Data Mining and Mathematical Programming
October 10-13, 2006

A Column Generation Approach for Support Vector Machines

Dolores Romero Morales
Dolores.Romero-Morales@sbs.ox.ac.uk
Saïd Business School
University of Oxford
Park End Street
Oxford, OX1 1HP
UNITED KINGDOM

Abstract

The widely used Support Vector Machine (SVM) method has shown to yield good results in Supervised Classification problems. Other methods such as Classification Trees have become more popular among practitioners than SVM thanks to their interpretability, which is an important issue in Data Mining.

In this work, we propose an SVM-based method, based on the concept of threshold functions, that automatically detects which predictor variables, and interactions are the highly relevant for the classifier. The proposed method is robust, i.e., it is stable in the presence of outliers and invariant to change of scale or measurement units of the predictor variables. The method involves the optimization of a Linear Programming problem with a large number of decision variables, for which we use the well-known Column Generation technique.

From the computational results we have run, we conclude that the classification ability of our method is comparable to standard SVM for different kernels and clearly better than Classification Trees. Finally, we illustrate how this method is promising in a hotel industry application.

Joint work with Emilio Carrizosa and Belén Martn Barragán.