Chen, W. et al. (Eds.) (2017). Proceedings of the 25th International Conference on Computers in Education.
New Zealand: Asia-Pacific Society for Computers in Education

Extraction of Relationships between Learners' Physiological Information and Learners' Mental States by Machine Learning

Yoshimasa TAWATSUJI[a*], Tatsuro UNO[b], Keita OKAZAKI[c], Siyuan FANG[d] & Tatsunori MATSUI[e]
[a, e]*Faculty of Human Sciences, Waseda University, Japan*
[b]*School of Human Sciences, Waseda University, Japan*
[c]*Comprehensive Research Organization, Waseda University, Japan*
[d]*Graduate School of Human Sciences, Waseda University, Japan*
*y.tawatsuji@aoni.waseda.jp

Abstract: The estimation of learners' mental states during the interaction between teachers and learners is a very important problem in improving the quality of teaching and learning. In this experimental study, we developed a deep learning neural network (DLNN) system that extracted the relationships between a learner's mental states and a teacher's utterances plus the learner's physiological information. The learner's physiological information consisted of the NIRS signals, the EEG signals, respiration intensity, skin conductance, and pulse volume. The learner's mental states were elicited through the learner's introspective reports using the Achievement Emotions Questionnaire (AEQ). According to the AEQ, the learner's mental states were divided into nine categories: Enjoy, Hope, Pride, Anger, Anxiety, Shame, Hopelessness, Boredom, and Others. In a simulation, the DLNN system exhibited the ability to estimate the learner's mental states from the learner's physiological information with high accuracy.

Keywords: Intelligent Mentoring System, physiological data, deep learning, mental state estimation

1. Introduction

1.1 Research Background and Objective

When evaluating the effects of teaching and learning, knowing learners' mental states during their learning processes is vital. Many fundamental studies in educational technology have investigated the relationships between learners' physiological information, such as eye movements and sweating, and learners' behaviors and mental states. Recently, with computers and biometric instruments becoming increasingly higher in performance and lower in cost, collecting a great amount of physiological data and processing them efficiently in real time has become easy. Therefore, many education support systems have been developed that can automatically estimate learners' mental states from learners' behavioral information and physiological information, to a certain extent. On the other hand, it is generally acknowledged in the field of pedagogy that the interactions between teachers and learners can influence learners' mental states. Therefore, it is interesting to formalize the relationships between teachers' behaviors and/or utterances and learners' mental states (or the factors related to learners' mental states). It is expected that the performance of the education support systems can be improved by integrating the formalized knowledge into the modules of the systems that serve to estimate learners' mental states. In our previous study, we tried to formalize the relationships between learners' mental states and teacher's utterances plus learners' physiological information collected during teacher–learner interactions (Takehana and Matsui, 2016). We used the technique of association rule mining to process the teachers' utterances, the learners' physiological information, and the learners' introspective reports. On the other hand, Horiguchi, Kojima, and Matsui (2010), Kojima, Muramatsu, and Matsui (2014), and Fujiyoshi, Yoshimura, Kunze, and Kise (2015) suggested the possibility of applying machine leaning techniques to the development of education support systems. Therefore, in the present study, we employed the technique of deep learning to analyze the data on teacher-learner interactions for the purpose of developing a system that can automatically estimate learners' mental states.

1.2 Synopsis

Section 2 describes the experiment that we conducted to obtain the empirical data that we used to train and test our machine learning systems. In this experiment, we recorded the teacher's utterances and the learners' physiological information and mental states. Section 3 introduces the construction of the first machine learning system by using a three-layer artificial neural network (3LNN). The inputs of the 3LNN are the data of a teacher's utterances and a learner's physiological information. The output of the 3LNN is the learner's mental states. The 3LNN was trained and tested using the experimental data. In Section 4, we redefined the inputs and output of the NN system as time-series data, and increased the number of hidden layers, producing a deep learning neural network (DLNN). The experimental data were used to train and test the DLNN. Section 5 describes the results of the construction of the two NN systems and discusses their significance. Section 6 summarizes this study and introduces some implications for future work.

