
DATA MINING IV

WIT*press*

WIT Press publishes leading books in Science and Technology.

Visit our website for new and current list of titles.

www.witpress.com

WIT*eLibrary*

Making the latest research accessible, the WIT electronic-library features papers presented at Wessex Institute of Technology's prestigious international conferences. To access the library and view abstracts free of charge please visit www.witpress.com

FOURTH INTERNATIONAL CONFERENCE ON DATA MINING
DATA MINING IV

CONFERENCE CHAIRMAN

N.F.F. Ebecken

COPPE, Federal University of Rio de Janeiro, Brazil

C.A. Brebbia

Wessex Institute of Technology, UK

A. Zanasi

TEMIS Text Mining Solutions, Italy

INTERNATIONAL SCIENTIFIC ADVISORY COMMITTEE

M. Costantino	H.M. Pires Capobianco
P. Giudici	S.O. Rezende
M. Gotttroy	R. Turra
T. Khabaza	D.E.N. Van den Poel
G. Lachtermacher	F. Wang
R. Maspons Bosch	R. Weber
N.M. Milic-Frayling	H.G. Zimmermann

Supported by:

www.ittoolbox.com



Organised by:

Wessex Institute of Technology, UK
COPPE, Federal University of Rio de Janeiro, Brazil

DATA MINING IV

Editors:

N.F.F. Ebecken

COPPE, Federal University of Rio de Janeiro, Brazil

C.A. Brebbia

Wessex Institute of Technology

A. Zanasi

TEMIS Text Mining Solutions, Italy

WITPRESS Southampton, Boston 

N.F.F. Ebecken

COPPE, Federal University of Rio de Janeiro, Brazil

C.A. Brebbia

Wessex Institute of Technology, UK

A. Zanasi

TEMIS Text Mining Solutions, Italy

Published by

WIT Press

Ashurst Lodge, Ashurst, Southampton, SO40 7AA, UK

Tel: 44 (0) 238 029 3223; Fax: 44 (0) 238 029 2853

E-Mail: witpress@witpress.com

<http://www.witpress.com>

For USA, Canada and Mexico

Computational Mechanics Inc

25 Bridge Street, Billerica, MA 01821, USA

Tel: 978 667 5841; Fax: 978 667 7582

E-Mail: infousa@witpress.com

<http://www.witpress.com>

British Library Cataloguing-in-Publication Data

A Catalogue record for this book is available
from the British Library

ISBN: 1-85312-806-6

ISSN: 1470-6326

*The texts of the papers in this volume were set
individually by the authors or under their supervision.
Only minor corrections to the text may have been carried
out by the publisher.*

No responsibility is assumed by the Publisher, the Editors and Authors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

© WIT Press 2004

Printed and bound in Great Britain by The MFK Group, Stevenage.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the Publisher.

Preface

The Fourth International Conference on Data Mining (“DATA MINING 2003”) follows the previous bi-annual conferences started in 1998 and hosted respectively in Rio de Janeiro, Cambridge and Bologna. Now the Conference is returning to Rio de Janeiro and changing the periodicity to annual. Despite competition from several other international conferences on data mining that started later, the research and applicative interest in our conference is continuously growing.

After each conference, the committee looks ahead to future conferences and redefines what it is that ICDM means to our attendees. Our vision for program content must reflect a perspective that is applicable for all participants and provide a critical mass across a broad spectrum in the multi-disciplinary field of data mining.

Data Mining today is recognized as a need to analyze and understand the information collected by government, business and scientific applications. Novel data, text and web mining application areas are fast emerging and these developments call for new perspectives and new approaches.

The objective of this conference was to discuss which insights about data structure make it possible to analyze the very large amount of information collected by Internet, business, scientific and government and develop innovative algorithms and data structures to speed computation.

In this conference many high quality papers were presented, coming from several countries, providing a meeting place for worldwide researchers from academia and industry and were organized in sections: Data and text mining, Clustering, Categorization, CRM, E-commerce and web data, Case studies, Post-processing and Knowledge Evaluation, Genomics and Bioinformatics, Novel Applications, Knowledge Management and Scalable Algorithms and High-Performance Platforms.

