World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:14, No:12, 2020

Parallel Fuzzy Rough Support Vector Machine for Data Classification in Cloud Environment

Authors : Arindam Chaudhuri

Abstract : Classification of data has been actively used for most effective and efficient means of conveying knowledge and information to users. The prima face has always been upon techniques for extracting useful knowledge from data such that returns are maximized. With emergence of huge datasets the existing classification techniques often fail to produce desirable results. The challenge lies in analyzing and understanding characteristics of massive data sets by retrieving useful geometric and statistical patterns. We propose a supervised parallel fuzzy rough support vector machine (PFRSVM) for data classification in cloud environment. The classification is performed by PFRSVM using hyperbolic tangent kernel. The fuzzy rough set model takes care of sensitiveness of noisy samples and handles impreciseness in training samples bringing robustness to results. The membership function is function of center and radius of each class in feature space and is represented with kernel. It plays an important role towards sampling the decision surface. The success of PFRSVM is governed by choosing appropriate parameter values. The training samples are either linear or nonlinear separable. The different input points make unique contributions to decision surface. The algorithm is parallelized with a view to reduce training times. The system is built on support vector machine library using Hadoop implementation of MapReduce. The algorithm is tested on large data sets to check its feasibility and convergence. The performance of classifier is also assessed in terms of number of support vectors. The challenges encountered towards implementing big data classification in machine learning frameworks are also discussed. The experiments are done on the cloud environment available at University of Technology and Management, India. The results are illustrated for Gaussian RBF and Bayesian kernels. The effect of variability in prediction and generalization of PFRSVM is examined with respect to values of parameter C. It effectively resolves outliers' effects, imbalance and overlapping class problems, normalizes to unseen data and relaxes dependency between features and labels. The average classification accuracy for PFRSVM is better than other classifiers for both Gaussian RBF and Bayesian kernels. The experimental results on both synthetic and real data sets clearly demonstrate the superiority of the proposed technique.

Keywords : FRSVM, Hadoop, MapReduce, PFRSVM
Conference Title : ICSRD 2020 : International Conference on Scientific Research and Development
Conference Location : Chicago, United States
Conference Dates : December 12-13, 2020