A Web-Based System for EEG Data Visualization and Analysis

A. Jonathan Garza, B. Sishir Subedi, C. Yuntian Zhang and D. Hong Lin
Department of Computer Science and Engineering Technology
University of Houston-Downtown
Houston, Texas, USA

Abstract - While many advances have been made in the understanding of the human brain, it still contains mysteries in its inner workings. The objective of this project is to help discover the answers to more of these mysteries by creating a model that can be applied to Electroencephalographic (EEG) brainwave data to predict what a person is doing or happening to them. The dependent variables of this study are the five major brain waves and the independent variable is the activities performed by the participants. We are creating an environment to capture and analyze EEG brainwave data using various custom developed tools, off the shelf software and hardware components. To help create this environment we are building a website to help facilitate access to and the analysis of collected EEG data. The types of analysis that can be currently performed on the data stored on the web server are wave analysis and statistical analysis. Also a mobile application is being developed to help facilitate the collection of EEG data that is then given to the web server and display the results of data analysis from the web server.

Keywords: Electroencephalographic data, Brain State Modeling, Web-Based System, EEG Data Visualization, EEG Data Analysis, computational biology

1 Introduction

The human brain is the result of many years of evolution. Its complexities have been studied for years and yet the research done has only begin to scratch the surface. To assist in the aid of that research, a system that can collect, store and processes Electroencephalographic (EEG) data is desirable. This is because EEG signals characterize the result of the neuron activities inside of a human brain. Naturally, they are used to study and understand human brain activities. In particular, EEG signals indicate that neural patterns of meanings in each brain occur in trajectories of discrete steps, whilst the amplitude modulation in EEG wave is the mode of expressing meanings [1]. The purpose of the system discussed in this paper is to allow people to easily store, analyze and collaborate on EEG data. The system takes EEG data and exposes it to various analytical techniques so the resultant brain states can be studied and predicted.

The vast implications of using EEG data to analyze brain states include designing brain-computer interfaces (BCI) where users can operate on a machine via brain activities, and using brain state models in healthcare related activities. Scientists now have the ability to measure and register electric potential of the human brain through the use of electroencephalographic technologies. The combination of electroencephalographic data with modeling methods in fields such as data mining and bioinformatics could be used to diagnose disease in advance to increase success of a cure. It could also be used to prove that subjects in a state of transcendental meditation are in a verifiable and observable state of mind that can be monitored and predicted [2]. Experiments found that cancer patients that practiced meditation experienced higher well-being levels, better cognitive function and lower levels of inflammation than a control group [3].

Therefore, a platform for comprehensive EEG data storage and processing is desirable to promoting applications of using EEG tools in both physiological (e.g., clinical uses, sleep evaluation, fatigue detection, etc.) and psychological (cognitive sciences, BCI, etc.) scopes. Such a platform consists of EEG data collection devices (viz., EEG headset), communication channels (e.g., smart phones), a web server that provides a web interface for users to access stored EEG data and activate data analysis algorithms, an online database for EEG data storage and processing and a forum for users to collaborate with each other while using the system. Figure 1 shows an outline of the proposed system.

Figure 1. EEG data analysis system architecture
2 Related Works

The technology of using the web for visualization purposes has been analyzed before. Nathan Holmberg, Burkhard Wunsche, and Ewan Tempero did a study on Interactive Web-Based Visualization in which they developed a framework to categorized different web-based technologies for 2D and 3D visualization [4]. DHTML which consists of HTML and JavaScript combined, performed well with disadvantages related to limited communication with servers at the time of the writing. Another point made was the popular use of this technology, such as its use by Google Maps. Finally the factor that this technology solely had was the fact that it was not a plugin that users have to install but is the only solution built natively into the web browser. This means users don’t need to install any special software to run the visualization software built in DHTML technology.

A study, done by Andrew V. Poliakov, Evan Albright, Kevin P. Hinshaw, et al., found that a major advantage, among others, in a server-client system setup is that the client’s hardware does not need to be particular powerful as most of the processing of large data is done on the server side. Servers also tend to be more powerful then personal computers, even at the inexpensive end of servers. Another important factor with regards to server hardware is the popular inclusion of more than one CPU which makes it possible to run parallel data processing methods reducing the overall needed processing time of large amounts of data.

