Multi-modal Biometric Emotion Recognition Using Classifier Ensembles

Ludmila I. Kuncheva[1], Thomas Christy[1],
Iestyn Pierce[2], and Sa'ad P. Mansoor[1]

[1] School of Computer Science, Bangor University, LL57 1UT, UK
[2] School of Electronic Engineering, Bangor University, LL57 1UT, UK

Abstract. We introduce a system called AMBER (Advanced Multi-modal Biometric Emotion Recognition), which combines Electroencephalography (EEG) with Electro Dermal Activity (EDA) and pulse sensors to provide low cost, portable real-time emotion recognition. A single-subject pilot experiment was carried out to evaluate the ability of the system to distinguish between positive and negative states of mind provoked by audio stimuli. Eight single classifiers and six ensemble classifiers were compared using Weka. All ensemble classifiers outperformed the single classifiers, with Bagging, Rotation Forest and Random Subspace showing the highest overall accuracy.

1 Introduction

Affective computing covers the area of computing that relates to, arises from, or influences emotions [14]. Its application scope stretches from human-computer interaction for the creative industries sector to social networking and ubiquitous health care [13]. Real-time emotion recognition is expected to greatly advance and change the landscape of affective computing [15]. Brain-Computer Interface (BCI) is a rapidly expanding area, offering new, inexpensive, portable and accurate technologies to neuroscience [21]. However, measuring and recognising emotion as a brain pattern or detecting emotion from changes in physiological and behavioural parameters is still a major challenge.

Emotion is believed to be initiated within the limbic system, which lies deep inside the brain. Hardoon et al. [4] found that the brain patterns corresponding to basic positive and negative emotions are complex and spatially scattered. This suggests that in order to classify emotions accurately, the whole brain must be analysed.

Functional Magnetic Resonance Imaging (fMRI) and Electro Encephalography (EEG) have been the two most important driving technologies in modern neuroscience. No individual technique for measuring brain activity is perfect. fMRI has the spatial resolution needed for emotion recognition while EEG does not. fMRI, however, offers little scope for a low-cost, real-time, portable emotion classification system. In spite of the reservations, EEG has been applied for classification of emotions [1, 5, 19, 20]. Bos [1] argues that the projections of positive and negative emotions in the left and right frontal lobes of the brain make

K.G. Mehrotra et al. (Eds.): IEA/AIE 2011, Part I, LNAI 6703, pp. 317–326, 2011.

these two emotions distinguishable by EEG. He also warns that the granularity of the information collected from these regions through EEG may be insufficient for detecting more complex emotions. Different success rates of emotion recognition through EEG have been reported in the recent literature ranging from moderate [2] to excellent accuracy [10, 13]. The reason for the inconclusiveness of the results can be explained with the different experimental set-ups, different ways of eliciting and measuring emotion response, and the type and number of distinct emotions being recognised.

Chanel et al. [2] note that, until recently, there has been a lack of studies on combination of biometric modalities for recognising affective states. Some physiological signals can be used since they come as spontaneous reactions to emotions. Among other affective states, stress detection gives a perfect example of the importance of additional biometric modalities. It is known that stress induces physiological responses such as increased heart rate, rapid breathing, increased sweating, cool skin, tense muscles, etc. This gives stress detection systems good chances of success [9]. Considerable effort has been invested in designing low-power and high-performance readout circuits for the acquisition of biopotential signals such as EEG/EMG electrodes [16, 24, 25], skin conductance sensors [12], temperature sensors and muscle tightness gauges. Finally, combination of EEG and other biometric modalities has proved to be a successful route for affective state recognition [1, 2, 10, 20].

Here we present a system for multi-modal biometric emotion recognition (AMBER) consisting of a single-electrode headset, an EDA sensor and a pulse reader. These modalities were selected due to their low cost, commercial availability and simple design. We evaluate state-of-the art classifiers, including classifier ensembles, on data collected from the system. The goal is to assess the ability of the classification methodologies to recognise emotion from signals spanning several seconds.

