Human Factors in Virtual Environments and Game Design, Vol. 96, 2023, 47–54
https://doi.org/10.54941/ahfe1003866

Active and Passive Machine Learning Predictors to Build Adaptive Virtual Environments

Timothy McMahan[1,2] and Thomas D. Parsons[3,4]

[1]Department of Learning Technologies, University of North Texas, Denton, TX 76207, USA
[2]Mixed Realities Laboratory, University of North Texas, Denton, TX 76207, USA
[3]Grace Center, Edson College, Arizona State University, Tempe, AZ 85281, USA
[4]Computational Neuropsychology and Simulation, Arizona State University, Tempe, AZ 85281, USA

ABSTRACT

Virtual environments are increasingly used for assessment and training. While virtual environments offer ecologically valid stimulus presentations, they still follow a one-size fits all model. Technological innovation provides opportunities to transform the virtual environments into a customized experience for each individual user. This allows for the personalization of the virtual environment to the unique capabilities of a user. Active and passive data logging systems provide data necessary for adaptive virtual environments. Currently, most adaptive systems apply either active or passive data collection for building an adaptive virtual environment. The goal of the current research is to identify an optimal methodology for integrating both active and passive data into an adaptive virtual environment that can employ user data for fine tuning stimulus presentations. The framework suggested provides optimal performance parameters for identifying user cognitive and affective states and keeping users in a flow state. The result is a customized experience that is personalized to the user.

Keywords: Adaptive virtual environments, Electroencephalography, Machine learning, Psychophysiology, Neuropsychology, Cognitive

INTRODUCTION

Virtual environments cover a variety of human-computer interaction delivery systems including video games, virtual reality, augmented reality, and/or mixed reality. While these simulation platforms are often promoted for their entertainment value, they also offer important promise for research and practice. Significant increases in applications for learning, cognitive training, psychological assessment, and rehabilitation are found in the literature (Bohil et al., 2011; Parsons et al., 2020). Technological advances have enhanced the multisensory presentations of virtual environments. However, these environments primarily follow a one-size fits all development model. Recent technological advancements provide the possibility for dynamically adaptive virtual environment (AVE) platforms that adapt to users in a manner that offers enhanced interactive experiences (Zahabi & Abdul, 2020, Scott et al. 2016, Shute & Towle 2018).

Converting a virtual environment into an AVE requires logging specific user performance data. Currently, the two best methods of identifying user performance are active and passive user metrics. Passive metrics include logging of physiological metrics for passive detection of changes in neurocognition (e.g., electroencephalography (EEG)) and autonomic arousal (e.g., electrodermal activity (EDA); electrocardiography (ECG); electrodermal activity (EDA); respiration, and electromyography (EMG). Active metrics are assessed via logs of user behaviours while immersed in the virtual environment. Active measures mainly focus on timing to complete specific tasks, error rate, responses to cuing, and individual task reactions. Most AVEs implement either active or passive monitoring to perform the adaptation. The ideal AVE would implement both active and passive metrics to create a fluid user experience.

An adaptive Virtual School Environment (VSE) with an interactive virtual human teacher and interactive virtual students has been designed that includes a Virtual Classroom Stroop assessment (McMahan et al. 2022). The VSE includes multiple scenarios (e.g., Stroop test, continuous performance test, and picture naming) presented to the user while immersed within a virtual classroom, hallway, and playground. All scenarios include the option to incorporate social cues from the virtual human teacher (see Figure 1). The Virtual Classroom Stroop Task within the VSE includes both high and low levels of distractions (e.g., visual, auditory) as well as multiple methods for delivering stimuli (e.g., visual, auditory, and bimodal). Utilizing the virtual teacher in the VSE provides multiple quantitative measures including accuracy, average reaction time for correct and incorrect answers—all with and without teacher directed attention and/or distractions.

Another virtual environment is the Virtual Environment Grocery Store (VEGS), which offers a high-fidelity virtual environment-based cognitive assessment (see Figure 1). The goal of VEGS, is to have users interact within the environment while performing a shopping task. VEGS, as a cognitive task, has the capability of measuring a person's learning, memory, navigation, and executive functions. VEGS is built to have multiple distraction capabilities.

