Synthesis Lectures on
Information Concepts, Retrieval, and Services

SYNTHESIS
COLLECTION OF TECHNOLOGY

Weili Guan · Xuemeng Song · Dongliang Zhou · Liqiang Nie

Advanced Multimodal Compatibility Modeling and Recommendation

Third Edition

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, 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.

Weili Guan · Xuemeng Song ·
Dongliang Zhou · Liqiang Nie

Advanced Multimodal Compatibility Modeling and Recommendation

Third Edition

Springer

Weili Guan
School of Electronics and Information Engineering
Harbin Institute of Technology
Shenzhen, Guangdong, China

Dongliang Zhou
School of Electronics and Information Engineering
Harbin Institute of Technology
Shenzhen, Guangdong, China

Xuemeng Song
School of Computer Science and Technology
Shandong University
Qingdao, Shandong, China

Liqiang Nie
Department of Computer Science and Technology
Harbin Institute of Technology
Shenzhen, Guangdong, China

ISSN 1947-945X ISSN 1947-9468 (electronic)
Synthesis Lectures on Information Concepts, Retrieval, and Services
ISBN 978-3-031-81047-3 ISBN 978-3-031-81048-0 (eBook)
https://doi.org/10.1007/978-3-031-81048-0

Originally Published by Morgan & Claypool Publishers, 2020

1st edition: © Morgan & Claypool Publishers 2020
2nd edition: © The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
3rd edition: © The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2025

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

If disposing of this product, please recycle the paper.

Preface

As society evolves, fashion is not merely a means to clothe the body but an essential facet of human culture, identity, and self-expression. Creating visually harmonious outfits has been a daily routine for people in modern society. However, crafting harmonious ensembles involves more than just following the latest trends; it requires a deep understanding of how different fashion elements—such as color, texture, style, and proportion—interact to form a cohesive and aesthetically pleasing look. For many, achieving this harmony can be daunting, given the vast array of choices available in today's fashion market. The emergence of fashion compatibility modeling represents a significant advancement in assisting individuals with navigating this complexity. By leveraging cutting-edge multimodal processing technologies, fashion compatibility modeling provides a systematic approach to evaluating the compatibility of a set of fashion garments and enables fashion recommendation. This book aims to explore these innovative methods, offering insights into how advanced neural networks, knowledge-driven frameworks, and multimodal integration can revolutionize fashion compatibility assessments and recommendations.

Throughout the book, we address the multifaceted challenges inherent in fashion compatibility modeling. These include integrating the category modality of fashion items, understanding the practical implications of try-on appearance, and the need for fine-grained fashion compatibility modeling. We also explore the challenges of incorporating side information, promoting the model efficiency, as well as the challenges involved in generating compatible fashion items. Each challenge is met with a targeted solution, ranging from category-aware attention networks to sophisticated generative models, all designed to enhance our understanding of fashion compatibility. Our exploration is structured across six chapters, each dedicated to a specific aspect of fashion compatibility modeling. From the foundational principles of multimodal compatibility modeling to the advanced techniques of generating multiple compatible items, the book provides a comprehensive overview of the latest developments in this field. By consolidating these methods, we aim to advance both the academic study and practical application of fashion

compatibility modeling, ultimately empowering individuals to make more informed and stylish clothing choices.

We invite you to embark on this journey through the intersection of fashion, technology, and creativity. Whether you are a researcher, a practitioner, or simply a fashion enthusiast, we hope this book will provide valuable insights and inspire new perspectives on the ever-evolving world of fashion.

Shenzhen, China — Weili Guan
Qingdao, China — Xuemeng Song
Shenzhen, China — Dongliang Zhou
Shenzhen, China — Liqiang Nie
October 2024

Acknowledgements

This book would not have been completed, or at least not in its current form, without the invaluable support of many colleagues, especially those from the iLearn Centers at the Harbin Institute of Technology, Shenzhen, and Shandong University. We would like to take this opportunity to express our deepest gratitude for their contributions to this extensive and time-consuming project.

First, we extend our sincere thanks to several colleagues who made significant contributions to specific chapters of this book: Dr. Peiguang Jing from Tianjin University, and Dr. Xue Dong and Dr. Na Zheng from Shandong University. Their participation in technical discussions, along with their constructive feedback and insightful comments, greatly benefited the development of this book.

Second, we are deeply grateful to the anonymous reviewers, who read the manuscript with great care and provided many thoughtful and constructive suggestions. Their input significantly enhanced the quality of the book.

