

Dimensionality Reduction Techniques for Processing Epileptic Encephalographic Signals

R.HARIKUMAR* and P.SUNIL KUMAR

Department of ECE, Bannari Amman Institute of Technology, India.

DOI: <http://dx.doi.org/10.13005/bpj/587>

(Received: February 03, 2015; accepted: March 19, 2015)

ABSTRACT

Epilepsy is a chronic neurological disorder of the brain, approximately 1% of the world population suffers from epilepsy. Epilepsy is characterized by recurrent seizures that cause rapid but reversible changes in the brain functions. Temporary electrical interference of the brain roots epileptic seizures. The occurrence of an epileptic seizure appears unpredictable. Various methods have been proposed for dimensionality reduction and feature extraction, such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Singular Value Decomposition (SVD). The advantages of the regularization dimension are that it is more precise than other approximation methods and it is easy to derive an estimator in the presence of noise due to the fully analytical definition. This paper gives an overall review of the dimensionality reduction techniques suitable for the application of electroencephalography signals from epileptic patients.

Key words: Epilepsy, Dimensionality, PCA, ICA, SVD

INTRODUCTION

One of the most important tasks in any pattern recognition system is to find an informative, yet small, subset of features with enhanced discriminatory power¹. Techniques that can introduce low-dimensional feature representation with enhanced discriminatory power are of paramount importance, because of the so called curse of dimensionality. Feature extraction consists of finding a set of measurements or a block of information with the objective of describing in a clear way the data or an event present in a signal. These measurements or features are the fundamental basis for detection, classification or regression tasks in biomedical signal processing and is one of the key steps in the data analysis process. Features constitute a new form of expressing the data, and can be binary, categorical or continuous: they represent attributes or direct measurements of the signal. For example, features may be the age, health status of the patient, family history, electrode

position or EEG signal descriptors (amplitude, voltage, phase, frequency, etc). The aim of extracting features is to identify "patterns" of brain activity: features can be used as input to a classifier². The performance of a pattern recognition system depends on both the features and the classification algorithm employed. More formally, feature extraction assumes there are N samples and D features, for a $N \times D$ data matrix. It is also possible to obtain a feature vector at the sample n form the feature matrix, that is, x is a uni dimensional vector $x=[x_1, x_2, \dots, x_D]$ called as pattern vector.

Our mental representation of the world is formed by processing large numbers of sensory inputs – including, for example, the pixel intensities of images, the power spectra of sound, and the joint angles of articulated bodies. While complex stimuli of this form can be represented by points in a high-dimensional vector space, they typically have a much more compact description. Coherent structure in the world leads to strong correlations

between inputs, generating observations that lie on or close to a smooth low-dimensional manifold. To compare and classify such observations- in effect, to reason about the world- depend crucially on modelling the nonlinear geometry of these low-dimensional manifolds.

Suppose the data lie on or near a smooth non-linear manifold of lower dimensionality $d \ll D$. To a good approximation then, there exists a linear mapping- consisting of a translation, rotation, and rescaling-that maps the high-dimensional coordinates of each neighbourhood to global internal coordinates on the manifold³. By design, the reconstruction weights W_{ij} reflect intrinsic geometric properties of the data that are invariant to exactly such transformations. Therefore expect their characterization of local geometry in the original data space is expected to be equally valid for local patches on the manifold. In particular, the same weights W_{ij} that reconstruct the i th data point in D dimensions should also reconstruct its embedded manifold coordinates in d dimensions.

Suitable Dimensionality Reduction Techniques **Principal Component Analysis**

It is a second order method and it is the best in the mean-square error sense. In different domains, it is called as Singular Value Decomposition (SVD), empirical orthogonal function (EOF), Hotelling transform and the Karhunen-Loeve transform. By finding a few orthogonal and linear combinations of the original variables, PCA finds to reduce the data dimension. In the field of engineering mathematics and statistics, PCA has been successfully applied in a variety of domains. For high dimensional data, the computation of the Eigen vectors might be infeasible since the covariance matrix is proportional to the data point's dimensionality.

Kernel PCA

In high dimensional data, PCA is used for the design and modeling of the linear variabilities. But the characteristic nature of the high dimensional data set is that it has a non-linear nature. In such cases PCA cannot determine the variability of data accurately. To address the problem of non-linear dimensionality reduction kernel PCA can be used. With the usage of kernels, some non

linear mapping is done so that the principal components can be computed efficiently in high – dimensional feature spaces. Kernel PCA finds the principal components where the latest structure is low dimensional and is easier to discover.

Sammon Map

It is a typical example of a perfect non-linear feature extraction technique. It is widely used for the 2-D visualization of data which has quite a high dimension. In order to minimize the error function, Sammon's algorithm uses gradient descent technique.

Multidimensional Scaling (MDS)

For the easy analysis of proximity data on a moderate set of stimuli, Multidimensional scaling can be used. It is used to express or reveal the hidden structure which underlies in the data. Here the points are mostly placed in a very low dimensional space and it is widely considered as an exploratory data analysis technique. The visualization of the structures in the data can be easily learnt by the researchers if the representation of the patterns of proximity is done in two or three dimensions. Multidimensional scaling is mostly classified into two types namely MDS analysis of interval of ratio data and MDS analysis of ordinal data.

Locally Linear Embedding

One of the widely used approaches for the problems addressing to the non-linear dimensionality reduction is Locally Linear Embedding. It is done by the computation of low-dimensional data which preserves the neighborhood embedding of high dimensional data⁴. The main aim here is to map the high-dimensionality regional data points to a single coordinate system which is assigned to a global position.

