

Contents

Abbreviations	xi
List of Figures	xiii
List of Tables	xix
1 Introduction	1
1.1 Research Objectives and Methods	5
1.2 Structure of This Thesis	7
2 Research Scope and Context	9
2.1 Industrial Sensor Data	10
2.1.1 Time Series Data	10
2.1.2 Supervised Machine Learning Tasks for Time Series	11
2.1.3 Data Management for Industrial Sensor Data	13
2.2 EDA and VIAL of Data for ML Methods	14
2.2.1 Exploratory Data Analysis	14
2.2.2 Interactive Labeling	18
2.2.3 Roles in Data Science: Process Experts and Data Scientists . .	19
2.3 Summary	20
3 Related Work	21
3.1 Exploratory Analysis of Time Series Data	22
3.2 Analytic Provenance	25
3.2.1 Provenance Data: Recording User Sessions	26
3.2.1.1 Interaction-Based Recording	28

- 3.2.1.2 Application State-Based Recording 29
 - 3.2.1.3 Query-Based Recording 31
 - 3.2.2 Quantifying Insight 32
 - 3.2.3 Visualizing Provenance Data 33
- 3.3 Guidance for Exploratory Data Analysis 35
 - 3.3.1 Visualization Recommendations 35
 - 3.3.2 Recommendations Based on User Preferences 37
 - 3.3.3 Generation of Full Exploratory Data Analysis Sessions and Action Recommendations 38
- 3.4 Labeling of Time Series Data 39
- 3.5 Guidance for Data Labeling 42
- 3.6 Summary 44

4 Guidance for EDA and VIAL of Industrial Sensor Data 47

- 4.1 A Unified Process for Guided EDA and VIAL in the Manufacturing Industry 48
- 4.2 Exemplary Use Cases 53
 - 4.2.1 Deep Drawing of Metal Sheets 53
 - 4.2.2 Metal Arc Welding 55
- 4.3 A System Design for Guidance in EDA and VIAL of Industrial Sensor Data 57
 - 4.3.1 Guidance for Exploratory Data Analysis of Industrial Sensor Data 58
 - 4.3.1.1 Step 1: Analysis Goals 58
 - 4.3.1.2 Step 2: Knowledge Gaps 59
 - 4.3.1.3 Step 3: Guidance Generation 60
 - 4.3.1.4 Step 4: Guidance Feedback Loop 62
 - 4.3.2 Guidance for Labeling of Industrial Sensor Data 63
 - 4.3.2.1 Step 1: Analysis Goals 63
 - 4.3.2.2 Step 2: Knowledge Gaps 63
 - 4.3.2.3 Step 3: Guidance Generation 65
 - 4.3.2.4 Step 4: Guidance Feedback Loop 67
- 4.4 Summary 67

5	Guided Exploratory Data Analysis of Industrial Sensor Data	69
5.1	A System for Exploratory Data Analysis of Industrial Sensor Data . .	71
5.1.1	Design Goals	71
5.1.2	Data: Data Formats and Forms	73
5.1.3	An Exploration UI for Custom Visualizations of Large Data Volumes	74
5.1.4	Prototype Implementation	79
5.1.5	Performance Evaluation	80
5.2	Domain Knowledge: Guidance for EDA of Industrial Sensor Data . .	82
5.2.1	Guidance Input: Tracking Interaction Logs	82
5.2.1.1	Design Goals	83
5.2.1.2	Gideon Replay Design	84
5.2.1.3	Example Applications	87
5.2.2	Guidance Input: Tracking Action Graphs	89
5.2.3	Guidance Generation: Action Sequences and Regions of Interest	91
5.2.4	Guidance Output: Guidance Visualization	95
5.3	Evaluation	96
5.3.1	Participants	97
5.3.2	Task and Design	97
5.3.3	Procedure	98
5.3.4	Results	99
5.3.5	Discussion of Results	101
5.4	Summary	102
6	Visual Interactive Labeling of Large Volumes of Industrial Sensor Data	103
6.1	Labeling Interface	105
6.1.1	Design Goals	105
6.1.2	Graphical User Interface	106
6.2	Guided Labeling	108
6.2.1	Semi-Automated Unsupervised Labeling	109
6.2.2	Active Learning Labeling	116

6.3	Evaluation	121
6.3.1	Semi-Automated Unsupervised Labeling	122
6.3.2	Active Learning Labeling	123
6.3.3	Performance	125
6.4	Summary	126
7	Transfer of Guidance for a Specific Use Case to Other Use Cases	129
7.1	Challenges for Transferability of Guidance	130
7.2	Identification of Similar Use Cases	131
7.3	Transfer of Guidance for EDA and VIAL	135
7.4	Evaluation Strategies	137
7.5	Summary	138
8	Critical Reflection and Outlook	139
9	Summary	145
	Bibliography	149