Third, we sincerely thank Springer Nature Publisher, particularly Dr. Gary Marchionini, the editor, Susanne Filler, and Jayarani Premkumar, the production editors, for their invaluable suggestions and support. Their efforts made the publishing process both smooth and enjoyable.

Finally, we reserve our heartfelt thanks for our beloved families, whose selfless consideration, endless love, and unwavering support made this endeavor possible.

Contents

Introduction

1

1.1 Background

The fashion industry is a dynamic and ever-evolving sector, continuously influenced by cultural transformations, technological advancements, and shifting consumer preferences. Globally, it remains a significant economic force. According to the McKinsey State of Fashion 2024 report,[1] the fashion industry is expected to continually grow despite challenges such as geopolitical instability and inflation. This resilience is driven by sustained consumer demand and the industry's capacity for adaptation, including innovations in sustainability and digital fashion. In this context, the art of crafting harmonious outfits has become increasingly crucial.

Fashion transcends its basic function of covering the body; it serves as a powerful medium for self-expression, enabling individuals to communicate their identity, mood, and personality. A visually-collocated ensemble can boost confidence, convey specific messages, and leave lasting impressions. Achieving such harmony in clothing selection requires more than just a superficial awareness of trends. It requires a comprehensive understanding of how various fashion elements interact to form a cohesive look. The concept of fashion compatibility is central to the creation of visually appealing outfits. This concept involves an understanding of how factors such as color, texture, style, and proportion can be integrated to produce a unified and aesthetically pleasing appearance. While some individuals may have an innate sense of what combinations are visually effective, many others find it challenging to assemble outfits that achieve balance and harmony. The vast array of choices in the modern fashion market can be overwhelming, even for those with a keen sense of style. To address these challenges, the task of fashion compatibility modeling has emerged, focusing on the development of systematic approaches to assess outfit compatibility, assist-

[1] https://www.mckinsey.com/industries/retail/our-insights/state-of-fashion.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025

W. Guan et al., *Advanced Multimodal Compatibility Modeling and Recommendation*,
Synthesis Lectures on Information Concepts, Retrieval, and Services,
https://doi.org/10.1007/978-3-031-81048-0_1

ing individuals in making more informed and aesthetically sound fashion choices. Fashion compatibility modeling presents inherent complexities due to the need of integrate various modalities—such as visual, textual, and categorical information, of fashion items. As each of these modalities significantly influences the perception of compatibility, existing methods focus on multimodal compatibility modeling.

1.2 Challenges

Although early studies have achieved remarkable progress, they suffer from the following key challenges.

The first prominent challenge lies in fully leveraging category information for outfit recommendations. Existing studies have focused solely on unified representation learning across various modalities, neglecting to effectively harness the category information as guidance to improve the dynamic representation of these items.

The second challenge is associated with the practical try-on appearance. Although numerous multimodal compatibility modeling methods yield promising results, they frequently overlook the practical considerations of fashion compatibility—specifically, how garments look when worn. This involves understanding the spatial arrangement and partial coverage of items according to their function or category. Nevertheless, evaluating compatibility based on actual try-on appearance is inherently challenging, especially in the absence of real try-on images.

Another significant challenge is fine-grained fashion compatibility modeling. In fact, fashion compatibility among items is typically influenced by multiple factors such as color, material, and style. Investigating these factors and fulfilling the fine-grained visual compatibility modeling is crucial for a deep understanding of fashion compatibility. However, we do not have explicit labels for supervising these factors modeling. Therefore, how to fulfill accurate fine-grained compatibility modeling without explicit supervision is challenging.

The incorporation of side information presents yet another critical challenge. While some methods enhance fashion item recommendations through visual features, many overlook additional side information, such as attributes like color and category. These attributes often convey key features not fully captured by an item's visual content alone, including material and brand. Integrating semantic attributes can promote the item characteristic learning and user preference learning, and hence being crucial for making accurate recommendations.

The fifth challenge lies in the model efficiency. Previous efforts focus on improving the recommendation effectiveness but overlook the model efficiency. Typically, modeling heterogeneous fashion entities—including users, outfits, items, and attributes—requires integrating these diverse elements within a unified graph. Traditional graph convolution methods often struggle with the inefficiencies facing with such a graph with diverse entity types and entity relationships. On the other hand, directly using conventional hashing for boosting model efficiency may cause much information loss. Therefore, how to boost the model

efficiency while maintaining the original data information as much as possible is another challenge.

