Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference

Adapting to Learners' Mental States Using a Physiological Computing Approach

Maher Chaouachi, Imène Jraidi and Claude Frasson
Département d'informatique et de recherche opérationnelle, Université de Montréal,
2920 Chemin de la Tour, Montréal, CANADA
{chaouacm, jraidiim, frasson}@iro.umontreal.ca

Abstract

In this paper we present a physiological computing approach based on electroencephalogram (EEG) signals to adaptively sequence the learning content according to the learners' mental states. The system draws on techniques from Brain Computer Interface and educational psychology to automatically select the next best learning activity according to changes in the learners' mental states such as attention and workload. The objective of this system is to maintain the learner in a positive mental state throughout the tutoring session.

Introduction

One of the main objectives of Intelligent Tutoring Systems (ITS) is to provide to the user an adapted and individualized learning environment. This adaptation can be operated with regards to several considerations (cognitive, educational, emotional, social, etc.), and can be related to different aspects of the system's interaction strategy (selection of the next learning step, providing an individualized feedback, or help, etc.). The emergence of affective computing over the last years has greatly enhanced the capabilities of these systems to understand the learners' needs and behaviors (Fairclough, 2009; Jraidi et al., 2013a). Research in the field of ITS is increasingly directed towards the integration of new techniques that can provide relevant indicators about the learners' affective states.

Furthermore, a growing body of research in the fields of artificial intelligence, intelligent user interface, psychology and physiological computing has been devoted to modeling and developing systems that are capable of assessing and adapting intelligently to the users' internal states using physiological (Jraidi et al., 2013b; Jraidi et al., 2013c). The aim of these systems is to enhance the user's performance and improve his interaction experience, by continuously analyzing, predicting and adapting to his internal state.

Mental workload and engagement are among the most commonly used indicators to dynamically assess changes in the users' states (Berka et al., 2004; Chaouachi et al., 2010). Several physiological sensors such as heart rate variability, oculomotor activity, pupillometry, body temperature, respiration and galvanic skin responses have been employed to detect mental state changes (Cain, 2007). However, the electroencephalography (EEG) is considered as the only physiological signal that can reliably and precisely track restrained changes in mental attention (or engagement) and workload, and that can be identified and quantified on a millisecond time-frame.

In this paper, we present a new ITS called MENTOR (MENtal tuTOR), which is entirely based on the analysis of the learners' engagement and workload extracted from the EEG data. The system uses these indexes in order to sequence the learning activities accordingly. An experimental study was conducted to evaluate our system. The goal was to verify the following two hypotheses:

1. The integration of the engagement and the workload brain indexes in an ITS can have a real impact on the learners' outcomes.
2. Using a mental state-based strategy in this type of adaptation can improve the learner's experience regarding their tutoring session.

Related work

Developing EEG indexes for workload and engagement assessment is a well-developed research domain. Several linear and non-linear classification and regression methods were used to measure these indexes in different kinds of cognitive tasks such as memorization, language processing, visual or auditory tasks. These methods rely mainly on a frequency processing approach using either the Power Spectral Density (PSD) or Event Related Potential (ERP) techniques to extract relevant EEG features(Berka et al., 2004). Wilson (Wilson, 2005) used an Artificial Neural Network (ANN) to classify operators' workload level by taking the users' EEG data as well as other physiological features as an input. The reported results showed up to 90% of classification accuracy. Gevins et al. (Gevins et al., 2005) used spectral features to feed a neural network classifying the user's workload while performing various memorization tasks. Berka and colleagues developed a workload index using Discriminant Function Analysis (DFA) for monitoring alertness and cognitive workload in different environments (Berka et al., 2007). In the educational context, the index developed by Berka and his colleague was used within a learning environment to analyze the students' behaviors while acquiring skills during a problem solving session (Stevens et al., 2007). In this paper, we propose to extend this approach by presenting a tutoring system whose adaptive strategy is entirely based on the values of the engagement and the workload indexes.

