About this Series

The series “Studies in Computational Intelligence” (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the worldwide distribution, which enable both wide and rapid dissemination of research output.
Educational Data Mining

Applications and Trends
Educational Data Mining (EDM) is a new discipline based on the Data Mining (DM) grounds (i.e., the baseline is composed of models, tasks, methods, and algorithms) to explore data from educational settings to find out descriptive patterns and predictions that characterize learners behaviors and achievements, domain knowledge content, assessments, educational functionalities, and applications.

This book introduces concepts, models, frameworks, tasks, methods, and algorithms, as well as tools and case studies of the EDM field. The chapters make up a sample of the work currently achieved in countries from the five continents, which illustrates the world labor of the EDM arena. According to the nature of the contributions accepted for this volume, four kinds of topics are identified as follows:

- **Profile** shapes a conceptual view of the EDM. It provides an introduction of the nature, purpose, components, processes, and applications. Through this section, readers are encouraged to: make an incursion in the EDM field, facilitate the extraction of source data to be mined, and acquire consciousness of the usefulness of this sort of approaches to support education policies.

- **Student Modeling** is an essential functionality of Computer-Based Educational Systems (CBES) to adapt their performance according to users needs. Most of the EDM approaches are oriented to characterize diverse student traits, such as: behavior, acquired domain knowledge, personality, and academic achievements by means of machine learning methods.

- **Assessment** evaluates learners’ domain knowledge acquisition, skills development, and achieved outcomes, as well as reflection, inquiring, and sentiments are essential subjects to be taken into account by CBES. The purpose is to differentiate student proficiency at the finer grained level through static and dynamic testing, as well as online and offline assessment.

- **Trends** focus on some of the new demands for applying EDM, such as text mining and social networks analysis. Both targets represent challenges to cope with huge, dynamical, and heterogeneous information that new generations of students produce in their every day life. These paradigms represent new educational settings such as: ubiquitous-learning and educational networking.
This volume is the result of one year of effort, where more than 30 chapters were rigorous peer reviewed by a team of 60 reviewers. After several cycles of chapter submission, revision, and tuning based on the Springer quality principles, 16 works were approved, edited as chapters, and organized according to the prior four topics. So the Part I corresponds to Profile that includes Chaps. 1–3; the Part II represents Student Modeling, which embraces Chaps. 4–8; the Part III concerns Assessment and has Chaps. 9–12; the Part IV is related to Trends through Chaps. 13–16. A profile of the chapters is given next:

1. Chapter 1 provides a bibliographic review of studies made in the field of Educational Data Mining (EDM) to identify diverse aspects related to techniques and contributions in the field of computer-based learning. Authors pursue to facilitate the use and understanding of Data Mining (DM) techniques to help the educational specialists to develop EDM approaches.

2. Chapter 2 overcomes the lack of data preprocessing literature through the detailed exposition of the tasks involved to extract, clean, transform, and provide suitable data worthy to be mined. The work depicts educational environments and data they offer; as well as gives examples of Moodle data and tools.

3. Chapter 3 illustrates how EDM is able to support government policies devoted to enhance education. The work shapes the context of basic education and how the government aims at reforming the current practices of evaluation to academics and students. Several findings extracted from surveys are shown to highlight the opinion of the community and provide an initial diagnostic.

4. Chapter 4 presents the Student Knowledge Discovery Software, a tool to explore the factors having an impact on the student success based on student profiling. Authors deeply outline how to implement the software to help educational organizations to better understand knowledge discovery processes.

5. Chapter 5 explains how to automate the detection of student’s personality and behavior in an educational game called Land Science. The work includes a model to learn vector space representations for various extracted features. Learner personality is detected by combining the features spaces from psycholinguistics and computational linguistics.

6. Chapter 6 attempts to predict student performance to better adjust educational materials and strategies throughout the learning process. Thus, authors design a multichannel decision fusion approach to estimate the overall student performance. Such an approach is based on the performance achieved in assignment categories.

7. Chapter 7 explores predictive modeling methods for identifying students who will most benefit from tutor interventions. Authors assert how the predictive capacity of diverse sources of data changes as the course progresses, as well as how a student’s pattern of behavior changes during the course.

