local error rate is added to train the last regression layer. The final regression results are obtained by integrating a linear dynamic system to smooth the results and improve the performance by capturing a more reasonable vigilance state switching, and eliminating excessive or unlikely state transitions. Comparison between CNN and traditional manual feature extraction shows that CNN yields models of significantly higher correlation coefficients.

The rest of the paper consists of four parts. We first give an overview of the state-of-the-art drowsiness detection model, which is used for comparison, and then details our methodology. In the third part, experiments setups and label acquisition are described. Finally, we present our results and analysis.

II. Drowsiness detection model

Drowsiness reflects human's mental and bodily states. The state-of-the-art methods mainly contain five steps: preprocessing, feature extraction, feature selection, feature processing and regression (classifier) training and prediction. Part of the complexity of drowsiness detection model lies in designing the proper features or feature combination. Although this paper is only concerned with the deep artificial network model, this section gives an overview of the state-of-the-art method we applied for comparison.

A. *Preprocessing*

Signals from electrodes are down sampled to 125 Hz at first. With the signals from four electrodes, we can easily obtain two EOG channels, the horizontal and the vertical, by substraction between the electrodes of the same color in Fig. 1. Then signals are filtered with a band pass between 0.3 Hz and 10 Hz. Afterwards, mean value is subtracted. In the final step, signals are saturated at a saturation constant and scaled to 0 to 1. When the preprocessing is finished, noise is eliminated and signals are normalized.

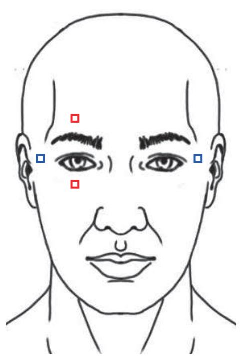


Fig. 1. Placement of electrodes. Squares represents electrodes. The two channel EOG signals, horizontal and vertical, are obtained by differences from the same color electrodes.

B. *Manual ad-hoc feature extraction*

The existing manual ad-hoc features extracted from EOG mainly contain SEM features, blink features and energy features. In our previous work, Ma *et al.* [10] and Wei *et al.* [17] employed various extraction techniques to get the EOG features. Their extraction procedure mainly contains three steps, eye movement detection, feature extraction, and feature processing. As is shown in [10][18], slow eye movement and blink are the most valuable eye movements correlated to fatigue. We use features from these two eye movements and energy to detect drowsiness.

Ma *et al.* applied the automatic detection technique of SEM developed by Magosso and colleagues [19][20][21]. This method is based on discrete wavelet transform (DWT) and includes three steps: wavelet decomposition, energy computation and discriminant function. Blink detection algorithm is an improved version of the double thresholds method. Two thresholds which represent the eyelid's closing speed and opening speed, respectively, are utilized to locate blink waveforms on the differences of vertical EOG.

Once the eye movements are detected, feature extraction is conducted. Table I is a list of features we extracted for comparison. Since the energy of different frequency bands in the EOG can implicitly express the intensity of different kinds of eye movements, we use wavelet transformation to extract the ratio of low and high frequency energy on both horizontal and vertical EOGs. We extracted features from EOG with a time window of 8 seconds, and the detailed description of features is shown in Table I.

TABLE I. List of features extracted from EOG.

feature	description
SEM proportion	SEM number
Closing time	duration of closing phase of blink
Closing PVe	peak velocity in closing phase of blink
Opening PVe	peak velocity in opening phase of blink
Closing MVe	mean velocity in closing phase of blink
Opening MVe	mean velocity in opening phase of blink
HEO LF/HF	PSD ratio between low and high frequency on horizontal channel
VEO LF/HF	PSD ratio between low and high frequency on vertical channel

C. *Feature processing and regression*

Due to the existence of noise in features, we introduce a linear dynamical system (LDS) approach [9], a semisupervised dynamic model, for smoothing and de-noising. Details of LDS will be introduced in section III-C.

As features are obtained in the processes above, we employ a support vector machine for regression. We utilized the library for support vector machines (LIBSVM) [22] and search the parameters with grid strategy to ensure the best results. For the need of comparison, we do a full leave-one-out cross-validation of the 22 participants.

