

Recommendation System Based on Correlation Analysis of User Behavior Data in Online Shopping Mall Environment

Yo Han Park[†] · Jong Hyeok Mun^{††} · Jong Sun Choi^{†††} · Jae Young Choi^{††††}

ABSTRACT

As the online commerce market continues to expand with an increase of diverse products and content, users find it challenging in navigating and in the selection process. Thereafter both platforms and shopping malls are actively working in conducting continuous research on recommendations system to select and present products that align with user preferences. Most existing recommendation studies have relied on user data which is relatively easy to obtain. However, these studies only use a single type of event and their reliance on time dependent data results in issues with reliability and complexity. To address these challenges, this paper proposes a recommendation system that analysis user preferences in consideration of the relationship between various types of event data. The proposed recommendation system analyzes the correlation of multiple events, extracts weights, learns the recommendation model, and provides recommendation services through it. Through extensive experiments the performance of our system was compared with the previously studied algorithms. The results confirmed an improvement in both complexity and performance.

Keywords : Recommendation System, User Behavior Data, Implicit Feedback Data, Co-occurrence, Target Data, Auxiliary Data

온라인 쇼핑몰 환경에서 사용자 행동 데이터의 상관관계 분석 기반 추천 시스템

박요한[†] · 문종혁^{††} · 최종선^{†††} · 최재영^{††††}

요약

매년 증가하는 온라인 상거래 시장과 점차 다양해지는 상품과 콘텐츠로 인해 사용자들은 선택 과정에 어려움을 느낀다. 이에 여러 기업들은 온라인 쇼핑몰에서 사용자가 선호할 상품을 선별하여 제공하기 위해 추천 시스템에 대한 지속적인 연구를 진행하고 있다. 대다수의 추천 시스템 연구에서는 비교적 획득하기 쉬운 사용자의 이벤트 데이터를 기반으로 연구를 진행하였으나 한 종류의 사용자 행동만을 고려하기 때문에 사용자의 선호도를 파악하는 것에 오차가 발생한다. 이에 본 논문에서는 여러 종류의 사용자 행동 데이터의 상관관계를 고려하여 사용자의 선호도를 분석하는 추천 시스템을 제안한다. 제안하는 추천 시스템은 사용자의 사용자 행동 데이터의 상관관계를 분석하고 가중치를 생성하여 추천 모델을 학습한다. 실험에서는 기존 연구의 알고리즘과의 성능 비교를 통해 제안하는 시스템의 복잡도와 성능 향상을 확인하였다.

키워드 : 추천 시스템, 사용자 행동 데이터, 암시적 피드백 데이터, 상관관계 분석, 목표 데이터, 보조 데이터

1. 서론

최근 코로나19 이후 전자상거래의 급격한 증가와 함께 다양한 상품과 콘텐츠들이 원하는 제품을 선택할 수 없을 정도로 많아졌다. 이로 인해 소비자들은 구매하고자 하는 제품을 선택하기가 더 어려워졌다. 수많은 온라인 아이템들 사이에서

원하는 제품을 찾는 것은 아이템이 늘어남에 따라 어려워지며, 사용자들은 더욱 제품을 선택하기 어려워진다[1]. 온라인 판매 플랫폼은 추천 시스템을 통해 사용자가 선호할 가능성이 있는 아이템을 선별하여 제공함으로써, 사용자들이 신속하게 제품을 구매할 수 있도록 도와준다[2].

사용자는 추천 시스템을 사용하여 원하는 상품 또는 콘텐츠를 탐색하는 시간과 노력을 줄일 수 있다. 실제로 구글 앱스토어, 유튜브, 넷플릭스 등 이커머스 기업들에서는 사용자의 만족도 향상을 위해 추천 시스템을 통한 사용자 맞춤형 콘텐츠 추천 서비스를 제공하고 있다[3-5].

추천 시스템의 추천 알고리즘은 크게 협업 필터링(Collaborative filtering, 이하 CF)과 콘텐츠 기반 필터링(Content-based filtering, 이하 CB)으로 분류된다. CF는 사용자들의 특

※ 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2022-0-00218).

† 비회원 : 송실대학교 컴퓨터학과 박사과정

†† 준회원 : 송실대학교 컴퓨터학과 박사과정

††† 정회원 : 송실대학교 컴퓨터학부 부교수

†††† 종신회원 : 송실대학교 컴퓨터학부 교수

Manuscript Received : November 10, 2023

Accepted : December 18, 2023

* Corresponding Author : Jong Sun Choi(jongsun.choi@ssu.ac.kr)

징들을 분석하여 유사한 사용자들을 찾고 유사한 사용자들이 선호했던 아이템들을 추천한다. 반면 CB는 사용자가 선호했던 아이템들의 특징들을 분석하여 유사한 아이템들을 찾고 유사한 아이템들을 추천한다. 이러한 두 가지 알고리즘은 공통적으로 사용자 데이터, 아이템 데이터, 그리고 인터랙션 데이터가 필요하며[6], 세부 내용은 다음과 같다.

- 사용자 데이터: 사용자의 성별, 위치 등의 사용자 특징을 나타내는 데이터
- 아이템 데이터: 아이템의 범주, 장르 등의 아이템 특성을 가진 데이터
- 인터랙션 데이터: 사용자가 해당 아이템의 선호 여부를 표현한 데이터

명시적 피드백 데이터(explicit feedback data)[7]는 사용자가 아이템을 소비한 후 자신이 해당 아이템을 얼마나 선호했는지를 직접적으로 표기한 데이터로, 영화 평점과 같은 점수 형태이다. 사용자가 직접 선호도를 표기했기 때문에 데이터의 신뢰도가 높아 추가적으로 가공하지 않고 추천 시스템에 사용할 수 있다. 하지만 사용자가 온라인 환경에서 아이템을 소비한 후에 자신의 선호도를 표현하지 않는 경우가 많아 암시적 피드백 데이터를 사용한 사용자의 선호도 유추 방법이 필요하다. 암시적 피드백 데이터(implicit feedback data)[8]는 사용자가 해당 아이템에 대해서 어떤 행동들을 취했는지에 대하여 수집한 로그 형식의 데이터이다. 암시적 피드백 데이터 자체로는 사용자가 해당 아이템을 선호하는지를 파악하기 어렵지만, 상관관계 분석과 같은 데이터 분석 기술을 활용하여 사용자의 아이템 선호도를 유추할 수 있다.

기존 추천 시스템 연구들을 살펴보면 사용자가 아이템에 대한 자신의 선호도를 직접 표현한 데이터인 명시적 피드백 데이터를 사용하는 연구를 진행하였다[9]. 사용자에게서 명시적 피드백 데이터를 얻기 위해서는 사용자에게 평가의 기준을 제시해야 하므로 사용자 선호도를 조사하기 위한 추가적인 평가 시스템이 필요하다. 그러나 사용자들은 평가 시스템을 제공하여도 아이템을 소비한 후 자신의 선호도를 기록하지 않는 경우가 많다. 이러한 이유로 명시적 피드백 데이터는 딥러닝 추천 시스템을 학습할 만큼 많은 데이터를 수집하는데 어려움이 있다[10].

암시적 피드백 데이터는 사용자의 선호도를 간접적으로 유추하기 위해 사용되고, 사용자들이 다양한 아이템에 대하여 어떤 행동들을 취했는지에 대한 로그 형식의 데이터로 저장된다.

