Neurocomputing 179 (2016) 152–160

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Kernel learning over the manifold of symmetric positive definite matrices for dimensionality reduction in a BCI application

Khadijeh Sadatnejad, Saeed Shiry Ghidary*

Computer Engineering and Information Technology Department, Amirkabir University of Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 19 April 2015
Received in revised form
26 September 2015
Accepted 22 November 2015
Communicated by W.Wu
Available online 17 December 2015

Keywords:
Brain computer interface
Nonlinear dimensionality reduction
Kernel learning
Riemannian geometry.

ABSTRACT

In this paper, we propose a kernel for nonlinear dimensionality reduction over the manifold of Symmetric Positive Definite (SPD) matrices in a Motor Imagery (MI)-based Brain Computer Interface (BCI) application. The proposed kernel, which is based on Riemannian geometry, tries to preserve the topology of data points in the feature space. Topology preservation is the main challenge in nonlinear dimensionality reduction (NLDR). Our main idea is to decrease the non-Euclidean characteristics of the manifold by modifying the volume elements. We apply a conformal transform over data-dependent isometric mapping to reduce the negative eigen fraction to learn a data dependent kernel over the Riemannian manifolds. Multiple experiments were carried out using the proposed kernel for a dimensionality reduction of SPD matrices that describe the EEG signals of dataset IIa from BCI competition IV. The experiments show that this kernel adapts to the input data and leads to promising results in comparison with the most popular manifold learning methods and the Common Spatial Pattern (CSP) technique as a reference algorithm in BCI competitions. The proposed kernel is strong, particularly in the cases where data points have a complex and nonlinear separable distribution.

1. Introduction

In brain–computer Interface systems that use motor imagery, brain activity is usually captured in the form of EEG signals and is transferred to an external device [27]. Extracting information from EEG signals is carried out by using different pattern recognition methods involving feature extraction, dimensionality reduction, and classification [32,49,51] to ultimately determine the user's mental state [28,36].

Several techniques are available for extracting features from EEG signals [25,4,7,8]. A common spatial pattern algorithm [38,7] and a spatial covariance matrix of a signal [4,5,9] are two major approaches to represent EEG signals in BCI applications. CSP can be considered to be a dimensionality reduction technique that learns spatial filters that maximize class separability. A spatial covariance matrix of the EEG signal, which lies in the space of symmetric positive definite matrices, can be formulated as a connected Riemannian manifold [2]. In recent years, methods using a spatial covariance matrix have attracted considerable attention [10,4,5,9].

In BCI application, samples are usually represented by large feature vectors. Therefore, these problems suffer from the curse of dimensionality [28]. Different research efforts have attempted to overcome the problem of the curse of dimensionality in the BCI literature. Zhang et al. [51] introduced Spatial-Temporal Discriminant Analysis (STDA) as a multiway extension of Linear Discriminant Analysis (LDA). They attempted to maximize the discrimination between two classes by finding two projections from the spatial and temporal information [52]. These projections reduce the dimensionality of the features that feed into the discriminant analysis. To overcome the problems of the curse of dimensionality and the bias-variance tradeoff for Event-Related Potential (ERP) classification in BCI applications, Zhang et al. [50] introduced Aggregation of Sparse Linear Discriminant Analysis (ASLDA). They introduced a sparse LDA to reduce the dimensionality. For this purpose, sparse discriminant vectors were learned by solving a l1-regularized Least Squares Regression (LSR). Sparse CSP that uses a linear combination of a subset of channels was introduced by Goksu et al. [20]. They proposed a generalized eigenvalue decomposition based on a greedy search to identify multiple sparse eigenvectors to compute spatial projections. They showed the effectiveness of the sparse CSP in comparison with the traditional CSP by examining the datasets in the BCI competition (2005). Wu et al. [46] used a statistical framework to provide a spatio-temporal representation of the EEG trials. They modeled the variance of source signals as random variables and proposed a

* Corresponding author. Tel.: +98 2164542737.
E-mail addresses: sadatnejad@aut.ac.ir (K. Sadatnejad), shiry@aut.ac.ir (S. Shiry Ghidary).

http://dx.doi.org/10.1016/j.neucom.2015.11.065
0925-2312/

hierarchical Bayesian model for retrieving the inter-trial variability of amplitude in a sparse way to provide a reduced representation of data [46].

In the case of representing EEG signals by spatial covariance matrices, although this representation reduces the length of the descriptors in comparison with the raw EEG, this reduction is not sufficient to overcome the curse of dimensionality. Dimensionality reduction over the space of SPD matrices by considering the Riemannian geometry of the SPD matrices has difficulties in comparison with treating the points as Euclidean objects [4]. Formulating covariance matrices as a connected Riemannian manifold [4] leads to a nonlinear relationship between observations and latent variables. Therefore, NLDR techniques are required to reduce the dimensionality over this manifold. Several techniques are adapted to the cases where the relationships between observations and latent variables are nonlinear [24]. Popular NLDR techniques, such as locally linear embedding (LLE) [39], local tangent space alignment (LTSA) [48], Laplacian Eigenmap (LE) [6], and Isomap [41], have been applied to the manifolds. However, these techniques all have shortcomings on the manifold of SPD matrices. These shortcomings stem from ignoring the geometrical structure of the manifold (i.e., living the manifold in the non-Euclidean space and performing computations by assuming that the data points are embedded in Euclidean space) [17].

