

Methods Comparison: Enhancing Diversity for Personalized Recommendation with Practical E-Commerce Data

Juryon Paik*

*Professor, Dept. of Digital Information and Statistics, Pyeongtaek University, Pyeongtaek-si, Korea

[Abstract]

A recommender system covers users, searches the items or services which users will like, and let users purchase them. Because recommendations from a recommender system are predictions of users' preferences for the items which they do not purchase yet, it is rarely possible to be drawn a perfect answer. An evaluation has been conducted to determine whether a prediction is right or not. However, it can be lower user's satisfaction if a recommender system focuses on only the preferences, that is caused by a 'filter bubble effect'. The filter bubble effect is an algorithmic bias that skews or limits the information an individual user sees on the recommended list. It is the reason why multiple metrics are required to evaluate recommender systems, and a diversity metrics is mainly used for it. In this paper, we compare three different methods for enhancing diversity for personalized recommendation - bin packing, weighted random choice, greedy re-ranking - with a practical e-commerce data acquired from a fashion shopping mall. Besides, we present the difference between experimental results and F1 scores.

▶ **Key words:** Recommender system, Evaluation metrics, Relevance, Diversity, Filter bubble, Greedy re-ranking

[요 약]

추천시스템은 소비자를 대신하여 소비자가 선호할 만한 아이템이나 서비스를 검색하여 구매할 수 있도록 한다. 추천시스템의 추천은 사용자들이 경험하지 않은 아이템들에 대한 선호 예측이기 때문에 완전하게 맞는 답이 도출되는 것은 불가능하다. 따라서 예측에 대한 평가가 수행되어야만 비로소 추천시스템이 정확한지 아닌지를 판단할 수 있다. 그러나 사용자 선호에 대한 예측 정확성만을 높이는 추천은 오히려 사용자의 만족도를 하락시킬 수 있는데 이는 사용자의 취향만을 반영한 편중된 결과로 사용자는 다양한 아이템들로 구성된 추천 결과를 받을 수 없는 필터버블 현상이 야기되기 때문이다. 품질 측정 지표의 다각화가 필요한 이유이고 대표적으로 다양성 지표가 사용된다. 본 논문에서는 추천 결과의 다양성 증대를 위한 3가지 기본 접근방법인 bin packing, weighted random choice, greedy re-ranking을 실제 e-커머스 데이터인 패션 쇼핑몰 데이터에 적용하여 도출된 결과와 F1 score에 기반을 둔 차이를 분석한다.

▶ **주제어:** 추천시스템, 평가지표, 연관성, 다양성, 필터버블, 탐욕적재순위

-
- First Author: Juryon Paik, Corresponding Author: Juryon Paik
 - *Juryon Paik (jrpaik@ptu.ac.kr), Dept. of Digital Information and Statistics, Pyeongtaek University
 - Received: 2022. 08. 10, Revised: 2022. 09. 14, Accepted: 2022. 09. 14.

I. Introduction

4차 산업혁명, 빅데이터 그리고 인공지능으로 대표되는 스마트 ICT 사회에서 '선택의 여지가 많아질수록 행복해지기보다는 포기한 선택에 대한 후회가 커진다'라는 심리학자 배리 슈워츠[1]의 말처럼 정보과다는 사회 구성원 모두가 직면한 가장 큰 문제 중 하나이다. 공급과 소비의 관점으로 접근하게 되면 문제는 더욱 구체화된다. 소비자 입장에서 어떤 상품 또는 서비스를 구매 및 이용하고자 할 때 정보가 부족한 경우보다는 너무 많아서 어떤 기준으로 판단하고 무엇을 선택해야 할지 판단할 수 없는 문제가 발생한다. 비즈니스로 대변되는 제3자 입장에서도 소비자와 시장에 대한 정보가 웹사이트, SNS 등 다양한 경로로부터 과도하게 발생될 뿐만 아니라 그 많은 데이터를 어떤 방법으로 분석하고 어떤 정보가 가치 있는지를 판단하기 어려워지는 문제가 발생한다.

정보 필터링 (Information Filtering: IF) 기술의 일종인 추천시스템 (Recommender System: RS)은 특정 시점에 특정 사용자가 관심을 가질만한 아이템들로 이루어진 리스트를 찾아주는 시스템으로, 과거 기업 내에서 주로 사용되었던 전문가시스템 (Expert System) 또는 의사결정시스템 (Decision System)이 정보화 시대의 개인화 IT 서비스를 위해 발전된 시스템으로 볼 수 있다. 전자상거래의 대중화, 그로부터 생산되는 다양한 종류의 데이터들, 이를 적극적으로 활용한 Amazon의 독자적인 추천시스템 A9을 기점으로 빅데이터, 머신러닝 더 나아가 딥러닝과 결부된 추천시스템은 IT B2C (Business-to-Customer) 생태계의 필수로 자리 잡았다고 해도 과언이 아니다. McKinsey와 Tech Emergency의 통계에 의하면, Amazon 매출의 35%, BestBuy의 23.7%가 추천시스템에 의해 이뤄지며, Netflix는 대여되는 영화의 최대 75%까지 추천을 경유한다고 한다. YouTube 또한 60%까지 추천시스템을 통해 기업 매출이 발생한다고 보고 있다[2-4]. 정보 검색 측면에서도 Google 뉴스의 38% 이상의 조회 수가 추천 결과에서 발생한다고 한다. 국내 경우 역시 다르지 않다. 국내 최대 포털인 네이버는 딥러닝 기반 개인추천 기능을 적용하여 전체 이용자의 80%를 AI 솔루션 사용자로, Netflix처럼 온라인동영상 서비스를 제공하는 SK브로드밴드 또한 인공지능 기술을 적용하여 고도화된 개인 맞춤형 추천서비스를 제공한다.