암시적 피드백 데이터는 사용자들의 행동에 대한 데이터를 모두 포함하고 있으므로, 사용자가 자신의 선호도를 직접적으로 표현하지 않은 아이템에 대한 사용자 행동의 간접적인 데이터도 포함한다. 이러한 이유로 암시적 피드백 데이터에는 불필요한 데이터도 많으며, 사용자가 선호하지 않는 아이템에 대

한 데이터가 없는 특징을 포함하고 있다. 이러한 특징들을 고려하여 사용자의 선호도를 간접적으로 유추하기 위해 사용자의 행동 데이터를 분석하는 연구들이 진행되고 있다[11,12].

실제 온라인 환경에서 수집되는 로그 데이터는 사용자 행동에 관한 여러 종류의 데이터를 포함한다. 사용자 행동의 종류를 '이벤트'로 지칭하며, 여러 종류의 이벤트가 포함된 암시적 피드백 데이터는 이기종 암시적 피드백 데이터로 지칭한다. 온라인 쇼핑물 환경에서 아이템에 대한 사용자의 행동을 다루는 대다수의 연구들에서는 영향도가 가장 높은 하나의 암시적 피드백 데이터를 사용하여 사용자의 선호도를 유추한다[13,14].

이기종 암시적 피드백 데이터를 이벤트에 따라 구분하지 않고 사용하거나 한 종류의 이벤트만을 사용하게 되면, 다른 암시적 피드백 데이터가 가지고 있는 사용자 선호 정보가 소실될 수 있다. 소실된 정보로 인해 아이템에 대한 사용자 선호도를 예측하기 어려워질 수 있으므로, 암시적 피드백 데이터를 전체적으로 활용하여 유추해석하기 위한 연구가 요구된다. 이에 본 연구에서는 사용자 행동 데이터인 이기종 암시적 피드백 데이터의 상관관계를 분석하여 가중치를 생성하고, 이를 바탕으로 사용자의 선호 아이템을 유추하여 제공하는 추천 시스템을 제안한다. 제안하는 추천 시스템은 사용자로부터 발생된 이기종 암시적 피드백 데이터의 손실을 최소화하여 추천 서비스의 품질을 향상시킬 수 있다. 논문의 2장에서는 대표적인 추천 시스템 연구들과 기존 암시적 피드백 데이터의 분석 방법에 대하여 작성하였으며, 3장에서는 사용자의 여러 행동 데이터를 분석하여 추천 시스템에 적용시키는 알고리즘을 소개한다. 4장에서는 암시적 피드백 데이터를 사용하는 다른 방법들과 결과를 비교하며 마지막 5장에서는 본 연구의 결론을 작성한다.

2. 관련 연구

본 장에서는 배경 지식과 기존의 대표적인 추천 시스템 연구들에서 암시적 피드백 데이터의 분석 방법 및 가중치 적용 등에 관한 내용을 살펴본 후, 제안하는 추천 시스템의 요구사항을 도출한다.

2.1 배경 지식

피드백 데이터란 온라인 환경에서 제공되는 서비스를 사용하는 사용자의 행동 데이터이다. 피드백 데이터의 종류는 아래와 같다.

- 명시적 피드백 데이터 : 사용자가 아이템에 대한 선호도를 수치적으로 표현한 데이터
- 암시적 피드백 데이터 : 사용자로부터 발생한 아이템에 대한 행동(클릭, 문의, 구매 등) 데이터

추천 시스템이 사용자가 원하는 아이템을 추천하기 위해서는 사용자가 소비한 아이템에 대한 사용자 선호도 정보가 필요하다. 명시적 피드백 데이터는 점수 형식의 데이터이므로 사용자 선호도로 사용할 수 있다. 이와 다르게 암시적 피드백 데이터는 사용자 행동에 대한 데이터로 데이터 분석을 통해 간접적으로 사용자 선호도를 구해야 한다[15].

피드백 데이터를 사용하는 추천 시스템은 행렬 분해 기반 협업 필터링과 딥러닝 기반 협업 필터링 연구로 구분된다[16, 17]. 행렬 분해 기반 추천 시스템의 사용자 선호도 측정은 사용자 사이 상관관계를 파악하는 부분에서 한계를 가지고 있으며, 이를 해소하기 위해 딥러닝 기반 추천 시스템을 사용한다. 딥러닝 기반 추천 시스템은 사용자 사이 상관관계를 학습하기 위해 많은 데이터가 필요하므로 명시적 피드백 데이터보다 암시적 피드백 데이터를 사용한다[18].

암시적 피드백 데이터를 사용하는 추천 시스템은 사용자의 행동이 존재하면 1, 존재하지 않으면 0으로 사용자 선호도를 표현하며, 사용자에게 사용자의 행동이 존재할 아이템을 추천한다. 추천 시스템에서 여러 종류의 사용자 행동 데이터를 모두 사용할 때는 사용자의 선호도가 가장 높은 사용자 행동을 목표 데이터로 설정하며, 나머지 행동들을 보조 데이터로 설정하여 데이터 사이 관계성을 구하여 사용한다[19,20].

2.2 이기종 암시적 피드백 데이터를 적용한 추천 시스템

He의 연구[18]와 같이 이기종 암시적 피드백 데이터에서 한 종류의 사용자 행동 데이터만을 사용하게 되면, 다른 사용자 행동 데이터가 추천 시스템에 반영되지 않는다. 이에 따라 Pan의 연구[22]와 Qiu의 연구[23]는 이기종 암시적 피드백 데이터를 적용하기 위해서 BPR(Bayesian Personalized Ranking)[21] 기반 알고리즘을 제안하였다. BPR은 아이템 선호도를 각 아이템마다 구하는 행렬 분해와 다르게 사용자가 두 개의 아이템 중에 어떤 아이템을 더 선호하는지를 측정한다. ABPR(Adaptive Bayesian Personalized Ranking for heterogeneous implicit feedbacks)[22]은 BPR을 사용하여 사용자의 목표 데이터를 가진 아이템과 보조 데이터를 가진 아이템을 비교하였다. 목표 데이터를 가진 아이템을 기준으로 보조 데이터가 사용자 선호도를 어느 정도 갖고 있는지를 수치로 표현하여 가중치로 사용하였다. Pan의 연구는 영화 데이터에서 영화 시청 데이터를 목표 데이터로 사용하고 영화 검색 데이터를 보조 데이터로 사용하여 실험을 진행하였으며, 영화 시청 데이터만 사용한 추천 시스템보다 성능이 좋아진 것을 확인하였다.