8. Chapter 8 predicts learner achievements by recording learner eye movements and mouse click counts. The findings claim: the most important eye metrics
that predict answers in reasoning questions include total fixation duration, number of mouse clicks, fixation count, and visit duration.

9. **Chapter 9** focuses on coherence expressed in research protocols and thesis. Authors develop a coherence analyzer that employs Latent Semantic Analysis to mine domain knowledge. The analysis outcomes are used to grade students and provide online support with the aim at improving their writings.

10. **Chapter 10** tailors an approach to automatically generate tests. It recognizes competence areas and matches the overall competence level of target students. The approach makes use of a concept map of programming competencies and a method for estimating the test item difficulty. The contribution is evaluated in a setting where its results are compared against a solution that randomly searches within the item space to find an adequate test.

11. **Chapter 11** outlines methods oriented to support teachers’ understanding of students’ activities on Exploratory Learning Environments (ELE). The work includes an algorithm that intelligently recognizes student activities and visualization facilities for presenting these activities to teachers. The approach is evaluated using real data obtained from students using an ELE to solve six representative problems from introductory chemistry courses.

12. **Chapter 12** adopts the concept of entropy from information theory to find the most dependent test items in student responses. The work defines a distance metric to estimate the amount of mutual independency between two items that is used to quantify how independent two items are in a test. The trials show: the approach for finding the best dependency tree is fast and scalable.

13. **Chapter 13** proposes ReaderBench, an environment for assessing learner productions and supporting teachers. It applies text mining to perform: assessment of the reading materials, assignment of texts to learners, detection of reading strategies, and comprehension evaluation to fostering learner’s self-regulation process. All of these tasks were subject of empirical validations.

14. **Chapter 14** analyzes a data set consisting of student narrative comments that were collected using an online process. The approach uses category vectors to depict instructor traits and a domain-specific lexicon. Sentiment analysis is also used to detect and gauge attitudes embedded in comments about each category. The approach is useful to instructors and administrators, and is a vehicle to analyze student perceptions of teaching to feedback the educational process.

15. **Chapter 15** introduces E-learning Web Miner, a tool that assists academics to discover student behavior profiles, models of how they collaborate, and their performance with the purpose of enhancing the teaching-learning process. The tool applies Social Network Analysis (SNA) and classification techniques.

16. **Chapter 16** depicts an approach to assess the students’ participation by the analysis of their interactions in social networks. It includes metrics for ranking and determining roles to analyze the student communications, the forming of groups, the role changes, and the interpretation of exchanged messages.
I express my gratitude to authors, reviewers, my assistant Leonor Adriana Cárdenas, the Springer editorial team, and the editors Dr. Thomas Ditzinger and Prof. Janusz Kacprzyk for their valuable collaboration to fulfill this work.

I also acknowledge the support given by the Consejo Nacional de Ciencia y Tecnología (CONACYT) and the Instituto Politécnico Nacional (IPN) of Mexico through the grants: CONACYT-SNI-36453, CONACYT 118862, CONACYT 118962-162727, IPN-SIP-20131093, IPN-COFAA-SIBE-DEBEC/647-391/2013.

Last but not least, I acknowledge the strength given by my Father, Brother Jesus, and Helper, as part of the research projects of World Outreach Light to the Nations Ministries (WOLNM).

August, 2013

Alejandro Peña-Ayala
### Contents

#### Part I  Profile

1 Which Contribution Does EDM Provide to Computer-Based Learning Environments? ........................................... 3  
   Nabila Bousbia and Idriss Belamri  
   1.1 Introduction ................................................................. 4  
   1.2 Educational Data Mining ............................................. 5  
       1.2.1 Definition ............................................................ 5  
       1.2.2 Areas in Relation to EDM ..................................... 6  
       1.2.3 Objectives of the EDM ....................................... 6  
       1.2.4 The Used Methods ............................................ 8  
       1.2.5 The Analyzed Data ........................................... 11  
       1.2.6 Process of Applying the EDM ................................ 12  
       1.2.7 Some Technological Tools Used in EDM .................... 13  
   1.3 Examples of EDM Applications in Computer-Based Learning Environments ........................................... 13  
       1.3.1 EDM Applications for Predicting and Evaluating Learning Performance ........................................... 15  
       1.3.2 EDM Applications for Analyzing Learners’ Behaviors ................................................................. 20  
       1.3.3 Discussion ............................................................ 23  
   1.4 Conclusions ................................................................... 24  
   References .......................................................................... 25

