# Continuous Vigilance Estimation Using LSTM Neural Networks

Nan Zhang[1], Wei-Long Zheng[1], Wei Liu[1], and Bao-Liang Lu[1,2,3(✉)]

[1] Center for Brain-like Computing and Machine Intelligence,
Department of Computer Science and Engineering,
Shanghai Jiao Tong University,Shanghai, China
vincentzn4@outlook.com, {weilong,liuwei-albert,bllu}@sjtu.edu.cn
[2] Key Laboratory of Shanghai Education Commission for Intelligent Interaction
and Cognition Engineering, Shanghai Jiao Tong University, Shanghai, China
[3] Brain Science and Technology Research Center,
Shanghai Jiao Tong University, Shanghai, China

**Abstract.** In this paper, we propose a novel continuous vigilance estimation approach using LSTM Neural Networks and combining Electroencephalogram (EEG) and forehead Electrooculogram (EOG) signals. We combine these two modalities to leverage their complementary information using a multimodal deep learning method. Moreover, since the change of vigilance level is a time dependent process, temporal dependency information is explored in this paper, which significantly improves the performance of vigilance estimation. We introduce two LSTM Neural Network architectures, the F-LSTM and the S-LSTM, to encode the time sequences of EEG and EOG into a high level combined representation, from which we can predict the vigilance levels. The experimental results demonstrate that both of the two LSTM multimodal structures can improve the performance of vigilance estimation models in comparison with the single modality models and non-temporal dependent models.

**Keywords:** EEG · Vigilance estimation · Multimodal · Deep learning · Recurrent neural network

## 1 Introduction

Brain-computer interaction (BCI) aims to translate brain activity or state into control signals for computer devices [1]. A lot of studies have been done on the assessment of human's brain states such as vigilance and emotion in order to develop affective brain-computer interaction systems [2]. Vigilance or alertness, which means the ability to maintain sustained concentration, is an important kind of mental state for user aware BCI systems. High vigilance is usually required for some occupations such as truck drivers or pilots. In these cases, low vigilance may bring tragedy to both themselves and other people. For example driving fatigue is believed to be a significant reason for most of the fatal traffic accidents [3]. Therefore a robust vigilance estimation system is desired to improve the transportation safety.


A. Hirose et al. (Eds.): ICONIP 2016, Part II, LNCS 9948, pp. 530–537, 2016.
DOI: 10.1007/978-3-319-46672-9_59

Various approaches have been proposed to estimate the vigilance level over the past years. Different kinds of signals are utilized such as video [4], speech [5] and physiological signals [3]. In these signals, EEG is considered as a good indicator of the transition from wakefulness to sleepiness. Eoh et al. showed that the proportion of different spectral components in EEG is related to the alertness level [3]. Davidson et al. introduced a warning system capable of detecting lapse with high temporal resolution [4]. In addition to EEG, EOG signal also contains information that has close relationship with vigilance status. Eye movement features such as slow eye movements (SEM) and blinks [8] have been shown to be good indicators of vigilance level. The traditional EOG are collected from electrodes placed around the eyes, which may distract subjects and cause discomfort. Zhang et al. proposed to place the electrodes on the forehead and extract features from forehead EOG to detect driving fatigue [9].

However, most of these methods ignore the time dependency property of the vigilance changing process. The subject's mental states are treated as independent points and the temporal dependency information are discarded in these models. Recurrent Neural Network (RNN) is a kind of artificial neural network where connections between units form a cycle which makes it suitable to process sequence data. RNN has been successfully applied to research domains such as machine translation [10] and speech recognition [11]. In this paper, we introduce the RNN as a multimodal encoder which can incorporate the temporal changes of EEG and EOG features to help with the estimation of vigilance. The mental state sequence is encoded into a fixed-dimensional vector representation which contains meaningful information to decode the vigilance level.

This paper is organized as follows. In Sect. 2, we describe the method used to build vigilance estimation system. Section 3 gives a detailed description about the setup of our simulated driving experiment and how we collect our data. In Sect. 4 we discuss the experiment results using different models. Finally in Sect. 5, conclusions are presented.

