Utilizing Machine Learning to Predict Flow State from Gameplay Interaction Data

1st Hsuan-Min Wang
Institute of Computer Science and Engineering
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
hmwang.cs04@nycu.edu.tw

2nd Kuan-Chun Hong
Institute of Data Science and Engineering
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
a0979478979@gmail.com

3rd Chuen-Tsai Sun
Institute of Computer Science and Engineering
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
ctsun@cs.nctu.edu.tw

2023 IEEE Gaming, Entertainment, and Media Conference (GEM) | 979-8-3503-1540-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/GEM59776.2023.10390248

***Abstract*—Flow, a profound psychological state associated with optimal experiences, serves as a pivotal gauge of engagement across various contexts. Traditionally evaluated through methods like surveys, interviews, and sampling surveys for qualitative analysis, these approaches aimed to decipher participants' experiences and perceptions. However, these methods are time-intensive and prone to memory and expressive limitations. In response, contemporary research increasingly leans towards physiological indicators for assessing flow. Utilizing physiological signals to detect flow reduces narrative inaccuracies and minimizes sensitivity to participants' subjective awareness, albeit demanding significant time, human resources, and specialized equipment. To surmount these challenges, this study introduces an innovative methodology for efficiently predicting flow states. utilizing gameplay and interaction data as inputs for machine learning models. By employing a real-time strategy game as the experimental environment, participants' gameplay recordings and interaction records are collected and paired with the Flow Short Scale questionnaire to establish a predictive model for flow state. The results demonstrate the success of this approach, achieving a significant prediction accuracy (MAE = 0.0623) and highlighting a strong correlation between objective gameplay records and subjective flow experiences. This streamlined methodology offers a promising avenue for quantifying and predicting flow state, contributing to a deeper understanding of engagement dynamics in digital environments.**

***Keywords*—Flow, Digital game application, Machine learning, Interactive device application.**

I. Introduction

The rapid evolution of information technology has ushered in an era of abundant high-performance, high-tech products. However, the sole emphasis on functionality has gradually waned. In today's landscape, consumers not only seek product features but also value the overall experience they provide. As a result, companies are increasingly pivoting towards user experience, employing various research methods to comprehend users' emotions and perceptions towards their offerings [1]. This paradigm, centered around user habits and preferences, is known as user-centered design. To gauge user experiences, past methodologies encompassed questionnaires, interviews, and experience sampling [2]. While these methods yield comprehensive and accurate insights, they lack the element of real-time feedback. Moreover, they are prone to individual memory bias and subjective judgments. Experience sampling, in particular, may disrupt the user's interaction flow and provide only fragmented insights. Furthermore, personal experiences can influence users' expressions, potentially masking their genuine feelings. Consequently, researchers are diligently exploring avenues to seamlessly grasp user experience without intruding on their interactions.

Recognizing that distinct emotions and experiences can trigger specific physiological signals or responses, the acquisition of immediate physiological data can illuminate users' experiences without disruption. This approach furnishes real-time information that is less vulnerable to individual consciousness or manipulation. Notably studied physiological signals encompass brainwaves, electrodermal responses, electrocardiograms [3], and heart rate variability [4]. Research has confirmed the high correlation between alpha and beta brainwave frequencies and cognitive control and immersion [5][6]. Emotional responses can manifest in changes in skin resistance [7], and scholars have employed eye-tracking and pupillometry to scrutinize human visual behavior [8]. Noteworthy successes have been achieved through studies rooted in physiological signals or responses, with applications ranging from dynamically adjusting game difficulty [9] and enhancing real-time learning experiences [10], to gauging developers' states during the development process [11] and assessing job satisfaction [12]. Currently, a burgeoning field of research seeks to enhance experiences [13] based on users' physiological signals. Despite the burgeoning literature in this domain, conducting such research entails significant time, manpower, and additional experimental equipment. This has limited the widespread adoption of experiments in diverse contexts, particularly within digital products catering to vast user bases numbering in the hundreds of thousands [14]. Familiarity with the experimental apparatus among the public could enhance participants' willingness to partake, consequently yielding more authentic outcomes.