BPRH(Bayesian Personalized Ranking for Heterogeneous implicit feedback)[23] 또한 BPR을 사용하여 보조 데이터의 가중치를 설정하였다. Qiu[23]의 연구는 온라인 쇼핑물 데이터에서 '아이템 구매'를 목표 데이터로 사용하였으며, '아이템 클릭'과 '아이템 좋아요'를 보조 데이터로 사용하였다. 목표 데이터와 보조 데이터의 상관관계를 정량화하기 위해 Equation

(1)을 사용하였으며, 정량화된 값을 사용자 행동의 가중치로 사용하였다.

$$C^{uu}(X) = \frac{2 * C_{ta}^{uu}(X) * C_{at}^{uu}(X)}{C_{ta}^{uu}(X) + C_{at}^{uu}(X)} \quad (1)$$

Equation (1)에서 u 는 목표 데이터를 의미하며, a 는 보조 데이터를 의미한다. BPRH는 상관관계에 기반한 가중치를 생성하여 적용하는 것이 추천 시스템 성능에 영향을 미치는 것을 증명하였다. 그러나 BPRH에서는 전체 데이터의 상관관계를 파악한 것이 아니라, 목표 데이터와 보조 데이터 사이의 상관관계만을 사용하여 상관관계를 정량화하였다. BPR을 사용한 연구들[22,23]은 데이터의 모든 상관관계를 구하는 것을 목표로 하지 않고 목표 데이터와 보조 데이터의 관계에만 집중하였다.

EHCF(Efficient Heterogeneous Collaborative Filtering without Negative Sampling for Recommendation)[24]는 모든 데이터들 사이의 상관관계를 구하기 위해 사용자 행동들 사이에 시간 순서를 학습하여 높은 성능을 보여준다. Chen의 연구[24]는 모든 데이터의 상관관계를 파악하는 것이 추천 성능에 어떤 영향을 미치는지를 증명하기 위해, 사용자 행동들을 시간 순서대로 시나리오를 구성하여 시나리오마다 다른 가중치를 설정하였다. EHCF는 온라인 쇼핑몰 데이터에서 '구매'를 목표 데이터로 사용하고, '클릭'과 '장바구니에 담기'를 보조 데이터로 사용하였다. 시나리오 구성은 [클릭 → 장바구니에 담기 → 구매], [클릭 → 구매], [장바구니에 담기 → 구매]와 같이 구성하였으며, 같은 '클릭'이라도 다른 시나리오에 있으면 다른 가중치를 부여했다. Chen의 연구는 기존의 목표 데이터와 보조 데이터의 상관관계와 함께 시간 관계성을 같이 사용하여 더 높은 추천 성능을 가질 수 있다는 것을 보여주었다.

2.3 머신러닝 기반 가중치 추천 시스템

머신러닝 알고리즘 기반 가중치 추천 시스템은 머신러닝 알고리즘을 사용하여 사용자 또는 아이템의 특징마다 다른 가중치를 부여하여 추천 정확도를 상승시킨다. 명시적 피드백 데이터를 사용하는 추천 시스템에서는 아이템의 특징들마다 다른 가중치를 구하여 적용하는 추천 시스템을 제안하였다 [25,26]. 이 연구에서는 추천 모델을 학습하기 전에 아이템의 서로 다른 특징들에 대해서 회귀 분석을 사용하여 가중치를 계산하였고, 아이템의 특징들에 대해서 관리할 수 있다.

기존 추천 시스템 구조에 머신러닝을 결합하여 가중치를 구하는 모듈을 추가한 추천 시스템 연구가 진행되었다. 영화 추천을 위한 Himel의 연구[27]는 사용자들 사이의 유사성을 측정하기 위해 머신러닝을 사용하였으며, 유사성 기반 가중치를 사용하여 높은 추천 성능을 보여주었다. 또한 Cui의 연구[28]는 [사용자, 아이템] 관계를 파악하는 추천 시스

템 구조에서 [사용자, 아이템, 시간] 관계를 파악하여 가중치를 생성하는 모듈을 추가한 추천 시스템 구조를 제안하였다. Bhimacarapu의 연구[29]에서는 환자에게 부작용 없이 의약품을 제공하기 위해 가중치 기반 추천 시스템을 제안하였다. 제안된 추천 시스템은 환자 데이터와 증상 데이터를 통계적 분석과 특이값 분해를 통해 가중치를 생성하였으며, 딥러닝 추천 모델을 사용하여 의약품을 추천하였다. 그러나 머신러닝 모델을 사용한 가중치 생성 연구는 머신러닝 모델의 학습 및 사용하는 데에 시간이 많이 소요된다.

온라인 쇼핑물 환경의 특성상 사용자의 이전 행동이 다음 아이템을 추천하는 알고리즘에 바로 반영되어야 한다. 그러나 머신러닝 알고리즘을 사용하면 사용자에게 아이템 추천이 늦어지기 때문에 Jiang의 연구[30]에서는 머신러닝 알고리즘을 사용하지 않는 추천 시스템을 제안하였다. Jiang의 연구에서는 머신러닝을 사용하지 않고 개인 사용자 평점 데이터에 신뢰도 분석 결과와 유사한 사용자 분석 결과를 결합하여 가중치를 생성하였고, 이를 통해 추천 성능을 높였다. Guo의 연구[31]에서는 실제 온라인 쇼핑물 데이터인 ‘Tabao’ 데이터 세트를 사용하여, 사용자 정보 분석 데이터와 여러 종류의 추천 알고리즘 결과를 사용하여 추천된 아이템의 가중치를 생성하는 접근 방식을 적용하였다. 이처럼 사용자의 행동 데이터를 다음 아이템에 바로 적용시키기 위해서 데이터 분석을 통한 가중치를 생성하여 사용한다.

가중치 기반 추천 시스템은 여러 종류의 암시적 피드백 데이터를 최대한 많이 사용하여 사용자 선호 정보가 소실되지 않아야 하며, 사용자에게 아이템 추천이 오래 걸리지 않기 위해 가벼운 가중치 생성 모듈이 필요하다. 이를 위해 이기종 암시적 피드백 데이터를 사용하여 가중치를 생성할 때는 Pan의 연구[22]와 Qiu의 연구[23]에서처럼 상관관계 기반 가중치를 사용해야 한다. Pan의 연구[22]와 Qiu[23]의 연구에서는 목표 데이터와 보조 데이터 사이의 1대1 상관관계에 초점을 두어 모든 데이터 사이의 상관관계를 파악할 필요가 있다. 또한 머신러닝 알고리즘을 사용하여 가중치를 생성하면 사용자에게 아이템 추천을 위한 시간이 증가하기 때문에 가중치 생성 모듈은 상관관계를 파악하지만 가벼워야 한다.

본 논문에서는 추천 과정에서 가중치를 생성하기 위해 머신러닝 알고리즘을 사용하는 대신 연산량을 줄이고 추천 모델의 성능 향상시키기 위해 상관관계 분석 알고리즘을 적용한다. 또한 간 상관관계를 파악하기 위해 목표 데이터와 보조 데이터의 관계와 사용자 행동의 빈도수를 사용하여 가중치를 생성한다. 제안하는 추천 시스템은 온라인 쇼핑물 환경에서 아이템 추천 서비스의 성능을 향상시키고자 한다.

3. 본 론

본 장에서는 상관 관계 분석 기반 추천 시스템을 제안한다.

