DATA-DRIVEN RECOMMENDATIONS FOR FASHION – AN INVESTIGATION OF PERSONALIZATION WITH SPARSE DATA

LIAO SHUIYING

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

School of Fashion and Textiles

Data-driven Recommendations for Fashion – An Investigation of Personalization with Sparse Data

LIAO Shuiying

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

December 2024

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____

_____LIAO Shuiying____

Abstract

The fashion industry is a dynamic and rapidly evolving field, characterized by diverse consumer preferences, behaviors, and styles. This ever-changing landscape has made innovative fashion recommendation systems a crucial area of study. In data-driven recommendation tasks, data sparsity issue is of significant concern, and it becomes even more critical in fashion recommendations, where a high level of personalization is required to cater to individual preferences. However, this emphasis on personalization exacerbates the data sparsity challenge, as user-specific interactions are often limited.

While existing data-driven recommendation methods have made significant strides in fashion recommendation research, they still face notable limitations. The most significant challenges arise from their inability to effectively address data sparsity, leading to suboptimal performance. This issue extends to their capacity to incorporate user-specific preferences effectively. Simple linear models, which are often employed in personalization tasks, fail to capture the complicated relationship between user preferences and item compatibility. Other limitations include challenges in tracking dynamic user preferences, understanding the multifaceted nature of fashion items, and providing context-aware recommendations that reflect the user's specific needs at any time. Capturing both enduring preferences and temporary needs is crucial for personalized recommendations. Long-term interests, such as style or brand affinity, and short-term interests, influenced by trends or seasons, both play a role in creating relevant suggestions. Ultimately, these challenges are fundamentally rooted in data sparsity.

This thesis addresses the above limitations by proposing multiple strategies specifically designed to manage sparse data in the context of data-driven fashion recommendations. By focusing on two core tasks—personalized clothing matching and

sequential recommendation—the study develops methods to better utilize sparse data and improve personalization and recommendation outcomes. Specifically, three alternative methods are designed for the personalized complementary recommendation task from three perspectives—consistency, coupled indirect personal compatibility, similar users or products—and can be flexibly replaced or combined depending on specific application requirements, and one method are proposed for the second sequential recommendation task. In Chapter 3, innovative constraints based on consistent user behaviors are introduced to better utilize interaction history, enhancing the model's ability to infer preferences with limited data. Chapter 4 presents a novel concept, Indirect Personal Compatibility, to balance personalization and compatibility in recommendations, achieving alignment with individual preferences through iterative training. Chapter 5 explores contrastive learning to capture latent representations in sparse environments, from user preferences and item compatibility views, and implements adaptive collaborative signal selectors to mitigate data noise when using historical interactions as auxiliary information. Last but not least, in product sequential recommendation in Chapter 6, besides a multi-scale transformer architecture, which captures both long-term and short-term preference within a unified framework. Data sparsity issue is addressed by identifying similarities among items the user has interacted with, filling the sparse interaction matrix across various behaviors. Besides, multiple data mining and augmentation techniques are explored to expand upon the available data, uncovering new chances for enhancing recommendation accuracy in sparse environments. Extensive experiments across two tasks and multiple open-access datasets to evaluate the effectiveness of all proposed methods in tackling data sparsity in data-driven personalized fashion recommendations. Results show significant improvements in recommendation performance, underscoring the potential of these approaches to enhance both robustness and accuracy in data-sparse environments.

Publications

1. **Liao Shuiying**, Ding Yujuan., and P. Y. Mok, “Recommendation of mix-and-match clothing by modeling indirect personal compatibility”. *In Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*, pp. 560-564, 2023.

2. **Liao Shuiying**, Ding Yujuan., P. Y. Mok, Huang Qiushi, and Cao Jialun, “Reproducibility Companion Paper: Recommendation of Mix-and-Match Clothing by Modeling Indirect Personal Compatibility”. *In Proceedings of the 2024 International Conference on Multimedia Retrieval*, pp. 1224-1227, 2024.

3. **Liao Shuiying** and P. Y. Mok, “Hypergraph-Enhanced Contrastively Regularized Transformer for Multi-Behavior E-commerce Product Recommendation”. *In Proceedings of the 2024 IEEE International Conference on Data Mining*, 2024.

4. **Liao Shuiying**, P. Y. Mok, Gerhard Flatz, and Li Li, “Consistency Regularization for Complementary Clothing Recommendations”. *Applied Soft Computing Journal*, 2025, Under Review.

5. **Liao Shuiying** and P. Y. Mok, “ICEnet: Attentive Preference Modeling Using Contrastive Learning for Personalized Fashion Matching Recommendations”. *For the 31st SIGKDD Conference on Knowledge Discovery and Data Mining,* 2025, Under Review.

6. **Liao Shuiying** and P. Y. Mok, “PCGNet: Unifying Shared and Specific Information for Fashion Matching Recommendations”. *For the 31st SIGKDD Conference on Knowledge Discovery and Data Mining,* 2025, Under Review.

7. **Liao Shuiying**, P. Y. Mok, Gerhard Flatz, and Li Li, “APCL: Attentive Preference Modeling with Contrastive Learning for fashion complementary recommendation”. *ACM Transactions on Multimedia Computing, Communications, and Applications,* 2025, Under Review.

8. **Liao Shuiying**, and P. Y. Mok, “Adaptive Preference Modeling for Sequential Recommendation”. *ACM Transactions on Recommendation Systems,* 2025, Under Preparation.

9. **Liao Shuiying**, and P. Y. Mok, “Shared-Specific Information Modeling with Mutual Information Maximization for Personalized Fashion Complementary Recommendations”, *IEEE Transactions on Multimedia,,* 2025, Under Preparation.

Acknowledgements

As I conclude the journey of my research studies, it is with a heart full of gratitude that I acknowledge the support and guidance I have received along the way.

First and foremost, I extend my deepest gratitude to my supervisor, Prof. Tracy Mok, for providing me with such a great opportunity to do my research in the fashion recommendation area. I appreciate all her unwavering support, invaluable insights, and patience throughout my research. Her expertise and guidance in my academic growth have been instrumental in my academic pursuits.

I would like to express my sincere appreciation to my fellow researchers and colleagues, I am grateful for the stimulating discussions and collaborative spirit. Your support has made my time in graduate school both enjoyable and productive. A special thank you should go to Dr. Yujuan Ding, who has been providing me with valuable assistance, guidance, and constructive feedback during my PhD.

I am indebted to my family, for their love, encouragement, and understanding throughout my PhD journey. Their belief in me has been a constant source of motivation, and their sacrifices have made this work possible.

As I move forward, I am eternally grateful for the opportunity to pursue my passion for knowledge and research under the guidance of such remarkable individuals.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Research Background

The fashion industry has been one of the fastest-growing businesses in the global economy and e-commerce plays an important role in the expansion process (Ding, Lai, et al., 2023a). According to market research, online fashion sales have seen a significant increase in recent years and will continue to grow at a substantial rate in the coming years (Bulović & Čović, 2020). This growth is attributed to the convenience and personalization offered by e-commerce platforms, which allow consumers to shop from anywhere and at any time, providing them with a seamless shopping experience. As digital platforms become the primary medium for fashion retail, the need for effective recommendation systems has become increasingly critical.

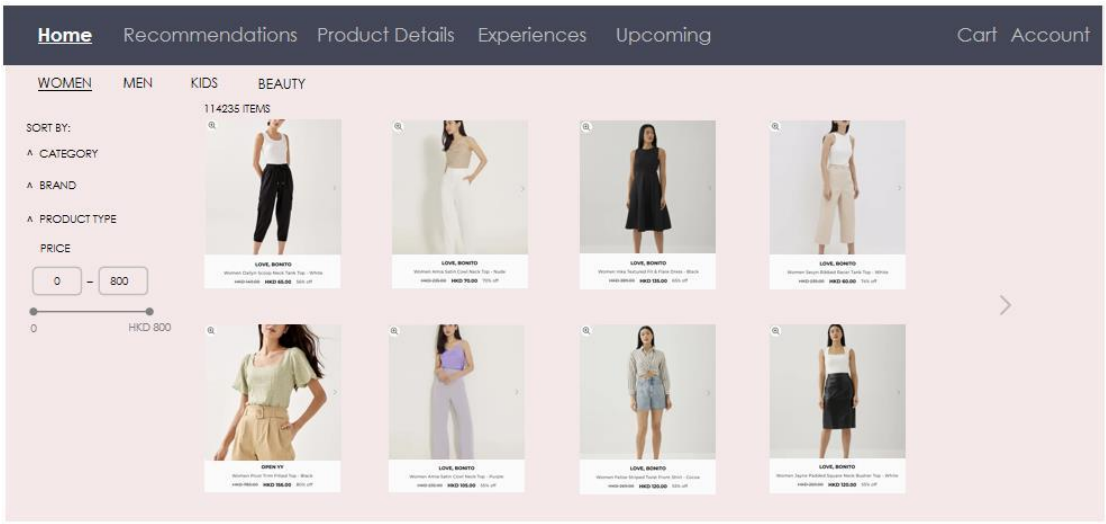

Figure 1-1 General fashion shopping page display.

Fashion recommendations aim to enhance customer experience by providing suitable suggestions related to fashion (Ding, Lai, et al., 2023a; Goti et al., 2023). The significance of personalized fashion recommendations lies in their potential to drive consumer engagement, increase sales, and improve customer satisfaction by offering recommendations that are not only accurate but also responsive to personal taste

(Chakraborty et al., 2021; Deldjoo et al., 2022). Distinguished from other areas, the fashion industry is vibrant and rapidly changing, characterized by its dynamic nature and the diverse preferences of consumers, which operate within a highly dynamic environment where both the items and user preferences are in constant flux. The evolving nature of fashion trends and the rapid change in consumer preferences make fashion data unique and challenging (Liao et al., 2023).

Fashion is highly personalized. For example, one person might prefer bright colors, while another might favor simple black and white. This illustrates that fashion is a reflection of our personalities. In addition, fashion recommendations are heavily reliant on transactional data to provide recommendations. This is because transactions are a strong indicator of personal preferences. For instance, if someone frequently purchases casual clothing, we can infer that casual attire is likely to be appealing to this user. Another unique aspect of fashion compared to other fields is that clothing often comes in pairs or sets. The way clothing is styled is a crucial part of personalization. Moreover, with the vast array of choices available online, customers can easily feel overwhelmed. Faced with so many options, users may experience confusion. Therefore, the key to effective fashion recommendations lies in understanding exactly what the user prefers based on their interactions and providing personalized suggestions.

However, user-item interaction histories are often sparse, as users typically interact with a relatively small fraction of the available fashion products. Despite the large number of available items, the low frequency of direct interactions between users and items poses a significant challenge to the development of effective personalized recommendation models.

The data sparsity issue is a pervasive issue across recommendation systems but is particularly severe in the fashion domain. The short-term availability of fashion items, combined with the limited direct interaction history between users and items, exacerbates the sparsity problem. Consequently, many items have limited or no interaction history, making it challenging to gather sufficient data for personalized

recommendations. Recent fashion recommendation methods mostly rely on deep learning and multimodal techniques (Guan et al., 2022; Liu et al., 2024; Song et al., 2019; X. M. Song et al., 2023; Zhou et al., 2022), which incorporate both visual and textual information, have gained traction in fashion recommendation due to their capacity to capture complex features in fashion data. While existing methods have made strides in fashion recommendation research, they exhibit notable limitations.

Many existing systems struggle to handle the sparsity of fashion data and incorporate user-specific preferences effectively, thus often failing to capture the complex relationship between user preferences or item compatibility. Consequently, fashion recommendation models are often unable to process and interpret the complex, multimodal nature of fashion data, resulting in recommendations that may lack relevance. In addition, with millions of products available for consumers, the ability to discover and select items that match individual tastes and preferences has become increasingly difficult (Liu et al., 2020). Moreover, fashion is inherently visual and subjective, the aesthetic appeal of items plays a crucial role in consumer decisions (Dong et al., 2020). Capturing and utilizing visual and textual information in recommendations is a complex task that requires advanced computational techniques. Recommendation systems must account for this diversity and subjectivity to provide relevant suggestions.

This thesis aims to address the limitations of current systems by exploring innovative solutions that leverage both the unique attributes of fashion data and the distinct requirements of fashion recommendation. Through extensive experimentation across two tasks and multiple open-access benchmark datasets, the proposed methods' effectiveness in addressing data sparsity in fashion recommendation systems will be measured, demonstrating significant improvements in recommendation performance.

1.2 Fashion Recommendation

Fashion in modern society has become one of the world’s largest industries.

Research advances in machine learning and artificial intelligence are significantly enhancing fashion-related applications, such as fashion analysis and recommendation (H.-J. Chen et al., 2023). In the dynamic and ever-changing world of fashion, recommendation systems have become indispensable tools for enhancing user experience and driving engagement. Fashion recommendation systems can be broadly categorized into several types, each serving a unique purpose: *Size and Fit Recommendation* (Abdulla & Borar, 2017); *Contextual Fashion Recommendation* (Hao et al., 2020); *Visual Search Recommendation* (Abluton, 2022); *Fashion Complementary Recommendation* (Fashion pair recommendation) (Ding, Lai, et al., 2023b; Huang & Huang, 2016; Song et al., 2021), and *Sequential Recommendations*. Among these, fashion complementary and sequential recommendations stand out as critical areas of focus due to their unique challenges and significant impact on user satisfaction.

Fashion complementary and sequential recommendations are crucial for several reasons. Firstly, they address the inherent complexity of fashion as a domain where aesthetic appeal and personal preference play significant roles. For example, in the field of fashion, products often do not appear individually like other products, but appear in several or a set (Jing et al., 2021). Complementary recommendations enhance the shopping experience by helping users create cohesive and stylish outfits, thereby increasing customer satisfaction and potentially boosting sales.

Sequential recommendations, on the other hand, are vital for capturing the temporal dynamics of user preferences. Fashion is characterized by rapid changes in trends and seasonal variations. By understanding the sequence of user interactions, these systems can anticipate future needs and preferences, offering timely and relevant recommendations that comfort to current trends. This capability is essential for maintaining user engagement and loyalty in a competitive market.

The significance of research in fashion complementary and sequential recommendations lies in their potential to overcome two major challenges: data sparsity

and personalization. Data sparsity is a common issue in fashion due to the vast diversity of products and the rapid turnover of trends, which result in limited interaction data for many items. Addressing this challenge requires innovative approaches that can learn from limited data, such as leveraging visual and contextual features through deep learning techniques. In addition, personalization is equally important, as fashion is inherently personal and subjective. Users have unique preferences and styles, recommendations that are comfortable to individual tastes. By focusing on complementary and sequential recommendations, researchers can develop systems that not only understand the aesthetic relationships between items but also adapt to the evolving preferences of users over time.

This thesis explores techniques to leverage sparse data environments to enhance recommendation efficacy, focusing on two key tasks: personalized clothing matching recommendation and product sequential recommendation. The two tasks will be introduced in the following sections in detail, discussing the methodologies and innovations that drive their effectiveness.

1.2.1 Fashion Complementary Recommendation

Fashion Complementary Recommendation focuses on suggesting items that go well with a given item, rather than suggesting similar items. In the context of fashion, it's akin to offering an accessory or piece of clothing that complements another item a user is interested in or has purchased. It is far more difficult to make complementary suggestions, which aim to offer products that are a good fit with a query product (Wang et al., 2022; Z. Yang et al., 2022; Yu et al., 2019), than it is to make substitute recommendations, which just take into consideration how similar two products are to one another.

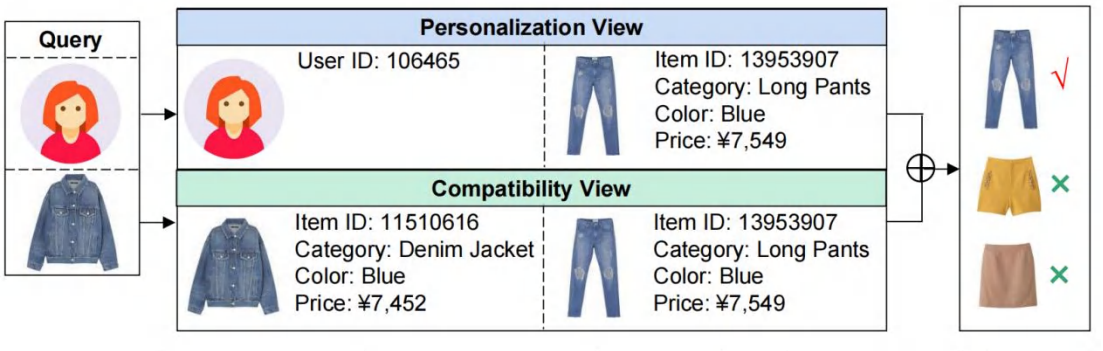

Figure 1-2 Personalized Fashion Complementary Recommendation Example.

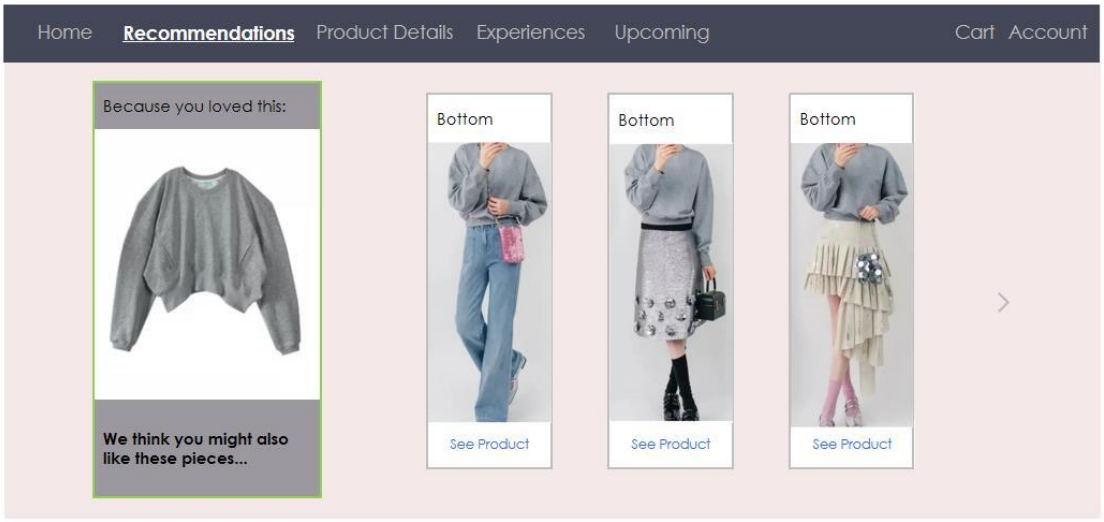

Figure 1-3 Personalized Fashion Complementary Recommendation Display.

The challenges of modeling fashion complementary products' compatibility are as follows: Fashion is subjective, and what one person believes complements another item might not match someone else's opinion; If not done subtly, users might feel overwhelmed or pressured by too many recommendations; Ensuring the recommended complementary items are in stock and available in the necessary sizes can be a challenge.

The vast majority of research (Cucurull et al., 2019; Jing et al., 2019) model item-item interactions in order to take advantage of complementing relations. Others suggested enhancing performance by taking advantage of the fine-grained compatibility of products on a number of different levels. For instance, Zhang et al. (Zhang et al., 2018) provided a quality-aware framework that would promote products of the highest quality based on the preferences of the user. Wang et al. (Wang et al.,

2018) carried out research using a path-constrained technique to learn the multi-relationships between various items. A knowledge-aware recommendation algorithm was used by Xu et al. (Xu et al., 2020) for complementary products that had asymmetric, non-transitive and higher-order interactions. Aspect-level compatibility was discovered by Wang et al. (Wang et al., 2022) to make intent-aware complimentary recommendations.

However, none of them has ever thought about the connection that exists between the potential items and the products that customers have traditionally queried.
In addition, personalized compatibility modeling places an emphasis not only on personal preferences but also on the compatibility of products (Chen et al., 2019; Li et al., 2020; Song et al., 2019).

Pioneering research such as (Song et al., 2019; Veličković et al., 2017) put out the idea of modeling inter-product interaction to determine product compatibility and combining this with user-product interaction to determine personal preference in a number of different ways. Recent research (Ai et al., 2018; Chen et al., 2018; McInerney et al., 2018; Rendle et al., 2012) has demonstrated that the integration of a customization component (that is, the user's past preferences) can significantly improve the effectiveness of a recommendation system. In the paradigm that was proposed by (Yan et al., 2022), product relations and user preferences were represented, respectively, by linearly combining a graph network and a transformer encoder. In order to capture personalized item-item relationships across prospective products, Lin et al. (Lin et al., 2022) further enhanced their model by incorporating intention self-attention.

However, the reasons why personalization and compatibility are modeled in a totally detached manner have not been properly examined.

1.2.2 Sequential Recommendation

Sequential recommendation focuses on predicting the next item a user is likely to interact with based on their historical sequence of interactions (J. Z. Chen et al., 2023). Sequential recommendation is particularly important in domains where user

preferences and behaviors are dynamically changing, such as in e-commerce and content consumption. Sequential recommendation systems aim to capture the temporal patterns and dependencies in user behavior, allowing them to provide timely and relevant suggestions that align with users' evolving interests. Particularly for e-commerce sites, it is crucial to consider both short-term activities and long-term shopping histories. Short-term activities can provide insights into a user's immediate needs, while long-term activities offer a window into their sustained preferences for certain items or categories over an extended period. Consequently, mining both long-term static preferences and short-term dynamic interests has become a popular focus in many e-commerce applications.

For example, as illustrated in the following picture, based on user historical behavior sequence, an online shopping scenario of multi-behavior sequential recommendation involves predicting the next item that a user might like to purchase. This includes actions such as viewing, favoriting, adding to cart, and purchasing. The goal of multi-behavior recommendation systems is to anticipate users' future activities by modeling the sequential dependencies of their interactions and understanding the correlations between different behavior patterns.

Figure 1-4 An online shopping example of sequential recommendation including multiple behavior types. The target is to predict the next item users would interact with, based on their historical behavior sequences.

Recently, there has been a noticeable trend toward using artificial intelligent enhanced sequential recommender systems to capture the complex connections between items from diverse transactions for product recommendation. GRU4Rec (Hidasi, 2015) first utilized recurrent neural networks to encode sequential user-item

interaction patterns chronologically. Next, the transformer, which has a more sophisticated architecture, is shown ideal for a sequential recommendation because its attention mechanism can effectively capture long-term dependencies (L. Liu et al., 2023). For instance, SASRec (Kang & McAuley, 2018) and BERT4Rec (Sun et al., 2019) employ attention-based mechanisms to model a global context for each element in the sequence. Furthermore, the introduction of graph neural networks (GNNs) has significantly improved the performance of sequential recommendation models (Ding et al., 2021). Models such as SR-GNN (Hidasi et al., 2015), MA-GNN (Ma et al., 2020), and GDERec (Qin et al., 2024) utilize graph message passing among neighboring items to capture sequential signals and global connections. Despite the encouraging outcomes, these models are primarily intended to handle a single type of interaction data and may not handle well a variety of user-item relationships. As a result, there is a growing interest in developing neural network models that can effectively explore the rich latent semantics encoded in both fine-grained and coarse-grained behavior-aware sequential information.

1.3 Data Sparsity

1.3.1 Data Source in Fashion Recommendations.

In fashion recommendation systems, a diverse array of data sources is utilized to provide personalized suggestions to users. Key data sources include user interaction data, which encompasses clicks, views, likes, purchases, and ratings. Product metadata, detailing attributes such as brand, category, color, size, material, and price, aids in understanding item characteristics for content-based recommendations. Visual data, derived from images and videos, is crucial for assessing style, color, and pattern similarities, while user profile data, including demographics like age, gender, and location, enhances personalization. Additionally, contextual data, such as time of day, season, or specific events, further refines recommendation relevance.

1.3.2 Definition and Impacts of Data Sparsity.

Data sparsity refers to the challenge in recommendation systems where the available data on user-item interactions is insufficient or unevenly distributed (Liu et al., 2020). This means that there are relatively few interactions, such as ratings, clicks, or purchases, recorded for a large number of items and users. As a result, the data matrix that represents these interactions is mostly empty, making it difficult for recommendation algorithms to accurately predict user preferences and generate reliable recommendations (Fan et al., 2023). However, data sparsity poses significant challenges, impacting the accuracy and personalization of recommendations. Sparse data can lead to reduced recommendation accuracy, as limited interactions hinder the prediction performance effectively. This sparsity exacerbates the cold start problem for new users or new items, limiting personalization and potentially delaying trend adaptation. Moreover, sparse data complicates model training, risking overfitting or underfitting due to insufficient data for robust pattern learning. Consequently, systems may rely more on auxiliary data sources, which might not fully capture user preferences (Xi et al., 2021), and face potential biases. Addressing data sparsity through techniques like collaborative filtering, content-based filtering, hybrid models, and data augmentation is essential for enhancing the effectiveness and relevance of fashion recommendation systems.

1.3.3 Causes of Data Sparsity in Fashion.

In fashion domain, the number of available items is vast and constantly growing. Users typically interact with only a small fraction of these items, and users may not frequently interact with the system, resulting in limited data about their preferences. What's more, new users or items entering the system have little to no interaction history, contributing to sparsity. This makes it challenging to recommend items to new users or to recommend new items to existing users. Additionally, interactions can be highly seasonal or trend-driven, causing fluctuations in data density over time (Bin et al., 2019). Items that are popular during certain seasons may have sparse interactions during off-

seasons.

1.4 Personalization in Fashion Recommendations

Fashion is a highly personalized domain, with individual tastes and preferences playing a significant role in purchase decisions. Personalized recommendations ensure that users are presented with items that align with their historical behaviors and expressed preferences, thereby facilitating the decision-making process and providing a more enjoyable shopping experience (Ding, Lai, et al., 2023b).

Moreover, personalization in fashion recommendations extends beyond simple preference matching; it involves understanding the contextual and situational factors that influence a user's choices. This includes considerations such as the occasion, weather, and seasonal trends, which are crucial for providing relevant and timely suggestions. Despite the clear benefits of personalization, fashion recommendation systems face significant challenges due to data sparsity. The fashion domain is characterized by a rapidly expanding catalog of items, with new products frequently introduced and older ones becoming obsolete. This results in a sparse dataset where many items have limited interaction history, making it difficult to accurately model user preferences and item popularity. He (He & Chua, 2017) and Deldjoo (Deldjoo et al., 2022) noted that sparse purchase data challenges the use of traditional recommender systems. The cold start problem, where new users or items lack sufficient historical data for effective recommendation, is particularly difficult in the fashion industry. Data sparsity also exacerbates the challenge of capturing suitable user preferences and the complex relationships between items.

In the context of fashion, where visual and stylistic aspects are paramount, traditional collaborative filtering approaches that rely heavily on user-item interaction data are often insufficient (Bin et al., 2019). As a result, fashion recommendation systems must leverage alternative data sources and advanced computational techniques to compensate for the lack of interaction data and to better understand the intricate

dynamics of fashion preferences and compatibility.

To overcome the challenges posed by data sparsity and to enhance personalization, fashion recommendation systems must effectively incorporate an understanding of both user preferences and item compatibility. User preferences refer to the stylistic, aesthetic, and functional inclinations of individual users, which can be inferred from their browsing, purchasing, and interaction behaviors. By analyzing these behaviors, recommendation systems can learn to predict which items are likely to appeal to a particular user. Item compatibility, on the other hand, aims to maintain stylistic and functional harmony between different fashion items. In the context of fashion, compatibility is crucial for tasks such as mix-and-match recommendations and outfit composition, where the goal is to suggest items that work well together. This requires the system to not only recognize individual item attributes but also to understand the combinatorial aspects of fashion. The integration of user preferences and item compatibility is thus essential for creating personalized fashion recommendations that are both relevant and stylistically coherent. Advanced techniques such as deep learning, computer vision, and natural language processing are increasingly employed to extract meaningful insights from user behavior data and item attributes, enabling recommendation systems to deliver highly personalized and context-aware recommendations.

1.5 Other Challenges

Another key challenge in personalized fashion recommendation is representation-related issues. Fashion recommendation is an area that highly relies on content and context information, such as visual, textual, and time information. Limited representation always has a significant effect on prediction accuracy. The prediction results would be impacted by the feature scales because the majority of current models use features to predict compatibility and personalization scores. More significantly, the majority of customized item-matching models forecast the recommendation scores

using multiple feature modalities, which inherently differ in magnitude because they come from various sources. The prediction scores at various levels will be influenced by multi-modal features with various scales.

Such unbalanced contributions directly resulted from scale difference is unreasonable since it improperly assigns the importance of different features. However, such a *feature scale* issue has been extensively ignored in existing methods (He & McAuley, 2016; Sagar et al., 2020; Song et al., 2019; X. Song et al., 2023), which therefore degrades their overall performance.

In addition, effectively capturing both long-term and short-term user interests is crucial for making personalized and contextually relevant suggestions. Long-term interests represent enduring preferences that persist over time, such as a user's general style preference or affinity for specific brands or colors. Short-term interests, on the other hand, reflect temporary or immediate needs, which could be influenced by current trends, seasonal changes, or special events (H. B. Liu et al., 2023; Xia et al., 2023; Xuan et al., 2023). A successful fashion recommendation system must combine these two types of preferences to provide suggestions that meet a user's stable tastes while also being responsive to dynamic, short-term interests.

1.6 Research Aim and Objectives

This research aims to address the above-mentioned challenges, especially focusing on critical challenges of personalization and data sparsity in the fashion recommendation domain. The goal is to provide personalized and context-aware fashion recommendations that adapt to dynamic user preferences and trends, even with limited interaction data, thereby improving the overall user experience and satisfaction in the ever-changing fashion landscape.

The specific objectives of this research study are as follows.

i. To conduct a thorough review of existing literature on personalized fashion

recommendation, especially focusing on fashion complementary recommendation tasks and sequential recommendation tasks.

ii. To tackle the pervasive issue of data sparsity in fashion recommendation by developing innovative strategies and systems, aiming to enhance the ability of these systems to provide personalized recommendations, leveraging advanced modeling techniques to extract meaningful insights from sparse transaction data.

iii. To develop innovative methodologies that effectively address the challenge of personalization in fashion recommendation systems. By leveraging advanced techniques such as multi-modal data integration, behavior-based constraints, and adaptive learning models, the study aims to capture the nuanced preferences of individual users.

iv. To design and develop a multi-scale transformer architecture to capture both long-term and short-term user preferences in the context of sequential recommendations. By identifying similarities among interacted items and employing data mining and augmentation techniques, the study seeks to fill gaps in the sparse interaction matrix, ultimately improving recommendation accuracy.

v. To assess the effectiveness of the proposed models for addressing personalization and data sparsity in fashion complementary recommendations and sequential recommendations, across various extensive experiments.

vi. To conclude the contribution of this thesis on personalized fashion recommendations, and further explore more advanced and cost-effective strategies in future work.

1.7 Methodology Overview

According to the defined research aim and objectives, the overall methodology for fashion complementary recommendation and sequential recommendation of this study

is illustrated below.

1.7.1 Overall Design of Personalized Fashion Recommendations

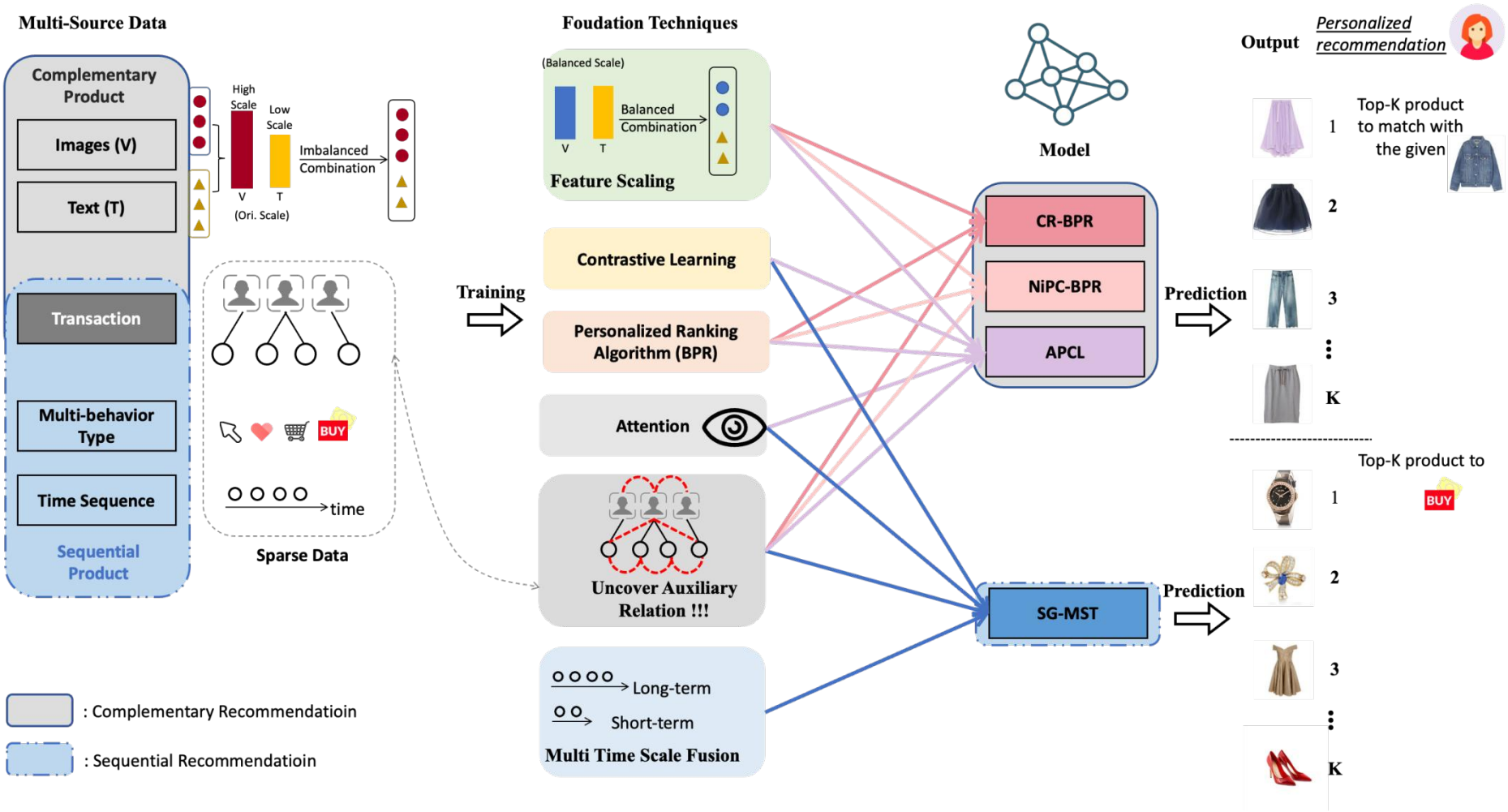

Figure 1-5 The overview of fashion recommendation systems developed in this study.

The overview of four proposed methods for fashion recommendation systems are illustrated in Figure 1-5. To address data sparsity challenges in fashion recommendation systems, this work proposes four frameworks leveraging multi-modal interaction data and diverse techniques to uncover auxiliary relationships from limited data. Specifically, three novel models are developed for the first fashion complementary recommendation task, and one for the second sequential recommendation task. Crucially, these three approaches for the fashion complementary recommendation, are designed as complementary and modular components rather than competing solutions. Each method addresses data sparsity from a unique perspective—consistency, coupled indirect personal compatibility, similar users or products—and can be flexibly replaced or combined depending on specific application requirements. This design philosophy emphasizes a collaborative exploration of the data sparsity problem by integrating different technologies that work together to alleviate data constraints while providing interpretability and scalability.

1.7.2 CR-BPR for Complementary Recommendation

The foundation of CR-BPR is a dual Bayesian Personalized Ranking (BPR) architecture that concurrently models product matching and user preferences. Hybrid collaborative filtering using latent and multi-modal feature inputs, which have proven successful in earlier research, is incorporated into each BPR branch. A feature scaling procedure is incorporated into CR-BPR to solve the problem of feature scale discrepancies impacting prediction. Before being utilized to predict user preferences and product matching scores, this approach normalizes a variety of feature types, such as latent features and multi-modal material. Additionally, consistency regularization branches are introduced by CR-BPR, which captures the similarity between target products and past selections from the standpoints of product matching and user preference. In addition to using more data than user-product interactions to suggest associated products that suit individual preferences, consistency modeling also serves as a regulation method.

The development of the CR-BPR model makes two contributions: first, it uses product consistency to address user preferences and product matching in clothing recommendations; second, to address the relative feature importance problem in multi-modal data features, it incorporates a feature scaling function, and which has been mainly ignored in earlier research. By integrating these components, CR-BPR aims to enhance personalized clothing matching and offers a more accurate and trustworthy recommendation system.

1.7.3 NiPC-BPR for Complementary Recommendation

Complementary recommendations, which recommend products that go well with other products, play a more significant role in fashion recommendations compared to substitute recommendations. Fashion products are unique because they are often better characterized by visual images or textual descriptions, making it challenging to integrate diverse factors from multi-modal data for predicting user preferences. While some existing work (Yang et al., 2021; Song et al., 2016; Wang et al., 2022) has improved co-purchased product compatibility modeling, they may overlook the signals from customers' previous possession of the product. Since the majority of current approaches to fashion complementary recommendation focus on compatibility modeling or personalization modeling independently, it may be possible to combine the two for recommendations that are more successful.

In this method, we propose a **N**ormalized **i**ndirect **P**ersonal **C**ompatibility modeling scheme based on **B**ayesian **P**ersonalized **R**anking (**NiPC-BPR**), Our methodology handles personalized compatibility modeling both directly and indirectly, in contrast to existing state-of-the-art personalized compatibility modeling methods such as GP-BPR(Song et al., 2019), which mainly directly and independently model product pairing compatibility and personal preference. Given this, in addition to taking use of the relationships between fashion products and customers' interactions with candidate products, we introduce an **i**ndirect **P**ersonal **C**ompatibility ('**iPC**') module to capture additional relations by modeling the compatibility between the complementary

products to be recommended and the user's historical query products. This can be thought of as an extension of directly exploiting relations between users' interactions with candidate products and the relations between fashion products.

1.7.4 APCL for Complementary Recommendation

In the third approach, we propose an innovative approach called **A**ttentive **P**reference modeling with **C**ontrastive **L**earning (**APCL**). To address the critical challenges of personalization and data sparsity in fashion recommendation systems, this method contributes to the following aspects:

Firstly, APCL introduces an Attentive Preference Similarity module that employs self-attention to combine interactions from other users who have chosen the same item. This innovative approach allows the model to capture the target user's latent interests more effectively, leading to more accurate personalized recommendations.

Secondly, APCL incorporates Contrastive Learning branches for enhanced representation learning of target items. This approach captures complementary information from multiple novel functional perspectives, which is particularly beneficial in scenarios with limited data, as it helps to learn transferable representations across different latent spaces.

In addition, to address the challenge of data sparsity, APCL utilizes implicit connections enriched through cross-view contrastive learning by applying Attentive Preference Similarity Module (AP Module). The AP Module identifies implicit connections between users based on their shared interactions with specific items. It does this by analyzing the global interaction data to find users who have shown interest in the same items as the target user. By considering the preferences of users who interact with similar items, the AP Module aims to uncover the target user's latent interests, which can then be used to generate more accurate personalized recommendations.

1.7.5 SG-MST for Sequential Recommendation

This method addresses the critical issues of data sparsity and the complexity of

modeling user preferences in e-commerce environments. The motivation behind this research stems from the need to enhance recommendation systems' ability to predict user interactions accurately, considering the diverse and dynamic nature of user behaviors and the inherent noise and limitations in transaction data.

SG-MST integrates a Similarity Augmented Multi-Behavior Hypergraph to capture complex dependencies among items and strengthen connections through item context similarities, leading to more informative latent representations. Additionally, a Contrastively Regularized Multi-Scale Transformer is introduced to capture enriched sequential patterns across both long-term and short-term dynamically sampled and augmented temporal scales, thereby improving the robustness of user behavior prediction. In addition, the SG-MST addresses data sparsity by leveraging a top-K similarity augmented hypergraph, which adds supplementary connections based on contextual correlations of items, thus creating a denser and more informative adjacent matrix. This approach helps to alleviate the limitations posed by sparse multi-behavior interactions and enhances the model's ability to extract meaningful patterns from limited data. Furthermore, the contrastive learning component focuses the model on meaningful user behavior patterns while suppressing noise and irrelevant information, enhancing the model's robustness and accuracy in the presence of noisy data.