In summary, our study opts for widely recognized experimental tools, the mouse and keyboard to enable participants to engage in the experiment under the most natural conditions. We select a real-time strategy game that necessitates extensive mouse and keyboard usage to swiftly establish an immersive experience as the experimental backdrop. Through the collection of data from mouse and keyboard interactions, as well as in-game activities, we construct a model capable of classifying and exploring various levels of user engagement. This research aims to contribute to the understanding and enhancement of user experiences in digital environments.

II. Literature Review

A. Flow Theory and Games

Flow theory, introduced by psychologist Csikszentmihalyi in 1975, integrates motivation, personality, and subjective experience. It characterizes a mental state where individuals become fully absorbed and deeply focused on an activity, resulting in heightened efficiency and creativity. In this state, individuals often lose track of time, overlook basic needs like hunger, and disregard unrelated sensory signals, leading to a profound sense of excitement and fulfillment [2].

Games, as designed activities, naturally lend themselves to fostering the flow experience. Consequently, researchers have long explored methods to predict a player's state of flow during gaming. Numerous studies have aimed to assess players' gaming conditions by examining indicators such as game outcomes, eye movements, and physiological signals..

B. Mouse, Keyboard, and Flow

At present, research utilizing input data from mouse and keyboard for flow analysis is relatively limited compared to studies based on physiological signals or self-reports. Zheng et al. [15] collected mouse and keyboard data from developers during their interactions with integrated development environments (IDEs) and systems to assess their non-intrusive work states. The study involved six participants, and data was collected over 17 days in 30-minute intervals. The preprocessed data was categorized into four flow-related aspects (Concentration, Involvement, Enjoyment, and Skill-Challenge Balance) using Random Forest classification. By inputting the resulting four-dimensional feature vectors into a Support Vector Machine (SVM), the study determined whether participants were in a positive or negative flow state.

III. Methods

A. Research Process

For this study, we employed commonly used input devices, namely the mouse and keyboard, as tools for detecting flow. We selected a real-time strategy (RTS) game with intricate events and demanding operations as the experimental platform. To create a balanced flow model, we designed three distinct difficulty levels of StarCraft maps for the participants, consisting of 37 students from National Yang Ming Chiao Tung University and Tsinghua University, all of whom had prior experience with RTS games.

To ensure a harmonious blend of challenge and skill within the flow model, participants engaged in three games of varying difficulty levels on custom-made maps. The difficulty was dynamically adjusted based on each player's in-game performance, inducing variations in the flow experience. We recorded the gameplay process and effective player behaviors using StarCraft's built-in recording feature. After each gaming session, participants completed the Flow Short Scale questionnaire to assess their flow experience. Subsequently, the collected data underwent preprocessing, and regression models were employed to predict the extent to which players entered the state of flow.

B. Research Tools

a) Custom-Made StarCraft Maps

The game maps employed in this study were crafted using StarCraft's official map editor. The objective of the game was for players to control their army and defeat the enemy camp. At the start of each game, participants received concise instructions. To elicit different levels of flow experience, participants engaged in three distinct versions of the game experiment, with variations limited to numerical values such as enemy unit damage and production speed, while other gameplay elements remained consistent.

b) Player Gameplay Records

The StarCraft game system featured an integrated mechanism for automatically recording all valid player actions in log files. Upon completing a game, the system autonomously recorded and stored the process as a file with the extension "SC2Replay" locally. Players could review their previous gameplay using the in-game replay player or utilize Python packages to convert detailed information, including game time, player names, ability usage, unit commands, camera movements, and click locations, into easily interpretable text files for subsequent analysis and review of the gameplay process.

c) Flow Short Scale

To ascertain whether players achieved a state of flow during gameplay, the Flow Short Scale [16] was employed in this experiment. This questionnaire subdivides flow into ten distinct psychological aspects, each accompanied by a corresponding question. Participants employed a 7-point Likert scale to respond to these questions. Additionally, to evaluate the balance between challenge and skill, three supplementary questions were included to address the potential underestimation of flow at low and high levels, as well as the distortion of flow questionnaire responses due to personal memory biases.