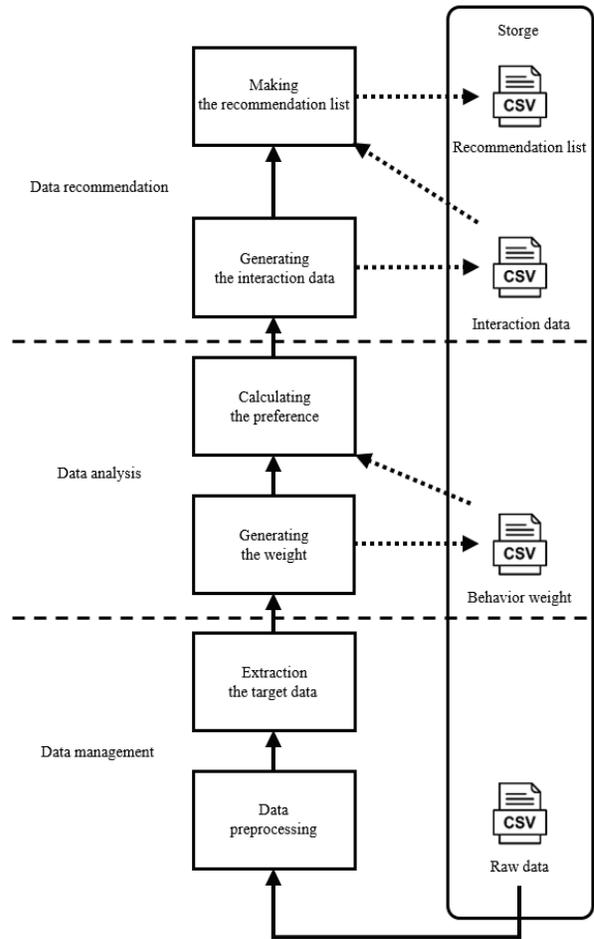


Fig. 1. Correlation Weight Recommendation System Overview

제안하는 추천 시스템은 암시적 피드백 데이터의 상관 관계를 분석을 통해 각 사용자 행동의 중요도를 파악하고, 행동마다 다른 가중치를 사용하여 성능을 개선하는 것을 목표로 한다.

3.1 제안하는 추천 시스템 개요

제안하는 추천 시스템은 사용자가 아이템에 대해 반응한 행동 데이터를 입력 받는다. 온라인 쇼핑물 기준으로 보았을 때 ‘클릭’, ‘장바구니에 담기’, ‘구매’ 등이 행동 데이터의 예시이다. 본 시스템은 행동 데이터를 통계 기반 분석을 통해 행동마다 서로 다른 가중치를 생성하고 가중치를 통해 사용자의 선호도를 간접적으로 유추하여 추천 정확도를 개선하는 것을 목표로 한다. 이를 위한 제안하는 추천 시스템은 데이터 분리 모듈, 데이터 분석 모듈, 가중치 생성 모듈, 인터렉션 데이터 생성 모듈, 아이템 추천 모듈 등으로 구성되며 전반적인 시스템의 개요는 Fig. 1과 같다.

제안하는 상관관계 분석 기반 추천 시스템의 전반적인 처리 과정은 Fig. 2와 같이 전처리, 목표 데이터 추출, 가중치 생성, 선호도 계산, 인터렉션 데이터 생성 및 추천 순으로 이루어진다. 먼저 사용자의 여러가지 행동 데이터들이 포함된 이기종

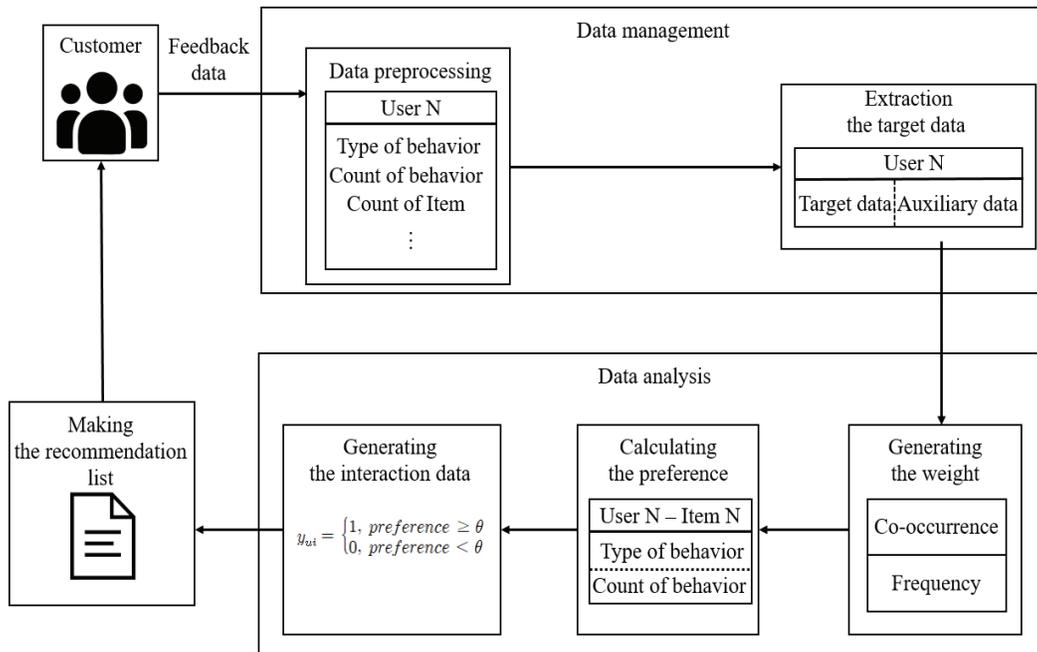


Fig. 2. Proposed Weight-based Recommendation System Service Configuration

암시적 피드백 데이터를 사용한다. 이기중 암시적 피드백 데이터 분석은 여러 행동 데이터를 분석해야하기 때문에 데이터 분리 모듈을 통해 목표 데이터(Target data)와 보조 데이터(Auxiliary data)로 나누는 작업이 선행된다. 이후에는 데이터들의 동시 발생 확률(Co-occurrence)과 발생 빈도(frequency)를 측정하며 목표 데이터와 보조 데이터를 통계적으로 분석한다. 데이터를 분리하고 나면 사용자의 각 행동들이 아이템 소비에 어느 정도 영향을 미치는지 수치적으로 나타내기 위한 통계적 분석에 기반한 가중치 정량화를 수행한다. 생성된 가중치와 사용자의 해당 아이템에 대한 행동 데이터는 아이템의 간접적인 선호도를 측정하는 데 사용된다. 선호도는 이진 분류기를 통해 분류되어 최종적으로 딥러닝 추천 모델에 학습되며 사용자는 추천 모델을 통해 아이템 추천 리스트를 제공받는다.

3.2 데이터 전처리 모듈

데이터 전처리 모듈(Data preprocessing)은 Fig. 3과 같으며, 이기중 암시적 피드백 데이터에서 제안하는 추천 시스템에서 사용되지 않는 항목들을 제거하고 사용자 기준으로 데이터를 전처리한다.

본 논문에서 사용자에게 빠르게 추천 아이템을 제공하기 위해 Fig. 3과 같이 시간 항목을 제거하며, 사용자와 아이템의 고유 번호를 결측치 없이 대응시킨다. 또한 각 사용자를 기준으로 행동, 행동 횟수, 사용자 행동이 발생한 아이템 수 등의 데이터를 정리한다.

3.3 목표 데이터 추출 모듈

목표 데이터 추출 모듈(Extraction the target data)은 Fig. 4

와 같다. 데이터 발생 유무 및 일부 데이터를 사용하는 것이 아닌 모든 데이터를 사용하기 위해 이기중 암시적 피드백 데이터를 목표 데이터를 추출한다. 목표 데이터는 사용자의 행동 중에 가장 높은 선호도를 갖는 행동으로 설정한다. 온라인 쇼핑몰에서는 직접적인 수익으로 연결되는 '아이템 구매'가 목표 데이터가 된다. 보조 데이터는 목표 데이터를 발생시키기 위한 사용자의 행동들을 의미하며, 목표 데이터가 아닌 사용자 행동들을 포함한다. 또한 보조 데이터를 통해 몇 개의 사용자 행동이 존재하는지와 각 행동마다의 희소성을 파악할 수 있다.