1.7.6 Theoretical Advantages

The theoretical advantages of each designed method are illustrated below figure and further elaborated.

Figure 1-6 Theoretical advantages of proposed methods.

(1) **Enhanced Personalization:** This research's first and important objective is to enhance personalization in fashion recommendations. Given the diverse consumer preferences and styles in the fashion industry, it is essential to develop systems that can cater to individual tastes. Personalization is achieved by modeling user preferences and leveraging user-specific interactions to provide recommendations. This research aims to address the challenge of data sparsity by integrating multi-modal information, such as visual and textual data, and employing techniques like feature scaling and contrastive learning to strengthen the system's robustness in data-scarce settings, allowing for richer feature extraction when user interaction data is limited.

(2) **Enhanced Compatibility:** Compatibility is an important research focus in fashion recommendation task, which refers to the ability to suggest items that not only satisfy individual preferences but also complement each other in terms of style, color, or design. This research focuses on developing different methods that can effectively capture the compatibility between given fashion products and the matching products to be recommended. By integrating consistency constraints and indirect personal compatibility modeling, the research aims to enhance the model's ability to infer preferences with limited data and provide recommendations that are both personalized and compatible.

(3) **Enhanced Representation/Feature Learning:** Enhancing feature learning is another core objective, aiming to capture more reasonable and comprehensive

features of fashion items and user preferences. This research employs techniques like normalized multi-purpose projection (NMPP) and cross-view contrastive learning to improve the representation of items in latent spaces. These methods help in capturing the multi-modal nature of fashion data, leading to more accurate and relevant recommendations. The research also explores the augmentation of representations with contrastive learning to better utilize sparse interaction data and uncover hidden patterns in user preferences and item compatibilities.

(4) **Enhanced Balance between Personalization and Compatibility:** Striking the right balance between personalization and compatibility is a significant challenge in fashion complementary recommendations, and which is overlooked by existing research. This research aims to develop models that can balance these two aspects effectively. By introducing concepts like indirect personal compatibility and utilizing attention mechanisms, the research seeks to align recommendations with individual preferences while ensuring that the suggested items are compatible with the given items.

(5) Improved Performance: The ultimate goal is to improve the performance of fashion recommendation systems. This includes enhancing accuracy, personalization and robustness, in data-sparse environments. The research employs a hybrid recommendation approach that combines the strengths of various techniques, such as BPR, feature scaling, contrastive learning, and attention mechanisms, to achieve this goal. To verify the effectiveness of all provided methods, this thesis provides diverse experiments on multiple benchmark datasets with multiple evaluation metrics. By addressing the challenges of personalization, compatibility, and data sparsity, the research aims to significantly boost the performance of fashion recommendation systems and provide users with a more satisfying and personalized shopping experience.

In short, data sparsity is the most important focus of this research, especially in personalized recommendations. This research aims to tackle data sparsity by employing

various strategies, including feature scaling, contrastive learning, and the integration of multi-modal information, and most importantly, exploring additional auxiliary connections within the limited provided data from different perspectives. These approaches help in making the most out of limited interaction data and enhance the model's ability to provide robust recommendations from sparse user interactions.

1.8 Organization of the Thesis

This dissertation is organized as follows:

Chapter 2 provides a comprehensive literature review on personalized fashion recommendation systems. It begins with an examination of evaluation metrics for recommendation systems and highlights the datasets commonly used for recommendation tasks, with a detailed look at datasets specifically designed for complementary and sequential recommendations. Then the related technologies in fashion recommendation are discussed, providing insights into the latest trends and innovations in the field.

In **Chapter 3**, to enhance the utilization of interaction history, this chapter introduces additional constraints by observing the consistency in user behaviors, which assists the model in making more accurate predictions of current user preferences. By leveraging these patterns, the system’s ability is improved to infer user interests even with limited interaction data, thereby addressing sparsity at the user-item interaction level.

In **Chapter 4**, recognizing the limitations of simple linear connections in balancing personalization and product compatibility, a novel concept termed Indirect Personal Compatibility is proposed. This approach enables the model to naturally intertwine compatibility and personalization over iterative training cycles, resulting in recommendations that more closely correspond with individual preferences while maintaining item compatibility coherently.

Additionally, in **Chapter 5**, to enhance latent representation modeling in a sparse

data environment, the implicit connection between user personalization and product compatibility within a homogeneous modality is explored through an innovative contrastive learning manner. Besides, adaptive collaborative signal selectors are designed to overcome data noise problems when exploring historical interaction as auxiliary information.

Last but not least, in product sequential recommendation in **Chapter 6**, besides a multi-scale transformer architecture, which integrates information across different time scales and allows the system to capture both long-term and short-term preference within a unified framework. The data sparsity issue is addressed by identifying similarities among items the user has interacted with, filling the sparse interaction matrix across various behaviors. Besides, multiple data mining and augmentation techniques are explored to expand upon the available data, uncovering new possibilities for enhancing recommendation accuracy in sparse environments.

In **Chapter 7**, through extensive experimentation across two tasks and multiple benchmark datasets, the effectiveness of all proposed methods in addressing data sparsity in fashion recommendation systems is measured. The results demonstrate significant improvements in recommendation performance, highlighting the potential of these approaches to enhance both the robustness and accuracy of fashion recommendations in data-scarce settings.

Chapter 2. Literature Review

In this chapter, we first give the classic evaluation metrics for recommendations systems, and present a comprehensive literature review on personalized data-driven fashion recommendations, providing a thorough overview of the latest advancements in this field. The review covers various perspectives, starting with a detailed examination of the technical aspects. We offer insights into the typical technologies and methodologies relevant to the creation of innovative solutions in personalized fashion recommendations.

2.1 Evaluations for Recommendations

Researchers commonly reserve a portion of the collected data for evaluation purposes, according to specific task requirements. In the context of fashion complementary recommendation, the test set comprises pre-defined outfits presumed to be compatible. Within each test sample, a single item is designated as the target for the positive sample and leaves a vacancy for the recommendation model to predict. The performance of the model is then assessed based on its ability to accurately predict the positive item. In this subsection, the 4 most commonly used evaluation metrics are introduced in detail.

AUC: The area under the ROC curve (AUC) (Zhang et al., 2013), which quantifies the model's capacity to distinguish between positive and negative samples, is utilized to evaluate the proportion of successful predictions relative to the ranking of all samples. This metric provides a comprehensive measure of the model's discriminative power across various threshold settings. A higher AUC value indicates better performance, with a value of 1.0 representing perfect ranking and 0.5 indicating random performance. It can be formulated as:

$$AUC = \frac{1}{|U|}\sum_{u\in U}\frac{1}{|P_u||N_u|}\sum_{i\in P_u}\sum_{j\in N_u} f(r_{ui} > r_{uj}) \tag{2-1}$$

where U is the set of users, P_u and N_u are the sets of positive and negative items for user u, respectively. r_{ui} is the predicted score for item i for user u. $f(\cdot)$ is the indicator function if the condition is true and 0 otherwise. AUC measures how well the model can distinguish between positive and negative matching products.

Figure 2-1 AUC example for complementary recommendations.

HR: Hit Ratio (HR@N), also known as Recall, measures the rate of relevant items that are successfully retrieved by the recommendation system. Specifically, HR is defined as the proportion of target items that appear in the top-N recommendations. HR is particularly useful for assessing a recommendation system to capture target items for a user. A higher HR indicates a better prediction ability.

HR is calculated as:

$$HR@N = \frac{1}{|U|}\sum_{u \in U} f(hit_u) \tag{2-7}$$

where hit_u is 1 if at least one of the relevant items for user u is the top-N recommendations, otherwise is 0.

NDCG: Normalized Discounted Cumulative Gain (NDCG@N), is a metric that evaluates the quality of recommendations by considering the position of relevant items in the ranked list. It is based on the premise that items appearing earlier in the recommendation list are more valuable than those appearing later. NDCG is calculated by normalizing the Discounted Cumulative Gain (DCG) at each position in the recommendation list. The normalization ensures that the metric is bounded between 0 and 1, with higher values indicating better ranking quality. NDCG is formulated as:

$$NDCG@N = \frac{1}{|U|}\sum_{u \in U}\frac{DCG@N_u}{IDCG@N_u} \tag{2-8}$$

where DCG is calculated as :

$$DCG@N_u = \sum_{i-1}^{N}\frac{2^{rel_i}-1}{\log_2(i+1)} \tag{2-4}$$

and $IDCG@N_u$ is the idea DCG, which is the maximum possible DCG for the user u.

MRR: Mean Reciprocal Rank (MRR) is a metric used to evaluate the effectiveness of recommendation systems by considering the rank of the first relevant item in the recommendation list. It is defined as the average of the reciprocal ranks of the first relevant item for each user. MRR is particularly useful in scenarios where the position of the first relevant item is of primary importance. A higher MRR value indicates that relevant items are ranked higher in the recommendation list.

$$MRR = \frac{1}{|U|}\sum_{u \in U}\frac{1}{rank_u} \tag{2-5}$$

where $rank_u$ is the rank position of the first relevant item in the recommendation list for user u.

2.2 Datasets

2.2.1 For Complementary Recommendation.

The proposed fashion complementary recommendation models introduced in this work are evaluated on two online open-access fashion datasets, which are composed of real interaction data from IQON3000 (Song et al., 2019) and Polyvore-519 (Lu et al., 2019). Table 2-1 shows the detailed statistics of the two datasets.

Table 2-1 Statistics of the Datasets.

Datasets	Polyvore-519	IQON3000
*Number of users	519	3236
*Number of top items	14721	85936
*Number of bottom items	15570	43086
*Total number of items	30291	129022
*Number of training samples	31933	170601

*Number of validation samples	5110	23095
*Number of testing samples	5197	23095
*Total number of samples	42240	216791

IQON3000 (Song et al., 2019)**:** Since most of the fashion complementary recommendation available datasets lack of user context, which make it difficult to generate personalized recommendations. Song *et al.* (Song et al., 2019) first introduce the IQON3000 dataset for personalized fashion recommendation tasks. The detailed samples we used in our experiment are given in the Table 2-1. Each product is associated with a visual image and textual description including categories, price, and item description. The available visual and textual features provided by Song *et al.*(Song et al., 2019) for IQON3000 are applied.

For visual pre-trained features, each fashion product image is fed into the 50-layer residual network (ResNet50) (He et al., 2016), and adopted the output of the last average pooling layer as the visual representation. Thereby, the visual modality of each product is represented by a 2048-D vector.

For textual modality, only the title description and category metadata is considered as the contextual information of the fashion item. To process the text, they employed the Japanese morphological analyzer Kuromoji [1] for tokenization, then represented each contextual description as a concatenated word vector, with each row representing a constituent word. For word representation, they utilized the 300-D vector from the Japanese word2vec Nwjc2vec, created from the NINJAL Web Japanese Corpus (Shinnou et al., 2017). The single-channel CNN consisted of a convolutional layer over the concatenated word vectors and a max pooling layer. With kernels of sizes 2, 3, 4, and 5, and 100 feature maps for each kernel, the rectified linear unit (ReLU) is applied as the activation function. Ultimately, a 400-D contextual representation for each fashion product is obtained.

[1] http://www.atilika.org/.

Polyvore-519 (Lu et al., 2019)**:** From user-outfit interaction data obtained from (Lu et al., 2019), triplet data samples of user and fashion product pairs were generated, excluding all other item categories and retaining only product pairs of one top and one bottom garment item in each outfit. The Polyvore-519 dataset was further improved by excluding user-product pair samples with fewer than two interactions because it contains a limited amount of interaction data. Table 2-1 provides comprehensive statistics for the two datasets.

For each fashion product, a 2048-D visual feature was obtained using the Resnet152 (He et al., 2016) pre-trained on ImageNet. The original dataset's 2400-D textural features, which were extracted using a pre-trained AlexNet (Krizhevsky et al., 2012) , are accessible for textual modality.

2.2.2 For Sequential Recommendation.

In the sequential recommendation task, we employ three public benchmark datasets that contain various interaction contexts and corresponding behavior information. These datasets are collected from real-world e-commerce platforms, reflecting the complexities and nuances of user-item interactions in practical applications. They have been widely used in the research community to benchmark the performance of multi-behavior recommendations. Table 2-2 provides a summary of these datasets' statistics.

Table 2-2 Statistics of datasets.

Datasets	Retailrocket	Taobao	IJCAI
Number of users	30691	437367	443924
Number of items	31240	99038	782695
Number of transactions	32690	472944	568250
Average Length	14.55	48.23	78.58
Sparsity	99.99%	99.99%	99.99%

Retailrocket i is produced by the website Retailrocket, which logs three different kinds of user behavior: auxiliary behaviors (page view and add-to-cart) and target

behaviors (purchase).

Taobao is one of the biggest e-commerce sites in China. It provides four different kinds of user-item interactions: auxiliary behaviors (add-to-cart, add-to-favorites, and page view) and target behavior (purchases).

IJCAI: In the 2015 IJCAI Contest, IJCAI was made available for the objective of predicting repeat customers. This dataset also includes the four kinds of interaction behaviors: [page view, add-to-favorites, add-to-cart, purchase].

2.3 Overview of Data-driven Recommendation Systems

Recommender systems have become an integral part of many online platforms and services, enhancing user experience and driving user engagement and retention. It primarily focuses on two core components: users and items. Users express their preferences for items in two ways: implicit or explicit feedback. Implicit feedback is inferred from a user's interactions, such as the time spent on a webpage or clickstream data. In contrast, explicit feedback is direct, where a user specifies their preference on a particular scale or interval, like a rating out of 5 stars. While many recommender systems leverage both implicit and explicit feedback, these are typically organized in a utility matrix, as demonstrated in Figure 2-2.

Frequently, the utility matrix contains numerous missing values, as shown in Figure 2-3. The main objective of recommender systems is to discover the missing values in the interaction matrix. Because consumers frequently rate a limited number of products, the initial matrix is typically relatively sparse, making this a difficult task. Additionally, as only those products will be suggested to users, we are only interested in those with high user ratings. The method used and the type of data source—contextual (weather, seasons, etc.), textual, visual, etc.—have a significant impact on how effective a recommender system is (Roy & Dutta, 2022).

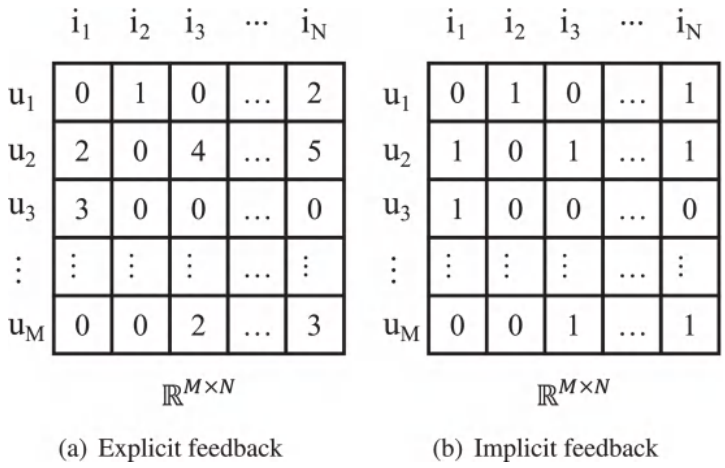

Figure 2-2 Interaction matrices of explicit and implicit feedback.

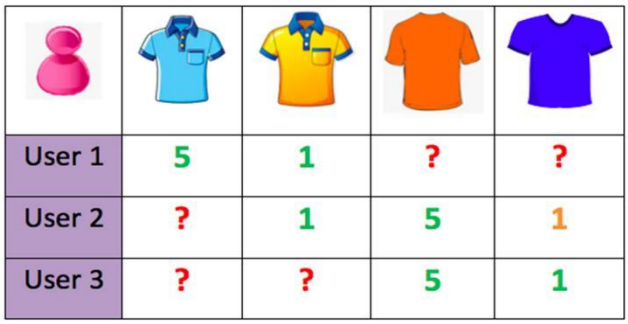

Figure 2-3 User-item interaction example (Chakraborty et al., 2021).

2.3.1 Content-based Recommendation.

A content-based recommender system suggests items with similar attributes by comparing the items' content and the user's preferences based on descriptions of items the user has interacted with. Several item profiles are made according to their features or description. For instance, the qualities of clothing will contain elements like color and brand. The other components in that item profile are merged to form a user profile when a user evaluates an item positively. All of the item profiles whose items the user has given a positive rating are combined into one user profile. (Roy & Dutta, 2022). Items in this user profile are then recommended to the user, as illustrated in Figure 2-4.

For **item profiles**, each item in the database is represented in terms of several descriptors or terms that are inherent to the item. This could be word expression in a piece of clothing in the case of fashion recommendations or visual attributes of a product (like the style or fabric).

For **user profiles**, they represent the preferences of a user in terms of the same descriptors or terms used for the items. User profile is built by analyzing the content the user has interacted with and gathering the terms from these items. For instance, if a user has liked several short dresses, the term "short" will have a high weight in the user's profile.

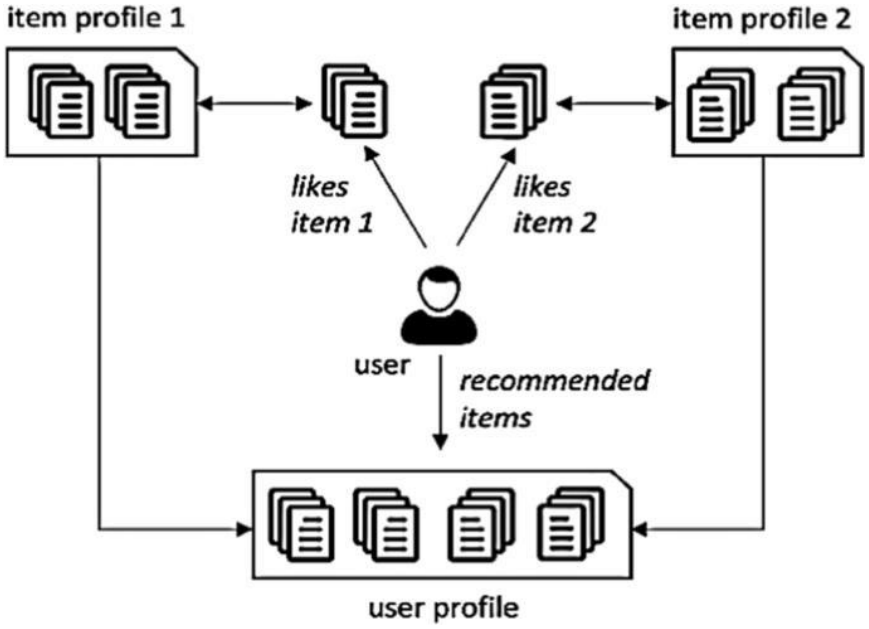

Figure 2-4 Content-based recommender system.(Roy & Dutta, 2022)

To make a recommendation, the system compares the user's profile with the item profiles and suggests items with the highest similarity. Various similarity measures like cosine similarity, Jaccard similarity (Jaccard, 1908), or Euclidean distance (Danielsson, 1980) can be used.

Advantages: (1) There is no cold start problem for items. Unlike collaborative filtering, content-based systems can recommend new items that have not been interacted with yet. (2) Additionally, since content-based recommenders are customized for each user based on their interactions, they are typically used in personalized suggestions. Because a user's profile is unique to them, this algorithm doesn't need other users' profile information because it doesn't influence their recommendation process. (3) Last but not least, we can explain why a user made a specific recommendation, because a user has shown interest in similar items in the past, as shown as transparency.

Disadvantages: (1) A limitation of this technique is that it necessitates a comprehensive understanding of the characteristics of the item in order to provide an appropriate recommendation. This information may not be always available for all items. Also, the extraction and utilization of features is crucial for the recommended results, and it is difficult to achieve good results if the special engineering is not mature enough. (2) Over-specialization (Roy & Dutta, 2022), if a user has shown interest in only a particular type of content, the system might end up recommending only that type of content and might miss out on diversifying the user's experience. (3) There is a cold start problem for new users, the system might not have enough information to make relevant recommendations, since content-based recommendation relies on user's historical interactions to give suggestions.

2.3.2 Collaborative filtering-based Recommendation.

Collaborative Filtering (CF) (Resnick et al., 1994) gives recommendation based on user-item interactions, especially highly relies on other users' interactions, and captures the idea that users who have agreed in the past tend to agree again in the future about the preference for certain items. For example, for a given user, the algorithm identifies a set of users as the given user's "neighborhood", who share similar preferences with this given user. The efficiency of a collaborative algorithm depends on how accurately the algorithm can find the neighborhoods of the target user (Roy & Dutta, 2022). There are several types of collaborative filtering approaches, which can be broadly categorized into Memory-Based Collaborative Filtering, Model-Based Collaborative Filtering.

2.3.2.1 Memory-Based Collaborative Filtering

The utility matrix is directly used for prediction in memory-based collaborative filtering techniques, which suggest new items based on user neighbors' preferences. The two categories of memory-based collaborative techniques are item-based collaborative filtering and user-based collaborative filtering. In this context, memory may refer to previously recorded interactions.

User-Based Collaborative Filtering (User-CF) recommends items based on the similarity between users. If user *A* and user *B* have rated items similarly in the past, then items liked by user *A* and user *B* has not yet interacted with will be recommended to user *B* and vice versa.

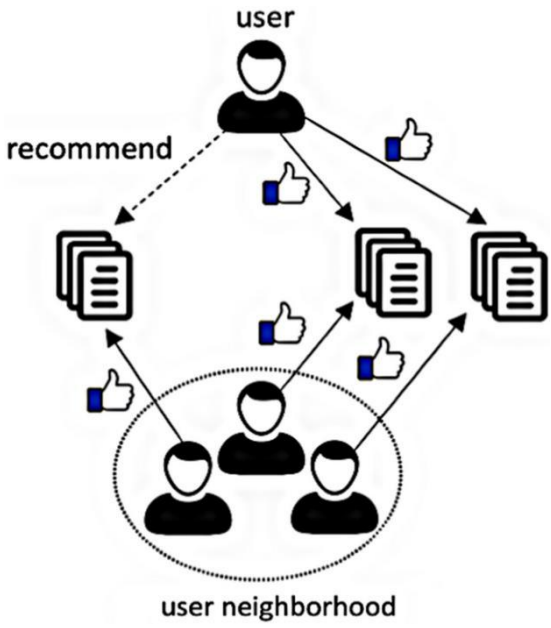

Figure 2-5 User-based collaborative filtering (Roy & Dutta, 2022).

Item-Based Collaborative Filtering (Item-CF) focuses on finding similar items based on users' interactions. If items *X* and *Y* have been liked or interacted with by the same set of users, then they are considered to be similar.

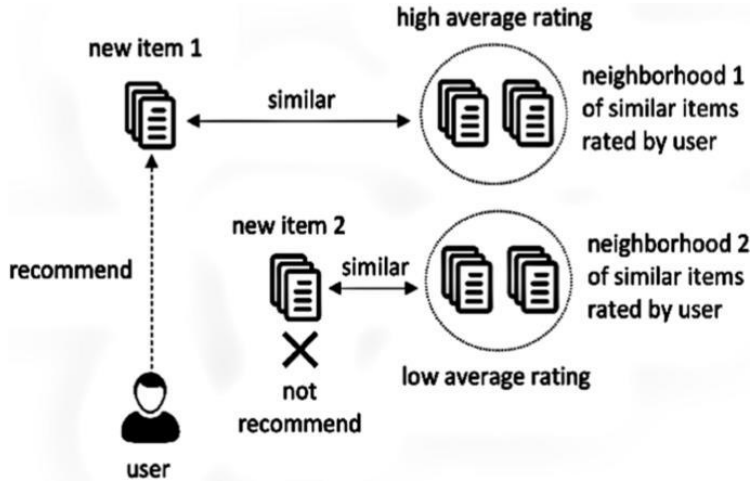

Figure 2-6 Item-based collaborative filtering (Roy & Dutta, 2022).

Advantages: (1) Memory-Based Collaborative Filtering is relatively simple and straightforward to implement, especially when dealing with user-item matrices. (2)

New user-item interactions can be incorporated into the recommendation system in real-time, which means that as soon as a user rates an item, the recommendation for that user and others can be updated instantly. (3) The recommendations are based directly on historical user-item interactions, making it easier to explain why a particular item was recommended (e.g., "Users who liked this item also liked...").

Disadvantages: (1) For large datasets, memory-based methods can be computationally expensive for large datasets require computing similarities between users or items (2) In the real world, user-item matrices are typically very sparse, which indicates that the majority of users have only rated a small portion of all available items, because of this sparsity, similarity measures may not be as accurate as they could be. (3) The Cold Start Problem may degrade the performance when there is no previous data to draw conclusions from, making it difficult to recommend new items or users who have not yet interacted with the system. (4) Additionally, popularity bias problem often occurs in memory-based CF methods, since there always is a tendency to recommend popular items, and there's a possibility that it will overlook niche items that might be of interest to certain users.

2.3.2.2 Model-Based Collaborative Filtering

In order to build a prediction model that can derive the user's rating for an item that hasn't been rated from user-item interaction data, model-based systems use a variety of data mining and machine learning technologies. Different from other methods, model-based methods do not include a new user's profile in the utility matrix before making predictions. Recommendations can be sent to users who are not yet supported by the model. A collection of products can be immediately recommended by them using the pre-trained model. The accuracy of this method is greatly influenced by the effectiveness of the underlying learning algorithm that was used to build the model.

The most popular model-based collaborative filtering method is low rank Matrix Factorization (MF) (Koren et al., 2009), which is a method used predominantly in recommendation systems and is closely related to Singular Value Decomposition (SVD)

(Golub & Reinsch, 1971) in linear algebra. The core idea is to represent a large user-item interaction matrix by the product of two lower-rank matrices, capturing the latent factors associated with users and items. As users and items grow larger, it is difficult for purely memory-based collaborative filtering to take on such a difficult computation. Take Figure 2-7 as an example.

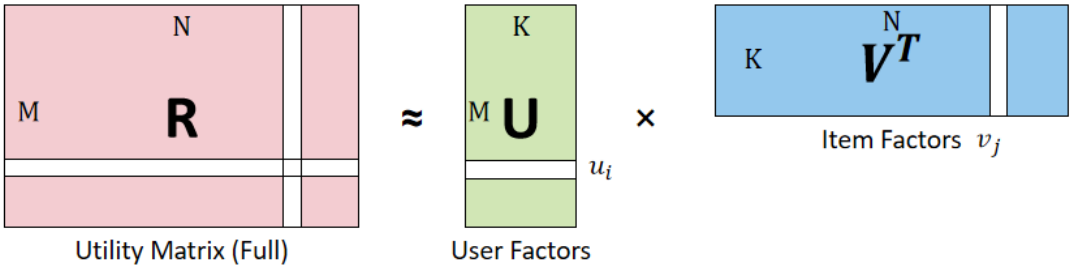

Figure 2-7 Matrix Factorization.

In many recommendation scenarios, we have a matrix R of size $M \times N$, where M is the number of users and N is the number of items. Most entries in R are missing, which represent unknown user-item interactions. The known entries are typically the ratings given by users to items. The goal of MF is to approximate R by factorizing it into two matrices U and V, such that:

$$R \approx U \times V^T \tag{2-6}$$

where U is an $M \times K$ matrix and V is an $N \times K$ matrix, representing the latent factors for users and items. K is the number of latent factors (typically $K \ll \min(M, N)$, hence "low rank"). The idea is that while we can have a large number of users and items, the actual number of latent factors influencing user-item interactions (like color in clothing, genres in movies, topics in articles, etc.) might be much smaller. By reducing our data to these latent factors, we can potentially discover patterns that are not immediately apparent. The rank of both the row space and the column space of R is K. Each column of V can be regarded as one of the row space basis vectors of R and the columns of U can be regarded as the corresponding coefficients. Here, the factorization of a matrix of rank K for different sets of basis vectors may have infinitely many solutions. Even if the rank of R is greater than K, it can be approximated as a product of factors whose rank is K. The error of this approximation is equal to $\|R - U \times V^T\|^2$.

For optimization, finding U and V typically involves minimizing the difference between the known entries in R and the corresponding entries in. A popular objective function to minimize is the Mean Squared Error (MSE) (Wang & Bovik, 2009):

$$Minimize\ J = \frac{1}{2}\|R - U \times V^T\|^2 \tag{2-7}$$

The smaller the objective function, the better the quality of the factorization. However, in contexts with a large number of missing values, only a subset of R is known, so the objective function is also uncertain. In order to learn U and V, the objective function needs to be rewritten based only on the observed values. The benefit, however, is that once the latent factors U and V have been learned, the entire scoring matrix can be reconstructed in one go using $U \times V^T$. Let S denote the set consisting of all user-item pairs (i, j) known in R. Then the predicted value of the (i, j) position of the matrix R will be as follows:

$$\hat{r}_{ij} = \sum_{s=1}^{K} u_{is} \cdot v_{js} \tag{2-8}$$

where $\hat{r}_{uj}$ is the predicted value not the observed value, then the optimization of the objective function of the incomplete matrix modified using the known values in S with added regularization will be shown below:

$$L(U, V) = \frac{1}{2}\sum_{(i,j)\in S}\left(r_{i,j} - \sum_{s=1}^{K} u_{is} \cdot v_{js}\right)^2 + \frac{\lambda}{2}\left(\|U\|_F^2 + \|V\|_F^2\right) \tag{2-9}$$

where λ is the penalty coefficient on the learned factors.

Advantages: (1) Once the model is trained, generating recommendations is typically faster and more scalable than computing similarities in memory-based approaches. (2) Model-based methods, especially matrix factorization techniques, can handle sparse data better by embedding users and items in a lower-dimensional latent space. (3) By learning latent factors, models trained from model-based collaborative filtering often capture deeper user-item interactions and can offer more personalized recommendations.

Disadvantages: (1) Building an effective recommendation model can be time-consuming, especially with complex models or large datasets. (2) While it's a challenge in both types, model-based collaborative filtering methods also struggle with new items or users that weren't in the training data. (3) Some model-based collaborative filtering methods, especially complex ones like deep learning models, can be harder to interpret compared to the transparent user-item based approaches of memory-based CF, as latent factors are not explicitly representing users' interest.

2.3.3 Knowledge-based Recommendation.

Knowledge-based recommendation is a type of recommender system that relies on explicit knowledge about users and items to generate personalized recommendations. The system typically uses explicit information provided by the user, either through a set of questions or direct input, to generate tailored recommendations, and which is particularly useful when user-item interactions are sparse and it's hard to compute reliable recommendations. For example, in many scenarios, especially in niche markets, users rate only a tiny fraction of the entire item set, in that scenario, knowledge-based methods can be beneficial.

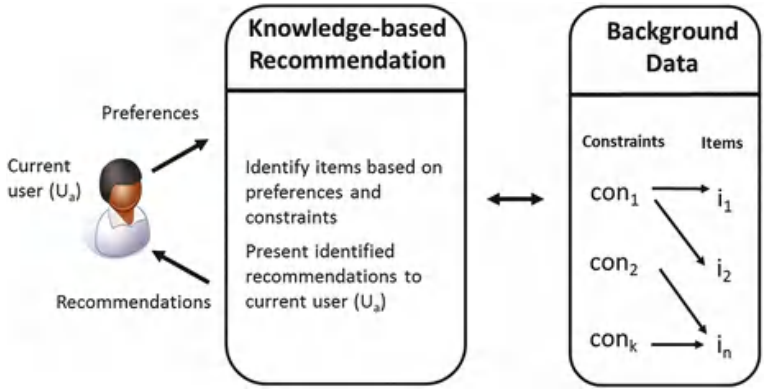

Figure 2-8 Knowledge-based recommendation (KBR) dataflow (Felfernig et al., 2014).

Knowledge-based recommendation systems suggest items to learners by leveraging domain-specific insights on how various items meet the needs of learners. These systems rely on three pivotal types of information: details about the learners, characteristics of the items, and insights on how well an item aligns with a learner's

requirements. Take Figure 2-8 as an example, the system collects explicit information from users, either through direct input, a questionnaire, or by observing user behavior over time to do a user profiling. Additionally, detailed information about each item is stored, allowing for robust matching against user profiles or queries. Then the system matches user profiles or queries against item profiles, often using sophisticated algorithms to determine the best fits. Within the e-learning context, these strategies compile information regarding both the learners and learning materials to enhance the recommendation process (Murtaza et al., 2022).

There are two types of knowledge-based recommendations, constraint-based and case-based recommendations. In constraint-based recommendations, users should provide explicit requirements or constraints, and the system recommends items that meet those constraints. For example, in a fashion recommendation system, a user might specify category, desired scenarios, and size, and the system would suggest fitting outfit or fashion pieces. In case-based recommendations, the system recommends items by finding similar past examples or cases and adapting them to the current user's needs. Usually, it then uses similarity metrics to find past cases that match the user's requirements.

Advantages: (1) Unlike collaborative or content-based systems, knowledge-based methods can handle situations where little to no data exists about users or items. (2) Since recommendations are based on explicit user requirements or queries, it's often easier to provide explanations for recommended items.

Disadvantages: (1) Users need to provide explicit information, which can be tedious and lead to lower engagement. (2) As the system relies heavily on user-defined criteria, there might be fewer unexpected or serendipitous recommendations.

2.3.4 Hybrid Recommendation.

A hybrid approach combines two or more recommendation methods to overcome the limitations inherent in individual techniques (Chakraborty et al., 2021). By blending different strategies, hybrid methods can be executed in several ways, such as weight the

output from each technique and combine them; or conduct a future combination where features from different sources or techniques are combined in a unified recommendation model; or mix the output of different recommender systems and the combined result is given as a recommendation. Take Figure 2-9 as an example, the hybrid filtering technique assumes the content-based filtering (R1) and collaborative filtering (R2) results. It then determines the weights of these results as R3, combines the results by influencing the higher weighted result, and suggests the final product as R4. (Chakraborty et al., 2021).

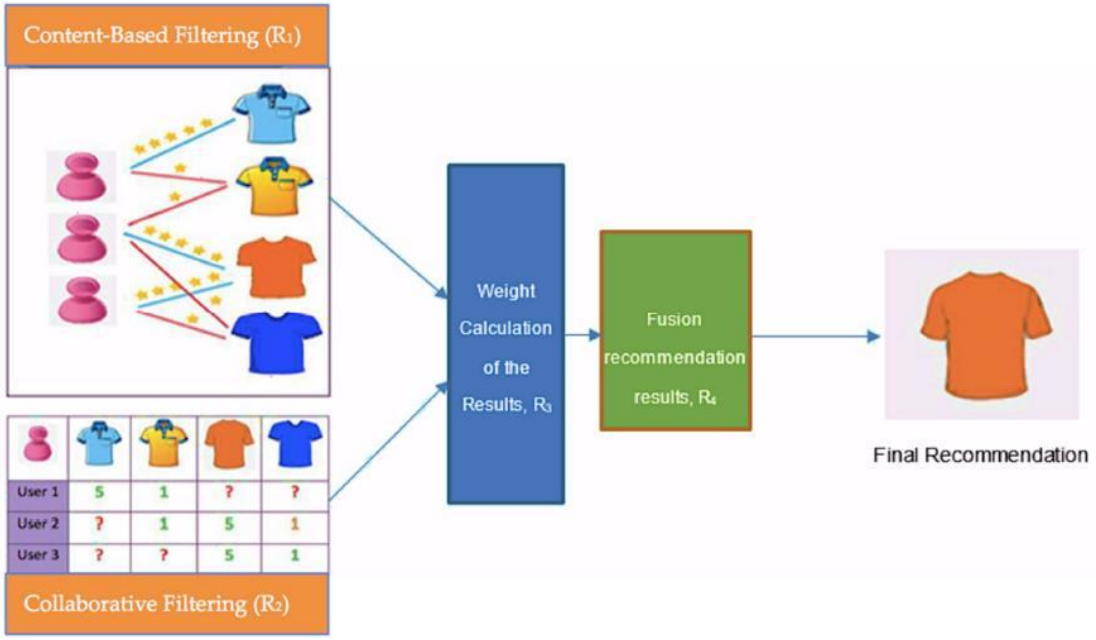

Figure 2-9 Hybrid filtering process (Chakraborty et al., 2021).

With the widespread use of deep learning and the maturity of various visual and text-based information extraction techniques, almost all existing recommender systems, especially in the research field, use hybrid recommender system algorithms. Existing recommender system approaches are also no longer limited to one or a collection of approaches, but rather modify known algorithms based on the available data and the purpose of the recommendation to achieve personalized and diverse recommendations. Take the following methods and those extension applications as examples:

Word2Vec (Church, 2017) and other extensions (Item2Vec (Barkan & Koenigstein, 2016), User2Vec (Hallac et al., 2019), Doc2Vec (Lau & Baldwin, 2016), Neu-Item2Vec

(Barkan et al., 2020)) are word embedding methods that commonly used in Natural Language Processing (NLP) tasks. However, in recommender systems, Word2Vec can be used to learn embedded representations of items or users. While Word2Vec is used in recommender systems, it can be used for both content-based and collaborative filtering-based recommendations. If we use Word2Vec to learn embeddings of item descriptions, titles, or any other text related to the content of an item, then it is content-based recommendation. For example, we can use Word2Vec to learn word vectors for each item description and recommend similar items to the user by calculating the similarity between the vectors. In other cases, it's more like collaborative filtering-based recommendation because it's based on the interaction between the user and the item. For example, Word2Vec can be used to learn the sequential behavior of users or items. We can think of the user's click stream (the sequence of items clicked) as a "sentence" where each item is a "word".

In addition, hybrid recommendation methods can also be utilized from the following intuitions:

- Factorization Machines (FM) (Xue et al., 2017) are a type of relation-based model that capture the interactions between user and item features in addition to the user-item interactions. FM can handle both sparse and dense feature interactions efficiently.
- Field-aware Factorization Machines (FFM) (Juan et al., 2016) is an extension of FM that considers feature interactions across different fields (e.g., user features and item features). It is particularly useful when dealing with feature interactions in large-scale sparse datasets.
- xDeepFM: Extreme version of the Deep Factorization Machine (Lian et al., 2018)extends DeepFM (Guo et al., 2017) by incorporating a cross network component, which models higher-order feature interactions in a more explicit way. The cross network captures feature interactions across different dimensions and helps improve the model's representation learning capability.

- Attentional Factorization Machines (AFM) (Xiao et al., 2017) enhances the FM component of DeepFM (Guo et al., 2017) by incorporating an attention mechanism to automatically learn the importance of different feature interactions. Then uses attention weights to weigh the importance of different feature interactions when predicting user-item preferences.
- Deep & Cross Network (DCN) (Wang et al., 2017; Wang et al., 2021) combines the deep neural network with cross network that explicitly models feature interactions.
- NFM: (He & Chua, 2017) Neural Factorization Machines is a variant of FM that replaces the dot product operation with a neural network layer, allowing the model to learn non-linear feature interactions, which combines the benefits of FM and deep learning to capture more complex feature interactions.
- Graph Neural Network recently has been explored as a powerful tool for fashion recommendation systems, leveraging their ability to model complex relationships and interactions within data. In the context of fashion recommendation, GNNs are particularly effective due to the inherent graph-like structure of fashion data, where items, users, and their interactions can be represented as nodes and edges in a graph.
- Attention-based Neural Networks, particularly transformers, have gained significant traction in various domains due to their ability to model complex dependencies and capture contextual information effectively. Attention-based networks can model intricate relationships by dynamically adjusting the focus on different aspects of the data, allowing for more sophisticated recommendations that consider multiple factors simultaneously.

For the limitation, the vast majority of these models often require a large amount of data to learn complex and meaningful representations effectively, which can be challenging for those systems that only have limited data availability.

