Engineering Applications of Artificial Intelligence 52 (2016) 161–167

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Label consistent semi-supervised non-negative matrix factorization for maintenance activities identification

Xiaodong Feng [a], Yuting Jiao [b,c,d,*], Chuan Lv [b,c,d], Dong Zhou [b,c,d]

[a] School of Political Science and Public Administration, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China
[b] Science and Technology on Reliability and Environmental Engineering Laboratory, Beijing 100191, China
[c] School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
[d] State Key Laboratory of Virtual Reality Technology and System, Beijing 100191, China

ARTICLE INFO

Article history:
Received 1 August 2015
Received in revised form
29 February 2016
Accepted 29 February 2016
Available online 24 March 2016

Keywords:
Non-negative matrix factorization
Semi-supervised learning
Label consistent regularization
Maintenance activities identification
PHM data challenge

ABSTRACT

Health prognostic is playing an increasingly essential role in product and system management, for which non-negative matrix factorization (NMF) has been an effective method to model the high dimensional recorded data of the device or system. However, the existing unsupervised and supervised NMF models fail to learn from both labeled and unlabeled data together. Therefore, we propose a label consistent semi-supervised non-negative matrix factorization (LCSSNMF) framework that can simultaneously factorize both labeled and unlabeled data, where the discriminability of label data is preserved. Specifically, it firstly incorporates a class-wise coefficient distance regularization term that makes the coefficients for similar samples or samples with the same label close. Moreover, a label reconstruction regularization term is also presented, as the classification error with coefficient matrix of labeled data is expected as low as possible, which will potentially improve the classification accuracy in maintenance activities identification for industrial remote monitoring and diagnostics. The experiment results on real maintenance activities identification application from PHM 2013 data challenge competition demonstrate that LCSSNMF outperforms the state-of-arts NMF methods and results provided by the competition.

1. Introduction

Prognostic and health management (PHM), as an integrated product management technique which focuses on the product health, is playing growingly essential role in product and system management. For complex system, effective health prognostic indicates great benefit for system safety, reliability and maintenance, as unexpected system breakdown will lead to huge loss for industry. Thus, it is vital to reduce system breakdown and assess system health states correctly and immediately. There are mainly two categories of methods applied in health prognostic, namely machine learning-based technique and statistical inference technique. The machine learning-based approaches, seeing the health prognostic as essentially a problem of pattern recognition, focus on more effective feature extraction method and more accurate classifier to obtain higher prognostic accuracy. They can be further divided into supervised (Alguindigue et al., 1993), unsupervised (Yang et al., 2011) and semi-supervised techniques (Jiang et al., 2013).

Actually, supervised learning could only utilize labeled data that cannot be automatically obtained in real application and the unsupervised learning is often imprecise without labeled data instruction if the different patterns partly overlap over health prognostic data. Semi-supervised learning, combining techniques of unsupervised and supervised learning, is one of the key tasks in pattern recognition and machine learning, which trains a model together with labeled and unlabeled data to improve recognition accuracy. Recently, semi-supervised learning is widely used in PHM filed. For example, Jiang et al. (2013) proposed a semi-supervised kernel marginal fisher analysis (SSKMFA) for feature extraction, which can discover the intrinsic manifold structure of dataset, and simultaneously considers the intra-class compactness and the inter-class separately; Yuan and Liu (2013) introduces manifold regularization based semi-supervised learning into fault detection by utilizing both labeled and unlabeled CM data. Besides, the observed data in prognostic and health management is often of high dimension, which often gives rise to the increase of storage space and computational cost. Thus, dimensionality reduction, helping to avoid the curse of dimensionality, has captured more

* Corresponding author at: Science and Technology on Reliability and Environmental Engineering Laboratory, China.
E-mail address: jytniu@hotmail.com (Y. Jiao).

http://dx.doi.org/10.1016/j.engappai.2016.02.016
0952-1976/

attention for data representation. Dimension reduction, which transforms the high-dimensional data into low dimensional space, can be regarded as a decisive step for intelligent fault diagnosis systems. Over past years, many techniques including principle component analysis (PCA) (Li et al., 2003), self-organizing map (SOM) (Yu et al., 2015), independent component analysis (ICA) (Guo et al., 2014), support vector machine (SVM) (Widodo et al., 2007), local linear embedding (LLE) (Demetgul et al., 2014), etc. have been employed as a dimensionality reduction technique for processing the fault diagnostic problems.

Different from the above methods, non-negative matrix factorization (NMF) (Lee and Seung, 1999) is another powerful technique for processing high-dimensional data, and has been successfully applied in a variety of areas such as text mining (Barman et al., 2006), facial image recognition (Xue et al., 2014), disease diagnostic (Zhang et al., 2013), sound source separation (Dessein et al., 2013), etc. Also it has been applied in PHM tasks. For example, Qinghua et al. (2009) combines NMF that is used to decompose multivariate data and neural network ensemble (NNE) for fault diagnostic of diesel valve trains; Li et al. (2011a) presents a novel feature extraction scheme for roller bearing fault diagnosis utilizing generalized S transform and two-dimensional non-negative matrix factorization, which can reduce the computation cost and preserve more structure information hiding in original 2D matrices compared to the NMF; Li et al. (2011b) also presents an algorithm based on S transform, non-negative matrix factorization (NMF), mutual information and multi-objective evolutionary for hybrid fault diagnosis of gearbox.

