

Contents

Part I Introduction

| | | |
|----------|---|----|
| 1 | Data Science in Action | 3 |
| 1.1 | Internet of Events | 3 |
| 1.2 | Data Scientist | 10 |
| 1.3 | Bridging the Gap Between Process Science and Data Science | 15 |
| 1.4 | Outlook | 20 |
| 2 | Process Mining: The Missing Link | 25 |
| 2.1 | Limitations of Modeling | 25 |
| 2.2 | Process Mining | 30 |
| 2.3 | Analyzing an Example Log | 35 |
| 2.4 | Play-In, Play-Out, and Replay | 41 |
| 2.5 | Positioning Process Mining | 44 |
| 2.5.1 | How Process Mining Compares to BPM | 44 |
| 2.5.2 | How Process Mining Compares to Data Mining | 46 |
| 2.5.3 | How Process Mining Compares to Lean Six Sigma | 46 |
| 2.5.4 | How Process Mining Compares to BPR | 49 |
| 2.5.5 | How Process Mining Compares to Business Intelligence | 49 |
| 2.5.6 | How Process Mining Compares to CEP | 50 |
| 2.5.7 | How Process Mining Compares to GRC | 50 |
| 2.5.8 | How Process Mining Compares to ABPD, BPI, WM, | 51 |
| 2.5.9 | How Process Mining Compares to Big Data | 52 |

Part II Preliminaries

| | | |
|----------|--|----|
| 3 | Process Modeling and Analysis | 55 |
| 3.1 | The Art of Modeling | 55 |
| 3.2 | Process Models | 57 |
| 3.2.1 | Transition Systems | 58 |
| 3.2.2 | Petri Nets | 59 |
| 3.2.3 | Workflow Nets | 65 |
| 3.2.4 | YAWL | 66 |

- 3.2.5 Business Process Modeling Notation (BPMN) 68
- 3.2.6 Event-Driven Process Chains (EPCs) 70
- 3.2.7 Causal Nets 72
- 3.2.8 Process Trees 78
- 3.3 Model-Based Process Analysis 83
 - 3.3.1 Verification 83
 - 3.3.2 Performance Analysis 85
 - 3.3.3 Limitations of Model-Based Analysis 88
- 4 Data Mining 89**
 - 4.1 Classification of Data Mining Techniques 89
 - 4.1.1 Data Sets: Instances and Variables 90
 - 4.1.2 Supervised Learning: Classification and Regression 92
 - 4.1.3 Unsupervised Learning: Clustering and Pattern Discovery 94
 - 4.2 Decision Tree Learning 94
 - 4.3 *k*-Means Clustering 100
 - 4.4 Association Rule Learning 104
 - 4.5 Sequence and Episode Mining 107
 - 4.5.1 Sequence Mining 107
 - 4.5.2 Episode Mining 109
 - 4.5.3 Other Approaches 111
 - 4.6 Quality of Resulting Models 112
 - 4.6.1 Measuring the Performance of a Classifier 113
 - 4.6.2 Cross-Validation 115
 - 4.6.3 Occam’s Razor 118
- Part III From Event Logs to Process Models**
- 5 Getting the Data 125**
 - 5.1 Data Sources 125
 - 5.2 Event Logs 128
 - 5.3 XES 138
 - 5.4 Data Quality 144
 - 5.4.1 Conceptualizing Event Logs 145
 - 5.4.2 Classification of Data Quality Issues 148
 - 5.4.3 Guidelines for Logging 151
 - 5.5 Flattening Reality into Event Logs 153
- 6 Process Discovery: An Introduction 163**
 - 6.1 Problem Statement 163
 - 6.2 A Simple Algorithm for Process Discovery 167
 - 6.2.1 Basic Idea 167
 - 6.2.2 Algorithm 171
 - 6.2.3 Limitations of the α -Algorithm 174
 - 6.2.4 Taking the Transactional Life-Cycle into Account 177
 - 6.3 Rediscovering Process Models 178
 - 6.4 Challenges 182

- 6.4.1 Representational Bias 183
- 6.4.2 Noise and Incompleteness 185
- 6.4.3 Four Competing Quality Criteria 188
- 6.4.4 Taking the Right 2-D Slice of a 3-D Reality 192
- 7 Advanced Process Discovery Techniques 195**
 - 7.1 Overview 195
 - 7.1.1 Characteristic 1: Representational Bias 197
 - 7.1.2 Characteristic 2: Ability to Deal With Noise 198
 - 7.1.3 Characteristic 3: Completeness Notion Assumed 199
 - 7.1.4 Characteristic 4: Approach Used 199
 - 7.2 Heuristic Mining 201
 - 7.2.1 Causal Nets Revisited 201
 - 7.2.2 Learning the Dependency Graph 202
 - 7.2.3 Learning Splits and Joins 205
 - 7.3 Genetic Process Mining 207
 - 7.4 Region-Based Mining 212
 - 7.4.1 Learning Transition Systems 212
 - 7.4.2 Process Discovery Using State-Based Regions 216
 - 7.4.3 Process Discovery Using Language-Based Regions 218
 - 7.5 Inductive Mining 222
 - 7.5.1 Inductive Miner Based on Event Log Splitting 222
 - 7.5.2 Characteristics of the Inductive Miner 229
 - 7.5.3 Extensions and Scalability 233
 - 7.6 Historical Perspective 236
- Part IV Beyond Process Discovery**
- 8 Conformance Checking 243**
 - 8.1 Business Alignment and Auditing 243
 - 8.2 Token Replay 246
 - 8.3 Alignments 256
 - 8.4 Comparing Footprints 263
 - 8.5 Other Applications of Conformance Checking 268
 - 8.5.1 Repairing Models 268
 - 8.5.2 Evaluating Process Discovery Algorithms 269
 - 8.5.3 Connecting Event Log and Process Model 272
- 9 Mining Additional Perspectives 275**
 - 9.1 Perspectives 275
 - 9.2 Attributes: A Helicopter View 277
 - 9.3 Organizational Mining 281
 - 9.3.1 Social Network Analysis 282
 - 9.3.2 Discovering Organizational Structures 287
 - 9.3.3 Analyzing Resource Behavior 288
 - 9.4 Time and Probabilities 290
 - 9.5 Decision Mining 294
 - 9.6 Bringing It All Together 297

