Abstract

Identification of Transcription Factor Binding Sites (TFBS) also called as motifs, from the promoter region of genes remains a highly important and unsolved problem of computational biology. Motifs are short, recurring patterns in DNA sequences that are presumed to have a biological function. In this paper, we propose an evolutionary approach to identify transcription factor binding sites. This approach is based on the genetic algorithm with population clustering. A simple genetic algorithm favors selection of fittest, and this selective pressure tends to remove the diversity of population. Sometimes promoter sequences of some genes consists multiple motifs that also need to be identified. The proposed algorithm uses clustering scheme to partition population in clusters and the mating is allowed only within cluster. This scheme enables algorithm to retain diversity of population over the generations, against the selection pressure and to find out multiple motifs in promoter sequences of co-regulated genes. We applied this approach on various data sets and the results show that it can find correct results
A Genetic Algorithm with Clustering for Finding Regulatory Motifs in DNA Sequences

for binding sites.

Reference

Index Terms

Computer Science

Artificial Intelligence

Key words

motif
regulatory binding sites

transcription factor

genetic algorithms

clustering