ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 12, December 2016

Soft Computing Applications in BioInformatics: A Succinct Study

[1] Satya Narayan Das [2] Sushruta Mishra [3] Brojo Kishore Mishra [4] Bijayalaxmi Panda
[1] Gandhi Institute of Engineering and Technology Gunupur, Odisha,
[2][3][4] C.V. Raman College of Engineering, Bhubaneswar, India.

Abstract :-- **Bioinformatics is recent trend of research in 21st century. Inspite of a large number of methodsimplemented in bioinformatics problems as well as many successful applications, we are in the beginning of a process to massively integrate the aspects and experiences in the different core subjects such as biology, medicine, computer science, engineering, chemistry, physics, and mathematics. Recently the use of soft computing tools for solving bioinformatics problems has started to gain momentum since it can handle imprecision, uncertainty in large and complex search spaces. Our study will focus on integrative research on soft computing paradigm in bioinformatics with particular emphasis on its wide applications.**

Key Words: **-- Bioinformatics, Soft computing paradigm, Artificial neural network, Fuzzy logic, Genetic algorithms, Bioinformatics tools**

I. INTRODUCTION

Recent growth in soft computing methodologies demonstrates the high standards of technology and tools in bioinformatics for dedicated purposes such as reliable and parallel genome sequencing, fast sequence comparison, search in databases, automated gene identification, efficient modeling and storage of heterogeneous data, etc. The basic problems in bioinformatics like protein structure prediction, multiple alignment, phylogenetic inference etc. falls into the category of NP-hard problems. For all these problems, soft computing offers on promising approach to achieve efficient and reliable heuristic solution. On the other side the continuous development of high quality biotechnology, e.g. micro-array techniques and mass spectrometry, which provide complex patterns for the direct characterization of cell processes, offers further promising opportunities for advanced research in bioinformatics. So bioinformatics must cross the border towards a massive integration of the aspects and experience in the different core subjects like computer science and statistics etc. for an integrated understanding of relevant processes in systems biology. This puts new challenges not only on appropriate data storage, visualization, and retrieval of heterogeneous information, but also on soft computing methods and tools used in this context, which must adequately process and integrate heterogeneous information into a global picture.

II. BIOINFORMATICS

Bioinformatics is the application of computer technology to the management of biological information. Computers are used to gather, store, analyze and integrate biological and genetic information which can then be applied to gene-based drug discovery and development. It is arguable that the origin of bioinformatics history can be traced back to Mendel's discovery of genetic inheritance in 1865.However, bioinformatics research in a real sense started in late 1960s, symbolized by Dayhoff's atlas of protein sequences and the early modeling analysis of protein and RNA structures. In fact, these early works represented two distinct provenances of bioinformatics: evolution and biochemistry, which still largely define the current bioinformatics research topics. The need for Bioinformatics capabilities has been precipitated by the explosion of publicly available genomic information resulting from the Human Genome Project. According to analysts, "the bioinformatics "industry", though in a fledgling condition at present, could in the next 20 - 30 years actually rival the drug industry in size.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 12, December 2016

III. TASKS OF BIOINFORMATICS

Different biological problems considered within the scope of bioinformatics involve the study of genes, proteins, nucleic acid structure prediction, and molecular design with docking. A broad classification of the various bioinformatics tasks is given as follows.

1) Alignment and comparison of DNA, RNA, and protein sequences.
2) Gene mapping on chromosomes.
3) Gene finding and promoter identification from DNA sequences.
4) Interpretation of gene expression and micro-array data.
5) Gene regulatory network identification.
6) Construction of phylogenetic trees for studying evolutionary relationship.
7) DNA structure prediction.
8) RNA structure prediction.
9) Protein structure prediction and classification.
10) Molecular design and molecular docking.

