Gestalt ofBioinformatics

Deepali Sanjay, Bhatnagar, Gurpreet Kaur
Computer Applications Deptt., Computer Applications Deptt., Computer Applications Deptt.
Giani Zail Singh, PTU, Giani Zail Singh, PTU, Giani Zail Singh, PTU
Campus, Bathinda Campus, Bathinda Campus, Bathinda
singladeepali88@gmail.com

Abstract-*Bioinformatics is a new and rapid evolving discipline that is emerging from the fields of experimental molecular biology and biochemistry. It integrates many core subjects likeBiology, computer science, biochemistry, statistics, mathematics and others. Bioinformatics is an application of computer technology that is skyrocketing in the recent years.Bioinformatics develops methods and software tools for understanding the concepts of biology .There is growing interest in the application of artificial intelligence (AI) techniques in bioinformatics. Bioinformatics, the subject of current review, is defined as the application of computational techniques to understand and organise the information associated with biological macromolecules. This paper will give a clear glance overview of bioinformatics, its definition, aims and applications in Artificial Intelligence in computer science.*

Keywords-*Artificial Intelligence (AI), Data mining, molecular biology, DNA sequence, Genome Annotation.*

I. INTRODUCTION

In 1970 PaulienHogeweg and Ben Hesper coined the term bioinformatics which to refers to the study of information processing in biotic systems.
According to the Oxford English Dictionary) (Molecular) Bioinformatics is conceptualizing biology in terms of molecules (in the sense of Physical chemistry) and applying "informatics techniques" (derived from disciplines such as applied math, computer science and statistics) to understand and organize the information associated with these molecules, on a large scale.

The primary goal of bioinformatics is to increase the understanding of biological processes. It is the science of melding molecular biology with computer technology.Bioinformatics is the combination of other fields like biology, computer science, statistics, biochemistry, mathematics and even more.

To study biological data different concepts of computer science, statistics, mathematics and engineering are combined together.

Biological data + Computer Calculations= Bioinformatics.Bioinformatics is new science to manage of biological information. To gather, store, analyse and integrate biological and genetic information, computers are used.
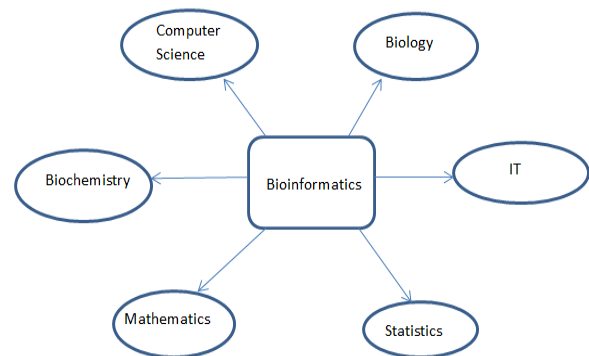

Fig 1. Bioinformatics multidisciplinary fields

Aims of bioinformatics are:
• To organize data that allows researchers to easily create and access information
• To develop tools that facilitates the analysis and management of data.
• To use biological data to analyse and interpret the results in a biologically meaningful manner.

II. APPLICATIONS OF BIOINFORMATICS

An area called computational biology preceded what is now called bioinformatics.Bioinformatics integrates mathematics, statistics and computer technology to solve complex biological problems. These molecular problems

cannot be solved by any other means. Some applications of bioinformatics are:

1. A. *Sequence Analysis*

   Sequence analysis determines the genes which encode regulatory sequences by using information of sequencing.For sequence analysis many powerful tools and computers are used.These computers can see the DNA mutations in an organism and also detects the sequence related to them.

   B. *Genome Annotation*
   Genome annotation is the process of attaching biological information to sequences.
   Genome annotation is very important part of human genome project as it marks the regulatory of sequence.

   C. *Comparative Genomics*
   It is the branch of bioinformatics which determines the genomic structure and function relation between different biological species.

   D. *Health and drug discovery*
   The basics tools of bioinformatics are helpful in drug discovery, diagnosis and disease management which enabled the scientists to make medicines and drugs to target more than 500 genes.

III. RELATION BIOINFORMATICS TO COMPUTER SCIENCE (ARTIFICIAL INTELLIGENCE)

Area of computer science plays a very important role in bioinformatics. With the goal of storing, organizing and analysing biological information Bioinformatics is interfaced with computer science and molecular biology. The combination of Biological data and Computer Calculations is Bioinformatics.

