Springer Theses

Recognizing Outstanding Ph.D. Research

Aims and Scope

The series "Springer Theses" brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student's supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today's younger generation of scientists.

Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

More information about this series at http://www.springer.com/series/8790

Zechao Li

Understanding-Oriented Multimedia Content Analysis

Doctoral Thesis accepted by
University of Chinese Academy of Sciences, Beijing, China

Springer

Author
Dr. Zechao Li
Nanjing University of Science and Technology
Nanjing
China

Supervisor
Prof. Hanqing Lu
National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences
Haidian District, Beijing
China

ISSN 2190-5053 ISSN 2190-5061 (electronic)
Springer Theses
ISBN 978-981-10-3688-0 ISBN 978-981-10-3689-7 (eBook)
DOI 10.1007/978-981-10-3689-7

Library of Congress Control Number: 2017939539

Printed on acid-free paper

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

This work is dedicated to my parents, and all my friends. Their support and encouragement keep me forward.

Supervisor's Foreword

Multimedia content analysis has attracted extensive attention in the multimedia and social media research communities. Its goal is to reveal the semantic information intelligently. Zechaos' Ph.D. work focuses on understanding-oriented multimedia content analysis from the low-level visual representation to the high-level semantic understanding. As a key member of my group, he made a number of significant contributions in his research work. He investigated advanced multimedia content analysis approaches and proposed understanding-oriented multimedia content analysis approaches, including data representation (feature selection and feature extraction), tag recommendation, and multimedia news services. He directly integrated the visual understanding and learning models into a unified framework. The visual understanding guides the model learning while the learned models improve the visual understanding. The inspiring idea of understanding-oriented multimedia content analysis has been recognized as opening up possibilities to challenging multimedia content and context understanding. The proposed structured subspace learning framework has been successfully generalized to social image understanding, (semi-) supervised classification and clustering. His work has brought in new thoughts and disruptive models in understanding multimedia data. I believe that this book will benefit researchers and students conducting research on multimedia computing and social multimedia analysis.

Beijing, China
January 2017

Prof. Hanqing Lu

Preface

The amount of today's multimedia contents explosively grows due to the popularization and rapid growth of digital mobile devices and social media tools. To efficiently analyze and understand the multimedia content is still a challenging task. Over the past decade, many advanced methods have been proposed in the literature, including a few books on this topic. However, there is no book offering a systematic introduction to multimedia content analysis towards an understanding-oriented approach. Therefore, this book will focus on a novel "understanding" framework for multimedia content interpretation. This book offers a systematic introduction to multimedia content analysis towards an understanding-oriented approach. It integrates the visual understanding and learning models into a unified framework, within which the visual understanding guides the model learning while the learned models improve the visual understanding. More specifically, the book presents multimedia content representations and analysis including feature selection, feature extraction, image tagging, user-oriented tag recommendation, and understanding-oriented multimedia applications. By providing the fundamental technologies and the state-of-the-art methods, this book will be of interest to graduate students and researchers working in the field computer vision and machine learning.

Chapter 1 introduces the background, challenges, and progresses of understanding-oriented multimedia content analysis. Chapters 2 and 3 introduce some works of understanding-oriented data representation. The personalized tag recommendation work is detailed in Chap. 4, followed by understanding-oriented multimedia news services in Chaps. 5 and 6. Chapter 7 concludes the book by summarizing the major points and identifying the future works.

Nanjing, China
January 2017

Zechao Li

Parts of this book have been published in the following articles:

- Zechao Li, Jing Liu, Jinhui Tang, Hanqing Lu. Robust Structured Subspace Learning for Data Representation. IEEE Trans. on Pattern Analysis and Machine Intelligence 37(10): 2085–2098, 2015.
- Zechao Li, Jinhui Tang. Unsupervised Feature Selection via Nonnegative Spectral Analysis and Redundancy Control. IEEE Trans. on Image Processing 24(12): 5343–5355, 2015.
- Zechao Li, Jing Liu, Yi Yang, Xiaofang Zhou, Hanqing Lu. Clustering-Guided Sparse Structural Learning for Unsupervised Feature Selection. IEEE Trans. Knowledge and Data Engineering 26(9): 2138–2150, 2014.
- Jing Liu, Zechao Li, Jinhui Tang, Yu Jiang and Hanqing Lu. Personalized Geo-Specific Tag Recommendation for Photos on Social Websites. IEEE Trans. on Multimedia 16(3): 588–600, 2014.
- Zechao Li, Jinhui Tang, Xueming Wang, Jing Liu, Hanqing Lu. Multimedia News Summarization in Search. ACM Trans. on Intelligent Systems and Technology 7(3): 33 (1–20), 2016.
- Zechao Li, Jing Liu, Meng Wang, Changsheng Xu, Hanqing Lu. Enhancing News Organization for Convenient Retrieval and Browsing. ACM Trans. on Multimedia Computing, Communications and Applications 10(1): 1 (1–20), 2013.
- Zechao Li, Jing Liu, Xiaobin Zhu and Hanqing Lu: Multi-modal Multi-correlation Person-centric News Retrieval. In ACM Conference on Information and Knowledge Management, 2010.

