Classification of video game players using EEG and logistic regression with ridge estimator

Gustavo A. Lujan-Moreno, Robert Atkinson, George Runger, Javier Gonzalez-Sanchez and Maria Elena Chavez-Echeagaray

Arizona State University, School of Computing, Informatics and Decision Systems Engineering, Tempe, Arizona. {gustavo.lujanmoreno, robert.atkinson, george.runger, jgonza24, mchaveze} @asu.edu

Abstract. The objective is to classify a group of subjects playing a video game as experts and novices using electroencephalogram (EEG) signals as inputs. Analytical methods applied to multi-channel EEG recording are described. A fast Fourier transform (FFT) is used to calculate the power spectral density for a number of bandwidths (delta, theta, alpha and beta) and ratios (e.g., theta/beta). A regularized logistic regression learning algorithm ($L_2$ penalty) was applied to the extracted features. We successfully classified 80% of the instances using a 10 fold cross-validation.

Keywords: data mining, electroencephalogram (EEG), gaming, linear regression, logistic regression, machine learning, ridge estimators.

1 Introduction

The utilization of headsets to gather brainwaves in the past was merely restricted for health purposes. Lately, the applications have expanded to other areas such as tutoring systems, learning environments and video games [5]. One study for example used Microsoft Kinect and EEG to isolate body movements of participants while playing a virtual ball game [2]. Concerns regarding usability, standardization, minimum set up time and ethical issues are still some of the challenges when using these type of devices [5].

Although different devices and methods exist to monitor brain activity such as: electrocorticography (ECoG), electroencephalography (EEG), functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), positron emission tomography (PET), magneto-encephalography (MEG), near-infrared spectroscopy (NIRS), and intracortical electrodes (ICE) [2][5], EEG has consistently showed to be accurate and convenient for practical purposes [5].

Furthermore, different areas of EEG applications have been identified such as: device control, user-state monitoring, evaluation, training and education, cognitive improvement, gaming and entertainment, [5]. Several commercial EEG devices even try to estimate the affective state of the users in real time such as: frustration, meditation, fatigue, stress, drowsiness, distraction, task engagement and mental workload [6][7].
Past studies have tried to classify or predict an outcome based on physiological information. For example, recent studies have used non-parametric tests and regression analysis on predictors based on heart rate, skin conductance and EEG to assess learners’ attention while overcoming obstacles [8].

In this study we use an EEG system to generate features that are later used to classify participants as experts or novices. For this purpose we use the video game Guitar Hero. This video game requires attention, rhythm and coordination. Video games have shown to have substantial influence in education, healthcare and even social change [3].

Our hypothesis is that experts and novices have a different cognitive process when playing a video game. Novices will normally go through a learning curve which includes the video game interface, the purpose of the game and finally the skill. Experts on the other hand normally try to achieve a “flow” state. Experts in different areas describe this feeling as if time and even the purpose of the game didn’t matter. Researchers have describe this “flow” zone as an enjoyable and satisfying experience [4].

We present the process to classify experts and novices using EEG signal as inputs. Furthermore, we approach the task from a statistical machine learning perspective applying a logistic regression algorithm that uses ridge estimators.

2 Methods

2.1 Game Environment

As previously mentioned, the Guitar Hero video game was used for this study. Guitar Hero is a game that involves holding a guitar interface while listening to music and watching a video screen. This type of video game uses a combination of graphics, multimedia and challenges that require to develop different skills. This gives the best context in which to elicit changes in the different brain waves frequencies that users generate. Guitar Hero provides a scenario where subjects are challenged in different ways that demands from them different skills that could be related with a learning process such as concentration as well as visual, motor, and auditory skills. The objective of the game is to hit the right button(s) by looking at the “notes” streaming on the screen. The user has five colored buttons to press on the guitar fingerboard. Both hands are needed to play since the left hand clicks the color buttons on the guitar arm and the right hand is used to depress a switch that resembles a guitar strum or string picking.

