Orthogonal Locality Sensitive Fuzzy Discriminant Analysis in Sleep-Stage Scoring

Rami N. Khushaba, Rosalind Elliott, Akram AlSukker, Ahmed Al-Ani, and Sharon McKinley
University of Technology, Sydney (UTS)
Rkhushab, alsukker, Ahmed@eng.uts.edu.au, Sharon.McKinley@uts.edu.au

Abstract

Sleep-stage scoring plays an important role in analyzing the sleep patterns of people. Studies have revealed that Intensive Care Unit (ICU) patients do not usually get enough quality sleep, and hence, analyzing their sleep patterns is of increased importance. Due to the fact that sleep data are usually collected from a number of Electroencephalogram (EEG), Electromyogram (EMG) and Electrooculography (EOG) channels, the feature set size can become large, which may affect the development of on-line scoring systems. Hence, a dimensionality reduction step is needed. One of the powerful dimensionality reduction approaches is based on the concept of Linear Discriminant Analysis (LDA). Unlike existing variants of LDA, this paper presents a new method that considers the fuzzy nature of input measurements while preserving their local structure. Practical results indicate the significance of preserving the local structure of sleep data, which is achieved by the proposed method, and hence attaining superior results to other dimensionality reduction methods.

1. Introduction

One of the most important tasks in any pattern recognition system is to find an informative low-dimensional feature representation, with enhanced discriminatory power, to overcome the so called curse of dimensionality [8]. This problem becomes more apparent when we deal with data formed using a number of signals that are collected from one or more channels, such as polysomnographic sleep data.

Various methods have been proposed for dimensionality reduction and feature extraction, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). PCA aims at preserving the global structure and works by projecting the data along the directions of maximal variances. The basis functions obtained by PCA are the eigenvectors of the data covariance matrix. LDA on the other hand, is most suitable for classification problems as it projects the data along the directions that maximize the ratio of the between class scatter matrix to the within class scatter matrix of the projected data.

Although both PCA and LDA were extensively used in the literature, a common factor that may limit their performance in certain applications is that both methods can only see the global Euclidean structure of data. In order to capture the local manifold structure many successful attempts were proposed in the literature including Neighborhood Preserving Embedding (NPE) [10], Locality Sensitive Discriminant Analysis (LSDA) [1], and Local Fisher Discriminant Analysis (LFDA) [6]. However, most biosignals generated by the human body tend to present patterns that are fuzzy in their nature, i.e., the samples belong to different classes with certain membership degrees. This may limit the performance of biosignal driven systems that utilize such methods.

A variation of Fisher's classical LDA is the Fuzzy Linear Discriminant Analysis (FLDA) [9]. There have been few attempts in the literature to propose variations to the original FLDA. Examples include the work proposed by Chen et al [2] and Kwak et al [4], where the latter is a fuzzy variation to the classical subspace LDA [7] utilizing PCA as a preprocessing step. Although such methods take the fuzzy nature of the data samples into consideration, but like the classical versions of LDA they only focus on preserving the global Euclidean structure.

Given the particular problem of sleep-scoring, it is quite common to have unbalanced distribution of the sleep stages with frequent changes from one sleep stage to another. This is particularly true with data collected from intensive care unit (ICU) patients. In such a case, adopting a feature projection method that maximizes the margins between the features belonging to different stages at each local area, while considering their the fuzzy nature of sleep patterns, may lead to better sepa-

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.49

ration of the different sleep stages. In order to bridge the gap between locality preserving and fuzzy discriminant analysis, a new method termed as Orthogonal Locality Sensitive Fuzzy Discriminant Analysis (OLSFDA) is presented in this paper. It is based on new derivations for the within class and between class scatter matrices. It also requires the basis functions to be orthogonal, which is a desired property in many applications.

The structure of this paper is as follows: Section 2 presents the proposed OLSFDA method. Section 3 presents the experimental results, and finally a conclusion is given in Section 4.

2 Orthogonal Locality Sensitive Fuzzy Discriminant Analysis (OLSFDA)

The first step in the proposed method employs PCA as a preprocessing step. The goal here is to remove any possible redundancy that can make S_W singular before starting the discriminant analysis. In such a step, we simply keep all the principal components to avoid any information loss. In the PCA stage, the transformation matrix related to PCA denoted as G_{PCA} is obtained by applying the eigen decomposition on the covariance matrix of the input. For the fuzzy locality preserving discriminant analysis stage, the first task is to compute the membership values of the samples in all of the classes.

