

Wil M.R. van der Aalst

# Process Mining

Discovery, Conformance and  
Enhancement of Business Processes

Q| UNIVERSITÄT  
<sup>m</sup> LIECHTENSTEIN  
Bibliothek

 Springer

# Contents

<b>1</b>	<b>Introduction</b>	<b>I</b>
1.1	Data Explosion	I
1.2	Limitations of Modeling	3
1.3	Process Mining	7
1.4	Analyzing an Example Log	11
1.5	Play-in, Play-out, and Replay	18
1.6	Trends	21
1.7	Outlook	23
<b>Part I Preliminaries</b>		
<b>2</b>	<b>Process Modeling and Analysis</b>	<b>29</b>
2.1	The Art of Modeling	29
2.2	Process Models	31
2.2.1	Transition Systems	31
2.2.2	Petri Nets	33
2.2.3	Workflow Nets	38
2.2.4	YAWL	40
2.2.5	Business Process Modeling Notation (BPMN)	42
2.2.6	Event-Driven Process Chains (EPCs)	44
2.2.7	Causal Nets	46
2.3	Model-Based Process Analysis	52
2.3.1	Verification	52
2.3.2	Performance Analysis	55
2.3.3	Limitations of Model-Based Analysis	57
<b>3</b>	<b>Data Mining</b>	<b>59</b>
3.1	Classification of Data Mining Techniques	59
3.1.1	Data Sets: Instances and Variables	60
3.1.2	Supervised Learning: Classification and Regression	62
3.1.3	Unsupervised Learning: Clustering and Pattern Discovery	64
3.2	Decision Tree Learning	64

3.3	£-Means Clustering . . . . .	70
3.4	Association Rule Learning . . . . .	74
3.5	Sequence and Episode Mining . . . . .	77
3.5.1	Sequence Mining . . . . .	77
3.5.2	Episode Mining . . . . .	78
3.5.3	Other Approaches . . . . .	81
3.6	Quality of Resulting Models . . . . .	82
3.6.1	Measuring the Performance of a Classifier . . . . .	83
3.6.2	Cross-Validation . . . . .	85
3.6.3	Occam's Razor . . . . .	88
<b>Part II From Event Logs to Process Models</b>		
<b>4</b>	<b>Getting the Data . . . . .</b>	<b>95</b>
4.1	Data Sources . . . . .	95
4.2	Event Logs . . . . .	98
4.3	XES . . . . .	107
4.4	Flattening Reality into Event Logs . . . . .	114
<b>5</b>	<b>Process Discovery: An Introduction . . . . .</b>	<b>125</b>
5.1	Problem Statement . . . . .	125
5.2	A Simple Algorithm for Process Discovery . . . . .	129
5.2.1	Basic Idea . . . . .	129
5.2.2	Algorithm . . . . .	133
5.2.3	Limitations of the or-Algorithm . . . . .	136
5.2.4	Taking the Transactional Life-Cycle into Account . . . . .	139
5.3	Rediscovering Process Models . . . . .	140
5.4	Challenges . . . . .	144
5.4.1	Representational Bias . . . . .	145
5.4.2	Noise and Incompleteness . . . . .	147
5.4.3	Four Competing Quality Criteria . . . . .	150
5.4.4	Taking the Right 2-D Slice of a 3-D Reality . . . . .	153
<b>6</b>	<b>Advanced Process Discovery Techniques . . . . .</b>	<b>157</b>
6.1	Overview . . . . .	157
6.1.1	Characteristic 1: Representational Bias . . . . .	159
6.1.2	Characteristic 2: Ability to Deal with Noise . . . . .	160
6.1.3	Characteristic 3: Completeness Notion Assumed . . . . .	161
6.1.4	Characteristic 4: Approach Used . . . . .	161
6.2	Heuristic Mining . . . . .	163
6.2.1	Causal Nets Revisited . . . . .	163
6.2.2	Learning the Dependency Graph . . . . .	164
6.2.3	Learning Splits and Joins . . . . .	167
6.3	Genetic Process Mining . . . . .	169
6.4	Region-Based Mining . . . . .	173
6.4.1	Learning Transition Systems . . . . .	174
6.4.2	Process Discovery Using State-Based Regions . . . . .	177