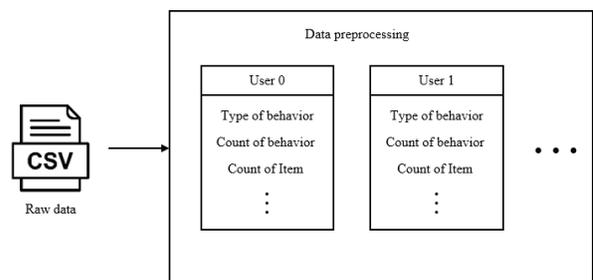


Fig. 3. Data Preprocessing

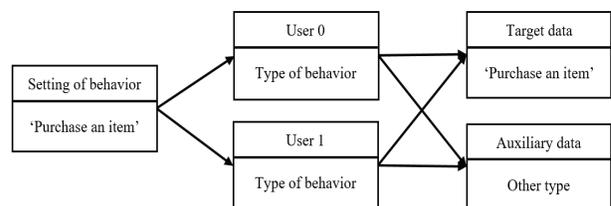


Fig. 4. Extraction the Target Data

3.4 가중치 생성 모듈

가중치 생성 모듈(Generating the weight)은 목표 데이터를 기준으로 사용자의 각 행동들이 목표 데이터를 발생시키는 영향도 측정한다. 이는 사용자의 해당 아이টে에 대한 선호도를 나타내는 지표로 활용된다.

보조 데이터에 포함된 사용자의 행동들이 목표 데이터 발생에 얼마나 영향을 미쳤는지를 수치화하기 위해 Table 1에 기술된 알고리즘을 수행한다. 목표 데이터와 보조 데이터들 동시 발생 확률을 계산한 이후에, 사용자 행동 빈도가 높고 동시 발생 확률이 낮으면 해당 행동의 가중치를 낮게 측정하고, 사용자 행동 빈도가 낮고 동시 발생 확률이 높으면 해당 행동의 가중치를 높게 측정한다. 발생 빈도에 따라 가중치를 조절하는 이유는 단순 클릭과 같이 발생 빈도가 높지만 의미없는 사용자 행동 데이터의 중요도를 낮추며 목표 데이터와 밀접하지만 발생 빈도가 낮은 사용자 행동 데이터의 중요도를 높여 정확한 사용자 선호도를 측정하기 위함이다. Table 1을 통해 생성된 가중치는 사용자의 행동마다의 사용자 선호도를 포함하고 있으므로, 가중치를 비교하여 어떤 행동이 사용자의 선호도가 높은 것인지를 수치상으로 비교할 수 있다.

Table 1. Algorithm of Generating the Weight

Input	: User feedback data(Csv file)
Output	: User behavior weight
1	TargetData, AuxiliaryData, behaviorType = readCsv(User feedback data)
2	targetItemIndex = index(TargetData)
3	auxiliaryItemIndex = index(AuxiliaryData)
4	auxiliaryTypeIndex = auxiliaryItemIndex[behaviorType]
5	for u ← 0 to length(User) do
6	for i ← 0 to behaviorType do
7	co-occurrence =
	count(targetItemIndex &
	auxiliaryItemIndex[i])/length(targetItemIndex)
8	frequencyRate =
	auxiliaryItemIndex[i])/length(auxiliaryItemIndex)
9	if co-occurrence > 0.5 then
10	if frequencyRate ≤ 0.5 then
11	weight[u][i] =
	co-occurrence*(1+frequencyRate)
12	else
13	weight[u][i] = co-occurrence
14	end if
15	else
16	if frequencyRate > 0.5 then
17	weight[u][i] =
	co-occurrence*(1-frequencyRate)
18	else
19	weight[u][i] = co-occurrence
20	end if
21	end for
22	end for
23	return weight

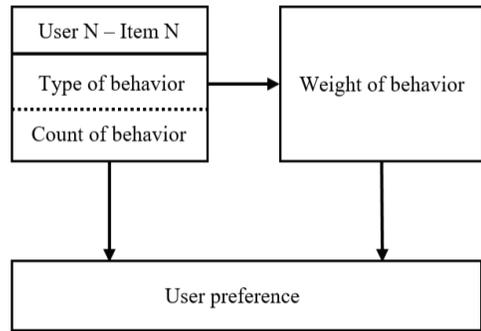


Fig. 5. Calculating the Preference

3.5 사용자-아이템 선호도 계산 모듈

사용자 선호도 계산 모듈(Calculating the preference)은 Fig. 5와 같으며, 가중치와 사용자의 여러 행동의 횟수를 분석하여 아이টে에 대한 사용자 선호도를 계산한다. 계산된 선호도는 사용자-아이템 쌍의 형식으로 저장되며, 사용자의 아이টে에 선호도를 수치화한다.

3.6 인터랙션 데이터 생성 모듈

인터랙션 데이터 생성 모듈(Generating the interection data)은 사용자의 선호도를 딥러닝 추천 모델에 입력되기 위한 인터랙션 데이터로 변환한다. 암시적 피드백 데이터를 사용하는 추천 모델은 인터랙션 데이터가 필요하며, Equation (2)를 통해 선호하는 아이টে와 선호하지 않는 아이টে으로 구분한다. 인터랙션 데이터는 아이টে을 구분하기 위해 임계값을 가지며 사용자 선호도가 임계값보다 높으면 사용자가 해당 아이টে을 선호하는 것으로 구분하고, 임계값보다 낮으면 사용자가 해당 아이টে을 선호하지 않는 것으로 구분한다. 해당 모듈을 통해 사용자가 해당 아이টে에 사용자 행동이 존재하더라도 선호하지 않는 아이টে 또한 구분할 수 있다.

$$y_{ui} = \begin{cases} 1, & preference \geq \theta \\ 0, & preference < \theta \end{cases} \quad (2)$$

3.7 추천 아이টে 리스트 생성 모듈

본 모듈에서의 추천 과정에는 암시적 피드백 데이터를 사용하는 대표적인 모델인 NCF(Neural Collaborative Filtering)를 사용한다. 추천 아이টে 리스트 생성(Making the recommendation list)은 사용자 인덱스 데이터와, 아이টে 인덱스 데이터, 그리고 인터랙션 데이터를 사용하며 NCF 모델을 학습한다. NCF 모델은 사용자가 선호하는 아이টে와 선호하지 않는 아이টে에 대한 관계성을 학습하고 사용자가 소비하지 않은 아이টে 중에 선호할 아이টে를 추천한다.

4. 실험 및 평가

본 장에서는 실제 온라인 환경에서 수집된 데이터 세트를 활용하여 가중치와 인터랙션 데이터를 생성하고 결과를 확인

한다. 가중치 생성은 머신러닝 알고리즘을 사용하는 것보다 연산량을 줄이기 위해 상관관계 분석 알고리즘을 사용하며, 제안하는 가중치 기반 추천 시스템의 성능을 확인하기 위해 기존 암시적 피드백 데이터를 사용하는 추천 시스템들과 비교한다.

