Indian Institute of Technology

Neural Network Course Project

Emotion Recognition from EEG Data

Group Members:
Nipun Shrivastava
2011CS50288
Nitesh Bharadwaj G.
2011EE20514

Supervisor:
Dr. Jayadeva
Teaching Assistant:
Sumit Soman

November 30, 2014

Contents

1 Introduction

Emotion recognition by computers is becoming increasingly popular. Our project focused on recognizing emotion from human brain activity, measured by EEG signals. We have tried to predict the level of valence, arousal, dominance and liking from the available EEG data using different learning models. The objective of this project was to develop a learning based algorithm that can be used to discriminate emotions using EEG data. We plan to classify the emotions primarily into valence, dominance, liking and arousal dimensions(high and low). Many different emotions like anger, disgust, fear, joy etc. can be plotted in these four dimensions. We can break down our project in few major steps as given below

1. Preprocessing Data
2. Feature Extraction
3. Feature Selection
4. Learning
5. Prediction and Verification

2 Dataset Description

We performed our work on DEAP dataset that has been made publicly available. The set contains the recordings of 7 physiological modalities for 32 users, each user viewing 40 video clips and grading them for emotional content. The DEAP dataset consists of two parts:

1. The ratings from an online self-assessment where 120 one-minute extracts of music videos were each rated by 14-16 volunteers based on arousal, valence and dominance.
2. The participant ratings, physiological recordings and face video of an experiment where 32 volunteers watched a subset of 40 of the above music videos. EEG and physiological signals were recorded and each participant also rated the videos as above. For 22 participants frontal face video was also recorded.

Though the report first talks about dataset and analysis methodology

3 Analysis Criterion

For analysing different learning models we went with the ***hold-out cross validation*** method (widely used technique to minimize generalization error) which is defined as follows:

1. Randomly split S into S_{train} (say, 70% of the data) and S_{cv} (the remaining 30%). Here, S_{cv} is called the hold-out cross validation set.
2. Train each model M on S_{train} only, to get some hypothesis h_i.

3. Repeat steps 4-5 times for each model and find the average accuracy and variance of accuracy for each for the learnt hypothesis h_i and then use it to compare the models on basis of the correctness as well as their consistency.

By testing on the remaining set of examples S_{cv} that the models were not trained on, we obtain a better estimate of each model's true generalization error. Usually, 70% is a typical choice for the amount of data to be used for training.

We further analysed each model by changing the amount of data used to learn the model, i.e., by varying the percentage of training data used to train the model and saw how well the model performed for various amount of training data.

4 PreProcessing Data

The data was downsampled (to 128Hz) and segmented. However, since brain activity only produces very weak signals, the eeg signals contain a lot of background noise. Before using the signals for emotion recognition, they have to be preprocessed, in order to remove unwanted noise using filters like bandpass, butterworth bandreject etc in frequency domain. Although eeg signals contain a lot of information, they also tend to result in very large amounts of data. We directly downloaded the downsampled Pre-processed data that was processed upon in order to make it free of all the noise and artefacts.
Sources of noise are static electricity or electromagnetic fields produced by surrounding devices. In addition to this external noise, the eeg signal is most of the times heavily influenced by artifacts that originate from body movement.

5 Feature Extraction

The data-source in DEAP dataset, [5], provides a 32 channel data for EEG apart from other peripheral physiological data. Apart from these channels, [1],[2] and [3] suggested that longitudinal and transverse bipolar montages can be constructed as virtual features. For e.g. C1-P1 is a virtual channel. 61 virtual channels are constructed making the total number of channels to be 93.

5.1 PSD

The common feature used in papers [2], [3] and [4] is power spectrum. We implemented the same in MATLAB using inbuilt functions.

1. Firstly, psd of each channel was calculated using a hamming window of 2 seconds with a 50% overlap.

2. Secondly, the spectral powers were calculated for each of the four bands (delta, alpha, beta and gamma bands).
   This makes the total number of features 372.

5.2 MSCE

Papers [1] and [2] suggest that the interaction between different regions of the brain plays a major role in triggering emotions rather than individually. So, the statistical features relating the different channels have to be taken into account. We have chosen *Magnitude-Squared Coherence Features* as it gives quantitative estimates of that relation. A similar method was also used in [1]. MSCE features for each pair of channels was calculated using the script we wrote on Matlab.There are ${}^{32}C_2 = 496$.

6 Feature Selection

Due to the large number of features, the problems of feature selection and removal of redundant and irrelevant features becomes pertinent. The feature selection problem in different papers is dealt with in different methods. For instance, paper [2] uses Pearson Correlation Coefficient while paper [1] uses Principal Component Analysis(PCA) to extract significant features. Therefore, we decided to go forward with Principal Component Analysis for reducing the total number of features. We even exploited other options like regularization for regression, but finally went with PCA.

