Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 11, November 2013, pg.129 – 134

SURVEY ARTICLE

Brief Survey on DNA Sequence Mining

NN Das[1], Poonam[2]

[1] *Computer Science and Engineering, Itm University, Gurgaon, India*
[2] *Computer Science and Engineering, Itm University, Gurgaon, India*

[1] nndas@itmindia.edu; [2] poonamr906@gmail.com

Abstract - Sequence Mining is one of the most commonly used technique in data mining. Sequence mining is the process of mining frequent patterns from a large datasets. The exiting algorithms have some limitations in predicting frequent patterns, in terms of time, space complexity and accuracy. To overcome these drawbacks, this paper made a study on existing sequence mining algorithms and generate a new algorithm for generating frequent patterns from the biological sequences(DNA)..This paper attempt to locate all the tandem repeats in a DNA sequence. A repeated substring is called a tandem repeat if each occurrence of the substring is directly adjacent to each other. The future scope of this paper is not only predicting the frequent patterns; but will also satisfy some factors such as: space complexity, time and predict accurate solution to the required problem. With the help of these three things into consideration an effective algorithm can be defined for predicting the tandem repeat in a given DNA sequence.

Keywords –Frequent patterns; DNA; Tandem Repeat; Motifs; KDD

Full Text: http://www.ijcsmc.com/docs/papers/November2013/V2I11201348.pdf