Abstract

Punctual identification of protein-coding regions in Deoxyribonucleic Acid (DNA) sequences because of their 3-base periodicity has been a challenging issue in bioinformatics. Many DSP (Digital Signal Processing) techniques have been applied for identification task and concentrated on assigning numerical values to the symbolic DNA sequence and then applying spectral analysis tools such as the short-time discrete Fourier transform (ST-DFT) to locate periodicity components. In this paper, first, the symbolic DNA sequences are converted to digital signal using the Z-curve method, which is a unique 3-D plot to illustrate DNA sequence and presents the biological behavior of DNA sequence. Then a novel fast algorithm is proposed to investigate the location of exons in DNA strand based on the combination of Linear Predictive Coding Model (LPCM) and Goertzel algorithm. The proposed algorithm leads to increase the
speed of process and therefor reduce the computational complexity. Detection of small size exons in DNA sequences, exactly, is another advantage of our algorithm. The proposed algorithm ability in exon prediction is compared with several existing methods at the nucleotide level using: (i) specificity - sensitivity values; (ii) Receiver Operating Curves (ROC); and (iii) area under ROC curve. Simulation results show that our algorithm increases the accuracy of exon detection relative to other methods for exon prediction. In this paper, we have also developed a useful user friendly package to analyze DNA sequences.

References

A Novel Fast Algorithm for Exon Prediction in Eukaryotic Genes using Linear Predictive Coding Model and Goertzel Algorithm based on the Z-Curve

September 2012.


Rabiner L. R., Schafer R. W., Digital Processing of Speech Signals, Prentice-Hall, Inc,
1987.

Index Terms

Computer Science Algorithms

Keywords
DNA sequence Protein coding regions Signal processing Exon Linear predictive coding model Goertzel algorithm
A Novel Fast Algorithm for Exon Prediction in Eukaryotic Genes using Linear Predictive Coding Model and Goertzel Algorithm based on the Z-Curve