Master’s Thesis of Data Science

Enhancing a Fashion-Specific Recommendation System via Integrating User Consumption Profile

사용자 소비 프로파일 반영을 통한
패션 특화 추천 시스템의 향상

BY

Chae Young Chung

August 2023

Graduate School of Data Science
Seoul National University
Data Science Major

Enhancing a Fashion-Specific Recommendation System via Integrating User Consumption Profile

Advisor Sanghack Lee

Submitting a master's thesis of
Data Science

June 2023

Graduate School of Seoul National University
Department of Data Science
Data Science Major

Chae Young Chung

Confirming the master's thesis written by

Chae Young Chung

July 2023

Chair	________________________	(Seal)
Vice Chair	________________________	(Seal)
Examiner	________________________	(Seal)

Abstract

Towards More Tailored Fashion Recommendations: An Improved Approach via Integration of User Consumption Profiles

Chae Young Chung

Data Science Department Data Science Major

The Graduate School

Seoul National University

Recommendation systems are crucial in today's digital platforms like Netflix and Amazon, enhancing personalization, customer loyalty, and revenue. Our study aims to develop a precise recommendation system rooted in a socio-scientific understanding of users, especially in fashion domain.

Further, we propose methods to create privacy-conscious user profiles in the face of strict regulations on personal data collection and usage. As part of our approach, we utilize the RFM (Recency, Frequency, Monetary) technique and develop a novel "Fashion Consciousness" index. These measures allow us to extract in-depth customer profiles, providing nuanced understanding of user preferences and behaviors, all while ensuring maximum privacy respect.

In order to capture the intricate factors like color, texture, and pattern influencing fashion product choices, we use Graph Neural Networks (GNN), particularly the Knowledge Graph Attention Network (KGAT). We enhance the performance of the system by creating the $\mathrm{KGAT_u}$ model, which broadens our graph structure to incorporate an item knowledge graph, a user-item bipartite graph, and importantly, an additional user knowledge graph. The pioneering integration of the user knowledge graph allows $\mathrm{KGAT_u}$ to effectively capture user-specific information, thereby improving the personalization and precision of our model.

We provides valuable insights by integrating the domain knowledge and abstract decision-making attributes on fashion products into the system, pushing the boundary of personalized user experience while also respecting data privacy and sensitivity.

keywords: Recommendation System, Fashion, Graph Neural Network, Domain Knowledge, User-centric, E-commerce, RFM
student number: 2021-21285

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

With the recent advances in technology, the era of big data has emerged, making it possible to collect vast amounts of data such as user search and purchase histories, ratings, and location information. As a result, the ability to analyze and utilize big data has become a key factor in enhancing a company's competitiveness. Therefore, companies are making various efforts to analyze and utilize big data to improve business operations and make more effective decisions.

Recommendation systems are a representative example of how companies are actively using big data technology. Recommendation systems are used for various purposes such as improving customer satisfaction and loyalty, product diversification, and increasing sales in giant platforms such as Netflix, YouTube, and Amazon. They have a significant impact on a company's profit. For instance, Netflix estimates that the combination of personalization and recommendation features has resulted in cost savings of more than $1 billion annually [2]. Recommendation systems are now widely applied in various fields such as e-commerce, search engine, finance, healthcare, and education, and their importance and scope of application are expected to expand in the future [3].

Recommendation systems have been developed based on item information, so far. These systems derive similarity between purchased items and recommend similar items to users, or recommend common preference items to other users who have similar tastes based on their preference information. However, research based on understanding individual consumer preferences or profiles has been relatively limited due to restricted sharing of sensitive user information and behavioral data. This is often due to regulations such as the Personal Information Protection Act, which aim to protect users' personal information.

Therefore, the purpose of this study is to implement a personalized recommendation system based on understanding of users, rather than solely focusing on item-user interactions. Our goal is to improve the performance of recommendation algorithms by integrating preferences and CRM (Customer Relationship Management) techniques, which are developed to better understand users, with deep learning.

1.1 Purpose of Study

1.1.1 GNN-Based Recommendation System Incorporating User Profiles

In this study, we propose to build a hybrid recommendation system based on user profiles and content-based approaches.

We plan to model complex relationships using GNN (Graph Neural Network) as our pipeline, which can simulate real-world scenarios where multiple factors can influence item selection [5]. GNN captures the interconnectivity between user profiles and explore the impact of various factors on item selection [6]. Additionally, we employ an attention mechanism that allows us to describe the impact of specific factors on the purchasing decision.

1.1.2 Domain-knowledge Recommendation System

Our recommendation system is designed specifically for fashion products, based on domain knowledge of the fashion industry. We focus on the unique characteristics of both fashion products and consumers. The fashion industry has a long production-to-sales time but a short product lifecycle, and demand is influenced by various external factors such as fashion trends, seasonality, and other complex factors [4]. To capture the characteristics of seasonality and fashion trends in the fashion industry, we use a variable called "trendiness." Additionally, to improve the system's performance, we quantify the consumption behavior of fashion-conscious consumers, who tend to spend more and shop frequently [7], using RFM levels and the trendiness of purchased items. We incorporate this information into the customer profile in the recommendation system to enhance its performance.

1.2 Significance of Study

With the widespread use of big data, concerns about privacy infringement and demands for corporate responsibility have increased. As a result, regulations and restrictions on the use of personal information have gradually been strengthened to prevent the violation of user rights.

In the United States and Europe, for example, laws have been enacted requiring that information that can identify individuals be de-identified. Moreover, there is a movement to expand the scope of identifiable information, as information that was previously classified as anonymized information can now be identified through advances in technology and analysis methods [8]. This indicates that there is a possibility that the data access and utilization will become narrower depending on the protection laws and the agreement of individual users, highlighting the need for the minimal use of de-identified information and analytical techniques that do not compromise user rights.

Therefore, this study aims to build an improved recommendation system that reflects users' consumption tendencies, explicitly derived based on minimal behavioral information, in order to suggest a positive direction for the development of recommendation systems that benefits both users and companies.

Additionally, an attention mechanism that assigns weight to characteristics that have a significant impact on purchasing decisions will be used to provide a numerical basis for demand prediction, as well as criteria for logistics and production planning. Ultimately, we hope to contribute to the promotion of sustainability and ESG (Environmental, Social, Governance) management by curbing overproduction, a persistent issue in the fashion industry.

1.3 Research Questions

To investigate the positive impact of consumer profiles focused on consumer behavior and the utilization of domain knowledge on recommendation performance, we specifically addressed the following four research questions:

RQ1: How does the application of consumer profiles using RFM (Recency, Frequency, Monetary) analysis compare to the baseline in terms of recommendation performance?

