Synthesis Lectures on
Information Concepts, Retrieval, and Services

SYNTHESIS
COLLECTION OF TECHNOLOGY

Lei Zhu · Jingjing Li · Weili Guan

Multi-modal Hash Learning

Efficient Multimedia Retrieval and Recommendations

Springer

Synthesis Lectures on Information Concepts, Retrieval, and Services

Series Editor

Gary Marchionini, School of Information and Library Science, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

This series publishes short books on topics pertaining to information science and applications of technology to information discovery, production, distribution, and management. Potential topics include: data models, indexing theory and algorithms, classification, information architecture, information economics, privacy and identity, scholarly communication, bibliometrics and webometrics, personal information management, human information behavior, digital libraries, archives and preservation, cultural informatics, information retrieval evaluation, data fusion, relevance feedback, recommendation systems, question answering, natural language processing for retrieval, text summarization, multimedia retrieval, multilingual retrieval, and exploratory search.

Lei Zhu · Jingjing Li · Weili Guan

Multi-modal Hash Learning

Efficient Multimedia Retrieval and Recommendations

Springer

Lei Zhu
Shandong Normal University
Jinan, China

Jingjing Li
University of Electronic Science
and Technology of China
Chengdu, China

Weili Guan
Monash University
Sydney, NSW, Australia

ISSN 1947-945X ISSN 1947-9468 (electronic)
Synthesis Lectures on Information Concepts, Retrieval, and Services
ISBN 978-3-031-37290-2 ISBN 978-3-031-37291-9 (eBook)
https://doi.org/10.1007/978-3-031-37291-9

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 Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated to every researcher who works on large-scale multimedia retrieval and recommendation.

Preface I

Heterogeneous multi-modal data are increasing explosively nowadays in the big data era. Multimedia retrieval and recommendation are facing unprecedented challenges on both computation speed and storage cost. The technique of hashing can project high-dimensional data into compact binary hash codes. With hashing, the most time-consuming semantic similarity computation during the multimedia retrieval and recommendation process can be significantly accelerated with fast Hamming distance computation, and meanwhile, the storage cost can be greatly reduced through binary embedding. Hence, multi-modal hashing has recently received considerable attention to support large-scale multimedia retrieval and recommendation.

This book is the first book dedicated to multi-modal hash learning, which learns binary representations in a low-dimensional Hamming space while preserving the heterogeneous multi-modal semantics for large-scale multimedia retrieval and recommendation. Multi-modal hash learning has become one of the promising techniques in recent years to support large-scale multimedia applications and has received great attention in both academia and industry.

This book serves as a systematic introduction to multi-modal hash learning for retrieval and recommendation, including a survey of current developments and the state-of-the-art in this research field. It not only comprehensively covers the key contents and recent advancements of multi-modal hashing, including context-aware hashing, cross-modal hashing, composite multi-modal hashing, and multi-modal discrete collaborative filtering, but also presents the hashing applications in multimedia retrieval and recommendation. Besides, the book provides a platform and practice of multi-modal hash learning. As the first book on this theme, it summarizes the latest developments and presents cutting-edge research on multi-modal hash learning for multimedia retrieval and recommendation. It may provide researchers with an understanding of the important problems and a good entry point for working on this research area.

The authors of this book have contributed substantial research on multi-modal hash learning and related topics. Lei Zhu is one of the leading researchers on multi-modal hashing. Jingjing Li is a long-term collaborator with Lei Zhu on research into multi-modal hashing. Weili Guan is a rising star scholar in multimedia retrieval and recommendation.

The book systematically summarizes their contributions in the direction of multi-modal hash learning for efficient multimedia retrieval and recommendation. It can be used as an excellent self-contained take-off point for beginning researchers in multimedia retrieval.

Jinan, China Lei Zhu
Chengdu, China Jingjing Li
Sydney, Australia Weili Guan

Preface II

In recent years, many multimedia applications such as search engines, social websites, and online shopping platforms have developed at an unprecedented speed. While these network services provide great convenience to our daily life, they generate a large amount of multimedia data, such as text, image, audio, and video. Multimedia data is not only large in quantity, but also complex in structure and diverse in content. These characteristics bring challenges to many research problems and great opportunities. In particular, the demands of users for multimedia data retrieval, content recommendation, and other technologies are increasing day by day. An urgent need is for efficient learning models to organize and manage large-scale multimedia data.