2.3.5 Advanced Techniques

2.3.5.1 Personalized Ranking Techniques in Recommender Systems

a) BPR:

BPR stands for Bayesian Personalized Ranking (Rendle et al., 2012), is a well-known algorithm that is widely used in collaborative filtering-based recommender systems (Hu et al., 2008; Liang et al., 2016), especially with implicit feedback. The primary objective is to address the personalized ranking problem, with the goal of rating products for each user according to their preferences. BPR's fundamental notion is to learn user and item embeddings (latent representations) using user-item interaction data, such as user-item ratings or binary interactions like clicks or purchases. BPR learns how to prioritize positive item interactions over negative ones for each user utilizing a pair-wise learning-to-rank approach. For example, combined with the mainstream idea of Matrix Factorization (MF) model (Koren et al., 2021), the preference based on user-item interactions with a classic dot-product based on user and item latent representation can be obtained as follows:

$$s_{ui} = e_u^T e_i + \alpha + \beta_u + \beta_i \quad (2\text{-}10)$$

where α is the global bias, β_u and β_i are bias for user u and item i, e_u and e_i are K-dimensional vectors representing latent factors associated with user u and item i, respectively. The inner product $e_u^T e_i$ quantifies the **compatibility** between user u and item i, indicating the level to which the product's inherent "properties" and the user's latent "preferences" coincide. In other words, it indicates how well the item fits the user's preferences based on the learned latent elements by measuring the degree of alignment between the user's preferences and the item's physical attributes (He & McAuley, 2016).

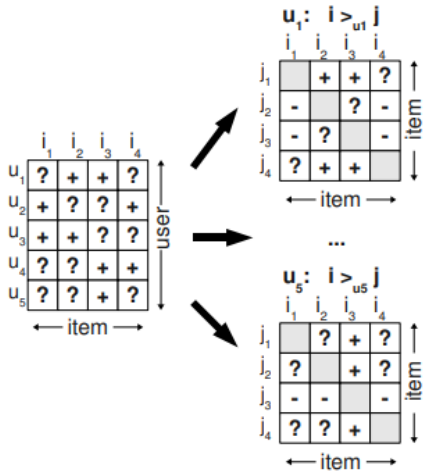

Figure 2-10 Basic Idea of BPR. (Rendle et al., 2012)

The MF-BPR approach aims to rank items that match a user's interests high and those that the user is not interested in low. Therefore, the model trained by BPR loss can be represented as:

$$BPR - opt = \sum_{D} ln\sigma(s_{ui} - s_{uj}) - \lambda\|\Theta_F\|^2 \tag{2-11}$$

The positive item set (i) represents the items that the user has interacted with, indicating their interest. In contrast, the negative item set (j) is assumed to be the complement of the positive set, as users do not explicitly label items they are not interested in. The objective is to rank the items that the user had interacted higher than those the user is not interested in. However, in the context of user-item interactions, the negative item set is not explicitly defined since users typically do not provide explicit feedback on items they are not interested in. Therefore, a common assumption made in this scenario is that items that have not been interacted with by the user form the negative set for that user. In other words, any item does not present in the user's interaction history is considered as a negative item for recommendation purposes. This assumption allows for the creation of a contrastive learning setup, where the model learns to distinguish positive interactions (items the user has interacted with) from negative interactions (items the user has not interacted with) in the training process, enabling effective personalized recommendations (Ding, Lai, et al., 2023b).

b) VBPR:

In the domain of fashion, where visual factors play a significant role in purchase decisions, personalized fashion recommendation methods have been developed based on Collaborative Filtering (CF) frameworks (Hu et al., 2008; He & McAuley, 2016), focusing on integrating visual information to improve recommendation performance.

Incorporating visual factors of fashion products not only enhances recommendation quality but also addresses the cold-start issue. Although latent factor models theoretically have the capability to capture any relevant dimensions in a recommendation system, they encounter a significant challenge with the presence of **cold items**. These cold items have too few associated observations, making it difficult to accurately estimate their latent dimensions. To address this issue, explicit features can serve as an auxiliary signal in such cases. He *et al.* (He & McAuley, 2016) firstly introduced **V**isual **B**ayesian **P**ersonalized **R**anking from implicit feedback (**VBPR**) with visual factors. They suggested partitioning the rating dimensions into two categories: visual factors and latent (non-visual) factors. By incorporating explicit visual features, representation of items with limited observations can be enhanced, leading to more effective recommendations and improved performance of the recommendation system.

The overall prediction can be summarized as follows:

$$s_{ui} = e_u^T e_i + \alpha + \beta_u + \beta_i + v_u^T v_i \quad (2\text{-}12)$$

where v_u and v_i represent D-dimensional visual factors, introduced to capture the visual interaction between user u and item i. As shown in Figure 2-11, these latent factors quantify the level of attraction or affinity that user u has towards each of the D visual dimensions present in item i.

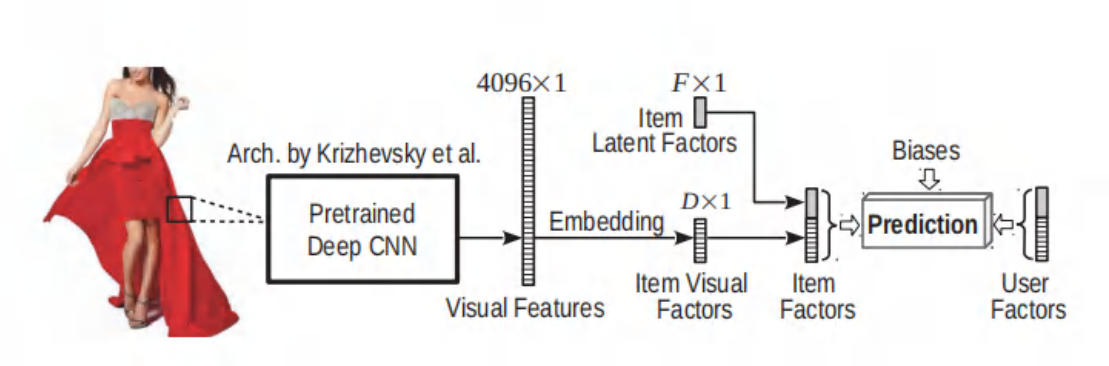

Figure 2-11 Basic Idea of VBPR (He & McAuley, 2016).

In their approach, they utilize Convolutional Neural Network (CNN) features, representing the high-dimensional characteristics of fashion items, and project them onto a lower-dimensional visual rating space to extract the visual factors of each item. This projection is achieved through the use of a projection matrix W: $\theta_i = W v_i$, transforming the original CNN features v_i into the visual factor θ_i The visual factor θ_i serves as a compact and informative representation capturing the essential visual attributes of the fashion item.

2.3.5.2 Training Technique

Contrastive Learning: The utilization of contrastive learning, a self-supervised learning technique, is prevalent in the domains of machine learning and deep learning (Chuang et al., 2020; Tian et al., 2020). The basic objective of contrastive learning is to obtain meaningful data representations by leveraging the relationships between comparable and non-comparable samples. In the context of contrastive learning, the model undergoes training to distinguish between positive pairings, which refer to samples that exhibit similarity, and negative pairs, which refer to samples that exhibit dissimilarity, within the dataset. The objective is to aggregate comparable sample representations in order to minimize their distance in the latent space, while ensuring that different sample representations are maximally separated from one other. This allows the model to reveal representations of the data that effectively capture its underlying structure and semantics. In the context of contrastive learning, the contrastive loss function holds significant importance. The objective of the method is

to minimize the spatial distance between positive pairs and maximize the spatial distance between negative pairs. Prominent examples of contrastive loss functions encompass "Information Noise Contrastive Estimation," also referred to as "InfoNCE," as proposed by (Gutmann & Hyvärinen, 2010), "triplet loss," introduced by Zhao *et al.* (Zhao et al., 2019), and "N-Pair loss," developed by (Chen & Deng, 2019), among other alternatives.

Contrastive learning has shown promising results in a variety of application domains, such as natural language processing, computer vision, and recommendation systems, among others. Image representation learning, text embedding, and collaborative filtering are some examples of the applications that have been developed using it. The ability of contrastive learning to learn powerful representations from unlabeled data, which can then be fine-tuned for downstream tasks with limited labeled data, is one of the most significant advantages of this method of machine learning. Because of this property, contrastive learning is an effective method for jobs that only have a few examples that have been labeled.

In this subsection, we only introduce InfoNCE (Information Noise Contrastive Estimation) (Gutmann & Hyvärinen, 2010) in detail. Deep neural networks are trained using the contrastive learning technique InfoNCE for a range of applications, such as recommendation systems, self-supervised learning, and unsupervised representation learning. By maximizing the agreement between similar pairs of data samples and limiting the agreement between dissimilar pairs, InfoNCE aims to discover meaningful representations. Recommendation systems employ InfoNCE to learn representations of individuals and items (fashion products, for example) that show how compatible and personal they are. It is particularly useful in situations with limited or insufficient interaction data, when traditional approaches might find it difficult to produce diverse and reliable recommendations.

The InfoNCE loss function aims to maximize mutual information between positive pairs (pairs of similar items or user-item interactions) while minimizing mutual

information between negative pairs (pairs of dissimilar items or user-item non-interactions). The loss function motivates the model to differentiate between positive and negative pairings, enabling it to learn meaningful representations that reflect the underlying relationships between items and users. Suppose we have a dataset with positive samples (denoted as z^+) and negative samples (denoted as z^-). Let $f(z)$ be the representation function that maps a sample z to its corresponding representation.

InfoNCE loss for a positive pair (z^+, z^{+1}) is given by:

$$loss = -log\left[\frac{exp\left(f(z^+)\cdot f\left(z^{+1}\right)/t\right)}{\left(exp\left(f(z^+)\cdot f\left(z^{+1}\right)/t\right)+\sum exp\left(f(z^+)\cdot f(z^-)\right)/t\right)}\right] \quad (2\text{-}2)$$

Here, represents the dot product between the representations, and the sum in the denominator is taken over all negative samples. The goal of this loss is to maximize the similarity between positive samples (z^+, z^{+1}) while minimizing the similarity between positive samples $(z^+,)$ and negative samples (z^-). By doing so, the model learns to distinguish positive samples from negative samples and obtains meaningful representations that capture the underlying structure of the data.

The temperature parameter t in InfoNCE is essential for adjusting the sharpness of the SoftMax function used to calculate the probabilities of positive and negative pairings. When t is large (e.g., greater than 1), the SoftMax function produces a gentler and more uniform distribution of probabilities, resulting in a more exploratory behavior. In this instance, the model becomes more uncertain and tends to attribute more similar probabilities to positive and negative pairs, which can encourage the model to investigate more diverse and meaningful representations. When t is small (e.g., less than 1), however, the SoftMax function becomes peakier, resulting in higher probabilities for positive pairs and lower probabilities for negative pairs. This can lead to more discriminatory behavior, where the model concentrates more on differentiating positive and negative pairs, which may result in overfitting and less diverse representations. The choice of the temperature parameter t in InfoNCE is dependent on the particular assignment and dataset. A larger value of t may be preferred when the

data is sparse or when the task requires more exploration and representational diversity. On the other hand, a lesser value of t may be appropriate for tasks where discriminative representations are more crucial.

2.3.5.3 Data Processing Technique

Normalization (Ioffe & Szegedy, 2015) has become a standard component in many modern deep learning architectures and has significantly contributed to the success and widespread adoption of deep neural networks in various applications. Given a mini-batch of activations, denoted as $B = \{x_1, x_2, x_3, \cdots, x_m\}$ in a layer, where m is the batch size, the first step is to calculate the mean (μ_B) and variance (σ_B^2) of the mini-batch:

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{2-3}$$

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2 \tag{2-4}$$

Then the normalized result of the batch using the mean (μ_B) and variance (σ_B^2) can be calculated as follows:

$$\bar{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \tag{2-5}$$

Here, ϵ is a small constant. After normalization, the activations are then rescaled and shifted using learnable parameters γ (scale) and β (shift) to introduce flexibility:

$$y_i = \gamma \bar{x}_i + \beta \tag{2-6}$$

The normalized activations y_i are then fed as input to the next layer. During the training process, the parameters γ and β are learned through backpropagation along with other model parameters. Batch Normalization offers several advantages, for example, the stable training: by reducing internal covariate shift, Batch Normalization ensures more stable and efficient training. This allows for higher learning rates, leading to quicker convergence and faster training. On the other hand, it could act as a regularization technique, as Batch Normalization reduces the need for other

regularization methods like dropout. It assists in preventing overfitting. Additionally, especially in deep neural networks, it minimizes the problem of exploding gradients during training. Neural networks become less sensitive to the learning rate selection as a result, which facilitates smoother and more reliable optimization and makes it simpler to determine the beneficial learning rate for training process.

2.3.5.4 State-of-the-art Method for Complementary Recommendation (GP-BPR)

Based on BPR and VBPR, Song *et al.* further introduces an innovative framework that leverages matrix factorization and Bayesian Personalized Ranking to tackle the challenges of personalized fashion compatibility modeling which called Personalized Compatibility Modeling for Clothing Matching (**GP-BPR**) (Song et al., 2019). By capturing the intricate relationships between fashion items and individual user preferences, GP-BPR offers a promising solution to enhance fashion recommendation systems and create more tailored and fashionable outfit suggestions for users.

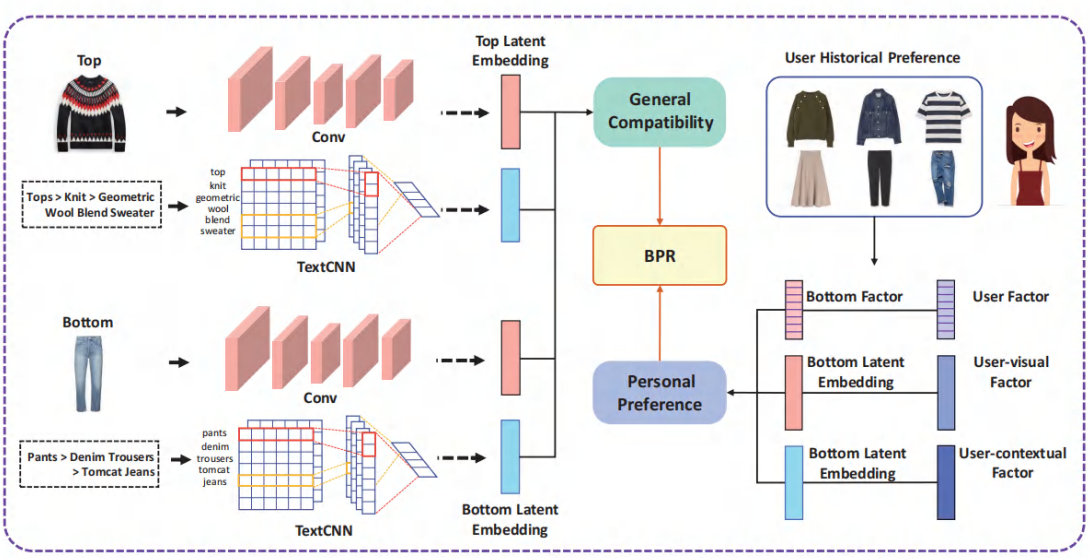

Figure 2-12 Basic Idea of GP-BPR (Song et al., 2019).

As shown in Figure 2-12, the general compatibility between fashion items is captured by using a latent space where the distance between their latent representations. To model the complex attribute interactions, a multi-layer perceptron (MLP) is employed for representation learning. Each fashion item is associated with multiple modalities, such as visual images and contextual information, which coherently

characterize the item. Both modalities are utilized to enhance the general compatibility between fashion items, and the measurement is designed as:

$$s_{gr} = \pi\left(v_g\right)^T v_r + \left(1-\pi\right)\left(w_g\right)^T w_r \quad (2\text{-}17)$$

where v_g and w_g are the visual and textual representation obtained from the MLP of the given product, similarly, the v_r and w_r are the latent corresponding feature of the target recommended product. As for the personal preference modeling towards the recommended product r, which resorts to the matrix factorization framework followed by VBPR(He & McAuley, 2016) plus textual factors (w_u, w_r), can be summarized as follows:

$$s_{ur} = (e_u)^T e_r + \eta(v_u)^T v_r + (1-\eta)(w_u)^T w_r + \beta_u + \beta_r + \alpha \quad (2\text{-}18)$$

The inner product between them conveys the visual preference interaction between the user and the recommended product r, and the final overall preference score combined both general compatibility and personal preference is linearly mapped by the trade-off manner:

$$p_{all} = \mu \cdot s_{gr} + \left(1-\pi\right) \cdot s_{ur} \quad (2\text{-}9)$$

where μ is the non-negative tradeoff parameter to control the relative importance of both components.

2.3.5.5 State-of-the-art Method for Sequential Recommendation (BERT4Rec)

BERT4Rec, which stands for BERT for Recommendation, is an adaptation of the original BERT (Bidirectional Encoder Representations from Transformers) model for sequential recommendations (Sun et al., 2019). BERT4Rec leverage the power of deep bidirectional transformers for recommendation systems, particularly in scenarios where sequential data plays a crucial role. The model aims to address the limitations of previous sequential recommendation methods, which often employ left-to-right unidirectional architectures. These unidirectional models are argued to be sub-optimal due to their restricted ability to capture the full context of user behavior sequences and their assumption of a rigidly ordered sequence, which may not always hold in real-world scenarios. The task objective is to randomly mask items

in the input sequence and predict masked items based on their surrounding context.

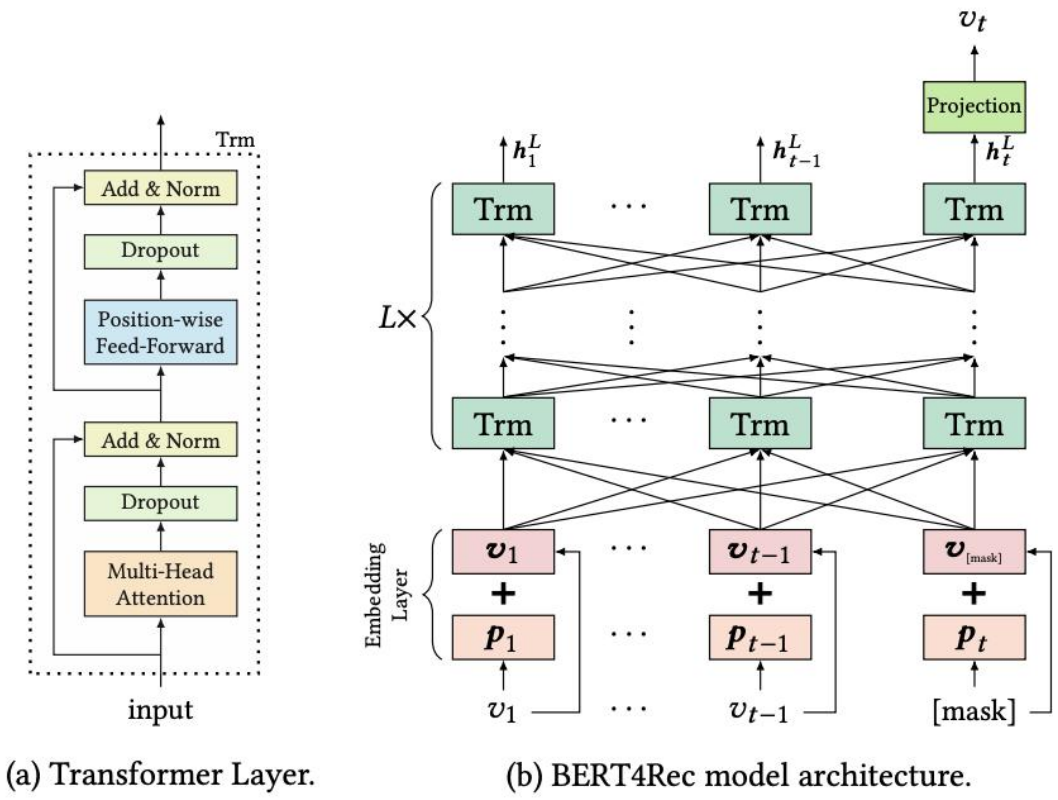

Figure 2-13 Overview of BERT4Rec

For an item in the sequence, its input representation is contrasted by:

$$h = v_n + t_n \tag{2-10}$$

where v_n is the input item embeddings, and t_n is the positional embeddings in the sequence organized by chronological order.

BERT4Rec is built using multiple bidirectional Transformer layers. Each layer iteratively revises the representation of every position by exchanging information across all positions at the previous layer. The Transformer layer has two sub-layers: Multi-Head Self-Attention, and Position-wise Feed-Forward Network.

The multi-head self-attention is formulated as:

$$MH(H^l) = [head1, head2, , , , headn]W^O \tag{2-11}$$

For each head, the attention output is calculated by:

$$headi = Attntion(H^l W_i^Q, H^l W_i^K, H^l W_i^V) \tag{2-12}$$

where the projection matrices for each head W_i^Q, W_i^K, W_i^V are learnable parameters.

The layer subscript l is omitted for simplicity. The attention function is Scaled Dot-Product Attentions as follows:

$$Attntion(Q,K,V) = softmax(\frac{QK^T}{\sqrt{\frac{d}{h}}})V \tag{2-23}$$

where query Q , key K , and value V are projected from the same matrix H^l with different projection matrices. $\sqrt{\frac{d}{h}}$ is the temperature.

The Position-wise Feed-Forward Network to the output of the self-attention sub-layer, which consists of two affine transformations with a GELU activation in between:

$$PFFN(H^l) = \left[FFN\left(h_1^l\right)^T; \dots; FFN\left(h_t^l\right)^T\right]^T \tag{2-24}$$

where each FFN is given by:

$$FFN(x) = GELU(xW^1 + b^1)W^2 + b^2 \tag{2-25}$$

where $W^1, W^2,\ b^1, b^2$ are learnable parameters and shared across all positions.

For the Stacking Transformer Layer, a residual connection around each of the two sub-layers, followed by layer normalization is designed. Also, the dropout to the output of each sublayer before normalization is applied. We obtain the final output H^L for all items in the input sequence following L layers that hierarchically exchange information across all positions in the preceding layer.

Then the final output distribution over target items is given by:

$$P(v) = softmax(GELU(h_t^l W^P + b^P)E^T + b^O \tag{2-26}$$

where W^P is the learnable projection matrix and b^P, b^O are bias terms. E is the embedding matrix for the item set.

2.4 Personalized Fashion Complementary Recommendation Task

2.4.1 Statements of the Problem

In the context of personalized fashion recommendation systems, complementary item recommendation refers to the task of identifying garments that are both aesthetically compatible with a given item and aligned with a user's personal preferences. Through preference modeling, we can predict users' preference scores for the matching garments, and sort them according to the scores, then recommend the top-k ranked items.

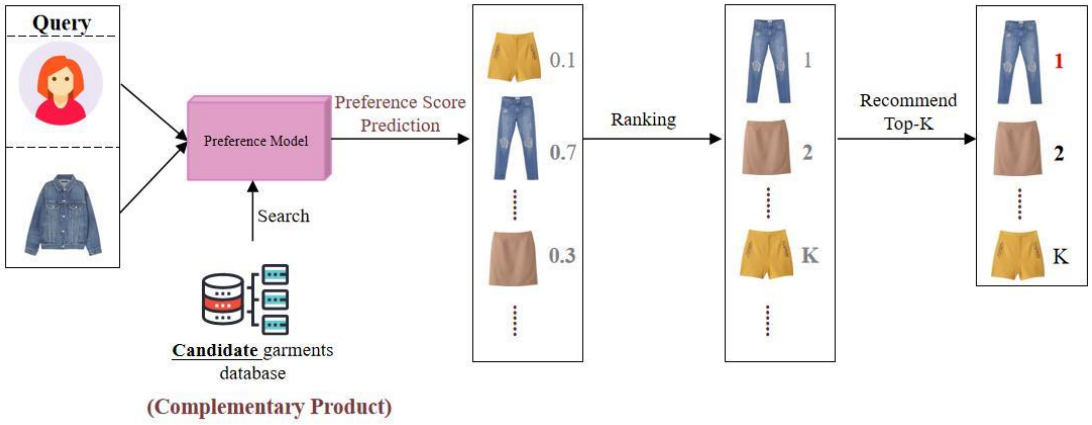

Figure 2-14 Personalized fashion complementary recommendation general model.

Suppose we have a set of users $U = \{u_1, u_2, u_3, \cdots, u_{|U|}\}$, a set of given product $G = \{g_1, g_2, g_3, \cdots, g_{|G|}\}$ and a set of recommended matching product $R = \{r_1, r_2, r_3, \cdots, r_{|R|}\}$, where $|U|, |G|, |R|$ denote the total number of users, given products, and recommended matching product, respectively. Each user $u \in U$ is associated with a set of triplets $< u, g, r >$ that represent the interaction records of user u, where $g \in G$ and $r \in R$.

To better capture contextual information and help the model make more relevant and trendy recommendations, we use multi-modal data to describe each product. For example, the given product g is represented by its visual feature $v_g \in \dot{R}^{dv}$ and visual

feature $w_g \in \dot{R}^{d_w}$, where d_v and d_w are the dimension of the corresponding features. Similarly, v_r and w_r denote the visual and textual features for the recommended matching product, where $v_r \in \dot{R}^{d_v}$ and $w_r \in \dot{R}^{d_w}$. The object is to derive a personalized complementary recommendation scoring function F, which effectively captures multi-modal based preference measurement between any triplet of $< u, g, r >$:

$$p_r^{u,g} = F(u, g, r|\theta) \quad (2\text{-}27)$$

where θ denotes the parameters to be learned in F. $p_r^{u,g}$ denotes the likelihood that the recommended matching product r will be preferred by the user u while still matching well with the specified given product g. An effective model F should be able to score compatible triplets higher than incompatible triplets based on personalization.

2.4.2 Compatibility Modeling

Research into topics related to clothing matching has garnered a significant amount of interest, both in terms of practical applications in everyday life and in the business world. In the context of this discussion, the compatibility criteria evaluate the degree to which several articles of clothing are compatible with one another (that is, match or fit together). The majority of the studies (Vasileva et al., 2018; Yang et al., 2020; Dong et al., 2020; Kaicheng et al., 2021) attempted to capture matching patterns that were visually compatible as well as functionally complementary, and they obtained validated results from their efforts.

Recent research has focused on modeling the fine-grained matching rules for clothing products, either at the attribute-level (Feng et al., 2018; Li et al., 2021; Zhou et al., 2022) or the semantic-level (Hou et al., 2019; Wang et al., 2022). This has been done in an effort to improve performance. For instance, Li et al. (Li et al., 2021) specifically evaluated the compatibility of garments by using attribute-level representations extracted from visual features. This was done in order to determine whether or not the garments were interchangeable. In order to improve the system's

overall performance, Feng et al. (Feng et al., 2018) constructed an outfit composition graph and an attribute matching map by combining functionality from the attribute partition and adversarial partition modules. Jing et al. (Jing et al., 2021) constructed a tripartite graph more recently, in order to provide a more informative representation, which significantly increased compatibility prediction. This graph encoded the correlations between features, clothing items, and outfits. Despite these efforts, however, there was not a successful tapping into the individualized preferences of the users.

2.4.3 Personalization Modeling

Studies (Ding et al., 2021; Trakulwaranont et al., 2022; Veličković et al., 2017) have shown that personalization data can significantly improve prediction accuracy. These studies benefited from ongoing improvements in computational capabilities. Hou et al. (Hou et al., 2019) proposed a semantic extraction network and a fine-grained preference attention module. Together, these two components are able to provide a description of the reasons for recommending clothing in a manner that is specific to the individual by semantically emphasizing the attributes of the clothing.

Zhan et al. (Zhan et al., 2021) developed a two-level attention to capturing user preferences and presented a method to create associations among outfit-level and product-level attributes. They also presented a method to create associations among attributes at the product level. In a study that was published not too long ago (Yan et al., 2022), a graph neural network was used to improve product representation, and a transformer was used to learn user purchase histories in order to make complementary recommendations. In spite of the fact that these studies have produced convincing results, their scope was primarily limited to the recommendation of specific outfits or fashion item pieces. GPBPR (Song et al., 2019) utilized user-item interactions to model personal preference and item-item interactions for the purpose of modeling compatibility. This allowed the researchers to address both product compatibility and user preference. By simulating the appropriateness of garments on an attribute level,

PAI-BPR (Sagar et al., 2020) further improved the model's interpretability. These methods are effective for matching personalized clothing, but none of them took into account the consistent relationship that is present in the interaction data. The aspect of consistency needs to be thoroughly investigated in addition to the aspects of user preference and general compatibility that need to be looked into.

Data issues such as limited data, the presence of noise, data sparsity, or cold start can all hurt the performance of recommendation models. In the context of data-driven models, regularization typically refers to the calibration techniques to avoid overfitting. Techniques such as L1 (Molnar, 2020) and L2 (Ribeiro et al., 2016) regularization are examples of typical regularization methods. Many different regularization methods have been successfully implemented in a variety of applications, including matrix factorization (Koren et al., 2009; Mnih & Salakhutdinov, 2007; Rendle et al., 2012), and implicit feedback (Lu et al., 2019; Zhu et al., 2017). One additional method of regularization or regulation is to model the consistency that is already present in the data that describes interactions.

2.4.4 Personalized Compatibility Modeling

Personalized fashion complementary recommendation combines both compatibility modeling and personalization modeling, which relies on many kinds of technologies, such as collaborative filtering(Gao et al., 2023; Liang et al., 2016), content-based filtering (Wang et al., 2022), and deep learning models, are commonly employed to capture item-item compatibility and user-item preferences. Various studies have investigated the extraction of useful auxiliary information from global sources to enhance the effectiveness of recommendations. For instance, Dong *et al.*(Dong et al., 2020) proposed a method to infer users' essential body measurements from their historical clothing purchases, providing insights into their body shape for better outfit recommendations. Similarly, Li *et al.* (Li et al., 2021)collected item information into an outfit representation and enhanced a user's representation using their historical outfits to further personalize recommendations.

However, publicly available fashion outfit datasets often lack timestamp information, which poses a challenge when utilizing users' interaction history to assist in personalized fashion recommendations. Nevertheless, our proposed techniques overcome this challenge by collecting interactions of users who have positively interacted with products and generating a feature that reflects users' latent interests based on the content information of the selected products. The application of attention networks effectively handles these unordered interactions without timestamps. As a result, our NiPC-BPR not only addresses the model learning challenges caused by data sparsity but also leverages additional similar interest information and global insights to improve the accuracy of fashion recommendations. By incorporating these advanced machines learning methods and considering auxiliary information, our research endeavors to develop more effective and personalized fashion recommender systems, empowering users to make fashionable choices that align with their individual tastes and preferences.

2.5 Personalized Multi-behavior Sequential Recommendation Task

2.5.1 Problem Formulation

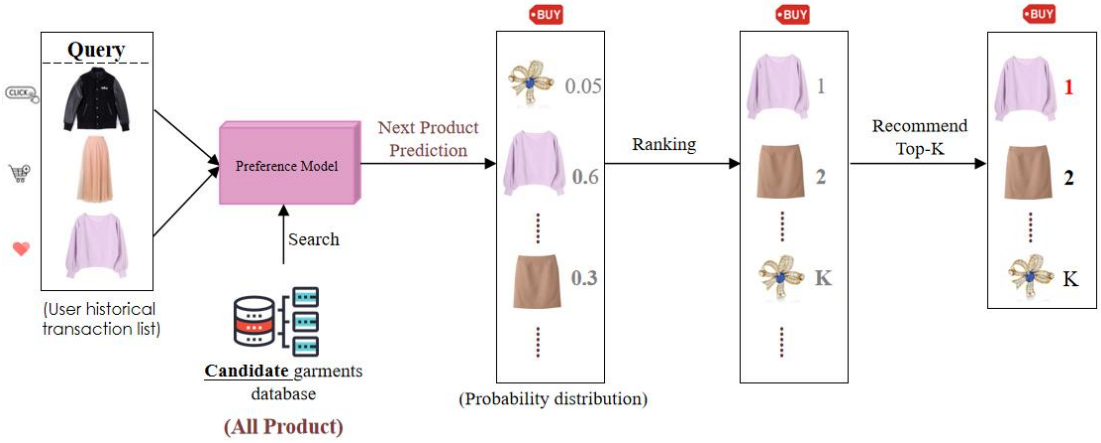

Figure 2-15 Personalized sequential recommendation general model.

Figure 2-15 is problem formulation of personalized sequential recommendation. With input of multi-behavior historical transaction or interaction list, the objective is to

predict the next item a customer is likely to purchase. The output from the preference modeling is a probability distribution over all candidate garments, based on this score distribution we rank and generate a recommendation list.

Let $U = \{u_1,\ u_2,\ u_3, \cdots, u_{|U|}\}$, $V = \{v_1,\ v_2,\ v_3, \cdots, v_{|V|}\}$ represent the set of users and psroduct items. The set of user interaction types is denoted as $A = \{a_1,\ a_2,\ a_3, \cdots, a_{|A|}\}$. Set A may include 'page view', 'add-to-favorite', 'add-to-cart', and 'purchase' actions. We denote the first three actions as *auxiliary behaviors* for assisting the prediction of *target behavior* of 'purchase' (set a_4 for illustration example). For each user u_i, we characterize the user interactions using a behavior sequence $S^{u_i} = \left[\left(v_1^{u_i}, a_1^{u_i}\right), \cdots, \left(v_n^{u_i}, a_n^{u_i}\right), \cdots, \left(v_N^{u_i}, a_N^{u_i}\right)\right]$, in which a behavior is represented by a 2-tuple $\left(v_n^{u_i}, a_n^{u_i}\right)$, i.e., a specific type of interaction $a_n^{u_i} \in A$ and its associated item $v_n^{u_i} \in V$. The behaviors in the sequence are sorted in time order and the sequence has a length of N time steps.

In multi-behavior sequential recommendation, given the sequential interaction histories S^{u_i} as input, the goal is to predict the probability P that user u_i will purchase the item $v_{N+1}^{u_i}$ at the next time step $\mathrm{N}+1$:

$$P\left(v_{N+1}^{u_i}, a_4 \middle| S^{u_i}\right) \tag{2-28}$$

2.5.2 Multi-behavior Sequential Recommendation

The primary goal of multi-behavior recommendation is to provide personalized recommendations based on their interactions across multiple types of behaviors or activities. Various approaches have been explored, e.g. by analyzing behavior connections through an attention mechanism (Xia et al., 2020; Xuan et al., 2023) or by employing graph-enhanced convolutional networks to understand correlations between different behaviors (Peng et al., 2023; Xia et al., 2021; Y. Yang et al., 2022). Recent advancements in multi-behavior recommender systems include new frameworks like NMTR (Gao et al., 2019), which is a recommendation framework that handles several tasks and incorporates specified connections between behaviors in an iterative manner.

Other examples include DIPN (Guo et al., 2019), which employs a hierarchical

attention network to recognize connections between various behaviors, and MATN (Xia et al., 2020), which leverages transformers to encode interactions among diverse behaviors. Furthermore, GNN-based techniques like MGNN (Zhang et al., 2020), MBGCN (Jin et al., 2020), and GHCF (Chen et al., 2021) utilize message passing on graphs to represent complex multi-behavioral dynamic data.

Moreover, some work has highlighted the importance of including additional information about users or items to generate enhanced correlations (Ding et al., 2022; Liao et al., 2023) for more accurate recommendations. However, these methodologies often struggle to effectively capture the dynamic user preferences that involve several behaviors within complex data environments. To address this issue, we propose a SG-MST model for recommendation by utilizing behavior-aware sequential patterns at various levels, from local to global, and incorporating data augmentation and self-learning to provide robust and accurate e-commerce product recommendations.

Chapter 3. Consistency Regulating Modeling for Personalized Clothing Matching Recommendation

As mentioned in Chapter 1, to achieve the objectives of enhancing personalization, compatibility, representation learning, and addressing the data sparsity issues, in this chapter, we propose the first method: CR-BPR, for the fashion complementary recommendation task. We first give a brief introduction of the proposed CR-BPR, and Then the detailed technologies and experiments will be reported after the introduction.

3.1 Introduction

Data analysis and common sense remind us that *consistency* exists in both user preference and item compatibility (He et al., 2020; Kim, 2009; Sun et al., 2018). This method recognizes the importance of consistency in user preferences and clothing matching, a factor often overlooked in previous research. Real-world fashion customers tend to favor products with the same brand, style, or color (Chen et al., 2019). Users may prefer clothing similar to their previous choices, while clothing matching can benefit from recommending items that go well with past selections. For example, as shown in Figure 3-1, When given a variety of options, the user (u) may choose clothing that is similar to what they have previously selected. From the perspective of clothing matching, it would also make sense to select bottom clothing that matches the sweater's top garment (g), if g has previously matched. When both factors are taken into account at the same time, the black skirt (b_1) would be the best option from a consistency perspective. Integrating both aspects can lead to more informed recommendations, making consistency a valuable addition to the model, especially when user behavior consistency outweighs fitting in with others (Kim, 2009). On the one hand, it broadens

the data set beyond the user-product interactions themselves. On the other hand, it could work as a function of regulation for collaborative filtering (CF)-based procedures, which are predicated mostly on searching for ways that reveal similarities between user and product interactions.

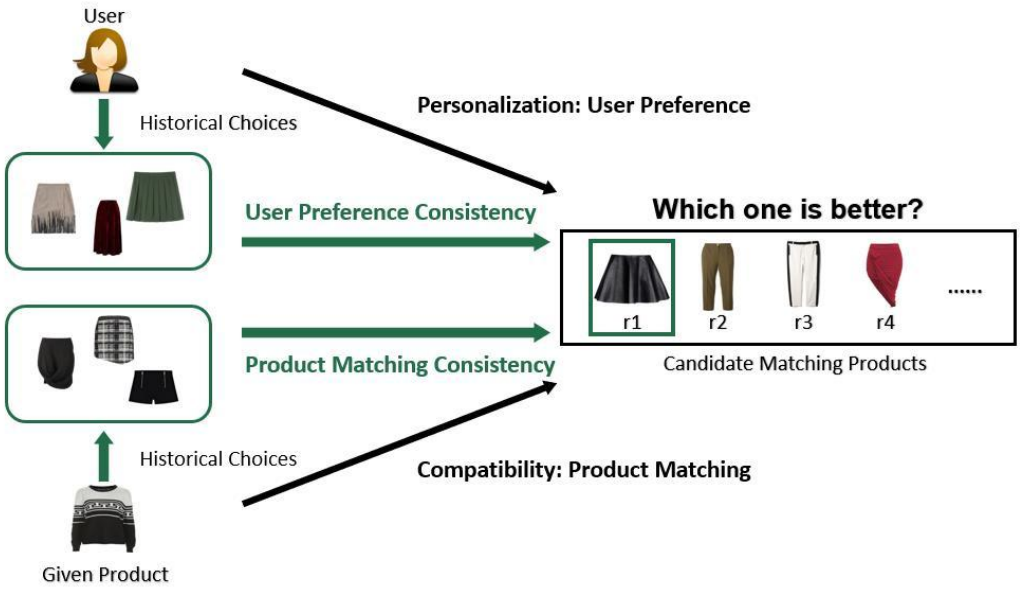

Figure 3-1 The key concepts.

In this chapter, we propose a novel approach **C**onsistency **R**egulating **B**ayesian **P**ersonalized **R**anking model with Feature Scaling (**CR-BPR**). It expands on the dual BPR framework, which simultaneously models compatibility and personalization. A hybrid collaborative filtering module with latent and multi-modal feature inputs is used by each BPR branch; this approach has been shown to be successful in earlier research (Song et al., 2019).

We present a feature scaling procedure for handling relative feature importance in order to handle variations in feature scales. Before predicting scores for compatibility and personalization, this normalization transformation is applied to various feature types, such as latent features and multi-modal content features. Furthermore, to make consistency modeling easier, two branches for consistency regulation are made to evaluate how similar the target product is to earlier choices from the perspectives of compatibility and personalization. A multi-branch BPR model that is jointly optimized

combines the two branches of consistency modeling with compatibility modeling and fundamental personalization.

3.2 Approach

3.2.1 Latent Representation with Feature Scaling (NMPP)

Multi-modal features derived from pre-trained deep neural networks, like CNN, are frequently used to represent clothing products in the personalized fashion complementary recommendation problem. This covers both textual and visual elements that highlight the primary attributes of the products. Nevertheless, neural network models that have already been trained for tasks like image classification are usually the basis of these features. These extracted product features go through a non-linear transformation to make them suitable for modeling user preferences or other purposes.