2.2 Participants and Design

We recruited 21 subjects from Arizona State University of which 14 were men and 7 were women. Age ranged from 18 to 28 years. Participants were compensated with $10 USD and they had the option to leave the study at any time. Participants were asked to self-report their experience playing video games. A total of 8 participants
were selected: 4 men and 4 women whose age range from 18 to 28 years. According to a self-report four of them were classified as novices and four of them as experts. In this study a novice is defined as a person who does not normally play video games. On the contrary, an expert is defined as a person who not only plays video games frequently but who also considers himself proficient at playing Guitar Hero. The self-report was validated by the final score of the participants while playing Guitar Hero. Participants played two songs: an easy and a difficult one. The easy song, “Story of my life”, had length of 5 minutes and 40 seconds, 19 segments, and a total of 511 notes. The difficult song, “One”, had length of 7 minutes and 3 seconds, 25 segments, and 2189 total number of notes. We had a total of 16 data sets, one for each player-difficulty possible combination. We used Weka 3.6.10 to perform the analysis. Also, to compute the PSD the EEGLAB toolbox in Matlab was used. EEGLAB is a Matlab toolbox for processing EEG data that performs time/frequency analysis.

2.3 Data Sources: EEG

Several studies have shown that physiological measures are reliable sources to measure learners’ attention [14]. In this case, we used a brain computer interface (BCI) device to capture neural oscillations also known as brain-waves signals. The neural oscillations are generated by the neural tissue and can vary by frequency, phase and amplitude.

For the EEG system we used Emotiv EEG. This device is a high resolution, multi-channel, wireless portable EEG system. It has 14 EEG channels with names based on the International 10-10 locations, these are: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4. The signal from the 14 channels has a sampling rate of 128 SPS.

2.4 Feature generation

The signal from the 14 channels was filtered with a band-pass filter (0.2–45Hz). Furthermore, Emotiv software applies digital notch filters at 50Hz and 60HZ in order to remove environmental artifacts. Once we had the decontaminated EEG raw signal we transformed the EEG data from time to frequency domain using a Fast Fourier Algorithm (FFT). In this study the Power Spectral Density (PSD) was calculated for the following bandwidths: delta, 0.1-4 Hz; theta, 4-8 Hz; alpha, 8-14 Hz; and beta, 14-30 Hz. The time/frequency decomposition was done for each of the 14 channels. Two ratios were also computed: the theta/beta and the delta/beta. Studies have shown that the theta/beta ratio may provide useful information in the study of affective and emotional regulation [2]. Additionally, power density ratios in frequency bands have been studied in neuroscience. For example, a studied showed that slow wave/fast wave (SW/FW) ratios increase in subjects with attention deficit hyperactivity disorder (ADHD) [2]. Since we have 8 participants and each one played 2 songs, we had a total of 16 combinations. However, one dataset was discarded due to problems with the timestamp. Consequently, with 15 rows of information we decided to average the power of each of the bandwidths and the two ratios in order to have only 6 variables.
Therefore, the logistic regression model would only need to compute 7 parameters (6 variables plus the intercept).

2.5 Logistic regression with ridge estimators

Logistic regression is a widely used method for classifying binary data. If we define $Y_i = 1$ as the event of classifying subject $i$ as an expert and $Y_i = 0$ otherwise then the probability that $Y_i = 1$ given the value $X_i = (X_{i1}, \ldots, X_{i7})$ can be defined as:

$$p(X_i) = \frac{\exp \left( \sum_{j=1}^{7} \beta_j X_{ij} \right)}{1 + \exp \left( \sum_{j=1}^{7} \beta_j X_{ij} \right)}$$

Normally, the optimal value for $\beta$ would be found maximizing the log likelihood function:

$$l(\beta) = \sum_i [Y_i \log p(X_i) + (1 - Y_i) \log (1 - p(X_i))]$$

The result will yield the well-known MLE $\hat{\beta}$ for $\beta$.

LeCessie and van Houwelingen (1992) proposed to use the same approach but this time including a penalty $\lambda$ to the norm of $\beta$:

$$l^\lambda(\beta) = l(\beta) - \lambda \|\beta\|^2$$

The ridge parameter in this case is $\lambda$ which controls the length of the norm of $\beta$. If we set $\lambda = 0$ this would be equivalent to the ordinary MLE. On the other hand as $\lambda \to \infty$ all parameters $\beta_j$ tend to 0. The effect in the logistic regression model is the same as in linear regression. We allow a small bias in the parameters $\beta_j$ but this will allow to reduce the variance and stabilize the model especially for predictions. The optimal value for the ridge estimator is found through cross-validation such that the mean error rate is minimal. For more details refer to [10].

