# CONTENTS

## NOTES ON CONTRIBUTORS

| xi |

## INTRODUCTION: EDUCATION AT COMPUTATIONAL CROSSROADS

Samira ElAtia, Donald Ipperciel, and Osmar R. Zaïane

xxiii

## PART I

### AT THE INTERSECTION OF TWO FIELDS: EDM

| 1 |

## CHAPTER 1

**EDUCATIONAL PROCESS MINING: A TUTORIAL AND CASE STUDY USING MOODLE DATA SETS**

3

Cristóbal Romero, Rebeca Cerezo, Alejandro Bogarín, and Miguel Sánchez-Santillán

1.1 Background 5

1.2 Data Description and Preparation 7

1.2.1 Preprocessing Log Data 7

1.2.2 Clustering Approach for Grouping Log Data 11

1.3 Working with ProM 16

1.3.1 Discovered Models 19

1.3.2 Analysis of the Models’ Performance 23

1.4 Conclusion 26

Acknowledgments 27

References 27

## CHAPTER 2

**ON BIG DATA AND TEXT MINING IN THE HUMANITIES**

29

Geoffrey Rockwell and Bettina Berendt

2.1 Busa and the Digital Text 30

2.2 Thesaurus Linguae Graecae and the Ibycus Computer as Infrastructure 32

2.2.1 Complete Data Sets 33

2.3 Cooking with Statistics 35

2.4 Conclusions 37

References 38

## CHAPTER 3

**FINDING PREDICTORS IN HIGHER EDUCATION**

41

David Eubanks, William Evers Jr., and Nancy Smith

3.1 Contrasting Traditional and Computational Methods 42

3.2 Predictors and Data Exploration 45

3.3 Data Mining Application: An Example 50
CHAPTER 4  
EDUCATIONAL DATA MINING: A MOOC EXPERIENCE  
Ryan S. Baker, Yuan Wang, Luc Paquette, Vincent Alevon, Octav Popescu, Jonathan Sewall, Carolyn Rosé, Gaurav Singh Tomar, Oliver Ferschke, Jing Zhang, Michael J. Cennamo, Stephanie Ogden, Therese Condit, José Diaz, Scott Crossley, Danielle S. McNamara, Denise K. Comer, Collin F. Lynch, Rebecca Brown, Tiffany Barnes, and Yoav Bergner

4.1 Big Data in Education: The Course  55
4.1.1 Iteration 1: Coursera  55
4.1.2 Iteration 2: edX  56
4.2 Cognitive Tutor Authoring Tools  57
4.3 Bazaar  58
4.4 Walkthrough  58
4.4.1 Course Content  58
4.4.2 Research on BDEMOOC  61
4.5 Conclusion  65
Acknowledgments  65
References  65

CHAPTER 5  
DATA MINING AND ACTION RESEARCH  
Ellina Chernobilsky, Edith Ries, and Joanne Jasmine

5.1 Process  69
5.2 Design Methodology  71
5.3 Analysis and Interpretation of Data  72
5.3.1 Quantitative Data Analysis and Interpretation  73
5.3.2 Qualitative Data Analysis and Interpretation  74
5.4 Challenges  75
5.5 Ethics  76
5.6 Role of Administration in the Data Collection Process  76
5.7 Conclusion  77
References  77

PART II  
PEDAGOGICAL APPLICATIONS OF EDM  

CHAPTER 6  
DESIGN OF AN ADAPTIVE LEARNING SYSTEM AND EDUCATIONAL DATA MINING  
Zhiyong Liu and Nick Cercone

6.1 Dimensionalities of the User Model in ALS  83
6.2 Collecting Data for ALS  85
6.3 Data Mining in ALS  86
6.3.1 Data Mining for User Modeling  87
6.3.2 Data Mining for Knowledge Discovery  88
6.4 ALS Model and Function Analyzing  90
CHAPTER 7  THE “GEOMETRY” OF NAÏVE BAYES: TEACHING PROBABILITIES BY “DRAWING” THEM

Giorgio Maria Di Nunzio

7.1 Introduction 99
7.1.1 Main Contribution 100
7.1.2 Related Works 101
7.2 The Geometry of NB Classification 102
7.2.1 Mathematical Notation 102
7.2.2 Bayesian Decision Theory 103
7.3 Two-Dimensional Probabilities 105
7.3.1 Working with Likelihoods and Priors Only 107
7.3.2 De-normalizing Probabilities 108
7.3.3 NB Approach 109
7.3.4 Bernoulli Naïve Bayes 110
7.4 A New Decision Line: Far from the Origin 111
7.4.1 De-normalization Makes (Some) Problems Linearly Separable 112
7.5 Likelihood Spaces, When Logarithms make a Difference (or a SUM) 114
7.5.1 De-normalization Makes (Some) Problems Linearly Separable 115
7.5.2 A New Decision in Likelihood Spaces 116
7.5.3 A Real Case Scenario: Text Categorization 117
7.6 Final Remarks 118
References 119