Even though some unsupervised NMF models (Zhang et al., 2012; Cai et al., 2011) perform well, they do not use any discriminative information contained in the dataset and cannot perform well for classification or recognition tasks due to the lack of discriminative information. Therefore, several supervised methods for enhancing the classification accuracy of NMF algorithm are presented (Zafeiriou et al., 2006; Nikitidis et al., 2012), and the experimental results have shown the successful application of supervised NMF algorithms. However, these NMF algorithms only factorize the training data and thus assume the input training data to be completely labeled, which are not the best for partial labeled data. In real world, there are labeled and unlabeled data. Thus, in order to utilize the labeled data and unlabeled data simultaneously, a host of semi-supervised NMF (SSNMF) algorithms have been proposed (Chen et al., 2008; Guan et al., 2011, 2012; Liu, 2012), which implement the dimension reduction on both the labeled and unlabeled data at the same time in an identical model. However, the labeled data is not completely exploited when training the model as only pair-wise discriminative information is preserved, motivating us to focus on more effective semi-supervised model for PHM.

In this investigation, a semi-supervised non-negative matrix factorization is proposed for health diagnostic of industrial assessment, which can incorporate the labeled data information and classification task. The proposed label consistent semi-supervised non-negative matrix factorization (LCSSNMF) is based on a novel label reconstruction regularization term, as the classification error with coefficient matrix of labeled data is expected as low as possible. The experiment results on real maintenance activities identification application demonstrate the improvement of LCSSNMF, compared to the state-of-arts NMF methods the results provided by the competition. The rest of the paper is organized as follows. Some representative and related works about NMF, i.e., non-negative matrix factorization, graph regularized non-negative matrix factorization and graph based semi-supervised non-negative matrix factorization, are introduced in Section 2. Proposed label consistent semi-supervised non-negative matrix factorization is detailed in Section 3 with optimization algorithm in Section 4. Section 5 shows the data reparation and experiment results of the proposed method with Section 6 concluding our work.

2. Related works and preliminaries

2.1. Non-negative matrix factorization

Given a set of n input samples $X=[x_1,x_2,\ldots,x_n]^T\in\mathbb{R}^{n\times m}$, where each sample $x_i\in\mathbb{R}^m$ is with m attributes. Non-negative matrix factorization (NMF) (Lee and Seung, 1999) tries to learn a base matrix $Q=[q_1,q_2,\ldots,q_r]^T\in\mathbb{R}^{r\times m}$ and a coefficient matrix $P=[p_1,p_2,\ldots,p_r]^T\in\mathbb{R}^{n\times r}$, where $q_i\in\mathbb{R}^m$ is the base vector and $p_i\in\mathbb{R}^r$ is the representation of each sample x_i on base matrix Q, so that:

$$X=PQ, \tag{1}$$

where, r, as the number of base vectors in Q, should satisfy that $(m+n)r\ll mn$. The goal of NMF is to make PQ as close as possible to X.

Thus, the objective of NMF can be described as:

$$\min_{P,Q} J_{NMF}=D(X\|PQ) \quad \text{s.t. } P\geq 0,\ \ Q\geq 0, \tag{2}$$

where $D(X\|PQ)$ represents the divergence between X and PQ, and if we use l_2 norm distance between two matrices, the loss function J_{NMF} in Eq. (2) is defined as:

$$J_{NMF}=\|X-PQ\|^2. \tag{3}$$

The aforementioned objective function can be minimized by the iterative update algorithm proposed by Lee and Seung (1999).

2.2. Graph regularized non-negative matrix factorization

Though NMF has been widely used in various pattern recognition tasks and gains well performance in application, original NMF ignores the statistical information such as geometry structure and discrimination information in the dataset. Therefore, graph regularized non-negative matrix factorization (GNMF), proposed by Cai et al. (2011), will minimize the distances between each sample and its k-nearest neighbors in the reduced coefficient space P. Based on manifold assumption in GNMF, if two data samples share similar geometry distribution in high dimensional space, the coefficient vectors of these two samples with respect to the new basis should also be close to each other, and vice versa. Studies prove that when the manifold assumption is unknown, nearest neighbor graph is an effective method to discrete approximation, illustrated as follows.

To preserve the local geometrical structure of the whole dataset, a graph $G=\{X,S\}$ can be built, where $S\in\mathbb{R}^{n\times n}$ is the weight matrix between pair-wise sample, defined as:

$$S_{ij}=\begin{cases}1, & x_i\in N_k(x_j) \quad \text{or} \quad x_j\in N_k(x_i)\\ 0, & \text{otherwise}\end{cases} \tag{4}$$

where $N_k(x_i)$ denotes k nearest neighbor set of x_i.

Then, with weight matrix S, the local structure can be obtained through the graph Laplacian regularization term to measure the smoothness of low-dimensional coding vector representations in P as below:

$$J_{\mathrm{Gr}}=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}S_{ij}\|p_i-p_j\|_2^2=\mathrm{Tr}(P^TL_SP), \tag{5}$$

where $L_S=D_S-S$ is the graph Laplacian matrix of G, and D is a diagonal matrix whose diagonal element is the sum of the current row vector of S, i.e., $(D_S)_{ii}=\sum_{j=1, j\neq i}^{n}S_{ij}$; $\mathrm{Tr}(\cdot)$ stands for the trace of a square matrix, meaning the sum of the diagonal entries.