A Study on User Perception and Experience Differences in Recommendation Results by Domain Expertise: The Case of Fashion Domains

Taehyung Noh
yestaehyung@hanyang.ac.kr
Hanyang University
Seoul, Republic of Korea

Haein Yeo
haeinyeo@hanyang.ac.kr
Hanyang University
Seoul, Republic of Korea

Myungjin Kim
myungkim331@hanyang.ac.kr
Hanyang University
Seoul, Republic of Korea

Kyungsik Han*
kyungsikhan@hanyang.ac.kr
Hanyang University
Seoul, Republic of Korea

ABSTRACT

To improve user satisfaction to the recommender systems (RS), it is essential to identify how users would perceive the system and what recommendation results they would prefer. Domain experts also use RS that relate to their work and decision-making, but existing studies have primarily focused on user experience from the general public and somewhat neglected the degree of user perception, understanding, and preference to the RS according to domain knowledge and interest and recommendation algorithm types. In this paper, we present My Own Style (MOS), a dashboard tool designed to analyze a given input fashion image and recommend outfit based on three recommendation algorithms that have different degrees of similarity and diversity. Based on the results from a large-scale user study with 166 participants, our results showed that the participants who have high fashion knowledge and interest (i.e., domain experts) well understood the results of RS and preferred the recommendation algorithm that provides similar outfit, while those who have low fashion knowledge and interest did not well understand the recommendation results as much as the expert group and preferred the algorithm that suggests diverse outfit. Our work is meaningful in providing empirical evidence on how to develop, select, and utilize recommendation algorithms according to domain expertise.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

User perception, Domain knowledge, Fashion recommendation

*Corresponding author

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585641

ACM Reference Format:
Taehyung Noh, Haein Yeo, Myungjin Kim, and Kyungsik Han. 2023. A Study on User Perception and Experience Differences in Recommendation Results by Domain Expertise: The Case of Fashion Domains. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3585641

1 INTRODUCTION

Recommender systems (RS) help users make decisions and improve their service experience by predicting and providing products or content that meet their needs. The direction of the development and use of the recommendation algorithm varies. For example, collaborative filtering (memory-based and model-based) recommends similar items based on behavior patterns (e.g., usage logs) among many users and items [5, 7, 30, 31]. A content-based filtering method provides recommendations based on the characteristics of the content [1, 25, 33]. A hybrid filtering method aims to increase the diversity of recommendation results by combining two or more recommendation algorithms [8, 28]. Recently, as deep learning technology advances, many research efforts have been made to embed and present data in a high dimensional space and construct relationships between users and items in complex graph formats. With such data representations, it has become possible to extract user-favored content (e.g., image, text) from deep learning models pre-trained with large-scale data [10, 20, 34].

Types of recommendation methods are diverse, meaning that the recommendation algorithm may be applied differently depending on the user and the context. Having a one-size-fits-all recommendation algorithm that can satisfy all types of users and contexts may be challenging and perhaps unrealistic. Thus, it is essential to understand how the user would perceive RS and what results the user would expect. In terms of user understanding, previous studies have mostly focused on consumers (e.g., product, movie, music) or only considered recommendation results from a single recommendation algorithm. Since domain experts tend to interpret recommendation results in conjunction with their domain knowledge, experience, and decision-making process [3, 14], their expectations of RS and interpretation of results may differ from those of consumers.

There have been some efforts to understand differences in perceptions and expectations between end users and domain experts regarding RS. However, the actual investigation and discussion of the results have been conducted in a limited fashion. Knijnenburg and Willemsen [16] presented that different preference elicitation (PE) methods should be applied depending on the difference in domain knowledge. However, since the experiment in their work was evaluated only with a single recommendation algorithm, there was only a preliminary discussion on the interpretation of various recommendation algorithm results due to differences in domain knowledge. Knijnenburg et al. [17] confirmed that the preference of recommendation results is influenced by domain knowledge in personal characteristics (PC) but did not discuss the difference about the preference in the recommendation results according to domain knowledge.

Our study targets fashion and design domains that rely heavily on subjective experiences and qualitative judgments and want to utilize quantitative indicators. We aim to develop several algorithms that recommend similar outfit results to a given outfit input and to discuss how to understand and interpret the recommendation algorithms' results according to the domain knowledge level.

