A Robust Approach to Estimating Vigilance from EEG with Neural Processes

Chen-Li Yao
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, China
yao09605@sjtu.edu.cn

Bao-Liang Lu*
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, China
bllu@sjtu.edu.cn

2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 978-1-7281-6215-7/20/$31.00 ©2020 IEEE DOI: 10.1109/BIBM49941.2020.9313095

***Abstract*—Robust vigilance estimation is essential for driving and other tasks that require a high degree of vigilance. Many approaches have been applied to estimating vigilance and much endeavor is made to improve the performance of vigilance estimating models. However, most of the existing approaches require test data to have similar quality as training data, which is difficult to be satisfied. In this paper, we adopt neural processes to deal with noise EEG data problem encountered in real scenarios. A publicly available dataset, SEED-VIG, is used to evaluate the performance and robustness of the proposed neural processes method. The dataset includes electroencephalography (EEG) and the corresponding vigilance level annotations during simulated driving. We compared the neural processes method with the existing regression models. The experimental results demonstrate that the neural processes are far better than all others in robustness, meanwhile maintain high accuracy.**

***Index Terms*—Neural processes, electroencephalography (EEG), vigilance estimation, noise EEG data**

I. Introduction

In human activities, there are many tasks that need people to maintain a high degree of vigilance, such as driving, flying, and security work. Reduced brain alertness not only affects work performance, but also may cause serious consequences. According to the National Highway Traffic Safety Administration, every year about 100,000 police-reported crashes involve drowsy driving, including more than 1,550 fatalities and 71,000 injuries [1]. Therefore, it is necessary to detect the vigilance. At present, there are three main types of vigilance estimation methods that have been mass-produced: detecting steering wheel operation characteristics, detecting vehicle trajectory, and judging by the driver's facial expression. But the common problem of these three methods is that when fatigue is detected, fatigue has already occurred. They all lack the function of predicting fatigue.

In recent years, the use of EEG signals to detect physiological and mental states has developed rapidly. Our previous studies show that using EEG signals to estimate vigilance has a better performance [2]. However, EEG acquisition equipments are generally expensive and inconvenient to wear. Often, high-quality EEG data can only be acquired in the laboratory environment, while the EEG data collected from real-world application environment are noisy, and even some data are missing. Therefore, a flexible and robust algorithm becomes necessary for the application of EEG-based vigilance estimation.

In 2018, Garnelo $et\ al.$ proposed a model called neural processes [3]. Neural processes combine the advantages of neural network and Gaussian process. They can not only learn efficiently like a neural network, but also make full use of data like a Gaussian process, and output the probability distribution of a prediction result.

In this paper, we apply neural processes to vigilance estimation based on EEG. We train a regression model to predict the degree of fatigue. We also compare the effectiveness and robustness of the neural processes method with support vector regression (SVR), multilayer perceptron (MLP), Gaussian process and continuous conditional neural fields (CCNF). Experimental results indicate that the neural processes method have much better robustness than all other models while maintaining high accuracy.

The remainder of this paper is organized as follows. Section 2 introduces the structure and the principle of neural processes. In section 3, we introduce the experiments settings. Section 4 presents experimental results and finally discussion and conclusions are given in section 5.

II. Model

A. Neural processes as stochastic processes

In probability theory and related fields, a stochastic or random process is a mathematical object usually defined as a family of random variables. We denote this random process as a random function, $\mathcal{F} : \mathcal{X} \rightarrow \mathcal{Y}$. For each finite sequence $x_{1:n} = (x_1, ..., x_n)$ with $x_i \in \mathcal{X}$, the corresponding function values $Y_{1:n} := (F(x_1), ..., F(x_n))$. Let $\rho_{x_{1:n}}$ be the marginal distribution of $(F(x_1), ..., F(x_n))$. Stochastic process needs to meet two conditions: exchangeability and consistency.

Exchangeability means the joint distribution is invariant to permutations of the elements in $x_{1:n}$. Consistency means if a part of the sequence is marginalized, the resulting marginal distribution is the same as it is defined on the original definition.