## 2 LSTM Neural Networks

Vigilance changing is a dynamic process. To incorporate the time dependency information, we introduce the Recurrent Neural Network (RNN) model as a temporal encoder. RNN contains cyclical connections in its hidden layers and can remember the history of its input. For a length $T$ input sequence $\boldsymbol{x}$, at time $t$ in forward pass, the hidden units states are updated as:

$$\boldsymbol{h_t} = f\left(\boldsymbol{W}\boldsymbol{x_t} + \boldsymbol{U}\boldsymbol{h_{t-1}} + \boldsymbol{b}\right) \tag{1}$$

where $\boldsymbol{h_t}$ and $\boldsymbol{x_t}$ are respectively the output vector and input vector of a hidden layer at time $t$, $f$ is the activation function, $\boldsymbol{W}$ and $\boldsymbol{U}$ are weight matrices, and $\boldsymbol{b}$ is the bias vector.

The problem of this simple RNN architecture is that only small range of contextual information can be used from the input sequence which will cause the vanishing gradient problem when applying to longtime sequence. Since we

need to learn information from longtime EEG/EOG sequences, the Long Short Time Memory (LSTM) neural network is applied. LSTM neural network is a RNN with LSTM blocks as units in hidden layers. Each LSTM block contains memory cells and input gate, output gate and forget gate, which provide write, read and reset operations for the cells. In this way, the LSTM cells can store states over long periods of time. The state of memory cells is updated at every time step $t$ as follows:
Input Gate:

$$\boldsymbol{i_t} = f\left(\boldsymbol{W_i x_t} + \boldsymbol{U_i h_{t-1}} + \boldsymbol{b_i}\right) \tag{2}$$

Forget Gate:

$$\boldsymbol{f_t} = f\left(\boldsymbol{W_f x_t} + \boldsymbol{U_f h_{t-1}} + \boldsymbol{b_f}\right) \tag{3}$$

Cells update:

$$\overline{\boldsymbol{C_t}} = g\left(\boldsymbol{W_c x_t} + \boldsymbol{U_c h_{t-1}} + \boldsymbol{b_c}\right) \tag{4}$$

$$\boldsymbol{C_t} = \boldsymbol{i_t} * \overline{\boldsymbol{C_t}} + \boldsymbol{f_t} * \boldsymbol{C_{t-1}} \tag{5}$$

Output Gate:

$$\boldsymbol{o_t} = f\left(\boldsymbol{W_o x_t} + \boldsymbol{U_o h_{t-1}} + \boldsymbol{b_o}\right) \tag{6}$$

$$\boldsymbol{h_t} = \boldsymbol{o_t} * k\left(\boldsymbol{C_t}\right) \tag{7}$$

where $f, g$ and $k$ are all activation functions, $\boldsymbol{i_t}, \boldsymbol{f_t}$ and $\boldsymbol{o_t}$ are outputs of gates, and $\overline{\boldsymbol{C_t}}$ is the candidate of cells' state.

The EEG and EOG feature sequences need to be adapted to the input of RNN architecture. First, the data is normalized to zero mean and unit variance, then the whole data sequence is divided into many short data sequences. Each data sequence is nearly one minute which, as we show in the experiment result, is long enough to estimate vigilance levels. In order to learn from multi modalities, we propose two LSTM network architectures that can fuse information from EEG and EOG sequences. One is to use two independent LSTM encoders to encode EEG and EOG sequences respectively and then combine their representations into one feature vector (F-LSTM) shown in Fig. 1(a). The other is to concatenate the feature vectors of EEG and EOG at each time step and then use stacked LSTM layers to encode the feature sequence into a compact feature vector (S-LSTM) shown in Fig. 1(b).

