Chen, W. et al. (Eds.) (2017). Proceedings of the 25th International Conference on Computers in Education. New Zealand: Asia-Pacific Society for Computers in Education

Modeling and Prediction of Reader's Affect based on EEG Signals

Kristine KALAW[a*], Ethel ONG[a+], and Judith AZCARRAGA[a]
[a]College of Computer Studies, De La Salle University, Manila, Philippines
*kristine_kalaw@dlsu.edu.ph, [+]ethel.ong@delasalle.ph

Abstract: Readers experience various emotions while reading, which may affect their overall enjoyment and comprehension of the material. The current work presents a study on brainwaves or EEG signals and their association to emotions while a person is reading literary fiction. EEG data from 32 participants, aged 18 years old and above, were collected with the use of an EEG headset. Features were then extracted and different datasets were built according to the sex, reading preference, and reading frequency profiles. Decision Trees were used to establish baseline performance results, and these were able to classify the Hourglass of Emotion model and Emotions of Literary Response models. Support Vector Machines and Multilayer Perceptrons were trained on the same datasets to see if there is an increase in performance. Results show that they indeed yielded better performance results than Decision Trees, however, only by a small degree. Principal Component Analysis was used as an approach for feature selection, and results show comparable performance as opposed to using the base feature set of all EEG features with an averaged ±5 margin of error.

Keywords: EEG, affect recognition, machine learning

1. Introduction

Various art forms, such as literature, music and film, evoke a variety of emotions in a person. The act of reading literary fiction is a profoundly emotional experience, and is part of a broader aspect of human growth and development based on understanding one's own experiences and the social world (Freire & Slover, 1983; Mar, Oatley, Djikic, & Mullin, 2011). It has been suggested that reading improves a person's empathy and theory of mind, as well as reduce prejudice towards minority groups (Kidd & Castano, 2013; Mar, Oatley, & Peterson, 2009; Vezzali, Stathi, Giovannini, Capozza, & Trifiletti, 2015). Emotions interplay in this human growth and development through our understanding of these experiences and how we react to them (Bechara, Damasio, & Damasio, 2000; Salovey & Mayer, 1990; Schwarz, 2000).

Humans usually perceive emotions by facial or vocal expressions. However, there are other physiological cues such as knowing that a person is nervous because their hand is clammy when held (relating to skin conductance) or determining that a person is excited when their pulse is felt (relating to heart rate). Hence, people naturally decipher emotions from many physiological signals (Picard, 2000).

Recent innovations in the design of intelligent agents have attempted to make these systems adaptive by enabling them to become affect-aware. This provides a richer and more natural human-computer interaction in particular situations. In order to reach this level, the agent must first be able to detect or recognize the affective state of the person. Such detection is possible through analysis of facial and/or vocal expressions (Zeng et al., 2006; Zeng, Pantic, Roisman, & Huang, 2007), a combination of physiological data (e.g. heart rate, skin conductance, muscle tension) (Picard, Vyzas, & Healey, 2001), or even something as unusual as mouse-click behavior (Azcarraga & Suarez, 2012).

People experience a variety of emotions while being engaged in certain activities. As shown in the studies of Azcarraga & Suarez (2012), Lin et al. (2010), Nie et al. (2011), and Yazdani et al. (2012), emotions evoked while being engaged in an activity–that is, answering math problems, listening to music, watching movies, and watching music videos, respectively–could be detected via a person's brainwaves or electroencephalogram (EEG) data, which is an example of physiological data. Miall & Kuiken (1994) and Cupchik et al. (1998) are empirical works in the culture, media, and arts that have established the relation between reading and emotional response. The current work presents an EEG-based detection and recognition of emotions while a person is reading literary fiction.

2. Challenges in Building the Reader Affect Model

Collecting the data, obtaining the ground truth, and defining the appropriate emotion model are some of the challenges in building any affective model (Picard, 2003). Picard (2000) listed factors to consider in gathering *good* affect data. It was observed that Azcarraga & Suarez (2012), Lin et al. (2010), Nie et al. (2011), Yazdani et al. (2012), and the current work is *event-elicited* (emotions are elicited with a stimulus), conducted in a *lab-setting* (in an area that is not the subject's usual environment), concerned with *feeling* (the subject feels the emotions internally), is *open-recorded* (the subject knows that they are being recorded), and is *emotion-purpose* (the subject knows that the experiment is about emotions).