*Corresponding author

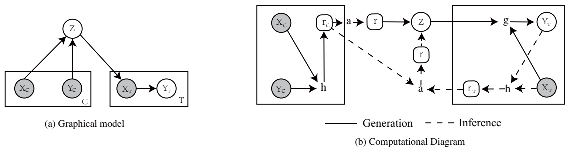


Fig. 1. (a) The graphical model of neural processes. x and y correspond to the data where $y = f(x)$. Group C and T refer to context points and target points, respectively. z is the global latent random variable. A filled circle means the corresponding variable is observed. (b) The computational diagram of neural processes. The solid part indicates the generation process. The dotted line indicates the inference process of the latent variable z. Variables in square boxes are intermediate representations. Letters 'g' and 'h' correspond to neural networks that act as an encoder and a decoder, respectively. Letter 'a' is an aggregator that aggregates the encoded representation of original x and y.

For a particular stochastic process f, the joint distribution is defined as:

$$\rho_{x_{1:n}}(y_{1:n}) = \int p(f)p(y_{1:n}|f, x_{1:n})\, df \tag{1}$$

To solve real-world problems, we represent the model as following:

$$p(x, y_{1:n}|x_{1:n}) = p(z)\prod_{i=1}^{n}\mathcal{N}(y_i|g(x_i, z), \sigma^2) \tag{2}$$

where $p(z)$ is a multivariate standard normal, and $g(x_i, z)$ is a neural network.

B. The neural processes model

To have a better performance at test phase, neural processes split the data into context data $\{x_{1:m}, y_{1:m}\}$ and target data $\{x_{m+1:n}, y_{m+1:n}\}$ as is shown in Fig. 1. A proposed distribution $q(z|x_{1:n}, y_{1:n})$ is introduced as variational posterior of the latent variables z. So the optimization problem turns into:

$$\arg\max_{\Theta} \log p(y_{m+1:n}|x_{1:n}, y_{1:m}) \tag{3}$$

where Θ represents all parameters of the model.

With the help of $q(z|x_{1:n}, y_{1:n})$, the evidence lower-bound (ELBO) is given by:

$$\log p(y_{m+1:n}|x_{1:n}, y_{1:m}) \geq \mathbb{E}_{q(z|x_{1:n},y_{1:n})}\left[\sum_{i=m+1}^{n}\log p(y_i|z, x_i) + \log\frac{p(z|x_{1:m}, y_{1:m})}{q(z|x_{1:n}, y_{1:n})}\right] \tag{4}$$

Note that in the above equation, conditional prior $p(z|x_{1:m}, y_{1:m})$ is intractable. So, neural processes approximate it using the variational posterior $q(z|x_{1:m}, y_{1:m})$. Finally, the evidence lower-bound (ELBO) turns out:

$$\log p(y_{m+1:n}|x_{1:n}, y_{1:m}) \geq \mathbb{E}_{q(z|x_{1:n},y_{1:n})}\left[\sum_{i=m+1}^{n}\log p(y_i|z, x_i)\right] - KL[q(z|x_{1:n}, y_{1:n})||q(z|x_{1:m}, y_{1:m})] \tag{5}$$

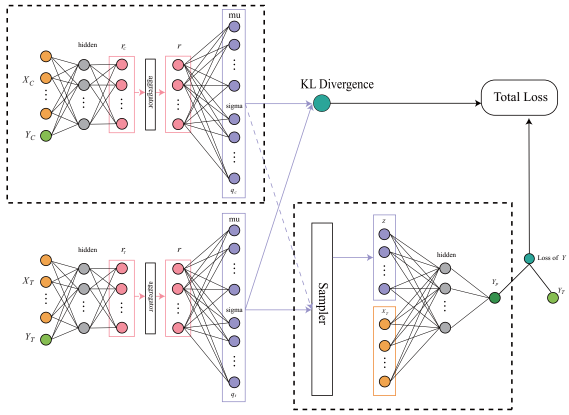


Fig. 2. The structure of neural processes. Context points and target points are encoded separately into representation r. The approximate multivariate gaussian distributions q_c and q_t are extracted from representations. The KL divergence of q_c and q_t are calculated, as part of the loss function which we wish to minimize. z is sampled from q_t. A decoder is constructed which generates Y_p from z and X_t. The distance between Y_p and Y_t is another part of the loss function.

