ART (Attractive Recommendation Tailor): How the Diversity of Product Recommendations Affects Customer Purchase Preference in Fashion Industry?

Hyokmin Kwon
hmkwon@lfcorp.com
LF Corporation
Seoul, South Korea

Jaeho Han
woghrnt2@ajou.ac.kr
Department of Computer Engineering, Ajou University
Suwon, South Korea

Kyungsik Han*
kyungsikhan@ajou.ac.kr
Department of Computer Engineering, Ajou University
Suwon, South Korea

ABSTRACT

This study examines the impact of the "diversity" of product recommendations on the "preference" of a customer, using online/offline data from a leading fashion company. First, through interviews with fashion professionals, we categorized the characteristics of customers into four types – gift, coordinator, carry-over, and trend-setter. Then, using a hybrid filtering method, we increased the accuracy and diversity of recommended products. We derived 13 salient features that reflect customer behavior based on the Purchase Funnel model and built a classification model that predicts a customer's preference rates. Second, we conducted two large-scale user tests with 20,000 real customers to verify the effectiveness of our recommendation system. Study results empirically demonstrated the importance of diversity of recommended products. The more diverse the product recommendations were, the higher the purchase rate, the average purchase amount, and the cross purchase rate were observed. In addition, we tracked the customers' purchase for two months after the user tests and found that diverse product exposure positively influenced customer retention (e.g., repurchase rate, amount).

CCS CONCEPTS

• **Information systems** → **Recommender systems**; **Personalization**; • **Computing methodologies** → Machine learning.

KEYWORDS

Fashion recommendation; Diversity; Feature engineering; Preference modeling; Large-scale user test

ACM Reference Format:
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 and Knowledge Management (CIKM '20), October 19–23, 2020, Virtual Event, Ireland*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3340531.3412687

*Corresponding author

CIKM '20, October 19–23, 2020, Virtual Event, Ireland

ACM ISBN 978-1-4503-6859-9/20/10…$15.00
https://doi.org/10.1145/3340531.3412687

1 INTRODUCTION

A recommendation system is more pronounced when it reflects the characteristics of the industry. Thus far, the study of fashion product recommendations has focused on eliciting similar products with better accuracy. These methods have the disadvantage of repeatedly recommending by and large similar products and increasing customers' fatigue. Studies have shown that in the fashion industry, a system recommending only similar products could reduce the number of choices and the quality of recommendation. Studies also pointed out that, without giving consideration to the diversity of the recommendations, the overall consumption of the customer could be reduced [7, 15]. Thus, in this study, we focus on "recommendation diversity," model a customer's "preference" for recommended products based on his/her online behaviors, and verify the effectiveness of the model based on the metrics (e.g., purchase amount, cross purchase rate, retention) used in fashion marketing through two large-scale user tests with 20,000 real customers.

First, we enhanced diversity through a combination of algorithms in which fashion customer characteristics are considered. Through a focus group interview (FGI) with seven fashion industry professionals, four types of fashion customers were identified : (1) *gift type*: customers who purchase products to give to others; (2) *coordinator type*: customers who purchase products associated with previous purchases; (3) *carry-over type*: customers who purchase products similar to existing purchases; and (4) *trend-setter type*: customers who are affected by the trends of other people's purchases. We selected algorithms that characterize each type (i.e., demographic filtering, apriori, item-based collaborative filtering and user-based collaborative filtering). We then employed *hybrid filtering*, a method of combining the results of algorithms independently [6, 8]. Through this method, we take advantage of each algorithm and increase the diversity of recommended products.

Second, we developed features that pertain to a customer's purchase behaviors that were used in modeling. Such features were determined by fashion professionals and based on Lewis's Purchase Funnel, which describes a four-stage of customers' decision-making steps [24, 26]: Awareness-Interest-Desire-Action. We initially developed 320 variables, considering the behavioral information related to the customer preference at each stage and had 13 final features used for making recommendation. With these features, the "preference probability" for each recommended product was predicted based on whether the customer clicks on the recommendation. We used various machine learning algorithms (e.g., eXtreme Gradient Boosting (XGBoost) [12], K-Nearest Neighbors, Random Forest, Logistic Regression), and the model with XGBoost yielded the best performance.

Third, with the developed model, we conducted two large-scale user tests with 20,000 real customers. The first user test examined whether the *existence* of the recommendation system influences the purchase rate and cross-purchase rate (i.e., purchasing a product that has not been previously experienced; which is considered very important in fashion industry). As a result, the experimental group (with recommendation) showed 15.7% higher response rate, $81 higher purchase amount per customer, and 8% higher cross purchase rate than the control group (without recommendation). The second user test examined whether having a *diversity* of products has a significant impact on customer satisfaction. The experimental group and the control group received a list of recommendation items based on hybrid filtering (greater diversity) and apriori (less diversity), respectively. Results showed that the experimental group had 3.2% higher purchasing rate, $13 higher average purchase amount per customer, and 1.1% higher cross purchase rate than the control group. In addition, we observed the customers who had received product recommendation after the test and found positive influence of recommendation diversity on customer retention, such as repurchasing and transaction.

