DEEP LEARNING-BASED FASHION ADVISING

DING YUJUAN

PhD

The Hong Kong Polytechnic University

2021

The Hong Kong Polytechnic University

Institute of Textiles & Clothing

Deep Learning-based Fashion Advising

Yujuan DING

A thesis
submitted in partial fulfilment of the requirements
for the degree of

Doctor of Philosophy

September 2020

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.

____________________ (Signed)

Yujuan DING (Name of student)

Abstract

Fashion advising provides proper fashion suggestions to facilitate the decision-making processes, which is helpful for both ordinary people and fashion business. In the recent years, with the accumulation of fashion-related data and the development of deep learning technologies, data-driven fashion analysis based on deep learning has attracted great research attention. This thesis focuses on the deep learning-based fashion advising for ordinary users/individuals, to which personalization and fashionability are two key perspectives. The two perspectives correspond to two basic standards in making desired fashion suggestions, to cater to the user preference/taste in fashion, to offer fashionable guidance and improve their aesthetic ability in fashion. In accordance with personalization and fashionability in fashion advising, this thesis works on two specific tasks which are personalized fashion recommendation and fashion trend foresting.

Existing works on personalized fashion recommendation limited to leverage more characteristic attributes or enhance the visual information of fashion items. However, the shopping patterns in user behaviors and the short-term behavior transition in fashion shopping were overlooked in developing the recommender systems. In the field of fashion trend forecasting, the research of data-driven fashion forecasting is still at its early stage. On the one hand, the fashion elements focused in previous studies are not specific, fine-grained and comprehensive to reveal real fashion trends. On the other hand, the proposed models were still based on statistical models which fall shorts in handling complicated fashion

trend signals. To address the limitations in existing studies, this thesis proposes three main objectives and fulfills them with three deep learning-based approaches accordingly.

First, to effectively capture the user preference in fashion and thus facilitate the performance of fashion recommendation, this thesis proposes to model the underlying shopping patterns in fashion shopping behaviors, which imply the diverse user preferences under specific aspects, such as *style*, *brand* or *print pattern*. For such goal, a Field-aware Graph Collaborative Filtering (FGCF) method is proposed to capture the fined-grained user shopping patterns which have been widely ignored in previous research. Specifically, the proposed FGCF method is able to model the factor field-level interactions and make overall recommendation prediction by aggregating the field-level results. It not only predicts the holistic user-item preference, but also infers the specific fashion preferences in different factor fields. Extensive experiments on real-life fashion purchase data demonstrate the effectiveness of the proposed FGCF method.

Second, to effectively model the short-term transition of user behaviors in fashion shopping and develop better sequential fashion recommender system, this thesis proposes an Attentional Content-level Translation-based Fashion Recommender (ACTR). Specifically, the ACTR leverages the item-item relationships (indicating the short-term intentions) in modeling the item-item interactions. To tackle the sparsity problem in item-item interactions, it introduces the content-level item transition modeling which decomposes the overall item-item interaction into different fashion aspects. Moreover, a user-aware content attention mechanism is devised in the ACTR to properly aggregate the content-level modeling results and generate the final recommendation results. Extensive experiments on real-life fashion shopping data demonstrate the effectiveness of the proposed ACTR method.

Third, towards meaningful fashion trend forecasting, this thesis aims to analyze fine-grained fashion elements which can effectively reveal fashion trends, in specific, to model and forecast the fashion trend of specific fashion elements for specific user groups. To this end, a large-scale fashion trend dataset (FIT) is firstly collected from Instagram and the time series popularity records of fashion elements as well as user information are extracted. To effectively model the time series data of fashion elements with rather complex patterns, a Relation Enhanced Attention Recurrent (REAR) network is proposed, which takes advantage of the capability of deep recurrent neural networks in modeling time-series data and connects specific fashion trends through the relations between user groups and fashion elements. Extensive experiments have demonstrated that REAR can make solid and meaningful fashion trend forecasting for a period of time in the future.

In summary, this thesis works on the deep learning-based fashion advising problem from two different key perspectives and studies three specific sub-tasks. With the output of the three studies, the personalized and fashionable fashion advice are able to be generated for specific users based on the specific approaches proposed. The research problems extracted and addressed in this thesis promote the development of personalized fashion recommendation and fashion trend forecasting. Moreover, the output of the thesis has a strong impact on the entire fashion industry which can specifically benefit the process of design, manufacturing and retailing.

Acknowledgement

First of all, I would like to thank my supervisor, Prof. Calvin Wong for providing me such a valuable opportunity to do my PhD in PolyU. I appreciate all his advising and guidance on my study and research over the years. I also want to thank Prof. Lai Zhihui in Shenzhen University who offered great help in leading me in the way of doing research.

I want to send my gratitude to Prof. Tat-Seng Chua in National University of Singapore, for providing me the great opportunity to study at NExT++. Thank Dr. Xun Yang for his kind and patient mentoring during my stay at NExT++. I also want to thank Ma Yunshan, for those insightful and inspiring discussions on research, as well as all other members at NExT++.

Special thanks should be sent to my friends, co-workers and roommates, Zhou Yangping and Zhou Yanghong who have been taking care of and helping me selflessly during my PhD. I also want to thank all my friends, who have been inspiring, encouraging and supportive all the time. Last but not the least, I want to thank my family for their constant understanding and supporting, and my cousins for helping take good care of my family over the years when I was far away from home.

The PhD degree might be nothing, but it would be no waste of time if it can bring some joy to people who love me and people that I love. Thank you all.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

Fashion is an essential aspect of culture and life, and the fashion industry is one of the most important parts of the global economy. At present, fashion industry accounts for 4% of the global economy, with a market value of $406 billion [1]. Driven by social demand and large profit potential, fashion has always been a hot research topic in academia. The fashion industry has been undergoing a digital transformation with the growing influence of social networks and the prosperity of e-commerce over the years. Such transformation can benefit the whole industry, but also requires strong technical support. In recent years, with the rapid development of artificial intelligence (AI), deep learning (DL) particular, researchers attempt to seek computational solutions to address various tasks in the fashion domain to meet technical demands from the industry. Meanwhile, the fashion industry is inclined to embrace new technologies, understand the science and engineering for new production methods and recognise the potential to innovate in the design of new systems and processes. It is reported that by 2023, 50% of large global companies will be using AI, advanced analytics and the Internet of Things in supply chain operations [2]. With such an industrial background, computational fashion studies have attracted increasing attention in computer vision, machine learning and multimedia communities.

The main reason that fashion has become an important application domain for DL is that many tools and innovative solutions based on DL can truly contribute to the industry, for example, by enabling businesses to improve their operations

[1] https://fashionunited.com/global-fashion-industry-statistics/
[2] https://www.gartner.com/smarterwithgartner/gartner-predicts-2019-for-supply-chain-operations/

and shopper experience. The studies that can produce top-level applicable solutions are usually grouped as high-level fashion applications. Compared with low-level research topics on fundamental computer vision problems in fashion, such as clothing parsing [188, 38, 108, 106, 173, 186] or fashion attribute recognition [127, 74, 13, 26, 148], high-level application studies are more task-oriented in seeking an applicable value to the industry [151]. Representative high-level fashion applications include fashion retrieval [104], fashion recommendation [193, 49], fashion synthesis [129, 24], visual try-on [25, 50], and others [110, 113].

One kind of fashion application research aims to acquire valuable insights or relevant information from large amounts of fashion-related data by proper analytics, thereby providing specific fashion advice. Such kind of research can be considered as data-driven fashion advising. For example, with the browsing and shopping records of customers, fashion retailers can analyze customers' shopping preference and therefore provide personalized and diverse services. In most cases, consumers can also benefit from those research outcomes for making better fashion choices and shopping decisions. Such fashion advising research has gained much research attention for its great significance to academia and industry, which brings technical challenges at the same time.

1.1.1 Fashion Advising

Fashion advising, which refers to the task of providing proper fashion suggestions to facilitate the decision-making process, is important in terms of helping ordinary people and the fashion business. Although technological change has infiltrated every aspect of modern life, people's desire to convey a sense of self through their appearance has not changed [10]. However, not everyone is a master of fashion, and those who lack the sense of fashion may need proper fashion advice in daily life. In other words, fashion advising technologies can benefit ordinary users in providing useful fashion suggestions such as personalized fashion item

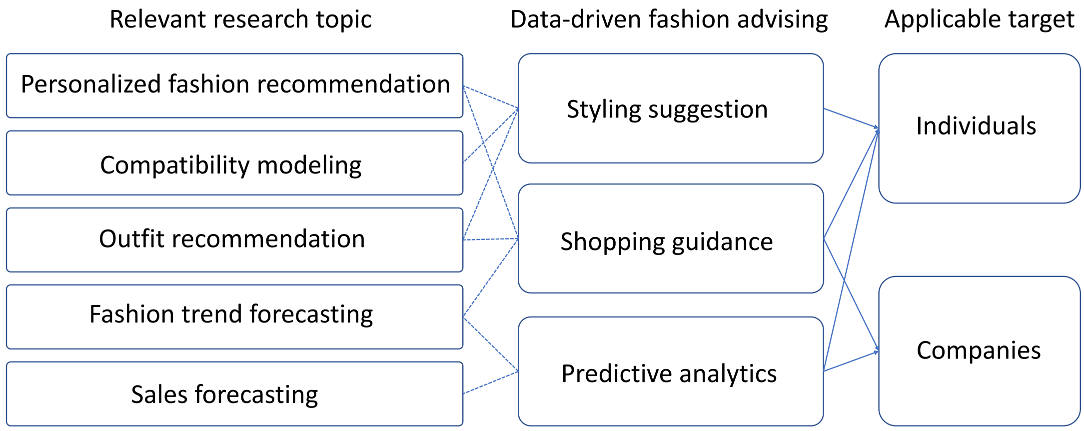

Figure 1.1: Illustration of data-based fashion advising, relevant research topics and its applicable targets

recommendation, mix-and-match guidance, or potential trend inspiration. In the business aspects, precise recommendation for users enables the fashion retailers, especially e-commerce retailers, to attract more customers, keep them longer in the platform and therefore increase the volumes of transactions. Meanwhile, fashion companies also directly benefit from fashion advising technologies which can act as consultants in establishing business strategies at any stage.

Fashion advising has been pervasive across the fashion business as well as the daily fashion-related lives of people. For example, fashion companies may consult with forecasting companies when formulating business strategies. Also, when shopping fashion, offline or online, people commonly receive advice from salespeople or the systems. Traditionally, fashion advice, for business or individuals, are generated by certain fashion experts based on their professional experience, knowledge and analytics. At present, the accumulation of huge amounts of fashion-related data with the advent of digital age has provided an alternative data-driven way for conducting fashion advising. This thesis attempts to employ advanced DL technologies to effectively generate fashion advice based on large-scale cross-media fashion data, which is classified under high-level fashion application research.

Fashion advising is a wide subject that can be divided into small research directions according to various application scenarios. As shown in Figure 1.1, the three main directions are styling suggestion, shopping guidance, and predictive analytics. Specifically, style suggestion aims to help individuals in styling by providing an effective matching and styling knowledge. Compatibility modeling is the topic that mostly focuses on styling suggestion. In addition, style suggestion can also be sub-objective in other research topics such as personalized fashion recommendation and outfit recommendation. The second direction, shopping guidance, is a specific application in fashion shopping scenario, which aims to help shoppers find their desired fashion items as well as retailers attract more buyers. Providing effective shopping guidance is the practical objective for most fashion recommendation studies. The third direction, predictive analytics, is more helpful in facilitating the business operations of fashion companies, while also inspiring individuals with predicted fashion trends. Figure 1.1 illustrates several important specific research topics in the field of computational fashion which are highly correlated to the topic of data-driven fashion advising. All listed research topics in Figure 1.1 focus on various aspects and try to partially address the task of fashion advising.

The main focus of this thesis is fashion advising for individuals, and mostly for online shopping scenarios. Specifically, this study aims to apply the powerful DL tools and develop effective models that can generate useful fashion advice based on available relevant fashion data. Three aspects are important in data-driven fashion advising.

(1) Users' personal fashion preference. Undoubtedly, personalization is one of the most important issues in fashion. In fashion advising, effective advice should cater to specific fashion tastes of target users. For example, if Sandy prefers street-style clothing, suggesting sexy items is not suitable for her. Therefore, the basic requirement for fashion advising is to effectively model users' personal preference from any forms of personalization-aware fashion data, such as purchase records.

(2) Users' short-term taste and behavior transition. The personal fashion preference pointed out above is usually a static, long-term standard to provide advice to different users. However, in some specific application scenarios, such as online browsing for fashion shopping, the user's temporary fashion taste swiftly changes, which usually happens along with his/her fashion selection behaviors. Such short-term patterns in their behavior, which is called behavior transition, also matters in capturing the user's real overall preference at a specific period.

(3) Fashion trend. The aforementioned two points are mainly from the personalization perspective, which aims to explore important personalized patterns from user behaviors. Besides personalization, fashionability is another key factor in making effective fashion suggestions. As discussed above, people tend to pick pretty, trendy pieces when making fashion choices. However, as fashion trends are constantly changing, in most cases, ordinary people experience difficulty in catching up with trends on their own. Under the circumstance, introducing the latest fashion trends is important in providing solid fashion advice.

The above three points are main perspectives to study the data-based fashion advising problem, which belong to two specific research topics: fashion recommendation and fashion trend forecasting.

Fashion recommendation generally refers to suggensting fashion items (mostly in e-commerce scenarios), while fashion trend forecasting focuses on mastering fashion trends and predicting the ups and downs of specific fashion elements. The two tasks explore different angles for solving the fashion advising problem. First, fashion appeals to everyone in the world on some level, but different people have different preferences and tastes. Fashion recommendations, therefore, aims to capture the various user preference so as to make more customized and personalized fashion suggestions. Second, as fashion constantly evolves, regardless of the difference between personal tastes, the fashionability of specific fashion elements (such as the color *white* or the pattern *stripe*) changes over time.

As people are naturally after fashionability in making fashion choices, fashion trend forecasting plays a significant role in offering preferable options, especially for those who are less fashion-conscious. Overall, an effective fashion adviser needs to simultaneously master the user's taste and the general fashion trend to make proper suggestions, which requires focusing on the techniques of both sides.

1.1.2 Personalized Fashion Recommendation

Fashion recommendations have been widely applied in fashion online shopping, from online store of fashion brands such as Nike [3] to multi-brand retailing platforms such as Amazon [4]. It is not an edge but an essential component in modern online systems. The potential commercial benefit of fashion recommendation has attracted a lot of research attention. According to different recommending targets, fashion recommendations can be further grouped into item-level [193] and outfit-level recommendations [66]. Based on the requirement, it can make recommendations based on a given fashion item [145] or user [14]. The outfit recommendation and item-based recommendations are reverent to but not the focus of this thesis, which are briefly reviewed in the next chapter. As the importance of personalization is widely acknowledged in fashion as claimed above, personalized fashion recommendations (recommending for users) are one main research target in this thesis.

Personalized recommendations aim to model the user preference based on user interaction feedback, such as click or purchase records [143]. These personalized recommendations in general domain or specific domains such as movies, books, music and news have been widely studied [175, 178, 101]. The basic rationale is to train the recommendation model to fit the user-item interaction history data and hopefully the users' future behavior can be effectively predicted by the trained

[3]https://www.nike.com/
[4]https://www.amazon.com/

model. Personalized fashion recommendation can be considered as a special application of general personalized recommendation problems. In the fashion domain, classic personalized recommendation strategies, such as collaborative filtering (CF) [144], have also been applied to tackle the personalized fashion recommendation problems [66, 193]. In most existing works, the user-item interactions are simply modeled in a classic and implicit manner [193], usually with the inner product (or MLP layer) of user and item representations [57].

Although a large number of general recommender systems are available, they can not necessarily be directly applied in the fashion domain perfectly. When studying the recommendation in the fashion domain, the unique characteristics of the fashion domain should be respected, which are mainly in the following aspects. First, fashion items are usually with abundant attributes and many of specific attributes can influence the users' appreciation of the items greatly. Important attributes, such as styles, colors, have been explored in previous studies and found effective in helping improve the recommendation performance. [133, 119]. Second, the aesthetic attributes matter more in the fashion domain. That is to say, the visual information may be more important for describing the recommending candidates, *i.e.*, fashion items, than in any other domains [193]. Third, the number of fashion items is extremely huge. Each item may have only been chosen by very few or even no users. Therefore, CF methods that only use user-item associations to characterize items may not be sufficient in the fashion domain [66].

Some effort has been made to develop personalized recommender systems that adapt to the fashion domain and respect the domain characteristics. One direction is to leverage more characteristic attributes [133]. The second direction is to enhance the visual information [193]. However, as we can see, most existing item-level personalized fashion recommendations are still based on the CF framework with minor revisions. Some of them might have leveraged more side information that is important in the fashion domain to enhance the entire model, but few dig deep into the characteristics of shopping behaviors in fashion. Two important aspects that have been overlooked in previous fashion recommendation research are summarised as follows:

(1) Shopping patterns in long-term behavior. Jannach *et al.* [72] investigated the user's browsing behavior in online fashion shopping and observed that on average, users inspected about 9 different items from 2.7 (of more than 330 available) categories, and considered 2.5 different colors and 3.6 brands in one session. Such a discovery suggests that users indeed often have specific shopping patterns when buying fashion items. Therefore, recommending items in accordance with their shopping patterns seems promising.

(2) Short-term behavior transitions and user intentions. Previous works on personalized fashion recommendations have attempted to model user's general fashion taste mainly, but neglected their short-term intentions. As we known, the key of recommendation is to master the users' preference. Although a user's fashion taste may be relatively stable, his/her preference evolves frequently in the real fashion shopping process as the short-term intention changes. Such phenomenon are obvious when users are browsing through fashion items. Therefore, to effectively explore users' preference, both long-term static and short-term dynamic parts should be involved.

1.1.3 Fashion Trend Forecasting

Compared with fashion recommendations that have been an active research topic, the data-driven fashion trend forecasting is still at its nascent stage [29]. Undoubtedly, fashion trend forecasting is of great demand for both fashion companies and users. Traditional fashion forecasting is mostly human-based, relying on fashion experts to examine artistic viewpoints, culture, societal attitudes and current events to predict the future [42]. In the recent decade, technological innovations such as the Internet has accelerated the changing rate of fashion, which makes fashion trend forecasting even more difficult. However, the advent of the digital age has also facilitated the accumulation of huge amounts of fashion related data, which provides an alternative data-driven way of addressing the fashion trend forecasting task [110].

The literature includes a few exiting data-driven fashion trend forecasting works, and most of them have twofold limitations. First, as preliminary attempts, only limited fashion elements with highly seasonal or simple patterns are investigated. The problem is that some elements can hardly reveal the real fashion trends. For example in two previous works [113, 3], *wearing jacket* was studied as the target fashion element, which is season-related rather than fashion-related. Second, as the fashion forecasting task is usually formulated as a time-series prediction problem, existing works still use statistical models [113]. Classic statistical models might be effective when modeling time-series signals with simple patterns, but for many fashion trend signals which are more complex and less cyclic, they can hardly achieve preferable performance in fashion trend forecasting.

Apparently, to conduct insightful trend forecasting with practical significance to the whole fashion industry, the research target should be specific and meaningful. It can be detailed design element such as *dotted pattern* or a general fashion style such as *sporty*. From the perspective of techniques, deep neural networks (DNNs) have shown their superiority in modeling sequential data in recent years [90, 116]. In particular, recurrent neural networks (RNNs) [62, 18] have achieved superior performance in relevant applications [33, 31]. These accomplishments have inspired us to develop more advanced fashion trend forecasting models based on DNNs.

Although personalized fashion recommendations and fashion trend forecasting are different tasks in terms of having different application purposes, both of them rely on effectively exploring historical data for making proper predictions. As mentioned, they are the two most important parts in the research area of fashion advising that aims to provide fashion advice from two perspectives, yet both benefit the fashion business and ordinary people simultaneously. By properly analyzing the current research status in the literature and considering the application demand, this thesis extracts three important research problems on the two tasks as the main research objectives.

1.2 Objectives and Challenges

This thesis seeks to employ deep learning technologies for effective fashion advising to help consumers or fashion lovers with their fashion choices as well as fashion companies with their businesses. Briefly, the research focus of this study lies in two perspectives: 1) develop more advanced personalized fashion recommender systems to enhance the personalization of fashion advising and 2) develop effective fashion trend forecasting models to ensure fashionability of fashion advising. More specifically, three specific research objectives and corresponding challenges are introduced as follows:

(1) To develop a recommendation approach that can model user shopping patterns in fashion shopping and therefore enhance the personalized fashion recommendation performance. Rather than leveraging more visual information or assistant features (such as aesthetic features), this study aims to go a step further in personalized fashion recommendations by exploiting users' interaction histories and analyzing the fashion shopping behaviors combined with the characteristic of the fashion domain. Certain shopping patterns are assumed to exist underneath these fashion shopping behaviors. For example, a user who loves *shoes* tend to buy shoes often, and a user who loves *white* is likely to buy more white items. The challenge is how to explore such shopping patterns from the users' historical behaviors and then boost the recommendation performance.

(2) To model content-level user behavior transitions in fashion shopping and accordingly develop a better sequential fashion recommendation approach. On top of the long-term preference of users (fashion taste in general), users' short-term intentions, which are usually implied in user behaviors, also play important roles in making their fashion decisions. To further improve the personalized fashion recommendation approach, both long-term preference and short-term intentions should be explored when modeling the users' interaction behaviors in fashion. In other words, the short-term transition in user behavior, which has been neglected in previous studies on fashion recommendations, should be respected and properly modeled. Such behavior transitions are especially important in some application scenarios such as instant recommendation when users randomly browse fashion items online. The main challenge in technique is how to leverage the effective model of such short-term patterns in the data along with the long-term user-item interactions considering the characteristics of fashion.

(3) To study fashion trend from the data perspective and effectively model the fashion trends and perform forecasting based on available relevant data. Towards meaningful fashion trend analysis, the first goal is to define specific fashion trends as research targets, which should have fashion significance and truly reveal fashion trends. Second, effective models should be designed to analyze fashion trend data, capture underlying important patterns and therefore enable solid forecasting.

1.3 Methodology

This study aims to improve the data-driven fashion advising from three aspects, resulting three research objectives as listed above. The specific approaches proposed corresponding to the three objectives are introduced as follows.

(1) A graph-based fashion recommendation method called Field-aware Graph Collaborative Filtering (FGCF) is developed to capture the fine-grained user shopping patterns. Specifically, all categorical factors (*e.g.*, *black*, *elegant*) are grouped into multiple fields (*e.g.*, *color*, *style*) and the interactions of factors at the field level [78] are modeled through factor field-level embedding propagation and aggregation on a fully-connected graph. Then, the field-aware interaction scores of user-item pairs in different factor fields are predicted, based on which the holistic score is further aggregated and used for pairwise training.

(2) An Attentional Content-level Translation-based Fashion Recommender (ACTR) is proposed to model both the user-item compatibility and the sequential dynamics among items. To effectively model the transition process of the user's behavior, that is, the interaction sequences of items, the content-level translation operation based on specific item-item relationships (substitution and mix-ant-match) is introduced in ACTR. In other words, it models the item-item relational interaction from different fashion aspects (also called factor field, such as color or *style*). The final recommendation results are generated based on the mixture of content-level translation models by an attentional combination, and the general user-item interaction model.

(3) A Relation Enhanced Attention Recurrent (REAR) network, which takes advantage of the capability of RNNs in modeling time-series data with the help of the sliding temporal attention, is proposed to address the time-series fashion trend forecasting task. To conduct insightful fashion trend forecasting with practical value in terms of the forecasting targets, this study focuses on fine-grained fashion element trends for specific user groups. To this end, it first contributes a large-scale fashion trend dataset (FIT) collected from Instagram with extracted time-series fashion element records and user information. Furthermore, to effectively model the time-series data of fashion elements with rather complex patterns, this study proposes the REAR model. It connects specific fashion trends through the relations between user groups and fashion elements rather than treating each fashion trend

signal in an isolated manner. Such relation leveraging is able to facilitate the deep learning model to capture the patterns of specific fashion elements.

1.4 Significance

The research in this thesis is important to both academia and industry. To academia, it extracts the key research problems from different aspects of fashion advising. Considering these various technical challenges, this study proposes three novel solutions to address the specific tasks. For industry, the research outcome of this thesis is applicable in several processes in the fashion value chain, thereby assisting the fashion business.

1.4.1 Significance to Academia

This study represents the first attempt to focus on the underlying pattern in fashion shopping behaviors at the first time. Instead of directly applying general recommendation algorithms to fashion domain, this study respects the characteristics of the fashion domain and the uniqueness of fashion shopping compared with other domains. A novel graph-based model is proposed to effectively explore specific shopping patterns with regard to various factors. With the proposed method, not only can the holistic user-item preference be predicted, but also the specific fashion preferences in different factor fields can be inferred. Extensive experiments demonstrate the effectiveness of the proposed approach which achieves better recommendation accuracy than existing recommender models.

This study explores users' short-term behavior transitions in the process of fashion shopping and aims to model both the user-item interaction and item-item interaction to further facilitate the fashion recommendation performance. This is first time in the fashion domain that the item-item relationships are leveraged in modeling the item-item transition. More importantly, a content-level translation-based approach for the item-item transition modeling is proposed, which is able to specify the item-item relationships from different aspects of fashion items. Such an improvement can alleviate the sparsity problem of item-item interaction when only applying item IDs [80], effectively explores transition relations between adjacent items and therefore provides more preferable next-item recommendations.

This study goes further into the fashion trend forecasting task and applies DNN-based technologies in modeling fashion trend signals. To facilitate the fashion trend forecasting research, this study contributes a large-scale dataset based on the social media platform Instagram, which contains the trend data of specific fashion elements and specific groups of people. The dataset spans five years and covers around 200 specific fashion elements and over 70 user groups based on over 680,000 raw images and other meta-data. Furthermore, a novel RNN-based model is proposed to effectively model the historical trends and thereby make solid trend predictions. Specifically, this study proposes to leverage multiple relations between different fashion trends, as well as apply temporal attention mechanism to improve the fashion trend forecasting performance. Extensive experiments have shown that our proposed model outperforms all baselines in terms of forecasting precision. Furthermore, the fashion trends predicted by our model are aligned with some professional human-based predictions.

1.4.2 Significance to Industry

Undoubtedly, the research task of fashion advising has a strong impact on the entire fashion industry throughout the fashion supply chain. In particular, fashion advising has the potential to benefit the process of design, manufacturing and retailing.

The fashion design and manufacturing process can use fashion advising to their advantage [5]. Traditionally, designers are those who present, create and lead fashion trends through the new items they produce. However, such traditional designer-leading mode involves high business risk because of designers' over-reliance on their artistic vision and too little input from a commercial perspective [70]. In this study, fashion trends can be effectively predicted based on big fashion data through DL models, which offer designers another view from the market, and therefore help them create products that are more appealing to consumers. In the manufacturing aspect, this study can help the manufacturers and developers make wiser business operations as well as reduce excess inventory [71].

The significance is greater in the retail process. First, brands such as Zara, H&M, TopShop, and Forever 21 have built their businesses on speed and agility. Once these retailers spot a new trend, they can deploy their hyper-rapid design and supply chain systems to bring the trends to market as quickly as possible. Second, every marketer knows that personalization is key to creating a marketing campaign, particularly in fashion. Retail companies, brands, or e-commerce platforms can attract more customers through well-designed recommender systems. Moreover, by analyzing the data of users' shopping behavior, they can better understand their consumers, master their shopping patterns, fashion taste, and even their instant short-term intentions. A promising direction for digital retailers is to improve conversion rates and boosts engagement if facilitated with effective recommender systems.

1.5 Thesis Structure

The rest of the thesis is structured as follows:

[5] https://www.cgsinc.com/blog/how-big-data-impacting-fashion-industry

Chapter 2 makes a comprehensive literature review on topics related to the personalized fashion advising from different aspects. First, from the technical perspective, several typical deep learning technologies that are relevant to this thesis in the development of new solutions are reviewed. Thereafter, the chapter briefly summarizes the different levels of computational fashion research in the literature. Finally, the chapter reviews in detail the research progress on the relevant tasks of personalized fashion recommendation and fashion trend forecasting, and summarizes the limitations of the current research on the two topics.

Chapter 3 illustrates the overall methodology of this thesis, which provides an overall picture of how the research problem is analyzed and handled. Furthermore, the chapter introduces the approach developed for each objective specifically and how these novel approaches address the sub-tasks and further fulfill the research purpose of this thesis.

Chapter 4 works on the task of modeling shopping patterns for personalized fashion recommendation. It first introduces the motivation that exploring shopping patterns for recommendation is important in the fashion domain as it can benefit the model preference modeling. A novel approach Field-aware Graph Collaborative Filtering (FGCF) is proposed for addressing the specific task. Extensive experiments prove the effectiveness of the model and demonstrate that the proposed FGCF can provide better fashion recommendation in terms of both accuracy and interpretibility.

Chapter 5 aims to model content-level relational transition in user interaction behaviors to facilitate the personalized fashion recommendation. It proposes to leverage the item-item relationships into the transition modeling to improve the model capability and explicability. A novel Attentional Content-level Translation-based Fashion Rcommender (ACTR) method is proposed to effectively model the relational item-item transitions combining characteristics of the fashion domain.

Preferable recommendation results are achieved by the proposed ACTR in extensive experiments based on real-world fashion shopping data, which validate the effectiveness of proposed method.

Chapter 6 focuses on the data-driven fashion trend forecasting problem based on social media. The chapter aims to explore effective patterns in the fashion trend of specific fashion elements for a certain group of people and perform future prediction based on history. A novel Relation-Enhanced Attentional Recurrent network (REAR) is proposed to model the fashion trend signals and make forecasting. Experimental results demonstrate the effectiveness of the proposed REAR method in terms of making solid long-period fashion trend forecasting.

Chapter 7 summarizes the whole thesis, emphasizes the contributions and points out the limitations. It also provides possible research directions in the future.

2 Literature Review

This thesis studies the personalized fashion advising problem, aiming to develop deep-learning-based solutions for solving specific tasks that relates to fashion advising from different aspects. This chapter first reviews several important deep learning advances that are involved in developing novel solutions in this thesis, including convolutional neural network (CNN), recurrent neural network (RNN), graph neural network (GCN) and knowledge graph (KG). It then introduces the development in computational fashion [151] analysis briefly to provide readers the whole picture at macro level, clarify the position of our research focus, *i.e.*, fashion advising, in the whole computational fashion field and how it relates with other computational research in the fashion domain. Thereafter, two specific topics closely related of this thesis, namely, fashion recommendation and fashion trend forecasting, are reviewed in detail. The overall research status on personalized fashion advising as well as the current limitations are summarized in the end.

2.1 Deep Learning Advances

Deep learning (DL) is a powerful technology which can discover intricate structures in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters [90, 39, 156]. In the past decade, deep learning, as well as artificial intelligence (AI), has achieved huge success and showed superior performance in many applications, from computer vision (CV), natural language processing (NLP) to acoustic modeling. In this part, the author reviews several important topics in the area of deep learning, including convolutional neural network (CNN), recurrent neural network (RNN) and graph neural network (GNN). All of these reviewed topics are highly related to this thesis and applied when we develop our own models for specific tasks.

