Knowledge-Based Systems 76 (2015) 148–165

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Automatic image annotation via compact graph based semi-supervised learning

Mingbo Zhao [a,b], Tommy W.S. Chow [b], Zhao Zhang [a,*], Bing Li [c]

[a] School of Computer Science and Technology, Soochow University, PR China
[b] Department of Electronics Engineering, City University of Hong Kong, Hong Kong Special Administrative Region
[c] School of Economics, Wuhan University of Technology, PR China

ARTICLE INFO

Article history:
Received 20 May 2014
Received in revised form 31 August 2014
Accepted 10 December 2014
Available online 23 December 2014

Keywords:
Graph based semi-supervised learning
Image annotation
Compact graph construction
Transductive and inductive learning
Label propagation

ABSTRACT

The insufficiency of labeled samples is major problem in automatic image annotation. However, unlabeled samples are readily available and abundant. Hence, semi-supervised learning methods, which utilize partly labeled samples and a large amount of unlabeled samples, have attracted increased attention in the field of image annotation. During the past decade, graph-based semi-supervised learning has been becoming one of the most important research areas in semi-supervised learning. In this paper, we propose a novel and effective graph based semi-supervised learning method for image annotation. The new method is derived by a compact graph that can well grasp the manifold structure. In addition, we theoretically prove that the proposed semi-supervised learning method can be analyzed under a regularized framework. It can also be easily extended to deal with out-of-sample data. Simulation results show that the proposed method can achieve better performance compared with other state-of-the-art graph based semi-supervised learning methods.

1. Introduction

In the real world, there are ever-increasing vision image data generated from Internet surfing and daily social communication. These metadata can be labeled or unlabeled, and accordingly be utilized for image retrieval, summarization, and indexing. In order to handle these datasets for realizing the above tasks, automatic annotation is an elementary step, which can be formulated as a pattern classification problem and accomplished by learning-based techniques. Traditionally, the learning-based methods can be categorized into two categories: (1) supervised learning, which aims to predict the labels of new-coming data samples from the observed labeled set, i.e., to handle the classification problem or to preserve the discriminative information which is embedded in the training set. Typical supervised learning methods include Support Vector Machines (SVM) [36,37], and Linear Discriminant Analysis (LDA) and its variants [1–3,17,18]; and (2) the other one is unsupervised learning, the goal of which is to handle the observed data with no labels and to grasp the intrinsic structure of the dataset. Typical unsupervised learning methods include clustering [19–22] and manifold learning methods, such as ISOMAP [23], Locally Linear Embedding (LLE) [24], and Laplacian Eigenmap (LE) [25]. In this paper, we mainly focus on the classification problem, which is traditionally a supervised learning task.

In order to handle the pattern classification problem, such as image annotation, the conventional supervised learning methods, such as Linear Discriminant Analysis (LDA) [1–3] and Support Vector Machine (SVM) [36,37], cannot deliver satisfactory classification accuracy when the number of labeled samples is not sufficient. However, labeling a large number of samples is time-consuming and costly. On the other hand, unlabeled samples are abundant and can be easily obtained in the real world. Hence, semi-supervised learning methods (SSL), which incorporate partly labeled samples and a large amount of unlabeled samples into learning, have become more effective than only relying on supervised learning. Recently, based on clustering and manifold assumptions, i.e., nearby samples (or samples of the same cluster or data manifold) are likely to share the same label [4–6], graph based semi-supervised learning methods have received considerable research interest in the area of semi-supervised learning. Typical methods include Manifold Regularization (MR) [15], Semi-supervised Discriminant Analysis (SDA) [16], Gaussian Fields and Harmonic Functions (GFHF) [4], Learning with Local and Global Consistency (LLGC) [5], and General Graph based Semi-supervised

* Corresponding author at: School of Computer Science and Technology, Soochow University, PR China. Tel.: +852 3442 2874.
E-mail addresses: mzhao4@cityu.edu.hk (M. Zhao), eetchow@cityu.edu.hk (T.W.S. Chow), cszzhang@gmail.com (Z. Zhang), lib675@163.com (B. Li).

http://dx.doi.org/10.1016/j.knosys.2014.12.014
0950-7051/

Learning method (GGSSL) [7,8]. All of these methods represent both labeled and unlabeled sets by a graph, and then utilize its graph Laplacian matrix to characterize the manifold structure [25,26].

In general, the abovementioned graph based semi-supervised learning can be divided into two categories: inductive learning methods and transductive learning methods. The inductive learning methods, such as MR [15] and SDA [16], try to induce a decision function that has a low classification error rate on the whole data space; while the transductive learning methods, also known as Label Propagation, aim to directly predict the label information from the labeled set to the unlabeled set along the graph, which is much easier to handle and less complicated than inductive learning methods. Two well-known transductive learning methods are GFHF [4] and LLGC [5]. GFHF has an elegant probabilistic explanation, and the output labels are the probabilistic values; however, it cannot detect outliers in data. In contrast, LLGC can detect outliers; however, its output labels are not probabilistic values. Both the problems in GFHF and LLGC have been eliminated by GGSSL [7,8], in which it can either detect the outliers or develop a mechanism to calculate the probabilities of data samples.

