International Journal of Innovative
Computing, Information and Control
Volume **6**, Number **4**, April **2010**

 ISSN 1349-4198
pp. 1605–1614

AN IMPROVED MULTI-LABEL CLASSIFICATION METHOD BASED ON SVM WITH DELICATE DECISION BOUNDARY

Benhui Chen, Liangpeng Ma and Jinglu Hu

Graduate School of Information, Production and Systems
Waseda University
Hibikino 2-7, Wakamatsu-ku, Kitakyushu-shi, Fukuoka 808-0135, Japan
bhchen@fuji.waseda.jp; jinglu@waseda.jp

Received February 2009; revised July 2009

Abstract. *Multi-label classification problem is an extension of traditional multi-class classification problem in which the classes are not mutually exclusive and each sample may belong to several classes simultaneously. Such problems occur in many important applications. Some researches indicate that the performance of classifier can be improved by introducing the information of multi-label training samples into learning procedure effectively. In this paper, we propose a novel method based on SVM with delicate decision boundary. For the basic overlapping problem of two labels, characteristics of double-label samples are utilized to obtain the range of overlapping sample space decided by two binary SVM classifier separating surfaces. And a bias model with delicate decision boundary is built for samples in overlapping sample space to improve the classification accuracy. Experimental results on the benchmark datasets of Yeast and Scene show that our proposed method improves the classification accuracy efficiently, compared with the basic binary SVM method and some existing well-known methods.*
Keywords: Multi-label classification, Support vector machine, Probabilistic outputs of SVM, Delicate decision boundary

1. **Introduction.** Multi-label classification problem is an extension of traditional multi-class classification problem in which its classes are not mutually exclusive and each sample may belong to several classes simultaneously. It is increasingly required by many real-world applications. For instance, in text categorization, a document generally has several different topics, such as *social*, *sport* and *health* [1, 2]; in bioinformatics, each gene may be associated with a set of functional classes, such as *metabolism*, *transcription* and *protein synthesis* [3]; in scene classification, each scene image may belong to several semantic classes, such as *beach* and *city* [4]. In all these cases, instances in the training set are associated with a set of labels and the task is to predict the label set for the unseen instances. The main challenge of this problem is that classes are usually overlapped and correlated. Generally, traditional multi-class learning algorithms cannot work with multi-label problem effectively.

Since Support Vector Machine (SVM) based methods have good generalization ability in single-label multi-class problem[5, 6], more attention also has been paid on such kind of techniques for multi-label problem. At present, there are mainly two types of SVM-based methods to solve the multi-label problem. One is to consider all samples and their labels simultaneously to construct one optimization formulation, for example, rank-SVM [3] and maximal margin labeling algorithm [7]. But it is time-consuming to solve such large scale optimization problems in numerical computation. The other is to decompose a multi-label problem into many binary class sub-problems and to solve them by using SVM-like methods. Generally speaking, the latter runs faster than the former does.