CNNs are designed to process data in the form of multiple arrays, for example, a colour image composed of three 2D arrays containing pixel intensities in the three colour channels. CNNs are special type of deep feedforward network that are much easier to train and generalized much better than networks with full connectivity between adjacent layers [90]. It has achieved great success in the computer vision community and brought back people's confidence in neural networks. According to Lecun *et al.* [90], there are four key ideas behind CNN: local connections, shared weights, pooling and the use of many layers. The first deep CNN architecture was LeNet [91] proposed by Lecun early in 1998, which is composed of all CNN essentials including convolution, subsampling and fully connected layers. LeNet has a groundbreaking significance for the development of deep CNNs, but it did not attract enough attention at that time due to the insufficient hardware computing and data [172]. Five years later, in 2012, as the computer hardware developed and the amount of available data increased, another newly proposed deep CNN architecture, AlexNet [87], achieved amazing results in the ILSVRC-2012 image classification competition. The success of AlexNet attracted an increasing number of researchers and the CNN was in high-speed growth in the following years. Some famous CNN architectures were subsequently proposed, such as VGG [147], GoogLeNet [157] and ResNets [51]. To date, CNN has been well-developed and widely applied to extract visual features in many high-level applications for its exceptional ability in image analysis. As visual information is extremely important in the fashion domain, visual feature extraction is an essential technical aspect of fashion recommendation and fashion trend forecasting.

RNNs are the main tools to handle sequential data involving variable length inputs or outputs, such as natural language and speech signals [140, 39]. The early RNN models suffered for the gradient vanish problem until a better model called

Long-Short Term Memory (LSTM) [62] was proposed in 2000. Another LSTM variation, Gated Recurrent Unit (GRU) [18] was proposed in 2014 and became popular in the research community for its lower computation cost and simplicity. So far, RNNs, mainly referring to LSTM and GRU, have been successfully applied in NLP [171, 155], time series analysis [89], speech and audio processing [150, 141] and traffic tracking and monitoring [5, 95]. RNNs have been successfully applied in recommender systems for developing sequence-aware recommendation algorithms [60, 28, 132]. RNNs are also powerful in modeling time-series data which are naturally sequential. They have achieved state-of-the-art performance in time series prediction [31, 97, 130] and have been successfully applied in specific tasks such as stock prediction [33] and sales forecasting [6]. In fashion trend forecasting, available data are usually in a time series, making them suitable to be processed by RNNs. However, few attempts have been made to study the fashion trend forecasting problem with RNNs so far.

GNNs are deep learning-based methods that operate in the graph domain [198]. Graph is a powerful tool to model a set of objects (nodes) and their relationships (edges), which has attracted increasing attention in the deep learning community for its great expressive ability. It shows great applicable value for modeling data in non-Euclidean structure across various areas such as social science (social networks) [48, 142], natural science (physical systems) [7, 35], knowledge graphs [47], and others [81]. In recommender systems, GNNs has been applied in various ways. Not only can they model graph data that widely exist in recommendation problems, but they can also model the information transformation and aggregation which is extremely important for representation learning in recommender systems. For example, in social recommendations, the social relations can be effectively exploited with the help of GNNs [32]. The graph can also model the high-order connectivity of users and items and make the traditional collaborative filtering recommendation model to become more expressive [177]. In session-based recommendation, GNNs can be helpful for exploring rich transitions among items and generate accurate latent vectors of items [184]. Another special graph, knowledge graph (KG), has been applied to incorporating richer relations between users and items, as well as other side information such as attributes [175, 170, 11].

KG Embedding is an important research direction which aims to embed components of a KG such as entities and relations into continuous vector spaces to simplify the manipulation while preserving the inherent structure of the KG [170]. This task has attracted massive research attention since it was proposed several years ago for its great significance in many high-level tasks in KG [8, 120, 179, 181]. The research on KG embedding also benefits the recommendation domain by providing another idea to model the historical data of user behaviors for the development of the recommender system. Except for incorporating additional relations as mentioned, KG embedding approaches, such as TransE [8] and TransR [98], model the users, items and their basic interactions in a translation manner [52, 37, 80, 160], which has achieved superior performance in various recommendation tasks.

Other deep learning technologies such as Generative Adversarial Net (GAN) [40], Transformer (self attention) [164], Autoencoder (AE) [12], Reinforcement Learning (RL) [117] are also frequently explored in recommendation systems. However, these technologies are not involved in the approach development in this thesis, and therefore are not reviewed in detail.

2.2 Computational Fashion

Deep learning-based fashion advising, the research target of this thesis, is a part of computational fashion research. Computational fashion [151, 45] refers to applying advanced computational methods, such as DL and AI, to address the specific application problems in the fashion domain. Based on the reviews on computational fashion developments [151, 45, 16], the tasks of fashion analysis in the literature can be roughly categorized into three groups: 1) low-level fashion recognition, 2) mid-level fashion understanding, and 3) high-level fashion analysis. The research in this thesis can be categorised under the last group. The related works in fashion advising, specifically, fashion recommendation and fashion trend forecasting, are discussed in detail in the following sections. In this section, the rest of computational fashion works that belong to different research levels are reviewed first.

The pixel-level fashion recognition tries to categorize each pixel in the fashion images into specific fashion categories, such as skirt or t-shirt. This task remains challenging due to the very fine-grained and wide range of fashion categories, as well as the combination and layers of fashion pieces in the image. Undoubtedly, pixel-level recognition is the foundation of the rest of fashion tasks for handling visual data. The specific tasks of pixel-level fashion recognition include clothing parsing and landmark detection. Researchers have applied various techniques, such as graphic, CNN and adversarial models, for addressing the pixel-level fashion recognition problems and obtaining pixelwise labels for fashion images [188, 38, 108, 106, 173, 186].

The pixelwise labels further facilitate the mid-level fashion understanding such as clothing detection, attribute and style recognition. For fashion attribute recognition, even the most basic and fundamental problem: fashion attribute definition, has been very tricky. Effort has been made in the literature to define the fine-grained fashion attribute professionally [23, 200] and employ various machine learning methods to conduct attribute classification [127, 74, 13, 26, 148]. Style classification is another widely studied mid-level fashion understanding task due to the significance of style in the fashion domain. When it comes to fashion style classification, the very basic issue, namely, the definition of style, is also a key issue. To apply deep learning methods, we have to define fashion styles in a definite manner. However, as fashion style is a subjective concept, providing explicit definitions is almost impossible. Despite the difficulties, researchers have also made some attempts, including defining basic fashion styles and linking them with handcrafted features [82], building fashion semantic space for describing clothing fashion styles[109], training polylingual topic models on outfit data to learn correspondences between fashion element and styles [162], and others [76, 75].

Based on lower-level fashion techniques, more higher-level applicable studies in fashion can be conducted. On top of fashion recommendation and trend forecasting, high-level fashion analysis includes fashion retrieval, fashion synthesis and others. Cross-scenario retrieval has attracted much research attention for having more practical significance. One representative cross-scenario retrieval task is the street-to-shop retrieval. To find the similar items from different domains, *i.e.*, street and shop, various solutions have been proposed in previous research based on metric learning, human pose estimation, sparse coding, and others [68, 96]. Some studies even aimed to retrieve the exact same item in the street photo from the online shop [104]. Whichever retrieval standard used, the key problem of street-to-shop retrievals is to train the model to learn similarities between the street and shop domains. A large number of works have studied on the street-to-shop retrieval task and proposed different solutions. For example, Wong *et al.* [174] proposed the Siamese Inception Network, Jiang *et al.* [77] proposed the Bi-directional cross-triplet embedding algorithm, and others [17]. Fashion synthesis has been a rapidly evolving topic in recent years with the great development of generative models such as GAN [40]. Among fashion synthesis research, one of the most eye-catching tasks is virtual try-on, in which some fascinating outputs have been produced [25, 192, 50, 169].

Both the low-level and mid-level fashion analysis are closely related to fashion advising research and many of them are important technical foundations. For example, accurate recognition of fine-grained fashion attributes can enhance the item representation greatly, thereby improving the recommendation performance. In some studies, fashion recommendation is directly based on visual analysis.

For instance, Zhou *et al.* [199] formulated the fashion recommendation as a cross-model retrieval problem and leveraged a human parsing model to enhance the visual feature extraction.

2.3 Fashion Recommendation

Fashion recommendation is an important application field for recommender systems, it has special challenges as well as significance due to the characteristics of fashion domain. According to different recommending targets, fashion recommendation can be grouped into item-level and outfit-level recommendations. Based on the recommending standard, it includes similarity-based and compatibility-based recommendations. The former is very similar to the task of similar fashion retrieval mentioned above. The latter, which is also known as mix-and-match task, focuses on discovering the fashion compatibility rules from fashion matching data. Based on whether personalization is involved in the recommender system or not, we have the personalized fashion recommendation and non-personalized one. In the rest of this section, we review several topics related to this thesis, including personalized recommendation, fashion compatibility modeling, outfit recommendation and personalized fashion recommendation.

2.3.1 Personalized Recommendation

Personalized recommender systems have been playing a vital and indispensable role in various information access systems to boost business and facilitate the decision-making process and are pervasive across numerous web domains such as e-commence and/or media websites [196, 73]. They try to estimate users' preferences on items based on historical user-item interactions, as well as the side information from both user and item sides. The recommendation methods are usually classified into three categories: collaborative filtering (CF), content-based and hybrid [1]. CF usually explores user preferences on specific items from historical user-item interaction records, which can be browsing history or purchase history. Content-based methods, instead, try to recommend an item to a user based on the description of the item and the profile of the user's interests [128].

CF methods can be further grouped into memory-based and model-based. The idea of memory-based CF methods is to make the estimation of user-item interaction directly from the user-item similarity matrix [22]. However, memory-based methods are not applicable when the size of the item set is extremely large but the interaction data is sparse. Comparatively, model-based CF methods are better in generalization and representation ability. Among model-based CF methods, matrix factorization (MF) is simple yet effective, and almost one of the most widely applied methods in industry. Despite the effectiveness, MF still lacks expressiveness in using inner product operation to model the user-item interaction. He *et al.* [58] proposed Neural Collaborative Filtering (NCF) which uses multilayer perceptron (MLP) to model the user-item interaction. However, the MLPs are not necessarily better than dot product for being the similarity function [138].

Another direction to improve the MF is to leverage more side information from both user and item sides rather than only using IDs. To this end, Factorization Machine (FM) [137] was proposed to model the feature interactions between an arbitrary number of entities. It achieved great success as it models second-order or even higher-order correlations between features to facilitate the user-item interaction model. Recently, the neural version of FM, NFM, was also proposed which applies the bilinear interaction pooling operation and stacks MLP to capture the non-linear relationship of users and items [55]. In real-world scenarios, rich heterogeneous side information is available, especially in the fashion domain, where abundant item attributes can be obtained by fundamental fashion attributes recognition tools. Studies have been conducted to incorporate user meta-data such as age, occupation and gender, and item meta-data such as price and categories through different latent factor models [154, 121, 176, 126]. However, few existing methods are able to dig into detailed user preference in terms of specific aspects of items such as a certain brand or style. This kind of specific user preference naturally exists in users' fashion shopping behaviors, which is important and would be helpful for overall user fashion taste modeling.

Another main stream of personalized recommendation research respects the order of the user behaviors, and tries to predict the next item that the user might be interested in. Compared with the non-sequential methods we have introduced above, the sequential recommendation methods are able to model the user-specific transition of interaction with items to master both the users' general interest and short-term preference [86, 131, 118]. A typical type is hybrid method which combines the sequence-learning methods with factorization-based matrix-completion techniques. The first hybrid sequential recommendation method is Factorized Personalized Markov Chain (FPMC) [136], which can be considered as a first-order Markov Chain whose transition matrix is jointly factorized with a standard two-dimensional user-item matrix factorization approach. In recent years, many deep learning-based methods have been proposed as well [191, 60, 59]. However, such sequence-aware recommendation studies has not attracted enough research focus in the fashion domain yet.

2.3.2 Fashion Compatibility Learning and Outfit Recommendation

Fashion compatibility learning is a domain-specific task which is also popularly studied in recent years. It is also termed as mix-and-match relationship modeling, cross-category fashion recommendation [190]. Some studies have focused on item-item compatibility modeling and cast it as a metric learning problem. In general, they try to project the fashion items into a latent space in which the representations of compatible items are close to each other [54, 165, 163]. To this end, different interaction functions were applied as pairwise compatibility metric, including data-independent functions such as inner product [153], Euclidean distance [115], data-dependent ones such as category-aware conditional similarity measures [163, 190] and others [189, 20]. Based on pairwise compatibility modeling, the compatibility of the whole outfit can be obtained accordingly [152]. Another group of outfit compatibility modeling approaches takes the whole fashion outfit as a set or an ordered sequence. Han *et al.* [49] employed the bidirectional LSTM to model the items in the outfit and explore the compatible relations between items in the matching outfit. However, such an approach was argued not reasonable to model items in order. To overcome this limitation, Chen *et al.* [14] proposed to use transformer [164] to model the item combinations, which achieved preferable performance. Cui [21] used the graph representation for the outfit modeling, which was also non-ordered and claimed to be more suitable to reflect the dense and complex relations among multiple items in an outfit.

Recently, outfit recommendation has attracted considerable attention in the area of fashion recommendation. The most common idea is to model the interaction between user and whole outfit under the condition that items in each outfit are compatible. Hu *et al.* [66] proposed a tensor factorization approach for the personalized outfit recommendation. Lin *et al.* [99] emphasised importance of visual analysis in fashion recommendation and employed the generative model to improve the visual understanding in outfit recommendation. Chen *et al.* [14] used user browsing history to model the user preference and addressed the outfit compatibility modeling and personalized outfit recommendation tasks all together with a transformer model. Lin *et al.* [100] formulated the fashion outfit recommendation as a multiple-instance-learning problem and proposed a two-stage neural network model, OutfitNet, to learn the compatibility between fashion items and users' taste for fashion outfits in two stages. Li *et al.* [94] proposed a hierarchical fashion graph network to integrate the information of the items composing the outfit to facilitate the representation of the fashion outfit, and then deployed the CF framework for the personalized outfit recommendation.

Capsule wardrobe creation is a special case of fashion outfit recommendation, which aims to maximize the mix-and-match popularity with minimum collection of fashion items in the wardrobe. This interesting task proposed by Hsiao *et al.* [65] can be regarded as an outfit compatibility evaluation problem which essentially tries to increase the compatibility of each fashion mix-and-match from the created capsule wardrobe. Further research on capsule wardrobe creation leveraged more attributes of people, such as body shape and personal preference [27]. The difference between capsule wardrobe creation and general outfit compatibility modeling is that the former needs to maximize the compatibility of multiple outfits simultaneously from the same group of item candidates whereas the latter evaluates each outfit independently.

2.3.3 Personalized Fashion Recommendation

Outfit recommendation has practical value in the fashion domain because it can offer a package of fashion suggestions and the mix-and-match advice. The item-level personalized fashion recommendation is also important especially for retailers. In most applicable scenarios, such as online shopping, users do not usually buy the whole outfit at a time. In previous works, researchers have realized that personalized fashion recommendation is not only applying available personalized recommender systems in the fashion domain but also considering the characteristics of the fashion domain [193]. The uniqueness of fashion recommendation lies in several aspects. First, the purpose of the recommendation is not only help users in their decision making for shopping, it can also provide styling suggestions to improve other people's taste in fashion. Second, compared with other domains such as books, music or video, the number of fashion items is extremely large while the interaction between user and fashion items can be very sparse. Third, specific characteristics in the fashion domain make some factors more important in fashion recommendation rather than others, such as aesthetics and popularity.

Kang *et al.* [79] built a visually-aware personalized recommender system based upon Bayesian Personalized Ranking (BPR) [135] and Siamese networks [46]. They further combined GAN to synthesize new item images to maximize personal objective value. Hwangbo *et al.* [69] built a recommender system for a real-world fashion e-commence platform based on typical item- and CF-based algorithm. They reflected several fashion domain characteristics such as seasonal preference change, complementary purchase. To help users search and find fashion products matching their taste among an increasing number of items, Ok *et al.* [122] proposed to derive implicit ratings from user log data and generated predicted ratings for item clusters through user-based CF. Zhang *et al.* [197] considered the influence of fashion bloggers on users' fashion purchase and proposed a model to learn personal implicit visual influence funnel from fashion bloggers to users for fashion recommendations. Liu *et al.* [102] emphasised *style* in their recommendation method. They proposed the DeepStyle method which extracted style features from item images and then incorporated them in the Visual BPR (VBPR) framework for personalized recommendation. Considering the importance of aesthetics in fashion, Yu *et al.* [193] proposed an aesthetic-based clothing recommendation method which employed another aesthetic evaluation model to extract aesthetic features from clothing images and further leveraged the feature in a CF-based frameworks.

The proceding review of existing works shows that even though visual information has been emphasised, most of the existing item-level personalized fashion recommendation approaches are still based on the CF framework. The improvements are limited in leveraging extra fashion-related features. None of them dig deep to analyze the fashion shopping behavior and the underlying fashion shopping patterns. Furthermore, most previous works assume that the user preference/taste in fashion to be static and few of them try to consider that the users' interests changing with time as in a practical situation. These works also do not take advantage of characteristics of fashion domain to enhance the sequential continuity modeling of user behavior to better explore user preference.

2.4 Fashion Trend Forecasting

Conventionally, fashion trends are envisioned and forecasted by fashion experts who examine the world around them - from culture, business, and arts to science and technology [36]. However, human-centric methods are usually inefficient, expensive, labor-intensive, highly dependent on the experts' background and may be biased because of personal preference. Due to the limitation of human-centric methods, in recent years, researchers have started to seek alternative data-driven method of fashion trend forecasting. Compared with other computational fashion tasks, fashion trend forecasting has not attracted as much research attention. Despite its research significance, fashion trend forecasting based on big data and DL technologies is still at its early stage and worth of exploration.

Hidayati *et al.* [61] studied runway fashion and tried to discover trendy patterns from fashion shows, the top fashion of which is far away from ordinary people's fashion preference. However, the problem is that most items shown in fashion shows did not become popular among the masses of people, which means they did not necessarily become trendy eventually. Vittayakorn *et al.* [167] later extended the task to a larger dataset and studied both runway and real-world fashion to produce quantitative analysis of fashion and trends. Al-Halah *et al.* [3] proposed of forecasting the fashion trends of specific styles based on Amazon's online shopping records. However, the online shopping activities may not reflect the real fashion trends as the purchase decisions are affected by multiple factors. Thereafter, Mall *et al.*proposed to forecast the fashion trends based on Instagram data by crawling millions of Instagram posts from 44 cities from around the world and analyzing the data with statistical models [113]. They modeled the fashion trend signals of each target element with a basic combination of linear and cyclical components, which were capable of capturing both coarse-level trends and fine-scale spikes. However, the limitation is that they only targeted a limited number of fashion elements which showed simple patterns in their trend signals (such as wearing hats or not) and did not include fine-grained fashion elements.

Moreover, statistical models are still the mainstream methods in current fashion trend modeling. However, in most cases, fashion trend signals can be very complex and with complicated patterns which are not easy to capture using traditional statistical methods such as linear regression or exponential smoothing [3, 113]. In the meantime, deep neural networks, especially recurrent neural networks, have shown superior performance in many time-series prediction tasks such as stock price prediction [33], sales forecasting [6] and others [31, 97, 130]. RNNs are promising to achieve preferable performance in fashion trend forecasting task but has not been studied.

2.5 Summary

The current research status and the main limitations of personalized fashion recommendation and fashion trend forecasting in the literature are summarized as follows.

Personalized Fashion Recommendation

(1) Personalized fashion recommendation has not attracted as much research attention as other recommendation tasks in the fashion domain such as outfit recommendation or compatibility modeling. Previous works on personalized fashion recommendation have not sufficiently combined the characteristics of fashion.

(2) Previous research on personalized fashion recommendation was limited to leveraging extra fashion-related features. None of them dig deep to analyze fashion shopping behaviors and the underlying patterns.

(3) The dynamic sequential continuity is rarely explored in the research on recommendation in the fashion domain. Majority of existing related works studied the non-sequential fashion commendation problem, which only focused on the static long-term user preference but neglected the interest drift of users over time.

Fashion Trend Forecasting

(1) Data-driven fashion trend forecasting is still at its nascent stage and has not attracted much research attention so far.

(2) The research targets in most existing works are not qualified to reveal the real fashion trends. Some studies were based on e-commence data, in which the popularity of fashion items are not mostly determined by the fashionability of items or the fashion trends. Some studies focused on the trends of specific fashion elements but with very simple and cyclic trend patterns, which are usually not related to fashionability, but related to seasonality instead.

(3) Existing works still applied traditional statistical approaches to analyze the time-series fashion trend signals. These approaches might have achieved preferable performance in modeling trends with simple patterns, but they fall short in making sound predictions for more complex trends.

3 Research Methodology

Towards effective fashion advising, two key perspectives are considered in this thesis: personalization and fashionability. Two specific research tasks are studied corresponding to the two perspectives, the personalized fashion recommendation task to explore the personalization, and the fashion trend forecasting task to explore the fashionability. As illustrated in Figure 3.1, for personalized fashion recommendation task, two specific aspects are emphasized, which are user preference modeling based on long-range fashion purchase data and content-level relational transition modeling for user behavior based on short-range fashion browsing data. The fashion trend forecasting research focuses on developing effective fashion trend modeling methods to predict the trends based on social media fashion-related data.

The study of fashion advising in this thesis is totally data-driven, and so are all the sub-tasks. Therefore, the main technical challenges are twofold. The first one is to choose proper and valid data for the specific research purpose. The second is to develop an effective deep learning-based approach to analyze the available data and mine effective information to address the corresponding research tasks. In the following part, the specific methodology for each sub-research task is introduced from the perspectives of tackling the two challenges.

3.1 Shopping Pattern Modeling for Personalized Fashion Recommendation

User preference is the main research focus for all personalized recommendation tasks which determines the decisions and behaviors of the user in purchasing.

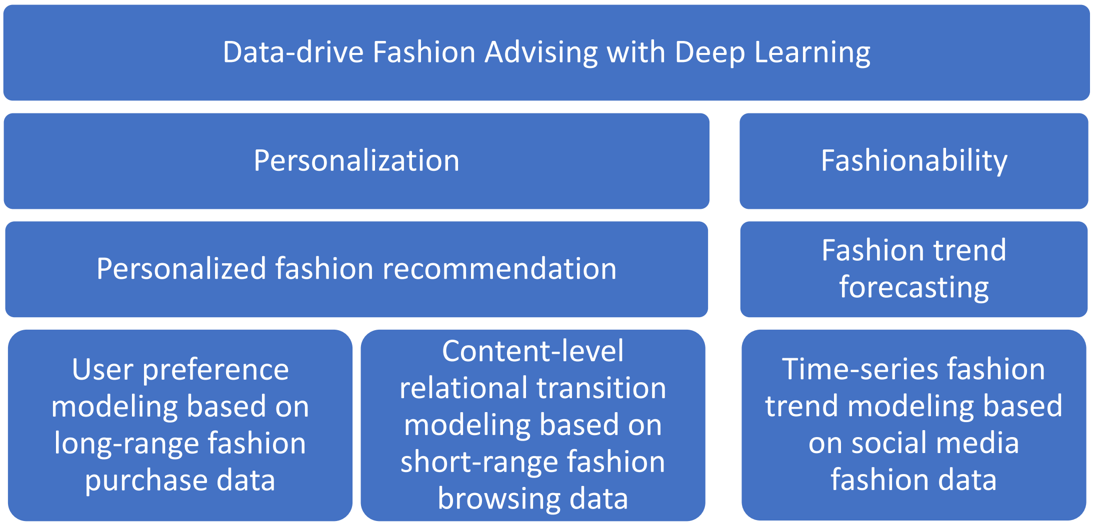

Figure 3.1: Research structure of data-driven fashion advising task with deep learning. The task has two key research perspectives and can be investigated from three important directions.

Different from other domains, user decisions in the fashion domain could be more diverse [107, 85], and more dependent on specific aspects of products, such as style, price or brand. Through observing real-world fashion shopping data, it has been found that diverse user preferences and specific shopping patterns are prevalent in the fashion domain. As a result, effectively modeling the diverse user preferences is key to fashion recommendation systems. In other words, we have to understand the shopping behaviors of different users, discover various shopping patterns and then derive preferable recommendation strategies. Therefore, the first research goal of this thesis is to discover personalized shopping patterns from historical user behavior for fashion recommendation. Specifically, beyond modeling the holistic user preference, it aims to predict fine-grained user fashion preferences in different factor fields [78], such as brand, price, style, color, which reflect specific user shopping patterns.

3.1.1 Data Preparing

Data on fashion shopping is used in the the study of user fashion preference and shopping patterns in developing a personalized fashion recommender system. The fashion shopping records contain information on users and items as well as user interaction feedback. It is the basic data usually used to analyze the user preference in developing recommender systems. Additional information is needed to investigate the detailed shopping patterns, which are related to specific attributes of items.

Different groups of influential attributes are defined as factor field in this study, which contains attributes describing the item from certain aspects (defined as factor). Considering the characteristics of fashion shopping, five factor fields are considered in analyzing the shopping patterns, which are *Color*, *Style*, *Brand*, *Price* and *Category*. In the data preparation, the factor information is obtained by analyzing the textual descriptions and images of the fashion item.

3.1.2 Approach

The **F**ield-aware **G**raph **C**ollaborative **F**iltering (FGCF) model is proposed to discover fine-grained personalized shopping patterns from historical user behaviors by considering the factor-field-level interactions for fashion recommendation. Compared with factor-based methods, such as FM [137], the factor field-based methods consider the difference between fields rather than treating all factors the same, which is therefore more promising to capture the underlying shopping pattern [78, 123]. The proposed model embraces three main parts: (1) a factor embedding layer for initializing factor embeddings and generating initial embeddings for factor fields, (2) factor field embedding propagation on a fully-connected graph, and (3) prediction layers for producing factor field-level and holistic interaction scores.

Specifically, the users and items are firstly represented as a set of factors from multiple influential factor fields in the factor embedding layer. To enable the higher-order embedding interaction between different factor fields, FGCF introduces a fully-connected graph $G = \{\mathcal{V}, \mathcal{E}\}$ whose nodes correspond to the factor fields of users and items, respectively. Each edge connects two different factor fields. The message passing is then conducted along the graph structure to model factor-field level embedding interaction. FGCF stacks multiple embedding propagation layers to explore the higher-order interaction between different factor fields. The node embeddings in the last layer are utilized as the final representation of each factor field. Finally, the aggregated holistic prediction is used to calculate the loss to train the entire recommendation model.

3.2 Content-level Relational Transition Modeling for Sequential Fashion Recommendation

Besides the user-item interaction modeling (main purpose in 3.1), capturing transition relationships between pairs of adjacent items in sequences (*i.e.*, sequential dynamics) is also important in predicting users' subsequent actions. The importance of transition modeling has been demonstrated in many existing sequence-aware recommendation works [28, 131]. However, such sequential dynamics modeling has not attracted much research attention in fashion recommendation yet. This type of modeling is important in fashion recommendation research, especially for application scenarios such as random browsing in online fashion shopping.

One typical sequential recommendation method, TransRec [52], proposes to employ a personalized translation operation to model the third-order interaction between user u, the previous interacted item i and the next item j. It has been validated to be effective and shown superiority in many cases than other basic methods which models user preference and sequential continuity of user behavior separately [52, 92]. Although existing sequential recommendation methods, such as TransRec, can be directly applied in the fashion domain, they cannot achieve preferable performance for two main reasons. On the one hand, the existing methods can be further improved technically. On the other hand, in the fashion domain, effective methods should incorporate the characteristics of fashion.

3.2.1 Data Preparing

For the study of sequential fashion recommendation in this section, short-range online fashion shopping data (randomly browsing) is applied. As the research focus of this part aims to effectively capture the sequential dynamics of the user behavior, it is more significant for continuous behaviors. Comparatively, purchase data has less continuous patterns compared with browsing data which record a short period of user behavior and in which the short-term dynamic patterns are more important to capture. A large-scale fashion shopping dataset POG from e-commence platform Taobao [14] is specifically adopted. This dataset contains the ordered interaction records of users with different items in a short period, which provides the basic sequential recommendation data for this study.

To generate a benchmark dataset for fairly evaluating different sequential fashion recommendation methods, sliding window strategy is applied in the original dataset to generate more short user-item interaction sequences with fixed sequence length. Moreover, to support the content-level item-item transition modeling in the proposed method, the associated attributes for each item are also extracted. A commercial fashion tagging tool[1] is specifically applied to extract three types of fashion elements (category, attribute and style) from the item images. A total of 225 different fashion element values (such as *dress*, *red* and others) are extracted in the entire dataset, which can be categorized into 24

[1] visenze.com

element groups based on a certain fashion taxonomy[2]. Element groups include *category*, *style* and specific attribute groups such as *color*, *pattern*, *neckline style* and *dress shape*.

3.2.2 Approach

To enhance the item-item transition modeling and further improve the performance of sequential fashion recommendation, the **A**ttentive **C**ontent-level **T**ranslation-based Fashion **R**ecommender (ACTR) is proposed. The proposed ACTR model is original from TransRec but innovatively combines the characteristics of fashion domain and leverages the content-level relational transitions. It contains three key components: item relationship prediction, content-level item-item transition modeling and recommendation based on mixture of transition.

It firstly leverages external relationships between fashion items to enhance the item-item transition modeling by predicting the next item relationship based on the user and previous item and based on that to model the relational item-item interaction. Inspired by the idea of factor field feature interaction [78], given the predicted relationship, the item-item transition is modeled by the combination of the sub-transition of specific fashion attribute field (such as *category*, *color* and others) between two items. A user-aware attention mechanism [166] is further introduced to organically combine different sub-transitions regarding different fashion content. Such a design is able to model the different levels that indicate the different fashion content which users are interested in. The final recommendation results are based on the mixture of the user-item transition modeling and relational context-level item-item transition modeling.

[2]ViSenze taxonomy is used here.

3.3 Fashion Trend Modeling Based on Social Media Data

Data-driven fashion trend forecasting is still at its nascent stage. At this stage, two main challenges have to be addressed for this task. First, although fashion trends can be an abstract topic which covers a wide range of fashion-related concept, in the forecasting task, the research target, that is, the specific fashion trends should be well defined. In previous works, the trend of some fashion elements that are less fashion-related are studied, such as *wearing jacket*. Such research objects are actually more season-sensible rather than fashion-sensible, which cannot truly reveal the fashion trend. Towards meaningful fashion trend forecasting, this work proposes to study the fashion trend of fine-grained fashion elements in specific groups of people. For example, the trend of *pattern:stripe* among *New York women aged 18-25*.

