Smart Cities: Big Data Prediction Methods and Applications

Hui Liu

Smart Cities: Big Data Prediction Methods and Applications

Science Press
Beijing

Springer

Hui Liu
School of Traffic and Transportation Engineering
Central South University
Changsha, Hunan, China

ISBN 978-981-15-2836-1 ISBN 978-981-15-2837-8 (eBook)
https://doi.org/10.1007/978-981-15-2837-8

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Smart city is to make full use of the new generation of information technology in various industries in the city based on the next generation of innovation in the knowledge society of the advanced form of urban information to achieve the deep integration of information, industrialization, and urbanization. The smart city in the twenty-first century can make the most of big data processing technology to analyze and mine the key information of the core system of urban operation, to make an intelligent response to a variety of urban needs, including transportation, environmental protection, public safety, and industrial and commercial activities, and to achieve a better urban life.

A smart city often intersects with big data city, intelligent city, ecological city, low-carbon city, and other regional development concepts, and even mixes with intelligent environment, intelligent traffic, smart grid, and other industry information concepts. At present, research on smart cities has different emphasis, mostly focused on technology application discussion, network construction analysis, wisdom effect research, etc. The research of smart cities is still in the new stage of vigorous development. Therefore, from the point of view of data science, the author has completed the work based on his theoretical achievements in the field of time series and the results of a large number of engineering experiments in smart grid, smart traffic, and smart environment.

Under the background of smart city, the book carries on big data prediction from three aspects of the power grid and building energy, road network traffic flow, environmental index data, which is of great practical significance to the construction and planning of smart city. With the help of big data prediction of power grid and building energy, it can further improve the power grid system and building energy management, the safety of power consumption, and the efficiency of energy utilization. With the help of big data prediction of road network traffic flow, the road network structure can be further improved, alleviate traffic congestion, and enhance the traffic capacity of the road network. With the help of big data prediction of the environment, it can further improve the quality of resident life and provide a healthy and green living environment for residents. The prediction methods proposed of smart cities based on big data in the book have a certain reference significance for the planning, construction, and development of green cities and smart cities.

Changsha, China
August 2019

Hui Liu

Acknowledgements

The studies in the book are supported by the National Natural Science Foundation of China, the National Key R&D Program of China, the Innovation Drive of Central South University, China. In the process of writing the book, Huipeng Shi, Zhihao Long, Hengxin Yin, Guangxi Yan, Zeyu Liu, and other team members have done a lot of model verification and other works; the author would like to express heartfelt appreciations.

About the Book

This book proposes theoretical methods and predictive models for large-scale data prediction and analysis in the smart cities using artificial intelligence and big data. At the same time, this book uses the method of case analysis to verify and analyze the proposed model. All content is divided into four parts, totaling 10 chapters. Part 1 summarizes the research background and development status of several key issues in smart city involved in the book and summarizes the related technologies. This part uses the statistical analysis of the literature to expound the research hotspots and research progress of smart cities with big data technology. It provides a theoretical basis for the study of theoretical algorithms in subsequent chapters. Part 2 uses the cross-correlation analysis to mine and analyze the regional correlation and time-domain variation in urban power load data. At the same time, this part uses ARIMA, SARIMA, ARCH, and SVM to predict the urban electricity consumption and household load consumption and constructs the framework of power load forecasting under the background of big data. Part 3 analyzes the traffic flow data and the vehicle trajectory data and establishes a traffic flow prediction model and a vehicle trajectory prediction model, respectively. Firstly, this part starts with single data-driven traffic flow data and establishes a deterministic prediction model and interval prediction model for traffic flow. Then, the multi-dimensional traffic flow data including the spatiotemporal mapping relationship is analyzed to realize the traffic flow prediction driven by multi-dimensional data. Part 4 starts with the three aspects of urban air quality, urban hydrological status, and urban noise and introduces the intelligent environment prediction technology in the context of big data. The ELM model, Bayesian model, RF algorithm, BFGS algorithm, and GRU model are used to predict the urban environmental time series.

This book can help scientists, engineers, managers, and students engaged in artificial intelligence, big data analysis and forecasting, computational intelligence, smart city, smart grids and energy, intelligent traffic systems, smart environment, air pollutant controlling, and other related research fields.

Contents

Part I Exordium

1 Key Issues of Smart Cities 3
1.1 Smart Grid and Buildings 3
1.1.1 Overview of Smart Grid and Building 4
1.1.2 The Importance of Smart Grid and Buildings in Smart City 5
1.1.3 Framework of Smart Grid and Buildings 6
1.2 Smart Traffic Systems 6
1.2.1 Overview of Smart Traffic Systems 6
1.2.2 The Importance of Smart Traffic Systems for Smart City 6
1.2.3 Framework of Smart Traffic Systems 8
1.3 Smart Environment 8
1.3.1 Overview of Smart Environment for Smart City 8
1.3.2 The Importance of Smart Environment for Smart City 10
1.3.3 Framework of Smart Environment 11
1.4 Framework of Smart Cities 11
1.4.1 Key Points of Smart City in the Era of Big Data 11
1.4.2 Big Data Time-series Forecasting Methods in Smart Cities 12
1.4.3 Overall Framework of Big Data Forecasting in Smart Cities 13
1.5 The Importance Analysis of Big Data Forecasting Architecture for Smart Cities 14
1.5.1 Overview and Necessity of Research 14
1.5.2 Review on Big Data Forecasting in Smart Cities 15
1.5.3 Review on Big Data Forecasting in Smart Gird and Buildings 18

1.5.4 Review on Big Data Forecasting in Smart Traffic Systems 21
1.5.5 Review on Big Data Forecasting in Smart Environment 22
References 23

