Chapter 13
Hybrid Genetics Algorithms for Multiple Sequence Alignment

John Tsiligaridis
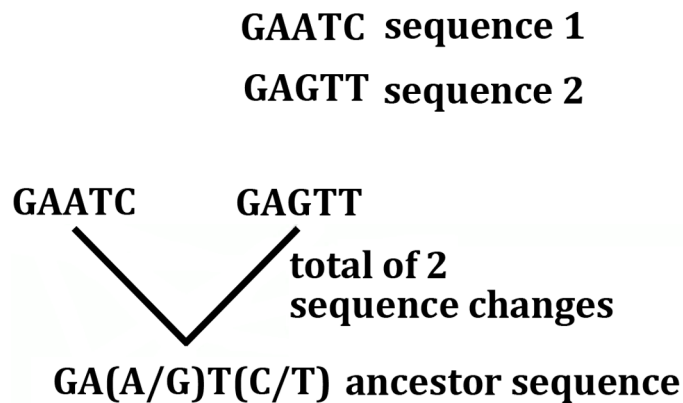*Heritage University, USA*

ABSTRACT

The purpose of this chapter is to present a set of algorithms and their efficiency for the consistency based Multiple Sequence Alignment (MSA) problem. Based on the strength and adaptability of the Genetic Algorithm (GA) two approaches are developed depending on the MSA type. The first approach, for the non related sequences (no consistency), involves a Hybrid Genetic Algorithm (GA_TS) considering also Tabu Search (TS). The Traveling Salesman Problem (TSP) is also applied determining MSA orders. The second approach, for sequences with consistency, deals with a hybrid GA based on the Divide and Conquer principle (DCP) and it can save space. A consistent dot matrices (CDM) algorithm discovers consistency and creates MSA. The proposed GA (GA_TS_VS) also uses TS but it works with partitions. In conclusion, GAs are stochastic approaches that are proved very beneficial for MSA in terms of their performance.

INTRODUCTION

Sequence Alignment is the procedure of comparing two (pair-wise alignment) or more (multiple sequence alignment) DNA or protein sequences by searching for a series of individual characters or character patterns that are in the same order in the sequences. Sequence alignment is useful for discovering functional, structural and evolutionary information in DNA or protein sequences. It is important to obtain the best possible, so-called "optimal", alignment to discover this information (Mount 2004). Sequences that are very much similar in the parlance of sequence analysis probably have the same function. Additionally, if two sequences from different organisms are similar, there may have been a common ancestor sequence, and the sequences are then defined as homologous. An alignment indicates the changes that could have occurred between the two homologous sequences and a common ancestor sequence during the evolution as shown in Figure 1.

Figure 1. The evolutionary relationship between two similar sequences

GAATC sequence 1

GAGTT sequence 2

GAATC   GAGTT

**total of 2
sequence changes**

GA(A/G)T(C/T) ancestor sequence

In Figure 1 there is an origin of similar sequences. Sequences 1 and 2 are each assumed to be derived from a common ancestor sequence. Some of the ancestor sequences can be inferred from conserved positions in the two sequences. For positions that vary there are two possible choices at these sites in the ancestor. In the phylogenetic analysis of three or more similar sequences, the separate distances from the ancestor can be estimated (Mount, 2004).

There are two principal methods of pair-wise alignment: the Dot Matrix analysis and the Dynamic Programming (DP) algorithm. Dot Matrix should be considered as a pair-wise (two sequences) alignment method that displays any possible sequence alignment in diagonal. The DP develops a process that generates a matrix of numbers that represents all possible alignments between two sequences (Jones, Pevzner, 2004). The highest set of sequential scores in the matrix defines an optimal alignment. For proteins, an amino acid substitution matrix, such as the Dayhoff percent accepted mutation matrix 250 (PAM250) is used for score matches and mismatches (Mount, 2004)

Multiple Sequence Alignment (MSA) is among the most important tasks in computational biology.

MSA is a very important extension of paiwise sequence alignment where there is a mutual alignment of three or more sequences. Usually we can find large families of similar sequences by identifying homologues in many different species (Lesk, 2012). In biological sequence comparison, emphasis is given to the simultaneous alignment of several sequences. Genetic Algorithms (GAs) are stochastic approaches for efficient and robust search that can play a significant role for MSA. GA is an evolutionary technique for large space search. It is based on principles inspired by the genetic and evolution mechanisms observed in natural systems (Goldberg,1989; Davis,1991). The near global optimal solution for objective function for Fuzzy Non-Linear Industrial Production Planning Problems is obtained by hybrid meta-heuristics optimization algorithms such as line search, genetic algorithms, and simulated annealing (Vasant, 2012).

Solutions for non linear programming problems that arise in industrial production planning in uncertain environments, using a methodology with GA and Hybrid GA are developed (Vasant, 2011). A very detailed discussion on the Fuzzy Rule-Based System is presented along with a procedure that evolves solutions using genetic operators and provide immediate medical decision (Mankand, 2014). A novel hybrid GA-GSA algorithm to permit the reliability analyst to increase the performance of the system by utilizing the uncertain data is presented. The optimal design parameters corresponding to each component of the system are evaluated by solving a nonlinear optimization problem and comparing their

Related Content

Planning of a Project with Imprecise Activity Time
Sk. Md. Abu Nayeemand Madhumangal Pal (2016). *Handbook of Research on Modern Optimization Algorithms and Applications in Engineering and Economics (pp. 505-515).*
www.irma-international.org/chapter/planning-of-a-project-with-imprecise-activity-time/147527

Sensitivity Analysis of GFRP Composite Drilling Parameters and Genetic Algorithm-Based Optimisation
Kanak Kalita, Ranjan Kumar Ghadaiand Ankur Bansod (2022). *International Journal of Applied Metaheuristic Computing (pp. 1-17).*
www.irma-international.org/article/sensitivity-analysis-of-gfrp-composite-drilling-parameters-and-genetic-algorithm-based-optimisation/290539

Scatter Search Applied to the Vehicle Routing Problem with Simultaneous Delivery and Pickup
Gladys Maquera, Manuel Laguna, Dan Abensur Gandelmanand Annibal Parracho Sant'Anna (2013). *Trends in Developing Metaheuristics, Algorithms, and Optimization Approaches (pp. 149-168).*
www.irma-international.org/chapter/scatter-search-applied-vehicle-routing/69723

Optimal Ordering of Activities of New Product Development Projects with Time and Cost Considerations
Hisham M. Abdelsalamand Amany Magdy (2013). *Meta-Heuristics Optimization Algorithms in Engineering, Business, Economics, and Finance (pp. 546-563).*
www.irma-international.org/chapter/optimal-ordering-activities-new-product/69897

The Position of Rough Set in Soft Set: A Topological Approach
Tutut Herawan (2012). *International Journal of Applied Metaheuristic Computing (pp. 33-48).*
www.irma-international.org/article/position-rough-set-soft-set/70237