Abstract

Longest Common Sequence problem is the most fundamental task in Computational Biology. This is not only a classical problem but also a challenging problem in bio sequences application. Many algorithms are being developed and these are discussed in terms of resource utilization efficiency. This paper proposes a model based on SRLCS algorithm [11] to obtain the possible
length of Longest Common Sequence (LCS). The model accounts the length of the sequences under consideration, the identity and similarity between them. The model is obtained by regressing the LCS results on the training data set by SRLCS. The model so obtained is a simple linear expression which gives the predicted length of LCS. The possible Length of LCS between the given sequences is a sufficient heuristic for biologists in decision making. Often such a result is useful while working on homology finding.

Reference

- The Pfam database is a large collection of protein families available at http://pfam.sanger.ac.uk/

Index Terms

Computer Science  
Bioinformatics
Key words
Pair wise
Longest Common Subsequence length

Fast LCS

Parallel Algorithm

SRLCS