Part II Smart Grid and Buildings

2 Electrical Characteristics and Correlation Analysis in Smart Grid 27
2.1 Introduction 27
2.2 Extraction of Building Electrical Features 28
2.2.1 Analysis of Meteorological Elements 29
2.2.2 Analysis of System Load 30
2.2.3 Analysis of Thermal Perturbation 31
2.3 Cross-Correlation Analysis of Electrical Characteristics 33
2.3.1 Cross-Correlation Analysis Based on MI 33
2.3.2 Cross-Correlation Analysis Based on Pearson Coefficient 35
2.3.3 Cross-Correlation Analysis Based on Kendall Coefficient 37
2.4 Selection of Electrical Characteristics 40
2.4.1 Electrical Characteristics of Construction Power Grid 40
2.4.2 Feature Selection Based on Spearman Correlation Coefficient 41
2.4.3 Feature Selection Based on CFS 43
2.4.4 Feature Selection Based on Global Search-ELM 45
2.5 Conclusion 46
References 48

3 Prediction Model of City Electricity Consumption 51
3.1 Introduction 51
3.2 Original Electricity Consumption Series 54
3.2.1 Regional Correlation Analysis of Electricity Consumption Series 54
3.2.2 Original Sequences for Modeling 55
3.2.3 Separation of Sample 56
3.3 Short-Term Deterministic Prediction of Electricity Consumption Based on ARIMA Model 58
3.3.1 Model Framework of ARIMA 58
3.3.2 Theoretical Basis of ARIMA 59
3.3.3 Modeling Steps of ARIMA Predictive Model 60
3.3.4 Forecasting Results 64

3.4 Power Consumption Interval Prediction Based on ARIMA-ARCH Model 69
3.4.1 Model Framework of ARCH 69
3.4.2 The Theoretical Basis of the ARCH 69
3.4.3 Modeling Steps of ARIMA-ARCH Interval Predictive Model 70
3.4.4 Forecasting Results 71
3.5 Long-Term Electricity Consumption Prediction Based on the SARIMA Model 76
3.5.1 Model Framework of the SARIMA 76
3.5.2 The Theoretical Basis of the SARIMA 77
3.5.3 Modeling Steps of the SARIMA Predictive Model 78
3.5.4 Forecasting Results 79
3.6 Big Data Prediction Architecture of Household Electric Power 81
3.7 Comparative Analysis of Forecasting Performance 84
3.8 Conclusion 86
References 88

4 Prediction Models of Energy Consumption in Smart Urban Buildings 89
4.1 Introduction 89
4.2 Establishment of Building Simulating Model 91
4.2.1 Description and Analysis of the BEMPs 91
4.2.2 Main Characters of DeST Software 94
4.2.3 Process of DeST Modeling 95
4.3 Analysis and Comparison of Different Parameters 101
4.3.1 Introduction of the Research 101
4.3.2 Meteorological Parameters 102
4.3.3 Indoor Thermal Perturbation 103
4.3.4 Enclosure Structure and Material Performance 105
4.3.5 Indoor Design Parameters 106
4.4 Data Acquisition of Building Model 108
4.4.1 Data After Modeling 108
4.4.2 Calculation of Room Temperature and Load 108
4.4.3 Calculation of Shadow and Light 108
4.4.4 Calculation of Natural Ventilation 109
4.4.5 Simulation of the Air-Conditioning System 110
4.5 SVM Prediction Model for Urban Building Energy Consumption 110
4.5.1 The Theoretical Basis of the SVM 110
4.5.2 Steps of Modeling 112
4.5.3 Forecasting Results 114
4.6 Big Data Prediction of Energy Consumption in Urban Building 115

4.6.1 Big Data Framework for Energy Consumption 117
4.6.2 Big Data Storage and Analysis for Energy Consumption 117
4.6.3 Big Data Mining for Energy Consumption 117
4.7 Conclusion 119
References 120

Part III Smart Traffic Systems

5 Characteristics and Analysis of Urban Traffic Flow in Smart Traffic Systems 125
5.1 Introduction 125
5.1.1 Overview of Trajectory Prediction of Smart Vehicle 125
5.1.2 The Significance of Trajectory Prediction for Smart City 126
5.1.3 Overall Framework of Model 127
5.2 Traffic Flow Time Distribution Characteristics and Analysis 129
5.2.1 Original Vehicle Trajectory Series 129
5.2.2 Separation of Sample 131
5.3 The Spatial Distribution Characteristics and Analysis of Traffic Flow 132
5.3.1 Trajectory Prediction of Urban Vehicles Based on Single Data 132
5.3.2 Trajectory Prediction of Urban Vehicles Based on Multiple Data 140
5.3.3 Trajectory Prediction of Urban Vehicles Under EWT Decomposition Framework 146
5.3.4 Comparative Analysis of Forecasting Performance 153
5.4 Conclusion 156
References 157

6 Prediction Model of Traffic Flow Driven Based on Single Data in Smart Traffic Systems 159
6.1 Introduction 159
6.2 Original Traffic Flow Series for Prediction 161
6.3 Traffic Flow Deterministic Prediction Driven by Single Data 162
6.3.1 Modeling Process 162
6.3.2 The Prediction Results 167
6.4 Traffic Flow Interval Prediction Model Driven by Single Data 167
6.4.1 The Framework of the Interval Prediction Model 167
6.4.2 Modeling Process 170
6.4.3 The Prediction Results 174

