

사용자 구매 내역을 활용한 딥러닝 기반 추천 시스템

오홍석*, 정재민**¹, 남윤중***, 서영덕****, 이의종**²

Deep Learning-based Recommendation System using User Purchase History

Hong-Seok Oh*, Jae-Min Jeong**¹, Yoon-Joong Nam***, Young-Duk Seo****, and Eui-Jong Lee**²

이 논문은 충북대학교 국립대학육성사업(2020)지원을 받아 작성되었음

요 약

오늘날 현대 사회에서 추천 시스템은 중요한 경쟁력의 요소로 자리를 잡았으며, 추천 시스템과 관련된 활용 및 연구가 활발하게 진행되고 있다. 특히 딥러닝을 활용한 맞춤형 추천 시스템은 전통적인 방식의 추천 시스템보다 향상된 결과를 제공하며 다양한 분야에서 편의성을 제공하고 있다. 따라서 본 연구는 사용자와 상품의 특징을 활용한 딥러닝 기반 추천 시스템 제안을 통해 다양한 분야에서 마케팅 및 추천 서비스 전략 수립 시 활용 가능한 솔루션을 제공하고자 한다. 이를 통해 다양한 분석 결과 및 유의미한 추천 결과를 도출하고자 한다. 연구 결과로 딥러닝을 활용한 데이터 분석 시 상품에 대한 사용자 특징 학습 및 다양한 선호도 추론이 가능함을 확인하였다.

Abstract

Recently, there has been an increase in studies related to recommendation systems in various fields such as video on demand (VOD), Internet protocol television (IPTV), and e-commerce. Also, various methods (e.g., collaborator filter, support vector machine, and k-means clustering) have been applied to the recommendation systems. In this paper, we proposed a deep neural network-based recommendation system for generating various recommendation results with purchase history. We performed experiments to show the effectiveness of the proposed system using industrial data. The experimental results show that the deep learning-based system can generate recommendation results with various points of view.

Keywords

recommendation system, artificial neural network, deep neural network, big data

* 충북대학교 컴퓨터과학과

- ORCID: <https://orcid.org/0000-0002-3229-1821>

** 충북대학교 소프트웨어학부(**² 교신저자)

- ORCID¹: <https://orcid.org/0000-0002-7308-7392>

- ORCID²: <https://orcid.org/0000-0003-2229-3898>

*** THINGS

- ORCID: <https://orcid.org/0000-0001-5194-1415>

**** 인하대학교 컴퓨터공학과

- ORCID: <https://orcid.org/0000-0001-8542-2058>

• Received: May 24, 2022, Revised: Jun. 21, 2022, Accepted: Jun. 24, 2022

• Corresponding Author: Euijong Lee

School of Computer Science, Chungbuk National University, Republic of Korea

Tel.: +82-43-261-3133, Email: kongjjagae@cbnu.ac.kr

I. 서론

빅데이터는 마케팅, 의료, 사물인터넷(IoT, Internet of Things)와 같이 오늘날의 핵심 비즈니스 분야에 활용되고 있다[1]. 데이터 또한 특정 분야의 데이터 뿐만 아닌 여러 분야의 데이터를 융합하여 현대 사회와 사용자에게 편의성을 제공하고 있다[2]. 빅데이터는 매 순간 대량으로 생성되지만 정제되지 않은 데이터를 유의미한 값으로 활용하기 위해 시간, 컴퓨터 자원, 인력을 소모하여 데이터 전처리 작업이 요구된다[3]. 따라서 효율적으로 빅데이터를 분석하고 활용할 수 있도록 다양한 방법론을 연구 및 개발되고 있다[4][5].

이들 중 많은 관심과 연구가 이루어지는 분야는 인공지능(AI, Artificial Intelligence)을 활용한 데이터 분석이다[6]-[8]. 빅데이터와 인공지능은 상호 보완적인 관계로 빅데이터는 데이터를 활용할 수 있게 제공하고 AI는 방대한 양의 데이터를 학습하여 의미 있는 결과를 도출한다[9]. 이러한 빅데이터와 AI의 상호 보완적인 관계 속에서 두 가지 기술을 융합한 가치 창출 또한 증가하고 있다[10][11]. 특히 추천 시스템의 경우 딥러닝(Deep learning)[12]을 활용한 수익모델 및 마케팅 등 다양한 서비스가 출시되고 있으며 관련 시장 및 연구 또한 증가하고 있다[13][14]. 딥러닝을 활용할 경우 기존의 추천 시스템 보다 개선된 추천 결과를 보였으며[15] 산업체에서 딥러닝 기반의 추천 시스템을 활용할 경우 개선된 추천 결과가 도출된 연구 또한 존재한다[16]. 앞선 결과를 통해 딥러닝을 활용한 추천 시스템은 매우 중요한 경쟁력의 요소로 자리 잡게 되었다[17].

본 연구는 기존 연구[18]를 토대로 사용자와 상품의 특징을 조합하고 딥러닝을 활용한 추천 시스템을 제안한다. 앞선 연구는 전체 상품특징을 활용하여 상품과 사용자 사이의 연관성을 확인하였다. 본 연구는 추가적인 모델 구성 및 학습을 바탕으로 조합된 상품의 특징을 입력값으로 활용하여 사용자와 상품의 관계를 파악하였다. 이와 더불어 사용자 집단 간 상품데이터를 딥러닝을 활용하여 분석하고자 한다. 따라서 학습된 추천 모델을 활용하여 상품과 사용자 간 연관성 파악을 통해 유의미한 결과를

도출할 수 있음을 보인다.

본 논문의 2장 관련 연구는 추천 시스템 동향 및 이를 활용한 다양한 연구를 소개하고 기존 연구와 본 연구와의 차별성을 소개한다. 3장 딥러닝 기반 추천 시스템은 본 연구에서 제안한 시스템 및 학습 모델을 소개한다. 4장 실험결과는 제안한 시스템에서 활용한 데이터 소개 및 분석, 전처리와 모델을 통한 실험결과를 제시한다. 5장 결론을 통해 실험결과와 한계점을 서술하고 향후 이러한 점을 보완하여 향후 진행할 연구를 소개한다.

