Classification of EEG for Affect Recognition: An Adaptive Approach

Omar AlZoubi[1], Rafael A. Calvo[1], and Ronald H. Stevens[2]

[1] School of Electrical and Information Engineering
The University of Sydney, Australia
http://www.weg.ee.usyd.edu.au
[2] IMMEX Project - University of California Los Angeles

Abstract. Research on affective computing is growing rapidly and new applications are being developed more frequently. They use information about the affective/mental states of users to adapt their interfaces or add new functionalities. Face activity, voice, text physiology and other information about the user are used as input to affect recognition modules, which are built as classification algorithms. Brain EEG signals have rarely been used to build such classifiers due to the lack of a clear theoretical framework. We present here an evaluation of three different classification techniques and their adaptive variations of a 10-class emotion recognition experiment. Our results show that affect recognition from EEG signals might be possible and an adaptive algorithm improves the performance of the classification task.

Keywords: Affective computing, EEG, Classification, Adaptive.

1 Introduction

New brain imaging technologies are opening the windows to new ways of looking at emotions and other affective states (i.e. *affects*). One of the longstanding psychological debates has been between categorical and dimensional models. In the former the assumption is that a discrete number of affects (e.g. 'anger') can be recognized through behavioral (e.g. facial actions or physiological measures) [1]. The latter assumes an underlying set of variables, often two, called valence (going from very positive feelings, to very negative) and arousal (also called activation, going from states like sleepy to excited).

In the studies that use EEG (recently reviewed by Olofsson [2]), most of the focus has been on Event Related Potentials (ERPs). Signal processing [3] and classification algorithms [4] for EEG have been developed in the context of building Brain Computer Interfaces (BCI), and we are seeking ways for developing similar approaches to recognizing affective states from EEG and other physiological signals. Very few of the affect recognition studies based on physiological data use EEG, most use EKG, EMG and skin conductivity [1,5].

These studies used traditional offline classification techniques, compared the performance of different classification algorithms, and evaluated different combinations of feature sets. The ultimate aim is to find an optimal combination of

A. Nicholson and X. Li (Eds.): AI 2009, LNAI 5866, pp. 52–61, 2009.

classifiers and feature sets that could deliver an optimal performance. In addition; offline classification is also useful in evaluating subject's specific features. However, real time affect recognition systems require a real time adaptive classification system that is necessary to cope with non-stationarities of EEG and other physiological data.

Non-stationarities are ubiquitous in EEG signals [6], occurring due to many factors such as 1) user fatigue, 2) electrode drift, 3) changes in the impedance of the electrodes, 4) user cognitive states modulation, such as attention, motivation, and vigilance.

This study provides new data on EEG based affect recognition, and presents a performance comparison of K-Nearest Neighbor (KNN), Support Vector Machines (SVM), and NaiveBayes using an adaptive classification technique. Section 2 discusses some of the literature on classification of EEG signals for affect recognition, and section 3 discusses the need for real time adaptive algorithms for non stationary data. Section 4 discusses the protocol, equipment and subjects used in our data collection. Section 5 presents the performance for both static and adaptive versions of KNN, SVM, and NaiveBayes, and section 6 presents our conclusions.

2 Background

It is hard to compare the results from different studies of affect recognition systems because researchers often use different experimental setups and data preprocessing techniques. Some of these studies [7,8] used a combination of EEG and other physiological signals for this task, while others [9] used EEG solely for affect detection.

In a study to detect the level of arousal from EEG and other physiological signals, Chanel et al [7] formulated this as a classification problem with two classes corresponding to 2 or 3 degree levels of arousal. The performance of two classification methods, NaiveBayes classifier and Fisher Discriminant Analysis (FDA) were evaluated on each EEG and physiological signal separately, and on combination of both. The study used the IAPS protocol for elicitation, 4 subjects, and the EEG was recorded from 64 electrodes with a sampling rate of 1024 Hz.