6.5 Traffic Flow Interval Prediction Under Decomposition Framework . . . 175
6.5.1 The Framework of the WD-BP-GARCH Prediction Model . . . 175
6.5.2 Modeling Process . . . 184
6.5.3 The Prediction Results . . . 187
6.6 Big Data Prediction Architecture of Traffic Flow . . . 190
6.7 Comparative Analysis of Forecasting Performance . . . 191
6.8 Conclusion . . . 193
References . . . 193

7 Prediction Models of Traffic Flow Driven Based on Multi-Dimensional Data in Smart Traffic Systems . . . 195
7.1 Introduction . . . 195
7.2 Analysis of Traffic Flow and Its Influencing Factors . . . 196
7.3 Elman Prediction Model of Traffic Flow Based on Multiple Data . . . 198
7.3.1 The Framework of the Elman Prediction Model . . . 198
7.3.2 Modeling Process . . . 198
7.3.3 The Prediction Results . . . 201
7.4 LSTM Prediction Model of Traffic Flow Based on Multiple Data . . . 202
7.4.1 The Framework of the LSTM Prediction Model . . . 202
7.4.2 Modeling Process . . . 205
7.4.3 The Prediction Results . . . 206
7.5 Traffic Flow Prediction Under Wavelet Packet Decomposition . . . 207
7.5.1 The Framework of the WPD-Prediction Model . . . 207
7.5.2 Modeling Process . . . 210
7.5.3 The Prediction Results . . . 214
7.6 Comparative Analysis of Forecasting Performance . . . 216
7.7 Conclusion . . . 220
References . . . 222

Part IV Smart Environment

8 Prediction Models of Urban Air Quality in Smart Environment . . . 227
8.1 Introduction . . . 227
8.2 Original Air Pollutant Concentrations Series for Prediction . . . 228
8.2.1 Original Sequence for Modeling . . . 228
8.2.2 Separation of Sample . . . 231
8.3 Air Quality Prediction Model Driven by Single Data . . . 232
8.3.1 Model Framework . . . 232
8.3.2 Theoretical Basis of ELM . . . 232

8.3.3 Steps of Modeling 233
8.3.4 Forecasting Results 233
8.4 Air Quality Mixture Prediction Model Driven by Multiple Data 234
8.4.1 Model Framework 234
8.4.2 Steps of Modeling 235
8.4.3 Forecasting Results 237
8.5 Air Quality Prediction Under Feature Extraction Framework 238
8.5.1 Model Framework 238
8.5.2 Theoretical Basis of Feature Extraction Method 238
8.5.3 Steps of Modeling 250
8.5.4 Forecasting Results 251
8.6 Big Data Prediction Architecture of Urban Air Quality 253
8.6.1 The Idea of Urban Air Quality Prediction Based on Hadoop 253
8.6.2 Parallelization Framework of the ELM 254
8.6.3 The Parallelized ELM Under the MapReduce Framework 254
8.7 Comparative Analysis of Forecasting Performance 256
8.8 Conclusion 258
References 259

9 Prediction Models of Urban Hydrological Status in Smart Environment 261
9.1 Introduction 261
9.2 Original Hydrological State Data for Prediction 262
9.2.1 Original Sequence for Modeling 262
9.2.2 Separation of Sample 265
9.3 Bayesian Classifier Prediction of Water Level Fluctuation 265
9.3.1 Model Framework 265
9.3.2 Theoretical Basis of the Bayesian Classifier 266
9.3.3 Steps of Modeling 267
9.3.4 Forecasting Results 268
9.4 The Elman Prediction of Urban Water Level 269
9.4.1 Model Framework 269
9.4.2 The Theoretical Basis of the Elman 270
9.4.3 Steps of Modeling 270
9.4.4 Forecasting Results 271
9.5 Urban River Water Level Decomposition Hybrid Prediction Model 272
9.5.1 Model Framework 272
9.5.2 The Theoretical Basis 272
9.5.3 Steps of Modeling 276
9.5.4 Forecasting Results 277

9.5.5 Influence and Analysis of Decomposition Parameters on Forecasting Performance of Hybrid Models 280
9.6 Comparative Analysis of Forecasting Performance 284
9.7 Conclusion 287
References 288

10 Prediction Model of Urban Environmental Noise in Smart Environment 289
10.1 Introduction 289
10.1.1 Hazard of Noise 289
10.1.2 The Significance of Noise Prediction for Smart City 290
10.1.3 Overall Framework of Model 291
10.2 Original Urban Environmental Noise Series 292
10.2.1 Original Sequence for Modeling. 292
10.2.2 Separation of Sample 294
10.3 The RF Prediction Model for Urban Environmental Noise 295
10.3.1 The Theoretical Basis of the RF 295
10.3.2 Steps of Modeling 295
10.3.3 Forecasting Results 296
10.4 The BFGS Prediction Model for Urban Environmental Noise 298
10.4.1 The Theoretical Basis of the BFGS 298
10.4.2 Steps of Modeling 299
10.4.3 Forecasting Results 299
10.5 The GRU Prediction Model for Urban Environmental Noise 302
10.5.1 The Theoretical Basis of the GRU 302
10.5.2 Steps of Modeling 303
10.5.3 Forecasting Results 304
10.6 Big Data Prediction Architecture of Urban Environmental Noise 305
10.6.1 Big Data Framework for Urban Environmental Noise Prediction 307
10.6.2 Big Data Storage for Urban Environmental Noise Prediction 308
10.6.3 Big Data Processing of Urban Environmental Noise Prediction 308
10.7 Comparative Analysis of Forecasting Performance 310
10.8 Conclusion 312
References 313

