Numerical algorithms for data processing and analysis

Chen, Chuan

Date of Award:
2016

Abstract

In this thesis, we investigate several numerical algorithms for data processing and analysis specifically the total variation (TV) based models for data processing and multi-domain data clustering.

The first part of the thesis is to study TV based methods for both image restoration and tensor decomposition problems. In Chapter 2, we introduce the image restoration problems which play an important role in many applications of sciences and engineering. Different from existing approaches, both total variation regularization and nonlinear least squares data-fitting are taken into consideration. By making use of the structure of the objective function, an efficient alternating direction method of multipliers (ADM) can be developed for solving the proposed model. Simulation study is carried out to examine the performance of the proposed method in nonlinear image restoration. The method is also applied to high dynamic rage imaging for illustration.

In Chapter 3, we further consider TV based tensor decomposition data with time dimension. Based on CANDECOMP/PARAFAC (CP) decomposition method, we employ the total variation regularization term in the time dimension to enforce the time factor constraint. We also employ ADM to solve the resulting optimization problem and study its convergence. Simulation studies are also carried out to assess the performance of our proposed method. Both video data and gene expression data analysis are conducted for illustrations.

For the second part of this thesis, we mainly focus on numerical methods for multi-domain data clustering, including multi-view data and multi-relation data. In Chapter 4, we study clustering for multiple undirected graphs over the same set of objects in different views, known as multi-view data. Different from existing approaches which seek for common clusters for all graphs, we search the clustering of objects within each graph while at the same time should be highly consistent across different graphs. We propose a simple but novel block spectral clustering method for multi-view data. We demonstrate that the multiplicity of the zero eigenvalue of the

constructed block Laplacian matrix is equal to the number of connected components in multiple graphs while the corresponding eigenvectors are solutions of the relaxation of multiple graphs cut problems. Simulations are conducted to examine the performance, while both publication data and gene expression data are analysed for illustration.

In Chapter 5, we further extent the clustering method for multi-view data to handle multi-relation data, i.e., multiple graphs over different sets of objects. With inter-relation among different graphs further considered in the block spectral clustering framework, we construct block Laplacian matrix with off-diagonal blocks representing inter-relation. By making use of its eigenvectors to perform clustering, global optimal solutions are obtained in the proposed method. Experiments on synthetic data sets are given to examine the clustering accuracy and computational time required by the proposed block clustering methods. Both publication data and gene expression data are conducted for illustration.

Table of Contents