



# Sample-dependent graph construction with application to dimensionality reduction

Bo Yang<sup>a,b</sup>, Songcan Chen<sup>a,\*</sup>

<sup>a</sup> Department of Computer Science and Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, PR China

<sup>b</sup> Department of Computer Science, Tonghua Normal University, Tonghua 134002, PR China

## ARTICLE INFO

### Article history:

Received 8 October 2009

Received in revised form

7 March 2010

Accepted 14 March 2010

Communicated by Y. Fu

Available online 5 May 2010

### Keywords:

Graph Laplacian

Graph construction

Dimensionality reduction (DR)

Graph embedding

Similarity neighborhood

Local neighbor

## ABSTRACT

Graph construction plays a key role on learning algorithms based on graph Laplacian. However, the traditional graph construction approaches of  $\varepsilon$ -neighborhood and  $k$ -nearest-neighbor need to predefine the same neighbor parameter  $\varepsilon$  (or  $k$ ) for all samples, which usually suffers from the difficulty of parameter selection and generally fail to effectively fit intrinsic structures of data. To mitigate these limitations to a certain extent, in this paper we present a novel and *sample-dependent* approach of graph construction, and name the so-constructed graph as Sample-dependent Graph (SG). Specifically, instead of predefining the same neighbor parameter for all samples, the SG depends on samples in question to determine neighbors of each sample and similarities between sample pairs. As a result, it not only avoids the intractability and high expense of neighbor parameter selection but also can more effectively fit the intrinsic structures of data. Further, in order to show the effectiveness of the SG, we apply it to the dimensionality reduction based on graph embedding, and incorporate it into the state-of-the-art off-the-shelf unsupervised locality preserving projection (LPP) to develop the sample-dependent LPP (SLPP). SLPP naturally inherits the merits of SG and maintains the attractive properties of the traditional LPP. The experiments on the toy and benchmark (UCI, face recognition, object category and handwritten digits recognition) datasets show the effectiveness and feasibility of the SG and SLPP with promising results.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

Graph Laplacian based learning has attracted growing interest; thus many corresponding algorithms have recently been presented, which widely involve dimensionality reduction (DR) [1–10], semi-supervised learning [11–15], label propagation [16–18], clustering [19–23], image match [18,24], etc. For these algorithms, the graph construction has become a heart. As claimed in [13], constructing a ‘good’ graph is even more important than choosing a ‘good’ algorithm. So, it is undoubtedly necessary and valuable to study graph construction and propose new construction approaches. In fact, some researchers have recently dedicated to study the graph construction [25–29]. Maier and Luxburg [25] showed that  $k$ -nearest-neighbor and  $\varepsilon$ -neighborhood graphs can produce extraordinarily different influences on unsupervised spectral clustering. Jebara et al. [26] presented a  $b$ -matching graph, which is alternative to  $k$ -nearest-neighbor graph. The authors in [27,28] focused on how to combine different graphs in a way that tends to give heavier weight to a better graph

while authors in [29] presented an in-depth study about the convergence of graph Laplacians on random neighborhood graphs. Obviously, these works provide some insight into the study of graph construction. However, it is still untouched how to construct a more effective graph depending on samples at hand; so, in this paper, we attempt to present an approach of sample-dependent graph construction.

More concretely, the sample-dependent graph construction is motivated by the following examinations: the determination of neighbors of each sample, i.e., construction of edges for a graph plays a key role in the process of graph construction. In traditional to graph constructions, neighbors of a sample are searched by the two main approaches of  $\varepsilon$ -neighborhood and  $k$ -nearest-neighbor [4,19]. Both approaches need to predefine the same neighbor parameter  $\varepsilon$  (or  $k$ ) for all samples. However, it is not easy to construct an effective graph for learning tasks by the two approaches due to the two main reasons:

- (i) The parameter is in advance set to the same value for all samples. Clearly, such setting is not very reasonable because local structures of each sample usually are different. Thus, the so-constructed graph generally fails to effectively capture and fit the intrinsic structures of data.

