IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Architectural Realm of Bioninformatics

K.Suganya[*1], SM.Sugithraa[2]
[*1,2]Associate Professor , Department of Software Engineering & IT[PG], A.V.C College of Engineering, Mayiladuthurai , India
ksuganya2002@yahoo.com

Abstract
Bioinformatics is the application of computer science and information technology to the field of biology and medicine. Bioinformatics deals with algorithms, databases and information systems, web technologies, artificial intelligence and soft computing, information and computation theory, software engineering, data mining, image processing, modeling and simulation, signal processing, discrete mathematics, control and system theory, circuit theory, and statistics, for generating new knowledge of biology and medicine, and improving & discovering new models of computation (e.g. DNA computing, neural computing, evolutionary computing, immuno-computing, swarm-computing, cellular-computing). Java, XML, Perl, C, C++, Python, R, SQL and MatLab are the programming languages popularly used in this field. A bioinformatician needs to have a basic and general sense of the ideas and approaches of science

Keywords: XML, IR SQL, GUI

Introduction

The term *bioinformatics* was coined by Pauline Hogeweg and Ben Hesper in 1978 for the study of informatic processes in biotic systems.Bioinformatics was applied in the creation and maintenance of a database to store biological information at the beginning of the "genomic revolution", such as nucleotide and amino acid sequences. The primary goal of bioinformatics is to increase the understanding of biological processes. What sets it apart from other approaches, however, is its focus on developing and applying computationally intensive techniques

Approaches for Bioinformatics

Both Static and dynamic approaches are available in the bioinformatics.
- ✓ Static
    - ➢ Sequences – Proteins, Nucleic acids and Peptides
    - ➢ Structures – Proteins, Nucleic acids, Ligands (including metabolites and drugs) and Peptides
    - ➢ Interaction data among the above entities including microarray data and Networks of proteins, metabolites
- ✓ Dynamic
    - ➢ Systems Biology comes under this category including reaction fluxues.
- ➢ Multi-Agent Based modelling approaches capturing cellular events such as signalling, transcription and reaction dynamics

Research areas in Bioinformatics Sequence Analysis

- ➢ The sequence information is analyzed to determine genes that encode polypeptides (proteins), RNA genes, regulatory sequences, structural motifs, and repetitive sequences
- ➢ Today, computer programs such as BLAST are used daily to search sequences from more than 260 000 organisms, containing over 190 billion nucleotides.
- ➢ The Institute for Genomic Research to sequence the first bacterial genome, does not produce entire chromosomes, but instead generates the sequences of many thousands of small DNA fragments (ranging from 35 to 900 nucleotides long, depending on the sequencing technology).
- ➢ Another aspect of bioinformatics in sequence analysis is annotation, which involves computational gene finding to search for protein-coding genes, RNA

genes, and other functional sequences within a genome.

Genome annotation:

Genomics, annotation is the process of marking the genes and other biological features in a DNA sequence. The first genome annotation software system was designed in **1995** by **Dr. Owen White**, who was part of the team at The Institute for Genomic Research that sequenced and analyzed the first genome of a free-living organism to be decoded, **the bacterium Haemophilus influenzae** White built a software system to find the genes (places in the DNA sequence that encode a protein), the transfer RNA, and other features, and to make initial assignments of function to those genes.

Most current genome annotation systems work similarly, but the programs available for analysis of genomic DNA are constantly changing and improving.

Computational evolutionary biology:

Evolutionary biology is the study of the origin and descent of species, as well as their change over time. Informatics has assisted evolutionary biologists in several key ways; it has enabled researchers to:

➢ trace the evolution of a large number of organisms by measuring changes in their DNA, rather than through physical taxonomy or physiological observations alone,

➢ more recently, compare entire genomes, which permits the study of more complex evolutionary events, such as gene duplication, horizontal gene transfer, and the prediction of factors important in bacterial speciation,

➢ build complex computational models of populations to predict the outcome of the system over time

➢ track and share information on an increasingly large number of species and organisms

Literature Analysis

Literature analysis aims to employ computational and statistical linguistics to mine this growing library of text resources. For example:

➢ abbreviation recognition - identify the long-form and abbreviation of biological terms,

➢ named entity recognition - recognizing biological terms such as gene names

➢ protein-protein interaction - identify which proteins interact with which proteins from text

The National Centre for Biotechnology Information

Bioinformatics is the field of science in which biology, computer science, and information technology merge into a single discipline.

There are three important sub-disciplines within Bioinformatics

➢ The development of new algorithms and statistics with which to assess relationships among members of large data sets.

➢ The analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures.

➢ The development and implementation of tools that enable efficient access and management of different types of information.

Bioinformatics Offer

❖ **Genomics**

Estimating the number of genes in an organism basing on the number of nucleotide base pairs was not reliable, due to the presence of high numbers of redundant copies of many genes. Genomics has corrected this situation. Useful genes can be selected from a gene library thus constructed and inserted into other organisms for improvement or harmful genes can be silenced.

❖ **Proteomics**

Proteomics involves the sequencing of amino acids in protein, determining its three-dimensional Structure and relating it to the function of the protein.

❖ **Medical informatics**

The terms **'medical informatics'** and '**health informatics**' have been variously defined, but can be best understood as the understanding, skills and tools that enable the sharing and use of information to deliver healthcare and promote health.

'Health informatics' is now tending to replace the previously commoner term **'medical informatics'**, Reflecting a widespread concern to define an information agenda for health services which recognizes the role of citizens as agents in their own care, as well as the major information-handling roles of the non-medical healthcare professions.