In collecting *good* affect data, there is the issue of determining with which sensor to obtain the data with. Sensors are typically expensive, invasive, and/or obtrusive. There can be difficulty in gathering accurate physiological data due to technical factors, such as where the sensors are applied or how much gel is used for the electrode. However, there are now advances in wearable technology that seamlessly integrate these sensors to what humans usually wear, such as the smart watch. With regard to this study, we used an Emotiv Insight EEG headset (https://www.emotiv.com/insight/), which can capture data from the *AF3*, *AF4*, *T7*, *T8*, and *Pz* channels, with the convenience of a dry sensor.

Apart from the sensors, another challenge regarding affect data is the ground truth to compare the classifications against. An observer cannot objectively say that a particular subject is feeling a certain emotion. Only the subjects can know what emotional state they are in. If the observer explicitly asks what the subjects are feeling, that may also compromise the ground truth depending on how comfortable the subjects are with expressing their feelings, how aware they are with their feelings, or if the subjects become irritated with the constant asking of how they feel.

The concern of determining the appropriate emotion model may be traced back to the domain or context of the experiment, and the unclear definition of emotions (Kleinginna Jr. & Kleinginna, 1981). For example, academic emotions pertain to the emotions experienced during learning (Pekrun, Goetz, Titz, & Perry, 2002), whereas product emotions are the emotional responses that result from the perception of consumer products (Hekkert & Desmet, 2002). Regardless of domain or context, there are two general kinds of emotion models. *Categorical models* are those that define a number of discrete basic emotions, such as the 6 basic emotions by Ekman (1972). In contrast, *dimensional models* describe the components of emotions and are often represented as a two- or three-dimensional space where the emotions are presented as points in the coordinate space of these dimensions, such as the arousal-valence scale by Russell (1980).

Given the difference in these two kinds of emotion models, Cambria et al. (2012) proposed a novel biologically inspired and psychologically motivated emotion categorization model they named as the *Hourglass of Emotion* (HoE). They described their model as one that is able to represent affective states both through its labels (categorical) and its 4 independent but co-occurring affective dimensions (dimensional)–namely, *aptitude*, *attention*, *pleasantness*, and *sensitivity*. In this way, their model can potentially describe a full range of emotional experiences. For this reason, this emotion model was chosen for this study.

Emotions are central to the experience of reading literary fiction. A person's affect and mood both *influence* and *are being influenced* before, during, and after the actual reading (Mar et al., 2011). Miall & Kuiken (2002) describe the *Emotions of Literary Response* (ELR) model, which are the emotions people feel *during* the reading process. This model describes if the emotions were caused by overall enjoyment in reading the text (*evaluative feelings*), by the events or characters in the fictional world (*narrative feelings*), or by the formal components of the text (*aesthetic feelings*). Hence, this emotion model was also chosen for this study.

3. Data Acquisition and Processing

The methodology is patterned after the aforementioned EEG-based emotion recognition studies (Azcarraga & Suarez, 2012; Lin et al., 2010; Nie et al., 2011; Yazdani et al., 2012). In this setup, the data acquisition is composed of a baseline EEG recording and presentation of the stimuli. For each stimuli, the participants were asked to annotate their emotions. Preprocessing, feature extraction and machine learning techniques were applied on the collected data.

EEG signals from 32 participants of various demographics were collected while they were reading *The Veldt* by Ray Bradbury. Prior to the experiment proper, the participants were asked to close their eyes and relax for 2 minutes. This recording serves as the baseline. Following the experimental set-up of Miall & Kuiken (1994) for the presentation of the stimuli, the story was manually divided into 72 segmentsusing phrase and sentence divisions while still retaining meaning and coherence by itself. The story segments were presented one by one via the data collector tool adapted from Azcarraga & Suarez (2012). After reading the segment, the participant would annotate what they felt and what caused it. The read-annotate process is repeated until the last segment is reached.