In this special task, the transformation of the product features involves the use of multiple multi-layer perceptions (MLPs) within different parts of the overall framework, as illustrated in Figure 3-2. Each MLP is responsible for processing product features of a specific modality, such as visual or textual. Utilizing the visual modality as an example, a sequence of linear projection and non-linear activation layers are applied to the pre-trained features $\mathrm{v_r}$ of the recommended matching product r in order to produce visual representations:

$$\begin{cases} v_r^1 = \sigma\left(W^1 v_r + b^1\right) \\ v_r^k = \sigma\left(W^k v_r^{k-1} + b^k\right), k = 2, \cdots, K \end{cases} \tag{3-1}$$

where W^k and b^k here are the learnable projection parameters and bias in the $k-th$layer, respectively, K indicates the total number of layers in the MLP network, and the K-th layer is the final output layer. The Sigmoid function is used to implement the activation function, σ. To keep things simple, the feature processing module mentioned in Eq. (3-1) is denoted by MLP_v, where the subscript v specifically refers to the visual features. The output layer's new visual representation as a result is:

$$v_r^* = v_r^K = MLP_v(v_r)\epsilon\dot{R}^{d_v^*} \tag{3-2}$$

where $R^{d_v^*}$ denotes the corresponding visual latent feature dimension. Similarly, the textual features are projected into another MLP module MLP_w to obtain the enhanced textual latent representation:

$$w_r^* = w_r^K = MLP_w(w_r)\epsilon\dot{R}^{d_w^*} \tag{3-3}$$

This procedure guarantees that the features of the product are transformed into an appropriate representation for the modeling tasks that follow. Through the use of modality-specific representations and MLPs, the framework is able to accommodate different kinds of product information and efficiently utilize the derived features. This method makes it possible to represent clothing items in a flexible and adaptive way, which supports various study objectives and makes it possible to model user preferences.

Additionally, batch feature scaling is applied to features before modeling user-product and product-product interactions in order to address the issue posed by different feature scales that may limit preference modeling. Using the illustration as an illustration, the visual representation tensor can be expressed as $V\ \epsilon\ \dot{R}^{d_v*l}$ for a single batch of matching items, where l represents the batch size. Then the feature scaling over dimension $m\ \epsilon\ [0, 1, \cdots, l]$: $V_m\ \epsilon\ \dot{R}^{1*l}$ is translated as follows:

$$\begin{cases} \bar{V}_m = \dfrac{V_m}{max\left(\|V_m\|_2, \varepsilon\right)} \\ \|V_m\|_2 = \sqrt{\sum_{n=1}^{l}(V_{nm})^2} \\ \varepsilon = 1e-12 \end{cases} \tag{3-4}$$

The batch representation transform into $\bar{V} = \left[\bar{V}_0^T, \bar{V}_1^T, \cdots, \bar{V}_m^T, \cdots, \bar{V}_{d_v}^T\right]^T\ \epsilon\ \dot{R}^{d_v*l}$ after the feature scaling process (Eq. (3-4)). Thereafter, the specific latent visual representation of the recommended matching product $\bar{v}_r$ can be obtained. For simplicity, we use $Norm(\cdot)$ to represent the whole feature scaling procedure, and the

overall latent representation after the MLP and feature scaling process is abbreviated as $NMPP(\cdot)$, such as:

$$\begin{cases} \bar{v}_r = Norm(MLP(v_r^*)) = NMPP(v_r^*) \\ \bar{w}_r = Norm(MLP(w_r^*)) = NMPP(w_r^*) \end{cases} \tag{3-5}$$

which denote the normalized visual and textual latent representation for the recommended product. However, for ID related embedding $(\mathrm{e_u}, \mathrm{e_g}, \mathrm{e_r})$, there is no MLPs procedure but only normalization, therefore the latent normalized embedding is abbreviated as:

$$\begin{cases} \bar{e}_u = Norm\left(\mathrm{e_u}\right) \\ \bar{e}_g = Norm\left(\mathrm{e_g}\right) \\ \bar{e}_r = Norm\left(\mathrm{e_r}\right) \end{cases} \tag{3-6}$$

where $e_u \in E^U \in \dot{R}^{|U|*d_e}$, and $e_g, e_r \in E^I \in \dot{R}^{|G+R|*d_e}$, and E denotes the initialization table.

3.2.2 CR-BPR Overall Scheme

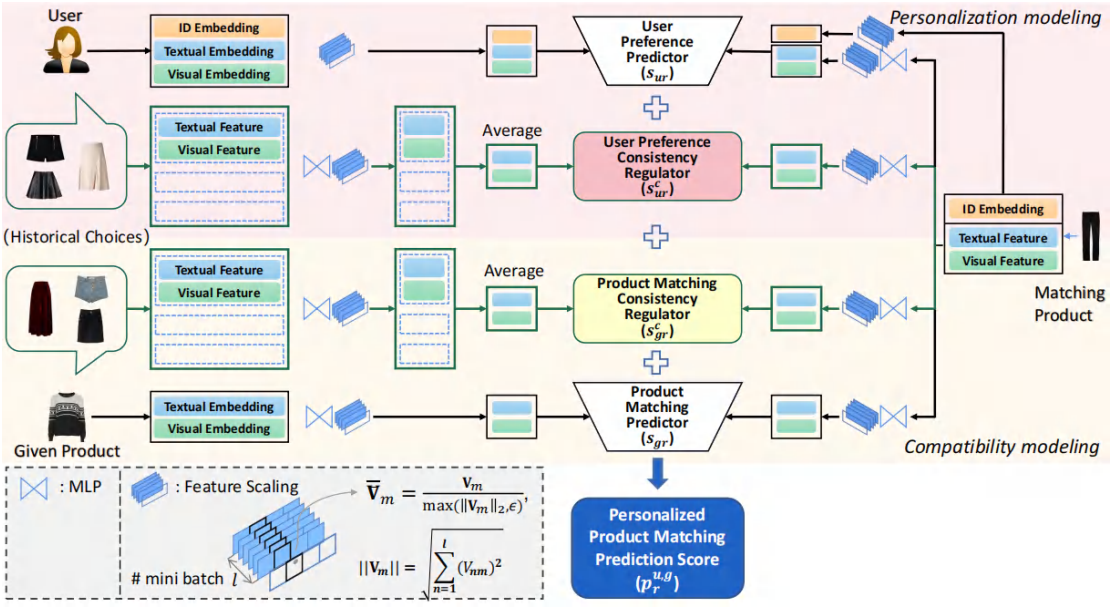

Figure 3-2 The Overview of CR-BPR Scheme.

As shown in Figure 3-2, the proposed **C**onsistency-**R**egulating dual **BPR** (**CR-BPR**) method addresses the personalized fashion complementary recommendation problem into four components, (1) user preference modeling, (2) product matching

modeling, (3) user preference consistency modeling, and (4) product matching consistency modeling. All four parts are based on the normalized representation module ($NMPP$). The overall input of the CR-BPR are the user; a given product; users' historical choices; the given product's previous choices; and the matching product, the output is the predicted preference score.

3.2.3 User Preference Modeling (Personalization)

The user preference predictor is based on **M**atrix **F**actorization (**MF**) (Koren et al., 2021), which has shown to be very effective in a variety of applications involving personalized recommendations. An estimate of the user-product interaction is encoded by the inner product, which is the basic idea behind splitting the user-product interaction matrix into user-latent and product-latent elements. To successfully integrate multi-modal product information, the user's preference for a particular product content is also examined. The user preference predictor that follows is built using the **V**isual **B**ayesian **P**ersonalized **R**anking (**V-BPR**) (He & McAuley, 2016) model:

$$s_{ur} = (\bar{e}_u)^T \bar{e}_r + \eta(\bar{v}_u)^T \bar{v}_r^p + (1-\eta)(\overline{w}_u)^T \overline{w}_r^p + \beta_u + \beta_r + \alpha \qquad (3\text{-}7)$$

In order to align the scale of various middle-stage representations, $\bar{e}_u$, $\bar{v}_u$ and $\overline{w}_u$are normalized latent embeddings for user u and the recommended item r. Their effectiveness in improving overall performance will be assessed. In the same way that $\bar{e}_u$, is normalized following embedding table initialization, $\bar{v}_u$ and $\overline{w}_u$ are two additional user representations linked to the user ID for modeling implicit user visual and textual preferences (Eq. (3-7)). The multi-modal representations of the suggested product r, $\bar{v}_r^p$ and $\overline{w}_r^p$, were derived from the NMPP module that was presented in the preceding section (3.2.1). The representations are unique to user preference modeling, as indicated by the subscript p.

Moreover, β_u and β_r are user and recommended product bias terms, which are standard in the score function of MF-BPR models. α is the global offset. All preference scores are measured with the inner product operation involving the corresponding user and recommended product representations. The trade-off parameter η is used to

balance the relative weight of visual and textual preferences while ensuring that both latent and content user preferences contribute Equally to the overall user preference score.

3.2.4 Product Matching Modeling (Compatibility)

The product matching modeling predicts the compatibility of a pair of products' matching connections. Potential matching cues from specific elements regarding fashion products, such as color, texture, style, and function (Liang et al., 2016), are investigated by modeling product-product interactions based on content information. The visual and textual compatibility of the two products is used to illustrate the present matching connection:

$$s_{gr} = \pi\left(\bar{v}_g^m\right)^T \bar{v}_r^m + \left(1-\pi\right)\left(\bar{w}_g^m\right)^T \bar{w}_r^m \quad (3\text{-}8)$$

In this Equation, $\bar{v}_g^m$ and $\bar{v}_r^m$, $\bar{w}_g^m$ and $\bar{w}_r^m$a re the NMPP-obtained normalized visual and textual latent representations of the specified and recommended products, respectively. The representations are limited to product matching modeling, as indicated by the subscript m. In each modality, vectors representing the matching relationships between two products are represented by the dot product, much like in user preference modeling. The visual and textual matching results are combined with a linear combination in order to efficiently capture the overall multi-modal matching score s_{gr}between the recommended item and the given product. In product matching modeling, π is used as a trade-off parameter to balance the significance of textual and visual modalities.

3.2.5 User Preference Consistency Modeling (UC)

Auxiliary data from the user preference consistency module helps the model make more accurate predictions. In order to assess user preference consistency, this article indicates calculating the average similarity between a target matching recommended product (r) and the user's previous selections (ur), specifically the recommended products the user has previously chosen or reviewed as recorded in user-

product interaction data. Greater user preference consistency and a higher likelihood that the target recommended product will be chosen as a match for the given product are indicated by a higher similarity score.

Several challenges present in ***U**ser Preference **C**onsistency (**UC**)* the fact that user preferences might change over time, as well as the different durations of available user-product interaction records. A timestamp indicating the sequence of each user-product interaction may not always be present in personalized fashion complementary records, though. In order to address these issues, we suggest combining each user's interactions with the target matching recommended product to create an unordered list of past selections, represented by the symbol ur, which stands for each user's overall preference. The products with original features that are most similar to the target product are filtered out in order to prioritize the consistency of the current preference. In this article, we'll call the filtered user-product previous selections the ***ur* list** throughout this article, while the term "original features" pertains to the pre-trained visual and textual features of each recommended product.

The following regulator shows how the *User Preference Consistency* is designed:

$$s_{ur}^{c} = \pi(\bar{v}_{ur})^{T}\bar{v}_{r}^{C_u} + \big(1-\pi\big)(\bar{w}_{ur})^{T}\bar{w}_{r}^{C_u} \tag{3-9}$$

Here $\bar{v}_{ur}$ and $\bar{w}_{ur}$ are abbreviations of the mean latent visual and textual representation of ur:

$$\begin{cases} \bar{v}_{ur} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\bar{v}_{ur}^{i} \\ \bar{w}_{ur} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\bar{w}_{ur}^{i} \end{cases} \tag{3-10}$$

N is the number of historical choices, $\bar{\mathrm{v}}_{\mathrm{ur}}^{\mathrm{i}}$ and $\bar{\mathrm{w}}_{\mathrm{ur}}^{\mathrm{i}}$ are the latent visual and textual representation of the i -th product in the ur list obtained from the $NMPP$ module, respectively. Similarly, $\bar{v}_{r}^{C_u}$ and $\bar{w}_{r}^{C_u}$ in the Eq.(3-9) are normalized visual representations and textual representations for a target recommended product also obtained from $NMPP$, where the subscript C_u means the representations are exclusive

to the modeling of *User Preference Consistency*. To reduce the effect of variable length of interaction and rapid changes in user preferences, the length of the ur lists for each user u is kept the same and relatively short.

s_{ur}^{c} is a content-based product similarity, which is derived as the dot product of the inherent characteristics product pairs of targets recommended product r and those in the ur list of historical choices. The superscript c is utilized to distinguish it from the user preference score s_{ur} in the user preference modeling (Eq. (3-7). For the purpose of the similarity calculation, the same trade-off parameters π and $(1-\pi)$ as in the previous branches are employed to reflect the relevance of visual and textual modality.

With user preference consistency, the historical choices of user u for target recommended product r will vary according to each triplet's similarity score s_{ur}^{c}. c(Hu et al., 2008).

3.2.6 Product Matching Consistency Modeling (GC)

Symmetric with *User Preference Consistency (UC)* modeling, another regulator named as *Product Matching Consistency (GC)* regulator is employed to capture product consistency from the perspective of product compatibility, a regulator is used. Products that are most similar to the current target recommended product rare then filtered based on pre-trained features. Similarly, all matching products that interact with the given product g are grouped. The historical selections that have been filtered are simply called the ***gr* list**, which characterizes the matching preference for the given product g. Here we define the *Product Matching Consistency* regulator as:

$$s_{gr}^{c} = \pi\left(\bar{v}_{gr}\right)^{T} \bar{v}_{r}^{C_g} + \left(1-\pi\right)\left(\bar{w}_{gr}\right)^{T} \bar{w}_{r}^{C_g} \tag{3-11}$$

Similar to Eq.(3-10), $\bar{v}_{gr}$ and $\bar{w}_{gr}$ are abbreviations of the mean latent visual and textual representation of gr:

$$\begin{cases} \bar{v}_{gr} = \dfrac{1}{N}\sum_{i=1}^{N} \bar{v}_{gr}^{i} \\ \bar{w}_{gr} = \dfrac{1}{N}\sum_{i=1}^{N} \bar{w}_{gr}^{i} \end{cases} \tag{3-12}$$

Also, N is the number of historical choices, $\bar{\mathrm{v}}_{\mathrm{gr}}^{\mathrm{i}}$ and $\bar{\mathrm{w}}_{\mathrm{gr}}^{\mathrm{i}}$ are the latent visual and textual representation of the i-th product in the grlist obtained from the $NMPP$ module, respectively. Furthermore, $\bar{v}_r^{C_g}$ and $\bar{w}_r^{C_g}$ in the Eq. (2-9) are normalized visual representations and textual representations which also obtained from $NMPP$ for the given product g. The subscript C_g means the representations are exclusive to the modeling of *Product Matching Consistency.* The dot product is applied to calculate the content-based similarity in terms of each visual and textual modality. The relative importance of the two modalities is controlled by the same trade-off parameters π and $(1-\pi)$. Thereafter, the average similarity (s_{gr}^{c}) between each target recommended product r and the products in the gr list is calculated, subscript c differentiates it from the product matching score s_{gr} in Eq.(3-8).

This product matching or compatibility relationship is derived from product-pair interactions by the CR-BPR scheme, which also acts as a consistency regulator to enhance prediction results, particularly in cases where the candidate recommended product r has a short history of product-pair interactions. When recommending a suitable complementary product, a higher similarity value denotes stronger regulation of product matching consistency.

3.2.7 Overall Preference Prediction

The proposed CR-BPR overall preference predictor linearly combines the four components, including s_{gr} – the user personal preference (i.e., personalization) score, s_{ur} – the product matching (i.e., compatibility) score, s_{gr}^{c} and s_{ur}^{c} – the two consistency regulating scores. The overall personalized complementary product preference score ($p_r^{u,g}$) indicates the preference score of users u towards the

recommended product r to match with a given product g, which is formulated as follows:

$$p_r^{u,g} = \mu \cdot s_{gr} + (1-\mu) \cdot s_{ur} + \emptyset \cdot s_{gr}^c + \varphi \cdot s_{ur}^c \quad (3\text{-}13)$$

The trade-off parameter μ is applied to balance the two main components of personalization and compatibility components, while $\emptyset$ and φ control the contributions of the two consistency regulating components.

3.2.8 Optimization

The CR-BPR model is optimized with the BPR loss, which is a pairwise loss to push the negative samples away from positive samples, thereby ranking the positive candidates higher, and which has proven to be powerful in implicit preference recommendation modeling (Cao et al., 2017; Loni et al., 2016). Let the preference scores for a pair of positive and negative samples, be obtained using Eq. (2-9) are denoted as $p_{r+}^{u,g}$ and $p_{r-}^{u,g}$, respectively. The BPR loss on the whole training set D is calculated by:

$$\mathcal{L} = \sum_D \left[-ln\left(\sigma\left(p_{r+}^{u,g} - p_{r-}^{u,g}\right)\right)\right] + \frac{\lambda}{2}\left\|\Theta_F\right\|^2 \quad (3\text{-}14)$$

where $D = \{(u, g, r+, r-) | u \in U \wedge g \in G \wedge (r+, r-) \in R\}$. λ is the non-negative hyperparameter and Θ_F represents the set of parameters of the model.

3.3 Experiment Preparation

This section introduces the experimental settings we prepared for verifying personalized fashion complementary recommendations, and experiment preparation. All the experiments are conducted on IQON3000 dataset and Polyvore dataset.

Baselines: Several representative recommendation models are selected as the baselines:

- **BPR-MF** (Rendle et al., 2012), which uses the Bayesian Personalized Ranking (BPR) algorithm when combined with Matrix Factorization (MF) to capture

the latent user-product relations.

- **V-BPR** (He & McAuley, 2016), which uses the factorization method to predict user preference by extracting visual features from product images.
- **T-BPR** (Song et al., 2019), which uses the same algorithm as V-BPR but substitutes textural features for visual features in order to incorporate textual information into BPR modeling.
- **VT-BPR** (Song et al., 2019), which further thoroughly characterizes user preferences based on both visual and textual factors, combining V-BPR and T-BPR.
- **GP-BPR** (Song et al., 2019), which is a personalized compatibility modeling approach based on multi-modal features. Using a joint BPR framework form, it efficiently models the relationship between user preference and product matching.
- **PAI-BPR** (Sagar et al., 2020), which is the state-of-the-art personalized product matching model. Compared to GP-BPR, it additionally leverages fine-grained attributes into the personalization and compatibility modeling.
- **PCE-NET** (Nie et al., 2023) leverages attention-based compatibility embedding modeling and personal preference modeling to capture fine-grained fashion compatibility features for enhanced personalized recommendation.
- **CP-TransMatch** (Ding, Mok, et al., 2023) addresses the personalized fashion matching task by leveraging a single-component translation operation to capture third-order user-item interactions and enhancing it with two graph learning modules focused on context and path perspectives.

Implementation Details: The three suggested models were trained using the same optimizer: Adam (Kingma & Ba, 2014) with the patient parameter set to 8 epochs and the maximum training epoch set to 80. An early stopping strategy was also used. From the training set, every triplet of data is a positive sample. Another bottom item will be

chosen at random from the matching item set as a negative sample for model training. The goal of the experiment was to first recommend a target matching item for the bottom item, and then recommend a target top product that would go well with a given bottom. To find the optimal settings for each training parameter, the grid search approach was used. More precisely, [64, 128, 256, 512] is the range in which the mini-batch size is searched.

To find the optimal settings for each training parameter, the grid search approach was used. More precisely, the following ranges are searched: [64, 128, 256, 512] for the mini-batch size; [0.001, 0.0001, 0.00001, 0.000001, 0.0000001] for the weight decay; [256, 512] for the hidden dimension; and [0.01, 0.001, 0.0001] for the learning rate. Based on empirical observations, the multi-purpose projection's layer (K) was set to 1, and the model's hyper-parameters were adjusted differently for the two datasets. Following (Song et al., 2019), In all experiments, the trade-off parameters π were set to 0.5 because it is assumed that the visual and textual modalities are equally important. The training set's user and product interaction set served as the foundation for consistency modeling.

3.4 Overall Recommendation Performance

The results of a comparison between the suggested CR-BPR and the current techniques for personalized clothing matching show how effective the former is, and the results are listed as:

Table 3-1 Performance comparison on IQON3000 and Polyvore datasets in terms of setting TOP garment as given product and BOTTOM garment as matching product. Larger numbers indicate better performance (↑). Bolded numbers indicate the best results, italicized and underlined indicate the second-best results.

Method	IQON3000				Polyvore			
	AUC	HR@10	NDCG@10	MRR	AUC	HR@10	NDCG@10	MRR
BPR-MF	0.8309	0.7024	0.5243	0.4687	0.7639	0.6916	0.5434	0.4973
TBPR	0.8316	0.6361	0.4301	0.3666	0.7839	0.6502	0.4868	0.4359

VBPR	0.8360	0.6941	0.5230	0.4694	0.8074	0.7141	*0.5848*	0.5443
VT-BPR	0.8384	0.7003	0.5134	0.4551	0.8105	0.6846	0.5266	0.4772
GP-BPR	0.8569	0.7396	0.5849	0.5366	0.8232	0.7121	0.5716	0.5277
PCE-NET	0.8341	0.6399	0.4110	0.3399	0.8235	0.4208	0.2380	0.1825
CP-TransMatch	*0.8842*	*0.8789*	*0.6453*	*0.5430*	*0.9001*	*0.8573*	0.5472	0.5144
CR-BPR	**0.9737**	**0.9505**	**0.8746**	**0.8518**	**0.9708**	**0.9582**	**0.8717**	**0.8361**

Table 3-2 Performance comparison on IQON3000 and Polyvore datasets in terms of setting BOTTOM garment as given product and TOP garment as matching product. Larger numbers indicate better performance (↑). Bolded numbers indicate the best results, italicized and underlined indicate the second-best results.

Method	IQON3000				Polyvore			
	AUC	HR@10	NDCG@10	MRR	AUC	HR@10	NDCG@10	MRR
BPR-MF	0.6849	0.5598	0.3776	0.3209	0.5724	0.2598	0.1331	0.0950
TBPR	0.7510	0.5315	0.3493	0.2929	0.6250	0.2938	0.1638	0.1243
VBPR	0.7626	0.5609	0.3724	0.3141	0.6644	0.2576	0.1316	0.0937
VT-BPR	0.7890	0.5341	0.3576	0.3032	0.6785	0.2523	0.1263	0.0885
GP-BPR	0.7913	0.6137	*0.4663*	*0.4196*	0.7075	0.2500	0.1253	0.0877
PCE-NET	0.7585	0.5111	0.3239	0.2669	0.7072	0.2222	0.1120	0.0789
CP-TransMatch	*0.8523*	*0.6251*	0.3756	0.2989	*0.7281*	*0.2859*	*0.1993*	*0.1419*
CR-BPR	**0.9031**	**0.8357**	**0.7863**	**0.7709**	**0.7637**	**0.3689**	**0.2026**	**0.1522**

We have the following observations:

(1) Our proposed CR-BPPR achieves better performances on both datasets and on both given top and given bottom conditions. CR-BPR is a comprehensive model that combines four essential branches: two main branches dedicated to personalization and compatibility, and two additional consistency regulating branches focused on user preference and product-matching consistency. By effectively integrating these branches, CR-BPR achieves an overall performance boost, providing enhanced recommendations in personalized fashion complementary scenarios.

(2) Because they only model user-product interaction patterns and ignore product-product compatibility patterns, MF-BPR, T-BPR, V-BPR, and VT-BPR perform the worst out of all the methods. In contrast, GP-BPR, which combines both product matching and user preference modeling, performs better than MF-BPR, T-BPR, V-

BPR, and VT-BPR. This shows that both relation modeling plays a significant role in the task of recommending personalized complementary clothing.

(3) Furthermore, the advantages of modeling user preference and user-product interactions in an integrated framework are confirmed by the fact that the GP-BPR and PCE-NET approaches were most similar to the current CR-BPR method.

(4) CP-TransMatch performs second best in the majority of settings when compared to other baselines and models multi-relational connectivity between fashion products that are subject to users using a multi-edge graph. This demonstrates how user preference modeling is greatly enhanced by thoroughly examining the user connections through user-product-product third order transactions. Though more computationally demanding, the complex path-based information translation can quickly result in overfitting and decreased efficacy, particularly in cases where the dataset is imbalanced.

3.5 Ablation Study

For proving the validation of our proposed NMPP module and designed three methods, we run the ablation study for each proposed scheme.

3.5.1 Effects of Feature Scaling

We discuss the effects of feature scaling from the parameter update perspective and experiment analysis perspective.

3.5.1.1 Feature scaling-based parameter update

In our baseline GP-BPR(Song et al., 2019), both user preference modeling and product matching modeling, it was assumed that the target recommended complementary product r would be represented identically, as shown by $v_r^p = v_r^m$ and $w_r^p = w_r^m$. To improve the matching product's representation in multiple latent spaces and capture different relations, these representations are deliberately different in the current formulation of CR-BPR.

Specifically, the representations v_g in the product matching branch and v_u in the user preference branch are adjusted accordingly to the following manner to achieve improvements:

$$\begin{cases} v_g \leftarrow v_g + \rho \cdot \left(\sigma\left(-p_{r+r-}^{u,g}\right) \cdot \left(v_{r+}^{m} - v_{r-}^{m}\right) - \lambda \cdot v_g\right) \\ v_u \leftarrow v_u + \rho \cdot \left(\sigma\left(-p_{r+r-}^{u,g}\right) \cdot \left(v_{r+}^{p} - v_{r-}^{p}\right) - \lambda \cdot v_u\right) \end{cases} \tag{3-15}$$

Model learning may be enhanced by this feature scaling-based parameter update. This will be compared to the baseline model GP-BPR (Song et al., 2019), in which v_g and v_u are updated without any feature scaling.

3.5.1.2 Feature Scaling Experiment Analysis

Table 3-3 Performance Comparison without and with (+) Feature Scaling (FS) in terms of AUC.

Method	Given Top (g)		Given Bottom (g)	
	Polyvore-519	IQON3000	Polyvore-519	IQON3000
BPR-MF	0.7639	0.8309	0.5724	0.6849
T-BPR	0.7839	0.8316	0.6250	0.7510
V-BPR	0.8074	0.8356	0.6644	0.7626
VT-BPR	0.8105	0.8384	0.6785	0.7890
GP-BPR	0.8232	0.8569	0.7075	0.7913
BPR-MF + FS	0.7375	0.8155	0.5975	0.6652
T-BPR + FS	0.7273	0.9463	0.6290	0.8653
V-BPR + FS	0.7443	0.8656	0.6523	0.7312
VT-BPR + FS	0.7549	0.9347	0.6454	0.8528
GP-BPR + FS	0.8863	0.9528	0.7341	0.8899

Based on the results shown in Table 3-3, it is clear that for the majority of models in the IQON3000 dataset, feature scaling resulted in better predictions. Interestingly, nevertheless, for a large number of models based on the Polyvore-519 dataset, it had the opposite effect. These results indicate that performance improvement cannot always be ensured by feature scaling. A key factor when evaluating the effect of feature scaling is the variation in the magnitude, range, and units of features among various datasets. Although feature scaling has proven beneficial in many machine learning applications, its efficacy varies depending on the task. For example, models like MF-BPR that only

handle one kind of data might not gain much from feature scaling. On the other hand, models that incorporate several kinds of data may provide significant advantages.

To further investigate the detailed impacts, GP-BPR's AUC with and without feature scaling was evaluated in relation to various trade-off parameter μ of Eq. (3-13). As GP-BPR only concludes compatibility score s_{gr} and personalization score s_{ur}. The horizontal axis μ of Figure 3-3 indicates the product-matching modeling weight, whereas $1-\mu$, in Eq. (3-13), symbolizes the importance of modeling user preferences. Experiments with various trade-off parameter settings demonstrated that, in the original GP-BPR, user preference predominated, but that, following feature scaling, the product-matching branch contributed more to the overall model performance. Both branches received equal contributions in the IQON3000 case, indicating better performance. However, performance significantly declined when the weight of product-matching equaled, indicating that general product-matching modeling is insufficient for matching personalized clothing. According to these results, compatibility and personalization modeling branches both significantly improve performance, while general product-matching modeling performs poorly in personalized clothing matching modeling.

.

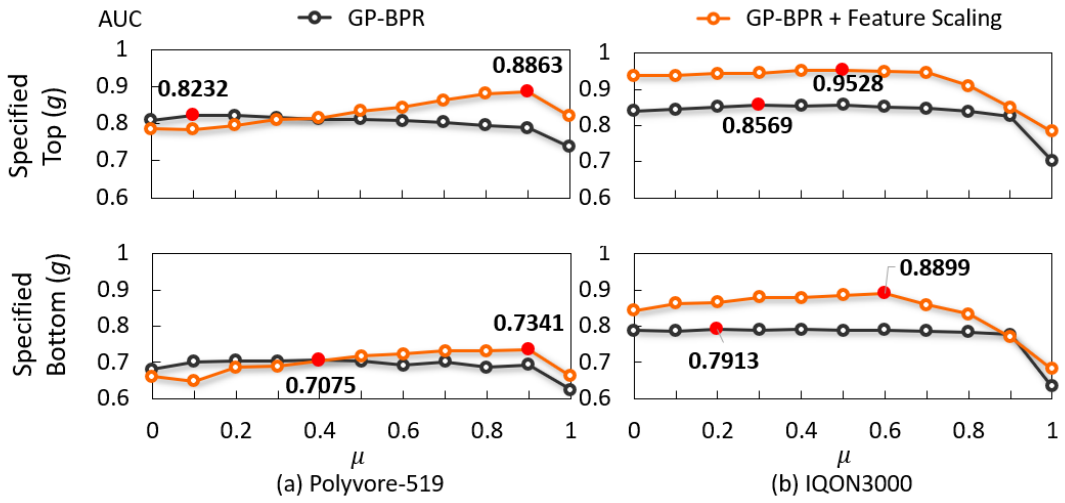

Figure 3-3 Comparing GP-BPR baseline with GP-BPR feature scaling using the (a) Polyvore-519 and (b) IQON3000 datasets.

3.5.2 Ablation of CR-BPR

This subsection reports the ablation study conducted on CR-BPR, including the validation experiment of effectiveness of consistency regulating and consistency regulator settings.

3.5.2.1 Effectiveness of consistency regulating

In Table 3-4, CR-BPR with User-preference Consistency regulator ('CR-BPR w/ UC') or CR-BPR with only product matching or product-compatibility Consistency regulator ('CR-BPR w/ GC') are the two consistency regulators that are used to compare the model's performance. The performance of the CR-BPR with GC and CR-BPR with UC models was assessed with the various weights of $\emptyset$ and ϕ (Eq. (3-13)) for the two datasets, and the results are shown in Figure 3-4. To better understand the contribution of each consistency regulating branch. Regardless of changing weights, the two consistency branches function in the same way. A small weight might not be enough to consistently regulate performance, but a large weight will cause similar historical choices to dominate the model while ignoring user preference and product matching modeling.

Table 3-4 Ablation Comparison for CR-BPR in terms of AUC.

Method	Given Top (g)		Given Bottom (g)	
	Polyvore-519	IQON3000	Polyvore-519	IQON3000
CR-BPR w/ UC	0.9609	0.9727	0.7554	0.8993
CR-BPR w/ GC	0.9609	0.9738	0.7606	0.9018
CR-BPR	0.9708	0.9737	0.7637	0.9031

The optimal performance of CR-BPR was evaluated by combining its two branches with varying weights ranging from one to ten for each branch. Figure 3-5 presents a summary of the overall model's performance with different branch weights. Notably, CR-BPR demonstrated better performance with IQON3000 compared to Polyvore-519, and it exhibited less sensitivity to changes in weights, resulting in

consistently even performance across both queries involving specified top and bottom settings. CR-BPR, equipped with single consistency regulating branches, effectively improved prediction performance, while CR-BPR with two consistency regulating branches showed slightly superior performance. Overall, the CR-BPR model exhibited robustness in the presence of rich data information, such as IQON3000, and successfully demonstrated the benefits of consistency regulation in enhancing both user-preference and product-matching consistency, ultimately leading to improved preference prediction accuracy.

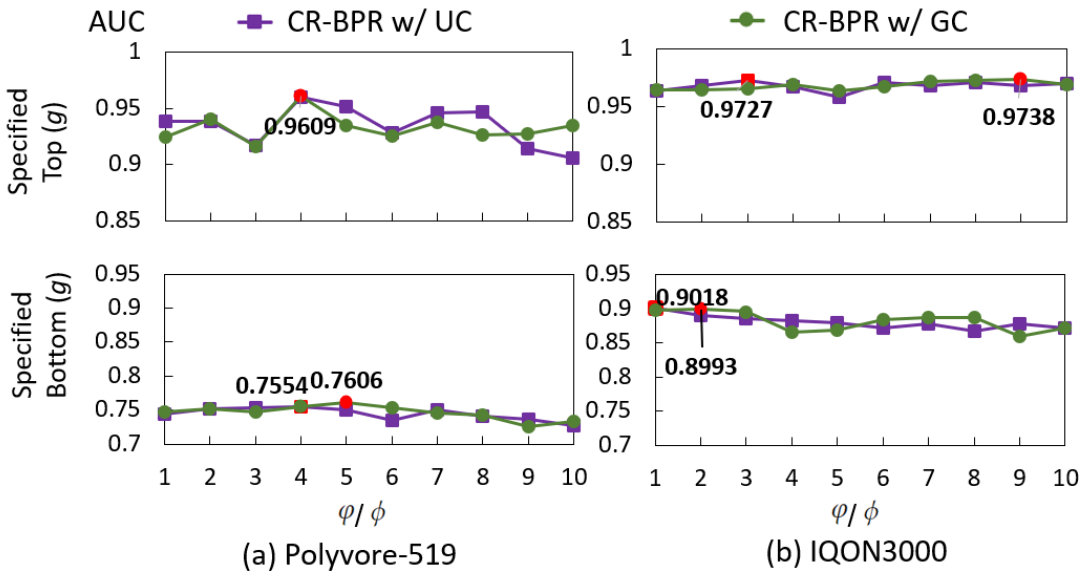

Figure 3-4 Performance of CR-BPR with GC and with UC branches with respect to the weight parameters φ and ∅, respectively, for the (a) Polyvore-519 and (b) IQON3000 datasets.

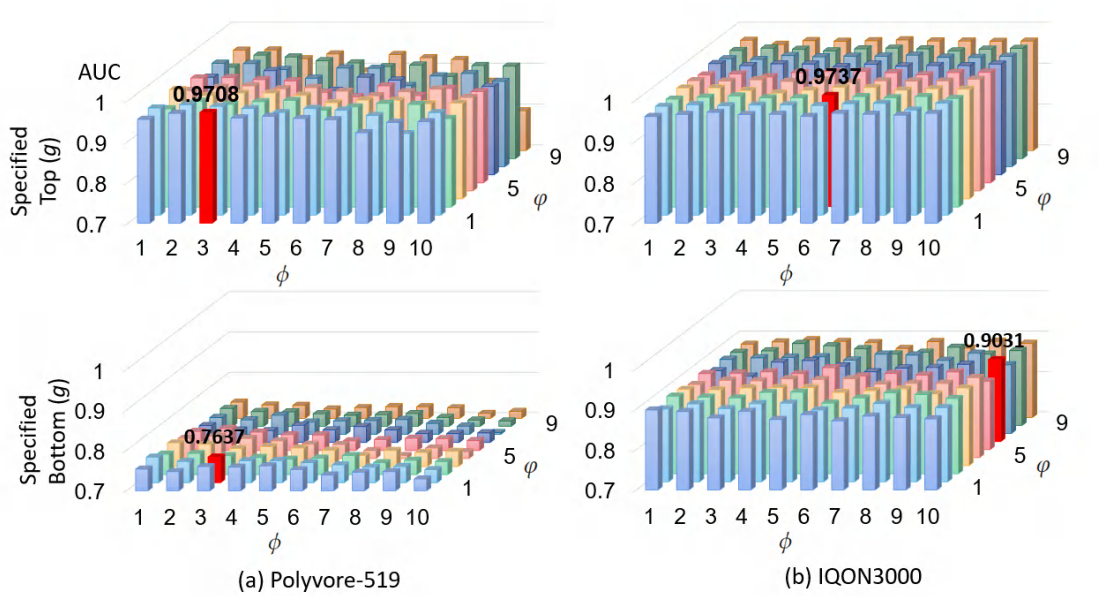

Figure 3-5 Performance of CR-BPR with different weights for the two consistency regulating branches: (a) Polyvore-519 and (b) IQON3000 datasets.

3.5.2.2 Consistency Regulator Settings

In this study, the number of historical choices (N) plays a crucial role in regulating user preference and product-matching modeling, as shown in Eq. (3-10) and Eq. (3-12). To investigate its impact, we conducted experiments using a wide range of historical choice numbers, ranging from 1 to 20, as shown in Figure 3-6. In cases where historical choice data was insufficient ($< N$), we filled in the remaining spaces of the user and product lists with the most similar product to the target matching garment.

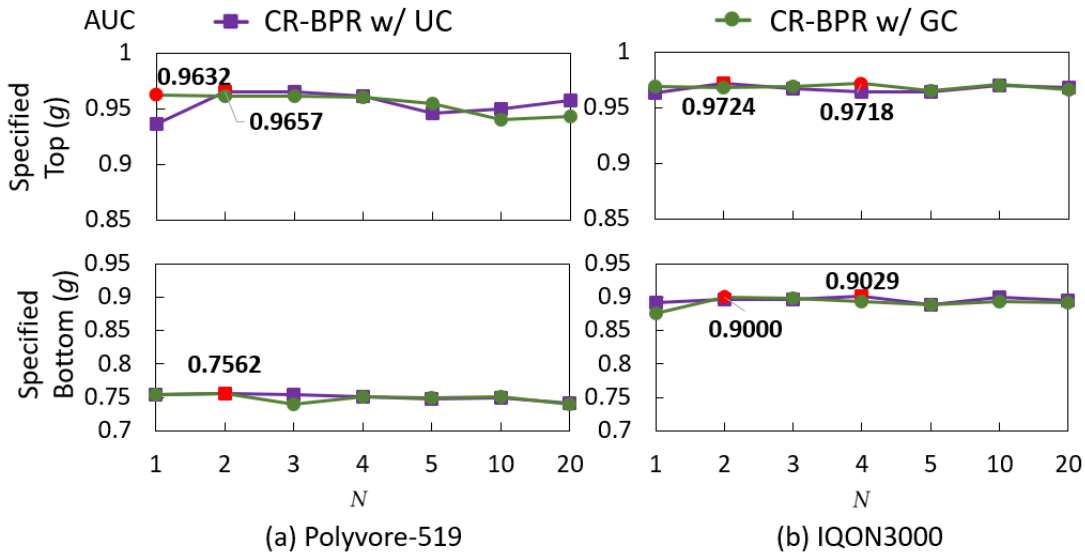

Figure 3-6 Performance of CR-BPR with GC and UC, respectively, using different number of historical choices N for the (a) Polyvore-519 and (b) IQON3000 datasets.

For datasets with data sparsity issues, such as Polyvore-519, the consistency branches were more sensitive to changes in N compared to IQON3000. Interestingly, the best performance was consistently achieved when $N = 2$, which strikes a balance between learning the current preference and considering the general preference. Too small a value of N may overly restrict the model, while a value may also introduce irrelevant noises. Therefore, appropriately using previous decisions, guided by preference and matching consistency, may significantly improves personalized clothing matching modeling.

3.5.3 Qualitative Evaluations

We performed qualitative evaluations by retrieving matching products for comparison and visualization of a few test examples in order to confirm the usefulness of our CR-BPR in real-world applications.

