Data Mining in E-Learning

WIT PRESS

WIT Press publishes leading books in Science and Technology.
Visit our website for the current list of titles.
www.witpress.com

WIT eLibrary

Home of the Transactions of the Wessex Institute, the WIT electronic-library provides the international scientific community with immediate and permanent access to individual papers presented at WIT conferences. Visit the WIT eLibrary at
http://library.witpress.com

Objectives of the Series

Advances in Management Information Series

Information and Communications Technologies have experienced considerable advances in the last few years. The task of managing and analysing ever-increasing amounts of data requires the development of more efficient tools to keep pace with this growth.

This series presents advances in the theory and applications of Management Information. It covers an interdisciplinary field, bringing together techniques from applied mathematics, machine learning, pattern recognition, data mining and data warehousing, as well as their applications to intelligence, knowledge management, marketing and social analysis. The majority of these applications are aimed at achieving a better understanding of the behaviour of people and organisations in order to enable decisions to be made in an informed manner. Each volume in the series covers a particular topic in detail.

The volumes cover the following fields:

Information
Information Retrieval
Intelligent Agents
Data Mining
Data Warehouse
Text Mining

Competitive Intelligence
Customer Relationship Management
Information Management
Knowledge Management

Series Editor

A. Zanasi

TEMIS Text Mining Solutions S.A.

Italy

Associate Editors

P.L. Aquilar
University of Extremadura
Spain

M. Costantino
Royal Bank of Scotland Financial Markets
UK

P. Coupet
TEMIS
France

N.J. Dedios Mimbela
Universidad de Cordoba
Spain

A. De Montis
Universita di Cagliari
Italy

G. Deplano
Universita di Cagliari
Italy

P. Giudici
Universita di Pavia
Italy

D. Goulias
University of Maryland
USA

A. Gualtierotti
IDHEAP
Switzerland

J. Jaafar
UiTM
Malaysia

G. Loo
The University of Auckland
New Zealand

J. Lourenco
Universidade do Minho
Portugal

D. Malerba
Università degli Studi
UK

N. Milic-Frayling
Microsoft Research Ltd
UK

G. Nakhaeizadeh
DaimlerChrysler
Germany

P. Pan
National Kaohsiung University of Applied Science
Taiwan

J. Rao
Case Western Reserve University
USA

D. Riaño
Universitat Rovira I Virgili
Spain

J. Roddick
Flinders University
Australia

F. Rodrigues
Poly Institute of Porto
Portugal

F. Rossi
DATAMAT
Germany

D. Sitnikov
Kharkov Academy of Culture
Ukraine

R. Turra
CINECA Interuniversity Computing Centre
Italy

D. Van den Poel
Ghent University
Belgium

J. Yoon
Old Dominion University
USA

N. Zhong
Maebashi Institute of Technology
Japan

H.G. Zimmermann
Siemens AG
Germany

Data Mining in E-Learning

Editors

C. Romero
University of Cordoba, Spain

S. Ventura
University of Cordoba, Spain

WIT*PRESS* Southampton, Boston

C. Romero
University of Cordoba

S. Ventura
University of Cordoba

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

WIT Press
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: 1-84564-152-3
ISSN: 1742-0172

Library of Congress Catalog Card Number: 2006920597

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 2006

Printed in Great Britain by Athenaeum Press Ltd., Gateshead.

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.

Contents

Preface xv

Biography xix

Part 1
An introduction to e-learning systems, data mining and their interactions 1

Chapter 1
Web-based educational hypermedia 3
P. De Bra

1 Introduction 3
2 Adaptive (educational) hypermedia 5
3 The AHAM reference architecture 8
4 A general-purpose adaptive web-based platform 10
4.1 Overall architecture of AHA! 10
4.2 The AHA! authoring tools 13
5 Questions, quizzes and tasks 14
6 Adapting to learning styles 15
7 Conclusions 16

Chapter 2
Web mining for self-directed e-learning 21
P. Desikan, C. DeLong, K. Beemanapalli, A. Bose & J. Srivastava

1 Introduction 21
2 Why self-directed e-learning? 23
3 Web-based self-directed e-learning applications 24
3.1 Google Scholar 24
3.2 Westlaw 25
3.3 CiteSeer 25
3.4 LexisNexis 27
3.5 Knowledge management systems 27
3.6 Dr. Spock's child care 28

4 Gaps in existing technology 29
4.1 Lack of community collaboration 29
4.2 Time management 30
4.3 Not self-improving 30
4.4 Implicit relationships not mapped 30
5 Web mining 31
5.1 Web mining taxonomy 32
5.1.1 Web content mining 32
5.1.2 Web structure mining 32
5.1.3 Web usage mining 33
5.2 Web mining research: state of the art 33
5.3 Web mining applicable to e-learning 34
6 Future directions of research 35
6.1 Usage rules 35
6.2 Keyword clustering: the conceptual thesaurus 35
6.3 Recommendation mining 36
6.4 Smart results: model of relevance 36
6.5 Intelligent knowledge management systems 36
7 Conclusion 37

