Classification of Different Cognitive and Affective States in Computer Game Players Using Physiology, Performance and Intrinsic Factors

Ali Darzi[1], Trent Wondra[2], Sean McCrea[2], and Domen Novak[1]

[1]Department of Electrical and Computer Engineering, University of Wyoming, Laramie, WY, United States
[2]Department of Psychology, University of Wyoming, Laramie, WY, United States
{adarzi, twondra1, smccrea, dnovak1}@uwyo.edu

Abstract. Intelligent systems infer human psychological states using three types of data: physiological, performance, and intrinsic factors. To date, few studies have compared the performance of the data types in classification of psychological states. This study compares the accuracy of three data types in classification of four psychological states and two game difficulty-related parameters. Thirty subjects played nine scenarios (different difficulty levels) of a computer game, during which seven physiological measurements and two performance variables were recorded. Then, a short questionnaire was filled out to assess the perceived difficulty, enjoyment, valence and arousal, and the way the participant would like to change two game parameters. Furthermore, participants' intrinsic factors were assessed using four questionnaires. All combinations of the three datasets were used to classify six aspects of the short questionnaire into either two or three classes using three types of classifiers. The highest accuracies for two-class and three-class classification were 98.4% and 81.5%, respectively.

Keywords: Affective computing · game difficulty adaptation · physiological measurements · task performance · intrinsic factors

1 Introduction

An intelligent cybernetic system can use affective computing techniques to infer the user's affective and cognitive states, then modify its behavior accordingly to ensure a positive user experience. For example, this approach is frequently used in computer games to intelligently adapt game difficulty to suit the player's mood and ability, thus providing a pleasant gameplay experience [1, 2]. The user's affective and cognitive states can be assessed using three types of data: physiological measurements, task performance, and intrinsic factors [3].

Physiological measures from either the central or peripheral nervous system can be used to quantitatively estimate psychological states in real time (during the task itself) without the user's active participation. They include the electroencephalogram (EEG), [4] which records the electrical activity of the brain, electrocardiogram (ECG) [5], which monitors electrical activity of the heart (specifically heart rate), galvanic skin response (GSR) [5], which records the activity of the skin's sweat glands, skin temperature, respiration rate [6], and eye movement. All of the above physiological signals were also analyzed in this study. **Task performance** is a task-specific concept and is thus not as generalizable as physiology, but is also frequently used to assess psychological states; in the case of games such as Pong, it is often defined simply as the in-game score [7]. In this study, participants' in-game score was recorded as well. Finally, a participant's **intrinsic factors** such as personality can provide significant information, but are generally combined with physiology or task performance to classify cognitive and affective states. In this study, participants were asked to fill out four questionnaires that assessed several intrinsic factors such as extraversion.

Playing a computer game may evoke several complex affective and cognitive states, such as mental workload [2], enjoyment, anger, hate, and love [6]. Alternatively, several studies have used two-dimensional models of emotion such as the valence-arousal system to asses in-game emotion [8]. In contrast, however, when designing systems that adapt game difficulty based on cognitive and affective states, most researchers only focus on a single psychological aspect such as perceived task difficulty. Thus, few studies have compared the ability of affective computing techniques to classify multiple different psychological states within the same game.

This study examines the accuracy of three different input datasets (physiology, performance, intrinsic factors) for classification of four different cognitive and affective variables (perceived difficulty, enjoyment, valence, arousal) and two different desired changes to game difficulty (ball speed, paddle size) in a computer-based game of Pong. Accurate classification of psychological states is perhaps the most critical scientific challenge of any game difficulty adaptation algorithm, and can be done using automated classification algorithms such as support vector machines (SVM), linear discriminant analysis [4]. Our ultimate goal is to have the computer game react to these states and adapt its difficulty to ensure the optimal game experience or the user; however, as the first step, this paper is limited to offline classification of psychological states. The objective is to find the most informative features for psychological state classification.

