Contents

About the Authors xi
Preface xiii
About the Companion Website xvii

Part 1 Big Data, Clouds and Internet of Things 1

1 Big Data Science and Machine Intelligence 3
1.1 Enabling Technologies for Big Data Computing 3
1.1.1 Data Science and Related Disciplines 4
1.1.2 Emerging Technologies in the Next Decade 7
1.1.3 Interactive SMACT Technologies 13
1.2 Social-Media, Mobile Networks and Cloud Computing 16
1.2.1 Social Networks and Web Service Sites 17
1.2.2 Mobile Cellular Core Networks 19
1.2.3 Mobile Devices and Internet Edge Networks 20
1.2.4 Mobile Cloud Computing Infrastructure 23
1.3 Big Data Acquisition and Analytics Evolution 24
1.3.1 Big Data Value Chain Extracted from Massive Data 24
1.3.2 Data Quality Control, Representation and Database Models 26
1.3.3 Big Data Acquisition and Preprocessing 27
1.3.4 Evolving Data Analytics over the Clouds 30
1.4 Machine Intelligence and Big Data Applications 32
1.4.1 Data Mining and Machine Learning 32
1.4.2 Big Data Applications – An Overview 34
1.4.3 Cognitive Computing – An Introduction 38
1.5 Conclusions 42
Homework Problems 42
References 43

2 Smart Clouds, Virtualization and Mashup Services 45
2.1 Cloud Computing Models and Services 45
2.1.1 Cloud Taxonomy based on Services Provided 46
2.1.2 Layered Development Cloud Service Platforms 50
2.1.3 Cloud Models for Big Data Storage and Processing 52
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.4</td>
<td>Cloud Resources for Supporting Big Data Analytics</td>
<td>55</td>
</tr>
<tr>
<td>2.2</td>
<td>Creation of Virtual Machines and Docker Containers</td>
<td>57</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Virtualization of Machine Resources</td>
<td>58</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Hypervisors and Virtual Machines</td>
<td>60</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Docker Engine and Application Containers</td>
<td>62</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Deployment Opportunity of VMs/Containers</td>
<td>64</td>
</tr>
<tr>
<td>2.3</td>
<td>Cloud Architectures and Resources Management</td>
<td>65</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Cloud Platform Architectures</td>
<td>65</td>
</tr>
<tr>
<td>2.3.2</td>
<td>VM Management and Disaster Recovery</td>
<td>68</td>
</tr>
<tr>
<td>2.3.3</td>
<td>OpenStack for Constructing Private Clouds</td>
<td>70</td>
</tr>
<tr>
<td>2.3.4</td>
<td>Container Scheduling and Orchestration</td>
<td>74</td>
</tr>
<tr>
<td>2.3.5</td>
<td>VMWare Packages for Building Hybrid Clouds</td>
<td>75</td>
</tr>
<tr>
<td>2.4</td>
<td>Case Studies of IaaS, PaaS and SaaS Clouds</td>
<td>77</td>
</tr>
<tr>
<td>2.4.1</td>
<td>AWS Architecture over Distributed Datacenters</td>
<td>78</td>
</tr>
<tr>
<td>2.4.2</td>
<td>AWS Cloud Service Offerings</td>
<td>79</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Platform PaaS Clouds – Google AppEngine</td>
<td>83</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Application SaaS Clouds – The Salesforce Clouds</td>
<td>86</td>
</tr>
<tr>
<td>2.5</td>
<td>Mobile Clouds and Inter-Cloud Mashup Services</td>
<td>88</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Mobile Clouds and Cloudlet Gateways</td>
<td>88</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Multi-Cloud Mashup Services</td>
<td>91</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Skyline Discovery of Mashup Services</td>
<td>95</td>
</tr>
<tr>
<td>2.5.4</td>
<td>Dynamic Composition of Mashup Services</td>
<td>96</td>
</tr>
<tr>
<td>2.6</td>
<td>Conclusions</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Homework Problems</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>103</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td><strong>IoT Sensing, Mobile and Cognitive Systems</strong></td>
<td>105</td>
</tr>
<tr>
<td>3.1</td>
<td>Sensing Technologies for Internet of Things</td>
<td>105</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Enabling Technologies and Evolution of IoT</td>
<td>106</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Introducing RFID and Sensor Technologies</td>
<td>108</td>
</tr>
<tr>
<td>3.1.3</td>
<td>IoT Architectural and Wireless Support</td>
<td>110</td>
</tr>
<tr>
<td>3.2</td>
<td>IoT Interactions with GPS, Clouds and Smart Machines</td>
<td>111</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Local versus Global Positioning Technologies</td>
<td>111</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Standalone versus Cloud-Centric IoT Applications</td>
<td>114</td>
</tr>
<tr>
<td>3.2.3</td>
<td>IoT Interaction Frameworks with Environments</td>
<td>116</td>
</tr>
<tr>
<td>3.3</td>
<td>Radio Frequency Identification (RFID)</td>
<td>119</td>
</tr>
<tr>
<td>3.3.1</td>
<td>RFID Technology and Tagging Devices</td>
<td>119</td>
</tr>
<tr>
<td>3.3.2</td>
<td>RFID System Architecture</td>
<td>120</td>
</tr>
<tr>
<td>3.3.3</td>
<td>IoT Support of Supply Chain Management</td>
<td>122</td>
</tr>
<tr>
<td>3.4</td>
<td>Sensors, Wireless Sensor Networks and GPS Systems</td>
<td>124</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Sensor Hardware and Operating Systems</td>
<td>124</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Sensing through Smart Phones</td>
<td>130</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Wireless Sensor Networks and Body Area Networks</td>
<td>131</td>
</tr>
<tr>
<td>3.4.4</td>
<td>Global Positioning Systems</td>
<td>134</td>
</tr>
<tr>
<td>3.5</td>
<td>Cognitive Computing Technologies and Prototype Systems</td>
<td>139</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Cognitive Science and Neuroinformatics</td>
<td>139</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Brain-Inspired Computing Chips and Systems</td>
<td>140</td>
</tr>
</tbody>
</table>
3.5.3 Google’s Brain Team Projects 142
3.5.4 IoT Contexts for Cognitive Services 145
3.5.5 Augmented and Virtual Reality Applications 146
3.6 Conclusions 149
Homework Problems 150
References 152