List of Figures

Fig. 1.1 Framework of smart grid and buildings 7
Fig. 1.2 Framework of smart traffic systems 9
Fig. 1.3 Framework of smart environment. 12
Fig. 1.4 Overall framework of big data forecasting in smart cities 14
Fig. 1.5 Citation report of subject retrieval on "TS= (smart cities) AND TS= (big data forecasting OR big data prediction)". 16
Fig. 1.6 The network diagram of documents based on the subject retrieval of "TS= (smart cities) AND TS= (big data forecasting OR big data prediction)". 17
Fig. 1.7 Research direction map of document based on the subject retrieval of "TS= (smart cities) AND TS= (big data forecasting OR big data prediction)". 18
Fig. 1.8 The overlay map of document based on the subject retrieval of "TS= (smart cities) AND TS= (big data forecasting OR big data prediction)" 19
Fig. 1.9 The item density map of document based on the subject retrieval of "TS= (smart cities) AND TS= (big data forecasting OR big data prediction)" 20
Fig. 1.10 The cluster density map of document based on the subject retrieval of "TS= (smart cities) AND TS= (big data forecasting OR big data prediction)". 20
Fig. 1.11 Subject retrieval of annual publication volume of various types of literature on "TS= (smart grid OR smart buildings) AND TS= (big data forecasting OR big data prediction)". 21
Fig. 1.12 Subject retrieval of annual publication volume of various types of literature on "TS= (smart traffic OR smart transportation) AND TS= (big data forecasting OR big data prediction)" 22
Fig. 1.13 Subject retrieval of annual publication volume of various types of literature on "TS= (smart environment) AND TS= (big data forecasting OR big data prediction)". 23

Fig. 2.1 Original meteorological factors series 29
Fig. 2.2 Original system load factors series 31
Fig. 2.3 Original thermal perturbation indoors series 32
Fig. 2.4 Heat map of cross-correlation result based on MI 34
Fig. 2.5 Heat map of cross-correlation result based on Pearson coefficient 36
Fig. 2.6 Heat map of cross-correlation result based on Kendall coefficient 38
Fig. 2.7 Original supply air volume series 39
Fig. 2.8 Original loop pressure loss series 39
Fig. 2.9 Original power consumption series 40
Fig. 2.10 Feature selection result based on Spearman correlation analysis . . 42
Fig. 2.11 Flowchart of CFS 43
Fig. 2.12 Flowchart of Global Search-ELM 46
Fig. 2.13 Feature selection result based on Global Search-ELM. 47

Fig. 3.1 Long-term electricity consumption series of different regions. . . . 54
Fig. 3.2 Short-term electricity consumption series 56
Fig. 3.3 Long-term electricity consumption series 57
Fig. 3.4 Separation of short-term electricity consumption series 57
Fig. 3.5 Separation of long-term electricity consumption series 58
Fig. 3.6 Modeling process of ARIMA predictive model 59
Fig. 3.7 Short-term electricity consumption first-order difference series 61
Fig. 3.8 Long-term electricity consumption first-order difference series 61
Fig. 3.9 Short-term series calculation results of autocorrelation coefficient and the partial autocorrelation coefficient 63
Fig. 3.10 Long-term series calculation results of autocorrelation coefficient and the partial autocorrelation coefficient 63
Fig. 3.11 Short-term series forecasting results of the ARIMA models in 1-step 65
Fig. 3.12 Short-term series forecasting results of the ARIMA models in 2-step 65
Fig. 3.13 Short-term series forecasting results of the ARIMA models in 3-step 66
Fig. 3.14 Long-term series forecasting results of the ARIMA models in 1-step 67
Fig. 3.15 Long-term series forecasting results of the ARIMA models in 2-step 67
Fig. 3.16 Long-term series forecasting results of the ARIMA models in 3-step 68
Fig. 3.17 Modeling process of the ARCH predictive model 69
Fig. 3.18 Short-term series predicted residuals series 71
Fig. 3.19 Long-term series predicted residuals series 72

Fig. 3.20 Short-term series interval forecasting results of the ARIMA models in 1-step . . . 74
Fig. 3.21 Short-term series interval forecasting results of the ARIMA models in 2-step . . . 74
Fig. 3.22 Short-term series interval forecasting results of the ARIMA models in 3-step . . . 75
Fig. 3.23 Long-term series interval forecasting results of the ARIMA models in 1-step . . . 75
Fig. 3.24 Long-term series interval forecasting results of the ARIMA models in 2-step . . . 76
Fig. 3.25 Long-term series interval forecasting results of the ARIMA models in 3-step . . . 76
Fig. 3.26 Modeling process of SARIMA predictive model . . . 77
Fig. 3.27 Short-term series relationship between minimum BIC detection and period . . . 79
Fig. 3.28 Long-term series relationship between minimum BIC detection and period . . . 80
Fig. 3.29 Short-term series forecasting results of the SARIMA models in 1-step . . . 81
Fig. 3.30 Short-term series forecasting results of the SARIMA models in 2-step . . . 82
Fig. 3.31 Short-term series forecasting results of the SARIMA models in 3-step . . . 82
Fig. 3.32 Long-term series forecasting results of the SARIMA models in 1-step . . . 83
Fig. 3.33 Long-term series forecasting results of the SARIMA models in 2-step . . . 83
Fig. 3.34 Long-term series forecasting results of the SARIMA models in 3-step . . . 84
Fig. 3.35 The MapReduce prediction architecture of electricity consumption . . . 85

Fig. 4.1 Connection diagram between smart city and smart building . . . 91
Fig. 4.2 Building energy modeling programs . . . 93
Fig. 4.3 Annual base temperature chart of 10 rooms . . . 94
Fig. 4.4 Daily dry-bulb temperature . . . 95
Fig. 4.5 2-D floor plan of 6 story building . . . 96
Fig. 4.6 3-D display of 6 story building . . . 97
Fig. 4.7 Building model plan . . . 99
Fig. 4.8 Shadow analysis of the building model from daily and annual solar path . . . 100
Fig. 4.9 The steps of modeling . . . 101
Fig. 4.10 Annual hourly load in the building . . . 102
Fig. 4.11 Annual hourly load per area in the building . . . 103
Fig. 4.12 The temperature of the hottest month . . . 103