RQ2: Can the integration of RFM analysis and domain knowledge in creating user profiles lead to enhanced recommendation performance?

RQ3: How does the recommendation system perform when additional attributes reflecting domain knowledge are applied?

RQ4: What is the optimal time interval for deriving customer consumption profiles that result in superior recommendation performance?

Chapter 2

Related Work

2.1 Recommendation System

The recommendation system has continued to evolve since it was recognized as a research field in 1970 at Duke University. Research on recommendation systems has progressed in three main directions: content-based algorithms, collaborative algorithms, and hybrid algorithms.

A Content-based algorithm directly analyzes contents of items and calculates the similarity between items or between items and user preferences [9]. The algorithm is built on the assumption that if a user likes an item, they will also like other items that have similar features. It has the advantage of that they do not require explicit user preference data so that they can be widely applied when there is a lack of information about other users [10]. However, recommendation systems still have some limitations, such as difficulty in providing various recommendations for users with vast purchase histories or diverse tastes, as well as the challenge of addressing the cold start problem for new users [11].

Collaborative filtering approaches have a significant advantage over content-based approaches, as they predict user preferences by utilizing information on items that the user has previously rated or selected, in addition to the decisions made by similar users, rather than solely on item characteristics as content-based algorithm does [10].

This approach is particularly useful in addressing the cold start problem, where limited information is available for new users. However, collaborative filtering may suffer from the sparsity problem, where users have limited interactions with items or with each other, leading to difficulty in accurately predicting user preferences.

In order to construct enhanced recommendation systems, hybrid recommendation systems have been introduced to complement the drawbacks of each approach and enhance the recommendation function by combining various recommendation algorithm [9].

2.2 GNN-Based Recommendation System

The GNN (Graph neural network)-based recommendation system is a hybrid approach that has been developed to address the limitations of previous research. Specifically, previous research treated each interaction as an independent data instance and ignored their relations, which made it difficult to extract attribute-based collaborative signals from the collective behaviors of users.

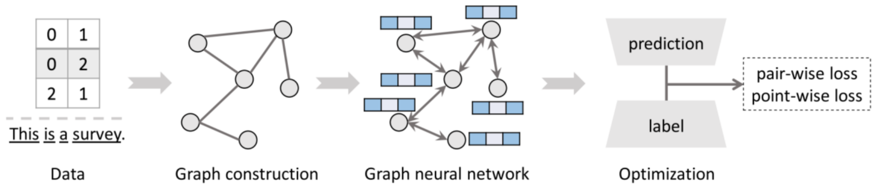

Figure 2.1: The process of implementing a GNN model: data-driven graph construction, customized GNN design, mapping representations, optimization using a loss function [12]

The GNN-based recommendation system, however, utilizes graph neural networks to capture the relationships between interactions and extract more meaningful signals from user behaviors [13]. Various GNN-based recommendation systems have been developed to model high-order connectivity information between users and items using graph neural networks, such as NGCF (Neural Graph Collaborative Filtering) [14] and

Light GCN (Graph Convolution Network) [15]. In addition, KGAT (Knowledge Graph Attention Network) applies an attention mechanism based on a graph structure that combines knowledge graphs and user-item graphs, while KGIN (Knowledge Graph-based Intent Network) focuses on relation paths [16].

CKG (Collaborative Knowledge Graph), which is the backbone of KGAT, refines the drawbacks of collaborative filtering, which cannot model side information, and supervised learning, which does not consider relationships between data.

However, CKG-based recommendation systems had drawbacks: path-based method and regularization-based method. Path-based methods, which feed-forwards paths specified based on domain knowledge for learning, suffer from a labor-intensive process of path specification and exhibit significant variations in performance depending on the specified path. A challenge with regularization-based methods is the lack of assurance regarding their ability to accurately capture high-order relations, as these relations are learned implicitly.

To address these issues, KGAT utilizes attention mechanisms and recursive embedding propagation. By combining the item-item entity graph and item-user graph, KGAT leverages attribute-based collaborative signals derived from users' collective behaviors to enhance recommendation performance [13]. However, despite these improvements, KGAT still struggles to fully utilize user-user connectivity based on user profiles.

2.3 Studies on Shared User Behaviors

In order to increase revenue and profits and improve user satisfaction, it is crucial to have a deep understanding of users.

Previous studies have designed recommendation systems utilizing customer segmentation and the RFM technique [22, 23]. The RFM technique is widely used for customer segmentation in CRM (Customer Relationship Management) for e-commerce industries [24], based on the recency, frequency, and monetary.

Fashion conscious consumers can be defined as a distinct group of consumers who are highly interested in fashion and are early adopters of new fashion trends.Several studies have found that fashion conscious consumers are willing to shop more frequently [19] and spend more money on fashion items. They also have a preference for online shopping [17]. Additionally, research by Cholachatpinyo, Padgett, & Crocker (2002) revealed that fashion-conscious consumers are more likely to experiment with new fashion items and tend to reject common norms of dress and style [18].

Chapter 3

Dataset

3.1 Data Description

We have decided to utilize a transaction dataset from a fashion retailer as it is more suitable for our objective of developing a recommendation system for fashion products, rather than relying on commonly used datasets like the Amazon dataset. Among various fashion retailer datasets available, we have chosen the "H&M Personalized Fashion Recommendations" dataset from the Kaggle competition, as it aligns closely with our goal of achieving highly personalized fashion product recommendations.

H&M dataset: The dataset for the H&M Personalized Fashion Recommendation Competition on Kaggle comprises article.csv, customer.csv, transaction.csv, and images for each article ID. The article.csv contains metadata about the products, including product number, category group, fabric, color, and detailed descriptions. The customer.csv provides metadata about the customers, including customer ID, subscription to fashion news, age, postal code, and other related information. The transaction.csv contains customer transaction data, including purchase date, customer ID, purchased product ID, price, and purchase channel (online or offline).

The dataset provides a valuable resource for conducting research on personalized fashion recommendation systems. The article metadata and images will be used to analyze the product features, such as category, color, pattern, that are most preferable to customers. The customer metadata can help to identify customer profile, as well. The transaction data is used to train machine learning models to predict future purchases and make personalized recommendations.

Period			
2018.09∼2020.09			
Text			
Type	**Columns**		**# of data**
Article	25	article_id, department_no, department_name, index_code, index_name, index_group_no, index_group_name, section_no, section_name, garment_group_n, garment_group_name, detail_desc, and etc.	105,542
Customer	7	customer_id, FN, Active, club_member_status, fashion_news_frequency, age, postal_code	1,371,980
Transaction	5	t_dat, customer_id, article_id, price, sales_channel	31,788,324
Image			
Type	**# of data**		
Article	105,440		

Table 3.1: H&M dataset description

Seasonal Trend Visuals: The image data depict the prevalent fashion trends for a given season, collected from external source.