추천시스템은 기본적으로 소비자의 취향 파악을 수행한다. 그리고 취향에 따라 제품 및 서비스를 추천하는데 중요한 것은 '추천된 아이템 및 서비스에 대해 소비자가 얼

마나 만족했는가'이다. 얼마나 정확하게 선호도를 예측했는가에 따라 소비자는 충성 고객이 될 수도 이탈 고객이 될 수도 있으며 이는 기업 이익과 직결되기 때문이다. 그러나 사용자가 선호할 만한 아이템만 제한적으로 제공할 때 추천이 한 방향으로 치우치는 필터 버블(Filter Bubble, 정보 여과 현상)을 초래한다. 사용자의 개인 정보 및 온라인 행동 패턴에 기반하여 선별적인 개인화된 추천 결과를 제공하기 때문에 사용자는 자신의 취향에 갇히게 되어 편중된 결과만 받게 되는 것이다[5]. 편향된 추천은 추천시스템에 대한 사용자의 만족도를 하락시키는 요인이 되기 때문에 다른 평가지표를 반영할 필요가 있다. 대표적으로 다양성 지표가 적용된다.

본 논문에서는 다양성 지표 개념을 정리하고 추천 결과의 다양성 증대를 위한 3가지의 기본 접근방식에 실무 데이터를 적용하여 구현 후 실행 결과의 차이를 비교한다. 본 논문의 구성은 다음과 같다. 2장에서는 추천시스템의 평가지표 중 연관성과 다양성에 관해서 설명하고 서로 다른 다양성 증대 방식의 특징을 정리한다. 3장에서는 실제 e-커머스 데이터를 사용한 구현 결과들과 각 결과의 차이점을 보인 후 정량지표 값인 F1 score와 도출된 결과와의 차이점을 기술한다. 마지막으로 다양성 지표에 시간 가중치 적용에 대한 필요성과 관련 연구 방향을 제시한다.

II. Preliminaries

1. Recommender System Evaluation Metrics

추천시스템이나 추천모델의 성능은 크게 두 가지 관점에서 평가가 이루어진다[6-8]. 비즈니스 또는 서비스 관점과 추천시스템의 품질 관점이다. 전자는 기업과 사용자 측면으로 나뉘는데, 기업 측 지표로는 해당 추천시스템이나 추천모델 적용으로 PV(Page View) 수치가 얼마나 증가했는지를 고려하며 사용자 측 지표로는 CTR(Click Through Rate) 수치가 새로운 추천 아이템으로 얼마나 상승했는지를 고려한다. 이때, PV는 사용자가 추천시스템 적용 사이트에서 몇 개의 페이지를 방문했는지에 대한 정량적 수치로 만약 사용자 10명이 해당 사이트를 방문하여 각 10개씩의 페이지를 방문했다면 PV 값은 100이 된다. CTR 수치는 간단하게 기술하면 노출된 광고에 대해서 사용자가 얼마나 클릭 했는가를 의미하는 정량적 값으로 일반적인 공식은 클릭수를 노출수로 나눈 백분율 값이다.

비즈니스/서비스 관점에서의 두 평가지표 모두 추천 시스템이 제공하는 추천 내용, 즉 추천 품질과 밀접한 관련이

있다. 사용자의 관심 범위 내의 상품이나 서비스 관련 페이지 및 광고를 해당 추천시스템이 제공한다면 페이지 방문 또는 클릭 수는 당연히 증가할 것이다. 이를 반영한 관점이 두 번째 관점인 추천시스템의 품질 관점에서의 평가 지표로 연관성(Relevance), 다양성(Diversity), 새로움(Novelty), 불예측성(Serendipity), 커버리지(Coverage) 등이 다각화된 품질 측정 특성들이다[9, 10]. 본 논문에서는 품질평가 지표 중 연관성과 다양성에 관해서 기술한다.

1.1 Relevance (Accuracy)

연관성은 추천된 아이템이 사용자의 선호 또는 취향과 얼마나 관련이 있는가를 측정하는 지표로 정확도(Accuracy)라고도 표현한다. 추천시스템 중에서 우수한 성능을 나타내는 협업필터링 기법에서 사용자의 만족도를 정량 표현하기 위해 사용된다. 사용자와 유사한 아이템을 구매한 다른 사용자를 찾거나 사용자가 선호하는 아이템과 유사한 아이템을 찾아 추천하는 기법으로 얼마나 정확하게 유사 사용자 또는 유사 아이템을 찾았는지가 평가의 주요 대상이 된다[8-12]. 사용자 전체를 U , 아이템 전체를 I , R 은 추천 아이템 집합이라 할 때, 각 사용자 u 에 대해서 목적함수의 최댓값을 구하는 아이템 i 를 선택하는 것으로 표현할 수 있다. 아래 식(1)에서 목적함수 f 값이 연관성 자체라면, $f(u, i) = \text{rel}(u, i)$, 정확도라고 한다.

$$u \in U, \hat{R}_u = \underset{i \in R}{\text{argmax}} \sum f(u, i) \dots (1)$$

그러나 연관성은 사용자가 좋아할 만한 아이템만 제한적으로 제공하면서 추천이 한 방향으로 치우치는 필터 버블(Filter Bubble) 현상을 야기하고 편향된 추천은 만족도를 하락시키는 요인이 되기 때문에 다른 평가지표를 반영하여 결과를 보정한다.

1.2 Diversity

다양성은 Top-k 추천 리스트에서 얼마나 다양한 아이템들이 추천되었는지를 측정하는 지표로 장르/범주나 태그/키워드 벡터들을 쌍으로 하여 유사도를 비교한다. Intra-list 다양성과 Inter-list 다양성으로 구분하여 정의한다 [9, 10, 13]. 전자는 추천되는 k개의 아이템이 비슷한 종류로만 구성되는 것을 지양하는 다양성이며, 후자는 동일 사용자가 같은 추천 요청을 여러 번 실행했을 때 똑같은 추천목록을 제시하는 것을 지양하는 다양성이다.