Given the universal set X = $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_l\}$, where $\mathbf{x}_k$ is a feature vector, $k = 1, 2, ..., l$ is the number of samples. For simplicity, it will be useful to describe the membership degree that the k^{th} data point has in the i^{th} class with the following notation

$$\mu_{ik} = \mu_i(\mathbf{x}_k) \in [0, 1] \tag{1}$$

In order to compute the above membership we start first by computing the distance between each sample point and its K nearest neighbors (K is different from the sample index k). The membership of sample k in class i is computed by dividing the sum of the exponential distance between sample k and its K_1 nearest neighbors ($K_1 \leq K$) that belong to class i divided by the sum of the exponential distance between sample k and all of its K neighbors, that is:

$$\mu_{ik} = \frac{\sum_{p_1=1}^{K_1} exp(-g\,|\mathbf{x}_k - \mathbf{x}_{p_1}|)}{\sum_{p_2=1}^{K} exp(-g\,|\mathbf{x}_k - \mathbf{x}_{p_2}|)} \tag{2}$$

where g is a suitably chosen positive constant. The description of the proposed OLSFDA proceeds with the fuzzy total scatter matrix given as:

$$S_T = \sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik} \left(\mathbf{x}_k - \overline{\mathbf{x}}\right) \left(\mathbf{x}_k - \overline{\mathbf{x}}\right)^T \tag{3}$$

where μ_{ik} is the membership of pattern k in class i, $\mathbf{x}_k$ is the k^{th} sample, c is the number of classes, l_i is the number of training samples for class i, $\overline{\mathbf{x}}$ is the mean of the training samples.

$$\overline{\mathbf{x}} = \frac{\sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik} \mathbf{x}_k}{\sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik}} \tag{4}$$

We proceed with the modification of the total scatter matrix starting from Eq.3 as follows

$$S_T = \sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik} \left(\mathbf{x}_k \mathbf{x}_k^T - \overline{\mathbf{x}} \mathbf{x}_k^T - \mathbf{x}_k \overline{\mathbf{x}}^T + \overline{\mathbf{x}} \overline{\mathbf{x}}^T\right) \tag{5}$$

$$\begin{aligned} S_T = \quad & \sum_{i=1}^{c} \left[\sum_{k=1}^{l_i} \mu_{ik} \mathbf{x}_k \mathbf{x}_k^T - \overline{\mathbf{x}} \sum_{k=1}^{l_i} \mu_{ik} \mathbf{x}_k^T \right. \\ & \left. - \overline{\mathbf{x}}^T \sum_{k=1}^{l_i} \mu_{ik} \mathbf{x}_k + \overline{\mathbf{x}} \overline{\mathbf{x}}^T \sum_{k=1}^{l_i} \mu_{ik} \right] \end{aligned} \tag{6}$$

Using Eq.4 we replace $\sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik} \mathbf{x}_k^T$ with $\overline{\mathbf{x}}^T \sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik}$, and $\sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik} \mathbf{x}_k$ with $\overline{\mathbf{x}} \sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik}$. Thus Eq.6 can be re-written as

$$\begin{aligned} S_T = \quad & \sum_{i=1}^{c} \left[\sum_{k=1}^{l_i} \mu_{ik} \mathbf{x}_k \mathbf{x}_k^T - \overline{\mathbf{x}} \overline{\mathbf{x}}^T \sum_{k=1}^{l_i} \mu_{ik} - \right. \\ & \left. \overline{\mathbf{x}} \overline{\mathbf{x}}^T \sum_{k=1}^{l_i} \mu_{ik} + \overline{\mathbf{x}} \overline{\mathbf{x}}^T \sum_{k=1}^{l_i} \mu_{ik} \right] \end{aligned} \tag{7}$$

which in turn simplifies to

$$S_T = \sum_{i=1}^{c} \left[\sum_{k=1}^{l_i} \mu_{ik} \mathbf{x}_k \mathbf{x}_k^T - \overline{\mathbf{x}} \overline{\mathbf{x}}^T \sum_{k=1}^{l_i} \mu_{ik} \right] \tag{8}$$

Using Eq.4 again we replace both $\overline{\mathbf{x}}$ and $\overline{\mathbf{x}}^T$ with their equivalent, thus the above equation turns into

$$\begin{aligned} S_T = \quad & \left[\sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik} \mathbf{x}_k \mathbf{x}_k^T - \left(\frac{\sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik} \mathbf{x}_k}{\sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik}} \right) \right. \\ & \left. \left(\frac{\sum_{p=1}^{c} \sum_{j=1}^{l_p} \mu_{pj} \mathbf{x}_j^T}{\sum_{p=1}^{c} \sum_{j=1}^{l_p} \mu_{pj}} \right) \sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik} \right] \end{aligned} \tag{9}$$

