



Application of Advanced Artificial Intelligence Techniques in Bioinformatics

Huixing Li^{1,2}(✉), Yan Xue^{1,2}, and Xiancai Zeng^{1,2}

¹ Henan Key Laboratory of Industrial Microbial Resources and Fermentation Technology, Nanyang Institute of Technology, Nanyang 473004, Henan, China

lihuixing@nyist.edu.cn

² School of Biological and Chemical Engineering, Nanyang Institute of Technology, Nanyang 473004, Henan, China

Abstract. The flexible application of computer technology in the development process of biology can effectively promote the innovative development of the discipline, improve the shortcomings of the discipline itself. It also promotes people to strengthen the cognition of bioinformatics through computing technology, and then give full play to the role of the technology itself to meet the needs of the current era. This paper is based on experience to conduct a detailed analysis and review on the application of computer technology in bioinformatics for the reference of relevant staff.

Keywords: Artificial Intelligence · Bioinformatics · application

1 Introduction

The concept of bioinformatics was introduced in the 1990s and is an emerging discipline, the essence of which is to use current information science, mathematical and theoretical knowledge, theoretical perspectives and other contents to study the phenomenon of life. To clarify the essence of the discipline of biology, this subjective is applicable for current practical needs. In the study of biological genetic material DNA, the actual sequence, structure and other related contents are organized by the flexible use of computers to fundamentally improve the efficiency of research.

Bioinformatics is affected by the gradual increase of its own data and information in the research process, and traditional research is gradually difficult to record and observe disciplinary information, which requires flexible application of computer technology to promote reasonable and efficient disciplinary research and meet the current practical needs [1]. For example, the application of computer technology has prompted the current research process related to retrieval, storage, query, observation and processing of biological disciplines to be efficient, helping researchers to analyze data reasonably, find out the laws hidden in bioinformatics from a large amount of data, and carry out reasonable simulations through computers, establish models to observe research, strengthen the analysis of data, and drive the discipline to develop positively. At this stage, the application of information technology in bioinformatics is more common, for example,

artificial intelligence technology, database, network technology, parallel computing technology and bio-computing and other related, etc., which prompt bioinformatics at this stage to break through the existing research bottlenecks, such as, conducting effective pattern recognition, prompting research growth points, prompting the gradual decomposition and simplification of relevant domain knowledge, and providing a good The foundation for human development. Based on the actual bioinformatics DNA research, we can flexibly apply the current computer image processing technology to complete the sequencing of genes by establishing a perfect three-dimensional model to clarify the arrangement of nucleotides and promote the three-dimensional structure, providing researchers with good research methods, directions and ideas to promote the development of the discipline. In the developed United States, Japan, Europe and other places, the use of computer technology for bioinformatics research is more advanced and its technology has gradually matured [2].

2 The Need for Computer Technology in Biology Applications

Currently, the development of information technology and computer technology innovation gradually involves all aspects of people's lives, especially the application in bioinformatics, which promotes the comprehensive development of the discipline. The development of bioinformatics is inseparable from the support of computer technology, and only on this basis can we lay a good foundation for the research of bioinformatics to ensure the stable development of its discipline. From a broad perspective, network technology and computer technology are the prerequisites for the stable development of bioinformatics and the basis for its compliance with modern requirements. For example, in the twentieth century, the United States put forward the high-performance Computer and Communication Program, which mainly refers to the use of the advantages of computers themselves to ensure the high performance of bioinformatics research work, especially compared with bioinformatics. The nature of the subject itself is more special, the number of cells or other contents studied is large, and its own small size and complex laws of change are difficult to be realized by human brain alone for memory and research, and the flexible application of computer technology can fundamentally accelerated research and satisfy the development (Fig. 1).