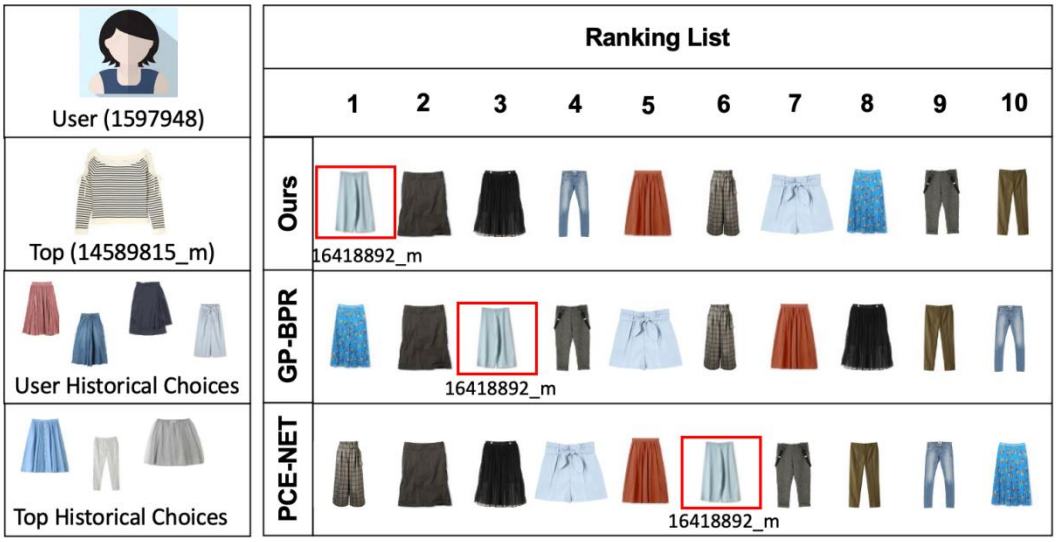

Figure 3-7 Illustration of the clothing matching recommendation results provided by three methods. The ground-truth are highlighted by red frame.

Figure 3-7 shows the results of the clothing matching ranking list produced by our CR-BPR, GP-BPR, and PCE-NET methods. Nine bottom garments were chosen at random and combined with the ground-truth matching product as candidates; the ground-truth was indicated in red frames. A user and the top pair of clothes in the testing dataset were chosen at random from IQON3000 as the query. The ranking results, which combine user and top historical choices, demonstrate that our method performs better than the other two methods by recommending products that are more similar to the target product in addition to ranking ground-truth products higher.

	u	g	r^+	r^-	Comparison
1	 user ID: 109153	 item ID: 12701822 price: 8,316 category: Knit color: Brown itemName: Over Long Turtleneck Knit Pullover	 item ID: 13579978 price: 81,800 category: Long Pants color: Black itemName: SKINNY RELAXED PANTS	 item ID: 2878469 price: 6,300 category: Skirt color: Blue itemName: Tulle Flare Skirt Delivery Time: within 7 days	GP-BPR × TransMatch √ CR-BPR w/o UC √ CR-BPR w/o GC × CR-BPR √
2	 user ID: 113489	 item ID: 14928928 price: 6,912 category: Knit color: Pink itemName: Hole Garment Relaxed Knit	 item ID: 15129866 price: 15,109 category :Long Pants color: Blue itemName: Standard Denim Pants	 item ID: 14918149 price: 22,680 category: Skirt color: Black itemName: ASTRAET Gazer Makishirt Skirt	GP-BPR × TransMatch √ CR-BPR w/o UC √ CR-BPR w/o GC √ CR-BPR √
3	 user ID: 109153	 item ID: 12421140 price: 6,469 category: Skirt color: Gray itemName: Botanical Net Check Trops	 item ID: 15188049 price: 27,000 category: Skirt color:Gray itemName: ZUCCa / S Lace Skirt / Skirt ZU71JG18425M	 item ID: 13756694 category: Long Skirt color: Blue itemName: Denim Long Skirt itemUrl:https://item.iqon.jp/13756694/	GP-BPR √ TransMatch × CR-BPR w/o UC √ CR-BPR w/o GC √ CR-BPR √
4	 user ID: 109153	 item ID: 9522707 price: 5,940 category: Tank Top color: Beige itemName: Published in Sale Magazine WAFFLE TANK TOP	 item ID: 6575965 price: 5,184 category: Skirt color: Blue itemName: jouetie Denim Skirt with Bandana	 item ID: 32905037 price: 14,456 category: Skirt color: Blue itemName: Piper Denim Skirt Material: Denim	GP-BPR × TransMatch √ CR-BPR w/o UC × CR-BPR w/o GC √ CR-BPR √
5	 user ID: 113489	 item ID: 12701468 price: 9,072, category: Tank Top color: Blue itemName: Maeda Riko Wearing Material: Twill × KamiSor	 item ID: 10416578 price: 7,776 category: Long Pants color: Blue itemName: Straight Denim Material: Denim	 item ID: 14947770 category: Long Pants color: Blue itemName: Lolais Skiing Denim Pants Material: Denim	GP-BPR × TransMatch √ CR-BPR w/o UC × CR-BPR w/o GC √ CR-BPR √

Figure 3-8 Comparison of the clothing matching recommendation results provided by different methods.

Figure 3-8 presents additional comparison results regarding recommendation accuracy. In particular, we chose a number of user-product-product transaction triplets (represented by the notation $\langle u, g, r^+ \rangle$ at random from the testing dataset. We selected a negative matching product r^- in each triplet by choosing a product with which the user had never interacted. The goal is to accurately match each user's provided top garment with the appropriate bottom garment. Only ID information is used, and user identities are anonymized to preserve privacy. Visual images and textual descriptions are used to represent each top and bottom piece of clothing. Figure 3-8 illustrates how well our CR-BPR works in every situation.

Furthermore, we discovered that matching predictions are simpler when there is a visual or textual connection between the provided tops and bottoms. This indicates

that the matching process is helped by the existence of unique and identifiable characteristics in both textual and visual representations. However, traditional compatibility models like GP-BPR and CR-BPR without user connections (i.e., CR-BPR w/o UC) find it difficult to differentiate between two matching products when their semantic features are very similar. This challenge results from the fact that these conventional models mainly rely on the intrinsic qualities of the products themselves, failing to take into account the extra context that user interactions can offer.

CP-TransMatch and our CR-BPR, which further incorporate user implicit connections, help the model grow more accurately. The outcomes further show how well our suggested approach handles the task of recommending personalized complementary clothing.

3.5.4 Application Examples

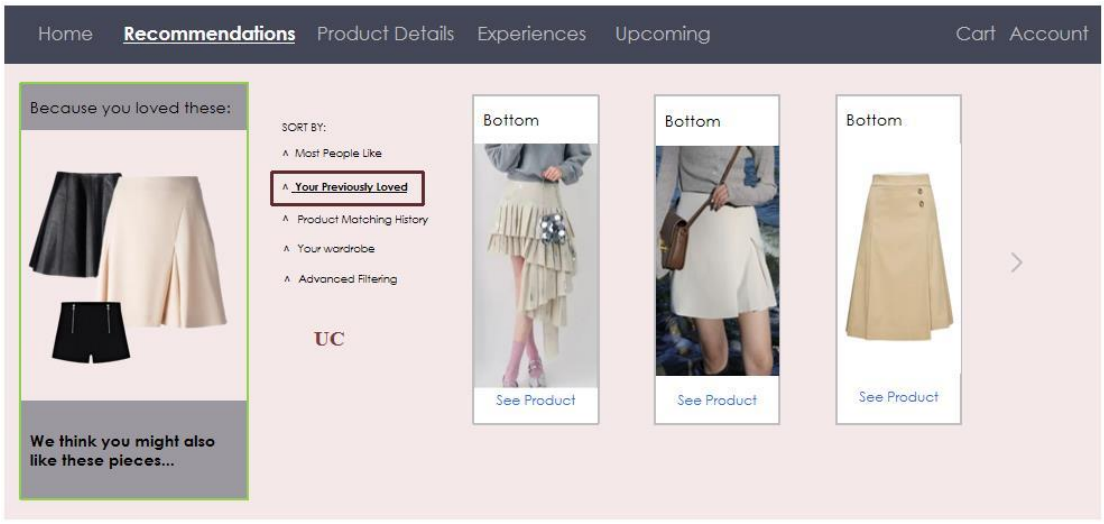

Figure 3-9 UC application example.

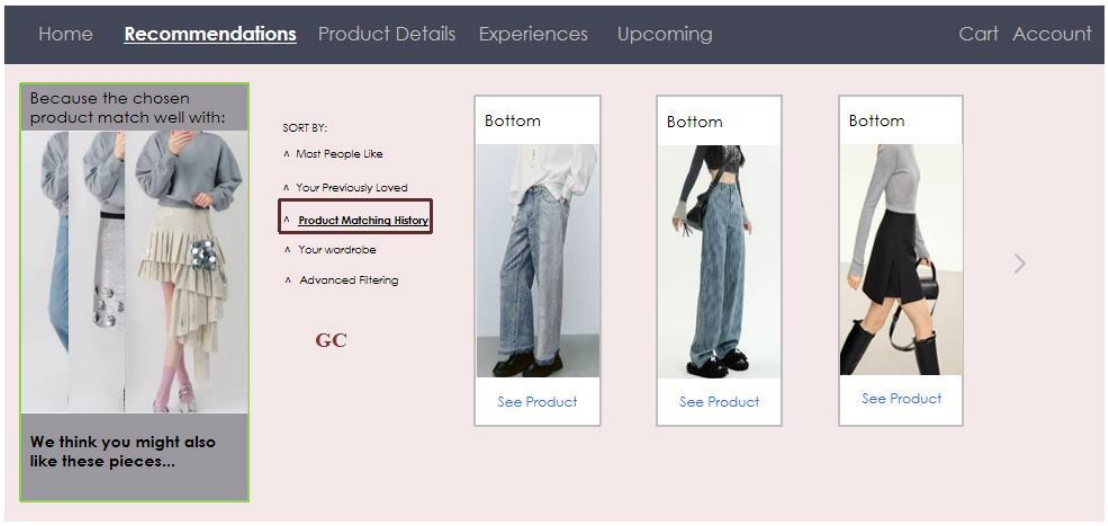

Figure 3-10 GC application example.

This subsection illustrates the potential application of our proposed two innovative tranches for personalized fashion complementary recommendations. In the example of the user preference consistency recommendation strategy, to match with a top a customer is browsing, the system may generate more recommendations that are similar to the customer's previously chosen bottoms, as similarity somehow represents consistency.

Symmetrically, Figure 3-10 shows that from another product-matching consistency perspective. For the current given top, the system would analyze the bottoms that go well with the top and provide recommendations similar to those previous matching items to keep consistency.

3.6 Summary

This chapter presents the CR-BPR model, which addresses the requirement for consistent product matching and user preference accordance while ensuring a balanced contribution from multi-source input data. Two collaborative filtering branches, each with its own consistency regularization, are incorporated into this method to model user preferences and product matching. Furthermore, multi-modal data is preprocessed using a feature scaling procedure before being fed into the model for learning.

The effectiveness of our suggested method has been confirmed through comprehensive experimental evaluations using two benchmark datasets, emphasizing the benefits of combining product-matching and user preference consistency when indicating complementary clothing items. Qualitative evaluations and comparative performance analyses show that our approach performs noticeably better than the alternatives.

Chapter 4. Modeling Indirect Personal Compatibility (NiPC-BPR) Scheme

In Chapter 3, we enhance representation learning by applying NMPP, and also improve the personalization and compatibility from the user preference and product compatibility consistency perspective. However, the balance between personalization and compatibility is overlooked, to address this issue, in this chapter, we further improve the personalized complementary recommendation by introducing a novel Indirect Personal Compatibility module to model personalization and compatibility in a coherent and coupled manner. In this Chapter, we give the brief introduction of the motivation followed by the detailed methodology and experiments.

4.1 Introduction

In this chapter, a ***N**ormalized **i**ndirect **P**ersonal **C**ompatibility* modeling scheme based on ***B**ayesian **P**ersonalized **R**anking* (**NiPC-BPR**) is proposed, which exploits direct and indirect personalization and compatibility relations from the user and product interactions, and effectively integrates various multi-modal data. As shown in Figure 4-1, an ***i**ndirect **P**ersonal **C**ompatibility* (**'iPC'**) module is added beyond personalization and compatibility, which was introduced in Chapter 3, to further exploit additional relations by modeling the compatibility between the complementary products to be recommended and the user's historical query products. Since this scheme uses the same user preference modeling, product matching modeling, and optimization method as CR-BPR does, the rest of the section only introduces the specific indirect Personal Compatibility module, and overall preference prediction.

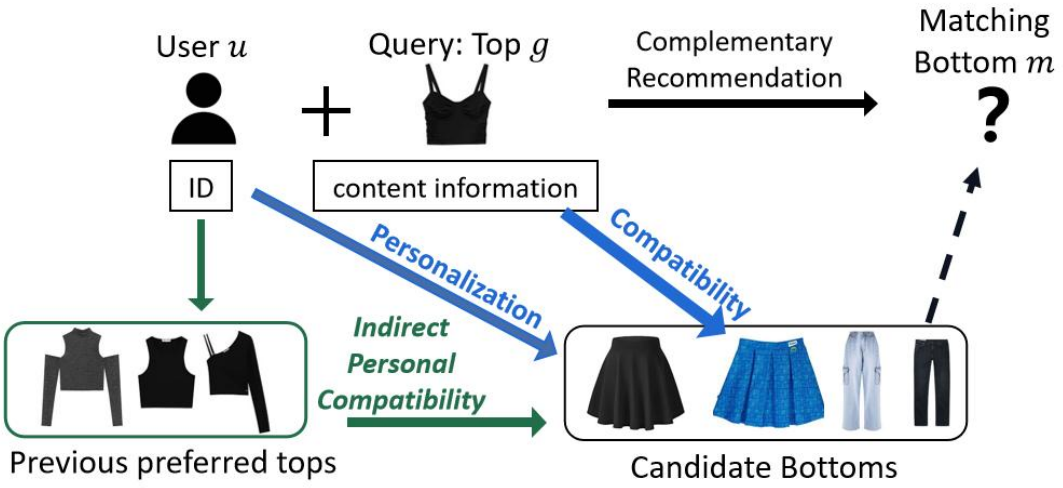

Figure 4-1 Example of Indirect Personal Compatibility.

4.1.1 Indirect Personal Compatibility Module

The additional indirect personal compatibility relation is derived from the target complementary product r and given products chosen by each user (ug), to comprehensively model the user preference and product matching preference:

$$s_{pc} = \pi\left(\bar{v}_{ug}\right)^T \bar{v}_r^{pc} + \left(1-\pi\right)\left(\bar{w}_{ug}\right)^T \bar{w}_r^{pc} \tag{4-1}$$

where $\bar{v}_r^{pc}$ and $\bar{w}_r^{pc}$ are the latent normalized visual and textual representation of the recommended complementary product r obtained from $NMPP$, and to distinguish from personalization and compatibility modeling, the superscript pc means the latent representations are exclusive to indirect personal compatibility modeling. $\bar{v}_{ug}$ and $\bar{w}_{ug}$ are to represent the latent mean representation of user's historical query products, and the **iPC** module aggregate the historical query products by:

$$\begin{cases} \bar{v}_{ug} = \dfrac{1}{N}\sum_{i=1}^{N} \bar{v}_{ug}^{i} \\ \bar{w}_{ug} = \dfrac{1}{N}\sum_{i=1}^{N} \bar{w}_{ug}^{i} \end{cases} \tag{4-2}$$

N denotes the number of each user's historical interactions over given products. The parameters π and $1-\pi$ in Eq. (4-1) serves the purpose of adjusting the importance of different modalities, similar to how the CR-BPR does in Eq.(3-7) and Eq.(3-8). By

using the proposed distance metric among the features of related products, the inner product defines quantifiable matching rules. Consequently, s_{pc} represents the degree of compatibility between the current complementary product and the user's historically preferred given products.

4.1.2 Overall Preference Prediction

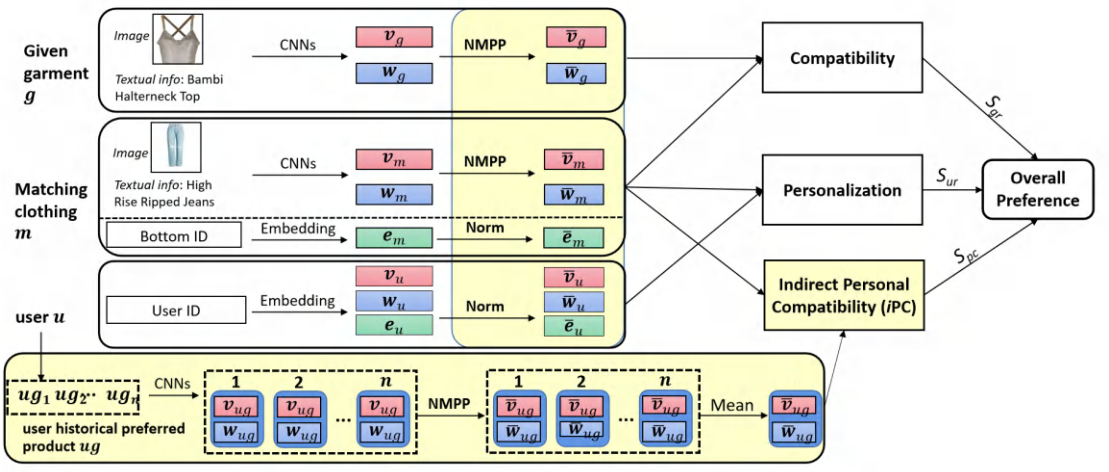

Figure 4-2 Overview of NiPC-BPR Scheme.

As shown in Figure 4-2, the proposed NiPC-BPR method represents F by three modules, including **compatibility module (s_{gr})(same as illustrated in Eq. (3-8))**, also, **personalization module (s_{ur})(same as illustrated in Eq. (3-8))**, and indirect personal compatibility module (s_{pc})(Eq.(4-1)). and represent inter-product compatibility between the given and matching products and user-product preference, s_{gr} and s_{ur} both are directly related to the target complementary product r; whereas s_{pc} indirectly models personal compatibility between a complementary product and similar products that those users previously queried (ug). Moreover, the method can be easily modified to generate personalized suggestions, effectively capturing both personalization and compatibility in a cohesive manner. Thus, the proposed NiPC-BPR method presents as:

$$p_r^{u,g} = \mu \cdot s_{gr} + (1-\mu) \cdot s_{ur} + \eta \cdot s_{pc} \tag{4-3}$$

μ, $1-\mu$ and η are weights corresponding to the three modules.

4.2 Overall Recommendation Performance

We conducted experiments on the two benchmark datasets: IQON3000, and Polyvore-519, and compared the performance of the different approaches in terms of prediction accuracy for either specifying a top item or a bottom item as the query.

Table 4-1 Performance Comparison on Polyvore-519 dataset.

Method	Given Top (g)	Given Bottom (g)
	AUC	AUC
BPR-MF	0.7639	0.5724
TBPR	0.7839	0.6250
VBPR	0.8074	0.6644
VT-BPR	0.8105	0.6785
GP-BPR	0.8232	0.7075
NiPC-BPR	0.9646	0.7631

Table 4-2 Performance Comparison on IQON3000 dataset

Method	Given Top (g)	Given Bottom (g)
	AUC	AUC
BPR-MF	0.8309	0.6849
T-BPR	0.8316	0.7510
V-BPR	0.8356	0.7626
VT-BPR	0.8384	0.7890
GP-BPR	0.8569	0.7913
MG-PFCM	0.7730	-
PAI-BPR	0.8502	-
A^3-FKG	0.9289	-
PS-OCM	0.9295	-
NiPC-BPR	0.9646	0.9044

From the two comparison Tables, the following observations can be found:

1) Among the models evaluated on both datasets, BPR-MF exhibited the poorest performance. This can be attributed to its reliance on limited information, such as user ID and item ID. On the other hand, VBPR and TBPR performed slightly better as they leveraged visual or textual information, which provided a slight advantage over the basic MF approach. Furthermore, VT-BPR combines both visual and

textual information and achieved a better performance than models that only used single modality. The findings revealed that both textual product descriptions and visual images play crucial roles in exploring user preferences.

2) GPBPR outperforms both the general compatibility and personal preference baselines, highlighting the importance of considering both product-product compatibility and user-product preference interactions in personalized fashion complementary recommendation. Meanwhile, A^3-FKG and PS-OCM achieve substantial improvements in prediction results, indicating that enhancing user or item representation through aggregating multiple relations and conducting information propagation can significantly benefit fashion recommendation.
3) The proposed NiPC-BPR method effectively estimates user preferences for matching products by incorporating both direct relations learned from user-product interactions and indirect relations inferred from users' historically preferred products within a unified framework. This approach outperforms other baseline methods, demonstrating the advantages of leveraging signals from users' past actions and the effectiveness of modeling personalization and compatibility coherently. The results highlight the benefits of considering both direct and indirect relations to enhance the recommendation performance in personalized fashion complementary recommendation.

4.3 Ablation of NiPC-BPR

This subsection reports the ablation study conducted on NiPC-BPR, including the validation experiment of effectiveness of iPC module, hyper-parameter settings, and effects on product interaction frequency.

4.3.1 *Effectiveness of iPC module*

Table 4-3 Ablation Comparison for NiPC-BPR in terms of AUC.

Method	Given Top (g)		Given Bottom (g)	
	Polyvore-519	IQON3000	Polyvore-519	IQON3000
NiPC-BPR w/o iPC	0.8863	0.9528	0.7341	0.8899
NiPC-BPR	0.9646	0.9746	0.7631	0.9044

Table 4-3 compares the proposed NiPC-BPR method without and with indirect Personal Compatibility (iPC) module on the two datasets in respect of AUC. NiPC-BPR outperforms which without iPC module, showing the benefit from users' historical interactions with query products to infer personalized compatibility and integrate disparate factors from various information cost-effectively.

4.3.2 *Hyper-parameter Settings Study*

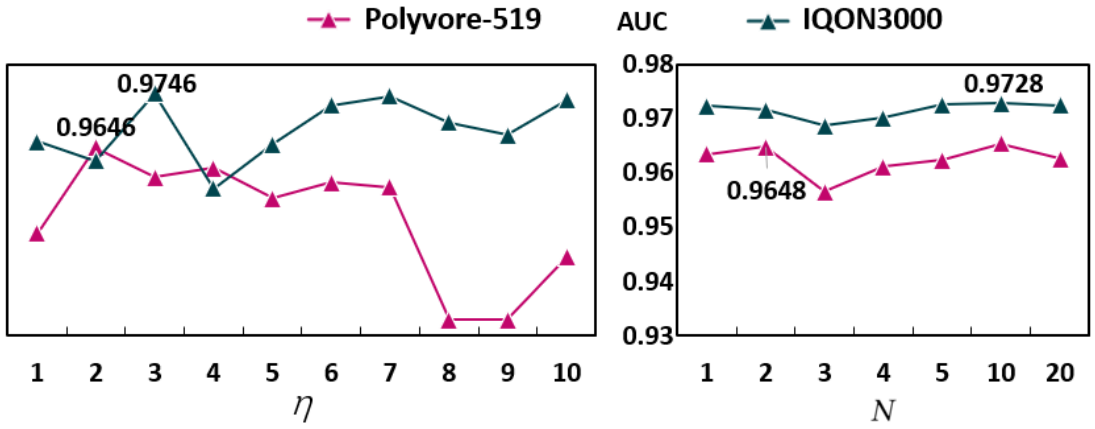

Figure 4-3 Performance of NiPC-BPR wrt parameter η and wrt different numbers (N) of user historical preferred given items.

In Figure 4-3, we present the performance analysis of our NiPC-BPR model concerning the weight parameter (η) for both datasets, with "top" as the query. Smaller values of η (e.g., $\eta = 2, 3$) demonstrate slightly improved complementary recommendation performance, while larger η values (greater than 4) lead to performance drops. Optimal weights for our iPC module enable balanced training without overly relying on current samples or historical data, resulting in better overall performance.

Furthermore, we investigated the impact of the number of historical preferred items (N) in the Eq.(4-2) on the overall performance. Figure 4-3 also displays two example curves illustrating the influence of changing N while maintaining constant weights for direct relations of s_{gr} and s_{ur} (as specified in Eq. 错误!未找到引用源。). Empirically, we observed that our NiPC-BPR model is not sensitive to changes in the historical choice of N for both datasets. Moreover, the AUC results of IQON3000 significantly outperformed those of Polyvore-519.

4.3.3 Effects on Product Interaction Frequency f.

Table 4-4 Performance comparison on two datasets in terms of AUC, under different product interaction frequencies f .

f	Given Top (g)				Given Bottom (g)			
	Polyvore-519		IQON3000		Polyvore-519		IQON3000	
	GP-BPR	Ours	GP-BPR	Ours	GP-BPR	Ours	GP-BPR	Ours
0 <= f <= 2	0.7476	0.8800	0.7585	0.4322	0.6323	0.5646	0.7466	0.6447
3 <= f <= 5	0.9059	0.9942	0.8294	0.8277	0.7741	0.8913	0.8396	0.9369
6 <= f <= 10	0.9537	1	0.8614	0.9773	0.8349	0.9615	0.8887	0.9811
11 <= f	0.9562	1	0.9295	0.9995	0.8770	0.9948	0.9351	0.9846

To assess the effectiveness of our indirect Personal Compatibility (iPC) modeling in user preference prediction, we divided the dataset into four subsets based on the frequency of product interactions (f). We then evaluated the performance of our approach in each subset. Table 4-4 shows that our approach outperforms the baseline GP-BPR (Song et al., 2019), particularly when the frequency of product interactions is higher. This suggests that considering historical patterns can significantly benefit personal compatibility modeling, especially when a substantial amount of high-quality historical data is available.

4.3.4 Application Example

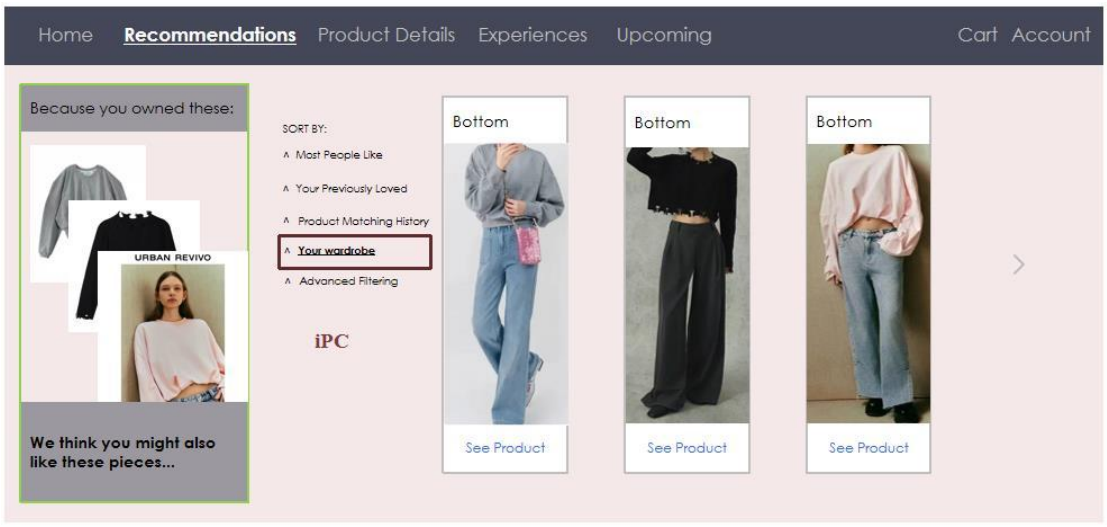

Figure 4-4 iPC application example.

Figure 4-4 shows the application example of the proposed second method for fashion complementary recommendations. For a given specific shirt, we combine the fashion garments a user has owned in their wardrobes, then to give the overall suggestions which can not only match with the current given top, but also match as many previously owned tops in their wardrobe, as possible.

4.4 Summary

We have crafted an innovative NiPC-BPR framework aimed at recommending coordinated outfits by capturing the compatibility of clothing items and individual user preferences through both independent and interconnected approaches. This framework capitalizes on users' past engagements with queried products to deduce personalized matching criteria and efficiently amalgamate diverse elements from assorted information sources in a cost-effective manner. Moving forward, our subsequent research will concentrate on optimizing the utilization of heterogeneous data within our NiPC-BPR framework to enhance its performance further.

Chapter 5. Attentive Preference Modeling with Contrastive Learning (APCL)

In this chapter, we further improve the representation learning by proposing two cross-functional views contrastive learning for personalized complementary recommendations mentioned in Chapter 1. In addition, compared to CR-BPR and NiPC-BPR introduced in Chapter 3 and Chapter 4, this chapter explores auxiliary connections between users and items from higher orders and deeper perspectives, to throughoutly utilize the interaction data. Detailed motivations and methodology will be introduced in this chapter. Similarly, extensive experiments on benchmark datasets are also conducted to evaluate the effectiveness of the proposed APCL.

5.1 Introduction

Consumers increasingly seek *personalized* experiences and interactions in various aspects of life, including fashion recommendations that align with users' unique preferences and styles. Fashion recommendation is also unique in the sense that clothing products are often presented in a coordinated outfit, including a top and a bottom garment, rather than as an individual clothing item itself. Therefore, the research of *personalized fashion complementary recommendations* (X. M. Song et al., 2023). has received a considerable amount of attention over the past decade because it not only addresses the needs of individual users for a convenient and efficient way of looking for fashion products that go well together, reducing the cognitive load of decision-making, but improves the conversion rates of e-commerce platforms, facilitating better user engagement. The goal of the task is to offer users curated lists of complementary fashion products that not only match the user's individual preferences personalization but also create a cohesive fashionable look, in terms of style, color, and shape, when combined and worn together compatibility (Abluton, 2022; Kim et al., 2024).

Existing methods can be grouped into the following categories: (a) Collaborative

Filtering based methods leverage user-product interactions, in form of ratings or implicit feedback indicating behaviourial similarity between users, where traditional mathematical modeling techniques, such as Factorization Machines, can be used to capture the relationship for user preference prediction. (b) Content-Based Filtering leverages product attributes, including visual and textual features, in order to create feature-rich representations so as to recommend similar or complementary products. Representation-based methods do not explicitly model user-production interaction relationships, rather they indirectly do so by learning informative representations and maximizing the similarity between the representations. (c) deep learning-based methods learn advanced representations that capture both complex relationships from interaction data as well as incorporate additional product content, such as attributes or textual descriptions. Bayesian Personalized Ranking (BPR) is one prominent approach that operate collaborative filtering in latent space where matrix factorization technique (Rendle et al., 2012) are used for preference prediction. Being a representation-based model, BPR has made it ideal to integrate with other representations for product textual (Song et al., 2019) an/or visual (He & McAuley, 2016)information. Examples include visual and textual BPR (V-BPR (He & McAuley, 2016), T-BPR (Song et al., 2019)) and their extensions, such as VT-BPR and GP-BPR (Song et al., 2019) in multi-modal context, where more complex deep learning based hybrid models are developed.

Regardless the types of method being used, existing solutions for fashion complementary product recommendations have not been very successful, and the challenges are threefold. First of all, as discussed earlier, fashion taste is something really personal that fashion has been regarded as a major means for individual expression across nations and centuries. Consequently, those content-based filtering methods focusing mainly on product compatibility modeling by taking into account of rich product information such as textual and visual information are indeed not a suitable approach for complimentary product recommendation, because of the ignorance of personalization modeling to uncover implicit user preferences buried in the user-

product interaction data. The second key challenge is the issue of data sparsity. It is very difficult, if not impossible, to uncover useful patterns or relationships from interaction data, when such data is sparse, specifically for personalization modeling with user-product interactions or for comparability modeling with product-product interactions. Data sparsity can significantly impact the learning process and the ability of the resulting model to learn informative representations. The problem of data sparsity is even more severe in fashion recommendations, because fashion products have a very short life cycle. In the age of fast fashion, fashion brands offer hundreds of new products per week, and we are now in the era of ultrafast fashion that retailers are launching 10,000 new designs per day. When fashion products are being launched in such a speed and volume, the transaction (interactions) data are very sparse. The third obstacle lie in the complex and multi-modal nature of fashion data. Fashion products are often presented to consumers in various formats over e-commerce websites, including product attributes in text descriptions, as well as in fashion images and videos, making it ideal to use representation-based methods for preference predictions, taking full advantage of all available data. Nevertheless, it is still a challenging research topic for the effective integration of multi-modal data in personalized complementary recommendations, especially in the context of fashion with sparse data.

To address the above problems in personalized fashion complementary recommendations in a cost-effective manner, we develop a novel **A**ttentive **P**reference with **C**ontrastive **L**earning model (**APCL**). The APCL scheme is a hybrid method that organically integrates both personalization and compatibility modeling within a unified framework.

Specifically, we propose novel attentive preference modules, simulating the self-attention mechanism of the transformer, to address the issue of limited interaction data by aggregating the interactions of *similar users* for an *implicit* personalization modeling. Here, *similar users* are defined as a group of other users who have chosen the same target product before, therefore by analyzing the user-product interactions of these

`similar' users an enhanced representation for the target user by capturing indirect user-product relationships. A similar concept is also used for *implicit* compatibility modeling when product-product interactions are limited, as depicted in Figure 5-1.

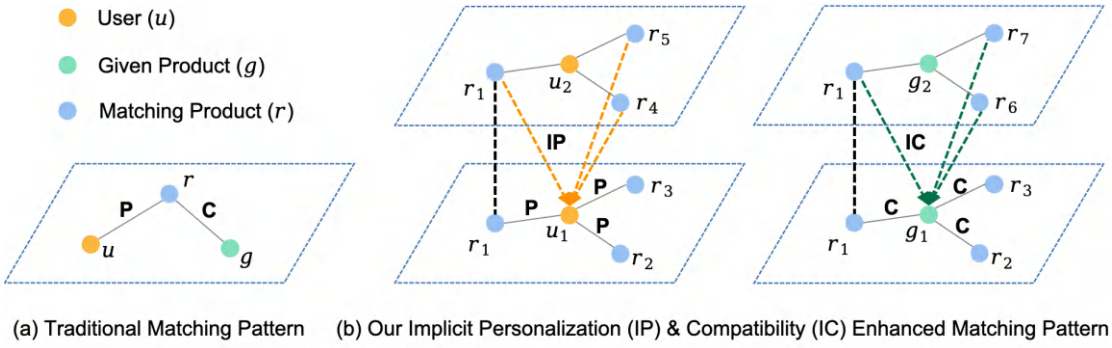

Figure 5-1 Personalized fashion recommendation modeling patterns comparison.

On the other hand, taking advantage of the rich multi-modal data available in fashion products, we propose a novel joint optimization strategy integrating BRP with contrastive learning, where the former uncovers complex relations from both direct and indirect interaction data while the latter learning informative representations from different modality of data, for optimized complimentary product recommendations that fit for individual users.

Different from existing contrastive learning-based recommendation studies that primarily focus on modality-specific alignments (e.g., aligning visual and textual features within a single product or between products), we propose novel contrastive learning losses in this section, aligning representations between functional views, namely between Personal Preference and Product Compatibility (P-C) and between two attentive preferences of implicit personalization and implicit compatibility (IP-IC), as shown in Figure 5-2, across textual and visual features.

The main contributions of this APCL are summarized as follows:

1. To address data sparsity issue, we develop a novel Attentive Preference module, using correlation sampling and attention mechanism, to uncover and learning higher-order relationship from implicit connections across users with shared preferences and across products with shared complementarity. To the

best of our knowledge, such innovative module was the first idea to defining indirect user-product and product-product interactions from which to mining enriched compatible representations for personalized complimentary recommendations.

2. We design two novel contrastive learning losses that can seamlessly integrate with BPR, ensuring adaptability and representation quality from multimodal data. The two contrastive learning losses align functional representations between direct personal preference and product compatibility as well as between indirect personal preference and compatibility, enabling the model to extract complementary information from different views, leading to robust and effective product representations.
3. Comprehensive experiments on two real-world benchmark datasets demonstrate the effectiveness of our APCL method, outperforming state-of-the-art methods in personalized fashion complementary recommendation tasks by large margins. By thorough ablation and comparative experiment, the results demonstrate the significance of the proposed AP module in capturing auxiliary hidden relations from sparse interaction data, and the proposed AP and CL modules can be adaptative and flexible to integrate with different baselines to improve recommendation accuracy.

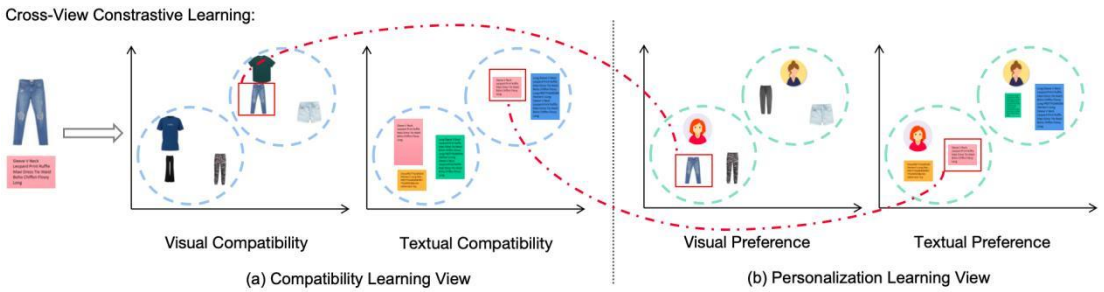

Figure 5-2 Illustration of cross-view contrastive learning.

By integrating user behavior, product relationships, and auxiliary contextual data, the proposed APCL is adaptable and flexible across various scenarios.

5.2 Approach

5.2.1 The Overall APCL scheme

The overall APCL scheme is shown as:

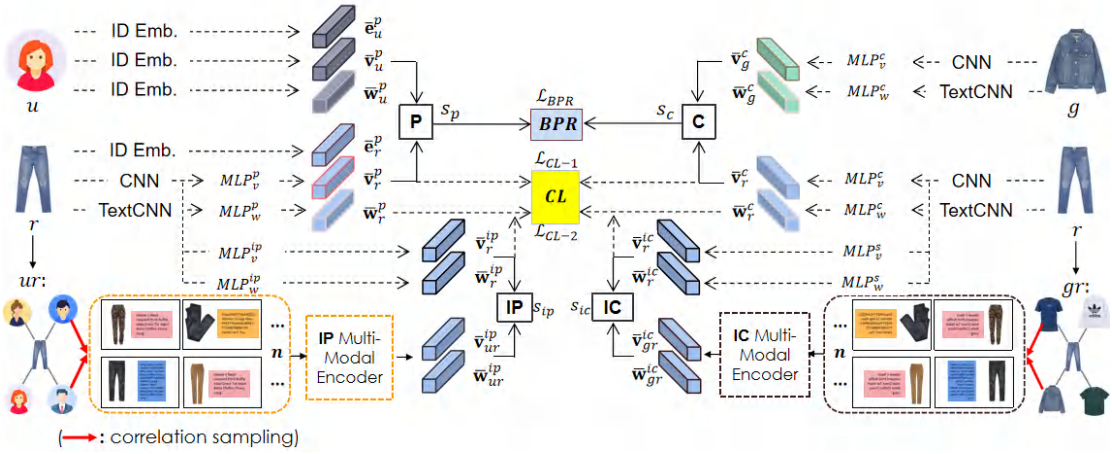

Figure 5-3 The overview of APCL.

As shown in Figure 5-3, the proposed APCL method (F) contains four main modules, including Personal preference modeling (s_p), multi-modal product Compatibility modeling (s_c), Indirect Personal preference Modeling (s_{ip}), and Indirect product Compatibility (s_{ic}). Multi-modal product Compatibility Learning view captures the degree of similarity between a target recommended product and a given product in terms of different content characteristics. While the representation from the Pezsonal preference Leaning view in the personalization space captures the relevant features of the target recommended product and users' personal preferences. Furthermore, Attentive Preference Modeling contains IP and IC can assist recommender systems in introducing a variety of implicit and higher-order user-item, item-item, and user-user interactions. This contributes to meeting consumers' personalization demands while keeping the diversity and thoroughness of the suggestion results. Thus, the overall output of the proposed APCL method is presented as:

$$p_r^{u,g} = s_p + s_c + \mu \cdot (s_{ip} + s_{ic}) \tag{5-1}$$

where μ is the corresponding weight, and the proposed APCL challenges the notion of

modeling personalization and compatibility as mutually exclusive factors. In other words, it rejects the idea that a high level of personalization should necessarily result in low compatibility, or vice versa, as traditional personalized complementary recommendations do, like GP-BPR (Song et al., 2019). Instead, the APCL model aims to treat these two components more equitably, blending them through a more comprehensive linear combination. This approach is theoretically more sound, although it introduces increased complexity in parameter selection during training.

