Neurocomputing 207 (2016) 476–487

Contents lists available at ScienceDirect

Neurocomputing

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

Graph construction based on data self-representativeness and Laplacian smoothness

Libo Weng [a,b], Fadi Dornaika [b,c,*], Zhong Jin [a]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
[b] Department of Computer Science and Artificial Intelligence, University of the Basque Country UPV/EHU, San Sebastian, Spain
[c] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

ARTICLE INFO

Article history:
Received 22 October 2015
Received in revised form
16 February 2016
Accepted 8 May 2016
Communicated by Haowei Liu
Available online 16 May 2016

Keywords:
Graph construction
Data self-representativeness
Laplacian smoothness
Semi-supervised learning
Label propagation
Image classification

ABSTRACT

Recently, graph-based semi-supervised learning (SSL) becomes a hot topic in machine learning and pattern recognition. It has been shown that constructing an informative graph is one of the most important steps in SSL since a good graph can significantly affect the final performance of learning algorithms. This paper has the following main contributions. First, we introduce a new graph construction method based on data self-representativeness and Laplacian smoothness (SRLS). Second, this method is refined by incorporating an adaptive coding scheme aiming at getting a sparse graph. Third, we propose two kernelized versions of the SRLS method. A series of experiments on several public image data sets show that the proposed methods can out-perform many state-of-the-art methods. It is shown that Laplacian smoothness criterion is indeed a powerful tool to get informative graphs.

1. Introduction

Most real world data can have a graph structure that describes the pairwise similarity or relationship among samples. It is realized that data graphs are a natural way to represent the data [1–5]. Graph-based methods operate on a data driven graph [6–13]. In the graphs, the nodes correspond to data samples and the weighted edges between nodes encode the similarity between two nodes.

The most common way to construct a graph, or equivalently to estimate its affinity matrix, is to construct k-nearest neighbor graphs [14] or ε-neighborhoods graphs. Then edge weights are estimated using a similarity function that quantifies the relationship between the sample and its neighbors. It is noticed that the parameter setting in these two kinds of methods will heavily influence the final task performance [15]. Indeed, there is no simple way that can predict the best parameter setting for these methods. Jebara et al. [7] propose a graph construction method by b-matching in order to force that all the nodes will have the same degree (degree means the number of the edges connected to the node). In the past decade, different coding schemes and code book generation methods have been proposed.

In [16], the authors aim at constructing hard graphs using a similar criterion used by [17]. In this work, the neighborhood selection and edge weighting are performed in a single step; the edge weight matrix is symmetric and only non-negative edge weights are allowed. The edge weights are computed using a constraint that forces the degree, or the weighted degree of every node, to be equal to or greater than one. The authors devise a quadratic program that computes the non-negative weights. In order to avoid the non-tractability of solution, the graph is incrementally constructed by solving quadratic programs with a subset of edges.

Wang et al. [8] use a similar criterion as Locally Linear Embedding (LLE) to construct the graph by calculating the weights between pairs of samples. Wei et al. [9] define a neighborhood preserving graph based on LLE criterion for semi-supervised dimensionality reduction. Sparse representation is a widely used technique which assumes that the complex signal can be represented by some basic signals. Qiao et al. [11] construct an ℓ_1 graph with weighted edges using the coefficients of the sparse coding based on the theory of sparse representation and apply it to a locality preserving projection method for human face recognition. Yan et al. [18] also use the coefficients of sparse representation to construct the graph for semi-supervised classification [19] and multi-label classification [20].

Moreover, different from the sparse representation, Waqas et al. [21] proposed collaborative neighborhood representation

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2016.05.021
0925-2312/

(CNR) based on ℓ_2 norm minimization. Thus, graph construction using the coefficients of CNR is also a good choice. Locality-constrained Linear Coding (LLC) [22] is another algorithm for data representation. Based on CNR and LLC, Dornaika et al. proposed a graph construction method named weighted regularized least square [2] and a two phase method in paper [23].

Many graph based semi-supervised learning (SSL) methods have been proposed [24–26] which can broadly fall into main two categories: (i) those that use the graph structure to spread labels from labeled samples to unlabeled ones and (ii) those that optimize a loss function based on smoothness constraints derived from the graph. The researches in graph based SSL have achieved many progresses in many real applications such as speech recognition, text categorization, protein structure prediction, face recognition, and so on.

Compared to the normal semi-supervised learning [27–29], graph based SSL methods have the following advantages over other approaches:

1. For many applications, graph based SSL performs better than most other SSL algorithms in comparative evaluations.
2. Most graph based methods have a convex objective thereby function providing convergence guarantees, making them attractive for solving large-scale problems.
3. For many graph-based SSL approaches, optimizing the objective can be achieved via message passing on graphs. Each iteration of the algorithm consists of a set of updates to each graph node. A node's updated value is computed based on the node's current value as well as on the neighbors' current set of values.
4. It is possible to derive simple fast heuristics that enable such algorithms to scale to large parallel machines with good machine efficiency.