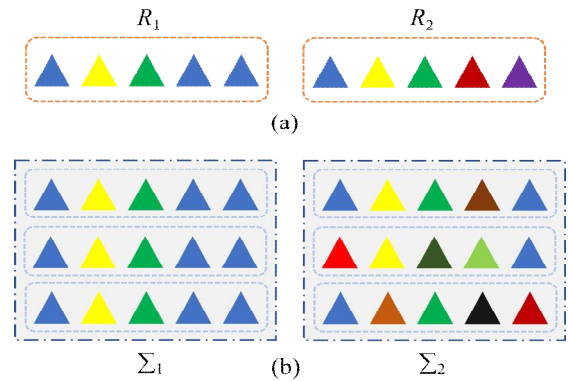


Fig. 1. Intra-list Diversity vs. Inter-list Diversity

Fig. 1은 번개장터 CTO 이동주 강의자료[13]의 내용을 재구성한 것으로 Fig. 1(a)는 Intra-list 다양성을 Fig. 1(b)는 Inter-list 다양성 예로 1(a)의 경우 R_2 의 다양성 값이 R_1 보다 높으며, 1(b)의 경우 Σ_2 의 다양성 값이 Σ_1 보다 높다. 추천목록 R 을 구성하는 아이템들의 유사도 값으로 표현하면 다음과 같다.

$$\text{sim}(R) = \frac{\sum_{i \in R} \sum_{j \in R \setminus i} \text{sim}(i, j)}{|R|(|R| - 1)} \dots (2)$$

$$\text{sim}(R_1, R_2) = \frac{\sum_{i \in R_1} \sum_{j \in R_2} \text{sim}(i, j)}{|R_1||R_2|} \dots (3)$$

식(2)는 Intra-list 아이템 간의 쌍에 대한 유사도 값을 합산하며 식(3)은 Inter-list 간 아이템 쌍들에 대한 유사도 값을 합산한다. 유사도는 0부터 1 사이의 값으로 정량화되기에 Intra-list 다양성과 Inter-list 다양성 모두 1에서 해당 유사도 합, $\text{sim}(R)$ 또는 $\text{sim}(R_1, R_2)$, 을 뺀 값이 추천 결과의 다양성 지표값이 된다. Intra-list 다양성이나 Inter-list 다양성 모두 연관성 지표로 인해 발생하는 사용자 선호도 편중 추천 결과를 취향이 다른 사용자와의 유사도를 높게 하여 추천 범위를 다양하게 확장한다.

2. Ways of Diversity Increasement

품질 관점 추천시스템의 평가지표 중 연관성 즉 정확도 지표와 같이 적용되어 사용자의 만족도를 높이면서도 다른 지표에 비해 상대적으로 정량화하기에 수월한 것은 다양성 지표이다. 따라서 다양성을 증대하기 위한 연구가 진행되었고 가장 많이 적용되는 연구 방법은 재순위(re-ranking) 방식이다[10, 15]. 본 장에서는 다양성 증대를 위한 가장 단순한 방식인 bin 채우기 방법부터 임의 선

택에 가중치를 부여한 방법 (Weighted Random Choice), 그리고 탐욕적 재순위 (Greedy Re-ranking) 방법에 대하여 참고자료 [10, 11, 14, 16]에 기반하여 정리한다.

2.1 Bin Packing

Intra-list 다양성을 향상하기 위한 경험적 방법 중 하나로 추천목록 R 을 구성하는 아이템들이 속한 클래스 분포가 사용자의 선호 클래스 분포와 유사하도록 추천한다. 사용자의 선호도만 고려하는 것이 아니라 클래스 선호 분포를 반영하는데 클래스별로 선호 분포만큼 정해진 수의 아이템이 추천되도록 조정한다.

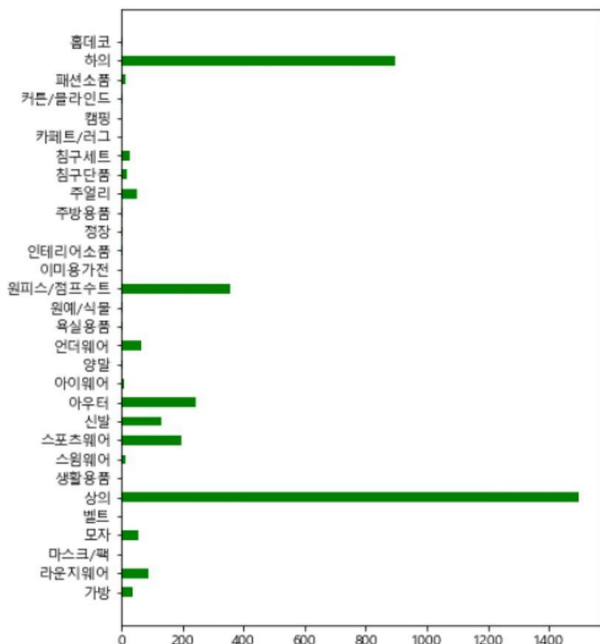


Fig. 2. Some User's Clicked Items' Category Distribution

Fig. 2는 어떤 사용자가 2021년 6월 3일부터 2021년 8월 4일까지 패션 쇼핑몰에서 판매하는 상품들 중 클릭한 상품들이 속한 중분류를 분석한 그래프로 클릭한 상품들의 대부분이 상/하의에 편중된 것을 알 수 있다. 연관성 즉 사용자 선호도 반영 정확도만 고려하게 될 경우 상/하의 상품들의 순위가 상위에 위치할 확률이 높기 때문이다. 이를 사용자가 클릭한 상품들이 속한 중분류 비중에 맞게, 예를 들면 Top-k 개중 상의가 10개이면 아우터는 2개, 주얼리나 모자는 1개 등으로 추천목록을 구성하는 것이다. 이것이 bin 채우기 방법으로 카테고리별로 사용자 선호 분포에 따라 bin 크기를 설정한 후 상위 카테고리별로 bin들을 채워가면서 bin이 다 채워지게 되면 그 bin에 속하는 아이템은 추천목록에서 제외하여 편중된 카테고리 추천을 일정 수준으로 방지하는 것이다.

2.2 Weighted Random Choice

동일 사용자에게 여러 번의 동일 추천 결과를 제공할 때 같은 추천목록이 여러 번 나오는 것이 아닌 조금씩 다른 추천목록을 제공하여 단조로움을 탈피하는 방법으로 Inter-list 다양성을 높인다. Top-k개의 추천목록을 결과로 제공한다면 k개보다 훨씬 많은 아이템으로 하나의 후보 추천목록을 구성한 후 해당 후보 추천목록에서 k개의 아이템을 임의 선택하여 제공한다. 이때 후보 추천목록을 구성하는 아이템들은 사용자 선호도에 따른 가중치(score)가 반영되는데, 가중치가 높으면 임의 선택 확률이 높아지고 가중치가 낮으면 임의 선택 확률이 낮아진다. Fig. 3에 도식화된 방식처럼, 가중값에 따라 영역이 다른 범위를 갖는 회전하는 과녁에 화살을 쏘아서 꽃힌 아이템을 선택하는 것과 같다고 할 수 있다.