The editors are grateful to the authors and members of the Scientific Advisory Committee for their contribution and support, which has ensured the quality of the contents of this book.

The Editors
Rio de Janeiro.

Contents

Section 1: Data and text mining

Scanning once a large distributed database to mine global association rules by growing a prefix tree for each local transaction <i>F. Wang & N. Helian</i>	3
Network mining for managing a broadband network <i>M. Chartrand & J. Johnson</i>	19
A comparison of some tree based prediction tools <i>P. S. Coelho, G. Lachtermacher & N. Ebecken</i>	33
An ontology engineering approach for knowledge discovery from data in evolving domains <i>P. Gottgroy, N. Kasabov & S. Macdonell</i>	43
Pattern distance of time series <i>D. WangDa & RongGang</i>	53
Learning patterns through artificial contrasts with application to process control <i>E. Tuv & G. Runger</i>	63
Preprocessing method and similarity measures in clustering-based text mining: a preliminary study <i>S. Iiritano, M. Ruffolo & P. Rullo</i>	73
Towards on an optimized parallel KNN-FUZZY classification approach <i>J. L. A. Rosa, N. F. F Ebecken & M. C. A. Costa</i>	81
Global model of distributed data mining is not summing of local models <i>M. Soliman & M. Kantardzic</i>	89

Management of R&D projects using the data mining strategy <i>M. V. Rodriguez y Rodriguez & N. F. F. Ebecken</i>	99
Expertise location: can text mining help? <i>A. Ibrahim</i>	109
Critical analysis reporting environment (CARE) <i>A. Parrish, D. Brown, R. Stricklin & D. Turner</i>	119
An instrument to support data mining projects <i>M. Santos & C. Sousa</i>	129
Mining text databases on clients opinion for oil industry <i>M. C. S. Lopes, G. S. Terra, N. F. F. Ebecken & G. G. Cunha</i>	139
Identification of blade vibration causes in wind turbine generators <i>P. R. J. Campbell & K. Adamson</i>	149
Association rule mining using list representation <i>F. Wang, N. Helian & Y. J. Yip</i>	159
 Section 2: Clustering	
Introducing prior knowledge into the clustering process <i>W. F. Castilho, H. A. do Prado & M. Ladeira</i>	171
A feature selection Bayesian approach for a clustering genetic algorithm <i>E. R. Hruschka Jr, E. R. Hruschka & N. F. F. Ebecken</i>	181
Genetic algorithm using iterative shrinking for solving clustering problems <i>P. Fränti & O. Virmajoki</i>	193
Starcluster: a visualization, clustering and classification tool <i>O. M. Machado Filho, A. G. Evsukoff & N. F. F. Ebecken</i>	205
Identifying the insolvency profile in a telephony operator database <i>C. A. Reis Pinheiro, A. G. Evsukoff & N. F. F. Ebecken</i>	215
CloNI: clustering of \sqrt{N} -interval discretization <i>C. Ratanamahatana</i>	225
Modelling retail bank customers' needs with Bayesian Belief Networks <i>P. Lakatos, B. Egri & P. Aszalós</i>	235

Section 3: Categorization

Data classification by a fuzzy genetic system approach <i>R. P. Espíndola & N. F. F. Ebecken</i>	245
One-against-all multiclass classification via discrete support vector machines <i>C. Orsenigo & C. Vercellis</i>	255
Alternative strategies for decision list construction <i>D. A. Newlands & G. I. Webb</i>	265
Getting a clearer picture: a business application of data mining <i>J. H. Schoeman, M. Ground & M. C. Matthee</i>	275
Convex hulls as an hypothesis language bias <i>D. A. Newlands & G. I. Webb</i>	285

Section 4: CRM

A screening tool based on neural networks <i>A. M. Cister & N. F. F. Ebecken</i>	297
Ontologies, CRM, Data Mining: How to integrate? <i>M. Gotttroy & P. Gotttroy</i>	307
Quantifying credit-scoring performance <i>G. Verstraeten & D. Van den Poel</i>	317
Investigating purchasing patterns for financial services using Markov and MTD models <i>A. Prinzie & D. Van den Poel</i>	327
The impact of product features and intermediaries on customer retention <i>B. Larivière & D. Van den Poel</i>	337
Using data mining to estimate customer's buying potential <i>G. Dragosavac</i>	347
Manufacturer-retailer promotion competition: customisation of coupon target selection <i>W. Buckinx & D. Van den Poel</i>	357

