Canterbury Christ Church University's repository of research outputs

http://create.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