

Student Research Symposium 2003

Jian Zhang

Modified Logistic Regression: An Approximation to SVM and Its Applications in Large-Scale Text Categorization

Logistic Regression (LR) has been widely used in statistics for many years, and has received extensive study in machine learning community recently due to its close relations to Support Vector Machines (SVM) and AdaBoost. In this paper, we use a modified version of LR to approximate the optimization of SVM by a sequence of unconstrained optimization problems. We prove that our approximation will converge to SVM, and propose an iterative algorithm called "MLR-CG" which uses Conjugate Gradient as its inner loop. Multiclass version "MMLR-CG" is also obtained after simple modifications. We compare the MLR-CG with SVM_light over different text categorization collections, and show that our algorithm is much more efficient than SVM_light when the number of training examples is very large. Results of the multiclass version MMLR-CG is also reported.