II. 관련 연구

본 장은 추천 시스템에 관련된 최신 연구를 소개한다. [19]는 인공지능을 사용한 개인의 재정 관리 시뮬레이터 시스템을 제안한 논문이다. 해당 연구에서는 재정과 관련된 문제 및 요구사항이 증가하고 있으며 개인 재무 관리의 중요성을 강조하였다. 또한, 전통적인 방식의 개인 재정 관리의 다양한 방법론과 한계점을 제시하였으며 최근 인공지능을 활용한 금융 관련 연구를 시사하였다. 따라서 사용자의 경제관념 확립을 위한 재정 관리 시뮬레이터를 제안하였다. 웹 서비스의 형태로 제공되는 재정 관리 시뮬레이터는 사용자의 자산과 관련된 개인 정보인 평균 연봉과 대출, 계좌, 환율 및 투자, 예금 적금에 대한 정보를 수집한다. 재정 관리를 위한 머신러닝 모델은 강화학습[20]의 한 분야인 Q-Learning[21]을 활용하였다. 모델은 사용자의 자산관리와 관련된 활동에 대한 예상 가능성을 보여주며 추천 서비스의 형태로 사용자에게 전략을 제시한다. 이를 활용해 사용자는 시뮬레이션을 활용하여 가상의 자산을 관리하는 연구가 수행되었다.

[22]는 추천 시스템에 대한 중요도가 증가함에 따라 Word2Vec의 다차원 특징을 이용한 머신러닝 기반 추천 정확도 개선을 제안하였다. 해당 연구에서는 Word2Vec 방식 중 Skip-gram 방식을 활용하여 단어 간의 연관성을 찾고 단어 데이터들을 벡터 상의 수치로 표현한다. 이를 통해 중심 아이템을 기준으로 이전과 이후에 클릭한 아이템의 연관성을 찾아낸다. 구매할 아이템을 예측하고 각각으로 나는

단어를 하나의 고유 아이템으로 인식 및 학습을 진행 하였다. 이를 통해 Word2Vec를 활용하기 전보다 높은 추천 정확도를 보임을 확인하였다. 또한, Word2Vec의 벡터의 차원이 증가할수록 추천의 정확도 또한 증가하였음을 보였다.

[23]은 추천 시스템 활용 시 사용자 선호도를 반영한 통합된 데이터 세트를 제안하였다. 연구진은 오늘날 추천 시스템의 중요성과 이를 위한 정제된 데이터의 활용을 역설하였으며 기존 정형 데이터와 비정형 데이터의 특징 및 각각의 데이터의 단점을 보완하기 위한 두 데이터를 통합한 데이터 모델을 제안하였다. 모델은 실제 산업체의 고객의 리뷰 및 평점, 장바구니 구매 데이터를 활용하여 데이터 모델을 구성하였다. 이를 통해 기존 정형 데이터와 연구진이 제안한 데이터 세트의 비교를 통해 추천 시스템에 적용 시 성능개선을 목표로 실험을 구성하였다. RMSE(Root Mean Square Deviation)와 MAE (Mean Absolute Error)를 활용한 성능평가 결과는 전체 알고리즘에서 통합된 데이터 세트를 활용할 경우 추천 시스템의 성능이 개선되었음을 보인다.

[24]는 딥러닝 기반 뉴스 추천시스템의 중요성과 현재 기술 동향을 분석하였다. 최근 뉴스 포털의 이용이 많아짐에 따른 뉴스 추천시스템의 중요성을 강조하며 뉴스 추천시스템 구성 시 고려 사항을 서술하였다. 이를 기반으로 현재 뉴스 추천 시스템에서 사용되는 딥러닝 기법을 소개하며 현재 시스템의 문제 및 이를 보완한 딥러닝 기반 뉴스 추천 시스템의 연구 동향을 소개하였다.

[25]는 고객의 검색 패턴과 상품 상세정보를 활용한 개선된 상품 추천 모델을 제안하였다. 해당 논문은 정형 데이터만을 사용하는 전통적인 협업 필터링 방식에 고객의 리뷰 및 검색 키워드 등의 비정형 데이터를 활용하여 추천 시스템을 제안하였다. 이는 최근 비정형 데이터를 활용한 추천 시스템 연구가 증가하고 있으며 비정형 데이터를 통해 정형 데이터를 활용한 전통적인 문제점인 콜드 스타트 (Cold start)와 데이터 희소성을 해결하고 추천 시스템의 성능을 개선하였다. 해당 시스템은 Doc2vec를 활용하여 문장 내부의 단어 간 연관성을 파악하는 방식으로 비정형 데이터를 추출하였으며, 고객의 검색 키워드 및 상품 상세정보를 활용하여 실험을 진

행하였다. 이를 통해 기존 추천 시스템보다 전반적으로 높은 추천 성능을 보였으며 추천 상품의 수가 일정 개수로 증가할수록 구매 성능이 더욱 개선되었음을 시사하였다.

본 연구와 기존 연구와의 차이점은 데이터 추출 시 상품의 특성과 사용자의 특징을 조합하여 딥러닝을 활용한 추천 시스템을 제안하는 것이다. 더불어 이를 가시화하여 상품의 특정 특징과 사용자 간 관계 결과를 보여주는 것을 목표로 한다.

III. 딥러닝 기반 추천시스템

본 장은 연구에서 활용한 시스템 구조 및 신경망의 생성 과정 및 추천시스템을 설명한다. 1절 딥러닝 기반 추천시스템의 구조는 연구에서 활용한 전체적인 시스템 흐름도를 통해 딥러닝 기반 추천 시스템을 설명한다. 2절 신경망 구성 및 활용에서는 학습에서 활용한 신경망을 설명한다.