In this work, we aim to answer the following two research questions:

- **RQ1**: Do domain experts make a difference in perceiving and understanding the results of recommendations?
- **RQ2**: Do domain experts have different preferred recommendation outcomes?

To answer our research questions, we developed three recommendation algorithms (i.e., image feature-based, fashion element-based, and random-based) that consider the degree of similarity and diversity differently based on previous work to improve a user's satisfaction in RS. Based on the three algorithms, we built the dashboard-based interface named My Own Style (MOS). The MOS interface recommends a set of different images from different algorithms based on the image selected by the user. We conducted a large-scale user study with 166 people to evaluate user perceptions of the recommended results.

Our analysis was based on two user groups clustered by the level of fashion characteristics (FC) that pertained to one's fashion knowledge and interest. As a result, the participants with high FC (high FC group) showed a higher understanding of recommendation results than those with low FC (low FC group). The high FC group preferred algorithms that provided similar results, while the low FC group preferred the diversity algorithm.

Based on these results, we highlight the importance of considering the development and application of recommendation algorithms depending on the degree of interest and understanding of the user and domain. Our findings are meaningful in that we provide empirical evidence from a large-scale user study for the practical direction of the utilization of recommendation algorithms from the user's point of view.

2 RELATED WORK

2.1 User experience in recommender systems

To increase user satisfaction with RS, research on the recommendation algorithm is being conducted using various approaches. Based on the study that user satisfaction increases when similarity and diversity are considered together, various user evaluation studies on recommendation algorithms have been conducted [6, 26, 29, 35]. Kwon et al. [19] showed that the diversity of recommendation results has a positive effect on customers' purchases, and the importance of the diversity of recommendation results has been shown through experiments using online and offline data from fashion companies. However, previous papers primarily investigated recommendation satisfaction only for the general public. Our study aims to identify how users depending on domain knowledge and interest perceive the results of RS and what results they prefer.

2.2 Domain knowledge in recommender systems

In the field of RS, studies comparing the perceptions and preferences of RS according to domain knowledge have not yet been sufficiently conducted. Knijnenburg and Willemsen [16] assumed that users with different domain knowledge prefer different PE methods in deriving recommendation results and confirmed this through experiments using different PE interfaces. Domain experts prefer attribute-based PE, novices prefer case-based PE, and this research showed that differences in domain knowledge affect PE preferences. However, this work has experimented on only a single algorithm. In our research, we developed several recommendation algorithms that reflect the degree of similarity and diversity differently to complement this and focus on checking how domain experts perceive recommendation results by algorithm. Knijnenburg et al. [17] confirmed how PC affect the RS results. The study recommended three different algorithms on the BBC MyMedia player and surveyed participants on the differences in the quality of the recommendation results and whether they preferred various results. This study is meaningful in that it used algorithms reflecting diversity in experiments and confirmed how PC affect them. However, there was no comparison between domain experts and general users, and it was simply confirmed that the PC had an impact on the recommendation results. To summarize, our study differs from previous studies in that we aim to group participants by domain knowledge and interest, namely "fashion characteristics (FC)," to see how each group understands the results from algorithms that have different degrees of recommendation similarity and diversity and satisfies with recommendation results. We also aim to discuss what should be reflected in RS for domain experts.

3 EXPERIMENT DESIGN

In this study, the participants experienced MOS and were asked to check and compare the results of three different recommendation algorithms we developed. Figure 1 shows the MOS interface. This section describes how the user experiment was designed.

3.1 Participants

We participated in the largest design festival held annually in the authors' country (anonymized) and ran a booth for four days. We asked the visitors to our booth whether they could participate in our experiment. A total of 166 participants joined the experiment, and Table 1a summarizes their occupation distributions. Due to the characteristics of the exhibition, the proportion of fashion industry workers, design industry workers, and people interested in fashion was high.

Table 1: (a) The distribution of occupations. (b) The distribution of the datasets used in the recommendation algorithms.