\* Corresponding author.

E-mail address: [s.chen@nuaa.edu.cn](mailto:s.chen@nuaa.edu.cn) (S. Chen).

- (ii) The selection of neighbor parameter is still intractable. It is well known that up to date, how to select parameter is still an open problem. Although there are some approaches to address it, which is *our first contribution*. Different from the traditional  $\varepsilon$ -neighborhood and  $k$ -nearest-neighbor graphs, instead of setting in advance the same neighbor parameter for all samples, the SG depends on samples in question to determine the neighbors of each sample and similarities between sample pairs. More specifically, for sample  $x_i$ , if the similarity between it and sample  $x_j$  is greater than the mean of similarities between sample  $x_i$  and all samples, then  $x_j$  becomes its neighbor. In other words, the Mean of Similarities (MS) for  $x_i$  defines its own similarity neighborhood, i.e., the hypersphere whose center is  $x_i$  and radius is MS. So the similarities of the samples within the hypersphere are greater than MS and they become automatically its neighbors. Correspondingly, the entry  $W_{ij}^S$  of adjacency matrix  $W^S$  is set to the similarity value. Notice that  $W_{ji}^S$  is not set to the same value because the defined similarity function (see its definition in subsection 3.1) varies from sample to sample and from data to data.

To mitigate the harassment of the traditional approaches, in this paper, we propose a *sample-dependent* approach of graph construction and name the so-constructed graph as Sample-dependent Graph (SG), which is *our first contribution*. Different from the traditional  $\varepsilon$ -neighborhood and  $k$ -nearest-neighbor graphs, instead of setting in advance the same neighbor parameter for all samples, the SG depends on samples in question to determine the neighbors of each sample and similarities between sample pairs. More specifically, for sample  $x_i$ , if the similarity between it and sample  $x_j$  is greater than the mean of similarities between sample  $x_i$  and all samples, then  $x_j$  becomes its neighbor. In other words, the Mean of Similarities (MS) for  $x_i$  defines its own similarity neighborhood, i.e., the hypersphere whose center is  $x_i$  and radius is MS. So the similarities of the samples within the hypersphere are greater than MS and they become automatically its neighbors. Correspondingly, the entry  $W_{ij}^S$  of adjacency matrix  $W^S$  is set to the similarity value. Notice that  $W_{ji}^S$  is not set to the same value because the defined similarity function (see its definition in subsection 3.1) varies from sample to sample and from data to data.

Compared with the traditional  $\varepsilon$ -neighborhood and  $k$ -nearest-neighbor graphs, the SG has some favorable and attractive properties:

- (i) The SG depends on samples in question to determine the neighbors of each sample and similarities between sample pairs. Such strategy avoids the stiff criterion as used in the two traditional graphs, i.e., the same neighbor parameter are predefined for all the samples.
- (ii) Although the SG is not yet parameter-free due to the similarity parameter, it reduces the number of parameters compared to the traditional graphs and thus is convenient and simple to use in learning algorithms. Undoubtedly, this can reduce the complexity in real applications.
- (iii) From the construction definition of the SG, it can be demonstrated (see Proposition 1 in Section 3 and the toy experiment in Section 4.1) that its adjacency matrix is generally asymmetric even if  $x_i$  and  $x_j$  are neighbors of each other, which is generally more reasonable and intuitive for effectively capturing and fitting the intrinsic structures of data.
- (iv) In the SG, generally, a sample in high-density area has more neighbors; on the contrary, a sample in low-density area has relatively less neighbors (this characteristic is simply shown in the toy experiment in Section 4.1), which reflects certain statistics property.
- (v) There are some similarities between the SG and the two traditional graphs, but the SG is not their special case (see (B) in Section 3.2).
- (vi) The SG is very general. It can potentially be adapted to many learning algorithms based on graph Laplacian.