Figure 3.1: Example of seasonal trend visuals [25, 26, 27]

Fashion Trend Terms: These are external data consisting of specific keywords and phrases that summarize the prominent fashion trends of a given season.

3.2 Collection Methods

We utilize the Pygooglenews library to conduct web scraping and retrieve URLs of web pages that included fashion trends such as “2020 Fall Trend Report” and “2020 Spring Trend” in their titles. We specifically target trend reports and fashion blogs related to fashion trends.

As fashion trend articles do not follow a standardized format or image layout, it was difficult to automatically scrape the article contents and trend images from each link. Therefore, the text and trend images are manually collected from each link.

The 300 Representative images for each season are extracted from multiple reports and fashion blogs. Based on fashion domain knowledge, we extract fashion keywords for each season, focusing on patterns, colors, design elements, categories, and moods, from the articles as in Figure 3.2. Furthermore, to calculate the similarity of items with trend articles, we integrate the item features in the H&M article metadata and created a new variable called “item description.”

Chunky Knits

Whether you're going super casual with jeans, or more dressed up in a silk midi skirt, this fall is all about the chunky knit. The bigger the better, whether fisherman knits, colorblocks, or solids.

Figure 3.2: Example of extraction fashion keywords from forecasting articles

Item description	**'product_type_name' + 'colour_group_name' + 'perceived_colour_value_name'+ 'perceived_colour_master_name' + 'graphical_appearance_name' + 'detail_desc'**

Figure 3.3: Item description generation by integrating item entities

3.3 Data Processing and Integration

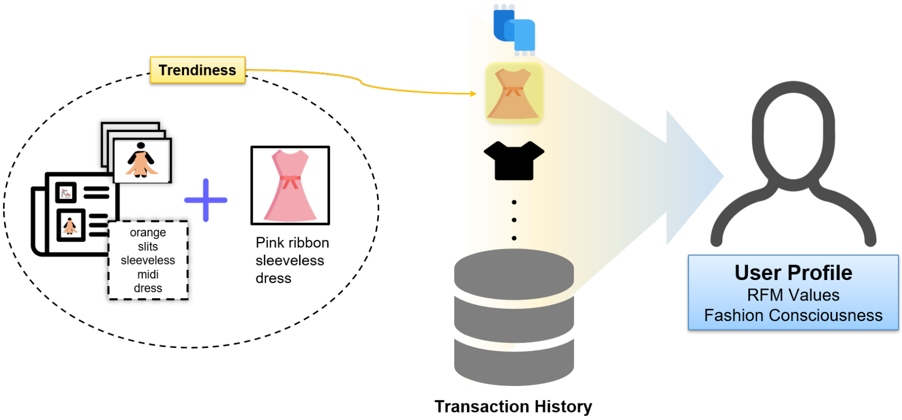

Figure 3.4: Data integration

Data integration is conducted through a process as Figure3.4. Based on six existing item entities, the integration leverages text and image data of the items to derive their trendiness. By combining the derived item trendiness with user transaction history, the fashion consciousness of each user is calculated. This information, along with the user's RFM level, is utilized as the user's consumption profile.

3.3.1 Feature selection

A total of 25 attributes are identified for the items. Among these, certain attributes such as the color group code and color group name has the same information in the different format, while others, such as the index group name and index name, exhibit a hierarchical relationship. To ensure the avoidance of redundant item information, careful selection is undertaken. Out of the original 25 attributes, six specific attributes are chosen as item entities for the recommendation system, namely, article ID, index group name, department name, product type name, graphical appearance name, color group name, perceived color value name, and trendiness.

3.3.2 RFM Levels

In RFM analysis, a fixed period is typically used, and in e-commerce, this period is usually six months, but it may vary depending on the situation. For this study, a one-month interval is selected as the reference period.

- Recency: the difference between the date of the recommendation and the date of the most recent purchase, representing how recently a customer made a purchase.
- Frequency: the number of purchases during the evaluation period, counting transactions that occurred on the same date as one purchase.
- Monetary: the total amount of money spent on all past transactions during the period

The RFM model is represented as RFM = aR + bF + cM [28], and adjustments to the ratings and weights can vary depending on the research subject. However, in this study, only adjustments to the rating system were adopted to derive the RFM attributes.

Traditional RFM typically uses a five-level rating system, but based on prior research indicating that a six-level rating system better reflects customer characteristics than a five-level or ten-level system, a six-level system was used in this study. The six-level system was proposed using the Rogers Innovation Adoption Curve, and each level is as follows [29]:

μ (mu): represents the mean (average) of the user distribution
σ (sigma): represents the standard deviation of the user distribution

- 6 (Top tier group): upper 2.5%

$$\mu + 2\sigma \sim$$

- 5 (Sub-upper tier group): upper 13.5%

$$\mu + \sigma \sim \mu + 2\sigma$$

- 4 (Upper tier group): upper 34%.

$$\mu \sim \mu + \sigma$$

- 3 (Lower tier group): lower 34%

$$\mu - \sigma \sim \mu$$

- 2 (Sub-lower tier group):lower 13.5%

$$\mu - 2\sigma \sim \mu - \sigma$$

- 1 (Sub-lower tier group): lower 2.5%

$$\sim \mu - 2\sigma$$

3.3.3 Trendiness and Fashion Consciousness

In the fashion industry, the typical year spans from February to January, traditionally marked by two major seasons: Spring/Summer (SS) and Fall/Winter (FW). These seasons usually follow a 4-5-4 week cycle within the four distinct periods of approximately 13 weeks each. However, H&M has adopted a different approach by releasing new products on a weekly basis, creating 52 seasons per year. Therefore, dividing the seasons into two broad categories is not suitable for calculating the trendiness of each item, given the extensive product launch schedule of H&M.

Additionally, trend analysis are often published based on four seasons: SPRING (Feb-Apr), SUMMER (May-Jul), FALL (Aug-Oct), and WINTER (Nov-Jan). To calculate the trendiness of items at the time of purchase, we divide the retail season into these four seasons. It should be noted that retail seasons may differ from the seasons that customers perceive.