Chapter 3
Data mining for the analysis of content interaction in web-based learning and training systems 41
C. Pahl

1 Introduction 41
2 Interaction and behavior 43
2.1 Learning and training interaction 43
2.2 Implementing interaction 43
2.3 An abstract model of content interaction 44
2.4 The interactive database learning environment 44
3 Data and web usage mining 45
3.1 Web usage mining in the educational context 45
3.2 Data and web mining techniques 46
3.3 Education-specific web usage mining 47
4 Session statistics 48
5 Session classification 49
6 Behavioral patterns 50
7 Time series 52
8 Conclusions 53

Chapter 4
On using data mining for browsing log analysis in learning environments 57
F. Wang

1 Introduction 57

2 Data mining 59
2.1 Association mining 59
2.2 Clustering 59
2.3 Web usage mining 60
3 Recommendation systems 60
3.1 Content-based filtering systems 60
3.2 Collaborative filtering systems 61
3.3 Recommendation systems based on association rules mining technologies 61
4 The research framework 62
5 Construction of browsing content structure 64
5.1 Data pre-processing 64
5.2 Model definition and construction 65
5.3 Model application 68
5.4 Summary statements 69
6 Personalized recommendation based on association mining 69
6.1 Model definition and construction 70
6.1.1 User browsing similarity in time-framed navigation sessions 71
6.1.2 The HBM clustering algorithm 71
6.1.3 Mining association rules 71
6.2 Model application 72
6.2.1 User classification 72
6.2.2 The window-sliding method 72
6.2.3 The maximal-matching method 73
6.3 Summary statements 73
7 Concluding remarks 73

Part 2
Case studies experiences of applying data mining techniques in e-learning systems 77

Chapter 5
Recommender systems for e-learning: towards non-intrusive web mining 79
O.R. Zaïane

1 Introduction 79
2 Collaborative filtering: how most systems work 81
3 Desired recommender systems in an online learning environment 82
4 Non-intrusive methods for recommendation 83
4.1 E-learning recommender with association rules 84
4.2 A model with clustering 85
5 Hybrid methods for recommendations 86
5.1 Architecture of a hybrid recommender system 86

5.2 User and visit session identification 87
5.3 Visit mission identification 87
5.4 Evaluating hybrid recommenders 90
6 Conclusion 93

Chapter 6
Active, context-dependent, data-centered techniques for e-learning: a case study of a research paper recommender system 97
T. Tang & G. McCalla

1 Introduction 98
2 A research paper recommender system 99
3 Two experiments in paper recommendation 102
3.1 What learners want: a survey 102
3.2 Evaluating pedagogy-oriented hybrid collaborative filtering 104
3.2.1 Simulation setup 104
3.2.2 Evaluation metrics and control variables 105
3.2.3 Experimental results and discussion 106
4 The ecological approach 107
5 Conclusion 111

Chapter 7
Applying web usage mining for the analysis of behavior in web-based learning environments 117
K. Becker, M. Vanzin, C. Marquardt & D. Ruiz

1 Introduction 117
2 The process of WUM 119
2.1 Pre-processing phase 119
2.2 Data mining phase 120
2.3 Pattern analysis phase 120
2.4 Support environments 121
3 WUM challenges in practice: a case study 121
3.1 Pre-processing phase 122
3.2 Data mining phase 122
3.3 Pattern analysis phase 123
3.4 Lessons learned 124
3.4.1 Pre-processing issues 124
3.4.2 Pattern analysis issues 124
4 LogPrep: a customizable pre-processing tool 125
4.1 Configuration language and configuration template 126
4.2 Customization features 127
5 OR3: ontology-based rule rummaging and retrieval tool 127
5.1 Ontology representation 128
5.2 Conceptual pattern representation 129

5.3 Pattern rummaging 130
5.4 Pattern clustering 131
5.5 Pattern filtering 131
5.5.1 Filter definition 131
5.5.2 Equivalence filtering 132
5.5.3 Similarity filtering 132
6 Discussions 133
7 Conclusions and future work 135

Chapter 8
Association analysis for a web-based educational system 139
B. Minaei-Bidgoli, P. Tan, G. Kortemeyer &, W.F. Punch