3 Experiments and results

As we previously mentioned, out of the 16 possible combinations we used 15 due to problems with the time stamp log in one of the expert-hard data sets. The inputs were standardized to have $\mu_i = 0$ and $\sigma_i = 1$. The ridge parameter was tuned up using cross-validation and we found out that the algorithm performed well with $\lambda = 0.001$. The logistic regression algorithm with ridge estimators with and 10-fold cross-validation was able to correctly classify 80% of the samples. The true positive rate for class 1 (expert) was 0.86, with 6 experts classified as experts and 1 expert classified as novice. On the other hand, the true positive rate for class 0 (novice) was 0.75, with
6 novices correctly classified and 1 novice classified as expert. These results can be seen in Table 1.

<table>
<thead>
<tr>
<th>No. Instances</th>
<th>Accuracy</th>
<th>Classified as:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
<td>1 6 1</td>
</tr>
<tr>
<td>Accuracy</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>TP Rate Class 1</td>
<td>0.86</td>
<td>0 2 6</td>
</tr>
<tr>
<td>TP Rate Class 0</td>
<td>0.75</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 shows the coefficients for the 7 parameters (6 variables plus intercept). The left part of the table shows the odds ratio. The odds ratio is the estimated increase in the probability of success \( p(Y_i = 1) \) with a one-unit change in the value of the variable \( x_i \). We observe that the largest odd ratio is given by the delta variable. Furthermore, the second largest odds ratio is caused by the predictor beta. We see the correspondence between the right and left table since the odds ratio can also be computed as \( e^\beta \).

Table 2. Coefficients and odds ratio for logistic regression with ridge parameter

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class 1</td>
<td>Class 1</td>
</tr>
<tr>
<td>delta</td>
<td>12.1738</td>
<td>193652.07</td>
</tr>
<tr>
<td>theta</td>
<td>-16.7003</td>
<td>0</td>
</tr>
<tr>
<td>alpha</td>
<td>1.1019</td>
<td>3.01</td>
</tr>
<tr>
<td>beta</td>
<td>4.8335</td>
<td>125.6488</td>
</tr>
<tr>
<td>Theta/Beta</td>
<td>2.3244</td>
<td>10.2202</td>
</tr>
<tr>
<td>Delta/Beta</td>
<td>-0.013</td>
<td>0.9871</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.5503</td>
<td></td>
</tr>
</tbody>
</table>

We can see that variables with the largest coefficients are delta and theta. The significance for delta is \( \text{Pr}(|t|) = 0.0207 \) and theta is \( \text{Pr}(|t|) = 0.0051 \). The interpretation of the coefficients in logistic regression is similar to that for the case of linear regression. This means that an increase in the delta variable will increase the chances of classifying a subject as an expert, assuming that all other predictors are held constant. On the other hand, an increase in the theta variable will decrease the chances of classifying a subject as an expert.

4 Conclusion and future work

We successfully classified experts from novices using logistic regression with ridge estimators. The final ridge parameter was tuned to be equal to 0.001. This value was selected while trying to minimize the cross-validation misclassification error. The results suggest that the cognitive process of an expert differs from a novice under the gaming context. The delta bandwidth (0.1-4 Hz) turned out to be significant and an
increment on this variable will increase the chances of classifying a subject as an expert. On the other hand, the largest coefficient in absolute value was the bandwidth theta (4-8 Hz). These conclusions shed some light regarding the importance of low frequencies when trying to classify experts from novices. Experts tend to have higher values of delta signals while novices are more inclined to have higher power in the theta bandwidth. It is necessary to conduct more studies to confirm these results using a larger sample size. One potential application of this finding is that we could certify a person as an expert not only by the number of training hours or final score but also by the analysis of brainwaves. This knowledge could be applied to know if, for example, a driver is really processing information as an expert during a driving test with no need of a human evaluator. A medicine student could be evaluated during a surgery practice to know if he is in the “flow” or still struggling with the procedure. The so called “flow state” now could be described in terms of a model and not only as an abstract definition.

References