In summary, the contributions of this study are as follows.

- We define four customer characteristics and apply hybrid filtering to increase product diversity.
- We present 13 salient features of purchase-related behaviors and develop a model that quantifies a customer's preference for recommended products.
- We demonstrate empirically the practical effectiveness (i.e., purchase rates, cross purchase rates, and retention rates) of our recommendation system from two large-scale user tests.

2 RELATED WORK

2.1 Recommendations in Fashion Industry

Relatively few studies on recommendations in the fashion domain have been done compared to other domains, and even those existing studies have mostly dealt with similarity measures for product images [30, 31]. While there have been some attempts to reflect *customer characteristics* as a key component of the recommendation system, they tend to reflect only a single attribute, such as the importance of coordination or the time to purchase after launch [10, 19, 25]. However, because fashion is inherently personal and social, it is important to consider not only individual preferences or past purchasing activities, but also the way in which the social environment plays a role such as through fashion rules and the latest trends [23]. Furthermore, if only individual preferences are considered in terms of recommendations, a customer's fatigue may increase due to the monotony of the recommended items [28]. To enhance the diversity of recommendations that are being overlooked, we selected and combined the four algorithms that reflect customer's purchasing characteristics.

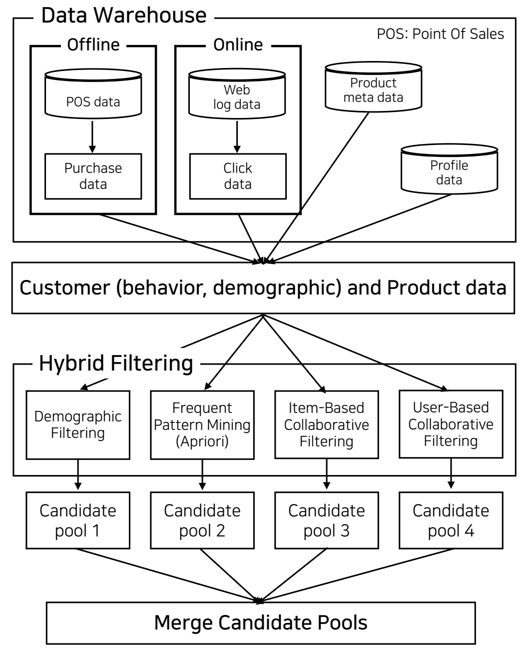

Figure 1: Data collection & candidate generation pipeline.

2.2 Feature Engineering

For developing a machine learning model that finely grasps user preferences, many researchers use domain knowledge for feature engineering. For example, based on the domain knowledge that young people prefer new songs and older people tend to listen only to songs that are often heard, Zhang et al. [34] created a feature called "Count of songs each user listened," and demonstrated its effectiveness. In a software crowdsourcing domain, Zhu et al. [35] presented the features that characterize similarity or relevance (i.e., work-to-developer similarity, work-to-work similarity, skill-based relevance, location-based relevance) between developers and their tasks. Mahsa et al. [5] conducted feature engineering to recommend relevant events to social media users who have different preferences. Their study analyzed the characteristics of a given dataset, divided the features into three groups (i.e., User-level, Event-level, and User-event-level), and saw their effectiveness on recommendation.

Overall, researchers in many domains have conducted feature engineering by reflecting the opinions and knowledge of domain professionals. However, applying the process of feature engineering to a fashion domain has not been well investigated. In this work, we present a list of salient purchase-related behavioral features based on the purchase Funnel theory and the feedback from fashion professionals through the interviews. We also empirically demonstrated the effectiveness of those features in predicting a customer's preference to fashion products, which were also validated by two large-scale user studies with 20,000 real customers.

3 DATA

In this study, we used both offline customer purchase history and online mall click history data from one of the leading fashion companies in South Korea [1] for the analysis. We used products' metadata

[1] http://lfmall.com/

(e.g., product types, brands) and customer profile information (e.g., gender, age, off/online registration date) for the product recommendation analysis (See Figure 1). We used the data collected for one year (from September 2017 to August 2018) to capture rapidly changing trends by season (Spring-Summer, Fall-Winter). We randomly selected 30,000 customers out of 1.8 million offline and online customers and extracted their purchasing history over one year. As a result, we obtained information on 163,890 product purchases and 1.1 million product clicks and used them as a *training dataset*. We collected additional data that covers the three months following the learning period (from September to November 2018). We used the three-month data for preference forecasting as a *test dataset*. We defined a preference as a click on a particular product (binary), a target variable Y. If a customer clicks on a particular recommended product, Y is set to 1, otherwise to 0.

4 RESEARCH METHOD

4.1 Focus Group Interviews (FGI)

Before developing a recommendation system, we first conducted focus group interviews (FGI) with fashion professionals to derive unique customer characteristics in the fashion domain. We interviewed seven professionals, including four customer experience management (CEM) professionals and three customer relationship management (CRM) professionals, twice. We aimed to derive (1) customer characteristics and (2) customer behavior information.