Because the reading activity is an experience that is never the same from one reading to the next (Tompkins, 1980), the participants could not go back and re-read the previous segments. They also have no inkling about selected reading material; thus, these collected data are the first impressions of the participants toward the short story. Note that the EEG recording while the participants are reading is called the event-related potentials (ERP).

Data preprocessing includes synchronizing the ERP recording with the emotion annotations, dividing the data into the corresponding story segments, and then further subdividing each segment into 2-second windows with 1-second overlap. The total number of these windows represents the number of instances for a particular participant.

4. Feature Extraction and Dataset Building

A moving average filter was applied to the EEG data. This smoothing technique removes the noise and allows us to see the trends in the data by showing the gradual change in the values. The *theta* (4Hz to 7Hz), *alpha* (8Hz to 12Hz), *low beta* (12Hz to 15Hz), *midrange beta* (16Hz to 20Hz), *high beta* (21Hz to 30Hz), and *gamma* (30Hz to 100Hz) frequency bands (NeuroSky, Inc., 2009) are extracted via a series of band pass filters. For each band, the signals were transformed to the frequency domain using a fast Fourier transform algorithm, and then the minimum, maximum, and mean values for each feature type were computed. The feature types are *magnitude*, which is the absolute value of the signal; *power spectral density* (PSD), which is the square of the signal in the frequency domain; and spectral power of asymmetric electrode pairs via *differential asymmetry* (DASM), or power subtraction, and *rational asymmetry* (RASM), or power division. This makes a total of 252 extracted features for each instance.

Each instance is labeled according to the self-reported emotion annotations the participants made. With regard to the HoE, the participants were asked to rate each of the 4 dimensions from -3 to +3. If the value is negative, the assigned label is *low*. If the value is positive, then the assigned label is *high*. Hence, *aptitude* pertains to whether the reader hated (low) or loved (high) what happened in the segment; *attention* pertains to the reader's level of interest to the events in the segment – that the reader is either amazed at the event that happened (low) or already expected it to happen (high); p*leasantness* pertains to the reader's amusement towards the events in the segment, on whether they were sad (low) or happy (high) about it; and s*ensitivity* pertains to whether the reader feared (low) or was angered (high) with the events in the segment. With regard to the ELR, whichever the participant chose is the assigned label. It is possible for the instance to have multiple assigned labels for ELR.

We have also identified certain *reader profiles* as a means of grouping the participants and building datasets around these groups. The profiles were based on what the participants reported after their reading session. Table 1 shows the number of participants per profile group.

Table 1: Number of participants per profile group.

Attribute	Profile Groups	# of Participants
P1: Sex	Female	23
	Male	9
P2: Reading Preference (Traditional books vs. eBooks)	RP1: Prefers reading traditional books only	8
	RP2: Prefers traditional books over eBooks	12
	RP3: Fine with both traditional books and eBooks	9
	RP4: Prefers eBooks over traditional books	3
	RP5: Prefers reading eBooks only	0
P3: Reading Frequency (How often do you read for fun?)	RF1: Almost all the time (approx. 1-2 books a week)	2
	RF2: Now and then (approx. 1-2 books a month)	14
	RF3: Not very often (approx. 1-2 books in 6 months)	14
	RF4: Never	2

The current work is concerned with the brainwaves of participants during reading time. These brainwaves (ERP) are isolated by employing baseline correction, which is simply subtracting the *i*th feature value of the baseline from all the ERP instances (Woodman, 2010). After performing baseline correction and building the datasets according to the profile groups, the datasets are standardized by computing for the z-score values. Extreme z-score values were clipped to at least -3 and at most +3.