5.2.2 Personal Preference Modeling View (P)

To achieve personalized fashion recommendation, similarly as previous methods, we apply BPR-MF (*Bayesian Personalized Ranking with Matrix Factorization)* (Rendle et al., 2012) The latent factors capture underlying patterns and preferences in the data. The prediction of users' personal preferences based on the implicit user-product interactions are as follows:

$$s_{ur} = \frac{\bar{e}_u \cdot \bar{e}_r}{\|\bar{e}_u\|\|\bar{e}_r\|} + \beta_u + \beta_r + \alpha \tag{5-2}$$

where $\bar{e}_u$ and $\bar{e}_r$ are normalized latent embeddings for user u and target recommended product r entities, respectively. Symmetrically with the compatibility modeling, here we use cosine similarity to replace the dot product between the user and the recommended product. Similar to Eq. (3-7), β_u and β_r are user and recommended product bias terms, and α is the global offset.

VBPR (He & McAuley, 2016) and GP-BPR(Song et al., 2019) further integrated the visual and textual features into the matrix factorization framework to learn personalized visual and textual preferences. Followed by those work, this module predicts the visual and textual personal preference as:

$$\begin{cases} \bar{v}_u^p = Norm(v_u) \\ \bar{v}_r^p = Norm(MLP_v^p(v_r)) \\ s_p^v = \dfrac{\bar{v}_u^p \cdot \bar{v}_r^p}{\|\bar{v}_u^p\|\|\bar{v}_r^p\|} \end{cases} \tag{5-3}$$

and

$$\begin{cases} \bar{w}_u^p = Norm(w_u) \\ \bar{w}_r^p = Norm(MLP_w^p(w_r)) \\ s_p^w = \dfrac{\bar{w}_u^p \cdot \bar{w}_r^p}{\|\bar{w}_u^p\| \|\bar{w}_r^p\|} \end{cases} \tag{5-4}$$

We use v_u and w_u embeddings associated with the user ID to help catch detailed user preferences on both visual and textual factors. Then the total user personal preference modeling can be obtained as:

$$s_p = s_{ur} + \pi \cdot s_p^v + (1 - \pi) \cdot s_p^w \tag{5-5}$$

Similarly, π is to control the contribution of each modality.

5.2.3 Multi-modal Product Compatibility Modeling View (C)

By leveraging both visual and textual modalities, the system can incorporate a wider range of features that contribute to the compatibility of complementary items. The visual information emphasizes visual coherence and aesthetics, while the textual information provides additional context and semantic understanding. Combining these modalities enables the system to consider various factors, including color coordination, pattern matching, style consistency, and adherence to specific fashion preferences, resulting in more accurate and comprehensive assessments of compatibility. To capture the latent representation between the two complementary products, this module employs two multi-layer perceptions (MLPs) followed by batch normalization layers as the feature encoders for each modality. Specifically, for visual compatibility learning, it defines the following function:

$$\begin{cases} \bar{v}_g^c = Norm(MLP_v^c(v_g)) \\ \bar{v}_r^c = Norm(MLP_v^c(v_r)) \\ s_c^v = \dfrac{\bar{v}_g^c \cdot \bar{v}_r^c}{\|\bar{v}_g^c\| \|\bar{v}_r^c\|} \end{cases} \tag{5-6}$$

The similarity between the given product and the recommended matching product in the learned visual compatibility space is evaluated using cosine similarity (s_c^v). The visual latent representations of the given product ($\bar{v}_g^c$) and the recommended complementary product ($\bar{v}_r^c$) obtained from the visual compatibility encoder are used for this calculation. Cosine similarity is a suitable metric in this context as it is scale-

invariant, meaning it is not affected by the magnitude of the vectors. This makes cosine similarity commonly used in recommendation systems. Similarly, for compatibility learning in the textual modality, a similar approach is defined as:

$$\begin{cases} \overline{w}_g^c = Norm(MLP_w^c(w_g)) \\ \overline{w}_r^c = Norm(MLP_w^c(w_r)) \\ s_c^w = \dfrac{\overline{w}_g^c \cdot \overline{w}_r^c}{\|\overline{w}_g^c\| \|\overline{w}_r^c\|} \end{cases} \tag{5-7}$$

Here, s_c^w denotes the textual compatibility score between the given product and the recommended complementary product, the superscript w is exclusive for textual modality. Thereafter, the results of the visual and textual compatibility scores are combined linearly to provide the comprehensive multi-modal compatibility score (s_c):

$$s_c = \pi \cdot s_c^v + \left(1 - \pi\right) \cdot s_c^w \tag{5-8}$$

5.2.4 AP Module

The proposed innovative Attentive Preference (AP) module contains two components: Attentive Indirect Personal Preference Modeling (IP) and Attentive Implicit Compatibility Modeling (IC), which will formulated in detail in following subsections.

5.2.4.1 Attentive Indirect Personal Preference Modeling View (IP)

In this subsection, we introduce the Attentive Indirect Personal Preference Modeling module, which calculates the similarity between products selected by other users who have selected the same target matching product (ur in Figure 5-3). Positive samples refer to products that have been chosen and positively evaluated by users. By computing the similarity between the target matching product and positive products selected by other users, we can identify additional products that share similar features or attributes to the target recommended complementary product. These similar products are likely to cater to similar needs or preferences, making them potential recommendations for the target users. For implementation details, we first filter users share the similar interest through a *Correlation Sampling Strategy* (will illustrate in the following subsection), then aggregate the n positive matching products chosen by

those users.

After that, we apply a cross modality attentive multi-modal attention encoder as shown in Figure 5-4, to generate a latent preference representation connected to the target matching product.

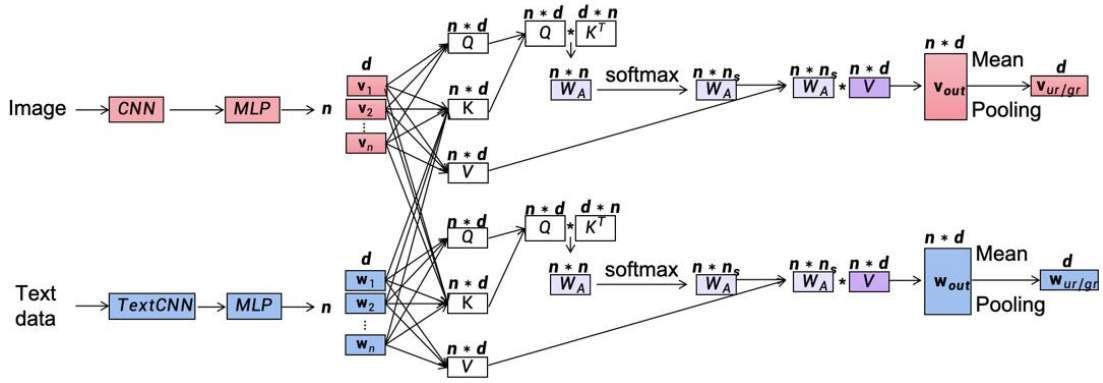

Figure 5-4 Implementation details in cross-modality Attentive Preference Encoder.

Given a set of query positive sample vectors $Q \in R^{n\times d}$, an attention function updates them via n key-value pairs, $K \in R^{n\times d}$ and $V \in R^{n\times d}$. To maintain simplicity, we ensure that the dimensions of each vector align with those of the query vectors. The output vector is obtained by taking a weighted sum of the value vectors V, where the weights are calculated based on the inner product between the query vector Q and the key vectors K:

$$Output = Attn(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{5-9}$$

To simplify the notation, the attentive preference encoding process (Eq.(5-9)) is denoted as $Attn(\cdot)$. For instance, the output generated from the visual attentive preference encoding is denoted as $Attn_v^s(v_{sn})$, where the superscript s represents the Preference Similarity, and the subscript v is specific to the visual modality. Similarly, the textual attentive preference encoding is denoted as $Attn_w^s(w_{sn})$, w is exclusive for textual modality. The overall Attentive Idirect Preference Similarity score s_{ip}, combining visual (s_{ip}^v) and textual (s_{ip}^w) information, can be predicted as follows:

$$s_{ip} = \pi \cdot s_{ip}^v + (1-\pi) \cdot s_{ip}^w \tag{5-10}$$

where the weight control parameter π is to balance the two modalities. Similarly, the

weighted linear combination is adopted to aggregate the visual (s_{ip}^{v}) and textual (s_{ip}^{w}) attentive preference modeling that is derived from the following functions:

$$\begin{cases} \bar{v}_{ur}^{ip} = Norm\left(MLP_{v}^{ip}\left(Attn_{v}^{ip}(v_{ur})\right)\right) \\ \bar{v}_{r}^{ip} = Norm\left(MLP_{v}^{ip}(v_{r})\right) \\ s_{ip}^{v} = \dfrac{\bar{v}_{ur}^{ip} \cdot \bar{v}_{r}^{ip}}{\|\bar{v}_{ur}^{ip}\|\|\bar{v}_{r}^{ip}\|} \end{cases} \tag{5-11}$$

$$\begin{cases} \bar{w}_{ur}^{ip} = Norm\left(MLP_{w}^{ip}\left(Attn_{w}^{ip}(w_{ur})\right)\right) \\ \bar{w}_{r}^{ip} = Norm\left(MLP_{w}^{ip}(w_{r})\right) \\ s_{ip}^{v} = \dfrac{\bar{w}_{ur}^{ip} \cdot \bar{w}_{r}^{ip}}{\|\bar{w}_{ur}^{ip}\|\|\bar{w}_{r}^{ip}\|} \end{cases} \tag{5-12}$$

where $\bar{v}_{ur}^{ip}$ and $\bar{w}_{ur}^{ip}$ denotes the visual and textual latent user attentive preference representation obtained from the *Attentive Indirect Preference Encoder* and an MLP layer for similarity projection followed by a normalization layer. The cosine similarity calculation method is adopted to measure the distance between the target recommended complementary product and user-attentive indirect preference representation in similarity latent space.

5.2.4.2 Correlation Sampling Strategy

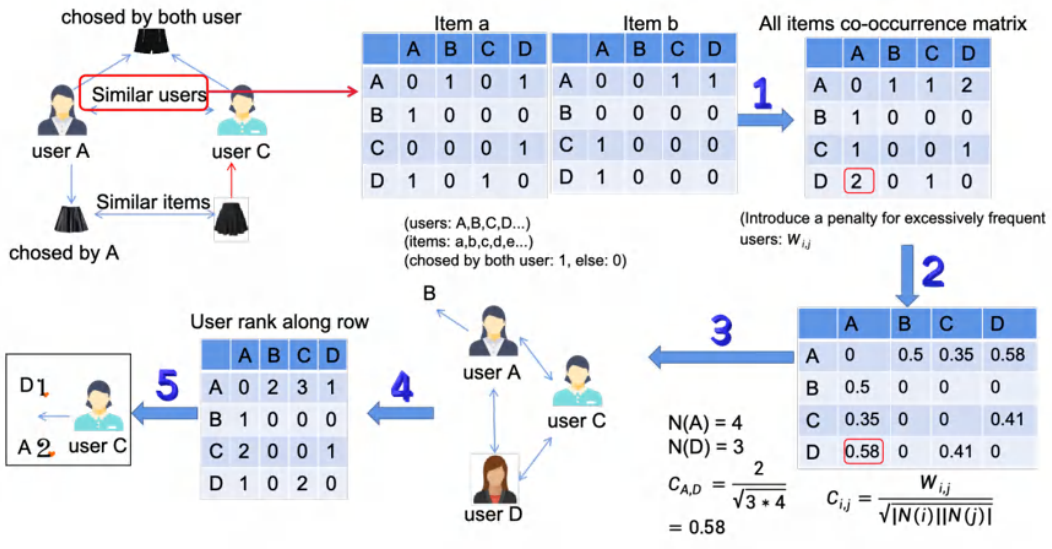

Figure 5-5 Simplified Example of Correlation Sampling Strategy.

The implementation details of the correlation sampling strategy we mentioned in

Figure 5-5 are shown as follows, and here we set the ur list selection for IP modeling as an example:

- Step 1: For each item in the interaction matrix, if both user A and user B have interacted with that item, we record the interaction as 1; otherwise, we record it as 0. This process is repeated for all items in the interaction matrix until all interactions between users and items have been recorded. By summing the matrices, we get a user-user co-occurrence ($\boldsymbol{W}_{ij}$) matrix (cumulative summation) based on each positive recommended product with a penalty for excessively frequent users as follows:

$$\boldsymbol{W}_{ij} = \sum_{u \in |N(i) \cap N(j)|} \frac{1}{\log\left(1 + |N(U)|\right)} \tag{5-13}$$

where N represents the number of occurrences in the whole training set. $N(i) \cap N(j)$ represents the frequency with which user u interacts with both item i and item j simultaneously. $N(U)$ denotes the total number of users.

- Step 2: Output user similarity score $\boldsymbol{C}_{ij}$ matrix by *Normalization* as following function:

$$\boldsymbol{C}_{ij} = \frac{\boldsymbol{W}_{ij}}{\sqrt{|N(i)||N(j)|}} \tag{5-14}$$

- Step 3: List similar users for each user.
- Step 4: Rank the similar users by similarity score ($\boldsymbol{C}_{ij}$).
- Step 5: Find top K similar users to each target user, and for each similar user, we use cross attention to select n related matching products ur from content-based learning manner.

5.2.4.3 Attentive Indirect Product Compatibility Modeling View (IC)

Symmetrical to IP module, the IC module calculates the similarity between products selected by other given products which have selected the same target matching

product (gr), following the same correlation sampling strategy in IP branch. The overall Attentive Indirect Compatibility Similarity score (s_{ic}) based on visual and textual information can be formulated as:

$$s_{ic} = \pi \cdot s_{ic}^{v} + \left(1 - \pi\right) \cdot s_{ic}^{w} \tag{5-15}$$

The weighted linear combination to aggregate the visual s_{ic}^{v} and textual s_{ic}^{w} attentive compatibility modeling that is derived from the following functions:

$$\begin{cases} \bar{v}_{gr}^{ic} = Norm\left(MLP_{v}^{ic}\left(Attn_{v}^{ic}(v_{gr})\right)\right) \\ \bar{v}_{r}^{ic} = Norm\left(MLP_{v}^{ic}(v_{r})\right) \\ s_{ic}^{v} = \dfrac{\bar{v}_{gr}^{ic} \cdot \bar{v}_{r}^{ic}}{\left\|\bar{v}_{gr}^{ic}\right\|\left\|\bar{v}_{r}^{ic}\right\|} \end{cases} \tag{5-16}$$

$$\begin{cases} \bar{w}_{gr}^{ic} = Norm\left(MLP_{w}^{ic}\left(Attn_{w}^{ic}(w_{gr})\right)\right) \\ \bar{w}_{r}^{ic} = Norm\left(MLP_{w}^{ic}(w_{r})\right) \\ s_{ic}^{v} = \dfrac{\bar{w}_{gr}^{ic} \cdot \bar{w}_{r}^{ic}}{\left\|\bar{w}_{gr}^{ic}\right\|\left\|\bar{w}_{r}^{ic}\right\|} \end{cases} \tag{5-17}$$

where $\bar{v}_{gr}^{ic}$ and $\bar{w}_{gr}^{ic}$ denotes the visual and textual latent preference representation under the IC space. The cosine similarity calculation method is adopted to measure the distance between the target matching product and Attentive Indirect Product Compatibility representation in another similarity latent space.

5.2.5 Optimization

For the optimization, APCL adopts a jointly optimization method, which combines both BPR pairwise learning, and two cross-view learning based contrastive optimization branches. The overall model objective is defined as following function:

$$\mathcal{L} = \gamma_1 \cdot \mathcal{L}_{BPR} + \gamma_2 \cdot \left(\mathcal{L}_{cl-1} + \mathcal{L}_{cl-2}\right) \tag{5-18}$$

where γ_1 and γ_2 are the weight of the BPR objective, and two contrastive learning objectives, the Cross P and C Views Contrastive Learning $\mathcal{L}_{cl-1}$, and the Cross IP and IC Views Contrastive Learning ($\mathcal{L}_{cl-2}$), respectively. The goal of the Bayesian Personalized Ranking (BPR) loss (Rendle et al., 2012) is to enable personalized ranking in recommender systems. It achieves this by optimizing the relative ordering of positive samples (items that the user likes) and negative samples (items that the user may not

like). On the other hand, the other two contrastive losses, inspired by *Information Noise-Contrastive Estimation* (InfoNCE) (Gutmann & Hyvärinen, 2010), are primarily used for unsupervised learning to learn deep feature representations. These losses optimize the distance between positive samples (similar data pairs) and negative samples (dissimilar data pairs). Both types of losses aim to optimize the comparison between positive and negative samples. By combining them through a weighted linear combination, we achieve joint optimization for our implicit content-based recommendation task.

For **BPR pairwise learning** objective ($\mathcal{L}_{BPR}$), let us identify the preference score for a pair of positive and negative samples obtained from Eq. 错误!未找到引用源。) as $p_{r+}^{u,g}$ and $p_{r-}^{u,g}$, respectively. The BPR loss on the entire training set D is determined as follows:

$$\mathcal{L}_{BPR} = \sum_{D}\left[-ln\left(\sigma\left(p_{r+}^{u,g} - p_{r-}^{u,g}\right)\right)\right] + \frac{\lambda}{2}\left\|\Theta_F\right\|^2 \tag{5-19}$$

where $D = \{(u, g, r+, r-)|u \in U \wedge g \in G \wedge (r+, r-) \in R\}$. λ is the non-negative hyperparameter and Θ_F represents the set of parameters of the model.

For **Cross P and C Views Contrastive leaning** objective ($\mathcal{L}_{cl-1}$), APCL proposes to aliment latent target recommended product representations obtained from Multi-modal Product Compatibility Modeling View ($\bar{v}_r^c$, $\bar{w}_r^c$) and Personal Preference Modeling View ($\bar{v}_r^p$, $\bar{w}_r^p$). The Cross-view Contrastive Learning for both visual and textual modalities is defined as follows:

$$\begin{cases} \mathcal{L}_{cl-1}^{v} = \sum_{V}\left[-\log\dfrac{exp\left(sim\left(\bar{v}_{r+}^{p}, \bar{v}_{r+}^{c}\right)/t\right)}{exp\left(sim\left(\bar{v}_{r+}^{p}, \bar{v}_{r+}^{c}\right)/t\right) + exp\left(sim\left(\bar{v}_{r+}^{p}, \bar{v}_{r-}^{p}\right)/t\right)}\right] \\ \mathcal{L}_{cl-1}^{w} = \sum_{W}\left[-\log\dfrac{exp\left(sim\left(\bar{w}_{r+}^{p}, \bar{w}_{r+}^{c}\right)/t\right)}{exp\left(sim\left(\bar{w}_{r+}^{p}, \bar{w}_{r+}^{c}\right)/t\right) + exp\left(sim\left(\bar{w}_{r+}^{p}, \bar{w}_{r-}^{p}\right)/t\right)}\right] \\ \mathcal{L}_{cl-1} = \mathcal{L}_{cl-1}^{v} + \mathcal{L}_{cl-1}^{w} \end{cases} \tag{5-20}$$

where t is the temperature hyper-parameter, and $sim(\cdot)$ denotes similarity calculation and here we adopt dot product. The subscript $r+$ and $r-$ represent the positive and negative pair samples. The proposed Cross-view Contrastive loss function is based on

a contrastive learning framework, aiming to enhance the feature representation by comparing the differences between positive and negative samples. Its objective is to maximize the similarity between positive samples generated from the Personal Preference Modeling View and those from the Multi-modal Product Compatibility Modeling View, while minimizing the similarity between positive and negative samples.

In personalized fashion complementary product recommendation, the target recommended products are independently mapped to personalization and compatibility spaces. By using the Cross-view Contrastive loss function, the model learns to strike a balance between personalization and compatibility representations. The temperature parameter (t) facilitates exploring various combinations of personalization and compatibility features, leading to more diverse recommendations. Simultaneously, this loss function emphasizes relevance by comparing the similarity of positive and negative samples, effectively filtering out recommendation products that match the user's personalization and are compatible with a given product.

For **Cross IP and IC Views Contrastive Learning** ($\mathcal{L}_{\mathrm{cl}-2}$), Symmetrically, the proposed IP and IC Views Contrastive Learning compares the target matching products latent representation derived from Attentive Indirect Personal Preference Modeling ($\bar{v}_r^{ip}$, $\bar{w}_r^{ip}$) and Attentive Implicit Compatibility Modeling ($\bar{v}_r^{ic}$, $\bar{w}_r^{ic}$) views. We define the Cross IP and IC Views Contrastive loss ($\mathcal{L}_{\mathrm{cl}-2}$) as follows:

$$\begin{cases} \mathcal{L}_{cl-2}^{v} = \sum_{V} \left[-\log \dfrac{exp\left(sim\left(\bar{v}_{r+}^{ip}, \bar{v}_{r+}^{ic}\right)/t\right)}{exp\left(sim\left(\bar{v}_{r+}^{ip}, \bar{v}_{r+}^{ic}\right)/t\right) + exp\left(sim\left(\bar{v}_{r+}^{ip}, \bar{v}_{r-}^{ip}\right)/t\right)} \right] \\ \mathcal{L}_{cl-2}^{w} = \sum_{W} \left[-\log \dfrac{exp\left(sim\left(\bar{w}_{r+}^{ip}, \bar{w}_{r+}^{ic}\right)/t\right)}{exp\left(sim\left(\bar{w}_{r+}^{ip}, \bar{w}_{r+}^{ic}\right)/t\right) + exp\left(sim\left(\bar{w}_{r+}^{ip}, \bar{w}_{r-}^{ip}\right)/t\right)} \right] \\ \mathcal{L}_{cl-2} = \mathcal{L}_{cl-2}^{v} + \mathcal{L}_{cl-2}^{w} \end{cases} \quad (5\text{-}21)$$

Similar to $\mathcal{L}_{\mathrm{cl}-1}$, we incorporate a temperature hyper-parameter t and utilize the dot product similarity represented by $sim(\cdot)$ for this loss function. The advantage of this loss is its ability to work without explicit labels, which is particularly valuable for content-based recommender systems where complete feedback on user preferences

may not be available. Content-based recommender systems often encounter the cold-start problem, where making recommendations for new users or items becomes challenging due to limited interaction data. However, by learning effective feature representations, we can compare new users or products with existing ones to make relevant recommendations, mitigating the impact of missing data. This loss aids in obtaining more similar interest representations for users with similar preferences, enhancing the performance of recommendation tasks in the presence of data gaps.

5.3 Experiments

To evaluate the performance of the proposed APCL model, we performed comprehensive experiments using publicly accessible fashion datasets. The experiments were designed to answer the following key research questions:

Q1: Does the APCL model outperform existing state-of-the-art methods in terms of performance?

Q2: Does the incorporation of AP and CL within the APCL model result in enhanced overall performance?

Q3: What is the impact of multi-modalities (visual and textual) and hyperparameters tuning on the model's effectiveness?

Q4: How effective is the APCL model in terms of personalized fashion complementary recommendation in different interaction environments, such as cold start?

5.3.1 Overall Performance Comparison (Q1)

Table 5-1 compares different benchmark methods' performance. From this table, we can obtain the following observations:

Table 5-1 The overall comparison of IQON3000 and Polyvore datasets in terms of setting Top as the given product and Bottom as the matching product to be

recommended. Bolded data are the best results.

Method	IQON3000				Polyvore			
	AUC	HR@10	NDCG@10	MRR	AUC	HR@10	NDCG@10	MRR
BPR-MF	0.8309	0.7024	0.5243	0.4687	0.7639	0.6916	0.5434	0.4973
TBPR	0.8316	0.6361	0.4301	0.3666	0.7839	0.6502	0.4868	0.4359
VBPR	0.8360	0.6941	0.5230	0.4694	0.8074	0.7141	*0.5848*	0.5443
VT-BPR	0.8384	0.7003	0.5134	0.4551	0.8105	0.6846	0.5266	0.4772
GP-BPR	0.8569	0.7396	0.5849	0.5366	0.8232	0.7121	0.5716	0.5277
PCE-NET	0.8341	0.6399	0.4110	0.3399	0.8235	0.4208	0.2380	0.1825
CP-TransMatch	*0.8842*	*0.8789*	*0.6453*	*0.5430*	*0.9001*	*0.8573*	0.5472	0.5144
APCL	**0.9739**	**0.9509**	**0.9103**	**0.8973**	**0.9832**	**0.8999**	**0.8343**	**0.8123**

1) The VT-BPR model outperforms its counterparts, such as V-BPR, T-BPR, and BPR-MF, highlighting the significant benefits of integrating multi-modal data for enhancing compatibility modeling, especially in fashion recommendations which heavily relies on semantic features.
2) GP-BPR and PCE-NET surpasses both the general compatibility and personal preference baselines, demonstrating that personalized fashion complementary recommendation should take into account both product-product compatibility and user-product preference modeling.
3) When specifying bottom to recommend matching tops, the overall results are lower compared to specifying tops to recommend matching bottoms. This is because the interaction data associated with tops has a higher distributional difference and is sparser. This highlights the critical importance of abundant interaction data for effective recommendations.
4) CP-TransMatch formulates the personalized fashion matching problem as a multi-relational connectivity problem and improves the prediction results by a large margin, which helps to better understand and exploit the complexity of user-item-item relationships. However, using a single-component translation operation to simulate target third-order interactions heavily relies on the diversity of user interaction data. When user behavior data is sparse, it not only reduces the

effectiveness of user relationship learning but also negatively affects the item representation learning, ultimately compromising the recommendation accuracy.

5) The proposed APCL outperforms other methods in both two datasets, and in all metrics, demonstrating the effectiveness of our proposed scheme for personalized fashion complementary recommendations.

5.3.2 Ablation Study (Q2)

To systematically evaluate the impact of the four key modules integrated into our proposed APCL framework, we performed ablation studies on variants of the model, specifically APCL-w/o-AP, APCL-w/o-CL, APCL-w/o-P, and APCL-w/o-C. These variants were created by removing the Attentive Preference (AP), Cross-View Contrastive Learning (CL), Personal Preference (P), and Product Compatibility (C) modules from the APCL framework. More specifically, APCL-w/o-CL means our proposed method without all contrastive learning loss but only keeps the BPR loss.

Additionally, to evaluate the individual contribution of visual and textual modalities, we conducted experiments on APCL-w/o-V and APCL-w/o-T models across both datasets. The notation -w/o-V represents a model that retains only textual information while excluding visual data, and vice versa for -w/o-T, which includes only visual information and omits textual data.

Table 5-2 presents a comprehensive comparison of the performance between the APCL model and its variants, measured across four metrics. The findings from these experiments clearly indicate that the integration of both visual and textual modalities, along with the Personal Preference, Product Compatibility, Attentive Preference, and Cross-View Contrastive Learning modules, substantially enhances the model's accuracy in comparison to the variant models.

Table 5-2 Ablation Experiment. -w/o refers to the evaluation results of the models WITHOUT applying according modules.

Method	IQON3000				Polyvore			
	AUC	HR@10	NDCG@10	MRR	AUC	HR@10	NDCG@10	MRR

-w/o-V	0.9541	0.9244	0.8589	0.8384	0.8016	0.4104	0.2354	0.1819
-w/o-T	0.9296	0.8780	0.7704	0.7370	0.9599	0.8730	0.7542	0.7167
-w/o-AP	0.9528	0.9128	0.7759	0.7308	0.9271	0.7741	0.6166	0.5670
-w/o-CL	0.9635	0.9345	0.8836	0.8675	0.9434	0.8328	0.6897	0.6447
-w/o-P	0.6654	0.4094	0.1597	0.0864	0.9248	0.8251	0.7156	0.6808
-w/o-C	0.9581	0.9117	0.8796	0.8688	0.9459	0.8355	0.6844	0.6363
APCL	**0.9739**	**0.9509**	**0.9103**	**0.8973**	**0.9832**	**0.8999**	**0.8343**	**0.8123**

The ablation study findings presented in Table 5-2 show that the performance of all examined metrics noticeably declines when any one module is left out of the combination. This illustrates how crucial each module is to raise the overall efficacy of the model. The synergistic combination of all four modules and two modalities provides the best results, indicating that the most accurate recommendations require the comprehensive integration of these elements.

5.3.3 Effect of Key Hyperparameters. (Q3)

We conduct a series of experiments to evaluate the impact of various key hyperparameters on the performance of our proposed APCL model. These hyperparameters include the temperature parameter t in Eq.(5-20) and Eq.(5-21), the contribution weight μ in Eq.(5-1), and the weight ratio γ_2/γ_1 in Eq.(5-18). As illustrated in Figure 5-6, each of these components plays a crucial role in enhancing the personalized fashion complementary recommendation performance. Notably, the model's performance exhibits only minor sensitivity to changes in these hyperparameters.

The temperature parameter t is particularly influential; a higher value encourages exploration, helping the model avoid local optima, while a lower value promotes exploitation, increasing the model's confidence in differentiating between positive and negative samples. Our findings, depicted in Figure 5-6, indicate that the APCL model achieves optimal performance with t values ranging from 1 to 5. This range fosters a

more exploratory behavior, as the model becomes more uncertain and tends to assign similar probabilities to both positive and negative samples.

Furthermore, when the BPR loss and Cross-view Contrastive loss are given equal importance (γ_2/γ_1 equals 0.5), the APCL model demonstrates its best performance. This balance appears to be critical in optimizing the model's effectiveness.

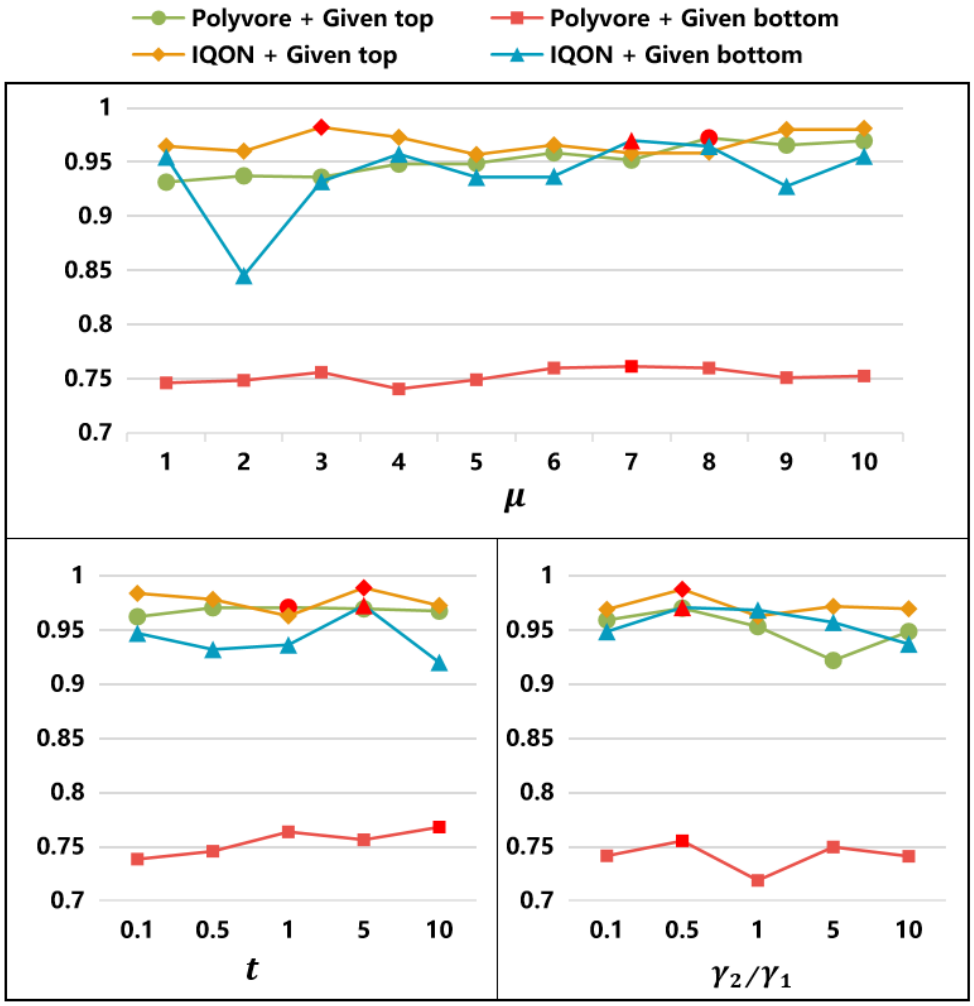

Figure 5-6 Impacts of different parameters.

5.3.4 On Different Interaction Frequency (Q4)

To verify the effectiveness and robustness of our proposed method, we conducted a series of experiments comparing the prediction performance of various baseline models across different interaction environments. Figure 5-7 illustrates the prediction accuracy of these baselines within distinct interaction frequency ranges for the target

matching item: [0, 2], (2, 5], (5, 10], and (10,]. For instance, the range (10,] indicates that the target matching item has more than ten interactions. Additionally, the Figure 5-7 includes purple bars representing the percentage of samples within each frequency range.

The sample distribution reveals that the majority of samples fall within the [0, 2] range, accounting for the majority of the total samples in all datasets. As the interaction frequency increases, the number of samples decreases, highlighting a skewed distribution where most predictions occur in the low interaction frequency range.

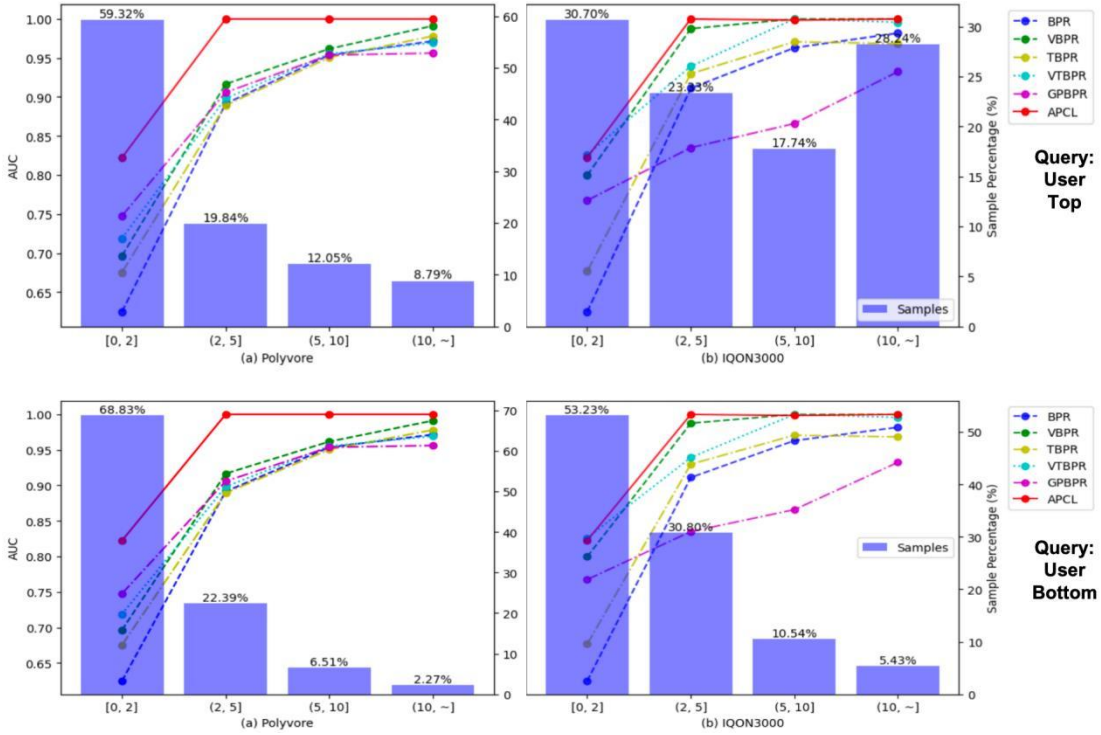

Figure 5-7 Comparison of Prediction Accuracy across Different Interaction Frequency Ranges for Various Methods.

Despite starting with a slightly lower initial accuracy, the APCL model outperforms all other methods. It quickly achieves perfect accuracy with minimal interaction, demonstrating its robustness and efficiency. This characteristic makes APCL particularly suitable for scenarios that demand high precision with fewer interactions.

5.3.5 On Cold Start (Q4)

To emphasize the effectiveness of the proposed adaptive AP and CL module in enhancing recommendation accuracy and robustness, especially in cold start scenarios. Figure 5-8 presents the AUC values of different recommendation algorithms when augmented with the AP and CL modules. On the Polyvore dataset, the addition of the AP module consistently improves the AUC values across all methods, with the most significant enhancement observed in the VBPR method. This substantial increase suggests that the AP module effectively captures user preferences and item relationships, leading to more accurate recommendations. The CL module also contributes positively, as seen in the VTBPR method, where the AUC improves by a large margin, indicating that cross-view contrastive learning strengthens the model's ability to discern between relevant and irrelevant features. Similarly, on the IQON3000 dataset, the integration of the AP module into the BPR method results in a notable increase, demonstrating the module's effectiveness in enhancing the model's prediction accuracy. The CL module, when combined with the APCL method, outperforming both the original model and the model with only the AP module, which underscores the complementary nature of the CL module in refining the recommendation quality.

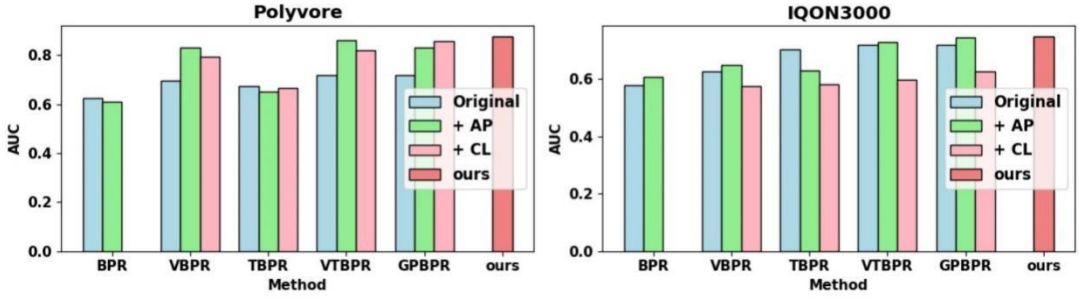

Figure 5-8 Performance comparison in terms of AUC Values for Baselines with and without AP and CL in Cold Start Scenario.

Notably, the GPBPR method exhibits a remarkable increase in AUC when both the AP and CL modules are applied. This result highlights the synergistic effect of combining both modules. In conclusion, the experiment results clearly demonstrate the

positive impact of incorporating the AP and CL modules into personalized fashion complementary recommendation systems. These two modules not only improve the AUC values but also contribute to the robustness of the existing recommendation algorithms, particularly in cold start scenarios where traditional methods may struggle due to the sparsity of interaction data. The consistent improvements across different methods and datasets underscore the adaptiveness and effectiveness of the AP and CL modules.

5.3.6 Case Study

We performed qualitative assessments by collecting matching products for comparison and exhibiting a few test examples in order to confirm the usefulness of our suggested approach in real-world applications. The comparison findings regarding the accuracy of recommendations are displayed in Figure 5-9.

	u	g	r^+	r^-	Comparison
1					GP-BPR ✗ CP-TransMatch √ **ours** √
2					GP-BPR ✗ CP-TransMatch ✗ **ours** √
3					GP-BPR √ CP-TransMatch √ **ours** √
4					GP-BPR ✗ CP-TransMatch √ **ours** √

Figure 5-9 Recommendation case results provided by different baselines.

Specifically, we randomly selected several user-product-product transaction triplets, denoted as $\langle u|g|r^+\rangle$, from the testing dataset. In each triplet, we chose a

negative matching product (r^-) by selecting a product that the user had not interacted with before. The objective was to correctly match the appropriate bottom clothing (r^+) for each user given the corresponding top garment (g).

