This book targets at publishing high-quality research, surveys, and case studies in the space of telecommunication networks and markets, leveraging data mining, statistics analysis, data analytics, and machine learning. The articles represented in this book cover a wide variety of topics related to telecommunication and network, big data, machine learning and date mining from a diversity of disciplinary viewpoints, including computer science, electrical engineering, statistics, and management. The studies in this book can be beneficial to researchers and practitioners for solving complex problems and synthesizing knowledge.

In Chapter 1, “Detecting Abnormal Traffic in Wireless Networks Using Unsupervised Models”, Alexis Huet discusses detection of abnormal traffic from continuous stream of data over wireless network, and leverages unsupervised learning method to describe structure of data and deduce patterns of the wireless network. The author identifies unexpected behaviors and detects new type of anomalies by combining traffic data with time stamps information and analyzing traffic in wireless network through different unsupervised models. He presents a new analysis model and compares it with other methods. The preliminary results show that the new model can provide a robust and early detection of anomalies as a fully automatic and data-driven solution.

Chapter 2 is titled “Evaluating Wireless Network Accessibility Performance via Clustering-Based Model” by Yan Wang and Zhensen Wu. The authors present a novel analytic methodology to evaluate the performance of wireless network accessibility through traffic measurement, network cells clustering, and regression models. The model leads to identification of where the accessibility remains high and when mobile network becomes congested.

In Chapter 3, “Modeling for Time Generating Network: An Advanced Bayesian Model”, Yirui Hu discusses modeling co-occurrence data in anomaly detection. The author presents and compares the Gaussian probabilistic latent semantic analysis (GPLSA) model to a Gaussian Mixture Model (GMM) using temporal network data. Simulation results indicate that this model performs very well compared with other well-known models.
In Chapter 4, “Identifying Dissatisfied 4G Customers from Network Indicators: A Comparison Between Complaint and Survey Data”, Xinling Dai analyzes feedback data directly collected from mobile users, to identify network problems which lead to customer dissatisfaction, and further provide references for the mobile operators to improve procedures for collection feedback answers. The author utilizes complaints data and survey data to create a classification model, and presents a quantitative methodology to predict dissatisfied users.

Chapter 5 is written by Mantian (Mandy) Hu with The Chinese University of Hong Kong in Hong Kong. The paper presents a targeting strategy for telecommunication companies to facilitate the adoption of 4G technology. The author applies machine learning technique and leverages information about individual’s local network to create a prediction model for consumer adoption behavior. The author implements and verifies the model in real-world scenario.

Chapter 6, “Mining of Leaders in Mobile Telecom Social Networks”, is also written by Mantian (Mandy) Hu from The Chinese University of Hong Kong. Her study seeks to exploit social network data collected by telecom operators and mine value from them. The author presents a marketing strategy to help telecom operators to improve marketing efficiency based on social network. This method utilizes influence of mutual peers in social network and identifies the influential users (leaders) who promote the information diffusion prominently.

In “Network-Based Targeting: Big Data Application in Mobile Industry”, Chapter 7, Chu Dang investigates targeting in mobile industry to understand how network effects and consumer behaviors- such as churning and adopting- interplays with each other. The author provides not only descriptions of churn prediction models, traditional innovation diffusion model and agent-based models, but also various applications implemented on them.

In Chapter 8, “Anomaly Detection in Wireless Networks: An introduction to Multi-Cluster Technique”, Yirui Hu proposes a novel quantitative algorithm of anomaly detection in wireless network based on model log-likelihood. The author applies multi-cluster based analysis to obtain the insights of human behaviors, and learn similar patterns in temporal traffic data, and further leverages the algorithm to automatically characterize the nature of traffic behavior and detect anomalies from normal behaviors.

In Chapter 9, “Continuous-Time Markov Chain-Based Reliability Analysis for Future Cellular Networks”, Hasan Farooq, Md Salik Parwez, and Ali Imran design a stochastic analytical model to analyze the effects of faults in a cellular network, and reduce network recovery time for improving the network reliability. The authors model the reliability behavior of a base station, and propose an adaptive fault predictive framework which can adapt the analytical model by dynamically learning from past database of failures. Numerical results are evaluated to demonstrate the applicability of the analytical model.
Md Salik Parwez, Hasan Farooq, Ali Imran, and Hazem Refai present a novel scheme for spectral efficiency optimization through users clustering in their article in Chapter 10, “Spectral Efficiency Self-Optimization Through Dynamic User Clustering and Beam Steering”. They decompose large-scale system-wide optimization problem into small-scale local sub-problems and thus provide a low complexity solution for dynamic system wide optimization. Every sub-problem involves clustering of users to determine focal point of the cell for given user distribution in time and space, and determining new values of azimuth and tilt that would optimize the overall system spectral efficiency performance. They propose three algorithms of user clustering to transform a given user distribution into the focal points that can be used in optimization: the algorithm based on received signal to interference ratio (SIR) at the user, the algorithm based on received signal level at the user, and the one based on relative distances of users from the base stations.

We would like to thank Mr. Zhongyuan Li and Ms. Claudia Woo, as well as the many authors and reviewers for their contribution to the articles in this book. The articles in this text span a great deal more of cutting edge areas that are truly interdisciplinary in nature. We hope that you will enjoy reading this book and find the articles informative.