The application of CRM in call centers <i>A. P. G. Serra & M. Martucci Júnior</i>	367
--	-----

Profiling in an augmented-commerce environment <i>O. Belo, A. Lourenço & J. Santos</i>	377
---	-----

Section 5: E-commerce and web data

A generic Data Mart architecture to support Web mining <i>J. D. Velásquez, H. Yasuda, T. Aoki & R. Weber</i>	389
---	-----

Qualifying Web data to meet expectations <i>A. Lourenço & O. Belo</i>	401
--	-----

The CRUD Cube <i>A. Sulaiman, J. Moreira de Souza & J. C. M. Strauch</i>	411
---	-----

Estimating term structure of interest rates: neural network vs one factor parametric models <i>F. Abid & M. B. Salah</i>	421
--	-----

Using data mining to personalize the e-mail marketing programs <i>A. Vana</i>	429
--	-----

Section 6: Case studies

A data mining approach for medium-term demand forecasting <i>G. S. Terra, M. C. S. Lopes & N. F. F. Ebecken</i>	437
--	-----

Airline financial efficiency <i>H. M. P. Capobianco, E. Fernandes & A. M. Cister</i>	447
---	-----

Analyzing a Brazilian convenience store sales tickets <i>M. G. Galindo, P. S. Coelho, G. Lachtermacher & N. Ebecken</i>	457
--	-----

Learning from the market: efficient lighting in China <i>K. Tiedemann</i>	469
--	-----

From IP port numbers to ADSL customer segmentation: knowledge aggregation and representation using Kohonen maps <i>F. Clérot & F. Fessant</i>	479
---	-----

Searching for indicators in educational processes through the use of a Learning Management System <i>L. Colazzo & A. Molinari</i>	489
---	-----

On the limitations of neural network techniques for analysing cause and effect relationships in manufacturing processes – a case study
R. S. Ransing, M. R. Ransing & R. W. Lewis 501

Data mining applied for analysis of fault sequences in electronic circuits
A. C. G. Oliveira, J. P. L. Bressan, L. E. Zárate & N. J. Vieira 511

Rough mereology for good circuit design
D. Kerr, B. Trottier & J. A. Johnson..... 523

Forecasting of the ultimate resistance of steel beams subjected to concentrated loads using data mining techniques
M. F. Santos, H. Quintela & P. Cruz..... 533

Section 7: Post-processing and knowledge evaluation

A critical review of rule surprisingness measures
D. R. Carvalho, A. A. Freitas & N. F. F. Ebecken..... 545

Precision and comprehensibility in the integration of regression rules
J. B. Pugliesi & S. O. Rezende..... 557

Section 8: Genomics and bioinformatics

Computer aided identification of biological specimens using self-organizing maps
E. J. Dean, A. P. Engelbrecht & A. Nicholas..... 569

A scalable data mining architecture for bioinformation
R. Li, Z. Zhang, S. Cao, Y. Zhu & Y. Li..... 583

Section 9: Novel applications

Data mining of chess chunks: a novel distance-based structure
A. Linhares..... 595

Independent components in telecommunication call-detail records
P. Lakatos, B. Egri & Z. Cseh 605

Section 10: Knowledge management

The importance of an adaptive gain variation method in back-propagation neural network learning algorithms <i>R. S. Ransing, M. R. Ransing & R. W. Lewis</i>	617
--	-----

Methodological problems in the Web Mining field <i>P. Mariani, M. Mezzanica & F. Verrecchia</i>	625
--	-----

Section 11: Scalable algorithms and high-performance platforms

An intelligent learning machine <i>S. Sayad, S. T. Balke & S. Sayad</i>	639
--	-----

Evaluating the scalability of data mining provider classifiers <i>C. L. Curotto & N. F. F. Ebecken</i>	651
---	-----

Parallel SVM for large data-set mining <i>L. Qian & T. Hung</i>	661
--	-----

Author Index	671
---------------------------	-----