

SpringerBriefs in Computer Science

Series editors

Stan Zdonik, Brown University, Providence, Rhode Island, USA

Shashi Shekhar, University of Minnesota, Minneapolis, Minnesota, USA

Xindong Wu, University of Vermont, Burlington, Vermont, USA

Lakhmi C. Jain, University of South Australia, Adelaide, South Australia, Australia

David Padua, University of Illinois Urbana-Champaign, Urbana, Illinois, USA

Xuemin Sherman Shen, University of Waterloo, Waterloo, Ontario, Canada

Borko Furht, Florida Atlantic University, Boca Raton, Florida, USA

V. S. Subrahmanian, University of Maryland, College Park, Maryland, USA

Martial Hebert, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Katsushi Ikeuchi, University of Tokyo, Tokyo, Japan

Bruno Siciliano, Università di Napoli Federico II, Napoli, Italy

Sushil Jajodia, George Mason University, Fairfax, Virginia, USA

Newton Lee, Institute for Education, Research, and Scholarships in Los Angeles, California, USA

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

More information about this series at <http://www.springer.com/series/10028>

Yuan Yao • Xing Su • Hanghang Tong

Mobile Data Mining

 Springer

Yuan Yao
State Key Laboratory for Novel Software
Nanjing University
Nanjing, China

Xing Su
Graduate Center
City University of New York
New York, NY, USA

Hanghang Tong
Arizona State University
Tempe, AZ, USA

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-3-030-02100-9 ISBN 978-3-030-02101-6 (eBook)
<https://doi.org/10.1007/978-3-030-02101-6>

Library of Congress Control Number: 2018958912

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

We have witnessed a fast-moving technological revolution due to the emergence of powerful smartphones. Smartphones are no longer limited to a texting-calling device, but a personal intelligent assistant with increasing abilities in sensing, computing, and networking. To make full use of this intelligent assistant, various data are collected and analyzed to provide better services for the end users.

In this book, we introduce the essential steps for mobile data mining tasks, including data collection (Chap. 2), feature engineering (Chap. 3), and learning models (Chaps. 4–6). We also discuss some key challenges and possible solutions during the introduction of each step. Overall, this book can serve as a primer for beginners to gain a big picture of mobile data mining. It also covers some useful information for further in-depth research on the topic.

Nanjing, China
New York, NY, USA
Tempe, AZ, USA
Aug 28, 2018

Yuan Yao
Xing Su
Hanghang Tong

Contents

- 1 Introduction** 1
 - 1.1 Background..... 1
 - 1.2 Typical Applications 2
 - 1.3 Steps, Characteristics, and Challenges 3
 - 1.4 Roadmap..... 5
- 2 Data Capturing and Processing** 7
 - 2.1 Smartphone Sensors 7
 - 2.2 Data Collection..... 8
 - 2.3 Data Denoising 12
 - 2.4 Summary 16
- 3 Feature Engineering** 17
 - 3.1 Data Segmentation 17
 - 3.2 Feature Extraction..... 18
 - 3.3 Feature Analysis and Sensor Selection..... 19
 - 3.4 Summary 23
- 4 Hierarchical Model** 25
 - 4.1 Problem Description 25
 - 4.2 A Hierarchical Framework 26
 - 4.3 Experimental Evaluations 28
 - 4.4 Summary 30
- 5 Personalized Model** 31
 - 5.1 Problem Description 31
 - 5.2 The Personalized Approach: Overview 32
 - 5.3 The Personalized Approach: Details 34
 - 5.3.1 Similarity Computation..... 34
 - 5.3.2 Distribution Estimation..... 35
 - 5.3.3 Sample Selection..... 36
 - 5.3.4 Sample Reweighting..... 36
 - 5.3.5 Algorithm Analysis 37

- 5.4 Experimental Evaluations 38
 - 5.4.1 Experiment Setup 39
 - 5.4.2 Experiment Results 39
- 5.5 Summary 41
- 6 Online Model 43**
 - 6.1 Problem Description 43
 - 6.2 Online Learning 44
 - 6.3 Experimental Evaluations 45
 - 6.3.1 Online Learning vs. Offline Learning..... 46
 - 6.3.2 How Much Data Are Enough to Train an Initial Model? 49
 - 6.3.3 Combining with the Hierarchical Model 49
 - 6.4 Summary 50
- 7 Conclusions 51**
 - 7.1 Summary 51
 - 7.2 Discussions 52
 - 7.2.1 More Combinations of Sensors 52
 - 7.2.2 More Usages of Smartphone Sensors 53
- References..... 55**

Acronyms

ARA	Average Resultant Acceleration
DFT	Discrete Fourier Transform
ETL	Extraction, Transformation, and Loading
FFT	Fast Fourier Transform
GMS	Global System for Mobile Communications
GPS	Global Positioning System
KMM	Kernel Mean Matching
Pegasos	Primal Estimated sub-GrAdient SOLver
SGD	Stochastic Gradient Descent
SMA	Signal-Magnitude Area
SVM	Support Vector Machine