We implement our model using python theano and decide all the hyper parameters by cross validation. In S-LSTM, we use 3 stacked hidden LSTM layers as encoder and one sigmoid neuron as output layer. Each LSTM layer has half number neurons comparing to the input layer. The internal weights in LSTM units are initialized from a standard Gaussian distribution followed by a SVD orthogonalization. The other weights are initialized from a uniform distribution with scale parameter determined by Xavier algorithm. The bias value of forget gates are initialized with ones. The activation functions of all the gates are sigmoid while tanh is used otherwhere in LSTM units. In F-LSTM, we append a

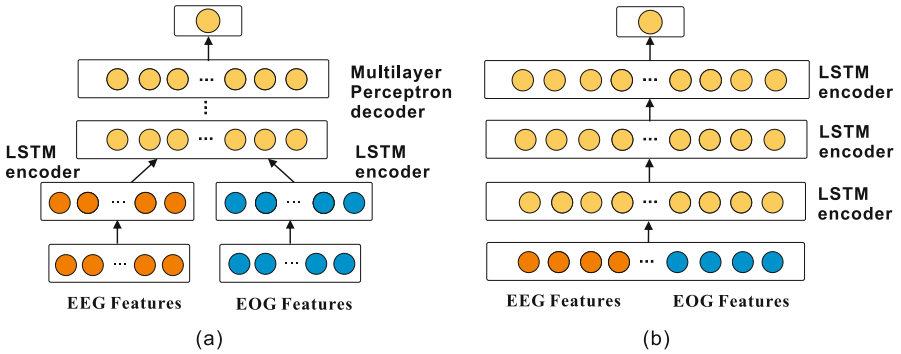


**Fig. 1.** Two LSTM structures adopted in this paper. Figure (a) depicts the F-LSTM model which combines two LSTM encoders designed respectively for EEG and EOG. Figure (b) depicts the S-LSTM model which merges the EEG and EOG at feature level.

Multilayer Perceptron (MLP) model after the last LSTM layer as a decoder. The activation function used in MLP is ReLU. In order to generalize our model, dropout with a probability 0.5 is added before the output layer. In training, RMSProp method is used instead of basic stochastic gradient descent method to optimize the loss function. Early stopping strategy is adopted when no improvement appears on the performance on validation set after 10 epochs.

## 3 Experiment Setup and Data Processing

### 3.1 Experiment Setup

The experiments were performed in a simulated driving environment. A four-lane high way scene was shown in front of a car. The subjects drove the car just like driving a real car outdoors. There were in total 21 subjects (12 men 9 women) at the age between 20 and 25, who participated in the experiments. Before the experiments started, a warm up session was performed to ensure every participant was familiar with the operation. All the experiments were conducted after lunch from 13:00 pm to 15:00 pm or after dinner from 21:00 pm to 23:00 pm. The participants were asked to drive the car for 2 h in the simulated driving environment. Both of the EOG and EEG signals were recorded simultaneously using the Neuroscan system with a 1000 Hz sampling rate. At the same time, eye movement data was recorded using SMI ETG eye tracking glasses.

### 3.2 Feature Extraction

**EEG Signals:** The EEG signals are down-sampled to 200 Hz to reduce computational complexity and preprocessed with a band-pass filter between 1 Hz and 75 Hz to reduce noise and artifacts. EEG signals from 17 channels

(FT7, FT8, T7, T8, TP7, TP8, CP1, CP2, P1, PZ, P2, PO3, POZ, PO4, O1, OZ, O2) located at temporal lobe and posterior lobe areas are recorded, since these areas have been shown to have high relevance with vigilance in previous findings [6] [7]. Short-time Fourier transform with a 8 s non-overlapping Hanning window is used to extract five frequency bands of EEG signals. Although a smaller time window can be used for EEG, but in order to align with EOG which needs a bigger window to detect eye movements, a 8 s window is selected. The five frequency bands are divided as follows, delta: 1–4 Hz, theta: 4–8 Hz, alpha: 8–14 Hz, beta: 14–31 Hz and gamma: 31–75 Hz. For each frequency band, we extract the differential entropy (DE) features, which has been shown superior performance compared to the power spectral density (PSD) features in our previous study [12].

