International Journal of Innovative Computing, Information and Control
Volume **6**, Number **3(A)**, March **2010**

ICIC International ©2010 ISSN 1349-4198
pp. **1089–1099**

THE OPTIMIZED SUPPORT VECTOR MACHINE WITH CORRELATIVE FEATURES FOR CLASSIFICATION OF NATURAL SPEARMINT ESSENCE

Xiaofeng Song[1], Saman K. Halgamuge[2], Dezhao Chen[3]
Shangxu Hu[3] and Bin Jiang[1]

[1]College of Automation Engineering
Nanjing University of Aeronautics and Astronautics
No. 29 Yudao St., Nanjing 210016, P. R. China
{ binjiang; xfsong }@nuaa.edu.cn

[2]Biomechanical Engineering Group
Department of Mechanical Engineering
Melbourne University
VIC 3010, Australia
saman@unimelb.edu.au

[3]Department of Chemical Engineering
Zhejiang University
Hangzhou 310027, P. R. China
{ dzchen; sxhu }@mail.hz.zj.cn

Received September 2008; revised February 2009

Abstract. *The performance of support vector machine (SVM) hybridized with two other methods for classification of chemical patterns was investigated. It was found that SVM for classification can be sensitive to noise and be affected by multicollinearity between attributes similar to other methods such as multivariable analysis and neural networks. The kernel function, its parameter and penalty factor C are the main factors affecting the classification performance of SVM. Correlative component analysis (CCA) was used to eliminate multicollinearity and noise of original sample data before classified by SVM. To improve the classification performance of SVM and obtain the optimal discriminate function, Eugenic Genetic Algorithm (EGA) was used to optimize the parameters of SVM. Finally, a typical example consisting of two classes of natural spearmint essence was employed to verify the effectiveness of the new hybridized approaches including CCA-EGA-SVM. The classification accuracy of this new method is much better than that obtained by SVM, CCA-SVM, CCA-SOM, and GA-CG-SVM.*
Keywords: Pattern recognition, Complex chemical patterns, Support vector machine, Correlative component analysis, Eugenic genetic algorithms

1. **Introduction.** Pattern classification or recognition is often used in chemical engineering and analytical chemistry, e.g., fault diagnosis, quality monitoring, manifold recognition and recipe design. Constituents of many organic materials, usually consisting of a great number of chemical components, are so complicated that it is difficult to ascertain their quantitative structure-property relationship (QSPR). For example, the determination of QSPR of spice, alcohol, pesticides, which cannot be described by chemical mechanism, depend on pattern recognition based on data collected through observations. Since the advanced analytical instruments make the precise description of chemical material possible, the observation sample data may contain abundant information, that is, data with a sample pattern can be stretched in high dimensionality space. Sometimes, severe collinearity