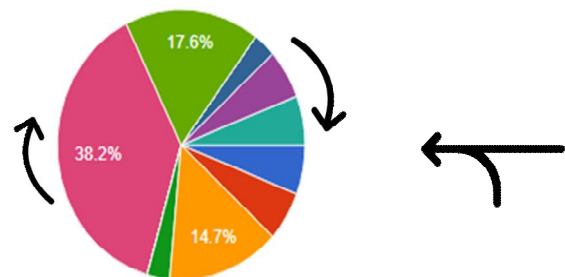


Fig. 3. The Way of Weighted Random Choice

가중치와 임의성을 사용하여 동일 사용자에게 같은 추천 질의에 대한 똑같은 추천목록 결과가 제공되는 경우의 수를 낮추는 방법이 가중치반영임의선택(weighted random choice)이다.

2.3 Greedy Re-ranking

추천시스템의 다양성을 향상하기 위해 가장 많이 사용되는 방법은 재순위(re-ranking) 방식이다[15]. 다양성 증대를 위해 크게 두 가지 접근이 존재하는데 목적함수 자체를 정의하는 것과 도출된 결과들의 순위를 다시 결정하는 재순위 방식이다. 처음부터 목적함수 자체를 정의하는 것은 어려움이 있기에 전자는 거의 이뤄지지 않는다. 그러나 재순위 방식은 기존 랭킹 방식을 그대로 활용하여 후보 추천목록을 선정하고, 후처리를 통해서 목적함수가 반영된 최종 추천목록을 선정하기 때문에 좀 더 합리적인 접근이다. 아이템에 대한 사용자의 연관성과 다양성을 조화시키기 위해 탐욕적 재순위 방식이 다양성 초기 연구부터 사용되었으며 핵심은 선호와 비선호 관계를 선형적 결합으로 구성하여 다양성을 증가시킨다.

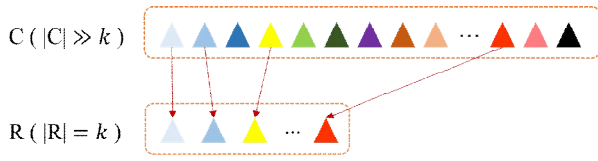


Fig. 4. Recommendation List by Greedy Re-ranking

$$f(u, i) = \alpha \cdot rel(u, i) + (1 - \alpha) \cdot g(u, i) \\ = \alpha \cdot rel(u, i) + (1 - \alpha) \cdot \frac{1}{|R|} \sum_{j \in R} (1 - sim(i, j)) \dots (4)$$

Fig. 4를 보면 리스트 C는 사용자 선호도와외의 연관성을 기준으로 k 개보다 훨씬 많은 아이템으로 구성된 후보 추천목록 리스트이다. C로부터 추천목록 R을 구성하는데 식(4)의 목적함수 값이 최대가 되도록 하는 아이템을 선정한다. 식(4)는 식(1)의 연관성으로부터 도출된 것으로 목적함수에 연관성 외에 다른 요소 $g(u, i)$ 즉 다른 평가지표 값을 포함하여 고려 가능하다는 것을 나타낸다. α 는 각 평가지표에 대한 가중값이다. $g(u, i)$ 대신 다양성 지표값을 고려하여 최종 추천목록을 생성하는 방법이 탐욕적 재순위 방식이다.

III. Methods Comparison

3.1 Data Configuration

본 장에서는 2장에서 기술한 다양성 향상 방법을 실제 e-commerce 데이터에 적용하여 추천목록 결과를 비교하고자 한다. 사용된 데이터는 패스트캠퍼스 교육[14]에서 사용한 어느 유명 패션 커머스 데이터로 익명화 등 전처리가 수행되었으며 2021.06.03.~2021.08.04. 기간의 자료이다. Table 1은 데이터 종류와 수를 Table 2는 각 개체의

주요 특징(feature)들을 보인다. 패션 커머스 아이템들은 모두 대분류(category1)와 소분류(category2)로 구분되며, 본 실험은 다양성을 반영할 클래스로 Fig. 2에서 제시한 소분류를 사용한다. 유효대분류 12개로 나뉘며 유효소분류는 90개이다. Fig. 5는 쇼핑물 사용자들의 가장 많은 관심을 받은 최상위 25개의 소분류 카테고리에 대한 분포 그래프로, ‘상의’ 카테고리로 사용자들의 관심이 집중된 것을 알 수 있다.

Table 1. Fashion Commerce Data’s Objects

Object	Number of Features	Total
User	3	254,958
Item	12	283,326
Event	13	5,880,407

Table 2. Main Features

Object	Main Features
User	id, birth date, gender
Item	id, name, category, price, brand
Event	session id, time stamp, event name, device, region

Event 데이터는 한 명의 사용자가 여러 개의 세션을 연결하여 상품 검색부터 구매까지 관련된 데이터이다. 필드 event_name은 해당 사용자가 세션별로 상품을 클릭했는지 (click_item), ‘좋아요’를 표시했는지 (like_item), 장바구니에 담았는지 (add_to_cart), 구매했는지 (purchase_success) 등 개인의 선호 및 구매 여부 등을 파악할 수 있는 특징을 갖는다. Fig. 6은 각 이벤트에 대한 사용자 세션 수를 카운트한 것으로, 가장 많은 세션에서 발생한 이벤트는 click_item으로 전체 세션 5,880,407개의 82% 이상을 차지하는 4,845,691개의 세션에서 발생했다.

