

SpringerBriefs in Computer Science

Di Wu



Robust Latent Feature Learning for Incomplete Big Data

 Springer

SpringerBriefs in Computer Science

Series Editors

Stan Zdonik, Brown University, Providence, RI, USA

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

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

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

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

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

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

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

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

Katsushi Ikeuchi, University of Tokyo, Tokyo, Japan

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

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

Newton Lee, Institute for Education, Research and Scholarships, Los Angeles, CA, 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.

**Indexing: This series is indexed in Scopus, Ei-Compendex, and zbMATH **

Di Wu

Robust Latent Feature Learning for Incomplete Big Data



Springer

Di Wu
College of Computer and Information Science
Southwest University
Chongqing, China

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-981-19-8139-5 ISBN 978-981-19-8140-1 (eBook)
<https://doi.org/10.1007/978-981-19-8140-1>

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

This work is subject to copyright. All rights are solely and exclusively licensed 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, expressed 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 Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

In the era of Big Data, high dimensional and incomplete (HDI) data are frequently encountered in many industrial applications, such as recommender systems, the Internet of Things, intelligent transportation, cloud computing, and so on. It is of great significance to analyze them for mining rich and valuable knowledge and patterns. Latent feature learning is one of the most popular representation learning methods tailored for HDI data due to its high accuracy, computational efficiency, and ease of scalability. The crux of analyzing HDI data lies in addressing the uncertainty problem caused by their incomplete characteristics. However, existing HDI methods do not fully consider such uncertainty.

In this book, I introduce several robust latent feature learning methods to address such uncertainty for effectively and efficiently analyzing HDI data, including robust latent feature learning based on smooth L_1 -norm, improving robustness of latent feature learning using L_1 -norm, improving robustness of latent feature learning using double-space, data-characteristic-aware latent feature learning, posterior-neighborhood-regularized latent feature learning, and generalized deep latent feature learning.

This is the first book that can help researchers and engineers fully understand how to employ robust latent feature learning to effectively and efficiently analyze HDI data. It is assumed that readers have a basic knowledge of mathematics, as well as a certain background in data mining. Readers can obtain an overview of the challenges of analyzing HDI data. In addition, this book provides several algorithms and real application cases, which can help readers quickly build their robust latent feature learning models to analyze HDI data.

Chongqing, China
September 2022

Di Wu

Acknowledgments

The work described in this book was supported by the National Natural Science Foundation of China under grant 62176070.

Contents

1	Introduction	1
1.1	Background	1
1.2	Symbols and Notations	2
1.3	Book Organization	3
	References	4
2	Basis of Latent Feature Learning	7
2.1	Overview	7
2.2	Preliminaries	8
2.3	Latent Feature Learning	8
2.3.1	A Basic LFL Model	8
2.3.2	A Biased LFL Model	12
2.3.3	Algorithms Design	13
2.4	Performance Analysis	13
2.4.1	Evaluation Protocol	13
2.4.2	Discussion	14
2.5	Summary	15
	References	17
3	Robust Latent Feature Learning based on Smooth L_1-norm	19
3.1	Overview	19
3.2	Related Work	20
3.3	A Smooth L_1 -Norm Based Latent Feature Model	21
3.3.1	Objective Formulation	21
3.3.2	Model Optimization	22
3.3.3	Incorporating Linear Biases into SL-LF	23
3.4	Performance Analysis	24
3.4.1	General Settings	24
3.4.2	Performance Comparison	25
3.4.3	Outlier Data Sensitivity Tests	27
3.4.4	The Impact of Hyper-Parameter	28

3.5	Summary	29
	References	30
4	Improving Robustness of Latent Feature Learning Using L_1-Norm	33
4.1	Overview	33
4.2	Related Work	34
4.3	An L_1 -and- L_2 -Norm-Oriented Latent Feature Model	35
	4.3.1 Objective Formulation	35
	4.3.2 Model Optimization	36
	4.3.3 Self-Adaptive Aggregation	37
4.4	Performance Analysis	39
	4.4.1 General Settings	39
	4.4.2 L^3F 's Aggregation Effects	40
	4.4.3 Comparison Between L^3F and Baselines	41
	4.4.4 L^3F 's Robustness to Outlier Data	41
4.5	Summary	43
	References	44
5	Improve Robustness of Latent Feature Learning Using Double-Space	47
5.1	Overview	47
5.2	Related Work	49
5.3	A Double-Space and Double-Norm Ensembled Latent Feature Model	49
	5.3.1 Predictor Based on Inner Product Space (D^2E -LF-1)	50
	5.3.2 Predictor on Euclidean Distance Space (D^2E -LF-2)	53
	5.3.3 Ensemble of D^2E -LF-1 and D^2E -LF-2	56
	5.3.4 Algorithm Design and Analysis	57
5.4	Performance Analysis	57
	5.4.1 General Settings	57
	5.4.2 Performance Comparison	59
5.5	Summary	59
	References	62
6	Data-characteristic-aware Latent Feature Learning	67
6.1	Overview	67
6.2	Related Work	68
	6.2.1 Related LFL-Based Models	68
	6.2.2 DPCLust Algorithm	68
6.3	A Data-Characteristic-Aware Latent Feature Model	70
	6.3.1 Model Structure	70
	6.3.2 Step 1: Latent Feature Extraction	71
	6.3.3 Step 2: Neighborhood and Outlier Detection	71
	6.3.4 Step 3: Prediction	73

- 6.4 Performance Analysis 76
 - 6.4.1 Prediction Rule Selection 76
 - 6.4.2 Performance Comparison 77
- 6.5 Summary 78
- References 81
- 7 Posterior-neighborhood-regularized Latent Feature Learning 85**
 - 7.1 Overview 85
 - 7.2 Related Work 86
 - 7.3 A Posterior-Neighborhood-Regularized Latent Feature Model 87
 - 7.3.1 Primal Latent Feature Extraction 87
 - 7.3.2 Posterior-Neighborhood Construction 88
 - 7.3.3 Posterior-Neighborhood-Regularized LFL 89
 - 7.4 Performance Analysis 91
 - 7.4.1 General Settings 91
 - 7.4.2 Comparisons Between PLF and State-of-the-Art Models 91
 - 7.5 Summary 93
 - References 93
- 8 Generalized Deep Latent Feature Learning 97**
 - 8.1 Overview 97
 - 8.2 Related Work 98
 - 8.3 A Deep Latent Feature Model 98
 - 8.4 Performance Analysis 101
 - 8.4.1 General Settings 101
 - 8.4.2 Effects of Layer Count in DLF 101
 - 8.4.3 Comparison Between DLF and Related Models 102
 - 8.5 Summary 103
 - References 106
- 9 Conclusion and Outlook 111**
 - 9.1 Conclusion 111
 - 9.2 Outlook 111

About the Author

Di Wu received a Ph.D. degree in Computer Science from Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China, in 2019. He was a visiting scholar from April 2018 to April 2019 at the University of Louisiana, Lafayette, USA. Currently, he is a Professor at the College of Computer and Information Science, Southwest University. His current research interests include data mining, artificial intelligence, and big data. He has published over 50 papers, including 12 IEEE TRANSACTIONS papers, three highly cited paper of ESI, and several top-tier conferences such as AAAI, ICDM, WWW, and IJCAI. His Google Scholar citations are more than 1800, and his H-Index is 23. He is an Associate Editor for Frontiers in Neurorobotics (SCI, IF 3.493). He received the Nomination Award for Excellent Doctoral Dissertation of the Chinese Association for Artificial Intelligence (CAAI).