2. Collection of Multifaceted Learning-Related Data

To collect multifaceted learning-related data, we conducted a biometric experiment. The subjects were one teacher and one learner of a private tutoring school. The learner was a junior high school student taking extra classes in the private tutoring school. We obtained the informed consents of participation from the student and the guardians of the student through the teacher. The physiological information collected in the experiment included near-infrared spectroscopy (NIRS) signals (recorded using Hitachi WOT-100), respiration intensity, skin conductance, and pulse volume (recorded using NeXus). The subject put on all the above instruments when taking a regular class as the experimental task. To align the recording time of these instruments, we placed time markers at the beginning and end of the measurement. The NIRS data were treated using the global average reference method (Nozawa and Kondo, 2009; Hirayama, Watanuki and Kaede, 2012). The entire experiment was recorded by three video cameras set at different locations in the classroom. Based on the video content, we divided the teacher's utterances into nine categories. We then revised part of the categories by referring to the categorization proposed by Fujie (2000), Shimizu and Uchida (2001), and Kishi and Nojima (2006). The final version of our categories is "Explaining," "Questioning," "Comprehension Checking," "Repeating," "Praising," "Task Fulfillment Checking," "Alerting," "Chatting," and "Others." Using these nine categories, we labeled the teacher's utterances while watching the video. After several days, the subject was asked to report the changes in the mental states during the course of the experiment while the subject watched the video. The Achievement Emotions Questionnaire (AEQ) (Pekrun, Goetz, Frenzel, Barchfeld and Perry, 2011) was used to divide the mental states into nine categories: "Enjoyment," "Hope," "Pride," "Anger," "Anxiety," "Shame," "Hopelessness," "Boredom," and "Others." The annotation of the video using the nine mental state categories was carried out on a computer program that we developed. The program allowed the subject to select from among nine buttons that represent the nine mental state categories.

3. Estimation of a Learner's Mental States Using a Three-Layer Neural Network

This section introduces the 3LNN that we constructed in the first try in this study to automatically estimate the learner's mental states from the learner's physiological information that we collected during the experiment. We selected a 63-s segment (19 min 37 s–20 min 40 s) from the 60-min video as the object of the data analysis because this segment contained frequent teacher–learner interactions.

3.1 Data Structure and Neural Network Architecture

The inputs of the 3LNN are 1) the learner's cerebral blood flow measured by NIRS signals (5Hz, 2) the learner's respiration intensity (32Hz), 3) the learner's skin conductance (32Hz), and 4) the teacher's utterances (labeled by the five categories "Explaining," "Questioning," "Comprehension Checking," "Alerting," and "Others"). To regulate the granularity of the input data types 1–3, linear interpolation was applied to the coarse-grained data types according to the granularity of the data type with the finest granularity. The result of this granularity regulation is a 2024*5 data matrix. The five columns represent respectively the five input data types and the learner's mental states. Then, the input data of types 1–3 were normalized to zero mean and unit variance. The 3LNN consists of one input layer, one hidden layer, and one output layer. The number of units in the hidden layer was determined though simulations. In the simulations, we tested the number of hidden-layer units from one to 25, and found that the number 19 had the minimum learning loss and the maximum accuracy. The results of the simulations are shown in Figure 1.

3.2 Methods and Results of System Performance Simulation

A simulation for testing the performance of the 3LNN was run on Python 3.5 with TensorFlow (ver 0.12.1). The activation function of the hidden layer was the tanh function, and the activation function of the output layer was the softmax function. The cost function was the cross-entropy error function. The gradient descent was used as the optimization method. The learning rate was set to 0.05. For cross-validation, 60 percent of the experimental data (totally 2049 data sets) were used as the training data, and the remaining 40 percent were used as the validation data. The learning phase had 5000 iterations. Part of the results of the learning phase is displayed in Figure 2. Figure 2 shows that the learning process converged at an early stage. We performed the cross-validation ten times and the accuracies obtained were 0.900, 0.917, 0.910, 0.917, 0.910, 0.914, 0.915, 0.899, 0.912, and 0.905. This implies that the 3LNN can estimate the learner's mental states from the learner's physiological information with high accuracy.

Figure 1. Changes in learning loss and accuracy that resulted from the adjustment of the number of the hidden-layer units.

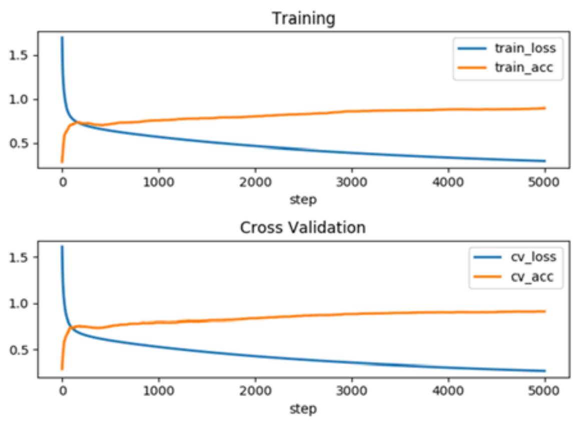

Figure 2. Results of the system performance simulation of the 3LNN.

4. Mental State Estimation Using Deep Learning

In the 3LNN described in Section 3, the temporal aspects of the input and output data were not taken into account. Only the relationships between the physiological information and the mental states at a single time point were discussed. Therefore, our next step is to increase the number of hidden layers in a try to improve the estimation accuracy of the NN system.

