Computational Scientific Discovery

Mark Addis[1]**, Fernand Gobet**[2]**, Peter Lane**[3] **and Peter Sozou**[4]

Abstract. This article introduces the Computational Scientific Discovery symposium and gives an overview of the contributed papers.

1 INTRODUCTION

Science is fundamentally about explaining phenomena in the world. It involves an interplay between theories (hypotheses describing mechanisms or rules governing processes) and data (observations and measurements). There have been important advances in computational power, software, logic and statistical analysis techniques, together with, in many instances, an increase in the availability of data. These developments are enabling computers and computational devices to play an increasing role in the process of science. Specific developments have included: the use of evolutionary computation techniques to develop scientific theories and models, automation of the collection and analysis of data in the lab, new techniques in interpreting and analysing data, and developments in the availability of data with application to advancing science. The papers cover several aspects of computational scientific discovery including computational methods for automating the collection and analysis of laboratory data (King), computational methods for automating the generation and refinement of scientific theories and models (Colton et al. and Lane et al.) and methods for generalising the representation of data to facilitate the development of scientific models (Godziszewski).

2 CONTRIBUTIONS

The keynote address *Robot Scientists*: *Automating Biology and Chemistry* by Ross King concerns the automation of discovery in biology and chemistry using robot scientists which are physically implemented robotic systems that apply techniques from artificial intelligence to execute cycles of automated scientific experimentation. Such robot scientists can automatically execute cycles of hypothesis formation, selection of efficient experiments to discriminate between hypotheses, execution of experiments using laboratory automation equipment, and analysis of results. Using robot scientists advances have been made including gene identification in yeast and the discovery of lead compounds in certain tropical diseases. *The HR3 Discovery System: Design Decisions and Implementation Details* by Simon Colton, Ramin Ramezani and Maria Llano deals with the implementation of automated theory formation as an approach for scientific discovery tasks. In this method entire theories are formed about a domain and then investigated for various purposes, instead of taking a problem solving perspective where a single answer is sought. How earlier versions of the system have been improved through design choices and implementation alterations is discussed and illustrated by particular reference to case studies about a synthetic domain mirroring an aspect of medical diagnosis, and invariant discovery in formal methods.

Evolving Process-Based Models from Psychological Data Using Genetic Programming by Peter Lane, Peter Sozou, Mark Addis and Fernand Gobet is about the development of computational models using the semi-automated search technique of genetic programming to provide explanations of psychological data. A system for representing psychological data, a class of process-based models, and algorithms for evolving models is considered. The problem of bloating is mitigated by including cognitive constraints as part of the fitness function. The modelling aims to engage with the increasingly important issue in psychology of 'big data' by facilitating the conversion of empirical results into theories. *Experimental Logic as a Model of Development of Mathematical Knowledge* by Michal Godziszewski investigates the learnability of certain kinds of mathematical concepts. Building on existing work in formal learning theory about experimental logics stemming from trial and error predicates, an identification between mathematical knowability and algorithmic learnability is made. A framework of experimental logics equivalent to the notion of learnability is presented as a context for the main result which is an improvement on an existing theorem by giving it a tighter bound. This result helps to explain the fact that cognitively accessible mathematical concepts fall precisely within the scope of algorithmic learnability.

[1] Faculty of Performance, Media and English, Birmingham City Univ., B42 2SU, UK. Email: mark.addis@bcu.ac.uk
[2] Dept. of Psychological Sciences, Univ. of Liverpool, L69 7ZA, UK. Email: fernand.gobet@liverpool.ac.uk
[3] School of Computer Science, Univ. of Hertfordshire, AL10 9AB, UK. Email: p.c.lane@herts.ac.uk
[4] Dept. of Psychological Sciences, Univ. of Liverpool, L69 7ZA, UK. Email: p.sozou@liverpool.ac.uk