2 A Survey on Pre-Processing Educational Data ......................... 29  
   Cristóbal Romero, José Raúl Romero and Sebastián Ventura  
   2.1 Introduction ................................................................. 30  
   2.2 Types of Educational Environments .............................. 31  
       2.2.1 Learning Management Systems ................................ 32  
       2.2.2 Massive Open Online Courses ............................... 32  
       2.2.3 Intelligent Tutoring Systems ................................. 33  
       2.2.4 Adaptive and Intelligent Hypermedia Systems ........... 33  
       2.2.5 Test and Quiz Systems .......................................... 33  
       2.2.6 Other Types of Educational Systems ...................... 34
2.3 Types of Data

2.3.1 Relational Data

2.3.2 Transactional Data

2.3.3 Temporal, Sequence and Time Series Data

2.3.4 Text Data

2.3.5 Multimedia Data

2.3.6 World Wide Web Data

2.4 Pre-Processing Tasks

2.4.1 Data Gathering

2.4.2 Data Aggregation/Integration

2.4.3 Data Cleaning

2.4.4 User and Session Identification

2.4.5 Attribute/Variable Selection

2.4.6 Data Filtering

2.4.7 Data Transformation

2.5 Pre-Processing Tools

2.5.1 General Purpose Data Pre-Processing Tools

2.5.2 Specific Purpose Data Pre-Processing Tools

2.6 Conclusions

References

3 How Educational Data Mining Empowers State Policies to Reform Education: The Mexican Case Study

Alejandro Peña-Ayala and Leonor Cárdenas

3.1 Introduction

3.2 Domain Study

3.2.1 A Glance at Data Mining

3.2.2 Educational Data Mining in a Nutshell

3.3 Related Works

3.4 Context

3.4.1 The Mexican State

3.4.2 Educational Community

3.4.3 National Assessments

3.4.4 The Constitutional Reform in Education

3.4.5 Community Reaction

3.5 Source Data

3.5.1 EXCALE Databases

3.5.2 Source Data Students’ Opinions

3.5.3 Framework

3.5.4 Exploration Analysis

3.6 Educational Data Mining Approach

3.6.1 Essential Mining

3.6.2 Supplementary Mining

3.7 Discussion
Part II  Student Modeling

4  Modeling Student Performance in Higher Education Using Data Mining ........................................ 105
Huseyin Guruler and Ayhan Istanbullu
4.1  Introduction ............................................................... 106
4.2  Background ............................................................... 109
  4.2.1 The Decision Tree Classification Model .......... 110
  4.2.2 The Decision Tree Mechanism ....................... 111
4.3  System Overview, Software Interface and Architecture ........ 112
4.4  Case Study: Modeling Student Performance ................ 115
  4.4.1 Data Description ............................................... 115
  4.4.2 Data Preparation ............................................. 116
  4.4.3 Analyzer Model ............................................... 117
4.5  Discussion of Results ............................................. 117
4.6  Conclusions ........................................................... 121
References ................................................................. 122

5  Using Data Mining Techniques to Detect the Personality of Players in an Educational Game ....................... 125
Fazel Keshtkar, Candice Burkett, Haiying Li and Arthur C. Graesser
5.1  Introduction ............................................................... 126
5.2  Literature Review ....................................................... 128
  5.2.1 Personality in Computer-Based Learning Environments ................. 128
  5.2.2 Emotion Detection Using Leary’s Rose Frameboard ................. 128
  5.2.3 Automatic Detection of Personality ....................... 128
  5.2.4 Personality and Student Behavior ....................... 129
  5.2.5 The Relationship Between Personality Traits and Information Competency .... 129
  5.2.6 Personality Traits and Learning Style in Academic Performance ........ 130
  5.2.7 A Neural Network Model for Human Personality .......... 130
  5.2.8 Relationships Between Academic Motivation and Personality Among the Students .... 130
5.2.9 Relation Between Learning from Errors and Personality........................................... 131
5.2.10 Academic Achievement and Big Five Model ......................................................... 131
5.2.11 The Big Five Personality, Learning Styles, and Academic Achievement..................... 132
5.2.12 Using Personality and Cognitive Ability to Predict Academic Achievement.................. 132
5.3 Leary’s Interpersonal Frame Board .............................................................................. 132
5.3.1 Land Science Game ................................................................................................. 133
5.3.2 Participants and Data Set Construction ..................................................................... 134
5.4 Annotation Scheme .................................................................................................... 135
5.4.1 Human Annotation .................................................................................................. 135
5.5 Model ......................................................................................................................... 137
5.5.1 Lexicon Resources ................................................................................................. 138
5.5.2 Feature Extraction ................................................................................................. 138
5.5.3 The Linguistic Inquiry and Word Count Features .................................................... 139
5.5.4 Automated Approaches to Personality Classification .............................................. 140
5.5.5 Classification Method ............................................................................................ 141
5.6 Experience and Results ............................................................................................... 143
5.6.1 Classification Results ............................................................................................. 143
5.7 Discussion and Analysis ............................................................................................. 145
5.7.1 Personality Trait Tracking Analysis ......................................................................... 145
5.7.2 ANOVA Analysis ................................................................................................... 146
5.8 Conclusion and Future Research ................................................................................. 148
References ....................................................................................................................... 149

