International Journal of Innovative Computing, Information and Control
Volume **6**, Number **5**, May **2010**
 ISSN 1349-4198
pp. **2245–2262**

MULTI-KERNEL SUPPORT VECTOR CLUSTERING FOR MULTI-CLASS CLASSIFICATION

Chi-Yuan Yeh, Chi-Wei Huang and Shie-Jue Lee*

Department of Electrical Engineering
National Sun Yat-Sen University
Kaohsiung 804, Taiwan
*Corresponding author: leesj@mail.ee.nsysu.edu.tw

Received November 2008; revised March 2009

Abstract. Applying *Support vector clustering (SVC) to multi-class classification problems has difficulty in determining the hyperparameters of the kernel functions. Multi-kernel learning has been proposed to overcome this difficulty, by which kernel matrix weights and Lagrange multipliers can be simultaneously derived with semidefinite programming. However, the amount of time and space required is very demanding. We develop a two-stage multi-kernel learning algorithm which conducts sequential minimal optimization and gradient projection iteratively. One multi-kernel SVC is constructed for the patterns of each class. The outputs obtained by all the multi-kernel SVCs are integrated and a discriminant function is applied to make the final multi-class decision. Experimental results on data sets taken from UCI and Statlog show that the proposed approach performs better than other methods.*
Keywords: Multi-class classification, Support vector clustering, Multi-kernel learning, SMO, Gradient projection

1. **Introduction.** Support vector machines have been shown to be an effective tool for solving classification problems [1, 2, 3, 4, 5, 6, 7], regression problems [8, 9, 10, 11] and clustering problems [13, 12, 14]. Support vector machines [1] were designed originally for binary classification. However, real-world applications usually involve multiple classes. To effectively extend binary support vector machines to solve multi-class classification problems is still an on-going research issue. Several solutions have been proposed to solve multi-class classification problems [1, 15, 16, 17]. Some solutions decompose a multi-class problem into a set of two-class sub-problems and then use a discriminant function to make the final decision. These include methods like One-Against-All (OVA) [15], One-Against-One (OAO) [1, 16] and Directed Acyclic Graph (DAG) [17]. Other methods have also been proposed, e.g., support vector clustering (SVC) [12], support vector data description (SVDD) [13] and one-class support vector machine (OCSVM) [14], for multi-class problems [18, 19]. In these methods, one SVC (SVDD, OCSVM) is derived from the training patterns of each class. The size of the kernel matrix and the training time of the method are less than that of OVA, OAO and DAG. However, discriminant functions for making decisions based on competing SVC (SVDD, OCSVM) classifiers are required.

Selecting proper kernel functions and hyperparameters is an important issue in kernel based applications [20, 21, 22, 23, 24]. Unsuitable kernel functions or hyperparameters can lead to a relatively poor performance. This problem is usually taken care of by a trial-and-error approach. Furthermore, a typical SVC (SVDD, OCSVM) application usually uses the same hyperparameter settings for each class. This may not be a good idea when pattern distributions are significantly different among different classes. For example, one class may contain patterns with a dense distribution where a kernel with a small variance