To extract the knowledge encoded in biological data advanced computational technologies, algorithms and tools needed. Basic problems in bioinformatics like protein structure prediction, multiple alignments of sequences,etc. are inherently non-deterministic polynomial-time. To solve these kinds of problems artificial intelligence (AI) approaches are used Researchers have used AI techniques like Artificial Neural Networks (ANN), Fuzzy Logic, Genetic Algorithms, and Support Vector Machines to solve problems in bioinformatics. Artificial Neural Networks is one of the AI techniques commonly in use because of its ability to capture and represent complex input and output relationships among data.

Many the Scalable algorithms are being developed by researchers for bimolecular simulation ,applying data mining, Natural language processing for information retrieval to analyse biological data ,DNA sequences(determine the precise sequence of nucleotides), protein sequences.In computer technology Bioinformatics entails the creation and advancement of databases, algorithms, computational and statistical techniques, and theory to solve formal and practical problems arising from the management and analysis of biological data.

With the rapid development in the field of information technology immense information related to molecular biology and genomic is available.

Developments of new algorithms, statistics measures and implementation of computer programs have helped inproviding information and developing new techniques in bioinformatics.

To produce meaningful information, biological data is analysed which involves writing and running software programs that use different algorithms .

IV. DATA MINING

Extracting or "mining" knowledge from large amounts of data is referred to as Data mining. Data Mining (DM) is the science of finding new interesting patterns and relationship in huge amount of data. Data mining is also called Knowledge Discovery in Databases (KDD).It is defined as "the process of discovering meaningful new correlations, patterns from large amounts of data stored in Warehouses".
In bioinformatics mining biological data helps to extract useful knowledge from massive datasets gathered in biology, and in other areas related life sciences.

V. TASKS OF DATA MINING

The primary goal of data mining is extracting meaningful new patterns from data. The different tasks performed by data mining are:

1. *Classification:* Classification is learning a function that classifies a data item into one of several predefined classes.
2. *Estimation:* is the process of finding an approximate value, which is a value that is usable for some purpose even if input data may be incomplete.
3. *Prediction:* the records are classified according to some future behaviour or estimated future value.
4. *Association rules*: Determining which things go together, also called dependency modelling.
5. *Clustering:* Segmenting a population into a number of subgroups.

6. *Description & visualization:* Representing the data using visualization techniques.

VI.   APPLICATIONS OF DATA MINING IN BIOINFORMATICS

DifferentApplications of data mining in bioinformatics include gene finding, protein function domain detection, protein function inference, disease diagnosis, disease treatment optimization, protein and gene interaction network reconstruction and data cleansing.

VII.   SOFTWARE AND TOOLS

The open source tools provide ideas for commercial applications.Simple software tool used forbioinformatics is command-line. Open-source software packages ranges from Bioconductor, BioPerl, Biopython, BioJava, BioRuby,Bi oclipse, EMBOSS, .NET Bio and UGENE.

Web services in bioinformatics

To run an application on one computer in one part of the world to use algorithms, data and computing resources on servers in other parts of the world, SOAP- and REST-based interfaces are being developed.

VIII.   CONCLUSION AND CHALLENGES

Bioinformatics and data mining are developing as interdisciplinary science. Data mining approaches are well suited for bioinformatics, since bioinformatics is data-rich but at molecular level it lacks a comprehensive theory of life's organization. Data mining in bioinformatics has hampered by many facets of biological databases. The integration of biological databases is main problem in bioinformatics. Data mining and bioinformatics are fast growing research area today's era. It is important to examine what are the important research issues in bioinformatics and to develop new data mining methods and techniques for scalable and effective analysis.

REFERENCES

[1] Jacques *Cohen "Bioinformatics-An introduction for Computer Scientists"*,ACM Computing Surveys,Vol. 36,No. 2,June 2004,pp.122-158.

[2] Janice Glasgow, Igor Jurisica"*AI Magazine, American Association for Artificial Intelligence*" Vol 25, No.1 (2004).

[3] Khalid Raza" Application Of Data Mining In Bioinformatics" Indian Journal Of Computer Science And Engineering, Vol 1 No 2(2010),Pp 114-118.

[4] Li, J.; Wong, L. and Yang, Q."*Data Mining in Bioinformatics*, IEEE Intelligent System, IEEE Computer Society" (2005).

[5] Muditha M. Hapudeniya"artificial Neural networks in Bioinformatics" *Sri Lanka Journal of Bio-Medical Informatics* 2010,Vol. 1(2),pp 104-111.

[6] N.M. Luscombe,D. Greenbaum" *What is bioinformatics? An introduction and overview"* Yearbook of Medical Informatics 2001.

[7] ZoheirEzziane" Applications of artificial intelligence in bioinformatics: A review" Elsevier, Expert Systems with Applications Vol30 (2006) pp-2–10.