Neurocomputing 120 (2013) 441–452

Contents lists available at ScienceDirect

Neurocomputing

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

# $L_1$-graph construction using structured sparsity

Guangyao Zhou, Zhiwu Lu, Yuxin Peng *

*Institute of Computer Science and Technology, Peking University, Beijing 100871, China*

ARTICLE INFO

*Article history:*
Received 4 June 2012
Received in revised form
2 January 2013
Accepted 24 March 2013
Communicated by Xianbin Cao
Available online 12 June 2013

*Keywords:*
Graph construction
Sparse representation
$L_1$-graph
Structured sparsity
Spectral clustering

ABSTRACT

As a powerful model to represent the data, graph has been widely applied to many machine learning tasks. More notably, to address the problems associated with the traditional graph construction methods, sparse representation has been successfully used for graph construction, and one typical work is $L_1$-graph. However, since $L_1$-graph often establishes only part of all the valuable connections between different data points due to its tendency to ignore the intrinsic structure hidden among the data, it fails to exploit such important information for the subsequent machine learning. Besides, the high computational costs of $L_1$-graph prevent it from being applied to large scale high-dimensional datasets. In this paper, we construct a new graph, called the $k$-nearest neighbor ($k$-NN) fused Lasso graph, which is different from the traditional $L_1$-graph because of its successful incorporation of the structured sparsity into the graph construction process and its applicability to large complex datasets. More concretely, to induce the structured sparsity, a novel regularization term is defined and reformulated into a matrix form to fit in the sparse representation step of $L_1$-graph construction, and the $k$-NN method and kernel method are employed to deal with large complex datasets. Experimental results on several complex image datasets demonstrate the promising performance of our $k$-NN fused Lasso graph and also its advantage over the traditional $L_1$-graph in the task of spectral clustering.



## 1. Introduction

Since graph is a powerful model to represent the data, it has served as foundation for lots of machine learning problems, such as spectral clustering [1,2], semi-supervised learning [3,4], dimension reduction [5] and so on. Although many graph-based methods have been developed for different machine learning tasks, graph construction still receives relatively little attention as pointed out in [6,7]. In the literature, there exist two commonly used strategies for graph construction, namely $k$-nearest neighbor ($k$-NN) and $\epsilon$−ball methods. Although these methods are easy both to understand and to implement, they suffer from inherent limitations, e.g. data dependency and sensitivity to noise.

Recently, to address these problems, sparse representation [8] has been successfully used for graph construction, among which one typical work is $L_1$-graph [9,10]. The success of $L_1$-graph lies in the sparse representation step, in which it seeks a sparse linear reconstruction of each data point with the other data points by exploiting the sparse property of the Lasso penalty [11]. This is, in fact, a new way that is fundamentally different from the traditional ones (like Euclidean distance, cosine distance, etc.) to measure the similarity between different data points. By inducing sparsity in the linear reconstruction process, it identifies the most relevant data points as well as their estimated similarity to the reconstructed data point, and by doing so gets a graph that proves effective in the subsequent graph-based machine learning tasks.

However, two interesting comments from previous works on the Lasso method and the $L_1$-graph attract our attention:

(1) As reported in [12], when faced with a group of highly correlated variables, Lasso method tends to randomly choose one of them.
(2) In [10], the authors stated that "for certain extreme cases, e.g. if we simply duplicate each sample and generate another new dataset of double size, $L_1$-graph may only connect these duplicated pairs".

With some simple mathematical derivations, as shown in Section 2, we can see the similarity between the sparse representation step of $L_1$-graph construction and the Lasso method. As a result, if we think of the data points and the similarity between data points in sparse representation step as variables and correlation between variables in Lasso method respectively, the first comment indeed suggests that the sparse representation does not connect all the data points that need to be connected. In the situation mentioned in the second comment, the similarity between the reconstructed data point and its duplicate (measured by the sparse representation method of $L_1$-graph) dominates others, which also makes the sparse representation step ignore many other valuable connections.

In addition, advances in technology have made large scale high-dimensional datasets common in many scientific disciplines,

* Corresponding author. Tel.: +86 10 82529699; fax: +86 10 82529207.
*E-mail addresses:* tczhouguangyao@gmail.com (G. Zhou), luzhiwu@pku.edu.cn (Z. Lu), pengyuxin@pku.edu.cn (Y. Peng).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.03.045

yet the construction process of $L_1$-graph, in which the computational costs become unbearable because of a huge matrix (details will be given in Section 2) constructed when dealing with these datasets, prevents it from being further applied to problems related to such large complex datasets.