Multi-modal hash learning can encode data from multiple different modalities into compact binary hash codes. It has the desirable advantages of fast retrieval speed and low storage cost, and can effectively support large-scale multimedia retrieval and recommendation. Therefore, multi-modal hash learning has recently gained increasing attention. Multi-modal learning to hash indeed has been well-studied in the past decade. However, multi-modal hashing for multimedia retrieval and recommendation in a big data environment has its unique properties and corresponding challenges, including but not limited to the following points:

(1) Heterogeneous modality gap. Multi-modal data features belong to different representation spaces, it is a challenge to directly build the correlation structures across heterogeneous modalities in the process of multi-modal hash learning.
(2) Ineffective multi-modal modeling. Existing methods usually exploit linear or simple nonlinear functions for multi-modal hash projection. They cannot effectively capture the intrinsic multi-modal data structure, which is important for modeling the multi-modal correlation and semantics.
(3) Inefficient hash optimization process. This problem leads to extremely high time and space complexity of multi-modal hashing in multimedia retrieval and recommendation. Such a large computational cost makes extending existing methods to large-scale scenarios difficult.

(4) Cold-start and explainable recommendation with binary hashing. Existing hashing-based recommendation systems employ user-item interactions and single auxiliary information to learn the binary hash codes. But the full interaction history is not always available and single auxiliary information may be missing. Moreover, existing hashing-based recommendation systems remain black boxes without any explainable outputs that illustrate why the system recommends the items.

In this book, to tackle the above research challenges, we present several state-of-the-art multi-modal hashing learning methods and verify them through extensive experiments. Specifically, we first introduce two context-aware hashing methods for large-scale image retrieval. One approach considers contextual social tags as a kind of semantic resource, and another approach considers exploring semantic information from image structure. We then present two cross-modal hashing learning frameworks to seek the multi-modal complementary space and learn hash functions to support unsupervised and supervised cross-modal retrieval, respectively. Following that, we work toward two composite multi-modal hashing methods. We not only design a self-weighted fusion strategy that adaptively preserves multi-modal feature information into hash codes by exploiting the complementarity of multi-modal features, but we also excavate bit-wise semantic concepts and align the heterogeneous modalities at the concept level for multi-modal hash learning. Thereafter, we introduce two hashing-based multi-modal recommendation methods: multi-modal discrete collaborative filtering and explainable discrete collaborative filtering. We finally conclude the book and present the future research directions in multi-modal hash learning, e.g., deep multi-modal modeling, multi-modal hash learning under open dynamic environment, lightweight hash model design, etc.

This book represents preliminary research on multi-modal hash learning for multimedia retrieval and recommendation. We hope that it can help beginners understand the field, and hope that it can arouse active researchers to work in this exciting field. If, in this book, we have been able to dream further than others have, it is because we are standing on the shoulders of giants.

Jinan, China	Lei Zhu
Chengdu, China	Jingjing Li
Sydney, Australia	Weili Guan
September 2022	

Acknowledgments

This book would not have been completed, or at least not be what it looks like today, without the support of many colleagues, especially those from the Big Media Data Computing Lab at Shandong Normal University. It is a pleasure to take this opportunity to acknowledge them for their contributions to this time-consuming book project. Their contributions have supplied ingredients for insightful discussions related to the writing of this book, and hence we are greatly appreciative.

Our first thanks undoubtedly go to Dr. Hui Cui, Dr. Yang Xu, Mr. Wentao Tan at Shandong Normal University, Mr. Xize Wu at Southeast University, Dr. Xu Lu at Shandong Agricultural University, Dr. Liqiang Nie, and Dr. Zheng Zhang at Harbin Institute of Technology (Shenzhen), Dr. Zhiyong Cheng at Shandong Artificial Intelligence Institute, Dr. Yang Yang at the University of Electronic Science and Technology of China, Dr. Junwei Han at Northwestern Polytechnical University, as well as Dr. Huaxiang Zhang at Shandong Normal University. We consulted with them on some specific technical chapters of the book and they are also the major contributors to some chapters. Their constructive feedback and comments at various stages have been significantly helpful in shaping the book. We also take this opportunity to thank Prof. Heng Tao Shen at the University of Electronic Science and Technology of China who never hesitated to offer his advice and share his valuable experience whenever the authors needed him.

We are grateful to Editor Mrs. Susanne Filler for her great efforts on this book. They also help to make the book published smoothly and enjoyable.

Last, but certainly not least, our thanks go to our beloved families for their selfless consideration, endless love, and unconditional support.