3.1 딥러닝 기반 추천시스템 구조

본 연구에서 제안하는 딥러닝 기반 추천 시스템의 구조는 그림 1과 같다. 시스템은 학습에 활용할 데이터와 관련된 데이터 모듈과 학습과 결과를 가시화하는 학습 및 추천 결과 모듈로 구성되어 있다. 데이터 모듈은 데이터 추출 및 전처리로 구성되어 있으며 학습 및 결과 모듈은 딥러닝 학습 모델과 추천 결과로 구성된다.

데이터 추출은 학습에서 활용되는 데이터를 추출하는 과정으로 본 연구는 사용자와 상품의 특징 데이터를 활용한다. 이는 상품과 사용자 특징 사이의 관계를 파악하여 연관성을 분석하기 위함이다. 데이터 전처리 과정은 데이터를 모델에 활용할 수 있도록 정제하는 과정으로 결측치 제거와 불균형 데이터 완화, 데이터 인코딩의 작업을 수행한다. 앞선 데이터 추출 및 전처리 작업을 통해 데이터를 도출한 후 이를 상품과 사용자 데이터로 구분하여 학습할 수 있는 형태로 분류한다. 모델과 추천의 과정은 앞서 추출된 데이터를 활용하는 단계로 신경망의 입력값에 따른 사용자 특징 데이터의 패턴을 학습하는 과정으로 구성된다. 이는 제품 특징과 사용자

의 특징 사이의 연관 관계를 파악하기 위함이다. 추천 결과는 학습된 신경망을 기반으로 사용자와 상품의 특징 관계를 백분율로 나타내며 이후 학습의 결과는 원그래프를 활용한 시각화를 통해 특징 간 연관 정도를 결과로 보여준다.

3.2 신경망 구성 및 활용

본 연구에서 활용한 신경망의 구조는 그림 2와 같다. 본 연구의 목적은 상품과 사용자 특징 간 연관 관계를 파악하는 것을 목표로한다. 따라서 두 데이터 사이의 연관 관계를 학습하기 위해 심층 신경망(DNN, Deep Neural Network)[25] 을 활용하여 모델을 구성하였다.

상품의 특징 데이터는 모델의 입력층에 활용된다. 사용자의 데이터는 입력된 상품데이터에 따른 사용자 특징을 도출하기 위해 학습 시 활용되며 출력값은 사용자의 특징을 사용한다. 따라서 데이터의 특징상 독립변수인 상품을 통해 사용자의 특징 변수와 관계를 확인하기 위해 학습을 진행한다. 신경망 모델의 은닉층(Hidden layer)[27]는 아래와 같이 구성하였다.

- 노드 개수 : 30개
- 배치 정규화(Batch normalization)[28]
- ReLU[29]
- Dropout[30]
- SoftMax[31]