The second challenge is how to effectively model historical fashion trend signals and conduct forecasting accordingly. This problem can be grouped as a time-series modeling and prediction problem, which is usually addressed by some classic statistical models. However, the fashion trend signals can have highly complex patterns that are not easy to capture by traditional methods. Some fashion trends may be seasonal (such as *neckline style: turtleneck*) while some are not (such as *color: blue*). Moreover, meaningful fashion trend forecasting requires predicting trends for a certain period, not only for one time point, which makes the task even more challenging. To properly address these challenges, a more powerful model is necessary.

3.3.1 Data Preparing

Selecting the data source choosing is important part in fashion trend forecasting. Social media data are finally selected as the research target for two main reasons. First, it sensitively and extensively records the fashion development with massive uploaded fashion-related images and comments everyday from multiple sources of end users, fashion bloggers and brands, *etc*. Second, rich information for both users and fashion items can be extracted from the images, metadata and other source data through well-developed computer vision or other machine learning techniques. As existing datasets cannot support this research, either from other data source (e-commence) or with less meaningful research targets, a new dataset called Fashion Instagram Trend (FIT) is contributed which is based on the widely used social platform Instagram.

To build the FIT dataset, user accounts from 14 main cities across the world are crowded, of which all pictures in the posts that uploaded from July 2014 to June 2019 are downloaded. The dataset spans five years, which is long enough for the fashion trend analysis. The information from the user and fashion item sides is properly extracted from images as well as metadata. The popularity of certain fashion elements for certain user groups over the time generates a specific fashion trend signal. In the end, the whole dataset generate around 8000 specific fashion element signals corresponding to fine-grained fashion elements and user groups. All generated fashion trend signals in FIT are regarded meaningful to reveal the real trends followed by large groups of people. The research goal is to effectively model these trend signals, explore important patterns with the proper approach, and therefore make accurate predictions.

3.3.2 Approach

To perform accurate data-driven fashion trend forecasting, we have to capture the underlying patterns in the historical fashion trend time series. Although traditional models such as exponential smoothing or linear regression have been effectively applied to model simple time-series data [3, 113], they fall short of hands in making sound predictions for more complex trends. Recent advances in deep learning have provided great solutions for many tasks [90]. In particular, recurrent neural networks (RNNs) have demonstrated its superiority in modeling time-series data [19, 180, 31]. However, such approaches have not been employed in the area of fashion trend analysis yet. From another perspective, most existing works model the pieces of fashion trend signals independently. However, according to both commonsense and fashion theories, the fashion trends are not independent but well-correlated with each other. For example, the fashion element *turtleneck* is an affiliated attribute to the fashion element *sweater*, therefore, the trends of these two elements should be closely related to each other. Similar relations also exist among user groups. These relations can be helpful for the forecast of the trends, yet were ignored in most existing models.

The solution proposed in this thesis is to employ the LSTM encoder-decoder as the basic model for the fashion trend forecasting task which leverages fashion element information and user group information. Two message passing modules are introduced in the model to explore the influence of correlated fashion elements or user groups in trend modeling. As the task is to forecast a period of time of future trend, it is more challenging than only one-step-ahead estimation. To effectively capture temporal patterns on future horizons, the proposed model is equipped with a sliding temporal attention module [31]. Specifically, at each time step, the decoder hidden state is taken to attend to several different periods of the history and generate the attention vector individually. The representation of the decoder hidden state is then updated by combining the attention results of all periods of the history. Therefore, the combined features can better describe the current time step as it incorporates both historical and future contextual information. The proposed model is called **R**elation **E**nhanced **A**ttention **R**ecurrent network, short in REAR.

3.4 Summary

This chapter presents the methodology of applying deep learning methods to address the data-driven fashion advising task. Specifically, towards three specific research objectives illustrated in Section 1.2, three deep learning-based models are proposed, namely, FGCF model for personalized fashion recommendation, ACTR for sequential fashion recommendation, and REAR for fashion trend forecasting.

4 Learning Personalized Shopping Patterns for Fashion Recommendation

4.1 Introduction

The fashion industry, ranging from global discount retailers to exclusive luxury brands, drives a significant part of global economy. It is now undergoing large-scale digital transformation, creating a growing demand for supportive technologies in online fashion retails, such as personalized fashion recommendation. Fashion recommendation system aims to recommend the suitable fashion items to users by modeling users' shopping preference based on historical behavior, which has received increasing research attention [14, 15, 64, 161] from both the industry and academic community in recent years.

Different from other domains, user decisions in the fashion domain could be more diverse [107, 85] and more dependent on some specific aspects of products, such as style, price or brand. Goswami and Khan [41] investigated the influence of consumer-decision making on online apparel shopping and stated that shoppers who are highly fashion and brand conscious are more inclined to buy stylish clothes, whereas value-conscious buyers would look for price benefits and best offers. For a better illustration, the upper part of Figure 4.1 provides an example in the fashion domain from the Amazon dataset [88]. The example reveals three representative shopping patterns from a given user's purchase history with regard to three user-specific aspects: category (*i.e.*, *shoes*), color (*i.e.*, *black*) and style (*i.e.*, *elegant*). Such diverse user preferences and specific shopping patterns are

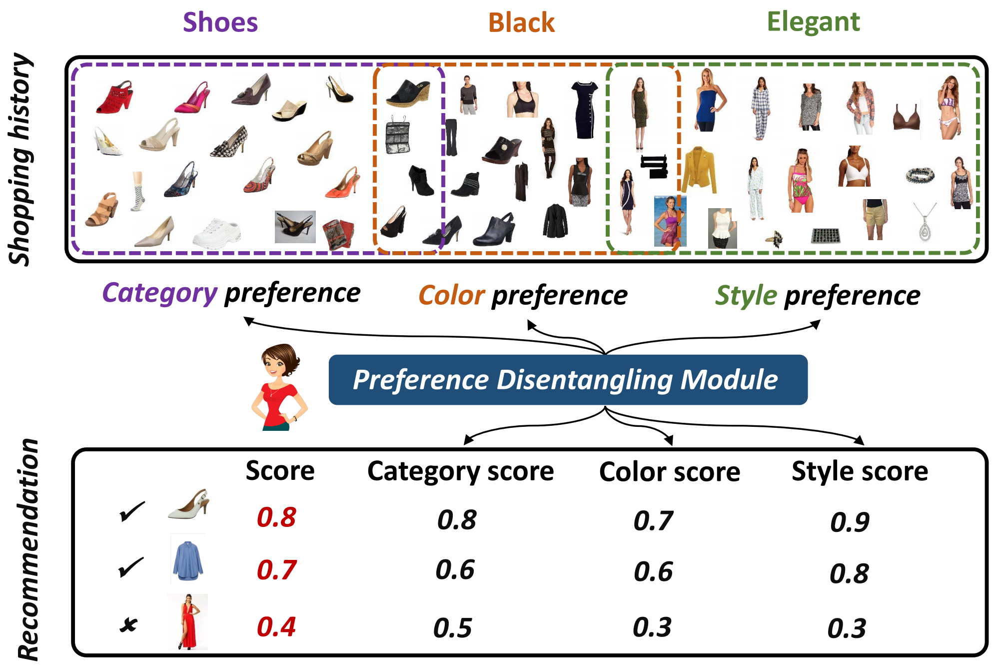

Figure 4.1: Illustration of diverse fashion shopping patterns and predicted user fashion preferences in various aspects of fashion recommendation

prevalent in the fashion domain. As a result, effectively modeling the diverse user preferences is key to fashion recommendation systems. In other words, understanding the shopping behaviors of different users is important for deriving preferable recommendation strategies.

Most existing efforts have primarily leveraged the visual appearance of fashion items to derive visually-aware fashion item representation, such as global image embedding [161, 53] and region-aware local representation [15, 64], before modeling the user-item interactions based on the inner product of user and item embeddings or some data-dependent interaction functions [56, 55]. Despite the progress, most existing fashion recommendation methods directly model the holistic user preference but do not reveal the diverse user preferences under different aspects of fashion items. Such operation leads to less convincing recommendation results because of the lack of interpretability.

To fill the research gap, this work develops a novel graph-based fashion recommendation method, named Field-aware Graph Collaborative Filtering (FGCF), which aims to capture fine-grained user shopping patterns. As influential factors in fashion shopping are categorical, these factors (*e.g., black, elegant*) are grouped into several factor fields (*e.g., color, style*) and field-level operations are further conducted, such as interaction modeling. Compared with factor-based methods, *i.e.*, FM [137], the factor field-based methods consider the difference between fields rather than treating all factors the same, which is a more promising approach to capture the underlying shopping pattern [78, 123].

As shown in the lower part of Figure 4.1, our goal is not only to predict the holistic user preference but also to infer the specific user fashion preferences in different factor fields. To achieve this, three main components are equipped in the proposed FGCF model: (1) a factor embedding layer that offers an initialization of low-dimensional embeddings of different user and item factors, and composes the embeddings of multiple predefined factor fields; (2) multiple factor field embedding propagation layers that refine the embeddings of factor fields on a fully-connected graph whose nodes are different factor fields of users and items; and (3) multiple prediction layers that aggregate the refined embeddings of factor fields, output the field-specific interaction scores next, and finally integrate the field-specific scores into the holistic score for pairwise training. By the proposed model, fine-grained user shopping patterns from diverse user historical behaviors are expected to be disentangled and therefore better fashion recommendations can be achieved. Extensive experiments on the large-scale dataset Amazon have demonstrated the effectiveness of our method.

4.2 Related Work

Most of user-preference based fashion recommendation works are based on general recommendation models. The most classical model, CF make recommendations by learning from user-item historical interactions, works based on CF and its family models include [67, 193, 44], such as previous shopping records.

FM is another effective basic recommendation model, which learns implicit latent representation for each feature and then considers all single and pairwise interactions of features to model the predictions of the user-item interactions. Unlike CF models which only exploit information from user and item IDs, FM can handle multiple tags or features, so that more side information is exploited and the recommendation performance can be improved. However, the FM model only considers the first-order interaction between features, which cannot fully exploit the rich information in real-world data with complex structures. The extension of FM, including attentional FM (AFM) [185] and neural FM (NFM) [55] have made improvements in this perspective, but they do not consider the effect across different features, which degrades the expressiveness of the model. Moreover, the nonlinear projection process is implicit, which causes the model to lack interpretability. Being capable of modeling high-order connective between users and items, graph network [84] has also been applied in personal recommendation. Combined with CF models, the graph model takes the users and items as the nodes in the graph and the high-order connectives can be modeled [177, 175]. However, existing works are limited in expanding connectives between users and items, and few studies have considered the connectives beyond user and item, for example, the specific attributes of items.

4.3 Problem Formulation

The goal of this work is to discover personalized shopping patterns from historical user behaviors for fashion recommendation. Specifically, beyond modeling the holistic user preference, we aim to predict fine-grained user fashion preferences in different factor fields [78], such as *brand*, *price*, *style* and *color*, which reflect specific user shopping patterns. For easy understanding, the specific task in this work is defined in this section.

Let u denotes a user from the entire user set $\mathcal{U} = \{u_t\}_{t=1}^{N_u}$ and i denotes an item from the whole item set $\mathcal{I} = \{i_t\}_{t=1}^{N_i}$. The interaction set between the users and items is defined as $\mathcal{R} = \{(u,i)\}$, which describes the historical shopping behaviors of users. In addition to the ID or dense visual features that distinguish a user or a fashion item, the associated influential *factors* in different *factor fields* [78] are exploited to represent the fashion items and users for the purpose of capturing fine-grained user fashion preferences. Formally, x, $\mathcal{X}^f = \{x_t^f\}_{t=1}^{Vf}$, and $\mathcal{X} = \{\mathcal{X}^{f_k}\}_{k=1}^{M}$ denote a factor, the factor set of a factor field f, and the whole factor field set respectively. A user and an item, with the associated factors are denoted as $\mathcal{X}_u = \{\mathcal{X}_u^{f_k}\}_{k=1}^{M_\mathcal{U}} = \{x_1^{f_k}, \cdots, x_{V_u^k}^{f_k}\}_{k=1}^{M_\mathcal{U}}$ and $\mathcal{X}_i = \{\mathcal{X}_i^{f_k}\}_{k=1}^{M_\mathcal{I}} = \{x_1^{f_k}, \cdots, x_{V_i^k}^{f_k}\}_{k=1}^{M_\mathcal{I}}$ respectively. Note that a factor field could be one-hot or multi-hot. For example, the factor field *Color* usually has only one factor (*i.e.*, value) for each fashion item, wheres the factor field *History* could have multiple factors for each user. Both the user ID and item ID can be treated as specific factor fields to capture the collaborative signal. Thus, the the problem is formulated as follows:

- **Input:** The user factor sets $\mathcal{X}_\mathcal{U} = \{\mathcal{X}_{u_t}\}_{t=1}^{N_u}$, the item factor sets $\mathcal{X}_\mathcal{I} = \{\mathcal{X}_{i_t}\}_{t=1}^{N_i}$, and the user-item interactions $\mathcal{R}$.

- **Output:** A predictive model which outputs not only the overall interaction score y_{ui} for a given user-item pair (u,i), but also specific interaction scores $\{y_{ui}^{f_k}\}$ in multiple influential factor fields of fashion products.

4.4 Approach

In this section, a Field-aware Graph Collaborative Filtering (FGCF) model is proposed to discover personalized shopping patterns from historical user behaviors for fashion recommendation. As illustrated in Figure 4.2, the proposed model embraces three main parts: (1) a factor embedding layer for initializing factor embeddings and generating initial embeddings for factor fields, (2) factor field

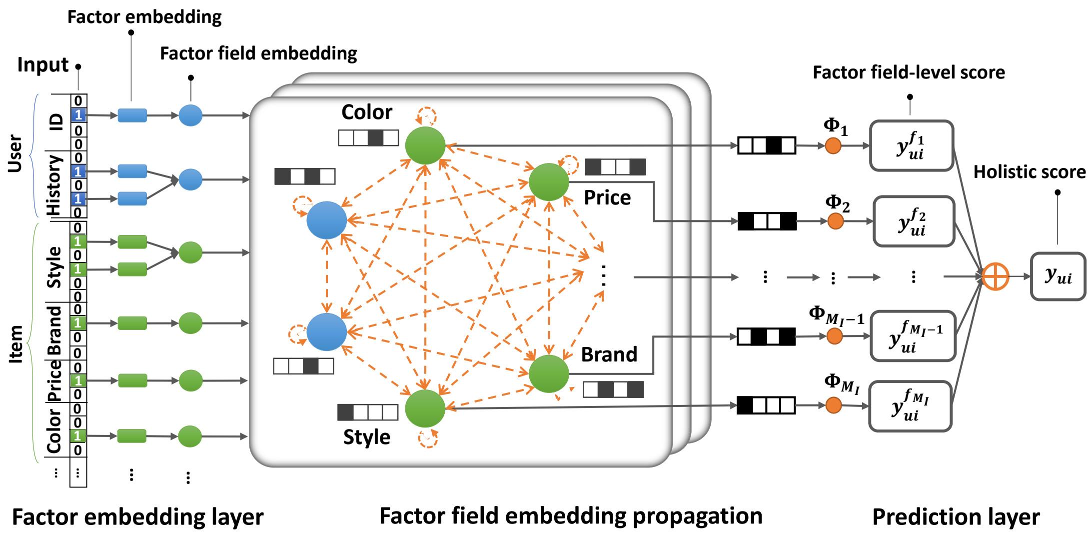

Figure 4.2: FGCF consists three main parts: factor embedding initialization, factor field embedding propagation on the graph and fashion preference prediction. The orange dots denote the MLPs for prediction.

embedding propagation on a fully-connected graph and (3) prediction layers for producing factor field-level and holistic interaction scores.

4.4.1 Factor Embedding Layer

Factor Embeddings: As mentioned in section 4.3, the users and items are represented as a set of factors from multiple influential factor fields. Following the strategies of mainstream recommendation models [185, 55, 177], a factor $x^f \in \mathcal{X}^f$ in a factor field f of a user $u \in \mathcal{U}$ (or an item $i \in \mathcal{I}$) is described by a low-dimensional dense embedding vector $\mathbf{e}_u^f \in \mathbb{R}^d$ ($\mathbf{e}_i^f \in \mathbb{R}^d$), where d denotes the embedding size. This idea can be easily implemented by an embedding look-up table: $\mathbf{E} = [\mathbf{E}_{\mathcal{U}}, \mathbf{E}_{\mathcal{I}}]$,

where $\mathbf{E}_{\mathcal{U}} = \{\mathbf{E}_{\mathcal{U}}^{f_k}\}_{k=1}^{M_{\mathcal{U}}}$ and $\mathbf{E}_{\mathcal{I}} = \{\mathbf{E}_{\mathcal{I}}^{f_k}\}_{k=1}^{M_{\mathcal{I}}}$ denote the factor embeddings in $M_{\mathcal{U}}$ user factor fields and $M_{\mathcal{I}}$ item factor fields, respectively.

Factor Field Embeddings: After initializing the factor embeddings of users and items, the next step is to obtain the initial representations of the user and item factor fields. One-hot factor fields, such as *color*, are represented directly as the embedding of the activated (nonzero) factor in this factor field. Multi-hot factor fields, such as *history*, are represented by the aggregation of the embeddings of the activated (nonzero) factors in this factor field with the average-pooling operation.

Based on the aforementioned solution, given a user-item pair (u, i), $M_{\mathcal{U}}$ user factor field embeddings $\{\mathbf{f}_u^k\}_{k=1}^{M_{\mathcal{U}}}$ and $M_{\mathcal{I}}$ item factor field embeddings $\{\mathbf{f}_i^k\}_{k=1}^{M_{\mathcal{I}}}$ can be obtained, which are denoted as $\mathcal{F}_{ui} = \{\mathbf{f}_{ui}^k\}_{k=1}^{M}$ for simplicity, where $M = M_{\mathcal{U}} + M_{\mathcal{I}}$. Thereafter, the $\mathcal{F}_{ui}$ will be fed into a fully connected graph for propagation, as described in section 4.4.2.

4.4.2 Factor Field Embedding Propagation on Graph

Graph: To enable the higher-order embedding interaction between different factor fields, FGCF introduces a fully-connected graph $G = \{\mathcal{V}, \mathcal{E}\}$ whose nodes correspond to the $M_{\mathcal{U}} + M_{\mathcal{I}}$ factor fields of users and items, respectively. Each edge connects two different factor fields. Motivated by the message-passing mechanism of GNN [198], FGCF performs message passing along the graph structure for factor-field level embedding interaction and updated from first-order propagation to higher-order propagation.

Layer-wise Embedding Propagation: The input to the graph G is a set of initial node (factor field) embeddings $\mathcal{F}_{ui} = \{\mathbf{f}_{ui}^k \in \mathbb{R}^d\}_{k=1}^{M}$. The embedding propagation layer updates the node embeddings as $\mathcal{F}_{ui}^{(1)}$, which tasks the following propagation rule as follows:

$$\mathbf{f}_{ui}^{k(l+1)} = \sigma\left(\lambda \mathbf{W}'^{(l)} \mathbf{f}_{ui}^k + (1-\lambda) \frac{1}{M} \sum_{j=1}^{M} \mathbf{W}^{(l)} (\mathbf{f}_{ui}^k \odot \mathbf{f}_{ui}^j)\right), \tag{4.1}$$

where the λ is a trade-off hyperparameter to balance the first-order and second-order signals. The $\mathbf{W}^{(l)} \in \mathbb{R}^{d_{l+1} \times d_l}$ and $\mathbf{W}'^{(l)} \in \mathbb{R}^{d_{l+1} \times d_l}$ are layer-specific trainable linear transformation matrices to distill useful information for propagation and d_l is the transformation size in the l-th layer. The σ denotes an element-wise activation function, such as Sigmoid$(\cdot)$ or Relu$(\cdot)$. Different from conventional graph convolutional networks which only consider the node-level embedding transformations $\mathbf{W}'^{(l)}\mathbf{f}_{ui}^{k}$, this method additionally leverages the interaction between two connected factor field nodes $\mathbf{f}_{ui}^{k} \odot \mathbf{f}_{ui}^{j}$ for message propagation and aggregation, where $\odot$ denotes the element-wise product.

Higher-order Embedding Propagation: FGCF stacks multiple embedding propagation layers to explore the higher-order interaction between different factor fields. Such high-order interactions are crucial to capture the factor field-aware user fashion preferences.

Compared with the classic factor-based methods, such as FM [137], the factor field embeddings updated by Eq. (4.1) represents the factor field-specific high-order interactions of the given user-item pair (u, i). According to a previous work [55], better modeling high-order interactions effectively helps capture the underlying nonlinear structure of data and mine more useful information. With multi-layer message propagation, higher-order interactions between factor fields are exploited, thereby effectively boosting the expressiveness of the model.

The node embeddings in the last layer are utilized as the final representation of each factor field, which are denoted by $\{\mathbf{F}_{ui}^{k} \in \mathbb{R}^{d'}\}_{k=1}^{M}$ where the d' are the embedding size of the final node embeddings.

4.4.3 Fashion Preference Prediction

Field-specific Preference: Given the final representation of each factor field after layer-wise embedding propagation on the graph, we can now predict the factor field-specific user fashion preferences.

Specifically, $M_{\mathcal{I}}$ predictors $h^k(\cdot)$ (only for item fields) are employed, which can be implemented by an MLP, to transform the final representation of each item factor field to a predictive field-specific preference prediction: $\hat{y}_{ui}^{f_k} = h^k(\mathbf{F}_{ui}^k)$, where $1 \leq k \leq M_{\mathcal{I}}$.

Holistic Preference: After obtaining the factor field-specific preference predictions, the holistic user preference $\hat{y}_{ui}$ is modeled as the aggregation of factor field-specific preference predictions: $\hat{y}_{ui} = 1/M_I \sum_{k=1}^{M_{\mathcal{I}}} \hat{y}_{ui}^{f_k}$, where $\hat{y}_{ui}$ is leveraged for pairwise training because the holistic supervision is available.

4.4.4 Model Training

Following previous works [135, 53], the pairwise Bayesian personalized ranking (BPR) loss is adopted in our approach to optimize the model parameters under the assumption that the observed interaction should have higher predictive score than unobserved interactions. The entire model is optimized by minimizing the following objective:

$$\min_{(u,i,j)\in\mathcal{O}} \sum -\ln s(\hat{y}_{ui} - \hat{y}_{uj}) + \eta \, \|\Theta\|, \tag{4.2}$$

where $\mathcal{O} = \{(u,i,j)|(u,i) \in \mathcal{R}^+, (u,j) \in \mathcal{R}^-\}$ denotes the pairwise training triples. $\mathcal{R}^+$ and $\mathcal{R}^-$ denote observed and unobserved interaction sets, respectively. Θ denotes all the trainable parameters and $s(\cdot)$ is the Sigmoid function. L_2 regularization is applied to avoid over-fitting.

4.5 Experiments

This section introduces the conducted extensive experiments on the real-world datasets to evaluate the effectiveness of the proposed FGCF model.

4.5.1 Dataset

A dataset is collected for this task based on the Amazon (Clothing, Shoes and Jewelry) dataset [88] and split into two sub-datasets: *Men* and *Women*[1]. The attributes (*i.e.*, *factors*) of items are mainly extracted from the *Color*, *Style*, *Brand*, *Price* and *Category* factor fields using their textual descriptions and product images. Specifically, brand and price information are directly extracted from the meta-data of the dataset.

The other attributes are extracted through the following processes:

- **Category factor extraction** Although category information is provided in the meta-data, it is of low quality with noise and coarse definition. In comparison, the category extracted from the product description is better defined and more accurate. As a result, the hierarchical keywords matching strategy is utilized to extract two levels of categories of items, coarse category and detailed category (denoted as Cate-C and Cate-D). Thereafter, manually check is conducted to further remove the noise in the extracted category information.

- **Color and style extraction** The color information of items is obtained by a commonly used tool called ColorThief [2], which can extract the dominant

[1] https://jmcauley.ucsd.edu/data/amazon/
[2] https://github.com/lokesh/color-thief

Table 4.1: Statistic of the two datasets

Dataset	#Interaction	#User	#Item
Amazon-Women	457,375	57,523	207,721
Amazon-Men	96,822	13,913	48,618

Table 4.2: Statistic of number of factors in each factor field in our datasets. Cate-C stands for coarse category and cate-D stands for fine-grained category.

Fields	Color	Style	Brand	Price	Cate-C	Cate-D
#Factor	30	11	1940	8	7/4*	51/24*

The former number is for Women dataset and the later is for Men dataset.

colors from the images of items. The style information is obtained by commercial fashion annotation tool ViSenze. [3]

Users with low interactions (those who purchased and reviewed less than 5 items) are further removed for the sake of data quality considering the interaction density. Table 4.1 lists the statistics of the two datasets, showing the number of users, items and interactions in each dataset. Table 4.2 presents the statistic of item factor collections. For both the Women and Men datasets, 80% of historical interactions of each user are randomly selected for training and 20% for evaluation. Each observed user-item interaction is treated as a positive instance, and paired with one random negative item that the user does not interact with. Two factor fields are employed for user representation: user *ID* and user *History* and seven factor fields for items: item *ID*, *Color*, *Style*, *Brand*, *Price*, coarse category *Cate-C*, and fine-grained category *Cate-D*. This work only considers the user fashion preferences in different item factor fields. The user factor field embeddings are only used for propagation and would not be applied for prediction.

[3] visenze.com

4.5.2 Experimental Settings

Baselines: Five representative recommendation models are selected as the baselines including MF [135], FM [137], NFM [55], AFM [185] and VBPR [53]. All baselines are reimplemented on our datasets and carefully tuned for the best performance, and are introduced in detail as follows:

- **Matrix Factorization (MF)** [135]: This is the basic Matrix Factorization model using BPR loss.
- **Factorization Machine (FM)** [137]: The factorization machine method takes all user and item information as features of an input interaction and can predict the score of the input by modeling second-order feature interactions.
- **Neural Factorization Machine (NFM)** [55]: Neural FM method is the state-of-the-art factorization method, which implements the FM with neural network.
- **Attentional Factorization Machine (AFM)** [185]: FM method with an attention mechanism.
- **Visual Bayesian Personalized Ranking (VBPR)** [53]: A state-of-the-art visual-based recommendation model which exploits extra visual features compared with the basic MF model.

Implementation Details: The proposed FGCF model and all baselines are implemented in Tensorflow. For all models, the optimizer is set to Adam optimizer, the batch size is set to 512, the embedding size is set to 64 and the Xavier initializer is selected to initialize all parameters.

Grid search strategy is applied to select the hyper-parameters for our FGCF model. The learning rate is searched in {0.00001,0.0005,0.0001,0.005}, and finally set to 0.0005 for both datasets. The l_2 regularization coefficient is set to $4e^{-6}$ for embedding, and $1e-4$ for other model parameters. The dropout ratio is set to 0.8 for both datasets. The default setting for depth of graph is two.

Evaluation Matrices: For evaluation, 99 negative items are sampled for each user in the test set [30, 58] from the candidate negative item pools which contain all unobserved items for the certain user. Five commonly used evaluation metrics are adopted to comprehensively evaluate the effectiveness of top-K recommendation and preference ranking [175, 58]: precision@K, recall@K, MAP@K, NDCG@K and MRR@K. K is set to 10 by default. The results for all matrices are reported averagely.

4.5.3 Overall Recommendation Performance

Tables 4.3 and 4.4 report the overall performance of the proposed FGCF model and baselines on two datasets, from which the following observations can be found:

(1) MF shows the worst performances on both datasets because it uses limited information, specifically user ID and item ID. VBPR is slightly better than MF for taking advantage of visual information.

(2) Comparatively, FM-family methods achieve better results because they exploit more attributes information. In particular, the NFM shows desired performance because it models the non-linear high-order feature interactions, and the attention mechanism adopted in AFM also helps improve the performance to identify the informative factor interactions.

Table 4.3: Overall Performance Comparison (Women)

	Precision	Recall	MAP	NDCG	MRR
MF	0.0390	0.3729	0.2136	0.2527	0.1540
VBPR	0.0424	0.4067	0.2179	0.264	0.1459
FM	0.0456	0.4367	0.2402	0.2841	0.1556
NFM	0.0457	0.4378	0.2405	0.2889	0.1625
AFM	0.0453	0.4332	0.2362	0.2844	0.1596
FGCF	**0.0471**	**0.4491**	**0.2473**	**0.2969**	**0.1688**
%Improv.	3.06	2.58	2.83	2.77	3.88

%Improv. denotes the percentage of the improvement compared to most competitive baseline.

Table 4.4: Overall Performance Comparison (Men)

	Precision	Recall	MAP	NDCG	MRR
MF	0.0274	0.2694	0.1371	0.1686	0.0886
VBPR	0.0343	0.3376	0.1699	0.2100	0.1054
FM	0.0367	0.3609	0.1868	0.2287	0.1185
NFM	0.0376	0.3695	0.1891	0.2324	0.1183
AFM	0.0371	0.3645	0.1876	0.2301	0.1177
FGCF	**0.0383**	**0.3764**	**0.1945**	**0.2381**	**0.1236**
%Improv.	1.86	1.87	2.86	2.45	4.30

(3) The proposed FGCF model consistently yields the best performance on both datasets evaluated. Compared with NFM that models the feature-level high-order interactions, FGCF captures more explicit high-order interactions in different item factor fields and then enables better expressiveness.

4.5.4 Effects of Embedding Propagation in FGCF

To investigate the effectiveness of embedding propagation (EP) and the effect of the graph depth on the overall performance, this section compares the performance of the FGCF model with different numbers of the embedding propagation layers, denoted by FGCF-K ($K \in \{0, 1, 2, 3\}$), as shown in Figure 4.3. In particular, $K = 0$ means predicting the field-specific user preferences using the original factor field embeddings. Figure 4.3 show the following observations:

(1) EP can significantly improve the performance because it helps enhance factor field representation by information exchange. It is reasonable because the field-specific user preference could not solely depend on the corresponding factor field, and is helpful to aggregate the second-order interactions between the factor field and other factor fields into the representation of this factor field by Eq. (4.1).

(2) Compared with the one-layer model, FGCF-2 achieves higher improvement because it models more higher-order factor field interaction, through which useful information from highly nonlinear data can be effectively exploited.

(3) The performance of FGCF-3 is worse compared with other settings, which shows that using too many EP layers might cause the over-smoothness of the node representation and performance degradation.