III. Deep Artificial network

In the traditional model of pattern recognition, a manual designed feature extractor gathers relevant information from the input and eliminates irrelevant variabilities. In order to bypass the manual ad-hoc feature extraction process, we try to incorporate a deep artificial neural network that transforms the raw signals into a set of features as input of a single linear regression layer. We investigate deep learning models popular at present and finally choose CNN based on the following two considerations. First, CNN model has been

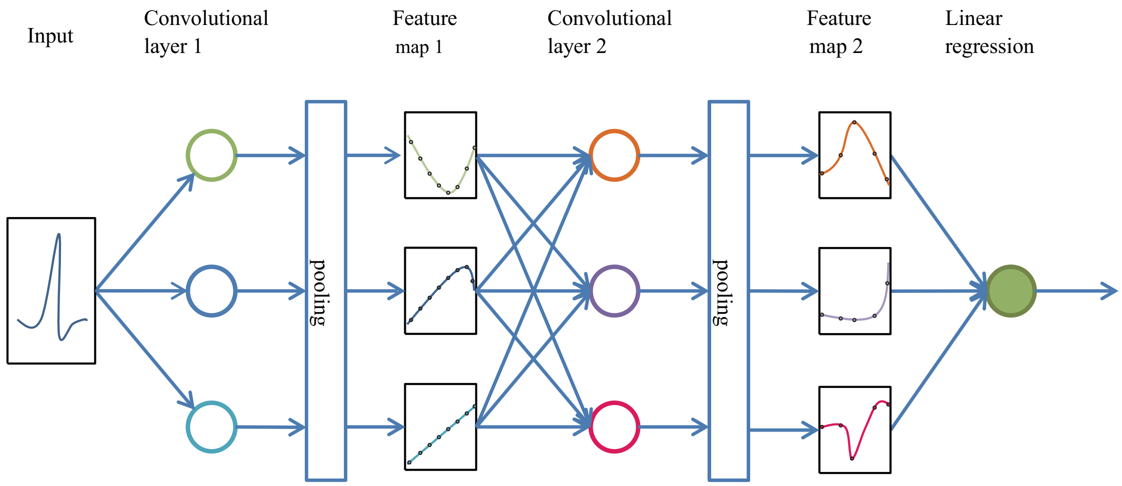


Fig. 2. The structure of a convolutional neural network adopted in this paper. The circles represent neurons and there are three neurons in each convolutional layers.

reported success in several fields, such as object recognition, and latest literature [12] shows effectiveness in the domain of psychophysiology, yielding efficient computational models of affect. Second, compared with other artificial neural networks, CNN is probably more suitable for EOG, which has abundant information in time domain and scarce in frequency domain. We believe that the automatic feature extraction stage via unsupervised learning process will yield physiological detectors, which can hardly be designed in manual ad-hoc methods, of more relevant to fatigue and thus generate a drowsiness detection model of higher accuracy. We train the CNN with an unsupervised learning method and convolutional auto-encoder in order to capture a distributed representation of its leading factors of variation. Further process includes a LDS approach to eliminate the unfavourable state switches.

A. Convolutional neural network

CNN is a hierarchical model with deep architecture that consists of a number of convolutional and sub-sampling layers followed by fully connected layers. The CNN scheme eliminates the feature extractor in the traditional model of pattern recognition, which is fed with 'raw' inputs, and relies on backpropagation learning algorithm to turn the layers into an appropriate feature extractor.

Figure 2 gives an intuitional description of the convolutional neural network. Convolutional and pooling layers are alternated to process large input signals into small features in small resolution. In the convolutional layers, there are a set of neurons that detect different patterns on a patch of the input. Each unit of a layer receives inputs from a set of units located in a small neighborhood in the previous layer. With local receptive fields, neurons can extract elementary features of input and these features are then combined by the higher layers. Each neurons contains a set of trainable weights and a bias and feature maps are calculated by applying an activation function to the weighted sum of the inputs plus the bias. With a pooling layer (max pooling in this paper), the resolution of the feature maps is reduced. The architecture itself realizes a form of regularization and keeps sparse by erasing all non-maximal values. To keep the model simple, we simply choose the linear regression as the predictor for drowsiness and the mean squared error as the cost function.