2 Component Devices of AMBER

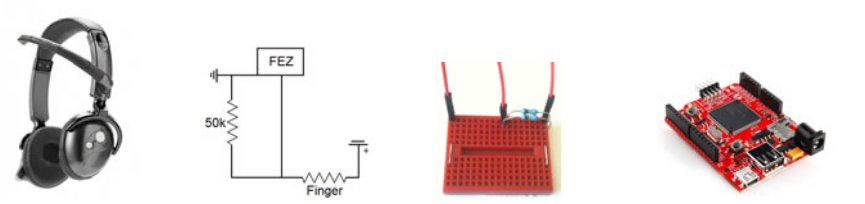
2.1 EEG Headset

We used a headset with single EEG electrode placed on the left of the forehead (The NeuroSky Mindset[1], Figure 1(a)). Mindset is a typical commercial EEG-headset of relatively low cost and good speed, suitable for real-time signal acquisition. It connects to a computer via a Bluetooth adapter, configured as a serial port. The data is received in variable sized packets and has to be reconstructed into readable form by a packet parser. A packet parser was written in Matlab to read and check the accuracy of the transmitted data.

2.2 EDA Sensor

Electro-Dermal Activity (EDA), also known as Galvanic Skin Response (GSR), is the measure of electrical resistance between two points across the skin. In its

[1] `http:\www.neurosky.com`



(a) "Mindset" (b) EDA Circuit (c) Open wire EDA sensor (d) Analogue-to-digital converter FEZ Domino

Fig. 1. Components of AMBER

most basic form, human skin is used as an electrical resistor whose value changes when a small quantity of sweat is secreted. Figure 1(b) depicts the circuit, and 1(c) shows the electronic breadboard used in AMBER.

To feed the signal into the system we used a low-cost analogue to digital converter, FEZ Domino, shown in Figure 1(d). The FEZ Domino enables electrical and digital data to be controlled using the .NET programming language. The digital output was transmitted to a computer using a TTL Serial to USB converter cable.

2.3 Pulse Reader

Pulse sensors can determine levels of anxiety and stress, thereby contributing to the recognition of emotion. A commercially available pulse monitor kit was used for AMBER. The monitor uses a phototransistor to detect variances in blood flowing through a finger. An infra-red light is emitted through a finger and the level of light able to pass through to the other side is detected by the phototransistor. The signal is fed to the FEZ Domino and further transmitted to the computer.

The pulse sensor was attached to the middle finger on the right hand. The EDA sensor was connected to the index finger and the ring finger of the left hand. The EEG was positioned as recommended by NeuroSky. The sampling rate of all three input devices of AMBER was set at 330 readings per second.

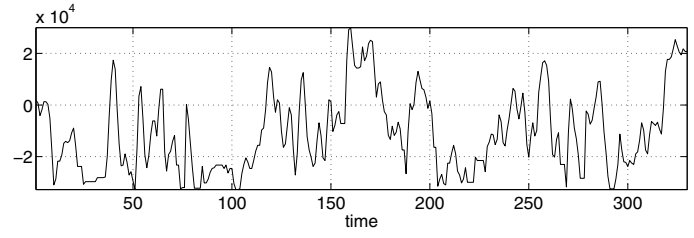
3 Data

3.1 The Data Collection Experiment

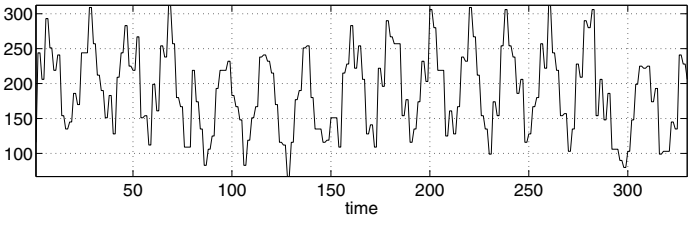
The experiment involved presenting auditory stimuli to the subject in twenty 60-second runs. The stimuli were selected so as to provoke states of relaxation (positive emotion) or irritation (negative emotion). The positive audio stimuli were taken from an Apple iPhone application called Sleep Machine. The composition was a combination of wind, sea waves and sounds referred to as Reflection (a mixture of slow violins tinkling bells and oboes); this combination was considered by the subject to be the most relaxing. The negative audio stimuli were

musical tracks taken from pop music, which the subject strongly disliked. The three biometric signals were recorded for 60 seconds for each of the 20 runs: 10 with the positive stimuli and 10 with the negative stimuli.

