Analysis and Visualization of EEG Data towards Academic Emotion Recognition

Dissertation Presented to
the Faculty of the College of Computer Studies
De La Salle University

in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Computer Science

by

Judith J. Azcarraga

Nelson Marcos, Ph.D.
Adviser

May 2014

Abstract

Brainwaves are analyzed, visualized, and are used to predict academic emotion based computational intelligence techniques, namely decision trees, Multi-Layered Perceptrons (MLP), k-means clustering, and Self-Organizing Maps (SOM). The brainwaves or EEG signals were collected from fifty six (56) academically-gifted students using an Emotiv EPOC sensor mounted on their head while they learn independently using two computer-based learning systems. Decision trees, as the baseline classifier, are able to classify the four academic emotions, namely frustrated, confused, bored, and interested, with prediction rates of only around 0.50. Multi-Layered Perceptrons are thus used to enhance the performance rates, achieving rates of 0.52 to 0.64 for Accuracy and 0.52 to 0.65 for Area Under the Curve (AUC). Prediction performance is also further improved by isolating the datasets based on gender, and/or handedness, and building separate classification models for each. Once separate classification models are built for the restricted sets, accuracy and AUC rates as high as 0.75 are achieved for datasets where all learners are only male and right-handed. Moreover, performance rates are shown to further improve when selective prediction is performed, i.e. only when the number of very high or very low feature values among the instances is high, such as when the number exceeds 20% of the 126 features used. As a further step in the analysis and visualization of the EEG data, unsupervised clustering is performed using k-means clustering and SOM. Using these two clustering methods, a new statistics-based method is proposed for determining the influence of the learners' profile, such as gender and handedness, on the formation of the clusters. Experimental results on the use of a novel structure for a SOM, referred to as structured SOM, are also presented. These experiments include a newly developed measure for quantifying the quality of a trained structured SOM. In addition to the new methods for the study of the k-means clusters, and the work on structured SOM, the detailed discussion on the manner by which the EEG data can be properly cleaned, prepared and pre-processed, including several methods for feature selection and dimensionality reduction, provide very useful information for future in-depth studies on EEG data and how they should be used to predict the academic emotions of learners.

Keywords : analysis of EEG data, emotion prediction, decision trees, Multi-Layered Perceptrons, k-means clustering, Self-Organizing Maps, structured SOM

Acknowledgment

This thesis would have remained a dream had it not been for the scholarship grant from the Philippine Council for Industry, Energy and Emerging Technology Research and Development of the Department of Science and Technology. Sincerest thanks to the Engr. Ermie Bacarra and and support staff of PCIEERD.

Special appreciation and thanks to my adviser Dr. Nelson Marcos and my mentor Dr. Merlin Suarez. I would like to thank you both for encouraging me in my research and for allowing me to grow as a researcher.

My deepest gratitude to the College of Computer Studies, particularly to the Center for Empathic Human-Computer Interaction and the Software Technology Department of De La Salle University for providing financial support for the presentation of my research papers at the various conferences that I attended, and for my summer training at Carnegie Mellon University.

Many thanks also to my colleagues/friends at CEHCI and the ST department, especially Ethel, Rhia, Jiji, Greg, Badong and Raffy, with whom I have spent some great time - with lots of laughter in spite of the serious and demanding work. In a special way, I also would like to thank Mac Cordell and my former students Nestor, John Boy and Ianne for helping me with some of the software modules I used in my thesis.

I would also like to thank my dissertation panel members, Dr. Raymund Sison, Dr. Prospero Naval, Dr. Florante Salvador, Dr. Nelson Marcos and Dr. Merlin Suarez for your brilliant comments and suggestions. Thanks also to Dr. Bob Legaspi and to my mentors at the Pittsburgh Science of Learning Center. Many thanks to Dr. Ryan Baker and Mr. Jean-François Nicaud for allowing me to use the Scooter and Aplusix software, respectively. And to Dr. Didith Rodrigo for providing the connection to Dr. Baker and Monsieur. Nicaud.

Special thanks to the students of De La Salle Canlubang (Integrated School) and of the Philippine Science High School (Main Campus), as well as to the administrators of these schools, especially Mr. Raffy Reloza, Mr. Mon Endriga, Dr. Helen Caintic, Ms. Jane Turingan, Ms. Tere Diaz, and the residence hall staff, for allowing me to collect my data.

Special thanks to my family. Words cannot express how grateful I am to my late father, my mother, my siblings, as well as to my friends and my relatives for your prayers and support.

And of course, to my three children Cid, Ken and Aly for all the sacrifices that you've made to free up time for me so I can finish my thesis. You were my inspiration, and your prayers and support were what sustained me. Finally, I would like to make special mention of my beloved husband Arnie who supported and guided me in writing this thesis, and most of all, who encouraged me to strive hard towards my goals.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1. Background of the Study

Students experience various emotions while engaged in learning activities and learning takes place through a complex interplay involving cognitive and affective dimensions (Pintrich & Schrauben, 1992). These emotions of students influence individual learning, memory and decision-making (Forgas, 1999) and are affected by the knowledge and goals of the student, and vice-versa (Mandler, 1999). A learner with positive affect tends to be more engaged and motivated to learn whereas a learner that experiences negative affect tends to be disengaged, unmotivated and would easily give up on the assigned task. Emotions experienced during learning, also referred to as academic emotions (Pekrun, 2002), may affect the flow of learning as well as the motivation to continue with the learning task. This has been the challenge for human tutors, as well as for designers and developers of learning systems. Indeed, effective tutors, whether human tutors or computer-based learning systems, are those who are not only aware of the cognitive needs of the students but also of their affective needs.

Affect recognition, therefore, is an important feature of computer-based tutoring systems (Mao & Li, 2010). Such systems can use the predicted affective state to control and time the various interventions in a tutoring scenario (Burleson, 2006). Moreover, awareness of the affective state of the learner can prevent a tutoring agent from intervening inappropriately (Conati & Maclaren, 2009) and from robbing the learner of the opportunity to discover the solution on their own (Burleson, 2006).

Indeed, recent developments in the design of tutoring systems or learning systems in general, have attempted to render these systems more adaptive not only to the learners' cognitive state but also to their affective state. In the case of these systems, usually referred to as affective tutoring systems, affective states such as "confident", "frustrated", "excited", "interested", "confused", "engaged" and "bored" may be recognized using the tutor-context and the user-profile (Baker et al, 2010; Beck, 2005; Chalfoun, Chaffar, Frasson , 2006) and sometimes in combination with signals from hardware sensors such as a camera, microphone (D'Mello & Graesser, 2010) and various other physiological sensors that capture EEG signals, electromyogram (EMG) signals, skin conductance levels, heart rate, and respiration rate (Arroyo et al, 2009; Conati & Maclaren, 2009; Blanchard, Chalfoun, Frasson, 2007; Burleson, 2006; Kapoor, Burleson & Picard 2007; Benadada et al 2008). Affective systems that use hardware sensors to recognize various affective states are based on emotion theories such as the Affective Infusion Model (AIM). According to this model, affective states can spread activation to related physiological and autonomic reactions, facial and postural expressions, verbal labels, action tendencies, and memories associated with that affect in the past (Forgas, 1999).

Natural episodes of emotions such as boredom, engagement and frustration cannot be reliably detected from the face (D'Mello & Graesser, 2010). For instance, frustration may not be obvious through facial expression and may be concealed due to some negative connotations to it and social display rules. Thus, alternate channels may be needed to detect the subtle expressions of such emotions (D'Mello & Graesser, 2010).

Physiological signals such as brainwaves and skin conductance may provide rich information about the current cognitive (Jausovec, 2000; Chanel, 2009; Kohlmorgen et al. 2007; Stevens et al., 2009) and emotional state (Arroyo et al, 2009; Benadada et al 2008 Burleson, 2006; Conati & Maclaren, 2009; Kapoor, Burleson & Picard 2007). Unlike a facial expression, such signals are difficult to mask and can capture natural emotional expression through physiological manifestations. A number of researches have attempted to use physiological signals such as brainwaves for detecting affective state (Chanel, 2009; Frantzidis et. al. 2010; Heraz, Razaki, & Frasson, 2007). An electroencephalogram (EEG) device is typically used for such purpose due to its low cost and portability. Such a device can read brainwaves and can measure the electrical activity in the brain induced by the electro-chemical processes related to the firing of neurons. Nevertheless, whether a given spike in neuron activities as captured by an EEG sensor is indeed induced by some emotion cannot be ascertained. Muscle movements near the eyes and forehead are typical noise/artifacts in EEG recording. Various other artifacts may also get (wrongly) captured. As explained later, serious care must be given to pre-processing EEG data in order to increase its signal-to-noise ratio and at some point be able to isolate segments of EEG signals, over some sustained period.

Brain activities have indeed been reported to provide some useful information as to emotion valence (positive or negative) and arousal (Frantzidis et. al. 2010; Heraz, Razaki, & Frasson, 2007; Chanel, 2009), user alertness and cognitive workload (Sanei & Chambers, 2007), stress level (Heraz et al., 2009), and level of frustration and distraction (Stevens, Galloway & Berka, 2007). In one neurophysiology research, negative emotions, such as "disgust" were found to be associated with right-sided activation in the frontal and anterior temporal regions whereas "happiness" was found to be associated with a left-sided activation in the anterior temporal region (Davidson, 1990).

As far as the methodology and experimental set-up for capturing EEG data from learners are concerned, most experiments on emotion-recognition based on brainwaves are conducted with the use of IAPS pictures (Chanel, 2009; Heraz, Razaki & Frasson, 2007; Petrantonakis & Hadjileontiadi, 2010), emotion-recall method and games (Chanel, 2009). Good accuracy results in emotion classification were achieved in these works. However, the experimental set-up in these studies are limited only to activities that are not related to learning, such as watching emotion-rich pictures, emotion recall and playing games.

Only a few have investigated the brainwaves of students while engaged in some learning systems. In (Stevens, Galloway & Berka, 2007), although the activity involves solving biology and math problems in a computer-based learning environment, there was no attempt to classify academic emotions as the study set out to measure cognitive states such as cognitive workload, engagement and distraction.

In a preliminary study, the level of academic emotions such as frustration, excitement and engagement was analyzed while students were engaged in solving mathematics and logic problems using turtle graphics. The subjects of this study were all academic achievers. However, in this preliminary study of (Azcarraga et al., 2010), it was the difficulty level of the task that was classified, based on the affective levels of excitement, frustration and engagement as indicated by the EEG system. Azcarraga et al. (2011a,b,c) show that indeed it may be possible to classify academic emotions based on EEG signals. However, the very specific features from brainwaves that may contribute to the onset of academic emotions have not yet been established and as such, a much more in-depth study of the brainwave patterns would be useful.

This in-depth study zeroes in on academic achievers, in order that other various academic-related variables such as general ability in mathematics and motivation as well as capacity for self-learning may be controlled. Indeed, the brainwaves pattern of academic achievers, who may be loosely referred to as the gifted learners, have been found to be different from average learners (Jausovec, 2000). This may not be surprising since a number of studies, particularly in psychology and education, have reported that gifted learners learn differently as compared to average ones. They are labeled as gifted because they tend to exhibit exceptional abilities compared to most other people of the same age. Gifted learners are characterized by high or advanced intellectual ability and/or outstanding talents. They have the potential to learn fast, to deal well with complex and abstract ideas and to have a large knowledge base (Diezmann, 2005; Feldhusen, 2005; Lee et al., 2006; Rayneri et al., 2006; Greene et al. ,2008).

With such abilities, these learners are typically given individualized instruction and can easily be engaged in self-regulated learning. Various computer-based learning systems can therefore be targeted at this special group of learners, since they can benefit much from self-regulated learning using individualized computer based learning systems. These learning systems, however, must be able to adapt to the academic emotions of their (academically gifted) users. And to be able to adapt to their academic emotions, these systems must be able to recognize them, and if possible, be able to determine in advance that a given user is likely to eventually get bored or frustrated or confused.

There is no current empirical work that has studied the brainwave patterns and their association with the affective learning behaviors of academically gifted students while engaged in some computer-based learning systems. This study fills this gap by doing an in-depth analysis and visualization of EEG signals collected from academically gifted students. The results of the study would certainly pave the way for designers of future learning systems to incorporate affect-related features in their system in order to respond positively to the special needs of learners. In addition, all the work on data collection, data preparation and pre-processing, as well as on classification and data visualization would be useful to any study of the use of EEG signals for the prediction of academic emotion.

1.2. Statement of the Problem

How should various pre-processing techniques and computational intelligence methods be employed, adapted and combined to predict learning-related emotions of academically gifted students based on their EEG data?

1.3. Research Questions

1. What are the relevant EEG features for classifying academic emotions and how can they be extracted from raw EEG signals?

2. How can patterns of extracted, selected and possibly transformed EEG signals be used to classify academic emotions?

3. As computational intelligence techniques for data analysis and visualization, how can supervised and unsupervised learning techniques be used, enhanced, and adapted to the kind of data extracted and processed from EEG signals?

1.4. Objectives of the Study

General Objective

To analyze and visualize EEG data collected from gifted learners, as they are using some computer-based learning system, and to associate patterns of EEG signals to specific academic emotions.

Specific Objectives:

1. To build a corpus of EEG signal data and user profile data from pre-selected gifted children using a polished data-collection system;

2. To recommend how best to adequately prepare the collected EEG data, such as the choice of sampling rate, normalization methods, increase in signal to noise ratio (removal of artifacts);

3. To implement various supervised and unsupervised machine learning and computational intelligence techniques for clustering and classification of the data in order to adequately classify the academic emotion of the learner based on EEG signals;

4. To characterize the relative importance of the different EEG channels based on unsupervised methods for clustering and visualization of the EEG data that would lead to an improvement in the performance of the supervised methods for classification;

5. To modify, enhance, and combine various computational intelligence techniques into a new affective model and to adapt the individual techniques to the specific nature of EEG signals; and

6. To assess how much a general affective model is able to classify academic emotions of individual learners (of possibly different personality profile).

1.5. Scope and Limitations

The participants for this study are first year (or Grade 7) and second year (or Grade 8) academically high-performing high school students with ages 11 to 14. EEG data and other useful information are gathered from fifty-six (56) high school students, 49 of which were students from the Philippine Science High School-Main Campus (PSHS-Main) while the other 7 belong to the top 10% of De La Salle-Canlubang. The collected data are subjected to various data pre-processing, feature extraction, and feature selection techniques prior to use in data analysis and visualization.

The data collection set-up is based on preliminary experiments conducted among a smaller number of college freshmen students from the College of Computer Studies of De La Salle University. Two different computer-based learning systems for high school mathematics, Aplusix and Scooter, are used during the data collection sessions. Aplusix is a learning system for algebra developed in Grenoble, France (Nicaud, Bouhineau & Huguet, 2002), while Scooter is a scatterplot learning software (Baker et al., 2006), developed as part of the Cognitive Tutor tutoring system. A software module for collecting simultaneous data signals from the EEG sensor and the self-reported emotions has been pretested and is the one deployed in the actual experiments.

The EEG sensor used in this study is the Emotiv EPOC sensor (http://www.emotiv.com). A commercial product typically used for gaming purposes, the Emotiv sensor is equipped with 14 channels based on the International standard 10-20 locations and is capable of assessing the level of several affective states (frustration, excitement, engagement, meditation). A less powerful version of this system was used in the preliminary studies.

Standard pre-processing and data preparation techniques are used, including computing deviations from baseline signals instead of using raw signals, Principal Components Analysis (PCA) (Sanei & Chambers, 2007) for feature extraction and selection, data normalization, and balancing of the datasets. The EEG signals are also carefully calibrated and synchronized with the other collected information, for purposes of tagging the signals with self-reported emotions.

The best sampling rate to be used (i.e. finding the appropriate window size for the computation of average EEG signals per segment) is set at 2.0 seconds. As for computational intelligence techniques, the classification performance rates are benchmarked against C4.5 decision tree, while Multi-Layered Perceptrons (MLP) are used to enhance the classification performance. For clustering, the *k*-Means method is used and would be combined with Self-Organizing Maps to enhance the data clustering, analysis, and visualization.

The RapidMiner is used to perform standard classification (http://www.rapidminer.com). Octave is used to program specific software modules for data normalization, pre-processing, and testing as well as for specific implementations of the Self-Organizing Maps and *k*-Means Clustering. In the study, the issues related to data storage, memory space, and computing time are considered but are not allowed to constrain the choice of methods or algorithms.

1.6. Significance of the Study

This study provides baseline data and information that would contribute significantly to the design of future affect-aware computer-based tutoring systems that adjust their choice and timing of tutoring interventions depending on both the cognitive and affective states of the learner. The affect-aware learning environment may automatically detect the level of engagement, boredom, frustration and other academic emotions of its intended users. These information may then be used by the system to provide the right feedback (i.e. appropriate system intervention and presentation of learning content) and to intervene at the proper time.

The study targeted at gifted learners, a specific group of learners that would benefit much from such systems since these types of learners often resort to self-regulated learning (Risemberg & Zimmerman, 1992). Considering the increasing number of specialized science and math schools (so called "science high schools") in the Philippines, the results of the study would be directly useful to the design of the delivery systems, and indirectly even the curriculum, of these specialized schools.

The contribution of this work is not only limited to the area of educational technology but also in the field of educational psychology as well. This study provides physiological basis for the study of the academic behavior of gifted learners, compared to data gathered from plain observation and self-reported emotions. In this study, every action performed or every affective state experienced by a learner is captured in real-time with the use of physiological sensors which are then recorded and automatically analyzed by the system.

Finally, although the study is targeting only the academically gifted students in mathematics of a very specific age group, the results of the study would open up a wide range of possible further studies on other types of learners, such as slow learners who may need help from alternative sources (e.g. intelligent tutoring systems specifically designed for them). The results of this study, particularly on the proper pre-processing of EEG signals and their use in a hybrid system (with components that supervised and/or unsupervised) for classifying academic emotion, provides important baseline information. It is expected that even among the group of slow learners, more specific studies would need to target specific age groups, or would need to tailor the system to a specific "cause for being a slow learner" (e.g. cannot focus attention on a task, low mental ability for specific subject areas, trauma from past learning experiences, inadequate preparation and training on prerequisite lessons).

Various other learning groups or sectors may need an affective model that would apply specifically to them, such as dyslexic or autistic children, artists and highly creative people, and adult learners like those who have retired from work and would like to learn about new areas of interest. Again, just like for the slow learners in school, the baseline information gathered from this study would provide essential and important information as to how EEG data must be prepared prior to classification, and how the pre-processed data may be used in some affective system that is designed and built following the general approach used in this study.

Chapter 2: Academic Emotions and Affect-Aware Learning Systems

2.1. Academic Emotions

Cognition affects emotion and emotion affects cognition. These relationships can be described by cognitive, attributional and network theories of emotions. One network theory that explains the link between affect and cognition is the Associative Network Model. This model suggests that 'affect and cognition are integrally linked within an associative network of cognitive representations' (Forgas, 1999). This means that affect influences how we perceive the world, process information and the way we use these two in the performance of cognitive tasks. This model assumes that some affective nodes are biologically wired into the brain, activated by a range of situational triggers which become greatly elaborated as a result of cultural learning. Also, affective states can spread activation to related physiological and autonomic reactions, facial and postural expressions, verbal labels, action tendencies, and memories associated with that affect in the past.

Another theory that describes the relationship between cognition and emotion is the Affect Infusion Model (AIM). Affect infusion is a process where affective information influences people's constructive processing, learning, memory, attention and associative processes and eventually the outcome of their "deliberations in an affect-congruent direction" (Forgas, 1999). The model defines four processing strategies an individual may choose in response to target, social judgment and situational features. These are 1) direct access, 2) motivational processing, 3) heuristic processing, and 4) substantive processing strategies. Factors influencing processing choices include task familiarity, complex and typicality of task, personal relevance, personal motivation, processing capacity and mood effects on processing. Processing asymmetry is present and may due to evolutionary/functional reasons, capacity effects and motivational effects.

Another cognitive theory that is commonly used in affective systems is the Ortony, Clore and Collins (OCC) Model. According to this theory, emotions are derived from the cognitive appraisal of the current situation, which consists of events, agents and objects, based on the OCC cognitive theory of emotion. The outcome of the appraisal depends on how the situation fits with one's goals and preferences (Steunebrink, Dastani & Meyer, 2009).

Emotions, in general, may be assumed to influence students' cognitive processes and performance, as well as their psychological and physical health. Pekrun (2002) refers to emotions that are directly linked to academic learning, classroom instruction and student achievement as academic emotions (as shown in Table 2.1), describes the domain of academic emotions in terms of positive and negative emotions and in relation to student achievement, instruction or process of studying. It was demonstrated that that academic emotions are significantly related to students' motivation, learning strategies, cognitive resources, self-regulation, academic achievement and to personality and classroom antecedents (Pekrun, 2002).

Table 2.1. The Domain of Academic Emotions as defined by Pekrun (2002)

The Domain of Academic Emotions: Examples

	Positive	*Negative*
Task-related and self-related		
Process	Enjoyment	Boredom
Prospective	Anticipatory joy	Hopelessness
	Hope	Anxiety
Retrospective	Joy about success	Sadness
	Satisfaction	Disappointment
	Pride	Shame and guilt
	Relief	
Social		
	Gratitude	Anger
	Empathy	Jealousy and envy
	Admiration	Contempt
	Sympathy and love	Antipathy and hate

The work on AMIBEL by Kort et al. (2001) models the impact of emotions on student learning, as shown in Figure 2.1, particularly working in the domains of science, math, engineering and technology (SMET). In a learning scenario, it is normal for the student to move from one quadrant to another. The typical movement is in a counter-clockwise direction. This implies that a typical learning experience involves a range of emotions, moving around the space, and moving from one quadrant to another, as they learn. According the model, it is also normal to be in two quadrants simultaneously (e.g. interested and frustrated).

The AMIBEL model recognizes that a student ideally begins in quadrant I or II when they are curious and fascinated on a new topic (quadrant I) or they might be puzzled and motivated to reduce confusion (quadrant II). While in any of these quadrants, the student is still in the process of constructing or testing knowledge.

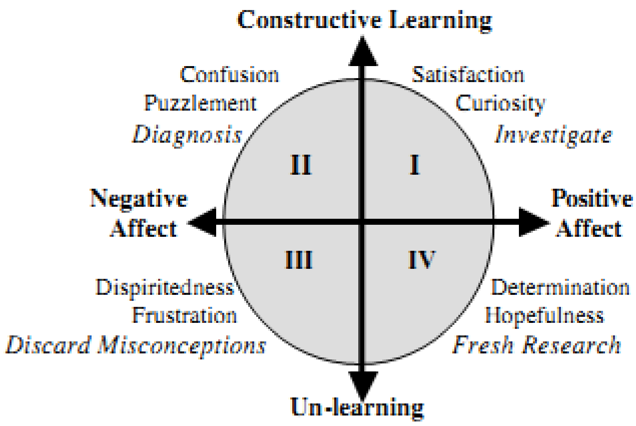


Figure 2.1 Affective Model of Interplay between Emotions and Learning (Kort et al 2001)

2.2. Affective Learning Systems

Affect recognition is an important feature of a computer-based tutoring system (Mao & Li, 2010). Systems can use the affective state to control and time intervention in a tutoring scenario (Burleson, 2006). Moreover, affective awareness can prevent a tutoring agent from intervening appropriately (Conati & Maclaren, 2009) and from robbing the learner's opportunity to discover the solution on their own (Burleson, 2006). Also, user satisfaction is seen as an important factor in an affective tutoring systems (Mao & Li, 2010).

In a computer-based learning environment, academic emotions may be recognized by using contextual information such as the details of the user-tutor interaction, and sometimes in combination with hardware and physiological sensors for better prediction accuracy. Table 2.2 presents recent works on affective systems. For example, Arroyo and his group were able to predict affective states such as "confident", "frustrated", "excited", and "interested" with high accuracy using such special devices such as a facial expression camera, posture chair, pressure mouse, skin conductance (Arroyo et al. 2009). A similar multimodal system is the "learning companion" (Kapoor, Burleson & Picard 2007) that fuses information from cameras, posture chair, pressure-sensitive mouse, skin sensor and task state to predict "frustration" - in order to determine when the user needs help. Finally, "Autotutor" uses information from conversation cues, posture and facial features to predict student boredom, flow/engagement, confusion and frustration (DMello & Graesser, 2010).

2.3. Brainwaves and Academic Emotions

The human brain is made up of about 100 billion of neurons that are interconnected through synapses. It is divided into four lobes, namely: the frontal, parietal, temporal and occipital lobes. Each of this is specialized in particular cognitive tasks. The frontal lobe is the region dedicated to movement, thinking initiation, reasoning or judgement, behaviour or emotions, memory and speaking. The parietal lobe is involved in sensation, reading, understanding of spatial relationship and knowing right from left. The temporal lobe is involved also in behaviour, memory, hearing and understanding language. The occipital lobe is the region involved in visual tasks.

Neurons communicate with other neurons through action potentials that produce some form of electrical signal. Such electrical signals can be captured and measured by an electroencephalogram (EEG) device.

Individuals who experience negative emotions particularly withdrawal behaviors were found to have greater right-hemisphere activation in the frontal and anterior temporal regions whereas those who experience positive emotions particularly approach behaviors were found to have greater left-hemisphere activation in the anterior temporal region (Davidson 1990).

Table 2.2 Recent Related Work on Affective Learning Systems

Reference	Type & number of participants	Elicitation methods	Sensors/ signals/ contextual info	Emotion classes	Methods (classifiers, feature selection, etc.)	Best results
Burleson (2006)	61 11-13 yr olds	Tower of Hanoi problem	camera, posture chair, skin conductance, pressure mouse	frustration, help-seeking	ANOVA, HMM, SVM, Dynamic Bayesian, kNN, Gaussian Process	79.17%
Arroyo et al. (2009)	38 high school 29 undergraduate students	Science problem (Wayang Outpost)	camera, posture chair, skin conductance, pressure mouse	confidence, frustration, excitement, interest	Linear Regression	69%
DMello & Graesser (2010)	28 students	(Autotutor)	conversation cues, posture and facial features	boredom, flow, engagement, confusion, frustration, neutral	Latent Semantic Analysis (LSA), Feature-level fusion LDA	50.6%
Conati & Maclaren (2009)	41 Grs. 6-7	Math problem (Prime Climb)	skin conductance, electromyogram (EMG), heart rate	Joy, distress, reproach, admiration	Dynamic Decision Network, conditional probability theory	73%
Kapoor & Picard (2005)	11-13 yo	Fripples Place	camera, posture chair	interest, disinterest	Gaussian Process	86%
Benadada, Chaffar & Frasson, (2008)	29 graduate students	learning sorting algorithms	skin conductance, EMG, heart rate, respiration, personality profile	Measures anxious, confident, boredom, intrigue, relief, disappointment, joy	ANOVA	n/a

Designed for clinical diagnosis, EEG sensors have recently been explored for measuring user alertness, cognitive engagement (Sanei & Chambers, 2007) and affect detection. In (Stevens et al., 2007), for example, the student's level of frustration, distraction and cognitive workload may be observed while the student is engaged in a multimedia learning environment. These cognitive and affective states seem to influence student problem solving skill. Heraz & Frasson (2009) explored the accuracy of using EEG and emotional dimensions in predicting the correctness of student's answer. In a similar work, Heraz, Razaki & Frasson (2007) investigated the use of brainwaves data to predict the emotional state of the student. Affect such as anger, boredom, confusion, contempt, curious, disgust, eureka, and frustration can be accurately detected (82%) from brainwaves (Heraz and Frasson 2007).

While raw EEG data usually contain artifacts such as eye-blinks and muscle movement, one approach to reduce such noise is to transform the data to different frequency bands (or frequency domain) (Heraz, 2007; Lee & Tan, 2006) and/or apply several filtering techniques (Petratonakis & Hadjileontiadi, 2010). Useful features may then be easily extracted from these transformed data.

One challenge of using EEG for affect detection is the number of channels (14-60) typically attached to an EEG headset. Not all channels provide useful information for emotion assessment. Petratonakis & Hadjileontiadi (2010) have limited the channels at the frontal lobe since EEG signals in this area tend to be more intense in the presence of positive or negative affect. Moreover, Ansari-Asl et al (2007) proposed an approach for automatically reducing the number of EEG channels to use based on synchronization likelihood.