Acknowledgements

First of all, I would like to express my gratitude to my supervisor, Prof. Hanqing Lu, for his long-term support and help. Professor Lu provided a well-equipped and active working environment for me and gave me full freedom to investigate any research problem of interest. His valuable suggestions and criticism play a significant role on my way toward a full-fledged researcher.

I would also thank Prof. Changsheng Xu (IEEE Fellow), for his guidance during my visiting study at China-Singapore Institute of Digital Media (CSIDM). Professor Jing Liu is another guider of my research. She gave me many pieces of advice in my research methodologies and her revision on my papers that I achieved remarkable progress. Her passion for academic research is of great importance to my decision to start my research career. I also thank Prof. Jian Cheng, Jinqiao Wang, Yifan Zhang, Meng Wang, Jinhui Tang, and Richang Hong. They spent a lot of time with me and provided me with a great deal of assistance. Much of my academic inspiration stems from discussions with them.

In addition, thanks are due to my colleagues in Image and Video Analysis (IVA) lab, including Chao Liang, Yang Liu (Male), Chunjie Zhang, Bo Wang, Chuanghua Gui, Xiao Yu Zhang, Si Liu, Tianzhu Zhang, Xiaobin Zhu, Peng Li, Yu Jiang, Yang Liu (Female), Wei Fu, Biao Niu, Yong Li, Cong Leng, Ting Yuan, among others, my colleagues in MultiMedia Computing (MMC) group, including Jitao Sang, Weiqing Min, Lei Yu, Zhaoquan Yuan, Ming Yan, among others. They helped me a lot during these years in many aspects. I will never forget the joyful and rewarding days spent with them. I am also grateful to all my good friends, wherever they are. My thanks go to everyone who contributed to my progress and happiness.

This book was partially supported by the 973 Program of China (Project No. 2014CB347600), and the National Natural Science Foundation of China under Grant No. 61402228.

Contents

Notations

Throughout this book, the lowercase italic letters (i.e., i, j, n, etc.) and the uppercase italic letters (i.e., A, B, M, etc.) denote scalars, while the bold uppercase characters (i.e, $\mathbf{W}$, $\mathbf{X}$, etc.) and the bold lowercase characters (i.e, $\mathbf{a}$, $\mathbf{x}$, etc.) are utilized to denote matrices and vectors, respectively. For any matrix $\mathbf{A}$, $\mathbf{a}^i$ means the i-th column vector of $\mathbf{A}$, $\mathbf{a}_i$ means the i-th row vector of $\mathbf{A}$, A_{ij} denotes the (i,j)-element of $\mathbf{A}$ and $\mathrm{Tr}[\mathbf{A}]$ is the trace of $\mathbf{A}$ if $\mathbf{A}$ is square. $\mathbf{A}^T$ denotes the transposed matrix of $\mathbf{A}$. The Frobenius norm of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is defined as $\|\mathbf{A}\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n A_{ij}^2 = \mathrm{Tr}[\mathbf{A}^T\mathbf{A}]$. The $\ell_{2,p}$-norm ($p \in (0,1]$) of $\mathbf{A}$ is defined as

$$\|\mathbf{A}\|_{2,p} = \left(\sum_{i=1}^{r}\left(\sqrt{\sum_{j=1}^{t} A_{ij}^2}\right)^p\right)^{\frac{1}{p}} = \left(\sum_{i=1}^{r}\|\mathbf{a}_i\|_2^p\right)^{\frac{1}{p}}. \tag{1}$$

Note that in practice, $\|\mathbf{a}^i\|_2$ could be close to zero. For this case, we can follow the traditional regularization way and define $D_{ii} = \frac{1}{\|\mathbf{a}^i\|_2+\varepsilon}$, where ε is very small constant. When $\varepsilon \to 0$, it is easy to verify that $\frac{1}{\|\mathbf{a}^i\|_2+\varepsilon}$ approximates $\frac{1}{\|\mathbf{a}^i\|_2}$. Furthermore, let $\mathbf{I}_m$ denote the identity matrix in $\mathbb{R}^{m\times m}$.