Performance Analysis of k-NN on High Dimensional Datasets

Year of Publication: 2011

Authors:
Pradeep Mewada
Jagdish Patil

Abstract

Research on classifying high dimensional datasets is an open direction in the pattern recognition yet. High dimensional feature spaces cause scalability problems for machine learning algorithms because the complexity of a high dimensional space increases exponentially with the number of features. Recently a number of ensemble techniques using
different classifiers have proposed for classifying the high dimensional datasets. The task of these techniques is to detect and exploit relevant patterns in data for classification. The k-nearest neighbor (k-NN) algorithm is amongst the simplest of all machine learning algorithms. This paper discusses various ensemble k-NN techniques on high dimensional datasets. The techniques mainly include: Random Subspace Classifier (RSM), Divide & Conquer Classification and Optimization using GA (DCC-GA), Random Subsample ensemble (RSE), Improving Fusion of dimensionality reduction (IF-DR). All these approaches generates relevant subset of features from original set and the results is obtain from combined decision of ensemble classifiers. This paper presents an effective study of improvements on ensemble k-NN for the classification of high dimensional datasets. The experimental result shows that these approaches improve the classification accuracy of the k-NN classifier.

Reference


Index Terms

Computer Science

Pattern Recognition

Key words

k-Nearest Neighbor

Ensemble Classifiers

High Dimensional

Feature Space