4.1 Real-Time Data Processing

As described in Section 3, we used a 3LNN to extract the relationships between the learner's physiological information and mental states that transpired at a single time point. Problems with regard to the following two points remain. First, from the perspective of application, normalizing the input data is difficult. More specifically, when real-time data processing is required, e.g., when developing an Intelligent Mentoring System (IMS), it is hard to apply the global average reference method to the NIRS data and normalize the physiological data. Hence, in the case of IMS development, using the original NIRS data and the original physiological data is necessary. Second, the learner's mental state at a certain time point (t_0) may influence the learner's physiological information and the teacher's utterances. This indicates the possibility that the learner's physiological data and the teacher's utterances after t_0 may also contain the information about the learner's mental states at t_0. This time lag was not considered in the construction of the 3LNN. Hence, we wish to integrate this time lag in the NN construction.

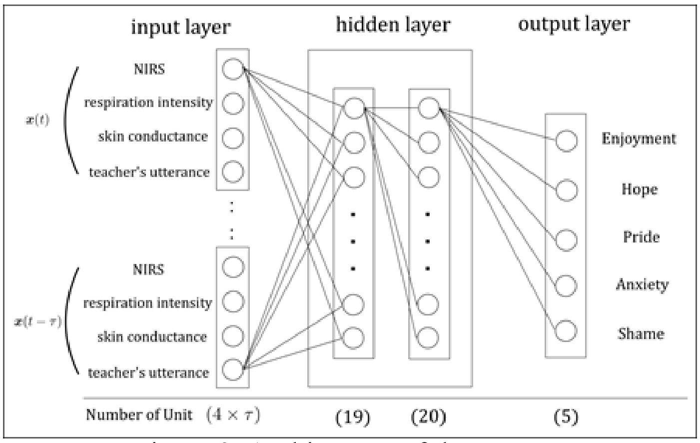

Figure 3. Architecture of the DLNN.

4.2 Data Structure and Neural Network Architecture

The input and output data of this deep learning neural network (DLNN) were the same as those of the 3LNN. However, the physiological data were not normalized in the case of the DLNN. The DLNN consists of one input layer, two hidden layers, and one output layer. The input data are composed of the input data types 1-4 that range from the time point $\tau - t$ to the time point t, and the output data are the mental state at the time point $\tau - t$. In this way, the time lag was integrated in the input and output data structures of the DLNN (shown in Figure 3). As in the case of the 3LNN, the number of the units in the two hidden layers was determined through simulations. In the simulations, we gradually changed the numbers of the hidden-layer units, and found that the DLNN had the best performance in terms of learning loss and accuracy when the first hidden layer had 19 units and the second hidden layer had 20 units. Hence, we set the number of the units in the first hidden layer to 19, and that of the second hidden layer to 20. Figure 4 shows how learning loss and accuracy change when the number of the units in the first hidden layer changes from 10 to 19, with the number of the units in the second hidden layer fixed at 20. Besides, time lags of nine different lengths were tested in the simulations; that is, τ took the values 2, 3, 4, 5, 6, 7, 8, and 9. The results of the simulation show that irrespective of the manner in which we adjusted the number of the units in the first hidden layer, the accuracy always reached its maximum when τ was 9. Hence, in the system performance simulation (described in Section 4.3), τ was set to 9.

Figure 4. Change in learning loss and accuracy that resulted from the adjustment of the number of the units in the first hidden layer from 10 to 19.

4.3 Methods and Results of System Performance Simulation

The performance of the DLNN was tested through a simulation run on Python 3.5 with TensorFlow (ver 0.12.1). The activation function of the two hidden layers was the tanh function, and that of the output layer was the softmax function. The cost function was the cross-entropy error function. The gradient descent was used as the optimization method. The learning rate was set to 0.05. As explained in Section 4.2, the number of the units in the first hidden layer was set to 19, and that of the second hidden layer was set to 20. The cross-validation was conducted ten times. For each time, 60 percent of the experimental data were used as the training data, and the remaining 40 percent were used as the validation data. The learning phase, which sought the minimum cross-entropy, contained 5000 iterations. Part of the results of the learning phase is shown in Figure 5. Figure 5 shows that the learning process converged at an early stage. The accuracies achieved during the cross-validations were 0.799, 0.842, 0.797, 0.793, 0.764, 0.811, 0.789, 0.829, 0.768, and 0.808 (mean = 0.8; sd = 0.023). This implies that the DLNN can estimate the learner's mental states from the learner's physiological information with an accuracy of approximately 80% even without data standardization.