Trendiness of items at the moment of purchase: In this study, we employ a pre-trained Xception model [30] to calculate the cosine similarity between 300 images per season and item images.Then, we calculated the trendiness of each item by measuring the jaccard similarity as below, between the fashion keywords, which are extracted from the trend articles and blog posts for each season, and the item description variable created by combining the H&M article metadata. By employing jaccard similarity, which focuses on shared elements, our aim is to determine the level of similarity between item descriptions and the fashionable keywords [31]. Consequently, items with a higher number of common elements between fashion keywords and item descriptions are considered more trendy, reflecting their alignment with the prevailing fashion trends discussed in the articles and blog posts.

$$f : \text{the fashion keywords} \quad i : \text{the item description}$$

$$\mathrm{sim}(f, i)_{\text{Jaccard}} = \frac{|I_f \cap I_i|}{|I_f \cup I_i|}$$

To derive the trendiness value as the average of trendiness derived from text and trendiness derived from seasonal images, it was necessary to mitigate the distortion caused by the difference in scale between the two values. Therefore, in order to align the scales, the trendiness values derived from text were transformed using logarithmic transformation to range between 0.1 and 1. For items without product images, the lowest value from the overall similarity was used as a replacement. In cases where the trendiness derived from text was 0, it was replaced with the minimum similarity value derived from text.

The average value of the two trendiness is used as the trendiness value for each item in the season. The products with high trendiness, in Figure 3.5 exhibit the reflection of fall fashion trends trends such as “sequin” and “faux leather.” Additionally, these products effectively capture the seasonality-“Fall” with its heavy fabric and design details such as long sleeves. Accessories generally have lower trendiness, and this can be attributed to the higher proportion of garment images in trend representative images when performing cosine similarity calculations to determine trendiness.

Figure 3.5: Example of fall trendy item base on “trendiness”

Fashion consciousness: Trend consciousness is evaluated by considering the item’s trendiness at the time of purchase, using the same data intervals as RFM. To align the trendiness with the scale of the RFM level and maintain the differences in trendiness between each item, the trendiness values are rescaled by multiplying them by 10 and then rounded to the second decimal place. This adjustment ensures consistency in the trendiness scale while preserving the relative variations among different items. It is represented by the mean value of the trendiness of the purchased items during the given period.

Customer ID	R	F	M	Fashion Consciousness
0000423b00	2	1	4	4.21
000058a12d	1	4	6	3.54
00006413d8	3	1	6	2.54
00007d2de8	2	4	3	2.49
00009c2aea	4	1	3	2.05

Article ID	Group	Department	Product Type	Graphical Appearance	Color	Perceived Color	Trendiness
458543003	Ladieswear	Jersey Basic	Skirt	Stripe	Blue	Dark	0.2407
458543001	Divided	UW	Cardigan	Solid	Beige	Light	0.2505
458428053	Divided	Basic 1	Dress	Metallic	Black	Metallic	0.4051
458428047	Children Size	Young Girl Jersey Basic	Dress	Dot	Yellow	Dot	0.3515
458428046	Ladieswear	Woven bottoms	Skirt	Denim	Blue	Denim	0.3514

Figure 3.6: Examples of user profiles and item entities

The dataset has been carefully prepared, encompassing a diverse range of variables essential for constructing fashion-oriented recommendations. By integrating data and deriving new indices such as ”trendiness,” ”fashion consciousness,” and ”RFM,” customer profiles were created. This thorough approach ensures the reliability and comprehensiveness of the dataset, enabling the development of the fashion recommendation system.

Chapter 4

Methodology

Our study aims to improve the performance of recommendation systems by incorporating user profile information and relationships into KGAT construction. To achieve this goal, we propose a novel graph structure based on the fashion domain knowledge. Our proposed approach represents a promising direction for improving the accuracy and efficiency of recommendation systems, and can be applied to various domains beyond the scope of this research.

4.1 Graph Construction

In the previous work, KGAT utilized a collaborative knowledge graph that combined a single user-item bipartite graph with a single knowledge graph.

In this study, we introduce a novel approach called $\mathrm{KGAT_u}$ (KGAT with user knowledge) that distinguishes itself from previous KGAT, by incorporating two separate knowledge graphs: one for users and one for items, along with a user-item bipartite graph. This design allows us to leverage transaction history, user knowledge, and item knowledge information, enhancing the recommendation process.

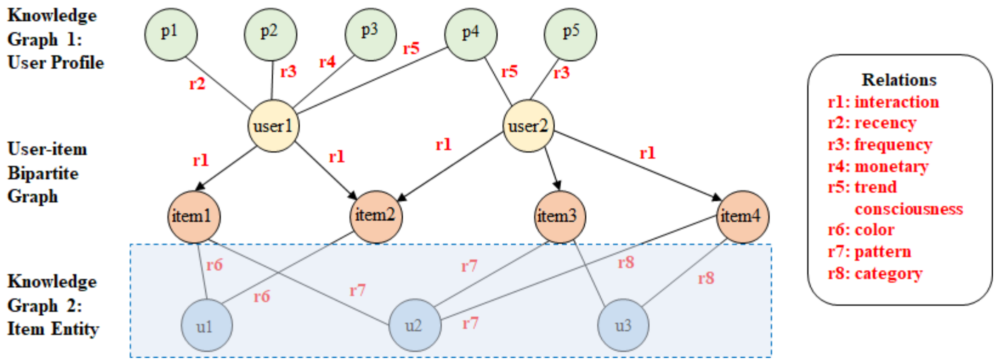

Figure 4.1: KGAT_u graph construction, expanding KGAT's graph structure to include a user knowledge graph

4.2 KGAT-Based Recommendation Model

4.2.1 Embedding

Knowledge graph embedding is a popular technique for transforming entities and relations in a knowledge graph into vector representations while preserving the graph structure.

TransR is a popular knowledge graph embedding model that is commonly used in collaborative knowledge graph (CKG) applications. It separates the entity and relation spaces to better capture the complex relationships between entities and relations. A plausibility score, as in the formulation below, is computed based on the distance between the projected representations of the head and tail entities in the relation space. A lower score of $g(h, r, t)$ indicates a higher likelihood that the triplet is true. Conversely, a higher score suggests a lower likelihood of truth. Therefore, the score provides a measure of how well the triple can be explained by the embedding model [13].

W_r : transformation matrix of relation, projecting head and tail entities into relation space r's space

e_h : embedding of the head h, e_t :embedding of the tail t,

e_r : embedding of the relation r,

$$g(h,r,t) = \|\mathbf{W}_r\mathbf{e}_h + \mathbf{e}_r - \mathbf{W}_r\mathbf{e}_t\|_2^2$$

In contrast to prior studies that focused on embedding only item-item entities and user-item relations using TransR, our study aims to embed all relations between user-user profiles, user-item, and item-item entities.By doing so, we seek to capture the complex and heterogeneous relationships that exist in collaborative knowledge graphs, and improve the performance of recommendation [13].

4.2.2 Attentive Embedding Propagation

The concept of an ego-network refers to the neighborhood information that is combined onto a given node. The aggregation process for ego-network condenses information related to the relationships and entities connected to the given node, and as the topological order of layer increases, the range of connected information expands.