B. Convolutional auto-encoder

An auto-encoder (AE) [23] is an unsupervised learning method to detect and remove input redundancies and preserve only essential aspects of the data in robust and discriminative representations. Generally, AE network converts an input vector into a code vector, using a set of weights. Then the code is used to reconstruct the approximate input with a set of generative weights. Using the stochastic descent method, we can gradually minimize the reconstruction error and get the proper codes of the inputs. Once a layer is trained, its code is fed to the next, to better model highly non-linear dependencies in the input.

The structure of convolutional auto-encoder (CAE) is similar to AE. However, the convolutional auto-encoder [14] shares weights among all locations in the input, preserving spatial locality, which differs from conventional AEs. The latent representation of the k-th feature map for an input x given by

$$z^k = \sigma(x * W^k + b^k) \tag{1}$$

where b is the bias, σ is an activation function (sigmoid function in this paper), and $*$ denotes the convolution operation. The reconstruction y is given by,

$$y = \sigma(\sum_{k \in H} z^k * \tilde{W}^k + c) \tag{2}$$

where c represents the bias and H identifies the group of latent feature maps; $\tilde{W}$ is the flip operation over the weights. We adopted the mean squared error (MSE) as the cost function:

$$E(\theta) = \frac{1}{2n} \sum_{i=1}^{n} (x_i - y_i)^2 \tag{3}$$

We use the backpropagation algorithm to minimize the MSE. We can easily calculate the gradient of the error function with respect to the parameters:

$$\frac{\partial E(\theta)}{\partial W^k} = x * \delta z^k + \tilde{z}^k * \delta y. \tag{4}$$

where δz and δy are the deltas of the hidden states and the reconstruction, respectively. Weights are updated using stochastic gradient descent method.

Auto-encoder serves as an unsupervised learning technique that makes remarkable improvements to gradient-descent supervised learning.

C. Linear dynamic system

Generally, there are unfavourable status switching in the prediction value from the linear regression layer. To eliminate the negative factor, we introduce a linear dynamic system (LDS) approach [9], a semi-supervised dynamic model. The LDS approach is suitable for both off-line and on-line processing, which is recognised as an advantage over moving average, and has been successfully used in our previous work [7][24].

The LDS approach is capable to use only previous data to evaluate the current state without time delay. LDS is a dynamic model of state space, by making use of the time dependency of state changes without any labels, to further reduce the influence of fatigue-unrelated EOG. The structure of the state space model is shown in Fig. 3, where z_i means hidden variable and x_i means observed variable. Arrows in the figure show the internal relations between hidden variables. However, the internal relation is recognized weak or nonexistence in observed variables, due to the noise and unfavourable state switches. As shown in figure 3, the key idea is to deduce the probability distribution at the current state based on the relation among hidden variables and the relation between hidden variable and current observed variable. We assume that the noise in observed variable x_i and the time transition among the hidden states are both gaussian. Thus we can calculate the expectation and variation of these gaussian relations and then get z_i by knowing x_i.

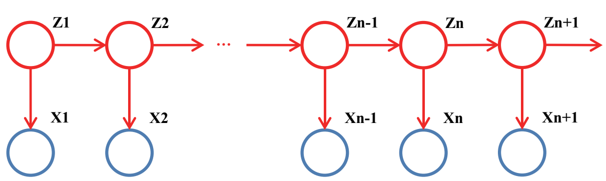


Fig. 3. The structure of state space of LDS. The red circles are hidden states, and the blue circles are observed states. Arrows mean inferable relations.