Fig. 4.13 The temperature of the coldest month 104
Fig. 4.14 Annual direct solar radiation 104
Fig. 4.15 The original data of 60 days load 105
Fig. 4.16 The temperature of 60 days after adjustment of meteorological parameter 105
Fig. 4.17 Indoor thermal perturbation setting of a room in a week 106
Fig. 4.18 The temperature of 60 days after adjustment of indoor thermal perturbation. 106
Fig. 4.19 The material change of the exterior wall 107
Fig. 4.20 The temperature of 60 days after adjustment of enclosure structure and material 108
Fig. 4.21 The construction of a 3-floor building with elevator 109
Fig. 4.22 The temperature of 60 days after adjustment of building construction 110
Fig. 4.23 Forecasting results of 1-step strategy 115
Fig. 4.24 Forecasting results of 2-step strategy 115
Fig. 4.25 Forecasting results of 3-step strategy 116
Fig. 4.26 The framework of Hadoop-SVM model. 119

Fig. 5.1 The general framework of the chapter 128
Fig. 5.2 D1: (**a**) Original longitude series; (**b**) Original latitude series; (**c**) Original speed series; (**d**) Original angle series 129
Fig. 5.3 D2: (**a**) Original longitude series; (**b**) Original latitude series; (**c**) Original speed series; (**d**) Original angle series 130
Fig. 5.4 D3: (**a**) Original longitude series; (**b**) Original latitude series; (**c**) Original speed series; (**d**) Original angle series 130
Fig. 5.5 The prediction principle of BP network 133
Fig. 5.6 The general framework of this section 134
Fig. 5.7 Trajectory prediction results of the ELM model (single data) for D1 135
Fig. 5.8 Trajectory prediction results of the ELM model (single data) for D2 135
Fig. 5.9 Trajectory prediction results of the ELM model (single data) for D3 136
Fig. 5.10 Trajectory prediction results of the BPNN (single data) for D1. . 138
Fig. 5.11 Trajectory prediction results of the BPNN (single data) for D2. . 138
Fig. 5.12 Trajectory prediction results of the BPNN (single data) for D3. . 139
Fig. 5.13 The general framework of this section 140
Fig. 5.14 Trajectory prediction results of the ELM model (multiple data) for D1 141
Fig. 5.15 Trajectory prediction results of the ELM model (multiple data) for D2 142
Fig. 5.16 Trajectory prediction results of the ELM model (multiple data) for D3 142
Fig. 5.17 Trajectory prediction results of the BPNN (multiple data) for D1 144

Fig. 5.18 Trajectory prediction results of the BPNN (multiple data) for D2 . . . 144
Fig. 5.19 Trajectory prediction results of the BPNN (multiple data) for D3 . . . 145
Fig. 5.20 The general framework of this chapter . . . 148
Fig. 5.21 Trajectory prediction results of the EWT-ELM model for D1 . . . 149
Fig. 5.22 Trajectory prediction results of the EWT-ELM model for D2 . . . 150
Fig. 5.23 Trajectory prediction results of the EWT-ELM model for D3 . . . 150
Fig. 5.24 Trajectory prediction results of the EWT-BPNN for D1 . . . 152
Fig. 5.25 Trajectory prediction results of the EWT-BPNN for D2 . . . 152
Fig. 5.26 Trajectory prediction results of the EWT-BPNN for D3 . . . 153

Fig. 6.1 Distribution of traffic flow series . . . 161
Fig. 6.2 The input/output data structure of single data . . . 163
Fig. 6.3 The training set and testing set . . . 163
Fig. 6.4 The results of the 1-step prediction for the traffic flow series . . . 164
Fig. 6.5 The results of the 2-step prediction for the traffic flow series . . . 164
Fig. 6.6 The results of the 3-step prediction for the traffic flow series . . . 165
Fig. 6.7 The results of the 4-step prediction for the traffic flow series . . . 165
Fig. 6.8 The results of the 5-step prediction for the traffic flow series . . . 166
Fig. 6.9 Comparison of prediction results based on different prediction steps . . . 166
Fig. 6.10 Interval prediction model flow . . . 168
Fig. 6.11 The division of datasets . . . 169
Fig. 6.12 The results of the 1-step interval prediction for the traffic flow series . . . 171
Fig. 6.13 The results of the 2-step interval prediction for the traffic flow series . . . 172
Fig. 6.14 The results of the 3-step interval prediction for the traffic flow series . . . 172
Fig. 6.15 The results of the 4-step interval prediction for the traffic flow series . . . 173
Fig. 6.16 The results of the 5-step interval prediction for the traffic flow series . . . 173
Fig. 6.17 The structure of the WD . . . 175
Fig. 6.18 The prediction process of the WD-BP-GARCH model for traffic flow series . . . 176
Fig. 6.19 The eight sub-series of the traffic flow series based on the WD . . . 178
Fig. 6.20 The prediction results of 3-step for each sub-series based on the BP prediction model . . . 179
Fig. 6.21 The prediction results of 3-step for each sub-series based on the BP-GARCH interval model . . . 180
Fig. 6.22 The results of the 1-step prediction for the traffic flow series . . . 181
Fig. 6.23 The results of the 2-step prediction for the traffic flow series . . . 181
Fig. 6.24 The results of the 3-step prediction for the traffic flow series . . . 182

