

# 사용자의 선호도 정보를 활용한 직무 추천 시스템 연구

이청용\* · 전상홍\*\* · 이창재\*\*\* · 김재경\*\*\*\*

## A Study on the Job Recommender System Using User Preference Information

Qinglong Li\* · Sanghong Jeon\*\* · Changjae Lee\*\*\* · Jae Kyeong Kim\*\*\*\*

### ■ Abstract ■

Recently, online job websites have been activated as unemployment problems have emerged as social problems and demand for job openings has increased. However, while the online job platform market is growing, users have difficulty choosing their jobs. When users apply for a job on online job websites, they check various information such as job contents and recruitment conditions to understand the details of the job. When users choose a job, they focus on various details related to the job rather than simply viewing and supporting the job title. However, existing online job websites usually recommend jobs using only quantitative preference information such as ratings. However, if recommendation services are provided using only quantitative information, the recommendation performance is constantly deteriorating. Therefore, job recommendation services should provide personalized services using various information about the job. This study proposes a recommended methodology that improves recommendation performance by elaborating on qualitative preference information, such as details about the job. To this end, this study performs a topic modeling analysis on the job content of the user profile. Also, we apply LDA techniques to explore topics from job content and extract qualitative preferences. Experiments show that the proposed recommendation methodology has better recommendation performance compared to the traditional recommendation methodology.

Keyword : Topic Modeling, Latent Dirichlet Allocation(LDA), Job Recommender System

Submitted : March 31, 2021

1<sup>st</sup> Revision : May 16, 2021

Accepted : June 14, 2021

\* 경희대학교 대학원 빅데이터응용학과 박사과정

\*\* 경희대학교 대학원 빅데이터응용학과 석사

\*\*\* 경희대학교 대학원 경영학과 박사수료, 교신저자

\*\*\*\* 경희대학교 경영대학/대학원 빅데이터응용학과 교수

## 1. 서론

정보기술의 발전과 모바일 기기 대중화로 인해 온라인 전자상거래 시장의 규모는 매년 꾸준히 증가하고 있다. 이로 인해 다양한 유형의 제품과 서비스가 꾸준히 등장하지만, 사용자들은 의사결정 과정에서 오히려 많은 시간을 소모하고 필요한 정보를 선택할 때 어려움을 경험하고 있다(Su and Khoshgoftaar, 2009; Park et al., 2012). 이와 같은 정보 과부하(Information Overload) 문제를 해결하기 위해 개인화 추천 서비스의 중요성이 대두되고 있으며, Amazon, Netflix, Google 등 세계적인 기업들은 사용자 맞춤형 추천 서비스를 제공하여 기업의 지속 가능한 경쟁력을 강화하고 있다(Linden et al., 2003; Bennett and Lanning, 2007; Das et al., 2007). 최근에는 실업 문제가 사회적 해결과제로 등장하고 구직에 대한 수요가 높아지면서 온라인 취업 사이트가 활성화되고 있다. 이에 따라 사용자 맞춤형 직무를 제공해주는 추천 서비스에 대한 필요성이 꾸준히 제기되고 있다(박수상, 2016).

온라인 취업 사이트에서 사용자들이 직무를 지원할 때 해당 직무에 대한 구체적인 내용을 파악하기 위해 직무 내용, 채용 조건 등 다양한 정보를 확인한다(Al-Otaibi and Ykhlef, 2012). 이는 사용자가 특정 직무를 지원할 때 단순히 직무 자체를 보고 지원하는 것이 아니라 직무에 관련된 다양한 세부적인 정보에 중점을 두고 있다는 것을 의미한다. 하지만 기존의 온라인 취업 사이트는 주로 광고 형태의 직무 추천, 선호도 평점, 지원 여부 등 정량화된 선호도 정보만을 사용하여 직무를 추천하고 있다(Reusens et al., 2017). 이처럼 정량화된 정보만을 사용하여 추천 서비스를 제공하면 사용자의 개인 선호도를 반영하지 못하기에 추천 성능이 저하되는 문제가 꾸준히 제기되고 있다(Zhang et al., 2014; Seo et al., 2017). 예를 들어, 사용자 A는 “데이터 사이언스”라는 직무에서 주로 데이터 분석 기획 및 분석 수행 관련 업무를 수행하였다. 사용자 B는 “데이터 분석가”라는 직무를 수행하지만, 업무 내

용은 데이터 분석을 위한 기획과 분석을 수행하는 것이다. 이처럼 두 직무의 업무 내용은 같지만, 실제 직무 자체는 서로 다르다는 것을 확인할 수 있다. 따라서 직무 추천 서비스를 제공할 때 직무 자체만을 고려하는 것이 아니라 직무 내용을 바탕으로 유사한 주제를 나타내는 직무를 정교하게 분류하여 이를 추천 시스템에 반영할 필요성이 있다.

