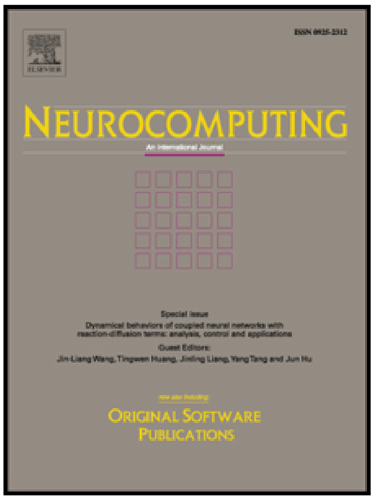
Accepted Manuscript



Graph regularized Multilayer Concept Factorization for Data Representation

Xue Li , Xiaobo Shen , Zhenqiu Shu , Qiaolin Ye , Chunxia Zhao

PII: S0925-2312(17)30119-4
DOI: 10.1016/j.neucom.2017.01.045
Reference: NEUCOM 17944

To appear in: *Neurocomputing*

Received date: 21 September 2015
Revised date: 5 December 2016
Accepted date: 17 January 2017

Please cite this article as: Xue Li , Xiaobo Shen , Zhenqiu Shu , Qiaolin Ye , Chunxia Zhao , Graph regularized Multilayer Concept Factorization for Data Representation, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.01.045

Graph regularized Multilayer Concept Factorization for Data Representation

Xue Li [a], Xiaobo Shen [a*], Zhenqiu Shu [a], Qiaolin Ye [b], Chunxia Zhao [a]

[a] *School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China*

[b] *School of Information Technology, Nanjing Forestry University, Nanjing, China*

[*]Corresponding author Tel: +86 15850749531
E-mail address: njust.shenxiaobo@gmail.com (X.Shen)

Abstract: Previous studies have demonstrated that matrix factorization techniques, such as Nonnegative Matrix Factorization (NMF) and Concept Factorization (CF), have yielded impressive results in image processing and data representation. However, conventional CF and its variants with single layer factorization fail to capture the intrinsic structure of data. In this paper, we propose a novel sequential factorization method, namely Graph regularized Multilayer Concept Factorization (GMCF) for clustering. GMCF is a multi-stage procedure, which decomposes the observation matrix iteratively in a number of layers. In addition, GMCF further incorporates graph Laplacian regularization in each layer to efficiently preserve the manifold structure of data. An efficient iterative updating scheme is developed for optimizing GMCF. The convergence of this algorithm is strictly proved; the computational complexity is detailedly analyzed. Extensive experiments demonstrate that GMCF owns the superiorities in terms of data representation and clustering performance.

Keywords: Concept Factorization, Multilayer factorization, Manifold learning, Dimensionality reduction, Data representation

1. Introduction

Recent years have witnessed large volumes of high dimensional data in face recognition [1~5], data clustering [6~10] and image retrieval [11~13], which have strongly motivated the development of dimensionality reduction. Generally speaking, processing high dimensional data is still a big challenge. Subspace learning, which dominates dimensionality reduction, has been widely exploited in image understanding and data representation. It is helpful to reveal low dimensional structure embedded in high dimensional space. Principal Component Analysis (PCA) [14] and Linear Discriminant Analysis (LDA) [15] have been the two most popular linear algorithms due to their relative simplicity and effectiveness. To detect the underlying lower dimensional manifold structure, some manifold learning methods have been proposed, such as ISOMAP [16], Locally Linear Embedding (LLE) [17], Laplacian Eigenmap [18], Locality Preserving Projections (LPP) [19] and Neighborhood Preserving Embedding (NPE)