C. The algorithm for training and prediction

The detail processes of training and prediction are shown in Fig. 2.

1) Training process: In the training process, input data $\{X_{1:n}, Y_{1:n}\}$ are organized into context set $\{X_{1:m}, Y_{1:m}\}$ and target set $\{X_{1:n}, Y_{1:n}\}$. Then a neural network acts as an encoder that transforms $\{X_{1:m}, Y_{1:m}\}$ and $\{X_{1:n}, Y_{1:n}\}$ into deep feature representations r_c and r_t, respectively. r_c and r_t are aggregated into r by an aggregator which calculates the means.

Another two neural networks are used to look for a multi-dimensional Gaussian distribution close to r. These two neural networks transform r into μ and Σ of the posterior distribution of global latent variable z. Based on context points and target points, we can draw two posterior distributions q_c and q_t. Then we calculate the KL divergence of these two distributions.

Next, we sample q_t to get z and construct another neural network as a decoder to generate the distribution of marginal distribution $p(y|z, x)$. Then we calculate the sum of log-probability of real y on the marginal distribution $p(y|z, x)$. Only $y_{m+1:n}$ are taken into account according to equation (8) this time.

2) Prediction process: In the prediction process, some context points x_c, y_c are needed. Context points are encoded into the posterior distribution q using the trained encoder and aggregator. Then z is sampled from q. Finally X_t and z are used as inputs, and y_p is obtained through the trained decoder.

III. Data Description and Preprocessing

A. The SEED-VIG Dataset

To collect EEG data, a simulated driving system was developed in our previous work. The system is composed of a large LCD screen and a real vehicle without unnecessary engine and other components. The vehicle was modified so the participants can operate the vehicle in the screen in front of them through the steering wheel. The LCD screen displays a four-lane highway scene which is primarily straight and monotonous, so as to induce fatigue in the subjects.

A total of 23 volunteers with an average age of 23.3 years participated in the fatigue driving experiments and 12 of them are female. All of the subjects had normal or corrected vision. Before the experiment, caffeine, tobacco, and alcohol that affect the nervous system were prohibited. Subjects were required to participate in the experiment in the afternoon or late at night to easily cause driving fatigue. The experiment lasted about 2 hours, during which the data were recorded. Both EEG and forehead EOG signals were recorded during the fatigue driving experiment. But only EEG data are used in this paper. For the EEG signals, 11-channel EEG signals from the posterior site and 6-channel EEG signals from the temporal site were recorded according to the international 10-20 electrode system.

B. Vigilance annotations

An annotation called PERCLOS measure is adopted to measure subjects' fatigue. PERCLOS indicates percentage of eye closure. So the value of PERCLOS is between 0 (high vigilance) and 1 (low vigilance).

C. Feature extraction from EEG

The raw EEG data are split into 885 segments. Each segment contains 8 seconds of raw data. A band-filter was applied to transform the raw data into 25 2-Hz frequency bands, that is frequency bands of 1∼2 Hz, 2∼4 Hz, ... , and 48∼50 Hz. Differential entropy (DE) [4], [5] features are extracted from each frequency band. The calculation formula for differential entropy is:

$$h(X) = \frac{1}{2}\log 2\pi e \sigma^2 \tag{6}$$

Then we applied linear dynamic system to make the DE features smoother.

D. Experiment settings and metrics

In order to make the neural processes method learn the EEG features of all of the 23 subjects' sober and fatigue state, the data are divided into 10 parts without shuffling, of which parts 1, 3, 5, 7, and 9 are used as the training set and parts 2, 4, 6, 8, and 10 are used as the test set.

We selected three evaluation indicators, Pearson's correlation coefficient (PCC), root-mean-square error (RMSE), and coefficient of determination (R^2). Typically, a small RMSE, and a large PCC and R^2 indicate a better performance.

