Sistem de analiza si predictia secventelor genomice bazat pe metode neuro-fuzzy de data-mining

Lucian Iulian Fira (director de grant)

Grant de cercetare cu CNCSIS, tip TD, 2004, cod 149, tema 99.

Coordonator: Prof. Dr. ing. Horia-Nicolai Teodorescu

Rezumat:

Bioinformatica, ramura informaticii cu cea mai rapida dezvoltare in ultimul deceniu, are ca unul din principalele obiective analiza datelor adunate din secventele genomice, prin metode de data mining.

In ultimii ani, au fost publicate baze de date imense cu secvente genomice provenind, in mare parte, din secventierea completa a genomului uman, dar si al virusilor, microbilor, parazitilor.

Castigarea unui avantaj competitiv, in epoca post-genomica, va depinde de realizarea de adnotari rapide si precise ale secventelor colectate.

In cadrul unui plan amplu propus de conducatorul de doctorat, pe baza experientei acumulate si fructificand metodologia noua de abordare a secventierii, propunem o structura hibrida de inteligenta artificiala, capabila sa invete secvente genomice si sa detecteze componente specifice sau tipare cunoscute.

Prin aceasta structura si prin metodologia originala propusa, se urmareste dezvoltarea de metode rapide de identificare a tipului de dinamici asociate cu o secventa genomica si utilizarea acestor metode pentru caracterizarea secventelor si pentru a ajuta la detectarea componentelor lor specifice.

Ne propunem incapsularea structurii dezvoltate intr-un sistem software complex, impreuna cu alte instrumente suplimentare, care sa poata identifica boli.

Pe baza metodologiei anterior elaborate, analiza secventelor genomice se face separat, pe fiecare tip de nucleotida, prin sisteme specifice, reprezentate de predictori neuro-fuzzy. Pe nivel ierarhic superior, un sistem neuronal de luare a deciziilor primeste informatii de la sistemele anterioare.

Instrumentele software realizate vor fi testate pe genomuri pentru recunoasterea speciilor, pentru identificarea genomurilor sau pentru identificarea si localizarea unor secvente specifice in cadrul unui aceluiasi genom.

Se vor compara rezultatele cu cele obtinute prin alte metode si se vor evidentia avantajele sistemului dezvoltat. Pe baza experientei ne asteptam in primul rand la o precizie mai buna.

Abstract:
Bioinformatics is the fastest development part of informatics during the last decade. One of the main applications of bioinformatics is the analysis by data mining of the data gathered in genome projects.
In the last years was published huge databases with genomic sequences deriving from the complete sequencing of the human, virus, microbe and parasite genome.
Gaining a competitive advantage in this post-genome age will depend on the capacity to perform rapid and precise annotation of collected sequences.
In an ample program proposed by the doctoral coordinator, based on the accumulated experience and fructifying the new methodology to approach the sequencing, we propose a hibrid artificial intelligence structure, able to learn genomic sequences and to detect specific components or known patterns.
By means of this structure and by the proposed original methodology, we aim to develop fast methods to identify the type of dynamics associated with a genomic sequence and use those methods to further characterize the sequences and to help detecting their specific components. Also, we aim to encapsulate the developed structure into complex software system, together with other additional tools able to identify diseases.
Based on already developed methodology, genomic sequence analysis is achieved separately on each nucleotide type, with specific systems, like neuro-fuzzy predictors. On the superior hierachical level, a neural decision-making system, receives informations from the preceding systems.
The achieved software tools will be tested on genomes to recognize species, to identify genomes, or to identify and locate specific sequences into the same genome. The results will be compared with results of oher methods and will be dignified the advantages of developed system. Based on experience, we first expected a better precision.