4.2 Fashion Customer Characteristics (FGI #1)

We led the discussion with seven fashion professionals and asked the following question, "what types of customers do you have?" As a result, the professionals defined four characteristics of the fashion industry customers: *gift type*, *coordinator type*, *carry-over type*, and *trend-setter type*, depending on the customer's purchases.

First, *gift type* customers make purchase as a gift to someone else other than themselves. According to the data, the portion of female customers who buy men's products or young customers purchase brands for others who are in a different age group (e.g., 60's) was very high. We found about 43% of total purchases were in this type from our data. Second, *the coordinator type* customers purchase additional products in coordination with the previously purchased ones. This means that the time difference association between the previous and subsequent purchases is important. According to our data, of the 160,000 customers who purchased suits during the year, 58% purchased shirts, with 49% purchased accessories such as belts and ties. Third, *the carry-over type* customers continuously purchase products similar to prior ones. The fashion industry strategically produces carry-over styles for a continuously selling product from the previous season, to reflect these customers' needs. Among 26,000 customers who purchased products last season, 25% of those who purchased the product this year have a "carry-over style," which is a traditional caribou style, with a very high rate of repurchase for similar products. Lastly, *the trend-setter type* customers are greatly affected by purchasing trends. Of a total 40,000 launched product groups, the ratio of customers who purchased the 50 top products is 12% of all customers who purchased the same top products, indicating that trend-setters are greatly affected by popular products.

4.3 Product Diversity with Hybrid Filtering

4.3.1 Algorithm for each customer type. First, for gift type customers, we recommend products suitable for the customer's profile by utilizing demographic information to reflect their needs. Customer's gender, age, and brand-product groups were formed into k-Means clustering as variables and were divided into five clusters (i.e., hardcore loyal, brand loyal, cherry picker, early bird and switcher). Finally, the products recommended to each customer are the best products of the season [2], which reflect the characteristics of the corresponding cluster. We called this approach Demographic Filtering (DF).

Second, for the coordinator type, we considered the timing of the purchase. We chose to quantify how strongly the purchase of one product affects that of the other and extracted the preferred product by the time of purchase [4]. We used an apriori algorithm, which removes patterns that fail to pass a certain level of a numerical threshold, enabling rapid calculation [1]. To reduce the computations, we set the minimum support (i.e., probability of purchasing items A and B at the same time) to 0.4.

Third, to reflect the carry-over type, we used Item-Based Collaborative Filtering (IBCF), because it captures the characteristic of purchasing same and/or similar products from the previous season. Because the implementation of IBCF for all products can result in overloading [14], we considered the top 20% customers (VIPs) in sales over the past year and extracted the products with the highest similarity [3, 6].

Lastly, to reflect trend-setter type, we used User-Based Collaborative Filtering (UBCF), because this type of customers are greatly influenced by other customers. Unlike IBCF, which reduces the amount of computation by only analyzing the purchase patterns of the top-of-sale VIPs, UBCF applies dimension reduction to reduce computation overloads. We used the Google's MinHash clustering method to reduce customer buying patterns [13, 21]. The Jacquard method was used to calculate the similarity among customers, and the most similar customer purchases were chosen as the recommended products [22].

4.3.2 Hybrid filtering. Each algorithm described above has the advantage of reflecting each type of the fashion customers, but there are also shortcomings. For example, demographics filtering does not reflect individual tastes [29], IBCF only tends to recommend products similar to the past purchasing history [17], and UBCF does not well recommend new product groups or products with few sales, which may reduce the diversity of recommended items [20]. Therefore, we employed hybrid filtering, which combines four methods, to overcome the shortcomings of each recommendation algorithm and increase diversity. In this study, among the hybrid filtering methods introduced by Burke et al. [9], we chose to show the recommended results of each algorithm. This approach can increase the diversity of recommendations [9]. Each algorithm returns top 5 products; thus, 20 products were recommended to a customer. The reason for choosing 20 is because this is the average number of VIP customers (in the top 4% who purchase more than $5,000 per year) purchases per season. If there exist overlapping products among the algorithms, the total will be less than 20. In this case, we filled the products in order of the algorithms with high Click Through Ratio (CTR). To verify the diversity of recommendations among

