

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; 8(1): 01-03
© 2023 Stats & Maths
<https://www.mathsjournal.com>
Received: 07-10-2022
Accepted: 05-12-2022

Arun Kumar Gali
Associate professor,
Department of Mathematics,
Government First Grade College
Sector no-43, Navanagar,
Bagalkote, Karnataka State,
India

Corresponding Author:
Arun Kumar Gali
Associate professor,
Department of Mathematics,
Government First Grade College
Sector no-43, Navanagar,
Bagalkote, Karnataka State,
India

Applications of Fuzzy in the field of Bio informatics

Arun Kumar Gali

DOI: <https://doi.org/10.22271/math.2023.v8.i1a.913>

Abstract

In this paper, Bioinformatics is deeply rooted in three traditional disciplines, i.e., biology, computer science, and statistics. Both biology and computer science often claim bioinformatics as a sub-discipline. Furthermore, bioinformatics has strong ties to physics, biophysics, mathematics, chemistry, and engineering. On the other hand, bioinformatics is becoming an independent discipline by itself, with its own theoretical foundations, analytical approaches, and computational techniques. This emergence is similar to biophysics, which evolved from an interdisciplinary field between biology and physics to an integral science.

Keywords: Fuzzy logic, fuzzy bioinformatics

1. Introduction

Fuzzy logic means approximate reasoning, information granulation, computing with words and so on. Fuzzy logic provides an inference structure that enables human reasoning capabilities to be applied to artificial knowledge-based systems. It provides a means for converting linguistic strategy into actions and thus offers a high-level computation.

Fuzzy logic provides mathematical strength to the emulation of certain perceptual and linguistic attributes associated with human cognition, whereas the science of neural networks provides a new computing tool with learning and adaptation capabilities.

The theory of fuzzy logic provides an inference mechanism under cognitive uncertainty; computational neural networks offer exciting advantages such as learning, adaptation, fault tolerance, parallelism, and generalization. Fuzzy logic has been applied to many different and diverse applications, including categorization of weather patterns and of seagull behaviors, control of cement kilns, passenger trains, and elevators, scheduling of subway trains and service technicians, and as a prediction mechanism in risk management. Empirically, five general categories have emerged within which a fuzzy logic based system is beneficial, and often even necessary:

Complex systems, where an adequate system model is difficult or impossible to define.

- Human expert controlled systems.
- Systems with moderately to very complex continuous (or semi-continuous) inputs and outputs, for example PID based control systems.
- Systems with human observations as control rules and/or inputs.
- Systems where vagueness is common, for example in economic systems, natural sciences, and behavioral sciences.

As a mathematical foundation, a generalization on the age-old but quite limiting concept of absolute truth, fuzzy logic can be successfully applied across a broad range of disciplines, and has the potential of having as significant an impact on the types of systems developed. Applications of fuzzy logic may also be found, among others, in the areas of Environmental Protection, Economy, Picture Processing, Power Systems, Social Sciences, Music, Hardware and Telecommunications. Special mentions deserve the many applications of fuzzy logic in Medicine like example in support of diagnosis, in Medical Image Processing, in medical data mining, and in medical modeling. This list is by all means not exhaustive; it only pretended to show the wide spectrum of applications that have been developed in the last decade.

2. Fuzzy Logic in Bio-informatics

As we enter the information age, we witness the impact of computers and computation in almost every corner of our lives. Many people in the world retrieve and broadcast information through the Internet. The weather forecast is made through extensive computation on supercomputers. Stocks are traded electronically. Airplanes are designed completely on computers before the first component is ever manufactured. We also witness substantial impact of computers and computation on biological and medical research, and this impact led to the birth of bioinformatics.

Although bioinformatics is a popular term in science and technology, there is no consensus for its definition. As a new field, its precise definition will take many years to finalize. A current semi-official definition for bioinformatics by the US National Institutes of Health (NIH) is "Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, represent, describe, store, analyze, or visualize such data". A related field, computational biology, is defined by NIH as "the development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems".

From these definitions, bioinformatics is focused on technology (Engineering) for developing tools and infrastructure, while computational biology is more about science (Biology) to generate hypotheses in understanding nature. Although the distinction between bioinformatics and computational biology is made by NIH and others, there is no doubt that the two fields are tightly coupled. Hence, the terms bioinformatics and computational biology are sometimes used interchangeably. For example, the definition of bioinformatics by Luscombe *et al.*, [2001] ^[1] includes some scope of computational biology specified by NIH, but restricts itself to the bimolecular aspect: "bioinformatics is conceptualizing biology in terms of macromolecules (In the sense of physical-chemistry) and then 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." Bioinformatics is deeply rooted in three traditional disciplines, i.e., biology, computer science, and statistics. Both biology and computer science often claim bioinformatics as a sub-discipline. Furthermore, bioinformatics has strong ties to physics, biophysics, mathematics, chemistry, and engineering. On the other hand, bioinformatics is becoming an independent discipline by itself, with its own theoretical foundations, analytical approaches, and computational techniques. This emergence is similar to biophysics, which evolved from an interdisciplinary field between biology and physics to an integral science.