2 Materials and Methods

Study Setup: In the study, we evoked different cognitive and affective states in 30 healthy university students (24.2 ± 4.4 years old, 11 females) using different difficulty levels of a computer game that was reused from our previous arm rehabilitation study [7]. It is a Pong game consisting of two paddles and a puck on a board (Fig. 1, left). The bottom paddle is controlled by the participant while the top paddle is s controlled by the computer. If the puck passes a player's paddle and reaches the top or bottom of the screen, the other player scores a point and the puck is instantly moved to the middle of the board, where it remains stationary for a second before moving in a random direction. The game difficulty can be adjusted using two parameters: the ball speed and the paddle size (with the paddle size being the same for both paddles at all times). The player moves their paddle left and right by tilting the Bimeo (Kinestica, Slovenia) arm rehabilitation device (Fig. 1, right) left and right.

Measured Data: For classification of the different affective and cognitive states, three types of data were collected: game performance, physiology, and intrinsic factors. The performance dataset includes two game performance measures: in-game score and the amount of arm movement, which is recorded by the Bimeo. To monitor the impact of intrinsic user factors on performance and physiology, four questionnaires were filled out: the learning and performance goal orientation measure [9], behavioral inhibition/activation scales [10], a self-efficacy scale [11], and a Big Five personality measure [12]. Two g.USBamp signal amplifiers and associated sensors (g.tec Medical Engineering GmbH, Austria) were used to record six types of physiological signals: 8-channel EEG, 2-channel electrooculogram (EOG), ECG, respiration [13], GSR, and ST. All physiological signals were sampled at 256 Hz. The EEG channels were recorded from prefrontal, frontal and central areas of brain based on the 10-20 placement system [14]: AF3, AF4, F1, F2, F5, F6, C1, and C2. As EEG signals are severely affected by eye activity, a 2-channel EOG was recorded to not only provide more physiological information but also to use as a reference signal with which to denoise the EEG signals. One EOG channel reflected up-down movement while the other one reflected left-right movement of the eyes. To record the EOG, small ECG electrodes (Kindall) were placed according to suggestions in the literature [4]. Finally, a seventh physiological signal (point of gaze on the screen in two dimensions) was recorded using an eye tracker (Gazepoint, Canada).

Study Protocol: The study protocol started with a 2-minute baseline recording of physiological signals, during which participants did not do anything and were instructed to relax. The main part of experiment then consisted of nine trials (test periods), each two minutes long. The nine trials consisted of all possible combinations of ball speeds (slow, medium, fast) and paddle sizes (small, medium, large), played in random order. After each trial, a short questionnaire was filled out to assess six parameters: perceived difficulty (1-7), enjoyment (1-7), valence (1-9, with 1 being very positive and 9 being very negative), arousal (1-9), desired changes to ball speed (-2 to 2, where -2 means decrease by 2 levels), and desired changes to paddle size (-2 to 2, where 2 means increase by two levels). It should be noted that the order of difficulty settings was preset, and that the participant's desired changes to the ball speed and paddle size were not actually used to adapt difficulty.

Contribution of this study: The perceived difficulty, enjoyment, valence and arousal obtained from the questionnaires were classified into either two (low/high) or

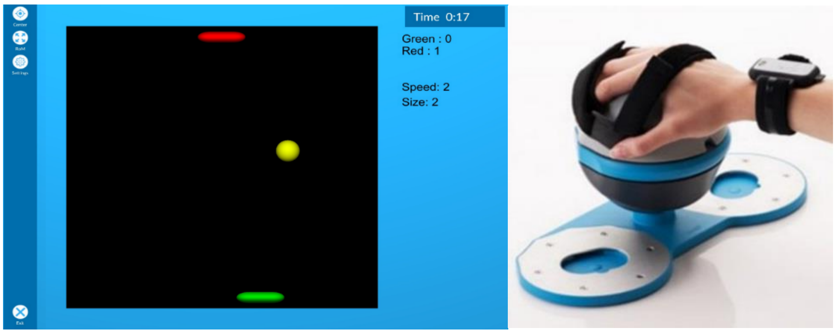

Fig. 1. The Pong game (left) and the BIMEO device (right).

three (low/medium/high) classes based on all combinations of the three recorded datasets (performance, physiology, intrinsic factors). All three datasets also included the current ball speed and paddle size. For 2-class classification, the class "low" was defined as 1-3 for all categories while the class "high" was defined as 5-7 for perceived difficulty and enjoyment or 7-9 for valence and arousal. In 3-class classification, the-low, medium and high ranges were 1-2, 3-5 and 6-7 for perceived difficulty and enjoyment; they were 1-3, 4-6 and 7-9 for valence and arousal. Similarly, the participants' desired changes to game difficulty settings were mapped into either two (increase /decrease) or three (increase/no change/decrease) classes. For both ball speed and paddle size, 1 and 2 were mapped to the class "increase" while -1 and -2 were mapped to "decrease". For two-class classification, the "no change" class was dropped.