The ego-network information is represented as a combination of the tail's embedding value and $\pi(h,r,t)$. $\pi(h,r,t)$ is learned through an attention mechanism, and is designed to promote greater propagation as the distance between the embedded head and tail becomes closer.

The resulting ego-network is then aggregated with the information of the reference node using the bi-interaction method, which has shown to be the most effective in previous research.

h : entity h (a given head), r : relation entity of h, t : tail entity of h,

N_h : neighborhoods of entity, eN_h :ego-network of h,

e_t : embedding of the tail t,

$\pi(h,r,t)$:the decay factor on each propagation on edge (h,r,t)

$$\mathbf{e}_{\mathcal{N}_h} = \sum_{(h,r,t)\in\mathcal{N}_h} \pi(h,r,t)\mathbf{e}_t$$

$$f_{\text{Bi-Interaction}} = \text{LeakyReLU}\left(\mathbf{W}_1\left(\mathbf{e}_h + \mathbf{e}_{\mathcal{N}_h}\right)\right) + \text{LeakyReLU}\left(\mathbf{W}_2\left(\mathbf{e}_h \odot \mathbf{e}_{\mathcal{N}_h}\right)\right)$$

To capture high-order connectivity information, additional propagation layers are added, allowing us to gather information propagated from more distant neighbors. This recursive process enables us to update the representation of an entity at each step (l-th step).

$$\mathbf{e}_h^{(l)} = f\left(e_h^{(l-1)}, e_{\mathcal{N}_h}^{(l-1)}\right)$$

We can define the information propagated within the l-ego network for the entity h as follows:

$$\mathbf{e}_{\mathcal{N}_h}^{(l-1)} = \sum_{(h,r,t)\in\mathcal{N}_h} \pi(h,r,t)\mathbf{e}_t^{(l-1)}$$

4.3 Model Prediction and Evaluation Metrics

The information obtained at each layer is concatenated to calculate the final representations of the user and item.

$$e_u^* = e_u^{(0)} \| \cdots \| e_u^{(L)}, \quad e_i^* = e_i^{(0)} \| \cdots \| e_i^{(L)}$$

The matching score is generated by taking the inner product of the final representations of the user and item.

$$\hat{y}(u,i) = e_u^* \top e_i^*$$

Two widely used metrics in recommender systems, recall@k and NDCG@k, are selected for evaluation.When making recommendations to users, it is typical to recommend the top N items that are most relevant. The parameter k represents the number of items to be recommended, and in our experiments, we set k to 50, considering the total number of items available, which is 105,440.

Recall measures the proportion of top k recommended items that match the user's preferences among all items the user is interested in. In our study, recall is calculated as the ratio of recommended items that were actually purchased by the user to the total number of purchased items. The recall value ranges from 0 to 1, with higher values indicating better recommendation performance.

NDCG evaluates the quality of search or recommendation results by considering both the rank of the result and the relevance score associated with that rank. It is computed by dividing the DCG (Discounted Cumulative Gain) by the IDCG (Ideal DCG) to normalize the result. Similar to recall, NDCG also takes values between 0 and 1, with values closer to 1 indicating better performance.

DCG is the sum of relevance scores of recommended items, considering their order, with higher weights assigned to items at higher ranks.

IDCG represents the maximum achievable DCG value in an ideal recommendation scenario where the precise ranking is known. It serves as a reference value for evaluating the relative performance of the system.

$$DCG_k = \sum_{i=1}^{k} \frac{2^{rel_i} - 1}{\log_2(i+1)}$$

$$IDCG_k = \sum_{i=1}^{k} \frac{rel_i^{opt}}{\log_2(i+1)}$$

$$NDCG_k = \frac{DCG_k}{IDCG_k}$$

Chapter 5

Experiment & Results

The dataset used in this study consists of transaction data from users who had records of purchasing five or more products.

The experiments are conducted for two seasons, and the results are averaged. Fashion trends typically divide into two major periods: S/S (Spring/Summer) and F/W (Fall/Winter). Based on these divisions, we select the periods of Spring and Fall when the season transitions occur. We designate the intervals for our test set to be the midpoint of Spring and Fall, which exhibit similar statistics in transaction records. Specifically, these periods are from March 18th to March 24th and from September 16th to September 22nd, considering that the retail season begins in February for Spring and August for Fall. The train/validation dataset intervals are defined as three months, one month, and two weeks prior to the target period. Within each interval, a random selection of 90% of the data is used as the train set, while the remaining data serve as the validation set.

Model performance is evaluated by training the models on the train set and selecting the best model based on its performance on the validation set after 150 epochs of training. The predictions made by the best model are then compared to the actual purchased items in the test set to measure performance.

Since we had to work with datasets containing multiple combinations and train both the KGAT and KGAT_u models alternately, it was challenging to adjust hyperparameters for each combination. Therefore, we only made few adjustments to the learning rate and early stopping step based on the baseline, while keeping the rest of the hyperparameters consistent with previous studies.

Hyperparameters set as follows: learning rate of 0.005, batch size of 1024 for collaborative filtering and batch size of 2048 for knowledge graph. The embedding size for all models is 64. In this study, we selected the 3-hop high connectivity approach, where each layer had dimensions of [64, 32, 16], considering both the results of previous research [29] and the time cost for the model training. The dropout rates are [0.1, 0.1, 0.1] for each layer. The early stopping step is set to 10 and Adam optimizer is implemented. All models are trained on a single NVIDIA RTX A6000 GPU.

5.1 Performance Comparisons

The baseline model is based on KGAT and utilizes six item-entity attributes for training: index group name, department name, product type name, graphical appearance name, color group name, and perceived color value name.

We compared BPR MF with the KGAT model, aiming to assess the advantages of the GNN model structure in the context of fashion recommendation. BPR MF (Bayesian Personalized Ranking Matrix Factorization) is a powerful algorithm widely used in personalized recommendation systems based on implicit data [32]. It is a type of collaborative filtering that utilizes user-item interactions to make personalized recommendations. The findings, as depicted in Table 5.1, indicated that the KGAT model is comparable with BPR MF model in terms of recommendation performance.

When the $\mathrm{KGAT_u}$ model is used, it is indicated by an asterisk (*).

		Recall@50	NDCG@50
1month	Baseline	**0.0469**	**0.0194**
	BPR MF	0.0458	0.0202

Table 5.1: Baseline and BPR MF performance comparison

RQ1: How does the application of consumer profiles using RFM (Recency, Frequency, Monetary) analysis compare to the baseline in terms of recommendation performance?

We compared the performance improvement achieved by applying the generally-used RFM (Recency, Frequency, Monetary) approach, which enables the construction of user consumption profiles. The evaluation scores indicating performance improvement are highlighted in bold. The results showed a 3.0% improvement in performance, based on Recall@50 as in table 5.2 , compared to the baseline.