However, most works have observed EEG data of subjects involved in audio and/or visual stimuli such as pictures (Heraz et al, 2007; Petrantonakis et al, 2010) and video clips in (Murugappan et al, 2009). Some have investigated EEG patterns on different cognitive tasks while at rest, performing mental arithmetic and mental rotation, playing in solo and with an opponent (Lee & Tan, 2006). Stevens et al (2007) have developed a system that allows EEG data to be observed while students are in an actual multimedia learning environment. However, this research has only conducted preliminary investigations of the EEG data from the participants. At the moment, no work has been done yet that developed a model that correlates brainwaves patterns with academic emotions such as boredom, frustration, engagement, confusion, excitement based on the data from learners in an actual learning environment. Table 2.3 summarizes the works on brainwaves analysis that use EEG.

2.4. Brainwave Pattern of Gifted Students

Studies reveal that gifted learners tend to be different from non-gifted learners in terms of brain signal patterns. Different patterns of brain activities between these two kinds of learners were reported in the work of Jausovec (2000). High alpha power (efficient neural firing and less mental effort) were seen on high IQ students while low alpha power (increase mental effort) were seen on low to average students while solving a closed-type problem. Table 2.4 presents studies that show brainwaves pattern of gifted learners.

2.5. Some Critical Gaps in the Existing Literature

Academic emotions are emotions commonly experienced by a student in a learning scenario. It can be described based on their dimensions (valence, arousal and dominance) or in quadrant. It is typical for students to experience a range of emotions and move around the emotion space as they learn. Thus, it is it is very important for a tutor to identify at which quadrant(s) the student is to be able to control and time the appropriate intervention.

In a computer-based learning environment, academic emotions may be recognized by using contextual information such as the details of the user-tutor interaction, and sometimes in combination with hardware and physiological sensors (EEG, EMG, skin conductance, respiration, heart rate, posture chair, biometric mouse) for better prediction accuracy. Designed for clinical diagnosis, EEG sensors are recently explored for measuring user alertness, cognitive engagement and affect detection. However, at the moment, no work has been done yet that developed a model that correlates brainwave patterns with academic emotions such as boredom, frustration, engagement, confusion, excitement based on the data from learners in an actual learning environment.

Despite the positive results that have been reported on the successful use of affective tutoring systems, much remain to be explored. In particular, affective tutoring systems must take into account the profile of the learner, such as whether they are fast learners, slow learners, or average learners. Indeed, students such as the gifted group or the fast learners must be provided with a differentiated program that caters to their unique learning needs and capabilities. Expectedly, differences in brainwaves pattern were seen among highly intellectual individuals and their counterpart.

Table 2.3 Recent Works on Brainwaves Analysis

Reference	Type & number of participants	Elicitation methods	Emotion classes	Methods (classifiers, feature selection, etc.)	Best results
Chanel (2009)	4 adults 11 adults 20 adults	International Affective Picture System (IAPS) Self-induced emotions computer game (Tetris)	Low and high emotion arousal and valence	ANOVA, FCBF, SFFS, Fisher criterion, Naïve Bayes, (D/Q)LDA, (L/R/RBF) SVM	72% 80% 56%
Heraz, Razaki & Frasson (2007)	17 undergraduates	IAPS	anger, boredom, confusion, contempt, curious, disgust, eureka, frustration	kNN	82.27%
Stevens, Galloway & Berka (2007)	12 Grs. 8-10	biology and math problems (IMMEX)	Measures cognitive workload, engagement, distraction	Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA)	n/a
Murugapann, Ramchandran & Sazali (2010)	25 university students	video clips	disgust, happy, surprise, fear, neutral	Surface Laplacian filtering, Discrete Wavelet Transform, KNN, LDA	83.26
Petrantonakis & Hadjileontiadi (2010)	16 adults	pictures	Happiness, surprise, anger, fear, disgust & sadness.	Hybrid Adaptive Filtering, Higher-Order Crossings Analysis, QDA, SVM, kNN, Mahalanobis Distance (MD)	83.33%
Lee & Tan (2007)	8 adults	Relaxation, mental arithmetic and rotation game in solo and with opponent (Halo shooter game)	Task classification	Wrapper-based feature selection, Bayesian Network	84% 92.4%
Jausovec & Jausovec (2000)	27 undergraduates	Solving problems with different complexities and type	Analysis of EEG power Differences in cognitive processes	EEG power (upper and lower alpha) General linear model	n/a
Kohlmorgen et al. (2007)	17 adults	Simulated driving with a) auditory task b) mental calculation task	Low and high mental workload	LDA	(1 subject) 95.6% & 91.8% Unspecified average for all

Table 2.4 Brainwaves Pattern of Gifted Learners

Reference	Research	Type & number of participants	Elicitation methods	Methods (classifiers, feature selection, etc.)
Seung-Hyun, Kwon, Jeong, Kwon, & Shin(2006)	Differences in brain information transmission between gifted and normal children during scientific hypothesis generation	11-12 yo gifted and non-gifted	problem solving (Quail Egg)	ANOVA, Averaged-Cross Mutual Information
Seung-Hyun, Soo, Kyung & Kil-Jae (2007)	Differences in EEG between gifted and average students and brain areas related to memorizing.	gifted & same age average students	memory problem (ROCF)	functional clustering
Jausovec (2000)	Differences in cognitive processes between gifted, intelligent, creative, and average individuals while solving complex problems	48 & 49 gifted, intelligent, creative, and average university students	creativity problems & problems with different complexities	ANOVA

Chapter 3: Data Collection and Pre-Processing

3.1. Data Collection from Gifted Students

The first preliminary study conducted among freshmen and sophomore undergraduate students who excelled in their mathematics courses was reported in (Azcarraga, Suarez & Inventado, 2010). In this study, the academic achievers used as human subjects in the experiments were asked to learn LOGO, a programming language that teaches turtle geometry, while an EEG sensor was attached to their head. Subsequent studies have investigated the EEG signals as well as the mouse clicks and movement of undergraduate students as they learn Scooter, a scatterplot tutoring system (Azcarraga et al., 2011a) and Aplusix, an algebra learning software (Azcarraga et al., 2011b; Azcarraga et al., 2011c; Ibañez, Lim & Lumanas, 2011). Using the preliminary results from the initial runs on college students, the set-up for the collection of data from the target gifted learners was designed.

The participants for this new study were first year (or Grade 7) and second year (or Grade 8) academically high-performing high school students with ages 11 to 14. They mostly came from the Philippine Science High School – Main Campus. All the participants in the study were presumed to be academically gifted because to enter this school, the students have been selected on the basis of their excellent overall academic performance in Grade School, and for having passed a highly selective entrance examination.

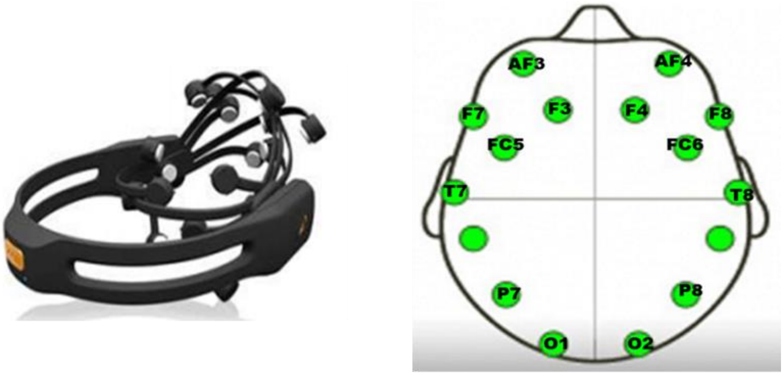
Fifty six (56) high school students participated in the experiment. Forty-nine (49) of them were students from the PSHS-Main Campus while the other 7 belong to the top 10% of De La Salle-Canlubang. Table 3.1 presents the number of students from each school, their gender and their hand-dominance. However, only 52 of them were found to be useful since 4 of them are not useable due to noise and incomplete self-report.

Table 3.1 Demographics of the participants

	Male	**Female**	**Right-Handed**	**Left-Handed**	**Total**
DLSU-STC	4	3	4	3	7
PSHS	32	17	44	5	49
Total	36	20	48	8	**56**

3.2. Experimental Set-up

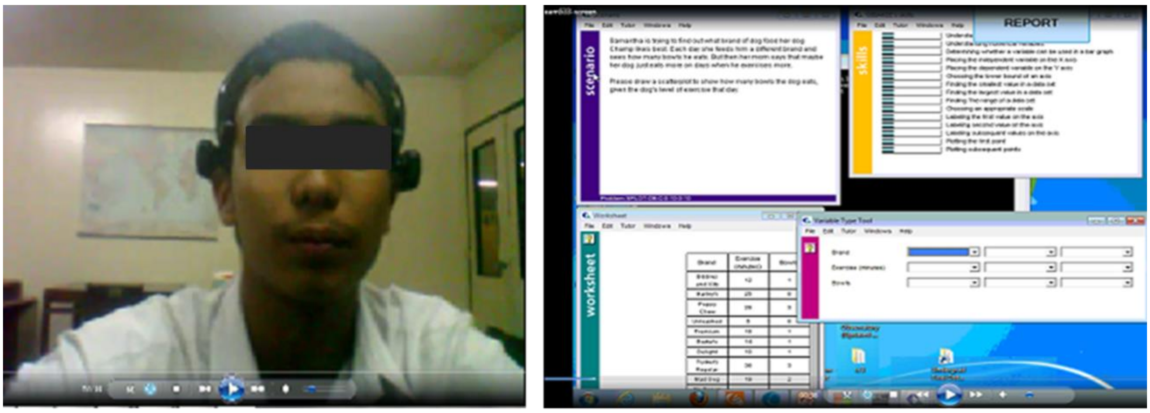
The EEG sensor that is used in this study is the Emotiv EPOC sensor. A commercial product typically used for gaming purposes, the Emotiv EPOC sensor is equipped with 14 channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) based on the International standard 10-20 locations. It is also capable of assessing the level of several affective states such as frustration, excitement, engagement and meditation. Figure 3.1 shows the (a) Emotiv EPOC headset and the (b) sensory channels location and label.



a) b)

Figure 3.1 a) Emotiv EPOC EEG headset and b) Emotive EPOC Channel Locations

Modules for capturing raw EEG data and the level of affective states produced by Emotiv have been implemented prior to the experiment and were used during the data collection experiments. Figure 3.2 shows a screenshot of a student wearing an EEG headset while solving a scatterplot problem. To establish the user profile of each student subject, questionnaires were given for the assessment of their personality type (e.g. introvert or extrovert), self-theory of intelligence, and to obtain other relevant information, such as gender and hand-dominance (left-handed or right-handed).



a) b)

Figure 3.2 a) Screenshot of a student wearing an Emotiv EPOC headset; and b) his current task in the Scooter scatterplot system

Since this study involves the participation of minors as human subjects, informed-consent forms were sent out to students and their parents following the ethical research guidelines of DLSU when human subjects are used in experiments. Prior to this, the school principal and administrators were asked of their approval to conduct the experiment. Please refer to the Appendix for a sample Informed-Consent form and the official letters to the school.

Two different computer-based learning systems for high school mathematics, Aplusix and Scooter, were used during the experiments. Aplusix (Nicaud, 2002) is an algebra learning software, while Scooter (Baker, 2008) as mentioned earlier, is a scatterplot learning software. Modules for collecting simultaneous data signals from the EEG sensor, the user screen, and for self-reporting of emotions were deployed.

The selected samples of academically-gifted students were subjected to a calibrated, video-taped learning session using the two learning systems, equipped with emotion-sensing devices that are all connected to a computer to record all the simultaneous data signals. Sensing devices which were used in order to capture various student affect include an EEG sensor (Emotiv EPOC set mentioned above) and a video-camera. Events of the learning systems (i.e. task presented, difficulty of the task, etc.) and student actions (i.e. response time, requested hints) were automatically being recorded in log files. The built-in camera of the laptop was capturing all the facial expressions, while another camera was set-up at the back of the learner to record the user-machine interaction and the screen displays.

Each experiment session with one student would last for about an hour. Below is the sequence of activities for each session.

1. Brief introduction about the research and its purpose, and what to expect during the experiment (3 mins)
2. Personality test and Self-theory of intelligence test (5 mins)
3. Setting-up of EEG sensor and modules for capturing the EEG signals, screen video and facial expression (3-7 mins)
4. Neutral activity (staring at a black screen with minimal movement) to capture the resting-state EEG of the participant (3 mins)
5. Slide presentation on how to make a scatterplot (10-13 mins)
6. Solving a scatterplot problem using the Scooter software (10 mins)
7. Slide presentation on how to use the Aplusix algebra software (3 mins)
8. Solving of 6 algebra problems ordered from easy to difficult (10 mins)

For the activities from #5-#8, each participant was asked to report the level of their confusion, boredom, frustration, interest and difficulty of the task by clicking on a sliding bar. The participants were instructed to self-report every 2 minutes, after solving each problem and whenever they sense that there is a change of emotion or task difficulty. The use of this slide bar had been pre-tested and the length of time interval (two minutes) had been calibrated so as not to be too intrusive, but at the same time would be frequent enough to detect any changes in academic emotion. The students have been asked whether they were bothered by the sliding bars and they have consistently responded that they were not, and that the sliding bars were intuitive enough to use with ease. No visible irritation with the sliding bar had been observed in the course of the experiments.

The EEG features together with user profile features and self-reported task difficulty and emotions were collected to form the gifted learners' corpus – on which all the subsequent data analysis, clustering and visualization were performed. The four affective states that are typically experienced in a learning system scenario, namely *boredom, frustration, interest* and *confusion* were used. These four affective states are referred to as academic emotions and are the target of classification and prediction.

As for clustering and data visualization, the two learner profile information that are used to complement the EEG signals collected using the Emotiv Epoc headset are the 1) gender (male or female); and 2) hand-dominance (right-handed or left-handed). These two plus the individual EEG signal features are the object of data clustering and visualization. Note that various other data, such as facial expressions using the front camera video, and the personality profile (introvert and extrovert), were not used.

3.3. Data Preprocessing and Data Preparation

The various data pre-processing, preparation, feature extraction, feature selection and computational intelligence techniques employed prior to emotion classification and data analysis were mostly well-known and have been applied to different types of datasets. The procedural flow from data capture to data preparation to feature extraction and selection up to classification, clustering and visualization is shown in Figure 3.3.

The EEG signals are sampled by the Emotive device around 2,048 times per second (per channel). However, the system automatically filters the samples and reduces the original 2,048 samples to 128 samples per second by performing moving average.

According to psycho-physiological literature (Levenson, 1998; Ekman, 1984), emotions persist for about 0.5 to 4 seconds. Guided by this, we used 2-second window samples. All the pre-processed EEG data and self-reported emotion tag were carefully synchronized, merged and uniformly segmented into 2-second windows with 1-second overlap. Each sample was transformed into alpha, beta and gamma frequency bands. Each sample or segment was treated as a single instance in each subject's dataset.

Brainwave signals are typically within the frequency range of 0-45 Hz. There are five major frequency bands, namely: alpha (α), theta (θ), beta (β), delta (δ) and gamma (γ). Sanei & Chambers (2007) describe the relevance of each of the frequency ranges and what each type of wave would generally account for. In particular, we have the following:

1. Delta waves lie within the range of 0.5-4 Hz. These waves are associated with deep sleep.

2. Theta waves lie within the range of 4-7.5 Hz. These are associated with deep meditation. A theta wave is often accompanied by other frequencies and seems to be related to the level of arousal.

3. Alpha waves lie within the range of 8-13 Hz. Characterized with round or sinusoidal shaped signal, alpha waves are associated with a relaxed awareness without any attention

or concentration. The alpha wave is the most prominent rhythm in the whole realm of brain activity.

4. Beta waves lie within the range of 14-26 Hz. These waves are associated with active thinking, active attention, focus on the outside world or solving concrete problems.

5. Gamma waves occur above 30Hz, mainly up to 45 Hz. These are typically used for confirmation of certain brain diseases.

Since EEG data usually contain non-desirable artifacts/noise that seriously degrade the quality of the data signals, high-pass, low-pass filtering and Moving Average (Hyndman, 2009) techniques were employed to clean up the data. Otherwise, if used directly for classification of the academic emotion, these signals would be incorrectly classified. Note that a low-pass filter is a filter that passes low-frequency signals but attenuates or reduces the amplitude of signals with frequencies higher than the cut-off frequency. A high-pass filter does the opposite.

Artifacts may also be removed by EEG data transformation i.e. frequency transform followed by the application of low-pass and high-pass filters. Artifacts due to actions such as eye movement/blinking and heart-functioning are most dominant below 4Hz. Muscle artifacts affect the EEG spectrum above 30Hz while non-physiological artifacts caused by power lines are clearly above 30 Hz (50-60 Hz).

The filtered, cleaned up EEG data were encoded into different frequency bands for subsequent feature extraction and once pre-processed, the pertinent features were extracted and analyzed. Since the data from the EEG signals were voluminous, the features were pre-selected and rendered in a form and format that would satisfy the requirements of the data classification, clustering and visualization techniques that would be employed later.

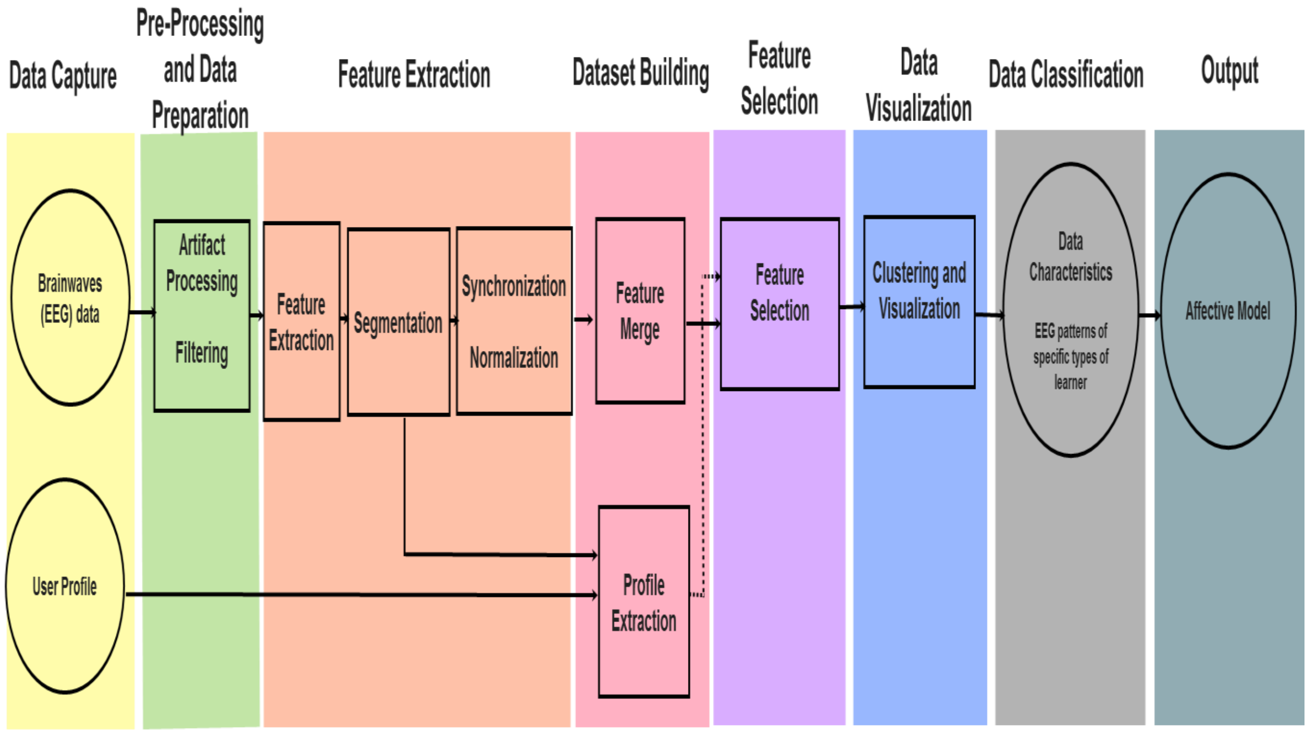


Figure 3.3 Procedural Flow of Data Preparation, Data Analysis and Data Visualization

3.4. Feature Extraction and Data Normalization

Feature extraction methods require parameters such as window length, frequency and topography of channels, and the ideal settings are typically unknown (Pfurtscheller et al., 2007). To identify such setting, different feature selection methods were employed by defining a subset of features from a large pool of features extracted from different feature extraction methods with different parameter settings and channels. With this, feature selection was applied to identify suitable feature extraction methods and to determine their parameter settings as well as to find the appropriate electrode positions.

Guided by previous studies on the analysis of EEG data, only the alpha, beta, and gamma frequency bands were used. Each signal coming from each the 14 channels positioned at specific locations on a student's head, are separated into 8 specific signals, which includes the Peak Magnitude and the Mean Spectral Power for each of the alpha, beta high, beta low, and gamma frequency bands.

The specific feature extraction method is as follows. To extract the specific EEG features, raw data from the 14 EEG channels were segmented into 2-second window samples, with 1-second overlap. Each sample was filtered and transformed into four frequency bands: alpha (8-12 Hz), low beta (12-21 Hz), high beta (21-30 Hz) and gamma (31-50 Hz), using built-in functions of Octave, an open-source high-level language for numerical computations.

Two features were extracted from each frequency band of each EEG channel: the peak magnitude and the mean spectral power. The peak magnitude is the highest magnitude or amplitude of the sample. The mean spectral power is the average power spectrum of the signals in the sample. This feature was used in one brainwaves research of Jausovec (2000). Aside from these, other features such as the average energy over brain areas (frontal, parietal, temporal, occipital), brain asymmetry scores or lateralization of the 7 right-left electrode pairs from the alpha band and the energy of beta for all the electrodes, were extracted (Chanel, 2009). All told, a total of 126 features were extracted from each segmented raw sample.

Furthermore, two sets of EEG data for each participant were processed. The first set is the EEG data taken during the 'resting-state' while the other is gathered during learning session. During the resting-state period, the values of each FFT-transformed feature for each student/subject were averaged. The average value serves as the baseline EEG of that particular subject (Davidson et al., 1990; Tomarken et al., 1992). Guided by the methodology in psycho-physiological research on emotion (cf. Davidson et al., 1990), the data taken during the learning sessions were converted into deviations from the baseline EEG of the 'resting-state' session. Figure 3.4 illustrates how raw signals are converted into positive or negative deviations from the baseline EEG signal.

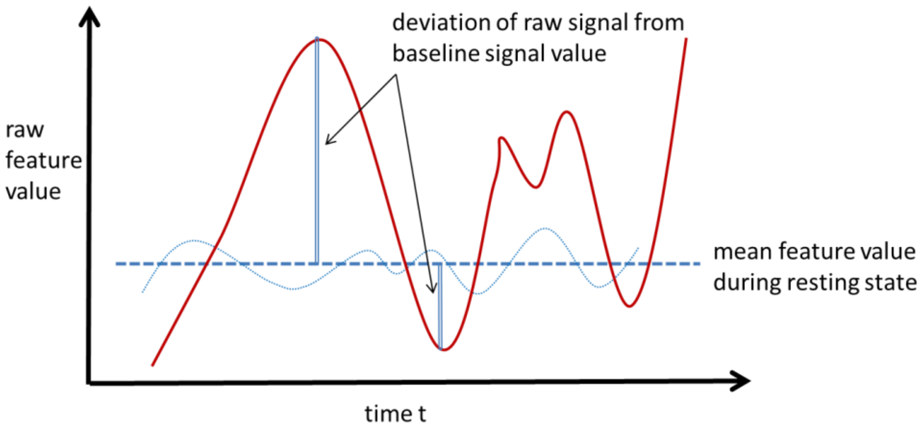


Figure 3.4 The raw signals per feature (red curve) are converted to deviations from the baseline EEG signal value for the given feature. The baseline EEG signal value of a given feature, and for a given student/subject, is the average value (horizontal dashed line) of the given feature taken during the entire 'resting state' session.

Each converted EEG signal value $\mathbf{f_d(i,s)}$ of student **s**, encoded as the deviation of the **i**th feature value from the baseline signal value, are then further processed and normalized. They are converted to standardized z-scores $\mathbf{f_z(i,s)}$ by computing the standard value according to eq. 3.1, and then extreme z-scores are clipped to at least -3 or at most +3. The standardized z-scores are thus normalized values within the range [-3, +3], with a mean of 0 and a standard deviation of 1.

$$f_z(i,s) = (f_d(i,s) - \mu(i)) \,/\, \sigma(i) \tag{3.1}$$

where $\mu(i)$ is the mean of the ith feature of all students in a given session, and $\sigma(i)$ is the standard deviation of the ith feature for all students in a given session.

3.5. Self-Reported Emotions and Assignment of Instance Labels

Each EEG segment or sample, with a total of 126 feature values, is considered an instance in each subject's dataset and each instance is labeled according to the self-reported emotions. To be more specific, let the first, second and third self-reports occur at time t(i), t(j), and t(k), respectively. Let t(q) and t(r) be half-way between t(i) and t(j) and half-way time between t(j) and t(k), respectively. The instances that occurred after t(i) but before t(q) take the self-rated emotion reported at t(i). The instances after t(q) but before t(r) take the self-rated emotion reported at t(j), and so on.

Self-rated emotions range in value from 0 to 100. Emotion rates for Frustrated, Confused, Bored and Interested were discretized to either 'low' or 'high' which are then used to label each instance. The label was 'low' if emotion rate was below 50, otherwise, the label was set to 'high'. Another labeling technique was performed in some experiments, particularly in SOM. The most *dominant emotion* was also used to label the instances. The most dominant emotion is identified as the emotion with the highest rate among the 4 emotions, with the condition that the value of that particular emotion is above 50. The labels are **F**, **C**, **B** or **I** if the dominant emotion is Frustrated, Confused, Bored or Interested, respectively. In cases when all emotion values are below 50, the label was set to **L** (low). On the other hand, if there are multiple dominant emotions, when there are more than 1 highest emotion values and are all above 50, the label was set to **M** (multiple).

3.6. Feature Selection and Principal Components Analysis

Various feature selection techniques were employed in the experiments. One technique performed the selection of features by statistical means. The other one adapted the Principal Components Analysis for feature selection.

Principal Components Analysis (PCA) is a widely used technique for EEG data filtering and feature extraction (Sanei & Chambers, 2007; Dornhege, et al., 2007). It is also used to reduce the dimensionality of the feature space. When used in filtering applications, the subspaces for signals and noise are separated by means of some classification techniques and the data are reconstructed from only the eigenvalues and eigenvectors for the actual signals (Figure 3.4). It may also be used for blind source separation of correlated mixtures if the original sources are statistically uncorrelated. The basic concepts of PCA described here were based mainly on (Sanei & Chambers, 2007).

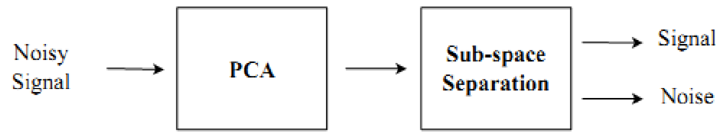


Figure 3.5. The general application of Principal Components Analysis (PCA) to separate the signal from the noise

To reduce the dimensionality of the feature space to p dimensions, optimal approximation of the data x_k where $x_k \in \mathbb{R}^m$ for $k = 1,..., n$ can be computed as

$$x_k \approx b + W a_k \quad \text{where} \quad b \in \mathbb{R}^m, \quad a_k \in \mathbb{R}^p, \quad p \leq m, \text{ and } \quad W \in \mathbb{R}^{m,p} \tag{3.2}$$

If this optimization is done by minimizing the squared error of eq. 3.3

$$\sum_{k=1}^{n} \| x_k - (b + W a_k) \|_2 \tag{3.3}$$

and simultaneously fixing the diagonal of $W^{\top}W$ to 1, the solution is found by choosing b

$$b = \frac{1}{n} \sum_{k=1}^{n} x_k \qquad (3.4)$$

W being the eigenvectors of the highest p eigenvalues (suitably scaled) of the so-called scatter matrix

$$\sum_{k=1}^{n} (x_k - b)(x_k - b)^{\mathrm{T}}, \text{ and} \qquad (3.5)$$

$$a_k = W^{\mathrm{T}} (x_k - b) \qquad (3.6)$$

Consequently, W is composed of orthogonal vectors, describing the p-dimensional subspace of $\mathbb{R}^m$, which shows the best approximation to the data. For normal distributed data, one finds the subspace by examining the covariance matrix, which indicates the direction with the largest variation in the data.

PCA transforms the 126-feature space of the EEG data set into an input space that is as small as a 3-component space, where each component is a combination of several of the original features. These components can be used as the new input features, and the size of the dataset is thus substantially reduced. The problem, however, is that once components are used, we are no longer able to visualize how these inputs would relate to the output. In the experiments reported in Chapter 5, the top 10 features in each of the top 3 components extracted through PCA are used as the reduced feature set, with less than 30 features adequately representing the original 126 features.