6.4.3 Process Discovery Using Language-Based Regions . . . . . 180  
 6.5 Historical Perspective. . . . . 183

**Part III Beyond Process Discovery**

**7 Conformance Checking** . . . . . 191  
 7.1 Business Alignment and Auditing. . . . . 191  
 7.2 Token Replay. . . . . 194  
 7.3 Comparing Footprints. . . . . 205  
 7.4 Other Applications of Conformance Checking. . . . . 209  
     7.4.1 Repairing Models. . . . . 209  
     7.4.2 Evaluating Process Discovery Algorithms. . . . . 210  
     7.4.3 Connecting Event Log and Process Model. . . . . 211

**8 Mining Additional Perspectives** . . . . . 215  
 8.1 Perspectives. . . . . 215  
 8.2 Attributes: A Helicopter View. . . . . 217  
 8.3 Organizational Mining. . . . . 221  
     8.3.1 Social Network Analysis. . . . . 222  
     8.3.2 Discovering Organizational Structures. . . . . 227  
     8.3.3 Analyzing Resource Behavior. . . . . 228  
 8.4 Time and Probabilities. . . . . 230  
 8.5 Decision Mining . . . . . 234  
 8.6 Bringing It All Together. . . . . 237

**9 Operational Support** . . . . . 241  
 9.1 Refined Process Mining Framework. . . . . 241  
     9.1.1 Cartography. . . . . 243  
     9.1.2 Auditing. . . . . 244  
     9.1.3 Navigation. . . . . 245  
 9.2 Online Process Mining. . . . . 245  
 9.3 Detect . . . . . 247  
 9.4 Predict . . . . . 251  
 9.5 Recommend. . . . . 256  
 9.6 Process Mining Spectrum. . . . . 258

**Part IV Putting Process Mining to Work**

**10 Tool Support** . . . . . 261  
 10.1 Business Intelligence?. . . . . 261  
 10.2 ProM. . . . . 265  
 10.3 Other Process Mining Tools . . . . . 270  
 10.4 Outlook . . . . . 274

**11 Analyzing "Lasagna Processes"** . . . . . 277  
 11.1 Characterization of "Lasagna Processes". . . . . 277  
 11.2 Use Cases. . . . . 281  
 11.3 Approach . . . . . 282

11.3.1	Stage 0: Plan and Justify. . . . .	283
11.3.2	Stage 1: Extract . . . . .	285
11.3.3	Stage 2: Create Control-Flow Model and Connect Event Log . . . . .	285
11.3.4	Stage 3: Create Integrated Process Model . . . . .	286
11.3.5	Stage 4: Operational Support . . . . .	286
11.4	Applications. . . . .	286
11.4.1	Process Mining Opportunities per Functional Area . . . . .	287
11.4.2	Process Mining Opportunities per Sector. . . . .	288
11.4.3	Two Lasagna Processes. . . . .	292
<b>12</b>	<b>Analyzing "Spaghetti Processes". . . . .</b>	<b>301</b>
12.1	Characterization of "Spaghetti Processes". . . . .	301
12.2	Approach . . . . .	305
12.3	Applications. . . . .	309
12.3.1	Process Mining Opportunities for Spaghetti Processes . . . . .	309
12.3.2	Examples of Spaghetti Processes. . . . .	310
<b>Part V Reflection</b>		
<b>13</b>	<b>Cartography and Navigation. . . . .</b>	<b>321</b>
13.1	Business Process Maps. . . . .	321
13.1.1	Map Quality. . . . .	322
13.1.2	Aggregation and Abstraction. . . . .	322
13.1.3	Seamless Zoom. . . . .	324
13.1.4	Size, Color, and Layout. . . . .	328
13.1.5	Customization. . . . .	330
13.2	Process Mining: TomTom for Business Processes?. . . . .	331
13.2.1	Projecting Dynamic Information on Business Process Maps. . . . .	331
13.2.2	Arrival Time Prediction. . . . .	333
13.2.3	Guidance Rather than Control. . . . .	334
<b>14</b>	<b>Epilogue. . . . .</b>	<b>337</b>
14.1	Process Mining: A Bridge Between Data Mining and Business Process Management . . . . .	337
14.2	Challenges. . . . .	339
14.3	Start Today!. . . . .	340
<b>References. . . . .</b>		<b>341</b>
<b>Index. . . . .</b>		<b>349</b>