Part 2 Machine Learning and Deep Learning Algorithms 155

4 Supervised Machine Learning Algorithms 157
4.1 Taxonomy of Machine Learning Algorithms 157
4.1.1 Machine Learning Based on Learning Styles 158
4.1.2 Machine Learning Based on Similarity Testing 159
4.1.3 Supervised Machine Learning Algorithms 162
4.1.4 Unsupervised Machine Learning Algorithms 163
4.2 Regression Methods for Machine Learning 164
4.2.1 Basic Concepts of Regression Analysis 164
4.2.2 Linear Regression for Prediction and Forecast 166
4.2.3 Logistic Regression for Classification 169
4.3 Supervised Classification Methods 171
4.3.1 Decision Trees for Machine Learning 171
4.3.2 Rule-based Classification 175
4.3.3 The Nearest Neighbor Classifier 181
4.3.4 Support Vector Machines 183
4.4 Bayesian Network and Ensemble Methods 187
4.4.1 Bayesian Classifiers 188
4.4.2 Bayesian Belief Networks 191
4.4.3 Random Forests and Ensemble Methods 195
4.5 Conclusions 200
Homework Problems 200
References 203

5 Unsupervised Machine Learning Algorithms 205
5.1 Introduction and Association Analysis 205
5.1.1 Introduction to Unsupervised Machine Learning 205
5.1.2 Association Analysis and A priori Principle 206
5.1.3 Association Rule Generation 210
5.2 Clustering Methods without Labels 213
5.2.1 Cluster Analysis for Prediction and Forecasting 213
5.2.2 K-means Clustering for Classification 214
5.2.3 Agglomerative Hierarchical Clustering 217
5.2.4 Density-based Clustering 221
5.3 Dimensionality Reduction and Other Algorithms 225
5.3.1 Dimensionality Reduction Methods 225
5.3.2 Principal Component Analysis (PCA) 226
5.3.3 Semi-Supervised Machine Learning Methods 231
## Contents

