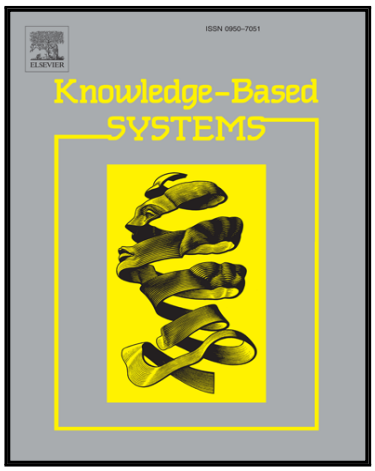

Robust co-clustering via dual local learning and high-order matrix factorization

Shiping Wang, Wenzhong Guo

PII: S0950-7051(17)30452-5
DOI: 10.1016/j.knosys.2017.09.033
Reference: KNOSYS 4054

To appear in: *Knowledge-Based Systems*

Received date: 12 December 2016
Revised date: 23 September 2017
Accepted date: 30 September 2017

Please cite this article as: Shiping Wang, Wenzhong Guo, Robust co-clustering via dual local learning and high-order matrix factorization, *Knowledge-Based Systems* (2017), doi: 10.1016/j.knosys.2017.09.033

Robust co-clustering via dual local learning and high-order matrix factorization

Shiping Wang[a,b], Wenzhong Guo[a,b,*]

[a] *College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China*
[b] *Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou University, Fuzhou 350116, China*

Abstract

Co-clustering is to group features and samples simultaneously and has received increasing attention in data mining and machine learning, particularly in text document categorization and gene expression. In this paper, two effective co-clustering algorithms are proposed to exploit the joint advantages of local learning and matrix factorization. First, the co-clustering problem is formulated as a form of matrix tri-factorization which embeds local structure learning and orthogonality constraints for clustering indicators. Using high-order matrix factorization, an effective algorithm is proposed for co-clustering problems and its convergence is proved. Second, symmetric co-clustering problems are studied, where the sample affinity matrix serves as the input matrix. Analogous high-order matrix factorization is used to develop an effective convergent algorithm for that problem. Finally, the two proposed algorithms are validated in eight publicly available real-world datasets from machine learning repository. Extensive experiments demonstrate that the proposed algorithms achieve competitive performance over existing state-of-the-art co-clustering methods in all tested datasets.

Keywords: machine learning, co-clustering, local learning, graph Laplacian, manifold learning, nonnegative matrix factorization.

1. Introduction

Clustering is one of the most important research topics in unsupervised learning and has received widespread attention in data mining, machine learning and computer vision. It is about the learning task to divide objects into clusters while minimizing cluster-within similarity as well as maximizing cluster-between discriminability. Many efficient clustering algorithms have been proposed in recent years, such as K-means clustering [15], spectral clustering [8, 37, 46] and density-based clustering [21, 23]. These clustering techniques can be mainly categorized into two types: constraint-based

*Corresponding author.
Email address: guowenzhong@fzu.edu.cn (Wenzhong Guo)