Existing graph construction methods have the following drawbacks and advantages:

- Graphs that are based on K Nearest Neighbor (KNN) paradigms such as KNN graphs and Locally Linear Embedding (LLE) graphs suffer from the fact that they need to specify the neighborhood size in advance. This imposes that every node in the graph should have the same number of edges. Constructing KNN graph is efficient but the performance of post-graph learning tasks can be poor.
- Graphs that are based on data self-representativeness with ℓ_2 regularization are very often providing dense graphs that are not very informative in general.
- Sparse graphs adopting ℓ_1 regularization provide sparse graphs that proved to be very powerful in many real learning problems. However, their computational load can be very expensive. Moreover, it is not clear if sparse graphs take into account data locality.

In this paper, we propose a new graph construction method based on data self-representativeness and Laplacian smoothness (SRLS). More precisely, we introduce a natural constraint for graph construction adopting data self-representativeness. Indeed, Laplacian smoothness criterion seems to be a powerful graph regularizer that has not been exploited in existing graph construction methods. In papers [30] and [31], the authors introduce a Laplacian sparse coding. The objective of their work is to obtain a dictionary and the corresponding sparse representation matrix of the data assuming that the data graph is known in advance. Unlike their work, our proposed method addresses the estimation of the affinity matrix of the graph. We assume that the dictionary is given by the original data set since we adopt data self-representativeness principle for pairwise similarity. The Laplacian smoothness is a part of the criterion to be optimized. Thus, the goal of SRLS is to obtain the unknown graph that simultaneously respects data self-representativeness and Laplacian smoothness. To the best of our knowledge, our work is the first attempt to construct graphs that adopt both criteria.

The remainder of the paper will be organized as: in the second section, some related work will be introduced. In the third section, the proposed method is presented. In the fourth section, we introduce two kernel versions of the proposed method. In the fifth section, we show some experiments on several real data sets to prove the efficiency of the proposed methods. In the last section, some conclusions will be drawn.

2. Related work

2.1. Traditional graph construction methods

k-nearest neighbor graphs and ε-neighborhoods' graphs are two traditional graph construction methods. Let the original data set be denoted by $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n]\in\Re^{d\times n}$.

kNN graphs: Samples $\mathbf{x}_i$, $\mathbf{x}_j$ are connected by an edge if $\mathbf{x}_i$ is in $\mathbf{x}_j$'s k-nearest neighborhood or vice versa. k is a parameter that controls the density of the graph.

ε-neighborhoods graphs: Samples $\mathbf{x}_i$, $\mathbf{x}_j$ are connected by an edge if the distance $d(\mathbf{x}_i,\mathbf{x}_j)\le\varepsilon$. The parameter ε controls neighborhood radius.

After the edges are decided, similarity will be measured to weigh the edges. The formula is shown in Eq. (1). For instance, $sim(\mathbf{x}_i,\mathbf{x}_j)$ is set to 1 as a constant or $sim(\mathbf{x}_i,\mathbf{x}_j)=e^{-\frac{\|\mathbf{x}_i-\mathbf{x}_j\|^2}{\sigma^2}}$, where σ is a parameter.

$$W_{ij}=\begin{cases} sim(\mathbf{x}_i,\mathbf{x}_j), & \mathbf{x}_i\in\delta_k(\mathbf{x}_j) \quad \text{or } \mathbf{x}_j\in\delta_k(\mathbf{x}_i),\\ 0, & \text{otherwise.}\end{cases} \tag{1}$$

where $\delta_k(\mathbf{x}_i)$ represents the set of $\mathbf{x}_i$'s k-nearest neighbors.

2.2. LLE graph construction

Locally Linear Embedding (LLE) is a classic manifold learning method. LLE preserves the neighborhood relationships of input samples [17]. It exploits the local linear reconstructions by minimizing the reconstruction error of the set of all local neighborhoods in the input space. It is discovered that the linear coding used by LLE can be used for the graph weight matrix construction. The weigh matrix $\mathbf{W}$ can be obtained by minimizing the reconstruction error:

$$\sum_i \|\mathbf{x}_i-\Sigma_{\mathbf{x}_j\in\delta_k(\mathbf{x}_i)}W_{ij}\mathbf{x}_j\|^2. \tag{2}$$

2.3. ℓ_1 Graph construction

In the traditional graph and the LLE based graph, it is needed to choose the parameters to control the radius of neighborhood while it is parameter-free in sparsity representation based graph. Qiao et al. [11] and Yan and Wang [18] proposed sparsity representation based graph construction methods in which every sample is represented as a sparse linear combination of the rest of input samples and the coefficients are considered as weights.

$$\begin{aligned}&\min \quad \|\mathbf{s}_i\|_1,\\ &\text{s.t.} \quad \mathbf{x}_i=\mathbf{X}\mathbf{s}_i,\end{aligned} \tag{3}$$

where $\mathbf{s}_i=[s_{i1},\ldots,s_{i,i-1},0,s_{i,i+1},\ldots,s_{in}]^T$ is an n-dimensional vector in which the i-th element is equal to zero (implying that the $\mathbf{x}_i$ is removed from $\mathbf{X}$), and the elements $s_{ij}, j\neq i$ denote the contribution of each $\mathbf{x}_j$ to reconstructing $\mathbf{x}_i$.