System design

MENTOR is a tutoring system that uses indicators extracted from the EEG physiological data to adjust the learning activities according to the learner's mental state. The system uses the Emotiv EEG headset (www.emotiv.com) to collect EEG raw signals. The reasons of choosing this EEG device is that it can be connected wireless to any machine through the receiving USB. The Emotiv device is also light, easy to use and does not require any particular material configuration. The Emotiv headset contains 16 electrodes located according to the 10-20 international standards. It allows recording simultaneously 14 regions (O1, O2, P7, P8, T7, T8, FC5, FC6, F3, F4, F7, F8, AF3 and AF4). Two additional electrodes are used as references, which correspond respectively to the P3 region (called DRL for Driven Right Leg) and the P4 region (called CMS for Common Mode sense). The system's sampling rate is 128 Hz. Two brain indexes are derived in real-time by MENTOR, namely mental engagement and workload.

Engagement

The engagement index used comes from the work of Pope and colleagues (Pope et al., 1995) at the National Aeronautics and Space Administration (NASA). This work is based on neuroscientific research on attention and vigilance (Pope et al., 1995; Prinzel Iii et al., 2002). It was found that the user's performance improved when this index is used as a criterion for switching between manual and automated piloting mode. This index is computed from three EEG frequency bands: θ (4-8 Hz), α (8-13 Hz) and β (13-22 Hz) as follows.

$$Eng\_Index = (\beta)/(\theta + \alpha)$$

The engagement index is computed each second from the EEG signal. In order to reduce the fluctuations of this index, we use a moving average on a 40-second mobile window. Thus, the value of the index as the time t corresponds to the total average of the ratios calculated on a period of 40 seconds preceding t. The extraction of the θ, α and β frequency bands is performed by multiplying one second of the EEG signal by a Hamming window (in order to reduce the spectral leakage) and applying a Fast Fourier Transform (FFT). As the Emotiv headset measures 14 regions at the same time, we used a combined value of the θ, α and β frequency bands by summing their values over all the measured regions.

Workload and MENTOR's Training Mode

Unlike the engagement index, there is no a common established method to directly assess mental workload from the EEG data. Therefore, we propose to build an individual mental workload predictive model for each learner. This model is trained using data collected from a training phase during which the learner performs a set of brain training exercises. This training phase involves three different types of cognitive exercises, namely: digit span, reverse digit span and mental computation.

The objective of theses training exercises is to induce different levels of mental workload while collecting the learner's EEG data. The manipulation of the induced workload level is done by varying the difficulty level of the exercises: by increasing the number of the digits in the sequence to be recalled for digit span and reverse digit span, and the number of digits to be added or subtracted for the mental computation exercises (see (Chaouachi et al., 2011) for more details about this procedure). After performing each difficulty level, the learner is asked to report his workload level using the subjective scale of NASA Task load index (NASA_TLX)(Hart et al., 1988). Once this training phase completed, the collected EEG raw data are cut into 1-second segments and multiplied by a Hamming window. A FFT is applied to transform each EEG segment into a spectral frequency and generate a set of 40 bins of 1 Hz ranging from 4 to 43 Hz (EEG pretreated vectors). These data are then reduced using a Principal Component Analysis (PCA) to 25 components (the score vectors). Next, a Gaussian Process Regression (GPR) algorithm with an exponential squared kernel and a Gaussian noise (Rasmussen, 2006) is run in order to train a mental workload predictive model (the EEG workload index) from the normalized score vectors. Normalization is done by subtracting the mean and dividing by the standard deviation of the all vectors. In order to reduce the training time of the predictive model, we used the local Gaussian Process Regression algorithm, which is a faster version of the GPR (Nguyen-Tuong et al., 2008).

Analysis of the computed indexes

In order to evaluate the learner's mental state, the system analyses the behavior of the engagement and workload indexes throughout the current learning activity. A slope of each index is computed using the least squared error function of the index's values from the beginning of the activity. For the engagement index, if the slope value is positive, then learner is considered as mentally engaged. Otherwise, the learner is considered as mentally disengaged. For the workload index, if the slope value is between - 0.03 and + 0.03, than the workload is considered as moderate (or fair). Otherwise, if the slope value is above 0.03, the learner is considered as overloaded, and if the slope is below -0.03 the learner is considered as under loaded.