Figure 1: Virtual school environment utilizing the STROOP; Virtual environment grocery store.

Recent psychometric validation found VEGS to have good construct validity for evaluating both young and older adults (Barnett et al. 2022; Weitzner et al. 2021). VEGS was found to be primarily an episodic and prospective memory assessment in the low distraction version (Parsons & Barnett 2017). Adding more distractions to the environment like NPCs, environmental noise, and visual elements showed that a user's performance was linked to both memory and executive function measures (Parsons & McMahan 2017).

ACTIVE METRICS

Active metrics are the metrics that we can extrapolate from the virtual environment itself. Because the user is constantly and actively completing tasks and objectives within the virtual environments, we have several metrics that we can be utilized to categorize the user's performance. Table 1 lists potential active metrics for virtual environments. These metrics are collected throughout the assessment process. One issue with active metrics is that while the active data is immediately logged, it must be postprocessed offline. In Table 1, several variables are calculated based upon the average over a given time period (e.g., average time to perform an operation in the VEGS, mean response time over a stimulus presentation condition in the VSE). The user must progress through the tasks before the system is able to make predictions about how the user is performing. This postprocessing delay causes the adaptability to be delayed until enough data is collected to perform the correct adaptation for the user.

Table 1. Possible active metrics extrapolated from virtual environments.

Virtual Environment Grocery Store	Virtual School Environment STROOP
# of times looked at shopping list	Mean Response Time Box Correct (Interference/Congruent)
Travel time to pharmacist	Mean Dwell Time Box Incorrect (Interference/Congruent)
Time spent shopping	Mean Response Time Word Correct (Interference/Congruent/Incongruent)
ATM withdraw time	Mean Dwell Time Word Correct (Interference/Congruent/Incongruent)
First item picked up	Mean Response Time Word Incorrect (Interference/Congruent/Incongruent)
Average pick up time	Mean Dwell Time Word Incorrect (Interference/Congruent/Incongruent)

PASSIVE METRICS

Passive AVE metrics are any metrics in which we can passively collect data in real-time. This includes eye-tracking, facial action coding, speech emotion recognition, and psychophysiology measures. Some psychophysiology measures lend themselves better to AVE's because they provide closer to real-time data (e.g., EEG, EMG and HR). One of the greatest benefits of psychophysiological signals is that they passively log user experience in a continuous manner that does not break the users sense of presence (Allanson & Fairclough 2004). The passive logging of psychophysiological data is done without the user's awareness. This provides objective measures of cognitive workload (Dehais et al. 2020, Tao et al 2019), task engagement (Pope et al. 1995), arousal (Cuthbert et al. 2000), and stress levels (Panicker et al., 2019). The goal of the current research is to identify a methodology to integrate the active metric results with our passive metrics measurements, specifically EEG. Table 2 shows various metrics we can gather from EEG to use for AVEs.

Table 2. Active metrics from EEG.

Passive Metrics	EEG Measurement
Delta	1 – 4 Hz Dominant wave during deep sleep
Theta	4 – 7 Hz Dominant wave during light sleep
Alpha	7 - 13 Hz Dominant wave when in a state of relaxation
Beta	13 – 25 Hz Dominant wave when active, busy, or concentrating
Gamma	25 – 43 Hz Dominant wave when forming ideas or memory processing
Engagement	$\frac{\text{Global Beta}}{(\text{Global Alpha} + \text{Global Theta})}$ (Freeman et al. 1999)
Arousal	$\frac{(\text{BetaF3} + \text{BetaF4})}{(\text{AlphaF3} + \text{AlphaF4})}$ (Giraldo & Ramirez 2013)
Valence	$\frac{\text{AlphaF4}}{\text{BetaF4}} - \frac{\text{AlphaF3}}{\text{BetaF3}}$ (Giraldo & Ramirez 2013)

MACHINE LEARNING ALGORITHMS

Machine learning offers a solution to categorizing how a user is performing in real-time within a virtual environment. Several machine algorithms exist which have demonstrated the potential to work in real-time including Support Vector Machine (SVM), Naïve Bayes (NB), and k-Nearest Neighbor (kNN). SVMs employ hyperplanes to portion the data into two or more classes. SVMs train via data known to be associated with each category and makes an attempt to position the data points into a higher dimensional space. The NB classifier is based upon Bayes theorem and uses computed probability that a given collection of data points is associated with a specific class. Classification of data points is done by selecting the result with the greatest probability. During the training period, kNN employs vectors containing features to store each category of data. Classification of new data requires kNN to compute the shortest distance between possible categories. Whichever category has the shortest distance is selected as the most possible category.