According to the assumptions mentioned above, we can obtain the equations,

$$\begin{aligned} x_i &= z_i + w_i, \\ z_i &= A z_{i-1} + v_i, \end{aligned} \tag{5}$$

where x_i is the observed state, z_i is the hidden state, A is the transition matrix, and w_i and v_i are independent gaussian random disturbance with zero mean. Eq. (5) can also be expressed in an equivalent form in terms of Gaussian conditional distributions,

$$\begin{aligned} p(x_i|z_i) &= \mathcal{N}(x_i|z_i + \bar{w}, Q), \\ p(z_i|z_{i-1}) &= \mathcal{N}(z_i|A z_{i-1} + \bar{v}, R). \end{aligned} \tag{6}$$

The initial state is assumed to be,

$$p(z_1) = \mathcal{N}(z_1|\pi_0, S_0). \tag{7}$$

The LDS model described above can be parameterized by $\theta = \{A, Q, R, \bar{w}, \bar{v}, \pi_0, S_0\}$. With the observation sequence x_i, θ can be determined using maximum likelihood through EM algorithm [25]. To inference the hidden states z_i from the observation sequence x_i, the marginal distribution, $p(z_i|X)$, is calculated. Then the hidden state can be expressed as,

$$z_i = E(z_i|X), \tag{8}$$

where E means expectation. This marginal distribution can be achieved by using messages propagation methods [25]. For on-line inference, X is set as $x_1, ..., x_i$, while for off-line inference, X is set as $x_1, ..., x_n$, including the data observed after x_i.

LDS approach is implemented in both manual ad-hoc feature extraction model and deep artificial model, however, in different ways. In the manual ad-hoc model, we implement LDS on EOG features for feature processing for that drowsiness-related parts are slow varying whereas the drowsiness-unrelated parts change irregularly. In the deep artificial model, LDS is implemented to smooth the regression results for that drowsiness-unrelated parts contribute to the unfavourable disturbance in the regression results. After LDS process, the drowsiness-related parts are reserved and the noises corresponding to the drowsiness-unrelated parts are removed.

IV. Experiment setup

A. experiment description

The whole experiment is about 70 minutes and 22 sessions are recorded from 22 different subjects. Subjects are required to get enough sleep at the night before the experiments. We conducted the experiments after lunch for the purpose that the subject is awake in the beginning and sleepy after about half an hour later.

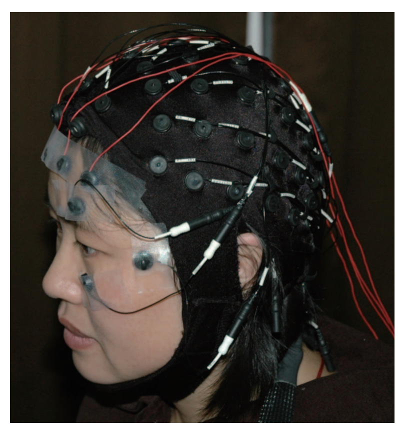


Fig. 4. The subject wearing the electrodes for recording both EEG and EOG signals.

The subject's task is to simply push the correct button as soon as possible according to the color of the image displayed on the screen in a quiet room with soft lights. We use traffic signs in four colors which are red, yellow, blue and green.

640 different signs are selected with each color 160. On the screen, signs are shown for 0.5 seconds and screen turns black for a 5-7 seconds interval. When the experiment performs, the system automatically records the correctness of each response (no response will be considered as incorrect). We suppose to get higher error rate when the subject is drowsier, and a curve of error rate is obtained throughout the whole experiment. The local error rates are calculated by a 2-minute time window with a step of 8s.

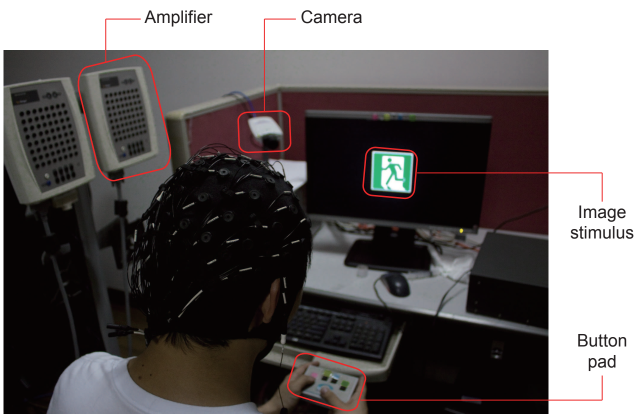


Fig. 5. The scene of our image identification experiment. The EOG data is collected by four electrodes on the EEG cap. The subject's task is to push the button which shares the same color with image stimulus.