To evaluate the effectiveness and robustness of the proposed neural processes method, we chose support vector machine regression (SVR), Gaussian process with RBF kernel (GP), multilayer perceptron with 5 hidden layers and each layer with 200 units (MLP) and continuous conditional neural field (CCNF) to perform a systematic comparison study on the SEED-IVG[1] dataset.

[1]http://bcmi.sjtu.edu.cn/~seed/download.html

TABLE I
AVERAGE PERFORMANCE AND STANDARD DEVIATIONS OF THREE COMMONLY USED MODELS, CONTINUOUS CONDITIONAL NEURAL FIELD, AND NEURAL PROCESSES

		SVR	GP	MLP	CCNF	NP
PCC	AVG	0.7413	0.5932	0.7614	0.7719	0.7646
	STD	0.0247	0.1165	0.029	0.0371	0.0225
RMSE	AVG	0.1411	0.1714	0.1292	0.1195	0.1271
	STD	0.0021	0.0068	0.0031	0.0017	0.002
R2	AVG	0.403	-0.1545	0.463	0.5046	0.5275
	STD	0.1527	1.6273	0.2464	0.1708	0.055

IV. Experimental Results

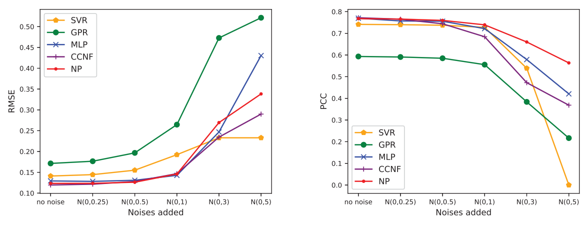
A. Performance without noise

In this section, we evaluated the performance of three conventional models (SVR, MLP, GP) and continuous conditional neural field (CCNF) to compare with the neural processes method by using all data as input without any noise. As we can see in Table 1, the proposed neural processes (NP) method achieves the best results on R^2 which is 0.5275, and achieves close results to the best score on PCC and RMSE.

B. Performance with noise

In real scenarios, the collected EEG signals often contain noises. In order to examine the robustness of the proposed neural processes method, we use two ways to add different proportions of noises. In order to make the experiment more reliable, all data go through a Standard Scaler change, making its value range similarly.

1) Adding Gaussian noises: We add noises that obey Gaussian distribution $N(0, 0.25)$, $N(0, 0.5)$, $N(0, 1)$, $N(0, 3)$, and $N(0, 5)$ to the test data sequentially. The performance of the five models is shown in Fig. 3. The neural processes method has much better robustness while maintaining high accuracy. Even with the addition of 5 times the standard deviation of the input data, our neural processes method can still obtain an acceptable accuracy.



(a) RMSE (b) PCC

Fig. 3. After adding different levels of Gaussian noises, the performance of the five methods on RMSE (a) and PCC (b). The horizontal axis denotes the deviation of the Gaussian distribution added as noise to the test data.

2) Replacing EEG with Gaussian noise: In real scenario, there will be some situations, such as the electrode falling off and causing part of the input to be completely replaced by noise. In order to verify the robustness of the model in this case, we gradually replaced 2, 4, 6, ..., and 16 of the 17 channels with white noise obeying the Gaussian distribution of

$N(0, 1)$. The experimental results are shown in Fig. 4. It can be seen that the neural processes method performs significantly better than other models in this case.

3) Performance with loss input: Since it is expensive and inconvenient to wear all 17 channels in practical applications, we try to gradually reduce the number of channels and investigate whether the neural processes method can still perform vigilance estimation. Since random sampling is involved in the calculation of neural processes, we make ten repeated predictions for each group of inputs (the red line in Fig. 5), and calculate the average of the ten prediction results as the final prediction result (the black line in Fig. 5).

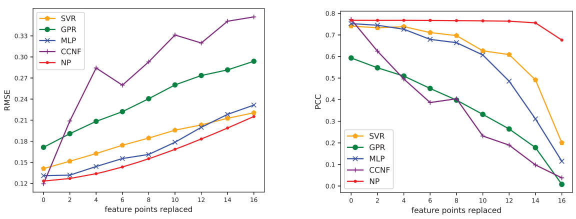
The experimental results are shown in Fig. 5. As the number of input channels decreases, the volatility of neural processes's prediction increases each time, but the average of the prediction can still be very close to the true value (blue line in Fig. 5).