Chapter 4: Classification, Clustering and Visualization Techniques

4.1. Classification and Prediction

Appropriate computational intelligence and machine learning techniques are applied on the extracted data in order to classify them into appropriate categories and classes, and to predict what categories and classes unknown data instances would belong to. Supervised training of classifiers are first trained with samples, instances or sample data that have accompanying class labels. Once trained, appropriate classification models would have been built which can now be used to predict the classes on unknown data.

The performance of such classifiers essentially boils down to measuring how much correct and incorrect predictions are made, although just on performance measures, a whole range of possible metrics to used are discussed in the literature.

A benchmark classification method hat is already well studied and has been described in the literature for several decades now is the Decision Tree method. This method is discussed in greater detail in the next section. In this method, a training dataset is used to build a decision tree in which the nodes of the tree are the different input features used, and the branching at each node is based on simple conditions (<, =, >) based on the feature associated with the node. Finally, the leaves of the tree are the different class labels.

To enhance the classification performance on the dataset, other classification methods are used with the performance of the Decision Tree as the benchmark rate. One of such advanced classifiers is the Multi-Layered Perceptrons (MLP) that is described in greater detail in Section 4.3. The MLP is also an extensively studied method, dating back to the early 80's (Werbos, 1974; Le Cun 1985, Rumelhart, Hinton and Williams, 1986; Haykin, 2009), but has since been refined, studied, and applied to a wide range of fields of application. The MLP is a supervised Neural Network model that is also trained using a labeled dataset like for Decision Trees. Once trained, they can be used to predict the classes of unknown instances or samples.

4.2. Decision Trees

Decision Trees, having been previously studied extensively, provide the baseline performance for classification and prediction, while Multi-Layered Perceptrons (MLPs) are used to enhance the prediction performance. A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a classification or decision. The Decision Tree method is widely popular for baseline classification for the following reasons:

- A Decision Tree is simple and easy to interpret. It can provide rules which can be easily translated to a Boolean logic format such as nested If-Then-Else statements.

- Decision trees can describe the data according to the conditions about the feature values from the root to the leaves leading to classes or categories. The important features are easily identified since they appear as a root node or nodes nearer to the root in the tree. Such information can be used for feature selection and for improving classification performance.

- Training a Decision tree, which is essentially just a one-time building of the tree, is typically much faster than most computational intelligence classifiers, such as those based on neural networks, which would typically require numerous iterations on the entire training set. Given the voluminous nature of EEG data, decision trees are thus the best choice for preliminary classification and data analysis, and to provide baseline classification performance rates.

By far the most important decision tree algorithm used in data mining is the ID3 (Quinlan, 1986) and its successor, the C4.5 algorithm (Quinlan, 1993). The tree is built top-down starting from the root node and recursively partitioning the tree or splitting on the values of attributes. The nodes of the tree are features (attributes) based on which classification is performed. At each node of the tree, the feature that is most useful for classifying is selected. In this case, "usefulness" is in terms of what is defined as the "information gain" measure. The feature with the highest normalized information gain is chosen at a specific node of the tree, and an IF-THEN condition involving the feature is set at that node. The "information gain" is essentially related to the size of the subset of data that would be correctly classified when a certain condition (<, =, >) is performed on a given feature.

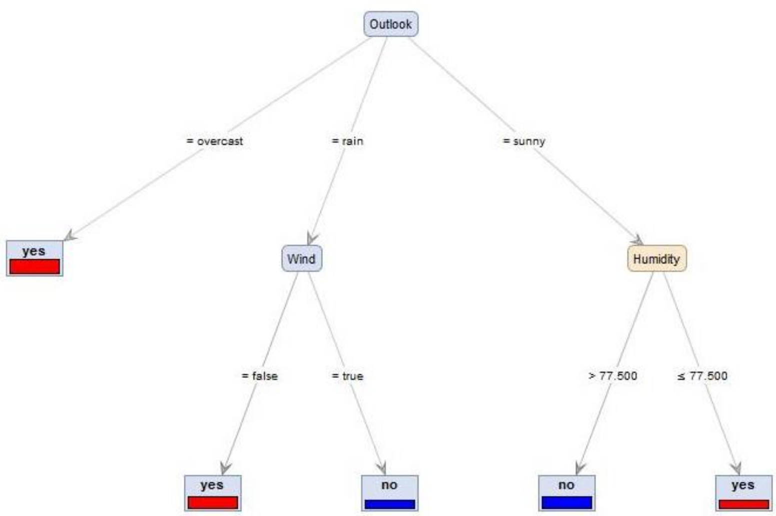


Figure 4.1 Decision Tree for Golf dataset

To illustrate how a decision tree works, a decision tree on whether to play golf or not, is presented in Figure 4.1. The figure is generated using Rapid Miner, based on the software's sample Golf dataset. The class attribute is 'Play' with values either 'yes' or 'no'. These are going to be predicted based on 4 features or attributes, namely Outlook ('sunny', 'overcast', 'rain'), Wind ('yes', 'no'), Humidity, and Rain ('true', 'false') The root node is 'Outlook'. Based on the value of 'Outlook', we would be able to traverse the tree and determine if 'Play' is 'yes' or 'no'. If the 'Outlook' attribute value is 'overcast', the 'Play' attribute will have the value 'yes'. If the 'Outlook' attribute value is 'rain' and the 'Wind' attribute has value 'false', then the 'Play' attribute will have the value 'yes'.

The leaf nodes are the class labels or categories, which provide the value of prediction. The non-leaf nodes, on the other hand, are the features or input attributes. For nominal attributes, the number of outgoing edges from a particular node is the number of possible nominal values for that particular feature or attribute. For example, the 'Outlook' attribute has three outgoing edges, namely: 'overcast', 'rain' and 'sunny'. For numerical attributes, such as 'Humidity', outgoing edges are labelled with disjoint ranges of values.

4.3. Multi-Layer Perceptron (MLP)

A Multi-Layer Perceptron (MLP) is a feedforward artificial neural network that can be trained in a supervised manner to map a given input set to output set. The input-output mapping need not be known to the user, but numerous examples of the correct input-output pairs must be available so that the network can learn to make the proper association. It has been proven in (Hornik, Stinchcombe & White, 1989) that MLPs are universal approximators – that is, some MLP can be trained to learn any mathematical function, possibly non-linear, between input and output sets, except that it is not known how many hidden units would be needed for every input-output mapping to be learned.

The architecture of a MLP consists of several layers of nodes (or neurons): one input layer, one or more hidden layers, and one output layer. Each node of one layer, whether an input or a hidden layer, is fully connected to the nodes in the next layer. The connections are assigned synaptic weights, or simply weights, which determine how much of the input from the previous layer actually reaches the other end of the connection in the next layer. These sets of connections weights are the free parameters of the system that are the ones that are tuned or updated during training/learning.

Each node of the MLP, except the nodes in the input layer, operates like a McCulloch and Pitts (1943) artificial neuron as shown in Figure 4.1. The weighted sum of input signals is referred to as the activation value, and the output of a node, or its response, is a function of the activation value, subject to some threshold value that may be unique for each node. The output (transform) function used in most MLPs is either of just two types : the logistic function and the hyperbolic tangent function, both of which are normalizable and differentiable (continuous functions).

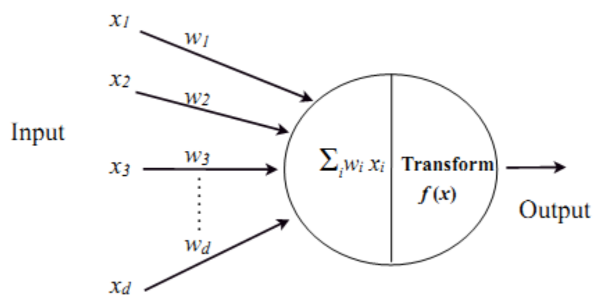


Figure 4.2 McCulloch and Pitts Model of a neural network node

The logistic function (as shown in Figure 4.3a) is defined as $f(x) = (1 + \exp(-x))^{-1}$. The hyperbolic tangent (as shown in Figure 4.3b) function is defined as

$$f(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \tag{4.1}$$

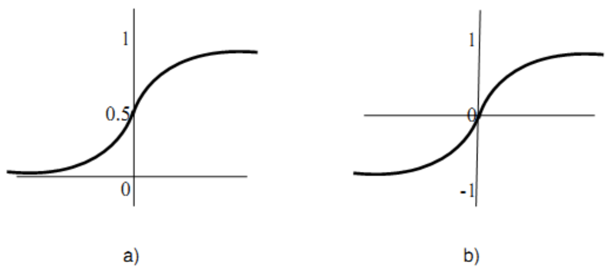


Figure 4.3 The output function, in terms of the activation value x, is either of the two mostly used functions, a) Logistic function and b) Hyperbolic tangent

Figure 4.4 presents the architecture of a MLP composed of one input layer, one hidden layer, and one output layer. Input signals from the input layer are fed to the first hidden layer, then the next hidden layer, and so on until the activation reaches the output layer. This process is called feed-forward activation.

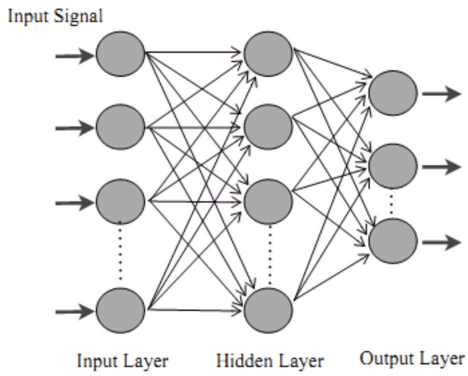


Figure 4.4 Architecture of a multilayer perceptron with one hidden layer

One learning method typically used for the training of a MLP is the backpropagation algorithm, or more formally the retro-propagation of error method. The algorithm had been independently discovered and popularized by Yann Le Cun (1985) and the team of Rumelhart, Hinton, and Williams (1986). However, the algorithm was published for the first time, more than a decade earlier, by Paul Werbos (1974).

There are two phases involved in the learning process by retro-propagation of error:

1. *Feed-Forward Activation phase:* the synaptic weights of the network are fixed and each input signal is propagated through the network, passing through each layer at a time from the input layer, to each of the hidden layers, up to the output layer.

2. *Backward Error Propagation phase:* the error signal is computed by comparing the output of the network with the desired output that accompanies the input element currently being presented to the network. The computed error signal is (back)propagated through the network, layer by layer from the output layer, to each of the hidden layers, until it reaches the input layer. Synaptic weights of the network are adjusted as a function of their contribution to the error computed at each node of the network.

For training the network, two methods may be applied: batch learning and on-line learning. In the batch method of supervised learning, adjustment of the synaptic weights of the the MLP is performed after the presentation of all the N examples in the training sample T and the entire training constitutes one epoch. In the on-line method, adjustments to the synaptic weights of the MLP are performed on an example-by-example basis (Haykin, 2009). For the on-line learning, the algorithm iterates through the training samples, each sample n being represented as a vector of input signals x(n) and a desired output d(n) representing the correct or expected output for input element n.

$$T = \{(x(n), d(n))\}^{N}_{n=1} \tag{4.2}$$

Training or learning is supervised and is implemented as follows (Haykin, 2009):

1. *Initialization.* The synaptic weights and thresholds are ideally defined from a uniform distribution whose mean is zero and whose variance is chosen so that the standard deviation of the induced local fields of the neurons lie at the transition between the linear and standard parts of the sigmoid activation. In many implementations of the MLP, it would suffice to set the weights to small random initial values.

2. *Presentation of Training Examples.* A training example is presented to the network where the training samples are selected at random.

3. *Forward Computation.* Given that the training example is described as a pair (x(n), d(n)) where x(n) is the input vector sent to the input layer and d(n) represents the desired output presented to the nodes of the output layer. The activation values and and output signals of the network are computed by proceeding forward through the network, layer by layer as described earlier. The activation value $v_j^{(l)}(n)$ for neuron j in layer l is defined as:

$$v_j^{(l)}(n) = \sum_i w_{ji}^{(l)}(n)\, y_j^{(l-1)}(n) \qquad (4.3)$$

   where $y_j^{(l-1)}(n)$ is the output signal of neuron i in the previous layer l-1 at iteration n, and $w_{ji}^{(l)}(n)$ is the synaptic weight of neuron j in layer l that is fed from neuron i in the previous layer l-1.

   When i = 0, $y_j^{(0)}(n) = +1$. Instead of thresholds per neuron, we use a bias. The bias applied to the neuron j in layer l is denoted by $w_{j0}^{(l)}(n) = b_j^{(l)}(n)$. This way, since thresholds are treated like connection weights, the training/learning processes will only compute for the correct values for connection weights - and the threshold can be deduced from the special connection weight, referred to in the literature as a bias.

   Using a sigmoid function, the output signal of neuron j in layer l is

$$y_j^{(l)} = \varphi_j\,(\,v_j\,(n)\,) \qquad (4.4)$$

   If neuron j is the first hidden layer (i.e. l = 1),

$$y_j^{(L)}(n) = x_j\,(n) \qquad (4.5)$$

   where $x_j\,(n)$ is the jth element of the input vector x(n).

   If neuron j is the output layer (i.e. l = L, where L represents as the depth of the network), then

$$y_j^{(L)} = o_j\,(n) \qquad (4.6)$$

Once the output of the nodes of the output layer have been computed, the errors can be computed. At the output layer, the error signal can be directly computed from the difference between the actual and the desired outputs

$$e_j(n) = d_j(n) - o_j(n) \tag{4.7}$$

where $d_j(n)$ is the jth element of the desired output vector d(n).

3. *Backward Computation.* As for the errors in the interior layers (hidden layers), the error cannot be computed directly and must be inferred from the errors of the nodes of the "preceding" layer. Note that direction is now backwards from the output layer, through the hidden layers, down to the input layer. The local gradients δs of the network are set as:

$$\delta_j^{(l)}(n) \begin{cases} e_j^{(L)}(n)\, \varphi_j'(v_j^{(L)}(n)) & \text{for neuron j in output layer L} \\ \varphi_j'(v_j^{(L)}(n) \sum_k \delta_k^{(l+1)}(n)\, w_{kj}^{(l+1)}(n) & \text{for neuron j in hidden layer l} \end{cases} \tag{4.8}$$

wher φ_j' represents differentiation (first derivative of $\varphi_{j)}$.

The synaptic weights of the network in layer l is adjusted by applying the generalized delta rule

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \propto [w_{ji}^{(l)}(n-1)] + \eta\, \delta_j^{(l)}(n)\, y_i^{(l-1)}(n) \tag{4.9}$$

4. *Iteration.* Forward and backward computations are performed iteratively while new training examples are presented to the network. These are applied until the chosen stopping criterion is met. The momentum and learning rate are usually adjusted and decreased as the number of training iteration increases.

4.4. Clustering and Data Visualization

Another route for the analysis of collected data, as opposed to classification and prediction, is to do unsupervised clustering and then data visualization. For example, instead of using the EEG dataset to build a model for predicting the academic emotion associated to each instance of a set of 126 EEG signals, it is also possible to group the EEG instances into a small number of "clusters" such that the instances found in each cluster are as close as possible to each other, and that the instances from the different clusters are as different as possible. Once clustered, we can try to visualize the clusters by determining the characteristics of the EEG data that constitute the individual clusters. For example, one cluster may have EEG data that have mostly high values among the signals collected from the left and right anterior frontal regions of the brain, or that another cluster may have EEG signals that are mostly 0 in all the right hemisphere region of the brain.

Further visualization can also be done according to the profile of the students from whom the EEG signals that belong to a given cluster have been collected. For example, a given cluster of

EEG signals may be predominantly coming from left-handed female learners. Note that in clustering and visualization, the academic emotions need not be the focus of the analysis when characterizing the different clustering.

One baseline clustering method is the *k*-means clustering method, which is described in greater detail in section 4.5. This is a family of clustering methods among the non-hierarchical clustering methods. There are numerous other hierarchical clustering methods, but these are not used in the study. Aside from *k*-means clustering, the Self-Organizing Map (SOM) by Teuvo Kohonen (2001; 2007) is used to enhance the visualization aspect of clustering. The SOM methodology is described at length in Section 4.6.

Azcarraga et al (2008) demonstrated how SOM labeling, which had always been supervised despite the fact that SOM training is unsupervised, can be done without the use of labeled datasets. After the unsupervised training of the SOM, the nodes are subjected to *k*-means clustering. These *k*-Means clusters are then the basis for delineating the regions in the trained map. Once the regions have been delineated, the clusters are then analyzed to know which are the most important features per cluster, and these features are then the basis for labeling the regions (clusters). Note that in this new method of labeling the trained SOM nodes, is there is no need for labeled dataset. A further refinement of this methodology that combines SOM and *k*-means clustering is presented in Chapter 6.

4.5. *k*-Means Clustering

k-Means clustering, being a well-established and extensively studied family of clustering methods, provides the baseline model for data clustering, while Self-Organizing Maps are employed to put some order among the clusters that are formed, and to visualize the inter-relationships among the data items.

k-means clustering (MacQueen, 1967) is one of the simplest and one of the most popular unsupervised learning techniques. In this approach, data points are grouped into *k* clusters, each of which has a centroid, or loosely referred to here as the "mean". Figure 4.5 shows *k*-means *(k*=5) clusters with centroids in blue circles. The algorithm proceeds as follows:

1. Select *k* points in the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Each data point is assigned or associated to the nearest centroid.
3. When all data points have been assigned to a centroid, the positions of the *k* centroids are re-computed.
4. Repeat steps 2 and 3 until the centroids no longer change their position in the data space.

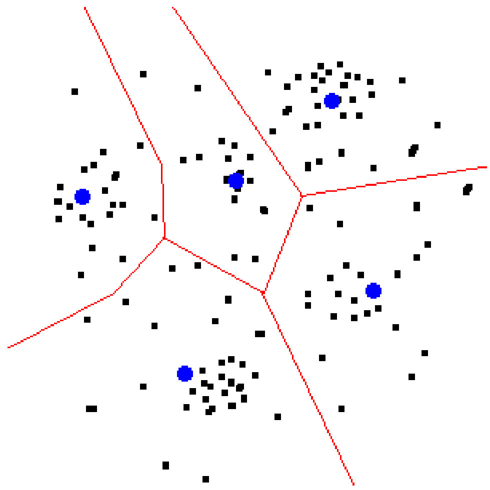


Figure 4.5 *k*-means clusters with k=5. The five blue circles are the centroids while the black circles are the data points. The red lines partition the clusters (from http://mnemstudio.org)

There are many variants of and numerous studies on the basic *k*-means algorithm described above. The choice of the most suitable value of k, for example, is already the subject of many studies. In this study, since k cannot be too large anyway, k values from 4 to 8 have been tried, and a value of k=6 has been used throughout the report because larger values of k has simply resulted in smaller clusters (with few samples) instead of any of the bigger clusters being split into smaller clusters.

The choice of the initial k centroids, also has been shown to affect the results, and various methods have been previously proposed in the literature. The most straightforward methods are 1) assign random values to the individual features of the k initial centroids, 2) use the first k instances/samples as the "seed" initial centroids; and 3) randomly choose k instances as the seed centroids.

The choice of the metric of similarity or distance for step 2 is also critical, and once again numerous studies on the topic have been done in the past. In all the simulations performed in this study, the Euclidean distance (largely the default metric to use) was used. Depending on the choice of the similarity or distance measure, and perhaps with the aim of speeding up the clustering process, various methods have been tried concerning the manner by which the centroids are re-computed as clustering progresses. At least two options exist, which are 1) to compute the new centroid every time a new instance is assigned to a given cluster, or 2) to do batch updating of the centroids after every pass through the entire dataset.

Finally, the criteria for determining when to stop – and related issues such as detection of cycles or loops where instances, for example, might just do a never-ending switch between one cluster in one pass, and the other cluster in the next pass, and back to the former cluster in the next pass and so on. In fact, it is also possible to connect the issue of finding the proper stopping criterion with the very first issue of finding the appropriate value of k. In this case, the value of k is systematically increased (or decreased) until a suitable value of k is arrived at.

4.6. Self-Organizing Maps (SOM)

Kohonen's Self-Organizing Maps (SOM) are neural network models commonly used for dimension reduction and data clustering (Kohonen & Honkela, 2007; Kohonen, 2001; Maimon et al, 2007). It can learn from complex, multidimensional data and can project these data to a topological map of fewer dimensions. The map is typically a 2D rectangular or hexagonal lattice of nodes. Visualization of clusters and their similarities become much easier and more insightful since the data are projected onto simple, plain 2D plots.

Unlike the MLP, the training for SOM networks is unsupervised since there are no known target outputs associated with each input pattern. Like in MLPs, the SOM training/learning process refers to the adjustment of the free parameters of the network, which are also the connection weights. However, the training mechanism and the learning rule used in SOMs are very different from those employed by MLPs.

A SOM network typically is composed of just an input layer and an output layer. Each node in the input layer is fully connected to the nodes in the (usually two-dimensional) output layer. Figure 4.6 presents a 4x4 SOM network. For such network, the output layer is consists of a 4x4 rectangular array of 16 neurons. The number of nodes in the input layer corresponds to the number of input variables whereas the number of output nodes depends on the specific problem and is defined by the user.

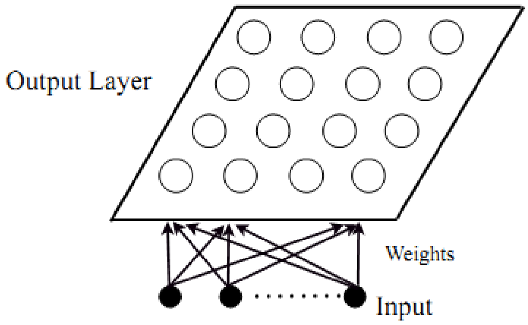


Figure 4.6 A 4x4 SOM network

Training of SOM networks involves the iterative presentation of input pattern x, which is randomly selected from the training set. For every training sample presented to the network, each neuron i in the output layer calculates how similar the input is to its weights w_i. Similarity is computed using various measures of similarity or distance, with a variety of methods currently in use. The most popular is the cosine of the angle between two vectors (a similarity measure) and the Euclidean distance (a distance measure). The unit in the output layer that is most similar to the input training sample is usually referred to as the winning node or the best-matching unit (bmu). The weights of the winning node (represented as a codebook weight vector) as well as the weights of neighbouring nodes in the output layer are adjusted.

The details of the training of a SOM network are as follows (Maimon et al, 2007):

1. Initialize the weights to small random values and the neighborhood size large enough to cover half the nodes.

2. Select an input pattern x randomly from the training set and present it to the network.

3. Find the best matching or "winning" node k whose weight vector w_k is closest to the current input vector x using the vector distance

$$\|x - w_k\| = \min_i \|x - w_i\| \tag{4.10}$$

where ‖·‖ represents the Euclidean distance.

4. Update the weights of nodes in the neighborhood of k using the Kohonen learning rule:

$$w_i^{new} = w_i^{old} + \propto h_{ik}(x - w_i) \quad \text{if i is in } N_k \tag{4.11}$$
$$w_i^{new} = w_i^{old} \quad \text{if i is not in } N_k \text{ (10)} \tag{4.12}$$

where ∝ is the learning rate between 0 and 1 and h_{ik} is a neighborhood kernel centered on the winning node and can take Gaussian form as

$$h_{ik} = exp\left[\!\left[\frac{|r_i - r_k|^2}{2\sigma^2}\right]\!\right] \tag{4.13}$$

where r_i and r_k are positions of neurons i and k on the SOM grid and σ is the neighborhood radius.

5. Decrease the learning rate slightly.

Repeat steps 1-5 with a number of cycles and then decrease the size of the neighborhood. Repeat until weights are stabilized.

After training the network, the topological relations of the samples in the input environment is reflected by the spatial distance between the nodes in the map to which the samples are closest to. Once adequately labeled, the map can serve as a visualization tool for examining the data set.

Manalili (2011) has worked on a 3-D Map that was used to visualize music data. In his work, the SOM had a structure so that specific genres of music were assigned to specific regions of the 3D SOM. In this case, the large SOM has some structure that can then be exploited in order to use the SOM as a kind of classification system.

This idea of giving some pre-defined structure to the SOM is further explored in this study of EEG signals. Each emotion (confused, frustrated, bored, and interested) is assigned to one 6x6 sub-SOM in each of the corners of a 16x16 SOM. The training method has been slightly modified to consider the structure that is being imposed during training. This "structured SOM' and the results are discussed in Chapter 6.

Chapter 5: Predicting Academic Emotion based on EEG Signals

5.1 Baseline Classification Performance based on Decision Trees

Any attempt at classifying EEG data and at predicting the academic emotion of students based on these EEG data should begin with establishing the baseline classification and prediction performance. This is done by using one of the most established classifiers in Machine Learning, which is the C4.5 classifier. The C4.5 classifier builds a decision tree with one feature as a node of the tree. The leaves of the tree are the different classes – in effect arriving at a decision or classification as to the class of a current instance depending on the label of the leaf that is arrived at following the values of the different instance features, and according to the <, +, > relations imposed at each node of the decision tree.

In evaluating the prediction performance of the classifiers for each of the 4 academic emotions, two performance measures were used, namely Accuracy and Area Under the Curve (AUC). The AUC relates the hit rate to the false alarm rate. These measures are all based on the contingency table (confusion matrix) and are computed based on the True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) (Fawcett, 2006). Given the items depicted in the confusion matrix of Table 5.1, eq. 5.1 and 5.2 compute the True Positive Rate and False Positive Rate, respectively. These in turn are used to compute the Area Under the Curve (AUC), given in eq. 5.3, and eq. 5.4 computes the Accuracy.

$$True\ Positive\ Rate\ (TPR) = \frac{\sum TP}{\sum TP + \sum FN} \tag{5.1}$$

$$False\ Positive\ Rate\ (FPR) = \frac{\sum FP}{\sum FP + \sum TN} \tag{5.2}$$

$$AUC = \frac{(1-FPR)+TPR}{2} \tag{5.3}$$

$$Accuracy = \frac{\Sigma TP + \Sigma TN}{\Sigma TP + \Sigma FP + \Sigma TN + \Sigma FN} \tag{5.4}$$

Table 5.1. Confusion Matrix for the computation of the performance measures

	True N	True P
Predicted N	True Negative (TN)	False Negative (FN)
Predicted P	False Positive (FP)	True Positive (TP)

Table 5.2 shows the classification performance on the EEG dataset for each of the academic emotions, namely *frustrated, bored, confused,* and *interested*, and for all the four learning sessions combined, namely *Aplusix, Aplusix Slide, ScatterPlot Slide,* and *Scooter*. Based on the results, it is

clear that the baseline accuracy for predicting the academic emotion of learners while they are using a computer-based learning module is low – hovering just above 50%, and sometimes registering accuracy rates that are even lower than 50% as shown in the *all combined* section of Table 5.2.

The results are consistent with the more recent reported results in the literature owing to various artifacts introduced during EEG signal capture, such as muscle twitches, moving of the head, distractions during the session, as well as the sometimes weak association between EEG signals academic emotion. Note also that learners do not always feel strong emotions, as either bored, confused, frustrated or interested. Their academic emotions may have been at borderline levels (around 50, in a scale of 0 to 100) when their EEG signals were captured and so predicting the emotion can be truly difficult for any classifier.

Table 5.2 shows the individual performance measures yielded by the Decision Tree for each of the four learning sessions and for each of the four academic emotions. Note that the performance rates vary a great deal from one learning session to another, given the same academic emotion. For example, the Accuracy of the *Interested* emotion can be as low as 0.16 for the Scooter scatterplot session and as high as 0.77 for the Aplusix Slide session. Similarly, the AUC for the *Bored* emotion can be as low as 0.43 for the Scooter session and as high as 0.69 for the Aplusix Slide session.

Note that, as explained in the previous chapter, all test results reported here are based on student-fold cross validation. In this modern approach to testing, all the instances of each individual student are isolated into a rolling test set, one train-test episode for each student, and the results on all the test sets are combined to determine the prediction accuracy. This avoids the common error in the experimental methodology involving multiple instances per subject, where the instances of a given student are found in both train and test sets. This would weaken the results of the experiments even if the train and test sets are drawn randomly using a much-studied 10-fold cross validation. When the number of instances is very large, and when the number of subjects is big, this may not be too much of a drawback. But when the number of instances and the number of subjects are few, the results would start to be questionable.