1 Introduction 140
2 Background 142
2.1 Association analysis 142
2.2 Data mining for online education systems 143
2.3 Related work 143
3 Contrast rules 144
3.1 Example 1: cr_1 (difference of confidence) 145
3.2 Example 2: cr_2 (difference of proportion) 146
3.3 Example 3: cr_3 (correlation and chi-square) 147
3.4 Contrast rules and interestingness measures 147
4 Algorithm 148
5 Experiments 149
5.1 Data model and attributes 149
5.2 Data sets 151
5.3 Results 152
5.3.1 Difference of confidences 152
5.3.2 Difference of proportions 153
5.3.3 Chi-square 154
6 Conclusion 154

Chapter 9
Data mining in personalizing distance education courses 157
W. Hämäläinen, T.H. Laine & E. Sutinen

1 Introduction 157
2 General paradigms for ITSs 158
3 Data description 160
4 Correlations and Linear Regression Models 162
5 Association rules and probabilistic models 164
6 Evaluating the predictive power by cross-validation 167
7 Conclusions 169

Chapter 10
Rule mining with GBGP to improve web-based adaptive educational systems 173
C. Romero, S. Ventura, C. Hervás & P. González

1 Introduction 173
2 Data mining in e-learning systems 174
3 Students' usage data 176
4 Knowledge discovery process 178
4.1 Rule discovery with GBGP 179
5 EPRules tool 181
6 Experimental results 184
6.1 Description of the discovered information 185
7 Conclusions and future work 186

Chapter 11
Identifying gifted students and their learning paths using data mining techniques 191
S. Bae, S.H. Ha & S.C. Park

1 Introduction 191
2 Data mining in education 192
2.1 Gifted education: a short review 192
2.2 Web mining 193
3 Identification of gifted students using neural network and data mining 193
3.1 Design of questionnaire 194
3.2 Clustering and classification of gifted students 195
3.3 Creating a giftedness quotient using neural networks 195
3.4 Applications 196
3.4.1 General and specific test: identifying of giftedness and their type 197
3.4.2 Evaluating the results of the identification test 199
4 Web mining for extracting learning path 199
4.1 Customized education 201
4.2 Virtual knowledge structure 203
4.3 Application of web mining to a web-based education system 203
5 Conclusions 205

Chapter 12
Data mining to support tutoring in virtual learning communities: experiences and challenges 207
E. Gaudioso & L. Talavera

1 Introduction 207
2 Data mining 208

3 Defining data mining tasks for supporting tutoring 210
3.1 A working problem 212
4 Data pre-processing 212
5 Building predictive models 215
5.1 Supporting tutors in course assessment 216
5.2 Supporting tutors in anticipating student activity levels 219
6 Building descriptive models 220
6.1 Supporting tutors in determining behavioral patterns 220
6.2 Supporting tutors in course assessment 222
7 Challenges and lessons learned 223
7.1 Definition of data mining tasks 223
7.2 Data preparation 223
7.3 Model building 224
8 Concluding remarks 225

Chapter 13
Analysis of user navigational behavior for e-learning personalization 227
E. Mor, J. Minguillón & J.M. Carbó

1 Introduction 227
2 E-learning environments 230
2.1 The UOC virtual campus 231
2.2 Virtual campus architecture and services 232
3 Navigational behavior analysis 233
3.1 Navigational levels 234
4 Experimental results 235
4.1 Server log files 236
4.2 Data pre-processing and feature extraction 238
4.3 Web mining 241
4.3.1 Variable relevance 241
4.3.2 Unsupervised clustering 242
4.4 Data fusion 243
5 Conclusions 243

Chapter 14
Automatically constructing an e-textbook via web mining 247
J. Chen & Q. Li

1 Introduction 247
2 System architecture 248
3 Building concept hierarchies 250
4 Topic content identification 251
4.1 Segmenting web pages 251
4.2 Identifying the topic content for topic pages 252
5 Ranking algorithm 253
5.1 Original rank 254

5.2 Cue phrases ... 254
5.3 Weighty tags ... 254
5.4 Concept descriptions and definitions ... 255
5.4.1 Pattern set mining ... 255
5.4.2 Concept discovery ... 258
5.5 Integrating extracted features ... 259
6 Conclusions ... 259

Chapter 15
Online outlier detection of learners' irregular learning processes ... 261
M. Ueno

1 Introduction ... 261
2 Learning management system 'Samurai' ... 263
3 Online outlier detection ... 264
3.1 Data ... 264
3.2 Model ... 265
3.3 Model ... 266
3.4 Outlier detection curves and examples ... 266
4 Simulation experiments ... 267
5 System ... 269
6 Evaluation ... 270
7 Animated agent to enhance learning ... 272
8 Conclusions ... 274

Chapter 16
Use of data mining to examine an outreach call center's effectiveness and build a predictive model for classifying future marketing targets ... 279
J. Luan, C. Summa & M. Wieland

