FROM PHYSIOLOGICAL SIGNALS TO EMOTIONS: IMPLEMENTING AND COMPARING SELECTED METHODS FOR FEATURE EXTRACTION AND CLASSIFICATION

Johannes Wagner, Jonghwa Kim, Elisabeth André

Institute of Computer Science, University of Augsburg, Germany

ABSTRACT

Little attention has been paid so far to physiological signals for emotion recognition compared to audio-visual emotion channels, such as facial expressions or speech. In this paper, we discuss the most important stages of a fully implemented emotion recognition system including data analysis and classification. For collecting physiological signals in different affective states, we used a music induction method which elicits natural emotional reactions from the subject. Four-channel biosensors are used to obtain electromyogram, electrocardiogram, skin conductivity and respiration changes. After calculating a sufficient amount of features from the raw signals, several feature selection/reduction methods are tested to extract a new feature set consisting of the most significant features for improving classification performance. Three well-known classifiers, linear discriminant function, k-nearest neighbour and multilayer perceptron, are then used to perform supervised classification.

1. INTRODUCTION

Emotion recognition is one of the key steps towards emotional intelligence in advanced human-machine interaction. Although many efforts have been taken recently to recognize emotions using facial expressions, speech and physiological signals [1, 2, 3, 4, 5], current recognition systems are not yet advanced enough to be used in realistic applications. Particularly little attention has been paid so far to physiological signals for emotion recognition compared to audiovisual emotion channels. Reasons are some significant limitations resulting from the use of physiological signals for emotion recognition. The main difficulty lies in the fact that it is very hard to uniquely map physiological patterns onto specific emotion types and that physiological data are very sensitive to motion artefacts. On the other hand, physiological signals have considerable advantages. We can continuously gather information about the users' emotional changes while they are connected to biosensors. Moreover, physiological reactions should be more robust against possible artefacts of human social masking since they are directly

controlled by the human autonomous nervous system. Work done in psycho physiology provides evidence that there is a strong relationship between physiological reactions and emotional/affective states of humans. One interesting suggestion in phylogenetic is that emotions develop as biological processes recede as determinants of behavior and that in species where biological processes directly and strongly determine behavior, emotions are absent or rudimentary [6].

Some previous works in physiological signal-based emotion recognition are summarized in Table 1. Physiological data sets used in most of the works are obtained by using visual elicitation methods in a lab setting where subjects intentionally express desired emotion types while looking at selected photos or watching movie clips. A recognition ac-

Author	Data Set	#Feat.	Classi.	Sel/Red	Results
Healey [2]	8 emotions using guided imagery technique [7]	11	DFA QDF	Fisher	80% - 90% for different subsets
Picard [3]	8 emotions using guided imagery technique [7]	40	DFA QDF	SFFS Fisher	81.25% for all 8 emotions
Haag et al. [4]	3 positive and negative states under variable arousal level using IAPS [8]	13	MLP	none	Arrousal with 96.58% Valence with 89.93%
Nasoz et al. [5]	6 emotions induced by movie clips	n/a	kNN DFA MBG*	none	kNN: 71.6% DFA: 74.3% MBG: 83.7% for all 6 emotions

* Marquardt Backpropagation

Table 1. Related works using physiological signals

curacy of over 80% on the average seems to be acceptable for realistic applications. However, it can be clearly observed that the accuracy strongly depends on the data sets (which were obtained in laboratory conditions). That is, the results were achieved for specific users in specific contexts. Moreover, it is very difficult to label emotion classes in physiological signals (waveforms) without uncertainty. In view of a generally applicable recognition system for realistic online applications, it is desirable to automatically select the most significant features and tune specific classifiers to manifold data sets obtained from different natural contexts.

In this paper, we describe the most important stages of a fully implemented emotion recognition system including data analysis and classification. First, we describe our physiological data set that we acquired under relatively natural conditions. We then investigate different feature extraction methods including feature selection/reduction algorithms and test selected classifiers. A hybrid method for feature extraction is considered and evaluated by combining it with each classifier. The most relevant features with respect to the emotion classes in our experiment are presented. Unlike the work above, we consider a larger variety of methods for pattern recognition in combination with different techniques for feature reduction in order to allow for easily tuning the recognition system to the requirements of a specific application.

