عنوان مقاله:

A Novel Classification Method: A Hybrid Approach Based on Large Margin Nearest Neighbor Classifier

نویسندگان:

Alieh Ashoorzadeh - *Department of Information Technology Management, Science and Research Branch, Islamic Azad University, Tehran, Iran*

Abbas Toloie Eshlaghy - *Department of Industrial Management, Science and Research Branch, Islamic Azad University, Tehran, Iran*

Mohammad Ali Afshar Kazemi - *Department of Industrial Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran*

خلاصه مقاله:

Classification is the operation of dividing various data into multiple classes where they share quantitative and qualitative similarities. Classification has many use cases in engineering fields such as cloud computing, power distribution, and remote sensing. The accuracy of many classification techniques such as k-nearest neighbor (k-NN) is highly dependent on the method used in the calculation of distances between samples. It is assumed that samples close to each other belong to the same class while samples that belong to different classes have a large distance between them. One of the popular distance calculation methods is the Mahalanobis distance. Many methods, including large margin nearest neighbor (LMNN), have been proposed to improve the performance of k-NN in recent years. Our proposed method aims to introduce a cost function to calculate data similarities while solving the local optimum pitfall of LMNN and optimizing the cost function determining distances between instances. Although k-NN is an efficient classification technique that is simple to comprehend and use, it is costly to compute for large datasets and sensitive to outlier data. Another difficult feature of k-NN is that it can only measure distance in Euclidean space. The distance metric should ideally be modified to fit the specific needs of the application. Due to the disadvantages in k-NN and LMNN methods, to optimize the objective function to calculate distances for the test data and to improve classification accuracy, we initially use the genetic algorithm to reduce the range of the solution space and then by using the gradient descent the optimal values of parameters in the cost function is obtained. Our method is carried out on different benchmark datasets with varying numbers of attributes and the results are compared to k-NN and LMNN methods. Misclassification rate, precision, f۱ score, and kappa score are calculated for different values of k, mutation rate, and crossover rate. Overall, our proposed method shows superior performance with an average accuracy rate of ۸۷.۸۱% which is the highest among all methods. The average precision, f۱ score, and kappa score of our method are ۰.۸۴۵۳, ۰.۸۵۱۳, and ۰.۶۹۷۶ respectively

کلمات کلیدی:

Classification, large margin nearest neighbor, Genetic Algorithm, Optimization

لینک ثابت مقاله در پایگاه سیویلیکا: