Sequential Patterns Mining For DNA Sequences Based On Divided and Conquers Approach

Aung Thet Htein

University of Computer Studies,Yangon

aungthethtein@gmail.com

Abstract

Data mining is process of pattern extraction from a large collection of datasets.Main goal of data mining is to discover the frequent itemsets(patterns).Sequential pattern mining is an important data mining problem that generates a combinatorial explosive number of intermediate subsequences.Sequential pattern mining generates patterns based on item occurrence order.PrefixlSpan is one of the fast sequential pattern mining algorithms based on divide and conquer approach.PrefixSpan algorithm partitons databases based on currently identified frequent patterns and grow to longer ones using projected databases.This paper presented mining DNA sequential patterns based on divide and conquers approach.Divide and conquer strategy process is partitioning method.By using PrefixSpan method,projected databases are processed in parallel,therefore processing time can be reduced and it will support the bioinformatics field.