Our work mainly aims to overcome these shortcomings of $L_1$-graph. To avoid $L_1$-graph's failure to establish all valuable connections between different data points, we seek to incorporate structured sparsity into the $L_1$-graph construction process. The main idea is to exploit the local structure across the dataset by making the reconstruction coefficients of every data point and its nearest neighbors also close to each other in value in the linear reconstruction process of the sparse representation step. To achieve this, we propose a novel regularization term, which makes use of the information provided by traditional ways of measuring similarity between data points, for the sparse representation step of $L_1$-graph construction to induce structured sparsity and reformulate it in matrix form to fit in our new graph (which we call $k$-NN fused Lasso graph) construction process. And in order to deal with large scale high-dimensional datasets, we employ the $k$-NN method and kernel method in our new graph construction process. To be more specific, we reconstruct each data point and construct the corresponding new regularization term, both with only its $k$ nearest neighbors to handle large scale datasets. And when solving the linear reconstruction problem, we use the kernel matrix instead of the original data vectors to handle high-dimensional datasets. The effectiveness of $k$-NN fused Lasso graph is verified by the experimental results on several large complex image datasets in the task of spectral clustering. Specifically, to gain a first impression of its effectiveness, the similarity (i.e. weight) matrix of our new graph on the doubled soybean dataset (the soybean dataset, which contains 47 35-dimensional instances, can be downloaded from the UCI Machine Learning Repository [20], and the doubled soybean dataset is generated by making an exact duplicate of each data point in the original dataset) are illustrated on Fig. 1(b). Comparing it with Fig. 1(a), we can easily see the tremendous advantage of our new graph over the $L_1$-graph.

Our main contribution is the development of the new $k$-NN fused Lasso graph construction method. To be more specific, our contributions can be summarized as follows:

(1) We proposed a novel regularization term to induce structured sparsity.
(2) We designed a reformulation strategy to incorporate the new regularization term into the graph construction process.
(3) We successfully employed the $k$-NN method and kernel method to make our graph construction method applicable to large scale high-dimensional datasets.

The idea of linearly reconstructing a given data point by its neighbors is also used in some other works, e.g. the locally linear embedding [22] method for dimension reduction. However, unlike our method, these works did not pay much attention to the reconstruction process itself. In [23], the authors proposed a unifying framework for dimension reduction called patch alignment. In our graph construction process, by using the $k$-NN method, we also construct a patch for each data point, and conducting the sparse representation step is similar to the part optimization in [23]. By unlike [23], we do not have a whole alignment step. We run the sparse representation step for each of the patches, and unifying them in the end to get the similarity matrix by symmetrizing the original similarity matrix constructed by the sparse representation steps. The idea of exploring the dataset structure in a pairwise manner is also present in some previous works, e.g. the max-min distance analysis [24]. But in [24], the authors used the pairwise distance between different classes, while our method focuses on the pairwise distance of the reconstruction coefficients of different data points. Also, our method is unsupervised in nature. We do not need such prior information as class labels, which makes our method applicable to many unsupervised or semi-supervised problems, and distinguishes our work from previous works like [24,14], as well as some other works, like the Group Sparse MahNMF in [25]. Like the elastic net [12], our new regularization term also has certain grouping effect. But we promote such grouping effect in a pairwise manner with the $L_1$ norm, which makes our method performs differently from the elastic net [12] as well as some other elastic net based works, such as Elastic Net Inducing MahNMF [25] and Manifold Elastic Net [26]. To the best of our knowledge, we have made the first attempt to incorporate the structured sparsity into the $L_1$-graph construction process, and the fact that our new $k$-NN fused Lasso graph outperforms the traditional $k$-NN graph and $L_1$-graph (see later experimental results in Section 6) when applied to spectral clustering on large complex image datasets demonstrates the great value of the structured sparsity information we utilize in our new method.

The rest of the paper will be organized as follows. In Section 2, we briefly review the $L_1$-graph construction method. In Section 3, we describe in detail how we overcome the shortcomings of

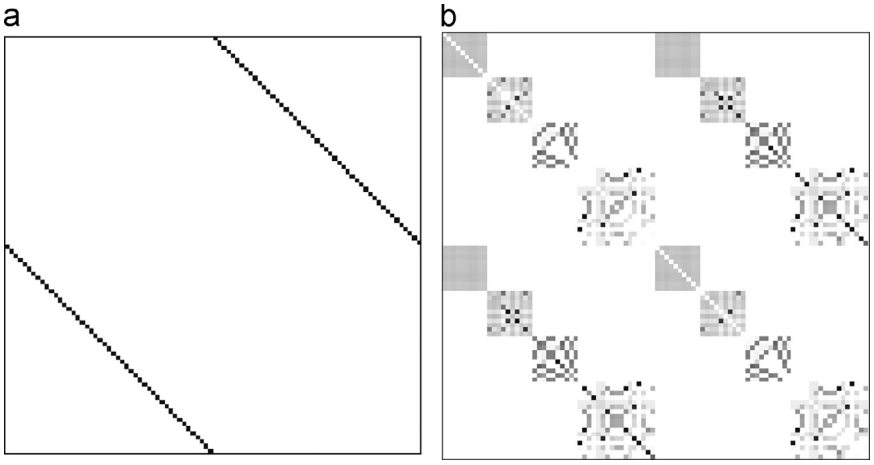


**Fig. 1.** Comparison between the similarity (i.e. weight) matrices of the $L_1$-graph and our $k$-NN fused Lasso graph on the doubled soybean dataset. For illustration purpose, the first and second half of the new dataset are identical copies of the original dataset, and each copy of the original dataset is rearranged such that data points within a class appear consecutively. More notably, the darker is a pixel, the larger is the similarity. (a) $L_1$-graph and (b) our graph.