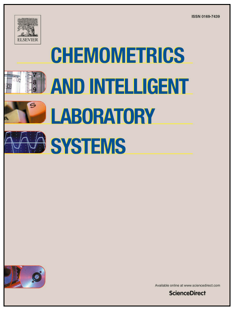
Accepted Manuscript



Active learning based semi-supervised exponential discriminant analysis and its application for fault classification in industrial processes

Jun Liu, Chunyue Song, Jun Zhao

PII: S0169-7439(18)30184-9

DOI: 10.1016/j.chemolab.2018.07.003

Reference: CHEMOM 3650

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received Date: 27 March 2018

Revised Date: 31 May 2018

Accepted Date: 5 July 2018

Please cite this article as: J. Liu, C. Song, J. Zhao, Active learning based semi-supervised exponential discriminant analysis and its application for fault classification in industrial processes, *Chemometrics and Intelligent Laboratory Systems* (2018), doi: 10.1016/j.chemolab.2018.07.003.

Active Learning Based Semi-supervised Exponential Discriminant Analysis and Its Application for Fault Classification in Industrial Processes

Jun Liu, Chunyue Song* and Jun Zhao

State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, China

Abstract: For the industrial fault classification, exponential discriminant analysis (EDA) requires that all the training samples should be labeled; however, only a minority of the training samples are randomly labeled in real industrial processes. This motivates the formulation of the active learning based semi-supervised exponential discriminant analysis in this paper. Firstly, to make EDA applicable to the semi-supervised industrial scenario, scatter matrices are transformed into their regularization variants through combining unlabeled training samples. Moreover, to reduce the adverse effect of random labeling of training samples, the best versus second-best rule is employed to select more informative training samples in an active way to upgrade the model classification performance. And the obvious performance improvement of the proposed method is demonstrated with extensive experiments on synthesized data, the TE process and a real industrial process.

Key Words: discriminant analysis; semi-supervised; matrix exponential; active learning; industrial fault classification

1. Introduction

In real industrial processes, process monitoring plays a significant role in ensuring the manufacturing safety and product quality, and its topics usually includes fault detection, fault diagnosis and fault classification. The related methods can be divided into three categories, i.e., the model-based, knowledge-based and data-driven methods [1-3]. Because the rapid development of information technologies such as the distributed control system (DCS) and internet of things (IoT) has laid a solid foundation for the acquisition, transmission, storage and analysis of process data, data-based process monitoring techniques have been receiving significant interests recently [4]. Usually, the multivariate statistical process monitoring (MSPM) methods such as principal components analysis (PCA), partial least squares (PLS), independent component analysis (ICA) and their corresponding variants can achieve satisfying fault detection and diagnosis performance in complex real industrial processes [5-8]; however, they ignore the relationships between different process operation modes during dimensionality reduction, and thus are not fit for fault classification. In contrast, discriminant analysis, which is also called Fisher discriminant analysis (FDA) in most cases, can find the projection matrix that maximizes