Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

Deep Manifold Learning of Symmetric Positive Definite Matrices with Application to Face Recognition

Zhen Dong,[1] Su Jia,[2] Chi Zhang,[1] Mingtao Pei,[1] Yuwei Wu[1]*
1. Beijing Laboratory of Intelligent Information Technology,
School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
2. Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, USA
dongzhen@bit.edu.cn, su.jia@stonybrook.edu, {zhangchi, peimt, wuyuwei}@bit.edu.cn

Abstract

In this paper, we aim to construct a deep neural network which embeds high dimensional symmetric positive definite (SPD) matrices into a more discriminative low dimensional SPD manifold. To this end, we develop two types of basic layers: a 2D fully connected layer which reduces the dimensionality of the SPD matrices, and a symmetrically clean layer which achieves non-linear mapping. Specifically, we extend the classical fully connected layer such that it is suitable for SPD matrices, and we further show that SPD matrices with symmetric pair elements setting zero operations are still symmetric positive definite. Finally, we complete the construction of the deep neural network for SPD manifold learning by stacking the two layers. Experiments on several face datasets demonstrate the effectiveness of the proposed method.

Introduction

Symmetric positive definite (SPD) matrices have shown powerful representation abilities of encoding image and video information. In computer vision community, the SPD matrix representation has been widely employed in many applications, such as face recognition (Pang, Yuan, and Li 2008; Huang et al. 2015; Wu et al. 2015; Li et al. 2015), object recognition (Tuzel, Porikli, and Meer 2006; Jayasumana et al. 2013; Harandi, Salzmann, and Hartley 2014; Yin et al. 2016), action recognition (Harandi et al. 2016), and visual tracking (Wu et al. 2015).

The SPD matrices form a Riemannian manifold, where the Euclidean distance is no longer a suitable metric. Previous works on analyzing the SPD manifold mainly fall into two categories: the local approximation method and the kernel method, as shown in Figure 1(a). The local approximation method (Tuzel, Porikli, and Meer 2006; Sivalingam et al. 2009; Tosato et al. 2010; Carreira et al. 2012; Vemulapalli and Jacobs 2015) locally flattens the manifold and approximates the SPD matrix by a point of the tangent space. The kernel method (Harandi et al. 2012; Wang et al. 2012; Jayasumana et al. 2013; Li et al. 2013; Quang, San Biagio, and Murino 2014; Yin et al. 2016) embeds the manifold into a higher dimensional Reproducing Kernel Hilbert Space (RKHS) via kernel functions. On new

*corresponding author

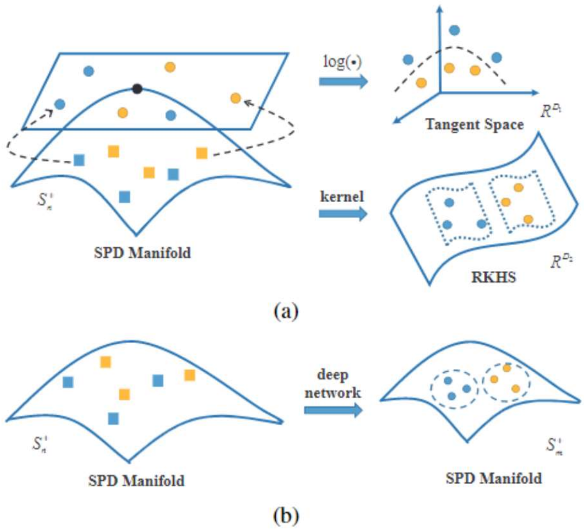

Figure 1: The comparison between our method and previous methods on analyzing the SPD manifold. (a) Previous methods either locally flatten the SPD manifold via tangent space approximation, or embed the manifold into a higher dimensional reproducing kernel Hilbert space. (b) Our method aims to find a non-linear mapping that projects high dimensional SPD matrices into a lower dimensional SPD manifold.

spaces, both methods convert the SPD matrix into a vector and learn a corresponding discriminative representation. However, both local approximation and kernel methods face two problems. First, the SPD matrices are high dimensional, which brings the problem of high computational cost. Second, the vectorization operation on SPD matrices might give rise to the distortion of the manifold geometrical structure.

To overcome the two problems mentioned above, we focus on learning a non-linear mapping which projects high dimensional SPD matrices to a low dimensional discriminative SPD manifold, as shown in Figure 1(b). Recently, the deep neural network has shown strong capability of describing complex non-linear maps and been successfully applied on many vision tasks, such as image classification (Krizhevsky, Sutskever, and Hinton 2012; Simonyan and Zisserman 2014; Szegedy et al. 2015; He et al. 2016) and face recognition (Sun, Wang, and Tang 2015; Taigman et al. 2015; Schroff, Kalenichenko, and Philbin 2015; Parkhi, Vedaldi, and Zisserman 2015). Motivated by these achievements of deep networks, we advocate modeling the non-linear mapping which reduces the dimensionality of high dimensional SPD matrices via a deep neural network.