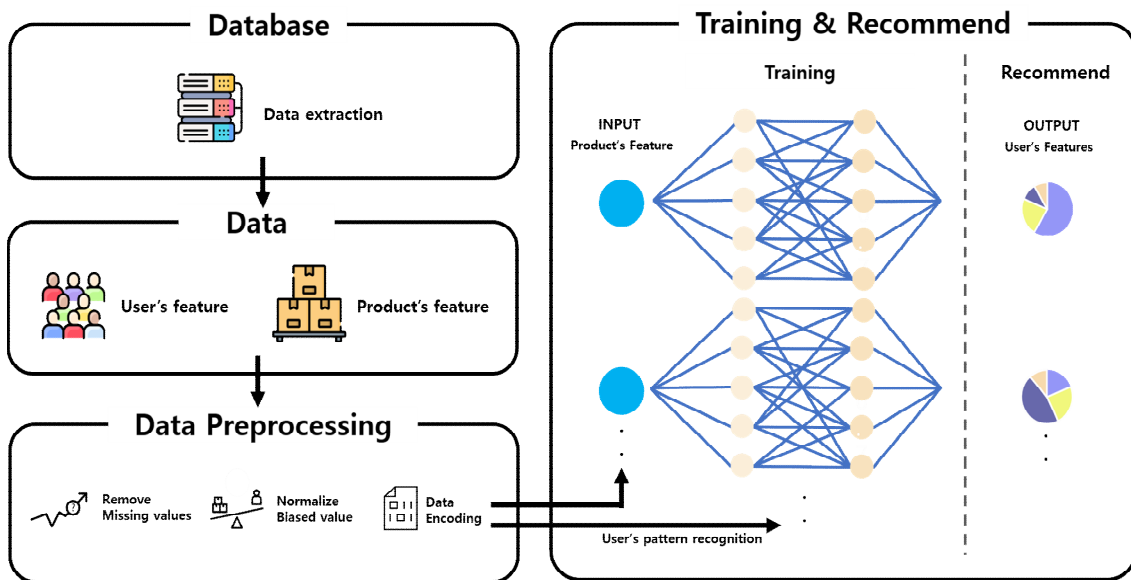


그림 1. 딥러닝 기반 추천 시스템 구조

Fig. 1. Deep learning based recommendation system structure

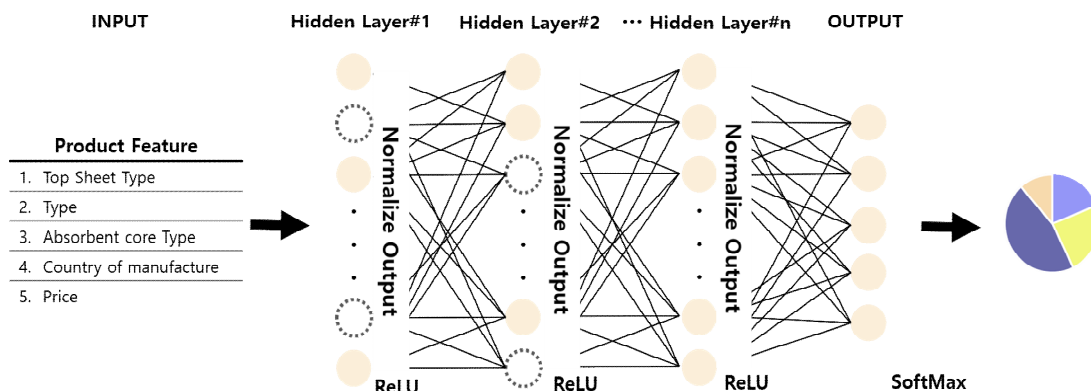


그림 2. 딥러닝 기반 추천시스템 모델 구성도

Fig. 2. Model structure of deep learning based recommendation system

추천 결과는 앞선 모델의 출력 결과를 가시화하는 과정으로 원그래프를 활용하여 보여준다. 제안하는 시스템의 추천 결과의 예시는 그림 3과 같다. 그림 3은 학습된 신경망에 흡수체(제품 특징) 데이터를 입력하여 생리 냄새(사용자 특징) 데이터를 입력한 결과이며 학습결과를 통해 고분자 흡수체를 사용한 사용자의 경우 생리 냄새 약함(54.5%)과 연관성이 가장 높음을 알 수 있다.

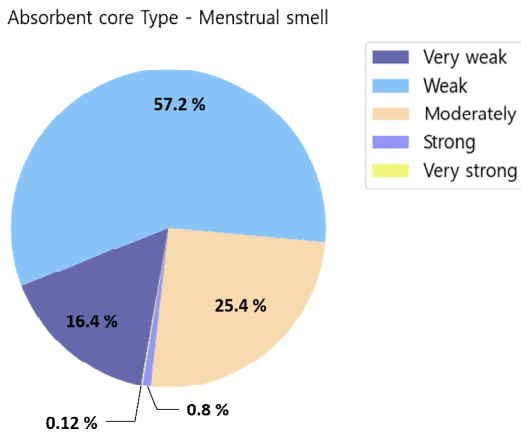


그림 3. 흡수체(고분자 흡수체)- 냄새 학습결과 예시
Fig. 3. Example of learning result(Absorbent core - smell)

IV. 실험 결과

본 장은 1절 데이터 구성을 통해 본 실험에서 활용한 데이터 세트의 개요 및 특징을 설명한다.

2절 데이터 추출은 앞선 데이터 구성을 기반으로 추출한 데이터 세트를 설명한다. 3절 데이터 전처리는 앞서 추출한 데이터 세트를 실험에서 활용하기 위한 전처리 과정 및 결과를 설명한다. 4절 실험을 위한 DNN 최적화 과정은 모델의 파라미터 값 조절을 통한 최적의 모델을 구성하기 위한 모델 전처리를 설명한다. 5절 실험결과를 통해 도출한 결과를 설명 및 분석한다.

4.1 데이터 구성

본 연구에서 활용된 데이터는 실제 산업체에서 활용되고 있는 데이터를 활용하였다. 여성 헬스케어 기업인 ‘먼슬리썸’에서 제공한 2018년부터 2022년까지 생리대의 특징과 기업에서 제공하는 서비스에 입력된 약 100,000명의 완벽하게 익명화된 사용자의 생리 특징 데이터를 사용하였다. 제공받은 사용자 특징 데이터는 표 1과 같으며 각각의 특징별 세부 데이터를 입력할 수 있다.

세부 데이터는 정도를 나타내는 데이터와 분류를 선택하는 정형적 형태의 데이터로 구성되어 있다. 예를 들어 생리통 생리량, 생리 냄새의 경우 1~5단계의 정도를 나타내고 신체 활동과 실내 활동은 자전거 타기, TV 보기, 헬스 등 선택할 수 있는 다양한 활동으로 구성되어 있어 각 특징에 대해 정도 혹은 활동의 분류를 선택할 수 있다.

표 1. 상품 및 사용자 특징

Table 1. Product and user features

User and product feature		
Type of feature	Describe	
User feature	Amount of menstruation	Very little, Few, Moderately, Many, Very many
	Menstrual smell	Very weak, weak, Moderately, Strong, Very strong,
	Menstrual pain	No pain, Slightly painful, Moderately, Painful, Very painful
	Outdoor activity	Running, Weigh training Dancing, Ride a bike, Yoga
	Indoor activity	Watching TV, Reading a book, Watching video clip, Caring for children, Phone calls to friends, Study
Product feature	Type	General type, disposable sanitary PAD, applicator, disposable liner
	Topsheet type	Organic cotton, Regenerated fiber, Cotton texture, Silk texture
	Absorbent core type	Natural absorbent, High polymer absorbent, Absorbers other than nature, Organic absorbent
	Country of manufacture	South Korea, China, Finland, Italy, Slovenian, Denmark
	Price	More than 0 ~ Less than 2500, More than 2500 ~ Less than 3000, More than 3000 ~ Less than 5500, More than 5500

상품 특징의 데이터는 생리대에 대한 데이터로 총 5가지로 구성되어 있으며, 제공받은 데이터는 표 1과 같다. 각각의 상품 특징은 세부적으로 분류되어 있으며 분류 번호를 통해 구성되어 있다. 예를 들어 탐시트의 경우 종류에 따라 순면감촉, 유기농 순면, 재생섬유, 등이 있으며 각각의 고유한 번호로 저장되어 있다.

4.2 데이터 추출

앞선 데이터 구성을 기준으로 연구에서 제안한 모델의 학습에 활용하기 위해 상품의 특징, 상품과 사용자의 특징 데이터를 추출했다. 모델 적용 및 연구 가설을 증명하고 연관성을 파악하기 위한 데이터는 다음 제시된 조건으로 추출하였다.