6 Students’ Performance Prediction Using Multi-Channel Decision Fusion ......................... 151
H. Moradi, S. Abbas Moradi and L. Kashani
6.1 Introduction ................................................................................................................. 152
6.2 Student Modeling ...................................................................................................... 153
6.3 Performance Prediction ............................................................................................. 156
6.3.1 Performance Prediction in ITS ............................................................................... 156
6.3.2 Data Mining Approaches for Prediction .................................................................. 157
6.4 Multi-Channel Decision Fusion Performance Prediction ............................................. 158
6.4.1 Determining the Performance Level in Assignment Categories ................................ 159
6.4.2 Determining Overall Performance Levels .............................................................. 161
6.4.3 Mapping from the Performance in Assignment Categories to Overall Performance .... 162
6.4.4 The Characteristics of Assignment Categories ....................................................... 163
6.5 Experimental Results and Discussion ......................................................................... 164
6.6 Conclusion and Future Work ..................................................................................... 172
References ....................................................................................................................... 173
7 Predicting Student Performance from Combined Data Sources

Annika Wolff, Zdenek Zdrahal, Drahomira Herrmannova and Petr Knoth

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>176</td>
</tr>
<tr>
<td>7.2</td>
<td>Defining the Problem</td>
<td>177</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Problem Specification 1</td>
<td>178</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Problem Specification 2</td>
<td>179</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Problem Specification 3</td>
<td>179</td>
</tr>
<tr>
<td>7.2.4</td>
<td>Problem Specification 4</td>
<td>179</td>
</tr>
<tr>
<td>7.3</td>
<td>Sources of Student Data</td>
<td>179</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Student Activity Data from the Virtual Learning Environment</td>
<td>180</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Demographic Data</td>
<td>180</td>
</tr>
<tr>
<td>7.3.3</td>
<td>Past Study</td>
<td>181</td>
</tr>
<tr>
<td>7.3.4</td>
<td>Assessment Data</td>
<td>181</td>
</tr>
<tr>
<td>7.4</td>
<td>Feature Selection and Data-Filtering</td>
<td>181</td>
</tr>
<tr>
<td>7.5</td>
<td>Classifiers for Predicting Student Outcome</td>
<td>184</td>
</tr>
<tr>
<td>7.5.1</td>
<td>Support Vector Machines and Decision Trees</td>
<td>184</td>
</tr>
<tr>
<td>7.5.2</td>
<td>General Unary Hypotheses Automaton</td>
<td>188</td>
</tr>
<tr>
<td>7.5.3</td>
<td>Bayesian Networks and Regression</td>
<td>189</td>
</tr>
<tr>
<td>7.6</td>
<td>Evaluation Framework</td>
<td>192</td>
</tr>
<tr>
<td>7.7</td>
<td>Real-Time Prediction</td>
<td>196</td>
</tr>
<tr>
<td>7.8</td>
<td>Revisiting the Problem Specification in Light of Results</td>
<td>196</td>
</tr>
<tr>
<td>7.8.1</td>
<td>Problem Specification 4 (Revised)</td>
<td>197</td>
</tr>
<tr>
<td>7.8.2</td>
<td>Problem Specification 5</td>
<td>198</td>
</tr>
<tr>
<td>7.8.3</td>
<td>Problem Specification 6</td>
<td>199</td>
</tr>
<tr>
<td>7.9</td>
<td>Developing and Testing Models on Open University Data (A Case Study)</td>
<td>199</td>
</tr>
<tr>
<td>7.10</td>
<td>Beyond OU: Applying Models on Alternative Data Sources</td>
<td>200</td>
</tr>
<tr>
<td>7.11</td>
<td>Conclusions</td>
<td>201</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>201</td>
</tr>
</tbody>
</table>

