Electroencephalography based Emotion Recognition using Fisher's Linear Discriminant Analysis on Support Vector Machine

I N Yulita¹, D Novita², A Sholahuddin³ and Emilliano⁴

¹,³ Department of Computer Science, Universitas Padjadjaran, Jl. Raya Bandung – Sumedang Km. 21, Sumedang 45363, Indonesia
²,⁴ Department of Electro, Universitas Padjadjaran, Jl. Raya Bandung – Sumedang Km. 21, Sumedang 45363, Indonesia

Abstract. Emotions as intense feelings for reactions to something affect someone in interacting with others such as in determining choices, actions, and perceptions. The emotional state of an individual can be seen clearly through facial expression and tone of speech. Apart from facial features or voice features, identification of emotions can also be done through brain waves. This study used an electroencephalogram signal as an input to recognize types of emotions. The electroencephalogram signal was chosen because it can record the true emotions of individuals. The recognition of emotions based on Support Vector Machine (SVM). To improve the performance, this method was combined with Fisher's Linear Discriminant Analysis (FLDA). The experiments showed the SVM performance increased above 30%. As a comparison, this research also implemented Multi-Layer Perceptron (MLP). The results showed that the performances of SVM and FLDA-SVM were higher than MLP or FLDA-MLP. It showed that FLDA-SVM was the best method of this research in recognizing emotions.

1. Introduction

Emotion is a feeling that encourages individuals to respond or behave towards stimuli, both from within and from outside themselves [1]. It, as a reaction, is closely related to cognitive activity as a result of the perception of a situation. Many scientists and psychologists try to make models to represent emotions. The model can be divided into two approaches namely category models and dimensional models. The dimensional model uses several axes to map emotions to a plane. The commonly used dimension is a model using two axes (arousal-valence or energy stress). In this model, the affect is described as a combination of the valence dimension that moves from pleasant (positive) to unpleasant (negative) feelings and arousal dimensions that refer to energetic feelings. For example, feelings of joy and peace are both aroused by stimuli with positive valence, but feelings of joy operate at high arousal levels while feelings of peace operate at low arousal levels. On the other hand, a combination of negative feelings and high arousal triggers feelings of anger, a combination of negative feelings and low arousal triggers feelings of sadness [2]. Affective stimuli can be people, conditions, things, or events; real or imaginary, in the past, future, or present.

In the world of robotics, the introduction of emotions plays a role in helping robots to accommodate humans will better. The existence of this ability, of course, will make the work of robots more leverage
in helping human tasks. In recognizing these emotions, you can use sound and images. But this study uses brainwave signals or electroencephalograph (EEG). The brain works a lot when feelings emerge. So this is the reason this data is used in this study. EEG has a lot to play in research in the field of health, including sleep problems [3]. To be classified, EEG data undergoes preprocessing [4], and feature extraction [5].

The methods commonly used for classification of human emotions are machine learning [6]. Some methods are Hidden Markov Models [7], Naive Bayes, boosting, and support vector machines (SVM). Of the several classification methods, SVM excels in emotional recognition because it produces the highest accuracy. Therefore, this study used SVM as a classifier. In order for SVM to work more optimally, this study proposed fisher's linear discriminant analysis (FLDA) to reconstruct features from wavelet results [8].

2. Literature study

2.1. Fisher's linear discriminant analysis (FLDA)

FLDA aims to transform features from data into more optimal features [9]. This method applies statistical operations to achieve the goal. The same objective is also the target of principal component analysis (PCA). But the drawback is a failure to separate data between classes. It is what can be done more optimally with FLDA. The mechanism is by optimizing the distribution between data in different classes. Also, separation focuses on data classification. Unlike PCA, which focuses on feature classification.

The following are the steps of the feature extraction process using FLDA:

a) Convert a two-dimensional matrix into one dimension or into row vectors or column vectors.
b) Group training data into a matrix of classes.
c) Calculate the average value of each class.
d) Calculate the total mean of all classes.
e) Calculate the matrix between-class scatter and within-class scatter matrix.
f) Calculate the covariance matrix value
g) Calculate eigenvalue (v) and eigenvector (d).
h) Calculate the projection matrix and the weight matrix.