Fuzzy set have partial memberships in multiple sets, which is used in processing of information from the field of molecular biology (For analyzing various properties of the transcript me and proteome of several organism), clinical practice guidelines, automobile and other vehicle subsystems (Automatic transmission), pattern recognition, image processing, remote sensing, language filters etc.

Fuzzy inference is an effective tool for the expression of guideline recommendations, and that it can be useful for the management of imprecision and uncertainty, fuzzy logic (FL) as an approach of logic-based modeling with the easy interpretability of Boolean models but significant advantages

including the ability to encode intermediate values for inputs and outputs. Fuzzy inference technologies are repeatedly applied in bioinformatics.

For example, increase the suppleness of protein motifs and learn about the distinction among polynucleotide, utilizing the fuzzy adaptive resonance theory for the analysis of experimental expression data, applying the dynamic programming algorithm for the alignment of the sequences based on fuzzy recast, fuzzy k-nearest neighbors algorithm used to identify the proteins sub-cellular locations from their dipeptide composition.

Applying fuzzy c-means and partitioning method for characteristic cluster relationship values of genes, analysis of gene appearance data, functional and ancestral relationships between amino acids with the help of fuzzy alignment method, fuzzy classification rules generated by neural network architecture for the analysis of affairs between genes and decipher of a genetic set-up to process micro-array images, use of fuzzy vector filtering framework in the classification of amino acid sequences into different superfamilies etc.

Needleman and Wunsch [1970] ^[15] have proposed a dynamic programming algorithm for performing a global alignment of two sequences. Smith and Waterman [1981] ^[2] have proposed an algorithm to find a pair of segments one from each of two long sequences such that there is no other pair of segments with greater similarity (homology).

In this local alignment algorithm, similarity measures allowed arbitrary length deletions and insertions. Cai, Juedes, and Liakhovitch [2000] ^[3] have proposed to combine existing efficient algorithms for near optimal global and local multiple sequence alignment with evolutionary computation techniques to search for better near optimal sequence alignments.

Tomida *et al.*, [2002] ^[16] have analyzed the experimental expression data using fuzzy adaptive resonance theory. Schlosshauer and Ohlsson [2002] ^[17] have proposed a novel approach to align sequences based on a fuzzy recast of a dynamic programming algorithm.

Fuchs [2002] ^[4] has recommended that bioinformatics is intensifying for research and development of new technology. Chang and Halgmuge [2002] ^[5] have proposed a method to increase the flexibility of protein motifs. Bourbakis [2003] ^[6] has analyzed that bioinformatics combines the multi-disciplinary area such as computer science, biology, physical and chemical principles, designing of tools utilized for the analysis and modeling of large biological data sets, chronic diseases management, learning of molecular computing and cloning etc.

Torres and Nieto [2003] ^[7] have studied the differences between polynucleotides. Protein localization data are a valuable information resource helpful in elucidating protein functions. It is highly desirable to predict a protein's sub cellular locations automatically from its sequence. Ying Huang and Yanda Li [2004] ^[9] have introduced the fuzzy k-nearest neighbors (*k*-NN) algorithm to predict proteins' sub cellular locations from their dipeptide composition. The prediction was performed with a new data set derived from version 41.0 SWISS-PROT databank, the overall predictive accuracy about 80% has been achieved in a jackknife test and the result demonstrated the applicability of this relative simple method and possible improvement of prediction accuracy for the protein sub cellular locations.

They also applied this method to annotate six entirely sequenced proteomes, namely *Saccharomyces cerevisiae*,

Caenorhabditis elegans, *Drosophila melanogaster*, *Oryza sativa*, *Arabidopsis thaliana* and a subset of all human proteins. Cordon *et al.*, [2004]^[18] have studied and analyzed DNA sequencing using genetic fuzzy systems. A new algorithm for local alignment of DNA sequences had been proposed by Paul and Konar [2005]^[10] have proposed direct comparison methods to obtain global and local alignment between the two sequences. They also proposed an alternate scoring scheme based on fuzzy concepts. Chen *et al.*, [2006]^[11] have proposed a partitioning approach, based on ant-colony optimization algorithm that significantly improved the solution time and quality by utilizing the locality structure of the problem.

Chang *et al.*, [2006]^[11] have established fuzzy PAM matrix using fuzzy logic and then estimated score for fitness function of genetic algorithm using fuzzy arithmetic. Their experimental results evidenced fuzzy logic useful in dealing with the uncertainties problem and applied to protein sequence alignment successfully.