Category		Key question	Primary feature	Count
Customer information		What is the customer characteristics?	Number of assembly members, gender, age distribution	4
		How long is the existing customer maintained?	Active members, dormant customers	6
		How many new customers are coming in?	Number of new customers, proportion of new customers	5
		How good are new customers coming in?	Number of orders per new customer, unit price, survival rate	14
		How is the customer using the channel?	Order channel configuration ratio, number of multi-channel orders	7
		Who is the customer group that needs to be targeted?	Dormant conversion customers, contribution of sales by member grade	10
Inflow channel		Which inflow channel does the customer enter?	Unique visitor/login visitor by inflow path, visit ratio by route	7
		Which inflow channel is the most effective?	Rate of order conversion by inflow path, sales by cost	25
Promotion		What is the response rate and effect of the campaign?	Sensitivity by campaign, SMS order conversion rate	36
		What promotion is cost effective?	Purchasing conversion rate by promotion and cost-effectiveness	21
Customer behavior	Visit	How many customers are visiting?	Number of visitors (unique visitor/login visitor), visit count	5
		What is the visit pattern? (frequency, time)	Number of visits per person/period, page count per visit	5
		Which channel do you shop at?	Visit ratio and sales ratio by visiting channel	17
		Where do you go after your visit?	Moving page by first page, rate of deviation immediately	5
	Explore	How active are you in your search?	Explore product type per person, click count	13
		What course (store) does the product go through?	Page view by planning competition and page view by brand/search	19
		How long will it lead to order/departure after navigation?	Forward/exit ratio to product details page	6
	Order	What is the customer's order pattern?	Number of orders by period and unit price by order pattern	31
		How much do customers order when they visit the store?	Turnover/departure rate by Funnel step	9
		How do I use a shopping cart before ordering?	Shopping basket utilization rate, matching rate	7
	Cancel/Return	What is the cancellation rate and return rate after ordering?	Cancellation rate by product, net order rate by product	26
	Pain point	What is the customer's pain point?	Number of failures per person, number of incidents per customer complaint type	14
Product		What is the order conversion rate by product group?	By product category - order conversion rate by store	13
		What is the conversion rate?	By brand-order conversion rate by store	10
		What is the efficiency of exhibition by product?	Average exhibition efficiency by product category	5

Table 1: Examples of 320 customer behavior indicators identified through FGI.

the five algorithms, we calculated the prediction coverage for each algorithm. The prediction coverage refers to the size of the entire set of items compared to the size of the items recommended by the system (as shown Eq. 1) and has been widely used as an indicator of diversity in recent recommendations [16].

$$\textit{prediction coverage} = \frac{I_p}{I} \tag{1}$$

As shown in Figure 2, the prediction coverage ranges from 24% to 78%, and the hybrid filtering method was the highest among the algorithms.

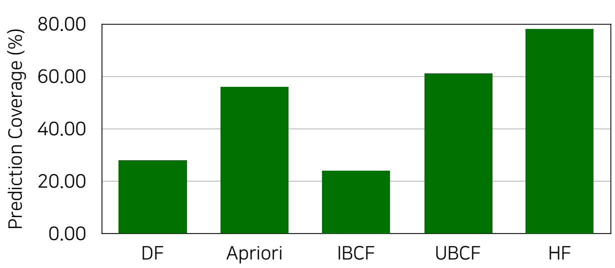

Figure 2: Result of diversity measure (HF: Hybrid Filtering).

4.4 Feature Engineering (FGI #2)

We conducted FGI once again to decide features that reflect the customer's preferences to products. It was conducted five times (2 hours each) with the same seven professionals. The feature engineering process went through four phases (See Figure 3). First, we extracted 320 indicators of key customer behaviors. Second, we reduced the 320 indicators to 120 through a correlation analysis with sales. Third, we reduced the number of indicators to 35, so that preferences could be analyzed in accordance with the Purchase Funnel theory. Finally, we chose 13 features using three criteria (i.e., feasibility, association with preferences, and information related to customers and products).

Extraction of key customer behavior information by type. Professionals were asked to discuss customers' characteristics and behaviors that pertain to customer preferences, followed by primary features that could give answers to the key questions. The professionals identified 320 indicators of customer behavior information, which were grouped by five categories: customer information, inflow channel, promotion, customer behavior and product. Table 1 shows some examples of these customer behaviors.

Correlation with Sales. We went through the process of determining which variables are highly correlated with purchases. We eliminated the indicators with low correlations with purchases through this process. Specifically, the multinomial chi-square test removed an indicator with a p-value greater than 0.05. As a result, 120 variables remained.

Selection of Feature based on Purchase Funnel. We employed the Purchase Funnel theory in order to obtain features that explain customer's preference based on the theory. This theory is a widely-used marketing model that theoretically presents the steps of a customer's product purchase [24]. This model explains Customer

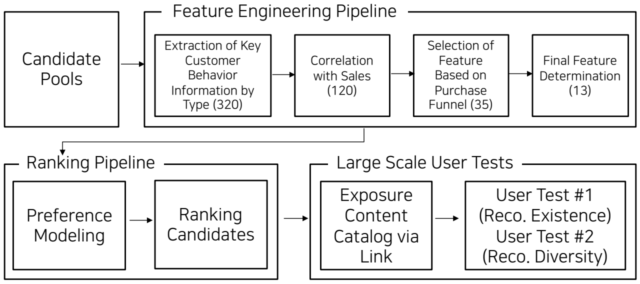

Figure 3: Feature engineering, ranking pipeline, and large-scale user tests (with 20,000 customers).

Decision Journey (CDJ), which contains four phases: Awareness, Interest, Desire, and Action. Consumers get Awareness about the product, then Interest to Desire, and lead to Action of purchase.