Passive and Active Machine Learning Results

Research performed previously has shown that passive measures are capable of categorizing a user into either high or low performer category in both the

VSE and VEGS (McMahan et al. 2021, McMahan et al. 2021). Participants were classified as either high or low performers based upon the calculated throughput in the VSE. Within the VEGS, high and low performers were categorized based on the number of items the user found while shopping. Two categories were chosen initially to best determine which classifier worked best with the type of data being fed to it. SVM always performed the best; kNN was closely behind and way above NB. However, it was suggested that the most optimal solution would utilize both the SVM and kNN classifier.

Utilizing EEG as a passive metric has been demonstrated to be more then capable of classifying users within a specific state in a virtual environment (McMahan & Parsons 2020). Utilizing engagement, arousal, and valence indices user states were classified utilizing SVM, kNN and NB. Similarly like the active results, it was determined that the ideal solution would be a combination of all the classifiers. The results from the classifiers were applied into a flow model designed to balance the intensity and complexity vs the user's ability level. The goal of the flow model is to keep the user in the perfect immersive state not to stressed and not bored with the current task. Utilizing EEG provided a solution for the creation of a flow model built upon the coordinate system of "Task Engagement" and "Arousal-Valence".

Combining Active and Passive Metrics in Machine Learning

The combination of multiple machine learning algorithms, also known as ensemble machine learning, provides a better solution to increase the performance overall for the AVE utilizing them (Gupta et al., 2020; Sun et al., 2007). Using both passive and active metrics in machine learning should increases the overall accuracy in classifying users into the correct category. The major issue that must be resolved is the interval nature of active metrics which only becomes available during certain times of the AVE and the continuous nature of passive metrics which are available immediately at the start of the AVE. Majority voting is one potential technique that uses multiple classifiers to vote on the correct category (Rojarath et al., 2016). Stacking classifiers is another approach in which the results of classifiers are fed in as new predictors to additional classifiers (Chatterjee & Byun, 2022).

Figure 2 demonstrates these two possible solutions for ensemble machine learning systems to take advantage of both types of data within an AVE. Figure 2A implements a voting scheme to determine the overall correct classification between the continuous data and the interval data. The voting scheme requires that the AVE implement four independent classifiers (e.g. kNN continuous data, SVM continuous, kNN interval, SVM interval). The continuous classifiers (bottom two classifiers in figure 2a) would be run based upon the availability of epoch of data from the passive EEG. The interval classifiers (upper two classifiers in figure 2a) would execute either on preset intervals that are coordinated with expected task completions or event driven based upon the user completing specific parts of the AVE. The result from each classifier is used for majority voting. The category with the highest number of votes from each time frame will be considered the current users performance level.

