Performance Estimation using Deep Learning Based Facial Expression Analysis

Cho Woo Jo, Young Ho Chae, and Poong Hyun Seong*
Dept. Nuclear & Quantum Eng, Korea Advanced Institute of Science and Technology
291, Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea
chowoo0701@kaist.ac.kr *Corresponding author: phseong1@kaist.ac.kr

1. Introduction

Scientists now widely accept that it is important for nuclear accident analysis to consider human error in addition to the failure of safety device. Investigating human factors in nuclear accidents is a continuing concern within the field of nuclear safety and human engineering. To reduce human error and to improve human performance, there have been a number of notable works to estimate operator performance objectively [1]; however, previous work has focused on post accidental analysis from over 50 years ago, and only a limited number of analysis contained the required information.

In the present study we propose facial expression based performance estimation system which solves these problems and provides immediate analysis non-intrusively. The study was conducted in the form of experimental simulation in nuclear accident diagnosis situations, and representative results from the experiment are presented. This work will generate fresh insight into the previous performance estimation system.

2. Methods and Results

In this section, experimental details in nuclear accident analysis are described. Through experiment, nuclear accident diagnosis performance and time sequence of facial expression data were collected at the same time.

2.1 Nuclear Accident Diagnosis

The experiment was subjected 83 students in Korea Advanced Institute of Science and Technology (KAIST). The subjects were to diagnose nuclear accidents in a private room with the help of a human observer.

In order to simulate nuclear accident, compact nuclear simulator was used for experiment. Compact nuclear simulator (CNS) is a nuclear power plant simulator developed by KAERI with the model of Westinghouse 3-loop Pressurized Water Reactor.

Five nuclear accidents were given out of design based nuclear accidents for diagnosis [2]: Loss of coolant accident, Steam generator tube rupture, Loss of feed water accident, and Main steam line break inside and outside of containment. The five nuclear accidents were repeated including instrumentation error to simulate accidents that is unavailable to diagnose [2].

In the experiment, participants were asked to diagnose total of 10 nuclear accidents in CNS screen with and without instrumentation failures. They could have enough time to diagnose accidents and there was no time pressure during the experiment. There were several instrumentations to check to differentiate which nuclear accident occurred. Then their diagnosis results were later scored for performance estimation.

Based on the performance diagnosis results, the participants were divided into high error and low error group. People whose number of correct accuracy in nuclear accident is more than average (8/10) were considered as low error while the others considered as high error group.

2.2 Facial Expression Analysis

Facial expression is a representative of mental and affective states although it lasts less than 4 seconds [3]. Even so, the short time of facial expression changes, which is called micro expressions, contain information of performance impairing stress [3], [4]. Thus, this research used facial expression analysis for performance estimation.

During the experiment, two Logitech web cameras were installed on computer screen (30 frames per second video record), and real-time facial expressions were analyzed by using iMotions software [5]. iMotions is one of automatic facial action unit coding system which provides analyzed data of facial emotions and action units. In this experiment, 7 basic facial emotions, 20 action units around eyes and mouth, and engagement level were analyzed.

As such, facial expression changes over time during accident diagnosis were recorded. To simplify the data containing important facial expression changes, the time range of facial expression was later adjusted around the moment of maximum facial expression changes. We eventually considered 2 seconds (60 frames) of facial expressions around maximum facial movements.

2.3 Performance Estimation System

For our analysis, two acquired data (nuclear accident diagnosis results and facial expression analysis data) were modeled for estimation system using Long Short Time Memory (LSTM). We utilized LSTM which is one of deep learning techniques that upgraded the previous Recurrent Neural Network (RNN). LSTM strengths in time sequence analysis and can remember the states for a long time by using its memory. Fig. 1 represents LSTM model that used in this study for developing facial expression based performance estimation system. As can be seen in Fig. 1, 60 frames of facial expression data were given as input data, then 3 LSTM layers with 60 nodes were used for performance estimation. At the last stage, flattened data with softmax activation layer were used for final binary classification (low error and high error group).

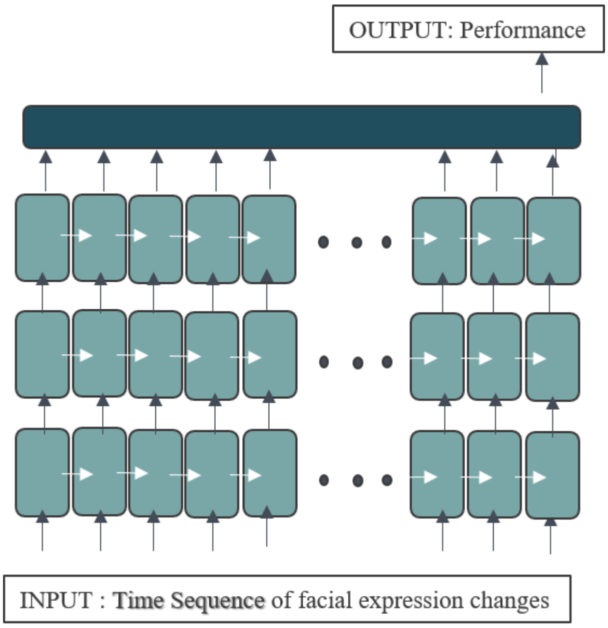


Fig. 1. LSTM based performance estimation system

Before training the estimation model, facial expression data coupled with accident diagnosis performance were split into train and test data. Train data were used to make an optimal performance estimation model with certain hyper parameters while test data were used for checking its model accuracy.

2.5 Performance Estimation per Problems

To solve the imbalanced data set problems, we matched the number of low error and high error group, and discarded several participants' data. This train test split tasks were repeated for total of 10 problems each. Table I shows the number of train and test dataset per problems from 83 participants.