**EOG signals:** EOG features are also extracted with a 8 s non-overlapping window on EOG signals. For traditional EOG, the electrodes are placed around eyes as shown in Fig. 2(a). This will distract subject from the driving process and bring discomfort to the subject. In this work, all electrodes are placed on the forehead as Fig. 2(b) and we extract features from forehead EOG. For traditional EOG, the vertical EOG (VEO) and horizontal EOG (HEO) are extracted by subtracting electrodes four from three and electrodes one from two, respectively.

For forehead EOG, forehead VEO is extracted from electrodes four and seven by using independent component analysis (ICA). Forehead HEO is extracted by simply subtraction from electrode five and six. After preprocessing forehead EOG signals, we detect eye movements such as blinks and saccades using wavelet transform method. Continuous wavelet coefficients at scale 8 with Mexican mother wavelet are calculated. The blinks and saccades are then detected from VEO and HEO, respectively. The statistical parameters such as blink/saccade duration, mean, maximum, variance and derivative are extracted as EOG features. A total of 36 EOG features are used in this paper. The detailed descriptions of EOG features are described in [9].


**Fig. 2.** Electrode placements for the traditional and forehead EOG setups. The yellow and blue colors indicate the electrode placements of the traditional EOG and forehead EOG, respectively. The electrode four is the shared electrode of both setups. (Color figure online)

### 3.3 Vigilance Labeling

In this study, the ground truth vigilance values are obtained using eye tracking glasses proposed in [13]. PERCLOS, which refers to the percentage of eyelid closure over time, is used as the index of alertness level. PERCLOS is defined as [13]:

$$PERCLOS = \frac{blink + CLOS}{interval} \tag{8}$$

$$interval = blink + fixation + saccade + CLOS \tag{9}$$

where 'CLOS' denotes the duration of eye closure. We calculate the PERCLOS values using eye tracking glasses as the labels of vigilance levels. It should be noted here that although the eye tracking glasses can estimate the vigilance level precisely, it's not a good choice to use in real world applications due to its expensive cost and longtime delay. So we only use it as a vigilance labeling method and obtain the labels to train our models.

## 4 Experiment Results

We use the support vector regression (SVR) with radial basis function (RBF) kernel as a baseline in this paper. To evaluate the experiment results, we divided our whole data sequence from one experiment into five segments and evaluated the performance with 5-fold cross validation. The Root Mean Square Error (RMSE) and Correlation Coefficient (COR) are used as metrics for the experiment results.

First we investigate whether multiple modalities are helpful for the result of vigilance estimation. We used the S-LSTM model for the two single modalities, which means instead of the concatenation of EEG and EOG features either the EEG or EOG feature was used as input to S-LSTM model. For multiple modalities, the S-LSTM and F-LSTM network architectures were used to fuse the two modalities. The mean and standard deviation for RMSE and COR are shown in Table 1. We can see that both of the two multimodalities models achieved better results than single modality methods.

Next we will examine the importance of time dependency information in estimating vigilance. The SVR model used doesn't take time dependency into

**Table 1.** Experiment results for different models. Each single modality uses S-LSTM model in first two columns. Last three columns are models fusing mulimodalities

| Model | | EEG | EOG | S-LSTM | F-LSTM | SVR |
|---|---|---|---|---|---|---|
| COR | Mean | 0.8237 | 0.8203 | 0.8329 | **0.8363** | 0.7958 |
| | Std | **0.0831** | 0.1191 | 0.0961 | 0.1009 | 0.1131 |
| RMSE | Mean | 0.0927 | 0.0935 | 0.0816 | **0.0807** | 0.1186 |
| | Std | 0.0259 | 0.0215 | 0.0189 | **0.0135** | 0.0515 |