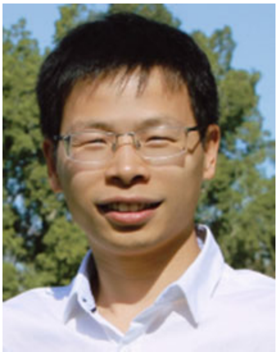
This work was supported in part by the National Natural Science Foundation of China under Grant 62172263, in part by the Natural Science Foundation of Shandong, China, under Grant ZR2020YQ47 and Grant ZR2019QF002, in part by the Youth Innovation Project of Shandong Universities, China, under Grant 2019KJN040.

September 2022

Lei Zhu
Jingjing Li
Weili Guan

Contents

About the Authors



Lei Zhu is currently a professor with the School of Information Science and Engineering, Shandong Normal University. He received his B.Eng. and Ph.D. degrees from Wuhan University of Technology in 2009 and Huazhong University Science and Technology in 2015, respectively. He was a Research Fellow at the University of Queensland (2016–2017). His research interests are in the area of large-scale multimedia content analysis and retrieval. Zhu has co-/authored more than 100 peer-reviewed papers, such as ACM SIGIR, ACM MM, IEEE TPAMI, IEEE TIP, IEEE TKDE, and ACM TOIS. His publications have attracted more than 6,200 Google citations. At present, he serves as the Associate Editor of IEEE TBD, ACM TOMM, and Information Sciences. He has served as the Area Chair of ACM MM/IEEE ICME, Senior Program Committee for SIGIR/CIKM/AAAI. He won ACM SIGIR 2019 Best Paper Honorable Mention Award, ADMA 2020 Best Paper Award, ChinaMM 2022 Best Student Paper Award, ACM China SIGMM Rising Star Award, Shandong Provincial Entrepreneurship Award for Returned Students, and Shandong Provincial AI Outstanding Youth Award.

Jingjing Li is currently a professor at the School of Computer Science and Engineering, University of Electronic Science and Technology of China (UESTC). He received his B.Eng., M.Sc., and Ph.D. degrees from UESTC in 2010, 2013, and 2015, respectively. His research interests are in the area of domain adaptation and zero-shot learning. He has co-authored more than 70 peer-reviewed papers, such as IEEE TPAMI, IEEE TIP, IEEE TKDE, CVPR, ICCV, AAAI, IJCAI, and ACM Multimedia. He won the Excellent Doctoral Dissertation Award from the Chinese Institute of Electronics in 2018.

Weili Guan is currently a final year Ph.D. student with the Faculty of Information Technology, Monash University Clayton Campus, Australia. Her research interests are multimedia computing and information retrieval. She received her bachelor's degree from Huaqiao University. She then obtained her master's degree from the National University of Singapore. After that, she joined Hewlett Packard Enterprise Singapore as a software engineer and worked there for around five years. She has published dozens of papers at the top conferences and journals, like ACM SIGIR, IEEE TIP, and IEEE TPAMI.

Abbreviations

ABinCF	Adversarial Binary Collaborative Filtering framework
ABQ	Adaptive Binary Quantization
ACR	Adjacent Correlation Reconstruction
AGCH	Aggregation-based Graph Convolutional Hashing
AGH	Anchor Graph Hashing
ALM	Augmented Lagrangian Multiplier
AP	Average Precision
ASCSH	Asymmetric Supervised Consistent and Specific Hashing
ATanh	Adaptive Tanh
BATCH	scalaBle AsymmeTric discrete Cross-modal Hashing
BoVW	Bag of Visual Words
BoW	Bag of Words
BP	Back-Propagation
BSTH	Bit-aware Semantic Transformer Hashing
CCA	Canonical Correlation Analysis
CCA-ITQ	ITerative Quantization with Canonical Correlation Analysis
CCQ	Composite Correlation Quantization
CCR	Coding Consistency Reconstruction
CDL	Collaborative Deep Learning
CF	Collaborative Filtering
CIRH	Correlation-Identity Reconstruction Hashing
CMFH	Collective Matrix Factorization Hashing
CM-MAN	Cross-Modal Message Aggregation Network
CNNH	Convolutional Neural Network Hashing
CPAH	Consistency-Preserving Adversarial Hashing
CSA	Cross-modal Semantic Aggregation
CTR	Collaborative Topic Regression
CVH	Cross-View Hashing
DBN	Deep Belief Network
DBRC	Deep Binary ReConstruction