Fig. 6.25 The results of the 4-step prediction for the traffic flow series . . . 182
Fig. 6.26 The results of the 5-step prediction for the traffic flow series . . . 183
Fig. 6.27 Comparison of prediction results based on different prediction steps . . . 183
Fig. 6.28 The results of the 1-step interval prediction for the traffic flow series . . . 185
Fig. 6.29 The results of the 2-step interval prediction for the traffic flow series . . . 185
Fig. 6.30 The results of the 3-step interval prediction for the traffic flow series . . . 186
Fig. 6.31 The results of the 4-step interval prediction for the traffic flow series . . . 186
Fig. 6.32 The results of the 5-step interval prediction for the traffic flow series . . . 187
Fig. 6.33 The MapReduce prediction architecture of traffic flow . . . 189

Fig. 7.1 Traffic flow motion diagram . . . 197
Fig. 7.2 The principle of the Elman network . . . 199
Fig. 7.3 The map of Rd1, Rd2, and Rd3 . . . 200
Fig. 7.4 The sample distribution of the three traffic flow series . . . 200
Fig. 7.5 The input/output data structure of one-step prediction . . . 201
Fig. 7.6 The results of the 1-step prediction for the traffic flow series . . . 202
Fig. 7.7 The results of the 2-step prediction for the traffic flow series . . . 203
Fig. 7.8 The results of the 3-step prediction for the traffic flow series . . . 203
Fig. 7.9 The results of the 4-step prediction for the traffic flow series . . . 204
Fig. 7.10 The results of the 5-step prediction for the traffic flow series . . . 204
Fig. 7.11 The principle of LSTM network . . . 205
Fig. 7.12 The results of the 1-step prediction for the traffic flow series . . . 206
Fig. 7.13 The results of the 2-step prediction for the traffic flow series . . . 207
Fig. 7.14 The results of the 3-step prediction for the traffic flow series . . . 207
Fig. 7.15 The results of the 4-step prediction for the traffic flow series . . . 208
Fig. 7.16 The results of the 5-step prediction for the traffic flow series . . . 208
Fig. 7.17 The structure of the WPD . . . 209
Fig. 7.18 The principle of WPD-prediction model . . . 210
Fig. 7.19 The sample distribution of the three traffic flow series . . . 211
Fig. 7.20 The eight sub-series of Rd1 traffic flow series based on the WPD . . . 212
Fig. 7.21 The eight sub-series of Rd2 traffic flow series based on the WPD . . . 212
Fig. 7.22 The eight sub-series of Rd3 traffic flow series based on the WPD . . . 213
Fig. 7.23 The input/output data structure of 1-step prediction . . . 213
Fig. 7.24 The results of the 1-step prediction for the traffic flow series . . . 215
Fig. 7.25 The results of the 2-step prediction for the traffic flow series . . . 215
Fig. 7.26 The results of the 3-step prediction for the traffic flow series . . . 216

Fig. 7.27 The results of the 4-step prediction for the traffic flow series . . . 216
Fig. 7.28 The results of the 5-step prediction for the traffic flow series . . . 217
Fig. 7.29 The results of the 1-step prediction for the traffic flow series . . . 218
Fig. 7.30 The results of the 2-step prediction for the traffic flow series . . . 218
Fig. 7.31 The results of the 3-step prediction for the traffic flow series . . . 219
Fig. 7.32 The results of the 4-step prediction for the traffic flow series . . . 219
Fig. 7.33 The results of the 5-step prediction for the traffic flow series . . . 220

Fig. 8.1 Schematic diagram of the prediction process of the proposed prediction models . . . 229
Fig. 8.2 Original AQI series . . . 230
Fig. 8.3 Original air pollutant concentrations series . . . 230
Fig. 8.4 Modeling flowchart of ELM prediction model driven by single data . . . 232
Fig. 8.5 The forecasting results of AQI time series by ELM (single data) in 1-step . . . 234
Fig. 8.6 The forecasting results of AQI time series by ELM (single data) in 2-step . . . 234
Fig. 8.7 The forecasting results of AQI time series by the ELM (single data) in 3-step . . . 235
Fig. 8.8 Modeling flowchart of the ELM prediction model driven by multiple data . . . 236
Fig. 8.9 The forecasting results of AQI series by the ELM (multiple data) in 1-step . . . 236
Fig. 8.10 The forecasting results of AQI series by the ELM (multiple data) in 2-step . . . 237
Fig. 8.11 The forecasting results of AQI series by the ELM (multiple data) in 3-step . . . 238
Fig. 8.12 Modeling flowchart of the ELM prediction model under feature extraction framework . . . 239
Fig. 8.13 Box diagram of original feature data . . . 241
Fig. 8.14 Box diagram of standardized feature data . . . 241
Fig. 8.15 Identification results of PCA . . . 243
Fig. 8.16 Sample series of selected principal components . . . 243
Fig. 8.17 The identification results of the Gaussian KPCA . . . 247
Fig. 8.18 Identification results of the FA . . . 250
Fig. 8.19 Sample series of selected common factors . . . 251
Fig. 8.20 The forecasting results of AQI time series by the PCA-ELM in 1-step . . . 252
Fig. 8.21 The forecasting results of AQI time series by the KPCA-ELM in 1-step . . . 252
Fig. 8.22 The forecasting results of AQI time series by the FA-ELM in 1-step . . . 253
Fig. 8.23 Structure-based parallelized ELM . . . 254
Fig. 8.24 The data-based parallelized ELM . . . 255

Fig. 8.25 The forecasting results of AQI time series by proposed models in 1-step 256
Fig. 8.26 The forecasting results of AQI time series by proposed models in 2-step 257
Fig. 8.27 The forecasting results of AQI time series by proposed models in 3-step 257