5. Results and Discussion

The experiments conducted are binary classifications of the HoE and ELR models on different datasets and classification methods. The different variables are listed below:

- **Dataset**: Female, Male, RP1 (prefers reading traditional books only), RP2 (prefers reading traditional books over eBooks), RP3 (fine with reading on both traditional books and eBooks), RP4 (prefers reading eBooks over traditional books), RF1 (reads books for fun almost all the time), RF2 (reads books for fun every now and then), RF3 (does not read for fun very often), RF4 (never reads for fun), Sex-merged, RP-merged, RF-merged
- **Classification Method**: Decision Tree (DT), Support Vector Machine (SVM), Multilayer Perceptron (MLP)
- **Hourglass of Emotions class labels (high/low)**: Aptitude (AP), Attention (AT), Pleasantness (PL), Sensitivity (SE)
- **Emotions of Literary Response class labels (true/false)**: Aesthetic feelings (AE), Evaluative feelings (EV), Narrative feelings (NA)

Across all classification experiments, the classifiers were trained with a *participant-fold cross-validation* (can also be known as *leave-one-participant-out cross-validation*), in which one participant from the dataset serves as the test set and the rest of the participants are the training sets. The validation is repeated until all of the participants have been chosen as a test set. This type of validation scheme was employed to maintain mutually exclusive training and test sets.

5.1 Baseline Classification Performance Based on Decision Trees

The goal of this experiment was to set the baseline performance using Decision Trees and to see whether there is an improvement in performance among the profile-specific datasets. The classifiers were trained using all 252 EEG features. As shown in Figure 1 to Figure 6, it is observed that on an average basis, there is no significant improvement in the performance between the general profile dataset and the profile-specific datasets. However, it is worth noting to go over the results per classification label, as discussed in Table 2 and Table 3.

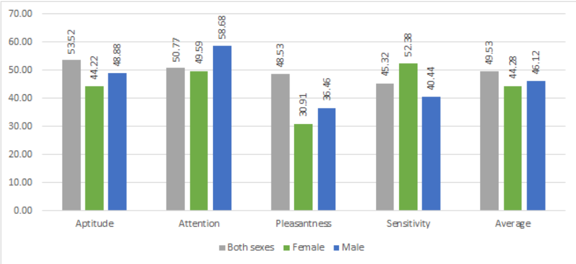

Figure 1. Comparison of DT f-measure values per HoE class label of the Sex-merged, Female, and Male datasets.

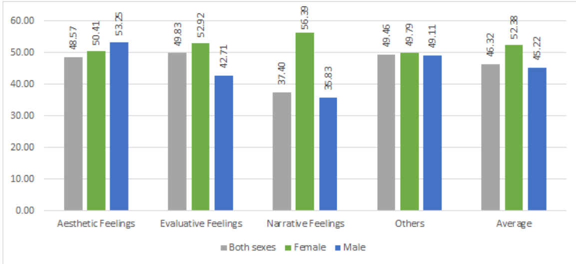

Figure 2. Comparison of DT f-measure values per ELR class label of the Sex-merged, Female, and Male datasets.

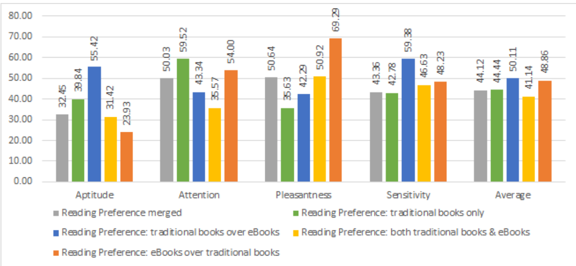

Figure 3. Comparison of DT f-measure values per HoE class label of the RP-merged, RP1, RP2, RP3 and RP4 datasets.

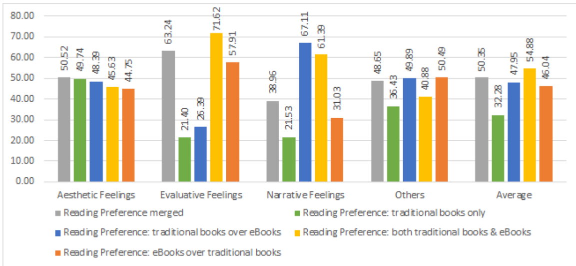

Figure 4. Comparison of DT f-measure values per ELR class label of the RP-merged, RP1, RP2, RP3 and RP4 datasets.