4.1 실험 데이터 및 실험

실험은 이기종 암시적 피드백 데이터인 BeiBei 데이터 세트[32]와 Sobazaar 데이터 세트[33]를 사용한다. 두 개의 데이터 세트들은 모두 실제 온라인 쇼핑물 환경에서 수집한 데이터이며, 각 데이터 세트의 사용자 수, 아이템 수, 사용자 행동의 종류는 Table 2와 같다. 실험은 데이터 분리 모듈을 통해 사용자의 행동을 분리하며, 분리된 사용자 행동의 종류는 Table 3과 같다. 목표 데이터는 사용자 행동 중에 가장 높은 선호도를 갖는 Beibei의 'purchase'와 Sobazaar의 'purchase:buy_clicked'로 설정한다.

Table 3에서 Sobazaar의 보조 데이터 8가지 사용자 행동 중 목표 데이터인 'purchase:buy_clicked'과 상관관계가 낮은 'pixel-order'와 'pixel-order-no-reference'는 추천 모델에 사용하지 않는다.

가중치 생성 모듈을 통해 목표 데이터와 보조 데이터의 상관관계를 구하기 위해 Beibei dataset은 [purchase, view], [purchase, addtocart]의 동시 발생 확률을 구한다. 보조 데이터의 발생 빈도에 기반한 [view, addtocart]의 가중치를 생성한다. 이후 동시 발생 확률로 생성된 가중치에 보조 데이터의

Table 2. Data Sets

	User	Item	Recorded	Event
Beibei	21,716	7,977	3,359,784	3
Sobazaar	17,126	24,785	488,996	9

Table 3. Event Type

	User behavior type
Beibei	view
	addtocart
	purchase
Sobazaar	content:product_clicked
	content:product_detail_viewed
	content:product_wanted
	pixel-init
	pixel-order
	pixel-order-no-reference
	product_detailed_clicked
	product_wanted
purchase:buy_clicked	

발생 빈도에 기반한 가중치를 추가하여 중요도가 높은 행동의 가중치를 증가시키고, 중요도가 낮은 행동의 가중치를 낮추어 생성한다. 마찬가지로 Sobazaar dataset에서 목표 데이터와 보조 데이터의 동시 발생 확률을 구하며, 보조 데이터의 발생 빈도에 기반하여 생성된 가중치를 변경한다. 생성된 가중치는 사용자의 여러 종류의 행동들이 아이템 선호도에 대한 영향도를 나타낸다.

4.2 가중치 생성 결과

Table 4는 Beibei 데이터 세트에서 하나의 아이템에 대한 각 사용자의 상관관계 분석과 보조 데이터의 발생 빈도이다. Table 4를 통해 'purchase'를 기준으로 'view'와 'addtocart'의 상관관계를 수치화해서 표현할 수 있으며, 'addtocart'가 'view'보다 'purchase'와 더 큰 상관관계를 가지는 것을 확인할 수 있다. 또한 V-AC 발생 빈도를 통해 'addtocart'가 'view'보다 더 중요한 보조 데이터임을 확인할 수 있다. 이는 'addtocart'가 'view'보다 더 유의미한 데이터인 것을 확인할 수 있다.

Table 5는 Sobazaar 데이터 세트에서 하나의 아이템에 대한 사용자 세 명의 'product_detail_viewed', 'pixel-init', 'product_wanted' 세 가지 행동 가중치이다. Table 5를 통해 세 가지 사용자 행동의 가중치를 비교하면 대부분 'purchase'를 기준으로 'pixel-init'이 가장 높은 상관관계를 가지고, 'product_wanted'가 다음으로 높은 상관관계를 가지며, 'product_detail_viewed'가 가장 낮은 상관관계를 보인다. 이를 통해 사용자의 행동은 'pixel-init' > 'product_wanted' > 'product_detail_viewed'의 순서대로 유의미하다고 판단할 수 있다.

가중치 생성 모듈을 통해 생성된 가중치와 사용자 행동의 횟수를 기반으로 해당 아이템을 얼마나 선호하는 지를 수치화한다. 수치화된 사용자 선호도를 통해 인터랙션 데이터를 생성하며 인터랙션 데이터를 서비스 추천 모듈의 입력값으로 사용한다.

Table 4. Some of the Weights Generated Using Beibei

	View	Addtocart	V-AC
User1	0.094	0.291	0.2
User2	0.24	0.395	0.3
User3	0.11	0.553	0.2

Table 5. Some of the Weights Generated Using Sobazaar

	Product_detail_viewed	Pixel-init	Product_wanted
User1	0.37	0.71	0.47
User2	0.31	0.56	0.49
User3	0.27	0.43	0.41

4.3 인터랙션 데이터 생성 결과

사용자 선호도가 임계값 이상이면 인터랙션 데이터가 1로 생성되며, 임계값 이하이면 0으로 생성된다. 실험결과, 임계값의 경우 실험을 진행하였을 때 0.5까지는 성능이 상승하였지만 0.5 이상의 경우 성능이 감소하는 경향을 보였다. 따라서 본 논문에서 임계값은 0.5로 설정하여 실험을 진행한다.

Table 6는 임계값에 따른 사용자의 행동이 존재하는 데이터와 인터랙션 데이터가 1로 표현된 데이터의 비율 변화이다. 임계값이 0.1일 경우 가중치가 적은 사용자 행동이 존재할 때에도 인터랙션 데이터가 1로 표시되는 것을 확인할 수 있었다. 그리고 0.3을 임계값으로 설정했을 경우 Beibei 데이터 세트에서 가중치가 가장 낮은 사용자 행동인 'view'만 존재하는 데이터가 제외되는 경향이 있었다.

마찬가지로 Sobazaar 데이터 세트에서 가중치가 낮은 사용자 행동인 'product_clicked'만 존재하는 데이터가 제외되는 경향이 있었다. 반대로 0.9와 같이 높은 수를 임계값으로 설정했을 때는, 인터랙션 데이터에 1로 표시된 데이터가 Beibei 데이터 세트에서 대부분 'purchase'가 발생한 데이터로 확인되었다. 또한 Sobazaar데이터 세트에서 대부분 'buy_clicked'가 발생한 데이터로 확인되었다.

본 논문의 실험에서는 임계값을 0.5를 사용하여 2개 이상의 사용자 행동들이 존재하는 아이템을 사용자가 선호한다고 판단하며, 인터랙션 데이터를 1로 설정하여 추천 모델을 학습시키는 것을 목표로 했다.

4.4 성능 평가

추천 시스템 성능 평가는 Recall@K 방법과 NDCG@K 방법을 적용한다. Recall@K는 상위 K개의 재현율을 성능 평가 지표로 사용한다. 재현율이란 Table 7과 같이 혼동 행렬(confusion matrix)과 Equation (3)을 사용하여 사용자가 추천 시스템이 추천한 K개의 아이템 중에 실제 소비한 아이템을 비교하여 평가한다.

