# *ESF Exploratory Workshop on Pattern Detection and Discovery*

***London, UK, September 2002***

# *Final Report*

David J. Hand, Niall M. Adams, Richard J. Bolton
Department of Mathematics
Imperial College
London SW7 2BZ

## 1. Executive Summary

Pattern detection and discovery is the new data analytic discipline concerned with finding anomalous, interesting, unusual, or unexpected structures in large data sets. The discipline, a subdiscipline of data mining, has arisen as a consequence of increased computer power. Researchers from several different application domains, including text mining, bioinformatics, commercial transaction databases, and many others have independently carried out work in this area, typically without being aware that other workers, in other domains were also tackling what was fundamentally the same problem. In addition, researchers from several different theoretical backgrounds, notably statistics and computer science, have simultaneously turned their attention to such problems. The differences in emphasis between these background has the potential to provide a research synergy if the communication barriers can be broken down. It was the aim of this

Workshop to bring together people from different application areas and to begin to break down the communication barriers.

A total of 19 papers were presented at the meeting, each with a named discussant who had seen the paper before, and with each being followed by a general discussion. The papers were available on the Website prior to the meeting and were published in a book given to the participants at the meeting. The entire meeting concluded with an open discussion on the achievements of the Workshop and what might follow it. There was considerable support for further similar meetings, research collaborations, and further publications.

**2. Scientific Content**

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 characterise more soundly the fundamental nature of their aims. 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 recognised 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 characterise 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 mining more broadly, is likely to suffer a backlash as users find that the inflated promises are not being fulfilled.

The Workshop lasted for four days, and took the form of 19 invited presentations of 45 minutes, followed by discussions of 15 minutes, launched by a 5 minute opening contribution from someone given a particular remit to study the paper beforehand and prepare a discussion contribution. This system paid off handsomely, and some very lively discussions resulted. (One of the discussants remarked that he had learnt more from this small workshop, with its opportunity for extended discussions, than he had from any of the large conferences he had been to.) The papers appeared in a book which was presented to participants at the meeting (Hand DJ, Adams NM, and Bolton RJ (2002) *Pattern Detection and Discovery ESF Exploratory Workshop*. Berlin: Springer). The papers were also available on the website for people to study prior to the meeting. The Workshop concluded with an open discussion, in which participants outlined their views on both the technical and organisational aspects of the meeting.

The participants at the Workshop were diverse in several ways:

- many of the participants expressed pleasure that the Workshop had pulled together people working in related problems from different disciplines. Future meetings might also usefully invite someone from the image processing community.
- the presence of people from different theoretical bases provided a rich source of interaction.
- in particular, having both statisticians and computer scientists, with their different approaches to the problems, proved especially enlightening. One of the discussants remarked that he now recognised the importance of statistics and ‘now

knew what he needed to learn in statistics', while another commented that she was grateful for the opportunity to listen to statisticians since she had been aware that data mining researchers had in the past proposed 'solutions' which violated basic statistical principles. Another remarked that the meeting had brought home the importance of being aware of the achievements of statistics and of teaching statistics to computer scientists.

- there was a mix of age and experience in the participants, including PhD students and very experienced researchers.

The papers presented a mix of high level investigations of the meaning of 'pattern' in the context of the Workshop, papers presenting applications of discovery and detection in various areas (including text, music, biology, medicine, and other areas), and papers describing aspects of pattern detection algorithms.

Several of the papers adopted frequent itemsets and association rules as the basis of the definition of 'pattern' and described ways to find these in databases. One of the discussants wondered why so many of the papers had taken this approach, and asked if association rules were the best way of defining patterns. These issues were complemented by a discussion of whether such tools had found many useful practical applications. There was a feeling among some participants that association rule analysis may have entered the realm of interesting mathematics for its own sake, though this was strongly resisted by others. Some participants laid down a challenge to others to provide some convincing examples of association. Since the meeting, there has been an exchange of emails concerning this issue and it seems likely that an edited book of examples of successful real applications of association analysis will result.

As far as defining 'pattern' went, the relative importance of the probability that a configuration would occur and the entropy of that configuration was noted. Clearly the representation used for the event space is critical, and the difficulty of finding a good representation was noted. One of the discussants said he would like to see more discussion of the key issues at future meetings. Perhaps future meetings could include particular invited talks with titles 'The key issues of pattern discovery in domain X.'