In order to simplify the above equation, we provide the following definition

Definition-1: N is the fuzzy amount of elements in all fuzzy classes, and this is given by:

$$N = \sum_{i=1}^{c} \sum_{k=1}^{l_i} \mu_{ik}, \tag{10}$$

Then by using some of the mathematical series identities Eq.9 can be proved to be equal to

$$S_T = \frac{1}{2N} \sum_{i=1}^{c} \sum_{k=1}^{l_i} \sum_{p=1}^{c} \sum_{j=1}^{l_p} \mu_{ik}\mu_{pj} \left(\mathbf{x}_k - \mathbf{x}_j\right) \left(\mathbf{x}_k - \mathbf{x}_j\right)^T \tag{11}$$

Given that the total scatter matrix is equal to the summation of the within class scatter matrix and the between class scatter matrix, that is

$$S_T = S_B + S_W \tag{12}$$

Thus, Eq. 11 can be decomposed into the following two equations

$$S_W = \frac{1}{2N} \sum_{i=p=1}^{c} \sum_{k=1}^{l_i} \sum_{j=1}^{l_p} \mu_{ik}\mu_{pj} \left(\mathbf{x}_k - \mathbf{x}_j\right) \left(\mathbf{x}_k - \mathbf{x}_j\right)^T \tag{13}$$

$$S_B = \frac{1}{2N} \sum_{i=1}^{c} \sum_{\substack{p=1 \\ p \neq i}}^{c} \sum_{k=1}^{l_i} \sum_{j=1}^{l_p} \mu_{ik}\mu_{pj} \left(\mathbf{x}_k - \mathbf{x}_j\right) \left(\mathbf{x}_k - \mathbf{x}_j\right)^T \tag{14}$$

Then we can get the transformation matrix G_{LSFDA} without orthogonalization property by

$$\mathbf{G}_{LSFDA} = arg \max_{G} trace \left(\frac{\mathbf{G}^T S_B \mathbf{G}}{\mathbf{G}^T S_W \mathbf{G}} \right), \tag{15}$$

The complete transformation matrix $\mathbf{G}$ is then computed by, $\mathbf{G} = \mathbf{G}_{LSFDA}\mathbf{G}_{PCA}$. Then a QR-decomposition is applied on the resultant matrix to acquire a new transformation matrix $\mathbf{Q}$, that is $\mathbf{G} = \mathbf{QR}$. In such an equation, $\mathbf{R}$ is an upper triangular matrix and $\mathbf{Q}$ is an orthogonal matrix, i.e., one satisfying $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$, where $\mathbf{Q}^T$ is the transpose of $\mathbf{Q}$ and $\mathbf{I}$ is the identity matrix. Finally, the resultant transformation matrix for the proposed two stages algorithm is $\mathbf{G}_{OLSFDA} = \mathbf{Q}$.

3 Practical Experiments and Results

Conventional manual sleep stage scoring is a rule based art that uses the electric potentials generated by the human body including the brain waves measured by means of Electroencephalogram (EEG) signal, Electrooculography (EOG) of eye movements, and Electromyography (EMG) of facial muscle activity. Normal human sleep is comprised of two distinct states known as rapid eye movement (REM) and non-rapid eye movement (NREM) sleep. NREM sleep is subdivided into four stages: stage 1, stage 2, stage 3, and stage 4. Sleep states and stages are defined using the Rechtschaffen and Kales (R&K) criteria [5] that describe waveform configurations and frequencies over 30 second intervals using EEG/EMG/EOG signals in a study known as Polysomnography (PSG). Unfortunately, PSG analysis is time consuming and cost ineffective as it requires a technician to visually inspect the recorded EEG/EMG/EOG signals to decided on the occurrences of the different sleep stages. Thus, an automated sleep stage scoring system is highly desirable. To achieve such a system, the proposed OLSFDA is employed within a pattern recognition framework to extract the most important small feature subset (thus reducing the computational cost) for such a system.

The sleep datasets utilized in this paper were collected from nine subjects in the ICU in the hospital in Australia. The collection period was about 24 hours during which two EEG channels (C3 and C4), two EOG channels (both eyes) and one EMG channel (Masseter muscle) were utilized. The collected signals were sampled at either 128 Hz or 256 Hz. These signals were used for automatic sleep scoring, then compared to scoring by a sleep expert of 30 second epochs according to R&K rules. A sliding window approach was utilized from which features were extracted, with windows size of 15 sec and an increment of 7.5 seconds.

Since there is no agreement in the literature on a suitable feature set to represent these signals, then an ensemble of features was extracted. Specifically, the following features were extracted from the EOG/EMG signals: AutoRegressive (AR) model parameters (10 features), integral absolute value (1 feature), mean absolute value (1 feature), root mean square (1 feature), skewness (1 feature), waveform length (1 feature), number of zero crossings (1 feature), the energy of the signal (1 feature). On the other hand, only the following features were extracted from each of the two EEG signals: AutoRegressive (AR) model parameters (10 features), Hjroth parameters described in [3] (6 features), and the energy of wavelet coefficient in the following frequency bands: 0-4 Hz, 4-8 Hz, $\cdots$, 60-64 Hz, 0-8 Hz, 8-16 Hz, $\cdots$, 56-64 Hz, 0-16 Hz, 16-32 Hz, $\cdots$, 48-64 Hz, 0-32 Hz and 32-64 Hz.