NDCG@K는 추천 시스템이 추천한 아이템의 순서가 사용자의 소비 순서와 일치하는지 평가한다. 추천 시스템이 K개의

Table 6. Percentage of Interaction Data Based on Threshold (%)

(A) Beibei dataset					
	0.1	0.3	0.5	0.7	0.9
User1	100	87.0	64.7	45.2	26.0
User2	96.1	94.5	68.8	47.5	41.3
User3	94.7	85.3	70.2	55.2	31.3

(B) Sobazaar dataset					
	0.1	0.3	0.5	0.7	0.9
User1	93.0	88.2	75.7	61.8	48.1
User2	100	87.3	67.8	58.1	49.7
User3	97.5	90.1	69.6	61.6	51.4

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

- TP : 추천 시스템이 추천한 아이템을 사용자가 실제로 소비함
- FN : 추천 시스템이 추천하지 않은 아이템을 사용자가 소비함

$$DCG = \sum_{i=1}^n \tag{4}$$

$$NDCG = \frac{DCG}{IDCG} \tag{5}$$

Table 7. Confusion Matrix

		Actual values	
		True	False
Predictive values	True	TP (True positive)	FP (False positive)
	False	FN (False Negative)	TN (True Negative)

아이템을 추천했을 때, 추천된 아이템이 순위별로 가중치를 주어 사용자가 실제 소비한 순서와 일치할 경우 높은 평가 점수를 획득하게 되는 방법이다.

따라서 NDCG 평가 방법을 사용하여 추천 순서를 반영한 결과를 Equation (4)와 (5)를 사용하여 평가한다. Relevance는 관련성 점수로 사용자와 아이템이 얼마나 관련 있는지를 나타내는 수치이며 Equation (4)인 DCG(Discout cumulative gain)를 보면 순위가 낮아질수록 을 통해 평가지표에 미치는 영향이 줄어들게 된다. 즉, 높은 순위의 아이템을 잘 추천하는 것이 낮은 순위의 아이템을 잘 추천해 주는 것보다 평가에 중요한 요소가 된다. 이러한 DCG를 정규화(Normalized)한 것을 NDCG라고 부르며 Equation (5)와 같다. NDCG는 가장 이상적인 값(Ideal DCG)으로 실제 NCG값을 나눠 0~1 사이 값으로 표현한다.

Table 8의 결과를 통해 기존 추천 시스템들보다 재현율이 높은 것을 확인할 수 있었으며, 추천 아이템 순위 예측도 성능이 개선되었다. 이는 논문에서 제안하는 가중치를 사용하여 사용자의 선호도를 측정하고 구매하지 않은 아이템 중에서 사용자가 선호하는 아이템과 사용자가 선호하지 않는 아이템을 구분해서 추천 모델을 학습했기 때문이다.

사용자 행동의 개수가 3개인 (A) Beibei 데이터 세트와 7개인 (B) Sobazaar 데이터 세트의 결과를 비교해 보면 행동의 수가 많아질수록 대부분 추천 성능이 낮아졌지만 제안하는 가중치 기반 추천 시스템에서는 50개의 아이템을 추천했을 경우의 성능이 유지되었다. 이를 통해 제안하는 시스템은 안정

Table 8. Performance Comparison
(A) Beibei dataset

	BPRH	EHCF	NCF	Proposed system
Recall@10	0.0512	0.1006	0.1668	0.1706
Recall@50	0.2528	0.3422	0.5125	0.5221
NDCG@10	0.0203	0.0623	0.0742	0.0833
NDCG@50	0.0724	0.1242	0.1382	0.1492

(B) Sobazaar dataset

	BPRH	EHCF	NCF	Proposed system
Recall@10	0.0324	0.0635	0.1245	0.1426
Recall@50	0.2417	0.3425	0.5116	0.5282
NDCG@10	0.0245	0.0537	0.0604	0.0732
NDCG@50	0.0682	0.1294	0.1409	0.1538

성 부분에서 실효성이 있다고 판단하였다. 이는 NCF와 성능 차이는 크지 않았지만 사용자의 단일 행동에 대한 결괏값이 아니라 다양한 행동들을 포함한 데이터라는 점에서 그 결괏값이 유의미하였다.

이는 제안하는 가중치 기반 추천 시스템의 데이터들 사이의 시간 순서를 제외하고 사용자 행동들 사이 상관관계 분석 기반으로 생성된 가중치가 행동의 종류가 많아져도 효율적으로 가중치를 생성 및 관리할 수 있으며 다양한 행동들을 분석하여 사용할 수 있다는 것을 의미한다.

5. 결 론

본 논문에서는 사용자 행동 데이터 상관관계 분석을 통해 사용자와 아이템 간 선호도를 예측한 추천 시스템을 제안한다. 제안하는 시스템에서는 상관관계 분석을 통한 행동 데이터별 가중치를 생성하였고 이 가중치를 기반으로 추천 모델을 학습하였다. 목표 데이터와 보조 데이터의 독립적인 상관관계 분석을 통해 계산의 복잡도를 낮추었으며, 추천 성능을 높였다. 데이터 분석을 통해 생성된 가중치는 보조 행동들 중에 더 중요한 행동을 구분할 수 있으며 사용자의 선호 아이템을 예측하여 추천하기 위해 사용된다. 이러한 이기종 암시적 피드백 데이터를 이용한 가중치 기반 추천 시스템은 다른 추천 시스템과 비교했을 때 개선된 성능을 보인다. 본 연구에서 실험에 사용한 데이터 셋은 3가지와 9가지의 사용자 행동만을 포함하지만, 실제 암시적 데이터는 그 이상의 행동 종류를 포함할 수 있다.

본 논문에서는 사용자 행동의 수에 제약을 두지 않았으므로 실험에서 사용한 행동 수보다 많은 종류에 대해 적용이 가능할 것이라고 기대한다. 다만 행동의 수가 많아지면 그에 따른 복잡도가 증가할 것이므로 처리량과 더불어 유의미한 관계 도출을 고려하기 위한 추가 연구가 필요하다. 또한 인터랙션 데이터를 생성하기 위한 임곗값 설정은 본 연구에서 진행한 것과 같이 실험 결과를 통해 최적의 임곗값을 사용하였지만 자동적으로 최적의 임곗값을 찾기 위한 방법이나 시나리오에 대한 추가 연구가 필요하다.

References

- [1] D. Jannach, A. Manzoor, W. Cai, and L. Chen, "A survey on conversational recommender systems," *ACM Computing Survey*, Vol.54, No.105, pp.1-36, 2021.
- [2] S. Milano, M. Taddeo, and L. Floridi, "Recommender systems and their ethical challenges," *Ai & Society*, Vol.35, pp.957-967, 2020.
- [3] P. Convington, J. Adams, and E. Sargin, "Deep neural networks for YouTube recommendations," *Proceedings of the 10th ACM Conference on Recommender Systems*, pp.191-198, 2016.
- [4] C. A. Gomez-Urbe and N. Hunt, "The netflix recommender system: Algorithms, business value, and innovation," *ACM Transactions on Management Information Systems*, Vol.6, No.13, pp.1-19, 2015.
- [5] H. Cheng et al., "Wide & deep learning for recommender systems," *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, pp.7-10, 2016.
- [6] P. Nagarnaik and A. Thomas, "Survey on recommendation system methods," *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, 2015.
- [7] Y. Koren, "Recommender system utilizing collaborative filtering combining explicit and implicit feedback with both neighborhood and latent factor models," US patents, 2011.
- [8] D. Yang, T. Chen, W. Zhang, Q. Lu, and Y. Yu, "Local implicit feedback mining for music recommendation," *Proceedings of the sixth ACM Conference on Recommender Systems*, pp.91-98, 2012.
- [9] W. Pan, S. Xia, Z. Liu, X. peng, Z. Ming, "Mixed factorization for collaborative recommendation with heterogeneous explicit feedbacks," *Information Sciences*, Vol.332, pp.84-93, 2016.
- [10] C. Zhang and C. Li, "Neural collaborative filtering recommendation algorithm based on popularity feature," *Culture-oriented Science & Technology (ICCST), International Conference on*, 2021.