Yue and Tang [2007]^[12] have applied the divide-and-conquer strategy to align three sequences so as to reduce the memory usage from $O(n^3)$ to $O(n^2)$. They used dynamic programming so as to guarantee optimal alignment. Nasser *et al.*, [2007]^[20] have provided a hybrid approach of dynamic programming and fuzzy logic to align multiple sequences progressively. They computed the optimal alignment of subsequences based on several factors such as quality of bases, length of overlap, gap penalty. Naznin *et al.*, [2009]^[19] have designed an iterative progressive alignment method for multiple sequence alignment by using new techniques for both generating guide trees for randomly selected sequences as well as for rearranging the sequences in the guide trees. In this work, the main objective of the researchers had been to apply different techniques in order to provide efficient alignment algorithms in terms of time and memory requirements.

3. Conclusion

In this paper, Bioinformatics is deeply rooted in three traditional disciplines, i.e., biology, computer science, and statistics. Both biology and computer science often claim bioinformatics as a sub-discipline. Furthermore, bioinformatics has strong ties to physics, biophysics, mathematics, chemistry, and engineering. On the other hand, bioinformatics is becoming an independent discipline by itself, with its own theoretical foundations, analytical approaches, and computational techniques. This emergence is similar to biophysics, which evolved from an interdisciplinary field between biology and physics to an integral science.

4. References

1. Luscombe NM, Greenbaum D, Gerstein M. What is Bioinformatics? A proposed definition and overview of the field. Year book of Medical Informatics. 2001;40(04):346-58.
2. Smith TF, Waterman MS. Identification of Common Molecular Subsequence. Journal of Molecular Biology. 1981;147(1):195-197.
3. Cai L, Juedes D, Liakhovitch E. Evolutionary Computation Techniques for Multiple Sequence Alignment. Proceedings of the 2000 Congress on Evolutionary Computation. 2000;2:829-835.
4. Fuchs R. From sequence to biology: The impact on bioinformatics. Bioinformatics. 2002;18(4):505-506.

5. Chang BC, Halgmuge SK. Protein motif extraction with neuro-fuzzy optimization. Bioinformatics. 2002;18(8):1084-1094.
6. Bourbakis NG. Bio-imaging and bioinformatics IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics. 2003;33(5):726-727.
7. Torres A, Nieto JJ. The fuzzy polynucleotide space: Basic properties. Bioinformatics. 2003;19(5):587-592.
8. Vasantha Kandasamy WB, Devadaoss AV. Some New Fuzzy Techniques. Journal of Mathematical and Computer Science. 2004;17(2):157-160.
9. Huang Y, Li Y. Prediction of protein subcellular locations using fuzzy K-NN method Bioinformatics. 2004;20(1):21-28.
10. Bandyopadhyay SS, Paul S, Konar A. Improved Algorithms for DNA Sequence Alignment and Revision of Scoring Matrix. Proceedings of International Conference on Intelligent Sensing and Information Processing; c2005, p. 485-490.
11. Bai SM, Chen SM. A new method for students learning achievement using fuzzy membership functions. Proceedings of the 11th conference on artificial intelligence, Kaohsiung, Taiwan, Republic of China; c2006.
12. He Yue, Guo Yue, Guo Yi. Application study in decision support with fuzzy cognitive map. International journal of Computer. 2007;12(12):324-331
13. Hata Y, Syojikobashi, Nakajima H. Human health care system of systems. IEEE System Journal. 2009;3(2):231-238.
14. Markowski A, Sam Mannan M, Bigoszewska A. Fuzzy logic for process safety analysis. Journal of Loss Prevention in the Process Industries. 2009;22(6):695-702.
15. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology. 1970;48(3):443-453.
16. Ogiso Y, Tomida A, Tsuruo T. Nuclear localization of proteasomes participates in stress-inducible resistance of solid tumor cells to topoisomerase II-directed drugs. Cancer Research. 2002;62(17):5008-5012.
17. Schlosshauer M, Ohlsson M. A novel approach to local reliability of sequence alignments. Bioinformatics. 2002;18(6):847-54.
18. Wendel HG, Stanchina ED, Fridman JS, Malina A, Ray S, Kogan S, Cordon-Cardo C, *et al.* Survival signalling by Akt and eIF4E in oncogenesis and cancer therapy. Nature. 2004;428(6980):332-337.
19. Naznin F, Sarker R, Essam D. Iterative progressive alignment method (IPAM) for multiple sequence alignment. In 2009 International Conference on Computers & Industrial Engineering; c2009. p. 536-541.
20. Mohd Yasin N, Nasser Noor M, Mohamad O. Does image of country-of-origin matter to brand equity?. Journal of Product & Brand Management. 2007;16(1):38-48.