d) Establishing Prediction Models

In this experiment, player gameplay records are divided into segments of 480 frames (30 seconds) each. The player's actions are categorized into 12 different items as features for the gameplay records model. Additionally, the player's actions using the mouse and keyboard are separately classified as features for the player operation model. For the gameplay records model, six regression models (Linear Regression, Random Forest, Support Vector Machine, Ada Boost, K-Nearest Neighbor, XGBoost) provided by the scikit-learn library are used for training. For the player operation model, three models (Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU)) also provided by the scikit-learn library are employed for training. Finally, the training results are evaluated and compared to assess the accuracy of each model. The best-performing model will be selected for prediction.

IV. Results

A. Models Derived from Gameplay Records

a) Regression Model Training

To construct predictive models from gameplay records, we adopted a stratified approach, splitting the feature data of gameplay records into training and testing sets, with a testing set proportion of 0.2. Subsequently, we evaluated six models, with noteworthy performance exhibited by the XGBoost Regressor and Random Forest models, both achieving errors below 0.1 (as shown in Figure 1).

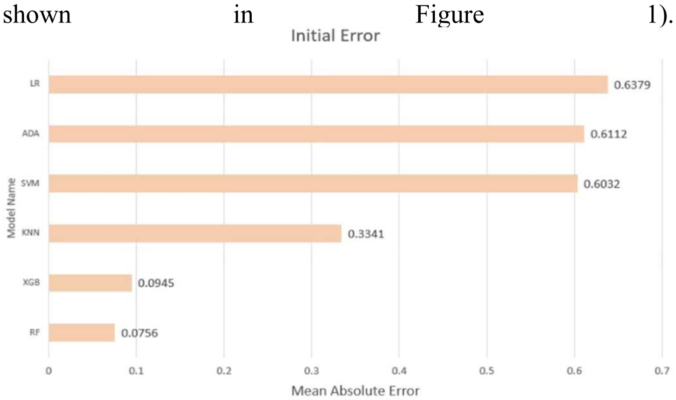

Fig. 1. Initial Performance of Regression Models

b) Parameter Tuning and Comparison of Regression Models

In our pursuit of enhanced predictive accuracy, we conducted parameter tuning using the GridSearchCV function from the scikit-learn library. This function systematically performed parameter tuning for each model after configuring the desired parameter options. Following ten rounds of cross-validation, we observed that the Random Forest and XGBoost Regressor models continued to excel (Figure 2).

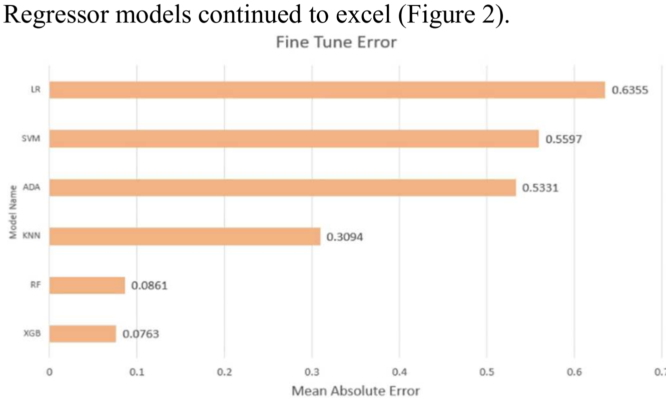

Fig. 2. Performance of Regression Models after Parameter Tuning

c) Prediction Outcomes

Following parameter tuning, we evaluated the models using the initial test dataset. The XGBoost model demonstrated exceptional performance. The graph below represents the complete prediction results of the XGBoost Regression model. The horizontal axis denotes the number of data points, while the vertical axis signifies the flow experience. The blue section corresponds to the actual flow experience of the participants (Truth), and the red section represents the prediction outcomes of the XGBoost model (Predict). Notably, the model displayed remarkable accuracy, with only a few data points deviating slightly from the predicted values. Furthermore, the Mean Absolute Error (MAE) was impressively low at 0.0623 (Figure 3).

