Dynamics of Continuous, Discrete and Impulsive Systems
Series **B:** *Applications & Algorithms* **15 (2008)** 675-682
Copyright ©2008 Watam Press http://www.watam.org

NONLINEAR KERNEL SUPPORT VECTOR MACHINE BASED RECURSIVE FEATURE SELECTION ALGORITHM FOR KEY VARIABLE IDENTIFICATION

Y. Mao[1,2] and Y.Y. Cheng[2] and Y.X. Sun[1]

[1]National Laboratory of Industrial Control Technology,
Zhejiang University, Hangzhou, P. R. China

[2]Pharmaceutical Informatics Institute, College of Pharmaceutical Sciences,
Zhejiang University, Hangzhou, P. R. China

Abstract. Key variable identification is related to many trouble-shooting problems in process industries. Linear kernel support vector machine based recursive feature elimination is proved to be an effective algorithm to do key variable identification. But when variables are nonlinear relative, it is difficult to work. In this paper, nonlinear kernel support vector machine based recursive feature selection is proposed to select key variables with nonlinear relation or linear relation. The experimental results on dataset from Tennessee Eastman process (TEP) simulator proved the effectiveness of our method.

Keywords. Fault diagnosis, Support vector machine, Recursive feature elimination.

AMS (MOS) subject classification: 68T37,62H30,62H20.

1 Introduction

Key variable identification is a hot topic in industrial data mining field; selecting the key variables driving the difference between fault and normal system status allows rapid recovery from process faults so it is crucial for production of consistent quality product. So it is very important for running chemical plants.

Support vector machine based recursive feature elimination (SVM-RFE) is an effective feature selection method successfully applied in chemical process fault diagnosis to identify key variables [1], when it is compared with some other key variable identification method, such as PCA combined contribution charts or genetic algorithm combined Fisher discriminant analysis propose in [2]. In [1], only linear kernel SVM-RFE is mentioned, it is work well when key variables are linearly relative. But when key variable are nonlinearly relative, it doesn't work. In this paper, Gaussian kernel SVM-RFE is proposed to deal with this problem. Cross-validation is used to evaluate the effectiveness of specific number of ranked variables. Gaussian kernel SVM-RFE and linear kernel SVM-RFE are compared on two cases from Tennessee