2. EXPERIMENTAL SETTING AND DATA COLLECTION

To induce our subject to unaffectedly feel different emotions, we used four music songs that he himself carefully handpicked in respect of four targeted emotion classes, joy, anger, sadness and pleasure. Thereby, we advised him to choose songs that can bring back some special memories to him. An advantage of this method is that most people are used to listening to music during other activities and for this reason tend to associate different moods with specific songs. Those links should help our subject to easily switch into the desired affective state. We use a 2D-emotion model [11], see Figure 2. While there is no unanimous agreement on the basic emotions suggested by theorists, such as [9], this model provides a simplified representation of human emotions by two dimensions: arousal and valence. Actually, the specification of the underlying emotions is a matter of pragmatic choice and depends on the applications. While the subject listens to the music songs, four-channel biosensors are used to record electromyogram (EMG), electrocardiogram (ECG), skin conductivity (SC) and respiration change (RSP). Overall, 25 recordings (25 days) for each emotion were collected. The length of the recordings depends on the length of the songs.
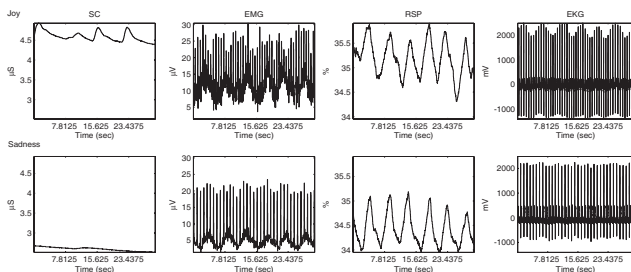

Fig. 1. An example of data collected from four sensors in two affective states: joy (top) and sadness (bottom).
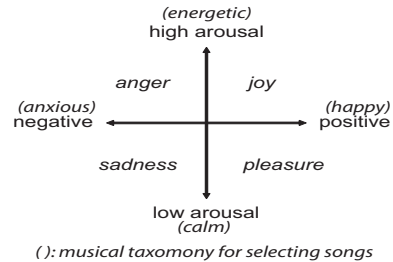

Fig. 2. The model of emotion for our experiment

3. FEATURE EXTRACTION AND CLASSIFICATION

Firstly, the raw signals were trimmed to a fixed length of two minutes. A lowpass filter was used to smooth the signals and the values were normalized to adjust to day-dependent differences in the baseline levels. The baseline of the SC signal was calculated and subtracted to consider only relative amplitudes. Artefacts of respiration and heart beat were deducted from the EMG signal. The breathing rate and amplitude were computed from the RSP signal. By detecting the R wave the heartbeat was calculated from the ECG signal. Afterwards, typical statistical values, such as mean value and standard deviation, were computed. Overall 32 features were extracted from the four signals.

In the next step, we tried to determine which features are most relevant to differentiate each affective state. Reducing the dimension of the feature space has two advantages. First of all, the computational costs are lowered and secondly the removal of noisy information may lead to a better separation of the classes. Several techniques were tested and compared with each other. Analysis of Variance (ANOVA) is a statistical method used to decide whether a feature shows a significant difference between two or more classes. Just the d most significant features are taken. The sequential forward selection (SFS) method is another popular algorithm starting with an empty set. The feature which fits best is inserted in every step. As an alternative one can also start with a full set and reduce the worst features. The latter method is known as sequential backward selection (SBS). Figure 3 shows the classification error against the number of features selected using the three feature selection methods. Instead of reducing the dimensionality by taking out some feature vectors in a high dimensional feature array, a new set of features can be extracted from the original one. Two popular methods are principal component analysis (PCA) and Fisher projection. Differently from feature selection, these reduction methods consider all information in the feature vector to create the new space. This means that discriminant parts of noisy features are also included. PCA involves feature transformation and obtains a set of transformed features rather than a subset of the original fea-
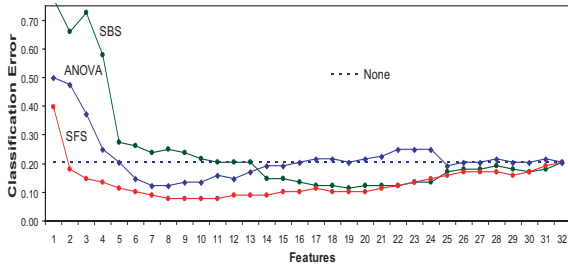
Fig. 3. Classification error against number of features selected using ANOVA, SFS and SBS.
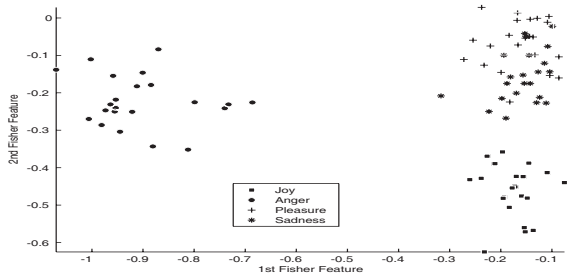

