International Journal of Innovative
Computing, Information and Control
Volume **6**, Number **8**, August **2010**

 ISSN 1349-4198
pp. 3763–3773

INSTANCE-LEVEL BASED DISCRIMINATIVE SEMI-SUPERVISED DIMENSIONALITY REDUCTION WITH CHUNKLETS

Na Wang[1], Xia Li[1], Yingjie Cui[1] and Jeng-Shyang Pan[2]

[1]College of Information Engineering
Shenzhen University
Nanhai Ave 3688, Shenzhen, P. R. China
{ wangna; lixia }@szu.edu.cn; cuiyingjie1984@163.com

[2]Department of Electronic Engineering
National Kaohsiung University of Applied Sciences
Kaohsiung 807, Taiwan
jspan@cc.kuas.edu.tw

Received May 2009; revised November 2009

Abstract. *Dimensionality reduction, an essential preprocessing step for high-dimensional data mining, has been well studied in both unsupervised and supervised learning. Semi-supervised dimensionality reduction, as the name implies, makes use of both unlabeled data and such domain knowledge as pairwise constraints which are easily available in many practical applications. Pairwise constraints specify whether a pair of instances belongs to the same class (must-link constraints) or different classes (cannot-link constraints). In this paper, an instance-level based discriminative semi-supervised dimensionality reduction method with chunklets named IDSDRC is proposed, which aims to simultaneously use both instance-level and chunklet-level pairwise constraints together with unlabeled data for dimensionality reduction. The instance-level pairwise constraints intend to exploit discrimination information, while the chunklets manage to discover clustering information. Meanwhile, unlabeled data are utilized to preserve the inherent geometrical structure of high-dimensional sample data. The proposed IDSDRC algorithm has a closed form solution. Experimental results on two standard face databases for face recognition and face clustering show that IDSDRC is efficient and superior to several established semi-supervised dimensionality reduction methods.*

Keywords: Semi-supervised learning, Dimensionality reduction, Face analysis, Manifold learning, Pairwise constraints

1. **Introduction.** Dimensionality reduction is a critical problem in many fields of information processing, such as machine learning, data mining, information retrieval, and pattern recognition [1-3]. Over the last few years, there has been considerable research on dimensionality reduction in the literature. Two of the most well-known methods are Principal Components Analysis (PCA) and Multi-dimensional Scaling (MDS), both of which are simple to implement and ensure to find the intrinsic geometry structure of data sets in the linear subspace of high-dimensional vector space. However, the linear nature of these algorithms makes them difficult to reveal the complex nonlinear manifold structures. A commonly known fact is that features used for specific purpose are not always the linear combination of features extracted from the real data [4-7]. Therefore, nonlinear dimensionality reduction sounds more appropriate to deal with data points in high-dimensional space.

Manifold learning falls into the category of nonlinear dimensionality reduction which has attracted more and more attention recently. Several efficient manifold learning techniques have been proposed, such as Isometric feature mapping (ISOMAP), Locally Linear