As a basis for classification into two or three classes, we first used the stepwise feature selection algorithm [13] to find the most informative set of features. Then, three different classifiers (SVM with a linear kernel, decision tree, or ensemble decision tree) were used to classify combinations of the different datasets (performance, physiology, intrinsic factors) into two or three classes for each of the six possible outcome variables (perceived difficulty, enjoyment, valence, arousal, desired ball speed and paddle size change) separately. The classifiers were validated using 10-fold crossvalidation method using all selected data points (independent from participant).

3 Results

Table 1 presents the 2-class classification accuracies for all combinations of the input datasets. The highest accuracy is obtained for classification of desired changes of ball speed using physiological measurements. For the other five classification cases, the combination of all datasets yields the most accurate classifier, with the lowest classification accuracy (86.9%) obtained for perceived difficulty.

Table 2 presents the 3-class classification accuracies for all combinations of the input datasets. The highest accuracy is obtained for emotional valence using the combination of all datasets. Physiology yielded the most accurate classifier for three of the six classification cases; the other three classification cases, the combination of all three datasets yielded the highest classification accuracy. The lowest classification accuracy was obtained for the desired paddle size change.

Table 1. Two-class classification accuracies all combinations of datasets. If the classification method is not mentioned, the support vector machine was used. (Ph: Physiology, In: Intrinsic factors, Pe: Performance, *: Ensemble decision tree is used)

Classification cases	Ph	In	Pe	Ph & In	Ph & Pe	In & Pe	All
Difficulty level	85.8%	*83.5%	*79.1%	85.7%	86.6%	*83.0%	**86.9%**
Enjoyment	**87.8%**	*80.9%	73.3%	85.7%	85.4%	*81.7%	86.5%
Valence	89.2%	*91.1%	88.8%	92.8%	94.9%	*93.0%	**93.9%**
Arousal	89.0%	*87. %	76.8%	89.8%	87.7%	*86.3%	**89.4%**
Speed change	**98.4%**	*97.2%	92.0%	96.5%	97.9%	*95.3%	96.6%
Paddle size Change	**98.3%**	92.2%	91.4%	97.5%	98.8%	92.4%	97.8%

Table. 2. Three-class classification accuracies for all combinations of datasets. If the classification method is not mentioned, the support vector machine was used. (Ph: Physiology, In: Intrinsic factors, Pe: Performance, *: Ensemble decision tree is used)

Classification cases	Ph	In	Pe	Ph & In	Ph & Pe	In & Pe	All
Difficulty level	76.6%	*70.0%	65.2%	77.4%	77.9%	*71.1%	**81.1%**
Enjoyment	68.8%	*69.1%	51.8%	70.1%	65.5%	64.8%	**71.4%**
Valence	70.7%	*67.0%	55.6%	75.6%	73.3%	*66.3%	**76.2%**
Arousal	**68.8%**	*67.9%	54.8%	67.8%	66.7%	*63.3%	66.7%
Speed change	80.0%	*77.0%	*70.7%	80.0%	77.4%	*78.5%	**81.5%**
Paddle size change	74.1%	*72.2%	55.9%	75.6%	72.2%	73.0 %	**78.1%**

4 Discussion

The obtained results compared the classification accuracy of game players' psychological states using all combinations of physiological signals, performance, and intrinsic factors. The physiological dataset was the most informative of the three individual datasets, dataset of the three, and the combination of all three datasets yielded the best accuracy for 8 of the 12 classification cases. As the classifiers are highly accurate, our next step will be to use them in a real-time manner: the participant's psychological state will be classified, and the game will then adapt its difficulty in a way that is expected to increase player motivation. Since the classifiers are not computationally demanding, a real-time version of the classification procedure is feasible.
Prior to real-time implementation, the training dataset should be expanded to include more than three possible discrete values of ball speed and paddle size, thus allowing the psychological state classification to also be useful for very high and very low difficulties. Furthermore, it may be possible to further increase classification accuracy and improve the user experience by including a history of previous difficulty levels and psychological states that they evoked, thus allowing the computer to estimate how participants reacted to certain difficulty levels in the past.