8 Predicting Learner Answers Correctness Through Eye Movements with Random Forest

Alper Bayazit, Petek Askar and Erdal Cosgun

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>Introduction</td>
<td>204</td>
</tr>
<tr>
<td>8.2</td>
<td>Background</td>
<td>205</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Related Work</td>
<td>205</td>
</tr>
<tr>
<td>8.2.2</td>
<td>Cognitive Processes</td>
<td>206</td>
</tr>
<tr>
<td>8.2.3</td>
<td>Eye Movement Data</td>
<td>207</td>
</tr>
<tr>
<td>8.2.4</td>
<td>Random Forest</td>
<td>209</td>
</tr>
<tr>
<td>8.3</td>
<td>Method</td>
<td>210</td>
</tr>
</tbody>
</table>
### Part III  Assessment

9  **Mining Domain Knowledge for Coherence Assessment of Students Proposal Drafts**  
Samuel González López and Aurelio López-López

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1 Introduction</td>
<td>230</td>
</tr>
<tr>
<td>9.2 Background</td>
<td>231</td>
</tr>
<tr>
<td>9.2.1 Global Coherence</td>
<td>231</td>
</tr>
<tr>
<td>9.2.2 Latent Semantic Analysis</td>
<td>232</td>
</tr>
<tr>
<td>9.2.3 Related Work</td>
<td>232</td>
</tr>
<tr>
<td>9.3 Analyzer Model of Global Coherence</td>
<td>233</td>
</tr>
<tr>
<td>9.4 Data Description (Corpus)</td>
<td>236</td>
</tr>
<tr>
<td>9.5 Experiments</td>
<td>236</td>
</tr>
<tr>
<td>9.5.1 Experimental Design</td>
<td>237</td>
</tr>
<tr>
<td>9.5.2 Agreement Evaluation</td>
<td>238</td>
</tr>
<tr>
<td>9.5.3 Across Section Exploration</td>
<td>247</td>
</tr>
<tr>
<td>9.6 Analysis and Discussion of Results</td>
<td>248</td>
</tr>
<tr>
<td>9.6.1 Across Section Exploration</td>
<td>248</td>
</tr>
<tr>
<td>9.7 System Overview</td>
<td>249</td>
</tr>
<tr>
<td>9.7.1 Intelligent Tutoring System</td>
<td>249</td>
</tr>
<tr>
<td>9.7.2 Web Interface</td>
<td>251</td>
</tr>
<tr>
<td>9.8 Conclusions</td>
<td>252</td>
</tr>
</tbody>
</table>

References | 252 |

10  **Adaptive Testing in Programming Courses Based on Educational Data Mining Techniques**  
Vladimir Ivančević, Marko Knežević, Bojan Pušić and Ivan Luković