Learning mode

The MENTOR tutoring system has been designed to help learners understand the Reverse Polish Notation (RPN), which is also known as the postfix notation. The lesson presented by the system includes four successive parts. The first part presents a set of formal definitions of the algebraic expressions as well as their structures and constituent elements. The second part explains how to determine the priorities between the operators and how to evaluate an algebraic expression without parentheses. The third part focuses on the concept of the RPN. The basics of the postfix notation are introduced and explained. The fourth part details the techniques used for the assessment of an RPN expression.

After the learner finishes each part of the lesson, the system presents four pedagogical activities so that the learner puts into practice the concepts seen in the previous part of the lesson and enhances his understanding. Each activity uses one of the two following pedagogical resources:

- **Questions:** each question presents a problem that the learner has to resolve. Hints are provided with each problem in order to help the learner find the solution and improve his knowledge acquisition. At the end of each question, the system informs the learner whether his answer was correct or not. In case of a wrong answer, the solution of the problem is given without presenting any explanation of the resolution process.
- **Worked examples:** a worked example describes a problem statement with the detailed steps and explanations leading to the solution. The learner is simply asked to read and understand these examples.

MENTOR's adaptive rules

MENTOR's decisional process lies mainly in the selection of the type of the pedagogical resource (a question or a worked example) to be provided as a next activity. In summary, 16 decisions (4 parts × 4 activities) have to be made by the system according to the learner's mental state.

This choice between worked examples and problems has often been discussed in educational psychology. On one hand, worked examples tend to have a lower mental load impact compared to problems (Paas et al., 2004). Indeed, a worked example provides all the required steps of the problem resolution process. The only effort that a learner has to produce is to understand these steps. On the other hand, the problems are more demanding in terms of mental efforts as the learner has to resolve the problem and in case of a wrong answer, he must also understand the solution.

Providing only worked examples to the learners can have a negative impact. The learner may not identify the relevant information pertaining to the worked example, and focuses rather on useless or secondary information. Another phenomenon that frequently occurs when the learning activities are only based on worked examples is the phenomenon of the illusion of understanding. This phenomenon arises when the learner thinks that he understood the example, while it is not really the case. This generally occurs when the learner browses the elements of the example superficially without producing a minimum effort to understand the goal of each step of the resolution process (Kalyuga et al., 2001).

Besides, presenting a worked example does not guarantee that the learner will be able to generalize from the shown example. Indeed, some learners do not spontaneously engage efforts in analyzing, reproducing and comparing the resolution steps of the example, as compared to the efforts that they would have made if they had to resolve the problem by themselves.

The advantage of using problem solving activities in a learning session is therefore to avoid these risks. The questions are always considered as an efficient educational instrument to assess the learner's knowledge and to help him rapidly acquire new skills. However, using a pedagogical approach exclusively based on solving problems can also hinder the learning process. In fact, as the mental effort is being more important compared to worked examples, the learner can be easily tired and overloaded. Moreover, if the learner fails to solve the problems, he can be frustrated, demotivated and even disengaged from the task.

The decision of presenting a worked example or a problem within MENTOR is based on a continuous analysis of the learner's mental engagement and workload. The goal is to select the pedagogical resource that maintains the learner in a positive mental state. More precisely, the system has to keep the learner mentally engaged and avoid overload and under load. If the system detects a negative mental state caused by an engagement drop, an overload or an under load, it will then try to correct this state by switching the type of the next pedagogical activity.

A total of seven adaptive rules are used by MENTOR:

(1) If the learner's mental state is positive (mentally engaged and neither overloaded nor under loaded), then the system selects a question for the next activity. This rule is applied whatever the current activity is (question, worked example or reading a part of the lesson).

(2) At the end of a question, if the learner's mental state is negative (disengaged, overloaded or under loaded), then the system provides a worked example in the next activity.