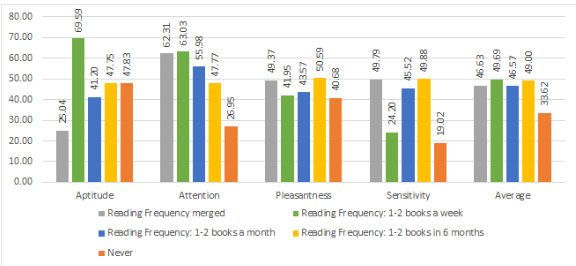

Figure 5. Comparison of DT f-measure values per HoE class label of the RF-merged, RF1, RF2, RF3 and RF4 datasets.

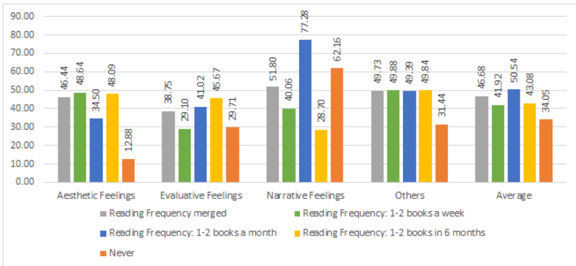

Figure 6. Comparison of DT f-measure values per ELR class label of the RF-merged, RF1, RF2, RF3 and RF4 datasets.

Table 2. Analysis on the Hourglass of Emotions DT classification results.

<table>
<tr><th></th><th>Aptitude</th><th>Attention</th><th>Pleasantness</th><th>Sensitivity</th></tr>
<tr><td>Female</td><td rowspan="2">Men and women have no distinguishing difference in loving or hating a story.</td><td></td><td></td><td>It is consistent with the women's debriefing reports wherein they expressed more anger or fear on certain elements of the story, such as the characters or the plot events, than the males – that is women are more expressive than males (Kret & De Gelder, 2012).</td></tr>
<tr><td>Male</td><td>Consistent with Kret & De Gelder (2012), in which men respond better to threatening cues than women.</td><td></td><td></td></tr>
</table>

<table>
<tr><td>RP1</td><td rowspan="2">Yielded better performance results than the general RP dataset but need further research as to why.</td><td></td><td rowspan="4">People who prefer reading eBooks over traditional books (RP4) yielded the best performance among all the PL DT classifications. Since the story segments were presented via a software program on a laptop (i.e. not a traditional book), it can be presumed that those who prefer reading traditional books have a certain bias clouding their enjoyment of the story. The said presumption is consistent because of the increase in performance of the f-measure values.</td><td></td></tr>
<tr><td>RP2</td><td></td><td>The RP2 dataset classifier may have yielded the best performance, but it may be because it was the only balanced dataset among the RP profiles.</td></tr>
<tr><td>RP3</td><td></td><td></td><td></td></tr>
<tr><td>RP4</td><td></td><td></td><td></td></tr>
<tr><td>RF1</td><td>Yielded better performance results than the general RF dataset but need further research as to why.</td><td rowspan="4">Despite the results of RF3 predicting all instances as high and RF4 predicting all instances as low, it is interesting to note the decline in the classification performances. People who always or frequently read for fun are accustomed to keeping track of the events in the story, and this accustomedness was reflected in their brainwaves.</td><td></td><td></td></tr>
<tr><td>RF2</td><td></td><td></td><td></td></tr>
<tr><td>RF3</td><td></td><td></td><td></td></tr>
<tr><td>RF4</td><td>Yielded better performance results than the general RF dataset but need further research as to why.</td><td></td><td></td></tr>
</table>

Table 3. Analysis on the Emotions of Literary Response DT classification results.

	Aesthetic Feelings	**Evaluative Feelings**	**Narrative Feelings**
Female	It can be assumed that the response or emotion evoked towards formal components of the text is the same on all the reading profiles, similar to that of the findings of Miall & Kuiken (1994). Further testing could be done to prove or disprove this assumption due to the imbalanced datasets and that these are the first impressions of	The Female dataset yielded the best performance in predicting this class label than the Male and general sex datasets. This is consistent with the notion that women are more expressive than men (Kret & De Gelder, 2012).	The Female dataset gave the best performance than the Male and general sex dataset. It is inferred that women's sympathy, empathy, or identification with the character and responses towards the plot events are registered more on their brainwaves rather than men's, a notion consistent with Kret & De Gelder (2012).