As shown in Figure 5-9, in every scenario, our model performs nicely. Furthermore, we found that traditional models like GP-BPR had difficulties differentiating between two matching products when their semantic features were extremely similar. This limitation results from these models' heavy reliance on the inherent attributes of the products themselves, which ignores the context that user interactions provide. Models such as CP-TransMatch and our APCL can achieve higher prediction accuracy than general compatibility modeling by integrating user implicit connections. The outcomes also show how well our suggested approach works for the task of recommending personalized complementary clothing.

5.3.7 Application Example

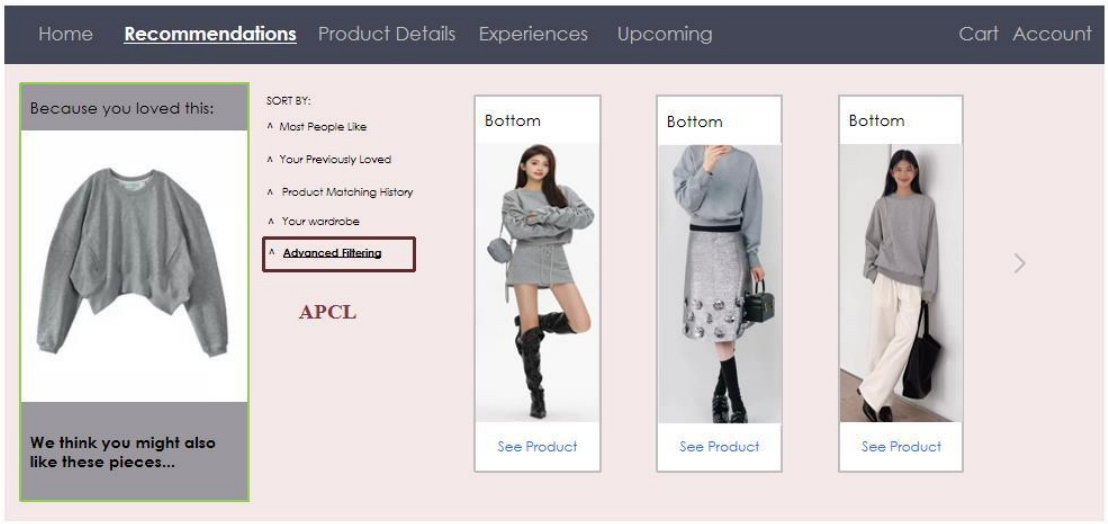

Figure 5-10 The application example of APCL.

Figure 5-10 shows an application example of the proposed APCL. Even for new users with limited transaction history, personalized recommendations can still be effectively generated by leveraging APCL. Specifically, the system can identify similar users. Moreover, the system can customize the connection order by exploring higher-order relationships in the user-item interaction graph. This approach allows for a more

comprehensive understanding of user preferences, even in the absence of extensive transactional data. By integrating these strategies, the system ensures that new users receive meaningful and personalized recommendations, thereby enhancing their engagement and satisfaction.

5.4 Summary

To overcome the limitations of data sparsity, inefficient representation learning, and limited adaptability, the APCL framework are proposed, this is a novel approach that integrates multi-modal data, implicit relationships, and cross-view contrastive learning to achieve robust and contextually relevant recommendations. The key components of the APCL framework: Personal Preference Modeling (P), Product Compatibility Modeling (C), Attentive Indirect Personal Preference (IP), and Attentive Indirect Compatibility (IC) work synergistically to enhance the model's performance. By combining direct and indirect modeling of user preferences and product compatibility, the framework addresses data sparsity while enriching representation quality. Furthermore, the integration of cross-view contrastive learning ensures alignment between personalization and compatibility, enhances product representation learning effectively and without heavily relies on multi-modal information.

Empirical evaluations on the Polyvore and IQON3000 datasets demonstrate the superiority of the proposed framework over state-of-the-art methods. The APCL framework achieves significant improvements, particularly excelling in cold-start and sparse data scenarios. Ablation studies confirm the critical contributions of each module, with results showing that removing any single component leads to noticeable performance degradation. These findings highlight the importance of the unified and comprehensive design of the APCL framework.

Chapter 6. Hypergraph-Enhanced Contrastively Regularized Transformer for Multi-Behavior Recommendation

In the previous chapters, we delved into the key challenges of personalization and data scarcity from three different methodological perspectives. We now turn to explore these issues in the context of sequential recommendation tasks. This chapter serves as a bridge between theoretical frameworks and practical applications, further elucidating the importance of addressing personalization and data scarcity in domains where user preferences dynamically change over time.

The importance of addressing personalization in sequential recommendation cannot be overstated. It is in this domain that the true complexity of user behavior is revealed, as preferences are not static but are influenced by numerous factors, including past interactions, current trends, and even temporal changes. On the other hand, data scarcity limits the availability of historical data points, thereby constraining the model's ability to learn and make accurate predictions, exacerbating these challenges. Therefore, this chapter is dedicated to deepening our understanding of how to effectively capture and leverage user behavior patterns to enhance recommendation systems, even in the face of limited data. In this Chapter, we will further delve into the complexities of multi-behavior sequential recommendation, leveraging the power of state-of-the-art Graph Neural Networks (GNNs) and transformers. As we progress through this chapter, we will provide a detailed overview of the implementation methods for these advanced models and discuss how they can be utilized to address the specific challenges of data sparsity in sequential recommendation.

6.1 Introduction

On various e-commerce platforms, users' interests are often hidden in the wealth

of user interaction data, not limited to purchases, but also includes browsing, clicking, favoriting, adding to cart, and many other types of behavior. By integrating various behavior data, some multi-behavior sequential recommendation systems were developed to identify users' purchasing intentions and provide more personalized and precise product predictions. Since historical user behaviors are often recorded in sequences of different time lengths, and users' behavioral habits may change over time (J. Z. Chen et al., 2023). Especially for e-commerce sites, which may need to consider both users' short-term activities and long-term shopping history. Since short-term activities may reflect a user's immediate needs, long-term activities reveal more users' sustained preferences for certain items or categories over an extended period. Thus mining users' both long-term static preferences and short-term dynamic interests (Xia et al., 2023) becomes a popular target in many e-commerce applications. This is also exactly the pursuit of multi-behavior recommendation systems, namely to predict user's future activities by modeling the sequential dependencies of users' interaction and the correlation of users' behavior patterns (Xia et al., 2023; Y. Yang et al., 2022).

Figure 6-1 An online shopping example of multi-behavior sequential recommendation. The target is to predict the next item users would interact with, based on their historical behavior sequences.

To capture complex multi-behavioral patterns, some graph neural-based networks are developed to integrate multiple behaviors into the same structure, unlike traditional graphical models that can only deal with single-behavior relationships. On the other hand, to capture both short-term and long-term preference, and inspired by attention-based frameworks like SASRec (Kang & McAuley, 2018), BERT4Rec (Sun et al., 2019), many multi-scale transformer recommendation methods model behavior

through different time (Y. Yang et al., 2022) or frequency scales (Shao et al., 2024). Despite the advancement, these methods are still commonly restricted by transaction data limitations, such as complex data environments like limited and noisy data.

To address above mentioned problems, a novel model named ***S**imilarity **G**raph-enhanced **M**ulti-**S**cale **T**ransformer* (**SG-MST**), seamlessly integrates the proposed ***S**imilarity Augmented Multi-Behavior Hyper-**G**raph* (**SG**) structure and the *Contrastively Regularized **M**ulti-**S**cale **T**ransformer* (**MST**) module.

As in a multi-behavior sequential recommendation structure, a multi-behavior hypergraph is prevalently utilized to effectively capture different behaviors and leverage higher-order relationships. While sparse multi-behavior interaction may cause limited connection in building adjacent multi-behavior matrices. To enhance the hypergraph construction and to create a denser and more informative adjacent matrix, we adopt a ***S**imilarity Augmented Multi-Behavior Hyper-**G**raph* (**SG**) structure. Specifically, beyond multi-behavior hypergraph, a top-K similarity augmented hypergraph is used to add supplementary connections, based on their top-K similar contextual correlated items. The proposed hypergraph tries to generate an adjacency matrix from multi-behave sequence interactions to define neighbor structures for the users that suffer from sparse interactions. This approach allows a sparse adjacency matrix to be filled with similar items rather than zeros (Fan et al., 2023).

Moreover, to capture long and short-term preference embedded in various sequences, sparse data and noisy data may negatively affect the Multi-Scale Transformer, making the model unstable in extracting behavioral patterns at different scales. In this section, we introduce a *Contrastively Regularized **M**ulti-**S**cale **T**ransformer* (**MST**) module. Specifically, we perform additional augmentation strategy and contrastive regularization within dynamically sampled temporal scales in a multi-scale transformer to uncover more informative short-term and long-term behavior patterns. Embedding data augmentation such as reordering, inserting, and replacing within temporal scales obtain diverse sequential representation, and helps

mitigate the impact of data constraint during attention training (Xie et al., 2022). Contrastive learning on different augmented views, enables the model to focus more on meaningful user behavior patterns while suppressing noise and irrelevant information, thereby enhancing robustness and accuracy.

6.2 Approach

As the detailed task formulation is illustrated in Chapter 2.5.1, here we start with the overall structure. The architecture of SG-MST is illustrated in Figure 6-2, which contains two modules: Similarity Augmented Multi-Behavior Hypergraph and Contrastively Regularized Multi-Scale behavior-aware Transformer encoder. The SG-MST first injects behavior sequence by an embedding layer combining ID, time and behavior type information. Then (a) the Similarity Augmented Multi-Behavior Hypergraph (SG), is designed to capture complex behavior and contextual dependencies between items to enhance sequential representation. (b) The Contrastively Regularized Multi-Scale Transformer encoder (MST) with augmentation is designed to capture sophisticated and informative behavior-sequential patterns from both coarse and fine-grained scales.

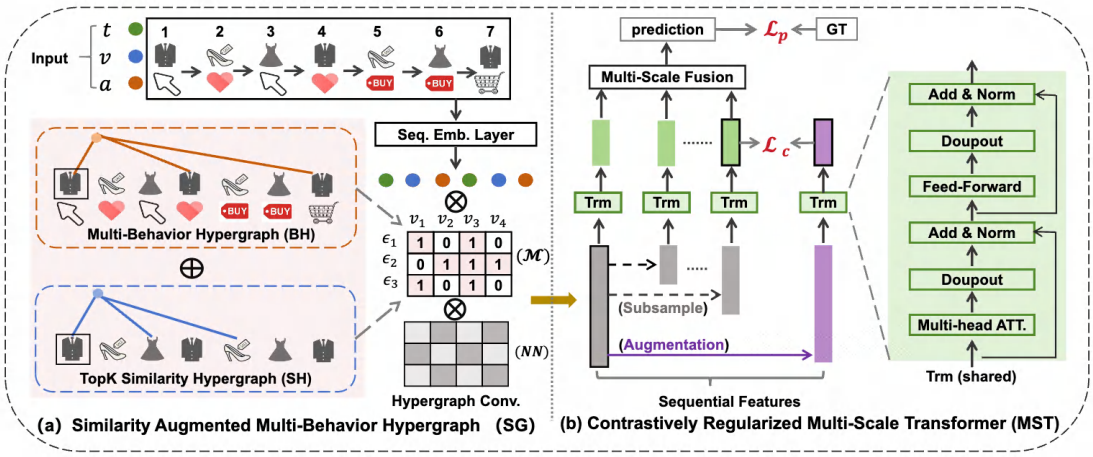

Figure 6-2 SG-MST structure overview.

In this section, before introducing the framework of SG-MST and the two new modules.

Sequence Embedding Layer:

We combine three different kinds of embedding to comprehensively represent each given item v_n in the input sequence. Specifically, we concurrently encode the item information associated with the interaction behavior type, and position information. The combined item context embedding is defined as:

$$x_n = v_n + a_n + t_n \tag{6-1}$$

where $v_n, a_n, t_n \in R^d$ are the item embedding, behavior type embedding, and position embedding initialized from embedding tables $E_V \in R^{|V|*d}$, $E_A \in R^{|A|*d}$, and $E_P \in R^{|N|*d}$. We then combine all item representations obtained from the embedding layer to represent the behavior-aware interaction sequence S^{u_i} of user u_i as $X^{u_i} \in R^{N*d}$.

6.2.1 Similarity Augmented Multi-Behavior Hypergraph (SG)

We design an augmented hypergraph structure to capture complex dependencies between items in user sequences. Within this hypergraph structure, we first create an item-wise hypergraph based on user multi-behavior interaction sequences. The adjacent matrix is constructed directly from multi-behavior sequence interactions to define the neighborhood structure for items, that suffer from sparse interactions making it less representative for inactive items (Fan et al., 2023).

To leverage a more informative connection matrix, we also introduce a top-K behavior-aware similarity hypergraph as a complementary augmentation based on latent context information.

Multi-Behavior Hypergraph (BH):

To capture complex interactions and high-order connections among all items, we construct an item-wise multi-behavior hypergraph G_v by connecting each items with hyperedges ε_q, where $\left|\varepsilon_q\right|$ is the number of unique items in the sequence S^{u_i}. Each item in the sequence is assigned with a hyperedge $\epsilon \in \varepsilon_q$.

The item-wise multi-behavior dependency connection matrix M_v is generated, where each element m_v is defined as:

$$m_v(v_n, \epsilon_n) = \begin{cases} 1, v_n \in \epsilon_q^n \\ 0, otherwise \end{cases} (m_v \in M_v) \quad (6\text{-}2)$$

where ϵ_n represents the hyperedge assigned to item v_n, ϵ_q^n denotes the set of multi-typed interactions of $u_i - v_n$, and $M_v \in R^{N*|\epsilon_q|}$.

Top-K Similarity Hypergraph (SH):

To address the challenge of limited connections in item-wise multi-behavior hypergraphs, the SH utilizes the contextual similarity score to augment the intricate relationships between items. Specifically, we construct a behavior-aware similarity hypergraph by identifying the K most similar items and utilizing the hidden relationships of these similar items to infer the item with limited multi-behavior interactions. This approach allows the previously sparse adjacency matrix to be filled with inferred similarity values, potentially enhancing the item latent representation.

We build the contextual similarity metrics by:

$$C^K = \underset{\{v_1, v_2, v_3, \cdots, v_k \in V\}}{\text{argmax}} x^T X \quad (6\text{-}3)$$

where x^T is the item context embedding obtained by the Eq.(6-1), and X is the corresponding embedding of all items in the sequence. To build a top-K behavior-aware similarity hypergraph G_s we define the similarity connection matrix M_s, where each element $m_s \in M_s$ is defined as:

$$m_s(v_n, \epsilon_s) = \begin{cases} \beta_{n,s}, v_s \in C^K \\ 0, otherwise \end{cases} (m_s \in M_s) \quad (6\text{-}4)$$

where for any item v_n, ϵ_s is the hyperedge for v_s, one of the top-K similarity list C^K obtained by Eq.(6-3), and $\beta_{n,s}$ denote the similarity score between item v_n and v_s. By incorporating contextual similarities among correlated items, the similarity hypergraph enhances the connections within the multi-behavior hypergraph. This is accomplished by assigning top-K weights to the hyperedges of each node, leveraging item contextual similarity.

Hypergraph Convolution:

To obtain the overall *Augmented Hypergraph* G, we combine the item-wise multi-behavior dependency connection matrix M_v from *Multi-Behavior Hypergraph*, and the behavior-aware similarity connection matrix M_s from *TopK Similarity Hypergraph* as $\mathsf{M} = M_v \| M_s$, and M denotes the overall augmented connection matrix. To enhance item latent representations, we adopt hypergraph convolutional layers to capture item dependencies over time.

Let D_v and D_e be the diagonal normalization matrices based on vertex and edge, respectively, and l_g is the number of convolutional layers. The information-passing process along with the augmented connection matrix M is defined as :

$$H^{(l_g+1)} = D_v^{-1} \cdot M \cdot D_e^{-1} \cdot M^T \cdot H^{(l_g)} \tag{6-5}$$

where $H^{(l_g+1)}$ is the updated latent item representation encoded from the previous layer $H^{(l_g)}$ of hypergraph convolution, and the initial input features $H^0 = X$, and X is the sequence embeddings where each item representation is obtained from the Eq.(6-1).

6.2.2 Contrastively Regularized Multi-Scale Transformer (MST)

Multi-Scale (MS):

We utilize a multi-scale transformer to enhance sequence representation learning following the *SG* structure. The interaction features are randomly subsampled into various sub-temporal representations $\left(H^i, H^n, \cdots, H^N\right)$, with H^N denoting the output from Eq.(6-5) and others as subsampled results. The indices i and n change dynamically during training to provide adaptive temporal scales.

Next, we employ *Transformation Layers* to capture information across these granularities, enabling the model to grasp both fine-grained details and broader patterns. Each scale shares the same transformer layers, promoting parameter efficiency and reducing model complexity while leveraging learned representations effectively across different scales.

Transformer (Trm):

Analogous to the classic BERT4Rec architecture (Sun et al., 2019), colored as

green in Figure 6-2, the multi-head attention mechanism is adopted to capture complex sequence relationships by focusing on different aspects of the input through multiple attention heads. For each transformer layer l, the multi-head attention is computed by projecting the input H^n into h subspaces, with each head processing the query, key, and value matrices $W_{Q_i}^l, W_{K_i}^l, W_{V_i}^l$ in parallel. The outputs are then concatenated and linearly projected to the final result using W_O^l. Scaled dot-product attention is employed, where the queries and keys are normalized by $\sqrt{d/h}$ to stabilize gradients. Following attention, feed-forward layers apply non-linear transformations using GELU activation, with layer normalization and residual connections. The parameters are unique to each layer, ensuring flexible representation learning across different transformer layers.

Multi-Scale Fusion:

To integrate the dynamic patterns across multiple sequential scales, we introduce an aggregation process, through a fusion layer, as follows:

$$\hat{O} = W_s\left(\bar{H}^i \| \bar{H}^n \| \cdots \| \bar{H}^N\right) + b_s \tag{6-6}$$

where the W_s and b_s represent the learnable weights and bias for fusion layer projection. For simplicity, $\bar{H}^i$, $\bar{H}^n$ represent the hidden representation obtained from transformer layers from various scales, where $i, n \in N$. $\bar{H}^N$ is representation obtained from the entire sequence.

Output Layer:

After propagating through previous layers, we get the output $\hat{O}$ generated for all elements in the input sequence. In the context of multi-behavior sequential recommendation, we mask all items in the sequence with the target purchase behavior type with a special token. For masked items, we generate their embedding through the average pooling strategy of contextual neighbors surrounding the masked position. Assuming we mask an item $v_m^{u_i}$ at the time step n, simplified as v_m, we predict the masked item through the two-layer feed-forward network, incorporating the GELU activation function in between, to produce an output distribution over the target items:

$$p_{(v)} = softmax\big(GELU\big(W_p \cdot \hat{O} + b_p\big)E_V^T + b_o\big) \tag{6-7}$$

where W_p is a learnable weight matrix while b_p and b_o are bias, $E_V \in R^{|V|*d}$ is the item embedding for all items V.

Contrastively Regularization and Augmentation:

To enhance the robustness and generalization capability of Multi-Scale Transformer, we employ an augmentation strategy on the enhanced sequential features. Specifically, we randomly apply *Augmentation* operations such as *reordering*, *insertion,* and *replacement* along the sequence features. This augmentation process aims to generate diverse representations of the sequences, facilitating the model to learn invariant features across different transformations. To distinguish whether two representations originate from the same historical sequence, we employ a contrastive loss function. This loss minimizes the difference between differently augmented views of the same sequence while maximizing the difference between augmented sequences from different users. We apply one random augmentation operations to each user's sequence to obtain an augmented sequence representation $\bar{H}^{Aug}$, and also randomly chose one of the subsampled represenation from $\left[\bar{H}^i, \bar{H}^n, \cdots, \bar{H}^N\right]$ to construct as the positive pair $(\bar{H}^n, \bar{H}^{Aug})$, while the remaining augmented sequences in the minibatch serve as negative samples $\bar{H}^-$. The similarity between representations is measured using the dot product. The contrastive loss function is defined as:

$$\mathcal{L}_C = -\log \frac{e^{sim(\bar{H}^n, \bar{H}^{Aug})}}{e^{sim(\bar{H}^n, \bar{H}^{Aug})} + \sum e^{sim(\bar{H}^n, \bar{H}^-)}} \tag{6-8}$$

6.2.3 Optimization

The total optimization function is given by:

$$\mathcal{L} = \mathcal{L}_P + \mathcal{L}_C + \frac{\lambda}{2}\|\Theta\|^2 \tag{6-9}$$

where $\mathcal{L}_C$ is the contrastive learning loss, Θ denotes all the trainable parameters in the model, and the recommendation loss $\mathcal{L}_P$ is as below:

$$\mathcal{L}_P = \frac{1}{|S_u^m|} \sum_{v_m \in S_u^m} -\log P\left(v_m = v_m^* \middle| S_u^{'}\right) \quad (6\text{-}10)$$

where $S_u^{'}$ is the masked behaviour-aware user interaction sequence, S_u^m ia all the m masked items in the sequence, and v_m^* is the ground truth of the masked item v_m.

6.3 Experiments

The experiments on three public benchmark datasets are conducted to verify the effectiveness of the proposed SG-MST, aiming to answer the following research questions:

Q1: How does our SG-MST perform as compared to various advanced recommendation methods?

Q2: How would the key modules and the specific settings in the method affect the overall performance?

Q3: How does multi-behavior sequential recommendation benefit from our methods?

6.3.1 Experiments Setup

Evaluation Protocol:

In our experimental setup, we employ the leave-one-out strategy for performance evaluation followed by (Y. Yang et al., 2022). Specifically, for each user, the last purchase in the temporally-ordered sequence is considered as the test sample, while the preceding purchases are used as validation samples. In addition, each positive sample is compared with 100 negative ones organized based on item popularity for evaluation. We use a combination of several evaluation metrics to measure the performance of our model: Hit Ratio (HR@N), Normalized Discounted Cumulative Gain (NDCG@N), and Mean Reciprocal Rank (MRR). For comparative purposes, we report the results of HR@N and NDCG@N with two cut-off values @N = 5 and 10. For all these metrics, the larger the value, the better the recommendation performance.

Implementation Details:

The Adam optimizer is adopted for the training process with the learning rate set as a learning rate of 0.001 and a weight decay of 0.01. The global batch size is selected from [256, 512, 1024] with NVIDIA GeForce RTX 4090 GPUs. The hidden sizes for latent representation were selected from [64, 256], and we searched the number of hypergraph propagation layers from [1, 2, 3]. We set the maximum training epoch as 200 and adopted the early stop training strategy with patience set as 10. The subsequence numbers were searched from [2, 3, 4]. The number of multi-head channels was set as 2 for the attention modules and we used the same maximum sequence length as 200 for all of the three datasets for padding operation.

6.3.2 Overall Performance Comparison (Q1)

We compare SG-MST with the following competitive baseline methods for sequential recommendation, utilizing various techniques:

- **GRU4Rec** (Hidasi, 2015) utilizes GRUs to capture the sequential behavior of users' interactions with items. It excels in session-based recommendation tasks by effectively modeling the temporal dynamics of user behavior.
- **BERT4Rec** (Sun et al., 2019) applies the BERT framework to capture the bidirectional interactions between items in users' historical sequences. It enhances the recommendation performance by understanding the context from both directions of a sequence.
- **SASRec** (Kang & McAuley, 2018) leverages the Self-Attention mechanism from the Transformer model to capture the long-range dependencies within users' interaction sequences.
- **SR-GNN** (Hidasi, 2015) employs Graph Neural Networks to model the entire session as a graph, capturing high-order connectivity patterns among items.
- **GCSAN (Xu et al., 2019)** combines graph convolutional networks with a self-

attention mechanism. It effectively captures both the global graph structure and local neighborhood interactions, providing more accurate and personalized recommendations.

- **BERT4Rec+B** (Y. Yang et al., 2022) is an augmented version of the BERT4Rec to effectively model the multi-behavior context by incorporating representations of behavior types into the input embeddings.
- **HyperRec** (Wang et al., 2020) excels in capturing complex user-item interactions and providing personalized recommendations by dynamically adjusting its parameters based on the input.
- **MB-GMN** (Xia et al., 2021) integrates multi-behavior pattern into a meta-learning paradigm, allowing it to uncover type-dependent behavior representations and capture interaction diversity and behavior heterogeneity.
- **MBHT** (Y. Yang et al., 2022) models user preferences by combining multiple behavioral types and also lineally combines Transformer and hypergraph neural networks to capture the dynamic heterogeneous relationships between users and items.
- **FEA-MB** (Du et al., 2023) incorporates behavior embedding into FEA, which integrates time-domain self-attention with frequency-domain attention to capture both low and high-frequency information and periodic characteristics to enhance prediction performance.

Table 6-1 Performance comparison on Retailrocket. Bolded numbers indicate the best results, italicized and underlined indicate the second best results.

Method	Retailrocket			
	HR@5	HR@10	NDCG@5	NDCG@10
GRU4Rec	0.640	0.708	0.575	0.597
BERT4Rec	0.707	0.763	0.665	0.663
SASRec	0.669	0.689	0.644	0.650
SR-GNN	0.848	0.891	0.780	0.793
HyperRec	0.860	0.833	0.705	0.820
GCSAN	0.872	0.890	0.846	0.81

BERT4Rec+B	0.908	0.916	0.898	0.901
MBHT	0.910	0.915	0.900	0.902
FEA-MB	*0.933*	*0.943*	*0.918*	*0.921*
SG-MST	**0.953**	**0.966**	**0.936**	**0.940**

Table 6-2 Performance comparison on Taobao.

Method	Retailrocket			
	HR@5	HR@10	NDCG@5	NDCG@10
GRU4Rec	0.147	0.209	0.105	0.125
BERT4Rec	0.195	0.255	0.154	0.215
SASRec	0.150	0.206	0.110	0.128
SR-GNN	0.102	0.153	0.071	0.087
HyperRec	0.145	0.224	0.130	0.133
GCSAN	0.217	0.188	0.160	0.188
BERT4Rec+B	0.246	0.313	0.194	0.215
MB-GMN	0.269	0.390	0.189	0.225
MBHT	0.293	0.370	0.231	0.256
FEA-MB	*0.323*	*0.405*	*0.254*	*0.282*
SG-MST	**0.336**	**0.417**	**0.267**	**0.286**

Table 6-3 Performance comparison on IJCAI.

Method	Retailrocket			
	HR@5	HR@10	NDCG@5	NDCG@10
GRU4Rec	0.141	0.200	0.100	0.119
BERT4Rec	0.297	0.340	0.196	0.223
SASRec	0.146	0.191	0.110	0.124
SR-GNN	0.072	0.118	0.048	0.062
HyperRec	0.140	0.236	0.109	0.144
GCSAN	0.119	0.175	0.086	0.104
BERT4Rec+B	0.257	0.342	0.189	0.216
MB-GMN	0.272	0.405	0.184	0.232
MBHT	*0.323*	*0.411*	*0.249*	*0.278*
SG-MST	**0.344**	**0.438**	**0.265**	**0.295**

The overall performance of all methods compared on three datasets are listed above, from which we can observe:

1) The experimental results show that SG-MST outperforms all other methods across all datasets. This highlights the validity of the proposed model as a robust solution

to multi-behavior sequential recommendations, and demonstrates the introduced enhancements. Comparisons between different datasets also show that prediction of user preferences in more complex interaction environments is more difficult.

2) The discrepancy in performance across datasets can be attributed to the calculation basis of metrics. With a smaller dataset like Retailrocket, these metrics may reflect a higher performance due to the limited interaction sequences, resulting in more straightforward predictions. Conversely, larger datasets like Taobao and IJCAI involve more diverse and complex user behaviors, making it challenging to achieve high scores on these metrics.
3) Behavioral sequence modeling helps to improve the performance of a single behavioral sequence recommendation model. MBHT, MB-GMN and FEA-MB outperform other sequential recommenders with single behavior modeling. Additionally, BERT4Rec+B, which incorporates representations of behavior types into the input embeddings, generally improves the recommendation performance compared to the original BERT4Rec and other methods in most cases. This experiment result highlights the significance of incorporating multi-behavior context into user preferences learning.
4) Graph-based sequential recommendation models, like SR-GNN and GCSAN, leverage graph convolutions to effectively capture global item dependencies and long-term user interests. However, in datasets like Taobao and IJCAI in our case, the complex multi-behavior interactions may lead to challenges in accurately constructing meaningful graphs.

6.3.3 Ablation Study (Q2)

To evaluate the contribution of the key modules in our proposed SG-MST, we conduct the ablation study on all datasets. The results are shown in Table 6-4, where (a) **-w/o SH*** represents *removing* the TopK Similarity Hypergraph from SG-MST, and (b) **-w/o SG*** means *removing* the whole Similarity Augmented Multi-Behavior

Hypergraph module, and only using the original sequence embedding as the input for the multi-scale transformer module. In contrast, (c) **-w/o MS*** means *removing* the subsampled multi-scale sequence but directly using H^N as the input for transformer layers, and (d) **-w/o MST*** means *removing* the whole Multi-Scale Transformer module and only using SG to get the updated target representation prediction.

Table 6-4 Ablation study with key modules.

Methods	Retailrocket (@5)		Taobao (@5)		IJCAI (@5)	
	HR	NDCG	HR	NDCG	HR	NDCG
-w/o SH*	0.938	0.909	0.322	0.257	0.368	0.281
-w/o SG*	0.909	0.880	0.303	0.239	0.324	0.245
-w/o MS*	0.925	0.899	0.319	0.250	0.321	0.244
-w/o MST*	0.862	0.837	0.241	0.194	0.265	0.204
SG-MST	**0.953**	**0.936**	**0.336**	**0.267**	**0.344**	**0.265**

It can be observed from the ablation comparison that removing each submodule results in a loss of model performance, and the full SG-MST model outperforms the ablated models in all datasets. From the experiments of -w/o SH and removing the whole augmented hypergraph module (-w/o SG), the prediction accuracy drops noticeably, which indicates the model would struggle to capture informative interaction collections for user preference learning with limited transactions. This observation demonstrates a more informative correlation modeling helps alleviate the sparsity challenges and increase the prediction accuracy.

Besides, -w/o MS* experiments leads to diminished performance, emphasizing its effectiveness in capturing diverse temporal patterns. From -w/o MST* experiments, we can observe that both short-term and long-term behavior dependencies are significant in exploring users' personalized preferences in the sequential recommendation.

The findings indicate that both the Similarity Augmented Multi-Behavior Hypergraph module and multi-scale transformer are proven effective and work well on three datasets, and their combination in the full SG-MST model results in further improved performance.

6.3.4 *Effect of Key Hyperparameters. (Q2)*

Two hyper-parameters in two modules are further investigated, specifically the top-k value K from Eq.(6-3) in the Similarity Augmented Multi-Behavior Hypergraph (SG), and the number of layers L in the Multi-Scale Transformer layers (MST). These parameters are crucial as the top-k value directly affects the connections in the hypergraph, and Ldetermines the depth of feature extraction and abstraction in the transformer, influencing the model's capacity to learn complex patterns.

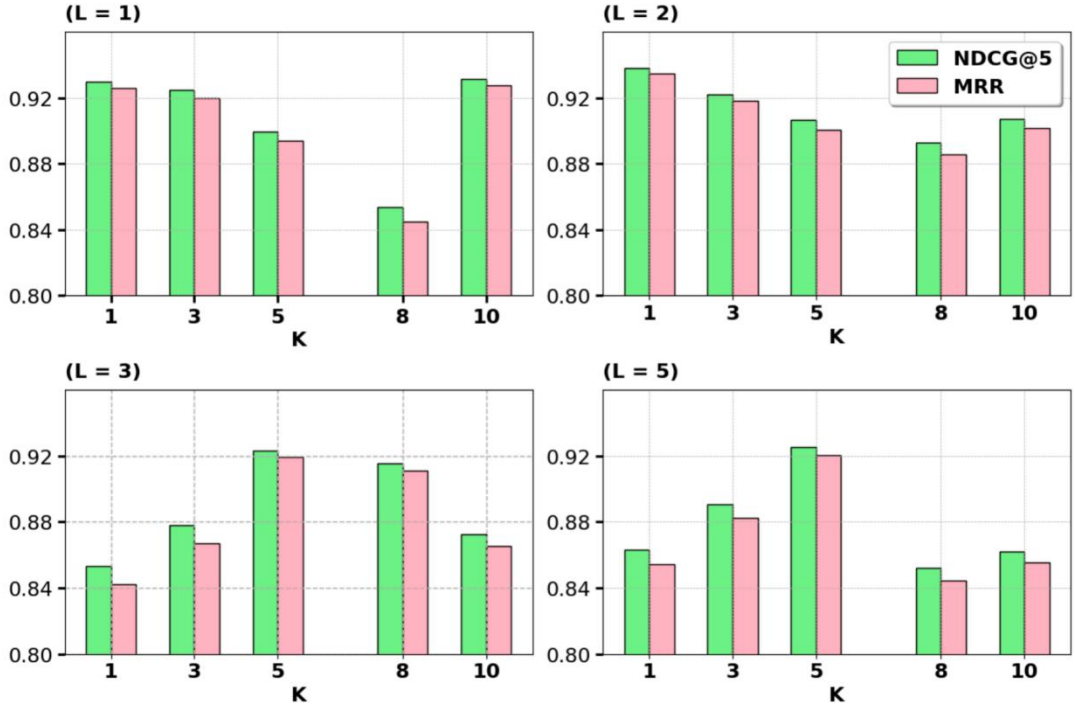

Figure 6-3 Performance comparison investigation on hyper-parameters in SG (K) and MST (L) modules.

Based on the performance results in Figure 6-3 and the training loss comparison in Figure 6-4, we can conclude that a large K may lead to an excessive number of edges in the hypergraph, and further lead to model overfitting. On the contrary, a small K may lead to an insufficient connection in the hypergraph, which may not adequately capture the correlations in the data. For the number of layers L in the Transformer model, fewer layers may result in insufficient model representation to capture complex features and patterns, thus limiting the performance of the model. Whereas a large L number increases the complexity and may lead to overfitting. An optimal value for these

parameters should balance model complexity and expressiveness.

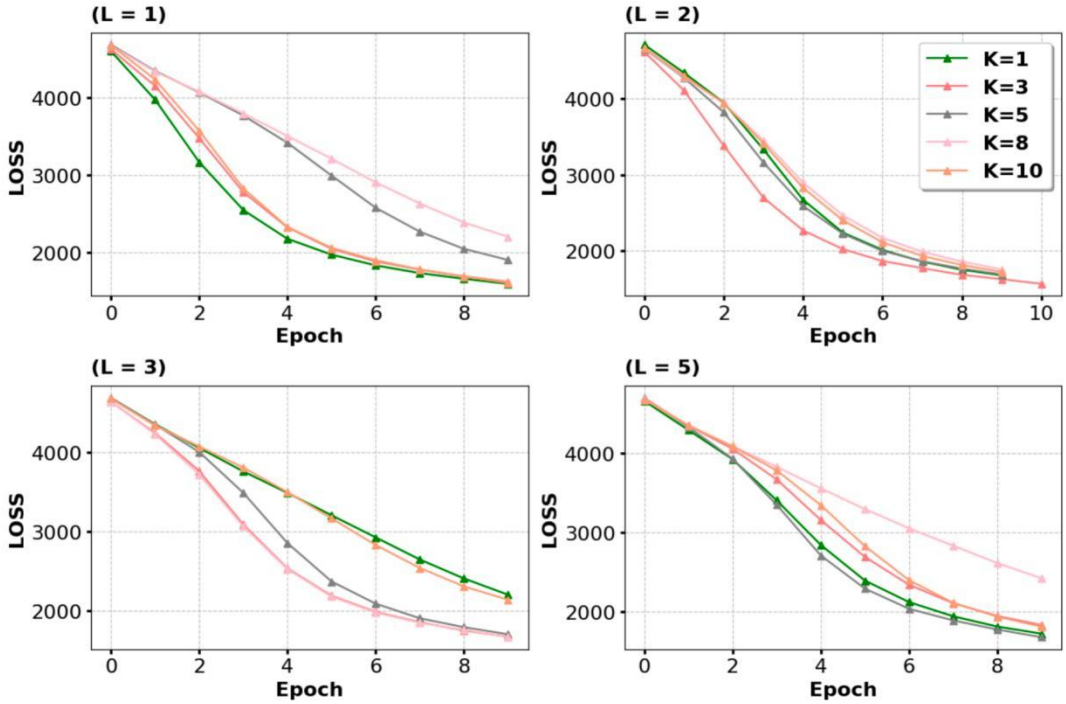

Figure 6-4 Loss comparison investigation on hyper-parameters in SG (K) and MST (L) modules.

6.3.5 Model Benefit Study (Q3)

6.3.5.1 On Various Interaction Richness

In this subsection, we analyze how our proposed SG-MST can benefit the multi-behavior sequential recommendation tasks in extremely limited interaction scenarios and interaction scenarios with longer time. Specifically, we filtered a subset focusing on sequences with a length of less than 10 from the Retailrocket dataset, and another subset filtered from Taobao with interaction longer than 70. We conducted extended experiments and analyzed the performance comparison in these two scenarios. The comparison with sequences of length less than 10 serves as a surrogate for situations where limited historical data is available. While the interaction sequences longer than 70, provide rich and sophisticated information for us to generate long-term preference predictions.

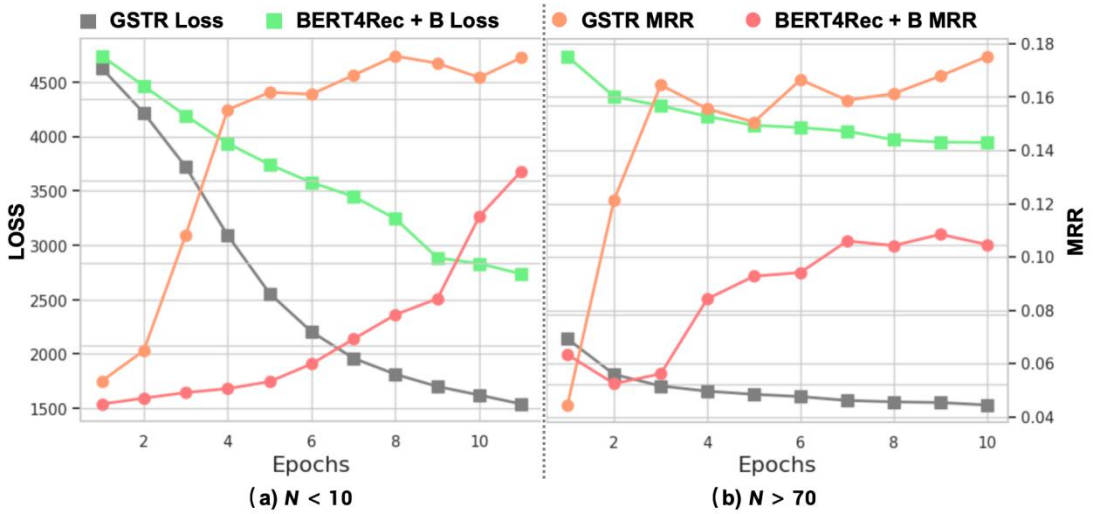

Figure 6-5 Comparison of training loss and evaluation curves in filtered sparse and long sequences.

In Figure 6-5, we compare the convergence characteristics of our SG-MST with that of BERT4Rec+B in the filtered sparse dataset. Along the model training process, SG-MST achieves a faster convergence rate. For example, SG-MST obtains its best performance at epoch 8, while BERT4Rec is still in the very early stages of training. For the final evaluation comparison shown in Table 6-5, the results show that our proposed model outperforms the baseline by a large margin.

The results in these two settings indicate the model's ability to extract meaningful information and make accurate recommendations even in extremely limited conditions. On the other hand, from the results in Table 6-5 where the sequence is longer than 70, our SG-MST still performs much better, which indicates the effectiveness of our model in diverse interactive scenarios.

Table 6-5 Performance Comparison in filtered scenarios.