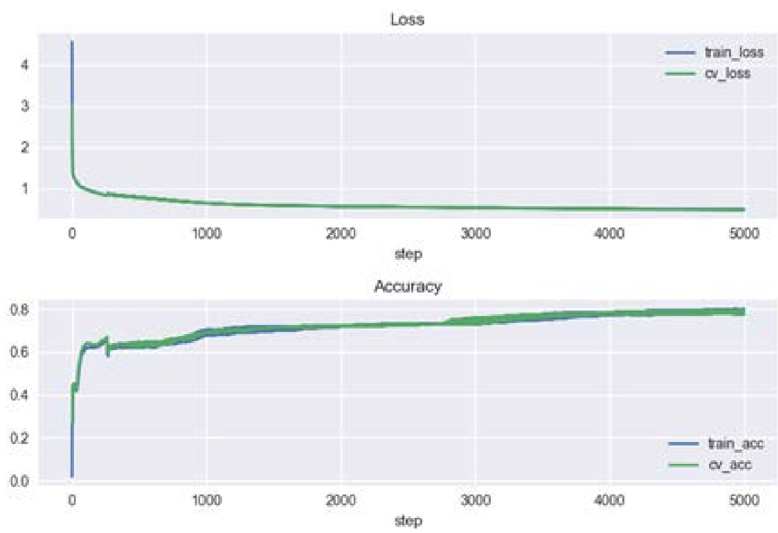

Figure 5. Change in learning loss and accuracy during the 5000 learning iterations.

5. Discussion

We experimentally verified that the DLNN had the ability to estimate the learner's mental states from the learner's physiological information with an accuracy of approximately 80% even without data standardization. One reason for this high performance may be that in the simulation, the possible estimation results were limited to only five mental state categories. Another point to note is that in the simulation, the learning processes converged at a very high speed. This suggests the risk of over-fitting.

6. Summary and Implications for Future Work

In this study, we developed a machine learning system that could automatically estimate a learner's mental states from the learner's multifaceted learning-related information, to a large extent. The physiological information was recorded in an experiment using biometric instruments. We constructed a DLNN to extract the relationships between the learner's physiological information and the learner's mental states, and carried out a simulation to test the performance of the DLNN. The results of the simulation suggest that the DLNN has the ability to estimate the learner's mental states from the learner's physiological information with an accuracy of approximately 80% even without data standardization. Nevertheless, because in the simulation the training and test data contained only a small number of mental state categories, there exists the possibility of over-fitting. To solve this problem, in our future studies, we plan to use the data of the entire video, instead of the 63-s segment that was used in this study. Another way to improve the system performance is to find a better method to optimize the number of the hidden layers and the numbers of the units in each hidden layer. In addition, it is important to understand the relationships extracted by the DLNN in the domains of educational technology and pedagogy. Therefore, in future we will try to read what the hidden layers have learned by analyzing and visualizing the weights in the DLNN.

References

Fujie, Y. (2000). Role of teacher's repetition in classroom teaching, Japan Journal of Educational Technology, 23(4), 201-212.

Fujiyoshi, H., Yoshimura, K., Kunze, K., & Kise, K. (2015). English ability estimation using eye fixation information during English tests, Technical Report of the Institute of Electronics, Information and Communication Engineers, 115(24), 49-54.

Hirayama, K., Watanuki, K., & Kaede, K. (2012). Brain activation analysis of voluntary movement and passive movement using near-infrared spectroscopy, Transactions of the Japan Society of Mechanical Engineers Series C, 78(795), 3803-3811.

Horiguchi, Y., Kojima, K., & Matsui, T. (2010, March). A method to estimate learners' impasses based on features in Low-Level Interactions by using MRA, In Proceedings of the 58th SIG on Advanced Learning Science and Technology of the Japanese Society for Artificial Intelligence, 1-6.

Kishi, T., & Nojima, E. (2006). A structural analysis of elementary school teachers' and children's utterances in Japanese classes, The Japanese Journal of Educational Psychology, 54(3), 322-333.

Kojima, K., Muramatsu, K., & Matsui, T. (2014). Experimental study on description of eye-movements among choices in answering to multiple-choice problems. Transactions of Japanese Society for Information and Systems in Education, 31(2), 197-202.

Nozawa, T., & Kondo, T. (2009, September). Comparison of artifact reduction methods for real-time analysis of fNIRS data, In Proceedings of the 24th Symposium on Biological and Physiological Engineering of the Society of Instrument and Control Engineers, 381-384.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). Contemporary educational psychology, 36(1), 36-48.

Shimizu, Y., & Uchida, N., (2001). How do children adapt to classroom discourse?: Quantitative and qualitative analyses of first grade homeroom activities. The Japanese Journal of Educational Psychology, 49(3), 314-325.

Takehana, K. & Matsui, T. (2016, July). Association rules on relationships between learner's physiological information and mental states during learning process, In Proceedings of HCI International (HCII2016), LNCS 9735, 209-219.