JSON has been shown as a viable way to transmit data from the server to the client software or browser. Many programming languages support JSON messages now and it is the native data representation present in JavaScript which provides convenience when developing website using JavaScript. Web-based system have also shown that it is possible to display multiple records of data together, allowing users to better compare interpersonal difference and similarities between different records as well [5]. This provides a greater aspect of analysis possible then just the displaying of individual records.

Another aspect of web-based system which has shown is that most users found a well-designed system to be easy to use and has a very quick learning curve [6]. This means that non-technical users can easily focus on the analysis of the data and less on the learning on how to use the system.

3 EEG Data Collection and Storage

The brain emits electrical signals that are caused by neurons firing in the brain. The patterns and frequencies of these electrical signals can be measured by placing a sensor on the scalp. For example, the EEG sensor by NeuroSky is able to measure the analog electrical signals commonly referred to as brainwaves and process them into digital signals to make the measurements available for further analysis. Table 1 lists the most commonly recognized frequencies that are generated by different types of brain activity.

### 3.1 EEG Headset

We briefly describe how a simple EEG headset can be built using open source materials. The prototype multi-functional headset we built consists of an EEG sensor, a pulse sensor, a temperature sensor, a microprocessor, and a microprocessor Bluetooth shield.

<table>
<thead>
<tr>
<th>Brainwave Type</th>
<th>Frequency Range</th>
<th>Mental States and Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>0.1 Hz to 3 Hz</td>
<td>Deep, dreamless sleep, non-REM sleep, unconscious</td>
</tr>
<tr>
<td>Theta</td>
<td>4 Hz to 7 Hz</td>
<td>Intuitive, creative, recall, fantasy, imaginary, dream</td>
</tr>
<tr>
<td>Alpha</td>
<td>8 Hz to 12 Hz</td>
<td>Relaxed, but not drowsy, tranquil, conscious</td>
</tr>
<tr>
<td>Low Beta</td>
<td>12 Hz to 15 Hz</td>
<td>Formerly SMR, relaxed yet focused, integrated</td>
</tr>
<tr>
<td>Midrange Beta</td>
<td>16 Hz to 20 Hz</td>
<td>Thinking, aware of self and surroundings</td>
</tr>
<tr>
<td>High Beta</td>
<td>21 Hz to 30 Hz</td>
<td>Alertness, agitation</td>
</tr>
</tbody>
</table>

The assembled headset is shown in Figure 2, where the three sensors are mounted on the tips of the three legs on the forehead supports. The microprocessor and the microprocessor Bluetooth shield are mounted on the back, and the ear lobe is used as an electrical ground base for the EEG sensor.

In order to test and validate that the headset is working properly and that all the sensors are functioning, a test environment had to be constructed. To simulate a real world environment, a mobile smart phone application was developed on the Apple iPhone platform. This platform was chosen for ease of access to development tools and availability of software development kits (SDK) from all the hardware and chipset vendors. Both NeuroSky and Red Bear Labs included sample applications that were then easily transferred to a...
custom application using a simple view to display all the sensor values.

3.2 Storage

To provide a reliable storage option that can handle the large amount of data from EEG recording sessions, we store the data in a relational database. The database itself is hosted on our web server. The EEG data is stored in its own separate tables apart from other tables necessary for the website. This partly for security reasons but mostly for clarity of which tables are for EEG data and which are not.

The data collected using the NeuroSky headset produces a comma separated values (CSV) file that can be quite large for about 3 minutes of data collection. On average the resulting file is 8 to 9 MB in size which was imported into the database manually in the beginning. To import the EEG data into the database required that the data file first be transferred to the server and then access the database to import the file’s contents. We have implemented on the website a method to upload EEG data files seamlessly. After users login, they can choose to upload a new file. Then they will choose which local file they want to upload and the unique name of the data set. After which the file will be uploaded in the background, displaying a progress bar for the user (Figure 3). Once the upload has been completed successfully, the file is stored in an upload folder and then immediately parsed into the database with the name given. Once completed, the EEG data is ready for immediate analysis through the website. If for some reason the upload fails at any point, the user receives an appropriate message. Also on failure of completion, if any incomplete data has been uploaded, it is removed to ensure consistency.