The EEG was then bandpass filtered between 4-45 Hz, artifacts such as eye blinks were identified and removed from the signals. Using a 6s epoch length, the bandpower at six frequency bands were computed, yielding 6 features from the EEG. According to the authors, most of the EEG features involve the Occipital (O) lobe, since this lobe corresponds to visual cortex and subjects are stimulated with pictures. Using only EEG features and the one leave-out method, a classification accuracy of 72% for NaiveBayes was achieved and 70% for FDA for one subject. Their results suggested that EEG could be used to assess the arousal level of human affects.

In a similar study Khalili et al [8] used EEG recorded from the scalp together with other physiological signals, which was then used to assess subject's arousal and valance levels. Three classes were assessed, Calm (C), Positively Excited (PE), and Negatively Excited (NE). The stimuli to elicit the target affects were IAPS images; each stimulus consists of a block of 5 pictures which assured stability of the emotion over time. Each picture was displayed for 2.5 seconds for a total of 12.5 seconds for each block of pictures. The data was acquired from 5 subjects, with 3 sessions of 30 trials per session from each subject. EEG recorded from 64 electrodes at 1024 sampling rate.

The preprocessing and feature extraction first involved segmenting EEG data into 40s frames. EEG was bandpass filtered between 4-45 HZ, and then applying a local laplacian filter to obtain a better localization of brain activity. The study used a set of features such as, mean, STD, Skewness, Kurtosis, mean of the absolute values of the first difference of raw signals, Mean of the absolute values of the first difference of normalized signal. These six features were computed for each electrode of the 64 electrodes yielding 6*64 = 380 features. This dimension was reduced using genetic algorithms (GA), and classification using KNN and Linear Discriminant analysis (LDA) by applying a leave-one out method. The investigators achieved a classification accuracy of 40% for LDA and 51% for KNN for 3 classes. For a two classes discrimination of PE, and NE they achieved a better results of 50% for LDA and 65% for KNN. However the best classification accuracy according to the authors was achieved using EEG time-frequency features of 70% for KNN.

Horlings et al [9] used EEG alone for classifying affective states into 5 classes on two affective dimensions: valance and arousal. They used the database of the Enterface project [10], and extended it with their own data. 10 subjects were chosen for the task of EEG acquisition using a Truescan32 system; emotion elicitation performed by using the IAPS protocol. The SAM Self-Assessment was also applied where subjects rate their level of emotion on a 2D arousal and valance scale. They performed two recording sessions consisted of 25-35 trials each, with a pause of 5 minutes in between, each trial consists of 5 pictures, and each picture is shown for 2.5 seconds.

The EEG data was then filtered between 2-30 Hz to remove noise and artifacts from the signal. The baseline value was also removed from each EEG signal. Feature extraction involved computing EEG frequency bandpower, Cross-correlation between EEG bandpower, Peak frequency in alpha band and Hjorth parameters, this resulted in 114 features. The best 40 features were selected for each of the valance and arousal dimensions based on the max relevance min redundancy (mRMR) algorithm [11]. Two classifiers were trained on this feature set, one classifier for arousal dimension, and another classifier for valance dimension. According to the authors, each classifier can use different features to obtain optimal performance; using an SVM classifier with 3-fold cross validation performed the best with 32% for the valance and 37% for the arousal dimension.

Most of these studies used offline *non-adaptive* classifiers, and to our knowledge this is the first time adaptive algorithms are evaluated in this context. The next section discusses the need for classifier adaptation, especially if the data source is non-stationary in nature.

3 Online Learning and Adaptation

Most classification methods are based on the hypothesis that data comes from a stationary distribution, this is not particularly true in real life situations, where the underlying concepts of stationarity are violated, by what is known as concept drift in the data mining community [12]. This is particularly the case in EEG signals, where it always changes its nature with time. A stationary signal on the other hand maintains its statistical properties all the time, or over the observation time.

This non-stationary nature of the signals means that a classification model built earlier using a particular set of physiological data is not going to reflect the changes that have already taken place to the signals. Consequently the classification accuracy will degrade with time, unless an update to the classification model is made, or in other words the model is adapted to reflect pattern changes in physiological signals.