1 Background ... 279
2 Three key questions addressed ... 281
3 Data sources ... 281
4 Design and method ... 282
5 Findings ... 283
5.1 Yield ... 283
5.1.1 Step one for answering question one: overall effect of the presence of call center ... 283
5.1.2 Step two for answering question one: computing specific yield rates ... 284
5.2 Predictive modeling ... 285
5.2.1 Data mining rationale and discoveries ... 291
6 Discussion ... 296

Index **299**

Preface

The design and implementation of web-based education systems have grown exponentially in the last years, spurred by the fact that neither students nor teachers are bound to a specific location and that this form of computer-based education is virtually independent of any specific hardware platforms. These systems accumulate a vast amount of information which is very valuable in analyzing students' behavior and to assist authors in detecting possible errors, shortcomings and improvements. However, due to the vast quantities of data these systems can generate daily, it is very difficult to manage manually, and authors demand tools which assist them in this task, preferably on a continuous basis. A very promising area to attain this objective is the use of data mining.

In the last years, researchers have begun to investigate various data mining methods to help teachers improve e-learning systems. These methods allow them to discover new knowledge based on students' usage data. The same idea has already been successfully applied in e-commerce systems and is now very popular. Comparatively little work in this direction has yet been released in e-learning systems. However, the number of contributions in this area have grown, both in international conferences (International Conference on Computers in Education, International Conference on Web-based Learning, World Conference on Open Learning and Distance Education, International Conference on Adaptive Hypermedia and Adaptive Web-based Systems, International Conference on User Modeling, International Conference on Intelligent Tutoring Systems, Pacific-Asia Conference on Knowledge Discovery and Data Mining, Genetic and Evolutionary Computation Conference, etc.) and in scientific journals (International Journal on E-Learning, IEEE Education, Computers & Education, Journal of Educational Technology Systems, Journal of Interactive Learning Research, User Modeling and User-Adapted Interaction, etc.). The main purpose of this book is to show the current state of this research area.

This book consists of openly solicited and invited chapters, written by international researchers and leading experts on the application of data mining techniques in e-learning systems. The book consists of 16 chapters organized in two parts. In the first part of the book (Chapters 1–4) we present an introduction to e-learning systems, data mining and the interaction between the two areas. In the

second part of the book (Chapters 5–16) we present several case studies and experiences of applying data mining techniques in e-learning systems. In particular, the chapters cover the following:

Chapter 1 describes recent and ongoing research in web-based education systems, in particular adaptive web-based educational hypermedia.

Chapter 2 describes specific examples of self-directed e-learning and how their functionality and utility can be improved through the use of web mining technology.

Chapter 3 proposes the use of web usage mining for the analysis and evaluation of learner interactions with contents in web-based learning and training systems.

Chapter 4 describes some models and methods of analyzing browsing log data to construct a browsing behavioral model which is helpful in supporting e-learning applications.

Chapter 5 suggests the use of web mining techniques as non-intrusive method to build an agent that could recommend actions, resources or simply links to follow, in a e-learning environment.

Chapter 6 proposes an e-learning system that recommends research papers to students wishing to study an area of research.

Chapter 7 describes a case study and an extensible and customizable preprocessing and pattern analysis tools for supporting the web usage mining process.

Chapter 8 introduces an approach for predicting student performance by the discovery of interesting contrast rules within a web-based educational system.

Chapter 9 introduces general paradigms for tackling intelligent tutoring systems and applies various data mining schemes to describe and predict student performance.

Chapter 10 proposes the use of evolutionary algorithms as an association rule mining method for discovering interesting relationships in student's usage data.

Chapter 11 proposes a neural network model for identification of gifted students and a web mining framework for distance education to provide their learning path.

Chapter 12 reviews some experiences using data mining to analyze data obtained from e-learning courses based on virtual communities.

Chapter 13 describes a framework for studying the navigational behavior of the users in an e-learning environment integrated in a virtual campus to include the concept of recommended itinerary.

Chapter 14 proposes the construction of an e-textbook automatically using data mining methodologies for a user-specified topic hierarchy and examines how web content mining can be applied to aid e-learning experiences.

Chapter 15 proposes a method of online outlier detection of learners' irregular learning processes using their response time to e-learning content.

Finally, Chapter 16 proposes the use of data mining in enrollment management.

In conclusion, we hope the reader will find this book a truly helpful guide and a valuable source of information about the application of data mining techniques in e-learning systems.

Cristóbal Romero & Sebastián Ventura
Córdoba, July 2005

Biography

Dr. Cristóbal Romero is an Assistant Professor in the Computer Science Department of the University of Córdoba, Spain. He received his Ph.D. in Computer Science from the University of Granada in 2003. His research interests lie in artificial intelligence and data mining in education.

Dr. Sebastián Ventura is an Associate Professor in the Computer Science Department of the University of Córdoba, Spain. He received his Ph.D. in Sciences from the University of Córdoba in 1996. His research interests lie in soft-computing and its applications.