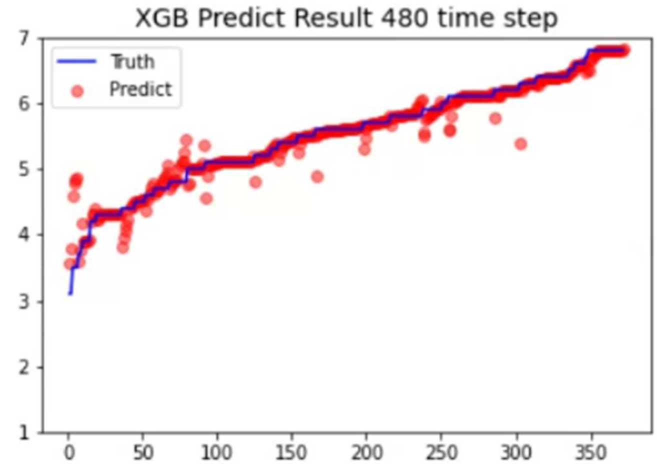

Fig. 3. XGBoost Prediction Results

B. Models Constructed Using Player Action Records

a) Time Series Model Training

In this experiment, three types of time-series models provided by the scikit-learn library will be employed for modeling: The Recurrent Neural Network (RNN) model, Long Short-Term Memory (LSTM) model, and Gated Recurrent Unit (GRU) model. To ensure that the training models align more closely with the characteristics of neurons and to prevent issues like gradient vanishing and overfitting in neural networks, various activation functions will also be employed during the model creation process.

b) Prediction Results

To minimize prediction errors arising from data splitting, we employed stratified cross-validation (K = 7). The data was divided into seven subsets with equal composition ratios. Training and testing sets were selected from these subsets, ensuring non-repetition of the chosen test data across iterations. The average of the results from seven predictions (MAE) was calculated. The findings revealed a MAE of approximately 0.62 for RNN and GRU, with marginal differences. In contrast, LSTM achieved a substantially lower MAE of 0.41, markedly distinct from the other two neural networks (as shown in Table 1).

Table1: Predictive Results of Time Series Models

	RNN	LSTM	GRU
1	0.5846	0.4234	0.5851
2	0.6406	0.3987	0.5792
3	0.6359	0.4077	0.6023

4	0.6154	0.3957	0.6185
5	0.6217	0.3884	0.6073
6	0.6488	0.4277	0.5818
7	0.5981	0.4439	0.6191
Average Error	0.6208	**0.4123**	0.5991

4. CONCLUSIONS AND FUTURE DIRECTIONS

The present study utilized Starcraft II as the experimental environment, collecting gameplay recordings of participants as training data for machine learning, combined with short-form Flow Short Scale questionnaires as indicators of engagement levels. The research findings revealed that predicting engagement levels based on player actions still holds substantial room for improvement. This is possibly attributed to the individual variances in operating methods, which lead to slightly distinct representations when entering the state of flow. Hence, when analyzing information that carries personal traits, a greater volume of data specific to each player is needed to differentiate individual differences and enhance predictive accuracy.

The successful application of a regression model to predict engagement levels using gameplay records (MAE = 0.0623) indicated a strong correlation between objective gameplay records and subjective gaming experiences. This suggests that predicting engagement levels based on gameplay records is viable. Looking at the predictive plots of the regression model, the results closely align with the original data provided by the participants.

Compared to building predictive models based on gameplay records, using player actions enables real-time assessment and prediction, providing greater flexibility in adapting to different experimental conditions and adjustments related to the flow state. Collecting a substantial amount of data and conducting model training across various contexts should yield more accurate predictive outcomes.