The non-stationarity of EEG signals can be seen as a shift in feature space as described by Shenoy et al [6]. The distinguishing patterns of interest of the physiological data are still there, what is really needed is to update or adapt the classification model in real-time to reflect the changes of data distribution. This type of change in the probability distribution of the data is also known as virtual concept drift [13], where the current model error rate is not any more acceptable given the new data distribution.

Online classifier learning and adaptation is particularly important in real time systems based on non stationary data sources in order to maintain the classification accuracy and overall performance of the system. Traditional classification systems learn inefficient models when they assume erroneously that the underlying concept is stationary while in fact it is drifting [14].

One possible solution to the problem is to repeatedly apply a traditional classifier to a fixed sliding window of examples. In this approach a similar number of examples are removed from the end of the window, and the learner is retrained, making sure the classifier is up to date with the most recent examples [15]. Other approaches apply a dynamic training window size strategy, by increasing the window size whenever the concept drift is not detected, and shrinking the window size whenever a concept drift is detected [12]. However, this is a challenging task, especially considering real time systems where memory requirements -especially if the window size is sufficiently large-, and speed/response time are issues [12]. Computationally expensive algorithms are not desired as it might slow the overall performance of the system. Other challenges may exist such as the availability of sufficient real time data as well as the lack of supervised data in actual real life applications. The next section discusses the experimental protocol used here for EEG acquisition.

4 Data and Methods

The system used in the recording was a wireless sensor headset developed by Advanced Brain Monitoring, Inc (Carlsbad, CA). It utilizes an integrated hardware and software solution for acquisition and real-time analysis of the EEG, and it has demonstrated feasibility for acquiring high quality EEG in real-world environments including workplace, classroom and military operational settings. It includes an easily-applied wireless EEG system that includes intelligent software designed to identify and eliminate multiple sources of biological and environmental contamination.

Data was recorded at 256 Hz sampling rate from multiple EEG scalp bi-polar sensor sites: F3-F4, C3-C4, Cz-PO, F3-Cz, Fz-C3, Fz-PO. Bi-polar recordings were selected in order to reduce the potential for movement artifacts that can be problematic for applications that require ambulatory conditions in operational environments. Limiting the sensors (seven) and channels (six) ensures the sensor headset can be applied within 10 minutes, making the tool more feasible in practical scenarios. Further exploratory studies should probably be performed with equipment that allows high density EEG.

Three subjects were asked to self-elicit a sequence of emotions and where recommended to use recollections of real-life incidents. Numerous studies support the notion that this can serve as a sufficient condition for emotion elicitation [16]. Each emotion trial lasted for 3 minutes with a 1 minute rest in between. The power spectral density (PSD) values in each of the 1-Hz bins (from 1 Hz 40 Hz) were calculated from each 1 second epoch. The first and second frequency bins are not considered since they are mostly contaminated by EEG artifacts, which mostly occur at low frequencies.

The end dataset therefore will have (38 frequency bins * 6 EEG Channels) 228 features, and (180 rows * 10 emotions) 1800 instances for each of the three subjects. Based on these datasets, a number of classification techniques were compared, together with an online simulation experiment that incorporated an adaptive classification technique. The next section discusses the classifiers and the results of the two experiments.

5 Results and Discussion

5.1 Offline Analysis

The offline analysis was done using Weka [17], Table 1 lists the classifiers used and their description. The performance of the three classifiers compared in Table 2 are based on 10-fold cross validation. All classifiers are set to their default parameter values as implemented in Weka. The ZeroR classifier represents the baseline accuracy that the 10 affects studied which was 10% (the difference is based on some session being less than the default 3 minutes). The best classification accuracy was achieved using a KNN classifier with k=3 and Euclidian distance measure, and this was nearly uniform across all subjects. An SVM classifier with a linear kernel, which is based on John Platt's sequential minimal optimization algorithm for training a support vector machines classifier [18] was less accurate than KNN; however its performance was comparably better than that of the NaiveBayes classifier. An explanation for the performance of KNN comes from the work done by Cieslak et al [19], where they found that KNN is less sensitive