Male	the participants towards the story. Note that most of the participants reported the other two ELR class labels as the cause of their emotions during their debriefings.		
RP1		Those who prefer traditional books (RP1, RP2) have classifiers with a lower performance as opposed to those who prefer eBooks (RP3, RP4). It can be surmised that displaying the story segments via a laptop screen presents an unconscious bias in the reader.	
RP2			
RP3			Yielded better performance results than the general RP dataset but need further research as to why.
RP4			
RF1			
RF2			Yielded better performance results than the general RF dataset but need further research as to why.
RF3			
RF4			Yielded better performance results than the general RF dataset but need further research as to why.

5.2 Improving Classification Performance with Support Vector Machines or Multilayer Perceptrons

The goal of this experiment is to attempt to improve the classification performance with SVM and MLP classifiers. The same conditions in the DT experiment were subjected to the SVM and MLP training. As discussed in the previous section, there is no significant improvement in the performance between the general profile dataset and the profile-specific dataset; therefore, the general profile datasets are the ones used in this experiment. As shown in Table 4 and Table 5, it is observed that SVM or MLP yields a slightly better performance, or at least on par, in predicting class labels than DT.

5.3 Feature Selection with Principal Component Analysis

The goal of this experiment is to see whether reducing the number of features via PCA could yield a result that is higher than or at least at par with the base feature set. As shown in Table 6 and Table 7, on average, it is observed that the performance of the classifiers with PCA feature sets yield subpar results to that of its counterpart with the base set of 252 EEG features. The average difference in the f-measure value of the Base classifiers and PCA classifiers for both SVM and MLP is ±5. Note that the processing of the SVM and MLP classifiers takes much of the computer's resources. Thus, if it is acceptable to have a ±5 margin of error, then using the PCA feature set would suffice as compensation for faster processing time.

Table 4. Comparison of the DT, SVM, and MLP f-measure values of the general Sex, RP, and RF datasets for the HoE class labels.

		DT	SVM	MLP
SEX	AP	53.52	49.18	49.94
	AT	50.77	48.22	49.37
	PL	48.53	43.70	51.63
	SE	45.32	44.97	57.20
	AVG	*49.53*	*46.52*	***52.04***
RP	AP	32.45	41.18	44.52
	AT	50.03	50.76	43.45
	PL	50.64	51.58	46.27
	SE	43.36	63.85	46.47
	AVG	*44.12*	***51.84***	*45.18*
RF	AP	25.04	46.38	42.86
	AT	62.31	45.50	48.49
	PL	49.37	58.60	53.69
	SE	49.79	52.87	57.91
	AVG	*46.63*	***50.84***	*50.74*

Table 5. Comparison of the DT, SVM, and MLP f-measure values of the general Sex, RP, and RF datasets for the ELR class labels

		DT	SVM	MLP
SEX	AE	48.57	45.59	56.18
	EV	49.83	49.94	46.79
	NA	37.40	37.21	53.07
	AVG	*45.27*	*44.25*	***52.01***
RP	AE	50.52	46.11	49.36
	EV	63.24	62.92	56.40
	NA	38.96	42.65	45.46
	AVG	***50.91***	*50.56*	*50.41*
RF	AE	46.44	46.44	50.69
	EV	38.75	38.76	47.59
	NA	51.80	39.78	60.20
	AVG	*45.67*	*41.66*	***52.83***

Table 6. Comparison of SVM (Base) to SVM (PCA) and MLP (Base) to MLP (PCA) f-measure values of the general Sex, RP, and RF datasets for the HoE class labels.