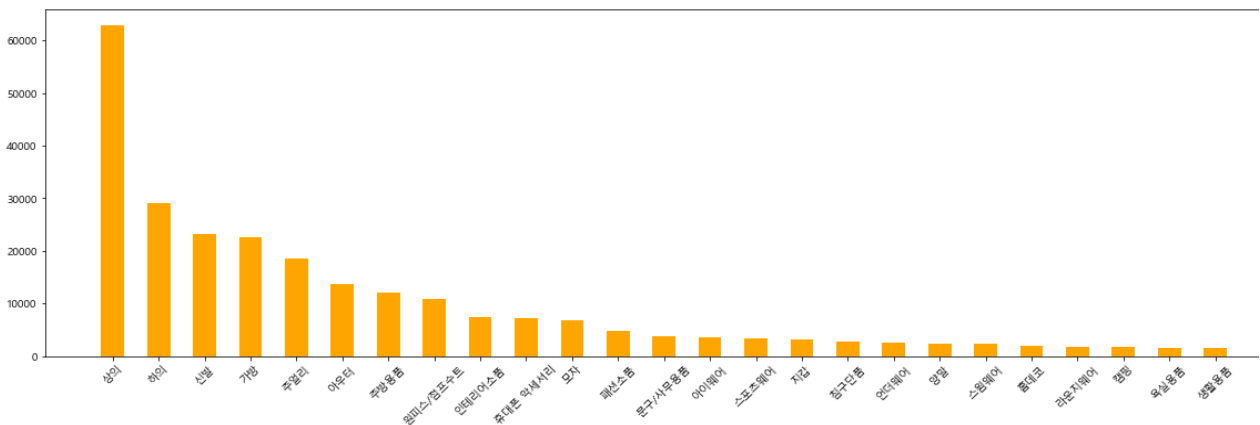


Fig. 5. Item Numbers on Top 25 of Category2

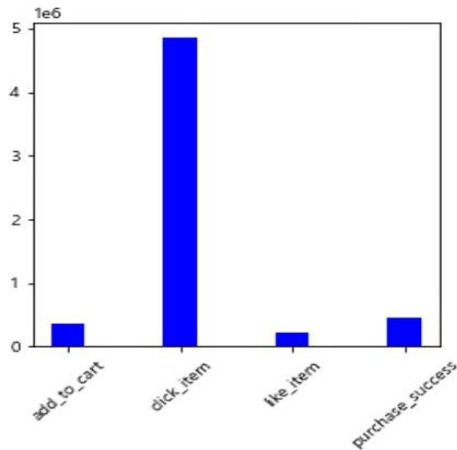


Fig. 6. Session Count of Each Event Type

3.2 Experimental Results

우선 아이템 기반 개인화 추천 방식을 활용하여 세 종류의 다양성 증대 방식을 구현 후 결과로 얻어지는 추천목록의 차이점을 분석한다. 그 후 F1 score를 사용하여, 다양성을 반영한 추천 결과에 포함된 아이템들이 실제로 사용자의 관심을 얼마나 받았는지를 평가한다. 사용자가 구매한 아이템 유사도는 사용자 기준이 아닌 세션 기준의 df-idf 방식으로 계산하였다. 이는 세션을 문서로 간주하고 아이템을 단어로 간주하여 아이템이 세션에 얼마나 많이 출현했는지를 반영한다. 본 실험에서는 세션의 이벤트 종류에는 가중치를 부여했지만 시간 가중치는 고려하지 않고 진행하였다. 추천목록 크기는 20개로 하고 후보 추천목록 크기는 100개로 정한다.

Fig. 7은 단순 개인화 추천목록으로, 어느 사용자 p의 7월 한 달 동안의 로그 기록을 기반으로 한다. 사용자 p가 클릭한 아이템들과 같은 세션에서 발견되는 아이템들의 유사도를 계산 후, 아이템별 상위 20개의 유사 아이템들을 선정하여 유사도 순으로 정렬한 20개의 추천 아이템들이다. 20중 12개가 소분류 ‘상의’에 편중된 것을 알 수 있다. Intra-list 및 Inter-list 다양성 반영에 따라, Fig. 7 결과와 비교하여 차이점이 존재하는지 알아본다.

item_name	category2_name
Draped Midi Dress_BLUE	원피스/점프수트
프렌치 프릴 뽀뽀 셔츠 시어셔커 순면 여름 차림이불(SS/Q&K)-3color	침구단품
시원한 슬리드 시어셔커 여름 빼개기바(50x70)-8color	침구단품
BTS Run TinyTAN Tumbler Jin	주방용품
리틀샘머 풍기인건 키즈 험대패드(SS/Q)	침구단품
CLWP9100 프레쉬 원 마일 레깅스 네이비	스포츠웨어
겔단 2000 울트라코튼 USA핏	상의
겔단 2000 울트라코튼 USA핏 사파이어	상의
핑크 시리즈 쿠션 커버 (2 colors)	홈데코
Calm Half Shirts Sand Shell	상의
화이트 오버핏 빅포켓 튜닉 셔츠 TNSH1E711WT	상의
퍼프 라운드 블라우스	상의
히든 포켓 오버핏 반팔 셔츠 크림 옐로우	상의
[더블유비프로젝트][세트] 미니 스웨트 세트 (티셔츠MJST7497 + 숏팬츠MJSP7498)	상의
Oversize Pin Tuck Shirts [Sherbet]	상의
[더블유비프로젝트] 시티샘머 반팔셔츠 아이보리 KMSS7493	상의
A RIBBON BROOCH PLEATS DR_IVORY	원피스/점프수트
[재밋] 슬린 오버핏 반팔셔츠 화이트 YHSS2351	상의
[타이퍼블유엔] 샘머코튼 반팔셔츠 화이트 SHSS3366	상의
[더블유비프로젝트] 시티샘머 반팔셔츠 라이트카키 KMSS7493	상의

Fig. 7. Top-20 Items from Simple Personalized Recommendation

사용자가 아이템을 구매했는지 아니면 단순히 클릭만 했는지 등 세션에서 발생한 이벤트에 따라 중요도에 차이를 두기 위해, Table 3처럼 가중치를 반영한다. 해당 가중치는 경험적 방법으로 정해지기 때문에, 본 실험에서 구매는 최고 가중치인 4로 정한 후 다른 이벤트들은 순차적으로 1씩 작아지는 값을 부여하였다.

Table 3. Weight Value of Each Event Type

Event	Weight
purchase_success	4
add_to_cart	3
like_item	2
click_item	1

첫 번째 다양성 증대 방식인 ‘빈 채우기’는 사용자의 세션 기록을 바탕으로 카테고리 가중치, 본 실험에서는 소분류 category2, 에 따라 전체 빈 사이즈 (추천목록 크기 20) 대비 카테고리별 빈 개수를 할당한다. Table 4는 사용자 p의 카테고리 선호 가중치에 따라 할당된 빈 개수이다.