Several of the discussants remarked that they would like to see the momentum of the Workshop maintained, and there was discussion of another such meeting probably next year. Germany seemed to be the most popular place, with two suggestions for possible venues. In general, it was felt that the meeting provided a great beginning, and there was considerable enthusiasm for maintaining the contacts established by the meeting.

The abstracts of the papers are given below. The full papers appear in Hand *et al* (op. cit.).

## *ABSTRACTS*

**David J. Hand:** ***Pattern detection and discovery***

Data mining comprises two subdisciplines. One of these is based on statistical modelling, though the large data sets associated with data mining lead to new problems for traditional modelling methodology. The other, which we term *pattern detection*, is a new science. Pattern detection is concerned with defining and detecting local anomalies within large data sets, and tools and methods have been developed in parallel by several applications communities, typically with no awareness of

developments elsewhere. Most of the work to date has focused on the development of practical methodology, with little attention being paid to the development of an underlying theoretical base to parallel the theoretical base developed over the last century to underpin modelling approaches. We suggest that the time is now right for the development of a theoretical base, so that important common aspects of the work can be identified, so that key directions for future research can be characterised, and so that the various different application domains can benefit from the work in other areas. We attempt to describe a unified approach to the subject, and also attempt to provide theoretical base on which future developments can stand.

**Katharina Morik:** ***Detecting interesting instances***

Most valid rules that are learned from very large and high dimensional data sets are not interesting, but are already known to the users. The dominant mode! of the overall data set may well suppress the interesting local patterns. The search for interesting local patterns can be implemented by a two step learning approach which first acquires the global models before it focuses on the rest in order to detect local patterns. In this paper, three sets of interesting instances are distinguished. For these sets, the hypothesis space is enlarged in order to characterize local patterns in a second learning step

**Arno Siebes and Zbyszek Struzik:** ***Complex data: mining using patterns***

There is a growing need to analyse sets of complex data, i.e. data in which the individual data items are (semi-) structured collections of data themselves, such as sets of time-series. To perform such analysis, one has to redefine familiar notions such as similarity on such complex data types. One can do that either on the data items

directly, or indirectly, based on features or patterns computed from the individual data items. In this paper, we argue that wavelet decomposition is a general tool for the latter approach

**Richard J. Bolton, David J. Hand, and Niall M. Adams:** ***Determining hit rate in pattern search***

The problem of spurious apparent patterns arising by chance is a fundamental one for pattern detection. Classical approaches, based on adjustments such as the Bonferroni procedure, are arguably not appropriate in a data mining context . Instead methods based on the false discovery rate - the proportion of flagged patterns which do not represent an underlying reality - may be more relevant. We describe such procedures and illustrate their application on a marketing dataset

**Paul Cohen , Brent Heeringa , and Niall M. Adams:** ***An unsupervised algorithm for segmenting categorical timeseries into episodes.***

This paper describes an unsupervised algorithm for segmenting categorical time series into episodes. The VOTING-EXPERTS algorithm first collects statistics about the frequency and boundary entropy of ngrams, then passes a window over the series and has two "expert methods" decide where in the window boundaries should be drawn. The algorithm successfully segments text into words in four languages. The algorithm also segments time series of robot sensor data into subsequences that represent episodes in the life of the robot. We claim that VOTING-EXPERTS finds meaningful episodes in categorical time series because it exploits two statistical characteristics of meaningful episodes.

**Antony Unwin:** ***If you can't see the pattern, is it there?***

Analytic methods are capable of finding structure in even complex data sets and, indeed, some methods will find structure whether it is there or not. Confirming and understanding analytic results can be difficult unless some way of visualising them can be found. Both global overviews and displays of local details are required and these have to be blended intelligently together. This paper discusses the development of coherent graphical tools for exploring and explaining large multidimensional data sets, emphasising the importance of using an interactive approach.

**Marek Wojciechowski and Maciej Zakrzewicz:** ***Dataset filtering techniques in constraint-based frequent pattern mining.***

Many data mining techniques consist in discovering patterns frequently occurring in the dataset. Typically the goal is to discover all the patterns whose frequency in the dataset exceeds a user-specified threshold. However, very often users want to restrict the set of patterns to be discovered by adding extra constraints on the structure of patterns. Data mining systems should be able to exploit such constraints to speed up the mining process. In this paper, we focus on improving the efficiency of constraint-based frequent pattern mining by using dataset filtering techniques. Dataset filtering conceptually transforms a given data mining task into an equivalent one operating on a smaller dataset. We present transformation rules for various classes of patterns: itemsets, association rules, and sequential patterns, and discuss implementation issues regarding integration of dataset filtering with well-known pattern discovery algorithms.