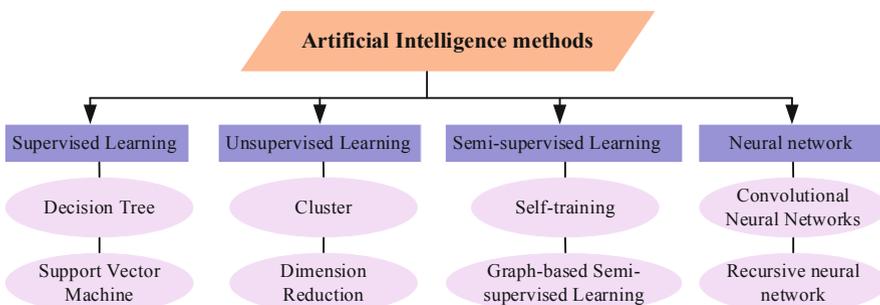


Fig. 1. Artificial Intelligence Techniques

3 Analysis of the Application of Computer Technology in Bioinformatics

3.1 Visualization Technology

Visualization technology has a wide range of applications in bioinformatics, especially for structural research, which can help researchers to deepen their understanding and analysis of the structure and present its contents clearly. For example, in the process of DNA research, DNA's own molecular spatial helix structure is abstract for a single theoretical study, at this time, computer visualization technology can be flexibly applied to display the DNA decomposition mechanism in the form of a view, and its own protein folding structure, helix, spatial shape, etc., to provide researchers with a good basis for research and accelerate research efficiency. The application of this visualization technology is the most basic one, but as people's research on bioinformatics deepens, it is necessary to innovate the existing visualization technology and make it deeper, so as to apply reasonable visualization technology in the study of DNA mutation, molecular hydrophilic study, surface area study, and related curve change study to meet the actual research requirements. For example, compared with bioinformatics, the molecular data itself has more complex relationships and its relationships have obvious correlation, influenced by this feature, in the research process, the current visualization technology is mainly gradually developed in the direction of information correlation, and then meet the current practical needs. At this stage, the more commonly used technologies are information wall display technology, which mainly serves to present biomolecular information itself; semantic mirror technology, which mainly refers to the presentation of biological phenomena generated by the combination of multiple factors based on the current actual database technology to simplify the complex knowledge; gene regulation network technology, which mainly uses the current network technology to effectively organize different data. The gene regulation network technology, mainly using the current stage of network technology to effectively organize different data, find out the combination law in a large amount of data, and then prompt researchers to research to a deeper level and clarify the more complex relationship and meaning of bioinformatics. Self-organizing feature mapping is widely used in genome sequence analysis, and a traditional mapping model is shown in Fig. 2.

3.2 Database Technology

Influenced by the nature of bioinformatics itself, in the process of continuous development, people gradually conduct deeper research on the composition, development, innovation and laws of life phenomenon itself, which leads to an explosive growth trend of research data in fact, for example, in the research process of DNA, its components accounting for a large number of molecules with proteins, the composition of nucleotides alone includes more than three billion compositions, based on this, The application of database technology is inevitable for the development of bioinformatics. With the continuous development of the times, people's research on genetic engineering continues to deepen, prompting the huge amount of bioinformatics information, showing an explosive growth trend. The flexible application of database information technology can organize