Over time, variants with different steps have emerged, but the model has been validated in many areas (e.g., business, consumer behaviors) and the basis for promotions or customer relationship management (CRM) strategies [26]. We grouped the indicators selected in the previous step. For example, we looked at whether the recommended product conforms to the customer's needs (awareness), whether it is placed in the shopping cart (interest), the degree of current desire to purchase seen through the purchase cycle (desire), and finally whether the action actually has a history of recent purchases of the related product (action). As a result, 35 indicators were selected in this phase.

Final feature determination. With 35 indicators, the seven professionals selected final candidates based on "realizability" of data collection, "relationship with preference," and the question, "Is it an indicator that comprehensively considers customer and product information?". Table 2 shows the final 13 features.

Type of CDJ		Feature
Awareness	AW1	Whether a customer's needs match a specific item
	AW2	Period from the last click date
	AW3	Number of page views in last 1 month
	AW4	Number of page views in last 3 months
Interest	IN1	Number of specific brand items in shopping cart in last 1 month
	IN2	Number of specific brand items in shopping cart in last 3 months
Desire	DE1	Period from the last purchase date
	DE2	Total purchase quantity of specific item in last 1 month
	DE3	Total purchase quantity of specific item in last 3 months
Action	AC1	Specific brand purchase amount in last 1 year
	AC2	Specific brand purchase quantity in last 1 year
	AC3	Specific item purchase amount in last 1 year
	AC4	Specific item purchase quantity in last 1 year

Table 2: Final 13 CDJ features used in modeling.

4.5 Preference Modeling

We prepared recommendation items for each customer from hybrid filtering, and had 13 features ready for preference modeling. As previously explained, the target value (dependent variable) for predicting a customer's preference was defined as a click on those items within a given season (three months). We examined four well-known machine-learning algorithms: eXtreme Gradient Boosting (XGBoost) [12], K-Nearest Neighbors (KNN), Random Forest (RF), and Logistic Regression (LR). We compared common performance indicators by measuring accuracy, F1-score, precision, and recall for each model. Table 3 presents the results. XGBoost showed the greatest performance (Acc: 91.75%, F1: 90.54%), followed by RF, KNN, and LR. The parameters of the XGBoost model with the best performance are as follows: max depth=6, learning rate=0.01, gamma=default, and n_estimators=1,200.

Model	Accuracy (%)	F1-score (%)	Precision (%)	Recall (%)
LR	89.12	88.24	88.70	87.78
KNN	90.45	89.31	89.44	89.18
RF	90.39	89.42	89.34	89.50
XGBoost	**91.75**	**90.54**	**90.75**	**90.33**

Table 3: Evaluation of four models.

4.5.1 Feature Importance. To examine how the 13 features affected the performance of the model, we measured the feature importance based on the XGBoost model. According to Figure 4, two features (purchases in the past month: DE2, three months: DE3) in the Desire group showed the highest, followed by item and brand sales in the Action group. Those features are related to recency, frequency and monetary (RFM) variables, which are the most frequently used in predicting future purchasing behavior. This means that our model also confirms the representation and importance of such variables in predicting preferences. On the other hand, two features related to the number of items (IN1 and IN2) in the shopping cart show the lowest importance. The reason for this can be interpreted as a desire reduction strategy among the consumption control strategies presented by Hoch et al. [18]. In other words, the process of adding a product into a shopping cart is a postponement in the four methods of reducing desire (e.g., avoidance, postponement, distraction, substitution), as opposed to an immediate purchase or a purchase that is not delayed too long.

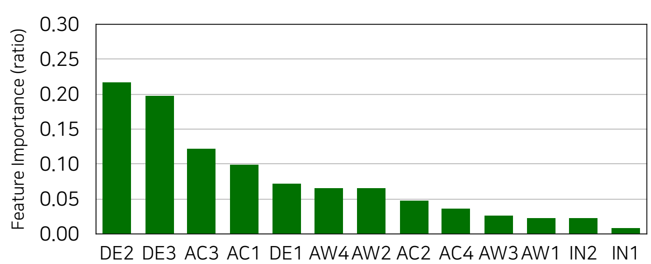

Figure 4: Feature importance ranking of the XGBoost model.

5 LARGE-SCALE USER TESTS

The goal of the two large-scale user (customer) tests was twofold. The first study goal was to investigate the customer response result with or without the "existence" of recommendations. The second study goal was to investigate the customer response result with or without consideration of the "diversity" of the recommended

Figure 5: Example of recommended products that is actually exposed to customers.

products. Figure 5 shows an example of product images used in the user test. Each customer received a set of different images of recommended products via multimedia messaging service (MMS).

The two tests looked to answer the following four hypotheses.

- *H1: Existence of recommendation has a significant effect on customer purchasing response.*
- *H2: Existence of recommendation has a significant effect on customer retention and satisfaction.*
- *H3: Diversity of product recommendations has a significant effect on customer response.*
- *H4: Diversity of product recommendations has a significant effect on customer retention and satisfaction.*

Figure 6 illustrates the offline test period in which hypotheses 1 to 4 were conducted.