IV. APPLICATIONS OF BIOINFORMATICS

Bioinformatics has found its applications in many areas. It helps in providing practical tools to explore proteins and DNA in number of other ways. Bio-computing is useful in recognition techniques to detect similarity between sequences and hence to interrelate structures and functions. Another important application of bioinformatics is the direct prediction of protein 3-Dimensional structure from the linear amino acid sequence. It also simplifies the problem of understanding complex genomes by analyzing simple organisms and then applying the same principles to more complicated ones. This would result in identifying potential drug targets by checking homologies of essential microbial proteins. Bioinformatics is useful in designing drugs.

The aims of Bioinformatics are:

1) To organize data in a way that allows researchers to access existing information and to submit new entries as they are produced
2) To develop tools and resources that aid in the analysis and management of data.
3) To use this data to analyze and interpret the results in a biologically meaningful manner.
4) To help researchers in the pharmaceutical industry in understanding the protein structures to make the drug design easy.

Algorithms in Bioinformatics

The following are some of the most important algorithmic trends in bioinformatics:

1. Finding similarities among strings (such as proteins of different organisms).
2. Detecting certain patterns within strings (such as genes, introns, and α-helices).
3. Finding similarities among parts of spatial structures (such as motifs).
4. Constructing trees (called phylogenetic trees expressing the evolution of organisms whose DNA or proteins are currently known.
5. Classifying new data according to previously clustered sets of annotated data.
6. Reasoning about microarray data and the corresponding behavior of pathways

V. USE OF SOFT COMPUTING IN BIOINFORMATICS

As soft computing [10,11,12] are considered to handle imprecision, uncertainty and near optimality in large and complex search spaces use of soft computing tools for solving bioinformatics problems have been gained the attention of researchers. Our literature survey of recent research papers (2006-2011) shows role of soft computing in modeling various aspects of bioinformatics, it involves genomic sequence, protein structure, gene expression microarray, and gene regulatory networks [11]. Most of the researches are woven around the tasks of pattern recognition and data mining like clustering, classification, feature selection, and rule generation, while classification pertains to supervised or unsupervised learning, clustering corresponds to unsupervised selforganization into homologous partitions. Feature selection techniques [16] aim at reducing the number of irrelevant and redundant variables in the dataset. Rule generation enables efficient representation of mined knowledge in humanunderstandable form. Many intangible parameters are mathematically modeled.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 3, Issue 12, December 2016

Soft Computing Techniques in Bioinformatics

(a) An expert system is designed by collecting knowledge from specific experts. With the help of expert system, a biologist may decision [14].The expert system formulates the decision by rule selection and by deciding factors and by assessing a situation. This problem can be resolved by soft computing techniques. Soft computing mechanism can extract those factors and then fire rules that match the expert's behavior.

(b) Systems often produce results different from the desired ones. This may be caused by unknown properties or functions of inputs during the design of systems [17]. This situation always occurs in the biological world because of the complexities and mysteries of life sciences. However, with its capability of dynamic improvement, soft computing can cope with this problem.

(c) The molecular biology is ever changing, Those new data and concepts update or replace the old ones. Soft computing can be easily adapted to a changing environment. This benefits system designers, as they do not need to redesign systems whenever the environment changes [18].

(d) Missing and noisy data is one characteristic of biological data. The conventional computer techniques fail to handle this. Soft computing based techniques are able to deal with missing and noisy data by deciding hedges in the data [13].

(e) Soft computing is capable to find the unknown relation between huge volumes of ever evolving biological data. It is possible that important hidden relationships and correlations exist in the data. Soft computing methods are designed to handle very large data sets, and can be used to extract such relationships [19].

5.1. Relevance of Artificial Neural Network in Bioinformatics

Neural networks have been widely used in biology since the early 1990s. They can be used to:

(a) Prediction and the translation sites initiation in DNA sequences and proteins [17,18].
(b) Explain the theory of artificial neural networks using applications in biology [2].
(c) Predict immunologically interesting peptides by combining an evolutionary algorithm [20].
(d) Study human TAP transporter [21].
(e) Carry out pattern classification and signal processing successfully in bioinformatics [22].
(f) Perform protein sequence classification [3,23,24].
(g) Predict protein secondary structure prediction [4].