4.5.5 Analysis of Field-specific Fashion Preferences

This study aims to predict the factor field-specific fashion preferences from diverse user shopping behaviors. Each field-specific fashion preference is dominated by a latent user fashion shopping pattern. Detailed discussion is provided in this section to clearly illustrate the effectiveness of the proposed method. First, the performance of fashion recommendation in different factor fields of items is listed in Table 4.5, from which the following observations can be obtained: (1)

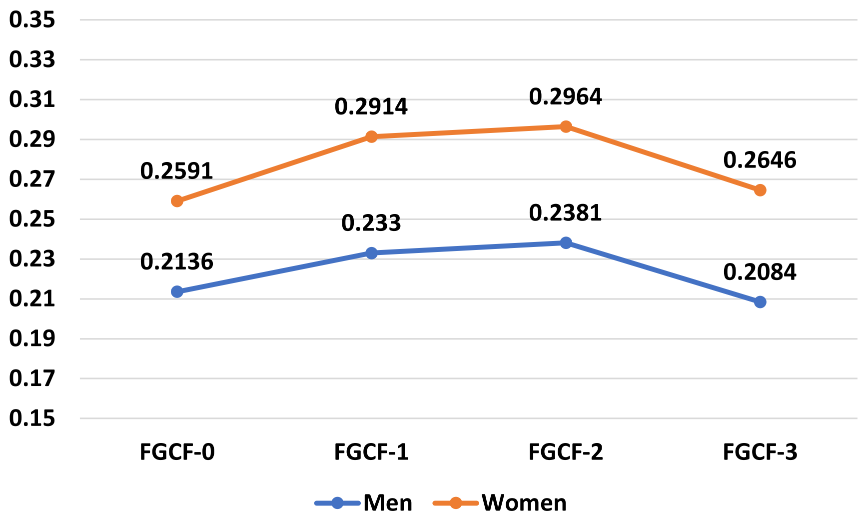

Figure 4.3: Performance comparison (NDCG) of FGCF-K model with different numbers (K) of embedding propagation layers in FGCF.

Overall, different item factor fields show diverse recommendation performance. Among the seven item factor fields, the *ID* field yields the highest performance. The result is reasonable because the ID embedding, which is widely applied in existing personalized recommendation works, has strong representation ability, which can effectively capture the collaborative filtering signal. (2) The *Brand* field performs the worst because more than 1,900 are in this field, and thus, the user preference in the field would be too diverse to capture. (3) Except the *ID* field, the *Cate-C* field yields the best performance on the Men dataset, whereas on the Women dataset, the *Style* and *Color* fields hit the top-2 best performance. This result shows that the consumers' shopping style could vary in gender, and usually, women are much more fashion-conscious than the men.

Then, two cases of disentangled factor field-based recommendation results (Top 10) are illustrated in Figure 4.4. The items in each row are ranked according to the corresponding factor field-level prediction scores. The first observation is that the ranking results in different factor fields are diverse and each ranking result shows distinct patterns (marked with squares in the same color). In the first case, the shopping pattern with regard to the factor field-*Style* (second row) is rather marked as 7 out of 10 items have the style *Casual*. A similar phenomenon can

Table 4.5: Performance of user preference prediction in different item factor fields. The performance of *Overall* prediction is listed at the bottom of the table as a baseline.

	Men		Women	
	Recall	NDCG	Recall	NDCG
ID	0.3545	0.2102	0.3846	0.2413
Color	0.1919	0.102	0.2821	0.1418
Style	0.2379	0.1263	0.2987	0.1632
Brand	0.1257	0.0735	0.0819	0.0572
Price	0.1640	0.0886	0.1664	0.0964
Cate-C	0.2608	0.1409	0.2132	0.1079
Cate-D	0.1775	0.0842	0.1981	0.1188
Overall	0.3764	0.2381	0.4491	0.2969

be observed from the factor field *Price* and *Cate-C*. Specifically, most top-ranking items based on price are in the price range *[50-100]* and *[30-50]*, and are shoes and bags if based on course category. Notably, the shopping pattern about color is not so obvious for this particular user, which implies that this user might not have a specific color preference. The second user, on the contrary, has preferred colors, which are *black* and *beige* according to the color-based ranking results. Our model also predicts that this user prefers *Casual* and *Feminine* fashion items. The most likely items he/she would buy are *Clothing* and *Underwear*, particularly *Pumps* or *Skirt*. From the preceding two cases and discussion, we can observe that certain shopping patterns related to specific factor fields are effectively mined by the proposed model. Another important observation is that most top-ranked items showing various shopping patterns are positive to the user, consistent with the true preference of the users (marked with a red tick), which demonstrates that our method can discover not only shopping patterns but also discover effective patterns.

Figure 4.4: Examples of top 10 recommendations based on factor field-level scores. Colorful squares show underlying patterns with regard to certain factor field. Ticked items are positive to the user.

4.6 Summary

This chapter focuses on the task of learning personalized shopping patterns for fashion recommendation, aiming not only to predict the historical fashion preference of users but also to reveal the diverse fashion preference with respect to specific aspects such as style, brand or price, that is, learn specific fashion shopping patterns. A novel graph-based model FGCF is proposed, which applies a fully-connected graph to effectively model the interaction of user and item related to specific factors through factor-field embedding propagation and aggregation. Extensive experiments on two datasets were conducted to evaluate the proposed approach for the specific task and the results show that the proposed FGCF model achieves better recommendation performance compared with baselines. More importantly, it effectively captures diverse fined-grain fashion shopping patterns from complex user purchasing behaviors.

5 Modeling Content-level Relational Transition for Sequential Fashion Recommendation

5.1 Introduction

The basic idea of most classic recommender systems is to model the compatibility between user-item pairs (i.e., user preference), such as MF [135] and its variants [53]. However, studies have shown that capturing transition relationships between pairs of adjacent items in sequences (*i.e.*, sequential dynamics) is also important in predicting user's next actions, such as purchasing or clicking [136, 52]. To explore sequential dynamics for better recommendation performance, researchers have devoted considerable efforts to sequential recommendation in recent years. A majority of sequential recommendation methods aim to model the relations between the user u, the item i that u recently picked and the item j, which would be picked next. A typical method, TransRec [52], proposes to employ a personalized translation operation to model the third-order interaction between u, i, and j. This method has been validated to be effective and has shown superiority in many cases than other basic methods which models user preference and sequential continuity of user behavior separately [52, 92].

The key idea behind TransRec is to model the transition process of user behaviors. Specifically, each user u acts as a translation vector, through which the previous item i is 'translated into' the next item j. However, such translation is determined by the user-item relations/interactions only and overlook the item-item relations, as illustrated in Figure 5.1 (a). Existing research has demonstrated that exploring

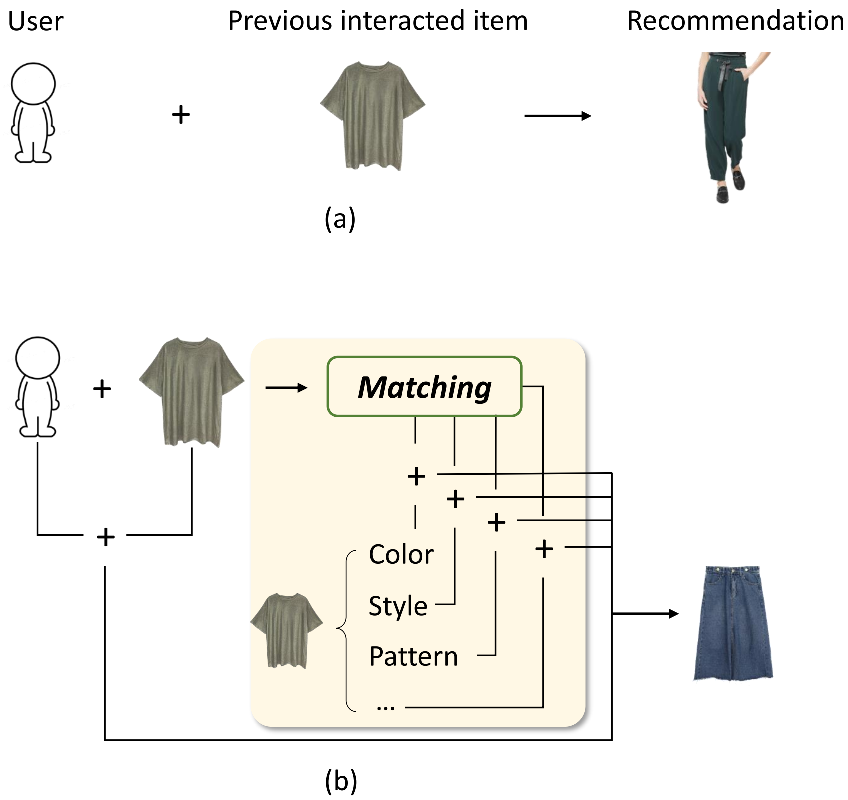

Figure 5.1: Instead of solely modeling single-component user-item-item transition as the basic translation-based sequential recommendation method does (showing in (a)), this study proposes to leverage item-item relationships (matching or substitution), which also indicate the user's intention for the next action. Based on the predicted relationship, the content-level transition is modeled to enhance the item-item interaction modeling, which accordingly improves the overall recommendation performance.

heterogeneous item relationships can help the item-to-item interaction model and facilitate the sequential recommendation results [80]. Particularly in the fashion domain, abundant contextual relationships exist between items. For example, two clothes with similar design details or the same style. Such item-to-item transition is particularly important in situations where the user behavior is highly continuous. For example, when doing online fashion shopping, most people click on items they find attractive continuously. In such circumstance, the users can been easily affected by the items he/she recently visited for their next click.

As we known, two common application scenarios exist for fashion recommendation: to recommend items for substitution and to recommend them for mix-and-match. These two recommenders are widely equipped by most fashion e-commence platforms, which reminds us of two most important relationships between fashion items, *substitution* and *mix-and-match*. Inspired by the idea of leveraging item-item relations in recommender system, this work aims to incorporate the two most common yet important relationships between fashion items into fashion recommendation, which can be helpful from two aspects. First, it is promising to improve the recommendation accuracy. Second, it makes the recommendation results more explainable by linking the recommendation item with the previous ones with a certain explicit relationships.

Kang *et al.* [80] proposed a general-purpose approach which is able to leverage heterogeneous item relationships and make recommendations based on the mixture of multiple relationships. The recommendation results in this method consist of two parts, general user-item translation and specific relationship-item translation, and both parts are modeled by TransE [8], the same as in the basic TransRec model. However, such general translation operation might not be effective enough to model relational item-item transition in this fashion recommendation task for two reasons. First, the number of fashion items is extremely huge, and many items only have a minor difference, which causes the sparsity problem when trying to model the item-item interactions. Second, unlike other utility-domain items, such as electronic devices, of which the the relationships are mainly determined by their functional properties, the relationships between fashion items are more sophisticated and specific. Even if we consider only the most basic relationships, *i.e.*, *substitution* and *mix-and-match*, the item-item transition can be different when focusing on various aspects. For example, two items may be able to match in terms of their categories, such as top and pants, but they do not necessarily hold the mix-and-match relationship as they can be not matching in terms of color, or style, or other design details.

In this chapter, to enhance the item-item transition modeling and further improve the recommendation performance, it is proposed to combine the characteristics of fashion domain and leverage the content-level relational transitions. Specifically, inspired by the idea of factor field feature interaction [78], the sub-transition of specific fashion attribute field (such as *category*, *color* and others) between two items is modeled. Such content-level transition modeling is promising to handle the item-item interaction sparsity problem and enhance the item transition modeling. A user-aware attention mechanism is further introduced to organically combine different sub-transitions regarding different fashion content. Such design is able to model the different levels that users care about with regard to different content [166]. The proposed model is called **A**ttentional **C**ontent-level **T**ranslation-based Fashion **R**ecommender (ACTR).

The contributions of this chapter are summarized as follows:

1) It proposes to incorporate item-item contextual relationships, specifically, substitution and mix-and-match, in the sequential recommendation models to improve the fashion recommendation performance in terms of recommendation accuracy and interpretability.

2) Considering the domain characteristics of fashion, a novel content-level interaction method is proposed in modeling the relational translation to take better advantage of rich associated information of fashion items, as well as alleviate the item-item sparsity problem. A attention mechanism is further devised to determine the importance of different attribute fields to different users and combine all content-level transitions into the overall recommendation results.

3) Extensive experiments on real-world e-commence fashion dataset iFashion under two settings demonstrate the effectiveness of the proposed method in making preferable personalized fashion recommendation. With the introduction of basic relationships between fashion items, the proposed model also aims to make relational recommendation that is suitable for various applicable scenarios.

5.2 Related Work

Two main types of sequential recommendation methods focus on sequence modeling of the ordered historical observations, Markov model-based and RNN-based. The Markov-based methods assume that the next user actions depend only on a limited number of the most recent preceding actions [131], or just the very last one. Factorized Personalized Markov Chain (FPMC) [136] is a highly representative sequential recommendation method, which can be considered as a first-order Markov Chain whose transition matrix is jointly factorized with a standard two-dimensional user-item matrix factorization approach. PRME [34] is another representative Markov-based method that shows better performance than FPMC. Inspired by translational metric embeddings [8] in modeling the transition of the user behaviors, many translation-based sequential recommendation methods have been developed in recent years. Specifically, TransRec [52] unifies item-item transition and user-item interaction to predict the next item user might be interested. MoHR [80] proposes to leverage heterogeneous item relationships in the transE-based translation framework and achieved preferable recommendation performance. CKE [194] utilizes another translation model transR [98] and incorporates multimedia knowledge to further boost the performance.

As RNNs are designed to process sequential data, they are suitable to handle the sequential recommendation problem. RNNs can model the dynamics of interactions and sequential patterns of user behaviors, as well as various multi-media side information along with the sequential signal. Variants of RNNs such as LSTM [43] and GRU [18] have been widely applied for developing a sequential recommender. Typical solutions include GRU4REC [60], NARM [93], STAMP [103] and others [158, 132, 183]. CNN-based methods also show competitive performance on sequential and session-based recommendation [159].

In the fashion domain, sequence modeling has been included in many tasks already. For example, in compatibility learning, Han *et al.* [49]employed the bidirectional LSTM to model the items in the outfit and exploit the compatible relations between items in the matching outfit. However, the sequence-aware fashion recommendation has barely been explored in literature. Most existing studies on fashion recommendation are non-sequential personalized fashion recommendation or fashion matching recommendation. For example, Yu *et al.* [193] proposed a aesthetic-based clothing recommendation method which leverage aesthetic features to enhance the fashion recommendation performance based on the CF frameworks. Hu *et al.* [66] proposed a tensor factorization approach for the personalized outfit recommendation. Although sequential recommendation in the general domain has been widely studied for its great practical value as reviewed above, in the fashion domain such a task is of exploration and faces new challenges and possible solutions.

5.3 Problem Formulation

Let $\mathcal{U} = \{u\}$ denote the whole user set and $\mathcal{I} = \{i\}$ denote the whole item set, all items that u has interacted with generate an item sequence $\mathcal{S}^u$. Given a user u and the item i, which he/her previously interacted with, this study aims to predict the interaction probability between u and item j for the next action of u, and generate the recommendation lists based on the probability scores. This task is formulated as a ranking problem, which requires the probability scores of positive next items to rank higher than that of negative items.

On top of user ID u and item ID i, associated information in the fashion domain (category, attribute, and style) is also leveraged in the proposed methods for modeling the content-level item-item transition. Each unique associated information is defined as a factor (such as *red*, *casual*), and the group containing factors describing the same aspect of the item as factor field (such as *color*), similar as defined in Chapter 4. A factor, the factor set of a factor filed f, and the whole factor field set are denoted by x, $\mathcal{X}^f = \{x_t^f\}_{t=1}^{Vf}$, and $\mathcal{X} = \{\mathcal{X}^{f_k}\}_{k=1}^{M}$ respectively. V^f is the number of factors in the field f, and M is the number of factor field.

5.4 Approach

A novel **A**ttentional **C**ontent-level **T**ranslation-based Fashion **R**ecommender (ACTR) model is proposed in this study to address the sequential fashion recommendation problem. The ACTR model is original from TransRec which applies single translation operation to effectively model third-order interaction between u, i and j. This model consists of three important parts: item relationship prediction, content-level item-item transition modeling, and recommendation based on mixture of transition.

5.4.1 Item Relationship Modeling and Prediction

Item-item relationships have been demonstrated effective in helping the next item prediction and have been explored in several works previously [80, 125, 187]. However, no attempt has ever been made in the fashion domain to employ fashion item relationships into fashion recommendations. Moreover, existing works in the general recommendation domain are limited to explore the external relationships, such as "also-viewed". By contrast, this work seeks to leverage the domain-specific relationships between fashion items, starting with the two most basic ones: substitution and mix-and-match.

From another aspect, such a relationship between adjacent items also reflects the temporal intention of users behind their actions. For example, if a user picks a T-shirt after the pick of pants, he/she is trying to find something for matching. To predict the next item-item relationship, we predict the intention of the user in his next action. The user intention is assumed to be determined by the user himself and his previous picks, and the next item is always related to his previous interacted items, implicitly or explicitly. The transnational operation [8] is applied to model the interaction between u, i and the specific relationship r inspired by knowledge graph-embedding techniques:

$$R(r|u,i) = b_r - d(\boldsymbol{\theta}_u + \hat{\boldsymbol{\theta}}_i, \boldsymbol{\theta}_r), \tag{5.1}$$

where $\boldsymbol{\theta}_u$, $\hat{\boldsymbol{\theta}}_i$, $\boldsymbol{\theta}_r \in \mathbb{R}^c$ are the embeddings of u, i, r respectively and f is the embedding size. $d()$ denotes distance measurement, which uses $L2$ distance in the specific implement of this work. b_r is the bias term. To obtain the interacted probability of different relationships under the condition of certain user u and historical interaction(s) i, a probability function P over all relationships is defined as follows:

$$P(r|u,i) = \frac{\exp(R(r|u,i))}{\sum_{r' \in \mathcal{R}} \exp(R(r'|u,i))}, \tag{5.2}$$

where $\mathcal{R}$ denotes the set of all relationships, in this case three relationships: *substitution*, *mix-and-match*, and *others*. The specific $P(r|u,i)$ can be interpreted as the possibility that the user u wants the next item to be related to i under the relationship r.

5.4.2 Content-level Relational Item Transition Modeling

Given a relationship between i and next item j, previous works model the relational item-item transition with a simple translation operation as described in sub-section 5.1. However, in our case, it is not effective enough because that item set is large and therefore item-item interaction is sparse. To enhance the item-item interaction, it is necessary to leverage more context information of the fashion items rather than using item IDs only. In fact, compared with other utility-oriented items, fashion items can be described by more details. Meanwhile, two items can be related with each other in many dimensions. For example, A, B and C are interchangeable in the same category, but A and B may be more similar to B and C if they have common details, such as the same color or same material.

Therefore, the relationship r between two items i and j should be applicable not only at the item-level, but also at the detailed content-level. Previous works usually represent the item with a latent embedding without considering any context information, which is the ID-based embedding. In contrast, this study aims to leverage more descriptive information on the fashion items. For clarity and consistency with the previous chapter, each piece of descriptive information is defined as a factor. As introduced above, each factor describes the item from one certain angle (for example, *red* describes *color*), each 'description angle' is defined as a factor field (*color* is a factor field in the case). As a result, for an item i, we have a factor set consisting of factor value of all factor fields $\mathcal{X}_i = \{x_i^f\}_{f \in \mathcal{F}}$. The item ID is treated as a special factor field, which means each specific item ID is one factor. The relational interaction between items i and j in specific factor field f is therefore modeled as:

$$R(j_f | i_f, r) = b_j - d(\boldsymbol{\theta}_i^f + \boldsymbol{\theta}_r, \boldsymbol{\theta}_j^f), \tag{5.3}$$

where $\boldsymbol{\theta}_i^f, \boldsymbol{\theta}_j^f \in \mathbb{R}^{c \times c}$ denote the representation of factor field f for items i and j.

5.4.3 Attentional Content-level Transition Aggregation

Several solutions can be applied to aggregate the transition results of different factor fields, such as average, sum or max pooling. Each operation represents an approach to treat different fields. For example, if we apply average or sum pooling, we assume that factors that belong to different fields contribute equally to the item transition modeling. In comparison, max pooling only respects the dominant factor field and ignore less influential fields in the process of transition modeling. Instead of using straightforward operations, this work proposes a user-aware attention mechanism for the content-level transition aggregation.

The relational item-item transition $R(j_f|i_f, r)$ of factor field f actually measures the interaction probability of item i and j under the relationship r in the aspect f. Intuitively, the importance of different aspects should be different, and also, it would be affected by the specific user u who actually interacts with the two items. For example, when seeking for a mix-and-match item for the previous picked one, u might care more about the *style* but less about the detailed *patterns*. Based on such consideration, we have to consider the user and previous item information to determine importance coefficients of different factor fields in the specific transition process. Specifically, inspired by graph attention network [166], the user-item-aware importance of factor field f is designed to be calculated as:

$$e_f = \boldsymbol{a}^T[\boldsymbol{W}\boldsymbol{\theta}_u, \boldsymbol{W}\boldsymbol{\theta}_i^f], \tag{5.4}$$

where $\boldsymbol{W} \in \mathbb{R}^{c\times c}$ is the trainable weight matrix, $\boldsymbol{a} \in \mathbb{R}^c$ is the trainable mapping vector. $[,]$ denotes the concatenation operation of two vectors. The final importance coefficient of f is calculated by applying a probability function over all factor fields:

$$\alpha_f = \frac{\exp(e_f)}{\sum_{f'\in\mathcal{F}} \exp(e_{f'})}. \tag{5.5}$$

Then, the overall relational item-item transition is modeled as:

$$R(j|i, r) = \sum_{f\in\mathcal{F}} \alpha_f R(j_f|i_f, r). \tag{5.6}$$

5.4.4 Sequential Recommendation

The sequential recommender is finally designed as the combination of the general third-order interaction modeling between u, i and j which emphasizes the long-term user preference, and the relational item-item interaction which focuses on the short-term sequential transition of the user behaviors. Specifically, the general user preference-based third-order interaction is captured by a personalized translation operation as follows:

$$R(j|u,i) = b_j - d(\boldsymbol{\theta}_u + \boldsymbol{\theta}_i, \boldsymbol{\theta}_j). \tag{5.7}$$

Note that the item ID is the only factor field applied in modeling the user-item interaction, which means $\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^{id}$. Moreover, the embedding of $\boldsymbol{\theta}_i$ is from different embedding sets as $\hat{\boldsymbol{\theta}}_i$ in Eq. 5.1. All relational item-item transitions are probabilistically mixed based on the predicted relationship probability determined by user u and the previous item i as introduced in Eq. 5.2. The final recommender is designed as:

$$R^*(j|u,i) = R(j|u,i) + \gamma \sum_{r \in \mathcal{R}} P(r|u,i) \times R(j|i,r), \tag{5.8}$$

where γ is the hyper-parameter that balances the importance of two terms.

5.4.5 Model Training

As introduced in the last section, the sequential fashion recommendation is formulated as a ranking problem in this study. Therefore, the final goal is to rank the ground truth next-item j higher than irrelevant item j^-. The S-BPR [136] loss is used as the sequential recommender loss, which is finally defined as:

$$L_s = - \sum_{(u,i,j,j^-) \in \mathcal{D}_s} \ln(\sigma(R^*(j|u,i) - R^*(j^-|u,i))), \tag{5.9}$$

where $\mathcal{D}_s = \{(u, \mathcal{S}_k^u, \mathcal{S}_{k+1}^u, j^-) | u \in \mathcal{U} \cap k \in [|\mathcal{S}^u| - 1] \cap j^- \in \mathcal{I} - \mathcal{S}^u\}$. As the item-item relationships is predicted in the process of user-item interaction modeling, the relationship learning is also involved in the overall model learning with a similar ranking loss specifically for item-item relationships, which aims to increase the probability of ground truth relationships. The loss for the relationship learning is:

$$L_r = -\sum_{(u,i,r,r^-)\in\mathcal{D}_r} \ln(\sigma(P(r|u,i) - P(r^-|u,i))), \tag{5.10}$$

where $\mathcal{D}_r = \{(u, \mathcal{S}_k^u, r, r^-)|u \in \mathcal{U} \cap k \in [|\mathcal{S}^u| - 1] \cap r = \text{rel}(\mathcal{S}_k^u, \mathcal{S}_{k+1}^u) \cap r^- \in \mathcal{R} - r\}$. $\text{rel}(a,b)$ means the relationship of a, b. Similarly, the item-relation-item interaction loss is introduced in the objective loss to model the pairs of relational items:

$$L_i = -\sum_{(i,r,j,j^-)\in\mathcal{D}_i} \ln(\sigma(R(j|i,r) - R(j^-|i,r))), \tag{5.11}$$

where $\mathcal{D}_i = \{(i, r, j, j^-)|i \in \mathcal{I} \cap j \in \mathcal{I}_{i,r} \cap j^- \in \mathcal{I} - \mathcal{I}_{i,r}\}$. $\mathcal{I}_{i,r}$ denotes the item set that consists of items having relationship r with item i. Finally, the problem becomes a multi-task learning problem, with the overall loss as:

$$L = L_s + \alpha L_r + \beta L_i, \tag{5.12}$$

where α and β are hyper-parameters that balance the importance of different tasks.

5.4.6 Discussion

The proposed ACTR model is similar to the MoHR [80] model in terms of translation-based interaction modeling and the basic multi-task learning framework. It is actually inspired by MoHR to a certain extent in relationship modeling. However, ACTR is also different from MoHR, mainly in three aspects. First, MoHR uses the same item embedding to predict the next relation and to model the user-item interaction. In other words, with the same user representation $\boldsymbol{\theta}_u$ and item representation $\boldsymbol{\theta}_i$, the model tries to predict the next item j and next relationship r in the same manner, which is unreasonable and limits the ability of the entire model. In ACTR, to avoid this problem, two different item representations are employed for different tasks respectively ($\hat{\boldsymbol{\theta}}_i$ for predicting relationship in Eq. 5.1 and $\boldsymbol{\theta}_i$ for predicting next item in Eq. 5.7). Second, the content-level item-item transition is modeled in ACTR to alleviate the sparsity problem in item-item interaction in the fashion domain. Moreover, an attention mechanism is introduced in content-level transition aggregation through which the importance of different factor fields are determined by the interacted user and previous item. The MoHR method is included as one of the baselines in this study; in the later experiment part, the experimental results show the different performance of the proposed ACTR and MoHR methods.

5.5 Experiments

5.5.1 Dataset

The dataset for the research in this study is derived from the large-scale fashion shopping dataset i from the well-known e-commence platform Taobao [**chen2019iFashion**]. The original iFashion dataset is composed of two parts: one is the user clicking records in a short session, in which each record contains the user and item list he interacted with in order. Another part of iFashion is the user-outfit interaction data, which contains the information of the user, the outfit he interacted with, and the items in the outfit. Only the first part is employed in this study, aka, the user clicking records of fashion items. It needs to explain that we do not use the Amazon dataset in the last study for several reasons. First, iFashion dataset contains successively clicking records of users in a fixed period of time, in which behaviors are more continues and suitable for the user intent exploring. Second, the sequence length of iFashion data is much longer than that of Amazon, which is better for the study of sequential fashion recommendation. Third, the new proposed iFashion dataset provides higher-quality product images, which reduces the noise in fashion element extraction.

As the original dataset is extremely large (with over 3M users and over 4M items), a subset dataset is generated for this study, which is termed as iFashion-Sequential

Recommendation (iFashion-SR). To become a benchmark for fairly evaluation of all sequential recommendation methods, the dataset should be applicable for different sequential recommendation methods. Considering that most deep learning sequential recommendation methods takes fixed length of historical interaction sequence as model input, very short sequences in the original dataset is abandoned.

Specifically, two settings of sub dataset are designed with the minimal sequence length being 7 and 10. Approximately 50,000 qualified (sequence longer than minimal length) user-item interaction sequences are picked under both settings. The sliding window strategy is further applied to generate short user-item sequences with fixed length (the length of sliding window). The window length of two settings are set to five and eight respectively, which means each long sequence can generate three short sequences at least. The last short sequences of all long sequences are used for validation and the second last ones are for testing. All the remaining short sequences are used for training. The specific data statistics of two settings are shown in table 5.1. Note that the preprocessing of the data in this work does not filter items based on their interaction frequency as some previous recommendation studies did, which means that small-sample items are still included in the processed dataset. Such a process is appropriate as in practical scenarios, small-sample items are important components in the whole item set, which should not be abandoned easily to reduce the challenge of the recommendation problem. As shown in Table 5.1, the scale of item set for both settings are clearly larger than that of the user set, which shows that the item-item sparsity problem aforementioned in the above sections literally exist in real-life dataset.

After settling two basic datasets, images of all interacted items are downloaded and then a commercial fashion tagging tool[1] is applied on these images to extract three types of fashion elements (category, attribute and style). A total of 225

[1]visenze.com

Table 5.1: Dataset Statistics

Seq Length	Five	Eight
#User	36,752	36,797
#Item	458,642	460,596
#Train sample	1,324,637	1,188,988
#Test sample	50,001	50,002
#Valid sample	50,001	50,002

different fashion element values (such as *dress*, *red*) are found in the entire dataset, which belong to 24 element groups based on a certain fashion taxonomy[2]. Element groups include *category*, *style*, and specific attribute groups such as *color*, *pattern*, *neckline style* and *dress shape*.

5.5.2 Experimental Settings

Baselines: Several competitive baselines are selected to be compared with the proposed method, which are specifically introduced as follows:

- **Matrix Factorization (MF)** [135]: This is the basic Matrix Factorization model using BPR loss. MF is the classic non-sequential recommendation model, which is easy to use but effective in certain applications.

- **Factorized Markov Chain (FMC)** [136]: It focuses on modeling the sequential dynamics by factorizing the item-item transition matrix, which ignores the personalization in the sequence modeling.

- **Factorized Personalized Markov Chain (FPMC)** [136]: Compared with FMC, FPMC models both personalized user-item transitions and the 'global'

[2]ViSenze taxonomy is used here.

item-item transitions by MF and factorized Markov Chains respectively, which therefore captures the personalized Markov behavior.

- **Hierarchical Representation Model (HRM)** [171]: This method extends FPMC by using aggregation operations such as max pooling to model more complex interactions. Here, we specifically use the max pooling as the performance is more competitive than average pooling according to the experimental results.

- **Personalized Ranking Metric Embedding (PRME)** [34]: It models the personalized Markov behavior by the summation of two Euclidean distances rather than the inner product used in FPMC.

- **Convolutional Sequence Embedding Recommendation (Caser)** [159]: Caser treats the embedding of fixed length of historical interacted items as an 'image' and proposes to use CNN to capture both personalized preference and sequential patterns.

- **Session-based Recommendation with RNN (GRU4REC)** [60]: This is an RNN-based sequential recommendation method. It also processes fixed length of historical items and applies GRU to handle the sequential data. It does not model user information, which therefore can only explore the sequential signal of the session behavior yet ignores personalized preference.

- **Translation-based Recommendation (TransRec)** [52]: This is the basic translation-based recommendation approach which unifies user preferences and sequential dynamics in a single translation operation.

- **Mixtures of Heterogeneous Recommenders (MoHR)** [80]: MoHR uses various recommenders to capture long-term preferences and item transitions in a unified transnational metric space.

In summary, all baselines can be classified into three groups, matrix factorization-based methods, translation-based methods and typical deep learning-based methods. MF, FMC, FPMC, HRM, PRME are matrix factorization-based methods, which are also the mainstream directions of recommender systems. They model the user preference component and sequential continuity component (if applicable) separately, which has been criticized as less inherent of the two parts. Caser and GUR4REC are two most representative deep learning-based recommenders that employ CNN and RNN respectively. TransRec and MoHR are both translation-based methods which are close to the proposed ACTR model. Specifically, TransRec can be considered as the basic framework of the ACTR model, and MoHR is the most similar model to ACTR as both are based on TransRec and both leverage item-item relationships to enhance the sequential dynamics modeling.