Table 4. Bin Size of Each Category

Category	Bin
상 의	10
하 의	5
원피스 / 점프수트	2
신 발	2
그 외	1

100개의 아이템으로 구성된 후보 추천목록에서 유사도가 높은 아이템을 선택할 때, 만약 해당 카테고리의 빈이 다 채워졌다면 선택하지 않고 건너뛰는 빈 채우기 원리에 따라 Fig. 8의 결과가 얻어진다. Fig. 7에서는 보이지 않던 ‘홈웨어’, ‘침구세트’, ‘라운지웨어’ 카테고리에 포함되는 아이템이 추천되어 intra-list 다양성이 증가한 것을 알 수 있다.

item_name	category2_name
Draped Midi Dress_BLUE	원피스/점프수트
프렌치 프릴 뱀부 시어셔커 순면 여름 차림이불(SS/Q&K)-3color	침구단품
BTS Run TinyTAN Tumbler Jin	주방용품
CLWP9100 프레쉬 원 마일 레깅스 네이비	스포츠웨어
길단 2000 울트라코튼 USA핏	상의
길단 2000 울트라코튼 USA핏 사파이어	상의
핑크 시리즈 쿠션 커버 (2 colors)	홈데코
CaIm Half Shirts Sand Shell	상의
화이트 오버핏 빅포켓 튜닉 셔츠 TNSH1E711WT	상의
퍼프 라운드 블라우스	상의
히트 포켓 오버핏 반팔 셔츠_크림 옐로우	상의
[더블유비이프로젝트][세트] 미니 스웨트 세트 (E/셔츠HJST7497 + 숏팬츠HJSP7498)	상의
Oversize Pin Tuck Shirts [Sherbet]	상의
[더블유비이프로젝트] 시티섬머 반팔셔츠 마이보리 KMSS7493	상의
A RIBBON BROOCH PLEATS DR_IVORY	원피스/점프수트
[재킷] 슬린 오버핏 반팔셔츠 화이트 YHSS2351	상의
빈티지플라워 순면60수 차림이불세트(SS/Q)-3color	침구세트
BALI PAJAMA COUPLE SET_Stripe	라운지웨어
[21RESORT]RESORT CROCHE SHORT PANTS	하의
#3311 데님 점프수트 [서현,청하,백지영,달사벳수빈 착용]	하의

Fig. 8. Top-20 Items from Bin Packing

두 번째 다양성 증대 방식은 inter-list 다양성 지표를 향상하는 weighted random choice로, 후보 추천목록 아이템들의 누적 분포를 계산하여 선호도가 높은 아이템들은 선택될 확률을 높게 한다. 그러나 랜덤하게 추천되기 때문에 동한 조건에서의 추천 결과일지라도 매번 다른 추천목록을 제공한다. Fig. 9와 Fig. 10은 동일 추천 요청을 2회 실행한 결과로, 서로 다른 추천목록이 제시됨을 알 수 있다. 즉 inter-list 다양성이 증가한 것을 확인할 수 있다.

item_name	category2_name
Draped Midi Dress_BLUE	원피스/점프수트
시원한 슬리드 시어셔커 여름 베게커버(50x70)-8color	침구단품
BTS Run TinyTAN Tumbler Jin	주방용품
[더블유비이프로젝트] 시티섬머 반팔셔츠 마이보리 KMSS7493	상의
[재킷] 슬린 오버핏 반팔셔츠 화이트 YHSS2351	상의
[더블유비이프로젝트] 시티섬머 반팔셔츠 라이트카키 KMSS7493	상의
[티따블유엔] [세트] 퍼먼트를 반팔셔츠 6중 STSS3281 + 퍼먼트 홈 숏팬츠 6중 STSP3282	상의
BIO 60s OPEN COLLAR SHIRTS (OFF WHITE)	상의
[정해진 착용]카라폴오버세미오버반팔셔츠 MWVS3112	상의
[재킷] 릴렉스 벌룬핏 반팔셔츠 마이보리 OYVS2359	상의
ruffle sleeveless top	상의
[더블유비이프로젝트] 시티섬머 반팔셔츠 네이비 KMSS7493	상의
LINEN ROUND HALF T_MOCHA PINK	상의
TAILORED SHORT JACKET	마우터
STITCH HALF PANTS	하의
WAVE NECK SLIM TOP [BROWN]	상의
HALF SLEEVE COTTON SHIRT_LIGHT GREEN	상의
스트라이프 로고 셔츠	상의
(CTD2) 니니 크레용 룰 스커트_레인보우	하의
RAINBOW HALF KNIT CARDIGAN_HOT PINK	상의

Fig. 9. Top-20 Items 1 from Weighted Random Choice

item_name	category2_name
Draped Midi Dress_BLUE	원피스/점프수트
시원한 슬리드 시어셔커 여름 베게커버(50x70)-8color	침구단품
BTS Run TinyTAN Tumbler Jin	주방용품
길단 2000 울트라코튼 USA핏	상의
핑크 시리즈 쿠션 커버 (2 colors)	홈데코
[더블유비이프로젝트] 시티섬머 반팔셔츠 마이보리 KMSS7493	상의
[재킷] 슬린 오버핏 반팔셔츠 화이트 YHSS2351	상의
[티따블유엔] 썸머코튼 반팔셔츠 화이트 SHSS3366	상의
빈티지플라워 순면60수 차림이불세트(SS/Q)-3color	침구세트
팝 핑크 베게 커버	침구단품
[21RESORT]RESORT CROCHE SHORT PANTS	하의
나이스룸탑 메쉬 조직 카라 니트 폴오버 N212PMK537	상의
ruffle sleeveless top	상의
[더블유비이프로젝트] 시티섬머 반팔셔츠 네이비 KMSS7493	상의
#3311 데님 점프수트 [서현,청하,백지영,달사벳수빈 착용]	하의
marine print jacket_deep orange	마우터
Elliott wind breaker	마우터
HALF SLEEVE COTTON SHIRT_LIGHT GREEN	상의
Tone-on-tone Loose Tee_Lavender	상의
STRIPE OPEN COLLAR KNIT_IVORY	상의

Fig. 10. Top-20 Items 2 from Weighted Random Choice

세 번째는 탐욕적 재순위 방식으로 후보 추천목록에 있는 모든 아이템의 유사도를 계산하여, 아이템 쌍에 대한 유사도 사전 생성 후 유사도 가중치와 다양성 가중치))를 반영하여 후보 추천목록에서 아이템을 선택한다.