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1 Introduction</td>
<td>259</td>
</tr>
<tr>
<td>10.2 Related Work</td>
<td>261</td>
</tr>
<tr>
<td>10.3 Background</td>
<td>263</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>10.3.1</td>
<td>Environment</td>
</tr>
<tr>
<td>10.3.2</td>
<td>Data Set</td>
</tr>
<tr>
<td>10.4</td>
<td>Modeling Programming Knowledge</td>
</tr>
<tr>
<td>10.4.1</td>
<td>Programming Knowledge Overview</td>
</tr>
<tr>
<td>10.4.2</td>
<td>Modeling Programming Competencies</td>
</tr>
<tr>
<td>10.4.3</td>
<td>Modeling Programming Concepts of the C Language</td>
</tr>
<tr>
<td>10.5</td>
<td>Estimating Test Difficulty</td>
</tr>
<tr>
<td>10.5.1</td>
<td>Estimating Test Item Difficulty</td>
</tr>
<tr>
<td>10.5.2</td>
<td>Estimating Student Capacity</td>
</tr>
<tr>
<td>10.6</td>
<td>Test Generation Algorithm</td>
</tr>
<tr>
<td>10.7</td>
<td>Application and Results</td>
</tr>
<tr>
<td>10.8</td>
<td>Conclusion and Future Work</td>
</tr>
<tr>
<td></td>
<td>References</td>
</tr>
<tr>
<td>11</td>
<td>Plan Recognition and Visualization in Exploratory Learning Environments</td>
</tr>
<tr>
<td></td>
<td>Ofra Amir, Kobi Gal, David Yaron, Michael Karabinos, and Robert Belford</td>
</tr>
<tr>
<td>11.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>11.2</td>
<td>Related Work</td>
</tr>
<tr>
<td>11.2.1</td>
<td>Plan Recognition</td>
</tr>
<tr>
<td>11.2.2</td>
<td>Assessment of Students’ Activities</td>
</tr>
<tr>
<td>11.3</td>
<td>The Virtual Labs Domain</td>
</tr>
<tr>
<td>11.4</td>
<td>Plan Recognition in Virtual Laboratories</td>
</tr>
<tr>
<td>11.4.1</td>
<td>Actions, Recipes, and Plans</td>
</tr>
<tr>
<td>11.4.2</td>
<td>The Plan Recognition Algorithm</td>
</tr>
<tr>
<td>11.4.3</td>
<td>Empirical Methodology</td>
</tr>
<tr>
<td>11.4.4</td>
<td>Complete Algorithms</td>
</tr>
<tr>
<td>11.5</td>
<td>Visualizing Students’ Activities</td>
</tr>
<tr>
<td>11.5.1</td>
<td>Visualization Methods</td>
</tr>
<tr>
<td>11.5.2</td>
<td>Empirical Methodology</td>
</tr>
<tr>
<td>11.5.3</td>
<td>Results</td>
</tr>
<tr>
<td>11.5.4</td>
<td>Discussion</td>
</tr>
<tr>
<td>11.6</td>
<td>Conclusion and Future Work</td>
</tr>
<tr>
<td>11.7</td>
<td>Experimental Problems</td>
</tr>
<tr>
<td>11.8</td>
<td>The Recipe Library for the Dilution Problem</td>
</tr>
<tr>
<td>11.8.1</td>
<td>Dilution Problem Recipes</td>
</tr>
<tr>
<td>11.8.2</td>
<td>Recipes Explanation</td>
</tr>
<tr>
<td>11.9</td>
<td>User Study Questionnaire</td>
</tr>
<tr>
<td></td>
<td>References</td>
</tr>
</tbody>
</table>
12 Finding Dependency of Test Items from Students’ Response Data ...................................... 329
  Xiaoxun Sun
  12.1 Introduction ........................................... 330
  12.2 Related Work ......................................... 330
  12.3 Mutual Independency Measure ....................... 331
    12.3.1 Preliminaries ..................................... 331
    12.3.2 Mutual Information Measure ..................... 332
    12.3.3 Finding the Best Dependency Tree ................. 333
    12.3.4 An Example ...................................... 335
    12.3.5 Extensions ...................................... 337
  12.4 Proof-of-Concept Experiments ....................... 337
    12.4.1 Data ........................................... 337
    12.4.2 Results on Synthetic Data Sets .................. 338
    12.4.3 Results on Real Data ............................ 340
  12.5 Conclusions and Future Work ....................... 341
  References .............................................. 341

Part IV Trends

13 Mining Texts, Learner Productions and Strategies with ReaderBench .................................... 345
  Mihai Dascalu, Philippe Dessus, Maryse Bianco, Stefan Trausan-Matu and Aurélie Nardy
  13.1 Introduction ......................................... 346
  13.2 Data and Text Mining for Educational Applications .... 347
    13.2.1 Predicting Learner Comprehension ................. 348
  13.3 Textual Complexity Assessment for Comprehension Prediction .................................. 348
    13.3.1 The Impact of Reading Strategies Extracted from Self-Explanations for Comprehension Assessment .... 350
  13.4 Cohesion-Based Discourse Analysis: Building the Cohesion Graph ............................... 352
  13.5 Topics Extraction ..................................... 355
  13.6 Cohesion-Based Scoring Mechanism of the Analysis Elements ................................. 359
  13.7 Identification Heuristics for Reading Strategies ................................................ 360
  13.8 Multi-Dimensional Model for Assessing Textual Complexity ........................................ 363
  13.9 Results .............................................. 364
  13.10 A Comparison of ReaderBench with Previous Work ................................................ 370
  13.11 Conclusions ......................................... 372
  References .............................................. 373
Maximizing the Value of Student Ratings Through Data Mining