2.2. Multi-layer perceptron (MLP)

MLP is a part of artificial neural network (ANN). ANN is a technology that was born from human efforts to find out how animal coordination systems occur, how nerves work, optimize themselves, and be able to become the center of all animal biological system [10]. Information processing in the ANN can be abbreviated as follows: Signals (either action or potential) appear as input units (synapses); the effect of each signal is expressed as a form of multiplication with a weight value to indicate the strength of the synapses. All signals given this weight multiplier are then added together to produce an activation unit. If this activation exceeds a certain threshold, the unit will provide output in the form of a response to an input. Based on the activation unit and its comparison with the threshold value, the results are then entered into the transfer function (non-linear function), which will produce output.

ANN has a layer of neurons. Neurons will be connected. Inter-layers are related to each other except the input layer can only be connected to the layers afterward, and vice versa with the output layer. Between the input layer and the output layer, there is a hidden layer. One architecture of ANN is MLP [11]. This method has more than one layer. With so many layers, this method is more optimal in solving more complex problems.

2.3. Support vector machine (SVM)

Initially, SVM works for linear operations [12]. Over time the SVM method was developed not only to be able to solve linear classification problems but also to solve non-linear classification problems. The mechanism is done by applying the concept of kernel tricks. This concept has proven to be effective in
high-dimensional data. SVM can also be used for regression where the output is a real or continuous number, called Support Vector Regression (SVR). This SVR method can be used to make predictions.

SVM uses the kernel to transform input space into feature space or implement models to a higher dimension [13]. This kernel function has many types including radial basis function (RBF) kernel, linear kernel, and Polynomial kernel. The differences in the three kernel functions are found in the mapping function into the feature space, and each kernel has advantages and disadvantages in each particular case so an experiment is needed in finding the best kernel function used in a case.

3. Research methodology
Research methodology is shown in Figure 1.

![Figure 1. The proposed research methodology in this study](image)

The stages of research in this study were:

a) Data collecting
   This study used secondary data from DEAP Queen Mary University London's research team. Electrodes were installed in the participant's scalp using a set based on 10-20 international systems rules [14]. There were 32 channels used in research. The dataset consists of 32 files. Each file contains 40 EEG records. During the recording process, each respondent was stimulated with several music videos and asked to do their assessment of valence and arousal levels. These levels were then to be the reference in the dataset. This study then used four classes taken from the data namely High Arousal High Valence (HAHV), Low Arousal High Valence (LAHV), Low Arousal Low Valence (LALV), High Arousal Low Valence (HALV).

b) Preprocessing
   This stage processes the signal with a downsample so that the sampling frequency becomes 128 Hz. The next is the elimination of noise and artifacts. The signals were decomposed into alpha, beta, and gamma waves. The final stage is implementing PCA.

c) Fisher's linear discriminant analysis (FLDA)
   It was used to classify individuals into one of two or more groups in cases where the independent variable is metric data (interval or ratio) and the dependent variable is nonmetric data (nominal or ordinal).

d) Classification
   This study used SVM and MLP as classifiers. SVM was run with eps = 0.001, gamma = 0, and degree of kernel = 3. Two types of kernel types were used in this study, namely linear and sigmoid. MLP worked based on learning rate = 0.3, momentum = 0.2, batch size. MLP was also tested by using two hidden layer values, i and t. i showed the numbers of features and classes then divided by 2. t was 2 times i.
4. Results
The accuracy of each method is shown in Table 1.

