Pattern Recognition 78 (2018) 154–166

Contents lists available at ScienceDirect

Pattern Recognition

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

Multi-view manifold learning with locality alignment

Yue Zhao [a], Xinge You [a,*], Shujian Yu [b], Chang Xu [c], Wei Yuan [a], Xiao-Yuan Jing [d], Taiping Zhang [e], Dacheng Tao [c]

[a] School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, Hubei, China
[b] Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA
[c] Centre for Artificial Intelligence, University of Technology Sydney, 81 Broadway Street, Ultimo, Sydney, NSW 2007, Australia
[d] State Key Laboratory of Software Engineering, School of Computer, Wuhan University, Wuhan, Hubei, China
[e] College of Computer Science, Chongqing University, Chongqing 400030, China

ARTICLE INFO

Article history:
Received 25 April 2017
Revised 8 January 2018
Accepted 14 January 2018

Keywords:
Manifold learning
Multi-view learning
Locality alignment

ABSTRACT

Manifold learning aims to discover the low dimensional space where the input high dimensional data are embedded by preserving the geometric structure. Unfortunately, almost all the existing manifold learning methods were proposed under single view scenario, and they cannot be straightforwardly applied to multiple feature sets. Although concatenating multiple views into a single feature provides a plausible solution, it remains a question on how to better explore the independence and interdependence of different views while conducting manifold learning. In this paper, we propose a multi-view manifold learning with locality alignment (MVML-LA) framework to learn a common yet discriminative low-dimensional latent space that contain sufficient information of original inputs. Both supervised algorithm (S-MVML-LA) and unsupervised algorithm (U-MVML-LA) are developed. Experiments on benchmark real-world datasets demonstrate the superiority of our proposed S-MVML-LA and U-MVML-LA over existing state-of-the-art methods.

1. Introduction

Real world objects, like images, texts or videos, are often represented with high-dimensional data [1]. Given the high dimensionality, the sample size required to estimate the function of several variables to a given degree of accuracy will grow exponentially with the increasing number of variables which implies the so-called *curse of the dimensionality* problem [2]. Dimensionality reduction is a common approach to decrease the demanding on the training samples and can reveal the intrinsic structure of the distribution of original high-dimensional measurements [3].

Manifold learning is an effective approach for nonlinear dimensionality reduction [4,5]. It learns an embedded low-dimensional nonlinear manifold through precisely described and preserved local geometric information of the original high-dimensional space [6,7]. Representative manifold learning algorithms include Isometric Feature Mapping (ISOMAP) [8], Local Linear Embedding (LLE) [9], Laplacian Eigenmaps (LE) [10], Hessian-based Local Linear Embedding (HLLE) [11], Local Tangent Space Alignment (LTSA) [12], etc. ISOMAP uses the geodesic distance to measure the geometric information within manifold. LLE assumes that a high-dimensional manifold can be approximated by small area in the Euclidean space, and the reconstruction coefficients of the local neighbors can be preserved in the low-dimensional space. LE manipulates on an undirected weighted graph to preserve the local neighbor relationships. HLLE obtains the low-dimensional representations through applying eigenanalysis to the matrix built by Hessian coefficients. LTSA utilizes local tangent information to represent the local geometry information and then provides a global coordinate by this local tangent information. Following these methods, locality preserving projections (LPP) [13], neighborhood preserving embedding (NPE) [14], orthogonal neighborhood preserving projections (ONPP) [15] and linear local tangent space alignment (LLTSA) [16] have been proposed respectively to solve the out-of-sample problem [17]. Recently, [5] proposes patch alignment framework which constructs local patches using their nearest neighbor relationships to capture the local geometry. Patch alignment framework unifies all the aforementioned manifold learning algorithms, it can also be applied for discriminative dimensionality reduction by imposing the discriminative information between nearest neighbors in the local optimization stage. Although the developed

* Corresponding author.
E-mail addresses: zhaoyue91@hust.edu.cn (Y. Zhao), youxg@hust.edu.cn, youxg@mail.hust.edu.cn, 53954664@qq.com (X. You), yusjlcy9011@ufl.edu (S. Yu), c.xu@sydney.edu.au (C. Xu), yuanwei@mail.hust.edu.cn (W. Yuan), jingxy_2000@126.com (X.-Y. Jing), tpzhang@cqu.edu.cn (T. Zhang), dacheng.tao@sydney.edu.au (D. Tao).

https://doi.org/10.1016/j.patcog.2018.01.012
0031-3203/

Discriminative Locality Alignment (DLA) [5] reduces to Linear Discriminant Analysis (LDA) [18,19] under specific parameter setting,[1] it is also feasible to deal with nonlinear distributed measurements, because of the utilization of local information.

In many practical pattern classification applications, the raw data are often collected from different source domains or different descriptors [20]. The different sources or descriptors reveal the fundamental characteristics and properties of the objects from different perspectives and are often treated as views of the objects. Different from single-view data which only contains partial information of the object, multi-view data usually carries complementary information between different views [21,22]. Thus, multi-view data can be used to learn the object sufficiently and multi-view learning has draw broad attentions in machine learning community [23–30]. Existing multi-view learning algorithms can be divided into three categories [20]: co-training [31], multiple kernel learning [32] and subspace learning [33]. Co-training trains separate yet correlated learners on each view to maximize the agreement and minimize the disagreement between the predictions from each view on the validation set. Multiple kernel learning treats each view as a single kernel, it aims to choose suitable multi-view sets and appropriate view combination methods to achieve better learning performances. Different from co-training and multiple kernel learning, subspace learning attempts to obtain a latent subspace shared by all views.

