Novel Adaptive Exon Predictor for DNA Analysis Using Singular Value Decomposition

Srinivasareddy Putluri, Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India
Shaik Yasmin Fathima, Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

ABSTRACT

This article describes how a realistic prediction of the exon regions in deoxyribonucleic acid (DNA) is a key task in the field of genomics. Learning of the protein coding regions is a key aspect of disease identification and designing drugs. These sections of DNA are known as exons, that show three base periodicity (TBP) which serves as a base for all exon locating methods. Many techniques have been applied successfully, but development is still needed in this area. We develop a novel adaptive exon predictor (AEP) using singular value decomposition (SVD) which notably reduces computational complexity and provides better performance in terms of accuracy. Finally, the exon locating capability of proposed SVD based AEP is tested using a real DNA sequence with accession AF099922, obtained from the National Center for Biotechnology Information (NCBI) database and compared with the existing LMS methods. It was shown that proposed AEP is more efficient for locating the exon regions in a DNA sequence.

KEYWORDS

Accuracy, Adaptive Exon Predictor, Computational Complexity, Deoxyribonucleic Acid, Genes, Genomic Sequence, Genomics, Protein Coding Regions

1. INTRODUCTION

The extensive area of research in the field of Bioinformatics is locating the exon regions in a genomic sequence. Vital genes form a subset in organisms which are needed for development, survival and fertility (Ning et al., 2014; Min et al., 2014). Hence, the identification of exons has pragmatic importance.

DOI: 10.4018/IJMTIE.2017010103
to spot human diseases (Dickerson, Zhu, Robertson, & Hentges, 2011) and drug target discovery in new pathogens (Inbamalar & Sivakumar, 2013; Cole, 2002). The protein coding regions and non-protein coding regions are present in a genomic sequence. The Subsection of genomics that focuses on locating the protein coding regions in a genomic sequence is known as gene prediction. The study of prime protein region structure helps the secondary and tertiary structure of protein coding regions for detection of all anomalies, cure diseases and design drugs, as soon as the entire structure of protein regions is analyzed. These studies support in knowing the assessment of phylogenetic trees (Maji & Garg, 2013; Hamidreza, Shamsi, Hamed, & Sedaaghi, 2013). Based on the elemental structure of molecules, the living organisms are divided into two types termed as prokaryotes and eukaryotes. The sections which code for proteins are continuous and long in prokaryotes; examples of prokaryotes are bacteria and archaea. The genes are a combination of coding sections divided by long non-protein coding sections in eukaryotes. These sections which code for proteins are also called as exons, whereas the non-protein coding sections are termed as introns. All living organisms other than bacteria and archaea come under this category. The coding sections reside in human eukaryotes are only 3% of the sequence and the remaining 97% are non-coding regions. Hence the identification of protein coding sections is a significant task (Maji & Garg, 2013; Wazim, Yuzhen, & Haixu, 2014). Almost in all DNA sequences, a three base periodicity (TBP) is exhibited by the protein coding regions. This is obvious by a sharp peak at a frequency f=1/3 in the power spectral density (PSD) plot (Ghorbani & Hamed, 2015). Several techniques for predicting exon regions are presented in literature based on various signal processing methods (Gangchen & Yihui, 2014; Yusuke, A., & Shuichi, 2014; VenkataSrikanth & Rahman, 2016). But, the length of the sequence in real-time gene sequence is extremely long and also the location of the exons varies from sequence to sequence. Existing signal processing techniques are not so accurate in the prediction of protein coding regions. Adaptive signal processing techniques are found to be favorable techniques to process very long sequences in several iterations and can change weight coefficients in accordance to the statistical behavior of the input sequence (VenkataSrikanth & Rahman, 2016). In this paper, efficient Adaptive Exon Predictors (AEPs) are developed using adaptive algorithms for locating protein coding sections. Least mean squares (LMS) algorithm is the fundamental adaptive technique. This algorithm is popular because of its simplicity in implementation. But this algorithm suffers problems like gradient noise amplification, weight drift and poor convergence. So, we put forward singular value decomposition (SVD) technique to improve
9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/article/novel-adaptive-exon-predictor-for-dna-analysis-using-singular-value-decomposition/201049?camid=4v1

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Engineering, Natural, and Physical Science, InfoSci-Knowledge Discovery, Information Management, and Storage eJournal Collection, InfoSci-Physical Sciences, Biological Sciences, and Engineering eJournal Collection, InfoSci-Select. Recommend this product to your librarian:

www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

Leadership Practicies Inventory
www.igi-global.com/chapter/leadership-practicies-inventory/20262?camid=4v1a

New Ultrasound-Based Methods for Early Osteoporosis Diagnosis and Fracture Risk Estimation
Theory, Analysis and Measurements of Rectangular Patch Resonator Sensors

A Real Time Attachment Free, Psycho Physiological Stress and Heart Rate Measurement System
[www.igi-global.com/article/real-time-attachment-free-psycho/58067?camid=4v1a](www.igi-global.com/article/real-time-attachment-free-psycho/58067?camid=4v1a)