- 1) 특징이 입력된 사용자
- 2) 1번 조건을 만족하는 사용자 중 구매한 생리대(상품)의 특징

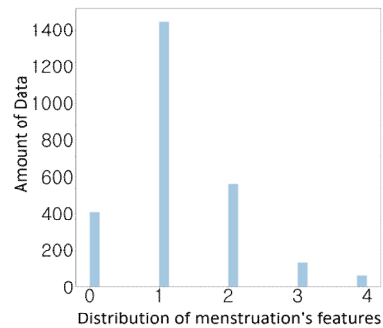
조건 1번의 입력된 사용자 특징은 ‘먼슬리썩’에서 제공하는 애플리케이션 서비스에 입력된 사용자 데이터를 활용하였다. 연구에 활용한 사용자 특징의 종류는 총 5가지이며 선정 근거는 여성 생리케어용품 사용평가를 기준으로 사용자의 특성을 선정하였다. 기존 연구[31][32]에 따르면 사용자가 생리 케어용품을 구매할 경우 고려하는 주요 요인으로 생리대 착용 시 활동성과 생리량을 상품 선정 시 중요한 요소로 고려하였으며, 냄새와 흡수율의 요인 또한 생리대의 주요선택 요소로 고려되었음을 알 수 있었다. 따라서 본 연구에 활용할 사용자의 특성으로 생리 냄새, 생리량, 생리통, 신체 활동, 실내활동을 사용자의 5가지 생리 특성으로 선정하였다. 조건 2번의 구매한 상품 특성의 경우 업체의 전자상거래 서비스에서 판매되는 상품 중 생리케어용품들의 특징들을 선정 및 활용하였다. 업체에서 제공하는 상품의 특징은 총 5가지로 유형, 탐시트, 흡수체, 제조 국가, 가격을 학습에 활용 하였다.

4.3 데이터 전처리

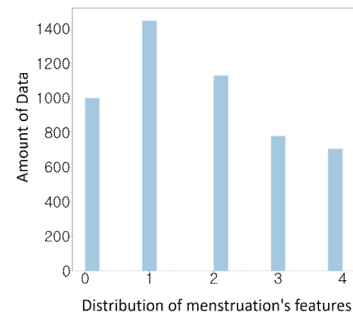
앞선 조건을 통해 데이터를 추출 후 사용자 데이터에서 두 가지 특징이 발견되어 이에 대한 전처리를 진행했다.

첫 번째 특징으로 사용자 특성을 모두 입력한 것이 아닌 부분 입력을 한 사용자가 많아 결측치 데이터가 있음을 확인하였다. 다른 특징으로는 사용자의 다섯 가지의 특징이 그림 4의 1번 분포 그래프와 같이 데이터의 분포가 특정한 값으로 편향된 것을 확인했다. 예를 들어 그림 4는 사용자 특성 중 냄새에 관한 데이터 분포 그래프로 1번 도표는 1차로 추출된 전처리 전 데이터의 분포특성이다. 추출된 데이터의 분포는 1번 특징(냄새 없음)으로 데이터가 편향되어 있음을 확인하였고, 결측치 데이터를 제거하는 과정과 데이터 분포를 고르게 만들기 위해 기존의 비율에 맞추어 다른 특성 데이터를 무작위 복사하여 데이터를 추가하는 작업을 진행했다 (그림 4의 2번 도표 참고).

앞선 데이터 전처리 과정을 통해 특징을 입력한 사용자 중 상품을 구매한 사용자는 약 1,400명이 도출되었으며, 해당 사용자가 구매한 생리대의 개수로는 약 4,400개의 데이터를 추출, 이를 신경망 학습에 활용하였다.



(a) ① 타입 냄새 전처리 전
(a) Before pre-processing



(b) ② 타입 냄새 전처리 후
(b) After pre-processing

그림 4. 타입 - 냄새 데이터 분포도
Fig. 4. Distribution of type - smell data

4.4 심층신경망 최적화

실험에 앞서 최적의 하이퍼 파라미터를 적용하기 위한 실험을 진행하였으며, 학습의 조건은 24x24 배치 사이즈와 러닝 레이트는 0.01로 적용하여 실험을 진행하였다. 옵티마이저의 경우 Adagrad[34], Adam[35], SGD[36], RMSprop[37]를 비교하여 모델에 적용하였다. 그림 5는 실험에서 사용한 옵티마이저 간 손실값 비교 그래프이다. 실험결과 Adagrad를 적용하여 학습을 진행할 때 다른 하이퍼 파라미터와 비교하여 손실값이 더 낮음을 확인하였고 실험 결과를 활용하여 모델학습에 적용하였다.

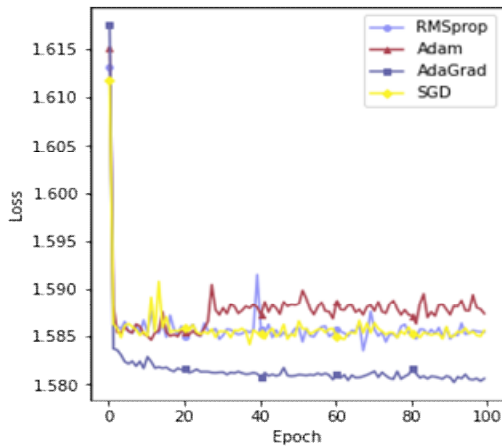


그림 5. 옵티마이저 간 Loss 그래프 비교
Fig. 5. Comparison of loss graphs between optimizers

4.5 실험 결과

학습된 내용을 기반으로 한 추천 시스템의 실험 결과는 다음과 같다. 상품 특징과 사용자 특징을 학습했을 때 도출되는 손실값 그래프는 그림 6이며, 각각의 그래프는 상품 특징과 사용자 특징을 학습한 결과이다. 해당 그래프의 추세선을 통해 특징별 학습이 적절하게 되었음을 알 수 있다. 또한, 특정 상품 특징뿐만이 아닌 전반적인 상품 특징에 대해서도 학습이 되었음을 확인할 수 있다.