		SVM (Base)	SVM (PCA)	MLP (Base)	MLP (PCA)
SEX	AP	49.18	49.02	49.94	44.88
	AT	48.22	49.24	49.37	44.21
	PL	43.70	48.32	51.63	50.20
	SE	44.97	43.44	57.20	56.13
	AVG	*46.52*	***47.51***	***52.04***	*48.86*
RP	AP	41.18	40.44	44.52	34.74
	AT	50.76	40.32	43.45	52.44
	PL	51.58	49.67	46.27	43.78
	SE	63.85	43.36	46.47	49.27
	AVG	***51.84***	*43.45*	***45.18***	*45.06*
RF	AP	46.38	56.01	42.86	45.05
	AT	45.50	49.67	48.49	46.75
	PL	58.60	58.19	53.69	48.42
	SE	52.87	51.36	57.91	42.16
	AVG	*50.84*	***53.81***	***50.74***	*45.60*

Table 7. Comparison of SVM (Base) to SVM (PCA) and MLP (Base) to MLP (PCA) f-measure values of the general Sex, RP, and RF datasets for the ELR class labels

		SVM (Base)	SVM (PCA)	MLP (Base)	MLP (PCA)
SEX	AE	45.59	45.59	56.18	46.14
	EV	49.94	46.78	46.79	49.17
	NA	37.21	37.09	53.07	53.48
	AVG	***44.25***	*43.15*	***52.01***	*49.60*
RP	AE	46.11	46.11	49.36	48.21
	EV	62.92	62.27	56.40	62.16
	NA	42.65	42.38	45.46	39.47
	AVG	***50.56***	*50.26*	***50.41***	*49.94*
RF	AE	46.44	46.44	50.69	46.43
	EV	38.76	38.76	47.59	52.74
	NA	39.78	40.30	60.20	58.90
	AVG	*41.66*	***41.83***	***52.83***	*52.69*

6. Conclusion and Future Work

EEG data from 32 participants were collected while they were reading *The Veldt*, a short story by Ray Bradbury. These EEG signals were collected with the use of an Emotiv Insight EEG headset while the participants were reading the story segments presented via the developed data collector tool. The data was processed and 252 EEG features were extracted. Different datasets were built according to sex, reading preference, and reading frequency profiles to see if whether there is a difference in classification performance per profile group. Decision Trees, Support Vector Machines, and Multilayer Perceptrons were used to classify the emotions based on the Hourglass of Emotion model, which describes what the reader is feeling, and the Emotions of Literary Response model, which describes what caused the emotion.

Prediction results show that, there is no significant improvement in the performance between the general profile datasets and its corresponding profile-specific datasets. However, in reviewing the results on a case-by-case basis, it is observed that they are consistent and in line with the debriefing interviews and other reviewed studies. SVM or MLP yielded better performance results than DT by a small degree. Classification experiments with feature selection via PCA yielded comparable performance results with an average ±5 margin of error. If such a margin is acceptable, then using the PCA feature set would suffice as compensation for faster processing time.

Certain stories can evoke only certain kinds of emotions. For example, a suspense or thriller story is geared towards the extreme levels of attention, whereas a horror story is geared towards low sensitivity. The same methodology may be repeated except that the stories presented subscribe to one of the 6 core emotional arcs (Reagan, Mitchell, Kiley, Danforth, & Dodds, 2016). Given this, a participant would read at least 6 different stories. The experiment would then become an *intra-subject* classification, instead of the *inter-subject* classification that the current work does. On the other hand, applying the current methodology with stimuli of another genre could also be done. Poems or news articles can be read in one sitting, wherein the latter could be considered as a story told in a factual narrative. These two present different writing styles, which are aligned with aesthetic feelings.

The current work makes use of the first impressions of the participants towards the story. Following what Tompkins (1980) said that reading is an experience that is never the same from one reading to the next, this could be tested by having the participant re-do the data acquisition process for a number of times. In this way, the fourth domain in the ELR model, *self-modifying feelings*, which involves the restructuring of the reader's understanding of the textual narrative, could potentially be mapped. The trajectory in the change of emotions for the same stimuli could be observed.

Other possible future work involve means of improving classification performance results, further research in the results noted in Table 2 and Table 3, and means of visualizing and showing the trajectory to discover patterns in reading behavior and preferences. For intelligent tutors and embodied conversational agents, the resulting models can be used as basis for the conversation topic with the reader, to address factors affecting one’s engagement with the reading task and comprehension of the reading materials.