5 Conclusions

In this study, three sets of classifiers are used to classify four affective/cognitive parameters of Pong game players into either 2 or 3 classes. The proposed classifiers can also determine how participants would like to change the game difficulty to make it more fun. Three data sets (physiological signals, game performance, and intrinsic factors are used as the input of the classifiers. Among the 2-class classifiers, the highest accuracy was obtained for desired ball speed change (98.4%) while the lowest was obtained for perceived difficulty level (86.9%). Among the 3-class classifiers, the highest accuracy was obtained for desired ball speed change (81.5%) while the lowest was obtained for psychological arousal (68.8%). As the next steps, additional improvements will be made to increase the classifiers' robustness, and the classifiers will then be used to adapt game difficulty in response to players' psychological states, thus improving the gameplay experience.

Acknowledgment. Research supported by the National Science Foundation under grant no. 1717705 as well as by the National Institute of General Medical Sciences of the National Institutes of Health under grant no. P20GM103432.

References

1. Tan, C.H., Tan, K.C., Tay, A.: Dynamic Game Difficulty Scaling Using Adaptive Behavior-Based AI. IEEE Trans. Comput. Intell. AI Games. 3, 289–301 (2011).
2. Zhang, X., Lyu, Y., Hu, X., Hu, Z., Shi, Y., Yin, H.: Evaluating Photoplethysmogram as a Real-Time Cognitive Load Assessment during Game Playing. Int. J. Human–Computer Interact. 34, 695–706 (2018).
3. Darzi, A., Gaweesh, S.M., Ahmed, M.M., Novak, D.: Identifying the Causes of Drivers' Hazardous States Using Driver Characteristics, Vehicle Kinematics, and Physiological Measurements. Front. Neurosci. 12, (2018).
4. Ma, J., Zhang, Y., Cichocki, A., Matsuno, F.: A Novel EOG/EEG Hybrid Human–Machine Interface Adopting Eye Movements and ERPs: Application to Robot Control. IEEE Trans. Biomed. Eng. 62, 876–889 (2015).
5. Rodriguez-Guerrero, C., Knaepen, K., Fraile-Marinero, J.C., Perez-Turiel, J., Gonzalez-de-Garibay, V., Lefeber, D.: Improving Challenge/Skill Ratio in a Multimodal Interface by Simultaneously Adapting Game Difficulty and Haptic Assistance through Psychophysiological and Performance Feedback. Front. Neurosci. 11, (2017).
6. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence: analysis of affective physiological state. IEEE Trans. Pattern Anal. Mach. Intell. 23, 1175–1191 (2001).
7. Goršič, M., Cikajlo, I., Novak, D.: Competitive and cooperative arm rehabilitation games played by a patient and unimpaired person: effects on motivation and exercise intensity. J. Neuroeng. Rehabil. 14, 23 (2017).
8. Reuderink, B., Mühl, C., Poel, M.: Valence, arousal and dominance in the EEG during game play. Int. J. Auton. Adapt. Commun. Syst. 6, 45 (2013).
9. Kim, T.T., Lee, G.: Hospitality employee knowledge-sharing behaviors in the relationship between goal orientations and service innovative behavior. Int. J. Hosp. Manag. 34, 324–337 (2013).
10. Carver, C.S., White, T.L.: Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. J. Pers. Soc. Psychol. 67, 319–333 (1994).
11. Hsia, L.-H., Huang, I., Hwang, G.-J.: Effects of different online peer-feedback approaches on students' performance skills, motivation and self-efficacy in a dance course. Comput. Educ. 96, 55–71 (2016).
12. Gosling, S.D., Rentfrow, P.J., Swann, W.B.: A very brief measure of the Big-Five personality domains. J. Res. Pers. 37, 504–528 (2003).
13. Darzi, A., Gorsic, M., Novak, D.: Difficulty adaptation in a competitive arm rehabilitation game using real-time control of arm electromyogram and respiration. In: 2017 International Conference on Rehabilitation Robotics (ICORR). pp. 857–862. IEEE (2017).
14. Klem, G.H., Lüders, H.O., Jasper, H.H., Elger, C.: The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. Electroencephalogr. Clin. Neurophysiol. Suppl. (1999).