Neurocomputing 267 (2017) 447–454

Contents lists available at ScienceDirect

Neurocomputing

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

Multiple kernel clustering with corrupted kernels

Teng Li*, Yong Dou, Xinwang Liu, Yang Zhao, Qi Lv

National Laboratory for Parallel and Distributed Processing, National University of Defense Technology, Changsha 410073, PR China

ARTICLE INFO

Article history:
Received 18 April 2016
Revised 14 March 2017
Accepted 11 June 2017
Available online 29 June 2017

Communicated by K. Li

Keywords:
Kernel method
Clustering
Multiple kernel clustering
Corrupted kernels

ABSTRACT

Multiple kernel clustering (MKC) algorithms usually learn an optimal kernel from a group of pre-specified base kernels to improve the clustering performance. However, we observe that existing MKC algorithms do not well handle the situation that kernels are corrupted with noise and outliers. In this paper, we first propose a novel method to learn an optimal consensus kernel from a group of pre-specified kernel matrices, each of which can be decomposed into the optimal consensus kernel matrix and a sparse error matrix. Further, we propose a scheme to address the problem of considerable corrupted kernels, where each given kernel is adaptively adjusted according to its corresponding error matrix. The inexact augmented Lagrange multiplier scheme is developed for solving the corresponding optimization problem, where the optimal consensus kernel and the localized weight variables are jointly optimized. Extensive experiments well demonstrate the effectiveness and robustness of the proposed algorithm.

1. Introduction

Clustering algorithms aim to find a meaningful grouping of samples x in an unsupervised manner for machine learning [1], computer vision [2,3] and data mining [4,5] fields. Kernel-based clustering methods, such as kernel k-means [6], has the advantage of handling non-linear separable clusters and usually achieve improved clustering performance. The performance of these kernel-based clustering methods is highly dependent on kernel selection. In a practical scenario, we can construct various kernels because many different kernel functions exist with different parameters. Since the information of these pre-specified kernels are unknown in advance, we have difficulty in selecting suitable kernels for the clustering task. Multiple kernel clustering methods, which extend the traditional single kernel method into a multiple kernel method, have been studied actively and have shown state-of-the-art results in recent years. Traditionally, multiple kernel clustering methods learn the optimal kernel through the linear or nonlinear combination of multiple base kernels. Huang et al. [7] seek an optimal combination of affinity matrices so that it is more immunity to ineffective affinities and irrelevant features. Huang et al. [8] propose a multiple kernel clustering algorithm by incorporating multiple kernels and automatically adjusting the kernel weights. In [9], the kernel weights are assigned to the information of the corresponding view and a parameter is used to control the sparsity of these weights. In [10], they propose a localized multiple kernel clustering method, which is dedicated to the dataset with varying local distributions. Gönen and Margolin [11] combine kernels calculated on the views in a localized way to better capture sample-specific characteristics of the data. Du et al. [12] present a robust multiple kernel k-means algorithm by replacing the sum-of-squared loss with $\ell_{2,1}$-norm.

Nevertheless, in a practical scenario, data may be corrupted with noise and outliers, which results in the corresponding kernel being corrupted as well. Besides, once these pre-specified kernels are corrupted with noise and outliers, the optimal kernel is likely to be corrupted, which may degrade the clustering performance consequently. Moreover, in multiple kernel clustering method, we usually construct a number of kernels to exploit the advantage of the kernel method as much as possible. If the original data have been corrupted by noise and outliers, the number of corrupted kernels will increases, which may worsen the performance of multiple kernel clustering. Since the optimal kernel is learned from these pre-specified kernels, once the number of corrupted kernels increases, a fatal effect may occur on the final optimal kernel learning. In addition, we have observed that existing multiple kernel learning methods do not sufficiently consider the corrupted situation among these kernels. This could result in learning the optimal kernel inaccurately and degrading the clustering performance. In this paper, we propose a robust multiple kernel clustering with corrupted kernels method. Compared with previous studies, our method is better at capturing the underlying structure of the data and can effectively address the situation where considerable

* Corresponding author.
E-mail addresses: liteng09@nudt.edu.cn, liteng09@163.com (T. Li), yongdou@nudt.edu.cn (Y. Dou), 1022xinwang.liu@gmail.com (X. Liu), zhaoyang10nudt@163.com (Y. Zhao), lvqi@nudt.edu.cn (Q. Lv).

http://dx.doi.org/10.1016/j.neucom.2017.06.044
0925-2312/

kernels are corrupted with noise and outliers. To the best of our knowledge, this study is the first to learn the optimal consensus kernel based on that each pre-specified kernel from a different kernel function can be decomposed into the optimal consensus kernel matrix and the sparse error matrix. In summary, we highlight the main contributions of this paper as follows:

- We propose a novel method for learning an optimal consensus kernel based on that each pre-specified kernel from a different kernel function can be decomposed into the optimal consensus kernel matrix and the sparse error matrix.
- To address the problem of considerable corrupted kernels, we enable the utilization of each given kernel to be adaptively adjusted by assigning a weight variable to each error matrix.
- The optimal consensus kernel and the localized weight variables are obtained by solving a constrained nuclear norm and an ℓ_1-norm minimization problem, in which each subproblem is convex and can be solved efficiently when optimizing one variable while fixing other variables through inexact augmented Lagrange multiplier scheme.
- We conduct comprehensive experiments to compare the proposed approach with existing state-of-the-art multiple kernel clustering methods on three benchmark datasets. The experimental results demonstrate the superiority of the proposed method.