- 10 Operational Support 301**
 - 10.1 Refined Process Mining Framework 301
 - 10.1.1 Cartography 303
 - 10.1.2 Auditing 304
 - 10.1.3 Navigation 305
 - 10.2 Online Process Mining 305
 - 10.3 Detect 307
 - 10.4 Predict 311
 - 10.5 Recommend 316
 - 10.6 Processes Are Not in Steady State! 318
 - 10.6.1 Daily, Weekly and Seasonal Patterns in Processes 318
 - 10.6.2 Contextual Factors 318
 - 10.6.3 Concept Drift in Processes 320
 - 10.7 Process Mining Spectrum 321
- Part V Putting Process Mining to Work**
- 11 Process Mining Software 325**
 - 11.1 Process Mining Not Included! 325
 - 11.2 Different Types of Process Mining Tools 327
 - 11.3 ProM: An Open-Source Process Mining Platform 331
 - 11.3.1 Historical Context 331
 - 11.3.2 Example ProM Plug-Ins 333
 - 11.3.3 Other Non-commercial Tools 337
 - 11.4 Commercial Software 339
 - 11.4.1 Available Products 339
 - 11.4.2 Strengths and Weaknesses 345
 - 11.5 Outlook 352
- 12 Process Mining in the Large 353**
 - 12.1 Big Event Data 353
 - 12.1.1 $N = \text{All}$ 354
 - 12.1.2 Hardware and Software Developments 356
 - 12.1.3 Characterizing Event Logs 364
 - 12.2 Case-Based Decomposition 368
 - 12.2.1 Conformance Checking Using Case-Based Decomposition 369
 - 12.2.2 Process Discovery Using Case-Based Decomposition 370
 - 12.3 Activity-Based Decomposition 373
 - 12.3.1 Conformance Checking Using Activity-Based Decomposition 374
 - 12.3.2 Process Discovery Using Activity-Based Decomposition 376
 - 12.4 Process Cubes 378
 - 12.5 Streaming Process Mining 381
 - 12.6 Beyond the Hype 384

- 13 Analyzing “Lasagna Processes”** 387
 - 13.1 Characterization of “Lasagna Processes” 387
 - 13.2 Use Cases 391
 - 13.3 Approach 392
 - 13.3.1 Stage 0: Plan and Justify 393
 - 13.3.2 Stage 1: Extract 395
 - 13.3.3 Stage 2: Create Control-Flow Model and Connect Event Log 395
 - 13.3.4 Stage 3: Create Integrated Process Model 396
 - 13.3.5 Stage 4: Operational Support 396
 - 13.4 Applications 397
 - 13.4.1 Process Mining Opportunities per Functional Area 397
 - 13.4.2 Process Mining Opportunities per Sector 398
 - 13.4.3 Two Lasagna Processes 402
- 14 Analyzing “Spaghetti Processes”** 411
 - 14.1 Characterization of “Spaghetti Processes” 411
 - 14.2 Approach 415
 - 14.3 Applications 418
 - 14.3.1 Process Mining Opportunities for Spaghetti Processes 418
 - 14.3.2 Examples of Spaghetti Processes 420
- Part VI Reflection**
- 15 Cartography and Navigation** 431
 - 15.1 Business Process Maps 431
 - 15.1.1 Map Quality 432
 - 15.1.2 Aggregation and Abstraction 432
 - 15.1.3 Seamless Zoom 434
 - 15.1.4 Size, Color, and Layout 438
 - 15.1.5 Customization 440
 - 15.2 Process Mining: TomTom for Business Processes? 441
 - 15.2.1 Projecting Dynamic Information on Business Process Maps 441
 - 15.2.2 Arrival Time Prediction 444
 - 15.2.3 Guidance Rather than Control 444
- 16 Epilogue** 447
 - 16.1 Process Mining as a Bridge Between Data Mining and Business Process Management 447
 - 16.2 Challenges 449
 - 16.3 Start Today! 451
- References** 453
- Index** 463



<http://www.springer.com/978-3-662-49850-7>

Process Mining

Data Science in Action

van der Aalst, W.

2016, XIX, 467 p. 250 illus., 13 illus. in color., Hardcover

ISBN: 978-3-662-49850-7