Occupation	Count
Fashion expert group (designer, MD, marketer)	35
Designer (UX, graphic, product, etc.)	59
The general public (other job)	72
Total	166

(a)

Data source of images	Image count
Fashion runway data	401,825
Public fashion data	48,533
Street fashion data	249,255
Total	699,613

(b)

The participants first answered the questions (Table 2) that measured their FC. Then, they were asked to use MOS, experiencing the process of finding the fashion images they preferred. Next, they answered the post-survey questions based on the recommendation results of the three algorithms for the selected image. They were only told that the results were from different recommendation algorithms, and not from the relationship between the results and the algorithms. The average study participation time was about 15 minutes, and the participants received a mobile giftcon equivalent to $5.00 as a reward for participation.

3.2 Developing MOS: dataset

We collected about 699,613 free-of-copyright images on the web. Since fashion runway gives the biggest inspiration for the fashion trend [2], we used fashion runway data labeled with year, season, and brand information between 2010 and 2022 [12]. We also collected and refined datasets produced to recognize fashion and identify trends in the authors' country from the public data repository. Finally, we collected street fashion data on social media to recommend the outfits of various people in their daily lives. MOS shows a wide range of fashion images and recommendation results for the image selected by users. Table 1b shows the specific number of images of the recommended dataset.

3.3 Developing MOS: interface

On the main page of the MOS, 100 images are randomly displayed (Figure 1). If participants want to see fashion images with specific criteria (e.g., year, season, fashion element), they can search for keywords through the filter on the left panel. When the participants clicks on one fashion image, a new window opens showing the information of element details of the selected fashion image. On the same page, the participant can also see the results of the three algorithms (showing 10 images each).

3.4 Developing MOS: recommendation algorithms

For this experiment, we developed three algorithms that reflected different levels of outfit similarity and diversity, based on previous studies [18, 19]:

- **Algorithm 1 (image feature-based recommendation; high similarity)**: To consider outfit similarity, we used the feature vector of images containing many types of visual information (e.g., mood, color, pattern). Considering such information will result in suggesting visually similar images [9, 15, 22]. For better feature extraction from the image, we used the Swin Transformer v2 [23] model pre-trained on the ImageNet-21K dataset and fine-tuned on the ImageNet-1K dataset at resolution 256×256. An input image will be embedded into the 1,537-dimensional vector space from the last hidden layer of the model and the embedding will be used for similarity calculations.
- **Algorithm 2 (element-based recommendation; mid similarity and diversity)**: To consider both similarity and diversity to some extent, we utilized fashion elements that include fashion categories (e.g., sweater, coat, dress) and attributes (e.g., sleeves, length, neckline). Compared to Algorithm 1 that uses visual information, recommended results based on fashion elements will be less visually similar to the selected image. The fashion elements used in Algorithm 2 were defined in Fashionpedia [13], and we used a total of 287 fashion elements (46 categories and 241 attributes) for model training. We used the Swin Transformer base model [24] as the backbone to train the Attribute mask R-CNN using the Fashionpedia dataset. With the model, we labeled 287 fashion elements into 699,613 images. We then embedded labeled fashion elements as 287 dimensional vectors using one-hot encoding.
- **Algorithm 3 (random-based recommendation; high diversity)**: Compared with the results driven by specific algorithms, a random selection algorithm is generally considered to increase diversity [4, 11, 21, 35]. Thus, Algorithm 3 randomly selects 10 fashion images from the image pool that has at least one match of the keywords (e.g., year, season, color, categories, attributes) to the selected image by the user.

For the recommendation of images similar to the one chosen by a user, MOS calculates cosine similarities between the embedding vectors of 699,613 images from Algorithms 1 and 2 and recommends the top 10 images with the highest cosine similarity. Figure 2 illustrates the results suggested by each algorithm.

3.5 Pre-study questionnaire

To group participants' FC, we surveyed their interest in fashion, knowledge of the fashion domain, and occupations. Fashion domain knowledge and interest in fashion are important factors that influence experts' decision-making [27, 32]. We thus measured those aspects through the five-point Likert-scale questions (1: strongly disagree, 5: strongly agree) as shown in Table 2. Each participant's answers were averaged, and if the average was higher than the median, the participant was assigned to the high FC group; otherwise, to the low FC group. The high FC group has 100 participants, while the low FC group has 66 participants. Figure 3 illustrates the results of analyzing the relationship between FC values and occupations (i.e., fashion experts, designers (UX, grapic, etc.), and the public).