Table 1. A description of the classifiers used in this study and their parameters

Classifier	Description and Parameters
ZeroR (baseline)	Predicts the majority class in the training data; used as a baseline.
NaiveBayes	A standard probabilistic classifier, the classifier assigns a pattern to the class that has the maximum estimated posterior probability.
KNN	A classical instance-based algorithm; uses normalized Euclidean distance with k=3. KNN assigns the class label by majority voting among nearest neighbors.
SVM	It combines a maximal margin strategy with a kernel method to find an optimal boundary in the feature space, this process is called a kernel machine. The machine is trained according to the structural risk minimization (SRM) criterion [20]. We used Weka's [17] SMO with linear kernel for the offline analysis. The online analysis used a SVM with linear kernel as implemented in PRTools 4.0. [21] Default parameters are used for both methods.

Table 2. Classification accuracy of EEG data using 10-fold cross validation for three subjects A,B,C

Classifier/Subject	A	B	C
ZeroR (baseline)	9.96%	9.93%	9.94%
NaiveBayes	42.83%	28.16%	33.48%
KNN(3)	**66.74%**	39.97%	57.73%
SVM	54.57%	40.80%	33.48%

to non-stationarities than SVM and NaiveBayes. Subject A data showed good separation tendency across all classification methods compared to the other two subjects B,C. The classification performance on subject C data achieved the second best classification accuracy across classifiers except in the case of SVM where subject B data achieved a 40.8% performance compared to 33.48% for subject C. These results suggest that accuracy can change considerably between subjects.

5.2 Online Simulation

This experiment involved comparing the performance of a basic adaptive algorithm [15] in combination with a KNN classifier with k=3 and Euclidian distance measure, SVM with a linear Kernel, and NaiveBayes classifiers as implemented in PRTools 4.0 [21], the classifiers were used with their default parameters. A description of the algorithm is listed in Table 3.

The algorithm was applied with three different training window sizes to compare the effect of window size on classification performance. The static versions

Table 3. Listing of the adaptive algorithm

1. *Choose an initial fixed training window size*
2. *Train a classifier w on the examples of the training window*
3. *On the arrival of new examples, update the training window by:*
   a. *Inserting the new examples into the training window*
   b. *Deleting an equal number of examples from the end of the training window.*
4. *Train the classifier on the new training window*

Table 4. Average error rate and standard deviation for the different classifiers, static and adaptive classifiers over the training sessions, with different window size

Method	**Static**		**Adaptive**	
Classifier/window size	**AvgErrorRate**	**STD**	**AvgErrorRate**	**STD**
Knn/250	0.710	0.140	**0.207**	**0.134**
Knn/450	0.714	0.143	0.247	.145
Knn/900	0.662	0.158	0.288	0.155
NaiveBayes/250	0.694	0.132	0.464	0.153
NaiveBayes/450	0.660	0.124	0.492	0.141
NaiveBayes/900	0.616	0.131	0.507	0.142
SVM/250	0.716	0.129	0.437	0.147
SVM/450	0.704	0.138	0.493	0.159
SVM/900	0.707	0.144	0.542	0.156

of the classifiers were also evaluated with the same window sizes. Static classifiers are those which initially trained on the first training window of examples, but are not updated later on. Training window sizes of 250, 450, and 900 were chosen, which account for 0.15, 0.25, and 0.5 of total dataset size. The training window was updated every 10 examples as it would be inefficient to update the training window and retrain the classifiers on the arrival of every example; 10 is also the number of classes in our dataset. The experiment was done on one subject's data (subject A), and is meant as a demonstration for the need for an adaptive classification technique for real time affect recognition systems, where the physiological data are continuously changing its behavior with time.