**Marzena Kryszkiewicz:** ***Concise representations of association rules***

Strong association rules are one of the basic types of knowledge. The number of rules is often huge, which limits their usefulness. Applying concise rule representations with appropriate inference mechanisms can lessen the problem. Ideally, a rule representation should be lossless (should enable derivation of all strong rules), sound (should forbid derivation of rules that are not strong) and informative (should allow determination of rules' support and confidence). In the paper, we overview the following lossless representations: representative rules, Duquenne-Guigues basis, proper basis, Luxenburger basis, structural basis, minimal non-redundant rules, generic basis, informative basis and its transitive reduction. For each representation, we examine whether it is sound and informative. For the representations that are not sound, we discuss ways of turning them into sound ones. Some important theoretical results related to the relationships among the representations are offered as well.

**Baptiste Jeudy and Jean-Francois Boulicaut*: Constraint-based discovery and inductive queries: application to association rule mining.***

Recently inductive databases (IDBs) have been proposed to afford the problem of knowledge discovery from huge databases. Querying these databases needs for primitives to: (1) select, manipulate and query data, (2) select, manipulate and query "interesting" patterns (i.e., those patterns that satisfy certain constraints) , and (3) cross over patterns and data (e.g., selecting the data in which some patterns hold). Designing such query languages is a long-term goal and only preliminary approaches have been studied, mainly for the association rule mining task. Starting from a discussion on the MINE RULE operator, we identify several open issues for the design of inductive databases dedicated to these descriptive rules. These issues concern not only the

offered primitives but also the availability of efficient evaluation schemes. We emphasize the need for primitives that work on more or less condensed representations for the frequent itemsets, e.g. the (frequent) δ-free and closed itemsets. It is useful not only for optimizing single association rule mining queries but also for sophisticated post-processing and interactive rule mining.

**Bart Goethals and Jan Van den Bussche:** ***Relational association rules: getting WARMeR.***

In recent years, the problem of association rule mining in transactional data has been well studied. We propose to extend the discovery of classical association rules to the discovery of association rules of conjunctive queries in arbitrary relational data, inspired by the WARMeR algorithm, developed by Dehaspe and Toivonen, that discovers association rules over a limited set of conjunctive queries. Conjunctive query evaluation in relational databases is well understood, but still poses some great challenges when approached from a discovery viewpoint in which patterns are generated and evaluated with respect to some well defined search space and pruning operators.

**M.Delgado, M.J. Martin-Bautista, D.Sanchez, and M.A. Vila:** ***Mining text data: special features and patterns***

Text mining is an increasingly important research field because of the necessity of obtaining knowledge from the enormous number of text documents available, especially on the Web. Text mining and data mining, both included in the field of information mining, are similar in some sense, and thus it may seem that data mining techniques may be adapted in a straightforward way to mine text. However, data

mining deals with structured data, whereas text represents special characteristics and is basically unstructured. In this context, the aims of this paper are three:

- to study particular features of text.
- to identify the patterns we may look for in text.
- to discuss the tools we may use for that purpose.

In relation with the third point we overview existing proposals, as well as some new tools we are developing by adapting data mining tools previously developed by our research group.

**Myra Spiliopoulou and Carsten Pohle:** ***Modelling and incorporating background knowledge in the web mining process***

The incorporation and exploitation of background knowledge in KDD is essential for the effective discovery of useful patterns and the elimination of trivial results. We observe background knowledge as a combination of beliefs and interestingness measures. In conventional data mining, background knowledge refers to the preferences and properties of the population under observation. In applications analysing the interaction of persons with a system, we identify one additional type of background knowledge, namely about the *strategies* encountered in pursuit of the interaction objectives. We propose a framework for the modelling of this type of background knowledge and us a template-based mining language to exploit it during the data mining process. We apply our framework on Web usage mining for Web marketing applications.

**Dunja Mladenić:** ***Modeling information in textual data combining labeled and unlabeled data.***

The paper describes two approaches to modeling word normalization (such as replacing "wrote" or "writing" by "write" ) based on the re-occurring patterns in: word suffix and the context of word obtained from texts. In order to collect patterns, we first represent the data using two independent feature sets and then fnd the patterns responsible for a particular word mapping. The modeling is based on a set of hand-labeled words of the form (word, normalized word) and texts Since the hand-labeling is a demanding task we investigate the possibility of improving our modeling by gradually adding unlabeled examples. Namely, we use the initial model based on word suffix to predict the labels. Then we enlarge the training set by the examples with predicted labels for which the model is the most certain. The experiment shows that this helps the context-based approach while largely hurting the suffix-based approach. To get an idea of the influence of the number of labeled instead of unlabeled examples, we give a comparison with the situation when simply more labeled data is provided.