The results confirm the fundamental concept that incorporating user understanding into the recommendation system can offer valuable insights for enhancing recommendation performance. Additionally, the universal applicability of RFM, which can be used in any e-commerce domain, suggests the potential for the versatile use of $\mathrm{KGAT_u}$.

		Recall@50	%Improv	NDCG@50	%Improv
1month	Baseline	0.0469	-	0.0194	-
	Baseline + RFM(*)	0.0483	**3.0%**	0.0197	**1.7%**

Table 5.2: Baseline and Baseline + RFM performance comparison

Despite showing some improvement compared to the baseline, the enhancement in performance was relatively modest. This can be attributed to the utilization of a 6-level RFM system, which may not fully capture the actual variations in RFM values. For example, when examining the graphs Figure 5.1, Figure 5.2, and Figure 5.3, which represent RFM analysis for the fall season, significant differences can be observed in the actual values within the same RFM level. This suggests that the implemented RFM categorization may not fully reflect the true distribution of RFM. In the case of Frequency, the actual values ranged from a minimum of 4 to a maximum of 26 within level 6. Although the 6-level categorization approach used in previous RFM studies has shown good performance [29], it is expected that better performance improvements can be achieved by refining the RFM categorization. Future research can focus on improving the RFM categorization to further enhance performance.

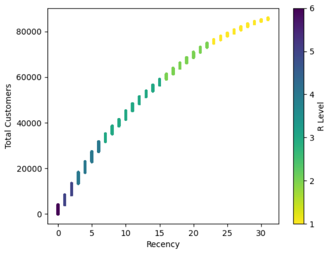

Figure 5.1: Recency

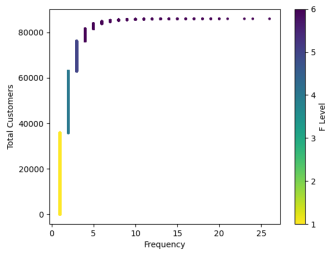

Figure 5.2: Frequency

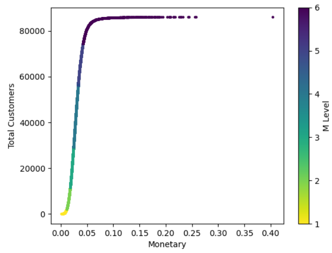

Figure 5.3: Monetary

RQ2: Can the integration of RFM analysis and domain knowledge in completing user profiles lead to enhanced recommendation performance?

The performance comparison revealed that incorporating Fashion Consciousness, in addition to the universal user behavior analysis of RFM, resulted in superior performance compared to both the baseline model and the RFM-only model. The improvements achieved were 3.9% in terms of Recall@50 and 3.1% in terms of NDCG@50 when compared to the baseline model.

		Recall@50	%Improv	NDCG@50	%Improv
1month	Baseline	0.0469	-	0.0194	-
	Baseline + RFM(*)	0.0483	**3.0%**	0.0197	**1.7%**
	Baseline + RFM + FC(*)	0.0487	**3.9%**	0.0200	**3.1%**
	Baseline + TR	0.0469	0%	0.0189	-2.6%
	Baseline + TR + RFM + FC(*)	0.0459	-2.2%	0.0188	-3.1%

Table 5.3: Performance comparison

In order to further examine the impact of incorporating Fashion Consciousness, a comparison was made between the recommendations for fall season generated by the RFM-only model as in figure 5.4 and the model that additionally considered Fashion Consciousness for User 3031, who could be classified as a fairly fashion-conscious customer. The user had an actual RFM level of 5 for Recency, 6 for Frequency, 5 for Monetary value, and a Fashion Consciousness score of 3.24 (in the top 72.23% of users)

The results showed that the recommendations with the incorporation of Fashion Consciousness displayed slightly more trendy items as in figure 5.5, highlighted with red boxes. Examples of trendy items are 0864929003, which featured the popular 2020 Fall trend of animal patterns, and 089478001, faux leather leggings, which aligned with the material trend.

Figure 5.4: u3031 Baseline+RFM

Figure 5.5: u3031 Baseline+RFM+FC

For User 263708, who exhibits relatively lower fashion consciousness with a Recency score of 1, Frequency score of 1, Monetary value score of 2, and Fashion Consciousness score of 2.91 (ranking in the bottom 34.41% of users), the recommendations generated solely based on the RFM model included some fashion items such as a leopard blouse and faux leather leggings marked with the red boxes. However, upon incorporating Fashion Consciousness (FC), the recommendations predominantly comprised of safe and basic items. Notably, there was an increased inclusion of accessories in the recommendation list compared to the baseline results. This can be attributed to the generally lower trendiness associated with accessories.

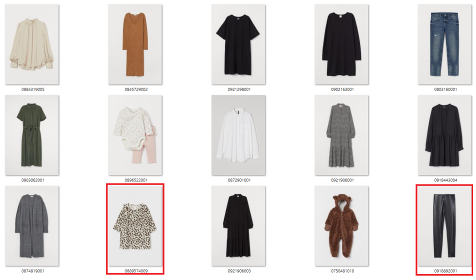

Figure 5.6: u263708 Baseline+RFM

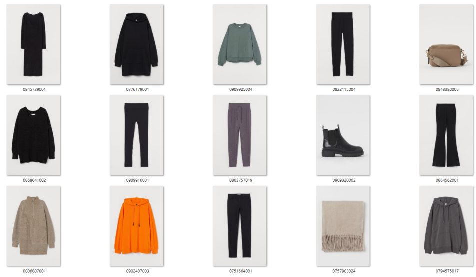

Figure 5.7: u263708 Baseline+RFM+FC

With regard to this, conducting additional precise calculations on trendiness can be expected to provide more accurate recommendation lists tailored to individual users. By refining the assessment of trendiness, the recommendation system can offer more personalized suggestions that align with the specific fashion preferences and consciousness of each user. This can enhance the overall user experience and increase the relevance and satisfaction of the recommendations provided.

RQ3: How does the recommendation system perform when additional attributes reflecting domain knowledge are applied?

The incorporation of trendiness attribute, derived from both image and text data, into the recommendation system resulted in a 2.6% decrease in performance in terms of NDCG@50 compared to the baseline as in table 5.3. Even when trendiness attribute was applied to the previously top-performing model, which includes RFM (Recency, Frequency, Monetary) and FC (Fashion Consciousness) attributes, a decrease in performance was observed. This phenomenon can be attributed to the presence of duplicated information across different attributes.

When calculating trendiness based on text data, it appears that the item entity information, such as index group name, department name, product type name, graphical appearance name, color group name, and perceived color value name, is included in the text data, resulting in duplicated item information.