5.2 Selective Prediction based on Unusual EEG Signal Strength

One approach to improving prediction accuracy that was previously explored using a different dataset, was first reported in (Azcarraga, Suarez & Inventado, 2011) and subsequently explored further and described in (Azcarraga & Suarez, 2013). In this approach, prediction of academic emotions are only performed when a certain pre-determined number of EEG feature values deviate significantly from the baseline mean. The hypothesis is that when recordings of an EEG sensor are picking up something "unusual", then the learner is experiencing some heightened emotion that may be easier to predict. As such, when prediction is limited to only those instances, where a given number of feature values deviate significantly from their baseline mean, the emotion prediction accuracy would tend to increase.

We thus refined the previous work following this line of thought, taking into more stringent account of what constitutes "number of feature values deviate significantly from their baseline mean". Also, pre-processing and data normalization has by now been more thoroughly and methodically executed. Finally, training and testing where performed under a more methodical, statistically-controlled environment.

Table 5.2. Individual and average performance of the Decision Tree (DT) over all four sessions (*Aplusix Slide, Aplusix, Scatterplot Slide, Scooter*) and according to each academic emotion

Session	Emotion	Accuracy	AUC
Aplusix Slide	Frustrated	0.45	0.45
	Confused	0.52	0.53
	Bored	0.72	0.69
	Interested	0.77	0.48
Aplusix	Frustrated	0.50	0.55
	Confused	0.59	0.52
	Bored	0.45	0.52
	Interested	0.36	0.53
Scatterplot Slide	Frustrated	0.62	0.63
	Confused	0.50	0.49
	Bored	0.55	0.51
	Interested	0.21	0.49
Scatterplot	Frustrated	0.50	0.49
	Confused	0.28	0.48
	Bored	0.29	0.43
	Interested	0.16	0.39
all combined	Frustrated	0.52	0.53
	Confused	0.47	0.51
	Bored	0.50	0.54
	Interested	0.38	0.47

The previous works described in (Azcarraga, Suarez & Inventado, 2011) and (Azcarraga & Suarez, 2013) were also based on an entirely different set of experimental conditions. The past works employed raw EEG values from 14 sensors which were not filtered and were not transformed to a frequency format, as they were done more systematically here. The human subjects were also a different set, and of a different age range. Finally, the academic emotions that were tested were different and the learning modules used for the experimental sessions were also quite different.

In the current work, the feature values are now standardized, ranging in values from -3 to +3, as described in the previous chapter. The data have been fully normalized and so therefore, the feature values are now comparable from one learner (subject) to another. Also, the validation tests employed in the previous works used a 10-fold cross validation, i.e. the original dataset was randomized and divided into 10 folds prior to validation, and each segment composed of 10% of the dataset was used for testing. Although this is a standard approach to machine learning and computational intelligence experiments, this is a weaker approach to testing and so in the follow-up experiments described here, we followed a stricter train-and-test methodology. As explained in section 5.1, all the instances of each student are isolated into a rolling test set, one train-test episode for each subject, and the results on all the test sets are combined to determine the prediction accuracy.

Table 5.3 shows the number of students/learners (subjects) according to increasing percentages of feature values that deviate from the baseline mean by at least one-standard deviation. The number of learners are shown for each of the four sessions, namely *Aplusix, Aplusix Slide, Scatter Plot Slide* and *Scatter Plot*. For example, 22 students in at least one of their EEG recordings have at least 10% of the feature values (at least 13 of the 126 features) that have values of less than -1.0, or higher than +1.0. Note that since the feature values have been normalized and transformed into standardized scores (z-scores), the feature value already represents the deviation from the mean (normalized to 0.0) in terms of the number of standard deviations (normalized to 1.0) from the mean.

Obviously, at 0%, all the 49 students with all their instances are included in the Aplusix sessions, while all 52 are included in the Scatter Slide Session. As explained in the previous chapter, the (maximum) number is different per session because although all 56 human subjects participated in all the sessions, some of their recordings for an entire session have been dropped due to technical failures (like when the Emotiv set got disengaged).

As the % of special event features increases, the number of students involved for testing decreases, since there would be fewer and fewer instances that deviate significantly (one standard deviation) from the mean. Of course, not all students have EEG recordings that show at least one instance with a minimum of 30% of the features having significantly high or significantly low values, i.e. values less than -1.0 or greater than +1.0. Note that by 30%, there would only be from 4 to 6 students that can be used for testing.

Table 5.3. Number of students per session according to percentage of *special event features* whose values deviate from the mean by at least one standard deviation.

	Percentage (%) of Special Event Detector Features									
Session	**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**
Aplusix	49	15	7	5	5	5	2	2	1	1
Aplusix Slide	49	22	8	4	4	3	3	1	1	1
Scatter Slide	52	25	9	6	5	4	4	2	2	1
Scooter	51	21	7	4	4	4	3	2	1	1

Such features whose values deviate from the mean by at least one standard deviation are referred to here as “special event” features. In the previous work, these were referred to as “outliers”, since mathematically, they truly are the outliers of a normal curve. In this study, the term “special event feature” is instead used to refer to “outlier” features since these are the features that may signal that something special may have triggered an unusually low or unusually high EEG signal. In some statistical experiments, “outliers” are sometimes removed (or treated with caution) as they tend to spoil the statistical process.

For the test sets, different datasets were formed for each percentage level of “special event” features, each of which is classified using the Decision Tree algorithm. Student-fold cross validation was employed to maintain mutually exclusive train and test sets. Moreover, for each validation process, the train set is balanced according to emotion tags (high or low). Each student becomes a test set only once and only those instances that have the required percentage of “special event features” are included in the test set. This applies to 10% to 90% detectors, whereas for the 0%, all instances of that particular student are included in the test. As for the train set, all instances of all students in the train set are included (other than the instances of the specific student in the test set).

Tables 5.4 to 5.7 present the classification performance of a Decision Tree classifier based on different "special event" percentages during the *Aplusix, Aplusix Slide, Scatterplot Slide* and *Scooter* sessions. The performance measures used in the evaluation are Accuracy and Area Under the Curve (AUC) as discussed in section 5.1. Note that for the Aplusix Slide session, as shown in Table 5.5, the Accuracy has improved from 0.45 to 0.84 for Frustrated, 0.52 to 0.94 for Confused, 0.72 to 0.81 for Bored and 0.77 to 0.86 for Interested. As for the Aplusix session, as shown in Table 5.6, not counting the perfect accuracies due to the very low number of students in the dataset, classification accuracy improved from 10% up to 90% compared with the baseline 0% in which all instances were included. In particular, the Accuracy for Confused increased from 0.59 to 0.92, 0.50 to 0.66 for Frustrated, and 0.45 to 0.98 for Bored.

A similar pattern of results were also observed for the Scatterplot Slide session. The Accuracy for Frustrated has increased from 0.62 to 0.97, 0.50 to 0.99 for Confused, 0.55 to 0.95 for Bored, and 0.21 to 0.33 for Interested. For the Scooter session, the accuracy for Frustrated increased from 0.50 to 0.99, 0.28 to 0.82 for Confused, 0.29 to 0.84 for Bored and 0.16 to 0.22 for Interested. Note that results at 60% to 90% are no longer reliable since there are only 1 to 4 students included in the set at those higher levels of percentages of special event features. In fact, even at levels of 30%, there are already less than 7 students in the dataset, and thus already making the results questionable in terms of their validity.

Table 5.4. Performance measures for increasing percentages of special event features, for each of the four emotions during the Aplusix-Slide session.

Aplusix Slide	**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**
	Frustrated									
Accuracy	*0.45*	0.63	0.57	0.84	0.67	0.36	0.19	1.00	1.00	1.00
AUC	*0.45*	0.62	0.53	0.79	0.69	0.63	0.56	0.00	0.00	0.00
	Confused									
Accuracy	*0.52*	0.29	0.38	0.39	0.55	0.88	0.94	0.00	0.00	0.00
AUC	*0.53*	0.25	0.45	0.3	0.38	0.57	0.57	0.00	0.00	0.00
	Bored									
Accuracy	*0.72*	0.81	0.68	0.61	0.67	0.42	0.55	0.00	0.00	0.00
AUC	*0.69*	0.79	0.50	0.56	0.68	0.40	0.46	0.00	0.00	0.00
	Interested									
Accuracy	*0.77*	0.86	0.70	0.55	0.67	1.00	1.00	1.00	1.00	1.00
AUC	*0.48*	0.49	0.50	0.50	0.50	0.00	0.00	0.00	0.00	0.00

Table 5.5. Performance measures for increasing percentages of special event features, for each of the four emotions during the Aplusix session.

Aplusix	**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**
	Frustrated									
Accuracy	*0.50*	0.53	0.55	0.57	0.56	0.48	0.64	0.66	1.00	1.00
AUC	*0.55*	0.65	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Confused									
Accuracy	*0.59*	0.60	0.75	0.76	0.75	0.87	0.92	0.92	0.88	0.88
AUC	*0.52*	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	Bored									
Accuracy	*0.45*	0.41	0.66	0.92	0.92	0.98	1.00	1.00	1.00	1.00
AUC	*0.52*	0.50	0.50	0.94	0.94	0.99	0.00	0.00	0.00	0.00
	Interested									
Accuracy	*0.36*	0.22	0.18	0.07	0.07	0.04	0.06	0.00	0.00	0.00
AUC	*0.53*	0.26	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.6. Performance measures for increasing percentages of special event features, for each of the four emotions during the Scatterplot Slide session.

Scatterplot Slide	**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**
	Frustrated									
Accuracy	*0.62*	0.57	0.55	0.66	0.66	0.69	0.87	0.92	0.97	1.00
AUC	*0.63*	0.61	0.62	0.66	0.66	0.36	0.47	0.50	0.50	0.00
	Confused									
Accuracy	*0.50*	0.61	0.66	0.77	0.77	1.00	0.99	1.00	1.00	1.00
AUC	*0.49*	0.53	0.39	0.69	0.69	0.50	0.50	0.00	0.00	0.00
	Bored									
Accuracy	*0.55*	0.60	0.72	0.71	0.71	0.73	0.95	1.00	1.00	1.00
AUC	*0.51*	0.50	0.50	0.50	0.50	0.50	0.50	0.00	0.00	0.00
	Interested									
Accuracy	*0.21*	0.19	0.20	0.24	0.24	0.33	0.13	0.08	0.03	0.00
AUC	*0.49*	0.46	0.40	0.50	0.50	0.50	0.50	0.50	0.50	0.00

Table 5.7. Performance measures for increasing percentages of special event features, for each of the four emotions during the Scooter session.

Scooter	***0%***	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**
	Frustrated									
Accuracy	*0.50*	0.57	0.74	0.89	1.00	0.99	0.93	1	1	0
AUC	*0.49*	0.53	0.50	0.77	0.50	0.99	0.88	0	0	0.5
	Confused									
Accuracy	*0.28*	0.32	0.10	0.14	0.02	0.82	0.06	0	0	0
AUC	*0.48*	0.0.50	0.50	0.15	0.48	0.00	0.03	0	0	0
	Bored									
Accuracy	*0.29*	0.29	0.41	0.68	0.84	0.76	0.69	0.52	1	1
AUC	*0.43*	0.49	0.50	0.72	0.83	0.51	0.50	0.5	0	0
	Interested									
Accuracy	*0.16*	0.13	0.16	0.22	0.19	0.01	0.05	0	0	0
AUC	*0.39*	0.29	0.09	0.29	0.11	0.34	0.50	0	0	0

Since from Table 5.4 we can see that in most of the sessions there is a sudden drop in the number of students from 0% to 10%, and from 10% to 20% of special event detectors, another set of experiments were conducted. In this new set of experiments, separate datasets were formed for 2%, 4%, 6%, 8%, 10%, 12%, 14%, 16%, 18%, and 20% for the Aplusix and Scooter sessions. Note that these two sessions involve some interactive tasks as students learn algebra (Aplusix) and as they learn how to plot using a specially-designed software module (Scooter). The other two sessions are passive sessions in which the students are merely asked to review the tutorial slides. Due to the very tedious nature of these new sets of experiments, these two latter sessions were no longer included. Table 5.8 provides the number of students for each of the new datasets.

For the Aplusix session, as shown in Table 5.9, we can confirm at a more detailed level the general pattern that as the percentage of the number of special event detectors increases, the prediction accuracy also increases. The accuracy for Frustrated increased from 0.49 to 0.60, Confused from 0.59 to 0.74, Bored from 0.45 to 0.65, and Interested from 0.35 to 0.4. For the Scooter in Table 5.10, the most notable result is the increase in Accuracy for the Frustrated emotion that increased from 0.50 to 0.74.

Table 5.8. Number of students per session according to percentage of *special event features* whose values deviate from the mean by at least one standard deviation.

	Percentage (%) of Special Event Detector Features										
Session	**0%**	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**	**14%**	**16%**	**18%**	**20%**
Aplusix	49	40	28	25	22	15	11	11	9	9	7
Scooter	47	40	33	32	25	20	18	15	10	10	9

Table 5.9. Performance measures for increasing percentages of special event features, for each of the four emotions during the Aplusix session.

Aplusix	***0%***	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**	**14%**	**16%**	**18%**	**20%**
	Frustrated										
Accuracy	*0.50*	0.50	0.49	0.56	0.6	0.53	0.58	0.55	0.52	0.60	0.55
AUC	*0.55*	0.59	0.61	0.64	0.67	0.65	0.64	0.59	0.59	0.66	0.33
	Confused										
Accuracy	*0.59*	0.63	0.67	0.63	0.55	0.60	0.62	0.71	0.71	0.70	0.75
AUC	*0.52*	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	Bored										
Accuracy	*0.45*	0.47	0.47	0.41	0.42	0.41	0.42	0.48	0.50	0.56	0.66
AUC	*0.52*	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	Interested										
Accuracy	*0.36*	0.41	0.35	0.27	0.27	0.22	0.15	0.18	0.20	0.25	0.18
AUC	*0.53*	0.55	0.49	0.43	0.31	0.26	0.21	0.23	0.24	0.27	0.11

Table 5.10. Performance measures for increasing percentages of special event features, for each of the four emotions during the Scooter session.

Scooter	***0%***	**2%**	**4%**	**6%**	**8%**	**10%**	**12%**	**14%**	**16%**	**18%**	**20%**
	Frustrated										
Accuracy	*0.50*	0.51	0.57	0.51	0.55	0.57	0.57	0.72	0.72	0.73	0.74
AUC	*0.49*	0.49	0.53	0.49	0.53	0.53	0.51	0.59	0.51	0.51	0.51
	Confused										
Accuracy	*0.28*	0.26	0.28	0.32	0.3	0.32	0.28	0.17	0.10	0.11	0.10
AUC	*0.48*	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	Bored										
Accuracy	*0.29*	0.3	0.33	0.26	0.27	0.29	0.29	0.33	0.37	0.40	0.41
AUC	*0.43*	0.48	0.50	0.49	0.49	0.49	0.49	0.48	0.50	0.50	0.50
	Interested										
Accuracy	*0.16*	0.11	0.13	0.12	0.11	0.13	0.11	0.14	0.19	0.17	0.16
AUC	*0.39*	0.34	0.34	0.26	0.28	0.29	0.12	0.08	0.10	0.09	0.09

5.3 Improving Classification Performance using Multi-Layered Perceptrons

Instead of a decision tree, as presented in the previous sections, Multi-Layered Perceptrons (MLP) were then employed to try to improve on the prediction accuracy. MLPs have been shown to be powerful classifiers, and have been proven mathematically to be "universal approximators". This implies that MLPs when provided with enough hidden units, can approximate any input-output relation that exist in the dataset, including complex, non-linear relations (Hornik,1989).

The same train-test datasets were subjected to MLP training and testing, following the same student-fold cross validation method of section 5.1. The performance results for the Aplusix Dataset are shown in Table 5.11, where the same figures of Table 5.2 (Aplusix) are included for ease of comparison. Note that as far as accuracy is concerned, the MLP performance rates are consistently higher than the Decision Tree's for each of the four academic emotions. In the case of the AUC rate, however, the performance rates are essentially the same, except for a drop in performance for the *Interested* emotion.

Table 5.11. Performance on the *Aplusix* session of the Multi-Layered Perceptron (MLP) classifier, as compared to a Decision Tree

	Decision Tree		**MLP**	
	Accuracy	***AUC***	***Accuracy***	***AUC***
Frustrated	*0.50*	*0.55*	*0.52*	*0.53*
Confused	*0.59*	*0.52*	*0.64*	*0.65*
Bored	*0.45*	*0.52*	*0.59*	*0.59*
Interested	*0.39*	*0.53*	*0.63*	*0.52*

There were studies that found some correlation between brain asymmetry and gender as well as between brain asymmetry and handedness (Davidson, Schwartz, et. al.1976; Good, Johnsrude et. al., 2001). In one of the reviews of Good et. al, handedness may be structurally correlated in the sensorimotor. Hand preference tends to be associated with increased connectivity in the dominant hemisphere. Moreover, it was also mentioned that males tend to have increased leftward asymmetry compared to females. Guided by these, isolating the datasets based on gender, and/or handedness was performed to improve classification performance. Such datasets, referred to here as "restricted datasets", contain only the EEG instances of students who are of a certain gender, e.g. all male, or a certain hand-dominance, e.g. all left-handed.

The future scenario for emotion prediction is to first inquire whether the student is male or female, and whether left-handed or right-handed. Depending on the category, the appropriately trained classifier would be used to predict the academic emotion.

Table 5.12 shows the distribution of male and female students among those included in the Aplusix session. Also shown is the distribution of left-handed and right-handed students. Moreover, Table 5.13 shows the actual number of instances according to gender and handedness. Note that some restricted datasets are no longer reliable due to the very few subjects included in them. Such is the case of the Left-Handed Male dataset with only 6 subjects and 3,935 instances, and even more so, the Left-Handed Female dataset with only two remaining subjects and 1,077 instances.

Table 5.12. Number of students who are male or female, and are left-handed or right handed, based on the Aplusix dataset.

	Right-Handed	**Left-Handed**	Total
Male	25	6	31
Female	16	2	18
Total	41	8	49

Table 5.13. Number of instances who are male or female, and are left-handed or right-handed, based on the Aplusix dataset

	Right-Handed	**Left-Handed**	Total
Male	17,897	3,935	21,832
Female	11,068	1,077	12,145
Total	28,965	5,012	33,977

In Table 5.14, the figures that appear in bold-face are those performance rates of restricted datasets that are better than the performance rate for the entire dataset (*all*) (Azcarraga, Marcos & Suarez, 2014). For the MLP classifiers, prediction performance consistently yields an accuracy of higher than 0.60 to as high as 0.75 for each of the four academic emotions when train and test datasets have been restricted to all right-handed males or all right-handed females. In fact, compared to the 0.60 mean accuracy for the entire dataset, the mean accuracy rates for the even more restricted datasets All-Right-Handed-Male and All-Right-Handed-Female are 0.66 and 0.67, respectively. Note the from Table 5.13, the two restricted datasets, namely Left-Handed Male and Left-Handed Female, do not have enough students/subjects to have any statistical basis for the performance rates achieved. The results in Table 5.14 therefore do not include the performance rates for these two restricted datasets.

Table 5.14. Performance of Decision Tree and Multi-Layered Perceptron (MLP) for the Aplusix session

		Decision Tree		MLP	
		Accuracy	*AUC*	*Accuracy*	*AUC*
Frustrated	*All*	*0.50*	*0.55*	*0.52*	*0.53*
	All Male	**0.69**	**0.70**	**0.61**	**0.62**
	All Female	0.41	0.42	**0.75**	**0.73**
	All Right-Handed	0.49	0.52	0.52	0.52
	All Left-Handed	0.30	0.27	0.30	0.26
	All Right-Handed Male	**0.60**	**0.60**	**0.63**	**0.63**
	All Right-Handed Female	0.26	0.27	**0.70**	**0.69**
Confused	*All*	*0.59*	*0.52*	*0.64*	*0.65*
	All Male	**0.60**	**0.58**	**0.66**	**0.66**
	All Female	**0.71**	**0.74**	**0.66**	0.61
	All Right-Handed	0.57	0.50	**0.67**	**0.68**
	All Left-Handed	0.37	0.38	0.30	0.29
	All Right-Handed Male	0.51	0.51	**0.70**	**0.70**
	All Right-Handed Female	**0.81**	**0.77**	0.61	0.61
Bored	*All*	*0.45*	*0.52*	*0.59*	*0.59*
	All Male	**0.59**	0.476	**0.69**	**0.68**
	All Female	**0.49**	0.48	**0.71**	**0.70**
	All Right-Handed	**0.61**	**0.61**	**0.63**	**0.63**
	All Left-Handed	0.38	0.35	0.27	0.23
	All Right-Handed Male	**0.53**	0.47	**0.67**	**0.68**
	All Right-Handed Female	0.43	0.43	**0.75**	**0.75**
Interested	*All*	*0.39*	*0.53*	*0.63*	*0.52*
	All Male	**0.46**	0.43	**0.74**	**0.56**
	All Female	**0.54**	0.36	0.53	0.43
	All Right-Handed	**0.40**	0.50	0.63	**0.53**
	All Left-Handed	**0.44**	0.47	0.62	0.40
	All Right-Handed Male	**0.50**	0.46	0.63	**0.59**
	All Right-Handed Female	**0.62**	0.40	0.61	0.50

Aside from the accuracy, the Area Under the Curve (AUC), as an additional measure of performance, is also shown in Table 5.14. The same general trend can be observed with the AUC for the MLP being higher than that of the Decision Tree. Furthermore, whereas the AUC of the MLP for the entire dataset is 0.57, the AUC for the All-Right-Handed-Male and All-Right-Handed-Female are 0.65 and 0.64 respectively. The high performance results may also be attributed to the similarity of EEG features among the participants who have the same demographic profile.

Future studies should probe deeper into the difference between the EEG patterns of male and female learners, as well as of EEG patterns of left-handed and right-handed learners. More detailed study, particularly, on the differences in brainwaves asymmetry among the different group of learners. All of such future work would require much more subjects, so that test results can be more reliable. The above preliminary results do seem to point to some significant difference based on gender, as well as based on handedness.

5.4 Feature Selection based on Principal Components Analysis (PCA)

All the tedious experiments that were performed using Multi-Layered Perceptrons, although they yielded interesting results, required enormous computer resources and took a very long time to complete. Given the hundreds of experiments that were run based on all the 126 features for the over 130,000 instances in the complete dataset, it is useful to be able to conduct the same experiments and achieve comparable results using a significantly lower number of features.

The natural approach towards dimensionality reduction is through Principal Components Analysis (PCA). PCA is widely used in data decomposition, filtering, classification and whitening. It is also used to reduce the dimensionality of the feature space and since datasets studied here are high-dimensional, voluminous, and are composed of features that all based on similar EEG signals, PCA becomes a very attractive choice.

From the original total of 126 features, the features were transformed into a fewer number of "principal components", and these components are the ones used for classification. Table 5.15 presents the classification performance of the Decision Tree when PCA components are used as features, instead of the original 126 EEG signals. The details of each component are attached in the Appendix.

Table 5.15. Classification Performance on Aplusix dataset using the Top 5/10 PCA Components as Features

	Accuracy	AUC	Remarks
Interested	73.60	0.50	Predicted All as Interested
Confused	56.47	0.50	Predicted All as not Confused
Frustrated	65.32	0.50	Predicted All as not Frustrated
Bored	85.48	0.50	Predicted All as not Bored

Results in Table 5.15 shows rates higher than 50%, but upon manual inspection, it was determined that the decision tree classifier was making a single prediction. With a balanced dataset, this gives an AUC rate of 0.50 for all the four emotions. These results suggest that PCA components cannot be used as alternative features in classifying EEG data even if PCA has the capability to capture the variances in the data. In other words, the variances in the data that are the basis for the extracted components are not related to the academic emotions.

Another set of experiments were explored by using the top three components using PCA, and by using these three components for selecting the subset of the features to be included for training and testing. If classification and prediction performance based on a (smaller) subset of the features is similar, then there would be a gain in efficiency especially when MLP would be used for classification.

Since a PCA component is just a linear transformation of all features, the top 10 features with the highest factor from the top 3 components are selected as basis for the new training and test datasets. Table 5.16 shows the selected features from each of the top 3 components while Table 5.17 presents the classification performance of using all the features and using only the selected features from PCA components. Note that only 29 instead of 30 features were considered since 1 of the features appeared in the top 10 of two components (BL_O1_PM) appears both in PC1 and PC3 in both sessions.

Table 5.16. Selected Features based on the top 10 features in the top 3 PCA components for each of the Aplusix and Scooter datasets. The feature codes in blue refer to those features that appear in both Aplusix and Scooter subsets of top features.

Aplusix			**Scooter**		
PC1	PC2	PC3	PC1	PC2	PC3
AL_AF4_PM	BH_F3_PM	BH_AF4_MSP	AL_AF4_PM	BH_F3_PM	BL_AF3_PM
AL_F3_PM	BH_F4_MSP	BH_O1_MSP	AL_P8_MSP	BH_F4_MSP	BL_AF4_MSP
AL_P8_MSP	BH_T8_PM	BL_AF3_PM	BH_AF3_PM	BH_T8_PM	BL_FC5_MSP
Area_4	GA_AF3_MSP	BL_AF4_MSP	BH_F7_MSP	GA_AF3_MSP	BL_FC5_PM
BL_AF3_MSP	GA_F3_MSP	BL_O1_MSP	BH_P7_MSP	GA_F3_MSP	BL_O1_MSP
BL_F7_MSP	GA_F3_PM	*BL_O1_PM	BL_AF3_MSP	GA_F3_PM	*BL_O1_PM
BL_F7_PM	GA_F4_MSP	BL_O2_PM	BL_F7_PM	GA_F4_MSP	BL_O2_PM
*BL_O1_PM	GA_F7_MSP	BL_P7_MSP	*BL_O1_PM	GA_F7_MSP	BL_P7_MSP
BL_P7_MSP	GA_T8_PM	GA_AF4_MSP	BL_P7_MSP	GA_T8_PM	BL_T7_MSP
GA_AF3_MSP	Wt	GA_O1_MSP	GA_AF3_MSP	Wt	BL_T7_PM

Figure 5.1 displays the different brain regions with the corresponding EEG channels of the Emotiv Epoc EEG sensor set. For the Aplusix PCA Components, the first component (PC1) tends to have higher factors for Alpha (features codes that start with AL) and Beta Low (feature codes that start with BL) with frequencies ranging 9-13 and 13-21, respectively. This component is sensitive to two near-low frequencies. PC2, on the other hand tends to be more sensitive to higher frequencies Beta High (BH) and Gamma (GA) in the frontal and temporal regions of the brain. PC3 is also sensitive to higher frequencies Beta High and Gamma in the Anterior Frontal, Occipital and Parietal regions of the brain, as well as to Low Beta (BL) frequency. Interestingly, the top 10 features of the second component (PC2) of Aplusix and Scooter are exactly the same.

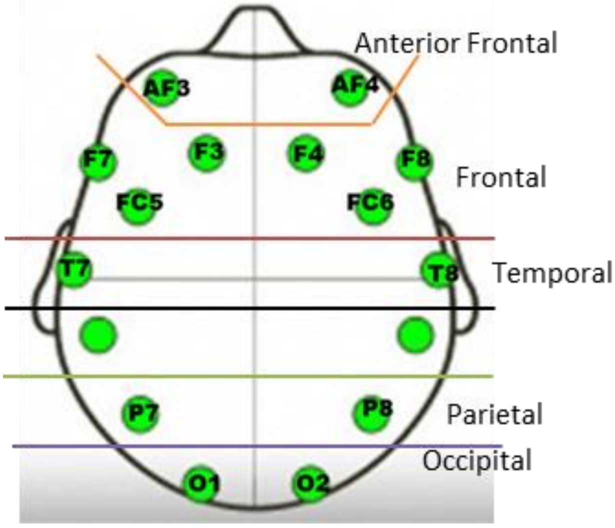


Figure 5.1 Brain Regions and EEG Channels

Table 5.17. Classification Performance of Aplusix and Scooter with Complete Features and PCA-based Selected Features

	Aplusix				**Scooter**			
	All Features (126)		PCA Based (29)		All Features (126)		PCA Based (29)	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Frustrated	0.50	0.55	0.52	0.57	0.46	0.50	0.46	0.50
Confused	0.59	0.52	0.59	0.52	0.27	0.50	0.27	0.50
Bored	0.45	0.52	0.45	0.52	0.65	0.48	0.68	0.50
Interested	0.39	0.53	0.16	0.50	0.86	0.50	0.86	0.50

Table 5.17 shows that the performance of PCA-based features (23% of the complete feature set) is almost the same as when all features are used. Same pattern was observed in two different sessions. This shows that 126 features can be reduced to smaller number of features (of around 23% of the original number of features) and still produce the more or less the same results. The implication is that the subset of features can be used in the place of the complete set, especially for computationally heavy classifiers such as the Multi-Layered Perceptrons.