Since the test data of the other four comparison models must maintain the same dimensions as the training data, comparison experiments cannot be performed.

V. Discussion and Conclusion

In this study, we introduced the neural processes method into vigilance estimation from EEG, compared it with three conventional methods, SVR, GP and MLP, and continuous conditional neural field (CCNF) and evaluated the robustness of the neural processes method. The main contributions of this paper are as follows: (1) The neural processes method has better performance than SVR, GP, MLP, CCNF on R^2 for vigilance estimation from EEG. (2) The neural processes method has very good anti-noises ability. (3) Neural processes does not require the test data and training data to have the same dimension.

As can be seen from Fig. 5, even if the inputs of test data contain very little information, the neural processes method can still accurately estimate vigilance. The reason is that neural processes believes that the vigilance level Y is determined by the input X (the EEG signals) and the hidden variables Z which represents the randomness of the EEG signals. During the training process, a random number of random input electrodes are chosen to optimize the distribution of Z in each iteration. Thus, in the prediction process, no matter which of the input electrodes are missing, the distribution of Z can participate in the prediction as the prior and provide a stable and accurate result.



(a) RMSE (b) PCC

Fig. 4. The performance of the five models with different numbers of feature points replaced by noise. The horizontal axis is the number of channels replaced by noise.

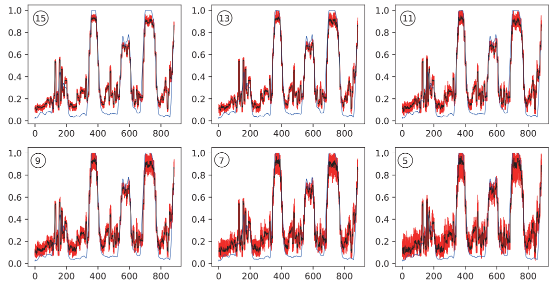


Fig. 5. The performance of the neural processes method when the number of input channels are gradually reduced. The blue line is the PERCLOS value. The red line indicates the prediction result obtained by ten samplings. The black line is the average of ten sampling results. The number in the upper left corner of each picture indicates the number of channels included in the input.

However the neural processes method still has its limitations. As learned from previews studies and also this study, the distribution of the EEG signals varies enormously from subject to subject. In practical applications, it is difficult to train a customized model for all the subjects. Therefore, some transfer learning techniques should be considered for neural processes to be better applied in practical applications.

To sum up, the results of three robustness estimating experiments demonstrated that the neural processes method has a relatively good performance in vigilance estimation and is outstanding in robustness. Due to its structural characteristics, we believe that the neural processes method is a promising approach to practical applications.

Acknowledgements

This work was supported in part by the National Key Research and Development Program of China (Grant2017YFB1002501), the National Natural Science Foundation of China (Grant No. 61673266 and No.61976135), the Fundamental Research Funds for the Central Universities, and the 111 Project.

References

[1] National Safety Concil, https://www.nsc.org/road-safety/safety-topics/fatigued-driving
[2] W. L. Zheng, and B. L. Lu, "A multimodal approach to estimating vigilance using EEG and forehead EOG." Journal of Neural Engineering, vol. 14, no. 2, p. 026017, (2017).
[3] M. Garnelo, J. Schwarz, D. Rosenbaum, F. Viola, D. J. Rezende, S. M. Eslami, and Y. W. Teh, "Neural Processes," ArXiv, abs/1807.01622. (2018)
[4] R. N. Duan, J. Y. Zhu, and B. L. Lu, "Differential entropy feature for EEG based emotion classification," 6th Int. IEEE/EMBS Conf. on Neural Engineering pp 81–84
[5] L. C. Shi, Y. Y. Jiao, and B. L. Lu, "Differential entropy feature for EEG based vigilance estimation," 35th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society pp 6627–30