Implementation Details: For fair comparison, the last item of each sequence is treated as the target item for recommendation in the proposed ACTR method and all baselines. For all translation-based and matrix factorization-based methods, including ACTR, only the last two items in the sequence are used as the previous and next items, as well as the user in the model training. For two deep learning-based methods, *i.e.*, Caser and GRU4REC, all items in the sequence are used. Specifically, all items except the last are used as the historical input, and the last item is the next item to recommend, similar to all other compared methods. Such implementation ensures that the quantity of training data and prediction targets are the same for all comparable methods. Hyper-parameters α, β and γ are set to 0.1, 0.1 and 0.5 by fine-tuning the model and referring to the MoHR. The embedding size is set to 10 for all types of embeddings. For the training process, the learning rate is set to 0.005 for the ACTR method, and the batch size is set to 5000. The maximum training epoch for ACTR, as well as all ablated models, is set to 4000, which is enough for all models to convergent. All baselines are carefully tuned with different training settings including the learning rate, batch size and weight decay for achieving preferable performance.

Table 5.2: Recommendation performance of ACTR model and baselines on iFashion-SR (sequence length=5)

	MRR	NDCG	Recall
MF	0.3602	0.4023	0.5377
FMC	0.2934	0.3259	0.4313
FPMC	0.3615	0.4037	0.5396
PRME	0.3306	0.3661	0.4804
HRM	0.3425	0.3842	0.5185
GRU4REC	0.3586	0.4021	0.5418
Caser	0.2596	0.2931	0.4011
TransRec	0.2921	0.3137	0.3843
MoHR	0.4177	0.4519	0.5629
ACTR	**0.4844**	**0.5229**	**0.6468**
%Improv.	**15.97**	**15.71**	**14.90**

Evaluation Metrics: For evaluation, 99 negative items are randomly sampled for each sample in the test and validation sets [30, 58] for the sake of reducing the computational cost in evaluation. All negative items are not interacted by the corresponding user. To comprehensively evaluate the effectiveness of top-K recommendation, three common evaluation matrices are employed: RECALL@K, NDCG@K and MRR@K. Specifically, k is set to 10 by default. Results for all matrices are reported averagely.

5.5.3 Recommendation Performance

Tables 5.2 and 5.3 show the overall recommendation performance of the proposed ACTR method as well as all baselines on the iFashion-SR dataset for two settings. The observations from the results are as follows:

Table 5.3: Recommendation performance of ACTR model and baselines on iFashion-SR (sequence length=8)

	MRR	NDCG	Recall
MF	0.3519	0.3934	0.5266
FMC	0.2895	0.3215	0.4253
FPMC	0.3540	0.3952	0.5277
PRME	0.3210	0.3556	0.4672
HRM	0.3369	0.3789	0.5131
GRU4REC	0.3574	0.4010	0.5413
Caser	0.2554	0.2885	0.3952
TransRec	0.2898	0.3106	0.3791
MoHR	0.4082	0.4415	0.5495
ACTR	**0.4751**	**0.5122**	**0.6314**
%Improv.	**16.39**	**16.01**	**14.90**

(1) The proposed ACTR outperforms all compared baselines in a large margin under all evaluation metrics for both experimental settings. In particular, the MRR and NDCG results are improved more significantly compared with Recall, which illustrates that the ACTR model is better at ranking correct recommender items in high positions.

(2) Comparing the results of MF, FMC and FPMC, we can discover that the performance of FMC is worse than that of MF, moreover. Moreover, although FPMC achieves the best performance among the three, it is very close to MF. Recall that FMC only models the item-item interactions while MF only models user-item interactions. Such experimental results show that item-item interactions are more difficult to model, which is probably due to the sparsity problem pointed out in the above analysis. That is also why FPMC only outperforms MF by a slight margin although it models both item-item and user-item interactions. These

results also justify the motivation of ACTR which is to enhance the item-item interaction modeling by leveraging the interaction in the factor-field level.

(3) Comparing the result of TransRec and MoHR, we can see that by incorporating item-item relationships and relational item-item transition modeling, the recommendation performance can be improved significantly. The TransRec shows the worst performance in the experiments, which indicate the very simple single component translation model might not be effective enough to model complex interactions. This also explains why the introduction of the relationships is necessary, as well as the content-level transitions. The experimental result shows such technical improvements are able to improve the expressiveness and effectiveness of the very basic translation model.

(4) Almost all methods perform better in the first experimental setting (sequence length=5) comparing two settings (Tables 5.2 and 5.3), although the performance difference is very minor. The first reason for such difference might be the difference in the number of training samples. From the dataset statistics shown in Table 5.1, we can see the first setting has approximately 150k more training samples than the second setting. Meanwhile, a smaller number of users and items makes the data distribution more dense and reduce the challenge. It is also notable the GRU4REC achieves close performance on two settings, which might imply that inputting more historical items can help capture the sequential patterns behind the user behavior and improve the recommendation performance.

5.5.4 Ablation Study

In this part, several ablation studies are conducted to validate the effectiveness of the different parts of the ACTR model. Three technical parts are important to investigate. First, compared with the design in MoHR, how does it help by applying two different sets of item embeddings for predicting relationships and modeling user-item interactions respectively. Second, whether the content-

Table 5.4: Ablation experiments (sequence length=5)

	MRR	NDCG	Recall
Vanilla	0.4177	0.4519	0.5629
V + T	0.4272	0.4600	0.5661
V + C	0.4399	0.4761	0.5935
V + T + C	0.4449	0.4791	0.5897
ACTR	**0.4844**	**0.5229**	**0.6468**

Table 5.5: Ablation experiments (sequence length=8)

	MRR	NDCG	Recall
Vanilla	0.4082	0.4415	0.5495
V + T	0.4155	0.4474	0.5508
V + C	0.4258	0.4609	0.5752
V + T + C	0.4311	0.4652	0.5758
ACTR	**0.4751**	**0.5122**	**0.6314**

level relational item-item transition indeed helps improve the recommendation performance as expected. Last, whether the attention mechanism designed to determine the importance of different factor fields to different users further improve the recommendation performance.

The results of the ablation study for two experimental settings are shown in Tables 5.4 and 5.5. **Vanilla** denotes the basic ACTR model that only uses one set of item embeddings, no content-level transition modeling, and no attention mechanism applied. **V+T** denotes the Vanilla model with two sets of item embeddings, **V+C** denotes the Vanilla model with content-level transition modeling and **V+T+C** denotes the model that has both components (refer to section 5.4.6 for a clearer introduction). The final ACTR model has both **T** and **C**, and also the attention mechanism for the content-level transition aggregation. The results show that all technical components specifically designed in the ACTR model work and improve the recommendation performance on both experimental settings. Specifically, two components innovatively introduced in the relational item-item transition process, the content-level transition modeling and the user-aware attention mechanism, greatly help the performance of the fashion recommendation model. Such observation demonstrates the effectiveness of the two technical components, which also reflects that enhancing the item-item interaction is important in sequential recommendation. in the fashion domain, such relational item-item interaction modeling can be improved by specifying and leveraging detailed information from different fashion aspects.

5.5.5 Qualitative Analysis

More detailed qualitative analysis and discussion on the proposed ACTR model for fashion recommendation is presented in this section. Figure 5.2 firstly illustrates six recommendation cases based on the ACTR model. To be clear, the user, previous item (on the left of the item pair) are input to the model, the user intention for next item (item-item relationship for the previous and next item) and the next item to recommend are both generated by the trained ACTR model. In all presented cases, the user intentions are correctly predicted and the recommended item is the target item the user would choose (ground truth), which demonstrates the effectiveness of the proposed ACTR model. Moreover, by clarifying the next intention, the recommendation results become more reasonable and therefore more convincing.

Note that the final ranking list for each user is based on the mixed result of long-term user preference and the relational short-term item-item transition, referring to Eq. 5.8. However, the recommendation list can be manipulated manually to only recommend the items that are in accordance with the predicted user's intention. For example in Figure 5.3, the user in the first case is predicted by the ACTR model that he wants items to match with his previous pick. Therefore,

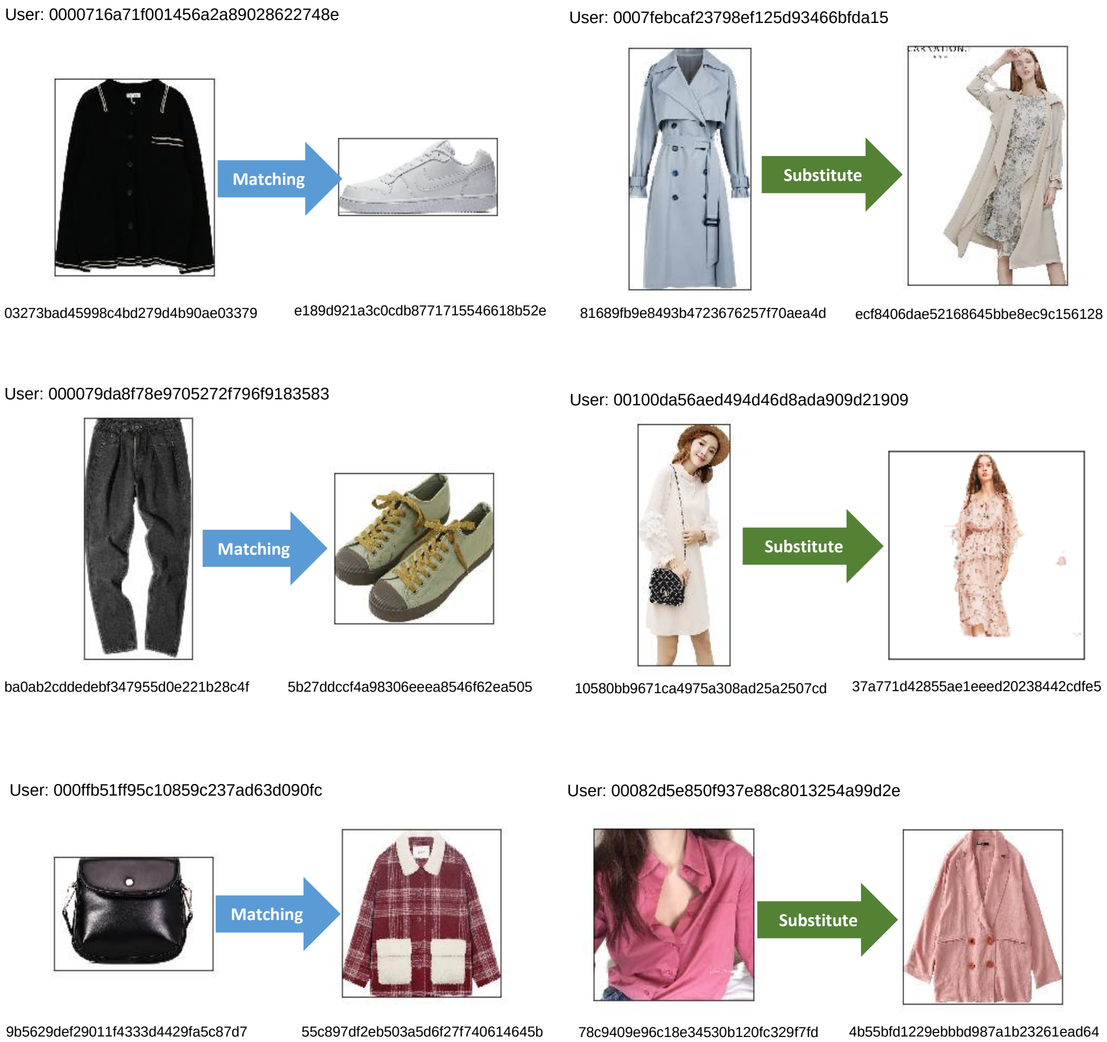

Figure 5.2: Six cases of successful recommendations. Each case shows the user name, item pairs containing the previous clicked item (in the left) and the recommended item (in the right), as well as the predicted user intention, which is the corresponding relationship of the item pair (indicated by the arrow in the middle of item pairs).

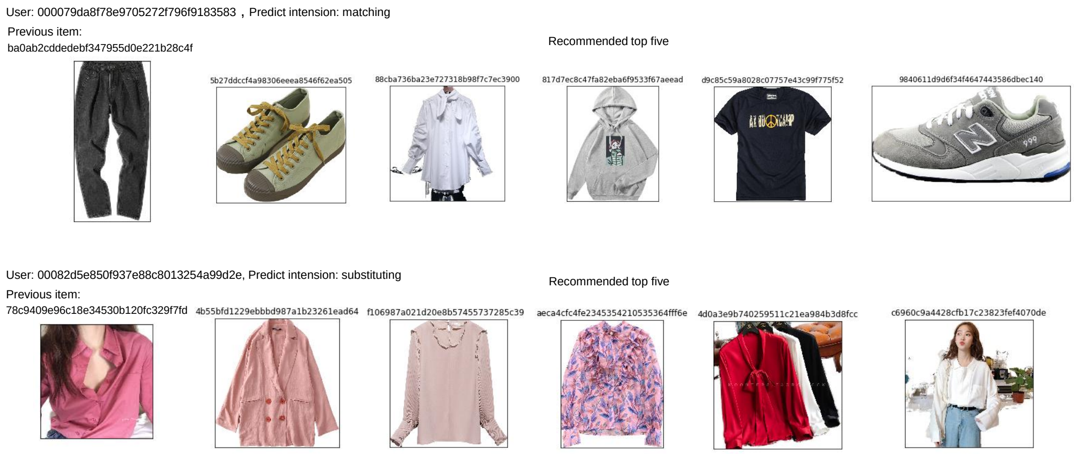

Figure 5.3: Two recommendation examples for two users for specific intentions, which are also predicted by the model.

the final recommendation list contains matching proposals only. Likewise, in the second case, the user intention for the next item is predicted to be a substitute. As a result, a recommendation list full of similar items is presented to him/her.

5.6 Summary

This chapter works on the sequential fashion recommendation task, aiming to model the content-level relational item-item transition to enhance the recommendation performance, as well as to bring more intrepretability. The item-item relationships are leveraged, which facilitates the item transition modeling and also suggests the user intention behind the action to some extent. An ACTR method is proposed to address the target task. Specifically, the relationship-aware item-item transition is specified into content-level and then obtained by aggregating the content-level transition results.

Extensive experiments have been conducted to evaluate the effectiveness of the proposed method. The experimental results show that the proposed ACTR model outperforms all competitive baseline methods in terms of recommendation accuracy, which demonstrates the effectiveness of all technical improvements in ACTR. Moreover, the qualitative analysis shows that the ACTR method can provide the user intention prediction along with the recommendation results, and also be able to provide special recommendation results given certain intentions.

6 Leveraging Multiple Relations for Fashion Trend Forecasting

6.1 Introduction

Fashion trend forecasting, aiming to master the changeability in fashion, is an increasingly important research field. Although rapid technological change has infiltrated every aspect of modern life, people's desire to convey a sense of self through their appearance has not changed. They need fashion guidance to develop good taste and catch up with the trends [10]. Additionally, for the fashion industry, valid forecasting enables fashion companies to establish marketing strategies wisely and continuously anticipate and fulfill their consumers' wants and needs. Traditionally, fashion experts need to travel and conduct surveys to determine people's real fashion tastes based on local culture and tradition, which usually affects the world's fashion trends [83]. However, such approaches are inefficient, expensive, highly dependent on the experts' background and are usually biased because of the expert's personal preference. Meanwhile, great progress on the Internet, big data, and artificial intelligence provides an alternative way to solve this challenging problem: automatically forecasting fashion trends based on fashion data.

In fact, fashion trend forecasting or analysis has attracted some research interests in the computing field, but is still at its nascent stage. One popular data source for fashion trend analysis are purchase records on either online or offline retail platforms [3]. However, purchase records directly reflect people's buying decisions which are influenced by many factors except real fashion preferences such as retailers' promotions, buying clothes for others (*e.g.*, family), and never wearing

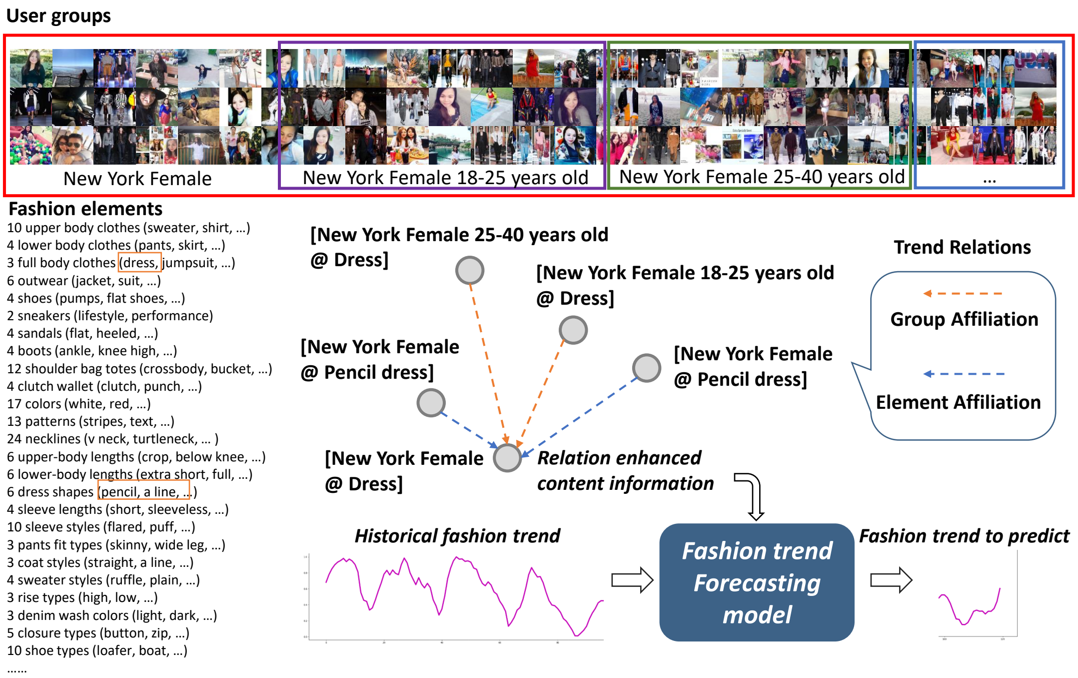

Figure 6.1: The fashion trend forecasting task aims to forecast the trends of meaningful fashion elements for specific user groups. Multiple relations (group and fashion element affiliations) between the time series are leveraged to enhance the prediction.

after buying. Thus, the real fashion taste of consumers cannot be properly captured solely based on retailing records. Social media now records daily life of people from all over the world and has turned into a platform for a growing number of users to show their fashion taste and opinions, which becomes a natural research basis for fashion trend analysis. Moreover, data from social media are massive, diverse, highly related to fashion and cover a long time span, which makes them applicable for insightful large-scale fashion trend analysis. One recent state-of-the-art study of fashion trends based on social media was conducted by Mall *et al.* [113]. They extracted fashion attributes from social media images with a CNN model and investigated the trend of specific attributes for each city. Compared with other works which target at implicit fashion styles by visual clustering [114, 2], the fashion trends demonstrated in Mall's work were more specific. However, the study has several limitations. First, the investigated fashion attributes are coarse-grained and of less significance, and most of the trends show simple seasonal patterns, such as *wearing jacket*. Second, they investigated the fashion trends with regard to cities, which actually attempts to distinguish fashion trends based on the locations of target user groups. However, only using city to categorize people is not enough, and more attributes from the aspect of users should be explored such as age and gender.

Similar to previous work [113], this chapter investigates fashion trends of specific fashion elements for various user groups. Towards more meaningful fashion trend analysis, three types of fashion elements are main targets, including category (*e.g.*, *dress*, *jeans*, *playsuit*), fashion style (*e.g.*, *feminine*, *sophisticated*) and detailed attributes (*e.g.*, *color*, *pattern*, *neckline*, *skirt shape*). More user profile information are explored and users are accordingly categorized into specific groups. In summary, a new dataset with extensive fine-grained fashion elements and user information is introduced for the purpose of this study. The time series in this dataset spans over five years from which the change and evolution of specific fashion trends in a long time horizon can be observed.

To perform accurate data-driven fashion trend forecasting, we have to capture the underlying patterns in the historical fashion trend time series. Although traditional models such as exponential smoothing or linear regression have been effectively applied to model simple time-series data [3, 113], they fall short in making sound predictions for more complicated trends. Recent advances in deep learning have provided great solutions for many tasks [90]. In particular, RNNs have demonstrated its superiority in modeling time series data [19, 180, 31]. However, such approaches have not been employed in the area of fashion trend analysis yet. From another perspective, most existing works model the pieces of fashion trend signals independently. However, according to both commonsense and fashion theories, the fashion trends are not independent but well-correlated with each other. For example, the fashion element *turtleneck* is an affiliated attribute to the fashion element *sweater*; therefore, the trends of these two elements should be closely related with each other. Al-Halah*et al.* [2] explored fashion influences between different cities, the conclusions of which to some extent coincide and support our hypothesis that the fashion trends of different user groups are also not independent. Specifically in our task, it is natural to consider that the trends of any fashion elements for the group *New York Female* can be affected by the corresponding trends for the group *New York Female 25-40*. As shown in Figure 6.1, each fashion trend should be connected to others based on the relation between the user groups and fashion elements. These relations are non-trivial to model yet can be helpful for the forecast of the trends. Most existing models failed to incorporate any kind of relations.

This chapter proposes to employ the LSTM encoder-decoder as the basic model for the fashion trend forecasting task which leverages fashion element and user group information. Two message-passing modules are introduced in the model to explore the influence of correlated fashion elements or user groups in trend modeling. As the task is to forecast a period of future trend, it is more challenging than only one-step-ahead estimation. To better capture temporal patterns on future horizons, the proposed model is equipped with a sliding temporal attention module [31]. Specifically, at each time step the decoder hidden state is attended to several different periods of the history and generate the attention vector individually, and then all periods of the history are combined to predict the current time step. The combined features therefore can better describe the current time step as it incorporates both historical information and future contextual information. The proposed model is named as **R**elation **E**nhanced **A**ttention **R**ecurrent network (REAR).

The contributions of this chapter are summarized as follows:

1) A new large-scale dataset for fashion trend forecasting based on Instagram, named Fashion Instagram Trending (FIT). The dataset is annotated with rich fashion elements and user information, which can benefit the research community towards specific and meaningful fashion trend forecasting.

2) REAR is proposed for addressing the fashion trend forecasting problem. The REAR model applies two message passing modules to leverage the relations between fashion elements and groups that affect the corresponding fashion trends. Also, a sliding temporal attention mechanism is devised to further improve the prediction capability for long-horizon forecasting.

3) Extensive experiments are conducted on the proposed FIT and the GeoStyle datasets [113]. Experimental results demonstrate that the REAR model is capable of capturing the complex patterns in time-series fashion trend data, and achieves preferable accuracy in fashion trend forecasting. Furthermore, qualitative analysis demonstrates that fashion trends generated by the REAR model based on the proposed dataset is consistent with the trends provided by authoritative fashion agencies.

6.2 Related Work

6.2.1 Fashion Trend Forecasting

Fashion trend forecasting, as an up-stream research task in the field of computational fashion analysis, is attracting increasing research attention. It is usually studied based on some fundamental related tasks such as fashion recognition, detection, retrieval and segmentation [105, 173, 104]. For example, Simo-Serra [146] studied the semantic outfit descriptions based on (non-visual) clothing meta-data which mainly consisted of color and coarse categories. They proposed the Fashion144k dataset for the research based on the fashion website chictopia.com which mainly explored meta-data. The descriptive words in the dataset to determine fashion trends are very sparse and less meaningful to some extent because they are edited casually by users. Al-Halah *et al.* [3] studied fashion trends based on fashion styles. To obtain fashion styles, they represented all fashion images with detected semantic attributes through a pre-trained deep attribute detection model, and then applied the nonnegative matrix factorization on all predicted attributes. They built a dataset based on the e-commence platform Amazon to support their research which consists of images and text of the purchased items for a fairly long period of time. First, the research target they chose, the fashion styles, were implicit and actually the clustering of images conditioned with attributes. Such lack of explanation causes the trend analysis to be less effective and convincing. Second, the e-commence data are not suitable for the fashion trend analysis because the purchase decision was affected by various factors, and fashionability of the item is only one of them. Another recent study [2] also adopted a similar strategy that it trains a neural network model to detect fashion attributes and then learns a set of fashion styles based on attributes through Gaussian mixture model. The dataset they adopted is the GeoStyle dataset proposed by Matzen *et al.* [114, 113]. In this dataset, more specific fashion attributes such as neckline shape or sleeve length are explored. However, although the attributes they investigated are more detailed, they are of less significance in terms of indicating real fashion and effectively revealing fashion trends. For instance, one fashion element in the dataset is *wearing jacket*, which apparently does not show the evolution of fashion but the change of seasons. In summary, existing fashion trend forecasting datasets are generally with less significance that cannot support the study of practical fashion trend and forecast. To address this research gap, this chapter proposes a large-scale fashion trend dataset based on Instagram, showing very specific, detailed and realistic fashion trends with respect to a large number of fine-grained fashion elements. The dataset collected in this chapter also contains rich user information so that the fashion trends among specific group of people can be further analyzed.

6.2.2 Time-series Prediction for Fashion Trend Forecasting

Fashion trend forecasting aims to predict the future value based on historical time series records, which is quite close to the time series prediction problem. Some classic and state-of-the-art time series prediction solutions are reviewed here. Autoregrassive (AR) [168] is a traditional, simple yet effective statistical model to address the time-series prediction problem. Based on AR, several advanced models are developed and also widely used to solve similar tasks, including moving averages (MA) [149], improved autoregressive integrated moving average (ARIMA) [9] and others [63, 182]. Although statistical models have achieved great success, they are still limited to modeling simple or cyclic patterns. In many real-life applications, such as the fashion trend forecasting in this paper, the data patterns are notorious for being highly volatile and are too complex to be captured by statistical models.

Recently, Neural Networks (NNs) have gradually become powerful techniques for many essential tasks. Specifically, RNNs, especially its variant LSTM [62], have achieved state-of-the-art performance in time series prediction [31, 97, 130] and have been successfully applied in many specific tasks such as stock prediction [33] and sales forecasting [6]. Even though fashion trend forecasting is also a type of time series prediction tasks, it is a domain-specific task, such that good solutions should be exclusive and leverage specific and beneficial knowledge in the specific domain. Such idea has also been proposed and implemented in other time series prediction tasks. In stock prediction, Feng *et al.* [33] incorporated domain knowledge of stocks and effectively improved stock price forecasting. However, in fashion trend forecasting, no such attempt has been made so far.

Table 6.1: Statistical Comparison between FIT and GeoStyle datasets

Dataset	City	Gender	Age group	Fashion Element	Time span
GeoStyle	44	N/A	N/A	46	3 years
FIT	14	2	4	197	5 years

6.3 Problem Formulation and Dataset

This chapter focuses on the fashion trend forecasting problem, which aims to make prediction of future popularity with regard to each fashion element (*e.g.*, *white*, *dress*, *off-shoulder* and others) for each user group (*e.g.*, *London female of age between 18 and 25*). Given a fashion element $f \in \mathcal{F}$ and a user group $g \in \mathcal{G}$, the temporal popularity of f for g is defined as a time series denoted as $\boldsymbol{y}_g^f = (y_1, \cdots, y_t, \cdots)$, where $\mathcal{F}$ is the set of all fashion elements; and $\mathcal{G}$ is the set of all user groups. The value of the time series at each time step t is defined as $y_t = N_t^{g,f}/N_t^g$, where $N_t^{g,f}$ is the number of the fashion elements f at time point t for group g; N_t^g is the number of all fashion items (*e.g, clothing, bags, shoes and others*) observed at time point t for group g. Given the historical inputs within the time span of $[1, T]$, the aim of this chapter is to forecast the future values of time $[T+1, T+T']$, where T is the historical sequence length or time span, and T' is called the forecast horizon (the number of steps ahead to forecast, $T' > 1$).

As most of the existing datasets are limited for this type of study, a new dataset based on the popular social media platform Instagram[1] is contributed, called FIT. Table 6.1 show the statistical comparison between FIT and Geostyle [113]. It shows that the FIT dataset has more user information, richer fashion elements and a longer time span.

Specifically, millions of posts uploaded by users were crowded from all over the world. To guarantee the quality of the crawled data, automated and manual

[1]instagram.com

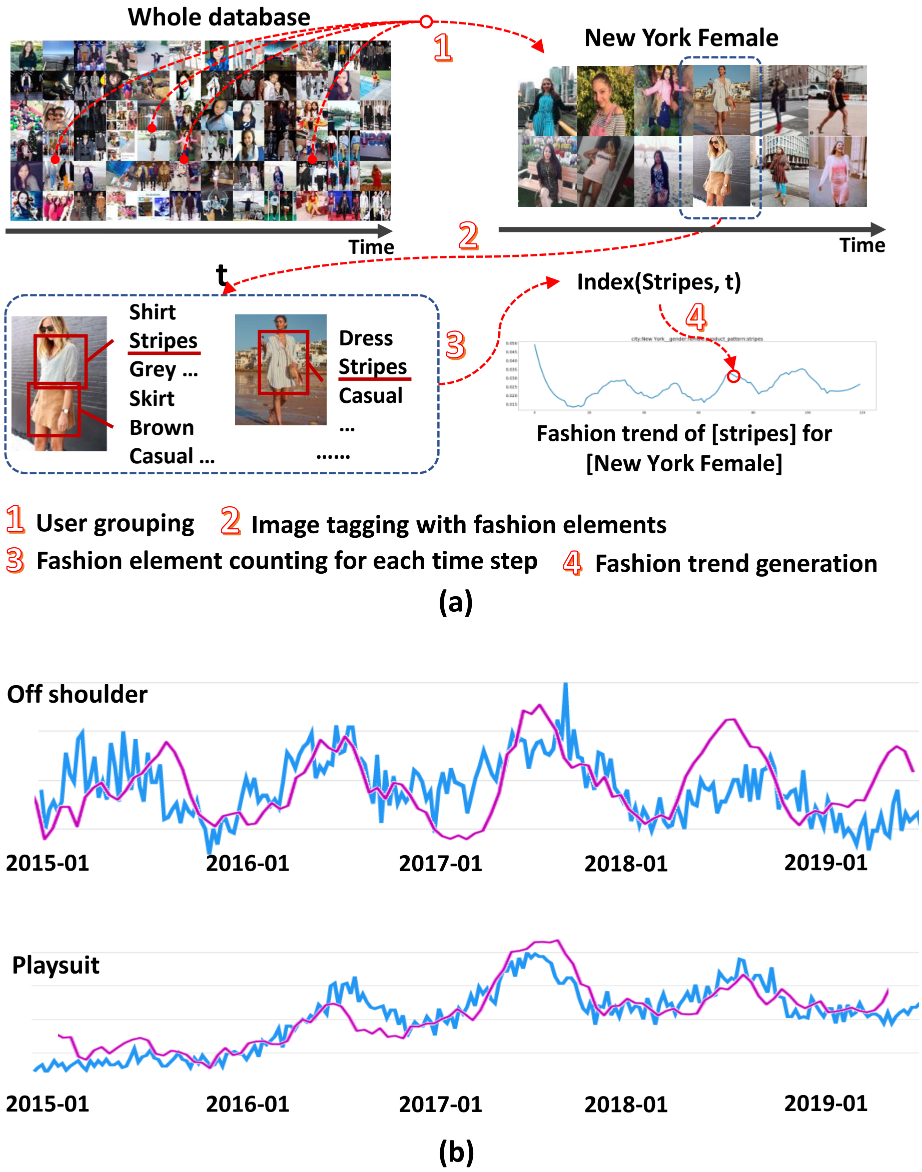

Figure 6.2: (a) An illustration of procedures of building the FIT dataset. (b) Two examples of the FIT dataset, where RED curves are from the FIT dataset and BLUE curves are from Google Trends (both examples belong to the group [New York, Female]).

filtering are conducted on the collected data, similar to that in [112, 111]. First, the pre-trained object detection model was applied to detect person body [134] and face [195]. Images without face or body, or with abnormal-sized face or body are filtered out. Then, images with "wrong people" are dropped, in which the bounding boxes containing people that are not the corresponding account owner by a majority vote mechanism over the predicted age and gender(see more details below). Finally, about *680K* images are kept in total. The annotation of the dataset is from two aspects: users and fashion elements, which will are introduced in detail below.