In our experiments, EOG signals are recorded by the NeuroScan system. As is shown in Fig. 4, four electrodes on the EEG cap are used to collect the data. The signal of horizontal channel is the electric potential difference of the left and right ones and the signal of vertical channel is from the top and the bottom ones. The signals are recorded using a 32-bit resolution and 500Hz sample rate.

B. EOG label

In our experiment, we use the local error rate as the index of fatigue. The local error rate $e(t)$ is derived by computing the target false recognition rate within a 2-min time window at 8-s step as

$$e(t) = \frac{NumF(ST + 2t - L/2, ST + 2t - 1 + L/2)}{NumT(ST + 2t - L/2, ST + 2t - 1 + L/2)}, \quad (9)$$

where ST is the start time for fatigue measurement, L is the 120-s window length, $NumF_{(i,j)}$ is the number of false responses within the time window (i, j), and $NumT_{(i,j)}$ is the number of total stimuli within the time window (i, j).

The result is represented by correlation coefficient γ of the regression result and the local error rate. γ ranges from -1 to 1. Higher absolute value indicates higher relevance. γ is calculated as follows:

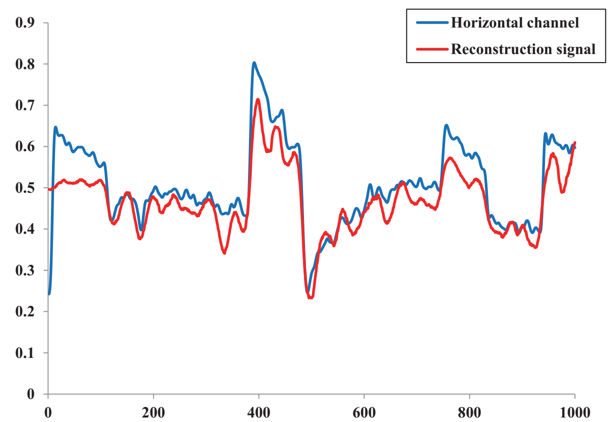
$$\gamma = \frac{\sum_t (f(t) - \overline{f(t)})(e(t) - \overline{e(t)})}{\sqrt{\sum_t (f(t) - \overline{f(t)})^2 * \sum_t (e(t) - \overline{e(t)})^2}}, \quad (10)$$

where $f(t)$ and $e(t)$ represent regression result and local error rate, respectively. $\overline{f(t)}$ and $\overline{e(t)}$ are their average over time.

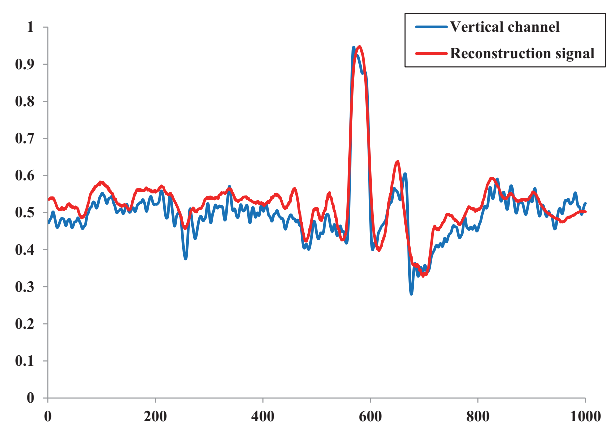
V. Result

To show the efficiency of our method, we applied the two methods described in section II and section III on the EOG data set we collected. More than 22 subjects have taken participate in the experiments. However, subjects who didn't show distinct fatigue during the experiments were eliminated. We performed the leave-one-out cross-validation to get the mean correlation coefficients which is regarded as the key index of drowsiness detection models. Comparison between the traditional model of pattern recognition and our models is presented to show that our model's advantages.

As is shown in section II, we processed the raw signals from the device and scaled the signals to the range of 0 to 1. We obtained two EOG channels, the horizontal and vertical, which are in the same range and high frequency removed. The training samples were randomly picked in the training sessions to balance the range of predictor variables.