Fig. 4. An example feature set projected onto the first two fisher features.

tures. However, in our case a great disadvantage of PCA is that it does not consider any class information. This can lead to a loss of important discriminating information. In fact, our analysis showed that it was practically impossible to improve the classification error by this method. In contrast, Fisher projection uses class information to minimize scatter within classes and maximize scatter between classes. Figure 4 shows our data projected onto the first two Fisher features. Fisher transformation is often used to get a good representation of multidimensional class data in a two dimensional space. After reducing the feature space, three pattern classification methods have been tested: k-nearest neighbour (kNN), linear discriminant function (LDF) and a multilayer perceptron (MLP). KNN is an instance based method. It saves all training examples and labels a new instance by looking at its nearest neighbours. LDF is a statistical method which builds a probability model for each class. For a new instance, the class whose model fits best is chosen. MLP is a neural network with one hidden layer containing several hidden units. Its input layer has enough cells to accept a whole feature vector. The output layer consists of one neuron for each class. Because of the relatively small amount of data the „leave one out" method was used to evaluate the classifiers.

4. RESULTS OF THE EXPERIMENT

This section presents some of the results we obtained by applying the methods described above. First of all, we tried to recognize all four emotions without any dimensional reduc-

tion. KNN was tested with $k = 3, 5, 10, 15$ (3NN, 5NN, 10NN, 15NN). Recognition rates of 79.55% and 80.68% were achieved which was equal to or slightly better than 79.55% with LDF. Using 4 and 6 hidden units (MLP4, MLP6) the neural network also provides 80.68%. For 2 and 8 hidden units (MLP2, MLP8), the recognition rate was only 78.41%. Overall, the average results for all four emotions were similar independent from the classifiers but differed among individual emotions. E. g. while joy was recognized by MLP4 and MLP6 with 95.46%, only 72.73% was achieved with 15NN. All algorithms had particular problems in separating pleasure and sadness. Table 2 provides an overview of the results. Next, we tested the three methods for feature

Method	Joy	Anger	Pleasure	Sadness	Quote
LDF	77.27%	100%	72.73%	68.18%	79.55%
3NN	90.91%	100%	72.73%	59.09%	80.68%
5NN	86.36%	100%	72.73%	59.09%	79.55%
10NN	81.82%	100%	77.27%	63.64%	80.68%
15NN	72.73%	100%	77.27%	68.18%	79.55%
MLP2	90.91%	90.91%	72.73%	59.09%	78.41%
MLP4	95.46%	95.46%	63.64%	68.18%	80.68%
MLP6	95.46%	100%	63.64%	63.64%	80.68%
MLP8	86.36%	100%	59.09%	68.18%	78.41%

Table 2. Recognition rates of four emotions without dimensional reduction.

reduction described earlier and a hybrid method of SFS and Fisher (SFS/Fisher). Using SFS in combination with LDF the recognition rate was raised from 79.55% up to 92.05%, by SFS/ Fisher in combination with 5NN from 79.55% to 90.91% and with MLP6 from 80.68% to 88.64%. Using ANOVA in combination with LDF and 5NN it gained just 7.97% and 7.08%. ANOVA has, however, the advantage of requiring less computation power. Table 3 gives a review of the results with dimensional reduction. We also tested

Method	none	SFS	Fisher	SFS/Fisher	ANOVA
LDF	79.55%	92,05%	79,55%	90,91%	87,50%
5NN	79,55%	86,36%	80,68%	90,91%	86,36%
MLP6	80,68%	87,50%	80,68%	88,64%	86,36%

Table 3. Recognition rates of four emotions with dimensional reduction.

discrimination between sets of emotions. We divided the emotions into groups of negative (anger/sadness) and positive (joy/pleasure) valence and into groups of high arousal (joy/anger) and low arousal (sadness/pleasure). It turned out that it was much easier to separate emotions along the arousal axis than along the valence axis. High and low arousal was recognized in about 95% of the cases. Negative and positive in only about 87% of the cases. Table 4 provides an overview of the results with the single sets of

emotions.

Method	4 emotions	Valence	Arousal
LDF (SFS)	92.05%	86.36%	96.59%
5NN (SFS/Fisher)	90.91%	86.36%	94.32%
MLP6 (SFS/Fisher)	88.64%	88.64%	94.32%

Table 4. Recognition results with sets of emotions.