(3) At the end of a worked example, if the system detects a negative mental state due to disengagement or under load, then it provides a question as a next activity.

(4) At the end of a worked example, if the system detects a negative mental state due to overload, then it provides a worked example in the next activity.

(5) After reading a part of the lesson, if the system detects a negative mental state due to disengagement or under load, then it provides a question as a next activity.

(6) After reading a part of the lesson, if the system detects a negative mental state due to overload, then it provides a worked example for the next activity.

Whatever the learners' mental state is, if he answers a question incorrectly, then the system provides a worked example in the next activity.

Experimental Study

To highlight the impact of using the learners' mental indicators as an adaptive criterion to manage the system's pedagogical resources, our experimental study relied on two different versions of MENTOR. The difference between these versions lies only in the adaptive logic of the decisional module. The first version leaves intact the adaptive logic with the seven basic intervention rules described previously. The selection of the resource to be provided is done according to the evolution of the learner's mental state. In particular, the system tends to privilege the questions in case of a positive mental state. In the opposite case, the selection of the type of the resource is made following heuristics that aim to correct the learner's mental state.

The second version of the system does not take into account the mental indexes of engagement and workload in selecting the type of the resource to be provided. Only the rule (7) is preserved in the adaptive logic of MENTOR, and the six other rules are ignored. The principle of this version is quite simple: after reading each part of the lesson, the system chooses to ask a question to the learner. As the learner answers correctly, the system continues to adopt the same strategy: asking questions. However, if an incorrect answer is given, the system chooses immediately a worked example as a next activity in order to fix the learner's reasoning. Once the learner finishes reading the example, the system automatically follows up with a question in order to increase his motivation and elicit a problem completion effect. So the unique parameter that can trigger an adaptation action in this version is an incorrect response of the learner.

The two used versions share a common point in their operation: if the adaptation parameters are positive, the two versions opt for a question as a next step. The mental state-based adaptive version of the system (the first) represents then an augmented version of the second, insofar as in addition to considering the accuracy of the response (through the 7th rule); it also applies other adaptive actions based on mental parameters.

In summary, we will compare two versions of the system, the first uses, in its adaptive logic, an analysis of the mental indexes in addition to the response of the learner, and the second is based solely on the response of the learner. Both versions use, in the same order, exactly the same pedagogical resources. That is the two versions will have to choose between the same pairs of resources including a question and a worked example.

Participants and Protocol

14 participants took part in our study. All were students in the University of Montréal in the same certification program in applied computer science. Each participant was randomly assigned to one of the two following groups. (1) The experimental group (N = 7) used the adaptive version of MENTOR: the learning activities are actively adapted to both the learners' brain indexes and answers. (2) The control group (N=7) used the second version of MENTOR that considers only the learners' answers.

For each participant, the experiment was conducted on two successive days. On the first day, the participant uses the training mode of MENTOR in order to create his individual workload model. In this phase, which lasts about an hour, the participant performs a set of 40 brain training exercises including digit span, reverse digit span and mental computation as described earlier.

On the second day of the experiment, the participant uses the learning mode of MENTOR. The duration of this phase is approximately one hour, including 20 to 30 minutes to learn the four parts of the Reverse Polish Notation lesson. The session starts with a pre-test followed by the lesson, then a post-test, and ends with a debriefing.

Pre-test and post-test. These tests use a set of 16 questions relative to the concepts of the lesson. Each of the four parts of the lesson is concerned with four different questions, and the same questions are asked in the pre-test and the post-test. For each question, the learner can answer true or false, or may choose not to respond. The score in each test is calculated as follows: a correct answer is worth 1 point and 0 point for a wrong answer.

Debriefing. During this phase, the learner is first asked to report his appreciation of his interaction with the learning environment by rating his *satisfaction level* regarding the lesson, using a scale of seven grades ranging from 1 (strongly disagree) to 7 (strongly agree) on how much he agrees with the following statement: "*Overall, I am satisfied with of my learning experience with the system*".