5.2. Relevance of Fuzzy Logic in Bioinformatics

Some of the important uses of fuzzy logic are listed below:

(a) Increasing flexibility of protein motifs [25].
(b) Studying differences between various polynucleotides [18].
(c) Analyzing experimental expression data [3] using fuzzy adaptive resonance theory [25].
(d) Studying aligning sequences based on a fuzzy dynamic programming algorithm [4,1].
(e) Mathematical modeling of complex traits influenced by genes with fuzzy-valued in pedigreed populations.
(f) Finding cluster membership values to genes applying a fuzzy partitioning method using fuzzy C-Means and fuzzy c-hard mean algorithms [25].
(g) Generating DNA sequencing using genetic fuzzy and neuro-fuzzy systems by anticipating disturbances due to intangible parameters [13].
(h) Identifying the cluster genes from micro-array data [5].
(i) Predicting protein's sub-cellular locations fuzzy k-nearest neighbors algorithm.
(j) Mapping specific sequence patterns to putative functional classes since evolutionary comparison leads to functional characterization of hypothetical proteins.
(k) Developing gene expression data [15].

5.3. Relevance of Genetic Algorithms in Bioinformatics

The most suitable applications of GAs in bioinformatics are:

(a) Alignment and comparison of DNA, RNA, and protein sequences [1,18,6].
(b) Gene mappings in chromosomes [9].
(c) RNA structure prediction [20].
(d) Protein structure prediction and clustering [27].
(e) Molecular design and molecular docking [8].

(f) Gene finding and promoter identification from DNA sequences [26].
(g) Interpretation of gene expression and micro array data [9].
(h) Gene regulatory network identification [9].
(i) Construction of phylogenetic tree for studying evolutionary relationship [7].
(j) DNA structure prediction [7].

VI. CONCLUSION

With an explosive growth of the annotated genomic sequences in available form, bioinformatics has emerged as a challenging and fascinating field of science. It presents the perfect harmony of statistics, biology and computational intelligence methods for analyzing and processing biological information in the form of gene, DNA, RNA and proteins. Soft computing algorithms on the other hand, have recently gained wide popularity among the researchers, for their amazing ability in finding near optimal solutions to a number of NP hard, real world search problems. A survey of the bioinformatics literature reveals that the field has a plethora of problems that need fast and robust search mechanisms. Problems belonging to this category include (but are not limited to) the multiple sequence alignment (MSA), protein secondary and tertiary structure prediction, protein ligand docking, promoter identification and the reconstruction of evolutionary trees. Classical deterministic search algorithms and the derivative based optimization techniques are of no use for them as the search space may be enormously large and discontinuous at several points.

REFERENCES

[1] N. Qian and T.J. Sejnowski; "Predicting the secondary structure of globular proteins using neural network models", J. Mol. Biol., Vol. 202(4), pp. 865-884, 1988.

[2] D. Wang and G. B. Huang; "Protein sequence classification using extreme learning machine", Proc. Int. Joint Conf. Neural Networks (IJCNN'05), Montreal, QC, Canada, pp. 1406-1411, August 2005.

[3] H. Saigo, J.P. Vert, N. Ueda, and T. Akutsu; "Protein homology detection using string alignment kernels," Bioinformatics, Vol. 20, pp. 1682-1689, 2004.

[4] M. Xiong, J. Li, and X. Fang; "Identification of genetic networks," Genetics, Vol. 166, pp. 1037-1052, 2004.

[5] C.I. Branden and J. Tooze; "Introduction to Protein Structure", Garland Publishing, New York, 2nd edition, 1999.

[6] Zhong Wei, AltunGulsah, Tian Xinmin, Harrison Robert, Tai Phang and Pan Yi; "Parallel protein secondary structure prediction schemes using Pthread and Open MP over hyperthreading technology", The Journal of Supercomputing, Vol. 41(1), pp. 1-16, 2007.

[7] B. Qian, S. Raman, R. Das, P. Bradley, A.J. McCoy, R.J. Read and D. Baker; "Highresolution structure prediction and the crystallographic phase problem", Nature, Vol. 450, pp. 259-264, 2007.