Therefore, it can be inferred that the duplicated information across attributes hindered the deep learning model's ability to fully learn the diversity of the data, leading to a degradation in performance.

RQ4: What is the optimal time interval for deriving customer consumption profiles that result in superior recommendation performance?

		Recall@50	%Improv	NDCG@50	%Improv
1month	Baseline	0.0469	-	0.0194	-
	Baseline + RFM(*)	0.0481	**3.0%**	0.0197	**1.7%**
	Baseline + RFM + FC(*)	0.0487	**3.9%**	0.0200	**3.1%**
2weeks	Baseline	0.0477	-	0.0197	-
	Baseline + RFM(*)	0.0463	-3.1%	0.0191	-3.0%
	Baseline + RFM + FC(*)	0.0456	-4.4%	0.0185	-6.3%

Table 5.4: Performance comparison based on time interval

While comparing the different time intervals of 1 month, 3 months, and 2 weeks, we faced challenges conducting the experiment for the 3-month interval due to memory limitations. The baseline model for the 3-month interval had to be stopped at 20 epochs using an early stopping rule, which made it less suitable for direct comparison with the 1-month interval. Therefore, the focus was shifted towards comparing the performance between the 1-month and 2-week intervals.

Based on the observations from the 1-month experiments, we expected that applying consumer profiling attributes to the 2-week baseline model would enhance its performance. Moreover, given that the baseline performance for the 2-week interval was superior to that of the 1 month, we hypothesized that the model incorporating Baseline + RFM + FC from the 2-week dataset would serve as the optimal choice.

However, contrary to our expectations, the empirical results illustrated a different trend. Both the Baseline+RFM and Baseline+RFM+FC models showed a performance decline of 3.1% and 4.4%, respectively, in terms of recall@50 compared to the 2-week baseline. This outcome indicated that the application of consumer profiling within the 2-week dataset did not result in performance improvement, as initially hypothesized. Additionally, among the six models featured in Table 5.4, the one-month Baseline+RFM+FC model emerged with the best performance. This suggests that while consumer profiling can improve a recommendation system's performance, a shorter data collection period might not sufficiently capture accurate customer consumption profile information.

In conclusion, our findings emphasize the necessity for careful consideration of the optimal timeframe for data collection when incorporating consumer profiling in model construction. It appears that a longer data collection period, such as a month, could provide a more accurate reflection of consumer behavior, and consequently, lead to superior model performance.

Chapter 6

Conclusion

In this study, we've introduced methods to boost recommendation system performance. Our fundamental idea is that a deep understanding of customers, built upon socio-scientific perspectives, leads to improved recommendation performance. We further suggested methods to create user profiles that respect privacy and data sensitivity.

To deal with the subjective factors that influence fashion choices, such as color, texture, and pattern, we utilized a Graph Neural Network (GNN), specifically the Knowledge Graph Attention Network (KGAT). By expanding the fundamental structure of the KGAT model to include a user knowledge graph, we constructed the $\mathrm{KGAT_u}$ model. This enhancement allowed us to capture a deeper understanding of users and their preferences within the recommendation system.

The RFM analysis technique was adapted to develop consumer profiles without relying on sensitive information. Through rigorous experimentation and evaluation, we demonstrated the versatility of the $\mathrm{KGAT_u}$ model, which facilitated personalized recommendations across diverse product categories. These findings provided compelling evidence to support the idea that gaining a deep understanding of users leads to new insights and enhances recommendation performance. We also incorporated fashion domain knowledge to design a fashion-centric recommendation system. By extracting a "trendiness" attribute from image and text data and integrating it with user transaction history, we determined each user's "fashion consciousness," resulting in significant performance improvements.

While our approach has resulted in performance improvements, there is still room for enhancement, particularly in the computation of ”trendiness.” The manual extraction of trend keywords for fashion trend analysis could be optimized through advanced deep learning models, such as fashion NER. Additionally, a lower trendiness observed for accessories may be due to their imbalanced representation in trend representative images. By recognizing and addressing these limitations, we could further enhance the performance of our recommendation system.

In conclusion, our research highlights the potential for improving recommendation system performance through a user-centric understanding. Nevertheless, there are still substantial possibilities for enhancing fashion recommendation systems, along with some challenges.

A limitation of our study lies in the fact that we conducted our experiments using the dataset only from the fashion brand H&M, which aligns with high fashion trends in a wearable manner. This potentially results in less differentiation in trendiness among individual items. Additionally, the nature of H&M’s customer base may already be somewhat segmented, meaning the impact of our user profile information on the recommendation system could potentially be minimal. It would be beneficial for future research to include datasets from various fashion brands or retailers with different market positions to validate the proposed methods more comprehensively.

Another limitation is grounded in the unique consumption patterns of fashion products. These often lead to distinct customer behaviors, such as avoiding repurchases of similar trendy items or refraining from buying products within the same category for a certain duration, irrespective of their trendiness. This poses challenges to the typical logic of recommendation systems, which suggest similar items based on customers’ past purchases and consumption patterns.Therefore, future research could focus on developing recommendation systems that effectively capture these distinctive fashion product consumption patterns. This will lead to the construction of recommendation systems that better accommodate the complexities associated with fashion product consumption.

Bibliography

[1] Das, D., Sahoo, L., and Datta, S. (2017),"A survey on recommendation system," *International Journal of Computer Applications*, 160(7)

[2] Gomez-Uribe, C. A., and Hunt, N. (2015), "The netflix recommender system: Algorithms, business value, and innovation," *ACM Transactions on Management Information Systems (TMIS)*, 6(4), 1-19.

[3] Lu, J., Wu, D., Mao, M., Wang, W., and Zhang, G. (2015), "Recommender system application developments: a survey," *Decision support systems*, 74: 12-32.

[4] Thomassey, S. (2013), "Sales forecasting in apparel and fashion industry: A review," *Intelligent fashion forecasting systems: Models and applications*, 9-27.

[5] Hamilton, William L., Rex Ying, and Jure Leskovec, "Representation learning on graphs: Methods and applications," *arXiv preprint arXiv:1709.05584*, 2017.

[6] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Philip, S. Y., "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, 32(1), 4-24, 2020.

[7] Iyer, R., and Eastman, J. K., "The fashion conscious mall shopper: An exploratory study," *Marketing Management Journal*, 20(2), 42-53, 2010.

[8] Kim, D. G., and Lee, H. "빅데이터 기반의 개인정보 비식별화 동향," *Review of Korean Society for Internet Information*, 16(2), 15-22, 2015.

[9] DAS, Debashis, SAHOO, Laxman DATTA, and Sujoy, "A survey on recommendation system," *International Journal of Computer Applications*, 160.7, 2017.