Isomap

It is a non-linear dimensionality reduction technique based on the Multi-Dimensional technique. The classical MDS can be represented in a generalized non-linear manner and that is called as an Isomap. In the input space, MDS is not performed, but it is performed at the non linear data manifold's geodesic space. Isomap does not apply

the classic MDS to the straight line distances. It always applies to the geodesic distances in order to find a low-dimensional mapping that preserves those pair wise distances.

Factor Analysis

It is a statistical technique which is closely related to the Principal Component Analysis (PCA) and dimensionality reduction. It is basically a generative model. From a set of latents, the observed data has been produced and through the analysis of mathematical equations the unobserved variables has been produced. The factors always follow a multivariate normal distribution and it is assumed to be more uncorrelated to noise.

Generalized SVD

It is a very versatile tool because it can be particularized to correspondence analysis under the appropriate choices of the constrain matrices. It can also be particularized to discriminant analysis and canonical correlation analysis. Generalized SVD performs a similar decomposition and it modifies or relaxes the conditions of orthogonality according the two constrain metrics.

Maximum Variance Unfolding

It is more similar to Isomap in a respect when upon the data , a neighbourhood graph is defined and the pairwise distances in the resulting graph is retained. Here the Euclidean distances are maximized between the data points under the condition that the corresponding distances in the neighbourhood graph is left unchanged. Then as a result of it, the optimization problem comes into existences and then it can be solved efficiently using semi definite programming.

Diffusion Maps

It is originating from the field of system dynamics. On the graph of a particular data, a Markov random walk is defined upon it. By performing the random walk for a number of times the data points proximity is obtained. Using this type of means only diffusion maps are defined. If the data is represented in a low dimension then the pair wise diffusion distances are retained there itself.

Independent Component Analysis (ICA)

ICA is a variant of principal component analysis (PCA) in which the components are assumed to be mutually statistically independent instead of merely uncorrelated. It is a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals. Independent Component Analysis describes a model for multivariate data describing large database of samples. The variables in the model are assumed non-Gaussian and mutually independent and they are called the independent components of the observed data.

Projection Pursuit

Unlike the PCA and FA, it is a complete linear method and it can easily incorporate higher than second order information. Projection Pursuit is used mostly for the analysis of non-Gaussian datasets. More than second order methods it is more computationally intensive. Any aspect of non-Gaussianity is usually measured through projection pursuits only. A projection index always defines the interestingness of a direction and looks for the directions that optimize that index.

Multi-unit Objective functions

To specify the objective functions there are many different ways. In spite of their different formulations they all are closely related and they even become equivalent in some conditions. Under certain conditions the non-linear PCA method is applicable where the mutual information method is most equivalent to maximum likelihood principle. Some of the most important Multi-unit Objective functions are Maximum likelihood and network entropy, Mutual information and Kullback-Leibler divergence, Non-linear cross-correlations, Non-linear PCA and Higher-order cumulant tensors.

Multi-layer autoencoders

They are feed-forward neural networks with hidden layers usually having an odd number. The network is trained well for the minimization process of the mean squared error between the input and the output. If the activation functions are linear and it is used in the neural network, then the performance of autoencoder will be more similar

to PCA⁵. To learn a non-linear mapping between the high dimensional and low dimensional data representation, the auto encoders use sigmoid activation functions. They have been used successfully for the application of problems related to the missing data and HIV analysis.

Laplacian Eigenmaps

It is more similar to LLE technique where the preservation of the local properties of the manifold is given a prior importance and thus the Eigen maps can easily find the low dimensional data representation. Between near neighbours the local properties are based on the pair wise distances. They always compute a low-dimensional representation of a particular data in which the distances between a datapoint and its nearest neighbors are minimized.

CONCLUSION

EEG signals are multidimensional, non stationary, and time domain biological signals, which are not reproducible. It is supposed to contain information about what is going on in the ensemble of excitatory pyramidal neuron level, at millisecond temporal resolution scale. Since scalp EEG contains considerable amount of noise and artifacts, and exactly where it is coming from is poorly determined, extracting information from it is extremely challenging. So far the two major paradigms used to understand scalp EEG are segregation and integration, both of which are computation intensive. The current explosion of interest in Epilepsy detection, on the other hand, underscores the need of online processing. This is compelling reason for the popularity of soft computing algorithms in human scalp EEG processing. This paper surveys the probable dimensionality reduction techniques for the processing of electroencephalography signals from epileptic patients.

REFERENCES

1. Adeli, H, Zhou, Z & Dadmehr, N, 'Analysis of EEG records in an epileptic patient using wavelet transform', *J. Neurosci. Methods*, **123**(1): 69-87 (2003).
2. Leon, D, Iasemidis, Deng-Shen Shiau, Chao Valit Wongse W, Sacellares, JC & Pardalos, PN, 'Adaptive Epileptic Seizure Prediction System', *IEEE Transactions on Biomedical Engineering*, **50**(5): 616-627 (2003).
3. Rabiner, R, 'Hidden Markov Models and Selected Application in Speech Processing", *proceedings of IEEE*, **77**(2): 20-42 (1989).
4. Schuyler, R, White, A, Staley, K & Cios, KJ, 'Epileptic seizure detection', *IEEE Eng. Med. Biol. Mag.*,: 74-81 (2007).
5. Subasi, A, 'EEG signal classification using wavelet feature extraction and a mixture of expert model', *Expert Systems with Applications*, **32**: 1084-1093 (2007).