Fig. 9.1 Original water level height series $\{X_1\}$ 263
Fig. 9.2 Original water level height series $\{X_2\}$ 263
Fig. 9.3 Fluctuation state of water level series $\{X_1\}$ 264
Fig. 9.4 Fluctuation state of water level series $\{X_2\}$ 264
Fig. 9.5 Modeling flowchart of Naive Bayesian classification predictor model 266
Fig. 9.6 Fluctuation trend prediction result of original water level series $\{X_1\}$ 268
Fig. 9.7 Fluctuation trend prediction result of original water level series $\{X_2\}$ 269
Fig. 9.8 Modeling flowchart of the Elman water level prediction model 269
Fig. 9.9 The forecasting results of water level time series $\{X_1\}$ by the Elman 271
Fig. 9.10 The forecasting results of water level time series $\{X_2\}$ by the Elman 272
Fig. 9.11 Modeling flowchart of the Elman water level prediction model under decomposition framework 273
Fig. 9.12 MODWT results of the original water level series $\{X_1\}$ 274
Fig. 9.13 The EMD results of the original water level series $\{X_1\}$ 275
Fig. 9.14 The SSA results of the original water level series $\{X_1\}$ 276
Fig. 9.15 The forecasting results of water level series $\{X_1\}$ by the MODWT-Elman 278
Fig. 9.16 The forecasting results of water level series $\{X_1\}$ by the EMD-Elman 278
Fig. 9.17 The forecasting results of water level series $\{X_1\}$ by the SSA-Elman 279
Fig. 9.18 Forecasting performance indices of different decomposition layer of the MODWT 281
Fig. 9.19 Forecasting performance indices of different mother wavelet of the MODWT 282
Fig. 9.20 Forecasting performance indices of different types of mother wavelet of the MODWT 282
Fig. 9.21 Forecasting performance indices of different window length of the SSA 283
Fig. 9.22 Forecasting results of water level series $\{X_1\}$ by optimal models in 1-step 284

Fig. 9.23 Forecasting results of water level series $\{X_1\}$ by optimal models in 2-step 285
Fig. 9.24 Forecasting results of water level series $\{X_1\}$ by optimal models in 3-step 285

Fig. 10.1 The general framework of the chapter 292
Fig. 10.2 Original public noise series 293
Fig. 10.3 Original neighborhood noise series 293
Fig. 10.4 Original traffic noise series 294
Fig. 10.5 Forecasting results of the RF model for D1 296
Fig. 10.6 Forecasting results of the RF model for D2 297
Fig. 10.7 Forecasting results of the RF model for D3 297
Fig. 10.8 Forecasting results of the BFGS model for D1 300
Fig. 10.9 Forecasting results of the BFGS model for D2 301
Fig. 10.10 Forecasting results of the BFGS model for D3 301
Fig. 10.11 The structure of the GRU 303
Fig. 10.12 The framework of the GRU in the section 303
Fig. 10.13 Forecasting results of the GRU model for D1 305
Fig. 10.14 Forecasting results of the GRU model for D2 306
Fig. 10.15 Forecasting results of the GRU model for D3 306
Fig. 10.16 The framework of Spark-RF model 309
Fig. 10.17 Forecasting results of proposed models for D1 311
Fig. 10.18 Forecasting results of proposed models for D2 311
Fig. 10.19 Forecasting results of proposed models for D3 312

List of Tables

Table 2.1 Statistical characteristics of the original meteorological factors series 29
Table 2.2 Statistical characteristics of the original system load factors series 31
Table 2.3 Statistical characteristics of the original thermal perturbation indoors series 32
Table 2.4 Cross-correlation coefficient based on MI 34
Table 2.5 Cross-correlation coefficient based on Pearson coefficient. 36
Table 2.6 Cross-correlation coefficient based on Kendall coefficient. 38
Table 2.7 Statistical characteristics of the original electrical characteristics series 40
Table 2.8 Spearman coefficient of feature selection 42
Table 2.9 Correlation coefficient of 8 groups of variables 43
Table 2.10 Calculation result based on forward selection search 44
Table 2.11 Calculation result based on backward elimination search. 44

Table 3.1 Cross-correlation coefficient of different regions based on Pearson 55
Table 3.2 Cross-correlation coefficient of different regions based on Kendall 55
Table 3.3 Cross-correlation coefficient of different regions based on Spearman 55
Table 3.4 Correlation coefficient characteristics of ARIMA model 62
Table 3.5 The forecasting performance indices of the ARIMA model for series 1 64
Table 3.6 The forecasting performance indices of the ARIMA model for series 2 64
Table 3.7 The forecasting performance indices of the ARIMA-ARCH model in 1-step 72
Table 3.8 The forecasting performance indices of the ARIMA-ARCH model in 2-step 73

Table 3.9 The forecasting performance indices of the ARIMA-ARCH model in 3-step 73
Table 3.10 The forecasting performance indices of the SARIMA model for series 1 80
Table 3.11 The forecasting performance indices of the SARIMA model for series 2 81
Table 3.12 The forecasting performance comparison of different model for series 1 86
Table 3.13 The forecasting performance comparison of different model for series 2 86

Table 4.1 The change of structure materials 107
Table 4.2 Output after modeling 110
Table 4.3 Output after room temperature calculation 111
Table 4.4 Output after room load calculation 111
Table 4.5 Output after calculation of building daily shadow 112
Table 4.6 Output after calculation of building illumination 112
Table 4.7 Output after calculation of natural ventilation. 113
Table 4.8 Output after calculation of air-conditioning scheme 113
Table 4.9 Output after calculation of wind network 113
Table 4.10 Output after calculation of AHC simulation 113
Table 4.11 The forecasting performance indices of the SVM model 117