The final challenge pertains to the compatible item generation, especially generating multiple compatible items simultaneously for a given item represents a notable challenge. Current methods are generally restricted to generating a single image of a compatible clothing item at a time, often depending on general image-to-image translation frameworks. In contrast, methods based on multi-modal image-to-image translation may encounter difficulties when there is significant spatial misalignment between source and target domains, often requiring additional information, such as attributes, user preferences, or reference masks, to guide the image generation process effectively.

1.3 Our Solutions

In response to the complex challenges of fashion compatibility modeling and recommendation, this book explores innovative approaches that introduce advanced methods to overcome the limitations of existing systems. To enhance the accuracy and efficiency of fashion compatibility assessments, we utilize multimodal data, advanced neural network architectures, and knowledge-driven frameworks. By combining theoretical insights with practical applications, this book offers a comprehensive examination of how cutting-edge technology can transform the way we perceive and interact with fashion. Ultimately, it empowers individuals to make more informed and stylish clothing choices.

To tackle the challenge of category-aware compatibility, we introduce a novel category-aware multimodal attention network, which enhances the compatibility modeling of fashion items by utilizing both category-specific visual and textual descriptions. This methodology improves the precision of fashion recommendations by considering contextual nuances inherent in different fashion categories.

To further advance the contextual understanding of fashion compatibility, we present the TryonCM2 framework, which uniquely combines discrete item interaction modeling with combined try-on appearance modeling. This integration allows for a more holistic view on how fashion items interact with one another, providing deeper insights into contextual fashion dynamics.

To address the challenge of fine-grained fashion compatibility modeling, we propose the CTO-Net approach, which also evaluates fashion compatibility from both discrete collocation and try-on perspectives. In particular, it presents a new disentangled graph learning scheme to investigate fine-grained fashion compatibility.

To enhance fashion item recommendation, we introduce MM-FRec, a multi-modal recommendation scheme that utilizes both visual and attribute data. This approach has been validated through extensive experimentation, demonstrating its effectiveness in improving the relevance and appeal of fashion item recommendations.

For addressing the challenge of computational efficiency, we explore a hashing-based outfit recommendation approach. This technique integrates diverse entity types and their relationships through a bi-directional graph convolution algorithm, significantly reducing computational costs while maintaining high-quality recommendations.

Additionally, to address the challenge of generating compatible fashion items, we present BC-GAN, a generative framework capable of synthesizing multiple compatible fashion items simultaneously. By employing contrastive learning, BC-GAN enhances the compatibility between original and synthesized items, a capability validated through experiments on the newly constructed DiverseOutfits dataset.

By consolidating these innovative methods, this book significantly advances the field of fashion compatibility modeling and recommendation. It provides researchers and practitioners with comprehensive insights into leveraging multimodal data, advanced neural networks, and knowledge-guided frameworks to develop more accurate and efficient fashion compatibility and recommendation systems.

1.4 Book Structure

The remainder of this book is organized into six chapters, each dedicated to a specific aspect of advanced multimodal compatibility modeling and recommendation systems in the fashion domain. Chapter 2 explores category-guided fashion compatibility modeling by introducing a category-aware multimodal attention network and a categorical dynamic graph convolutional network. These techniques are designed to enhance visual-semantic embeddings. Chapter 3 delves into try-on-enhanced fashion compatibility modeling through the TryonCM2 framework, where the combined fashion compatibility is modeled by generating a try-on template based on the given fashion items and analyzing its contextual structure with bidirectional long-short term memory (Bi-LSTM) network. Chapter 4 is dedicated to fine-grained and try-on-enhanced outfit compatibility modeling, utilizing the CTO-Net framework. This chapter introduces a disentangled graph learning approach coupled with integrated distillation learning to comprehensively evaluate compatibility from both collocation and try-on perspectives. Chapter 5 presents the MM-FRec framework for attribute-enhanced fashion item recommendation, which jointly models the attribute-enhanced latent preference and visual preference of a user based on GCN. In particular, a multitask learning-based relation-aware propagation method is designed to distinguish importances of neighbor connections with different relation types. Chapter 6 focuses on hashing-based efficient outfit recommendation. It describes a method grounded in heterogeneous graph learning, where a bi-directional graph convolution algorithm is introduced to optimize computational efficiency. To address information loss, dual similarity-preserving regularizations are proposed. The final chapter discusses a collocated clothing synthesis method using the BC-GAN frame-