However, most of the state-of-the-art manifold learning methods are single-view-based. A naive method for manifold learning on multi-view data is just concatenating multiple views into a single feature vector. This concatenation makes no sense as data from different views may lie in quite different distributions. In this paper, we propose multi-view manifold learning with locality alignment (MVML-LA) framework to realize manifold learning under multi-view scenario. A common low-dimensional latent space that can preserve sufficient information of input views is generated. More specifically, we integrate locality alignment in the latent space learning process to enhance its discriminating capability and develop two specific algorithms in supervised and unsupervised scenarios, respectively. To summarize, the main contributions of this work are twofold.

- In order to extend manifold learning into multi-view scenario, we propose multi-view manifold learning framework. Different from existing multi-view manifold learning methods, our framework learns a common low-dimensional latent space that carries sufficient information of multiple views. Furthermore, locality alignment based on neighbor relationship is appended to enhance the discriminating ability of the low-dimensional latent space.
- In practice, there is no guarantee that the data label information is always available. In order to improve the generalization ability of our framework, both supervised and unsupervised versions of MVML-LA are developed.

The rest of this paper is organized as follows. In Section 2, we formulate our problem and briefly describe the related works. Following this, we introduce our proposed S-MVML-LA and U-MVML-LA in Section 3. The experiments and results analysis are conducted in Section 4. Finally, Section 5 concludes this paper.

[1] DLA selects k_1 nearest neighbors from the same class of point x_i and k_2 nearest neighbors from the different classes of x_i to form its local patch. In this sense, DLA reduces to LDA if k_1 is equal to the sample number from the within-class of x_i and k_2 is equal to the sample number from the between-class of x_i.

Table 1
Frequently used notations and descriptions.

Notations	Descriptions
z_i	A single sample vector.
z_{ij}	The jth within-class neighbour of z_i.
z^{ip}	The pth between-class neighbour of z_i.
$Z=\{z_1, z_2, \ldots, z_n\}$	A n sample set.
$Z^v=\{z_1^v, z_2^v, \ldots, z_n^v\}$	The vth view of Z.
Q^v	The view generation function of view v.
ε^v	The construction error of view v.

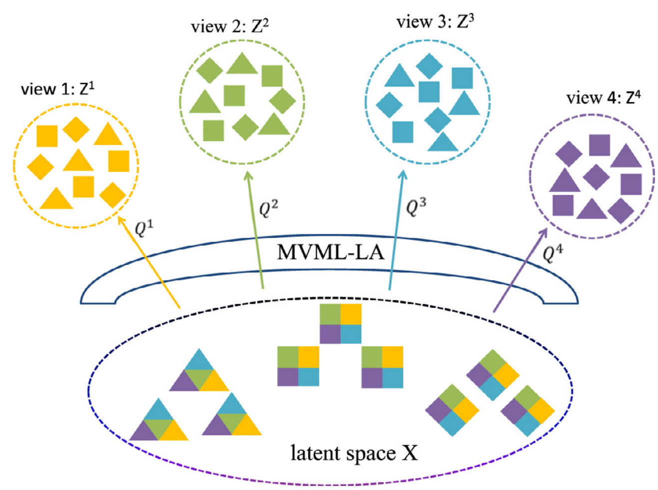

Fig. 1. Latent space formulation. Multiple views (different colors represent different views, different shapes represent different classes) are located on spaces which can be generated by the latent space we learned.

2. Problem formulation and related works

2.1. Notations

In this paper, scalars are represented by lowercase boldface letters (e.g., $\mathbf{z}$), vectors appear in lowercase letters (e.g., z) and matrices are indicated by uppercase letters (e.g., Z), specifically, I stands for the identity matrix. We represent a n sample set as $Z=\{z_1, z_2, \ldots, z_n\}$, where $z_i|_{i=1}^n$ denotes the ith column (or sample) in Z. Moreover, Z^v represents the vth view to sample set Z: $Z^v=\{z_1^v, z_2^v, \ldots, z_n^v\}$. Furthermore, z_{ij} represents the jth within-class neighbor of z_i and z^{ip} represents the pth between-class neighbor of z_i. In addition, Q^v and ε^v are the view generation function and construction error for the vth view, respectively. For $z \in \mathbb{R}^{n\times 1}$, $||z||_2 = \sqrt{\sum_{i=1}^n \mathbf{z}_i^2}$ is the L_2 norm of z, where $\mathbf{z}_i$ is the ith element of z. For clarity, we summarize the frequently used notations and their corresponding descriptions in Table 1.

2.2. Problem formulation and general framework

The motivation of multi-view manifold learning is that the learned low-dimensional embedding should be more informative than it learned from individual view or concatenated view, thus boosting the performance of classical machine learning tasks, like classification. In order to ensure the information of each view is preserved in the learned low-dimensional latent space, we assume that the multiple views (or observations) are embedded on different spaces of a common latent space (see Fig. 1). Specifically, suppose we are given m views of observations $Z^v|_{v=1}^m$ to form the joint feature space with n samples (i.e., $Z^v=\{z_1^v, z_2^v, \ldots, z_n^v\} \in \mathbb{R}^{K\times n}$, where K is the dimensionality of the vth view), then the common $M(M \ll K)$ dimensional hidden representation representation $X=\{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^{M\times n}$ of all views can be generated by the fol-