Table 5.1 Statistical characteristics of the original series 131
Table 5.2 The forecasting performance indices of the ELM (single data) model 137
Table 5.3 The forecasting performance indices of the BPNN (single data) model 139
Table 5.4 The forecasting performance indices of the ELM (multiple data) model 143
Table 5.5 The forecasting performance indices of the BPNN (multiple data) model 146
Table 5.6 The forecasting performance indices of EWT-ELM model 151
Table 5.7 The forecasting performance indices of EWT-BPNN model 153
Table 5.8 The comprehensive forecasting performance indices of the ELM models 154
Table 5.9 The comprehensive forecasting performance indices of the BP models 155

Table 6.1 The mathematical statistical information of original series 162
Table 6.2 The performance estimating results of the BP prediction model 167
Table 6.3 The performance estimating results of the interval prediction model 174
Table 6.4 The performance estimating results of the WD-BP prediction model 184

Table 6.5 The performance estimating results of the WD-BP-GARCH prediction model 188
Table 6.6 The performance estimating results of the deterministic prediction models 189
Table 6.7 The performance estimating results of the interval prediction model 190

Table 7.1 The specific meanings of Rd1, Rd2, and Rd3 199
Table 7.2 The performance estimating results of the Elman prediction model 205
Table 7.3 The performance estimating results of the LSTM prediction model 209
Table 7.4 The performance estimating results of the WPD-prediction model 217
Table 7.5 The performance estimating results of the involved prediction models 221

Table 8.1 Statistical characteristics of the original series 231
Table 8.2 The forecasting performance indices of the ELM (single data) model 235
Table 8.3 The forecasting performance indices of the ELM (multiple data) model 238
Table 8.4 Identification results of PCA 242
Table 8.5 The identification results of the Gaussian KPCA 246
Table 8.6 Identification results of the FA 250
Table 8.7 The forecasting performance indices under feature extraction framework 253
Table 8.8 The comprehensive forecasting performance indices of proposed models 258

Table 9.1 Statistical characteristics of the original water level series 265
Table 9.2 The forecasting performance indices of the Elman model 273
Table 9.3 The forecasting performance indices of three hybrid models under decomposition framework 279
Table 9.4 The forecasting performance indices of optimal hybrid models 286

Table 10.1 Statistical characteristics of the original series 294
Table 10.2 The forecasting performance indices of the RF model 298
Table 10.3 The forecasting performance indices of the BFGS model 302
Table 10.4 The forecasting performance indices of the GRU model 307
Table 10.5 The comprehensive forecasting performance indices of proposed models 312

Abbreviations

AHU	Air handling unit
AIC	Akaike information criterion
ANFIS	Adaptive network-based fuzzy inference system
ANN	Artificial neural network
AQI	Air quality index
AR	Autoregressive
ARCH	Autoregressive conditional heteroscedasticity
ARCH-LM	Autoregressive conditional heteroscedasticity-Lagrange multiplier
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
BEMPs	Building energy modeling programs
BFGS	which is made up of the initials C. G. Broyden, R. Fletcher, D. Goldfarb, and D. F. Shanno
BIC	Bayesian information criterion
BM	Bayesian model
BP	Back propagation
RFID	Radio frequency identification
BPNN	Back propagation neural network
CA	Cellular automata
CAD	Computer-aided design
CDH	Cloudera's distribution including apache hadoop
CFS	Correlation-based feature selection
CNN	Convolutional neural network
CWC	Coverage width-based criterion
CWT	Continuous wavelet transform
DAG	Directed acyclic graph
DOE	Department of energy
DWT	Discrete wavelet transform
ELM	Extreme learning machine
EMD	Empirical mode decomposition
ERM	Empirical risk minimization

ES	Exponential smoothing
ESP	Energy simulation program
EWT	Empirical wavelet transform
FA	Factor analysis
FITNETs	Function fitting neural networks
FTS	Fuzzy time series
GA	Genetic algorithm
GARCH	Generalized autoregressive conditional heteroscedasticity
GFS	Google file system
GIS	Geographic information system
GPS	Global positioning system
GRU	Gated recurrent units
GWO	Grey wolf optimization
HDFS	Hadoop distributed file system
HDP	Hortonworks data platform
HHT	Hilbert–Huang transform
HMM	Hidden Markov model
HVAC	Heating, ventilating, and air-conditioning
HVACSIM	HVAC simulation
IMF	Intrinsic mode function
IOT	Internet of things
ITS	Intelligent transportation system
KMO	Kaiser-Meyer-Olkin
KPCA	Kernel principal component analysis
LSTM	Long short-term memory
MA	Moving average
MAE	Mean absolute error
MAPE	Mean absolute percent error
MI	Mutual information
MODWT	Maximal overlap discrete wavelet transform
NARX	Nonlinear autoregressive with external input
PCA	Principal component analysis
PICP	Prediction interval coverage probability
PINAW	Prediction interval normalized average width
PJM	An US regional transmission organization named Pennsylvania–New Jersey–Maryland
PSO	Particle swarm optimization
RBF	Radial basis function
RF	Random forest
RMSE	Root mean square error
SD	Seasonal decomposition
SDE	Standard deviation of error
SPARK	Simulation problem analysis and research kernel
SQL	Structured query language
SRM	Structural risk minimization

SSA	Singular spectrum analysis
SVM	Support vector machine
SVR	Support vector regression
TRNSYS	Transient system simulation tool
WD	Wavelet decomposition
WPD	Wavelet packet decomposition
YARN	Yet another resource negotiator