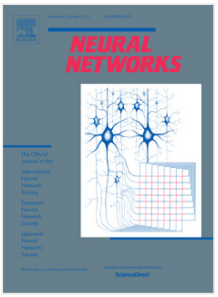
Accepted Manuscript



A divide-and-combine method for large scale nonparallel support vector machines

Yingjie Tian, Xuchan Ju, Yong Shi

PII: S0893-6080(15)00253-1
DOI: http://dx.doi.org/10.1016/j.neunet.2015.11.008
Reference: NN 3559

To appear in: *Neural Networks*

Received date: 7 April 2015
Revised date: 19 August 2015
Accepted date: 17 November 2015

Please cite this article as: Tian, Y., Ju, X., & Shi, Y. A divide-and-combine method for large scale nonparallel support vector machines. *Neural Networks* (2015), http://dx.doi.org/10.1016/j.neunet.2015.11.008

A Divide-and-Combine Method for Large Scale Nonparallel Support Vector Machines

Yingjie Tian[a], Xuchan Ju[a,b,*], Yong Shi[a]

[a]*Research Center on Fictitious Economy & Data Science, Chinese Academy of Sciences, Beijing 100190, China*
[b]*School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 101408, China*

Abstract

Nonparallel Support Vector Machine (NPSVM) which is more flexible and has better generalization than typical SVM is widely used for classification. Although some methods and toolboxes like SMO and libsvm for NPSVM are used, NPSVM is hard to scale up when facing millions of samples. In this paper, we propose a divide-and-combine method for large scale nonparallel support vector machine (DCNPSVM). In the division step, DCNPSVM divide samples into smaller sub-samples aiming at solving smaller subproblems independently. We theoretically and experimentally prove that the objective function value, solutions, and support vectors solved by DCNPSVM are close to the objective function value, solutions, and support vectors of the whole NPSVM problem. In the combination step, the sub-solutions combined as initial iteration points are used to solve the whole problem by global coordinate descent which converges quickly. In order to balance the accuracy and efficiency, we adopt a multi-level structure which outperforms state-of-the-art methods. Moreover, our DCNPSVM can tackle unbalance problems efficiently by tuning the parameters. Experimental results on lots of large data sets show the effectiveness of our method in memory usage, classification accuracy and time consuming.

Keywords: Support vector machine, nonparallel support vector machine, large scale, clustering, divide, combine.

*Corresponding author
Email addresses: tyj@ucas.ac.cn (Yingjie Tian), juxuchan@126.com (Xuchan Ju), yshi@ucas.ac.cn (Yong Shi)