(a) Horizontal channel



(b) Vertical channel

Fig. 6. The reconstruction of signals from CAEs. Fig. 6(a) and Fig. 6(b) show the reconstruction results of the horizontal and vertical channels of EOG, respectively. The length of signals in these two figures is 8 seconds.

The CNN we trained contains two convolutional layers with 8 and 4 neurons, respectively, as well as maxpooling layer over non-overlapping windows of size 8 and 2, respectively. On the consideration of overfitting, the topology of the network, which is competent to extract proper drowsiness-related fea-

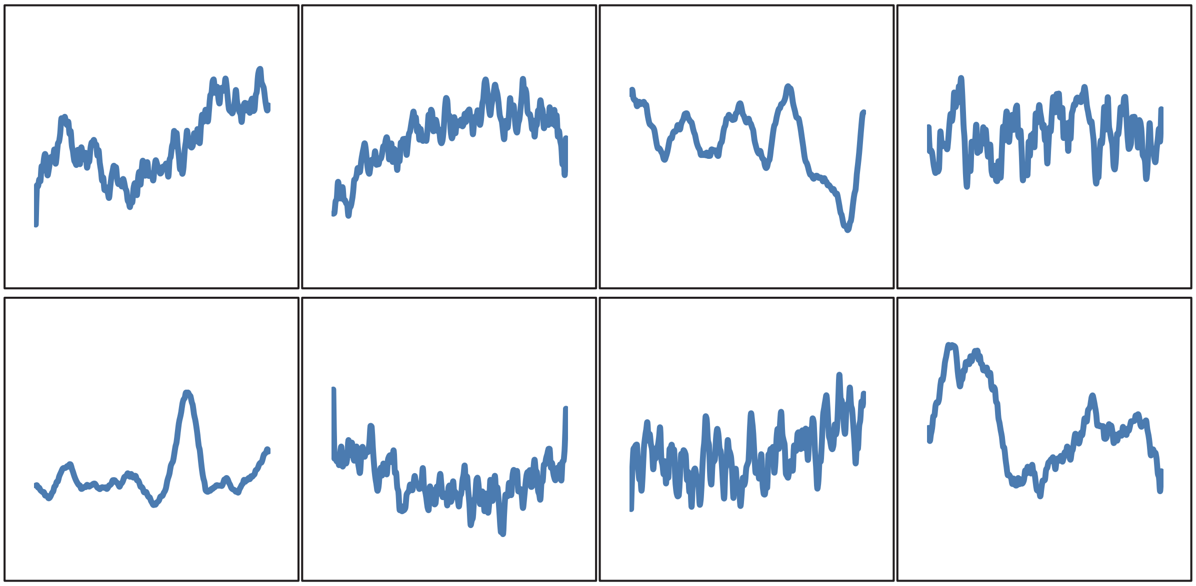


Fig. 7. Some of the learned feature maps of the first layer of CNN.

tures, was selected after preliminary experiments. Each neuron in the first and second convolutional layers has 201 and 26 inputs, respectively. We pretrained the convolutional neural network using stacked convolutional auto-encoders with the same structure as the CNN. The CAEs were kept training with entire training set without labels until the reconstruction error in Eq. (4) converged to a relatively small value. Fig. 6 depicts a reconstruction of the EOG signals in the first layer, both horizontal and vertical channels. Reconstruction error converges to a small value with training epoches of CAEs. We can intuitively see that the signals are perfectly reconstructed. Once the CAEs were trained, the weights were used to initialize a CNN. The linear regression layer was randomly initialized and then the local error rate labels were added to train the entire network using backpropagation learning algorithm with a relatively small number of epoches. Fig. 7 depicts a learned feature maps of the first layer. Part of the feature maps is similar to patterns of some eye movements, while part of the feature maps could be the patterns of artifacts. The second layer of the networks integrates the output of the first layer into higher features and the linear regression layer turns the higher feature into appropriate drowsiness predictor.