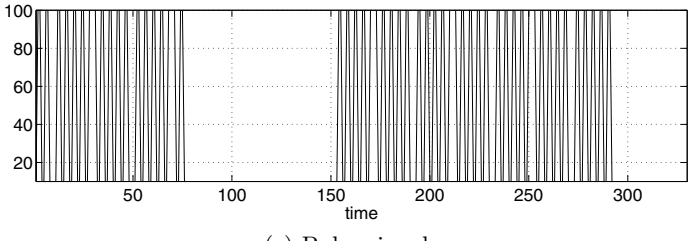
Typical examples of one-second run of the three signals is shown in Figure 2.



(a) EEG signal



(b) EDA signal



(c) Pulse signal

Fig. 2. Typical 1-second runs of the three input signals

3.2 Feature Extraction and Collating the Data Sets

Eight data sets were prepared by cutting the 60-second runs into sections of {3, 4, 5, 6, 10, 12, 15, and 20} seconds respectively. The sections were pooled to form a data set. All sections from a positive run were labelled as positive, and those from the negative runs were labelled as negative. For example, for the 5-second sections, there were $10 \times 60/5 = 120$ positive examples and 120 negative examples.

Ten features were extracted from each section. The power spectrum of the EEG signal was computed using the Welch method, and cut into 8 bands: delta

(1-3Hz), theta (4-7Hz), alpha 1 (8-9Hz), alpha 2 (10-12Hz), beta 1 (13-17Hz), beta 2 (18-30Hz), gamma 1 (31-40Hz) and gamma 2 (41-50Hz). The first 8 features for a particular section were the mean power within the respective frequency bands. Figure 3 shows the frequency powers for the 8 bands and the two classes, averaged across all examples from the respective class. The 4-second data set was used for this illustration.[2] The axes are omitted for clarity of the plot and error bars of the 95% confidence intervals are displayed. Significant differences between the curves for the two classes are observed in bands δ, γ_1 and γ_2.

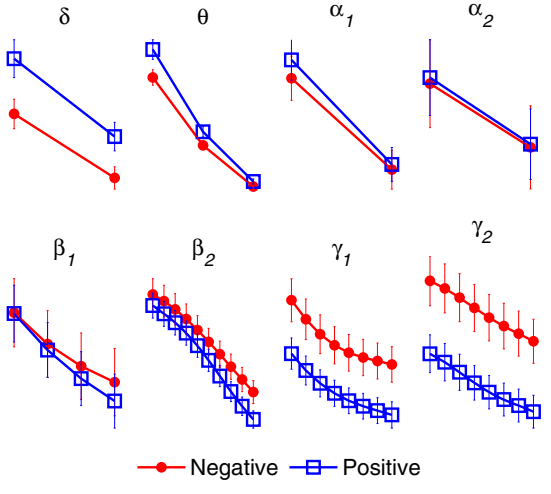


Fig. 3. Frequency powers for the 8 bands and the two classes

The remaining two features for the sections were the mean EDA signal and the mean pulse signal.

4 Classification Methods

The most widely used classification method in neuroscience analyses is the Support Vector Machine classifier (SVM) [3, 11, 18, 22]. Our previous research confirmed the usefulness of SVM but also highlighted the advantages of multiple classifier systems (classifier ensembles) [6, 7, 8].