As shown in Table I, train and test dataset size for each problems were different because participants showed different performance results for each accident diagnosis. To be specific, for loss of coolant accident, most participants diagnosed the accident correctly; therefore, not that much data for high error group data existed while many of low error group data were thrown out. However, for loss of feed water accident including instrumentation error, for example, there were similar number of participants of low and high error group; thus, most data were employed for analysis.

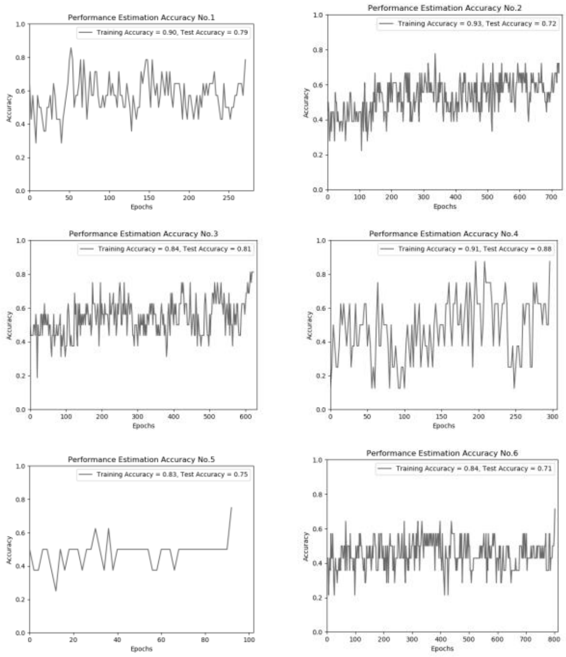
Table I: Problem Train Test Data Split

Problem Number	Train Dataset	Test Dataset
1	40	14
2	56	18
3	44	16
4	22	8
5	24	8
6	44	14
7	28	10
8	44	16
9	40	14
10	20	6

2.5 Performance Estimation Results

We set hyper parameters of the estimation model equally for 10 problems. From hyper parameter tuning, we finally chose certain hyper parameters: Adam optimizer with learning rate 0.0001, binary cross entropy loss function, 10 batch size, 50 cell size, and 3 LSTM layers. Moreover, we set early stopping criteria when train and test accuracy showed more than 70 % of accuracy while train accuracy is higher than test accuracy.

As a result, 10 problems' performance estimation results were obtained. In Fig. 2, the performance estimation accuracy from test dataset are represented.



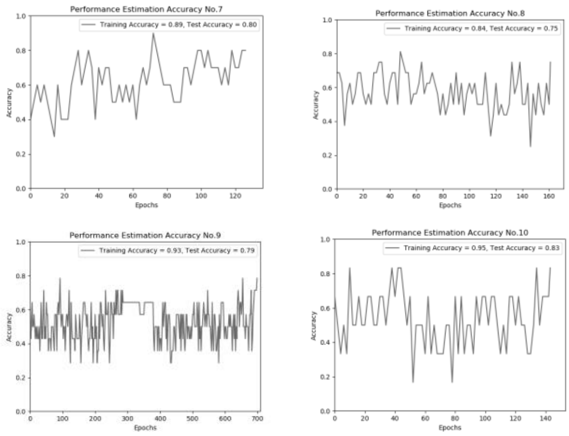


Fig. 2. Performance Estimation Results (10 problems)

As Fig 2 represents, the estimation results were differed from 71% to 88%; however, they showed quiet encouraging results. The early stopped epochs of each model were different for each problems' estimation results, but they were tend to have similar training tendency for each problems.

In general, therefore, it seems that facial expressions contain the information of performance impairing stress [4], and we can differentiate performance from facial expressions regardless of the difficulties of problems. An implication of this is the possibility that nuclear accident diagnosis performance estimation can be applied in other cognitive decision making situations. Moreover, the findings of this study suggest that disregarding the specific problems that participants face, facial expression reflects people's performance.

However, one concern about the findings of this results was that the performance estimation was really simple: low and high error group binary classification. As the number of train and test dataset was small compared to the data complexity (7 basic emotions, 20 action units, and engagement levels), simplifying the classification was inevitable. Thus, it suffers from the same limitations associated with a small number of dataset; the model accuracy has high fluctuations, and the model may not generalize the real situation.

3. Conclusions

Our results are inspiring and should be validated in a larger experimental studies. It is very much the key component in the future attempts to overcome the reliability of facial expression based performance estimation system in a complex situation. Future research should further develop and confirm this basic findings by conducting experiments in another realistic settings.

The present finding confirms that facial expression analysis can be applied in performance estimation. More generally, these basic findings are consistent with research showing that emotion-relevant facial expression reflects biological responses to performance impairing stress [3], [4]. With its non-intrusive and immediate analysis characteristics, facial expression based performance estimation is a desirable for future work.

We expect further application not only in real nuclear accident diagnosis but also in other highly cognitive decision making situations. We believe that apart from looking for post accidental human factor analysis, future research should look for real time human factor analysis like facial expression analysis.

REFERENCES

[1] https://www.oecd-nea.org/brief/brief-02.html
[2] A study on the systematic framework to develop an effective diagnosis procedure of the nuclear power plants, KAERI/TR 2597/2003
[3] https://www.paulekman.com/resources/micro-expressions/
[4] Dinges, D. F., Rider, R. L., Dorrian, J., McGlinchey, E. L., Rogers, N. L., Cizman, Z., ... & Metaxas, D. N. (2005). Optical computer recognition of facial expressions associated with stress induced by performance demands. Aviation, space, and environmental medicine, 76(6), B172-B182.
[5] Senechal, T., McDuff, D., & Kaliouby, R. (2015). Facial action unit detection using active learning and an efficient non-linear kernel approximation. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 10-18).