item_name	category2_name
Draped Midi Dress_BLUE	원피스/점프수트
프렌치 프릴 뱀부 시어셔커 순면 여름 차림이불(SS/Q&K)-3color	침구단품
시원한 슬리드 시어셔커 여름 베개커버(50x70)-8color	침구단품
리틀샘머 풍기민견 키즈 침대패드(SS/Q)	침구단품
BTS Run TinyTAN Tumbler Jin	주방용품
CLMP9100 프레쉬 원 마일 레깅스 네이비	스포츠웨어
길단 2000 울트라코튼 USA핏	상의
길단 2000 울트라코튼 USA핏 사파이어	상의
핑크 시리즈 쿠션 커버 (2 colors)	홈데코
Solid Linen Box Shirts [Cream]	상의
A RIBBON BROOCH PLEATS DR_IVORY	원피스/점프수트
[더블유비이프로젝트] 시티샘머 반팔셔츠 아이보리 KHSS7493	상의
퍼프 라운드 블라우스	상의
[티더블유엔] [세트] 퍼먼트플 반팔셔츠 6종 STSS3281 + 퍼먼트플 숏팬츠 6종 STSP3282	상의
[더블유비이프로젝트][세트] 미니 스웨트 세트 (티셔츠MJST7497 + 숏팬츠MJSP7498)	상의
[정해진 착용]카라콜오버세미오버반팔셔츠 MWSS3112	상의
빈티지플라워 순면60수 차림이불세트(SS/Q)-3color	침구세트
[제멋] 슬린 오버핏 반팔셔츠 화이트 YHSS2351	상의
[티더블유엔] 샘머코튼 반팔셔츠 화이트 SHSS3366	상의
PIERROT SS POLO SHIRT (WHITE)	상의

Fig. 11. Top-20 Items from Greedy Re-Ranking

다양성 증대 방식으로 새로운 아이템들이 추천되었을 때, 사용자가 그들 중에서 선택하는 아이템이 있을지를 예측하는 방식도 정확도를 측정하는 방법의 하나이다. 본 논문에서는 사용자가 클릭한 제품 중에서 추천목록에 있는 아이템들이 얼마나 포함되었는지를 확인하기 위해, 정밀도 (precision)와 재현율(recall)을 모두 반영한 F1 score로 평가한다.

재현율은 1로 수렴하는데 이는 추천목록 개수를 20개로 정했기에 때문에 사용자가 클릭한 아이템 수가 매우 큰 수가 되기 때문이다. 따라서 식(4)는 식(5)처럼 단순해진다.

$$F1\ score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \dots (4)$$

$$F1\ score = \frac{2 \cdot Precision}{Precision + 1} \dots (5)$$

Table 5. Category Diversity and Scores

Way	Category	F1 Score
simple top-20	{상의:12, 스포츠웨어:1, 원피스/점프수트:2, 주방용품:1, 침구단품:3, 홈데코:1}	0.182
bin packing	{라운드지워어:1, 상의:10, 스포츠웨어:1, 원피스/점프수트:2, 주방용품:1, 침구단품:1, 침구세트:1, 하의:2, 홈데코:1}	0.182
w. random choice 1	{상의:14, 아우터:1, 원피스/점프수트:1, 주방용품:1, 침구단품:1, 하의:2}	0.709
w. random choice 2	{상의:10, 아우터:1, 원피스/점프수트:1, 주방용품:1, 침구단품:2, 침구세트:1, 하의:2, 홈데코:1}	0.621
greedy re-ranking	{상의:11, 스포츠웨어:1, 원피스/점프수트:2, 주방용품:1, 침구단품:3, 침구세트:1, 홈데코:1}	0.182

단순 top-k 추천, 다양성 증대를 위한 빈 채우기 방식, 두 번의 weighted random choice 방식, 그리고 greedy re-ranking 방식 (정확도 가중치 0.7, 다양성 가중치 0.3) 결과들이 도출한 category2 종류와 F1 score를 Table 5에 제시한다. Inter-list 다양성 향상을 목적으로 하는 weighted random choice 방식의 F1 score가 다른 방식들에 비해 상대적으로 매우 높게 산출되었다. 이는 후보 추천목록의 고정된 100개 아이템에 대해서 누적 분포에 따라 확률적으로 아이템이 선택되도록 했기 때문에, 사용자 p가 클릭한 아이템들이 다른 접근방법들에 비해서 많이 포함된 것으로 추정된다. 이에 대한 정확한 원인을 위해 다음 연구에서는 다양한 크기의 추천목록 결과로 값을 비교해볼 필요가 있다.

Weighted random choice 방식을 제외한 나머지 방법들은 같은 F1 score를 보인다. 그러나 포함된 카테고리 개수를 보면, 단순 top-20은 6종류이지만, bin packing은 9종류, greedy re-ranking은 7종류로, 두 방식 모두 단순 추천보다 다양한 종류의 카테고리 아이템을 추천한다. 그런데도 F1 score가 같다는 것은 카테고리 다양성이 높아짐에도 정확도는 하락하지 않았다는 것을 의미한다.

IV. Conclusions

코로나19 바이러스(SARS-CoV-2) 팬데믹이 전 세계적으로 발생하면서 산업 대부분은 저성장을 보였지만 e-커머스 산업은 예외였다. 자의든 타의든 언택트 소비가 급속도로 확장되었고, 장기화한 코로나19 바이러스로 온라인 소비는 더욱 증가하고 있다. 범람하는 제품 및 서비스 정보를 소비자가 직접 비교하는 것은 이제는 불가능해졌고, 대신 추천시스템이 필수로 자리 잡아 소비자의 개인 정보 및 온라인 행동 패턴에 기반하여 소비자 자신도 알지 못했던 취향까지 반영하여 아이템이나 서비스를 추천한다. 이러한 추천 알고리즘은 점점 선별적인 개인화된 추천 결과