5.4 How to Choose Machine Learning Algorithms? 233  
5.4.1 Performance Metrics and Model Fitting 233  
5.4.2 Methods to Reduce Model Over-Fitting 237  
5.4.3 Methods to Avoid Model Under-Fitting 240  
5.4.4 Effects of Using Different Loss Functions 242  
5.5 Conclusions 243  

Homework Problems 243  
References 247

6 Deep Learning with Artificial Neural Networks 249  
6.1 Introduction 249  
6.1.1 Deep Learning Mimics Human Senses 249  
6.1.2 Biological Neurons versus Artificial Neurons 251  
6.1.3 Deep Learning versus Shallow Learning 254  
6.2 Artificial Neural Networks (ANN) 256  
6.2.1 Single Layer Artificial Neural Networks 256  
6.2.2 Multilayer Artificial Neural Network 257  
6.2.3 Forward Propagation and Back Propagation in ANN 258  
6.3 Stacked AutoEncoder and Deep Belief Network 264  
6.3.1 AutoEncoder 264  
6.3.2 Stacked AutoEncoder 267  
6.3.3 Restricted Boltzmann Machine 269  
6.3.4 Deep Belief Networks 275  
6.4 Convolutional Neural Networks (CNN) and Extensions 277  
6.4.1 Convolution in CNN 277  
6.4.2 Pooling in CNN 280  
6.4.3 Deep Convolutional Neural Networks 282  
6.4.4 Other Deep Learning Networks 283  
6.5 Conclusions 287  

Homework Problems 288  
References 291

Part 3 Big Data Analytics for Health-Care and Cognitive Learning 293

7 Machine Learning for Big Data in Healthcare Applications 295  
7.1 Healthcare Problems and Machine Learning Tools 295  
7.1.1 Healthcare and Chronic Disease Detection Problem 295  
7.1.2 Software Libraries for Machine Learning Applications 298  
7.2 IoT-based Healthcare Systems and Applications 299  
7.2.1 IoT Sensing for Body Signals 300  
7.2.2 Healthcare Monitoring System 301  
7.2.3 Physical Exercise Promotion and Smart Clothing 304  
7.2.4 Healthcare Robotics and Mobile Health Cloud 305  
7.3 Big Data Analytics for Healthcare Applications 310  
7.3.1 Healthcare Big Data Preprocessing 310  
7.3.2 Predictive Analytics for Disease Detection 312
Contents

7.3.3 Performance Analysis of Five Disease Detection Methods 316
7.3.4 Mobile Big Data for Disease Control 320
7.4 Emotion-Control Healthcare Applications 322
7.4.1 Mental Healthcare System 323
7.4.2 Emotion-Control Computing and Services 323
7.4.3 Emotion Interaction through IoT and Clouds 327
7.4.4 Emotion-Control via Robotics Technologies 329
7.4.5 A 5G Cloud-Centric Healthcare System 332
7.5 Conclusions 335

Homework Problems 336
References 339

8 Deep Reinforcement Learning and Social Media Analytics 343
8.1 Deep Learning Systems and Social Media Industry 343
8.1.1 Deep Learning Systems and Software Support 343
8.1.2 Reinforcement Learning Principles 346
8.1.3 Social-Media Industry and Global Impact 347
8.2 Text and Image Recognition using ANN and CNN 348
8.2.1 Numeral Recognition using TensorFlow for ANN 349
8.2.2 Numeral Recognition using Convolutional Neural Networks 352
8.2.3 Convolutional Neural Networks for Face Recognition 356
8.2.4 Medical Text Analytics by Convolutional Neural Networks 357
8.3 DeepMind with Deep Reinforcement Learning 362
8.3.1 Google DeepMind AI Programs 362
8.3.2 Deep Reinforcement Learning Algorithm 364
8.3.3 Google AlphaGo Game Competition 367
8.3.4 Flappybird Game using Reinforcement Learning 371
8.4 Data Analytics for Social-Media Applications 375
8.4.1 Big Data Requirements in Social-Media Applications 375
8.4.2 Social Networks and Graph Analytics 377
8.4.3 Predictive Analytics Software Tools 383
8.4.4 Community Detection in Social Networks 386
8.5 Conclusions 390

Homework Problems 391
References 393

Index 395