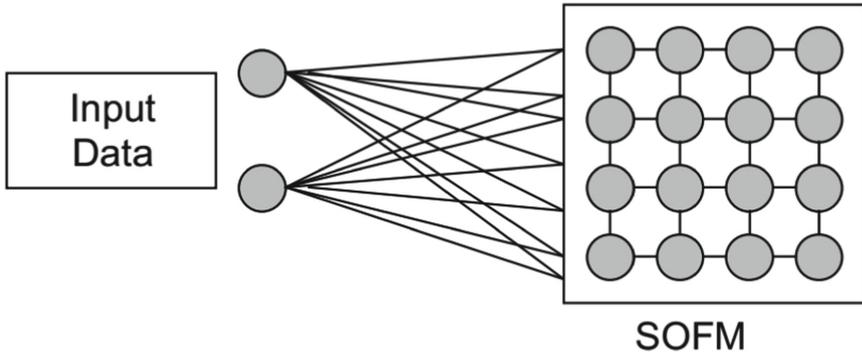


Fig. 2. Self-Organizing Feature Map

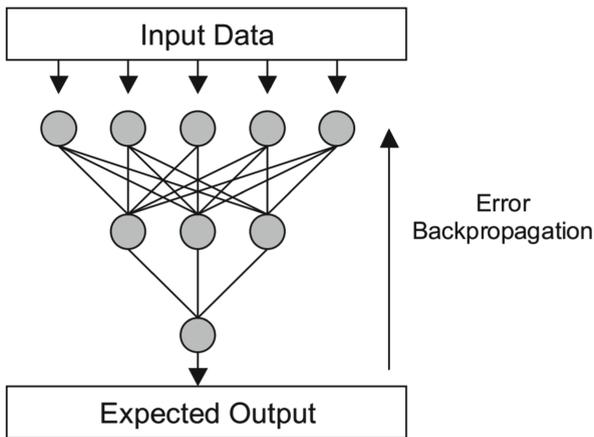


Fig. 3. Supervised learning model

and classify the relevant information reasonably, so that people can use and refer to it effectively. For example, a good database can be established to classify DNA research data into proteins, nucleotides, structures, etc. If the data are more massive, a secondary database can be established to divide them into finer details and ensure their rationality. The international application of database technology in bioinformatics is more advanced, for example, bacterial gene database, protein structure database, human gene database, glycan database, cell database, immunology database, plant gene database, and fungal gene database, covering almost all fields of bioinformatics to meet the current needs of the times. A typical supervised learning model topology is given in Fig. 3.

3.3 Artificial Intelligence Technology

Artificial intelligence technology is a more common technology at this stage, especially in the current era, people are in the era of information explosion, and to ensure reasonable research, people have established a large number of data and information base, how to find the reference value of information in the huge amount of information is the current bioinformatics researchers must pay attention to the problem, and then in the research process to find some kind of law, and rational explanation of life phenomena. The flexible application of current artificial intelligence technology can effectively help researchers to conduct reasonable data analysis and promote the innovative development of bioinformatics disciplines. For example, in the process of continuous development, people will gradually innovate the gene identification algorithm, currently there are mainly three kinds: the first, homology method, the second, based on statistical methods, the third, based on machine learning methods, in the actual application process, the more common application is the third method, mainly because the neural network itself has a strong non-linear ability, can fundamentally the data itself. The main reason is that the neural network itself has strong nonlinear ability, which can fundamentally analyze the data itself and ensure the efficiency and accuracy of analysis. In the research and analysis process, influenced by the actual nature of bioinformatics itself, there are also other related methods, for example, the Hidden Horse Model method [3], which can be reasonably combined with other related methods in the research process, such as, Bayesian methods [4], neural network methods, etc., which are widely used in biological sequence analysis and play a huge role, thus meeting the current practical needs. The application of artificial intelligence technology and neural network technology can ensure the effective development of bioinformatics, break through the existing research obstacles, and meet the needs of the times. A conventional neural network-based array gene expression model is depicted in Fig. 4.

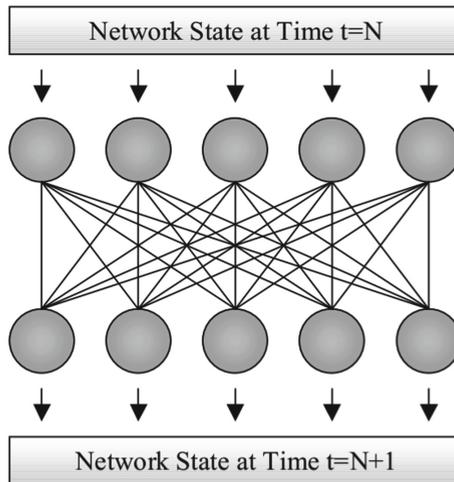


Fig. 4. Neural network application in Gene expression

4 Conclusion

The application of computer technology in biological research is the inevitable development of the times, which can help researchers flexibly use advanced technology to analyze the knowledge content, deeply understand its connotation, and change the traditional research concepts and ideas, so as to promote bioinformatics research gradually from quantitative to qualitative changes, and improve the level of biological science research.

References

1. Gauthier J, Vincent AT, Charette SJ, Derome N (2019) A brief history of bioinformatics. *Brief Bioinform* 20(6):1981–1996
2. Narayanan A, Keedwell EC, Olsson B (2002) Artificial intelligence techniques for bioinformatics. *Appl Bioinform* 1:191–222
3. De Fonzo V, Aluffi-Pentini F, Parisi V (2007) Hidden markov models in bioinformatics. *Curr Bioinform* 2(1):49–61
4. Neapolitan RE (2009) Probabilistic methods for bioinformatics: with an introduction to Bayesian networks. Morgan Kaufmann

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