[8] "Special Issue on Bioinformatics Part I: Advances and Challenges", Proc. IEEE, Vol. 90(11), November 2002.

[9] G. Fogel and D. Corne (eds.); "Evolutionary Computation in Bioinformatics", San Francisco, CA: Morgan Kaufmann, 2002.

[10] Aboul Ella Hassanien, Mariofanna G. Milanova and Tomasz G. Smolinski; "Computational Intelligence in Solving Bioinformatics Problems: Reviews, Perspectives, and Challenges", Comp. Intel. in Biomed. &Bioinform., SCI 151, pp. 3-47, Springerlink 2008.

[11] P. Baldi and S. Brunak; "Bioinformatics: The Machine Learning Approach", Cambridge, MA: MIT Press,1998.

[12] Rabindra Ku. Jena, Musbah M. Aqel, Pankaj Srivastava and Prabhat K. Mahanti; "Soft Computing Methodologies in Bioinformatics", European Journal of Scientific Research, Vol.26(2), pp. 189-203, 2009.

[13] J. Cheng, P. Baldi; "Improved residue contact prediction using support vector machines and a large feature set", BMC Bioinformatics, Vol. 8, pp. 113, 2007.

[14] Juan R. González, David A. Pelta, and José L. Verdegay; "Solving Bioinformatics Problems by Soft Computing Techniques: Protein Structure Comparison as Example", Intel. Sys.and Tech., SCI 217, pp. 123-136, Springer-Verlag Berlin Heidelberg 2009.

[15] A. Dal Palu, A. Dovier, and F. Fogolari; "Constraint logic programming approach to protein structure prediction", BMC Bioinformatics, Vol. 5, pp. 186, 30 November 2004.

[16] J. Setubal and J. Meidanis; "Introduction to Computational Molecular Biology", Boston, MA: Thomson, 1999.

[17] A. Narayanan, E. Keedwell, and B. Olsson; "Artificial Intelligence Techniques for Bioinformatics", Applied Bioinformatics, Vol. 1(4), pp. 191-222, 2003.

[18] Y. Huang and Y. Li; "Prediction of protein subcellular locations using fuzzy k-NN method", Bioinformatics, Vol. 20(1), pp. 21-28, 2004.

[19] Swagatam Das, Ajith Abraham, and Amit Konar; "Swarm Intelligence Algorithms in Bioinformatics", (SCI) 94, pp. 113-147, Springerlink., 2008.

[20] Zou Xiu-fen, Pan Zi-shu, Kang Le-shan and Zhang Chu-yu; "Evolutionary computation techniques for Protein structure prediction: A Survey", Wuhan University Journal of Natural Sciences, Vol. 8(1B), 2003.

[21] E.E. Snyder and G.D. Stormo; "Identification of protein coding regions in genomic DNA", J. Mol. Biol., Vol. 248, pp. 1-18, 1995.

[22] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman; "Gapped BLAST and PSI-BLAST: A new generation of protein database search programs", Nucleic. Acids Res., Vol. 25, pp. 3389-3402, 1997.

[23] J. Moult, K. Fidelis, A. Kryshtafovych, B. Rost, T. Hubbard and A. Tramontano; "Critical assessment of methods of protein structure prediction - Round VII", Proteins, Vol. 69(8), pp. 3-9, 2007.

[24] The UniProt Consortium; "The Universal Protein Resource (UniProt)", Nucleic. Acids Res., Vol. 35(D), pp. 193-197, 2007.

[25] L. A. Zadeh; "Fuzzy logic, neural networks, and soft computing", Commun. ACM, Vol. 37, pp. 77-84, 1994.

[26] D.E. Goldberg; "Genetic Algorithms in Search, Optimization and Machine Learning", Reading, MA: Addison-Wesley, 1989.

[27] A.P. Engelbrecht; "Fundamentals of Computational Swarm Intelligence", Wiley, 2005.