All experiments were run within Weka [23] with the default parameter settings. The individual classifiers and the classifier ensemble methods chosen for this study are shown in Table 1.[3] Ten-fold cross-validation was used to estimate

[2] The curves for the remaining 7 data sets were a close match.

[3] We assume that the reader is familiar with the basic classifiers and ensemble methods. Further details and references can be found within the Weka software environment at `http://www.cs.waikato.ac.nz/ml/weka/`

Table 1. Classifiers and classifier ensembles used with the AMBER data

	Single classifiers
1nn	Nearest neighbour
DT	Decision tree
RT	Random tree
NB	Naive Bayes
LOG	Logistic classifier
MLP	Multi-layer perceptron
SVM-L	Support vector machines withl inear kernel
SVM-R	Support vector machines with Radial basis function (RBF) kernel

	Ensembles
BAG	Bagging
RAF	Random Forest
ADA	AdaBoost.M1
LB	LogitBoost
RS	Random Subspace
ROF	Rotation Forest

the classification accuracy of the methods. All ensembles consisted of 10 single classifiers (the default value in Weka).

Since Rotation Forest (ROF) is a relatively recent ensemble method [17], we give a brief description here. ROF builds classifier ensembles using independently trained decision trees. Each tree uses a custom set of extracted features created in the following way. The original feature set is split randomly into K subsets (the default value in Weka is $K = 3$), principal component analysis (PCA) is run separately on each subset, and a new set of n linear extracted features is constructed by pooling all principal components. Different splits of the feature set will lead to different extracted features, thereby contributing to the diversity introduced by the bootstrap sampling. The average combination method is applied on the (continuous-valued) votes of the classifiers.

5 Results

Table 2 shows the correct classification (in %) for all methods and data sets. The highest accuracy for each data set is highlighted as a frame box, and the second highest is underlined. All highest accuracies are achieved by the ensemble methods. The individual classifiers reach only one of the second highest accuracies while the ensemble methods hold the remaining 7 second highest scores. This result confirms the advantage of using the classifier ensembles compared to using single classifiers, even the current favourite SVM. In fact, SVM-R was outperformed by all classifiers and ensembles, and SVM-L managed to beat only the logistic classifier. A series of pilot experiments revealed that none of the modalities alone were as accurate as the combination.

Table 2. Classification accuracy from the 10-fold cross-validation

Method	Data sets and number of instances							
	3s	4s	5s	6s	10s	12s	15s	20s
	400	300	240	200	120	100	80	60
1nn	62.84	64.89	63.44	62.04	61.11	60.87	56.48	59.93
DT	64.16	58.57	67.37	65.92	58.49	62.78	69.96	58.93
RT	60.02	63.02	61.9	62.63	57.11	66.66	66.75	57.10
NB	64.69	63.81	64.45	64.48	65.02	67.82	65.43	61.07
LOG	62.04	60.37	62.59	63.27	59.26	59.16	57.59	57.53
MLP	62.46	59.37	63.28	63.36	63.43	64.22	57.05	58.47
SVM-L	62.09	61.41	63.52	62.38	62.32	59.13	58.70	56.83
SVM-R	50.81	51.16	50.56	50.52	50.18	51.19	51.66	51.33
BAG	65.56	65.62	68.25	67.09	67.37	68.79	66.46	64.37
RAF	64.51	64.65	66.08	65.27	65.86	69.58	67.29	61.57
ADA	63.41	62.21	70.00	67.59	61.07	66.28	73.80	63.30
LB	65.34	62.92	68.78	68.05	62.04	64.02	68.27	60.70
RS	64.96	64.78	66.25	68.21	64.61	67.43	68.95	61.77
ROF	66.90	65.41	66.86	67.23	67.36	69.30	65.46	62.27

To visualise the results, Figure 4 shows the 14 ensemble methods sorted by their overall ranks. Each method receives a rank for each data set. As 14 methods are compared, the method with the highest classification accuracy receives rank 1, the second best receives rank 2 and so on. If the accuracies tie, the ranks are shared so that the total sum is constant ($1 + 2 + 3 + \cdots + 14 = 105$). The total rank of a method is calculated as the mean across all 8 data sets. The total ranks and the mean accuracies of the 14 classification methods are shown in the two columns to the right of the colour matrix in Figure 4.