**Helena Ahonen-Myka:** ***Discovery of frequent word sequences in text.***

We have developed a method that extracts all maximal frequent word sequences from the documents of a collection. A sequence is said to be frequent if it appears in more than $\sigma$ documents, in which $\sigma$ is the frequency threshold given. Furthermore, a sequence is maximal, if no other frequent sequence exists that contains this sequence. The words of a sequence do not have to appear in text consecutively. In this paper, we describe briefly the method for finding all maximal frequent word sequences in text and then extend the method for extracting generalized sequences from annotated texts, where each word has a set of additional, e.g. morphological, features attached to it. We

aim at discovering patterns which preserve as many features as possible such that the frequency of the pattern still exceeds the frequency threshold given.

**Pierre-Yves Rolland and Jean-Gabriel Ganascia:** ***Pattern detection and discovery: the case of music data mining.***

In this paper the problem of automatically detecting (or extracting, inducing, discovering) patterns from music data, is addressed. More specifically, approaches for extracting "sequential patterns" from sequences of notes (and rests) are presented and commented. Pecu1iarities of music data have direct impact on the very nature of pattern extraction and, correlatively, on approaches and algorithms for carrying it out. This impact is analyzed and paralleled with other kinds of data. Applications of musical pattern detection are covered, ranging from music analysis to music information retrieval.

**Frank Höppner:** ***Discovery of core episodes from sequences using generalization for defragmentation of rule sets.***

We consider the problem of knowledge induction from sequential or temporal data. Patterns and rules in such data can be detected using methods adopted from association rule mining. The resulting set of rules is usually too large to be inspected manually. We show that (amongst other reasons) the inadequacy of the pattern space is often responsible for many of these patterns: If the true relationship in the data is fragmented by the pattern space, it cannot show up as a peak of high pattern density, but the data is divided among many different patterns, often difficult to distinguish from incidental patterns. To overcome this fragmentation, we identify core patterns that are shared among specialized patterns. The core patterns are then generalized by

selecting a subset of specialized patterns and combining them disjunctively. The generalized patterns can be used to reduce the size of the set of patterns. We show some experiments for the case of labeled interval sequences, where patterns consist of a set of labeled intervals and their temporal relationships expressed via Allen's interval logic.

**Ursula Gather , Roland Fried , Michael Imhoff , and Claudia Becker:** ***Patterns of dependencies in dynamic multivariate data.***

In intensive care, clinical information systems permanently record more than one hundred time dependent variables. Besides the aim of recognising patterns like outliers, level changes and trends in such high-dimensional time series, it is important to reduce their dimension and to understand the possibly time-varying dependencies between the variables. We discuss statistical procedures which are able to detect patterns of dependencies within multivariate time series.

## 3. Final Programme

| | Monday 16th | Tuesday 17th |
|---|---|---|
| **09:00**-09:30 | **Registration** | **Session Chair: Paul Cohen**<br>Speaker: Arno Siebes<br>Discussant: Richard Bolton |
| 09:30-10:00 | **Introduction** | |
| **10:00**-11:00 | **Session Chair: Niall Adams**<br>Speaker: David Hand<br>Discussant: Antony Unwin | Speaker: Helena Ahonen-Myka<br>Discussant: Dunja Mladenic |
| **11:00**-11:30 | Tea & Coffee | Tea & Coffee |
| 11:30-12:30 | Speaker: Marzena Kryszkiewicz<br>Discussant: Jean-Francois Boulicaut | Speaker: Katharina Morik<br>Discussant: Myra Spiliopolou |
| 12:30-14:00 | **Lunch** | **Lunch** |
| **13:00** | | |
| | | |
| **14:00**-15:00 | **Session Chair: Heikki Mannila**<br>Speaker: Pierre-Yves Rolland<br>Discussant: Paul Cohen | **Session Chair: Pierre-Yves Rolland**<br>Speaker: Heikki Mannila<br>Discussant: Niall Adams |
| **15:00**-16:00 | Speaker: Frank Höppner<br>Discussant: Ursula Gather | Speaker: Daniel Sanchez<br>Discussant: Helena Ahonen-Myka |
| **16:00**-16:30 | Tea & Coffee | Tea & Coffee |
| 16:30-17:30 | Speaker: Antony Unwin<br>Discussant: Katharina Morik | Speaker: Bart Goethals<br>Discussant: Marek Wojciechowski |
| **17:00** | | |
| **18:00** | | |
| **19:00** | | |