The remainder of this paper is organized as follows. We introduce the proposed multiple kernel clustering with Corrupted Kernels algorithm (MKCCK) in Section 2. An efficient alternate optimization algorithm is proposed in Section 3, where the details of the algorithm are also provided. The discussion are presented in Section 4. We compare the clustering performance of MKCCK and state-of-the-arts multiple kernel clustering algorithms in Section 5, where the robustness study and parameter sensitivity are also presented. Finally, conclusions are drawn in the Section 6.

2. Multiple kernel clustering with corrupted kernels

In this section, we present multiple kernel clustering by learning an optimal consensus kernel from the pre-specified kernels. To address the problem of corrupted kernels, we further enable the utilization of each pre-specified kernel to be adaptively adjusted.

2.1. Formulation

Given N samples, we construct m kernels $K_1, K_2, \dots, K_m$, which are $N \times N$ matrices, from m different kernel formulations, to learn an optimal consensus kernel K from these pre-specified base kernels and cluster them into their respective groups.

To capture the underlying structure of the correct information while removing the noise information that may degrade the clustering performance from the consensus kernel, we consider the problem under the following conditions: (1) The consensus kernel is low-rank and tends to be block diagonal, with minimal noise. (2) Noise and outliers might be introduced in data acquisition, thus, a fraction of data entries might be corrupted in these base kernels. Considering these conditions, we formulate the unified optimization framework as follows:

$$\min_{K,E_p} rank(K) + \lambda \sum_{p=1}^{m} \|E_p\|_{\ell}, \quad s.t. \ \forall p, K_p = K + E_p, \tag{1}$$

where the norm $\|\cdot\|_{\ell}$ on the error matrix E depends the prior knowledge on the pattern or corruptions, and λ is a trade-off parameter between the two terms. In a practical scenario, the norm $\|\cdot\|_1$ represents the randomly element-wise corruptions. To address this problem generally, we formulate our framework based on the norm $\|\cdot\|_1$. Equipped with ℓ_1-norm of E, we can reformulate the unified optimization framework as follows:

$$\min_{K,E_p} rank(K) + \lambda \sum_{p=1}^{m} \|E_p\|_1, \quad s.t. \ \forall p, K_p = K + E_p. \tag{2}$$

The optimization problem (2) is difficult to solve because of the discrete nature of the rank function. By replacing the rank function with the nuclear norm, we convert the problem into the following convex optimization, which provides a good surrogate for the problem (2):

$$\min_{K,E_p} \|K\|_* + \lambda \sum_{p=1}^{m} \|E_p\|_1, \quad s.t. \ \forall p, K_p = K + E_p. \tag{3}$$

As the problem (3) shows, the consensus kernel K is learned from these pre-specified kernels K_p. However, we cannot guarantee that all these pre-specified kernels take the equal corrupted information in practice. Based on the assumption that some of the kernels are severely corrupted by noise and outliers, the norm value of the error matrix becomes extremely large and directly influences the objective function in problem (3). To alleviate the effect of the corrupted kernel on the consensus kernel learning and to recover the underlying structure of the consensus kernel correctly, we assign a localized weight variable α_p to each error matrix. Therefore, we formulate the unified optimization framework as

$$\min_{K,E_p,\alpha_p} \|K\|_* + \lambda \sum_{p=1}^{m} \alpha_p \|E_p\|_1 s.t. \ \forall p, \ K_p = K + E_p, \sum_{p=1}^{m} \alpha_p^{\gamma} = 1, \alpha_p \ge 0, 0 < \gamma < 1, \tag{4}$$

where $\alpha_p \ge 0$ is used to control the weight of each error matrix and $\sum \alpha_p^{\gamma} = 1$ is used to avoid a trivial solution. As observed in problem (4), once a pre-specified kernel is severely corrupted by noise or outliers, the corresponding α_p is assigned with a small weight so as to minimize the objective function. Also, once a pre-specified kernel has a good underlying data structure, the corresponding α_p is assigned with a large weight to reduce the influence induced by other corrupted kernels.

3. Alternate optimization

We propose to address the introduced equality constraints through an inexact augmented Lagrangian method (inexact ALM) [13] that can be formed as

$$L(K, E_p, \alpha_p, Y_p, \mu) = \|K\|_* + \lambda \sum_{p=1}^{m} \alpha_p \|E_p\|_1 + \sum_{p=1}^{m} \langle Y_p, K_p - K - E_p \rangle + \frac{\mu}{2} \sum_{p=1}^{m} \|K_p - K - E_p\|_F^2, \tag{5}$$

where μ is the penalty factor that controls the rate of convergence of the inexact ALM, Y_p is the Lagrange multiplier for the constrain $D_p = A + E_p$, $\langle \cdot \rangle$ denotes the inner-product operator and $\|\cdot\|_F$ is the Frobenius norm. The optimization problem can be solved by Inexact ALM. The optimization procedure is shown in Algorithm 1.

3.1. Optimizing K w.r.t E, α are fixed

When other variables are fixed, the subproblem is

$$K^{(k+1)} = \arg\min_{K} L(K, E_p^{(k)}, \alpha_p^{(k)}, Y_p^{(k)}, \mu^{(k)}) \tag{6}$$

which can be solved by Singular Value Threshold [14] method