Table 4 shows the average error rate and standard deviation over the training sessions of both static and adaptive classifiers, with different window sizes 250, 450, and 900. It can be seen that the adaptive KNN classifier with a window size of 250 samples has the lowest average error rate overall, and the lowest standard deviation among the adaptive classifiers which indicates that the classifier maintained a subtle performance over the training sessions. This can be also inferred from Figure 1 which shows the performance of the KNN classifier with a window size of 250; clearly the adaptive version of the classifier outperforms the static one by nearly 50%. KNN proves to outperform SVM and NaiveBayes with non-stationarity data, and this comes from the way KNN works by voting amongst nearest examples.

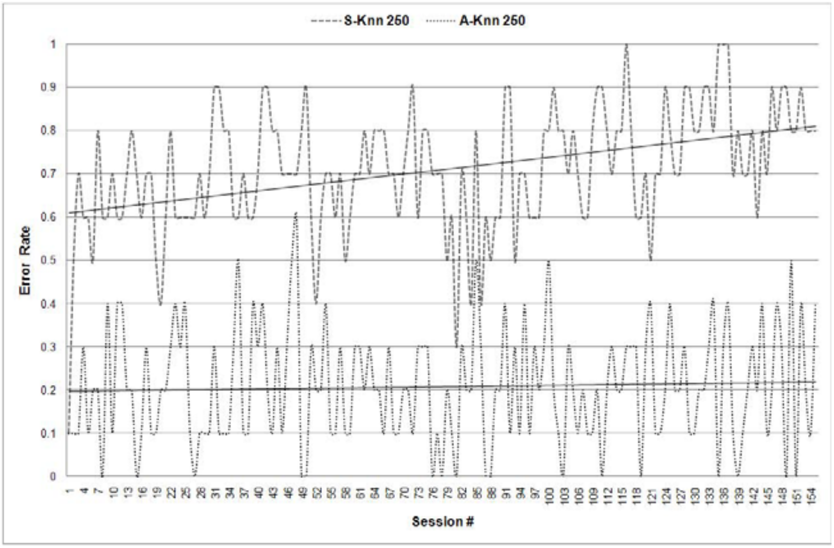

Fig. 1. Adaptive vs. static KNN classifier with a window size of 250 examples, the two solid lines in the middle show linear trend lines

The effect of window size on classifier performance can be inferred from Table 4, adaptive classifiers performance relatively enhanced with a smaller window size. An explanation for this comes particularly from the nature of the non-stationarity data; the smaller the window size, the more is the chance to build a model that can best classify unforeseen examples that are close enough in time, and get more localized information in time from the data, given that the data changes its behavior with time. On the other hand the performance of the adaptive classifiers is degraded with a larger window size, and this is due to the non-stationarity problem mentioned earlier, training the classifiers on a larger window size fails to build an efficient model for the fast changing data.

The average static classification performance was relatively improved with a larger window size, which was not surprising, given the dataset size, and this shouldn't be confused with the earlier discussion as the training and testing was done at different windows in time than the adaptive versions. However, a closer examination of Figure 1 shows the upward trend of the static classifier. That is, as time goes on the error rate goes upwards as well, and the classification performance degrades with time.

It is worth mentioning that the training time for each classifier varied greatly, while NaiveBayes, and KNN training time were relatively small especially if the window size is small, the training time for SVM was considerably higher since the classification task was a multiclass problem. This should be taken in consideration if it is going to affect the response time of the affect recognition system. On the other hand, if the window size is large, the memory requirements for KNN for example becomes larger, since it needs to store its distance matrix in memory, and classify instances as they arrive to their nearest neighbors; these are some of the design considerations that require attention.

6 Conclusions and Future Work

Despite the lack of strong neuroscientific evidence for correlates of brain activity at the cortical level with affective events, our recordings indicate that affect recognition from EEG might be possible. Rather this study did not focus on the neuroscience behind affects so we do not intend to speculate about its implications. Rather the study focused on the automatic classification techniques that could be used for EEG data, and they showed that accuracies well above the baseline are possible. We also evaluated an adaptive version of the algorithms showing that the error rate for the static versions of each algorithm was higher than that of the adaptive version. Future work would look at using a dynamic appraoch for updating the training window size.