Then, the learner evaluates the quality of the tutoring provided by the system by reporting his perceived level of *relevance* of the system's proposed activities, using another scale of seven grades ranging from 1 (strongly disagree) to 7 (strongly agree), on how much he agrees with the following statement: "*Overall, I am satisfied with the learning activities selected by the system. The examples and questions are presented at the right time and helped me to understand the lesson. The choice made between asking a question or presenting an example fits my level of understanding*".

Results and discussion

Learning Performance

A 2 (group: experimental vs. control) × 2 (time: pre-test vs. post-test) mixed-model analysis of variance (ANOVA) was conducted to compare the learners' outcomes of the two groups in terms of scores achieved in both tests. The group variable is a between-subject factor that compares the scores between the two experimental conditions, whereas the time variable is a within-subject factor that analyzes, for each participant individually, the score variation (changes) between the pre-test and the post-test.

First, the analysis yielded a main effect of the time variable, showing a significant difference of the learners' scores in both groups between the pre-test and the post-test: $F(1, 12) = 2253.353$ $p < 0.001$. Thus, there was significant a learning gain regardless of the group, and hence regardless of the version of the system which was used by the participants.

Second the analysis yielded a significant interaction effect of both factors (group × time) on the learners' outcomes: $F(1, 12) = 29.824$, $p < 0.001$. The results revealed that over time, that is between the pre-test and the post-test, the learners of the experimental group got significantly better learning performances compared to the control group. The means of scores obtained in the pre-test and the post-test for the both groups are listed in Table 1.

Table 1. Learners' outcomes in both groups before and after the tutoring session.

	Pre-test	Post-test
Experimental group		
M	4.86_a	13.86_b
SD	1.07	0.70
Control group		
M	3.57_a	10.71_c
SD	1.27	0.95

Values with different subscripts differ significantly.

The comparison of the learners' scores between the experimental group and the control group revealed that there was no statistically significant differences between the two groups in the pre-test: $F(1, 12) = 4.190$, p = n.s. The overall mean score in the pre-test was $M = 4.21$ ($SD = 1.31$). In contrast, the comparison of the learners' scores in the post-test showed that the scores achieved in the experimental group were significantly higher than the control group: $F(1, 12) = 50.069$, $p < 0.001$. The mean score of the experimental group was $M = 13.86$ ($SD = 0.67$) against $M = 10.71$ ($SD = 0.95$) for the control group. These results confirm our first hypothesis, that is using the workload and the engagement indexes as a main criterion to control the user's activities can have a positive impact on his learning performances. The learners' whose pedagogical resources were selected according to their mental states were able to provide an average of 86.6 % correct answers after the tutoring session. An increase of 22.7 % in terms of learning outcomes was achieved using this adaptive strategy.

Subjective Measures

An ANOVA was conducted in order to compare the learners' satisfaction levels between the experimental group and the control group. This ANOVA showed an almost significant difference between the two groups: $F(1, 12) = 4.545$, $p = 0.054$. The learners of the experimental group reported higher satisfaction ($M = 5.71$, $SD = 1.604$) in comparison to the control group ($M = 4.29$, $SD = 0.756$).

A second ANOVA was performed to compare the learners' ratings of the relevance of the activities proposed by the tutoring system in both groups. These ratings were significantly higher in the experimental group ($M = 5$, $SD = 1.414$) versus ($M = 2.43$, $SD = 0.787$) in the control group, $F(1, 12) = 17.673$, $p < 0.05$. These results confirm thus that incresing the system's adaptive logic with the EEG engagement and workload indexes has a positive effect on the users' satisfaction regarding their learning experience in general, and their appreciation regarding the relevance of the decisions taken by the system in the selection of the pedagogical resources more specifically.

Conclusion

In this paper we have presented an intelligent tutoring system called MENTOR (MENtal tuTOR) that adapts its tutoring content according to the user's brain activity. The goal was to show that enhancing the ITS adaptive logic with two physiological mental indicators, namely the engagement and the workload indexes, can improve the learners' outcomes and interaction experience.