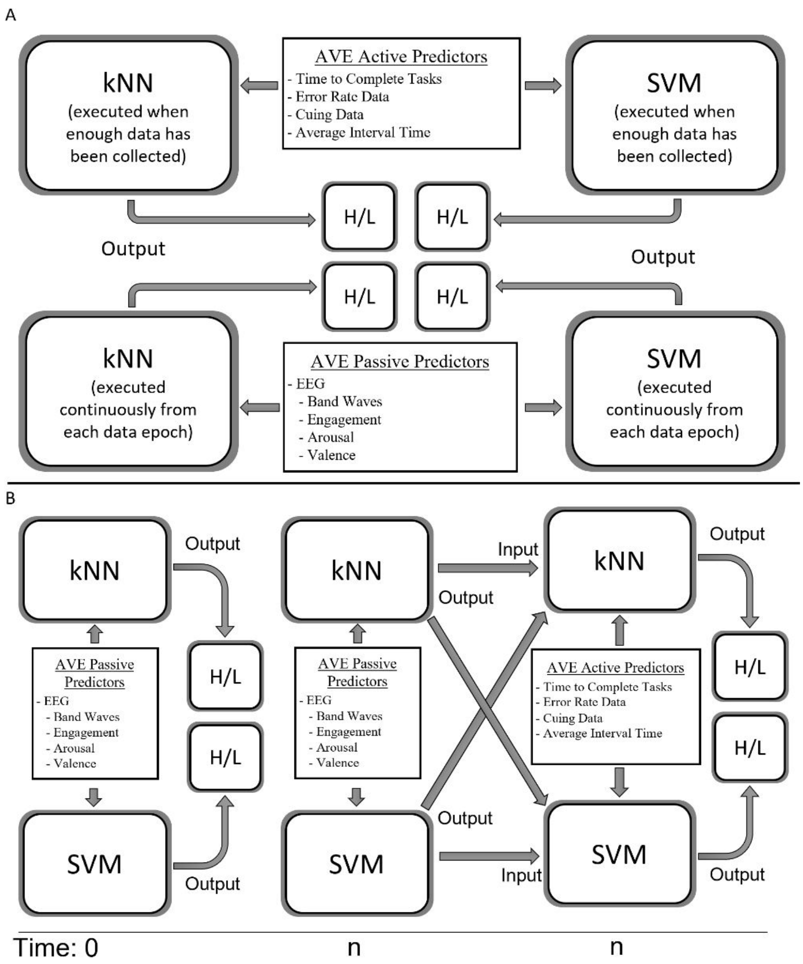

Figure 2: A: Voting based ensemble machine learning system split between continuous data and interval data. **Figure 2B:** Stacking based ensemble machine learning algorithms executed with only continuous data until interval data becomes available.

Figure 2B demonstrates a stacking method for implementing the ensemble machine learning system. At time 0 to 1-n in Figure 2B, the ensemble system uses only the continuous data as predictors to determine the performance of the user. Once active predictors become available at time n, the system switches to a stacking method in which the output from one classifier is fed into the input of another classifier. At time n in Figure 2B, the passive classifiers kNN and SVM identify what they believe is the correct performance level of the user. These results are then fed in as additional predictors for the Active classifiers and the final decision is made based upon the predictive probability of active classifier results.

An alternative solution is the application of machine learning that employs reinforcement learning. In this solution, the machine learning classifier uses the passive metrics at every epoch interval to classify the user's performance. As the AVE receives active data the results of the passive machine learning and active machine learning are compared. If they are found to be equal, the AVE continues using the same training data that it has been operating with. If the results are not equal, then the passive machine learning classifiers model can be updated to reflect data that better match what was measured during a particular event recently completed within the AVE. Throughout the entire AVE, the machine learning models would be being refined to fine tune the data to more accurately match the users' abilities.

CONCLUSION

In summary, we were able to identify common metrics that can be implemented in virtual environments that will aid in future development of AVEs. The framework we developed will allow for better detection of cognitive and affective states allowing the systems to fine tune difficulty settings for each individual user. The AVE will have the capability to accurately determine if the user is in a state of flow and make real-time adjustments to the tasks within the virtual environment. This results in a better user experience for the individual and a more accurate measurement of their performance in the environment.

REFERENCES

Allanson, J. and Fairclough, S. H., 2004. A research agenda for physiological computing. *Interacting with computers*, *16*(5), pp. 857–878.

Barnett, M. D., Chek, C. J., Shorter, S. S. and Parsons, T. D., 2022. Comparison of Traditional and Virtual Reality-Based Episodic Memory Performance in Clinical and Non-Clinical Cohorts. *Brain Sciences*, 12(8), p. 1019.

Bohil, C. J., Alicea, B. and Biocca, F. A., 2011. Virtual reality in neuroscience research and therapy. Nature reviews neuroscience, 12(12), pp. 752–762.