7 Learning

We used a variety of learning methods to construct an accurate and efficient predictor. Below is the description of all the methods that we used:

7.1 Neural Network

We formed a simple 2 layered feed forward neural network to build a classifier for the problem. The network consisted of h hidden layer neurons which were varied to give us the best possible accuracy. We actually used two type of neural networks one which learnt weights for all the 4 classes simulataneously with 4 output units and the other model that consisted of 4 separate independent neural networks, one for each parameter. The second model performed better when we checked the two models with varying number of hidden layer neurons. We were careful in choosing the number of hidden layer neurons so that not to perform overfitting or underfitting. We used 100 hidden units in our network.

7.2 SVM

For this we simply used the LIBSVM tool to perform classification after obtaining the x vectors and y labels for all the four cases. We used an RBF kernel for classification and tested on different gamma and cost values and finally found $gamma = 0.004$ *and* $C = 11$ as the best combination pair. We arrived at this value by performing a **co-ordinate descent on the two parameters**.

7.3 Logistic Regression

Logistic regression is an approach to prediction, like Ordinary Least Squares (OLS) regression. However, with logistic regression, we use sigmoid function on the linear combination of input features. A simple logistic regressor with gradient descent was implemented and used to learn the theta values as part of model which were then used to evaluate the accuracy of the model. Again we used different values for stopping criterion(change in error in succcessive loops) to get the best possible model from this domain for comparison.

8 Prediction and Verification

The learnt model was used to predict the labels of the cross validation set which were then cross checked with the ground truth. This was performed multiple times on randomly distributed data sets for each model in order to get an estimate of average accuracy and also to calculate the variance.

9 Result

Initially we tested our models using all the 93 channels*4 band powers (alpha, beta, theta and gamma) = 372 features and found out that none of the models were working very well and gave us the accuracy of less than *60%*. This was very disappointing as even a *random binary classifier can achieve an average accuracy of 50 percent for a binary label.* The possible reason that we inferred from the initial results was **overfitting**. This happened because we had *372 features and only 1280 data points* which are obviously not enough to develop a good general classifier (*primary motivation to apply PCA*).
Also after close analysis of the raw data we found that even though the scales were same, there was a *lot of variation between the EEG values of same electrode for 2 different persons when watching the same video.* Upon mapping the EEG data, we saw the wave structure to be completely different for different individuals. While one person had a continuous fluctuating EEG value while the other had periodic peaks for the same with ranges varying by order of magnitudes.

Upon adding the MSCE features which takes into account the statistical interaction between different channels, the accuracy improved.

Classification Method	Valence	Arousal	Dominance	Liking
C-SVM	56% ± 2.9%	59% ± 2.9%	64% ± 2.7%	67% ± 2%
C-SVM with MSCE	63% ± 2%	62% ± 2%	65% ± 2.5%	67% ± 2.7%
Error Back propagation	56% ± 0.2	54.3% ± 7.7%	62% ± 0.8%	60% ± 4.4%
Error Back Propagation with MSCE	57% ± 2.7%	58% ± 0.3%	62% ± 0.8%	60% ± 0.5%
Logistic regression	50% ± 5%	48% ± 7%	48% ± 7%	44% ± 11%
Logistic regression with MSCE	58% ± 3.3%	59% ± 3.1%	63% ± 1.7%	68% ± 2.8%

Figure 1: Final Results

Since, the main problem was overfitting, using PCA helped in increasing the accuracy to over 60%.

10 Conclusion

1. After applying PCA, average accuracy had a marginal increase(2-4.5%) for all the models helping us infer that PCA helps to curb over fitting to some extent but not greatly. Its the only explaination as PCA doesnot add any extra information.

2. The performance of logistic regression and linear SVM, with accuracy less than 50% for all cases(with and without PCA) using only PSD features show that emotions aren't linear functions of the spectral features.

3. But an accuracy of almost 56% (in case of PCA and MCSE) for linear SVM and logistic regression tells us that the MCSE feature (cross spectral feature) add to the linearly separable nature of the data and are more fitting features to be used for emotion classification.

4. Neural network performed better for separate networks for each label, suggesting minimum cross correlation between the 4 labels. Because of this reason, the weight vectors are influenced by all the states at the same time in a single network model bringing down the overall accuracy.

5. Of all the models, **C-SVM** performed the best and with minimum average standard deviation for each parameter giving us the best model to be used for this purpose.