5.5 Feature Selection Based on *k*-Means Clustering

A novel feature selection method was also developed, this time based on k-means clustering. The approach is initially based on past work on self-organizing maps developed for marketing research studies (Azcarraga, Setiono, Hsieh & Pan, 2008), and later also used in Belgium for other business applications (Seret & Baesens, 2014). In the past work, the scheme was used to identify the primary and secondary features that would characterize each cluster. The same basic idea is used to do feature selection. Essentially, the primary and secondary features that characterize each cluster would be the natural features to be selected for feature selection – this time with dimensionality reduction as the aim.

In the adaptation of the original scheme for tagging primary and secondary features, a more statistically coherent basis for tagging features is used. Also, the past works were based on Self-Organizing Maps (SOM) followed by *k*-means clustering of SOM nodes. In the scheme presented here, no SOM is done. Once the dataset is clustered using *k*-means, the population mean of each feature for data points belonging to all the clusters ($mean_{pop}$) is computed, as well as the individual mean per feature of all data points belonging to a given cluster ($mean_{in}$). Given these mean values, the ***z score*** (also referred to in the literature as *standard score*) is then computed for each feature in each cluster according to eq. 5.5.

$$z\ score = \frac{x-\mu}{\sigma} = \frac{mean_{in} - mean_{pop}}{\sigma_{pop}} \tag{5.5}$$

The features to be selected from each cluster is based on their z score values. If the z score of a particular feature is more than +δ or less than - δ, such a feature is selected. A smaller subset of features is obtained when only features with positive z scores (with value of at least +δ) are included. Indeed, in some applications, the positive signals are what are being tracked.

Based on this feature selection scheme, with δ = 0.5, a subset of 58 features is obtained for the Aplusix dataset when only positive z scores are included, and only 30 features when only positive z scores are included, at δ = 0.6.

Table 5.18 shows the classification performance of a Decision Tree for the Aplusix dataset using the three subsets of features, and compares the performance rates to the baseline figures just as in Table 5.17. Compared to a reduced set of 58 features, the classification performance when all features are used is essentially the same for all emotions except for *Interested* where the datasets using a reduced feature set produced slightly higher accuracy and AUC rates. With further reduction of features, to a subset of only 30 features (approximately the same number as the number of selected features when using PCA discussed in section 5.4), the Accuracy and AUC rates improves slightly for Frustrated emotion while it degrades slightly with the Interested emotion.

Table 5.18. Comparison of performance of a Decision Tree on the Aplusix dataset between experiments where all 126 features are used and with experiments when feature-selection based on z scores is used

	All Features (126)		58 Features +z, δ = 0.5		30 Features +z, δ = 0.60	
Emotion	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Frustrated	0.50	0.55	0.50	0.55	0.52	0.57
Confused	0.59	0.52	0.59	0.52	0.59	0.52
Bored	0.45	0.52	0.45	0.52	0.45	0.52
Interested	0.39	0.53	0.41	0.56	0.39	0.50

As a final set of experiments, the datasets with the two subsets of features are used to train several Multi-Layered Perceptrons in order to compare their performance rates with the baseline results given earlier in Table 5.12. Table 5.19 shows the classification performance of the MLP for the Aplusix dataset using the two subsets of features, and compares the performance rates to the baseline figures just as in Table 5.17. Essentially the same results are revealed in Table 5.19 as for Table 5.18. The implication, just as for PCA as basis for feature selection, is that the subset of features can be used in the place of the complete set, especially for computationally heavy classifiers such as the Multi-Layered Perceptrons.

Table 5.19. Comparison of performance of Multi-Layered Perceptrons on the Aplusix dataset between experiments where all 126 features are used and with experiments when feature selection based on z scores is used

	126 Features (all)		58 Features + z, $\delta = 0.5$		30 Features + z, $\delta = 0.60$	
Emotion	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Frustrated	0.52	0.53	0.56	0.56	0.55	0.55
Confused	0.64	0.65	0.66	0.67	0.66	0.66
Bored	0.59	0.59	0.61	0.62	0.59	0.59
Interested	0.63	0.52	0.67	0.62	0.47	0.53

Chapter 6: Data Clustering and Visualization

6.1 Unsupervised Clustering Methods

Two *supervised* classifiers, namely the Decision Tree C4.5 classifier and the Multi-Layered Perceptrons, have been used to predict academic emotions based purely on the EEG signals. These classifiers are considered "supervised" because they are trained using instances that have accompanying labels, i.e. academic emotion. These labels are the same ones that would be predicted during testing, and the performance results are based on whether or not the predicted labels match the accompanying labels. Following the established protocol for 5-fold, 10-fold, or student-fold cross validation approaches to testing, no instance that is used during testing would have been shown to the classifier during training.

To complement the results drawn from supervised classifiers, the same datasets were also analyzed, and visualized using two *unsupervised* clustering methods, namely the well-established *k-means* algorithm and the equally well-studied Self-Organizing Maps. These methods are unsupervised because the clusters that are formed are based purely on the input features, i.e. the EEG signals, and the emotion labels are not used during clustering. Later, during data visualization and in the analysis of the clusters, the academic emotions used as labels or tags for the training of supervised classifiers, are also going to be used in order to gain insights about the characteristics of the clusters that were formed.

6.2 *k*-Means Clustering

Employing the standard k-mean algorithm using the RapidMiner tool kit, the Aplusix dataset was subjected to clustering using various values for **k**. After experimenting with smaller and larger values of **k**, the results presented in this chapter is for clusters formed at **k**=6. At larger values of **k**, more and more small clusters are formed and thus making the analysis just more tedious without enhancing the results. A smaller value of **k** would have been possible, too, such as **k**=4 or 5, but six clusters provided a good balance between enough variety in the clusters formed, without having clusters with too few data points in them.

Table 6.1 shows the number of data points that belong to each of the six clusters. Each data point is an instance of 126 EEG signal readings. Numerous data points belong to each individual student and it is obviously possible, and even quite likely, that the EEG signal instances of a given student would appear in more than one clusters.

Table 6.1. Number of instances (data points) in each cluster based on the Aplusix dataset

Cluster	Number of data points
1	2,052
2	2,624
3	2,031
4	5,039
5	19,579
6	2,652

Although the gender and the hand-dominance are not part of the dataset used for clustering, each instance has a student identification number, and from the student identification number, the gender and the hand dominance of the student can be deduced. These two profile information are used for the data visualization discussion later in this chapter. There were other profile information that were recorded for each student, such as the personality profile, e.g. introvert or extrovert, but these information have not been used in the current study.

Table 6.2. gives the demographic profile of each cluster, listing the number of students, with their gender and hand dominance, who have at least one instance in each of the given clusters. Table 6.3 gives the same demographics, according to gender and hand-dominance, but the table shows the number of data points, instead of the number of users. Note how the ratios can vary significantly depending on whether we look at the number of students per cluster, as in Table 6.2, or the number of data points, as in Table 6.3. For example, in cluster 2 of Table 6.2, there are 71.4% right-handed students that have at least one EEG instance belonging to the cluster. In Table 6.3, however, only 48.1% of the EEG instances that belong to cluster 2 are from students who are right-handed.

Table 6.2. Demographics of each cluster based on the number of students with at least one EEG instance that belongs to each of the six clusters.

Cluster	Number of students	Male	Female	% Male	Right-Handed	Left-Handed	% RH
1	4	3	1	75.0%	3	1	75.0%
2	7	6	1	85.7%	5	2	71.4%
3	32	24	8	75.0%	28	4	87.5%
4	18	10	8	55.6%	17	1	94.4%
5	39	23	16	59.0%	33	6	84.6%
6	37	25	12	67.6%	33	4	89.2%

Table 6.3. Demographics of each cluster, according to the number of data points or EEG instances that belong to each cluster. The count for individual instances EEG would depend on whether the instance belongs to the gender and hand dominance of the student from whom the EEG instance was collected.

Cluster	Number of data points	Male	Female	% Male	Right-Handed	Left-Handed	% RH
1	2,052	1,321	731	64.4%	1,582	470	77.1%
2	2,624	2,010	614	76.6%	1,261	1,363	48.1%
3	2,031	1,341	690	66.0%	2,024	7	99.7%
4	5,039	2,771	2,268	55.0%	5,016	23	99.5%
5	19,579	11,782	7,797	60.2%	16,439	3,140	84.0%
6	2,652	2,607	45	98.3%	2,643	9	99.7%
Total	**33,977**	**21,832**	**12,145**	**64.3%**	**28,965**	**5,012**	**85.2%**

6.3 Characterizing the Clusters using Decision Trees

It would be interesting to be able to characterize each cluster in terms of the kind of EEG signals that get associated to each of them. It is possible that high values of a certain specific EEG signal might be strongly associated to a specific cluster. For example, it may turn out that high values of the peak magnitude of the alpha band of the left temporal region, which is feature F9 (AL-T7-PM), would tend to be mostly associated to cluster 5. Of course, we do not expect very strict one-to-one relationships between high values or low values of specific features to specific clusters. Nevertheless, general relationships between specific EEG signals and clusters would be useful information that would describe what each cluster might represent.

In order to determine which EEG features can describe and discriminate which cluster, we feed the same Aplusix dataset to a C4.5 decision tree classifier, but this time, the associated label is the cluster number. Table 6.4 shows the confusion matrix for the decision tree classifier. Note that the precision and recall for Clusters 1, 3, and 5 are high. This implies that if we convert the rules extracted from the decision tree that relate to these clusters, we may be able to have simple if-then rules that would characterize these three clusters in terms solely of the EEG signals (i.e. 126 features).

Figure 6.1 shows a snapshot of portion of the decision tree produced by RapidMiner, using the C4.5 *decision* tree method. The decision tree is large, with numerous branches and leaves and so it is quite tedious to convert the tree into specific IF-THEN rules for each of cluster 1, 3 and 5. For illustration purposes, Figure 6.2 shows an IF-THEN rule for class 5, which shows the specific features that are used to classify an instance as a cluster 5 datapoint. In the IF-THEN rule are also the conditions (<, =, >) and the thresholds for each of the features that were extracted. Note that since RapidMiner starts with class 0 instead of class 1, cluster 5 in Figure 6.2 corresponds to cluster_4 in Figure 6.1.

Table 6.4. Confusion matrix in Classifying Clusters

	true cluster 1	*true cluster 2*	*true cluster 3*	*true cluster 4*	*true cluster 5*	*true cluster 6*	***precision***
Predicted cluster 1	1,653	0	0	0	162	5	***0.91***
Predicted cluster 2	0	1,621	0	719	7	486	***0.57***
Predicted cluster 3	1	0	1,749	119	0	936	***0.62***
Predicted cluster 4	193	46	2	2,371	2,010	640	***0.45***
Predicted cluster 5	200	685	0	974	17,397	10	***0.90***
Predicted cluster 6	5	272	280	856	3	575	***0.29***
recall	***0.81***	***0.62***	***0.86***	***0.47***	***0.89***	***0.22***	

Figure 6.1 Decision subtree for classifying cluster 5 (cluster_4 in the tree)

```
If    Area_5 > -1.967
      AL_AF3_PM > -1.294
      BH_P7_PM > -0.367
      GA_F3_PM > -1.993
      GA_F7_PM > 0.040
      BH_F7_PM > -0.974
      BL_P7_MSP > -0.250
      AL_AF3_MSP > -1.083
      AL_AF4_MSP > -1.138
      AL_AF4_MSP > -1.055
      BH_P7_PM > -0.325
      Area_5 > -0.147
      Area_5 > -0.059
      BH_T7_MSP > -1.060
      BH_P8_MSP > -0.843
      BH_AF3_MSP > -0.990
      BH_P8_MSP > -0.759
      AL_P8_PM > -0.447
then
      cluster_4
```

Figure 6.2 If-Then rule for classifying cluster 5 (cluster_4 in the rule)

6.4 Characterizing Clusters based on Test of Difference of Proportions

Another method for characterizing the clusters is by applying standard statistical tests of Difference of Proportions. The test of proportions is used to determine whether the proportions of male and female or of right-handed and left-handed students are significant for any or all of the individual clusters. For example, from Table 6.3 we see that for there are a total of 28,965 EEG instances belonging to right handed students compared to only 5,012 EEG instances belonging to left-handed students. This represents an overall percentage of right-handed students of 85.2%. Cluster 2, however, has 1,261 EEG instances belonging to right-handed students compared to 1,363 EEG instances belonging to left-handed instances, which represents a percentage of right-handed students of 48.1%. Obviously, the proportion of right-handed students in cluster seems much lower, but is the difference in proportion statistically significant? The Test of Difference of Proportions would ascertain this, with the possibility of specifying difference confidence levels.

The test of proportion for gender and hand dominance is conducted as follows. Let

$$p_1 = \text{sample proportion from population 1}$$
$$p_2 = \text{sample proportion from population 2}$$
$$n_1 \& n_2 \text{ are sample sizes of } p_1 \& p_2 \text{ respectively}$$

Then, we set the null hypothesis as $p_1 = p_2$. For example, the null hypothesis is that the male proportion in a certain cluster (p1) is the same as the male proportion in all the other clusters (p2). The alternative hypothesis is that the proportions are different.

The null hypothesis is rejected if the *p-value* of a proportion is less than a certain pre-set α, e.g. 0.01. The *p-value* is defined as the probability, under the assumption of the null hypothesis H, of obtaining a result that is equal to or more extreme than what was actually observed. In other words, it is the probability of observing a sample statistic at least as extreme as the test statistic. To compute for the *p-value,* the proportion (p), the standard error (SE) and *t statistic* must be determined and computed as follows:

$$p = \frac{(p_1 * n_1) + (p_2 * n_2)}{n_1 + n_2} \tag{6.1}$$

$$SE = \sqrt{p * (1 - p) * \left[\frac{1}{n_1} + \frac{1}{n_2}\right]} \tag{6.2}$$

$$t\ statistic = \frac{(p_1 * p_2)}{SE} \tag{6.3}$$

The t-statistic, with the help of a two-tailed t-distribution calculator (NB: these used to come in table forms but are now available online as a "calculator"), can then be used to derive the *p-value*, with degrees of freedom set at **n1 + n2 - 2**. If the *p-value* is less than 0.01 (α = 0.01), then there is 99% confidence in rejecting the null hypothesis. When the null hypothesis is rejected, then the alternative hypothesis, that indeed the proportions are different, may be accepted. For example, in Table 6.5, the *t statistic* for cluster 1 and 3 are 0.1180 and 1.7178, respectively, which give p-values of 0.91 and 0.09 that are both not less than 0.01 and therefore the null hypothesis for clusters 1 and 3 cannot be rejected. In the other cases, the absolute value of the t-statistics are sufficiently large to give very small p-values, thus leading us to reject the corresponding null hypotheses.

Based on the procedure outlined above, Table 6.5 shows that for all clusters, except clusters 1 and 3, the null hypothesis that the proportions are equal can be rejected at α = 0.01, or a confidence level of 99%. Table 6.6 shows that for the hand-dominance proportions for all the six clusters, the null hypothesis that the proportions are equal can be rejected at $\alpha = 0.01$ or a confidence level of 99%. On manual inspection of the proportions, indeed most of the proportions vary greatly between the proportion within the cluster (p1) and the proportion in all the other clusters (p2).

All these imply that, indeed, gender and hand-dominance are important characteristics that need to be taken into account – since EEG signals seem to be clustered somewhat according to gender and hand-dominance. This result corroborates the findings on improved classification performance reported in section 5.3 on restricted datasets.

Table 6.5. Test of difference of gender proportions shows that all clusters (marked with *), except clusters 1 and 3, have significant proportion differences, at $\alpha = 0.01$

Cluster	**Male**	**Female**	**p1**	**p2**	**n1**	**n2**	**n**	**p**	**SE**	**t - statistic**	**p value (two-tailed)**
1	1,321	731	0.64	0.64	2,052	31,925	33,977	0.64	0.011	0.1180	0.91
2 *	2,010	614	0.77	0.63	2,624	31,353	33,977	0.64	0.01	13.74	7.98E-43
3	1,341	690	0.66	0.64	2,031	31,946	33,977	0.64	0.011	1.7178	0.08
4 *	2,771	2,268	0.55	0.66	5,039	28,938	33,977	0.64	0.007	-14.87	7.54E-50
5 *	11,782	7,797	0.60	0.70	19,579	14,398	33,977	0.64	0.005	-18.29	2.16E-74
6 *	2,607	45	0.98	0.61	2,652	31,325	33,977	0.64	0.009	38.10	0
totals	21,832	12,145									

Table 6.6. Test of difference of hand dominance proportions shows that all clusters (marked with *) have significant proportion differences, at $\alpha = 0.01$

Cluster	Right	Left	p1	p2	n1	n2	n	P	SE	t- statistic	p value (two-tailed)
1 *	1,582	470	0.77	0.86	2,052	31,925	33,977	0.85	0.008	-10.74	6.94E-27
2 *	1,261	1,363	0.48	0.88	2,624	31,353	33,977	0.85	0.007	-55.93	0
3 *	2,024	7	1.00	0.84	2,031	31,946	33,977	0.85	0.008	18.88	4.11E-79
4 *	5,016	23	1.00	0.83	5,039	28,938	33,977	0.85	0.005	31.01	3.6E-208
5 *	16,439	3,140	0.84	0.87	19,579	14,398	33,977	0.85	0.003	-7.8	6.48E-15
6 *	2,643	9	1.00	0.84	2,652	31,325	33,977	0.85	0.007	21.8	1.3E-104
totals	28,965	5,012									

6.5 Visualization of Significant EEG Features on a Brain Map

Towards further visualization of the clusters, the features are counted depending on whether they correspond to sensors positioned on the Left or Right hemispheres of the brain, and for each of the frequency bands *alpha, beta low, beta high*, and *gamma*. Table 6.7 shows the summary of the number of feature *z-scores*, as described in the previous chapter for the alternative basis used in feature selection, exceed the threshold set at $\delta = 0.5$.

Table 6.7. High, positive Features on Left and Right Brain Hemispheres, based on z scores with $\delta = 0.5$.

	Alpha		**Beta Low**		**Beta High**		**Gamma**	
Cluster	*LH*	*RH*	*LH*	*RH*	*LH*	*RH*	*LH*	*RH*
1	0	0	3	0	4	4	12	9
2	9	9	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	2	0	0	0	3	0	2	0
6	0	0	0	0	0	0	0	0

Figures 6.3 to 6.5 show the details of Table 6.7 above by mapping different features, whose z-scores are above 0.5, to specific locations in the brain, on a per cluster basis and for each of the frequency bands. Notice that Cluster 1 reveals mostly high positive feature values in the beta high and gamma bands, both for Peak Magnitude and Mean Spectral Power. Cluster 2 has mostly high positive feature values in the alpha band. Cluster 5 shows high positive feature values in the left hemisphere. All the other three clusters do not have special features associated with them that have high and positive z scores. Note that clusters 1 and 5 are the same clusters in Table 6.4 where the prediction of cluster labels is highest given the feature values of the EEG instances. The precision and recall for Cluster 1 are 0.91 and 0.81, respectively while Cluster 5 precision and recall are 0.9 and 0.89, respectively.

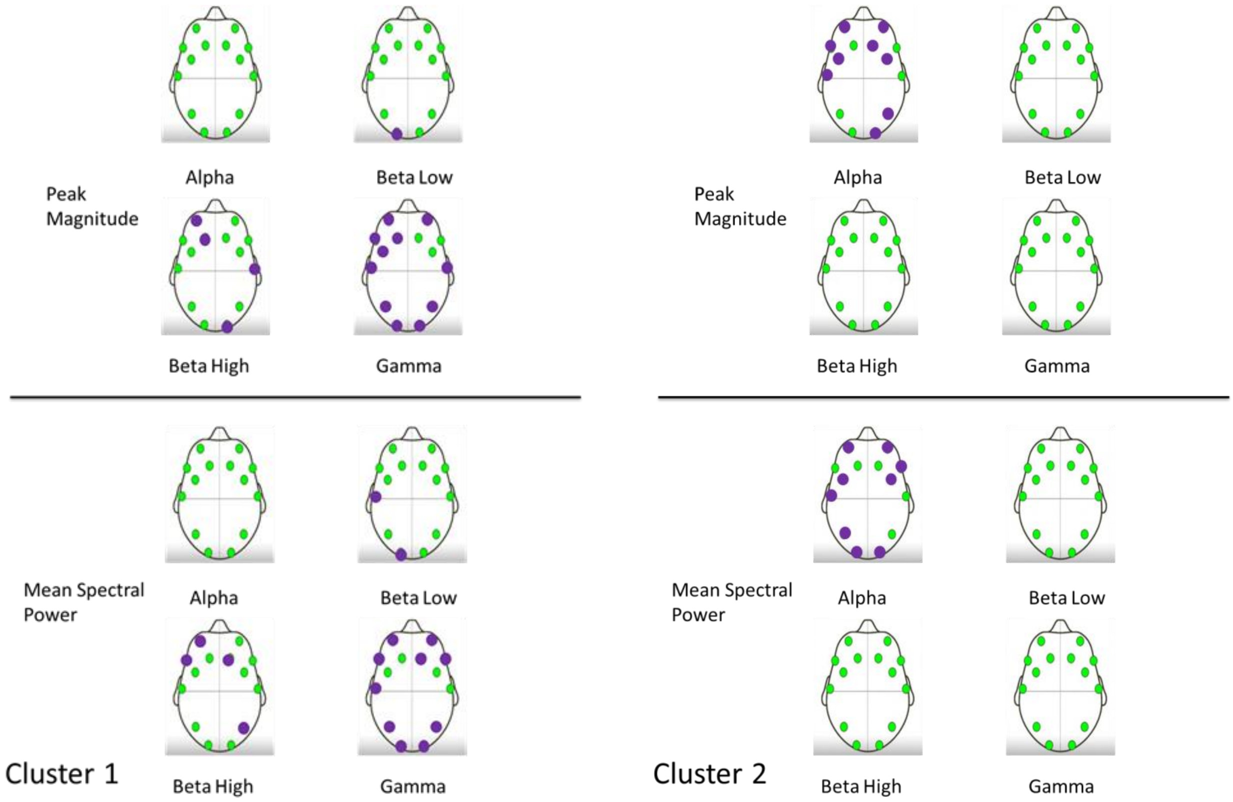


Figure 6.3 Significant features in Alpha, Beta and Gamma frequency bands in Clusters 1 and 2. Purple dots are significant features and green circles are non-significant features

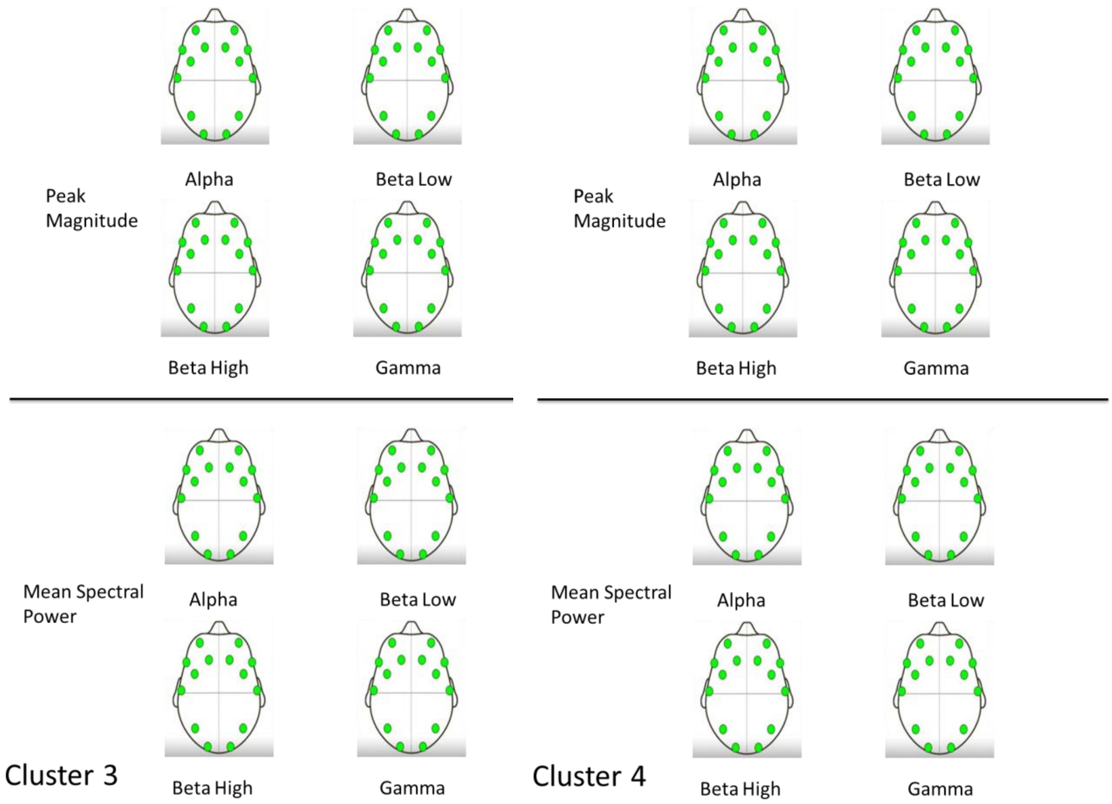


Figure 6.4 Significant features in Alpha, Beta and Gamma frequency bands in Clusters 3 and 4.

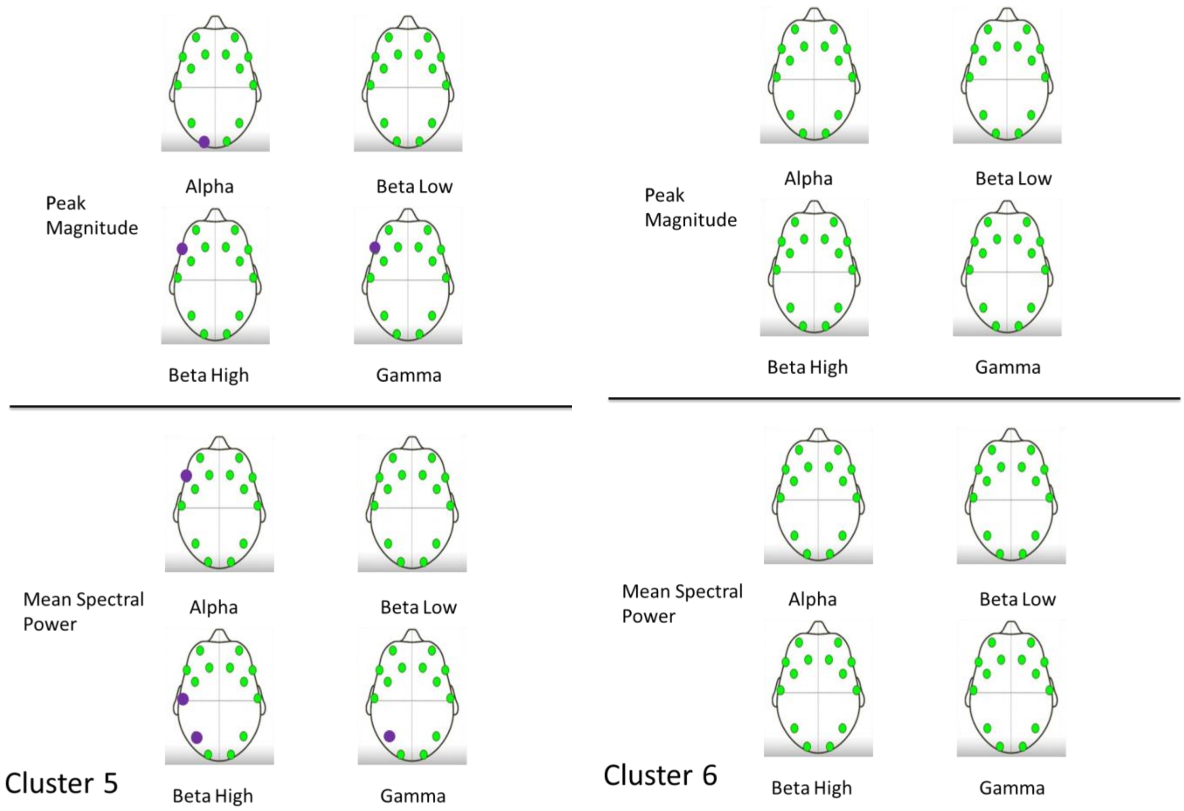


Figure 6.5 Significant features in Alpha, Beta and Gamma frequency bands in Clusters 5 & 6

6.6 Visualization using Self-Organizing Map (SOM)

Carrying the data visualization method one step further, we also depict the similarities between the clusters. Recall that the clusters that *k means* clustering would generate are labeled as cluster 1 to **k**, with no indication as to whether certain of these **k** clusters are similar to one another, relative to their distance to the other clusters. If the clusters can somehow be laid out in a 2D map where the geographic distance between two clusters as they are laid out in the map would indicate their relative similarity (or dissimilarity) vis a vis the other clusters, then an additional layer of visualization would be offered to the user.