Fig. 8 shows an example of regression result of a subject in experiment. We can institutively see that a growing fatigue about 13 minutes since the experiment started and the subject fell asleep about 5 minutes later. The predict drowsiness curve seizes the tendency and fits the the local error rate curve in red well. We can see that the subject refreshed himself in the 23 minute and got tired 4 minutes later which is all captured by the drowsiness curve. The entire fitting results are shown in Fig. 9 and our model gets a mean correlation coefficient of 0.73 over the statistical methods of 0.70. More than half of the correlation coefficients of the artificial network are slightly better or better than the statistical method and our method gets a smaller standard deviation. These results indicate that our model processes an equivalent or even better abilities than the corresponding models built on commonly used ad-hoc statistical features on drowsiness detection.

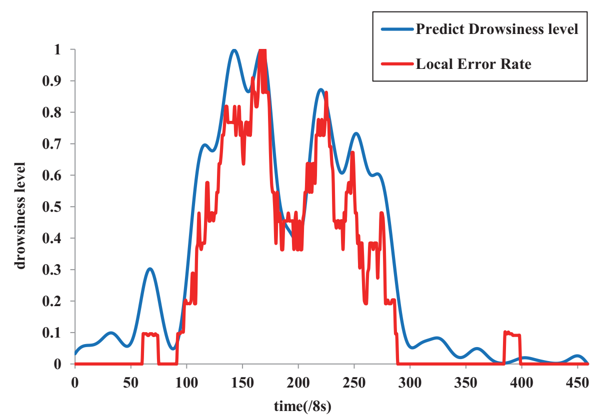


Fig. 8. An example of the regression result of a subject. The predicted drowsiness curve in blue seizes the tendency of local error rate and perfectly reflects the states of the subject.

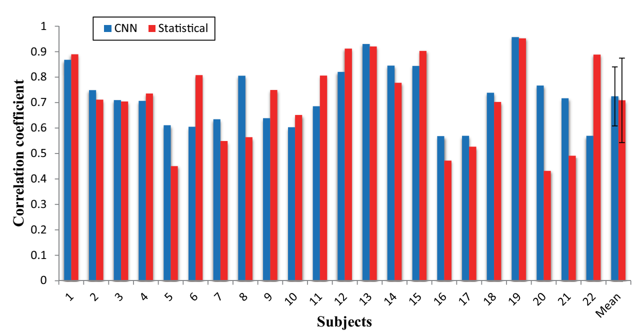


Fig. 9. Correlation coefficients of regression results for 22 testing sets.

VI. Conclusion

We introduced the application of convolutional neural network to EOG-based drowsiness detection and proposed a new reliable drowsiness detection model in this paper. The model we proposed employs two convolutional layers that learn to extract relevant features from the EOG signals. The EOG dataset is derived from 22 subjects of fatigue experiments. The experimental results on our EOG dataset showed that convolutional neural network possesses an equivalent or even better ability over manual ad-hoc feature extraction method on drowsiness detection task. Despite the manual designed features advantage in depicting eye movements and interpreting the physical properties of EOG, part of the features extracted through eye movements pattern detection by the deep neural networks is similar to the manual designed one, while other features are different from the manual designed and hard to design in practice. This model automatically provide a more complete and appropriate set of features and our results show that combination of these features yields a better predictor of drowsiness. In this model, the LDS approach performs an important step for reduction of unfavorable disturbance of drowsiness-unrelated parts. As a consequence, we obtain a remarkable increase in prediction accuracy.

This work also further demonstrates that convolutional neural network is applicable to physiological signals and deep learning methodologies are highly appropriate for drowsiness detection. This work also suggests that the trivial, tough and unstable feature extraction process in the traditional drowsiness detection model of pattern recognition can be redundant. With small modifications, the methodology proposed can be applied for online drowsiness detection model, which can be widely used in various scenarios. Future work includes a promotion of the topology of the network and tests on the parameter sets. Drowsiness detection models in other one-dimensional time-series physiological signals such as EEG and EMG should be done and more experiments in other scenarios will be performed to test the generality of the drowsiness detection model.

Acknowledgment

This work was partially supported by the National Natural Science Foundation of China (Grant No. 61272248), the National Basic Research Program of China (Grant No. 2013CB329401), the Science and Technology Commission of Shanghai Municipality (Grant No. 13511500200), and the Open Funding Project of National Key Laboratory of Human Factors Engineering (Grant No. HF2012-K-01).