Given that gameplay records solely consist of objective numerical values related to player combat actions (such as click locations) and feedback perceptions (such as enemy/ally unit eliminations), and do not possess personal characteristics, individual variations are absent. Consequently, the regression model can accurately predict a player's gameplay state. This conclusion is in line with the proposition by Alves et al. that a close relationship exists between game scores, physiological signals, and player gameplay status. In cases where physiological signal data is incomplete or insufficient, using game scores to predict gameplay status yields more accurate results [9].

REFERENCES

[1] L. Cooke, "Is Eye Tracking the Next Step in Usability Testing?," 2006 IEEE International Professional Communication Conference, 2006, pp. 236-242

[2] Csikszentmihalyi M., Beyond boredom and anxiety: Experiencing flow in work and play San Francisco: Jossey-Bass, 1975.

[3] M. Ueno, S. Wada and T. Takami, "Development of game player analysis with physiological indexes," 2015 IEEE 4th Global Conference on Consumer Electronics (GCCE), 2015, pp. 339-340

[4] Y. Bian et al., "A Physiological Evaluation Model for Flow-Experience in VR Games: Construction and Preliminary Test," 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI), 2015, pp. 244-249

[5] Katahira K, Yamazaki Y, Yamaoka C, Ozaki H, Nakagawa S, Nagata N. "EEG Correlates of the Flow State: A Combination of Increased Frontal Theta and Moderate Frontocentral Alpha Rhythm in the Mental Arithmetic Task. " J. Front Psychol. vol. 9, March, 2018

[6] Lennart Nacke and Craig A. Lindley. Flow and immersion in first-person shooters: measuring the player's gameplay experience. Proceedings of the 2008 Conference on Future Play, 2008, pp. 81-88

[7] X. Ye, H. Ning, P. Backlund and J. Ding, "Flow Experience Detection and Analysis for Game Users by Wearable Devices-based Physiological Responses Capture", IEEE Internet of Things Journal, vol. 8, pp. 1373-1381, February, 2021

[8] W. Yan and J. Vongphantuset, "Application of Virtual Reality Combined with Eye Tracking Technology for Design Flows," 2020 International Conference on Big Data and Informatization Education (ICBDIE), 2020, pp. 449-452

[9] T. Alves, S. Gama and F. S. Melo, "Flow adaptation in serious games for health," 2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH), 2018, pp. 1-8

[10] Sinha, R. Gavas, D. Chatterjee, R. Das and A. Sinharay, "Dynamic assessment of learners' mental state for an improved learning experience," Frontiers in Education Conference (FIE),2015, pp. 1-9

[11] S. C. Müller and T. Fritz, "Stuck and Frustrated or in Flow and Happy: Sensing Developers' Emotions and Progress", 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering, Florence, 2015, pp. 688-699

[12] Maeran, R. & Cangiano, Francesco. Flow experience and job characteristics: Analyzing the role of flow in job satisfaction. J. TPM - Testing, Psychometrics, Methodology in Applied Psychology. vol. 20, pp. 13-26, March, 2013.

[13] Eisenberger, Robert & Jones, Jason & Stinglhamber, Florence & Shanock, Linda & Randall, Amanda. Flow Experiences at Work: For High Need Achievers Alone?. J. Organizational Behavior, vol. 26, pp. 755-775, November, 2005.

[14] X. Ye, H. Ning, P. Backlund and J. Ding, "Flow Experience Detection and Analysis for Game Users by Wearable Devices-based Physiological Responses Capture", IEEE Internet of Things Journal, vol. 8, pp. 1373-1381, February, 2021

[15] Z. Zheng, L. Wang, Y. Cao, Y. Zhuang and X. Tao, "Towards Non-Invasive Recognition of Developers' Flow States with Computer Interaction Traces," 2019 26th Asia-Pacific Software Engineering Conference (APSEC), 2019, pp. 300-307

[16] Rheinberg, F., Vollmeyer, R., and Engeser, S. "Assessment of flow experiences", Diagnosis of Motivation and Self-Concept (Tests and Trends N.F.2), 2003, pp. 261–279