Classification of Protein By Using Random Forest Classifier

Su Myat Thwin, Mar Mar Min
Computer University (Pathein)
zephyrlay.mama@gmail.com, mptn28@gmail.com

Abstract

Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predtictions. Classification is a form of data analysis that can be used extract models describing important data classes or to predict future data trends. A random forest is an ensemble (ie. a collection) of unpurned decision trees. Random forests (Rfhenceforth) is a popular and very efficient algorithm, based on model aggregation ideas, for both classification and regression problems. A random forest model is typically made up of tens or hundreds of decision trees. The system is to studythe Random Forest Classifier and to classify class label of protein data using Random Forest Classifier. The system focuses on protein or not. So this system is intended to classify protein data. The experimental results show that the proposed method achieves high accuracy for testing data.