It should be noted that one important step of graph based SSL is to construct a graph with weights for characterizing the data structure. The graph is usually used to find the neighbors by k-neighborhood or ε-neighborhood in the whole data [23–26], and then to define the weight matrix on the graph. There are commonly two ways to define the weight matrix: one is to apply the Gaussian function [4,5,7,8,15,16] and the other is to employ the local linear reconstruction strategy [13,14]. The Gaussian function is easily manipulated in many graph based SSL, but estimating the optimal variance in the Gaussian function is very difficult [13]. The locally linear reconstruction strategy has no such problem, as it is based on the assumption that each sample can be reconstructed by a linear combination of its neighborhoods. The weight matrix is then automatically calculated when the neighborhood size is fixed. However, as pointed out in [10], using the neighborhoods of a sample to reconstruct it may not achieve the minimum result, which may not well capture the manifold structure of the dataset.

In this paper, motivated by the framework of GGSSL [7,8], we present an effective semi-supervised learning method, namely, Compact Graph based Semi-supervised Learning (**CGSSL**), for image annotation, which is based on a newly proposed compact graph. The newly proposed graph finds the neighbors of each sample and calculates the graph weights in a way as [13,14]. However, since the minimum reconstruction error of a sample may not be obtained by its own neighborhood, it aims to reconstruct the sample by using the neighbors of its adjacent samples and then preserve the graph weights corresponding to the minimum error. In this way, a more compact graph can be constructed, which can well capture the manifold structure of the dataset. In addition, in order to establish the connection to the normalized graph, we further symmetrize and normalize this compact graph. With these processes, the proposed CGSSL can be theoretically analyzed from the perspective of a graph, through which we show that the proposed CGSSL can be derived from a smoothness regularized framework. The proposed CGSSL can also be easily extended to its inductive out-of-sample version for handling new-coming data by using the same smoothness criterion. Finally, extensive simulations on image annotation and content based image retrieval show the effectiveness of the proposed CGSSL.

The main contributions of this paper are as follows: (1) we propose a new compact local reconstruction graph with symmetrization and normalization for semi-supervised learning. The new graph construction strategy can find a more compact way to approximate a sample with its neighborhoods, which can better grasp the manifold structure embedded in the dataset. In addition, with symmetrization and normalization processes, the proposed CGSSL can be analyzed theoretically under a regularized framework; (2) the proposed CGSSL is a transductive learning method, and it can also be easily extended to its inductive out-of-sample version for handling new-coming data by using the same smoothness criterion; (3) we analyze the relationships between the proposed CGSSL and other state-of-the-art graph based semi-supervised learning in terms of objectives, parameters and out-of-sample extensions, which are helpful for better understanding graph-based semi-supervised learning methods. Moreover, extensive simulations based on image annotation and content based image retrieval have verified the effectiveness of the proposed method; and (4) we further analyze the proposed CGSSL and other state-of-the-art methods from the time-to-process point of view and give an explicit implementation choice. Simulation results regarding computational time show that the proposed CGSSL can be more suitable and practical for handling large-scale image annotation tasks.

The rest of this paper is organized as follows: In Section 2, we will provide some notations and present the proposed CGSSL; in Section 3, we will give detailed analysis and out-of-sample extensions for the proposed CGSSL; extensive simulations are conducted in Section 4, and final conclusions are drawn in Section 5.

2. The proposed semi-supervised learning method

Let $X = [X_l, X_u] \in R^{d\times(l+u)}$ be the data matrix, where d is the number of data features, and the first l and the remaining u samples in X represent the labeled set X_l and unlabeled set X_u, respectively. Each sample in X_l is associated with a class label $c_i, i \in [1,2,\ldots,c]$, where c is the number of classes. The goal of graph based semi-supervised learning methods is to propagate the label information of the labeled set to the unlabeled set according to the distribution associated with both the labeled and unlabeled set [4,5], and through which the predicted labels of the unlabeled set, called soft labels, can be obtained.

2.1. Review of graph construction

In label propagation, a similarity matrix must be defined for evaluating the similarities between any two samples. The similarity matrix can be approximated by a neighborhood graph associated with weights on the edges. Officially, let $\widetilde{G} = (\widetilde{V}, \widetilde{E})$ denote this graph, where $\widetilde{V}$ is the vertex set of $\widetilde{G}$ representing the training samples, and $\widetilde{E}$ is the edge set of $\widetilde{G}$ associated with a weight matrix W containing the local information between two nearby samples. There are many ways to define the weight matrix. A typical way is to use the Gaussian function [4,5,7,8,15,16]:

$$w_{ij} = \exp\left(-\|x_i - x_j\|^2/2\sigma^2\right) \quad x_i \in N_k(x_j) \quad or \quad x_j \in N_k(x_i), \tag{1}$$

where $N_k(x_j)$ is the k neighborhood set of x_j, and σ is the Gaussian function variance. However, σ is hard to be determined, and even a small variation of σ can alter the results dramatically [13]. Wang et al. have proposed another strategy to construct $\widetilde{G}$ by using the neighborhood information of samples [13,14]. This strategy assumes that each sample can be reconstructed by a linear combination of its neighborhoods [24], i.e., $x_i \approx \sum_{j:x_j\in N_k(x_i)} w_{ij}x_j$. It then calculates the weight matrix by solving a standard quadratic programming (QP) problem as:

$$\min \left\| x_i - \sum_{j:x_j\in N_k(x_i)} w_{ij}x_j \right\|_F^2 \quad s.t.\ w_{ij} \geqslant 0, \quad \sum_{j\in N_k(x_i)} w_{ij} = 1. \tag{2}$$

The above strategy is empirically better than the Gaussian function, as the weight matrix can be automatically calculated in a closed