최근에는 정량화된 선호도 정보만을 사용하는 기존 개인화 추천 서비스 연구의 문제점을 극복하기 위해 다양한 분야의 데이터를 사용하는 연구가 활발히 진행되고 있다(Zhang et al., 2019). 대표적으로 온라인 쇼핑, 커뮤니티 등에서 사용자가 직접 작성한 구매 후기, 제품 정보 등 텍스트 데이터를 사용하는 연구가 꾸준히 진행되고 있다(Lei et al., 2016; Zheng et al., 2017; Cheng et al., 2018). 본 연구의 목적은 사용자들의 지원한 직무와 선호하는 직무 주제를 바탕으로 맞춤형 직무 추천 서비스를 제공하는 것이다. 이를 위해 사용자가 지원한 직무에 대해 토픽 모델링(Topic Modeling) 분석을 수행하여 비슷한 주제를 나타내는 직무들을 정교하게 분류하고 이를 바탕으로 맞춤형 직무 추천 서비스를 제공하는 방법론을 제안하고자 한다. 토픽 모델링은 직무 소개 등과 같은 텍스트 데이터에서 일정한 형태를 보이는 주제를 파악하는 분석 방법이다(Jelodar et al., 2019). 이를 위한 다양한 분석 기법이 존재하지만, 대표적으로 사용되는 것이 LDA(Latent Dirichlet Allocation) 기법이다(박상현 외, 2017; 이병현 외, 2020). 따라서 본 연구에서 제안하는 추천 방법론은 직무 소개에서 일정한 패턴을 나타내는 주제를 탐색하기 위해 LDA 기법을 적용하고 이를 통해 사용자가 선호하는 직무 주제를 파악한다. 이처럼 유사한 주제로 분류된 직무 주제를 바탕으로 최근 개인화 추천 서비스 연구에서 널리 사용되는 NCF 프레임워크를 적용하여 사용자별 직무 지원 정보 및 직무 주제 정보를 포괄적으로 고려하는 방법론을 제안한다. 즉, 직무 소개에 대해 토픽 모델링 분석을 수행하여 사용자가 지원한 직무를 연관성이 높은 토픽에 분류하고 분류한 결과를 바탕으로 사용자-토픽

행렬을 도출한다. 이를 바탕으로 사용자가 특정 주제를 포함하는 직무에 대한 지원 여부를 나타내는 암묵적 피드백 문제로 정의하고 맞춤형 직무 추천 서비스를 제공하고자 한다. 이를 통해 기존 사용자 맞춤형 직무 추천 서비스에서 시도하지 않은 직무 소개에서 연관성이 높은 직무들을 분류하고 이를 추천 시스템에 정교하게 반영하는 새로운 추천 방법론을 제안한다.

본 연구는 다음과 같은 구성으로 진행된다. 제 2장에서는 개인화 추천 서비스와 토픽 모델링 관련 이론적 배경을 구체적으로 설명한다. 제 3장에서는 본 연구에서 제안한 추천 방법론에 대하여 단계적으로 설명한다. 제 4장에서는 실험 데이터, 평가 지표 및 결과에 대하여 설명한다. 마지막 장에서는 결론을 설명하고 본 연구의 한계점과 추후 연구계획에 대하여 설명한다.

## 2. 이론적 배경

### 2.1 개인화 추천 서비스

개인화 추천 서비스 연구는 1990년대에 처음으로 소개된 이후 현재까지 다양한 분야에서 연구가 활발히 진행되고 있다. 개인화 추천 서비스는 사용자의 구매 기록 혹은 선호도 정보 등을 활용하여 사용자에게 가장 적합한 아이템 또는 서비스를 제공하는 것을 목적으로 한다(Park et al., 2012). 기존 개인화 추천 서비스 연구에서는 협업 필터링(Collaborative Filtering, CF) 알고리즘을 널리 사용하였으며 현재까지도 우수한 추천 성능을 나타내고 있다. CF 알고리즘은 사용자 혹은 아이템 간의 유사도를 바탕으로 사용자의 최종 선호도를 예측하고 이를 바탕으로 사용자 맞춤형 아이템 혹은 서비스를 제공한다(Su and Khoshgoftaar, 2009). 추천 대상 사용자가 정해지면 해당 사용자의 구매 기록이나 선호도 정보를 바탕으로 다른 사용자나 아이템 간의 유사도를 계산하여 사용자의 최종 선호도를 예측한다(Park et al., 2012). 하지만, 새로운 사용자가 나타나면 유사도를

계산할 수 있는 구매 기록 혹은 선호도 정보가 부족할 수가 있다. 이처럼 CF 알고리즘은 새로운 사용자가 나타나면 선호도를 예측하기 어려운 Cold Start 문제점이 존재하고 있다. 마찬가지로 사용자가 선호하는 아이템이지만 다른 사용자들이 해당 아이템에 대한 구매를 진행하지 않아서 추천을 제공할 수 없는 First Start 문제점도 존재하고 있다. 또한, 사용자의 구매 기록 혹은 선호도 정보가 꾸준히 증가하면 알고리즘의 연산량이 증가하게 되어 많은 연산 시간과 연산 비용이 발생하는 확장성(Scalability) 문제점도 존재한다(Herlocker et al., 2000).