Figure 3. EEG data upload in progress

4 Dynamic EEG Data Modeling

To study what EEG data analysis algorithms to be implemented on the web server, we have first begun to do the analysis using Microsoft Excel and the statistical computing programming languages R and Python. Component frequencies, including five major brain waves- Delta(1-3Hz), Theta(4-7Hz), Alpha Low(8-9Hz), Alpha High(10-12Hz), Beta Low(13-17Hz), Beta High(18-30Hz), Gamma Low(31-40Hz), and Gamma Mid(41-50Hz) were extracted from the raw dataset. These frequencies represent specific brain states including deep meditation and high anxiety.

The data in the headset reports brain wave frequencies as a function of its power spectrum. Fourier Transform analysis was implemented by the application software package to decompose the raw EEG time series into a voltage by frequency spectral graph (power spectrum). This power spectrum values obtained for specific brain waves was investigated for the numerical analysis of Quantitative EEG data.

The power spectrum data was normalized to reduce variability, which might have occurred due to difference in contact distance between the headset and the user, and changes in environmental condition. In addition, automatic scaling feature in the hardware accounts for noises that make the data values large. The normalized method used the sum of all the eight brain waves power spectrum data and divided each data point by the sum to scale it within the range of 0 to 1. The box plot (Figure 3) graphically depicts the numerical spread of the normalized data for the combined brain waves as well as meditation, movie watching, and reading aloud brain states. The standard deviation model was calculated for each brain state to investigate the variability of data and predict the percentage of values that are present within the one standard deviation from the mean (Figure 4).

Figure 4. Visualization of normalized data sets using box plot. (A) Combined data (B) Meditation (C) Movie watching (D) Reading aloud

Classification based advanced machine learning algorithm was implemented to further analyze the EEG data from different brain states. The main challenge in this process
was the problem of data separation for each brain wave at different brain states. Generally, the brain waves data from different states tend to cluster together, which becomes tedious for classification algorithms to draw a best fitting separation line.

Figure 5. Standard deviation model for (A) Meditation, (B) Movie watching, and (C) Reading aloud brain states.

The classification algorithms applied along with the obtained accuracy score is shown in Table 3. Due to the complexity of data clustering, as an initial step of modeling, two low frequency waves - delta and theta brain waves were chosen as two variables and brain states- meditation, movie watching, and reading aloud were used as three nominal class values discretized as 0, 1, and 2 respectively. The results show that the K-Nearest model contributed the best prediction score with three classes - 78% for meditation, 64% for reading aloud, and 71% for watching a movie. And for two classes - 89% for meditation combined with movie, and 85% for reading brain states.

Interestingly, when two class systems was used by combining meditation and movie watching as class 0, and reading aloud as class 1, the algorithm performed very well with lower error rate (Table 2). This finding highlights the fact that the majority of volunteers participated for EEG data collection are inexperienced meditator, and the data collected during movie watching and meditation are close to one another than reading aloud. Also, once again with two class system, K-Nearest model performed the best. Since, K-Nearest model depends on highest number of neighboring data point to classify itself to that particular group, the clustering effect of meditation and movie watching should have contributed to the superior performance of the K-Nearest algorithm (Figure 6).

Further, the dynamic fast Fourier transform analysis of EEG data was conducted to reveal the occurrence of dynamic frequency at a steady state along with the time series (Figure 7). The result shows comparison of FFL graph of three different brain states from two volunteers. In experiences meditator, the graphs show the localized energy of waves whereas in inexperienced meditator volunteer the localization of wave energy is low. This finding underscores the additional layer of complexity for analyzing EEG data.

