

Preface

It is widely recognized that the field of biology is in the midst of a 'data explosion'. A series of technical advances in recent years has increased the amount of data that biologists can record about different aspects of an organism at the genomic, transcriptomic and proteomic levels. This data is, of course, vital to advancing our knowledge. In recent years, the discipline of *bioinformatics* has allowed biologists to make full use of the advances in computer science and computational statistics in analysing this data. However, as the volume of data grows, the techniques used must become more sophisticated to cater for large-scale data and noise. Also, given the growth in biological data, there is a need to extract information that was not previously known from these databases to supplement current knowledge. Large databases may contain interesting patterns that, if identified and authenticated by further laboratory and clinical work, can lead to novel theories about the causes of various diseases and also possibly to new drugs for their treatment. The discipline of bioinformatics has reached the end of its first phase, and the motivation behind this book is to characterize the principles that may underlie second phase bioinformatics. That is, second phase bioinformatics is when the discipline, instead of being informed by just computer science and computational statistics, is also informed by artificial intelligence techniques.

As we show in this book, there are problems in bioinformatics and many other sciences that cannot be solved satisfactorily even with the fastest computers. Clearly, a more 'intelligent' approach is required to solve these increasingly difficult bioinformatics problems, such as gene expression analysis and protein structure prediction. This book attempts to address this by looking at the latest advances in artificial intelligence technology as applied to computational problems in biology. Artificial intelligence methods are often based on the ways in which humans solve

search and optimization problems, or how nature has solved its own problems, for example by using the principles of 'survival of the fittest' in evolutionary computation.

This book is divided into three parts, each containing a number of chapters. These parts are designed to allow readers to access the material most relevant to them. The first part, *Introduction*, introduces the material necessary to understand the technology and biology included in the later chapters. We recognize that bioinformatics is highly cross-disciplinary and therefore some, all or none of these chapters may be relevant to the reader, depending on their background. The next part, *Current Techniques*, describes the established artificial intelligence techniques in bioinformatics including probabilistic, nearest neighbour and genetic algorithm approaches. The final part, *Future Techniques*, is intended to give the reader an impression of the latest thinking in the area of intelligent bioinformatics. Some of these approaches may not have been widely applied to problems in bioinformatics, but algorithms such as genetic programming and various hybrid approaches can be expected to make a big impact in this domain if experience in other areas of science and technology is anything to go by.

In short, this book has been written to engage and interest readers from many disciplines. Biologists are provided for in that there is a full introduction to the challenges for computer science, and computer scientists should also find the chapters on biology and bioinformatics informative. Practicing bioinformaticians are also likely to find the book enlightening, as much of the material has previously only been included in specialist publications and a collection such as this provides a single resource for many intelligent problem-solving techniques in bioinformatics. However, as with any book of this type, not every technique can be included due to space restrictions and apologies are offered to researchers whose own favourite analytical techniques are not covered in this book.

Edward Keedwell

Ajit Narayanan