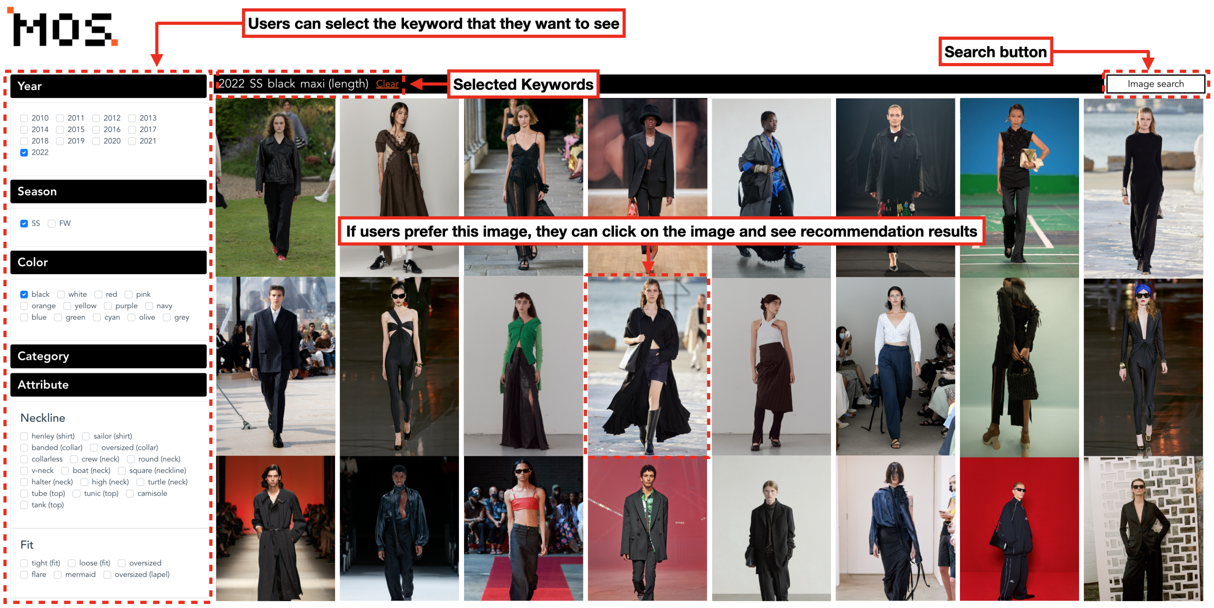

Figure 1: MOS interface for image selection. Users can search for keywords (e.g., 2022, SS, black, maxi length) and click on a fashion image to get recommended results.

Figure 2: Top 10 examples of each recommendation algorithm. (a) shows Algorithm 1 results that consider high similarity. (b) shows Algorithm 2 results that consider mid similarity and diversity. (c) shows Algorithm 3 results that consider high diversity.

Table 2: Questions about fashion characteristics

ID	Question	Implication
Q1	I am more interested in fashion than most other people	Fashion interest
Q2	I feel I know a lot about fashion	Fashion domain knowledge
Q3	I would classify myself as an expert on fashion	Fashion domain knowledge

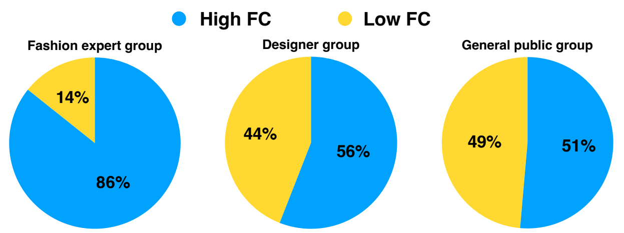

Figure 3: Frequency analysis of the FC groups according to occupation.

The FC values of the fashion expert group accounted for 86% of the high rate, 56% of the designer group, and 51% of the public group. The chi-square test confirmed a significant association between the FC values and occupational groups ($X^2(2)$=12.3, p=0.002).