본 연구에서 제안한 모델의 개인별 선호도 결과는 표 2와 같다. 실험결과 딥러닝을 활용한 데이터 분석 시 상품과 사용자 간 연관성 파악을 통한 다양한 선호도 추론이 가능함을 알 수 있었다. 예를 들어 표 2-1의 결과는 상품 특징인 탐시트를 입력값으로 사용자 특성(생리 냄새)과의 관계를 알아본 결과이다. 학습결과 재생섬유(탐시트)를 사용하는 사용자는 생리 냄새 아주 약함과 약함의 특징이 가장 많이 분포한 것을 유추할 수 있다. 표 2-2의 결과로 디지털형 타입의 생리대의 경우 생리량이 많은 경우 사용하지 않는 것을 유추할 수 있었다.

앞선 실험과 더불어 추천 결과 개선 및 사용자의 다양한 특징을 도출하기 위해 입력값의 특징 증가 [25]와 사용자 특징의 조합을 통한 실험을 진행하였다(표 2-3).

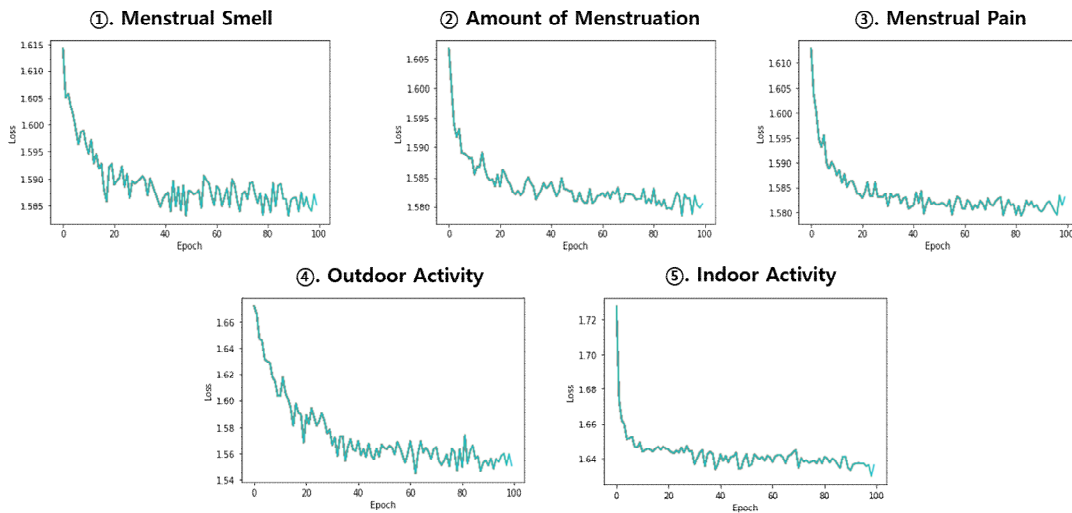


그림 6. 상품특징 - 사용자 특징 로스 그래프
Fig. 6. Loss graphs of product features and user features

표 2. 개인화 추천 결과

Table 2. Result of personalized recommendation

The result of personalized recommendation			
Input value	Topsheet type - Regenerated fiber	Input value	Type - digital
	<p>① Topsheet Type - Menstrual smell</p>		<p>② Type - Amount of Menstruation</p>
Input value	Type - normal	Input value	Type - normal, Absorbent core type - High polymer
	<p>③ Type - Menstrual smell</p>		<p>④ Type, Absorbent core type - Menstrual smell</p>
Input value	Type - applicator	Input value	Type - band
	<p>⑤. Activity(Indoor, Outdoor)</p>		<p>⑥. Activity(Indoor, Outdoor)</p>

결과는 일반형 생리대의 특징을 입력한 경우 냄새라는 사용자의 특징이 고루 분포되는 것을 볼 수 있다. 이는 일반형 타입의 생리대는 냄새라는 특징과 특별한 연관 관계가 없다고 유추할 수 있다. 하지만 일반형 생리대와 고분자 흡수체라는 특징을 함께한 결과(표 2-4)의 경우는 냄새가 강한 경우 사

용되지 않는 결과가 도출된다. 이는 일반형이며 고분자 흡수체의 특징을 가진 생리대의 경우 냄새가 적은 사용자가 사용한다는 결과를 도출해 낼 수 있다.

또한, 사용자 집단별 상품의 분포를 분석하기 위해 활동(신체 활동, 실내 활동)에 대한 신경망 학습을 진행하였으며 결과는 표 2-5, 2-6의 결과와 같다.

표 2-5, 표 2-6의 결과는 어플리케이션이라는 상품 특징은 활동적인 행동(헬스, 달리기 등)을 하는 경우 사용되는 것을 유추할 수 있다. 하지만 밴드형의 경우 실내활동(친구와 통화, 동영상 시청 등)의 정적인 활동과 연관 있는 것으로 결과가 도출된다. 따라서 딥러닝 기반 학습결과에 따라 생리대의 타입이라는 상품의 특징은 정적인 사용자와 동적인 사용자에게 따라 다르게 추천될 수 있음을 확인할 수 있다. 이를 기반으로 일반형, 일회용 생리대, 어플리케이션은 활동적인 신체 활동 시 선호하는 제품군으로 나왔으며 정적인 실내 활동 시 일회용 라이너, 디지털형, 밴드형의 제품군을 선호하는 것을 추가로 도출할 수 있었다.

본 연구 결과를 기반으로 딥러닝을 활용한 데이터 분석 결과는 상품 특성에 대한 사용자의 특징을 분석하는데 활용될 수 있음을 확인할 수 있었다.

V. 결 론

본 연구는 사용자 구매 내역을 활용한 딥러닝 기반 추천 시스템을 제안했고 이를 위해 제품의 특징과 사용자의 특징을 고려한 인공신경망을 설계했다. 본 연구의 실험 결과는 산업체의 데이터를 통해 진행되었다. 실험 결과를 통해 제품과 사용자의 특징의 연관 관계를 파악하고 다양한 분석 결과를 도출할 수 있음을 확인하였다. 이는 본 연구를 통해 기업 및 다양한 분야에서 비즈니스 모델 및 마케팅 전략을 세울 때 활용될 수 있을 것으로 기대된다.

