Abstract

Longest Common Subsequence (LCS) and Shortest Common Subsequence (SCS) problems are to find subsequences in given sequences in which the subsequence is as long as possible and as short as possible subsequence respectively. These subsequences are not necessarily contiguous or unique. In this paper we have proposed two new approaches to find LCS and SCS, of N sequences parallely, using DNA operations. These approaches can be used to find LCS and SCS, of any window size, from any number of sequences, and from any type of input data. The proposed work can be applied to finding diverging patterns, constraint LCS, redescription mining, sequence alignment, speech recognition, find motifs in genetic data bases, pattern recognition, mine emerging patterns, contrast patterns in both scientific and commercial databases. Implementation results shown the correctness of the algorithms. Finally, the validity of the algorithms are checked and their time complexity is analyzed.


- Nan Li and Tompa. M. Analysis of computational tools for motif discovery. Algorithms
    specification discovery. Int. Conf. on Knowledge Discovery and Data Mining, pages 460-469,
    2007.
  - Annila. H. M., Toivonen. H., and Verkamo. A. I. Discovery of frequent episodes in
  - Martinez. M. An efficient method to find repeats in molecular sequences. Nucleic Acid
  - Martinez. M. A flexible multiple sequence alignment program. Nucleic Acid Research,
  - Suyama. M., Nishioka. T., and Junichi. O. Searching for common sequence patterns
    requirement from sequences. SIGMOD Int. Conf. on Management of Data, pages 623-633,
    2005.
  - Bin Ma. A polynomial time approximation scheme for the closest substring problem.
  - Smith. H. O., Annau. T. M., and Chandrasegaran. S. Finding sequence motifs in
    groups of functionally related proteins. Proceedings of National Academy (USA), 87:826-830,
    1990.
  - Agarwal. R. and Srikant. R. Mining sequential patterns. Int. Conf. on Data
  - Agarwal. R. and Srikant. R. Mining sequential patterns: Generalizations and
    performance improvements. Extending DataBase Technology, pages 3-17, 1996.
  - Staden. R. Computer methods to locate signals in nucleic acid sequences. Nucleic
  - Isisdore Rigoutsos and Aris Floratos. Combinatorial pattern discovery in biological
  - Saurabh Sinha. On counting position weight matrix matches in a sequence, with
  - Yin-Te Tsai. The constrained longest common subsequence problem. Information
  - Qian Wan and Aijun An. Diverging patterns: Discovering signi cant dissimilarities in
  - Guan. X. and Uberbacher. E. C. A fast lookup algorithm for detecting repetitive DNA
  - Yan. X., Han. J., and Afhar. R. Colspan: Mining closed sequential patterns in large

Index Terms
Keywords
DNA operations  Motifs  LCS  SCS  CLCS  Pattern recognition  Diverging pattern
Exceptional mining