There are various ways to organize clusters in a 2D map, and even to organize the various individual data points on a two-dimensional map. For example, the visual representation in 2D of the relative distances between the data points can be done by Multi-Dimensional Scaling (MDS). In this work, we describe the results of visualizing the clusters in a 2D map using Self-Organizing Maps.

As a clustering method suitable for data visualization, SOM has the ability to do three different tasks that are relevant to this study : 1) preliminary clustering the data points into N centroids (where N is the number of nodes in the map); 2) clustering of the N nodes (centroids) into **k** clusters, and 3) laying out of the N centroids as well as the **k** clusters in a regular 2D map where the proximity of the nodes in the map (and of the centroids and clusters that the nodes represent) would represent their relative distance in the feature space. The smaller the distance between two centroids in the feature space, the smaller also is the geographic distance between the two nodes in the map.

Figure 6.6 shows a standard 16 x 16 SOM where the 256 nodes, after training, are clustered using *k-means* clustering at **k**=6. Note that in general, the nodes that belong to a given cluster are contiguous to each other. This owes to the general characteristic of the SOM method where neighbor nodes tend to have similar weight values, and these weight values are the expected means or centroids of the EEG instances that are associated to the node. Note also that one cluster, cluster 2 delineated as yellow cells, tends to dominate the map. This reflects what had earlier been noticed in pure *k-means* clustering (without SOM) that many EEG instances are associated to one cluster (cluster 5 from Table 6.1).

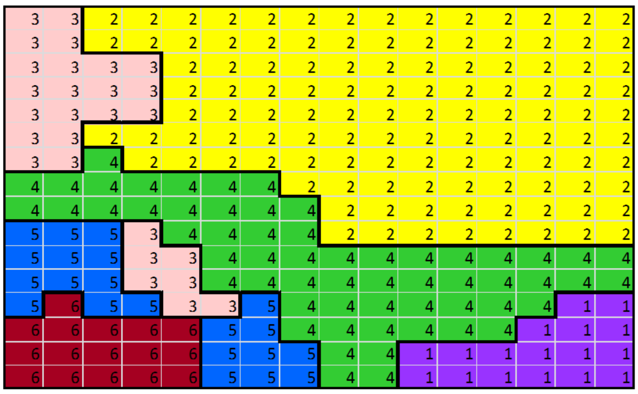


Figure 6.6 Standard SOM trained with Aplusix dataset showing six clusters, dominated by cluster 2, delineated by yellow cells.

After clustering of nodes, the nodes are labeled with the dominant emotion of the EEG instances that are nearest to each node, following the methodology described in Chapter 4. The distance is measured using the *Euclidean distance* formula for an *n*-dimensional space as shown below.

$$d(q,p) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \tag{6.4}$$

where q and p are Euclidean vectors in a Euclidian space. The formula can be applied in measuring the distance between a SOM node and an instance data point, since each is a vector with *n* dimensions and the training of the SOM essentially creates *prototype vectors* (also referred to as *exemplars*) among the weight vectors of the individual noades of the map.

The distance between each node and each instance data point is computed. The class of the instance that is nearest to the node is used as the label of the node. Note that each instance is labeled with the most *dominant* emotion - ***I*** for Interested, ***C*** for Confused, ***B*** for Bored and ***F*** for Frustrated. The emotion with the highest value is chosen as the dominant emotion, but only if it is above 50, otherwise there is no dominant emotion (label is ***L***). In cases when there are two or more dominant emotions that are equal and are above 50, the tag is set to ***M*** since there are multiple dominant emotions. Figure 6.7 shows the SOM map superimposed by (dominant) emotion classes, labeled as F for Frustrated, C for Confused, B for Bored, and I for Interested.

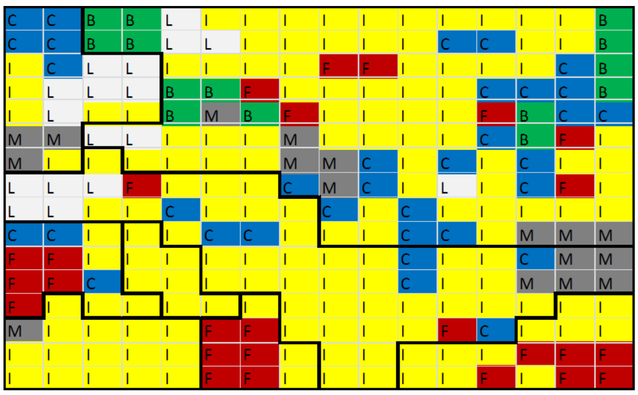


Figure 6.7 Standard SOM trained with Aplusix dataset showing six clusters, and each node labeled with the most dominant academic emotion. Frustrated (F) are red cells, Confused (C) are blue cells, Bored (B) are green cells, Interested (I) are yellow, Multiple dominant emotions (M) are grey cells and No dominant emotions (L) are white cells.

In Figure 6.7, the colors of the cells have nothing to do with the colors previously assigned to the six clusters of Figure 6.6. But the delineation among the 6 clusters are still shown with thick black lines. As expected, the emotions are scattered all over the map, and there is no single emotion that is assigned to only one cluster. Neither are the emotions located in just one region of the map. This reflects the earlier results yielded by Decision Trees and MLP that the accuracy of the prediction of the emotion based on the EEG data is not so high. And since the clustering is based solely on the EEG data, we cannot expect the emotions tags to be associated on a one-to-one basis with the clusters.

Most of the cells in Figure 6.7 are labeled as I (Interested - yellow) and this is the dominant tag in the map. Upon inspection of the data, indeed, the majority of the EEG instances have Interested as the dominant emotion.

The same test of difference in proportions, the way it was done for the *k-means* clusters in section 6.4, can also be done with the six clusters of SOM nodes. The test results are summarized in Table 6.8 for gender proportions and Table 6.9 for hand-dominance proportions.

Table 6.8. Test of Difference in Gender Proportions for the Aplusix dataset based on the six clusters of SOM nodes

Cluster	Male	Female	p1	p2	n1	n2	n	p	SE	t	p value (two-tailed)
1	1,318	730	0.64	0.64	2,048	31,929	33,977	0.64	0.0109	0.098	0.922
2 *	11,800	7,026	0.63	0.66	18,826	15,151	33,977	0.64	0.0052	-6.757	1.434E-11
3 *	2,494	563	0.82	0.63	3,057	30,920	33,977	0.64	0.0091	20.956	6.794E-97
4 *	3,395	3,046	0.53	0.67	6,441	27,536	33,977	0.64	0.0066	-21.477	1.192E-101
5 *	1,441	64	0.96	0.63	1,505	32,472	33,977	0.64	0.0126	26.076	1.952E-148
6	1,384	716	0.66	0.64	2,100	31,877	33,977	0.64	0.0108	1.628	0.103
totals	21,832	12,145									

Table 6.9. Test of difference in hand-dominance proportions for the Aplusix dataset based on the six clusters of SOM nodes

Cluster	Right	Left	p1	p2	n1	n2	n	p	SE	t	p value (two-tailed)
1 *	1,579	469	0.77	0.86	2,048	31,929	33,977	0.85	0.0081	-10.7281	8.29E-27
2 *	15,663	3,163	0.83	0.88	18,826	15,151	33,977	0.85	0.0039	-11.8786	1.77E-32
3 *	1,693	1,364	0.55	0.88	3,057	30,920	33,977	0.85	0.0067	-48.8164	0
4 *	6,441	0	1.00	0.82	6,441	27,536	33,977	0.85	0.0049	37.0841	4E-295
5 *	1,505	0	1.00	0.85	1,505	32,472	33,977	0.85	0.0094	16.5073	5.62E-61
6 *	2,084	16	0.99	0.84	2,100	31,877	33,977	0.85	0.0080	18.6638	2.38E-77
totals	28,965	5,012									

Aside from the Tables 6.8 and 6.9, which produce basically the same results as the tests of difference of proportions done on the six clusters derived directly from *k-means*, Figures 6.8 and 6.9 present the trained SOM, this time labeled by the students that each node is nearest to. These labels are generated by finding the EEG instance that is closest to a given node, and once identified, the student (student identification number) that is associated to the specific EEG instance is used as the label for the node. Once labeled, we use the gender and the hand dominance of each student and use these to color the map. In Figure 6.8, the green cells are the female students. The orange cells in Figure 6.9 correspond to left-handed students. The delineation of the six clusters are still shown in these maps.

Notice that the distribution of the colored cells in Figures 6.8 and 6.9 also reflects the relative proportions found in the data themselves. For example, there are 36.7% of the cells in Figure 6.8

that are green(female), and the actual proportion in the dataset is 39.5%. There are 16.3% left handed students in the dataset, and there are about 14.1% organ cells (left handed) in Figure 6.9.

The trained SOM is able to capture the same clusters found by *k-means*, except that in the SOM, each cluster is further subdivided into several prototype vectors that act as higher-resolution centroids. Instead of one centroid for an entire cluster in *k-means*, SOM can assign several centroids (one centroid per node) and the collection of these higher-resolution centroids represent the boundaries of each cluster. Subsequently, by laying out these clusters and the cells belonging to each cluster in a flat 2D map, it is possible to see the relationships between the clusters. Cluster 3, for example, has several higher resolution centroids (SOM nodes) near the regions occupied by clusters 4, 5, and 6. This shows that these four clusters have much more in common, than with cluster 1, for example.

16	16	10	10	15	15	42	42	40	40	5	5	5	9	9	47
16	16	10	10	15	15	42	42	40	40	40	5	5	9	9	14
16	16	46	46	43	43	43	48	48	48	18	18	18	12	14	14
4	46	46	46	35	35	48	48	17	17	18	18	30	30	30	14
4	46	41	41	35	35	35	48	17	17	17	11	30	30	30	3
4	4	41	41	49	49	49	19	17	17	17	33	30	30	2	2
4	28	41	41	49	49	49	19	19	33	33	33	21	21	2	2
37	37	37	37	49	49	49	38	19	38	33	20	21	21	24	2
37	37	32	32	36	36	36	38	38	20	20	13	13	13	24	24
39	39	32	32	32	36	36	22	22	20	20	20	13	6	6	6
39	39	32	28	28	7	7	22	22	22	20	44	44	20	6	6
39	39	39	28	28	7	7	22	22	22	20	44	44	6	6	6
39	34	34	34	28	28	7	26	26	26	45	44	44	44	1	1
19	34	34	34	34	8	8	26	26	26	45	45	45	1	1	1
7	34	34	34	34	8	8	27	27	44	29	29	29	25	25	25
31	31	31	31	31	8	8	27	27	27	29	29	25	25	25	25

Female: Green 18/49 Students (36.7%) 101/256 Cells (39.5%)
Male: Blue 31/49 Students (63.3%) 155/256 Cells (60.5%)

Figure 6.8 Standard SOM Gender map. Green cells are female students while blue cells are male students.

16	16	10	10	15	15	42	42	40	40	5	5	5	9	9	47
16	16	10	10	15	15	42	42	40	40	40	5	5	9	9	14
16	16	46	46	43	43	43	48	48	48	18	18	18	12	14	14
4	46	46	46	35	35	48	48	17	17	18	18	30	30	30	14
4	46	41	41	35	35	35	48	17	17	17	11	30	30	30	3
4	4	41	41	49	49	49	19	17	17	17	33	30	30	2	2
4	28	41	41	49	49	49	19	19	33	33	33	21	21	2	2
37	37	37	37	49	49	49	38	19	38	33	20	21	21	24	2
37	37	32	32	36	36	36	38	38	20	20	13	13	13	24	24
39	39	32	32	32	36	36	22	22	20	20	20	13	6	6	6
39	39	32	28	28	7	7	22	22	22	20	44	44	20	6	6
39	39	39	28	28	7	7	22	22	22	20	44	44	6	6	6
39	34	34	34	28	28	7	26	26	26	45	44	44	44	1	1
19	34	34	34	34	8	8	26	26	26	45	45	45	1	1	1
7	34	34	34	34	8	8	27	27	44	29	29	29	25	25	25
31	31	31	31	31	8	8	27	27	27	29	29	25	25	25	25

Orange: Left Handed 8/49 Students (16.3%) 36/256 Cells (14.1%)
Grey: Right Handed 41/49 Students (83.7%) 220/256 Cells (83.7%)

Figure 6.9 Standard SOM Hand dominance map. Orange cells are left-handed while grey cells are right handed students.

6.7 Structured Self-Organizing Map (SSOM)

Although the combination of SOM and *k-means* as discussed in the previous section is a powerful alternative to pure *k-means* clustering, it is sometimes useful to have a pre-defined structure in the SOM for some applications where some stability in the distribution of emotions in the trained SOM would be desirable. Recall that since the SOM methodology assigns random weights at the start of training, the academic emotions can be anywhere in the map after training. Although the general relationships among the clusters would be maintained, the trained SOM (even if trained with the same training parameters and with exactly the same dataset) would yield a different trained map each time.

We thus present initial experimentations on a Structured Self-Organizing Map (SSOM), that is based on earlier work on the topic, but applied to a different type of data (i.e. music files) (Azcarraga & Manalili, 2011). Before the start of the training, some regions of the map would be pre-assigned to specific class labels. A semi-labeled dataset is then used for training, where instances of a given label would only search for best-matching units (**bmu**) from among the nodes that have been pre-assigned to the given label. The resulting methodology is thus a semi-supervised SOM.

Figure 6.10 shows a SSOM with 4 emotion corners, each of which has 6x6 nodes. Unlike a standard SOM where data points can be associated to any node of the map, SSOM restricts the selection of bmu's for a particular region in the map, depending on the emotion tag. For example, instances with a label **F** (dominant emotion is Frustrated) are limited to the bmu's in the upper left corner of the map.

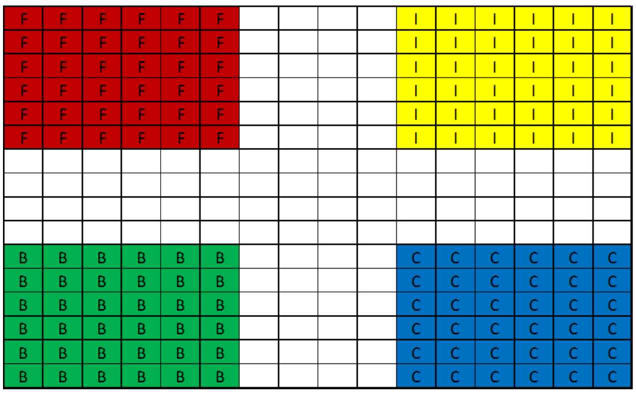


Figure 6.10 16 x 16 Structured SOM with 4 emotion corners (6x6 per corner)

In adapting the Strucured SOM methodology to the analysis and visualization of EEG signals as they relate to academic emotions, three different training schemes were studied, namely, Structured SOM *Cross LM*, Structured SOM *Anywhere LM,* and Structured SOM *Nowhere LM*. All experiments follow the same SOM pre-defined structure as depicted in Figure 6.10 but differ in terms of finding the bmu's during training for **L** (no dominant emotion) and **M** (multiple dominant) emotion tags.

For the SSOM *Cross LM*, the bmu's for datapoints with label **L** and **M** are selected only within the inner cross of the map (white cells). In SSOM *Anywhere LM*, bmu's of such data points can be anywhere in the map. In SSOM *Nowhere LM*, datapoints with labels L or M are not included during the training of the structured SOM (i.e. ignored during training). These are only included once the SOM nodes are being labeled.

The algorithm in Procedure 6.1 below formally describes the Structured SOM training algorithm. Let *map* be a matrix with *n* rows and *m* columns, and each element *map* (i,j) has a vector of weights *map(i,j).weights* and a set of pre-assigned labels denoted as *map(i,j).assignment*. EEG instances that are being used for training are denoted as *currentInstance* and have a vector of features and an assigned label. The vector of weights and the vector of features have the same length. The learning rate *a(t)* and neighborhood size *r(t)* are defined in the same way as such parameters are defined for standard SOMs, where *t* is the number of iterations, and *tMax* is the maximum number of training iterations. Note that the search for the best matching units is restricted among those map element or node whose pre-defined "set of assigned labels" (*map(i,j).assignment*) would include the assigned label of the current instance that has been selected during training.

```
a (0) = getInitialLearningRate ()
r (0) = getIinitial NeighborhoodSize ()
a (tMax) = getFinalLearningRate ()
r (tMax) = getFinalNeighborhoodSize ()

for t = 1 to tMax

   currentInstance = getNextInstance (t)
   minDistance = bigValue

    for i = 1 to n
        for j = 1 to m

          if currentInstance.label ∈ map(i,j).assignment
                distance = getDistance (map(i,j).weights, currentInstance.features)
             if distance < minDistance
                bmuRow = i
                bmuColumn = j
                   minDistance = distance
             endif
          endif

          map.weights   = computeWeights (bmuRow, bmuColumn, currentInstance.features, a(t), r(t))
          a(t+1) = updateLearningRate (a(0), a(tMax), t);
          r(t+1) = updateNeighborhoodSize (r(0), r(tMax), t);

         endfor
    endfor
endfor
```

Procedure 6.1 Training algorithm for a Structured SOM

6.8 Comparison of Training Schemes for Structured SOM

Given the three training schemes for the Structured SOM, we proceed now with comparing the relative quality of the respective SOMs that were produced after training. Note that the first two training schemes include the L and M tags, but the search for their bmu's is pre-determined differently, i.e. the Cross LM is limited to the inner cross section of the map, while the Anywhere LM has no restriction and can find bmu's anywhere in the map. The third training scheme, Nowhere LM, completely ignores during training the instances with tags L and M.

Tables 6.10 and 6.11 show the results using Structured SOM Cross LM for the Test of Difference of Proportions for gender and hand-dominance, respectively. Essentially the same results are derived for the trained map of Structured SOM Anywhere LM. These are shown in Tables 6.12 and 6.13, as the Test of Difference of Proportions for gender and hand-dominance, respectively. The distribution of clusters, labeled maps for the dominant emotions, and the specific maps for gender and hand-dominance are shown in Figures 6.11 to 6.14, for the Structured SOM Cross LM, and Figure 6.15 to 6.18 for Structured SOM Anywhere LM.

Note how the emotions are this time more regrouped in both Figure 6.12 and 6.16, compared to the standard SOM of Figure 6.7. On the upper-left corner are mostly the red cells that correspond to the Frustrated emotion. On the lower left corner are the green cells that correspond to the Bored emotion. Not so distinct are the blue cells, corresponding to the Confused emotion that are located at the right bottom corner. This is slightly more noticeable in Figure 6.16 of the Structured SOM Anywhere LM map. Finally, on the upper right corner are mostly the yellow cells that correspond to the Interested emotion.

Regarding the blue cells that correspond to the Confused emotion in both Figure 6.12 and 6.16, notice how a significant group of such cells are also found in the upper-left corner, corresponding to the Frustrated corner. This is true whether the Structured SOM Cross LM is used or the Structured SOM Anywhere LM. There are a number of blue cells in the diagonal region that bridges the two corners of Frustrated and Confused. Note also that in both Figure 6.12 and Figure 6.16, the Interested nodes (yellow cells) are also intermixed with the confused cells. This may signify that confusion among gifted learners may be a natural emotion felt by students when they are interested in a topic – and in fact, confusion can be a motivation for further trying out the software until the skill is mastered, or the topic is understood.

Tables 6.14-6.15 and Figures 6.19-6.22 present the very similar results for the SSOM Nowhere LM training scheme, where instances with labels L or M are not included during training of SOM Map.

Aside from the preceding tables and figures, which are rather subjective bases for visually comparing the results of the three training schemes, we discuss below a performance measure that would compute for the relative quality of the three training schemes. The performance measures are again based on the earlier work by Azcarraga and Manalili (2011), but some improvements have been made as discussed below.

The performance of a Structured SOM can be computed by computing the distance of a certain instance with a dominant tag Q (i.e. F, I, C, or B) with respect to the SOM corner pre-assigned to label Q. This distance is then compared to a similarly computed distance to the three other corners of the map, corresponding to the three other emotions. The instances with dominant tags L and M are not included in the computation of the performance measure.

In terms of the distance that was measured, two distances are considered, both using the euclidean distance of the bmu in the map. Given a trained SSOM, using any of the three training schemes, the bmu of an instance with tag Q is selected by comparing its distance to each and every node in the map, including those nodes outside of the pre-assigned corner that was used during training. The node with the smallest distance, is the standard basis for selecting the best matching unit (bmu).

Once the bmu is identified, the bmu's x and y coordinates in the map are used to compute first for its distance (Euclidean distance) to each of the four corners of the map. In the 16x16 SOM we used in all the experiments, we have the corners with (x,y) coordinates (1,16) for frustrated, (1,1) for Bored, (16,16) for Interested, and (16,1) for confused. Tables 6.16 – 6.18 show the quality of SSOM Cross LM, Anywhere LM and Nowhere LM, respectively, based on the distance of individual EEG instances to the 4 emotion corners. Note that the distances in the diagonal are the minimum values for all rows and for all columns, and this is true for all three training schemes.

The problem with the original performance measure described above, is that an EEG instance whose bmu is somewhere within the 6x6 sub-SOM of its correct emotion tag would register distances that are not zero if the bmu just happens to be not the exact corner node of the sub-SOM. This is an anomaly that is corrected by the new performance measure that is introduced in this study, which assigns a distance of 0 when the bmu is within the correct sub-SOM; and the distances outside of the correct sub-SOM are computed vis a vis any the nearest edge-node of the four sub-SOMs in each of the four corners of the map.

Tables 6.19 – 6.21 show the relative quality of SSOM Cross LM, Anywhere LM and Nowhere LM, respectively, based on the average distances of individual EEG instances to the edges of the four emotion sub-SOMs in the map. Obviously, the distances shown in Tables 6.19 to 6.21 are all smaller than the distances shown in Tables 6.16 to 6.18, since distances to the 4 corners are always the largest distances possible for any node location in the map.

The SOM is of good quality, as far as the association of the EEG instances to their dominant emotion tags is concerned, when the values in the diagonal elements in each of the matrices are smaller than any of the values in the respective row of column. For example, the average distance of the tag B must be smallest vis a vis the B sub-SOM, compared to the average distances to the sub-SOMs for tags F, C and I.

As can be seen in all the matrices depicting the relative SOM quality in Tables 6.19 to 6.21, the values in the diagonal are significantly lower than the rest of the values in the row or column in the matrix. In general, therefore, for all the three training schemes that were employed, the SSOM trained SSOM were of good quality. For emotions B and F, the smallest distances are registered by the Structured SOM *Nowhere L M*, while for emotion I and C, the smallest distances are registered by the Structured SOM *Anywhere LM*. The relative differences, however, between the distances to the correct node vis a vis the distances to the sub-SOM of the three other tags are all significantly large, implying that all three training schemes result in trained SOMs that re of good quality.

Table 6.10. Gender proportions for Structured SOM Cross L and M

Cluster	Male	Female	p1	p2	n1	n2	n	p	SE	t statistic	p value (two-tailed)
1 *	9,988	4,961	0.67	0.62	14,949	19,028	33,977	0.64	0.0052	8.723	2.841E-18
2	1,261	727	0.63	0.64	1,988	31,989	33,977	0.64	0.0111	-0.791	0.429
3 *	3,255	4,237	0.43	0.70	7,492	26,485	33,977	0.64	0.0063	-42.567	0
4 *	560	925	0.38	0.65	1,485	32,492	33,977	0.64	0.0127	-21.8266	6.831E-105
5 *	2,722	597	0.82	0.62	3,319	30,658	33,977	0.64	0.0088	22.4721	5.027E-111
6 *	4,046	698	0.85	0.61	4,744	29,233	33,977	0.64	0.0075	32.5864	2.197E-229
totals	21,832	12,145									

Table 6.11. Hand dominance proportions for Structured SOM Cross L and M

Cluster	Right	Left	p1	p2	n1	n2	n	p	SE	t	p value (two-tailed)
1 *	12982	1967	0.87	0.84	14,949	19,028	33,977	0.85	0.0039	7.3398	2.19E-13
2 *	1597	391	0.80	0.86	1,988	31,989	33,977	0.85	0.0082	-6.3713	1.9E-10
3 *	6605	887	0.88	0.84	7,492	26,485	33,977	0.85	0.0046	8.0502	8.54E-16
4 *	1479	6	1.00	0.85	1,485	32,492	33,977	0.85	0.0094	15.9432	5.12E-57
5 *	3303	16	1.00	0.84	3,319	30,658	33,977	0.85	0.0065	24.4041	2.1E-130
6 *	2999	1745	0.63	0.89	4,744	29,233	33,977	0.85	0.0056	-46.1347	0
totals	28965	5012									

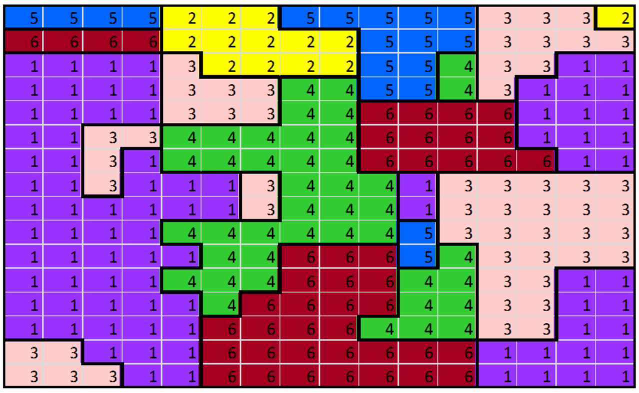


Figure 6.11 Structured SOM Cross LM trained with Aplusix dataset showing six clusters.

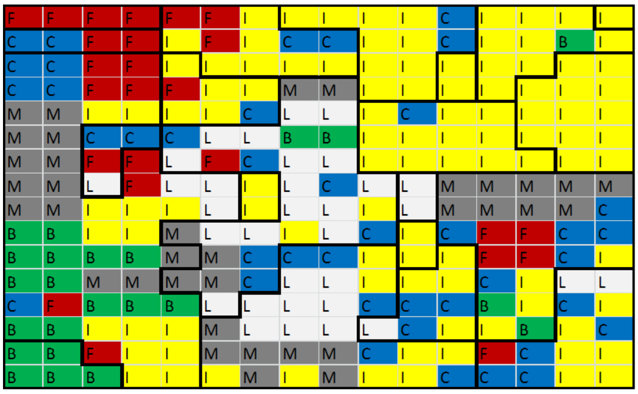
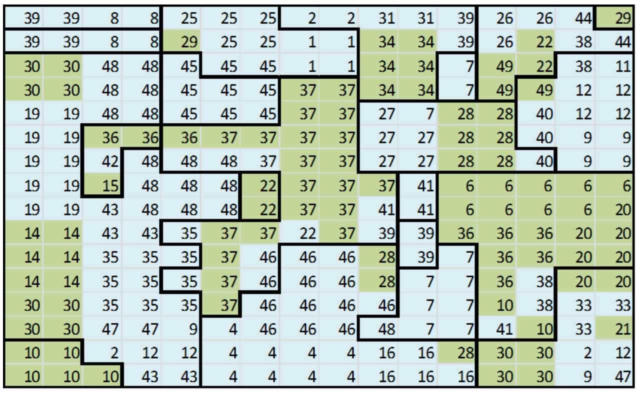
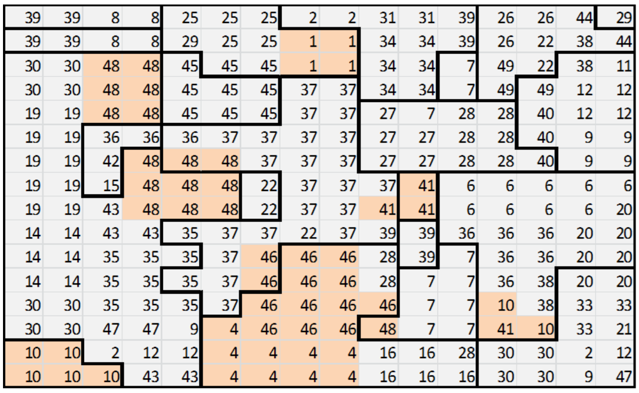


Figure 6.12 Structured SOM Cross LM trained with Aplusix dataset showing six clusters, and each node labeled with the most dominant academic emotion. Frustrated (F) are red cells, Confused (C) are blue cells, Bored (B) are green cells, Interested (I) are yellow, Multiple dominant emotions (M) are grey cells and No dominant emotions (L) are white cells.



