Abstract

Recombination has major influence on evolution. Recombination occurs at specific region on chromosomes more frequently than other regions. Chromosomal region where recombination occurs more frequently is hot recombination region, whereas, the region where recombination occurs less frequently is cold recombination region. In this paper, supervised machine learning model based on support vector machine and ensembles of support vector machine have been devised for the efficient and effective classification of hot and cold recombination regions based on the compositional features of nucleotide sequences. Models were validated using tenfold cross validation techniques. These models gave high classification accuracy of 87.0%, 91.58%, and 92.14% using support vector machine and its boosting and bagging ensembles respectively. Moreover, support vector machine ensemble with bagging gave remarkably high area under receiver operating curve of .9580. Furthermore, results indicate that bagging ensembles achieved the best result while used for the performance improvement of support vector machines.

References
On Support Vector Machine Ensembles for Classification of Recombination Breakpoint Regions in Saccharomyces cerevisiae


G. Marais, D. Mouchiroud, and L. Duret, "Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes," Proceedings of the...

Index Terms

Computer Science Applied Sciences

Keywords

Recombination Support Vector Machine Boosting Bagging Classification