2018-08-25-bioinfo

aphrodite

September 4, 2018

Exekutiv Sammanfattning

Innehåll

Powered by OrgMode and LaTeX

1 intro

1.1 classes

- bioinfomatics
- computational biology

1.2 algorithms

- smith-waterman
- needleman-wunsch
- dynamic programming

2 Smith-Waterman

The Smith‑Waterman algorithm performs local sequence alignment; that is, for determining similar regions between two strings of nucleic acid sequences or protein sequences. Instead of looking at the entire sequence, the Smith‑Waterman algorithm compares segments of all possible lengths and optimizes the similarity measure.

The algorithm was first proposed by Temple F. Smith and Michael S. Waterman in 1981. Like the Needleman‑Wunsch algorithm, of which it is a variation, Smith‑Waterman is a dynamic programming algorithm. As such, it has the desirable property that it is guaranteed to find the optimal local alignment with respect to the scoring system being used (which includes the substitution matrix and the gap-scoring scheme). The main difference to the Needleman-Wunsch algorithm is that negative scoring matrix cells are set to zero, which renders the (thus positively scoring) local alignments visible. Traceback procedure starts at the highest scoring matrix cell and proceeds until a cell with score zero is encountered, yielding the highest scoring local alignment. Because of its cubic computational complexity in time and quadratic complexity in space, it often cannot be practically applied to large-scale problems and is replaced in favor of less general but computationally more efficient alternatives such as (Gotoh, 1982), (Altschul and Erickson, 1986), and (Myers and Miller, 1988).

3 Needleman‑Wunsch algorithm

The Needleman-Wunsch algorithm is an algorithm used in bioinformatics to align protein or nucleotide sequences. It was one of the first applications of dynamic programming to compare biological sequences. The algorithm was developed by Saul B. Needleman and Christian D. Wunsch and published in 1970. The algorithm essentially divides a large problem (e.g. the full sequence) into a series of smaller problems and uses the solutions to the smaller problems to reconstruct a solution to the larger problem. It is also sometimes referred to as the optimal matching algorithm and the global alignment technique. The Needleman-Wunsch

algorithm is still widely used for optimal global alignment, particularly when the quality of the global alignment is of the utmost importance.

4   Dynamic programming

Dynamic programming is both a mathematical optimization method and a computer programming method. The method was developed by Richard Bellman in the 1950s and has found applications in numerous fields, from aerospace engineering to economics. In both contexts it refers to simplifying a complicated problem by breaking it down into simpler sub-problems in a recursive manner. While some decision problems cannot be taken apart this way, decisions that span several points in time do often break apart recursively. Likewise, in computer science, if a problem can be solved optimally by breaking it into sub-problems and then recursively finding the optimal solutions to the sub-problems, then it is said to have optimal substructure.

If sub-problems can be nested recursively inside larger problems, so that dynamic programming methods are applicable, then there is a relation between the value of the larger problem and the values of the sub-problems. In the optimization literature this relationship is called the Bellman equation.