As a construction strategy of graph, the proposed approach can potentially be applied to many learning algorithms based on graph Laplacian; however, in this paper, we just concern on an application of SG to DR to demonstrate the effectiveness of our

strategy mainly due to its attractive properties and popularity.<sup>1</sup> In fact, many DR algorithms based on graphs [1,2,4,10,20,22,23,31–44] have recently been proposed, typically like the locality preserving projection (LPP) [1], neighborhood preserving embedding (NPE) [33], Bayesian tensor analysis (BTA) [38,39], local coordinates alignment (LCA) [40], etc. Their implementations share a common routine, i.e., first construct an affinity graph and then embed it into an objective function to optimize the function for preserving some (locally geometrical or structural) properties of original high-dimensional space; in such a way, it get a projection matrix for dimensionality reduction. We follow the routine as well. Specifically, we incorporate the SG into the state-of-the-art off-the-shelf unsupervised LPP to develop the sample-dependent LPP (SLPP), which is *our second contribution*.

SLPP also has some attractive properties:

- (i) SLPP naturally inherits all merits of the SG, as mentioned in (i)–(vi) above.
- (ii) It is worth to point out that though the adjacency matrix of SG is generally asymmetric, the Laplacian matrix induced from it is still symmetric (see Proposition 2 in Section 3).
- (iii) SLPP shares some good characteristics of LPP, such as avoidance of the “out-of-sample” problem [21].
- (iv) The SG can be directly incorporated into the off-the-shelf LPP, hence does not bring much more computational complexity compared to LPP.

The rest of this paper is organized as follows: Section 2 briefly reviews the traditional approaches of graph construction, and takes LPP as an example to introduce unsupervised DR algorithms based on graph embedding. Section 3 firstly presents the novel SG, demonstrates its property and then develops the unsupervised SLPP algorithm. Section 4 performs the experiments to evaluate the effectiveness and feasibility of the SG and SLPP. Section 5 gives the conclusion remarks and discusses the future work.

## 2. Related works

### 2.1. Traditional graph construction

Given a set of  $n$  samples  $X = \{x_1, \dots, x_n\}$ ,  $x_i \in R^D$ , a weighted graph  $G = \{V, E, W\}$  can be constructed, where  $V$  corresponds to the sample points in set  $X$ ,  $E$  denotes the edge set between the sample pairs and entry  $W_{ij}$  of the adjacency matrix and  $W$  denotes similarity between  $x_i$  and  $x_j$ . The construction process of graph  $G$  can typically be decomposed into two steps: (1) construction of edges and (2) calculation of weight value for each edge. Step (1) mainly includes the two categories of  $\varepsilon$ -neighborhood and  $k$ -nearest-neighbor [19].

$\varepsilon$ -neighborhood:  $x_i$  and  $x_j$  are linked by an edge if  $\|x_i - x_j\|^2 < \varepsilon$  where  $\|\cdot\|$  is the usual Euclidean norm in  $R^D$  and  $\varepsilon$  is a local threshold parameter, which controls the size of a local neighborhood.

$k$ -nearest-neighbor:  $x_i$  and  $x_j$  are linked by an edge if  $x_i$  is among  $k$  nearest neighbors of  $x_j$  or  $x_j$  is among  $k$  nearest neighbors of  $x_i$ .

And at the same time, for Step (2), there are three frequent approaches:

- (i) Heat kernel

$$W_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2 / 2p^2), & \text{if } x_i \text{ and } x_j \text{ are neighbors} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where  $p$  is the width parameter of the heat kernel.

<sup>1</sup> The effect of proposed strategy on other learning algorithms based on graph Laplacian will be across-the-board demonstrated and exploited in the future work.

Download English Version:

<https://daneshyari.com/en/article/411068>

Download Persian Version:

<https://daneshyari.com/article/411068>

[Daneshyari.com](https://daneshyari.com)