	June '19 (30days)	Jul.~Sep.'19 (60days)
	H1	***H2***
First user test	Reco. Provided (experimental) Reco. Not provided (control)	Measuring repurchase
	H3	***H4***
Second user test	HF applied (experimental) Apriori applied (control)	Measuring repurchase

Figure 6: Large scale test period (HF: Hybrid Filtering).

5.1 User Test #1 (Recommendation Existence)

The first user test was done with 10,000 customers. We defined the experimental and control groups (5,000 customers per group). The experimental group received an instant message containing a promotion link that shows a list of 20 recommended items based on our model. On the other hand, the control group received an instant message without a promotion link to the recommendation web page. The only difference between the experimental and control group MMS was the presence or absence of the promotion link.

5.1.1 H1: Effect on customer purchasing response. Regarding H1, we found that inducing purchases through recommendation was effective, and the average purchase amount per customer also significantly increased (Table 4). According to the fashion professionals, it is by and large very difficult to generate cross-purchases that lead customers to experiencing new brands and products. However, in our test, the cross purchase rate increased as a result of the recommendation. This can be considered very valuable, because

Index	Reco. provide	Reco. not provide
Target customer	5,000	5,000
Purchase customer	2,407	1,621
Purchase amount	$1,180,867	$663,324
[a] Customer transaction	$491	$409
[b] Purchase response rate***	48.14%	32.42%
[c] Cross purchase rate***	19.44%	11.29%

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[a] Customer transaction = purchase amount / number of purchase customers
[b] Purchase response rate = number of purchase customers / number of target customers
[c] Cross-purchase rate = number of customers who purchased the first experienced brand product / number of purchase customers

Table 4: Results of the first user test for H1.

increased cross-purchase rates can positively influence customer retention [27].

Results indicate significant differences between the two groups. The experimental group presented a 15.72% higher response rate, an $82 higher average purchase amount per customer, and 8.15% higher cross purchase rate than the control group. Given that a 10% discount and other purchase inducement campaigns are usually judged to be successful even with a 2% cross-purchase increase, an 8.15% increase is significant.

5.1.2 H2: Effect on customer retention/satisfaction. Regarding H2, we measured the customer retention rate for two months after the test and compared the results between the experimental and control groups. Our assumption here was that the recommendation of various products would have a significant effect after purchase through the customer's cognitive satisfaction. Previously, Xiao et al. [33] showed that additional purchase information can be used as a important criterion to measure customer satisfaction by tracking their post-purchase behaviors. Wang et al. [32] also suggested that continuous customer retention implies the cognitive behavior of post-purchase sand the achievement of customer satisfaction. As such, we verified the numerical results for customer purchase satisfaction after recommendation.

The reason why we considered the customer repurchase information for the two-month period is that the total customer purchase cycle of the second purchase is 63 days on average. From a marketing perspective, if the customer makes the second purchase within the average purchase cycle, it can be determined that the customer is satisfied with the service or product and the customer's LifeTime Value (LTV) is high (Table 5).

Results indicate that there was a difference in the repurchase rate and average purchase amount per customer between the experimental and the control groups. We found that 39.93% of the customers in the experimental group made a second purchase while 24.43% of the customers in the control group made a repurchase. On average, the ratio of secondary purchases within 60 days after the first purchase is 23.8%. This indicates that the recommendation of various products has a positive effect on the customer's retention rate as well as contributes to the customer's improved purchasing power (customer transaction).

5.2 User Test #2 (Recommendation Diversity)

5.2.1 H3: Diversity effect on customer response. We examined H3 from the second user test. We used the same metrics used in the

Index		Reco. provide	Reco. not provide
Purchase customer		2,407	1,621
[a] Retention	Repurchase customer	961	396
	[b] Repurchase rate***	39.93%	24.43%
	[c] Customer transaction	$496	$382

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[a] Retention : repurchase for 60 days after recommendation

[b] Repurchase rate = number of repurchase customers / number of purchase customers through recommendation

[c] Customer transaction = repurchase amount / number of repurchase customers

Table 5: Results of the first user test for H2.

first test (i.e., response rate, per-customer purchasing unit cost, and purchase of new items) and compared the results between two customer groups. Unlike the first test, we targeted customers who tend to purchase items from the same category. These customers are likely to have high psychological resistance to other types of items, making cross-purchase inducement highly difficult. Thus, we wanted to see whether the variety of recommendations affects customers who are less open to new products.

Both customer groups received a promotion link that showed a list of recommendation items. An item list for the experimental group (5,000 customers) was based on our hybrid filtering model, whereas that for the control group (5,000 customers) was based on the apriori algorithm. We selected the apriori algorithm (17.51%), because it yielded the highest preference ratio (click through ratio), compared to other methods (demographic filtering: 15.26%, IBCF: 6.94%, and UBCF: 5.26%).

Table 6 summarizes the results, indicating that the experimental group showed 3.14% higher response, $13 higher per-customer purchasing unit cost, and 1.14% higher purchase on brand products that they had not experienced before, compared to the control group. By recommending various products to customers who mostly make purchases from the same category, we were able to obtain a higher purchasing response rate, compared to previous campaigns. According to the company, it is worth noting that the 1.14% difference in cross-purchase rates is promising, due to the nature of the target customers in the second test who are less open to new products.