Chatterjee, S. and Byun, Y. C., 2022. EEG-Based Emotion Classification Using Stacking Ensemble Approach. *Sensors*, *22*(21), p. 8550.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N. and Lang, P. J., 2000. Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biological psychology*, *52*(2), pp. 95–111.

Dehais, F., Lafont, A., Roy, R. and Fairclough, S., 2020. A neuroergonomics approach to mental workload, engagement and human performance. Frontiers in neuroscience, 14, p. 268.

Freeman, F. G., Mikulka, P. J., Prinzel, L. J. and Scerbo, M. W., 1999. Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biological psychology*, *50*(1), pp. 61–76.

Giraldo, S. and Ramirez, R., 2013. Brain-activity-driven real-time music emotive control. In *The 3rd International Conference on Music & Emotion, Jyväskylä, Finland, June 11-15, 2013*. University of Jyväskylä, Department of Music.

Gupta, A., Khan, R. U., Singh, V. K., Tanveer, M., Kumar, D., Chakraborti, A. and Pachori, R. B., 2020. A novel approach for classification of mental tasks using multiview ensemble learning (MEL). *Neurocomputing*, *417*, pp. 558–584.

McMahan, T., Duffield, T., & Parsons, T. D. (2021). Virtual Environment Grocery Store (VEGS) for assessing memory in persons with epilepsy: A Comparative Study on the Predictive Ability of the Support Vector Machine, K-Nearest Neighbors, and Naïve Bayes. *Proceedings of the 13th International Conference on Disability, Virtual Reality & Associated Technologies*. pp. 130–136.

McMahan, T. and Parsons, T. D., 2020. Adaptive Virtual Environments using Machine Learning and Artificial Intelligence. *Annual Review of Cybertherapy and Telemedicine*, p. 141.

McMahan, T., Duffield, T. and Parsons, T. D., 2021. Feasibility study to identify machine learning predictors for a virtual school environment: virtual reality Stroop task. *Frontiers in Virtual Reality*, *2*, p. 673191.

Panicker, S. S. and Gayathri, P., 2019. A survey of machine learning techniques in physiology based mental stress detection systems. Biocybernetics and Biomedical Engineering, 39(2), pp. 444–469.

Parsons, T. D. and Barnett, M., 2017. Validity of a newly developed measure of memory: Feasibility study of the virtual environment grocery store. *Journal of Alzheimer's disease*, *59*(4), pp. 1227–1235.

Parsons, T. D. and McMahan, T., 2017. An initial validation of the virtual environment grocery store. *Journal of neuroscience methods*, *291*, pp. 13–19.

Parsons, T. D., Gaggioli, A. and Riva, G., 2020. Extended reality for the clinical, affective, and social neurosciences. Brain Sciences, 10(12), p. 922.

Pope, A. T., Bogart, E. H. and Bartolome, D. S., 1995. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology*, *40*(1-2), pp. 187–195.

Rojarath, A., Songpan, W. and Pong-inwong, C., 2016, August. Improved ensemble learning for classification techniques based on majority voting. In *2016 7th IEEE international conference on software engineering and service science (ICSESS)* (pp.107–110). IEEE.

Scott, E., Soria, A. and Campo, M., 2016. Adaptive 3D virtual learning environments—A review of the literature. *IEEE Transactions on Learning technologies*, *10*(3), pp. 262–276.

Shute, V. and Towle, B., 2018. Adaptive e-learning. In *Educational psychologist* (pp. 105–114). Routledge.

Sun, S., Zhang, C. and Zhang, D., 2007. An experimental evaluation of ensemble methods for EEG signal classification. *Pattern Recognition Letters*, *28*(15), pp. 2157–2163.

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X. and Zhang, T., 2019. A systematic review of physiological measures of mental workload. International journal of environmental research and public health, 16(15), p. 2716.

Weitzner, D. S., Calamia, M. and Parsons, T. D., 2021. Test-retest reliability and practice effects of the virtual environment grocery store (VEGS). *Journal of Clinical and Experimental Neuropsychology*, 43(6), pp. 547–557.

Zahabi, M. and Abdul Razak, A. M., 2020. Adaptive virtual reality-based training: a systematic literature review and framework. *Virtual Reality*, *24*, pp. 725–752.