- [11] H. Ko, S. Lee, Y. Park, and A. Choi, "A survey of recommendation systems: Recommendation models, techniques, and application fields," *Electronics*, Vol.11, No.1, pp.141, 2022.
- [12] S. Sidana, M. Trofimov, O. Horodnytskyi, Y. Maximov, and M. R. Amini, "User preference and embedding learning with implicit feedback for recommender systems," *Data Mining and Knowledge Discovery*, Vol.35, pp.568-592, 2021.
- [13] X. Chen, L. Li, W. Pan, and Z. Ming, "A survey on heterogeneous one-class collaborative filtering," *ACM Transactions on Information Systems*, Vol.38, No.35, pp.1-54, 2020.
- [14] E. R. N. Valdez, D. Quintana, R. G. Crespo, P. Isasi, and E. H. Viedma, "A recommender system based on implicit feedback for selective dissemination of ebooks," *Information Sciences*, Vol.467, pp.87-98, 2018.
- [15] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in Artificial Intelligence*, 2009.
- [16] Y. Koren, S. Rendle, and R. Bell, "Advances in collaborative filtering," *Recommender Systems Handbook*, pp.91-142, 2021.
- [17] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," *2008 Eighth IEEE International Conference on Data Mining*, 2008.
- [18] X. He, L. Liao, H. Zhang, L. Nie, and X. Hu, "Neural collaborative filtering," *WWW '17: Proceedings of the 26th International Conference on World Wide Web*, pp.173-182, 2017.
- [19] Y. Li, S. Wang, Q. Pan, H. Peng, T. Yang, and E. Cambria, "Learning binary codes with neural collaborative filtering for efficient recommendation systems," *Knowledge-Based Systems*, Vol.172, pp.64-75, 2019.
- [20] S. Rendle, W. Krichence, L. Zhang, and J. Anderson, "Neural collaborative filtering vs. matrix factorization revisited," *Proceedings of the 14th ACM Conference on Recommender Systems*, pp.240-248, 2020.
- [21] S. Rendle, C. Freudenthaler, Z. Gantner, and L. S. Thieme, "BPR: Bayesian personalized ranking from implicit feedback," *arxiv:1205.2618*, 2012.
- [22] W. Pan, H. Zhong, C. Xu, and Z. Ming, "Adaptive Bayesian personalized ranking for heterogeneous implicit feedbacks," *Knowledge-Based Systems*, Vol.73, pp.173-180, 2015.
- [23] H. Qiu, Y. Liu, G. Guo, Z. Sun, J. Zhang, and H. T. Nguyen, "BPRH: Bayesian personalized ranking for heterogeneous implicit feedback," *Information Sciences*, Vol.453, pp. 80-98, 2018.
- [24] C. Chen, M. Zhang, Y. Zhang, W. Ma, and S. Ma, "Efficient heterogeneous collaborative filtering without negative sampling for recommendation," *AAAI-20 Technical*, Vol.34, 2020.
- [25] S. Debnath, N. Ganguly, and P. Mitra, "Feature weighting in content based recommendation system using social network analysis," *WWW '08: Proceedings of the 17th international conference on World Wide Web*, pp.1041-1042, 2008.
- [26] B. Wang, Q. Liao, and C. Zhang, "Weight based KNN recommender system," *2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2013.
- [27] T. Himel, M. N. Uddin, M. A. Hossain, and Y. M. Jang, "Weight based movie recommendation system using K-means algorithm," *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, 2017.
- [28] Z. Cui et al., "Personalized recommendation system based on collaborative filtering for IoT scenarios," *IEEE Transactions on Services Computing*, Vol.13, 2020.
- [29] U. Bhimavarapu, N. Chinalapudi, and G. Battineni, "A fair and safe usage drug recommendation system in medical emergencies by a stacked ANN," *Algorithms*, Vol.15, No.6, pp.186, 2022.
- [30] L. Jiang, Y. Cheng, L. Yang, J. Li, H. Yan, and X. Wang, "A trust-based collaborative filtering algorithm for E-commerce recommendation system," *Journal of Ambient Intelligence and Humanized Computing*, Vol.10, pp.3023-3034, 2019.
- [31] Y. Guo, M. Wang, and X. Li, "An interactive personalized recommendation system using the hybrid algorithm model," *Symmetry*, Vol.9, No.10, pp.216, 2017.
- [32] Beibei Dataset [Internet], <https://github.com/dingjingtao/NegativeSamplerBPR/tree/master/BPRplusView/data/beibei>
- [33] Sobazaar Dataset [Internet], <https://github.com/hainguyen-telenor/Learning-to-rank-from-implicit-feedback/tree/master/Data>



박요한

<https://orcid.org/0009-0009-5769-4617>
 e-mail : imjin3027@soongsil.ac.kr
 2021년 국가평생교육진흥원
 멀티미디어학과(학사)
 2023년 송실대학교 컴퓨터공학과(석사)
 2023년~현 재 송실대학교 컴퓨터학과
 박사과정

관심분야 : 빅데이터 처리, MLOps, Recommender System



문 종 혁

<https://orcid.org/0000-0002-4919-0058>
e-mail : jonghyeokmun@soongsil.ac.kr
2017년 한경대학교 컴퓨터공학과(학사)
2019년 송실대학교 컴퓨터학과(석사)
2019년~현 재 송실대학교 컴퓨터학과
박사과정

관심분야: 로봇 미들웨어, 빅데이터 처리, MLOps, XAI



최 종 선

<https://orcid.org/0000-0001-9648-0667>
e-mail : jongsun.choi@ssu.ac.kr
2000년 송실대학교 컴퓨터학부(학사)
2002년 송실대학교 컴퓨터학부(석사)
2008년~2010년 유한대학교
e-비즈니스과 전임교원

2010년 송실대학교 컴퓨터학부(박사)
2011년~2012년 송실대학교 지능형로봇연구소 연구원
2012년~2013년 서일대학교 인터넷정보과 전임교원
2013년~현 재 송실대학교 컴퓨터학부 부교수
관심분야: 로봇 소프트웨어 플랫폼, 클라우드 컴퓨팅



최 재 영

<https://orcid.org/0000-0002-7321-9682>
e-mail : choi@ssu.ac.kr
1984년 서울대학교 제어계측공학과(학사)
1986년 미국 남가주대학교 전기공학과
(컴퓨터공학)(석사)
1991년 미국 코넬대학교 전기공학부
(컴퓨터공학)(박사)

1992년~1994년 미국 국립오크리지연구소 연구원
1994년~1995년 미국 테네시 주립대학교 연구교수
1995년~현 재 송실대학교 컴퓨터학부 교수
관심분야: 로봇 미들웨어, 시스템소프트웨어, 병렬/분산처리,
고성능컴퓨팅