The learning mode of MENTOR provides a tutoring environment that adapts its content actively to the learner's brain indexes. The system evaluates the learner's mental state, and selects the pedagogical activity that best suits to his state. An experimental study was conducted to evaluate our system. This study showed the following results: (1) MENTOR can significantly improve the users' performances in terms of learning gains before and after using the tutoring environment, as compared to a control group where a non-adaptive version of the system was used. (2) The mental state-based adaptive logic of MENTOR has a positive influence on the users' interaction experience in terms of satisfaction, and positive reactions when using the tutoring environment.

Acknowledgements

We acknowledge SSHRC (Social Science and Human Research Council) through the LEADS project and NSERC (National Science and Engineering Research Council) for funding this research.

References

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., et al. 2004. Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *International Journal of Human-Computer Interaction*, 17(2), 151-170.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V., et al. 2007. EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. *Aviation, Space, and Environmental Medicine*, 78(5), B231-B244.

Cain, B. 2007. A review of the mental workload literature: DTIC Document.

Chaouachi, M., Chalfoun, P., Jraidi, I., & Frasson, C. 2010. Affect and Mental Engagement: Towards Adaptability for Intelligent Systems. *23rd International FLAIRS Conference*. Daytona Beach, Florida, USA: AAAI Press.

Chaouachi, M., Jraidi, I., & Frasson, C. 2011. Modeling Mental Workload Using EEG Features for Intelligent Systems. In J. Konstan, R. Conejo, J. L. Marzo & N. Oliver (Eds.), *User Modeling, Adaption and Personalization* (Vol. 6787, pp. 50-61): Springer Berlin Heidelberg.

Fairclough, S. H. 2009. Fundamentals of physiological computing. *Interacting with Computers,* 21(1-2), 133-145.

Gevins, A., & Smith, M. E. 2005. Assessing fitness-for-duty and predicting performance with cognitive neurophysiologic measures. Biomonitoring for Physiological and Cognitive Performance during Military Operations. *Proceedings of SPIE* (Vol. 5797-18, pp. 127-138).

Hart, S. G., & Staveland, L. E. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload*, 1(3), 139-183.

Jraidi, I., Chaouachi, M., & Frasson, C. 2013a. A dynamic multimodal approach for assessing learners' interaction experience. Paper presented at the *Proceedings of the 15th ACM on International conference on multimodal interaction.*

Jraidi, I., Chaouachi, M., & Frasson, C. 2013b. A hierarchical probabilistic framework for recognizing learners' interaction experience trends and emotions. *Advances in Human-Computer Interaction*.

Jraidi, I., & Frasson, C. 2013c. Student's Uncertainty Modeling through a Multimodal Sensor-Based Approach. *Journal of Educational Technology & Society*, 16(1), 219-230.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. 2001. When problem solving is superior to studying worked examples. *Journal of educational psychology*, 93(3), 579.

Nguyen-Tuong, D., Peters, J. R., & Seeger, M. 2008. Local Gaussian process regression for real time online model learning. Paper presented at the Advances in Neural Information Processing Systems.

Paas, F., Renkl, A., & Sweller, J. 2004. Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional science*, 32(1), 1-8.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. 1995. Biocybernetic system evaluates indices of operator engagement in automated task. Biological psychology, 40(1), 187-195.

Prinzel Iii, L. J., Pope, A. T., & Freeman, F. G. 2002. Physiological self-regulation and adaptive automation. *The International Journal of Aviation Psychology*, 12(2), 179-196.

Rasmussen, C. E. 2006. Gaussian processes for machine learning.

Stevens, R., Galloway, T., & Berka, C. 2007. Allocation of time, EEG-engagement and EEG-workload resources as scientific problem solving skills are acquired in the classroom. *Proceedings of 3rd Augmented Cognition International*, 22-27.

Wilson, G. 2005. Operator functional state assessment for adaptive automation implementation. Proceedings of SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations (pp. 100–104). Orlando, FL: SPIE: The International Society for Optical Engineering.