

Abstract

Ismail M El-Badawy

On the use of Pseudo-EIIP mapping scheme for identifying exons locations in DNA sequences

Identifying exons locations in DNA sequences is one of the significant applications of signal processing in bioinformatics. Mapping a DNA character string into a numerical sequence is a prerequisite prior to its analysis using signal processing techniques. Recently, the Pseudo-EIIP DNA symbolic-to-numeric mapping scheme shows a promising performance with the filter-based exons prediction method, as compared to the traditional EIIP scheme. This paper investigates the performance of the Pseudo-EIIP mapping scheme with different period-3 exons prediction methods, as compared to a number of existing one-dimensional mapping schemes. We conduct MATLAB simulation on the BG570 genomic dataset for the purpose of evaluating exons prediction performance utilizing the ROC curve, precision-recall curve and F-measure. The results reveal the superiority of the Pseudo-EIIP numerical representation over other traditional one-dimensional representations, when employed with different period-3 exons prediction methods.