Kathryn Gates, Dawn Wilkins, Sumali Conlon, Susan Mossing and Maurice Eftink

14.1 Introduction ........................................ 380
14.2 Description of the Data Set .......................... 382
  14.2.1 The Process for Collecting Evaluations and Presenting Results ................. 382
  14.2.2 Details About the Data Set ......................... 383
  14.2.3 Questions and Variables of Interest .................. 383
  14.2.4 Selected Results from the Statistical Analysis .............. 384
14.3 The Methodology ...................................... 385
  14.3.1 A High-Level View of the Process .................... 385
  14.3.2 Corpus Word Analysis ............................... 386
  14.3.3 Category Selection .................................. 387
  14.3.4 The Domain-Specific Lexicon ......................... 388
  14.3.5 The Assessment Process .............................. 389
  14.3.6 Refining the Lexicon ................................ 392
  14.3.7 The Algorithm ...................................... 393
14.4 Assessment Results ................................. 397
  14.4.1 Qualitative Validity Assessment of Category Scores by Teaching and Learning Specialists .............. 397
  14.4.2 Quantitative Assessment Through the Comparison of Summary Scores with Overall Instructor Performance Ratings ......... 398
  14.4.3 Quantitative Assessment Through the Comparison of Category and Summary Scores for Teaching Award Winners with All Instructors .............. 403
14.5 Applications of the Methodology ............... 404
  14.5.1 Evaluation of Instruction at the University of Mississippi ...................... 404
  14.5.2 Other Educational Applications ...................... 405
14.6 Future Work ........................................... 407
References .............................................. 409

Data Mining and Social Network Analysis in the Educational Field: An Application for Non-Expert Users

Diego García-Saiz, Camilo Palazuelos and Marta Zorrilla

15.1 Introduction ........................................ 412
15.2 Background and Related Work ..................... 414
  15.2.1 Social Network Analysis ............................ 415
  15.2.2 Classification Applied to the Educational Context: Students’ Performance and Dropout .................. 417
  15.2.3 Data Mining Tools for Non-Expert Users .............. 418
15.3 E-Learning Web Miner ........................................... 419
15.3.1 Description of E-Learning Web Miner .................... 419
15.3.2 General View of the E-Learning Web Miner
Architecture ...................................................... 420
15.3.3 New Services Provided .................................... 422
15.3.4 Mode of Working .......................................... 423
15.4 Case Study ....................................................... 426
15.4.1 Courses ...................................................... 426
15.4.2 Social Network Analysis in E-Learning Courses ....... 428
15.4.3 Prediction of Students’ Performance
and Dropouts .................................................... 432
15.5 Conclusions ....................................................... 435
References .......................................................... 436

16 Collaborative Learning of Students in Online Discussion
Forums: A Social Network Analysis Perspective ................ 441
Reihaneh Rabbany, Samira Elatia, Mansoureh Takaffoli
and Osmar R. Zaïane
16.1 Introduction .................................................... 442
16.2 Background and Related Works ............................. 444
16.2.1 On Collaborative Learning and E-Learning:
An Educational Perspective .................................... 444
16.2.2 Social Networks: A Data Mining Perspective ........... 445
16.2.3 Social Network Analysis of Online Educational Forums: Related Works .................. 448
16.3 Network Analysis in E-Learning .............................. 450
16.3.1 Students Interaction Network ............................. 450
16.3.2 Term Co-Occurrence Network ........................... 453
16.4 Case Studies ..................................................... 456
16.4.1 Extracting Networks ...................................... 458
16.4.2 Interpreting Students Interaction Network .............. 459
16.4.3 Interpreting Term Co-Occurrence Network ............. 462
16.4.4 Objective Evaluation ..................................... 463
16.5 Conclusions ..................................................... 464
References .......................................................... 464

Author Index .......................................................... 467