References

[1] E. Martin, "Breakthrough research on real-world driver behavior released," *National Highway Traffic Safety Administration*, 2006.

[2] D. F. Dinges and R. Grace, "Perclos: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance," *Federal Highway Administration Office of motor carriers, Tech. Rep. MCRT-98-006*, 1998.

[3] F. Friedrichs and B. Yang, "Drowsiness monitoring by steering and lane data based features under real driving conditions," *Signal Processing*, pp. 209–213, 2010.

[4] J. Berglund, "In-vehicle prediction of truck driver sleepiness: Steering related variables," Ph.D. dissertation, Linköping, 2007.

[5] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," 1968.

[6] L.-C. Shi and B.-L. Lu, "Dynamic clustering for vigilance analysis based on EEG," in *Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society,* Vancouver, BC Canada, pp. 54-57, 2008.

[7] L.-C. Shi and B.-L. Lu, "EEG-based vigilance estimation using extreme learning machines," *Neurocomputing*, vol. 102, pp. 135–143, 2013.

[8] L.-C. Shi, H. Yu, and B.-L. Lu, "Semi-supervised clustering for vigilance analysis based on EEG," in *Proceedinds of International Joint Conference on Neural Networks,* pp. 1518-1523, 2007.

[9] L.-C. Shi and B.-L. Lu, "Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning," in *Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society,* Buenos Aires, Argentina, pp. 6587-6590, 2010.

[10] J.-X. Ma, L.-C. Shi, and B.-L. Lu, "Vigilance estimation by using electrooculographic features," in *Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society,* Buenos Aires, Argentina, pp. 6591-6594, 2010.

[11] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, pp. 1106–1114, 2012.

[12] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, "Learning deep physiological models of affect," *IEEE Computational Intelligence Magazine,* , vol. 8, no. 2, pp. 20–33, 2013.

[13] H. Cecotti and A. Graser, "Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 433-445, Mar. 2011.

[14] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proceedings of International Conference of Artificial Neural Networks and Machine Learning 2011*. pp. 52-59, 2011.

[15] J. Weston, F. Ratle, H. Mobahi, and R. Collobert, "Deep learning via semi-supervised embedding," in *Proceedings of International Conference of Machine Learning,* pp. 1168-1175, 2008.

[16] N. Jaitly and G. Hinton, "Learning a better representation of speech soundwaves using restricted boltzmann machines," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing ,* Prague, Czech Republic, pp. 5884–5887, 2011.

[17] Z.-P. Wei and B.-L. Lu, "Online vigilance analysis based on electrooculography," in *The 2012 International Joint Conference on Neural Networks (IJCNN).* IEEE, pp. 1–7, 2012.

[18] R. Schleicher, N. Galley, S. Briest, and L. Galley, "Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?" *Ergonomics*, vol. 51, no. 7, pp. 982–1010, 2008.

[19] E. Magosso, F. Provini, P. Montagna, and M. Ursino, "A wavelet based method for automatic detection of slow eye movements: A pilot study," *Medical engineering & physics*, vol. 28, no. 9, pp. 860–875, 2006.

[20] E. Magosso, M. Ursino, A. Zaniboni, F. Provini, and P. Montagna, "Visual and computer-based detection of slow eye movements in overnight and 24-h eog recordings," *Clinical neurophysiology*, vol. 118, no. 5, pp. 1122–1133, 2007.

[21] E. Magosso, M. Ursino, A. Zaniboni, and E. Gardella, "A wavelet-based energetic approach for the analysis of biomedical signals: Application to the electroencephalogram and electro-oculogram," *Applied Mathematics and Computation*, vol. 207, no. 1, pp. 42–62, 2009.

[22] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[23] G. E. Hinton and R. S. Zemel, "Autoencoders, minimum description length, and helmholtz free energy," *Advances in neural information processing systems*, pp. 3–3, 1994.

[24] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "EEG-based emotion recognition during watching movies," in *The 5th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 667–670, 2011.

[25] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications: with R examples*, Springer, 2011.