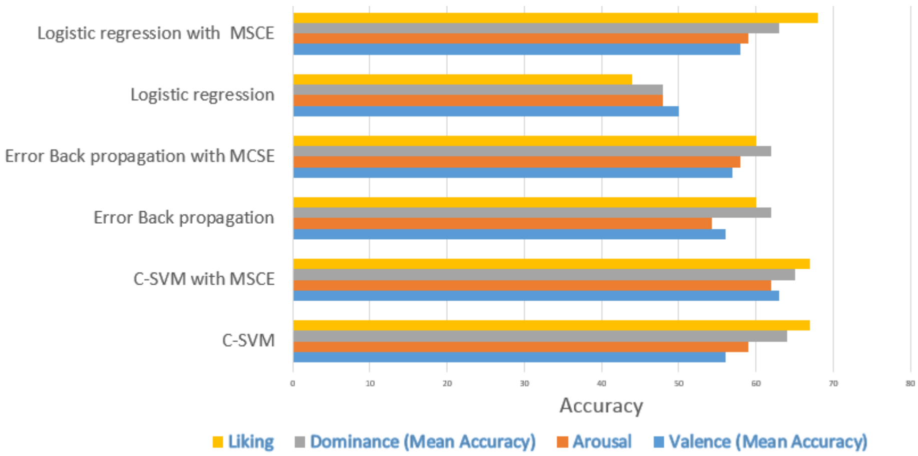


Figure 2: Performance of various methods

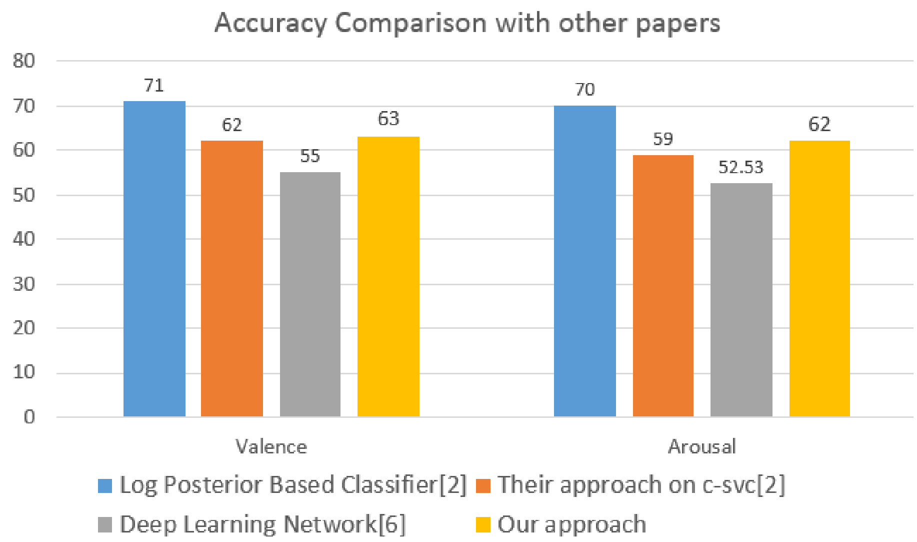
6. Bayesian weighted-log-posterior method and the feedforward network of ERNN boast an accuracy of 70%, but when they implemented them on

normal SVM and neural network their accuracy came down to 60%, making our approach much more efficient. Similarly [6] uses a deep learning methodology and could only get a 53% accurate predictions.

7. The scale difference which we mentioned briefly in the results suggests us that it might be a better idea to make the classifier more specific to subjects as the EEG data may attain completely different values for 2 persons feeling the same way. But since theories suggest that an individual's brain performs in a consistent way, a learner can still be used and it would be a better idea for the learner to initially train on some supervised data specific to the subject for which it had to predict the emotions in future, then the accuracy can increase greatly.

Classification Method	Valence	Arousal	Dominance	Liking
C-SVM with MSCE	57% ± 13%	47% ± 14.9%	63% ± 14.5%	86% ± 9%
Error Back Propagation with MSCE	50% ± 11%	49% ± 8%	59% ± 13%	85% ± 7.5%
Logistic regression with MSCE	50.5% ± 16%	59.5% ± 16%	69.5% ± 17%	91.5% ± 5.5%

Figure 3: Training and Testing on single person



8. An average accuracy of above 80% for some labels like liking suggests that a person specific learner can prove to be a much better approach to take

instead of making a general classifier. But we can not say that with much credibility as we had only 40 input patterns per person corresponding to 40 videos watched by him/her.

9. Our approach performed relatively well as compared to the approaches taken by other papers. Log posterior method was the only method to outperform us on the valence and arousal dimensions.

11 References

References

[1] Khosrowabadi, Reza, et al. *ERNN: A Biologically Inspired Feedforward Neural Network to Discriminate Emotion from EEG Signal.* (2014): 1-1

[2] Yoon, Hyun Joong, and Seong Youb Chung. *EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm.* Computers in biology and medicine 43.12 (2013): 2230-2237.

[3] Wang, Xiao-Wei, Dan Nie, and Bao-Liang Lu. *Emotional state classification from EEG data using machine learning approach.* Neurocomputing 129 (2014): 94-106.

[4] Chanel G., Ansari-Asl K., and Pun T. *Valence-arousal evaluation using physiological signals in an emotion recall paradigm.* InSystems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on(pp. 2662-2667). IEEE.

[5] S. Koelstra, C. Mühl, M. Soleymani et al. *DEAP: a database for emotion analysis; using physiological signals.* IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18–31, 2012.

[6] Jirayucharoensak, Suwicha, Setha Pan-Ngum, and Pasin Israsena. *EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation.* The Scientific World Journal 2014 (2014).