Table 1. The accuracies of all methods in this study

| Methods            | Accuracy (%) |
|--------------------|--------------|
| MLP (i)            | 39.84        |
| MLP (t)            | 40.70        |
| SVM (linear)       | 41.48        |
| SVM (sigmoid)      | 41.48        |
| FLDA-MLP (i)       | 86.48        |
| FLDA-MLP (t)       | 86.17        |
| FLDA-SVM (linear)  | 87.03        |
| FLDA-SVM (sigmoid) | 87.11        |

Table 1 shows that the lowest performance was in MLP, and the highest was FLDA-SVM. Both MLP and SVM show increased accuracy after being combined with FLDA. Without using FLDA, MLP with t hidden layer has a higher performance than MLP with i hidden layers. But after being combined with FLDA, performance shows the opposite. SVM by using a linear or sigmoid kernel has an accuracy of 41.48%. After being combined, the accuracy increases above 40%. In addition, the accuracy of FLDA-SVM using a sigmoid kernel is higher than linear.

Tables 2 and 3 show that MLP has a tendency to predict all data into HAHV. This also happens in SVM in Tables 4 and 5. Even in Table 5, none of the data is classified outside the HAHV class. Tables 6 and 7 show changes in performance from MLP after being combined with FLDA. Classes other than HAHV have also been classified more precisely. However, the portion of accuracy is still less than the combination of FLDA and SVM in Tables 8 and 9.

Table 2. Confusion matrix of MLP (i)

|             | Predicted: HAHV | Predicted: LAHV | Predicted: LALV | Predicted: HALV |
|-------------|-----------------|-----------------|-----------------|-----------------|
| Actual: HAHV | 440             | 46              | 3               | 42              |
| Actual: LAHV | 224             | 34              | 1               | 18              |
| Actual: LALV | 162             | 12              | 0               | 8               |
| Actual: HALV | 228             | 18              | 4               | 35              |

Table 3. Confusion matrix of MLP (t)

|             | Predicted: HAHV | Predicted: LAHV | Predicted: LALV | Predicted: HALV |
|-------------|-----------------|-----------------|-----------------|-----------------|
| Actual: HAHV | 446             | 52              | 4               | 29              |
| Actual: LAHV | 217             | 43              | 1               | 16              |
| Actual: LALV | 167             | 12              | 0               | 8               |
| Actual: HALV | 227             | 21              | 5               | 32              |
### Table 4. Confusion matrix of SVM (linear)

| Actual: HAHV | Predicted: HAHV | Predicted: LAHV | Predicted: LALV | Predicted: HALV |
|--------------|-----------------|----------------|----------------|----------------|
| HAHV         | 523             | 3              | 0              | 5              |
| LAHV         | 272             | 1              | 0              | 4              |
| LALV         | 180             | 3              | 0              | 4              |
| HALV         | 274             | 3              | 1              | 7              |

### Table 5. Confusion matrix of SVM (sigmoid)

| Actual: HAHV | Predicted: HAHV | Predicted: LAHV | Predicted: LALV | Predicted: HALV |
|--------------|-----------------|----------------|----------------|----------------|
| HAHV         | 531             | 0              | 0              | 0              |
| LAHV         | 277             | 0              | 0              | 0              |
| LALV         | 187             | 0              | 0              | 0              |
| HALV         | 285             | 0              | 0              | 0              |

### Table 6. Confusion matrix of FLDA-MLP (i)

| Actual: HAHV | Predicted: HAHV | Predicted: LAHV | Predicted: LALV | Predicted: HALV |
|--------------|-----------------|----------------|----------------|----------------|
| HAHV         | 498             | 11             | 14             | 8              |
| LAHV         | 29              | 237            | 5              | 6              |
| LALV         | 24              | 11             | 139            | 13             |
| HALV         | 38              | 7              | 7              | 233            |

### Table 7. Confusion matrix of FLDA-MLP (t)

| Actual: HAHV | Predicted: HAHV | Predicted: LAHV | Predicted: LALV | Predicted: HALV |
|--------------|-----------------|----------------|----------------|----------------|
| HAHV         | 495             | 14             | 13             | 9              |
| LAHV         | 28              | 238            | 5              | 6              |
| LALV         | 25              | 11             | 137            | 14             |
| HALV         | 38              | 7              | 7              | 233            |