Index	Hybrid filtering	Apriori (baseline)
Target customer	5,000	5,000
Purchase customer	614	457
Purchase amount	$205,017	$146,588
Customer transaction	$334	$321
Purchase response rate***	12.28%	9.14%
Cross purchase rate*	16.63%	15.49%

Table 6: Results of the second user test for H3.

5.2.2 H4: Diversity effect on customer retention/satisfaction. In the second test, we measured the customer retention rate for two months after the test and compared the results of the experimental group (with HF recommendation) and the control group (with apriori recommendation). Table 7 summarizes the results. The repurchase rate of the experimental group was 3.86% and $86 higher in guest prices than that of the control group. Although this had less of an effect compared to the first test (because the target customer

Index		Hybrid filtering	Apriori (baseline)
Purchase customer		614	457
Retention	Repurchase customer	107	62
	Repurchase rate***	17.43%	13.57%
	Customer transaction	$407	$321

Table 7: Results of the second user test for H4.

has experienced only a single item), from the marketing perspective in fashion industry, it is still considered a good strategy for new customer care as the result shows meaningful differences.

6 DISCUSSION

6.1 Study Results and Implications

Our FGI with the fashion professionals indicated that current recommendation systems have mainly focused on the accuracy of information filtering, thus resulting in recommending similar items. We strived to reflect the characteristics of the fashion industry and its customers to compensate for these limitations. We introduced a novel methodology of drawing and exposing recommended products based on collaboration with fashion professionals and demonstrated the effectiveness of our approach through large-scale user tests. The summary and contributions of our work are as follows.

First, we classified the unique characteristics of customers in the fashion industry into four types (i.e., gift type, coordinator type, carry-over type, and trend-setter type), and mapped the corresponding algorithm for each type (i.e., demographic filtering, apriori, IBCF, and UBCF). Second, by independently merging the results from the four recommendations through a hybrid filtering method, we increased the diversity of the recommended products while maintaining the strengths of each algorithm. Third, using our proposed method, we increased the accuracy of the recommendation. This reflects the characteristics of customers and products so that products are recommended according to the customer's preference. Based on fashion professionals' knowledge and the Purchase Funnel theory, we derived 13 salient features pertaining to the customer's purchasing behaviors. We built a highly accurate recommendation model that well predicts the probability of product preference for each customer.

Fourth, we empirically verified the actual effectiveness of the recommendation system through two large-scale user tests with 20,000 customers. Especially, we proceeded our research with metrics used in actual fashion domains, not the ones used in previous recommendation research; thus, presenting effectiveness of recommendation model more realistically. Results suggested that companies can achieve higher sales rates, cross-selling rates and per-customer purchase prices. From a customer's perspective, there was the benefit of gaining the cognitive satisfaction to new product experiences and of purchasing them easily. We also found that customer satisfaction led to second purchases after the recommendation.

Fifth, our study provided insights into how a recommendation system can be applied to customer management in fashion company. If a particular customer's preference can be recognized in advance, the customer's desire to make a purchase can be enhanced through specifically *tailored* marketing activities. Customers who purchase only similar products every season can experience new products through a recommendation system offering a variety of products. This have an impact on reducing customer churn and diversifying the category of products a company can sell to customers, thereby increasing the customer transaction. Salesmen from offline stores can use this recommendation system to propose customized products for specific customers, and this in turn will make customers more liable to purchase products. Online malls can increase their conversion rate by exposing custom-tailored recommended products rather than best-selling products-oriented ones.

6.2 Limitations and Future work

We observe limitations in our study, which are expected to be addressed in our future work. The customer's sensitivity and subjective experience, which are important for customer experience management (CEM), have not been reflected in our recommendation system. To realize customer satisfaction, many companies have focused on delivering positive experiences before and after a purchase and the use of a product, and have used qualitative methods such as interviews to read the customer's emotions about the purchase. However, it is now possible to infer customer's sensibility by analyzing customer's left-over product reviews and text on social networking sites. ASOS, a global fashion e-commerce company, considered the accuracy of its recommendations by drawing new features from unstructured data such as text and images related to products [11]. In the future, more meaningful results can be derived if the assessment and satisfaction of customers' products are reflected in the recommendation system. For example, a study of the correlation between unstructured data (e.g., purchasing product reviews in text, images) and customers' purchases, repurchase, and satisfaction can potentially give an indication for understanding customers more deeply.

ACKNOWLEDGEMENTS

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the program (IITP-2020-2018-0-01431) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation) and the program (2020R1F1A1076129) by the NRF (National Research Foundation).

REFERENCES

[1] Rakesh Agrawal, Ramakrishnan Srikant, et al. 1994. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, Vol. 1215. 487–499.

[2] Shyong K Lam Al Mamunur Rashid, George Karypis, and John Riedl. 2006. ClustKNN: a highly scalable hybrid model-& memory-based CF algorithm. *Proceeding of webKDD* 2006 (2006).