Female: Green 18/49 Students (36.7%) 97/256 Cells (37.9%)
Male: Blue 31/49 Students (63.3%) 159/256 Cells (62.1%)

Figure 6.13 Structured SOM Cross LM Gender map superimposed by the student ID. Green cells are female students while blue cells are male students



Orange: Left Handed 8/49 Students (16.3%) 53/256 Cells (20.7%)
Grey: Right Handed 41/49 Students (83.7%) 203/256 Cells (79.3%)

Figure 6.14 Structured SOM Cross LM Hand dominance map superimposed by the student ID. Orange cells are left handed while grey cells are right handed students.

Table 6.12. Gender proportions for Structured SOM Anywhere L and M

Cluster	Male	Female	p1	p2	n1	n2	n	p	SE	t	p value (two-tailed)
1 *	11,020	5,378	0.67	0.62	16,398	17,579	33,977	0.64	0.0052	10.9514	7.2825E-28
2 *	4,283	4,033	0.52	0.68	8,316	25,661	33,977	0.64	0.0060	-27.9212	1.2208E-169
3 *	2,805	715	0.80	0.62	3,520	30,457	33,977	0.64	0.0085	20.1785	5.10113E-90
4 *	2,381	396	0.86	0.62	2,777	31,200	33,977	0.64	0.0095	24.6532	5.0334E-133
5	1,266	730	0.63	0.64	1,996	31,981	33,977	0.64	0.0111	-0.7960	0.426057153
6 *	77	893	0.08	0.66	970	33,007	33,977	0.64	0.0156	-37.1326	7.044E-296
totals	21,832	12,145									

Table 6.13. Hand dominance proportions for Structured SOM Anywhere L and M

Cluster	Right	Left	p1	p2	n1	n2	n	p	SE	t	p value (two-tailed)
1 *	14,450	1,948	0.88	0.83	16,398	17,579	33,977	0.85	0.0038	14.4167	5.59E-47
2 *	6,936	1,380	0.83	0.86	8,316	25,661	33,977	0.85	0.0045	-5.4546	4.94E-08
3 *	3,505	15	1.00	0.84	3,520	30,457	33,977	0.85	0.0063	25.3139	4.5E-140
4 *	1,500	1,277	0.54	0.88	2,777	31,200	33,977	0.85	0.0070	-48.4361	0
5 *	1,604	392	0.80	0.86	1,996	31,981	33,977	0.85	0.0082	-6.3476	2.21E-10
6 *	970	0	1.00	0.85	970	33,007	33,977	0.85	0.0116	13.1445	2.28E-39
totals	28,965	5,012									

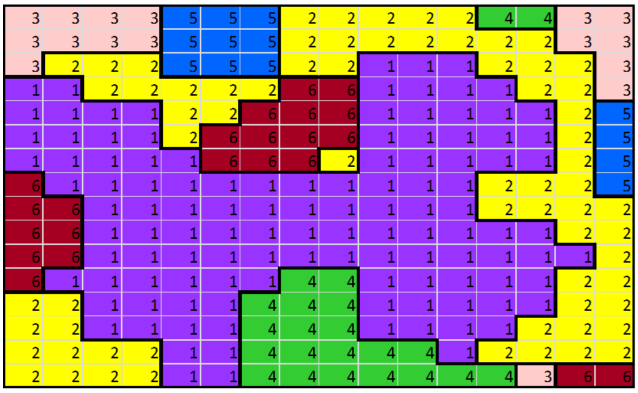


Figure 6.15 Structured SOM Anywhere LM trained with Aplusix dataset showing six clusters

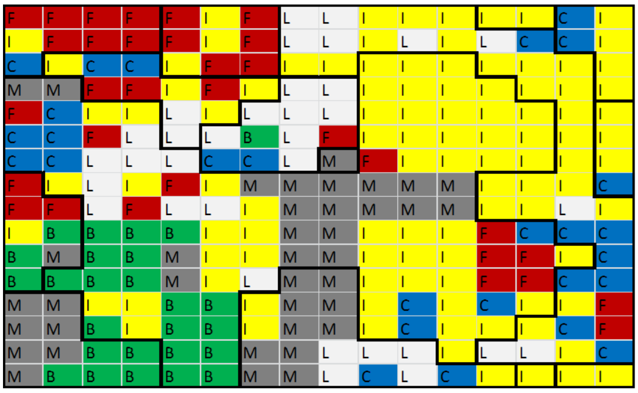


Figure 6.16 Structured SOM Anywhere LM trained with Aplusix dataset showing six clusters, and each node labeled with the most dominant academic emotion. Frustrated (F) are red cells, Confused (C) are blue cells, Bored (B) are green cells, Interested (I) are yellow, Multiple dominant emotions (M) are grey cells and No dominant emotions (L) are white cells.

39	8	8	8	25	25	25	41	41	27	27	28	28	28	39	31
39	8	8	8	25	25	25	41	41	40	41	41	41	7	39	34
39	7	7	7	29	25	25	41	41	9	9	9	49	49	7	34
19	19	42	42	45	45	45	37	37	9	9	9	22	22	26	26
30	30	48	48	48	45	37	37	37	23	2	12	12	38	44	44
30	30	48	48	48	37	37	37	37	23	23	12	12	38	44	29
30	30	48	48	48	37	37	37	19	24	23	12	18	18	44	29
37	48	48	48	48	48	19	19	19	19	19	19	22	22	22	1
37	37	48	48	48	48	43	19	19	19	19	19	22	22	22	20
37	35	35	35	35	43	43	19	19	33	33	33	30	30	20	20
37	35	35	35	35	43	43	19	19	33	33	33	30	30	20	20
37	35	35	35	35	43	43	4	4	33	33	33	30	30	20	20
6	6	47	9	30	30	4	4	4	47	47	47	38	38	38	36
6	6	10	2	30	30	4	4	4	47	47	47	15	15	36	36
6	6	10	10	14	14	4	4	46	46	46	40	41	15	7	7
6	10	10	10	14	14	4	4	46	46	46	16	28	39	7	7

Orange: Left Handed 8/49 Students (16.3%) 54/256 Cells (21.1%)
Grey: Right Handed 41/49 Students (83.7%) 202/256 Cells (78.9%)

Figure 6.17 Structured SOM Anywhere LM Hand dominance map superimposed by student ID. Orange cells are left handed while grey cells are right handed students.

39	8	8	8	25	25	25	41	41	27	27	28	28	28	39	31
39	8	8	8	25	25	25	41	41	40	41	41	41	7	39	34
39	7	7	7	29	25	25	41	41	9	9	9	49	49	7	34
19	19	42	42	45	45	45	37	37	9	9	9	22	22	26	26
30	30	48	48	48	45	37	37	37	23	2	12	12	38	44	44
30	30	48	48	48	37	37	37	37	23	23	12	12	38	44	29
30	30	48	48	48	37	37	37	19	24	23	12	18	18	44	29
37	48	48	48	48	48	19	19	19	19	19	19	22	22	22	1
37	37	48	48	48	48	43	19	19	19	19	19	22	22	22	20
37	35	35	35	35	43	43	19	19	33	33	33	30	30	20	20
37	35	35	35	35	43	43	19	19	33	33	33	30	30	20	20
37	35	35	35	35	43	43	4	4	33	33	33	30	30	20	20
6	6	47	9	30	30	4	4	4	47	47	47	38	38	38	36
6	6	10	2	30	30	4	4	4	47	47	47	15	15	36	36
6	6	10	10	14	14	4	4	46	46	46	40	41	15	7	7
6	10	10	10	14	14	4	4	46	46	46	16	28	39	7	7

Female: Green 18/49 Students (36.7%) 86/256 Cells (33.6%)
Male: Blue 31/49 Students (63.3%) 170/256 Cells (66.3%)

Figure 6.18 Structured SOM Anywhere LM Gender map superimposed by student ID. Green cells are female students while blue cells are male students

Table 6.14. Gender proportions for Structured SOM Anywhere L and M

Cluster	Male	Female	p1	p2	n1	n2	n	p	SE	t	p value (two-tailed)
1*	3075	3456	0.47	0.68	6,531	23,881	30,412	0.64	0.007	-31.67	1.20E-216
2*	10441	5199	0.67	0.62	15,640	14,772	30,412	0.65	0.005	8.45	3.09E-17
3*	2724	1196	0.69	0.64	3,920	26,492	30,412	0.64	0.008	7.22	5.38E-13
4*	690	0	1.00	0.64	690	29,722	30,412	0.64	0.018	19.78	1.53E-86
5*	1338	9	0.99	0.63	1,347	29,065	30,412	0.64	0.013	27.37	5.37E-163
6*	1566	718	0.69	0.64	2,284	28,128	30,412	0.64	0.010	4.43	9.41E-06
totals	19834	10578									

Table 6.15. Hand dominance proportions for Structured SOM Anywhere L and M

Cluster	Right	Left	p1	p2	n1	n2	n	p	SE	t	p value (two-tailed)
1*	5307	1224	0.81	0.86	6,531	23,881	30,412	0.85	0.005	-9.94	2.89E-23
2*	13458	2182	0.86	0.85	15,640	14,772	30,412	0.85	0.004	3.65	0.000262
3*	3656	264	0.93	0.84	3,920	26,492	30,412	0.85	0.006	14.98	1.41E-50
4*	690	0	1.00	0.85	690	29,722	30,412	0.85	0.014	11.04	2.86E-28
5*	1338	9	0.99	0.85	1,347	29,065	30,412	0.85	0.010	14.87	8.24E-50
6*	2279	5	1.00	0.84	2,284	28,128	30,412	0.85	0.008	20.26	1.09E-90
totals	26728	3684									

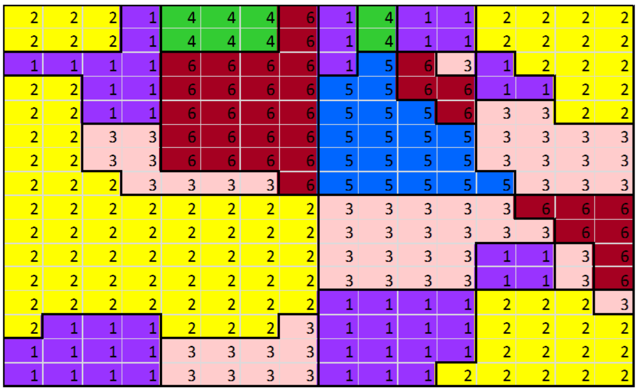


Figure 6.19 Structured SOM Nowhere LM trained with Aplusix dataset showing six clusters

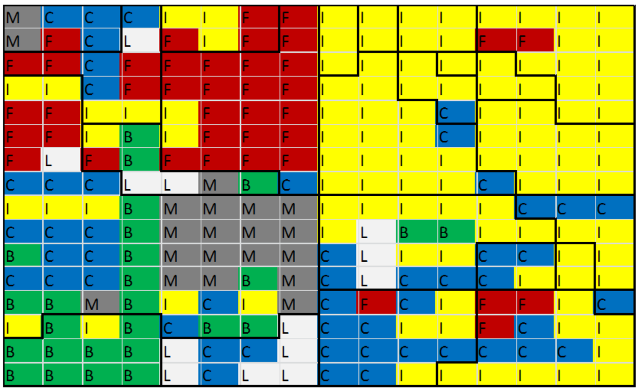


Figure 6.20 Structured SOM Nowhere LM trained with Aplusix dataset showing six clusters, and each node labeled with the most dominant academic emotion. Frustrated (F) are red cells, Confused (C) are blue cells, Bored (B) are green cells, Interested (I) are yellow, Multiple dominant emotions (M) are grey cells and No dominant emotions (L) are white cells.

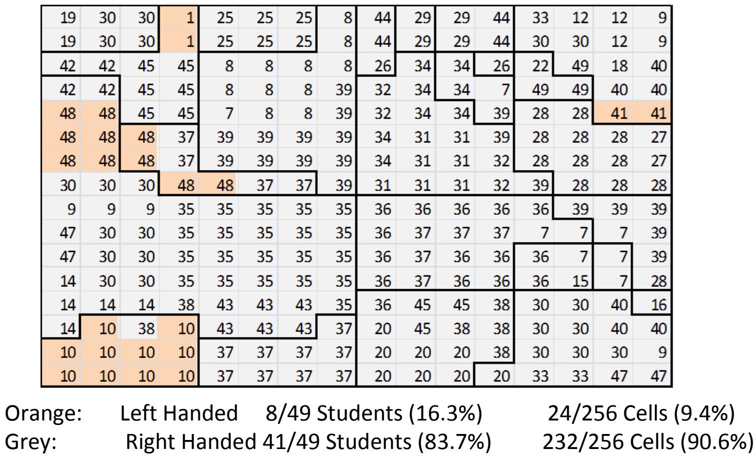


Figure 6.21 Structured SOM Nowhere LM Hand dominance map superimposed by student ID. Orange cells are left handed while grey cells are right handed students.

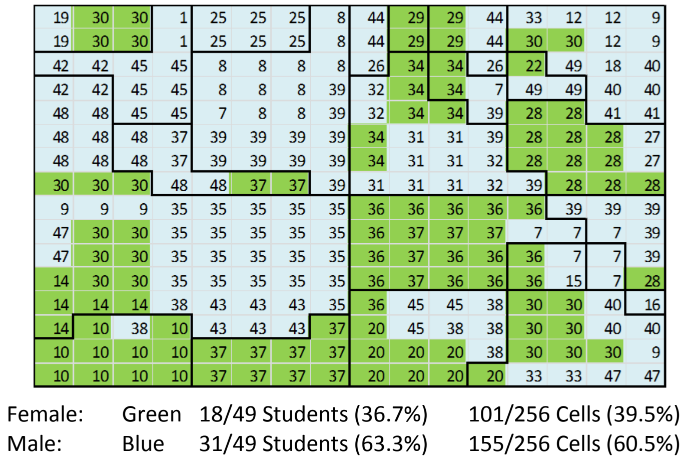


Figure 6.22 Structured SOM Nowhere LM Gender map superimposed by student ID. Green cells are female students while blue cells are male students

Table 6.16. Measure of the quality of SSOM Cross LM based on distance to the four emotion corners

	True I	True C	True F	True B
Pred I	7.75	13.88	12.15	17.93
Pred C	11.48	8.65	16.38	13.67
Pred F	13.10	14.66	6.35	12.81
Pred B	15.39	10.89	12.83	3.86

Table 6.17. Measure of the quality of SSOM Anywhere LM based on distance to the four emotion corners

	True I	True C	True F	True B
Pred I	7.46	12.97	12.08	16.56
Pred C	11.38	8.47	15.49	12.25
Pred F	12.66	14.24	6.65	12.72
Pred B	15.21	11.02	12.80	5.39

Table 6.18. Measure of the quality of SSOM Nowhere LM based on distance to the four emotion corners

	True I	True C	True F	True B
Pred I	7.41	14.53	11.50	16.84
Pred C	12.25	8.94	15.89	12.76
Pred F	12.53	14.28	6.63	12.31
Pred B	15.75	10.02	12.90	4.95

Table 6.19. Measure of the quality of SSOM Cross LM based on distance of individual EEG instances to the edges of the four emotion sub-SOM

	True I	True C	True F	True B
Pred I	3.15	8.09	6.76	11.02
Pred C	6.50	3.99	9.76	8.30
Pred F	7.61	8.85	2.02	7.52
Pred B	9.11	6.60	7.52	0.74

Table 6.20. Measure of the quality of SSOM Anywhere LM based on distance of individual EEG instances to the edges of the four emotion sub-SOM

	True I	True C	True F	True B
Pred I	2.60	7.09	6.49	9.73
Pred C	6.05	3.17	9.04	6.77
Pred F	7.00	8.26	2.11	7.19
Pred B	8.73	5.80	7.19	1.00

Table 6.21. Measure of the quality of SSOM Nowhere LM based on distance of individual EEG instances to the edges of the four emotion sub-SOM

	True I	True C	True F	True B
Pred I	3.08	8.68	6.24	9.86
Pred C	7.26	4.20	9.24	7.08
Pred F	7.39	8.28	1.79	6.70
Pred B	9.45	5.18	7.13	0.64

Chapter 7: Conclusion and Recommendations

7.1. Conclusion

Brainwaves, or EEG signals, collected from fifty six (56) academically-gifted students while they learn independently using two computer-based learning systems were analyzed, visualized, and used to predict the academic emotion of the students based on computational intelligence techniques. The EEG signals of the students were collected using an Emotiv EPOC sensor set, that was attached to their head during the 40-minute experimental session, while they were made to learn to use Aplusix and Scooter, both of which are mathematics oriented software designed to elicit various academic emotions in the learners.

The academic emotions that were considered in the study were *frustrated, confused, bored,* and *interested*. The 40-minute learning session was divided into 4 distinct learning phases, namely 1) Scatterplot Slide, 2) Scooter 3) Aplusix Slide, and 4) Aplusix. Sessions 1 and 3 are used to do a slide presentation about how the Scooter and Aplusix software are to be used. Sessions 2 and 4 are interactive sessions where the students actually use the learning software to learn about algebra and about how to plot graphs.

During the course of the entire 40-minute session, the learners were prompted by the system every two minutes to evaluate themselves in terms of the four emotions. These self-declared emotion tags were used as labels for the training and testing of the datasets, and a separate dataset was maintained for each of the four learning phases (also referred to as *sessions*).

The 40-minute session is preceded by an orientation as well as a period of relaxation where a "resting state" is used to determine the baseline EEG readings for each student. These baseline EEG readings are then used to calibrate the raw EEG signals that are collected per student and to convert these to standard and normalized scores prior to classification training and prediction.

The results, analysis and discussion of results, and the various new techniques that are proposed in this study are significant in several ways:

- There were numerous tedious experimentations performed on the various computational intelligence and data mining techniques that were employed. As a result of the numerous experiments that were tried and evaluated, this study provides detailed information on how best to prepare, clean, and pre-process EEG data. These information would be very useful for deeper studies in the future on how EEG data, and other similar data, should be used to classify and predict the academic emotions of learners.

- The results on Principal Components Analysis as well as the computation of z scores as basis for feature selection provide two distinct methods for reducing the dimensionality of the datasets. Dimensionality reduction is critical especially when computationally intensive classifiers are used, such as the Multi-Layered Perceptrons (MLP), given the large volumes of data that are being processed. It was shown in this study that comparable prediction performance rates can be achieved with just about 30 selected features, compared to the original 126 features used for calculating the baseline classification and prediction performance on the datasets.

- Decision trees were first used to predict the four academic emotions, using accuracy and Area Under the Curve (AUC) as the main performance metrics. The initial prediction performance rates were only slightly above 0.50. Given the baseline performance rates yielded by the Decision trees, Multi-Layered Perceptrons (MLP) were then used to enhance performance rates, achieving rates of 0.52 to 0.64 for accuracy and 0.52 to 0.65 for AUC.

- Classification performance was shown to further improve by isolating the datasets based on gender, and/or handedness, and building separate classification models for each. By having such restricted datasets and separate models on them, the Accuracy and AUC rates were as high as 0.75 for datasets where all learners are only male and right-handed.

- Performance rates are even further improved when selective predictions are made, where prediction of emotions need not be performed all the time. In particular, prediction is attempted only when the number of very high or very low feature values among the instances is about 20% or higher. These very high or very low feature values are referred to as "special event features". In the study, a detailed and methodical procedure is presented as to how such features can be detected.

- As a final step in the analysis and visualization of the EEG data, unsupervised clustering of the Aplusix datasets is performed. Using *k-means* clustering and Self-Organizing Maps (SOM), a new statistics-based method is presented for determining the influence of the students' profile, such as gender and handedness, on the formation of the clusters. The influence of the students' profile is computed using a Test of Difference of Proportions. The significance of the difference in proportions among the clusters can thus be detected using a method that is backed by pertinent concepts in statistics. Special features on a per cluster basis, based on feature z scores, are also plotted on a schematic diagram of a human brain which also very clearly show that certain clusters have specialized on specific regions of the map (i.e. left hemisphere for cluster 5), or on certain frequency bands (i.e. alpha band for cluster 2 and gamma and beta high bands for cluster 1).

- Experimental results on the use of a novel structure for a SOM are also presented. Instead of allowing the clusters of nodes to form anywhere on the SOM map, which is the standard SOM methodology, a pre-defined structure is assigned prior to training. In the experiments, 16x16 SOM maps were used, and each corner composed of a 6 x 6 sub-SOM was pre-assigned to one of the four academic emotions. Results on the experiments on these new structured SOM, combined with k-means clustering, pave the way for further analysis and visualization of the data.

- Three training schemes have been developed to build on the structured SOM (SSOM) approach to EEG visualization. To compare the quality of the trained SSOM, two performance measures have been proposed, building on an earlier work on the use of SSOM on a music archive.

For now, all the experiments that were performed only concern academically gifted students. Indeed, this is the sector of students who would benefit the most from independent learning – such as when using computer-based learning modules to learn about specific topics of interest on their own. The results of the study would certainly pave the way for designers of future learning systems to incorporate affect-related features in the system that respond positively to the special needs of learners.

Such future affect-aware learning systems should have effective predictors of the academic emotion of students – that are able to recognize, with high prediction accuracy, those critical points during learning sessions when students are getting confused about the topic, or even worse, when they are about to get frustrated or bored with the learning session. Such affect-aware systems could then prevent frustration or boredom to set in by altering the tutoring tasks, by reinforcing the lessons on the possible sources of confusion using examples and illustrations, by adjusting the level of difficulty of the problems and questions posed, and various other pedagogical and instructional interventions.

Clearly, the classifiers must have high prediction accuracies, otherwise any of the three negative emotions (boredom, confusion, and frustration) might go unnoticed, or the students may be mistaken to be bored, for example, when in fact he or she is just confused. When prediction fails, either as false negatives or false positives, the interventions in terms of tutoring adjustments would therefore become ineffective. Indeed, the timeliness of tutor interventions is critical in scaffolding the learning experience of the learner, and the interventions, obviously, must correspond correctly to the academic emotion of the learner. The various techniques that were presented in the study serve to improve the prediction accuracies of the classifiers that would be developed.

Accordingly, the profile of the learner, being male or female, or being right- or left-handed, would determine which specialized classifier would be used to predict academic emotions based on his/her EEG signals. In addition, prediction will not be attempted on a continuous basis. Rather, the values of the features would be monitored and only when the number of "special event features" has exceeded a certain threshold (e.g. 20%) would the system attempt to do a prediction of the academic emotion of the student.

7.2. Recommendations

The study has opened up a number of research directions that should be pursued as follow-up studies or as entirely new lines of research. Some are smaller studies that have to do with completing various experiments that have not been performed in this study and would complement the results reported here. Others are major lines of research that are truly interesting to pursue, and would greatly enhance whatever initial results have been produced by this study so far.

- Supervised clustering methods that were employed in this study are only Decision Trees and Multi-Layered Perceptrons (MLP). The natural extension for further studies would be to study other classifiers, notably Support Vector Machines (SVM). Other alternatives to Principal Components Analysis, such as Singular Value Decomposition (SVD), should also be pursued. Lastly, faster alternatives to Self-Organizing Maps, such as Multi-Dimensional Scaling (MDS) should be considered.

- The results of this study suggest that more restricted datasets like that of right-handed male students tend to yield higher prediction rates compared to unrestricted sets that include all students, both male and female, and both right-handed and left-handed. The results on purely left-handed students, however, are not reliable because of the small number of subjects. It would be very interesting to pursue this line of research by collecting significantly more data from left-handed students and completing a thorough analysis of restricted datasets. Other pertinent student profiles may also be considered, such as the basic personality profile of *introverted* or *extroverted*, or perhaps even having highly restricted datasets, such as female Mensa members who are visual artists and who are left handed.

- In most of the clustering experiments conducted in this study, one of the findings is that one or two clusters have relatively large numbers of data points relative to the other clusters. Further increasing the value of *k* in *k means* clustering does not always result in these big clusters being split into smaller clusters. A modification in the *k means* clustering method that targets specifically the large clusters in order to break them down into smaller clusters may be considered – without, of course, sacrificing the strict theoretical basis for splitting the clusters into smaller clusters.

- The visualization of EEG data using the combination of *k means* clustering and standard SOM or Structured SOM, has provided some insights as to the characteristics of each cluster, not only in terms of EEG features but in terms also of the dominant emotions and the gender and handedness profile of the students. However, this study explored only one kind of structure where each of the four corners of the map is trained to a specific emotion. Other predefined structures may be considered, such as when the three negative emotions (Boredom, Frustration, and Confusion) are on the left half of the SOM, while the positive emotion (Interested) would be assigned to the right half of the map. Such new structures can be used as the basis for a more refined methodology for a combination of unsupervised clustering and supervised classification.

- It would be interesting to be able to do an on-line plotting of the location of an EEG instance in the Structured SOM map. The current implementation in Octave of the Standard SOM and the Structured SOM can specifically locate the node in the map that is associated to each EEG instance in the dataset. It would be useful to synchronize the plotting of the node locations with the actual raw EEG signals as they are being captured from the Emotive EPOC headset so that the emotion associated to the nodes in the map, and more importantly the trajectory, can be visualized. With such a software application, one can view the facial expression of the student, the past and current answers provided to the learning software, as well as the hand and body movements as they might correlate with the past and current emotion readings displayed on the map.

- The trajectory of the dominant academic emotion should be studied more carefully as a sequence pattern of EEG signals. For example, certain patterns may be discovered that would be associated to specific cases of students being interested for some time and start to get confused, and when left unchecked proceed to being frustrated. Another pattern might refer to cases when a student starts by being interested but after some time starts to move into the SOM territory for boredom. In both of these examples, it would be useful to be able to recognize the occurrence of such patterns so that an affect-aware learning system can intervene in advance, and thus return the student back to the "interested" territory of the SOM.

- Finally, although the study had to concentrate on a specific age range of students and a specific sector of learners (i.e. academic achievers in mathematics), it would be useful to be able to see which of the results may be ported to other types of students, either of different age range or perhaps students who are poor in mathematics. Students who are poor in mathematics, for example, may benefit greatly from specialized computer-based learning modules, or outright intelligent tutoring systems, that are affect-aware.

References

Ansari-Asl, K., Chanel, G. & Pun, T. (2007). A Channel Selection Method for EEG Classification in Emotion Assessment based on Synchronization Likelihood. *15th European Signal Processing Conference (EUSIPCO)*, Poznan, Poland.

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. M. (2009). Emotion Sensors Go To School. In V. Dimitrova, R. Mizoguchi, B. Du Boulay, & A. Graesser (Eds.), *AIED* (Vol. 200, pp. 17–24). IOS Press.

Azcarraga, A., & Manalili, S. (2011). Design of a Structured 3D SOM as a Music Archive. 10th Workshop on Self-Organizing Maps *WSOM 2011, Helsinki, Finland* (pp. 188–197).

Azcarraga, A., Setiono, R., Hsieh, M. H., & Pan, S.-L. (2008). Improved SOM Labeling Methodology for Data Mining Applications. *Soft Computing for Knowledge Discovery and Data Mining* (pp. 45–75).

Azcarraga, J., Ibanez JR, J. F., Lim, I. R., & Lumanas JR, N. (2011). Predicting Student Affect Based on Brainwaves and Mouse Behavior. *11th Philippine Computing Science Congress*. Naga City, Philippines: Computing Society of the Philippines.

Azcarraga, J., Ibanez JR, J. F., Lim, I. R., Lumanas JR, N., Trogo, R., & Suarez, M. T. (2011). Predicting Academic Emotion based on Brainwaves Signals and Mouse Click Behavior. In T. Hirashima et al. (Ed.), *19th International Conference on Computers in Education (ICCE)* (pp. 42–49). Chiang Mai, Thailand: NECTEC, Thailand.

Azcarraga, J., Inventado, P. S., & Suarez, M. T. (2010). Predicting the Difficulty Level Faced by Academic Achievers based on Brainwaves Analysis. In S. L. Wong (Ed.), *18th International Conference on Computers in Education* (pp. 107–109). Putrajaya, Malaysia: Asia-Pacific for Computers in Education.

Azcarraga, J. J., Ibanez JR, J. F., Lim, I. R., & Lumanas JR, N. (2011). Use of Personality Profile in Predicting Academic Emotion Based on Brainwaves Signals and Mouse Behavior. *2011 Third International Conference on Knowledge and Systems Engineering*, 239–244.

Azcarraga, J. & Suarez, M. (2013). Recognizing Student Emotions using Brainwaves and Mouse Behavior Data. *International Journal of Distance Education Technologies*, *11*(2), 1–15.