본 연구는 딥러닝을 활용하여 상품과 사용자 데이터 사이의 연관성 파악과 이를 통한 사용자 집단 분석에서 활용 가능성을 보였다는 점에서 의의가 있다. 따라서 추후 연구에서는 본 연구를 바탕으로 추천 결과의 정확도 향상 및 사용자 간 관계 파악을 통한 상품과 상품 간 연관규칙 파악, 주기별 판매 상품 변화를 파악하여 개인 사용자를 위한 추천 시스템 연구를 진행해 보고자 한다.

References

- [1] Baehyun Kim, "Big Data Platform Case Analysis and Deployment Strategies to Revitalize the Data Economy", *Journal of Information and Security*, Vol. 21, No. 1, pp. 73-78, Mar. 2021. <https://doi.org/10.33778/kcsa.2021.21.1.073>.
- [2] Daniel E. O'Leary, "Artificial intelligence and big data", *IEEE intelligent systems*. Vol. 28, No. 2, pp. 96-99, Mar.-Apr. 2013. <https://doi.org/10.1109/MIS.2013.39>.
- [3] Qingchen Zhang, Laurence T. Yang, Zhikui Chen, and Peng Li, "A survey on deep learning for big data", *Information Fusion*, Vol. 42, pp. 146-157, Jul. 2018. <https://doi.org/10.1016/j.inffus.2017.10.006>.
- [4] Kibria and Mirza Golam, "Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks", *IEEE access*, Vol. 6, pp. 32328-32338, May 2018. <http://doi.org/10.1109/ACCESS.2018.2837692>.
- [5] Batko, Kornelia, and Andrzej Ślęzak, "The use of Big Data Analytics in healthcare", *Journal of big Data*, Vol. 9.1, pp. 1-24, Jan. 2022. <https://doi.org/10.1186/s40537-021-00553-4>.
- [6] Chen, Xue-Wen, and Xiaotong Lin, "Big data deep learning: challenges and perspectives", *IEEE access*, Vol. 2, pp. 514-525, May 2014. <http://doi.org/10.1109/ACCESS.2014.2325029>
- [7] Naumov and Maxim, "Deep learning recommendation model for personalization and recommendation systems", *arXiv preprint arXiv:1906.00091*, May 2019. <https://doi.org/10.48550/arXiv.1906.00091>.
- [8] Meng and Hongying, "Depression recognition based on dynamic facial and vocal expression features using partial least square regression", *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, pp. 21-30, Oct. 2013. <https://doi.org/10.1145/2512530.2512532>.
- [9] Leopold Schmetzing, "Trends in Artificial Intelligence and Big Data", *European Strategy and Policy Analysis System*, pp. 1-19, Jan. 2019.

- [10] Choi Hun., "Stock prediction analysis through artificial intelligence using big data", *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 25.10, pp. 1435-1440, Aug. 2021. <https://doi.org/10.6109/jkiic.e.2021.25.10.1435>.
- [11] Misra and N. N., "IoT, big data and artificial intelligence in agriculture and food industry", *IEEE Internet of Things Journal*, Vol. 9, pp. 6305-6324, May 2020. <https://doi.org/10.1109/JIOT.2020.2998584>.
- [12] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton, "Deep learning", *nature*, Vol. 521.7553, pp. 436-444, May 2015. <https://doi.org/10.1038/nature14539>.
- [13] Qinglong Li, Sanghong Jeon, Changjae Lee, and Jae Kyeong Kim, "A Study on the Job Recommender System Using User Preference Information", *Journal of Information Trchnology Services*, Vol. 20(3), pp. 57-73, Jun. 2021. <https://doi.org/10.9716/KITS.2021.20.3.057>.
- [14] Lee Kyung Yul and Lee Kyung A., "A Case Study on the Use of Artificial Intelligence (Ai) in Platform Business in the Era of Untact.", *Journal of Cultural Product & Design*, Vol. 62.0, pp. 111-120, Sep. 2020. UCI : I410-ECN-0102-2021-600-001348130.
- [15] Singhal, Ayush, Pradeep Sinha, and Rakesh Pant. "Use of deep learning in modern recommendation system: A summary of recent works.". *arXiv preprint arXiv:1712.07525*, pp. 1-6. Dec. 2017. <https://doi.org/10.48550/arXiv.1712.07525>.
- [16] Cheng. Heng-Tze, "Wide & deep learning for recommender systems", *Proceedings of the 1st workshop on deep learning for recommender systems*, pp. 7-10, Sep. 2016. <https://doi.org/10.1145/2988450.2988454>.
- [17] Aminu Da'u, and Naomie Salim, "Recommendation system based on deep learning methods: a systematic review and new directions", *Artificial Intelligence Review*, Vol. 53.4, pp. 2709-2748, Aug. 2020. <https://doi.org/10.1007/s10462-019-09744-1>.
- [18] Sehwa Choi, Eunji Shin, Seokjin Yang, Youngduk Seo, and Euijong Lee, "Deep Learning-based Recommendation System using User Purchase History", *Proceedings of KIIT Conference*, pp. 421-424. Nov. 2021.
- [19] Antoniuk and Dmytro S., "Development of a simulator to determine personal financial strategies using machine learning", *CEUR Workshop Proceedings*, Vol. 3077, pp.12-26, Dec. 2022. <http://ceur-ws.org/Vol-3077/paper02.pdf>.
- [20] Kaelbling, Leslie Pack, Michael L. Littman, and Andrew W. Moore, "Reinforcement learning: A survey", *Journal of artificial intelligence research*, Vol. 4, pp. 237-285, May 1996. <https://doi.org/10.1613/jair.301>.
- [21] Watkins, Christopher JCH, and Peter Dayan, "Q-learning", *Machine learning*, Vol. 8.3, pp. 279-292, May 1992. <https://doi.org/10.1007/BF00992698>.
- [22] Se-Joon Park, "Improving Recommendation Accuracy based on Machine Learning using Multi-Dimensional Features of Word2Vec", *Journal of KIIT*, Vol. 19, No. 3, pp.9-14, Dec. 2021. <https://doi.org/10.14801/jkiit.2021.19.3.9>.
- [23] Minyoung Yeo, Seungjoo Lee, Seokho Ahn, Euijong Lee, and Young-Duk Seo, "Data integration method reflecting user preferences in the personalized recommender system", *Proceedings of KIIT Conference*, pp. 378-381, Nov. 2021.
- [24] Yunju Lee, Jaejun Lee, and Hyunchul Ahn, "Improvement of a Product Recommendation Model using Customers", *Search Patterns and Product Details.Journal of the Korea Society of Computer and Information* , Vol. 26(1), pp. 265-274, Jan. 2021. <https://doi.org/10.9708/jksci.2021.26.01.265>.
- [25] L. Samek and Wojciech, "Evaluating the visualization of what a deep neural network has