For users, three types of user information (*i.e.*, age, gender and location) are collected. Based on these information, users are separated into different groups. To obtain this information, the off-the-shelf age and gender detector tools [139, 4, 124] are firstly applied on each person (face) bounding box of all posts (images) of the users, and then the dominant gender and average age are selected as the final gender and age. Person bounding boxes are detected as the opposite gender and with age differing from the detected age by over five years are dropped. Four age groups are defined: 0 to 18, 18 to 25, 25 to 40, and above 40. The information of location is based on the longitude and latitude data that comes with the post, and choose the most frequent one as the location of the user. Finally, 14 main cities across the world are included in the proposed dataset. The combination of the three types of user attributes forms a group, resulting in 74 groups in the whole dataset.

A commercial fashion tagging tool[2] is applied on the dataset to extract three types of fashion elements (category, attribute and style) from the images, resulting in a total 197 of different fashion elements for the whole dataset. Each image is labelled with user group, time, and fashion elements after the annotation in the end. After the information extraction process, the popularity of each fashion element for each user group for every half month is calculated and the popularity

[2]visenze.com

for the whole time span generates a time series data. The post time of FIT dataset ranges from July 2014 to June 2019, spanning five years, which means that each time series has 120 data points. To ensure the data quality, sparse time series with over *50%* of time points empty are further dropped. Finally, around *8000* time series are obtained in total. The main procedures of generating the specific fashion trends in FIT are illustrated in Figure 6.2 (a) Note that the tags (element information) come from an existing tagging tool, which might contain some noise and result in a small bias of real fashion trends. However, the recognition results were manually checked partly and found to have relatively satisfying average accuracy. More importantly, as each time series data is a statistical ensemble of a group of users' data, the noise of each user on the final time series is hugely weakened. Besides, the author comprehensively analyzes fashion trends in FIT and compares some of them with that from **Google Trends**[3], and observes highly similar patterns, which further validate the credibility of the FIT dataset (see examples in Figure 6.2) (b). It should be noted that the Google Trends is used to justify the reliability of the contributed dataset in certain cases. It is notable that Google Trends can only be used for certain popularly searched rough-level fashion elements, and it is unable to provide trends of most fine-grained fashion elements and specific groups of users.

6.4 Approach

This chapter aims to develop an end-to-end model to forecast the fashion trends given the historical input. The LSTM encoder-decoder is adopted as the basic framework, which is able to incorporate both time series inputs and the associated sequence information into a unified model and make multi-horizon forecasting. To leverage the multiple relations among sequences, the message passing mechanism is used. It is able to propagate information among related sequences thus obtain relation-enhanced representations for each sequence. Moreover, a

[3]trends.google.com

sliding temporal attention mechanism is deviced to enhance the forecasting performance in long-horizon scenario. The proposed attention mechanism adaptively attends to the historical sequences, thereby countering the error accumulation effects during decoding. The proposed framework is named as REAR as shown in Figure 6.3.

6.4.1 Relation Enhanced Historical Trend Encoding

Sequence Feature Embedding: Given a time series $(y_1, \cdots, y_T)$ indicating the past trend of fashion element f for group g within time period $[1, T]$, the task is to forecast the future values of the trend $(y_{T+1}, \cdots, y_{T+T'})$. The group g is defined by the combination of three attributes $g = [c, a, n]$, where $c \in \mathcal{C}$ is the city, $a \in \mathcal{A}$ is the age group ($\mathcal{C}$, $\mathcal{A}$ denote all cities and all age groups) and $n \in \{male, female\}$ is gender. Each of the group features c, a, and n is converted into into embedding (randomly initialized) $\boldsymbol{c} \in \mathbb{R}^D$, $\boldsymbol{a} \in \mathbb{R}^D$, and $\boldsymbol{n} \in \mathbb{R}^D$ separately to obtain the group representation, where D is the dimensionality of sequence feature embedding. All three group embeddings are then aggregated into one unified group representation via a linear layer:

$$\boldsymbol{g} = \boldsymbol{W}_g[\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{n}] + \boldsymbol{b}_g \tag{6.1}$$

where $\boldsymbol{W}_g \in \mathbb{R}^{D \times 3D}$, $\boldsymbol{b}_g \in \mathbb{R}^D$, and $\boldsymbol{g} \in \mathbb{R}^D$. The fashion element f is directly converted into an embedding $\boldsymbol{f} \in \mathbb{R}^D$.

Multiple Relation Modeling: As discussed in subsection 6.1, fashion trend signals are not independent but correlated with and influence each other via multiple relations. Such time-series correlations have also been emphasized in other time series modeling tasks such as stock price prediction [33] and water and air quality monitoring [97]. In terms of fashion trend, various complex relations exist, for example, the lead-lag influential patterns across different big cities [2]. However, such complex relations are difficult to specify and explore.

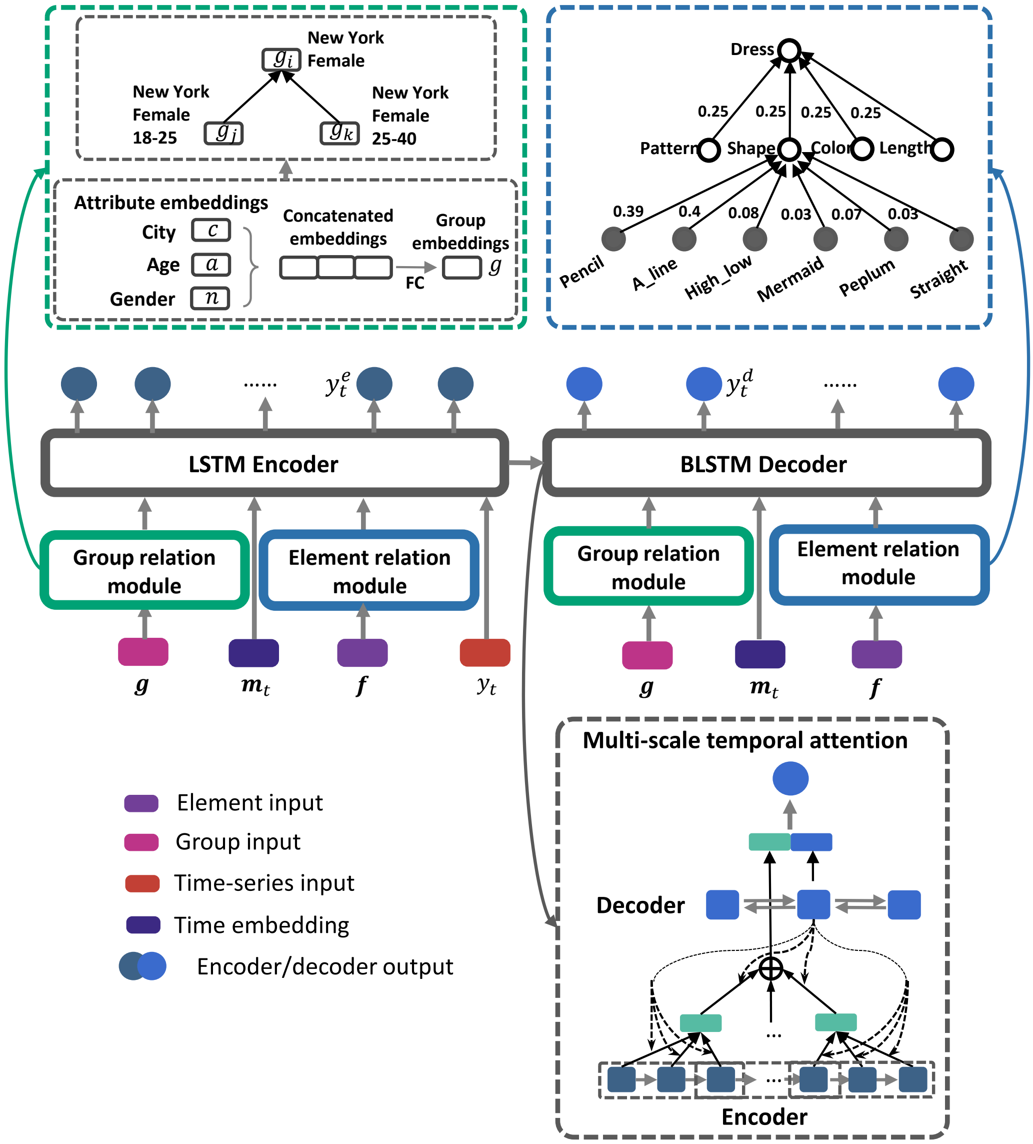

Figure 6.3: The **R**elation **E**nhanced **A**ttention **R**ecurrent network (REAR) framework. It is based on the LSTM encoder-decoder framework and incorporates relations between groups and fashion elements (shown in the green and blue dashed line boxes respectively). Moreover, a sliding temporal attention module (shown in the black dashed line box) is used at the decoder stage to enhance long-horizon forecasting performance.

In this work, the relation incorporation is started from the direct, intuitive but important relations among fashion trends: the affiliation relations determined by sequence features, *i.e.*, the group and fashion element.

(1) Relations between fashion elements: Each piece of time series data of fashion trend describes the specific trend of one fashion element, such as *dress shape: A_line*. There are three types of fashion elements considered in this study, namely category, attribute and style. Two of them (category and attribute) can be naturally organized with a tree-structured taxonomy. As part of the taxonomy shown in the blue dashed box in Figure 6.3, the category *Dress* has several attributes (*e.g.*, *Pattern, Shape, Color*) and the attribute *Shape* has several values (*e.g.*, *Pencil, A_line, High_low*). It is easy to understand that because many attributes are affiliated to certain categories, the corresponding attribute values (*i.e.*, *A_line*) are produced fully or partially based on one category (*i.e.*, *dress*). This relation indicates that the trends of the affiliated fashion attributes can reflect the trend of the related category to a certain extent. With such relations between categories and attributes, the child nodes are designed to affect the parent nodes.

The normalized portion of each child node out of its parent node is denoted in the edge of the taxonomy in Figure 6.3. Apparently, the trends of parent nodes will be consistent with the sum of the children nodes. For example, if the trend of the attribute *peplum* goes up, it is highly probable that the category *dress* also goes up. Note that this correlation of time series trends is directed, which only holds from children to parents and does not hold vice versa. For example, if the number of *Shape* goes up, the number of *Pencil* is not definitely goes up because it may be caused by the increasing of other attribute values like *A_line*.

The message passing mechanism is applied, which is called affiliation relation. Specifically, there are three types of nodes in this tree: *category*, *attribute* and *attribute value*, and the affiliation relations are between *attribute* and *category*, *attribute value* and *attribute*. As mentioned above, each fashion element f is converted into a vector representation $\boldsymbol{f}$. The message passing is conducted between those embeddings, *i.e.*, passing messages from child nodes to their parent nodes. The message passing for node i is as follows:

$$\boldsymbol{s}_i = \sum_{j \in \mathcal{E}_i} \alpha_j^i \boldsymbol{f}_j, \tag{6.2}$$

where α_j^i is the weight which is proportional to the impact of element j on element i. During implementation, α_j^i is set to the normalized portion of each child node *w.r.t* its parent node. $\mathcal{E}_i$ denotes the set of fashion elements that affect i, *i.e.*, elements affiliated to i. Therefore, $\boldsymbol{s}_i \in \mathbb{R}^D$ is the message passed from all affiliated child nodes. The propagated information is finally aggregated with the original node representation using a linear layer and the relation enhanced fashion element embedding is generated therefore.

$$\boldsymbol{f}_i^* = \boldsymbol{W}_e[\boldsymbol{f}_i, \boldsymbol{s}_i] + \boldsymbol{b}_e. \tag{6.3}$$

where $[,]$ denotes the concatenation operation of two vectors, $\boldsymbol{W}_e \in \mathbb{R}^{D \times 2D}$ and $\boldsymbol{b}_e \in \mathbb{R}^D$ are the parameters of this liner layer.

(2) Relations between user groups: Each time series data describes the fashion trend of a certain user group, which is defined by two required attributes (city and gender) and one optional attribute (age group). The groups with the attribute of *age group* are naturally more fine-grained than their corresponding groups without the attribute (but with same *city* and *gender*). The fine-grained groups can be treated as affiliations of their corresponding coarse-grained groups. As shown in Figure 6.1, there are three groups [*New York Female 25-40 years old*], [*New York Female 18-25 years old*] and [*New York Female*]. Apparently, the two fine-grained groups [*New York Female 25-40 years old*] and [*New York Female 18-25 years old*] are affiliations (subsets) of the coarse-grained group [*New York*

Female]. Message passing is conducted between group embeddings to leverage such relation, similarly as between fashion elements:

$$\begin{cases} \boldsymbol{r}_i = \sum_{j \in \mathcal{G}_i} \beta_j^i \boldsymbol{g}_j \\ \boldsymbol{g}_i^* = \boldsymbol{W}_g[\boldsymbol{g}_i, \boldsymbol{r}_i] + \boldsymbol{b}_g. \end{cases} \tag{6.4}$$

$\boldsymbol{r}_i$ is the propagated information from group $i's$ affiliated groups $\mathcal{G}_i$, β_j^i is message passing weight between two groups, which is initiated to 1 for all affiliation relations and 0 else. $\boldsymbol{W}_g$ is the trainable parameter.

LSTM Encoder: The LSTM is adopted to encode the historical fashion trend sequence as well as the enhanced sequence features. Specifically, the input of the encoder network at timestep t is generated by concatenating the group representation $\boldsymbol{g}^*$, the fashion element representation $\boldsymbol{f}^*$, the timestep feature $\boldsymbol{m}_t$ (the position of each point within one year, converted to vector representation thus $\boldsymbol{m}_t \in \mathbb{R}^D$):

$$\boldsymbol{v}_t^e = [\boldsymbol{g}^*, \boldsymbol{f}^*, \boldsymbol{m}_t, y_t], \tag{6.5}$$

where $\boldsymbol{v}_t^e \in \mathbb{R}^{3D+1}$. The output of the encoder LSTM is the hidden representations for the input sequence at timestep t, denoted as:

$$\boldsymbol{h}_t^e = LSTM^e(\boldsymbol{v}_t^e; \boldsymbol{h}_{t-1}^e), \tag{6.6}$$

where $\boldsymbol{h}_{t-1}^e, \boldsymbol{h}_t^e \in \mathbb{R}^H$, and H is the size of the hidden state. $\boldsymbol{h}_{t-1}^e$ is the encoder hidden state at timestep $t-1$,

6.4.2 Attended Future Trend Decoding

Bi-directional LSTM Decoder: The decoder network is a bi-directional LSTM (BiLSTM), of which the initial hidden state is $\boldsymbol{h}_T^e$, *i.e.*, the last hidden state of the encoder. At each decoding step, it takes the input features and outputs the forecasting value. The input feature of the decoder network at timestep

t is $\boldsymbol{v}_t^d = [\boldsymbol{g}^*, \boldsymbol{f}^*, \boldsymbol{m}_t]$, which is different from $\boldsymbol{v}_t^e$ by removing the trend value y_t and thus $\boldsymbol{v}_t^d \in \mathbb{R}^{3D}$. The BiLSTM can propagate information from forward and backward directions, and the final prediction should be made after the information propagation of BiLSTM [31]. Formally, the hidden state from forward LSTM is denoted as $\overrightarrow{\boldsymbol{h}_t^d}$ and from backward as $\overleftarrow{\boldsymbol{h}_t^d}$. The final hidden state $\boldsymbol{h}_t^d$ by is the concatenation of them as follows:

$$\begin{cases} \overrightarrow{\boldsymbol{h}_t^d} = \overrightarrow{LSTM^d}(\boldsymbol{v}_t^d; \overrightarrow{\boldsymbol{h}_{t-1}^d}) \\ \overleftarrow{\boldsymbol{h}_t^d} = \overleftarrow{LSTM^d}(\boldsymbol{v}_t^d; \overleftarrow{\boldsymbol{h}_{t+1}^d}) \\ \boldsymbol{h}_t^d = [\overrightarrow{\boldsymbol{h}_t^d}, \overleftarrow{\boldsymbol{h}_t^d}], \end{cases} \tag{6.7}$$

where $\overrightarrow{\boldsymbol{h}_t^d}, \overrightarrow{\boldsymbol{h}_{t-1}^d}, \overleftarrow{\boldsymbol{h}_t^d}, \overleftarrow{\boldsymbol{h}_{t+1}^d} \in \mathbb{R}^H$, and $\boldsymbol{h}_t^d \in \mathbb{R}^{2H}$.

Sliding Temporal Attention: Even though the LSTM encoder-decoder framework is able to model time series data, its performance deteriorates when the sequence length increases due to the memory update mechanism. Various temporal attention mechanisms have been tried before [19, 31] to address this problem. For example, Cinar *et al.* [19] proposed the position-based attention model over the entire history to capture pseudo-periods in the history. However, this model can be significantly diluted when applied over a long history. To counter this problem, Fan *et al.* [31] proposed to use the hidden state in each decoding step to attend to different parts of the history data, and utilized a multimodal fusion scheme to combine the attended results. The purposes of separating the whole history into parts are two-fold: (1) Better attention scores are learned on shorter sequences; (2) Each part of the history mimics the period in the sequence such as business cycles (one month or one quarter) [31]. However, since there are no overlaps between any two adjacent parts, some important time spans may be inappropriately separated into two attention parts. To address this problem, a sliding attention scheme is devised, which performs attention over the parts of the history generated by a sliding window.

Specifically, at the decoding stage, a window with fixed length (shorter than the history length) slides over the encoding history with a specific sliding step, thereby generating a list of sub-sequences. Let $\boldsymbol{h}_i^m$ denote the i-th hidden states in m-th encoder sub-sequence, the temporal attention weights of $\boldsymbol{h}_i^m$ for the t-th decoder step is computed as

$$p_{mi}^t = \boldsymbol{v}_p^T \tanh(\boldsymbol{W}_p \boldsymbol{h}_t^d + \boldsymbol{V}_p \boldsymbol{h}_i^m + \boldsymbol{b}_p), \tag{6.8}$$

$$\gamma_{mi}^t = \frac{\exp(p_{mi}^t)}{\sum_{j=1}^{T_a} \exp(p_{mj}^t)}, \tag{6.9}$$

where T_a is the length of temporal attention window. Then the attended content vectors $\boldsymbol{c}_t$ and the transformed $\boldsymbol{d}_t$ of sub-sequence m are:

$$\boldsymbol{c}_m^t = \sum_{t_{i=1}}^{T_a} \gamma_{mi}^t \boldsymbol{h}_i^m \tag{6.10}$$

$$\boldsymbol{d}_m^t = \mathrm{ReLU}(\boldsymbol{W}_d \boldsymbol{c}_m^t + \boldsymbol{b}_d) \tag{6.11}$$

As shown in the right-top in Figure 6.3, the temporal attention of each sliding window performs independently on the encoded history and the results of which are fused together with the multimodal attention [31]. The fusing weights $\phi_{1\ldots M}^t$ is specifically obtained as follows:

$$q_m^t = \boldsymbol{v}_q^T \tanh(\boldsymbol{W}_q \boldsymbol{h}_t^d + \boldsymbol{V}_q \boldsymbol{d}_m^t + \boldsymbol{b}_q), \tag{6.12}$$

$$\phi_m^t = \frac{\exp(q_m^t)}{\sum_{k=1}^{M} \exp(q_k^t)}. \tag{6.13}$$

Note that M indicates the number of sub-sequences generated by sliding window, which is determined by T_a and the sliding step l. Finally, the overall information obtained by the sliding attention from the history encoding is:

$$\boldsymbol{x}_t = \sum_{m=1}^{M} \phi_m^t \boldsymbol{d}_m^t. \tag{6.14}$$

The hidden states of the BiLSTM decoder and the attention information are then concatenated and generate the final enhanced hidden state at each decoding step as follows:

$$\boldsymbol{h}_t^{d*} = \boldsymbol{W}_x[\boldsymbol{h}_t^d, \boldsymbol{x}_t] + \boldsymbol{b}_x. \tag{6.15}$$

All $\boldsymbol{W}$, $\boldsymbol{V}$ are corresponding weight matrices with proper dimensions and all $\boldsymbol{b}$ are the bias.

6.4.3 Model Training

The prediction is made based on the final hidden state at each step for both encoder and decoder stages during training. However, for testing, predictions only happen at the decoder stage. Particularly, linear layers are applied to make predictions in the encoder and decoder respectively:

$$\begin{cases} y_t^e = \boldsymbol{W}_e \boldsymbol{h}_t^e + b_e \\ y_t^d = \boldsymbol{W}_d \boldsymbol{h}_t^{d*} + b_d, \end{cases} \tag{6.16}$$

where $\boldsymbol{W}_e, \boldsymbol{W}_d \in \mathbb{R}^{1\times 2H}$ and $b_e, b_d \in \mathbb{R}$ are the parameters for the linear layer; $y_t^e, y_t^d \in \mathbb{R}$ are the forecasting value at each time step for the encoder and decoder respectively. L1 loss is used to train the whole model, including the encoder loss $L_e(\cdot)$ and decoder loss $L_d(\cdot)$:

$$L = L_e(\boldsymbol{y}_e, \boldsymbol{y}_e^*, \boldsymbol{\theta}_e) + L_d(\boldsymbol{y}_d, \boldsymbol{y}_d^*, \boldsymbol{\theta}_d), \tag{6.17}$$

where $\boldsymbol{\theta}_e$, $\boldsymbol{\theta}_d$ are the model parameters for encoder and decoder respectively; $\boldsymbol{y}_e, \boldsymbol{y}_e^* \in \mathbb{R}^{(T-1)}$ are the prediction and ground truth of the encoder sequence; and $\boldsymbol{y}_d, \boldsymbol{y}_d^* \in \mathbb{R}^{T'}$ are the prediction and ground truth of the decoder sequence.

6.5 Experiments

To verify the effectiveness of the proposed REAR model, extensive experiments are conducted on two datasets. In particular, the following research questions are concerned:

- **RQ1**: Can the REAR model make better fashion trend forecasting compared with the current state-of-the-art models in terms of forecasting accuracy?
- **RQ2**: Can the introduced modules of incorporating multiple relations and sliding attention mechanism improve the performance?
- **RQ3**: How does the REAR model perform in trend forecasting in terms of specific fashion elements, and based on that, how can the model produce insightful fashion trend forecasting?

6.5.1 Experimental Settings

Experimental Setup: The proposed REAR model is evaluated on two fashion trend forecasting datasets, the proposed FIT dataset and the GeoStyle dataset [113]. Each time series in GeoStyle spans over three years and each week has one data point, while FIT dataset spans over five years and has one data point for every half month. Because of the different timespans and granularities of the two datasets, different schema is set to form the data samples. For GeoStyle, one year's historical time series are taken as input and the following half year's trends are forecasting target. For FIT, two year's historical time series are taken as input and two settings are set for the forecasting length: the following 9 months and 12 months. The sliding window strategy is applied on both datasets to generate the aforementioned data samples. As the timespan of GeoStyle is shorter than that of FIT, only the last sample of each full sequence is kept as the testing sample on GeoStyle, while the last 6 samples of each full sequence are kept as the testing samples on FIT. Note that the parts for prediction in all testing samples will not be included in training process. The Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are used as the evaluation metrics [113] as majority time-series prediction works do, which are specifically calculated as follows:

$$MAE = \frac{1}{m}\sum_{i=1}^{m} \|y_i^* - y_i\| \tag{6.18}$$

$$MAPE = \frac{1}{m}\sum_{i=1}^{m} \|\frac{y_i^* - y_i}{y_i^*}\| \tag{6.19}$$

Implementation Details: The embedding size of all feature embeddings is set to 10, including user attributes (including the embeddings of city, age, and gender), fashion elements and timestep. The hidden state size of both encoder and decoder LSTM network is set to 50. Since each sequence in GeoStyle only has one attribute (city) and does not have any user attributes of age and gender, the group embedding is solely composed by the city embedding. The hyper parameters T_a is selected from {12, 24} on both datasets and l is selected from {1, 2, 5, 10, 20} on GeoStyle and {1, 2, 4, 8, 12, 24} on FIT. T_a and l are finally set to $[12, 20]$, $[24, 2]$, $[24, 4]$ respectively for three experimental settings: [GeoStyle, 6-month prediction], [FIT, 9-month prediction], and [FIT, 12-month prediction].

Compared with the statistical models such as AR, VAR, and ES, which learn one model for each time series, deep learning based models learn a unified model for all the time series in the dataset. Thus the varying magnitudes of different time series will hugely affect the deep learning based models. To counter this problem, practical min-max normalization is applied on both datasets. For GeoStyle, to have a fair comparison with the previous methods which is based on the dataset without any normalization, the min-max normalization is firstly conducted before feeding data into the model (all deep learning-based methods), and after prediction, the predicted values are converted back to the original magnitudes using the pre-saved min and max values (*i.e.*, a min-max de-normalization operation). For FIT, the min-max normalization is conducted over the whole dataset and all the training and prediction for both baselines and the proposed model are based on the normalized dataset. During training, a batch of 400 different time series are randomly sampled for each iteration. For the evaluation of the REAR model, a validation set is curated by further separating the testing data: using the data points with odd indexes as the test set, and those with even indexes as the validation set.

To make the evaluation result more robust and convincing (to avoid the case that superb performance is achieved by chance), the model is evaluated in every epoch when it starts to converge. For each evaluation, the average validation results of the last 10 evaluation steps are calculated (starting from the 10th evaluation). When the average validation result achieves the best, the corresponding average testing results are taken as the final performance.

Baselines: The REAR model with several methods introduced as follows:

- **Mean** and **Last**: They use the mean value or the value of last point of the input historical data as the forecasting value.

- **Autoregression (AR)**: It is a linear regressor which uses the linear combination of last few observed values as the forecasting value.

- **Vector Autoregression (VAR)**: It is another stochastic process model which generalizes the univariate AR model by allowing for more than one evolving variable.

- **Exponential Smoothing (ES)** [3]: It aggregates all the historical values with an exponential decayed weight, and the more recent values have higher impact on the forecast.

- **Linear** and **Cyclic** [113]: They are linear or cyclical parametric models which let historical values to fit the specific predefined model.

- **Geostyle** [113]: It is a parametric model combining a linear component and a cyclical component. It is the state-of-the-art fashion trend forecasting method on Geostyle dataset.

- **MM-ATT** [31]: An LSTM-based encoder-decoder framework, which utilizes multimodal attention during decoding. It achieved the state-of-the-art performance for multi-horizon time series forecasting in several tasks not related to fashion.

6.5.2 Overall Performance on Fashion Trend Forecasting

The overall performance of the proposed REAR model for fashion trend forecasting is firstly evaluated by comparing it to that of baselines. The overall results on FIT and GeoStyle are shown in Tables 6.2 and 6.3 respectively. Based on the results, the following observations can be obtained:

(1) The proposed REAR model yields the best performance on both datasets under two evaluation metrics, especially on the FIT dataset where the REAR method outperforms all other competitors by a large margin. For the nine-month prediction, REAR is the only method that achieves MAE lower than **0.09** and

Table 6.2: Performance of REAR and baseline models for fashion trend forecasting on FIT (the lower is better)

Method	Nine months		Twelve months	
	MAE	MAPE	MAE	MAPE
Mean	0.1131	51.50	0.1205	54.99
Last	0.1531	62.20	0.1392	52.85
AR	0.1143	45.44	0.1186	46.69
VAR	0.1042	45.29	0.1021	41.41
ES	0.1531	62.20	0.1392	52.85
Linear	0.1749	58.11	0.1948	64.56
Cyclic	0.1626	56.76	0.1746	60.65
GeoStyle	0.1582	54.93	0.1735	60.20
MM-ATT	0.0898	32.79	0.0976	37.02
REAR	**0.0864**	**29.45**	**0.0951**	**32.26**

Table 6.3: Performance of REAR and baseline models for fashion trend forecasting on GeoStyle(the lower is better)

Method	Six months	
	MAE	MAPE
Mean	0.0292	25.79
Last	0.0226	21.04
AR	0.0211	20.69
VAR	0.0150	17.95
ES	0.0228	20.59
Linear	0.0365	24.40
Cyclic	0.0165	16.64
GeoStyle	0.0149	16.03
MM-ATT	0.0137	15.31
REAR	**0.0134**	**14.360**

MAPE lower than **30**. For the longer prediction of 12 months, REAR outperforms the most competitive baseline (MM-ATT) by over 10% in MAPE. As the FIT dataset covers much more fine-grained fashion elements, more user information, and fashion trend time series with more realistic and complex patterns, it is more challenging to model. Therefore, the baseline methods do not perform well. The proposed REAR method achieves better performance because it is able to more effectively model nonlinearity in data and leverage various relations between different fashion trends.

(2) For the GeoStyle dataset, the proposed REAR still achieves best MAE and MAPE results. However, the difference in performance of various methods is not as significant as for FIT. The possible reasons for such results are manifold. First, GeoStyle is an easier fashion trend dataset as most fashion trend signals in it are highly seasonal. As a result, traditional statistical models such as Cyclic can also achieve preferable prediction accuracy. Second, the prediction length (six months) is relatively shorter, which further lowers the difficulty of the trend forecasting. Moreover, the strength of the REAR method is that it leverages multiple relations to enhance the individual fashion trend forecasting. In the GeoStyle, there are no group relations nor element relations to exploit, which therefore limits the performance of the REAR method. Nevertheless, the basic model of REAR is still quite effective, which helps achieve the best performance, especially outperforming another LSTM-based model MM-ATT.

(3) Overall, most methods perform better for nine-month prediction than one-year prediction on the FIT dataset, including the REAR model. Such results are reasonable because of two reasons. First, the one-year prediction requires to forecast data with longer time horizon, which is apparently more difficult. Second, such setting reduces the quantity of training data. Comparatively, the REAR can achieve more desired performance for the long-period prediction compared to MM-ATT, which shows the strategy of considering trend relations and the sliding temporal attention mechanism especially helpful in more challenging cases.