[10] 손지은, 김성범, 김현중, and 조성준, "추천 시스템 기법 연구동향 분석," *대한산업공학회지*, 41(2), 185-208, 2015.

[11] Sharma, R., and Singh, R, "Evolution of recommender systems from ancient times to modern era: a survey," *Indian Journal of Science and Technology*, 9(20), 1-12, 2016.

[12] Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., and Li, Y., "A survey of graph neural networks for recommender systems: Challenges, methods, and directions," *ACM Transactions on Recommender Systems*, 1(1), 1-51, 2023.

[13] Wang, X., He, X., Cao, Y., Liu, M., and Chua, T. S.,"Kgat: Knowledge graph attention network for recommendation,"*In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 950-958, 2019.

[14] Wang, X., He, X., Wang, M., Feng, F., and Chua, T. S.,"Neural graph collaborative filtering,"*n Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 165-174, 2019.

[15] He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., and Wang, M.,"Lightgcn: Simplifying and powering graph convolution network for recommendation,"*In Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pp. 639-648, 2020.

[16] Wang, X., Huang, T., Wang, D., Yuan, Y., Liu, Z., He, X., Chua, T. S., "Learning intents behind interactions with knowledge graph for recommendation,"*In Proceedings of the Web Conference 2021*, pp. 878-887, 2021.

[17] Wan, F., Youn, S., and Fang, T., "Passionate surfers in image-driven consumer culture: Fashion-conscious, appearance-savvy people and their way of life," *ACR North American Advances*, pp. 266-274, 2001.

[18] Cholachatpinyo, A., et al.,“A conceptual model of the fashion process–part 2: An empirical investigation of the micro-subjective level.” *Journal of Fashion Marketing and Management: An International Journal* 6.1, 24-34, 2002.

[19] Iyer, R., and Eastman, J. K., “The fashion conscious mall shopper: An exploratory study,” *Marketing Management Journal*, 20(2), 42-53, 2010.

[20] Rodrigues, F., and Ferreira, B., “Product recommendation based on shared customer’s behaviour,” *Procedia Computer Science*, 100, 136-146, 2016.

[21] Jiang, T., and Tuzhilin, A. , “Improving personalization solutions through optimal segmentation of customer bases,” *IEEE transactions on knowledge and data engineering*, 21(3), 305-320, 2008.

[22] 이재성, 김재영, and 강병욱. , “RFM 다차원 분석 기법을 활용한 암시적 사용자 피드백 기반 협업 필터링 개선 연구,” *지능정보연구*, 25(1), 139-161, 2019.

[23] Cho, Y. S., Gu, M. S., and Ryu, K. H. , “Development of Personalized Recommendation System using RFM method and k-means Clustering,” *Journal of The Korea Society of Computer and Information*,17(6), 163-172, 2012.

[24] 민지홍, “e-비즈니스 업계에서 고객의 RFM 구매 이력이 차기 구매 시점에 미치는 영향,” *e-비즈니스연구*,16(4), 191-204, 2015.

[25] Pieri, K., Goldberg, C., Samaha, B., and Clinton, L. M., “ The Best Looks from Paris Fashion Week Fall 2020,” *harpersbazzar*, http://www.harpersbazaar.com/fashion/fashion-week/g31102501/paris-fashion-week-fall-2020/, 2020.

[26] Griswold, M., “The Biggest Fashion Trends You’ll See Everywhere This Fall, According To Fashion Week,” *stylecaster* http://stylecaster.com/fashion/fashion-trends/1085438/fall-fashion-trends-2020/, 2020.

[27] Joseph, S., “I Can’t Stop Obsessing Over These 6 Fall Trends,” *popsugar*, http://www.popsugar.com/fashion/photo-gallery/47780471/image/47781381/Divine-Heritage-Cuffed-Pleated-Shorts, 2020.

[28] 진병운, 조영성, and 류근호, “RFM 기법과 연관성 규칙을 이용한 개인화된 전자상거래 추천시스템,” *한국컴퓨터정보학회논문지*,15(12), 227-235, 2010.

[29] 유귀열, and 문영수, “RFM 에서 등급부여 방법에 관한 연구,” *한국데이터정보과학회지*,24(2), 245-255, 2013.

[30] Chollet, F.,“Xception: Deep learning with depthwise separable convolutions,” *In Proceedings of the IEEE conference on computer vision and pattern recognition*,1251-1258, 2017.

[31] Bag, S., Kumar, S. K., and Tiwari, M. K., “An efficient recommendation generation using relevant Jaccard similarity,” *Information Sciences*, 483, 53-64, 2019

[32] Rendle, Steffen, et al., “BPR: Bayesian personalized ranking from implicit feedback,” *arXiv preprint arXiv:1205.2618*, 2012

초 록

오늘날 추천 시스템은 초개인화 서비스, 고객 충성도 및 매출 향상에 기여하며, Netflix와 Amazon 등과 같은 디지털 플랫폼들에서 중추적인 역할을 하고 있다. 본 연구는 사용자에 대한 사회과학적 이해를 기반으로 더욱 정확한 추천 시스템을 개발을 목표로 하였으며, 특히 패션 이커머스 분야에 특화된 시스템 개발에 중점을 두었다.

개인 데이터의 수집 및 사용에 관한 규제가 강화되는 상황에서, 사용자의 민감 정보나 적극적인 상호작용 없이도 사용자 소비 프로파일을 생성하는 방법을 제안하였다. RFM (Recency, Frequency, Monetary) 기법을 활용하고 패션 상품에 특화된 소비 지수-”패션 민감도” 도출을 위한 방법 고안하여, 사용자의 소비 행동 특성을 추출하였다.

패션 제품 선택에 영향을 미치는 요소들, 즉 색상, 질감, 패턴 등 추상적 요인 간의 복잡한 관계를 그래프 구조로 포착하였고, 그래프 신경망(GNN) 중 KGAT(Knowledge Graph Attention Network)를 기반으로 도메인 지식을 적용하였다. 기존 KGAT 그래프 구조인 Item knowledge graph와 user-item bipartite graph에 user knowledge graph 구조를 추가하여 KGAT_u 모델을 구축하였다. user knowledge graph의 도입으로, 추천 시스템에 사용자의 소비 특성을 반영할 수 있도록하여 추천 성능의 정확도를 높였다.

본 연구를 통해 사용자의 민감 정보를 침해하지 않으면서도 패션 도메인 지식과 패션 상품의 추상적인 의사결정 요인을 추천시스템에 통합할 수 있는 새로운 방법을 제안함으로써 추천 시스템의 발전에 기여하고자 한다.

주요어: 추천시스템, 패션, 그래프 신경망, 사용자 중심, 도메인 지식, 이커머스, RFM

학번: 2021-21285