| | Wednesday 18th | Thursday 19th |
|---|---|---|
| **09:00**-09:30 | **Session Chair: Katharina Morik**<br>Speaker: Paul Cohen<br>Discussant: Xiaohui Liu | **Session Chair: Antony Unwin**<br>Speaker: Richard Bolton<br>Discussant: Arno Siebes |
| 09:30-10:00 | | |
| **10:00**-11:00 | Speaker: Dunja Mladenic<br>Discussant: Daniel Sanchez | Speaker: Xiaohui Liu<br>Discussant: Frank Höppner |
| **11:00**-11:30 | Tea & Coffee | Tea & Coffee |
| 11:30-12:30 | Speaker: Ursula Gather<br>Discussant: David Hand | |
| 12:30-14:00 | **Lunch** | **Lunch and Close** |
| **13:00** | | |
| | | |
| **14:00**-15:00 | **Session Chair: Daniel Sanchez**<br>Speaker: Jean-Francois Boulicaut<br>Discussant: Bart Goethals | |
| **15:00**-16:00 | Speaker: Myra Spiliopolou<br>Discussant: Francesco Bonchi | |
| **16:00**-16:30 | Tea & Coffee | |
| 16:30-17:30 | Speaker: Marek Wojciechowski<br>Discussant: Marzena Kryszkiewicz | |
| **17:00** | | |
| **18:00** | | |
| **19:00** | **Conference dinner** | |

| | |
|---|---|
| Talk 1 | David Hand: *Patterns* |
| Talk 2 | Marzena Kryszkiewicz: *Concise Representations Of Association Rules* |
| Talk 3 | Pierre-Yves Rolland: *Pattern Detection And Discovery: The Case Of Music Data Mining* |
| Talk 4 | Frank Hoeppner: *Rule Discovery From Sequential Data: Using Generalization For Defragmentation* |
| Talk 5 | Antony Unwin: *If You Can't See The Pattern, Is It There?* |
| Talk 6 | Arno Siebes: *Searching Patterns Over Patterns* |
| Talk 7 | Helena Ahonen-Myka: *Discovery Of Frequent Word Sequences In Text* |
| Talk 8 | Katharina Morik: *Detecting Interesting Instances* |
| Talk 9 | Heikki Mannila: *Patterns In Biological Data* |
| Talk 10 | Daniel Sanchez: *Mining Text Data: Special Features And Patterns*. |
| Talk 11 | Bart Goethals: *Relational Association Rules: Getting Warmer* |
| Talk 12 | Paul Cohen: *Are Patterns Real? Towards Universal, Objective Features Of Patterns* |
| Talk 13 | Dunja Mladenic: *Modeling Information In Textual Data By Using Sequential Modeling And Unlabeled Data* |
| Talk 14 | Ursula Gather: *Patterns Of Dependencies In Dynamic Multivariate Data* |
| Talk 15 | Jean-Francois Boulicaut: *Constraint-Based Discovery Of Patterns And Inductive Query Evaluation* |
| Talk 16 | Myra Spiliopoulou: *Embedding Background Knowledge Into The Web Usage Mining Process* |
| Talk 17 | Marek Wojciechowski: *Dataset Filtering Techniques In Constraint-Based Frequent Pattern Mining* |
| Talk 18 | Richard Bolton: *Multiplicity: Guarding Against The Discovery Of Spurious Patterns* |
| Talk 19 | Xiaohui Liu: *Patterns: From Detection To Explanation* |

**4. Assessment of the Results**

There was universal agreement amongst the Workshop participants that it had been successful. In particular, it succeeded in:

- Establishing communication between researchers in different application areas.

- Establishing communication between researchers with different theoretical backgrounds (especially statistics and computer science).
- Making explicit the common nature of the efforts, despite the different application areas and theoretical approach.

Particular areas to be explored in more detail in the future should include

- Further discussion of formal definitions of the concept of ‘pattern’.
- Establishment of a sounder theoretical basis.
- Demonstration of the role and importance of the association rule approach.