N<10	HR@5	HR@10	NDCG@5	NDCG@10
BERT4Rec+B	0.7070	0.7625	0.6446	0.6625
SG-MST	**0.9359**	**0.9522**	**0.9085**	**0.9136**
N>70	HR@5	HR@10	NDCG@5	NDCG@10
BERT4Rec+B	0.1309	0.1696	0.0988	0.1114
SG-MST	**0.2098**	**0.2759**	**0.1624**	**0.1835**

6.3.5.2 On Cold Start and Noisy Data

Table 6-6 Performance comparison in the cold start and noisy data scenarios in Retailrocket dataset.

Cold Start	HR@5	HR@10	NDCG@5	NDCG@10
BERT4Rec+B	0.8124	0.8445	0.7443	0.7547
SG-MST	**0.9281**	**0.9291**	**0.9256**	**0.9243**
Noisy Data	HR@5	HR@10	NDCG@5	NDCG@10
BERT4Rec+B	0.6505	0.7100	0.5816	0.6009
SG-MST	**0.9250**	**0.9380**	**0.8973**	**0.9016**

In this subsection, we evaluate the performance of three models under cold start and noisy data scenarios. Noisy data is simulated by randomly inserting items into the user interaction sequences and assigning these items random behaviors, excluding purchase actions, with a noise insertion rate of 10 percent.

Experimental results in Table 6-6 show that our model performs well in both cold-start and noisy data scenarios, highlighting the key advantages of our model. In the cold-start scenario, it is often difficult for the system to interact with new users due to the lack of interaction history, whereas our model can effectively handle sparse data and shows strong generalization. Under noisy data conditions, random items and behaviors are introduced in user interaction sequences, while our model maintains high accuracy and relevance, showing its robustness. These observations demonstrate the effectiveness of SG-MST in handling sparse and corrupted interaction sequences and ensure reliable recommendations.

6.3.6 Extended Study

6.3.6.1 Multi-Behavior Relationships Visualization

In order to show the model interpretation for multi-behavior correlations, we perform additional cross-behavior relationships visualization using a behavior type similarity heatmap projection, and all the type weights are learned from our SG-MST.

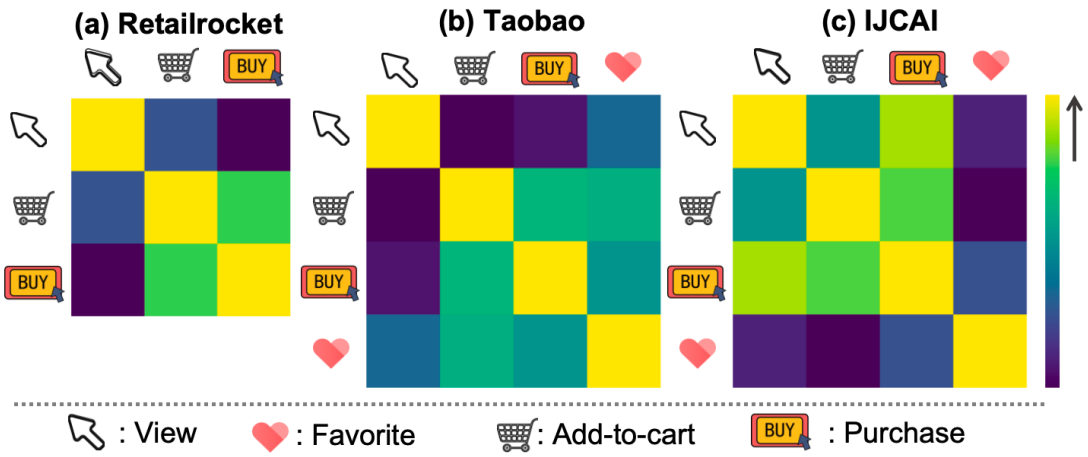

Figure 6-6 Learnable Behavior Type Embedding Similarity Heatmap obtained from SG-MST in three datasets.

Figure 6-6 illustrates the relationships within the embedding space between the different behavior types, with the horizontal and vertical axes representing different behaviors. Each cell in the heatmap indicates the cosine similarity value between the corresponding behavior types. Higher similarity values imply greater resemblance between the learned weight of respective behavior types, whereas lower one indicates larger disparities.

6.3.6.2 Case Study

In this subsection, we conduct further case studies to show the user-specific multi-behavior sequential interaction correlations learned from our model. In Figure 6-7, the left side of the heatmap shows the item ID information in the real interaction sequence and the corresponding behavioral patterns.

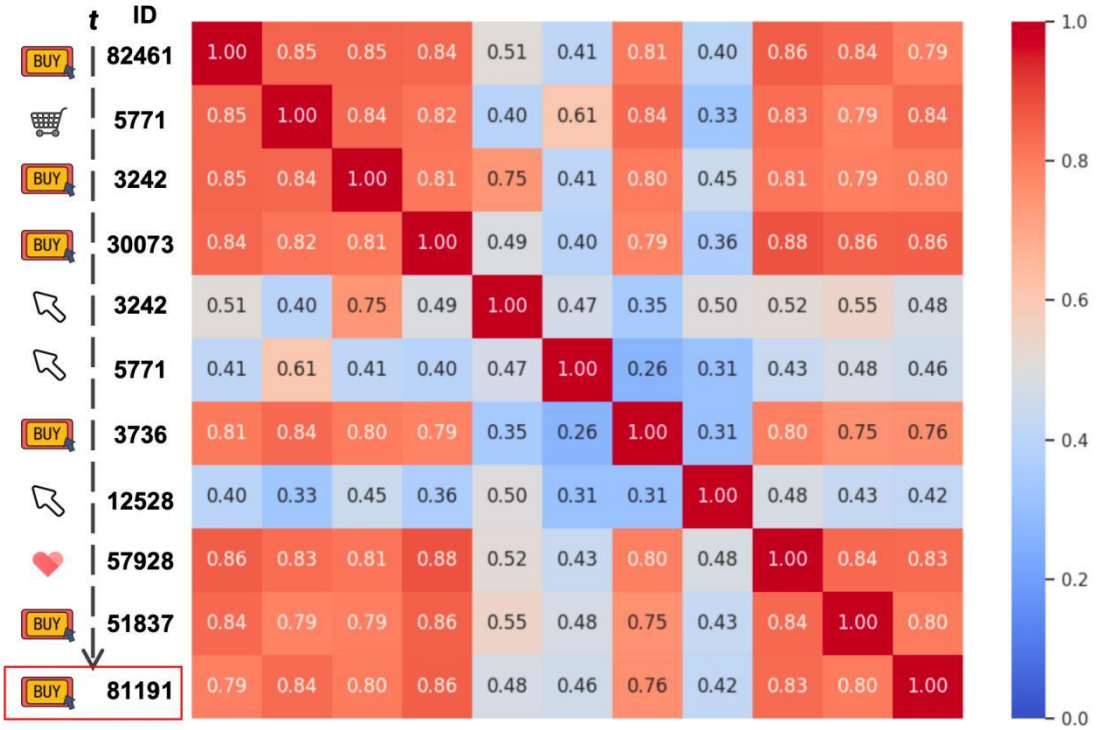

Figure 6-7 Model interpretation with the case studies on the learned multi-behavior interaction correlations.

This heatmap illustrates the dependencies between different types of interaction behaviors in a sequence, particularly the relationship between the predicted last item and others. Each value in the heatmap represents the score of dependency between corresponding types of interactions. Higher values indicate stronger dependencies, while lower values show weaker relationships. By leveraging the purchasing patterns of different behavior patterns, recommendation models can make more accurate and personalized next-item predictions.

6.3.7 Application Example

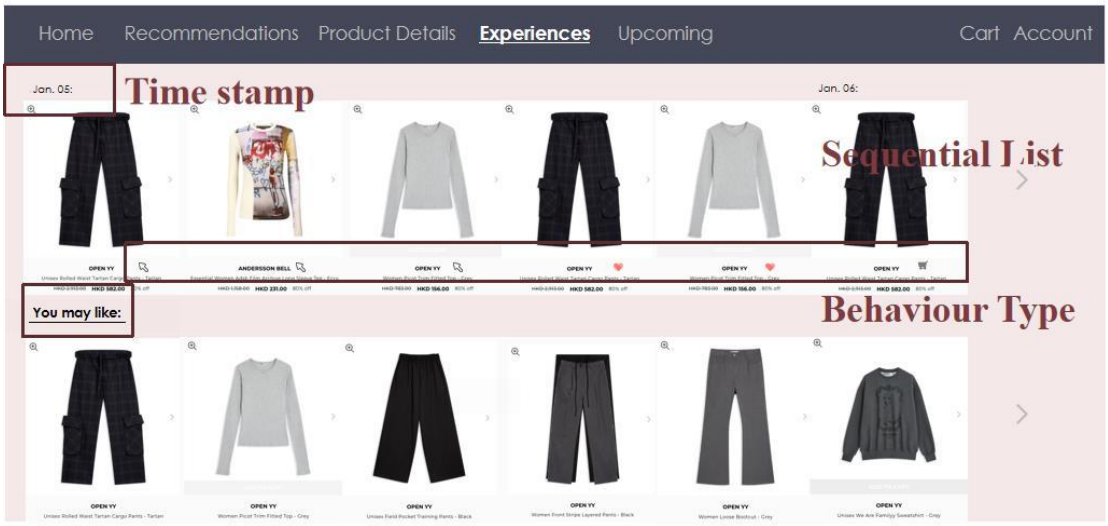

Figure 6-8 Personalized fashion sequential recommendation example.

Figure 6-8 shows the application example of SG-MST for personalized fashion sequential recommendation. We leverage multi-behavior sequential data to capture the dynamic preferences of users. Specifically, for each user, we record the timestamp and behavior type (e.g., clicks, add-to-cart, purchases) associated with the items they interact with. This rich behavioral data is then processed by the SG-MST model to generate a personalized recommendation list.

6.4 Summary

In this Chapter, we present a multi-behavior recommendation framework for e-commerce product recommendations that integrates the strengths of transformers and graph neural networks and combines three types of embeddings, called SG-MST. The innovative fusion of a Similarity Augmented Multi-Behavior Hypergraph for enhancing sequential representation learning, and a Contractively Regularized Multi-Scale Transformer for both long-term and short-term preference learning have led to an effective and efficient recommendation system. Extensive experiments across diverse datasets have shown the superior performance of SG-MST compared to other methods. Additionally, our ablation studies have demonstrated the individual contributions of the proposed two new modules, and the model benefit studies have also confirmed the

significance of SG-MST in achieving improved recommendation accuracy.

However, despite the effectiveness, our method can be improved further in several aspects. For example, in the current version, technical designs in different parts can be further advanced with more expressive designs. Further enhancements to the multi-scale fusion strategy, the sampling strategy and the sequence augmentation strategy are needed in future applications.

Chapter 7. Conclusions and Recommendations for Future Work

7.1 Conclusions

The fashion industry, with its dynamic nature and rapidly evolving trends, presents unique challenges and opportunities for recommendation systems. This thesis, titled "Data-driven Recommendations for Fashion: An Investigation of Personalization with Sparse Data," explores the complexities inherent in fashion recommendation systems and proposes innovative solutions to enhance their effectiveness. Through a comprehensive examination of the challenges posed by data sparsity and the need for personalization, this research has contributed to the advancement of recommendation technologies in the fashion domain.

One of the primary challenges identified in this research is the issue of data sparsity, which is particularly pronounced in the fashion industry. The transient nature of fashion items, coupled with the limited interaction history between users and items, exacerbates the sparsity problem. This thesis has demonstrated that traditional recommendation methods often struggle to provide accurate recommendations in such environments. By leveraging multimodal data, including visual and textual information, this research has shown that it is possible to mitigate the effects of data sparsity and improve recommendation quality. The exploration of personalized clothing matching and product sequential recommendation tasks has highlighted the importance of understanding both long-term and short-term user preferences. This thesis has proposed methods to effectively capture and integrate these dual aspects of user preferences, thereby enhancing the personalization of recommendations.

This research has made several significant contributions to the field of fashion recommendation systems. Firstly, it has advanced the understanding of how multimodal data can be utilized to address data sparsity. By incorporating images and text

descriptions into the recommendation process, the proposed methods have demonstrated improved robustness in data-scarce contexts. This approach not only enhances the accuracy of recommendations but also provides a richer understanding of user preferences and item attributes. Secondly, the introduction of Indirect Personal Compatibility represents a novel approach to balancing personalization and product compatibility. This innovation addresses a critical gap in existing recommendation systems, which often fail to capture the complex relationships between user preferences and item compatibility. Furthermore, by exploring implicit connections between user personalization and product compatibility with functional cross views contrastive learning, this thesis has demonstrated the potential applications and improved recommendation accuracy.

The findings of this research have significant implications for the fashion industry. The proposed methods offer practical solutions for improving the effectiveness of recommendation systems, particularly in environments characterized by data sparsity. By leveraging multimodal data and advanced machine learning techniques, retailers can enhance the personalization of their recommendations, leading to increased customer satisfaction and engagement. The integration of long-term and short-term user preferences into recommendation models also provides valuable insights for practitioners seeking to deliver contextually relevant suggestions. As the fashion industry continues to evolve, the insights and findings of this thesis will serve as a valuable foundation for future research and practice, driving innovation and enhancing the effectiveness of recommendation systems in the fashion domain.

Importantly to note, the three approaches for the fashion complementary recommendation are designed as complementary and modular components rather than competing solutions. Each method tackles the issue of data sparsity from a unique perspective: consistency, coupled indirect personal compatibility, and similar users or products. These models can be flexibly replaced or combined depending on specific application requirements. This design philosophy emphasizes a collaborative approach

to addressing the data sparsity problem by integrating different technologies. These technologies work together to alleviate data sparsity constraints while ensuring interpretability and scalability. In terms of the three innovative methods developed for the first complementary recommendation task, the overall AUC results comparison shows that the third method, APCL, achieves the highest performance. This demonstrates the effectiveness of exploring deeper and broader connections within the data. However, it does not necessarily mean the APCL would be perfect in other various fashion recommendation scenarios, and the three models are not designed to compete with each other; instead, they serve as alternatives that can complement each other. They also have the potential to be integrated into future designs to further enhance recommendation performance.

7.2 Recommendations for Future Work

While this thesis has addressed several key challenges in fashion recommendation systems, it also highlights areas for future research. One potential direction is the exploration of augmented reality and virtual try-on technologies, which offer exciting opportunities for enhancing the user experience. By integrating these technologies with recommendation systems, researchers can further personalize the shopping experience and provide users with more immersive and interactive options.

In addition, the exploration of cross-domain recommendation systems also presents a promising avenue for future research. By leveraging data from multiple domains, such as fashion, lifestyle, and entertainment, researchers can develop more comprehensive models that capture a broader range of user preferences.

Another most important exploration of future work would focus on LLM-empowered personalized fashion recommendation. Fashion recommendation systems must adeptly model the intricate relationships between users and items to effectively meet users' preferences. The fashion domain is profoundly influenced by contextual

factors, including cultural trends, seasonal shifts, and individual styles, which contribute to the complexity of fashion recommendation algorithms. Existing methods often fall short in capturing these nuances, resulting in suboptimal recommendation performance. This limitation is largely due to their inability to process and understand the diverse and multimedia-rich content inherent to the fashion industry.

Furthermore, fashion recommendation systems must adapt to the ever-evolving nature of the industry, necessitating continuous learning and adaptation mechanisms that can swiftly respond to new trends and user feedback. Consequently, there is an increasing demand for systems that can harness the power of large-scale data analytics and machine learning to deliver personalized and context-aware fashion recommendations that are constantly evolving.

Figure 7-1 Example of LLM-empowered information enhancement.

There is a growing interest in leveraging the capabilities of Large Language Models (LLMs). Recent advancements in LLMs have significantly impacted multiple fields, including natural language processing, healthcare, conversational agents, and content generation. Previous studies have demonstrated the potential of LLMs in enhancing recommendation systems. Thanks to advancements in large language models, collecting textual information has become much easier compared to the manual efforts required in the past. Figure 7-1 is an example of LLM-empowered information enhancement. The generated textual information is now more aligned with visual data, enabling a richer understanding of multimodal content. Moreover, He et al. explored the use of LLMs as zero-shot conversational recommenders, showing that LLMs can

outperform traditional models even without extensive fine-tuning. Similarly, Ren et al. highlighted how LLMs can be integrated with sequential recommendation models to improve the accuracy and relevance of recommendations by incorporating real-world knowledge. Numerous applications and academic studies have demonstrated that LLMs possess the ability to understand complex semantic meanings within texts. This capability enables them to process unstructured data sources such as fashion blogs, user reviews, and social media posts, which can be harnessed to analyze fashion trends and provide additional domain knowledge. Such insights significantly enhance the contextual relevance of fashion recommendation systems.

References

Abdulla, G. M., & Borar, S. (2017). Size recommendation system for fashion e-commerce. KDD workshop on machine learning meets fashion,

Abluton, A. (2022). Visual Recommendation and Visual Search for Fashion E-Commerce. International Conference on Similarity Search and Applications,

Ai, Q., Azizi, V., Chen, X., & Zhang, Y. (2018). Learning heterogeneous knowledge base embeddings for explainable recommendation. *Algorithms*, *11*(9), 137.

Barkan, O., Caciularu, A., Katz, O., & Koenigstein, N. (2020). Attentive item2vec: Neural attentive user representations. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),

Barkan, O., & Koenigstein, N. (2016). Item2vec: neural item embedding for collaborative filtering. 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP),

Bin, C. Z., Gu, T. L., Sun, Y. P., & Chang, L. (2019). A personalized POI route recommendation system based on heterogeneous tourism data and sequential pattern mining [Article]. *Multimedia Tools and Applications*, *78*(24), 35135-35156. https://doi.org/10.1007/s11042-019-08096-w

Bulović, V., & Čović, Z. (2020). The impact of digital transformation on sustainability in fashion retail. 2020 IEEE 18th International Symposium on Intelligent Systems and Informatics (SISY),

Cao, D., Nie, L., He, X., Wei, X., Zhu, S., & Chua, T.-S. (2017). Embedding factorization models for jointly recommending items and user generated lists. Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval,

Chakraborty, S., Hoque, M. S., Rahman Jeem, N., Biswas, M. C., Bardhan, D., & Lobaton, E. (2021). Fashion recommendation systems, models and methods: A review. Informatics,

Chen, B., & Deng, W. (2019). Deep embedding learning with adaptive large margin N-pair loss for image retrieval and clustering. *Pattern Recognition*, *93*, 353-364.

Chen, C., Ma, W., Zhang, M., Wang, Z., He, X., Wang, C., Liu, Y., & Ma, S. (2021). Graph heterogeneous multi-relational recommendation. Proceedings of the AAAI conference on artificial intelligence,

Chen, H.-J., Shuai, H.-H., & Cheng, W.-H. (2023). A Survey of Artificial Intelligence in Fashion. *IEEE Signal*

Processing Magazine, *40*(3), 64-73.

Chen, J. Z., Zheng, L., & Chen, S. T. (2023). User view dynamic graph-driven sequential recommendation [Article]. *Knowledge and Information Systems*, *65*(6), 2541-2569. https://doi.org/10.1007/s10115-023-01840-7

Chen, W., Huang, P., Xu, J., Guo, X., Guo, C., Sun, F., Li, C., Pfadler, A., Zhao, H., & Zhao, B. (2019). POG: personalized outfit generation for fashion recommendation at Alibaba iFashion. Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining,

Chen, X., Xu, H., Zhang, Y., Tang, J., Cao, Y., Qin, Z., & Zha, H. (2018). Sequential recommendation with user memory networks. Proceedings of the eleventh ACM international conference on web search and data mining,

Chuang, C.-Y., Robinson, J., Lin, Y.-C., Torralba, A., & Jegelka, S. (2020). Debiased contrastive learning. *Advances in neural information processing systems*, *33*, 8765-8775.

Church, K. W. (2017). Word2Vec. *Natural Language Engineering*, *23*(1), 155-162.

Cucurull, G., Taslakian, P., & Vazquez, D. (2019). Context-aware visual compatibility prediction. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,

Danielsson, P.-E. (1980). Euclidean distance mapping. *Computer Graphics and image processing*, *14*(3), 227-248.

Deldjoo, Y., Nazary, F., Ramisa, A., Mcauley, J., Pellegrini, G., Bellogin, A., & Di Noia, T. (2022). A review of modern fashion recommender systems. *arXiv preprint arXiv:2202.02757*.

Ding, Y., Lai, Z., Mok, P., & Chua, T.-S. (2023a). Computational technologies for fashion recommendation: A survey. *ACM Computing Surveys*, *56*(5), 1-45.

Ding, Y., Lai, Z., Mok, P., & Chua, T.-S. (2023b). Computational Technologies for Fashion Recommendation: A Survey. *arXiv preprint arXiv:2306.03395*.

Ding, Y., Ma, Y., Wong, W. K., & Chua, T.-S. (2021). Leveraging two types of global graph for sequential fashion recommendation. Proceedings of the 2021 International Conference on Multimedia Retrieval,

Ding, Y., Mok, P., Bin, Y., Yang, X., & Cheng, Z. (2023). Modeling Multi-Relational Connectivity for Personalized Fashion Matching. Proceedings of the 31st ACM International Conference on

Multimedia,

Ding, Y. J., Ma, Y. S., Wong, W. K., & Chua, T. S. (2022). Modeling Instant User Intent and Content-Level Transition for Sequential Fashion Recommendation [Article]. *IEEE Transactions on Multimedia*, *24*, 2687-2700. https://doi.org/10.1109/tmm.2021.3088281

Dong, X., Wu, J., Song, X., Dai, H., & Nie, L. (2020). Fashion compatibility modeling through a multi-modal try-on-guided scheme. Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval,

Du, X., Yuan, H., Zhao, P., Qu, J., Zhuang, F., Liu, G., Liu, Y., & Sheng, V. S. (2023). Frequency enhanced hybrid attention network for sequential recommendation. Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval,

Fan, Z., Xu, K., Dong, Z., Peng, H., Zhang, J., & Yu, P. S. (2023). Graph collaborative signals denoising and augmentation for recommendation. Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval,

Felfernig, A., Jeran, M., Ninaus, G., Reinfrank, F., Reiterer, S., & Stettinger, M. (2014). Basic approaches in recommendation systems. *Recommendation Systems in Software Engineering*, 15-37.

Feng, Z., Yu, Z., Yang, Y., Jing, Y., Jiang, J., & Song, M. (2018). Interpretable partitioned embedding for customized multi-item fashion outfit composition. Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval,

Gao, C., He, X., Gan, D., Chen, X., Feng, F., Li, Y., Chua, T.-S., & Jin, D. (2019). Neural multi-task recommendation from multi-behavior data. 2019 IEEE 35th international conference on data engineering (ICDE),

Gao, Y., Huang, Z.-W., Huang, Z.-Y., Huang, L., Kuang, Y., & Yang, X. (2023). Multi-scale broad collaborative filtering for personalized recommendation. *Knowledge-Based Systems*, 110853.

Golub, G. H., & Reinsch, C. (1971). Singular value decomposition and least squares solutions. In *Handbook for Automatic Computation: Volume II: Linear Algebra* (pp. 134-151). Springer.

Goti, A., Querejeta-Lomas, L., Almeida, A., de la Puerta, J. G., & López-de-Ipiña, D. (2023). Artificial Intelligence in Business-to-Customer Fashion Retail: A Literature Review [Review]. *Mathematics*, *11*(13), 32, Article 2943. https://doi.org/10.3390/math11132943

Guan, W. L., Jiao, F. K., Song, X. M., Wen, H. K., Yeh, C. H., Chang, X. J., & Acm. (2022, Jul 11-15). Personalized Fashion Compatibility Modeling via Metapath-guided Heterogeneous Graph Learning. [Proceedings of the 45th international acm sigir conference on research and development in information retrieval (sigir '22)]. 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Madrid, SPAIN.

Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). DeepFM: a factorization-machine based neural network for CTR prediction. *arXiv preprint arXiv:1703.04247*.

Guo, L., Hua, L., Jia, R., Zhao, B., Wang, X., & Cui, B. (2019). Buying or browsing?: Predicting real-time purchasing intent using attention-based deep network with multiple behavior. Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining,

Gutmann, M., & Hyvärinen, A. (2010). Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. Proceedings of the thirteenth international conference on artificial intelligence and statistics,

Hallac, I. R., Makinist, S., Ay, B., & Aydin, G. (2019). user2vec: Social media user representation based on distributed document embeddings. 2019 International Artificial Intelligence and Data Processing Symposium (IDAP),

Hao, X., Zhike, H., & Yichen, S. (2020). ClothNet: A Neural Network Based Recommender System. *Fuzzy Systems and Data Mining VI: Proceedings of FSDM*, *331*, 261.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition,

He, R., & McAuley, J. (2016). VBPR: visual bayesian personalized ranking from implicit feedback. Proceedings of the AAAI conference on artificial intelligence,

He, X., & Chua, T.-S. (2017). Neural factorization machines for sparse predictive analytics. Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval,

He, Y., Zhang, Y., Liu, W., & Caverlee, J. (2020). Consistency-aware recommendation for user-generated item list continuation. Proceedings of the 13th international conference on web search and data mining,

Hidasi, B. (2015). Session-based Recommendations with Recurrent Neural Networks. *arXiv preprint arXiv:1511.06939*.

Hou, M., Wu, L., Chen, E., Li, Z., Zheng, V. W., & Liu, Q. (2019). Explainable fashion recommendation: A semantic attribute region guided approach. *arXiv preprint arXiv:1905.12862*.

Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. 2008 Eighth IEEE international conference on data mining,

Huang, Y., & Huang, T. (2016). Outfit recommendation system based on deep learning. 2nd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2017),

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. International conference on machine learning,

Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.*, *44*, 223-270.

Jin, B., Gao, C., He, X., Jin, D., & Li, Y. (2020). Multi-behavior recommendation with graph convolutional networks. Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval,

Jing, P., Ye, S., Nie, L., Liu, J., & Su, Y. (2019). Low-rank regularized multi-representation learning for fashion compatibility prediction. *IEEE Transactions on Multimedia*, *22*(6), 1555-1566.

Jing, P., Zhang, J., Nie, L., Ye, S., Liu, J., & Su, Y. (2021). Tripartite graph regularized latent low-rank representation for fashion compatibility prediction. *IEEE Transactions on Multimedia*, *24*, 1277-1287.

Juan, Y., Zhuang, Y., Chin, W.-S., & Lin, C.-J. (2016). Field-aware factorization machines for CTR prediction. Proceedings of the 10th ACM conference on recommender systems,

Kaicheng, P., Xingxing, Z., & Wong, W. K. (2021). Modeling fashion compatibility with explanation by using bidirectional lstm. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Kang, W.-C., & McAuley, J. (2018). Self-attentive sequential recommendation. 2018 IEEE international conference on data mining (ICDM),

Kim, H.-D. (2009). Applying consistency-based trust definition to collaborative filtering. *KSII Transactions*

on Internet and Information Systems (TIIS), *3*(4), 366-375.

Kim, Y., Rome, S., Foley, K., Nankani, M., Melamed, R., Morales, J., Yadav, A., Peifer, M., Hamidian, S., & Huang, H. H. (2024). Improving Content Recommendation: Knowledge Graph-Based Semantic Contrastive Learning for Diversity and Cold-Start Users. *arXiv preprint arXiv:2403.18667*.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, *42*(8), 30-37.

Koren, Y., Rendle, S., & Bell, R. (2021). Advances in collaborative filtering. *Recommender systems handbook*, 91-142.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Lau, J. H., & Baldwin, T. (2016). An empirical evaluation of doc2vec with practical insights into document embedding generation. *arXiv preprint arXiv:1607.05368*.

Li, X., Wang, X., He, X., Chen, L., Xiao, J., & Chua, T.-S. (2020). Hierarchical fashion graph network for personalized outfit recommendation. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval,

Li, Y., Chen, T., & Huang, Z. (2021). Attribute-aware explainable complementary clothing recommendation. *World Wide Web*, *24*, 1885-1901.

Lian, J., Zhou, X., Zhang, F., Chen, Z., Xie, X., & Sun, G. (2018). xdeepfm: Combining explicit and implicit feature interactions for recommender systems. Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining,

Liang, D., Altosaar, J., Charlin, L., & Blei, D. M. (2016). Factorization meets the item embedding: Regularizing matrix factorization with item co-occurrence. Proceedings of the 10th ACM conference on recommender systems,

Liao, S., Ding, Y., & Mok, P. (2023). Recommendation of Mix-and-Match Clothing by Modeling Indirect Personal Compatibility. Proceedings of the 2023 ACM International Conference on Multimedia Retrieval,

Lin, Z., Zang, S., Wang, R., Sun, Z., Senthilnath, J., Xu, C., & Kwoh, C. K. (2022). Attention over self-attention: Intention-aware re-ranking with dynamic transformer encoders for recommendation. *IEEE Transactions on Knowledge and Data Engineering*.

Liu, H., Wang, Y., Peng, Q., Wu, F., Gan, L., Pan, L., & Jiao, P. (2020). Hybrid neural recommendation with joint deep representation learning of ratings and reviews. *Neurocomputing*, *374*, 77-85.

Liu, H. B., Ding, J. Y., Zhu, Y. M., Tang, F. L., Yu, J. D., Jiang, R. B., & Guo, Z. W. (2023). Modeling multi-aspect preferences and intents for multi-behavioral sequential recommendation [Article]. *Knowledge-Based Systems*, *280*, 13, Article 111013. https://doi.org/10.1016/j.knosys.2023.111013

Liu, J. H., Hou, L., Yu, X., Song, X. M., & Ren, Z. C. (2024). Unifying heterogeneous and homogeneous relations for personalized compatibility modeling [Article]. *Knowledge-Based Systems*, *290*, 11, Article 111560. https://doi.org/10.1016/j.knosys.2024.111560

Liu, L., Cai, L., Zhang, C., Zhao, X., Gao, J., Wang, W., Lv, Y., Fan, W., Wang, Y., & He, M. (2023). Linrec: Linear attention mechanism for long-term sequential recommender systems. Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval,

Loni, B., Pagano, R., Larson, M., & Hanjalic, A. (2016). Bayesian personalized ranking with multi-channel user feedback. Proceedings of the 10th ACM conference on recommender systems,

Lu, Z., Hu, Y., Jiang, Y., Chen, Y., & Zeng, B. (2019). Learning binary code for personalized fashion recommendation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,

Ma, C., Ma, L., Zhang, Y., Sun, J., Liu, X., & Coates, M. (2020). Memory augmented graph neural networks for sequential recommendation. Proceedings of the AAAI conference on artificial intelligence,

McInerney, J., Lacker, B., Hansen, S., Higley, K., Bouchard, H., Gruson, A., & Mehrotra, R. (2018). Explore, exploit, and explain: personalizing explainable recommendations with bandits. Proceedings of the 12th ACM conference on recommender systems,

Mnih, A., & Salakhutdinov, R. R. (2007). Probabilistic matrix factorization. *Advances in neural information processing systems*, *20*.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Murtaza, M., Ahmed, Y., Shamsi, J. A., Sherwani, F., & Usman, M. (2022). AI-based personalized e-learning systems: Issues, challenges, and solutions. *IEEE Access*.

Nie, X. Z., Xu, Z. J., Zhang, J. Q., & Tian, Y. (2023). Attention-Based Personalized Compatibility Learning for Fashion Matching [Article]. *Applied Sciences-Basel*, *13*(17), 19, Article 9638. https://doi.org/10.3390/app13179638

Peng, X., Sun, J., Yan, M., Sun, F., & Wang, F. (2023). Attention-guided graph convolutional network for multi-behavior recommendation. *Knowledge-Based Systems*, *280*, 111040.

Qin, Y., Ju, W., Wu, H., Luo, X., & Zhang, M. (2024). Learning graph ode for continuous-time sequential recommendation. *IEEE Transactions on Knowledge and Data Engineering*.

Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2012). BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. Proceedings of the 1994 ACM conference on Computer supported cooperative work,

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). " Why should i trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining,

Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, *9*(1), 59.

Sagar, D., Garg, J., Kansal, P., Bhalla, S., Shah, R. R., & Yu, Y. (2020). Pai-bpr: Personalized outfit recommendation scheme with attribute-wise interpretability. 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM),

Shao, Z., Wang, S., Lu, W., Zhang, W., Guan, H., & Zhao, L. (2024). Filter-Enhanced Hypergraph Transformer for Multi-Behavior Sequential Recommendation. ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),

Shinnou, H., Asahara, M., Komiya, K., & Sasaki, M. (2017). nwjc2vec: Word embedding data constructed from ninjal web japanese corpus. *Journal of Natural Language Processing*, *24*(5), 705-720.

Song, X., Fang, S.-T., Chen, X., Wei, Y., Zhao, Z., & Nie, L. (2021). Modality-oriented graph learning toward outfit compatibility modeling. *IEEE Transactions on Multimedia*.

Song, X., Han, X., Li, Y., Chen, J., Xu, X.-S., & Nie, L. (2019). GP-BPR: Personalized compatibility modeling for clothing matching. Proceedings of the 27th ACM international conference on multimedia,

Song, X., Wang, C., Sun, C., Feng, S., Zhou, M., & Nie, L. (2023). MM-FRec: Multi-Modal Enhanced Fashion Item Recommendation. *IEEE Transactions on Knowledge and Data Engineering*.

Song, X. M., Fang, S. T., Chen, X. L., Wei, Y. W., Zhao, Z. Z., & Nie, L. Q. (2023). Modality-Oriented Graph Learning Toward Outfit Compatibility Modeling [Article]. *IEEE Transactions on Multimedia*, *25*, 856-867. https://doi.org/10.1109/tmm.2021.3134164

Song, Y., Elkahky, A. M., & He, X. (2016). Multi-rate deep learning for temporal recommendation. Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval,

Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., & Jiang, P. (2019). BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. Proceedings of the 28th ACM international conference on information and knowledge management,

Sun, G.-L., Cheng, Z.-Q., Wu, X., & Peng, Q. (2018). Personalized clothing recommendation combining user social circle and fashion style consistency. *Multimedia Tools and Applications*, *77*, 17731-17754.

Tian, Y., Sun, C., Poole, B., Krishnan, D., Schmid, C., & Isola, P. (2020). What makes for good views for contrastive learning? *Advances in neural information processing systems*, *33*, 6827-6839.

Trakulwaranont, D., Kastner, M. A., & Satoh, S. i. (2022). Personalized Fashion Recommendation using Pairwise Attention. International Conference on Multimedia Modeling,

Vasileva, M. I., Plummer, B. A., Dusad, K., Rajpal, S., Kumar, R., & Forsyth, D. (2018). Learning type-aware embeddings for fashion compatibility. Proceedings of the European conference on computer vision (ECCV),

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Wang, J., Ding, K., Hong, L., Liu, H., & Caverlee, J. (2020). Next-item recommendation with sequential

hypergraphs. Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval,

Wang, R., Fu, B., Fu, G., & Wang, M. (2017). Deep & cross network for ad click predictions. In *Proceedings of the ADKDD'17* (pp. 1-7).

Wang, R., Shivanna, R., Cheng, D., Jain, S., Lin, D., Hong, L., & Chi, E. (2021). Dcn v2: Improved deep & cross network and practical lessons for web-scale learning to rank systems. Proceedings of the web conference 2021,

Wang, R., Wang, J., & Su, Z. (2022). Learning compatibility knowledge for outfit recommendation with complementary clothing matching. *Computer Communications*, *181*, 320-328.

Wang, Z., & Bovik, A. C. (2009). Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, *26*(1), 98-117.

Wang, Z., Jiang, Z., Ren, Z., Tang, J., & Yin, D. (2018). A path-constrained framework for discriminating substitutable and complementary products in e-commerce. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining,

Xi, W.-D., Huang, L., Wang, C.-D., Zheng, Y.-Y., & Lai, J.-H. (2021). Deep rating and review neural network for item recommendation. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(11), 6726-6736.

Xia, L., Huang, C., Xu, Y., Dai, P., Zhang, B., & Bo, L. (2020). Multiplex behavioral relation learning for recommendation via memory augmented transformer network. Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval,

Xia, L., Xu, Y., Huang, C., Dai, P., & Bo, L. (2021). Graph meta network for multi-behavior recommendation. Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval,

Xia, L. H., Huang, C., Xu, Y., & Pei, J. (2023). Multi-Behavior Sequential Recommendation With Temporal Graph Transformer [Article]. *IEEE Transactions on Knowledge and Data Engineering*, *35*(6), 6099-6112. https://doi.org/10.1109/tkde.2022.3175094

Xiao, J., Ye, H., He, X., Zhang, H., Wu, F., & Chua, T.-S. (2017). Attentional factorization machines: Learning the weight of feature interactions via attention networks. *arXiv preprint arXiv:1708.04617*.

Xie, X., Sun, F., Liu, Z., Wu, S., Gao, J., Zhang, J., Ding, B., & Cui, B. (2022). Contrastive learning for sequential recommendation. 2022 IEEE 38th international conference on data engineering (ICDE),

Xu, C., Zhao, P., Liu, Y., Sheng, V. S., Xu, J., Zhuang, F., Fang, J., & Zhou, X. (2019). Graph contextualized self-attention network for session-based recommendation. IJCAI,

Xu, D., Ruan, C., Cho, J., Korpeoglu, E., Kumar, S., & Achan, K. (2020). Knowledge-aware complementary product representation learning. Proceedings of the 13th International Conference on Web Search and Data Mining,

Xuan, H., Liu, Y., Li, B., & Yin, H. (2023). Knowledge enhancement for contrastive multi-behavior recommendation. Proceedings of the sixteenth ACM international conference on web search and data mining,

Xue, H.-J., Dai, X., Zhang, J., Huang, S., & Chen, J. (2017). Deep matrix factorization models for recommender systems. IJCAI,

Yan, A., Dong, C., Gao, Y., Fu, J., Zhao, T., Sun, Y., & McAuley, J. (2022). Personalized complementary product recommendation. Companion Proceedings of the Web Conference 2022,

Yang, X., Du, X., & Wang, M. (2020). Learning to match on graph for fashion compatibility modeling. Proceedings of the AAAI Conference on artificial intelligence,

Yang, X., Song, X., Feng, F., Wen, H., Duan, L.-Y., & Nie, L. (2021). Attribute-wise explainable fashion compatibility modeling. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *17*(1), 1-21.

Yang, Y., Huang, C., Xia, L., Liang, Y., Yu, Y., & Li, C. (2022). Multi-behavior hypergraph-enhanced transformer for sequential recommendation. Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining,

Yang, Z., Ye, J., Wang, L., Lin, X., & He, L. (2022). Inferring substitutable and complementary products with Knowledge-Aware Path Reasoning based on dynamic policy network. *Knowledge-Based Systems*, *235*, 107579.

Yu, H., Litchfield, L., Kernreiter, T., Jolly, S., & Hempstalk, K. (2019). Complementary recommendations: A brief survey. 2019 International Conference on High Performance Big Data and Intelligent

Systems (HPBD&IS),

Zhan, H., Lin, J., Ak, K. E., Shi, B., Duan, L.-Y., & Kot, A. C. (2021). $ A^ 3$-FKG: Attentive Attribute-Aware Fashion Knowledge Graph for Outfit Preference Prediction. *IEEE Transactions on Multimedia*, *24*, 819-831.

Zhang, H., Zha, Z.-J., Yang, Y., Yan, S., Gao, Y., & Chua, T.-S. (2013). Attribute-augmented semantic hierarchy: towards bridging semantic gap and intention gap in image retrieval. Proceedings of the 21st ACM international conference on Multimedia,

Zhang, W., Mao, J., Cao, Y., & Xu, C. (2020). Multiplex graph neural networks for multi-behavior recommendation. Proceedings of the 29th ACM international conference on information & knowledge management,

Zhang, Y., Lu, H., Niu, W., & Caverlee, J. (2018). Quality-aware neural complementary item recommendation. Proceedings of the 12th ACM conference on recommender systems,

Zhao, X., Qi, H., Luo, R., & Davis, L. (2019). A weakly supervised adaptive triplet loss for deep metric learning. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops,

Zhou, Z., Su, Z., & Wang, R. (2022). Attribute-aware heterogeneous graph network for fashion compatibility prediction. *Neurocomputing*, *495*, 62-74.

Zhu, J., Zhang, J., He, L., Wu, Q., Zhou, B., Zhang, C., & Yu, P. S. (2017). Broad learning based multi-source collaborative recommendation. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management,