- learned", IEEE transactions on neural networks and learning systems, Vol. 28, No. 11, pp. 2660-2673, Aug. 2016. <https://doi.org/10.1109/TNNLS.2016.2599820>.
- [26] Huang, Guang-Bin, Yan-Qiu Chen, and Haroon A. Babri, "Classification ability of single hidden layer feedforward neural networks", IEEE transactions on neural networks, Vol. 11, No. 3, pp. 799-801, May 2000. <https://doi.org/10.1109/72.846750>.
- [27] Cooijmans and Tim, "Recurrent batch normalization", arXiv preprint arXiv:1603.09025, pp. 1-13, Feb. 2016. <https://doi.org/10.48550/arXiv.1603.09025>.
- [28] Agarap and Abien Fred, "Deep learning using rectified linear units (relu)" arXiv preprint arXiv:1803.08375, pp. 1-7, Feb. 2019. <https://doi.org/10.48550/arXiv.1803.08375>.
- [29] Baldi, Pierre, Peter J., and Sadowski, "Understanding dropout", Advances in neural information processing systems, Vol. 26 No. 1, Jan. 2013. <https://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf>.
- [30] Gao Bolin and Lacra Pavel, "On the properties of the softmax function with application in game theory and reinforcement learning" arXiv preprint arXiv:1704.00805, Aug. 2018. <https://doi.org/10.48550/arXiv.1704.00805>.
- [31] Moon Jee Hyun and Eunkyung Jeon, "Analysis on User Sensitive Evaluations of Disposable and Reusable Sanitary Pads", Korean Society for Emotion and Sensibility, Vol. 17, No. 2, pp. 77-84, Apr. 2014. <https://doi.org/10.14695/KJSOS.2014.17.2.77>.
- [32] Yeon Soo Kim, Sujeung Kim, Sujin Lee, and Dong-eun Kim, "Analysis on the Shape and Size of Sanitary Pads and User Experience Evaluation", Journal of the Korean Society of Clothing and Textiles, Vol. 44, No. 3, pp. 485-498, Jun. 2020. <https://doi.org/10.5850/JKSCT.2020.44.3.485>.
- [33] Hadgu, Asmelash Teka, Aastha Nigam, and Ernesto Diaz-Aviles, "Large-scale learning with AdaGrad on Spark", 2015 IEEE International Conference on Big Data (Big Data). IEEE, pp. 2828-2830, Oct. 2015. <http://dor.org/10.1109/BigData.2015.7364091>.
- [34] Kingma, Diederik P., and Jimmy Ba, "Adam: A method for stochastic optimization", arXiv preprint arXiv:1412.6980, pp. 1-15, Jan. 2017. <https://doi.org/10.48550/arXiv.1412.6980>.
- [35] Nitish Shirish Keskar and Richard Socher, "Improving generalization performance by switching from adam to sgd", arXiv preprint arXiv:1712.07628, pp. 1-10, Dec. 2017. <https://doi.org/10.48550/arXiv.1712.07628>.
- [36] Kurbiel Thomas, and Shahrzad Khaleghian, "Training of deep neural networks based on distance measures using RMSProp" arXiv preprint arXiv:1708.01911, pp. 1-6, Aug. 2017. <https://doi.org/10.48550/arXiv.1708.01911>.

저자소개

오 흥 석 (Hong-Seok OH)



2022년 02월 : 충북대학교
소프트웨어학과 공학사
2022년 03월 ~ 현재 : 충북대학교
컴퓨터과학과 석사과정
관심분야 : 빅데이터 분석, IoT,

정 재 민 (Jae-Min Jeong)



2018년 03월 ~ 현재 : 충북대학교
소프트웨어학과 재학중
관심분야 : 빅데이터 분석, IoT

남 윤 중 (Yoon-Joong Nam)



2002년 02월 : 남서울대학교
컴퓨터학과 학사
2004년 02월 : 광운대학교
컴퓨터학과 석사
2019년 06월 ~ 현재 : (주)쌍즈 CTO
관심분야 : 빅데이터 분석,
인공지능 시스템

서 영 덕 (Young-Duk Seo)



2012년 02월 : 고려대학교
컴퓨터전과 통신공학과 공학사
2018년 02월 : 고려대학교
컴퓨터공학과 공학박사
2020년 09월 ~ 현재 : 인하대학교
컴퓨터공학과 조교수
관심분야 : 추천 시스템,
사물인터넷, 텍스트 마이닝

사물인터넷, 텍스트 마이닝

이 의 종 (Eui-Jong Lee)



2012년 02월 : 고려대학교
컴퓨터정보학과 이학사
2018년 09월 : 고려대학교
컴퓨터공학과 공학박사
2020년 09월 ~ 현재 : 충북대학교
소프트웨어학부 조교수
관심분야 : 소프트웨어 공학,
사물인터넷, 빅데이터 분석

사물인터넷, 빅데이터 분석