CHAPTER 8  EXAMINING THE LEARNING NETWORKS OF A MOOC

Meaghan Brugha and Jean-Paul Restoule

8.1 Review of Literature 122
8.2 Course Context 124
8.3 Results and Discussion 125
8.4 Recommendations for Future Research 133
8.5 Conclusions 134
References 135

CHAPTER 9  EXPLORING THE USEFULNESS OF ADAPTIVE ELEARNING LABORATORY ENVIRONMENTS IN TEACHING MEDICAL SCIENCE

Thuan Thai and Patsie Polly

9.1 Introduction 139
9.2 Software for Learning and Teaching 141
9.2.1 Reflective Practice: ePortfolio 141
9.2.2 Online Quizzes 143
9.2.3 Online Practical Lessons 144
CHAPTER 10  
INVESTIGATING CO-OCCURRENCE PATTERNS OF LEARNERS’ GRAMMATICAL ERRORS ACROSS PROFICIENCY LEVELS AND ESSAY TOPICS BASED ON ASSOCIATION ANALYSIS  
Yutaka Ishii 
10.1 Introduction  157  
10.1.1 The Relationship between Data Mining and Educational Research  157  
10.1.2 English Writing Instruction in the Japanese Context  158  
10.2 Literature Review  159  
10.3 Method  160  
10.3.1 Konan-JIEM Learner Corpus  160  
10.3.2 Association Analysis  162  
10.4 Experiment 1  162  
10.5 Experiment 2  163  
10.6 Discussion and Conclusion  164  
Appendix A: Example of Learner’s Essay (University Life)  164  
Appendix B: Support Values of all Topics  165  
Appendix C: Support Values of Advanced, Intermediate, and Beginner Levels of Learners  168  
References  169  

PART III  
EDM AND EDUCATIONAL RESEARCH  
CHAPTER 11  
MINING LEARNING SEQUENCES IN MOOCs: DOES COURSE DESIGN CONSTRAIN STUDENTS’ BEHAVIORS OR DO STUDENTS SHAPE THEIR OWN LEARNING?  
Lorenzo Vigentini, Simon McIntyre, Negin Mirriahi, and Dennis Alonzo 
11.1 Introduction  175  
11.1.1 Perceptions and Challenges of MOOC Design  176  
11.1.2 What Do We Know About Participants’ Navigation: Choice and Control  177  
11.2 Data Mining in MOOCs: Related Work  178  
11.2.1 Setting the Hypotheses  179  
11.3 The Design and Intent of the LTTO MOOC  180  
11.3.1 Course Grading and Certification  183  
11.3.2 Delivering the Course  183  
11.3.3 Operationalize Engagement, Personal Success, and Course Success in LTTO  184  
11.4 Data Analysis  184
CONTENTS ix

11.4.1 Approaches to Process the Data Sources 185
11.4.2 LTTO in Numbers 186
11.4.3 Characterizing Patterns of Completion and Achievement 186
11.4.4 Redefining Participation and Engagement 189
11.5 Mining Behaviors and Intents 191
11.5.1 Participants’ Intent and Behaviors: A Classification Model 191
11.5.2 Natural Clustering Based on Behaviors 194
11.5.3 Stated Intents and Behaviors: Are They Related? 198
11.6 Closing the Loop: Informing Pedagogy and Course Enhancement 198
11.6.1 Conclusions, Lessons Learnt, and Future Directions 200

References 201

CHAPTER 12 UNDERSTANDING COMMUNICATION PATTERNS IN MOOCs: COMBINING DATA MINING AND QUALITATIVE METHODS 207
Rebecca Eynon, Isis Hjorth, Taha Yasseri, and Nabeel Gillani

12.1 Introduction 207
12.2 Methodological Approaches to Understanding Communication Patterns in MOOCs 209
12.3 Description 210
12.3.1 Structural Connections 211
12.4 Examining Dialogue 213
12.5 Interpretative Models 214
12.6 Understanding Experience 215
12.7 Experimentation 216
12.8 Future Research 217

References 218

CHAPTER 13 AN EXAMPLE OF DATA MINING: EXPLORING THE RELATIONSHIP BETWEEN APPLICANT ATTRIBUTES AND ACADEMIC MEASURES OF SUCCESS IN A PHARMACY PROGRAM 223
Dion Brocks and Ken Cor

13.1 Introduction 223
13.2 Methods 225
13.3 Results 228
13.4 Discussion 230
13.4.1 Prerequisite Predictors 230
13.4.2 Demographic Predictors 232
13.5 Conclusion 234

Appendix A 234

References 236

CHAPTER 14 A NEW WAY OF SEEING: USING A DATA MINING APPROACH TO UNDERSTAND CHILDREN’S VIEWS OF DIVERSITY AND “DIFFERENCE” IN PICTURE BOOKS 237
Robin A. Moeller and Hsin-liang Chen

14.1 Introduction 237
14.2 Study 1: Using Data Mining to Better Understand Perceptions of Race 238
14.2.1 Background 238