

Preface

The collation of large electronic databases of scientific and commercial information has led to a dramatic growth of interest in methods for discovering structures in such databases. These methods often go under the general name of data mining. One important subdiscipline within data mining is concerned with the identification and detection of anomalous, interesting, unusual, or valuable records or groups of records, which we call *patterns*. Familiar examples are the detection of fraud in credit-card transactions, of particular coincident purchases in supermarket transactions, of important nucleotide sequences in gene sequence analysis, and of characteristic traces in EEG records. Tools for the detection of such patterns have been developed within the data mining community, but also within other research communities, typically without an awareness that the basic problem was common to many disciplines. This is not unreasonable: each of these disciplines has a large literature of its own, and a literature which is growing rapidly. Keeping up with any one of these is difficult enough, let alone keeping up with others as well, which may in any case be couched in an unfamiliar technical language. But, of course, this means that opportunities are being lost, discoveries relating to the common problem made in one area are not transferred to the other area, and breakthroughs and problem solutions are being rediscovered, or not discovered for a long time, meaning that effort is being wasted and opportunities may be lost.

The aims of this workshop were to draw together people from the variety of disciplines concerned with this common interest, and to attempt to characterize more soundly the fundamental nature of their common interests. That is, (i) we sought to break down barriers, so that advances in one area could be communicated to others, and so that the common nature of the efforts could be recognized and taken advantage of, and (ii) we aimed to distil the essence of the common problem of pattern detection, so that the discipline could advance on a sound footing.

The various literatures concerned with pattern detection have arisen from practical needs. Perhaps inevitably because of this, there has been a heavy emphasis on the development of algorithms and methodology, with very little attention being paid to the development of a sound theory. But a sound theoretical basis is important, if one is to characterize those areas which need new research, if one is to identify strategies for problems thought to be impossible, and if one is to be able to transfer methodology between entirely different application areas. Perhaps above all, a sound theory is important if one is to tackle the problems which bedevil pattern detection, such as the occurrence of spurious patterns by chance alone, patterns arising from data distortion, and the issue of patterns which, though real, are of no practical significance or which are already well-known. Without addressing issues such as these, pattern detection (and data

Organization

The ESF Exploratory Workshop on Pattern Detection and Discovery was organized by the Department of Mathematics, Imperial College of Science, Technology and Medicine.

Executive Committee

Chair: David J. Hand, Imperial College of Science,
Technology and Medicine
Organization: Richard J. Bolton, Imperial College of Science,
Technology and Medicine
Publication: Niall M. Adams, Imperial College of Science,
Technology and Medicine

Referees

Niall Adams
Richard Bolton
David Hand
Yannis Manolopoulos
Nicolas Pasquier
Graeme Richards

Sponsoring Institutions

European Science Foundation
The Royal Statistical Society
Barclaycard
Fair, Isaac

mining more broadly) is likely to suffer a backlash as users find that the inflated promises are not being fulfilled.

The workshop was made possible by generous funding from several sources:

- The European Science Foundation provided the bulk of the funding. Without its support, this workshop would not have been possible.
- The Royal Statistical Society saw the value of what we were trying to do, and its importance to statisticians, and to statistics as a discipline.
- Barclaycard recognized the relevance of our aims for organizations such as theirs.
- Fair, Isaac also saw the potential importance of what we were trying to do.

To all these organizations we are very grateful.

July 2002

David J. Hand
Richard J. Bolton
Niall M. Adams

Participants at the Workshop

Niall Adams

Department of Mathematics, Imperial College, UK
n.adams@ic.ac.uk

Helena Ahonen-Myka

Department of Computer Science, University of Helsinki, Finland
helena.ahonen-myka@cs.helsinki.fi

Richard J. Bolton

Department of Mathematics, Imperial College, UK
r.bolton@ic.ac.uk

Jean-Francois Boulicaut

INSA-Lyon Scientific and Technical University, France
jfboulic@lisi.insa-lyon.fr

Paul Cohen

Experimental Knowledge Systems Laboratory, Department of Computer Science,
University of Massachusetts, USA
cohen@cs.umass.edu

Ursula Gather

Fachbereich Statistik, University of Dortmund, Germany
gather@statistik.uni-dortmund.de

Bart Goethals

School for Information Technology, University of Limburg (tUL), Belgium
bart.goethals@luc.ac.be

David J. Hand

Department of Mathematics, Imperial College, UK
d.j.hand@ic.ac.uk

Frank Höppner

Department of Computer Science, University of Applied Sciences BS/WF,
Germany
hoepfner@et-inf.fho-emden.de

Xiao Hui Liu

Department of Information Systems and Computing, Brunel University, UK
xiaohui.liu@brunel.ac.uk

Marzena Kryszkiewicz

Institute of Computer Science, Warsaw University of Technology, Poland
mkr@ii.pw.edu.pl

Heikki Mannila
Department of Computer Science, University of Helsinki, Finland
Heikki.Mannila@hut.fi

Dunja Mladenić
Department for Intelligent Systems, J. Stefan Institute, Slovenia
Dunja.Mladenic@ijs.si

Katharina Morik
Fachbereich Informatik, University of Dortmund, Germany
morik@ls8.cs.uni-dortmund.de

Pierre-Yves Rolland
Atelier de Modélisation et de Prévision, Université d'Aix-Marseille III, France
p-y.rolland@yahoo.com

Daniel Sánchez
Department of Computer Science and Artificial Intelligence (DECSAI),
University of Granada, Spain
daniel@decsai.ugr.es

Arno Siebes
Centrum voor Wiskunde en Informatica, Amsterdam, The Netherlands
Arno.Siebes@cwi.nl

Myra Spiliopoulou
Department of E-Business, Leipzig Graduate School of Management (HHL),
Germany
myra@ebusiness.hhl.de

Antony Unwin
Institut für Mathematik, Universität Augsburg, Germany
Antony.Unwin@Math.Uni-Augsburg.de

Marek Wojciechowski
Politechnika Poznańska, Instytut Informatyki, Poland
Marek.Wojciechowski@cs.put.poznan.pl

Chris Workman
Center for Biological Sequence Analysis, Institute of Biotechnology,
The Technical University of Denmark, Denmark
workman@cbs.dtu.dk