Deep Manifold Learning of SPD Matrices

IZ Raad

2024-03-25

Symmetric positive definite (SPD) matrices are incredibly powerful tools in encoding visual information. Well established applications of this class of matrices include facial and action recognition, as well as visual tracking. However, current methods of data classification utilizing SPD manifolds have two major drawbacks: they have a high computation cost and risk distortion of the manifold's geometric structure.

SPD matrices form a Reimannian manifold, defined as a smooth topological space paired with a Riemannian metric that allows for maintenance of geometric notions of distance and angle. There are two methods through which these are typically analyzed: local approximation and kernel methods. Local approximation takes advantage of the locally Euclidean nature of manifolds. This can be thought of as "flattening" a neighborhood about an element of our manifold to a real vector space, tangent at this element. The kernel method embeds our manifold into a higher dimensional Hilbert space via kernel functions. Both methods vectorize our SPD matrices in order to gain a discriminative representation. The high computational cost of both methods is traced back to the fact that the SPD matrices themselves are often high dimensional. The conversion of SPD matrices to vectors gives rise to our second major concern surrounding the integrity of the manifold's geometric structure.

To mitigate these shortcomings of SPD matrices, Dong et al suggest an alternative method: nonlinear mappings of SPD matrices onto low dimensional manifolds. To achieve both reduction of dimension and nonlinear operation, Dong et al propose two layers: a connected two dimensional layer and a symmetrically clean layer. The former reduces dimensionality via a linear mapping, and the latter utilizes a deep neural network to model a nonlinear mapping. Concisely, nonlinearity is achieved by assigning zero values to symmetric pairs of elements. Matrices produced by these layers retain their SPD nature. Thus, the aforementioned shortcomings of SPD representations are offset. Indeed, the low-dimensional space decreases computational costs, and the manifold's geometric structure is fully appreciated because this method works on an SPD matrix rather than a vectorized form.

The effectiveness of this neural network as a facial recognition method was tested on two data sets: YouTube celebrities and ICT-TV. The YouTube celebrities data set contains snippets of YouTube videos centering 47 unique individuals. These clips vary greatly in frame number, lighting, and expression. Additionally, low resolution recordings with a high compression rate result in many frames being noisy and of low quality. Alternatively, the ICT-TV data set gathers frames that tend to be of higher quality, as they originate from the professionally filmed television shows Big Bang Theory and Prison Break. Frames from both data sets are pre-processed by resizing detected faces to the same pixel count and equalizing lighting effects. The faces are then flattened to a 2,880 element vector, which is reduced to 100 dimensions by a principal component analysis. These vectors are used to gain the kernel representation of the face, a 100×100 SPD matrix. For each individual, video clips are randomly split into testing and training sets with a one-to-one ratio. Various methods of facial recognition are then evaluated based on percentage accuracy.

Table 1: Comparison results with other methods on YTC and ICT-TV datasets.

Methods	YTC	ICT-TV-BBT	ICT-TV-PB
AIM (Pennec, Fillard, and Ayache 2006)	30.33 ± 3.72	37.64 ± 3.14	11.29 ± 2.46
SM (Sra 2012)	28.85 ± 3.41	38.07 ± 2.93	11.41 ± 2.85
LEM (Arsigny et al. 2007)	31.34 ± 3.64	40.89 ± 3.08	13.36 ± 2.72
SPDML (Harandi, Salzmann, and Hartley 2014)	40.86 ± 3.24	41.52 ± 2.14	17.94 ± 2.82
RSR (Harandi et al. 2012)	34.01 ± 3.06	47.93 ± 2.72	15.52 ± 2.30
LEK (Li et al. 2013)	33.81 ± 3.83	44.16 ± 2.71	16.74 ± 2.74
CDL (Wang et al. 2012)	31.84 ± 2.54	44.38 ± 2.28	15.26 ± 2.06
ITML-LEM (Vemulapalli and Jacobs 2015)	33.42 ± 3.42	46.62 ± 2.03	14.39 ± 2.52
LEML (Huang et al. 2015)	38.04 ± 2.11	49.60 ± 2.57	18.73 ± 2.20
DCC (Kim, Kittler, and Cipolla 2007)	32.84 ± 3.61	46.68 ± 3.04	15.31 ± 2.83
GDA (Hamm and Lee 2008)	32.09 ± 3.17	46.14 ± 2.98	17.03 ± 2.91
AHISD (Cevikalp and Triggs 2010)	31.16 ± 3.04	41.62 ± 2.72	15.88 ± 2.25
CHISD (Cevikalp and Triggs 2010)	32.08 ± 2.66	45.24 ± 2.58	16.52 ± 2.91
SSDML (Zhu et al. 2013)	34.77 ± 2.59	42.36 ± 2.47	13.71 ± 3.07
Our method	$\mathbf{46.37 \pm 3.07}$	$\mathbf{55.18 \pm 2.94}$	$\mathbf{24.18 \pm 2.05}$

The above table summarizes the results of this test. The first group contains basic methods used as a baseline for comparison. AIM and SM based supervised methods are contained in the second group. LEM based supervised methods are contained in the third group. The final group describes methods that do not center SPD manifolds. RSR and LEM use sparse representation based classifiers, while others use nearest neighbor classifiers. The three data sets unanimously see the proposed method outperforming all others, a testament to its effectiveness.

Bounds in the accuracy of facial recognition algorithms give rise to normative concerns regarding fairness, misuse, and data privacy. While this paper claims that the accuracy of this method is higher than others, it did not elaborate if this accuracy rate is similar among all demographic groups. If facial recognition methods become more prominent in surveillance and policing, unfair methods could result in certain demographics receiving a disproportionate amount of false accusations. In order for a method like deep manifold learning of SPD matrices to be used, it must prove to meet conditions of fairness. Beyond testing for fairness, this facial recognition method is still not always correct. While it is indeed more accurate than randomly selecting an individual, the method remains more often wrong than right in two of the three tested data sets. Someone falsely identified by a facial recognition algorithm may find it harder to gain resources like legal counsel, depending on who has access to the data. Additionally, information about people's whereabouts could beget data privacy issues. If this data is accessed by a malicious source, it opens the possibility of revealing sensitive information without the individual's consent. This also opens up the possibility of physical safety concerns like stalking. Thus, if facial recognition data is to be collected, there must be a way to ensure its safety. The development of technology like this also calls into question who should be able to utilize this technology. Some corporations with access to facial recognition data may use this to understand population trends or individual habits and preferences. One could argue that this treats individuals as a means to an end, as this utilizes their personal data for corporate benefit, likely without consent.

References

Dong, Z., Jia, S., Zhang, C., Pei, M., & Wu, Y. (2017). Deep Manifold Learning of Symmetric Positive Definite Matrices with Application to Face Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence, 31*(1). https://doi.org/10.1609/aaai.v31i1.11232