International Journal of Innovative
Computing, Information and Control
Volume **7**, Number **9**, September **2011**

ICIC International ©2011 ISSN 1349-4198
pp. 5133–5143

STRUCTURE SELECTION FOR DAG-SVM BASED ON MISCLASSIFICATION COST MINIMIZATION

Hui Yi, Xiaofeng Song and Bin Jiang

College of Automation Engineering
Nanjing University of Aeronautics and Astronautics
No. 29, Yudao Street, Baixia Distinct, Nanjing 210016, P. R. China
{ binjiang; xfsong }@nuaa.edu.cn

Received April 2010; revised September 2010

Abstract. *DAG-SVM is one of the most popular multi-class support vector machines, but it has k! kinds of decision structures while dealing with k-class problems. How to select an appropriate structure is a key problem in DAG-SVM. In this paper, the classification bias of DAG-SVM was first discussed, and then the concept of misclassification cost was introduced in structure selection for DAG-SVM. Based on the principle of Misclassification Cost Minimization, a structure selection method for DAG-SVM was proposed to minimize the risk of total misclassification cost. The proposed method was tested on several benchmark data sets, and the results indicated that MCM-based selecting method for DAG-SVM obtained satisfactory results.*
Keywords: Misclassification cost, DAG-SVM, Structure selection

1. **Introduction.** Support vector machine (SVM) proposed by Vapnik [1] is an intelligent classification method based on Structure Risk Minimization (SRM). Compared with neural network, SVM can be especially applied to high dimensional and small samples, and cannot fall into local minima. Therefore, SVM has been widely employed by scientists in fields like letter recognition, text categorization and fault diagnosis. In such fields, there are usually more than two types of samples, and multi-type classifications are required. However, SVM is originally designed for binary classification, and it could not be directly applied to such fields. In order to solve these problems, serval extension strategies for multiple-SVM [2-4], such as one against one, one against rest and decision tree were proposed, among which Support Vector Machine based on Decision Acyclic Graph (DAG-SVM) proposed by Plantt [5] is a high performance and efficient multi-class SVM classifier.

DAG-SVM benefits the essential point of Decision Directed Acyclic Graph in graph theory, and can avoid the decision redundancy, sample unbalance, and blind regions problems that usually exist in conventional extension strategies for multi-type classification [5]. However, for k-class classification problems ($k > 2, k \in R$), there are $k!$ kinds of candidate structures for selection in decision tree of DAG-SVM, and different decision tree structures can lead to different classification results. How to select a suitable decision structure is the key point when employing DAG-SVM method.

Currently, some selecting methods have already been proposed [6-8], and these methods effectively improved the classification accuracy. However, what should be noted is that even with the same classification accuracy, the costs of misclassification may vary a lot. Taking the diagnosis of cancer for example, the cost of false negative may be life which is much bigger than the cost of false alarm. In this work, we introduced the cost-sensitive