DCC	Discrete Cyclic Coordinate descent
DCD	Discrete Coordinate Descent
DCF	Discrete Collaborative Filtering
DCHUC	Deep Cross-modal Hashing with hashing functions and Unified hash Codes jointly learning
DCMF	Discrete Content-aware Matrix Factorization
DCMH	Deep Cross-Modal Hashing
DCMVH	Deep Collaborative Multi-View Hashing
DDL	Discrete Deep Learning
DeepMF	Deep Matrix Factorization
DFM	Discrete Factorization Machines
DIS	DIScretization
DJSRH	Deep Joint-Semantics Reconstructing Hashing
DMFH	Deep Multiscale Fusion Hashing
DMVH	Discrete Multi-View Hashing
DPSH	Deep Pairwise Supervised Hashing
DSDH	Deep Supervised Discrete Hashing
DSR	Discrete Social Recommendation
DSTDH	Dual-level Semantic Transfer Deep Hashing
DTMF	Discrete Trust-aware Matrix Factorization
EDCF	Explainable Discrete Collaborative Filtering
FastHash	Fast supervised Hashing
FC	Fully Connected layer
FDMH	Flexible Discrete Multi-view Hashing
FGCMH	Flexible Graph Convolutional Multi-modal Hashing
FOMH	Flexible Online Multi-modal Hashing
GAN	Generative Adversarial Network
GCH	Graph Convolutional Hashing
GCN	Graph Convolutional Network
HCG	Hash Code Generation
HMAH	Hierarchical Message Aggregation Hashing
HSS	Hierarchical Sequence-to-Sequence
ICM	Iterated Conditional Modes
IMH	Inter-Media Hashing
IM-MAN	Intra-Modal Message Aggregation Networks
ISR	Identity Semantic Reconstruction
ITQ	ITerative Quantization
JDSH	Joint-modal Distribution-based Similarity Hashing
KSH	Supervised Hashing with Kernels
LAGNH	Lightweight Augmented Graph Network Hashing
LLE	Locally Linear Embedding

LSH	Locality Sensitive Hashing
LSMH	Latent Semantic Minimal Hashing
LSSH	Latent Semantic Sparse Hashing
LSTM	Long Short-Term Memory
MAH	Multi-view Alignment Hashing
MAP	Mean Average Precision
MCGC	Multi-modal Collaborated Graph Construction
MDCF	Multi-modal Discrete Collaborative Filtering
MF	Matrix Factorization
MFDCF	Multi-Feature Discrete Collaborative Filtering
MFH	Multiple Feature Hashing
MFKH	Multiple Feature Kernel Hashing
MGRN	Masked visual semantic Graph-based Reasoning Network
MLP	Multi-Layer Perceptron
MRF	Markov Random Field
MTFH	Matrix Tri-Factorization Hashing
MvDH	Multi-view Discrete Hashing
MVLH	Multi-View Latent Hashing
NDCG	Normalized Discounted Cumulative Gain
NeuHash-CF	Neural Hashing-based Collaborative Filtering
NINH	Network In Network Hashing
NLL	Negative Log Loss
PCAH	Principal Component Analysis Hashing
PMF	Probabilistic Matrix Factorization
SADH	Similarity-Adaptive Deep Hashing
SAH	Semantic-Aware Hashing
SAPMH	Supervised Adaptive Partial Multi-view Hashing
SCADH	SCAlable Deep Hashing
SCRATCH	Scalable disCRete mATrix faCtorization Hashing
SDH	Supervised Discrete Hashing
SDMH	Supervised Discrete Multi-view Hashing
SGH	Scalable Graph Hashing
SH	Spectral Hashing
SKLSH	Locality-Sensitive Hashing with Shift-invariant Kernels
SMFH	Supervised Matrix Factorization Hashing
SMH-OQA	Supervised Multi-modal Hashing with Online Query-Adaption
SOTA	State-Of-The-Art
SRCH	Semantic-Rebased Cross-modal Hashing
SSDH	Semantic Structure-based Deep Hashing
SuperSDH	Supervised Semantics-preserving Deep Hashing
SVD	Singular Value Decomposition

TBH	Twin-Bottleneck Hashing
UDCMH	Unsupervised Deep Cross-Modal Hashing
UH-BDNN	Unsupervised Hashing with Binary Deep Neural Network
UMH-OQA	Unsupervised Multi-modal Hashing with Online Query-Adaption
WDHT	Weakly supervised Deep Hashing using Tag embeddings
WMH	Weakly supervised Multi-modal Hashing
ZSR	Zero-Shot Recommendation