Chapter 43

Insulin DNA Sequence Classification Using Levy Flight Bat With Back Propagation Algorithm

Siyab Khan
https://orcid.org/0000-0003-4455-9466
The University of Agriculture, Peshawar, Pakistan

Rehan Ullah
https://orcid.org/0000-0003-0738-9269
The University of Agriculture, Peshawar, Pakistan

Abdullah Khan
https://orcid.org/0000-0003-1718-7038
The University of Agriculture, Peshawar, Pakistan

Maria Ali
The University of Agriculture, Peshawar, Pakistan

Rahat Ullah
University of Malakand, Pakistan

ABSTRACT

Various nature-inspired algorithms are used for optimization problems. Recently, one of the nature-inspired algorithms became famous because of its optimality. In order to solve the problem of low accuracy, famous computational methods like machine learning used levy flight Bat algorithm for the problematic classification of an insulin DNA sequence of a healthy human, one variant of the insulin DNA sequence is used. The DNA sequence is collected from NCBI. Preprocessing alignment is performed in order to obtain the finest optimal DNA sequence with a greater number of matches between base pairs of DNA sequences. Further, binaries of the DNA sequence are made for the aim of machine readability. Six hybrid algorithms are used for the classification to check the performance of these proposed hybrid models. The performance of the proposed models is compared with the other algorithms like BatANN, BatBP, BatGDANN, and BatGDBP in term of MSE and accuracy. From the simulations results it is shown that the proposed LFBatANN and LFBatBP algorithms perform better compared to other hybrid models.

DOI: 10.4018/979-8-3693-3026-5.ch043

1. INTRODUCTION

In the field biological sciences the analysis of humans DNA is a key factor and it is essential to know and understand about the DNA, and its functionality because DNA having all the genetic information related to the functioning and reproduction of an organism (Nguyen et al., 2016). It is the genetic material of the cell (Chao, 2006). DNA is fundamentally made up of four types of similar chemicals called Adenine, Guanine. Thiamine, and Cytosine which are repeated millions and billions of times in the genome, called nucleotides or base pairs of the DNA sequence. Adenine makes a bond with Thiamine and Guanine made bond with cytosine (Chao, 2006). In order to understand and decode the biological information a new field came into being called Bioinformatics (Hapudeniya, 2010). Bioinformatics is a newly evolving research area in the 21st century, which combines numerous fields like biology, Mathematics, computer science and statistics etc. Problem in the field of Bioinformatics is hard because the ratio of data in Bioinformatics is growing exponentially (Hapudeniya, 2010). For extracting of knowledge of the huge amount of biological data to various advanced computer technologies, algorithms are needed to be used (Hapudeniya, 2010). In this regard various statistical and computational methods are attempted, Data mining methods like rule learning (RL), Naïve Bayes (NB), nonlinear integral classifier (NIC) are used for DNA sequence classification (Nurul Amerah Kassim1, 2017). The decision tree is used for the classification of DNA sequence (Tansim, 2018a). The traditional statistical and data mining techniques for the classification of DNA sequence classification having limitations with respect to accuracy. In order to solve the problem of low accuracy, advanced computational methods like machine learning, and hybrid methods with neural network are used for DNA sequence classification (Nurul Amerah Kassim1, 2017). An artificial neural network is a computational model (Wu-Catherine, McLarty, & biochemistry, 2000). The concept of neural networks is primarily taken from the biological neural system. Artificial neural network mimics the human brain, which is made up of small units called neurons. Each neuron has its cell body few short dendrites and single elongated axon (Hapudeniya, 2010). Numerous researchers work in deep neural network for DNA and proteins problems (Eickholt & Cheng, 2013). Various nature inspired metaheuristic optimization techniques are also used in the field of Bioinformatics to solve problems like cuckoo search methods are used for multiple DNA sequence alignment, which is one of the core issues in the field of Bioinformatics (Kartous, Layeb, & Chikhi, 2014). Therefore, this research proposed a new hybrid metaheuristic method levy flight Bat algorithm for the classification of insulin DNA sequences of a healthy homosephian (Human). In the proposed model Bat algorithm are hybrid with Levy flight and artificial neural network and Back propagation neural network in order to improve the accuracy and explains the role of optimization techniques with neural networks.

The remaining section of the paper is organized as follows. Section 2 will discuss the background. While section 3 will explain the methods and material, section 4 will explain proposed algorithm. And similarly, the next section 5 will elaborate the result and discussion. Finlay section 6 will conclude the results respectively.

2. BACKGROUND

There are numerous categories which are used for the problem of classifications in which of them the first one model based, similar to the Markov model (HMM), sequence-sequence classification and further statistical models like linear regression, logistic regression etc. Used for the classification of the biological

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/insulin-dna-sequence-classification-using/342561

Related Content

Domain-Based Approaches to Prediction and Analysis of Protein-Protein Interactions
Morihiro Hayashidaand Tatsuya Akutsu (2014). *International Journal of Knowledge Discovery in Bioinformatics (pp. 24-41).*
www.irma-international.org/article/domain-based-approaches-to-prediction-and-analysis-of-protein-protein-interactions/105099

Sentiment Based Information Diffusion in Online Social Networks
Mohammad Ahsan, Madhu Kumari, Tajinder Singhand Triveni Lal Pal (2018). *International Journal of Knowledge Discovery in Bioinformatics (pp. 60-74).*
www.irma-international.org/article/sentiment-based-information-diffusion-in-online-social-networks/202364

An Approach for Biological Data Integration and Knowledge Retrieval Based on Ontology, Semantic Web Services Composition, and AI Planning
Muhammad Akmal Remliand Safaai Deris (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications (pp. 1727-1744).*
www.irma-international.org/chapter/approach-biological-data-integration-knowledge/76144

Efficient and Robust Analysis of Large Phylogenetic Datasets
Sven Rahmann, Tobias Muller, Thomas Dandekarand Matthias Wolf (2006). *Advanced Data Mining Technologies in Bioinformatics (pp. 104-117).*
www.irma-international.org/chapter/efficient-robust-analysis-large-phylogenetic/4248

Finding Minimum Reaction Cuts of Metabolic Networks Under a Boolean Model Using Integer Programming and Feedback Vertex Sets
Takeyuki Tamura, Kazuhiro Takemotoand Tatsuya Akutsu (2010). *International Journal of Knowledge Discovery in Bioinformatics (pp. 14-31).*
www.irma-international.org/article/finding-minimum-reaction-cuts-metabolic/40969