The classification accuracy results were computed using different feature projection methods including:

FLDA [4], NPE [10], PCA [8], LSDA [1], and LFDA [6] in comparison to the proposed OLSFDA and the Baseline (when using the total feature set with 140 features without any dimensionality reduction). The number of extracted features was set to 5 only for OLSFDA, NPE, PCA, LSDA, and LFDA (as the rank of the scatter matrix is not limited to c - 1 like classical LDA), while the default for FLDA is c - 1, where c is the number of classes.

Samples from each subject were first randomized and then divided into a training set comprising 40% of the total samples and a testing set comprising the remaining 60%. A linear support vector machine classifier (acquired from `http://www.csie.ntu.edu.tw/~cjlin/liblinear/`) was utilized to compute the classification accuracy. The whole process was repeated 15 times and the average classification accuracies are presented for all of the methods as a boxplot in Figure 1. These results indicate the significance of the OLSFDA performance in comparison to all other methods. One possible justification for the performance of the NPE, PCA, LSDA, and LFDA is that such methods require more features to present better classification accuracies and that five features only were not enough for NPE, PCA, LSDA, and LFDA to achieve powerful results. On the other hand, both OLSFDA and FLDA were capable of achieving better results than NPE, PCA, LSDA, and LFDA. This may be justified by the fact that they share many properties. Both methods utilize PCA as a preprocessing step (that we modified for FLDA to include all principal components instead of a certain ratio only to present comparable performance to OLSFDA). Additionally, both methods take into consideration the membership of the samples in all classes. On the other hand, the proposed OLSFDA exhibited better performance than FLDA due to its locality preserving nature imposed by the new scatter matrices.

4 Conclusion

A new feature projection method termed OLSFDA is presented. OLSFDA is based on a mixture of the concepts of locality preserving and fuzzy discriminant analysis. Mathematical derivations of the proposed method were provided. In order to compare the performance of OLSFDA with a number of well-known dimensionality reduction methods, the challenging problem of sleep-stage scoring was considered. Results indicated the good performance of OLSFDA, which managed to achieve a classification accuracy of 90.7% across nine subjects.

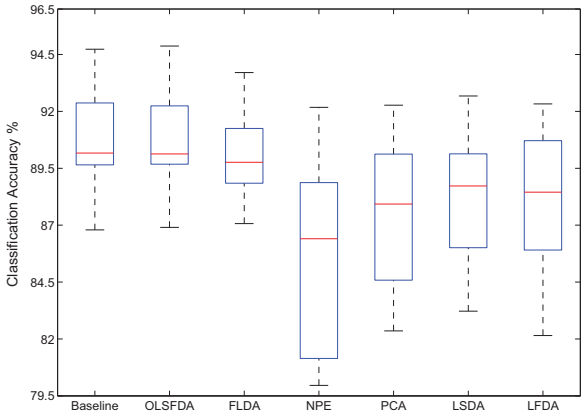

Figure 1. Box plot of the classification accuracy results

References

[1] D. Cai, X. He, K. Zhou, J. Han, and H. Bao. Locality sensitive discriminant analysis. In *Proceeding of the International Joint Conference on Artificial Intelligence (IJCAI'07), Hyderabad, India*, 2007.

[2] Z. P. Chen, J. H. Jiang, Y. Li, Y. Z. Liang, and R. Q. Yu. Fuzzy linear discriminant analysis for chemical data sets. *Chemometrics and Intelligent Laboratory Systems*, 45:295–302, 1999.

[3] B. Hjorth. Eeg analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29:306–310, 1970.

[4] K. C. Kwak and W. Pedrycz. Face recognition using a fuzzy Fisherface classifier. *Pattern Recognition*, 38:1717–1732, 2005.

[5] A. Rechtschaffen and A. Kales. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Washington Public Health Service US Government Printing Office 1968.

[6] M. Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of Machine Learning Research*, 8:1027–1061, 2007.

[7] D. L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):831–836, 1996.

[8] S. Theodoridis and K. Koutroumbas. *Pattern Recognition, 3rd ed.* Academic Press, San Diego, CA, USA, 2006.

[9] J. Watada, H. Tanaka, and K. Asai. Fuzzy discriminant analysis in fuzzy groups. *Fuzzy Sets and Systems*, 19:261–271, 1986.

[10] H. Xiaofei, C. Deng, Y. Shuicheng, and H. J. Zhang. Neighborhood preserving embedding. In *Tenth IEEE International Conference on Computer Vision (ICCV'2005), pages 1208-1213*, 2005.