The colour matrix represents the classification accuracies of the methods sorted by total rank. Warm colours (brown, red and yellow) correspond to higher accuracy while cold colours (green and blue) correspond to lower accuracy. The figure reveals several interesting patterns in addition to the already discussed superior accuracy of ensembles over individual classifiers. First, the cooler colours in the last column (20s data set) indicate relatively low accuracy compared to the middle columns. This seems counter-intuitive because the frequency spectrum and the EDA and pulse averages are calculated from larger bouts of the signal, and should be less noisy. The reason for this anomaly is most likely the smaller number of data points to train the classification methods. Note that the cooler colours in the first couple of columns is not unexpected. Three- and four-second sections may be insufficient to noisy estimates of the features, hence the lower accuracy. Second, the mixture of colours in the row corresponding to

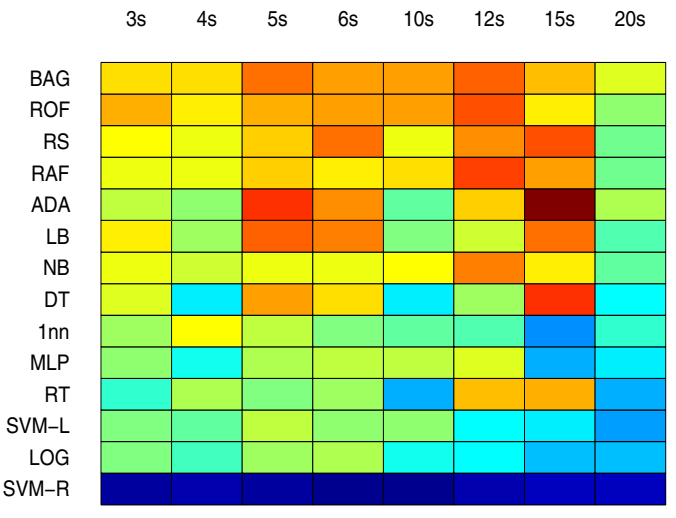


Fig. 4. Colour matrix for the classification methods sorted by their average ranks. Brown/red correspond to high accuracy; green/blue correspond to low accuracy.

AdaBoost (ADA) supports the finding elsewhere that AdaBoost's performance can vary considerably for noisy data. This row also contains the single dark brown cell corresponding to the highest accuracy of 73.8% achieved in the whole experiment.

6 Conclusion

This paper presents a case study of affective data classification coming from three biometric modalities: EEG electrode, electrodermal sensor (EDA) and pulse reader, embedded in a system called AMBER. The results indicate that positive and negative emotional states evoked by audio stimuli can be detected with good accuracy from a time segment spanning a few seconds. This work serves as a first step in a developing an inexpensive and accurate real-time emotion recognition system. Improvements on the hardware and the preprocessing of the signals are considered. We are currently working towards preparing an experimental protocol and the supporting software for gathering data from AMBER on a large scale. The new protocol will be based on a combination of visual, audio and computer-game type of stimuli.