3.6 User survey

After experiencing MOS, the participants were asked to answer the questions about their understanding of and satisfaction with the recommendation results. First, to confirm the differences in understanding of the participants' recommendation results, we asked them to choose the most similar algorithm and the most diverse algorithm. Next, to confirm the difference in the participants' satisfaction with the results, we asked them to choose the results that they were the most satisfied with. Table 3 shows the three questions.

4 RESULTS

For the questions of user experience in Table 2, we considered two user groups based on the FC values (low FC and high FC). We then measured statistical differences in the answers between the two FC groups, using the chi-square test.

4.1 RQ1: Do domain experts make a difference in perceiving (understanding) the results of recommendations?

Figure 4a shows the results of analyzing the responses to the algorithm, which is thought to be similar between the low and high FC groups. In this case, the recommendation results were obtained from Algorithm 1. Overall, the answers between the two groups were marginally statistically different ($X^2(2)$=5.00, p=0.082). The high FC group answered correctly (78%), which was higher than the low FC group (64%). The low FC group selected two other algorithms (2 and 3) with quite high ratios. This result indicates that the high FC group understood the recommendation results generally better than the low FC group.

On the other hand, for recommendation diversity (Figure 4b), no significant difference was found between the FC groups ($X^2(2)$=0.0153, p=0.992). This result indicates that the participants had generally consistent criteria when evaluating the diversity of recommendation results.

4.2 RQ2: Do domain experts have different preferred recommendation outcomes?

To answer RQ2, we analyzed the results of satisfactory recommendations according to the FC group (Figure 4c). The high FC group had the highest rate of satisfying Algorithm 1 at 71%. Algorithms 2 and 3 reflected diversity was 29%, significantly lower than Algorithm 1. On the other hand, the low FC groups selected 52% of Algorithms 2 and 3, reflecting diversity, and there was a slight difference from 48% of Algorithm 1. As a result of the statistical comparison between the two groups, we found a significant difference ($X^2(2)$=9.74, p=0.008). In other words, the preferred recommendation results differed between the two FC groups. This can be interpreted to mean that results reflecting diversity are preferred for low FC groups, and similar results are preferred for high FC groups. These results emphasize the importance of providing different recommendation results, considering the differences in domain knowledge and interest.

5 DISCUSSION

In this work, we identified how experts in the fashion domain perceive multiple recommendation algorithms and what results they prefer. By dividing the understanding of the recommendation algorithms and the satisfaction with the results according to the difference in FC (domain knowledge and interest), we found that algorithm understanding and satisfaction were different depending on FC. Our findings emphasize that when developing a RS, users of the system should be considered. In this section, we discuss the key findings of our study results and the implications for designing RS for domain experts.

5.1 Key findings

First, we found that the understanding of the recommendation results differed between the two FC groups. The high FC group was well aware of the results for the similarity and diversity intended by the algorithms, while the low FC group was not. Our study results also showed that the high FC group preferred similar recommendation results, while the low FC group preferred diverse results. These results show that user perceptions of recommendation results are quite different between the high FC and the low FC groups. This further means that different algorithmic approaches need to be considered depending on who the target user will be. It appears that the

Table 3: Questions about user perception of the recommendation results.

ID	Question	Implication
Q1	Which algorithm do you feel is the most "similar" to your chosen image?	Similarity
Q2	Which algorithm do you feel is the most "diversity" to your chosen image?	Diversity
Q3	Which algorithm do you feel is the most "satisfied" with your chosen image?	Satisfaction