Information Retrieval(IR)Systems

- IR is field concerned with organization and
- retrieval of knowledge-based information

(Hersh, 2003)

- Many resources are available for little or no cost

e.g., **MEDLINE,** full text of publications, practice guidelines,consumer health information resources

- Has become "mainstream" with Web
    - Over 80% of all Web users search for personal health information *(Fox and Rainie, 2000)*
    - Nearly 90% of US physicians use Web *(Taylor and Leitman, 2002)*

Challenges to IR systems

- Timeliness – Need information at point of care " just in time"

(Chueh and Barnett, 1997)

- Access – Affordable and archival
    - (Hersh and Rindfleisch, 2000)
    - Government funds research, scientists perform it, and publishers copyright reports of it Quality – Anyone can be a publisher, but is this good in fields like health care
    - There are serious problems with quality of information, especially on consumer-oriented sites
    - (Silberg et al., 1997 and many, many more)

Tuberculosis

- 3000-5000years BC evidence in Egyptian and southAmerican Mummies.
- 18-19th century major cause of death in Europe.
- Spread during Industrial revolution.
- WHO estimates 8 million new cases annually. 25 millionwould lose their life in coming decades.
- Therapy 6-8 months / combination of four drugs, Rif, Etb, pyrazenamide, INH in addition Quinolones.
- Diagnostics : culture test, immuno, PCR
- Vaccines : BCG or ?
- Pulmonary TB about 80%
- Extra Pulmonary TB about 20%

Bioinformatics Role in TB

- Development of Databases.
- Analysis of the genomic information.
- Identification of new drug target.
- Development of vaccines.
- Development of new diagnostics.

- Typing of *Mycobacterium* by DNA fingerprinting.

Medical Applications in Bioinformatics

❖ **Molecular medicine**

Molecular medicine is a spacious field, in which chemical, physical, biological and medical techniques are used to depict molecular structures and mechanisms, identify essential molecular and genetic errors of disease, and to develop molecular interventions to correct them.

The human genome completion implies that we can explore for the genes directly associated with different diseases and begins to understand the molecular basis of these diseases more clearly. By getting this new knowledge of the molecular mechanisms of disease it will enable better.

❖ **Personalized medicine**

Personalized medicine is a medical model emphasizing the efficient use of information about a patient and by using this information we can select or optimize patient's preventative and therapeutic care. As genetic code varies from person to person and affects the body's response to drugs, there will be increased risk of disease that will also vary person to person

With the development of pharmacogenomics clinical medicine will become more personalized.

Personalized medicine can largely define the scope and nature of human biological variation that allows the targeting of medical treatment which may include drugs or cell therapies.

As a result, potentially lifesaving drugs never make it to the marketplace. To find the best drug that react with particular patient doctors have to use trial and error as those with the same clinical symptoms can show a wide range of responses to the same treatment. In the future, doctors will be capable of analyzing patient's genetic profile and recommend the best available drug therapy and dosage from the beginning. eatments, cures and even preventative tests to be developed.

❖ **Preventative medicine**

Application of bioinformatics in which preventive medicine or preventive care are taken to prevent diseases, rather than curing them or treating their symptoms.

The development of diagnostic tests are needed which provides precise details of the genetic mechanism of disease to measure an

individual's inclination to different diseases which may become a distinct reality.

Pharmaceutical companies are looking forward for the preventive actions such as change of life or having treatment at the initial stages. It will more likely to be successful when treating with the disease.

Conclusion

➢ Today it is possible to perform (using heuristic algorithms) **80% accurate searche**s perhaps

   90-95% accuracy from the leading software Systems**.**

➢ **Sensitive algorithms** which improve the search

   Accuracy, such as hidden Markov models and

➢ **Smith-Waterman algorithm**, are also available but take more time to execute the search.

   Now, to handle these demanding needs, computers are being designed around the biologists.

➢ Leading bioinformatics companies are developing software systems which permit research

   scientists to integrate their diverse data and tools under Common Graphical User Interfaces (GUIs). It also permits scientists to share information and provides powerful solution to archive data. The whole area of biology can immensely benefit from the bioinformatic approach. Bioinformatics tools for efficient research will have significant implications in life sciences and betterment of human lives.

References

[1] "Bioinformatics Journal". http://bioinformatics.oxfordjournals.org/.

[2] Hogeweg, P. (1978). "Simulating the growth of cellular forms". Simulation **31** (3): 90–96. doi:10.1177/003754977803100305.

[3] Hogeweg, P. (2011). Searls, David B.. ed. "The Roots of Bioinformatics in Theoretical Biology". PLoS Computational Biology **7** (3): e1002021. Bibcode 2011PLSCB...7E0020H. doi:10.1371/journal.pcbi.1002021. PMC 3068925. PMID 21483479.

[4] Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M (February 1977). "Nucleotide sequence of bacteriophage phi X174 DNA". Nature **265** (5596): 687–95. doi:10.1038/265687a0.PMID 870828.

[5] Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (January 2008). "GenBank". Nucleic Acids Res. **36** (Database issue): D25–30. doi:10.1093/nar/gkm929. PMC 2238942.PMID 18073190.

[6] Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM (July 1995). "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd". Science **269** (5223): 496–512 doi:10.1126/science.7542800.PMID 7542800.

[7] "Open Bioinformatics Foundation: About us". Official website. Open Bioinformatics Foundation. http://www.open-bio.org/wiki/Main_Page. Retrieved 10 May 2011.

[8] "Open Bioinformatics Foundation: BOSC". Official website. Open Bioinformatics Foundation. http://www.open-bio.org/wiki/BOSC. Retrieved 10 May 2011.

[9] International Biometric Group, Independent Testing of Iris Recognition Technology, May