In practical terms, it is hoped that we will see

- further workshops
- research collaborations
- possibly a further book illustrating practical merits of the association rule approach

**5. The Participants** (YR indicates Young Researcher)

| Participant | Email Address | Mailing Address |
|---|---|---|
| Heikki Mannila | Heikki.Mannila@hut.fi | Department of Computer Science, PO Box 26 (Teollisuuskatu 23), FIN-00014 University of Helsinki, Finland |
| Marzena Kryszkiewicz | mkr@ii.pw.edu.pl | Institute of Computer Science, Warsaw University of Technology, Ul Nowowiejska 15/19, 00665 Warszawa, Poland |
| Marek Wojciechowski | Marek.Wojciechowski@cs.put.poznan.pl | Politechnika Poznanska, Instytut Informatyki, ul. Piotrowo 3a, 60-965 Poznan, Poland |
| Frank Hoeppner | hoeppner@et-inf.fho-emden.de | University of Applied Sciences BS/WF, Department of Computer Science, Salzdahlumer Str. 46/48, 38302 Wolfenbüttel, GERMANY |
| Myra Spiliopoulou | myra@ebusiness.hhl.de | Leipzig Graduate School of Management (HHL), Department of E-Business, Jahnallee 59, D-04109 Leipzig, Germany |
| Helena Ahonen-Myka | helena.ahonen-myka@cs.helsinki.fi | Department of Computer Science, P. O. Box 26 (Teollisuuskatu 23), FIN-00014 University of Helsinki, Finland |
| Dunja Mladenic | Dunja.Mladenic@ijs.si | J. Stefan Institute, Department for Intelligent Systems, Jamova 39, 1001 Ljubljana, Slovenia |
| Jean-Francois Boulicaut | jfboulic@lisi.insa-lyon.fr | INSA Lyon, LISI, Bat. Blaise Pascal, F-69621 Villeurbanne Cedex, France |
| Katharina Morik | morik@ls8.cs.uni-dortmund.de | Universität Dortmund, Fachbereich Informatik, Lehrstuhl für Künstliche Intelligenz, LS VIII, D-44221 Dortmund |
| Ursula Gather | gather@statistik.uni-dortmund.de | Fachbereich Statistik, Universität Dortmund, D - 44221 Dortmund, Germany |
| Pierre-Yves Rolland (YR) | p_y_rolland@yahoo.com | Université d'Aix-Marseille III, Atelier de Modélisation et de Prévision, 15-19 allee Claude Forbin, 13627 AIX-EN-PROVENCE, France |
| Antony Unwin | Antony.Unwin@Math.Uni-Augsburg.de | Universität Augsburg, Institut für Mathematik, Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse, D-86135 Augsburg |
| Francesco Bonchi (YR) | bonchi@di.unipi.it | Dipartimento di Informatica, Corso Italia 40, 56125 Pisa ITALY |
| Arno Siebes | Arno.Siebes@cwi.nl | Centrum voor Wiskunde en Informatica, Kruislaan 413, NL-1098 SJ Amsterdam, P.O. Box 94079, NL-1090 GB Amsterdam, Netherlands |
| Daniel Sánchez (YR) | daniel@decsai.ugr.es | Department of Computer Science and Artificial Intelligence (DECSAI ), E.T.S. Ing. Informática, University of Granada, C/ Daniel Saucedo Aranda s/n, 18071- Granada, SPAIN |
| Bart Goethals (YR) | bart.goethals@luc.ac.be | University of Limburg (tUL), School for Information Technology, Universitaire Campus, B-3590 Diepenbeek, Belgium |
| Xiaohui Lui | xiaohui.liu@brunel.ac.uk | Department of Information Systems and Computing, Brunel University , Uxbridge, Middlesex, UB8 3PH, UK |
| David Hand | d.j.hand@ic.ac.uk | Department of Mathematics, Imperial College, 180 Queen's Gate, London, SW7 2BZ, UK |
| Niall Adams | n.adams@ic.ac.uk | Department of Mathematics, Imperial College, 180 Queen's Gate, London, SW7 2BZ, UK |
| Richard Bolton (YR) | r.bolton@ic.ac.uk | Department of Mathematics, Imperial College, 180 Queen's Gate, London, SW7 2BZ, UK |
| Paul Cohen (Self funded) | cohen@cs.umass.edu | EKSL, Department of Computer Science, 140 Governors Drive, University of Massachusetts, Amherst, MA 01003-4610, USA |
| Veronica Vinciotti (YR) | v.vinciotti@ic.ac.uk | Department of Mathematics, Imperial College, 180 Queen's Gate, London, SW7 2BZ, UK |