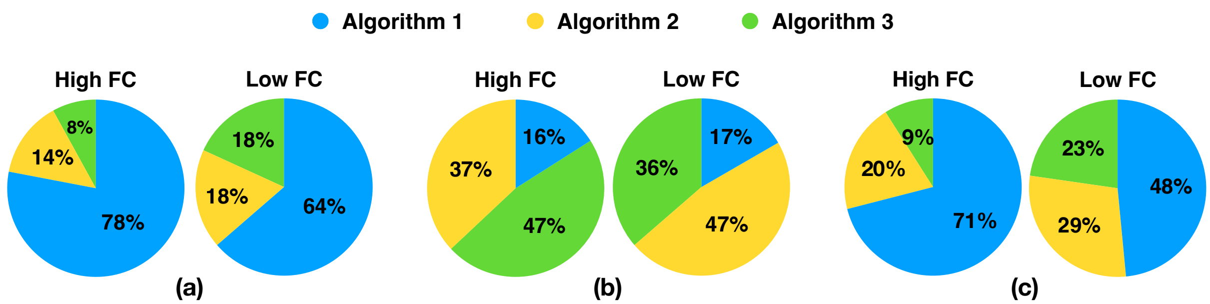

Figure 4: Frequency analysis of algorithm selection by the FC group. The participants were asked to choose (a) which algorithm generates the most similar outfits, (b) which algorithm generates the most diverse outfits, and (c) which algorithm generates the most satisfying results.

high FC group wants to see more details about how the algorithm works, because people in that group would think about how to leverage the results for their design task or decision making [12]. Domain experts are likely to take a close look at and interpret the recommendation results in conjunction with knowledge, experience, and decision-making processes. Prior research also highlights the importance of designing a decision-making system based on an understanding of domain experts' work practices and expectations [3, 14]. Our study results align with those insights from different experiment conditions (considered multiple recommendation algorithms and group differences) and suggest providing a clear explanation of the recommendation algorithms to be adopted by domain experts.

6 CONCLUSION

This study investigated how fashion domain knowledge and interest affect the understanding and preference of different recommendation results. We developed three recommendation algorithms with different degrees of similarity and diversity and designed a dashboard, called MOS, that displays different sets of recommendation results for a given fashion image. The results of the experiment with 166 participants showed that the high FC group understood the recommendation results and preferred similar results, while the low FC group did not understand the results as much as the high FC group and preferred diverse results. These results highlight the different perceptions of recommendation algorithms and results according to domain knowledge and interest. Our study contributes to providing empirical evidence from the large-scale study in algorithm development, selection, and utilization depending on user group. In our future work, we plan to identify other elements that experts consider in addition to similarity and diversity and to expand the study with more experts and a better designed interface. Our study will provide insights for researchers, practitioners, and designers who want to develop and utilize RS for domain experts.

ACKNOWLEDGMENTS

This research was supported by the National Research Foundation (NRF) and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021M3A9E4080780, No.2020-0-01373, No.IITP-2022-2018-0-01431).

REFERENCES

[1] Chumki Basu, Haym Hirsh, William Cohen, et al. 1998. Recommendation as classification: Using social and content-based information in recommendation. In *Aaai/iaai*. 714–720.

[2] Vertica Bhardwaj and Ann Fairhurst. 2010. Fast fashion: response to changes in the fashion industry. *The international review of retail, distribution and consumer research* 20, 1 (2010), 165–173.

[3] Carrie J Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S Corrado, Martin C Stumpe, et al. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–14.

[4] Pablo Castells, Neil Hurley, and Saul Vargas. 2022. Novelty and diversity in recommender systems. In *Recommender systems handbook*. Springer, 603–646.

[5] Chih-Ming Chen, Chuan-Ju Wang, Ming-Feng Tsai, and Yi-Hsuan Yang. 2019. Collaborative similarity embedding for recommender systems. In *The World Wide Web Conference*. 2637–2643.

[6] Tommaso Di Noia, Vito Claudio Ostuni, Jessica Rosati, Paolo Tomeo, and Eugenio Di Sciascio. 2014. An analysis of users' propensity toward diversity in recommendations. In *Proceedings of the 8th ACM Conference on Recommender Systems*. 285–288.

[7] Michael D Ekstrand, John T Riedl, Joseph A Konstan, et al. 2011. Collaborative filtering recommender systems. *Foundations and Trends® in Human–Computer Interaction* 4, 2 (2011), 81–173.

[8] G Geetha, M Safa, C Fancy, and D Saranya. 2018. A hybrid approach using collaborative filtering and content based filtering for recommender system. In *Journal of Physics: Conference Series*, Vol. 1000. IOP Publishing, 012101.

[9] Ming He, Shaozong Zhang, and Qian Meng. 2019. Learning to style-aware Bayesian personalized ranking for visual recommendation. *IEEE Access* 7 (2019), 14198–14205.

[10] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. 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*. 639–648.