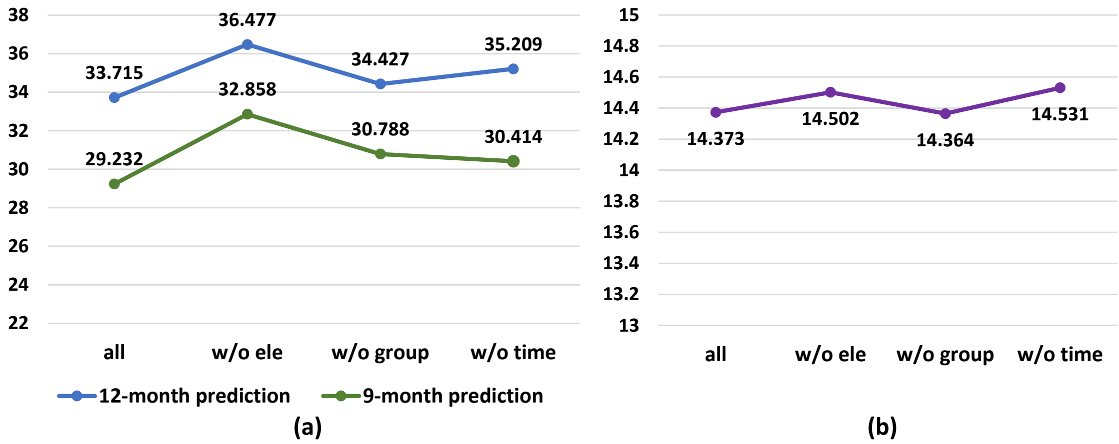

Figure 6.4: Performance of REAR model with and without three types embeddings (MAPE result, the lower the better) on two datasets, (a) FIT and (b) GeoStyle.

6.5.3 Ablation Study

In this section, several ablation experiments are conducted to validate the detailed parts in the proposed model. The effectiveness of three types of content embeddings used as the input in the sequential model are firstly evaluated, *i.e.*, time embedding, group embedding and element embedding (explained in Eq. 6.5). The MAPE results of four models for three experimental settings on two datasets are illustrated in Figure 6.4. The **all** is the model with all three types of embeddings in the basic LSTM encoder-decoder framework, while the remaining three models, *i.e.*, **w/o ele**, **w/o grp**, **w/o time**, represent removing the element embedding, group embedding and time embedding respectively. From the experimental results we can see that for FIT, removing any sequence feature embedding can degrade the overall performance. On GeoStyle, we can discover that both element and time embeddings are effective in improving the performance, yet the group embedding is not. As the group information in GeoStyle is limited, only with the city information, it is not enough to help the model. It is also notable that overall, the element embedding is more significant for the fashion trend modeling compared to the other two, which is probably because the fashion element is still the deterministic factor for fashion trending.

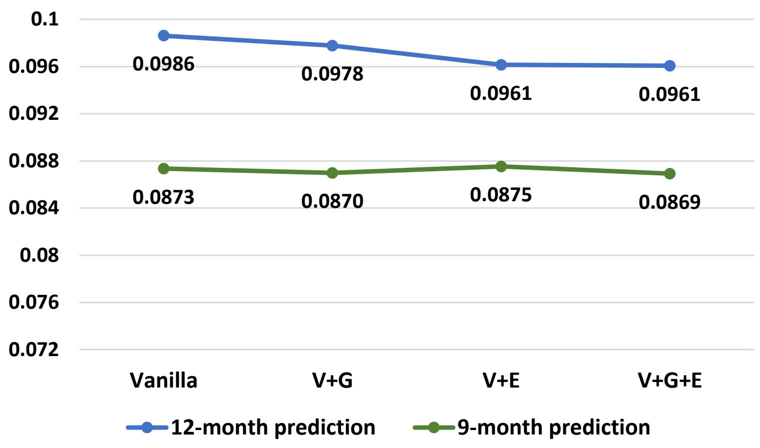

Figure 6.5: Performance of models with and without leveraging relations (MAE result, the lower the better). **V** denotes vanilla, **G** denotes group relation and **E** denotes element relation. Results on FIT dataset.

The effectiveness of incorporating multiple relations in the fashion trend forecasting model is discussed next. Specifically, two types of relations, element relations and group relations, are exploited in the proposed model. The experimental results of models with/without each type of relations are shown in Figure 6.5. The MAE results of **V+G** and **V+E** are better than that of **Vanilla** (no relation leveraged) for the 12-month fashion trend prediction. Such improvement is not quite marked in the 9-month prediction case. However, it is clear that leveraging both relations effectively improves the prediction performance. This proves the effectiveness of introducing relations in helping the sequence modeling of fashion trend signals and the trend prediction. In addition, the more significant improvement for the 12-month prediction concludes that the relation knowledge is particularly helpful for more challenging longer-distance prediction cases.

We discuss in detail the sliding attention mechanism in the REAR model. Specifically, the effectiveness of employing the attention mechanism and the effect of varying sliding steps on the performance of the whole model is analyzed. The performance of the models (MAE results) with varying sliding steps of the temporal attention on two experimental settings on the FIT dataset was reported in Table 6.4. From the results we can see that first, introducing the temporal attention

Table 6.4: Performance of models with temporal attention in different sliding steps. Results on FIT dataset.

sliding steps	Nine months	Twelve months
No Att	0.08692	0.09606
2	**0.08644**	0.09532
4	0.08646	**0.09514**
8	0.08674	0.09571
12	0.09685	0.09712
24	0.08742	0.09615

with sliding steps clearly reduce the MAE for both the 9-month prediction and 12-month prediction in most cases. The difference is that when the sliding step is two, the model performs best for the 9-month prediction while for 12-month prediction, the best setting is four sliding steps. For both prediction settings, the performance worsens when the sliding step becomes large. However, another noteworthy point is that the small sliding step causes large computational cost, which should be considered in comprehensive model evaluation.

6.5.4 Fashion Trend Analysis

This part shows some visualization results of the fashion trend prediction, as well as some in-depth performance analysis to further discuss the effectiveness of the REAR model, especially the multiple relations explored. Figure 6.6 illustrates five examples which are fashion trends of different fashion elements for different groups. To depict the whole picture of data, the entire fashion trend signals are shown and the prediction parts are highlighted in the colored square in the right part. The first two rows compare the prediction results of **Vanilla** model and models leveraging element relations. It is clear from the comparison that taking advantage of element relations makes the prediction of future trend more precise. In the third and fourth cases, we can see the difference of prediction results brought by leveraging group relations. Similar to the first two cases, the

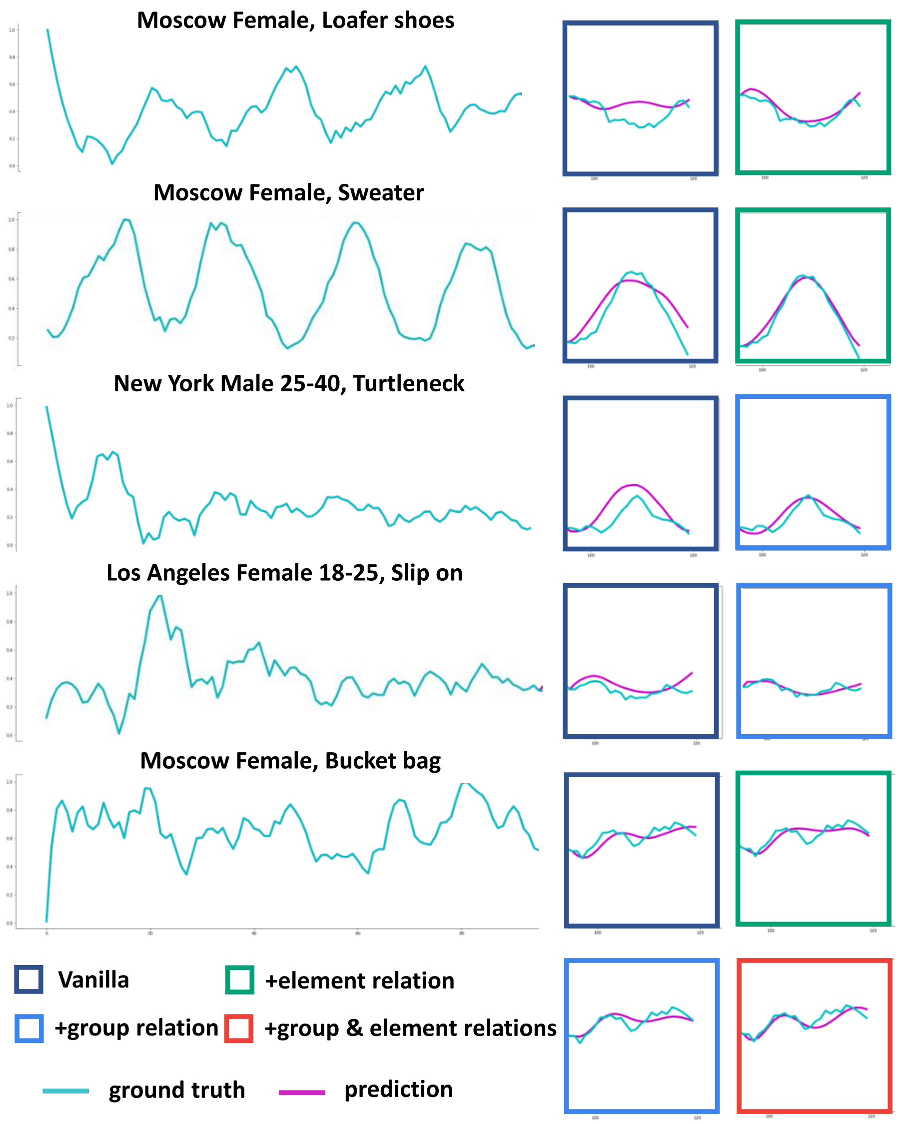

Figure 6.6: Fashion trend forecasting results of models w/o two kinds of relations.

Vanilla model is not good enough to make solid trend predictions, while the models incorporating relations perform better. Last case shows all four prediction results from models without relations, which validates the effectiveness of our assumptions about incorporating multiple relations to boost the fashion trend prediction performance. Such observations from the visualization results are consistent to the results of quantitative analysis in the ablation study part.

[5] A/W means the Fall/Winter season, which starts around late July and ends in December. S/S means the Spring/Summer season, which usually goes from January to June each year.

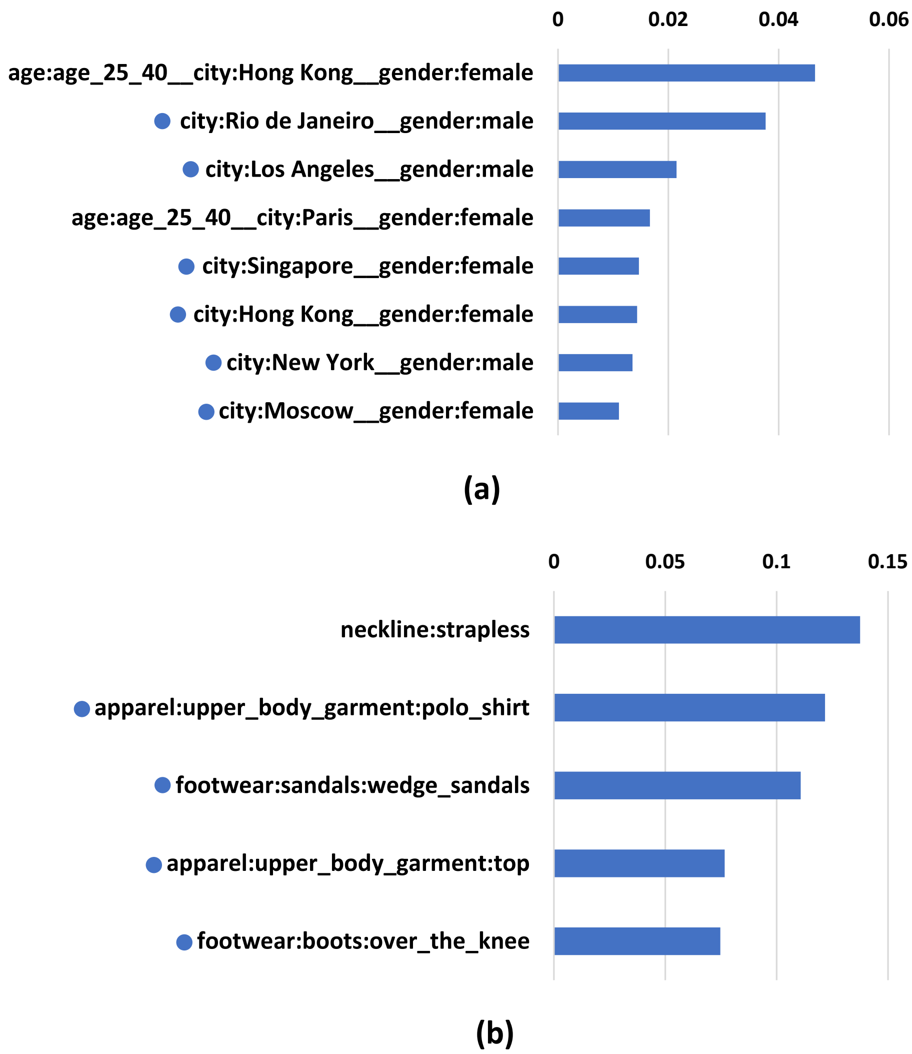

Figure 6.7: Performance improvement (relative improvement on MAE) by applying relations on specific groups and fashion elements. (a) Top 8 groups in terms of performance improvement by leveraging group relations; (b) Top 5 fashion elements in terms of performance improvement by leveraging element relations. Results on 9-month prediction on FIT dataset.

The importance of incorporation of two kinds of relations in fashion trend prediction is analyzed in a more detailed manner in this section. Specifically, the performance change of fashion trends belonging to **different groups** made by considering and modeling the **group relations** is investigated, with the top eight groups whose fashion trend prediction is improved the most shown in Figure 6.7 (a). From the result we can see that the most significantly improved groups with the help of group relation incorporation are those of coarse-grained (such

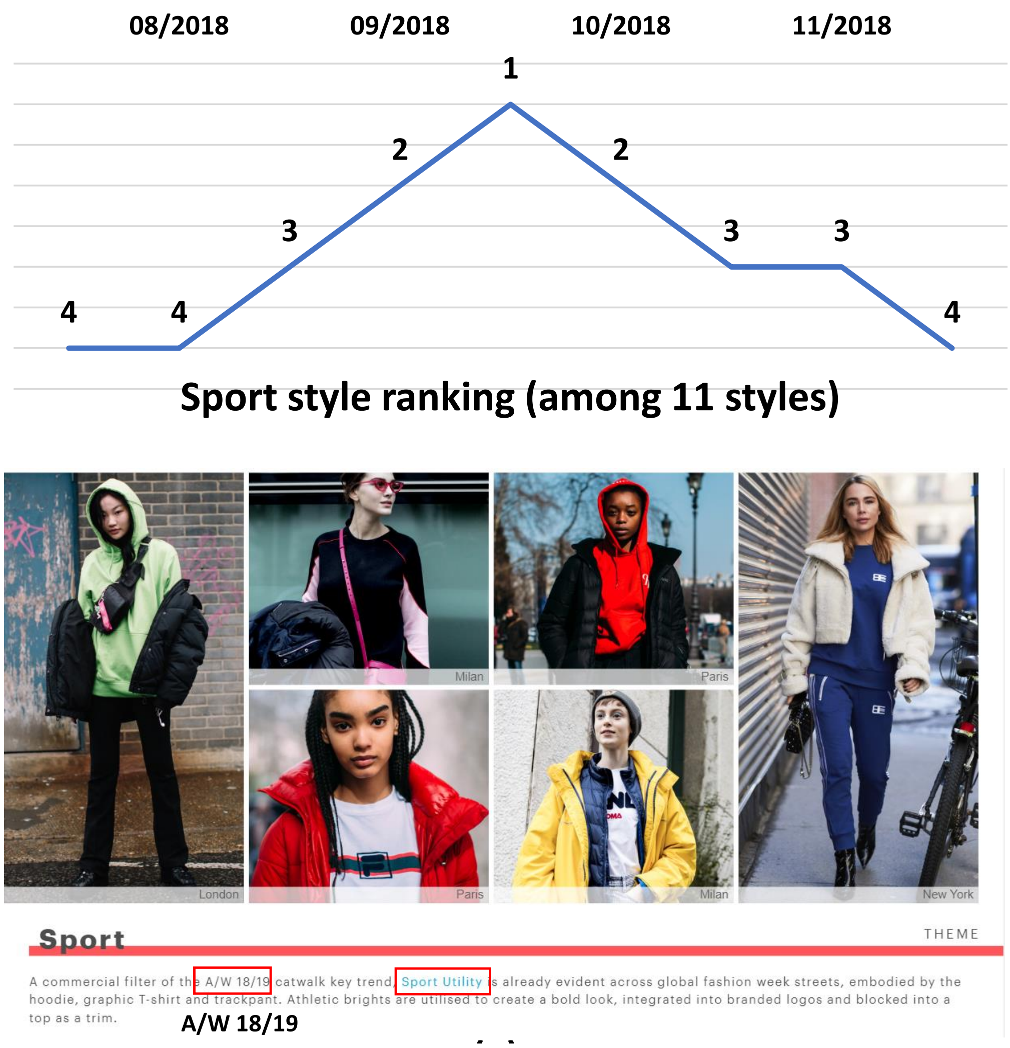

Figure 6.8: Trend prediction of the fashion style **sporty**. top: predicted ranking of **sporty** by REAR model; bottom: professional trendy style prediction produced by WSGN for A/W 18/19. [5]

as Rio de Janeiro Female, without age group). Such a result is reasonable as in REAR model, the coarse-grained groups adsorb the information from their affiliating groups, therefore the representation of these groups is enhanced. The impact of **element relations** on trends of **different fashion elements** is also investigated and similar results have been found. Figure 6.7 (b) illustrates the five most improved fashion elements in terms of fashion trend prediction. Most of these fashion elements belong to the higher-level node in the taxonomy (*e.g.*, category). Recall that the groups or elements relations are incorporated in REAR by effectively passing information from the fined-grained affiliation nodes to their

Figure 6.9: Color trend prediction. top: predicted top 5 colors among four major fashion metropolis in June 2018; bottom: professional trendy color prediction produced by WSGN for S/S 2018.

parent nodes. In practice, the parent nodes are exactly those wider covering groups and higher-level fashion elements. As discussed above, these parent nodes adsorb more information, and thus are better modeled to achieve more preferable performance.

The proposed REAR model is able to make fashion trend predictions regarding specific fashion elements for a wide range of time. Figures 6.8 and 6.9 show two cases of fashion trend prediction from the proposed REAR model and compare that with the corresponding future trends made by the professional fashion forecasting agency WGSN [4]. The trend forecasting from WGSN is expert-based while our forecasting is data-based. The chart on top in Figure 6.8 shows the ranking result of the style **sporty** based on the prediction result of REAR, from which we can see that the **sporty style** is predicted to become more popular since September 2018. Such forecasting results show consistency with WGSN, which forecasts that the **sport** style is the key trend in the A/W 18/19 season. Figure 6.9 shows the color forecasting results. Specifically, on the top, the top five popular colors in June 2018 for four major fashion cities predicted by the REAR model is shown, and the major cities are Paris, New York, London and Milan. We can see that based on REAR model, the most trendy colors in June 2018 include black, red, blue, and beige, while grey and pink are also popular in some cities. Meanwhile, the WGSN produces the S/S 18 street color trend, it is clear that most of them are in the top popular color list produced by REAR(matching predictions are marked by ticks). From the two cases, we can see that the proposed REAR model not only predicts the future trend, but also generates the forecasting results that are similar to the old-fashion expert-based fashion forecasting results produced by professional fashion forecasting agencies.

6.6 Summary

This chapter addresses the fashion trend forecasting problem based on social media data, aiming to mine the complex patterns in the historical time series records of fashion elements and accordingly predict the future trends. Specifically, over 190 fashion elements are studied and three types of user information are involved towards meaningful fashion trend forecasting. An effective model, REAR, is proposed to leverage multiple relations between specific fashion trends and therefore capture the complex patterns in the time-series data and effectively forecast fashion trends.

[4]WGSN is one of the most acknowledged professional fashion forecasting agency. The results of WGSN forecasting are obtained based on the fashion forecasting reports from the WSGN website, https://www.wgsn.com/fashion/

7 Conclusion and Future Work

This thesis works on the data-driven fashion advising task based on deep learning technologies. Targeting at two key perspectives of fashion advising, personalization and fashionability, it focuses on two research tasks, which are personalized fashion recommendation and fashion trend forecasting. Specific conclusions, the limitation of this thesis and the potential future directions are given in this chapter.

7.1 Conclusion

Three main research objectives are extracted from the two tasks: 1) modeling shopping patterns in fashion shopping to enhance the personalized fashion recommendation, 2) modeling content-level relational item-item transition for sequential fashion recommendation, and 3) fashion trend modeling and forecasting based on social media data. The following three works are specifically conducted in this thesis:

(1) A graph-based Field-aware Graph Collaborative Filtering (FGCF) method is developed to capture the fine-grained user shopping patterns. Leveraging associated fashion factors, which are grouped into different factor fields, the field-level interactions between the user and items are effectively modeled. Such operation specifies the user's preference and enhance the modeling of it, which further improves the performance of the fashion recommendation model. Based on the Amazon review data, we construct two fashion datasets: Amazon-women and Amazon-men for the evaluation of the proposed method. Through extensive experiments on the two datasets, our FGCF model is demonstrated effective in

making personalized fashion recommendation. It also able to discover some shopping patterns from user historical behaviors.

(2) An Attentional Content-level Translation-based Fashion Recommender (ACTR) is proposed to model both the user-item compatibility and sequential dynamics among items. To enhance the item-item transition modeling, the proposed method leverages the item-item relationships and model the content-level item interactions from different fashion aspects. The final recommendation results are obtained by aggregating the content-level transitions with the attention mechanism, as well as the long-term user preference. To facilitate this study, we apply the iFashion dataset and prepare two settings based on it with various sequence lengths. Extensive experiments on the two settings demonstrate the effectiveness of the proposed method.

(3) An RNN-based Relational Enhanced Attention Recurrent (REAR) network is proposed to model the specific fashion trends of fine-grained fashion element trends and specific user groups, which are deemed effective to reveal the real fashion trend. The proposed model takes advantage of the capability of deep recurrent neural networks in modeling time-series data and connects the fashion trends signals with multiple relations to further enhance the trend modeling. We collect a new FIT dataset based on Instagram and set different input for this study and forecasting lengths for different experimental settings. Experimental results on our FIT dataset and another GeoStyle dataset show the Superity of the proposed REAR method in fashion trend forecasting.

In summary, the first two models well address the personalized fashion recommendation problem in non-sequential and sequential manners respectively. The two methods focus on different aspects to improve the performance for personalized fashion recommendation, but both incorporate the characteristics of fashion domain. The third model focuses on the perspective of fashionability in fashion advising, which is able to effectively forecast a period of fashion trend in the future, with regard to specific fashion elements and user groups. All three methods address the corresponding tasks effectively and contribute to the research of data-driven fashion advising based on deep learning from two different key perspectives.

7.2 Future Work

Despite of the achievement of this thesis, there still exist limitations and several potential directions for future work.

(1) In the first work to model the user shopping patterns, only six main fashion domain factor fields are considered, which means that only the fashion patterns related to the involved factor fields can be modeled. In real-life applications, the shopping patterns of users can be also determined by external factors, such as promotions. Therefore, the shopping patterns in fashion domain can be investigated in-depth and comprehensively by leveraging more influential factors and therefore enhance the exploration of user preference.

(2) In the second work, only two simple relationships between fashion items are incorporated, namely, the substitution and mix-and-match, which are mostly defined from the functional perspective. Further improvement can be made by specifying the item relationships from fashion perspective, such as considering more aesthetic factors. Moreover, the proposed model is based on the translation-based sequential recommendation model, which has shown much improvement in performance with several technical contributions. In the future work, the main technical improvement in the proposed model, *i.e.*, the content-level relational transition modeling, can be applied in other basic sequential recommendation models and evaluated for the effectiveness.

(3) In the third work, the study on fashion trends is solely based on the social media data, which reveal the fashion trends among ordinary people. The forecasting

of the fashion trends are only based on modeling and analyzing the historical trend records among ordinary people. However, according to the fashion diffusion theory [10], fashion trends for masses can be largely affected by users with higher fashion-conscious and the marketing activities from fashion brands. If these exogenous influential factors which can lead the fashion trends to a certain extent can be explored, the forecasting of fashion trends would be further improved.

(4) In this thesis, the personalization and fashionability in fashion advising are studied as two independent research topics separately, which is the main limitation of this study. In the future work, the fashion trend forecasting results can be leveraged into the personalized fashion recommendation models to generate better recommendation results for users which not only cater to their personal fashion taste, but also provide them new and inspiring fashion suggestions.

References

[1] Gediminas Adomavicius and Alexander Tuzhilin. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions". In: *IEEE transactions on knowledge and data engineering* 17.6 (2005), pp. 734–749.

[2] Ziad Al-Halah and Kristen Grauman. "From Paris to Berlin: Discovering Fashion Style Influences Around the World". In: *arXiv preprint arXiv:2004.01316* (2020).

[3] Ziad Al-Halah, Rainer Stiefelhagen, and Kristen Grauman. "Fashion forward: Forecasting visual style in fashion". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 388–397.

[4] Grigory Antipov, Moez Baccouche, Sid-Ahmed Berrani, and Jean-Luc Dugelay. "Effective training of convolutional neural networks for face-based gender and age prediction". In: *Pattern Recognition* 72 (2017), pp. 15–26.

[5] Abdelhadi Azzouni and Guy Pujolle. "A long short-term memory recurrent neural network framework for network traffic matrix prediction". In: *arXiv preprint arXiv:1705.05690* (2017).

[6] Kasun Bandara, Peibei Shi, Christoph Bergmeir, Hansika Hewamalage, Quoc Tran, and Brian Seaman. "Sales demand forecast in e-commerce using a long short-term memory neural network methodology". In: *International Conference on Neural Information Processing*. Springer. 2019, pp. 462–474.

[7] Peter Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, et al. "Interaction networks for learning about objects, relations and physics". In: *Advances in neural information processing systems*. 2016, pp. 4502–4510.

[8] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. "Translating embeddings for modeling multi-relational data". In: *Advances in neural information processing systems*. 2013, pp. 2787–2795.

[9] George EP Box and Gwilym M Jenkins. "Some recent advances in forecasting and control". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 17.2 (1968), pp. 91–109.

[10] Evelyn L Brannon. *Fashion forecasting*. Fairchild Books, 2005.

[11] Yixin Cao, Xiang Wang, Xiangnan He, Zikun Hu, and Tat-Seng Chua. "Unifying Knowledge Graph Learning and Recommendation: Towards a Better Understanding of User Preference". In: (2019).

[12] Minmin Chen, Zhixiang Xu, Kilian Weinberger, and Fei Sha. "Marginalized denoising autoencoders for domain adaptation". In: *arXiv preprint arXiv:1206.4683* (2012).

[13] Qiang Chen, Junshi Huang, Rogerio Feris, Lisa M Brown, Jian Dong, and Shuicheng Yan. "Deep domain adaptation for describing people based on fine-grained clothing attributes". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 5315–5324.

[14] Wen Chen, Pipei Huang, Jiaming Xu, Xin Guo, Cheng Guo, Fei Sun, Chao Li, Andreas Pfadler, Huan Zhao, and Binqiang Zhao. "POG: Personalized Outfit Generation for Fashion Recommendation at Alibaba iFashion". In: *arXiv preprint arXiv:1905.01866* (2019).

[15] Xu Chen, Hanxiong Chen, Hongteng Xu, Yongfeng Zhang, Yixin Cao, Zheng Qin, and Hongyuan Zha. "Personalized Fashion Recommendation with Visual Explanations Based on Multimodal Attention Network: Towards Visually Explainable Recommendation". In: *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. Paris, France, 2019, pp. 765–774. ISBN: 978-1-4503-6172-9.

[16] Wen-Huang Cheng, Sijie Song, Chieh-Yun Chen, Shintami Chusnul Hidayati, and Jiaying Liu. "Fashion Meets Computer Vision: A Survey". In: *arXiv preprint arXiv:2003.13988* (2020).

[17] Zhi-Qi Cheng, Xiao Wu, Yang Liu, and Xian-Sheng Hua. "Video2shop: Exact matching clothes in videos to online shopping images". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 4048–4056.

[18] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. "Empirical evaluation of gated recurrent neural networks on sequence modeling". In: *arXiv preprint arXiv:1412.3555* (2014).

[19] Yagmur Gizem Cinar, Hamid Mirisaee, Parantapa Goswami, Eric Gaussier, Ali Aït-Bachir, and Vadim Strijov. "Position-based content attention for time series forecasting with sequence-to-sequence rnns". In: *International Conference on Neural Information Processing*. Springer. 2017, pp. 533–544.

[20] Guillem Cucurull, Perouz Taslakian, and David Vazquez. "Context-aware visual compatibility prediction". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 12617–12626.

[21] Zeyu Cui, Zekun Li, Shu Wu, Xiao-Yu Zhang, and Liang Wang. "Dressing as a whole: Outfit compatibility learning based on node-wise graph neural networks". In: *Proceedings of the International Conference on World Wide Web*. 2019, pp. 307–317.

[22] Christian Desrosiers and George Karypis. "A comprehensive survey of neighborhood-based recommendation methods". In: *Recommender systems handbook*. Springer, 2011, pp. 107–144.

[23] Wei Di, Catherine Wah, Anurag Bhardwaj, Robinson Piramuthu, and Neel Sundaresan. "Style finder: Fine-grained clothing style detection and retrieval". In: *Proceedings of the IEEE Conference on computer vision and pattern recognition workshops*. 2013, pp. 8–13.

[24] Haoye Dong, Xiaodan Liang, Ke Gong, Hanjiang Lai, Jia Zhu, and Jian Yin. "Soft-gated warping-gan for pose-guided person image synthesis". In: *Advances in neural information processing systems*. 2018, pp. 474–484.

[25] Haoye Dong, Xiaodan Liang, Xiaohui Shen, Bochao Wang, Hanjiang Lai, Jia Zhu, Zhiting Hu, and Jian Yin. "Towards multi-pose guided virtual try-on network". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, pp. 9026–9035.

[26] Qi Dong, Shaogang Gong, and Xiatian Zhu. "Multi-task curriculum transfer deep learning of clothing attributes". In: *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2017, pp. 520–529.

[27] Xue Dong, Xuemeng Song, Fuli Feng, Peiguang Jing, Xin-Shun Xu, and Liqiang Nie. "Personalized Capsule Wardrobe Creation with Garment and User Modeling". In: *Proceedings of the International Conference on Multimedia*. ACM. 2019, pp. 302–310.

[28] Tim Donkers, Benedikt Loepp, and Jürgen Ziegler. "Sequential user-based recurrent neural network recommendations". In: *Proceedings of the Eleventh ACM Conference on Recommender Systems*. 2017, pp. 152–160.