References

1. Boss, D.O.: EEG-based emotion recognition (2006), http://emi.uwi.utwente.nl/verslagen/capita-selecta/CS-oude=Bos-Danny
2. 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)
3. De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebela, R., Formisano, E.: Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. NeuroImage 43(1), 44–58 (2008)
4. Hardoon, D.R., Mourao-Miranda, J., Brammer, M., Shawe-Taylor, J.: Unsupervised analysis of fMRI data using kernel canonical correlation. NeuroImage 37(4), 1250–1259 (2007)
5. Ko, K., Yang, H., Sim, K.: Emotion recognition using EEG signals with relative power values and Bayesian network. International Journal of Control, Automation, and Systems 7, 865–870 (2009)
6. Kuncheva, L.I.: Combining Pattern Classifiers. Methods and Algorithms. John Wiley and Sons, N.Y (2004)
7. Kuncheva, L.I., Rodríguez, J.J.: Classifier ensembles for fMRI data analysis: An experiment. Magnetic Resonance Imaging 28(4), 583–593 (2010)
8. Kuncheva, L.I., Rodriguez, J.J., Plumpton, C.O., Linden, D.E.J., Johnston, S.J.: Random subspace ensembles for fMRI classification. IEEE Transactions on Medical Imaging 29(2), 531–542 (2010)
9. Liao, W., Zhang, W., Zhu, Z., Ji, Q.: A decision theoretic model for stress recognition and user assistance. In: Proceedings of the National Conference on Artificial Intellegence, vol. 20, PART 2, pp. 529–534 (2005)
10. Liao, W., Zhang, W., Zhu, Z., Ji, Q., Gray, W.D.: Towards a decision-theoretic framework for affect recognition and user assistance. International Journal of Man-Machine Studies 64(9), 847–873 (2006)
11. Mourao-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H., Stetter, M.: Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional mri data. NeuroImage 28(4), 980–995 (2005)
12. Nakasone, A., Prendinger, H., Ishizuka, M.: Emotion recognition from electromyography and skin conductance. In: Proc. of the 5th International Workshop on Biosignal Interpretation, pp. 219–222 (2005)
13. Petrantonakis, P., Hadjileontiadis, L.: Emotion recognition from EEG using higher-order crossings. IEEE Transactions on Information Technology in Biomedicine 14(2), 186–197 (2010)
14. Picard, R.W.: Affective computing. Technical Report 321, M.I.T Media Laboratory Perceptual Computing Section (1995)
15. Picard, R.W.: Emotion research by the people, for the people. Emotion Review (2010) (to appear)

16. Rebolledo-Mendez, G., Dunwell, I., Martínez-Mirón, E.A., Vargas-Cerdán, M.D., de Freitas, S., Liarokapis, F., García-Gaona, A.R.: Assessing neuroSky's usability to detect attention levels in an assessment exercise. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5610, pp. 149–158. Springer, Heidelberg (2009)
17. Rodríguez, J.J., Kuncheva, L.I., Alonso, C.J.: Rotation forest: A new classifier ensemble method. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(10), 1619–1630 (2006)
18. Sato, J.R., Fujita, A., Thomaz, C.E., Martin, M.G.M., Mourao-Miranda, J., Brammer, M.J., Amaro Junior, E.: Evaluating SVM and MLDA in the extraction of discriminant regions for mental state prediction. NeuroImage 46, 105–114 (2009)
19. Sherwood, J., Derakhshani, R.: On classifiability of wavelet features for eeg-based brain-computer interfaces. In: Proceedings of the 2009 International Joint Conference on Neural Networks, pp. 2508–2515 (2009)
20. Takahashi, K.: Remarks on emotion recognition from biopotential signals. In: 2nd International Conference on Autonomous Robots and Agents, pp. 186–191 (2004)
21. van Gerven, M., Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., Ramsey, N., Haselager, P., Vuurpijl, L., Gielen, S., Desain, P.: The brain-computer interface cycle. Journal of Neural Engineering 6(4), 41001 (2009)
22. Wang, Z., Childress, A.R., Wang, J., Detre, J.A.: Support vector machine learning-based fMRI data group analysis. NeuroImage 36(4), 1139–1151 (2007)
23. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)
24. Yasui, Y.: A brainwave signal measurement and data processing technique for daily life applications. Journal Of Physiological Anthropology 38(3), 145–150 (2009)
25. Yazicioglu, R.F., Torfs, T., Merken, P., Penders, J., Leonov, V., Puers, R., Gyselinckx, B., Hoof, C.V.: Ultra-low-power biopotential interfaces and their applications in wearable and implantable systems. Microelectronics Journal 40(9), 1313–1321 (2009)