Despite the experimental protocol we used is common in the literature, the analysis of the confusion matrix produced by most of the classification algorithms studied showed that fixing the order of the sequence of affects elicitation might be having an effect on their accuracy. Future work should consider using counterbalanced order for the affects elicited, these type of methodological issues can only be solved in larger studies.

References

1. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence: analysis of affective physiological state. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(10), 1175–1191 (2001)
2. Olofsson, J.K., Nordin, S., Sequeira, H., Polich, J.: Affective picture processing: an integrative review of erp findings. Biol. Psychol. 77(3), 247–265 (2008)
3. Bashashati, A., Fatourechi, M., Ward, R.K., Birch, G.E.: A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. J. Neural Eng. 4(2), R32–R57 (2007)
4. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B.: A review of classification algorithms for eeg-based brain-computer interfaces. Journal of Neural Engineering 4(2007) (2007)
5. Calvo, R.A., Brown, I., Scheding, S.: Effect of experimental factors on the recognition of affective mental states through physiological measures. In: Nicholson, A., Li, X. (eds.) AI 2009. LNCS (LNAI), vol. 5866, pp. 61–70. Springer, Heidelberg (2009)
6. Shenoy, P., Krauledat, M., Blankertz, B., Rao, R., Müller, K.: Towards adaptive classification for bci. Journal of Neural Engineering 3(1) (2006)
7. Chanel, G., Kronegg, J., Grandjean, D., Pun, T.: Emotion assessment: Arousal evaluation using eeg's and peripheral physiological signals. In: Gunsel, B., Jain, A.K., Tekalp, A.M., Sankur, B. (eds.) MRCS 2006. LNCS, vol. 4105, pp. 530–537. Springer, Heidelberg (2006)
8. Khalili, Z., Moradi, M.H.: Emotion detection using brain and peripheral signals. In: Cairo International on Biomedical Engineering Conference, CIBEC 2008, pp. 1–4 (2008)

9. Horlings, R., Datcu, D., Rothkrantz, L.: Emotion recognition using brain activity. In: Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing. ACM, New York (2008)
10. Savran, A., Ciftci, K., Chanel, G., Mota, J., Viet, L., Sankur, B., Akarun, L., Caplier, A., Rombaut, M.: Emotion detection in the loop from brain signals and facial images (2006)
11. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 1226–1238 (2005)
12. Last, M.: Online classification of nonstationary data streams. Intell. Data Anal. 6(2), 129–147 (2002)
13. Tsymbal, A.: The problem of concept drift: Definitions and related work. Technical report, Department of Computer Science, Trinity College (2004)
14. Hulten, G., Spencer, L., Domingos, P.: Mining time-changing data streams. In: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, California, pp. 97–106. ACM, New York (2001)
15. Widmer, G., Kubat, M.: Learning in the presence of concept drift and hidden contexts. Mach. Learn. 23(1), 69–101 (1996)
16. Coan, J.A., Allen, J.J.B.: Handbook of emotion elicitation and assessment. Oxford University Press, Oxford (2007)
17. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, San Francisco (2005)
18. Platt, J.: Machines using sequential minimal optimization. In: Schoelkopf, B., Burges, C., Smola, A. (eds.) Advances in Kernel Methods - Support Vector Learning (1998)
19. Cieslak, D., Chawla, N.: A Framework for Monitoring Classifiers' Performance: When and Why Failure Occurs? Knowledge and Information Systems 18(1), 83–109 (2009)
20. Vapnik, V.: Statistical Learning Theory. John Wiley and Sons, Inc., New York (1998)
21. Heijden, F., Duin, R., de Ridder, D., Tax, D.: Classification, parameter estimation and state estimation. John Wiley & Sons, Chichester (2004)