The Knowledge Engineering Review, Vol. 9:3, 1994, 305–316

Book reviews

Artificial intelligence and molecular biology by Lawrence Hunter (Ed.), AAAI, Menlo Park, CA and MIT Press, Cambridge, MA, 1993, pp 470, ISBN 0-262-58115-9.

Lawrence Hunter has edited 12 chapters (ten from the US and two from the UK) describing significant applications from the growing research community of Artificial Intelligence (AI) in molecular biology. These include knowledge based techniques such as qualitative simulations, as well as machine learning, computational linguistics and three neural network systems applied to a range of significant problems of biological data interpretation. The intended readership of this book appears to be the same community of unusual interdisciplinary experts and widely read computer scientists. It is unfortunately not particularly approachable for lab scientists, and perhaps some effort should address this, since these applications may have the most practical significance for this group.

Hunter identifies and conveys much of the characteristic holism of biology in his enthusiastic introduction of biological life to computer scientists, and describes, not entirely accurately, the potential of recent technological developments in molecular biology for genome analysis. Unfortunately for the biologist, he provides no comparable introduction to artificial intelligence. Although this would be difficult given the diversity of AI, the depth and quality of this set of papers require an introduction to unite them as a whole book, especially if its readership is to be extended to the biology community.

Each of the individual contributions provides an excellent summary of each biological domain. For the biologist with some computing knowledge, introductory AI can be obtained piecemeal from most of the chapters, such as Searls' on linguistics, Steeg's contrasting contribution on neural nets and search algorithms, and many others. Unfortunately, this bias towards an AI readership means that sometimes the authors' enthusiasm for a particular technique may dominate the discussion of its use in a biological context.

Lederberg's prologue, perhaps aptly mistyped "foreward", provides his wordy and not entirely clearly argued vision of AI interpreting molecular biology data. He confidently does not question the contribution that AI can make to molecular biology, and in his afterword (there is as yet no "afterwards"), his only reservations are the fear of computerising the current wisdom of scientists without question, and so preserving human error.

By contrast, real biological questions are addressed in the chapters of this book, including those of protein and RNA structure prediction, DNA sequence analysis, NMR and crystallography, genetic regulation and metabolism. The AI techniques employed for each problem seem appropriate and promising. Various contributions show how biological knowledge in a domain can be represented using different approaches, avoiding excess detail and exposing regularities and constraints. There is a lot of material here which could be used for introductory or survey purposes to widen the readership of this book, and thus the potential interest in these applications. Applications which offer declarative expression of the domain and symbolic reasoning are the best examples to use for consultation with and constructive criticism from biologists. Chapters such as Karp's, on simulation of the tryptophan operon, express expert biological knowledge clearly and demonstrate to the biologist principles of knowledge representation and use in AI. Other cases such as neural networks may be less appreciated by lab scientists, despite their successful analyses of data, because they are seen to lack explicit models of the problem domain. Only through understanding of the possibilities of AI can the lab users come to discern and help develop the applications best suited to their researches. If AI is to collaborate with the molecular biology community there must be user participation and education, which could well begin with a text such as this.

The quality and potential of the chapters may distract from hard questions about their real contribution to biological science or to advances in AI. The work described here is still largely that of AI applications research. The emphasis in most chapters is on tool development for biological research in challenging areas of interpretation and rationalization of observed phenomena. These areas require the intuition and experience of domain experts, but they also offer the opportunity for expert systems and automation. Unfortunately, there is as yet little evidence of successful laboratory use of this.

Hunter concedes that computation in biology has so far proceeded without AI, despite its use elsewhere. He believes that AI must now make its own new discoveries, rather than repeat results from conventional means, if it is to gain widespread and practical acceptance. This book may reinforce the view that AI techniques have contributed nothing new to biology, but demonstrate more potential than progress. This publication also undoubtedly represents the influence of a new opportunity for AI; the Human Genome Project, which may allow AI techniques to prove themselves in an area whose needs cannot otherwise be met. This global research endeavour to determine the molecular meaning of life is generating an abundance of data with a complexity which exceeds the capacity of conventional computational techniques. These data raise many problems of interpretation which cannot be easily addressed, perhaps most noticeably in the prediction of the protein structure from DNA sequence. The need for symbolic manipulation and reasoning argues strongly for the use of artificial intelligence techniques in these areas, as is clearly demonstrated by the contributors to this book. Thus a new opportunity is being provided by molecular biology for AI which only requires that the latter is sufficiently mature to begin a lively and productive relationship.

Reviewed by Catherine Hearne, ACL, ICRF, London, UK

Artificial intelligence and molecular biology by L. Hunter (Ed.), AAAI Menlo Park, CA and MIT Press, Cambridge MA, 1993, pp 470, £35.95, ISBN 0-262-581159.

Spare a thought for those who will soon have to tackle the biggest cryptographic puzzle of all time—deciphering the 3000 million "bits" of data found in the genetic blueprint of a human being. The majority of the complete genetic blueprint of a human—the genome—is stored on the strands of DNA found in the chromosomes of the cells of our body. The massive international program of the Human Genome Project plans to read the billions of "bits" of data recorded on the DNA strands. These bits are in the form of one of four chemicals (abbreviated to A, T, G or C) which are strung together to make a DNA strand. Some regions of DNA strands—called genes—direct the manufacture of proteins.

Proteins are essentially the molecules which give the cell structure and the ability to do the specialized jobs that different cells do. Proteins are made of amino acids which are strung together in an order directed by the gene. By obtaining the sequence of the entire genome, scientists will have access to the genes and proteins that regulate the processes of embryogenesis, death and the bit in between. Given the DNA sequence of a gene, it is easy to predict the order in which the amino acids are built into a protein. However, prediction of the actual 3D structure of this protein is incredibly difficult because the amino acids of the protein interact with themselves and the surrounding water in a complex manner. In water, proteins are not long strings of amino acids, but are folded into intricate structures. Thus, by reading the sequence of a gene, we cannot easily predict the structure—and the function—of a protein.

This book aims to introduce readers to the various approaches used to examine biological systems through computational techniques. It attempts—and succeeds—to draw together a range of reviews covering biology-oriented aspects of information processing and artificial intelligence.