References

Azcarraga, J., & Suarez, M. T. (2012). Predicting academic emotions based on brainwaves, mouse behaviour and personality profile. In P. Anthony, M. Ishizuka, & D. Lukose (Eds.), *PRICAI 2012: Trends in Artificial Intelligence* (pp. 728–733). Springer Berlin Heidelberg.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex*, *10*(3), 295–307.

Cupchik, G. C., Oatley, K., & Vorderer, P. (1998). Emotional effects of reading excerpts from short stories by James Joyce. *Poetics*, *25*(6), 363–377.

Ekman, P. (1972). Universal and cultural differences in facial expression of emotion. In *Nebraska symposium on motivation* (Vol. 19, pp. 207–284). University of Nebraska Press Lincoln.

Freire, P., & Slover, L. (1983). The importance of the act of reading. *Journal of Education*, 5–11.

Hekkert, P. P. M., & Desmet, P. M. A. (2002). The basis of product emotions. In *Pleasure with products: beyond usability*. CRC Press.

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*, *342*(6156), 377–380.

Kleinginna Jr., P. R., & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, *5*(4), 345–379.

Kret, M. E., & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia*, *50*(7), 1211–1221.

Lin, Y. P., Wang, C. H., Jung, T. P., Wu, T.-L., Jeng, S. K., Duann, J. R., & Chen, J. H. (2010). EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, *57*(7), 1798–1806.

Mar, R. A., Oatley, K., Djikic, M., & Mullin, J. (2011). Emotion and narrative fiction: Interactive influences before, during, and after reading. *Cognition & Emotion*, *25*(5), 818–833.

Mar, R. A., Oatley, K., & Peterson, J. B. (2009). Exploring the link between reading fiction and empathy: Ruling out individual differences and examining outcomes. *Communications*, *34*(4), 407–428.

Miall, D. S., & Kuiken, D. (1994). Foregrounding, defamiliarization, and affect: Response to literary stories. *Poetics*, *22*(5), 389–407.

Miall, D. S., & Kuiken, D. (2002). A feeling for fiction: Becoming what we behold. *Poetics*, *30*(4), 221–241.

NeuroSky, Inc. (2009). Brain Wave Signal (EEG) of NeuroSky, Inc. NeuroSky, Inc.

Nie, D., Wang, X. W., Shi, L. C., & Lu, B. L. (2011). EEG-based emotion recognition during watching movies. In *Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on* (pp. 667–670). IEEE.

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.

Picard, R. W. (2000). Toward computers that recognize and respond to user emotion. *IBM Systems Journal*, *39*(3.4), 705–719.

Picard, R. W. (2003). Affective computing: challenges. *International Journal of Human-Computer Studies*, *59*(1), 55–64.

Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *23*(10), 1175–1191.

Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, *5*(1), 31.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality*, *9*(3), 185–211.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition & Emotion*, *14*(4), 433–440.

Tompkins, J. P. (1980). The reader in history: The changing shape of literary response. *Reader-Response Criticism: From Formalism to Post-Structuralism*, *201*.

Vezzali, L., Stathi, S., Giovannini, D., Capozza, D., & Trifiletti, E. (2015). The greatest magic of Harry Potter: Reducing prejudice. *Journal of Applied Social Psychology*, *45*(2), 105–121.

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*, *72*(8), 2031–2046.

Yazdani, A., Lee, J.-S., Vesin, J.-M., & Ebrahimi, T. (2012). Affect Recognition Based on Physiological Changes During the Watching of Music Videos. *ACM Transactions on Interacive Intelligent Systems*, *2*(1), 7:1–7:26.

Zeng, Z., Hu, Y., Fu, Y., Huang, T. S., Roisman, G. I., & Wen, Z. (2006). Audio-visual emotion recognition in adult attachment interview. In *Proceedings of the 8th International Conference on Multimodal Interfaces* (pp. 139–145). New York, NY, USA: ACM.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2007). A survey of affect recognition methods: Audio, visual and spontaneous expressions. In *Proceedings of the 9th International Conference on Multimodal Interfaces* (pp. 126–133). New York, NY, USA: ACM.