To achieve this goal, two key issues need to be addressed: dimension reduction and non-linear operation. We introduce two basic layers, *i.e.*, the 2D fully connected layer and the symmetrically clean layer, to realize dimension reduction and non-linear operation, respectively. The 2D fully connected layer reduces the dimensionality of the SPD matrices via a linear mapping, and the symmetrically clean layer sets the symmetric pairs of elements in the SPD matrix as zeros to add non-linearity to the mapping. The two layers should ensure that the output matrices are symmetric positive definite. We thus provide the necessary and sufficient condition for the 2D fully connected layer, and prove that the symmetrically clean layer keeps the symmetry and positive definite properties of SPD matrices. Based on the two layers, the deep neural network for SPD manifold learning is constructed and evaluated on the face recognition tasks. Our network has several advantages compared with the traditional methods on analyzing the SPD manifold. First, learning discriminative representations in new learned low dimensional SPD space brings low computational cost. Second, our method works on the original SPD matrix instead of the vectorization form, which makes full use of the manifold geometrical structure.

This work is, to the best of our knowledge, the first to exploit the deep neural network to analyze the SPD manifold. The contributions of the paper are two-fold: (1) We propose an non-linear operation on the SPD manifold, and prove that SPD matrices with symmetrically clean operation are still symmetric positive definite. (2) The proposed deep neural network is able to project high dimensional SPD matrices to a low dimensional discriminative SPD manifold, and achieves good performances on the face recognition task.

Related Work

In this section, we briefly review several SPD manifold related work including two aspects: SPD manifold metrics and representative work of learning discriminative functions by these metrics.

Let's define the manifold of $n \times n$ SPD matrices as $\mathbb{S}_n^+$. The SPD matrix to the matrix space is similar as positive number to the real number space. A straightforward metric is the Frobenius norm between SPD matrices which is an extension of the Euclidean measure, but several undesirable effects may occur since the Frobenius norm ignores the manifold geometrical structure, such as the swelling of diffusion tensors (Arsigny et al. 2006; Pennec, Fillard, and Ayache 2006). To overcome the problem, several metrics on Riemannian manifold are introduced. The Affine Invariant Metric (AIM) proposed by (Pennec, Fillard, and Ayache 2006) is defined as

$$\begin{aligned}\delta_A(\boldsymbol{A},\boldsymbol{B}) &= \|\log(\boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2})\|_F \\ &= \Big(\sum_{i=1}^{n}(\log\lambda_i)^2\Big)^{1/2},\end{aligned} \tag{1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix, $\log(\cdot)$ is the matrix logarithm operator, $\boldsymbol{A},\boldsymbol{B} \in \mathbb{S}_n^+$ and λ_i is the generalized eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{B}$, *i.e.*, $\det(\lambda_i\boldsymbol{A}-\boldsymbol{B}) = 0$. Although the AIM is invariant to affine transformations, it is a high computational burden in practice (Arsigny et al. 2007). To reduce the computation cost, the Stein Metric (SM) is studied and introduced by Sra (2012):

$$\delta_S(\boldsymbol{A},\boldsymbol{B}) = \log\det\Big(\frac{\boldsymbol{A}+\boldsymbol{B}}{2}\Big) - \frac{1}{2}\log\det(\boldsymbol{A}\boldsymbol{B}). \tag{2}$$

The δ_S has several similar properties as δ_A and is less expensive to compute (Cherian et al. 2013). Furthermore, Harandi, Salzmann, and Hartley (2014) proved that the length of any curve is the same under δ_S and δ_A up to a scale of $2\sqrt{2}$. Another metric on $\mathbb{S}_n^+$ is the Log-Euclidean Metric (LEM) which is considered by endowing the SPD manifold a Lie group structure (Arsigny et al. 2006; 2007). The LEM is given by

$$\delta_L(\boldsymbol{A},\boldsymbol{B}) = \|\log(\boldsymbol{A}) - \log(\boldsymbol{B})\|_F. \tag{3}$$

Different from δ_A and δ_S, δ_L is a bi-invariance metric, *i.e.*, $\delta_L(\boldsymbol{A},\boldsymbol{B}) = \delta_L(\boldsymbol{B},\boldsymbol{A})$. Since LEM only needs matrix logarithm and Euclidean operations, its computation cost is much less than the AIM and the SM.

Based on these metrics, a few works are proposed to learn discriminative functions on the SPD manifold. One representative work is (Vemulapalli and Jacobs 2015). Their work first flattens the manifold by projecting SPD matrixes to the tangent space at the point of the identity matrix with the matrix logarithm operator $\log(\cdot)$ for local approximation, and then performs the information theoretic metric learning method for the corresponding vectors of the points on the tangent space. They further conduct experiments on face and object datasets, and obtain good performances.

To consider the local manifold structure of manifold data points, the kernel method embeds the SPD manifold into a higher dimensional RKHS via a kernel function and learns discriminative functions on the new space. Based on the LEM, the Covariance Discriminative Learning (CDL) (Wang et al. 2012) employs a new kernel function and conducts partial least squares or linear discriminant analysis in the new space. Besides, many attempts focus on sparse representation and dictionary learning on SPD matrix with appropriate kernels, such as Riemannian Sparse Representation (RSR) (Harandi et al. 2012; 2016) and online dictionary learning (Zhang et al. 2015) based on the SM, and Log-Euclidean Kernel (LEK) (Li et al. 2013) and Manifold Kernel Sparse Representation (MKSR) (Wu et al. 2015) based on the LEM. Yin et al. (2016) further proposed a sparse subspace clustering method for the SPD manifold via an LEM based kernel.

To handle the problems of high computation cost and manifold geometrical structure distortion which methods