In this paper, we attempt to overcome the curse of dimensionality in the SPD matrix space in BCI applications by learning a kernel that is adapted to the manifold by considering the Riemannian geometry of the manifolds. The main contribution of this paper is learning a kernel by minimizing a measure that shows the non-Euclidean characteristics of the manifold by changing the volume elements, while preserving the geometry, of the input space. This minimization is especially useful in the cases where the data points lie on a manifold with a nonzero intrinsic curvature. The proposed kernel, when applied in multi-dimensional scaling [21], provides an NLDR technique that is well adapted to the manifold of SPD matrices.

The rest of the paper is organized as follows. In Section 2, we describe mathematical preliminaries that are required for learning over Riemannian manifolds and understanding the proposed modifications in the feature space. Section 3 provides more details on learning a data-dependent kernel by preserving geometry. Section 4 reports our experiments on a BCI data set. Our findings are discussed in Section 5, and concluding notes are mentioned in Section 6.

2. Preliminaries

In this section we describe basic concepts of Riemannian geometry that are necessary to understand our proposed approach. We review the metric applied in the SPD matrix space, its associated log and exp map, and the kernel functions from a geometrical point of view [22,23].

2.1. Riemannian geometry

The Riemannian metric on the Riemannian manifolds is a positive definite metric that takes two tangent vectors as inputs and generates a real number, which is a generalization of the inner product, and allows the similarity or dissimilarity of two points on the manifold to be measured [13,16,45]. A common invariant Riemannian metric on the tangent space of the SPD matrices [15,33,34] is defined as

$$<y,z>_X = \mathrm{trace}(X^{-\frac{1}{2}} y X^{-1} z X^{-\frac{1}{2}}) \tag{1}$$

where X denotes a point on the manifold and y and z show tangent vectors in the tangent space formed at point X.

The length of the curves along the manifold is computed by integrating the metric tensor along the curve, which connects two points on the manifold [13,26]. The geodesic, which is the local distance-minimizing curve over the manifold of SPD matrices associated with a metric from Eq. (1), is computed as

$$d_G{}^2(X,Y) = <\log_X(Y), \log_X(Y)>_X = \mathrm{trace}\left(\left(\log^2\left(X^{-1/2} Y X^{-1/2}\right)\right)\right) \tag{2}$$

where X and Y are two points on the manifold, $\log_X(Y)$ is the Riemannian log map of point Y to the tangent space formed at point X, and d_G denotes the geodesic distance on the manifold of the SPD matrices [42]. The Riemannian log map projects a point on the manifold to a point in tangent space. Its inverse is Riemannian $\exp_X(y)$, which projects a tangent vector $y \in T_X M$ into a point Y on the manifold.

The Riemannian exponential and logarithmic mappings associated to the metric of Eq. (1) are defined as

$$\exp_X(y) = X^{1/2} \exp\left(X^{-\frac{1}{2}} y X^{-\frac{1}{2}}\right) X^{1/2} \tag{3}$$

$$\log_X(Y) = X^{1/2} \log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right) X^{1/2} \tag{4}$$

where exp and log are matrix exponential and logarithmic functions that are calculated as:

$$\exp \Sigma = \sum_{k=0}^{\infty} \frac{\Sigma^k}{k!} = U \exp(D) U^T, \Sigma = UDU^T$$
$$\log \Sigma = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{(\Sigma - I)^k}{k} = U \log(D) U^T,\ \Sigma = UDU^T \tag{5}$$

Eq. (5) assumes that Σ is decomposed into eigenvalues and vectors. Note that exp operator on the matrices always exists, while the log operator is defined only on symmetrical matrices with positive eigenvalues [15].

2.2. Kernel geometry

Kernel function $K(.,.)$ corresponds to the inner product in a high dimensional space H.

$$K(x,x') = \varphi(x).\varphi(x') \tag{6}$$

where φ is a projection of the input space S into the higher dimensional space H. The kernel function $K(.,.)$ induces a Riemannian metric to S using mapping φ, which is computed as [1,45]

$$g_{ij}(x,x') = \frac{\partial}{\partial x_i} \frac{\partial}{\partial x'_j} K(x,x')|_{x=x'} \tag{7}$$

where x_i denotes ith basis of vector x. Eq. (7) is written in Einstein summation notation. The volume element corresponding to the induced metric in input space is computed as [45]

$$dV = \sqrt{g(x)} dx_1 \dots dx_n \tag{8}$$

where $g(x)$ represents the determinant of the matrix whose elements are g_{ij} and dV denotes the volume element. The expression $\sqrt{g(x)}$ is a factor that controls the expansion and contraction of volume elements [44].

2.3. Kernel principal component analysis

Kernel Principal Component Analysis (KPCA) (Algorithm 1) [40], which is widely used in dimensionality reduction and denoising applications, is a nonlinear generalization of principal component analysis (PCA) [19]. Classical PCA is designed to reduce dimensionality in the cases where the manifold is linearly