Table 2. Summary of classification based prediction scores

<table>
<thead>
<tr>
<th>Model</th>
<th>Brain State</th>
<th>Standard Deviation</th>
<th>K Nearest Neighbor</th>
<th>Support Vector Machine</th>
<th>Naive Bayes</th>
<th>Logistic Regression</th>
<th>Naïve Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three Class</td>
<td>0. Meditation</td>
<td>64.70</td>
<td>78.57</td>
<td>78.57</td>
<td>50.80</td>
<td>78.57</td>
<td>50.07</td>
</tr>
<tr>
<td>1. Movie Watching</td>
<td>75.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Reading Aloud</td>
<td>65.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two Class</td>
<td>0. Meditation and Movie Watching</td>
<td>-</td>
<td>89.29</td>
<td>82.34</td>
<td>-</td>
<td>89.71</td>
<td>-</td>
</tr>
<tr>
<td>1. Reading Aloud</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. Data analysis of brain waves. Standard deviation model for (A) Logistic Regression model three class (color: blue-meditation, orange-movie, and brown-reading), (B) Logistic Regression model two class (color: blue-meditation plus movie, and brown-reading), (C) K-Nearest model two class (color: pink-meditation plus movie, and blue-reading), and (D) Support Vector Machine with Linear Kernel two class (color: blue-meditation plus movie, and yellow-reading).

5 Web Visualization of EEG Data

We want to take a new approach through affective computing, which employs EEG signals recorded when users perform some brain activities and apply analytical algorithms to captured EEG data to detect the brain state. EEG signals can be measured at any moment and are not dependent on feelings, emotions, or human behaviour. We are investigating an automatic EEG-based recognition system that can record the EEG signals from users and measure their brain states. The EEG data are filtered to get separate frequency bands which are then analysed then displayed via a web server and web user interface.
5.1 Web Interface Visualization

The web server provides a user interface that allows users to view EEG data in the database and perform data analysis. Figure 8 and Figure 9 show the web interface of the data which is rendered in wave form mode and statistical mode, respectively.

The wave form rendering seen in Figure 8, is generated dynamically within the user’s browser. This allows the graph to be zoomed in (Figure 10) or moved around to allowed focus on a specific section of the graph.

5.2 Mobile Application

We have also developed an iPhone app. iPhone users can use the app that will connect to the database to view data. Figure 5 shows the two functions, viz., “collect data” and “view data”, that a user can choose on the iPhone app. The user can display data in text mode by viewing individual data frames (Figure 11(a)), or display the wave form of recorded data in certain time period (Figure 11(b)).

5.3 Web Collaboration

We have developed a forum to provide a way for researchers to communicate with each other through the website. The forum has been built using phpBB because of it is open source software and also because of its features. The forum has been divided into categories to better focus discussions within a category (Figure 12). The categories are currently Website, Modeling, and Off-Topic. Users can then post their message or reply within the appropriate category for other members to see. Users can also include in their post pictures or files that might help further their discussion. For example, they might include in their post an image of a wave graph from an analysis of their data from our website to discuss with other members.
6 Conclusions

The EEG headset, mobile application, web server, and web interface proved to be a good starting foundation for a proof of concept setup and to be able to visually see the EEG data capture and display from start to finish. It showed that it would be possible to capture EEG data from anywhere, while on the move, and be able to immediately see some of the results. Although much more refinement is required in increasing the performance of the data capture and reducing the electrical noise interference during the EEG capture, the system as built provides a good foundation for future improvements.

6 Future Works

The system that we have developed currently shows great promise for future implementation of more analysis algorithms. We are working on incorporating more of those analysis algorithms after developing and testing them first locally before inclusion on the website. We will also develop more visualizations methods for the website as necessary for different algorithms so as to represent results in the best visual appropriate. We also will further develop the mobile application to be more tightly integrated with the website. We also aim to address the complexity of classification of brain waves data by modeling the major brain waves independently with clinically significant brain regions combined with the time-series analysis. This will achieve an efficient and predictable brain wave modeling system which has potential application in hospitality and clinical industry for self-controlled deep brain relaxation and early diagnosis of various brain abnormalities respectively.

7 Acknowledgements

Gregor Schreiber and Minghao Yang contributed to the work presented in this paper. This work is partially supported by NSF CAHSI grant, NSF UGI-CSTEM grant, and NSF Summer REU grant.

8 References