Finally, we tested our recognition system with a data-set[1] from MIT Media Lab [3] to evaluate the "universality" of the system and to see if correlations exist between the subjects. The MIT-dataset contains physiological data of four sensors: SC, EMG, RSP and BVP (blood volume pulse). 20 data sets of a single subject consecutively expressing eight emotional states were collected. In this experiment, the subject used specific images as cues during each emotion episode. To enable a comparison of both data sets only subsets which divide the emotions in groups of low/high arousal and positive/negative valence respectively were used.

Table 5 compares the results of both data sets. Although both data sets were collected in completely different experiments, similar recognition rates were achieved. In both data sets it was easier to distinguish emotions along the arousal axis than the valence axis. To find out which features are

Groups	valence		arousal	
Data set	MIT	Own	MIT	Own
LDF (SFS)	85,00%	86,36%	87,50%	96,59%
5NN (SFS/Fisher)	81,67%	86,36%	86,88%	94,32%
MLP6 (SFS/Fisher)	82,50%	88,64%	86,25%	94,32%

Table 5. Comparison of the results from both data sets.

significant for a specific emotion, the Tukey method [10] was applied to our data set. This method provides a pairwise comparison of features among multiple classes. It turned out that joy was characterized by high SC- and EMG-levels, deep and slow breathing and an increased heart rate. In contrast, anger was accompanied by flat and fast breathing. The SC- and EMG-levels were high as well. Pleasure and sadness are well identified by low SC- and EMG-signals, but pleasure was accompanied by a faster heart rate. Of course these results are highly user-dependent. When applying the same method to the MIT-Dataset, we observed perceptible differences in the physiological response of both subjects. While in our data set positive emotions were characterized by a low SC-level, the MIT-Dataset showed a high level of SC. Nevertheless, it turned out that high SC- and EMG-levels were also good indicators in general for high arousal. There was also a correlation between a higher breathing rate and emotions of negative valence.

5. CONCLUSION

In this paper, we investigated a variety of methods for pattern recognition in combination with different techniques for feature reduction to recognize affective state from physiological data. Physiological data was acquired in four different affective states and three pattern recognition methods have been tested: k-nearest neighbour, linear discriminant function and a multilayer perceptron. Recognition rates of about 80% were achieved for all three classifiers. By applying feature reduction, results could be improved up to 92%. Our recognition rates are comparable to those achieved by [4]. However, in contrast to them we rely on an automatically selected feature set. To compare the results, a second data set was used which had been recorded at MIT Media Lab. It turned out that for both data sets it was easier to distinguish emotions along the arousal axes than along the valence axes. Statistical methods were used to find out which features are significant for a specific emotion for each data set. Although differences in the physiological response of the subjects were noticed we also found similarities, e.g. a higher breathing rate for emotions with a negative valence.

6. REFERENCES

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz and J. G. Taylor: *Emotion recognition in human-computer interaction*, IEEE Sinal Process. Mag., 18, pp. 32-80, 2001

[2] J. A. Healey: *Wearable and Automotive Systems for Affect Recognition from Physiology*, PhD thesis, MIT, Cambridge, MA, May 2000

[3] R. W. Picard, E. Vyzas, and J. Healey: *Toward Machine Emotional Intelligence: Analysis of Affective Physiological State*, IEEE Transactions Pattern Analysis and Machine Intelligence, Vol.23, No.10, pp.1175-1191, Oct. 2001

[4] A. Haag, S. Goronzy, P. Schaich, J. Williams: *Emotion Recognition Using Bio-Sensors: First Step Towards an Automatic System*, Affective Dialogue Systems, Tutorial and Research Workshop, Kloster Irsee, Germany, June 14-16, 2004

[5] F. Nasoz, K. Alvarez, C. L. Lisetti, N. Finkelstein: *Emotion Recognition from Physiological Signals for Presence Technologies*, International Journal of Cognition, Technology and Work, Special Issue on Presence, Vol 6(1), 2003

[6] C. Ratner: *A Cultural-Physiological Analysis of Emotions*, Culture and psychology, vol. 6, pp. 5-39, 2000

[7] M. Clynes: *Sentics: The Touch of the Emotions*, 250 pp, Doubleday/Anchor, New York, 1977

[8] Center for the Study of Emotion and Attention: *The International Affective Picture System: Digitized Photographs*, University of Florida, Center for Research in Psychophysiology, 2001

[9] P. Ekman: *An argument for basic emotions. Cognition and Emotion*, In: Cognition and Emotion, 8, pp. 169-200, 1992

[10] *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/

[11] P. Lang: *The emotion probe: Studies of motivation and attention*, American Psychologist, 50(5), pp. 372-385, 1995