A Mental Workload Estimation for Visualization Evaluation Using EEG Data and NASA-TLX

Soobin Yim [1], Chanyoung Yoon[1], Sangbong Yoo[1], and Yun Jang [1†]

[1]Sejong University, Seoul, South Korea

Abstract

Mental workload is a cognitive effort felt by users while solving tasks, and good visualizations tend to induce a low mental workload. For better visualizations, various visualization techniques have been evaluated through quantitative methods that compare the response accuracy and performance time for completing visualization tasks. However, accuracy and time do not always represent the mental workload of a subject. Since quantitative approaches do not fully mirror mental workload, questionnaires and biosignals have been employed to measure mental workload in visualization assessments. The electroencephalogram (EEG) as biosignal is one of the indicators frequently utilized to measure mental workload. Since everyone judges and senses differently, EEG signals and mental workload differ from person to person. In this paper, we propose a mental workload personalized estimation model with EEG data specialized for each individual to evaluate visualizations. We use scatter plot, bar, line, and map visualizations and collect NASA-TLX scores as mental workload and EEG data. NASA-TLX and EEG data as training data are used for the mental workload estimation model.

CCS Concepts

• ***Human-centered computing*** → *Visualization design and evaluation methods;* • ***Computing methodologies*** → *Supervised learning by classification;*

1. Introduction

Mental workload is a cognitive effort felt by users while solving tasks. To evaluate the mental workload from visualization, researchers ask users to solve visual analytics tasks through visualizations. Many researchers say that visualization can be evaluated with accuracy and response time because good visualization makes successful completion of given tasks [Loh97, SML*17]. Nevertheless, it is difficult to achieve both high accuracy and low response time at the same time [Pla04]. Therefore, it is necessary to use additional metrics along with response time and accuracy to measure mental workload in visualization evaluation. Generally, questionnaires such as NASA Task Load Index (NASA-TLX) [AH21] have been utilized to measure mental workload. However, participants have the hassle of repeating the evaluation session every time after performing each task during entire the experiment to measure the mental workload. Therefore, researchers examine biosignals such as electroencephalogram (EEG) to avoid questionnaires in the evaluation process [GTL*21, CXL20]. However, as far as we know in visualization evaluation, there is only one study. Anderson et al. [APM*11] study visualization evaluation using mental workload. They employ Extraneous Cognitive Load (ECL) calculated with alpha and the frequency bands in the EEG. Their study is similar to our approach in that the visualization is evaluated using EEG. They estimate mental workload with mathematical analysis on EEG data and evaluate only the boxplot visualization. However, we apply deep learning on EEG data to estimate mental workload and evaluate the scatter, bar, line, and map visualizations. In this paper, we study the mental workload estimation model using EEG in the visualization evaluation. Since there is a difference in individual mental workload when extracting information from data visualizations [LKK17], we propose a mental workload estimation model with EEG data specialized for each individual to evaluate visualizations. We have participants perform visualization tasks to collect EEG data and NASA- TLX scores. After performing the visualization tasks, the participants answer the NASA-TLX questionnaires. In the preprocessing, band power data of EEG are extracted as train data, and the NASA- TLX scores as labeled data are converted to a 10-point Likert workload level using a weighted matrix. Then, we train the model with the train data and label data for mental workload estimation.

2. Experiment Design

In this section, we present the experiment for data collection to train our mental workload estimation model as visualization evaluation. In the experiment, we showed four visualization types, including scatter plot, bar chart, line chart, map for R datasets and

† Yun Jang is the corresponding author. jangy@sejong.edu

DOI: 10.2312/evp.20221117

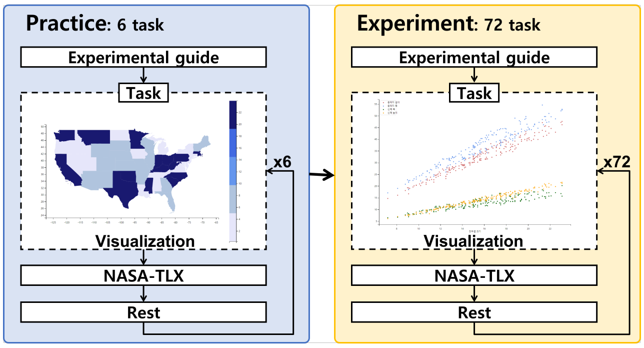

Figure 1: *Experiment procedure. 6 practice tasks and 72 experiment tasks are presented to the participants.*

geographic datasets [pro, osI, PAPB19]. The experimental procedure is demonstrated in Figure 1. The participant is presented with a visualization and a visualization task at the same time. The participant uses the visualization to answer the visualization task and records the answer. While the participant performs the visualization task, the response time is collected. Afterward, the participant takes a NASA-TLX survey and takes a break. After the rest, the next set of visualization and visualization task is given newly. The participant performs a total of 78 visualization tasks, including 6 practice tasks and 72 experiment tasks. During the experiment, EEG data and NASA-TLX score were collected. Since our study is to create a specialized model for each individual, one participant is sufficient for the validation. Nevertheless, we recruited 7 participants to demonstrate that our model works for diverse EEG data from numerous participants, including three male students and four female students. Among the participants, three major in data visualization (Ph.D., MS), and four are undergrads who took a data visualization class.

We employ the NASA-TLX to measure the mental workload. We estimate the weight array by comparing scores for subscales of NASA-TLX and adding weights to subscales with higher scores. After that, the NASA-TLX score is calculated by multiplying the weight array and the subscale scores.