[29] Mikayla DuBreuil and Sheng Lu. "Traditional vs. big-data fashion trend forecasting: an examination using WGSN and EDITED". In: *International Journal of Fashion Design, Technology and Education* 13.1 (2020), pp. 68–77.

[30] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. "A multi-view deep learning approach for cross domain user modeling in recommendation systems". In: *Proceedings of the International Conference on World Wide Web*. 2015, pp. 278–288.

[31] Chenyou Fan, Yuze Zhang, Yi Pan, Xiaoyue Li, Chi Zhang, Rong Yuan, Di Wu, Wensheng Wang, Jian Pei, and Heng Huang. "Multi-horizon time series forecasting with temporal attention learning". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2019, pp. 2527–2535.

[32] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. "Graph neural networks for social recommendation". In: *Proceedings of the International Conference on World Wide Web*. 2019, pp. 417–426.

[33] Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. "Temporal relational ranking for stock prediction". In: *ACM Transactions on Information Systems (TOIS)* 37.2 (2019), pp. 1–30.

[34] Shanshan Feng, Xutao Li, Yifeng Zeng, Gao Cong, Yeow Meng Chee, and Quan Yuan. "Personalized ranking metric embedding for next new poi recommendation". In: *Proceddings of the International Joint Conference on Artificial Intelligence*. 2015.

[35] Alex Fout, Jonathon Byrd, Basir Shariat, and Asa Ben-Hur. "Protein interface prediction using graph convolutional networks". In: *Advances in neural information processing systems*. 2017, pp. 6530–6539.

[36] Julia Gaimster. "The changing landscape of fashion forecasting". In: *International Journal of Fashion Design, Technology and Education* 5.3 (2012), pp. 169–178.

[37] Alberto Garcia-Duran, Roberto Gonzalez, Daniel Onoro-Rubio, Mathias Niepert, and Hui Li. "Transrev: Modeling reviews as translations from users to items". In: *European Conference on Information Retrieval*. Springer. 2020, pp. 234–248.

[38] Ke Gong, Xiaodan Liang, Dongyu Zhang, Xiaohui Shen, and Liang Lin. "Look into person: Self-supervised structure-sensitive learning and a new benchmark for human parsing". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 932–940.

[39] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[40] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial nets". In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

[41] Shubham Goswami and Shagufta Khan. "Impact of consumer decision-making styles on online apparel consumption in India". In: *Vision* 19.4 (2015), pp. 303–311.

[42] C Grammenos. *The evolution of "big data" in fashion forecasting*. 2015.

[43] Alex Graves. "Generating sequences with recurrent neural networks". In: *arXiv preprint arXiv:1308.0850* (2013).

[44] Sida Gu, Xiaoqiang Liu, Lizhi Cai, and Jie Shen. "Fashion coordinates recommendation based on user behavior and visual clothing style". In: *Proceedings of the 3rd International Conference on Communication and Information Processing*. ACM. 2017, pp. 185–189.

[45] Xiaoling Gu, Fei Gao, Min Tan, and Pai Peng. "Fashion analysis and understanding with artificial intelligence". In: *Information Processing & Management* 57.5 (2020), p. 102276.

[46] Raia Hadsell, Sumit Chopra, and Yann LeCun. "Dimensionality reduction by learning an invariant mapping". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 2. IEEE. 2006, pp. 1735–1742.

[47] Takuo Hamaguchi, Hidekazu Oiwa, Masashi Shimbo, and Yuji Matsumoto. "Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach". In: *arXiv preprint arXiv:1706.05674* (2017).

[48] Will Hamilton, Zhitao Ying, and Jure Leskovec. "Inductive representation learning on large graphs". In: *Advances in neural information processing systems*. 2017, pp. 1024–1034.

[49] Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S Davis. "Learning fashion compatibility with bidirectional lstms". In: *Proceedings of the International Conference on Multimedia*. ACM. 2017, pp. 1078–1086.

[50] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S Davis. "Viton: An image-based virtual try-on network". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7543–7552.

[51] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[52] Ruining He, Wang-Cheng Kang, and Julian McAuley. "Translation-based recommendation". In: *Proceedings of the eleventh ACM conference on recommender systems*. 2017, pp. 161–169.

[53] Ruining He and Julian McAuley. "VBPR: visual bayesian personalized ranking from implicit feedback". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2016.

[54] Ruining He, Charles Packer, and Julian McAuley. "Learning compatibility across categories for heterogeneous item recommendation". In: *Proceedings of the International Conference on Data Mining*. IEEE. 2016, pp. 937–942.

[55] Xiangnan He and Tat-Seng Chua. "Neural factorization machines for sparse predictive analytics". In: *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2017, pp. 355–364.

[56] Xiangnan He, Xiaoyu Du, Xiang Wang, Feng Tian, Jinhui Tang, and Tat-Seng Chua. "Outer product-based neural collaborative filtering". In: *arXiv preprint arXiv:1808.03912* (2018).

[57] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. "Neural collaborative filtering". In: *WWW*. 2017, pp. 173–182.

[58] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. "Neural collaborative filtering". In: *Proceedings of the International Conference on World Wide Web*. 2017, pp. 173–182.

[59] Balázs Hidasi and Alexandros Karatzoglou. "Recurrent neural networks with top-k gains for session-based recommendations". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 2018, pp. 843–852.

[60] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. “Session-based recommendations with recurrent neural networks”. In: *arXiv preprint arXiv:1511.06939* (2015).

[61] Shintami C. Hidayati, Kai-Lung Hua, Wen-Huang Cheng, and Shih-Wei Sun. “What Are the Fashion Trends in New York?” In: *Proceedings of the International Conference on Multimedia*. ACM. 2014, 197–200. ISBN: 9781450330633.

[62] Sepp Hochreiter and Jürgen Schmidhuber. “Long short-term memory”. In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[63] Charles C Holt. “Forecasting seasonals and trends by exponentially weighted moving averages”. In: *International journal of forecasting* 20.1 (2004), pp. 5–10.

[64] Min Hou, Le Wu, Enhong Chen, Zhi Li, Vincent W Zheng, and Qi Liu. “Explainable Fashion Recommendation: A Semantic Attribute Region Guided Approach”. In: *arXiv preprint arXiv:1905.12862* (2019).

[65] Wei-Lin Hsiao and Kristen Grauman. “Creating capsule wardrobes from fashion images”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 7161–7170.

[66] Yang Hu, Xi Yi, and Larry S Davis. “Collaborative fashion recommendation: A functional tensor factorization approach”. In: *Proceedings of the International Conference on Multimedia*. ACM. 2015, pp. 129–138.

[67] Yifan Hu, Yehuda Koren, and Chris Volinsky. “Collaborative filtering for implicit feedback datasets”. In: *2008 Eighth IEEE International Conference on Data Mining*. Ieee. 2008, pp. 263–272.

[68] Junshi Huang, Rogerio S Feris, Qiang Chen, and Shuicheng Yan. “Cross-domain image retrieval with a dual attribute-aware ranking network”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1062–1070.

[69] Hyunwoo Hwangbo, Yang Sok Kim, and Kyung Jin Cha. “Recommendation system development for fashion retail e-commerce”. In: *Electronic Commerce Research and Applications* 28 (2018), pp. 94–101.

[70] Ayelet Israeli and Jill Avery. “Predicting consumer tastes with big data at Gap”. In: *Harvard Business School Publication (July 10)* (2017).

[71] Tim Jackson. “The process of trend development leading to a fashion season”. In: *Fashion Marketing*. Routledge, 2007, pp. 192–211.

[72] Dietmar Jannach and Malte Ludewig. "Determining characteristics of successful recommendations from log data: a case study". In: *Proceedings of the Symposium on Applied Computing*. 2017, pp. 1643–1648.

[73] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender systems: an introduction*. Cambridge University Press, 2010.

[74] Menglin Jia, Yichen Zhou, Mengyun Shi, and Bharath Hariharan. "A deep-learning-based fashion attributes detection model". In: *arXiv preprint arXiv:1810.10148* (2018).

[75] Shuhui Jiang and Yun Fu. "Fashion style generator". In: *Proceedings of the International Joint Conference on Artificial Intelligence*. 2017, pp. 3721–3727.

[76] Shuhui Jiang, Ming Shao, Chengcheng Jia, and Yun Fu. "Consensus style centralizing auto-encoder for weak style classification". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2016.

[77] Shuhui Jiang, Yue Wu, and Yun Fu. "Deep bi-directional cross-triplet embedding for cross-domain clothing retrieval". In: *Proceedings of the International Conference on Multimedia*. ACM. 2016, pp. 52–56.

[78] Yuchin Juan, Yong Zhuang, Wei-Sheng Chin, and Chih-Jen Lin. "Field-aware factorization machines for CTR prediction". In: *RecSys*. 2016, pp. 43–50.

[79] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian McAuley. "Visually-aware fashion recommendation and design with generative image models". In: *Proceedings of the International Conference on Data Mining*. IEEE. 2017, pp. 207–216.

[80] Wang-Cheng Kang, Mengting Wan, and Julian McAuley. "Recommendation through mixtures of heterogeneous item relationships". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 2018, pp. 1143–1152.

[81] Elias Khalil, Hanjun Dai, Yuyu Zhang, Bistra Dilkina, and Le Song. "Learning combinatorial optimization algorithms over graphs". In: *Advances in Neural Information Processing Systems*. 2017, pp. 6348–6358.

[82] M Hadi Kiapour, Kota Yamaguchi, Alexander C Berg, and Tamara L Berg. "Hipster wars: Discovering elements of fashion styles". In: *Proceedings of the European Conference on Computer Vision*. Springer. 2014, pp. 472–488.

[83] Eundeok Kim, Ann Marie Fiore, and Hyejeong Kim. *Fashion trends: analysis and forecasting*. Berg, 2013.

[84] Thomas N Kipf and Max Welling. "Semi-supervised classification with graph convolutional networks". In: *arXiv preprint arXiv:1609.02907* (2016).

[85] Emine Koca and Fatma Koc. "A Study of Clothing Purchasing Behavior By Gender with Respect to Fashion and Brand Awareness". In: *European Scientific Journal* 12.7 (2016), pp. 234–248.

[86] Yehuda Koren. "Collaborative filtering with temporal dynamics". In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2009, pp. 447–456.

[87] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[88] Himabindu Lakkaraju, Julian McAuley, and Jure Leskovec. "What's in a name? understanding the interplay between titles, content, and communities in social media". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2013.

[89] Martin Längkvist, Lars Karlsson, and Amy Loutfi. "A review of unsupervised feature learning and deep learning for time-series modeling". In: *Pattern Recognition Letters* 42 (2014), pp. 11–24.

[90] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), p. 436.

[91] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[92] Hui Li, Ye Liu, Nikos Mamoulis, and David S Rosenblum. "Translation-based sequential recommendation for complex users on sparse data". In: *IEEE Transactions on Knowledge and Data Engineering* (2019).

[93] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. "Neural attentive session-based recommendation". In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 2017, pp. 1419–1428.

[94] Xingchen Li, Xiang Wang, Xiangnan He, Long Chen, Jun Xiao, and Tat-Seng Chua. "Hierarchical Fashion Graph Network for Personalized Outfit Recommendation". In: *arXiv preprint arXiv:2005.12566* (2020).

[95] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting". In: *arXiv preprint arXiv:1707.01926* (2017).

[96] Zongmin Li, Yante Li, Weiwei Tian, Yunping Pang, and Yujie Liu. "Cross-scenario clothing retrieval and fine-grained style recognition". In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE. 2016, pp. 2912–2917.

[97] Yuxuan Liang, Songyu Ke, Junbo Zhang, Xiuwen Yi, and Yu Zheng. "Geoman: Multi-level attention networks for geo-sensory time series prediction." In: *Proceedings of the International Joint Conference on Artificial Intelligence*. 2018, pp. 3428–3434.

[98] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. "Learning entity and relation embeddings for knowledge graph completion". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2015.

[99] Yujie Lin, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Jun Ma, and Maarten de Rijke. "Improving outfit recommendation with co-supervision of fashion generation". In: *Proceedings of the International Conference on World Wide Web*. 2019, pp. 1095–1105.

[100] Yusan Lin, Maryam Moosaei, and Hao Yang. "OutfitNet: Fashion Outfit Recommendation with Attention-Based Multiple Instance Learning". In: *Proceedings of The Web Conference 2020*. 2020, pp. 77–87.

[101] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. "Personalized news recommendation based on click behavior". In: *Proceedings of the 15th international conference on Intelligent user interfaces*. 2010, pp. 31–40.

[102] Qiang Liu, Shu Wu, and Liang Wang. "DeepStyle: Learning user preferences for visual recommendation". In: *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2017, pp. 841–844.

[103] Qiao Liu, Yifu Zeng, Refuoe Mokhosi, and Haibin Zhang. "STAMP: short-term attention/memory priority model for session-based recommendation". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2018, pp. 1831–1839.

[104] Si Liu, Zheng Song, Guangcan Liu, Changsheng Xu, Hanqing Lu, and Shuicheng Yan. "Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2012, pp. 3330–3337.

[105] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. “Deepfashion: Powering robust clothes recognition and retrieval with rich annotations”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 1096–1104.

[106] Ziwei Liu, Sijie Yan, Ping Luo, Xiaogang Wang, and Xiaoou Tang. “Fashion landmark detection in the wild”. In: *Proceedings of the European Conference on Computer Vision*. Springer. 2016, pp. 229–245.

[107] Fred C Lunenburg. “THE DECISION MAKING PROCESS.” In: *National Forum of Educational Administration & Supervision Journal*. Vol. 27. 4. 2010.

[108] Yawei Luo, Zhedong Zheng, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. “Macro-micro adversarial network for human parsing”. In: *Proceedings of the European Conference on Computer Vision*. 2018, pp. 418–434.

[109] Yihui Ma, Jia Jia, Suping Zhou, Jingtian Fu, Yejun Liu, and Zijian Tong. “Towards better understanding the clothing fashion styles: A multimodal deep learning approach”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2017.

[110] Yunshan Ma, Yujuan Ding, Xun Yang, Lizi Liao, Wai Keung Wong, and Tat-Seng Chua. “Knowledge Enhanced Neural Fashion Trend Forecasting”. In: *Proceedings of the International Conference on Multimedia Retrieval*. ACM. 2020, pp. 82–90.

[111] Yunshan Ma, Lizi Liao, and Tat-Seng Chua. “Automatic Fashion Knowledge Extraction from Social Media”. In: *Proceedings of the International Conference on Multimedia*. ACM. 2019, pp. 2223–2224.

[112] Yunshan Ma, Xun Yang, Lizi Liao, Yixin Cao, and Tat-Seng Chua. “Who, Where, and What to Wear? Extracting Fashion Knowledge from Social Media”. In: *Proceedings of the International Conference on Multimedia*. ACM. 2019, pp. 257–265.

[113] Utkarsh Mall, Kevin Matzen, Bharath Hariharan, Noah Snavely, and Kavita Bala. “GeoStyle: Discovering Fashion Trends and Events”. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, pp. 411–420.

[114] Kevin Matzen, Kavita Bala, and Noah Snavely. “Streetstyle: Exploring world-wide clothing styles from millions of photos”. In: *arXiv preprint arXiv:1706.01869* (2017).

[115] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. “Image-based recommendations on styles and substitutes”. In: *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2015, pp. 43–52.

[116] Tomáš Mikolov, Stefan Kombrink, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. “Extensions of recurrent neural network language model”. In: *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. 2011, pp. 5528–5531.

[117] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. “Human-level control through deep reinforcement learning”. In: *nature* 518.7540 (2015), pp. 529–533.

[118] Joshua L Moore, Shuo Chen, Douglas Turnbull, and Thorsten Joachims. “Taste Over Time: The Temporal Dynamics of User Preferences.” In: *ISMIR*. Citeseer. 2013, pp. 401–406.

[119] Hai Thanh Nguyen, Thomas Almenningen, Martin Havig, Herman Schistad, Anders Kofod-Petersen, Helge Langseth, and Heri Ramampiaro. “Learning to rank for personalised fashion recommender systems via implicit feedback”. In: *Mining Intelligence and Knowledge Exploration*. Springer, 2014, pp. 51–61.

[120] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. “A three-way model for collective learning on multi-relational data.” In: *Proceedings of the International Conference on Machine Learning*. Vol. 11. 2011, pp. 809–816.

[121] Xia Ning and George Karypis. “Sparse linear methods with side information for top-n recommendations”. In: *Proceedings of the sixth ACM conference on Recommender systems*. 2012, pp. 155–162.

[122] Minjae Ok, Jong-Seok Lee, and Yun Bae Kim. “Recommendation framework combining user interests with fashion trends in apparel online shopping”. In: *Applied Sciences* 9.13 (2019), p. 2634.

[123] Junwei Pan, Jian Xu, Alfonso Lobos Ruiz, Wenliang Zhao, Shengjun Pan, Yu Sun, and Quan Lu. “Field-weighted factorization machines for click-through rate prediction in display advertising”. In: *Proceedings of the International Conference on World Wide Web*. 2018, pp. 1349–1357.

[124] Gabriel Panis, Andreas Lanitis, Nicholas Tsapatsoulis, and Timothy F Cootes. “Overview of research on facial ageing using the FG-NET ageing database”. In: *Iet Biometrics* 5.2 (2016), pp. 37–46.

[125] Chanyoung Park, Donghyun Kim, Jinoh Oh, and Hwanjo Yu. “Do" Also-Viewed" Products Help User Rating Prediction?” In: *Proceedings of the 26th International Conference on World Wide Web*. 2017, pp. 1113–1122.

[126] Seung-Taek Park and Wei Chu. “Pairwise preference regression for cold-start recommendation”. In: *Proceedings of the third ACM conference on Recommender systems*. 2009, pp. 21–28.

[127] Seyoung Park, Bruce Xiaohan Nie, and Song-Chun Zhu. “Attribute and-or grammar for joint parsing of human pose, parts and attributes”. In: *IEEE transactions on pattern analysis and machine intelligence* 40.7 (2017), pp. 1555–1569.

[128] Michael J Pazzani and Daniel Billsus. “Content-based recommendation systems”. In: *The adaptive web*. Springer, 2007, pp. 325–341.

[129] Albert Pumarola, Antonio Agudo, Alberto Sanfeliu, and Francesc Moreno-Noguer. “Unsupervised person image synthesis in arbitrary poses”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 8620–8628.

[130] Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison Cottrell. “A dual-stage attention-based recurrent neural network for time series prediction”. In: *arXiv preprint arXiv:1704.02971* (2017).

[131] Massimo Quadrana, Paolo Cremonesi, and Dietmar Jannach. “Sequence-aware recommender systems”. In: *ACM Computing Surveys (CSUR)* 51.4 (2018), pp. 1–36.

[132] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. “Personalizing session-based recommendations with hierarchical recurrent neural networks”. In: *Proceedings of the Eleventh ACM Conference on Recommender Systems*. 2017, pp. 130–137.

[133] Hua Quanping. “Analysis of collaborative filtering algorithm fused with fashion attributes”. In: *International Journal of u-and e-Service, Science and Technology* 8.10 (2015), pp. 159–168.

[134] Joseph Redmon and Ali Farhadi. “Yolov3: An incremental improvement”. In: *arXiv preprint arXiv:1804.02767* (2018).

[135] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. "BPR: Bayesian personalized ranking from implicit feedback". In: *arXiv preprint arXiv:1205.2618* (2012).

[136] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. "Factorizing personalized markov chains for next-basket recommendation". In: *Proceedings of the 19th international conference on World wide web*. 2010, pp. 811–820.

[137] Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. "Fast context-aware recommendations with factorization machines". In: *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2011, pp. 635–644.

[138] Steffen Rendle, Walid Krichene, Li Zhang, and John Anderson. "Neural Collaborative Filtering vs. Matrix Factorization Revisited". In: *arXiv preprint arXiv:2005.09683* (2020).

[139] Rasmus Rothe, Radu Timofte, and Luc Van Gool. "Deep expectation of real and apparent age from a single image without facial landmarks". In: *International Journal of Computer Vision* 126.2-4 (2018), pp. 144–157.

[140] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. "Learning representations by back-propagating errors". In: *nature* 323.6088 (1986), pp. 533–536.

[141] Hasim Sak, Andrew W Senior, and Françoise Beaufays. "Long short-term memory recurrent neural network architectures for large scale acoustic modeling". In: (2014).

[142] Alvaro Sanchez-Gonzalez, Nicolas Heess, Jost Tobias Springenberg, Josh Merel, Martin Riedmiller, Raia Hadsell, and Peter Battaglia. "Graph networks as learnable physics engines for inference and control". In: *arXiv preprint arXiv:1806.01242* (2018).

[143] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. "Item-based collaborative filtering recommendation algorithms". In: *Proceedings of the 10th international conference on World Wide Web*. 2001, pp. 285–295.

[144] J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. "Collaborative filtering recommender systems". In: *The adaptive web*. Springer, 2007, pp. 291–324.

[145] Yong-Siang Shih, Kai-Yueh Chang, Hsuan-Tien Lin, and Min Sun. "Compatibility family learning for item recommendation and generation". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018.

[146] Edgar Simo-Serra, Sanja Fidler, Francesc Moreno-Noguer, and Raquel Urtasun. "Neuroaesthetics in fashion: Modeling the perception of fashionability". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 869–877.

[147] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[148] Krishna Kumar Singh and Yong Jae Lee. "End-to-end localization and ranking for relative attributes". In: *Proceedings of the European Conference on Computer Vision*. Springer. 2016, pp. 753–769.

[149] Eugen Slutzky. "The summation of random causes as the source of cyclic processes". In: *Econometrica: Journal of the Econometric Society* (1937), pp. 105–146.

[150] Hagen Soltau, Hank Liao, and Hasim Sak. "Neural speech recognizer: Acoustic-to-word LSTM model for large vocabulary speech recognition". In: *arXiv preprint arXiv:1610.09975* (2016).

[151] Sijie Song and Tao Mei. "When Multimedia Meets Fashion". In: *IEEE Transactions on Multimedia* 25.3 (2018), pp. 102–108.

[152] Xuemeng Song, Fuli Feng, Xianjing Han, Xin Yang, Wei Liu, and Liqiang Nie. "Neural compatibility modeling with attentive knowledge distillation". In: *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2018, pp. 5–14.

[153] Xuemeng Song, Fuli Feng, Jinhuan Liu, Zekun Li, Liqiang Nie, and Jun Ma. "Neurostylist: Neural compatibility modeling for clothing matching". In: *Proceedings of the International Conference on Multimedia*. ACM. 2017, pp. 753–761.

[154] David H Stern, Ralf Herbrich, and Thore Graepel. "Matchbox: large scale online bayesian recommendations". In: *Proceedings of the 18th international conference on World wide web*. 2009, pp. 111–120.

[155] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. "Sequence to sequence learning with neural networks". In: *Advances in neural information processing systems*. 2014, pp. 3104–3112.

[156] Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang, and Joel S Emer. "Efficient processing of deep neural networks: A tutorial and survey". In: *Proceedings of the IEEE* 105.12 (2017), pp. 2295–2329.

[157] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[158] Yong Kiam Tan, Xinxing Xu, and Yong Liu. "Improved recurrent neural networks for session-based recommendations". In: *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. 2016, pp. 17–22.

[159] Jiaxi Tang and Ke Wang. "Personalized top-n sequential recommendation via convolutional sequence embedding". In: *Proceedings of the International Conference on Web Search and Data Mining*. 2018, pp. 565–573.

[160] Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. "Latent relational metric learning via memory-based attention for collaborative ranking". In: *Proceedings of the International Conference on World Wide Web*. 2018, pp. 729–739.

[161] Hessel Tuinhof, Clemens Pirker, and Markus Haltmeier. "Image-Based Fashion Product Recommendation with Deep Learning". In: *International Conference on Machine Learning, Optimization, and Data Science*. 2018, pp. 472–481.

[162] Kristen Vaccaro, Sunaya Shivakumar, Ziqiao Ding, Karrie Karahalios, and Ranjitha Kumar. "The elements of fashion style". In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 2016, pp. 777–785.

[163] Mariya I Vasileva, Bryan A Plummer, Krishna Dusad, Shreya Rajpal, Ranjitha Kumar, and David Forsyth. "Learning type-aware embeddings for fashion compatibility". In: *Proceedings of the European Conference on Computer Vision*. 2018, pp. 390–405.

[164] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need". In: *Advances in neural information processing systems*. 2017, pp. 5998–6008.

[165] Andreas Veit, Balazs Kovacs, Sean Bell, Julian McAuley, Kavita Bala, and Serge Belongie. "Learning visual clothing style with heterogeneous dyadic co-occurrences". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015, pp. 4642–4650.

[166] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. "Graph attention networks". In: *arXiv preprint arXiv:1710.10903* (2017).

[167] Sirion Vittayakorn, Kota Yamaguchi, Alexander C Berg, and Tamara L Berg. "Runway to realway: Visual analysis of fashion". In: *2015 IEEE Winter Conference on Applications of Computer Vision*. IEEE. 2015, pp. 951–958.

[168] Gilbert Thomas Walker. "On periodicity in series of related terms". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 131.818 (1931), pp. 518–532.

[169] Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang. "Toward characteristic-preserving image-based virtual try-on network". In: *Proceedings of the European Conference on Computer Vision*. 2018, pp. 589–604.

[170] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. "Knowledge graph convolutional networks for recommender systems". In: *Proceedings of the International Conference on World Wide Web*. 2019, pp. 3307–3313.

[171] Pengfei Wang, Jiafeng Guo, Yanyan Lan, Jun Xu, Shengxian Wan, and Xueqi Cheng. "Learning hierarchical representation model for nextbasket recommendation". In: *Proceedings of the International ACM SIGIR conference on Research and Development in Information Retrieval*. 2015, pp. 403–412.

[172] Wei Wang, Yujing Yang, Xin Wang, Weizheng Wang, and Ji Li. "Development of convolutional neural network and its application in image classification: a survey". In: *Optical Engineering* 58.4 (2019), p. 040901.

[173] Wenguan Wang, Yuanlu Xu, Jianbing Shen, and Song-Chun Zhu. "Attentive fashion grammar network for fashion landmark detection and clothing category classification". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 4271–4280.

[174] Xi Wang, Zhenfeng Sun, Wenqiang Zhang, Yu Zhou, and Yu-Gang Jiang. "Matching user photos to online products with robust deep features". In: *Proceedings of the International Conference on Multimedia*. ACM. 2016, pp. 7–14.

[175] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. “Kgat: Knowledge graph attention network for recommendation”. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2019, pp. 950–958.

[176] Xiang Wang, Xiangnan He, Liqiang Nie, and Tat-Seng Chua. “Item silk road: Recommending items from information domains to social users”. In: *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2017, pp. 185–194.

[177] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. “Neural graph collaborative filtering.” In: *Proceedings of the International ACM SIGIR conference on Research and development in Information Retrieval*. 2019, pp. 165–174.

[178] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. “Explainable reasoning over knowledge graphs for recommendation”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 5329–5336.

[179] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. “Knowledge graph embedding by translating on hyperplanes.” In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 14. 2014. Citeseer. 2014, pp. 1112–1119.

[180] Ruofeng Wen, Kari Torkkola, Balakrishnan Narayanaswamy, and Dhruv Madeka. “A multi-horizon quantile recurrent forecaster”. In: *arXiv preprint arXiv:1711.11053* (2017).

[181] Jason Weston, Antoine Bordes, Oksana Yakhnenko, and Nicolas Usunier. “Connecting language and knowledge bases with embedding models for relation extraction”. In: *arXiv preprint arXiv:1307.7973* (2013).

[182] Peter R Winters. “Forecasting sales by exponentially weighted moving averages”. In: *Management science* 6.3 (1960), pp. 324–342.

[183] Sai Wu, Weichao Ren, Chengchao Yu, Gang Chen, Dongxiang Zhang, and Jingbo Zhu. “Personal recommendation using deep recurrent neural networks in NetEase”. In: *Proceedings of the International Conference on Data Engineering*. IEEE. 2016, pp. 1218–1229.

[184] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. “Session-based recommendation with graph neural networks”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 346–353.

[185] Jun Xiao, Hao Ye, Xiangnan He, Hanwang Zhang, Fei Wu, and Tat-Seng Chua. "Attentional factorization machines: Learning the weight of feature interactions via attention networks". In: *arXiv preprint arXiv:1708.04617* (2017).

[186] Sijie Yan, Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. "Unconstrained fashion landmark detection via hierarchical recurrent transformer networks". In: *Proceedings of the International Conference on Multimedia*. ACM. 2017, pp. 172–180.

[187] Carl Yang, Lanxiao Bai, Chao Zhang, Quan Yuan, and Jiawei Han. "Bridging collaborative filtering and semi-supervised learning: a neural approach for poi recommendation". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017, pp. 1245–1254.

[188] Wei Yang, Ping Luo, and Liang Lin. "Clothing co-parsing by joint image segmentation and labeling". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 3182–3189.

[189] Xun Yang, Xiangnan He, Xiang Wang, Yunshan Ma, Fuli Feng, Meng Wang, and Tat-Seng Chua. "Interpretable Fashion Matching with Rich Attributes". In: *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2019, pp. 775–784.

[190] Xun Yang, Yunshan Ma, Lizi Liao, Meng Wang, and Tat-Seng Chua. "Transnfcm: Translation-based neural fashion compatibility modeling". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 403–410.

[191] Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. "A dynamic recurrent model for next basket recommendation". In: *Proceedings of the International ACM SIGIR conference on Research and Development in Information Retrieval*. 2016, pp. 729–732.

[192] Ruiyun Yu, Xiaoqi Wang, and Xiaohui Xie. "Vtnfp: An image-based virtual try-on network with body and clothing feature preservation". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, pp. 10511–10520.

[193] Wenhui Yu, Huidi Zhang, Xiangnan He, Xu Chen, Li Xiong, and Zheng Qin. "Aesthetic-based clothing recommendation". In: *Proceedings of the International Conference on World Wide Web*. 2018, pp. 649–658.

[194] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. "Collaborative knowledge base embedding for recommender systems". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, pp. 353–362.

[195] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. "Joint face detection and alignment using multitask cascaded convolutional networks". In: *IEEE Signal Processing Letters* 23.10 (2016), pp. 1499–1503.

[196] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. "Deep learning based recommender system: A survey and new perspectives". In: *ACM Computing Surveys (CSUR)* 52.1 (2019), pp. 1–38.

[197] Yin Zhang and James Caverlee. "Instagrammers, fashionistas, and me: Recurrent fashion recommendation with implicit visual influence". In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 2019, pp. 1583–1592.

[198] Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. "Graph neural networks: A review of methods and applications". In: *arXiv preprint arXiv:1812.08434* (2018).

[199] Wei Zhou, PY Mok, Yanghong Zhou, Yangping Zhou, Jialie Shen, Qiang Qu, and KP Chau. "Fashion recommendations through cross-media information retrieval". In: *Journal of Visual Communication and Image Representation* 61 (2019), pp. 112–120.

[200] Xingxing Zou, Xiangheng Kong, Waikeung Wong, Congde Wang, Yuguang Liu, and Yang Cao. "Fashionai: A hierarchical dataset for fashion understanding". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2019, pp. 0–0.