를 제공하기 때문에, 소비자는 자신의 취향에 갇히게 되어 편중된 결과만 받게 된다. 이러한 치우친 추천은 추천시스템에 대한 사용자의 만족도를 하락시키는 요인이 되기 때문에, 다양한 관점에서의 추천 품질 측정 지표가 반영될 필요가 있다. 대표적으로 다양성이 적용되는데, 다양성 지표는 사용자의 만족도를 높이면서도 다른 품질 평가지표에 비교해 상대적으로 정량화하기에 수월하기 때문이다. 다양성 증대를 통해 사용자는 자신의 선호를 반영한 카테고리뿐만 아니라, 선호 여부를 자신조차도 알지 못하는 카테고리에 속하는 아이템이나 서비스를 적절한 비율로 추천받게 된다. 또한 같은 추천 요구를 여러 번 수행하더라도 서로 다른 추천 결과를 받을 수 있다. 이는 가능한 많은 종류의 아이템들을 커버하면서도 사용자에게 새롭으면서도 신선한 추천을 경험하게 하므로 만족도를 높게 된다. 따라서 다양성 지표는 지금의 추천시스템 구현에 있어서 매우 중요한 요소라고 할 수 있다.

본 논문에서는 다양성 증대를 위한 기본 3가지 접근방법인 bin packing, weighted random choice, greedy re-ranking을 실제 e-커머스 데이터인 패션 쇼핑몰에 적용하여 도출되는 결과와 F1 score에 기반한 차이를 제시하였다. 이번 연구는 초개인화 추천을 위한 다양성 증가 방법에 대한 시작으로 본 논문에서는 단순 방법론을 구현하여 실제 데이터에 적용하여 비교하였다. 이는 최신 추천과 과거 추천 비중을 동일시하여 구현한 것으로 시간 흐름에 따른 사용자의 선호 변화를 반영하지 않은 것이다. 시간 가중치를 반영하여 다양성 증대 방법 구현 비교를 후속 연구로 진행하고 있으며, 더 나아가 사용자 프로파일링 구성을 통한 추천 결과의 연관성 및 다양성 향상 방안 연구를 진행하고자 한다.

ACKNOWLEDGEMENT

This paper was supported by the Research Fund, 2021, Pyeongtaek University in Korea.

REFERENCES

- [1] S. Barry, "The Paradox of Choice: Why More Is Less," Ecco Press, Revised ed., ISBN 9780062449924, 2016.
- [2] A. I. Hariad, and D. Nurjanah, "Hybrid attribute and personality based recommender system for book recommendation," 2017 International Conference on Data and Software Engineering (ICoDSE), November 2017, pp. 1-5, DOI: 10.1109/ICODSE.2017.8285874.
- [3] X. Xie, J. Lian, Z. Liu, X. Wang, F. Wu, H. Wang, and Z. Chen "Personalized Recommendation Systems: Five Hot Research Topics You Must Know," Microsoft Research Lab - Asia, November 2018.
- [4] D. Lee and K. Hosanagar, "How Do Recommender Systems Affect Sales Diversity? A Cross-Category Investigation via Randomized Field Experiment," Information Systems Research, Vol. 30, No. 1, Mar 2019, pp. 239-259, DOI: 10.1287/isre.2018.0800
- [5] Gil Park, "Recommendation Algorithms and Hyper-Personalization," MOBIINSIDE, February 2021, <https://www.mobiinside.co.kr/2021/02/01/algorithm-personalization/>
- [6] J. Choi and H. J. Lee, "An Integrated Perspective of User Evaluating Personalized Recommender Systems: Performance-Driven or User-Centric," Journal of Society for e-Business Studies, Vol. 17, No. 3, 2012, pp. 85-103, DOI: 10.7838/JSEBS.2012.17.3.085.
- [7] M. Kim, "Evaluation Metrics for Recommender Systems," March 2021, mskim.log.
- [8] H. Park and K. Kim, "Recommender system using BERT sentiment analysis," Journal of Intelligence and Information Systems, Vol. 27, No. 2, June 2021, pp. 1-15. DOI: 10.13088/JIIS.2021.27.2.001.
- [9] P. Castells, N. J. Hurley, S. Vargas, "Novelty and Diversity in Recommender Systems," In: F. Ricci, F., L. Rokach, B. Shapira, (eds) Recommender Systems Handbook. Springer, 2015, pp. 881-918, DOI:10.1007/978-1-4899-7637-6_26.
- [10] M. Kaminskas and D. Bridge, "Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems," ACM Transactions on Interactive Intelligent Systems Vol.7, No.1, March 2017, pp.1-42, DOI:10.1145/2926720.
- [11] S. Vargas and P/ Castells, "Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems," Proceedings of the fifth ACM conference on Recommender systems (RecSys'11), October 2011, pp. 109-116, DOI:10.1145/2043932.2043955.
- [12] F. O. Isinkayea, Y. O. Folajimib, and B. A. Ojokoher, "Recommendation systems: Principles, methods and evaluation," Egyptian Informatics Journal, Vol. 16, No. 3, Nov. 2015, pp. 261-273, DOI: 10.1016/j.eij.2015.06.005.
- [13] D. Jannach and M. Jugovac, "Measuring the Business Value of Recommender Systems," ACM Transactions on Management Information Systems, Vol. 10, No. 4, December 2019, pp 1-23, DOI:10.1145/3370082.
- [14] D. Lee, "A to Z for Implementation of a Recommender System with Practical Data," 2021, fastcampus.co.kr.
- [15] H. Na and K. Nam, "Application of diversity of recommender system according to user preference change," Journal of Intelligence and Information Systems, December 2020, Vol. 26,

No. 4, pp. 67-86, DOI: 10.13088/JIIS.2020.26.4.067.

- [16] J. P. Kelly and D. Bridge, "Enhancing the diversity of conversational collaborative recommendations: a comparison," *Artificial Intelligence Review* Vol. 25, 2006, pp. 79-95, DOI:10.1007/s10462-007-9023-8.

Author



Juryon Paik received the B.E. degree in Information Engineering from Sungkyunkwan University, Korea in 1997. She worked at the Samsung SDS company about a year. She achieved her M.E. and Ph.D. degrees in

Computer Engineering from the same university in 2005 and 2008, respectively. Dr. Paik joined the faculty of the Department of Digital Information and Statistics at Pyeongtaek University, Pyeongtaek-si, Korea, in 2016. She is currently an assistant professor. She is interested in tree mining, big data mining, information retrieval and recommender systems.