We collected EEG data using the Emotiv epoc+ device [HSb] and the API [HSa] provided by Emotiv company. We removed artifacts in the EEG data with the Emotive API. We placed the Emotiv epoc+ sensors on the participants' heads so that the sensor contact quality is at least 85% or higher in all channels. The band power data for each channel include delta (0.5 $\sim$ 8Hz), theta (4 $\sim$ 8Hz), alpha (8 $\sim$ 12Hz), low beta (12 $\sim$ 16Hz)), high beta (16 $\sim$ 25Hz), and gamma (25 $\sim$ 45Hz). Since the band power data has a wide range of values, it is not suitable for use as an input to a neural network. Therefore, we use MinMax normalization to normalize the dataset into a range from 0 to 1.

3. Mental Workload Estimation Model for Visualization

In this section, we present our mental workload estimation model using only EEG data. The proposed model classifies the mental workload level as 0$\sim$10 with EEG band power data. In the model, we utilize the EEG data preprocessed in Section 2 as input. We also convert the NASA-TLX score calculated in Section 2 to mental workload level and utilize the level as a label. Then, we estimate the mental workload level utilizing with various models, including Support Vector Machine (SVM), which is the most used machine learning model in previous studies, and Deep Neural Network (DNN), Convolutional Neural Network (CNN), and Long-Short Term Memory (LSTM), which are deep learning models.

We perform the classification as a model for estimating mental workload. Since classification is the task of classifying data into appropriate labels, the NASA-TLX score calculated in Section 2 is used as the label. However, the NASA-TLX score is a 100-point Likert scale, and the range is too broad. Therefore, it is difficult to achieve good performance because the number of scores corresponding to each label is small. Hence, the NASA-TLX score is reduced to a 10-point Likert scale, which is used as a label for the classification. Note that the larger Likert scale indicates more mental workload. We measure the F1-scores when applying the test set in the models trained using the train set. The mental workload estimation average accuracies of the models for the 7 participants are 26.22% for SVM, 88.57% for DNN, 82.67% for CNN, and 80.76% for LSTM.

4. Conclusion

In this paper, we proposed a mental workload estimation model for visualization evaluation using EEG data. EEG data and NASA-TLX score measured from 7 participants were preprocessed and used to train models, including SVM, DNN, CNN, and LSTM. The performances were compared with F1 scores, and the DNN model produced the best performance. From this study, we believe it is possible to evaluate visualizations with our proposed model. While EEG data is collected when a participant performs visualization tasks, the mental workload is predicted instantly. Since the EEG signal patterns vary depending on the participants, we trained the model separately with individual EEG data and obtained satisfactory performances. However, there exist differences in the prediction accuracies. Therefore, we plan a study to improve the prediction accuracy of the model by finding the changing patterns of EEG data according to mental workload through additional data collection. Also, we plan to improve the model performance with various EEG data preprocessing methods, such as the corresponding PSD analysis and feature extraction. We also examine more diverse visualization type to distinguish various mental workload levels.

5. Acknowledge

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2022-0-00305, Development of automatic data semantic information composition/expression technology based on augmented analysis for diagnosing industrial data status and maximizing improvement, 50%) and (No.2021-0-00469, Development of heterogeneous fusion data detection and tracking technologies, 50%).

References

[AH21] AMES-HOME T. N.: Nasa-tlx. https://humansystems.arc.nasa.gov/groups/tlx/, 2020 (accessed 28 December, 2021). 1

[APM*11] ANDERSON E. W., POTTER K. C., MATZEN L. E., SHEPHERD J. F., PRESTON G. A., SILVA C. T.: A user study of visualization effectiveness using eeg and cognitive load. *Computer Graphics Forum 30*, 3 (2011), 791–800. 1

[BP12] BALDWIN C. L., PENARANDA B.: Adaptive training using an artificial neural network and eeg metrics for within-and cross-task workload classification. *NeuroImage 59*, 1 (2012), 48–56.

[CXL20] CHEN J., XUE L., LIU Z.: A pilot workload evaluation method based on eeg data and physiological data. In *2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)* (2020), IEEE, pp. 1–6. 1

[GTL*21] GUPTA S. S., TAORI T. J., LADEKAR M. Y., MANTHALKAR R. R., GAJRE S. S., JOSHI Y. V.: Classification of cross task cognitive workload using deep recurrent network with modelling of temporal dynamics. *Biomedical Signal Processing and Control 70* (2021), 103070. 1

[HSa] HARDWARE E. . B. D. M., SOFTWARE: emotivapi. URL: https://emotiv.gitbook.io/cortex-api/ [cited 01 December 2021]. 2

[HSb] HARDWARE E. . B. D. M., SOFTWARE: emotivepoc+. URL: https://www.emotiv.com/epoc/ [cited 28 February 2022]. 2

[LKK17] LEE S., KIM S.-H., KWON B. C.: Vlat: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 551–560. 1

[Loh97] LOHSE G. L.: The role of working memory on graphical information processing. *Behaviour & Information Technology 16*, 6 (1997), 297–308. 1

[osI] OF /SPACETIMECOOR/DATASETS ILDA I.: Map datasets. URL: https://ilda.saclay.inria.fr/spacetimecorr/datasets/ [cited 01 12 2021]. 2

[PAPB19] PEÑA-ARAYA V., PIETRIGA E., BEZERIANOS A.: A comparison of visualizations for identifying correlation over space and time. *IEEE transactions on visualization and computer graphics 26*, 1 (2019), 375–385. 2

[Pla04] PLAISANT C.: The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces* (2004), pp. 109–116. 1

[pro] PROJECTS V. A.-B. G.: Rdatasets. URL: https://vincentarelbundock.github.io/Rdatasets/ [cited 01 December 2021]. 2

[SML*17] SHENHAV A., MUSSLICK S., LIEDER F., KOOL W., GRIFFITHS T. L., COHEN J. D., BOTVINICK M. M.: Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience 40*, 1 (2017), 99–124. 1