[3] Amir Albadvi and Mohammad Shahbazi. 2009. A hybrid recommendation technique based on product category attributes. *Expert Systems with Applications* 36, 9 (2009), 11480–11488.

[4] Sarabjot Anand, AR Patrick, John Hughes, and David Bell. 1998. A data mining methodology for cross-sales. *Knowledge-based systems* 10, 7 (1998), 449–461.

[5] Mahsa Badami, Faezeh Tafazzoli, and Olfa Nasraoui. 2018. A case study for intelligent event recommendation. *International Journal of Data Science and Analytics* 5, 4 (2018), 249–268.

[6] Ana Belén Barragáns-Martínez, Enrique Costa-Montenegro, Juan C Burguillo, Marta Rey-López, Fernando A Mikic-Fonte, and Ana Peleteiro. 2010. A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition. *Information Sciences* 180, 22 (2010), 4290–4311.

[7] Keith Bradley and Barry Smyth. 2001. Improving recommendation diversity. In *Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science, Maynooth, Ireland*. Citeseer, 85–94.

[8] Robin Burke. 2002. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction* 12, 4 (2002), 331–370.

[9] Robin Burke. 2007. Hybrid web recommender systems. In *The adaptive web*. Springer, 377–408.

[10] Pedro G Campos, Fernando Díez, and Iván Cantador. 2014. Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction* 24, 1-2 (2014), 67–119.

[11] Ângelo Cardoso, Fabio Daolio, and Saúl Vargas. 2018. Product characterisation towards personalisation: learning attributes from unstructured data to recommend fashion products. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 80–89.

[12] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 785–794.

[13] Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 271–280.

[14] Mukund Deshpande and George Karypis. 2004. Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems* 22, 1 (2004), 143–177.

[15] Daniel Fleder and Kartik Hosanagar. 2009. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management science* 55, 5 (2009), 697–712.

[16] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. 2010. Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the conference on Recommender systems*. ACM, 257–260.

[17] Georg Groh and Christian Ehmig. 2007. Recommendations in taste related domains: collaborative filtering vs. social filtering. In *Proceedings of the 2007 international ACM conference on Supporting group work*. ACM, 127–136.

[18] Stephen J Hoch and George F Loewenstein. 1991. Time-inconsistent preferences and consumer self-control. *Journal of consumer research* 17, 4 (1991), 492–507.

[19] Yang Hu, Xi Yi, and Larry S Davis. 2015. Collaborative fashion recommendation: A functional tensor factorization approach. In *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 129–138.

[20] Zan Huang, Hsinchun Chen, and Daniel Zeng. 2004. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM Transactions on Information Systems (TOIS)* 22, 1 (2004), 116–142.

[21] Sergey Ioffe. 2010. Improved consistent sampling, weighted minhash and l1 sketching. In *2010 IEEE International Conference on Data Mining*. IEEE, 246–255.

[22] Paul Jaccard. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaudoise Sci Nat* 37 (1901), 547–579.

[23] Nick Landia. 2017. Building recommender systems for fashion: industry talk abstract. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, 343–343.

[24] E St E Lewis. 1908. *Financial advertising, for commercial and savings banks, trust, title insurance, and safe deposit companies, investment houses*.

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

[26] Priyanka Rawal. 2013. AIDA Marketing Communication Model: Stimulating a purchase decision in the minds of the consumers through a linear progression of steps. *International Journal of Multidisciplinary research in social & management sciences* 1, 1 (2013), 37–44.

[27] María Teresa Salazar. 2010. Modelling cross-sales to promote customer retention in the financial services industry: the'who-what-when framework'. Two case studies. (2010).

[28] Oren Sar Shalom, Noam Koenigstein, Ulrich Paquet, and Hastagiri P Vanchinathan. 2016. Beyond collaborative filtering: The list recommendation problem. In *Proceedings of the 25th international conference on world wide web*. International World Wide Web Conferences Steering Committee, 63–72.

[29] Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl, et al. 2000. Analysis of recommendation algorithms for e-commerce. In *EC*. 158–167.

[30] Devashish Shankar, Sujay Narumanchi, HA Ananya, Pramod Kompalli, and Krishnendu Chaudhury. 2017. Deep learning based large scale visual recommendation and search for e-commerce. *arXiv preprint arXiv:1703.02344* (2017).

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

[32] Weiquan Wang and Izak Benbasat. 2007. Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems* 23, 4 (2007), 217–246.

[33] Bo Xiao and Izak Benbasat. 2007. E-commerce product recommendation agents: use, characteristics, and impact. *MIS quarterly* 31, 1 (2007), 137–209.

[34] Jianyu Zhang and Françoise Fogelman-Soulié. 2018. KKbox's music recommendation challenge solution with feature engineering. In *11th ACM International Conference on Web Search and Data Mining WSDM*.

[35] Jiangang Zhu, Beijun Shen, and Fanghuai Hu. 2015. A learning to rank framework for developer recommendation in software crowdsourcing. In *2015 Asia-Pacific Software Engineering Conference (APSEC)*. IEEE, 285–292.