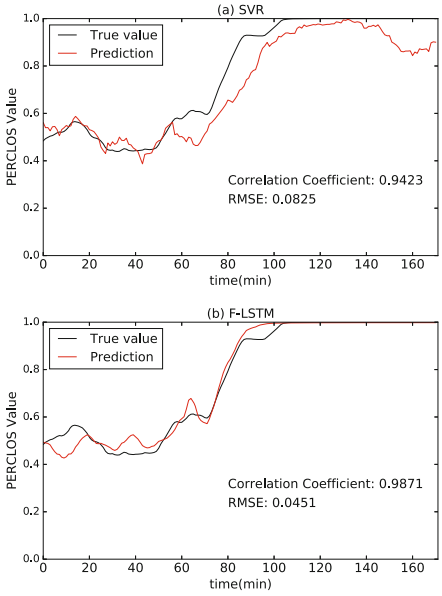


**Fig. 3.** The vigilance level prediction curves obtained by SVR and F-LSTM models.

consideration. The input of SVR model is the concatenation of EEG and EOG features. The mean and standard deviation for RMSE and COR are shown in Table 1. We can see from the results that LSTM models can significantly improve the estimation results compared to SVR. Figure 3 shows the vigilance prediction curves of SVR and M-LSTM models. The curves of M-LSTM model is more smooth comparing to SVR model. This means incorporating the time dependency information to vigilance estimation can make the system more robust to noise and predict the trend of vigilance levels more accurately.

## 5 Conclusion

In this paper, we have proposed a vigilance estimation approach combining two modalities and incorporating time dependency information. Two LSTM neural network structures were proposed to encode longtime signal sequences. The experimental results show that our proposed multimodal LSTM based methods can achieve significant improvement on vigilance estimation comparing to the traditional models.

**Acknowledgment.** This work was supported in part by the grants from the National Natural Science Foundation of China (Grant No. 61272248), the National Basic Research Program of China (Grant No. 2013CB329401) and the Major Basic Research Program of Shanghai Science and Technology Committee (15JC1400103).

## References


1. Brunner, C., et al.: BNCI horizon 2020 – towards a roadmap for brain/neural computer interaction. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2014, Part I. LNCS, vol. 8513, pp. 475–486. Springer, Heidelberg (2014)
2. Lu, Y., Zheng, W.-L., Li, B., Lu, B.-L.: Combining eye movements and EEG to enhance emotion recognition. In: IJCAI 2015, pp. 1170–1176 (2015)
3. Eoh, H.J., Chung, M.K., Kim, S.-H.: Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. Int. J. Ind. Ergon. **35**(4), 307–320 (2005)
4. Davidson, P.R., Jones, R.D., Peiris, M.T.R.: EEG-based lapse detection with high temporal resolution. IEEE Trans. Biomed. Eng. **54**(5), 832–839 (2007)
5. Krajewski, J., Batliner, A., Golz, M.: Acoustic sleepiness detection: framework and validation of a speech-adapted pattern recognition approach. Behav. Res. Methods **41**(3), 795–804 (2009)
6. Khushaba, R.N., Kodagoda, S., Lal, S., Dissanayake, G.: Driver drowsiness classification using fuzzy wavelet-packet-based feature extraction algorithm. IEEE Trans. Biomed. Eng. **58**(1), 121–131 (2011)
7. Shi, L.-C., Bao-Liang, L.: EEG-based vigilance estimation using extreme learning machines. Neurocomputing **102**, 135–143 (2013)
8. Ma, J.-X., Shi, L.-C., Lu, B.-L.: Vigilance estimation by using electrooculographic features. In: 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6591–6594 (2010)
9. Zhang, Y.-F., Gao, X.-Y., Zhu, J.-Y., Zheng, W.-L., Lu, B.-L.: A novel approach to driving fatigue detection using forehead EOG. In: 2015 7th International IEEE/EMBS Conference on Neural Engineering, pp. 707–710 (2015)
10. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems, pp. 3104–3112 (2014)
11. Deng, L., Li, J., Huang, J.-T., et al.: Recent advances in deep learning for speech research at Microsoft. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 8604–8608 (2013)
12. Shi, L.-C., Jiao, Y.-Y., Lu, B.-L.: Differential entropy feature for EEG-based vigilance estimation. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6627–6630 (2013)
13. Gao, X.-Y., Zhang, Y.-F., Zheng, W.-L., Lu, B.-L.: Evaluating driving fatigue detection algorithms using eye tracking glasses. In: 2015 7th International IEEE/EMBS Conference on Neural Engineering, pp. 767–770 (2015)