[11] Min Hou, Le Wu, Enhong Chen, Zhi Li, Vincent W Zheng, and Qi Liu. 2019. Explainable fashion recommendation: A semantic attribute region guided approach. *arXiv preprint arXiv:1905.12862* (2019).

[12] Youngseung Jeon, Seungwan Jin, and Kyungsik Han. 2021. FANCY: human-centered, deep learning-based framework for fashion style analysis. In *Proceedings of the Web Conference 2021*. 2367–2378.

[13] Menglin Jia, Mengyun Shi, Mikhail Sirotenko, Yin Cui, Claire Cardie, Bharath Hariharan, Hartwig Adam, and Serge Belongie. 2020. Fashionpedia: Ontology, segmentation, and an attribute localization dataset. In *European conference on computer vision*. Springer, 316–332.

[14] Seungwan Jin, Hoyoung Choi, and Kyungsik Han. 2022. AI-Augmented Art Psychotherapy through a Hierarchical Co-Attention Mechanism. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 4089–4093.

[15] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian McAuley. 2017. Visually-aware fashion recommendation and design with generative image models. In *2017 IEEE international conference on data mining (ICDM)*. IEEE, 207–216.

[16] Bart P Knijnenburg and Martijn C Willemsen. 2009. Understanding the effect of adaptive preference elicitation methods on user satisfaction of a recommender system. In *Proceedings of the third ACM conference on Recommender systems*. 381–384.

[17] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User modeling and user-adapted interaction* 22, 4 (2012), 441–504.

[18] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems–A survey. *Knowledge-based systems* 123 (2017), 154–162.

[19] Hyokmin Kwon, Jaeho Han, and Kyungsik Han. 2020. ART (Attractive Recommendation Tailor) How the Diversity of Product Recommendations Affects Customer Purchase Preference in Fashion Industry?. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2573–2580.

[20] Richard P Larrick and Daniel C Feiler. 2015. Expertise in decision making. *The Wiley Blackwell handbook of judgment and decision making* 2 (2015), 696–721.

[21] Neal Lathia, Stephen Hailes, Licia Capra, and Xavier Amatriain. 2010. Temporal diversity in recommender systems. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. 210–217.

[22] Qiang Liu, Shu Wu, and Liang Wang. 2017. Deepstyle: Learning user preferences for visual recommendation. In *Proceedings of the 40th international acm sigir conference on research and development in information retrieval*. 841–844.

[23] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. 2022. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12009–12019.

[24] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10012–10022.

[25] Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. 2011. Content-based recommender systems: State of the art and trends. *Recommender systems handbook* (2011), 73–105.

[26] Christian Matt, Alexander Benlian, Thomas Hess, and Christian Weiß. 2014. Escaping from the filter bubble? The effects of novelty and serendipity on users' evaluations of online recommendations. (2014).

[27] Aron O'Cass. 2004. Fashion clothing consumption: antecedents and consequences of fashion clothing involvement. *European journal of Marketing* (2004).

[28] Ashish Pal, Prateek Parhi, and Manuj Aggarwal. 2017. An improved content based collaborative filtering algorithm for movie recommendations. In *2017 tenth international conference on contemporary computing (IC3)*. IEEE, 1–3.

[29] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. 157–164.

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

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

[32] Douglas J Tigert, Lawrence J Ring, and Charles W King. 1976. Fashion involvement and buying behavior: A methodological study. *ACR North American Advances* (1976).

[33] Robin Van Meteren and Maarten Van Someren. 2000. Using content-based filtering for recommendation. In *Proceedings of the machine learning in the new information age: MLnet/ECML2000 workshop*, Vol. 30. 47–56.

[34] Chi-Man Wong, Fan Feng, Wen Zhang, Chi-Man Vong, Hui Chen, Yichi Zhang, Peng He, Huan Chen, Kun Zhao, and Huajun Chen. 2021. Improving conversational recommender system by pretraining billion-scale knowledge graph. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. IEEE, 2607–2612.

[35] Mi Zhang and Neil Hurley. 2008. Avoiding monotony: improving the diversity of recommendation lists. In *Proceedings of the 2008 ACM conference on Recommender systems*. 123–130.