### Table 8. Confusion matrix of FLDA-SVM (linear)

| Actual: HAHV | Predicted: HAHV | Predicted: LAHV | Predicted: LALV | Predicted: HALV |
|--------------|-----------------|----------------|----------------|----------------|
| HAHV         | 504             | 10             | 10             | 7              |
| LAHV         | 29              | 237            | 5              | 6              |
| LALV         | 25              | 11             | 137            | 14             |
| HALV         | 38              | 6              | 5              | 236            |
Table 9. Confusion matrix of FLDA-SVM (sigmoid)

|        | Predicted: HAHV | Predicted: LAHV | Predicted: LALV | Predicted: HALV |
|--------|-----------------|-----------------|-----------------|-----------------|
| Actual: HAHV | 504             | 10              | 10              | 7               |
| Actual: LAHV  | 29              | 236             | 6               | 6               |
| Actual: LALV  | 25              | 8               | 142             | 12              |
| Actual: HALV  | 38              | 6               | 8               | 233             |

5. Conclusion
This study proposed the use of FLDA in SVM for emotion recognition. The results were known that SVM has increased accuracy when combined with FLDA. As a comparison, MLP was also applied in this study. FLDA also provides improved performance on MLP. But the value is lower than FLDA SVM. This shows that FLDA SVM is the best method in this study, especially by using a sigmoid kernel.

6. References

[1] Niedenthal PM and Ric F 2017 Psychology of emotion Psychology Press
[2] Moksnes UK, Espnes GA and Haugan G 2014 Stress, Sense of Coherence And Emotional Symptoms Adolescents Psychology & Health 291 32-49
[3] Yulita IN, Fanany MI and Arymurthy AM 2017 Fuzzy Clustering and Bidirectional Long Short-Term Memory for Sleep Stages Classification Proc. Int. Conf. Intelligent System and Information Technology (Indonesia: IEEE) p 11-16
[4] Yulita IN, Rosadi R, Purwani S and Suryani, M 2018 Multi-Layer Perceptron for Sleep Stage Classification Journal of Physics: Conference Series Vol 1028 No. 1 (IOP Publishing) p 012212
[5] Fadhilullah MU, Resahya A, Nugraha DF and Yulita IN 2018 Sleep Stages Identification in Patients With Sleep Disorder Using K-Means Clustering Journal of Physics: Conference Series 1013(1) 012162
[6] Atkinson J and Campos D 2016 Improving BCI-Based Emotion Recognition by Combining EEG Feature Selection and Kernel Classifiers Expert Systems with Applications 4 35-41
[7] Borst JP and Anderson JR 2015 The discovery of processing stages: Analyzing EEG data with hidden semi-Markov models NeuroImage vol 108 p.60-73.
[8] Pham BT and Prakash I 2019 Evaluation and comparison of LogitBoost Ensemble, Fisher’s Linear Discriminant Analysis, logistic regression and support vector machines methods for landslide susceptibility mapping Geocarto International Vol 34(3) p. 316-333.
[9] Shinmura S 2016 New theory of discriminant analysis after R Springer Vol 10 p.978-981
[10] Rodrigues E, Pereira LD, Gaspar AR et al 2017 Estimation of classrooms occupancy using a multi-layer perceptron arXiv preprint
[11] Fan X, Wang L and Li S 2016 Predicting chaotic coal prices using a multi-layer perceptron network model Resources Policy Vol 50 p. 86-92
[12] Yu B et al 2016 Hybrid dual-tree complex wavelet transform and support vector machine for digital multi-focus image fusion Neurocomputing Vol. 182 p. 1-9
[13] Jaramillo F et al 2018 On-line estimation of the aerobic phase length for partial nitrification processes in SBR based on features extraction and SVM classification Chemical Engineering Journal Vol. 331 p.114-12
[14] Koelstra SE 2012 DEAP: A Database For Emotion Analysis; Using Physiological Signals IEEE Transactions on Affective Computing Vol. 3(1) p. 18-31
Acknowledgments
This work supported by Padjadjaran University Internal Grant funded by Directorate of Research and Service Community of Padjadjaran University