Azcarraga, J., Marcos, N. & Suarez, M. (2014). Modelling EEG Signals for the Prediction of Academic Emotions. *Utilizing EEG Input in Intelligent Tutoring Systems of the 12th International Conference on Intelligent Tutoring Systems (ITS 2014).*

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why Students Engage in " Gaming the System " Behavior in Interactive Learning Environments. *Journal Of Interactive Learning Research*, *19*(2), 185–224.

Benadada, K., Chaffar, S., & Frasson, C. (2008). Towards Selection of Tutorial Actions Using Emotional Physiological Data. *WEC ITS*.

Blanchard, E., Chalfoun, P., & Frasson, C. (2007). Towards Advanced Learner Modeling: Discussions on Quasi Real-time Adaptation with Physiological Data. *Advanced Learning Technologies 2007 ICALT 2007 Seventh IEEE International Conference on* (pp. 809–813). Ieee. doi:10.1109/ICALT.2007.264

Burleson, W. (2006). *Affective Learning Companions : strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance. System*. Massachusetts Institute of Technology.

Chanel, G. (2009). *Emotion assessment for affective computing based on brain and peripheral signals. University of Geneva*. Citeseer.

Cherkassky, V., & Mulier, F. (2007). Learning From Data: Concepts, Theory, and Methods Second Edition. *Learning from Data* (Vol. 19, pp. 711–720). John Wiley & Sons, Inc.

Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, *19*(3), 267–303. Retrieved from http://www.springerlink.com/index/10.1007/s11257-009-9062-8

Davidson, R. J., Schwartz, G. E., Pugash, E., & Bromfield, E. (1976). Sex Differences in Patterns of EEG Asymmetry. *BiologicalPsychology 4*, *4*, 119–138.

Davidson, R. J., Ekman, P., Saron, C. D., Senulis, J. A., & Friesen, W. V. (1990). Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology. I. *Journal of Personality and Social Psychology*, *58*(2), 330–41.

Diezmann, C. (2005). Challenging Mathematically Gifted Primary Students. *Australian Journal of Gifted Education*, *14*(1), 50–57.

Dornhege, G., Millán, J., Hinterberger, T., McFarland, D. & Müller, K.B. (2007). Toward Brain-Computer Interfacing. Massachusetts Institute of Technology.

D' Mello, S. K. D., & Graesser, A. (2009). Multimodal Affect Detection from Conversational Cues , Gross Body Language , and Facial Features. *User Modeling and Useradapted Interaction*, 1–27.

D' Mello, S. K., & Graesser, A. (2010). M ultimodal Semi-Automated A ffect Detection from Conversational Cues , Gross Body Language, and Facial Features. *User Modeling and User Adapted Interaction*, *20*(2), 147–187.

Ekman, P. (1984). Expression and the nature of emotion. In K. Scherer & P. Ekman (Eds.), *Approaches to Emotion* (pp. 319–344). Hillsdale, NJ.: Erlbaum.

Emotiv EPOC Headset. (n.d.). Retrieved from http://www.emotiv.com

Fawcett, T. (2006). An Introduction to ROC Analysis. *Pattern Recognition Letters*, *27*, 861–874.

Feldhusen, J. (2005). Giftedness, Talent, Expertise and Creative Achievement. *In the Conception of Giftedness* in (Eds.) Sternberg & Davidson 2005

Forgas, J. P. (1999). Network Theories and Beyond. *Handbook of Cognition and Emotion* (pp. 589–611). John Wiley & Sons, Ltd.

Frantzidis, C. A., Bratsas, C., Klados, M. A., Konstantinidis, E., Lithari, C. D., Vivas, A. B., Papadelis, C. L., et al. (2010). On the Classification of Emotional Biosignals Evoked While Viewing Affective Pictures : An Integrated Data-Mining-Based Approach for Healthcare Applications. *IEEE Trans Inf Technol Biomed*, *14*(2), 309–318.

Furnham, A., Monsen, J., & Ahmetoglu, G. (2009). Typical intellectual engagement, Big Five personality traits, Approaches to learning and cognitive ability predictors of academic performance. *British Journal of Educational Psychology*, *79*(Pt 4), 769–82.

Good, C. D., Johnsrude, I., Ashburner, J., Henson, R. N. A., Friston, K. J., & Frackowiak, R. S. J. (2001). Cerebral Asymmetry and the Effects of Sex and Handedness on Brain Structure : A Voxel-Based Morphometric Analysis of 465 Normal Adult Human Brains, *700*, 685–700. doi:10.1006/nimg.2001.0857

Greene, J., Moos, D., Azevedo, R., & Winters, F. (2008). Exploring differences between gifted and grade-level students use of self-regulatory learning processes with hypermedia. *Computers and Education*, *50*, 1069–1083.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software. *SIGKDD Explorations*, *11*(1).

Haykin, S. (2008). *Neural Networks and Learning Machines*. *Pearson Prentice Hall New Jersey USA 936 pLinks* (p. 906). Prentice Hall.

Heraz, A, Razaki, R., & Frasson, C. (2007). Using machine learning to predict learner emotional state from brainwaves. *Advanced Learning Technologies 2007 ICALT 2007 Seventh IEEE International Conference on* (Vol. 0, pp. 853–857). IEEE Computer Society.

Heraz, Alicia, & Frasson, C. (2009). Predicting Learner Answers Correctness through Brainwaves Assesment and Emotional Dimensions. *Artificial Intelligence in Education*, *200*, 49–56. Retrieved from http://www.iro.umontreal.ca/~frasson/FrassonPub/AIED-2009-Heraz-Frasson.pdf

Heraz, Alicia, Jraidi, I., Chaouachi, M., & Frasson, C. (2009). Predicting Stress Level Variation from Learner Characteristics and Brainwaves. (V. Dimitrova, R. Mizoguchi, B. Du Boulay, & A. C. Graesser, Eds.)*Artificial Intelligence in Education*, *200*, 722–724.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/0893608089900208

Hyndman, R. J. (2009). Moving averages. *International Encyclopedia of Statistical Science* (pp. 866–869).

Ibanez JR, J. F., Lim, I. R., & Lumanas JR, N. (2011). *Affect Recognition Using Brainwaves and Mouse Behaviour for Intelligent Tutoring Systems*. De La Salle University, Manila, Philippines.

Jausovec, N. (2000). Differences in Cognitive Processes Between Gifted, Intelligent, Creative, and Average Individuals While Solving Complex Problems: An EEG Study. *Intelligence*, *28*(3), 213–237.

Jausovec, N., & Jausovec, K. (2000). EEG activity during the performance of complex mental problems. *International Journal of Psychophysiology*, *36*(1), 73–88.

Kapoor, A, Burleson, W., & Picard, R. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies*, *65*(8), 724–736.

Kapoor, Ashish, Mota, S., & Picard, R. W. (2001). Towards a Learning Companion that Recognizes Affect. *AAAI Fall symposium*, (543), 2–4.

Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K.-R., Curio, G., Hagemann, K., et al. (2007). Improving human performance in a real operating environment through real-time mental workload detection. *Toward Brain-Computer Interfacing* (pp. 409–422). Cambridge, MA: MIT Press.

Kohonen, T. (2001). *Self-Organizing Maps* (Third Edit.). Springer.

Kohonen, T. & Honkela, T. (2007). Kohonen Network. *Scholarpedia*, *2*(1), 1568.

Kort, B., Reilly, R., & Picard, R. (2001). An Affective Model of interplay between emotions and learning: Reengineering Educational Pedagogy-Building a Learning Companion. *International Conference on Advanced Learning Technologies*.

Le Cun, Y. (1988). A Theoretical Framework for Back-Propagation. In D. Touretzky, G. Hinton & T. Sejmnowski (Eds) *Proceedings of the 1988 Connectionist Models Summer School* , 21-28, CMU, Pittsburgh, Pa. Morgan Kaufmann.

Lee, C.J. & Tan, D. (2006). Using a Low-Cost Electroencephalograph for Task Classification in HCI Research. *19th annual ACM Symposium on User Interface Software and Technology* (Vol. 7, pp. 81–90).

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297). Berkeley: University of California Press.

Maimon, O. & L. Rokach (2007). *Soft Computing for Knowledge Discovery and Data Mining*. Springer- Verlag.

Manalili, S. (2011*). i3DMO: Interactive 3D Music Organizer*. Masteral Thesis, De La Salle University.

Mandler, G. 1999. *Handbook of perception and cognition 2nd ed*., B. M. Bly & D. E. Rumelhart Eds., San Diego, CA: Academic Press. pp. 367-384

Mao, X. and Li, Z. (2010). Agent based affective tutoring systems: A pilot study. *Computers and Education, 55*, 202-208.

McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics, 5,* 115-133.

Murugappan, M., Ramachandran, N., & Sazali, Y. (2010). Classification of human emotion from EEG using discrete wavelet transform. *Journal of Biomedical Science and Engineering*, *3*(April), 390–396.

Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neuroscience & Biobehavioral Reviews*, *33*(7), 1004–1023.

Nicaud, J.-F., Bouhineau, D., & Huguet, T. (2002). The Aplusix Editor: A new kind of software for the learning of algebra. In S. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *Lecture notes in computer science* (Vol. 2363, pp. 178–187). Berlin/Heidelberg: Springer.

Ortony, A., Collins, A., & Clore, G. L. (1988). *The cognitive structure of emotions*. Cambridge [England]; New York: Cambride University Press.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, *37*(2), 91–105.

Petrantonakis, P.C. and Hadjileontiadis, L. J. (2010). Emotion Recognition from Brain Signals Using Hybrid Adaptive Filtering and Higher-Order Crossings Analysis. *IEEE Transactions on Information Technology in Biomedicine*, *14*(2), 186–197.

Plotkin, W. B. (1976). On the Self-Regulation of the Occipital Alpha Rhythm : Control Strategies , States of Consciousness , and the Role of Physiological Feedback. *Journal of Experimental Psychology: General*, *105*(1), 66–99.

Pfurtscheller, G., Müller-Putz, G., Graimann, B. et al (2004). Graz-Brain-Computer Interface: State of Research. Toward Brain-Computer Interfacing. 63-84.

Pintrich, P.R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in clasroom academic tasks. In D.H. Schunk & J.Meece (Eds.), *Student Perceptions in the Classroom*, 149-179, Hillsdale, NJ: Lawrence Erlbaum.

Rapidminer Tool. (n.d.). Retrieved from http://www.rapidminer.com

Rayneri, L.J, Gerber, B.L. & Wiley, L.P (2006). The Relationship Between Classroom Environment and the Learning Style Preferences of Gifted Middle School Students and the Impact on Levels of Performance. The Gifted Child Quarterly, Spring 2006, 50(2). pp. 104-117.

Risemberg, R., and Zimmerman, B. (1992). Self-Regulated Learning in Gifted Students., Roeper Review, 15 (2), 98-101

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing* (Vol. 1, pp. 318–362). MIT Press.

Sanei, S., & Chambers, J. A. (2007). *EEG signal processing*. *Book* (Vol. 1, p. 289). Wiley-Interscience.

Seitz, R. J., Schäfer, R., & Scherfeld, D. (2008). Valuating Other People ' S Emotional Face Expression : A Combined Functional Magnetic Resonance Imaging And Electroencephalography Study. *Neuroscience*, *152*, 713–722. doi:10.1016/j.neuroscience.2007.10.066

Seret, Alex; vanden Brouke, K.L.M; Baesens, Bart; Vanthienen, J. (2014). A dynamic understanding of customer behavior processes based on clustering and sequence mining. *Expert Systems Application*, *41*(10), 4648–4657.

Setiono, R., & Liu, H. (1995). Understanding neural networks via rule extraction. *14th International Joint Conference on Artificial Intelligence* (pp. 480–485).

Setiono, R., Member, S., Baesens, B., & Mues, C. (2008). Recursive Neural Network Rule Extraction for Data With Mixed Attributes. *Network*, *19*(2), 299–307.

Sternberg, R. J., & Davidson, J. E. (2005). *Conceptions of giftedness*. *Education* (p. 467). Cambridge Univ Press.

Stevens, R.H., Galloway, T. and Berka, C. (2007). EEG-Related Changes in Cognitive Workload, Engagement and Distraction as Students Acquire Problem Solving Skills, In C. Conati, K. McCoy, and G. Paliouras (Eds.): *UM 2007*. LNAI 4511, pp. 197–206. Springer-Verlag

Tan, D., & Nijholt, A. (Eds.). (2010). *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*. Springer.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. *Springer* (Vol. 26, p. 188). Springer.

Werbos, P.J. (1974). *Beyond Regression: New tools for prediction and analysis in the behavioral sciences.* PhD Thesis. Harvard University, Cambridge, MA.

Williamson, S. J., Kaufman, L., Lu, Z., Wang, J., & Karron, D. (1997). Study of human occipital alpha rhythm : the alphon hypothesis and alpha suppression. *International Journal of Psychophysiology*, *26*(1-3), 63–76.

Zhang, L. F. (2006). Thinking styles and the big five personality traits revisited. *Personality and Individual Differences* (p. 11).

Appendix A

Publications

Journal

Azcarraga, J. & Suarez, M. (2013). Recognizing Student Emotions using Brainwaves and Mouse Behavior Data. *International Journal of Distance Education Technologies*, *11*(2), 1–15.

Conferences and Workshops

Azcarraga, J., Ibanez JR, J. F., Lim, I. R., & Lumanas JR, N. (2011). Predicting Student Affect Based on Brainwaves and Mouse Behavior. *11th Philippine Computing Science Congress*. Naga City, Philippines: Computing Society of the Philippines.

Azcarraga, J., Ibanez JR, J. F., Lim, I. R., Lumanas JR, N., Trogo, R., & Suarez, M. T. (2011). Predicting Academic Emotion based on Brainwaves Signals and Mouse Click Behavior. In T. Hirashima et al. (Ed.), *19th International Conference on Computers in Education (ICCE)* (pp. 42–49). Chiang Mai, Thailand: NECTEC, Thailand.

Azcarraga, J., Inventado, P. S., & Suarez, M. T. (2010). Predicting the Difficulty Level Faced by Academic Achievers based on Brainwaves Analysis. In S. L. Wong (Ed.), *18th International Conference on Computers in Education* (pp. 107–109). Putrajaya, Malaysia: Asia-Pacific for Computers in Education.

Azcarraga, J. J., Ibanez JR, J. F., Lim, I. R., & Lumanas JR, N. (2011). Use of Personality Profile in Predicting Academic Emotion Based on Brainwaves Signals and Mouse Behavior. *2011 Third International Conference on Knowledge and Systems Engineering*, 239–244.

Azcarraga, J., & Suarez, M. (2012). Predicting Academic Emotions Based on Brainwaves, Mouse Behaviour and Personality Profile. *12th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2012): Trends in Artificial Intelligence* (pp. 728–733).

Azcarraga, J., Marcos, N. & Suarez, M. (2014). Modelling EEG Signals for the Prediction of Academic Emotions. *Utilizing EEG Input in Intelligent Tutoring Systems of the 12th International Conference on Intelligent Tutoring Systems (ITS 2014).*

Appendix B

CONSENT TO PARTICIPATE IN A RESEARCH STUDY

Your permission is being asked to allow your child/ward to participate in a research study, entitled

Analysis and Visualization of EEG and Mouse-Click Data towards Understanding the Affective Learning Behavior of Academic Achievers

Please read the information below together with your child/ward, and please do not hesitate to ask any question about anything that may not be clear to you, before deciding whether or not to allow your child/ward to participate in the study.

The study is conducted by Judith J. Azcarraga as part of her work towards a PhD in Computer Science at DLSU.

PURPOSE OF THE STUDY

The study aims to analyze and predict academic-related emotions of academic achievers based on their brainwaves signals and mouse-click behaviour while doing a specific learning task using some learning software.

BASIS FOR PARTICIPATION IN THE STUDY

Being a student of the Philippine Science High School, your child/ward has a proven record of outstanding academic achievement and is therefore a target participant (subject) of the study.

PARTICIPATION AND WITHDRAWAL

The participation of your child/ward is completely voluntary. Even if you decide to grant permission for your child/ward to participate in the study, you may still subsequently withdraw from it at any time without penalty or consequences of any kind.

EXPERIMENTAL PROCEDURE

Each session will last for about 45 minutes. If you allow your child/ward to participate in this study, this would imply the following:

1) Your child will be asked to solve mathematical problems using two different Math Tutoring Systems. The problems are of different difficulty levels to be able to induce various emotions from the participants.
2) During the experiment, an Emotiv electroencephalogram (EEG) sensor will be attached to the head of your child/ward and a camera will be used to capture his/her facial expressions.
3) Your child/ward will be asked about his/her emotions during the course of the experiment, i.e. whether he/she feels *interested*, *bored*, *confused* or *frustrated*.
4) The personality profile of your child/ward, through a test administered by the Guidance Office of PSHS, would also be used as part of the information collected from each participant.

POTENTIAL RISKS AND DISCOMFORT (IF ANY)

Emotiv EEG sensors are commercial products marketed worldwide that are typically used for modern computer games, instead of a gamepad, keyboard, or joystick. These sensors are safe to use since they just read brainwaves signals and then send these signals to a computer thru a USB connection. They are not connected to any electrical outlet while attached to the head of the user.

These sensors do not emit any radiation nor any magnetic signals that will harm the user.

These sensor devices have been used in a number of University research laboratories in the US and Japan.

In case your child finds the EEG sensor uncomfortable to use, or for any reason whatsoever that he/she would like to stop the session, the experiment will be halted immediately.

POTENTIAL SOCIAL AND PERSONAL BENEFITS

Since the study will help analyze and detect learner emotions while engaged in a learning activity based on brainwaves signals and mouse-click behavior, the results of the study would provide useful baseline data for designers of future computer-based learning systems. We expect more and more of such learning software in the future, and bright children like your child/ward would benefit the most when they engage in self-regulated learning.

In addition, your child/ward may learn something about themselves particularly in the awareness of their emotions during the learning process. For example, they would understand better the types of software features that make them learn better, and perhaps those features that tend to bore them.

CONFIDENTIALITY

Any information that is obtained in connection with this study and that can be identified to be those of your child/ward will remain confidential and will be disclosed only with your permission or as required by law.

The video recordings of your child's (ward's) facial expression during the learning session will be used solely for experimental purposes, and after the data collection is over, they will be stored in a private archive.

Portions of these video-recordings may be published and/or presented in scientific journals and/or scientific conference proceedings, but will never be published in a non-scientific venue. Further, no information, such as name, address, or other private information, will be included in these publications.

Apart from this possible usage, such data will only be viewed/used for experimental purposes. At any time during or after the experiment, you may request to review or edit the tapes and/or request that your files be destroyed.

SIGNATURE OF VOLUNTARY AND INFORMED CONSENT

Please sign below to signify that you understood the procedures described above and that you agree to allow your child/ward to participate in the study.

____ ____________________

Name of Child/Ward and Signature Date

____ ____________________

Name of Parent/Guardian and Signature Date

SIGNATURE OF INVESTIGATOR

I attest that, based on my judgment, the Parent/Guardian is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to allow his/her child/ward to participate in this research study.

I further attest that, based on my judgment, the child/ward, being the subject of the study, is voluntarily and knowingly giving informed consent to participate in this research study.

____ ____________________

JUDITH J. AZCARRAGA Date

IDENTIFICATION OF INVESTIGATOR

If you have any questions or concerns about the research, please email or contact:

JUDITH J. AZCARRAGA
Center for Empathic Human-Computer Interactions (CEHCI)
Rm. G402, Gokongwei Bldg.
De La Salle University, 2401 Taft Avenue, Manila
Email address: jay.azcarraga@delasalle.ph

This consent form is based on the research ethics protocol of DLSU and is patterned after the consent form used by the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology

October 15, 2011

INVITATION TO PARTICIPATE IN A RESEARCH STUDY

Dear ________________________:

I am Judith J. Azcarraga, a PhD student at De La Salle University. I am doing my dissertation entitled "*Analysis and Visualization of EEG and Mouse-Click Data towards Understanding the Affective Learning Behavior of Academic Achievers*" which aims to analyze and predict academic-related emotions of students who do well in school based on their brainwaves signals and mouse-click behavior while using some learning software.

You are being invited to participate in the study because as a student of Philippine Science High School, you are considered to have demonstrated outstanding academic achievement. If you agree to be in this study, you will be solving mathematical problems with the use of two different Math learning systems while an electroencephalogram (EEG) sensor capable of capturing your brainwaves signals will be attached to your head. A camera will also be used to capture your facial expressions and gestures during the experiment. During the course of the experiment, which will take about 45 minutes, you will also be asked to give your emotion, like whether you feel *interested*, *bored*, *confused* or *frustrated*. Furthermore, I will need your personality profile, possibly through a test run by the Guidance Office of PSHS.

The kind of EEG sensor to be used in the study is certainly safe and is in fact being used for modern computer games instead of a gamepad, keyboard, or joystick. I can assure you that there are no expected safety and health risks involved.

Please talk this over with your parents/guardians before you decide whether or not to participate in the study. I will also ask them to give their permission for you to take part in this study. But even if your parents or guardian say "yes", you can still decide not to participate. Being in this study is up to you and no one will be upset if you do not want to participate or even if you change your mind later and would not want to proceed with the experiment after all.

You can ask any questions that you may have about the study. If you have a question, you can text/call me thru my HP #0920-907-9830 or email me at jay.azcarraga@delasalle.ph.

Signing your name at the bottom means that you agree to be in this study. You and your parents/guardians will be given a copy of this form after you have signed it.

Very truly yours,

JUDITH J. AZCARRAGA

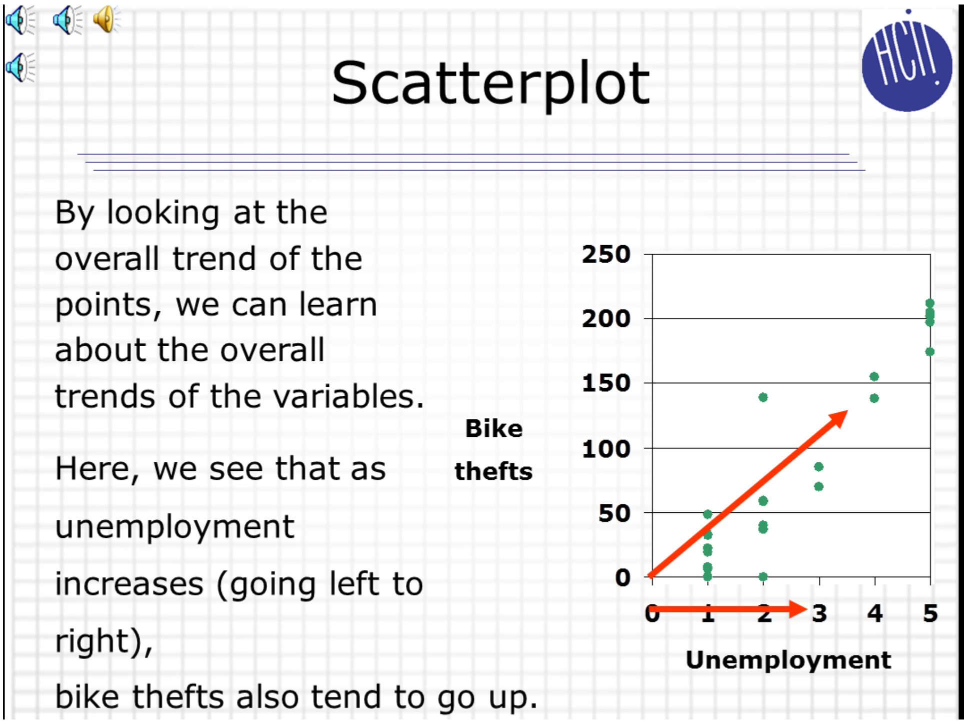
Having understood what the study is about, and having conferred with my parents/guardian about my participation in the study, I am signing this invitation to say that I am willing to participate in the study.

__________________________________ __________________

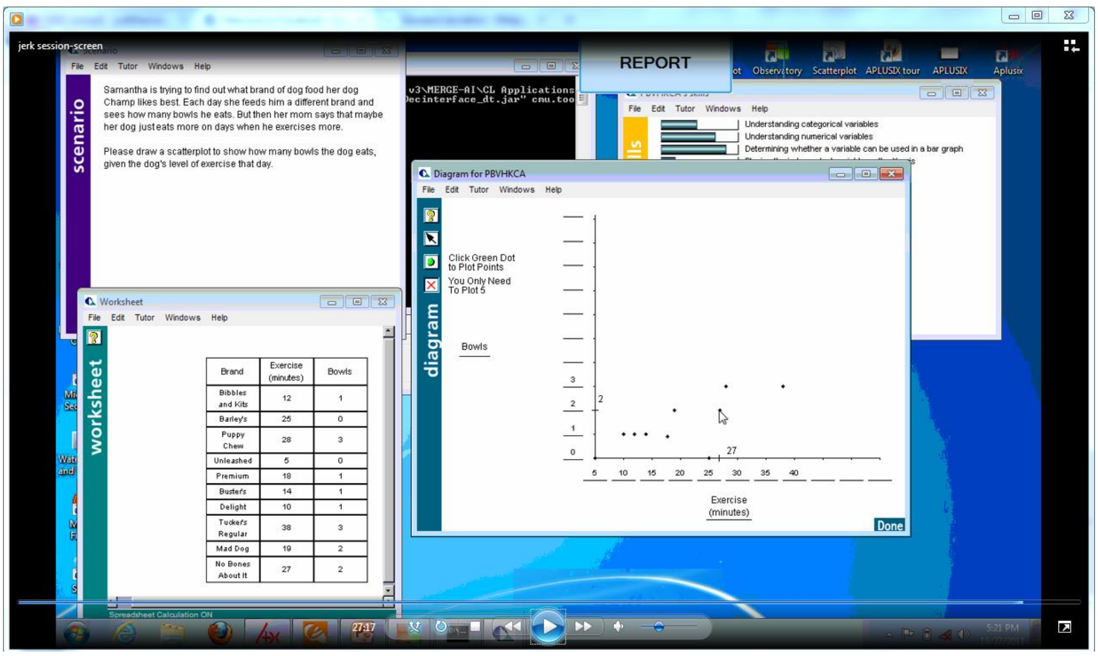
Name of Participant and Signature Date

Appendix C

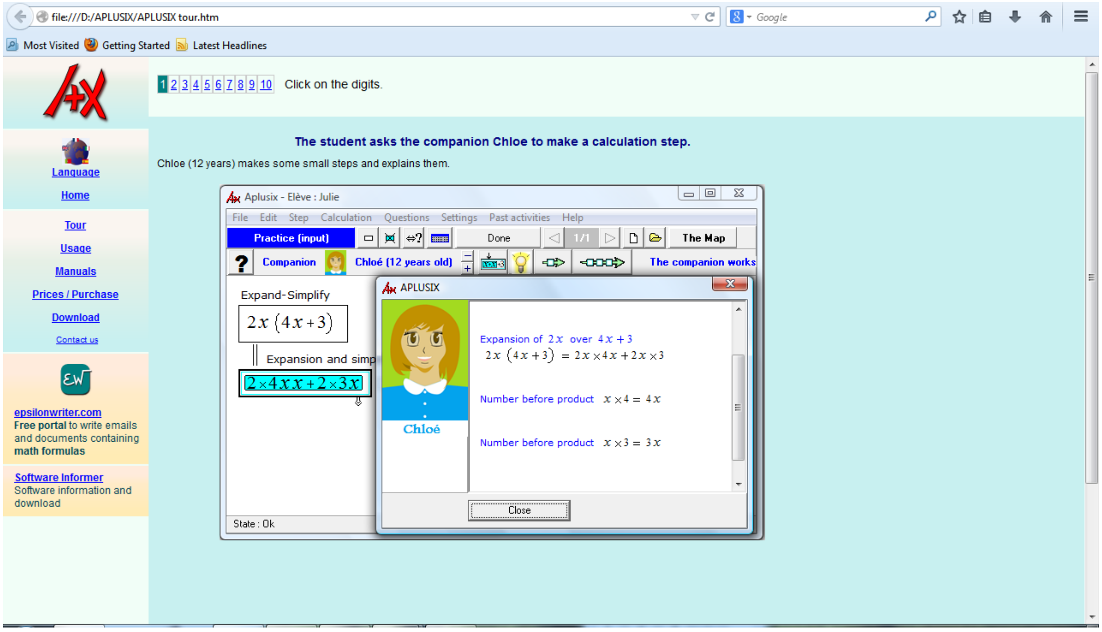
Screenshot of Scatterplot Slide Presentation



Screenshot of Scooter (Scooter) software



Screenshot of Aplusix Slide Presentation



Screenshot of Aplusix (Algebra) software