



# CREaTE

Canterbury Research and Theses Environment

Canterbury Christ Church University's repository of research outputs

<http://create.canterbury.ac.uk>

Please cite this publication as follows:

Guo, W., Khalid, N., Liu, Y., Li, M. and Qi, M. (2015) A resource aware MapReduce based parallel SVM for large scale image classification. *Neural Processing Letters*, 44 (1). pp. 161-184. ISSN 1370-4621.

Link to official URL (if available):

<https://dl.acm.org/citation.cfm?id=J580&picked=prox>

This version is made available in accordance with publishers' policies. All material made available by CReaTE is protected by intellectual property law, including copyright law. Any use made of the contents should comply with the relevant law.

Contact: [create.library@canterbury.ac.uk](mailto:create.library@canterbury.ac.uk)



## A Resource Aware MapReduce based Parallel SVM for Large Scale Image Classifications

### Abstract

Machine learning techniques have facilitated image retrieval by automatically classifying and annotating images with keywords. Among them support vector machines (SVMs) are used extensively due to their generalization properties. However, SVM training is notably a computationally intensive process especially when the training dataset is large. This paper presents RASMO, a resource aware MapReduce based parallel SVM algorithm for large scale image classifications which partitions the training data set into smaller subsets and optimizes SVM training in parallel using a cluster of computers. A genetic algorithm based load balancing scheme is designed to optimize the performance of RASMO in heterogeneous computing environments. RASMO is evaluated in both experimental and simulation environments. The results show that the parallel SVM algorithm reduces the training time significantly compared with the sequential SMO algorithm while maintaining a high level of accuracy in classifications

full article attach later