최근에는 이와 같은 기존 개인화 추천 서비스의 한계를 개선하기 위해 심층 신경망(Deep Neural Network, DNN)을 CF 알고리즘에 적용하는 다양한 연구가 활발하게 수행되고 있다. 기존 사용자와 아이템 간의 상호작용은 주로 내적을 통한 선형 학습을 통해 나타내고 있다. He et al.(2017)는 기존 선형관계의 한계를 개선하기 위해 사용자와 아이템 잠재적 요인 사이의 비선형 관계를 학습하기 위해 DNN을 적용하는 NCF(Neural Collaborative Filtering) 추천 방법론을 제안했다. 실험 결과를 통해 연구에서 제안하는 방법론이 전통적인 추천 방법론이 비해 추천 성능 측면에서 우수한 성과를 보여주었다. Tay et al.(2018)은 개인화 추천 서비스를 위한 새로운 LRML(Latent Relational Metric Learning) 추천 방법론을 제안했다. LRML는 메모리 모델과 어텐션 메커니즘(Attention Mechanism)을 사용하여 잠재적 요인 대신 잠재적 관계를 학습하여 모델의 성능을 향상했다. Netflix 및 MovieLens 데이터를 사용하여 실험한 결과 기존의 추천 방법론과 비교했을 때 우수한 성능을 나타내고 있음을 입증했다. Hutterer(2011)와 Reusens et al.(2017)은 CF 알고리즘을 적용하여 사용자와 직무 간의 명시적 피드백(Explicit Feedback)과 암묵적 피드백(Implicit Feedback)을 분석하여 사용자에게 직무를 추천하는 연구를 제안하였으며, 암묵적 피드백을 사용했을 때 명시적 피드백만큼 추천 성능이 우수하다는 것을 입증했다. 또한, CF 알고리즘보다 하이브리드

CF 알고리즘이 사용자에게 적합한 직무를 추천하는데 우수한 성능을 나타내고 있음을 입증한 연구가 있으며, 사용자 정보와 암묵적 피드백에 DNN을 적용하여 사용자에게 맞춤형 직무를 추천하는 다양한 연구가 수행되었다(Yang et al., 2017; Qin et al., 2020).

기존의 선행연구를 통해 CF 알고리즘을 적용하면 사용자에게 적합한 직무를 추천할 수 있음을 확인할 수 있다. 또한, 기존 CF 알고리즘에 DNN 결합하여 사용하면 전통적인 추천 방법론보다 더 우수한 추천 성능을 나타냄을 확인할 수 있다. 본 연구에서는 사용자 맞춤형 직무 추천 알고리즘으로 DNN과 CF 알고리즘을 결합한 NCF 프레임워크를 적용하여 제안한 방법론의 추천 성능을 측정하고자 한다.

## 2.2 토픽 모델링

토픽 모델링은 텍스트 마이닝 연구에서 대표적으로 사용되는 기법으로 텍스트 데이터에서 의미가 있는 주제를 탐색하는 방법이다(Jelodar et al., 2019). 토픽 모델링 연구에서 주로 사용하는 기법에는 PCA(Principal Component Analysis), LSA(Latent Semantic Analysis) 등 다양하게 존재하고 있다. 그 중에서 대표적으로 많이 사용되는 방법은 확률 분포를 사용하는 베이지 추론 기반 LDA(Latent Dirichlet Allocation) 기법이다(이병현 외, 2020). LDA는 문서의 검색이나 주제를 파악하기 위한 대표적인 알고리즘으로 전체 텍스트에서 일정한 패턴을 발견하는 기법이다(Lu et al, 2011; Blei, 2012). LDA는 텍스트에 토픽이 포함되어 있으며, 텍스트 내의 토픽은 확률 분포를 바탕으로 단어를 구성한다고 가정한다(Blei et al., 2010). 토픽 모델링 기법은 유사한 주제를 나타내는 텍스트를 하나의 토픽 그룹으로 구성한다는 점에서 기존의 군집화 기법과 비슷하지만 하나의 텍스트가 여러 토픽에 대응될 수 있다는 점에서 기존 군집화 기법과 차별성을 가진다. 따라서 문서, 뉴스 등 텍스트에서 토픽을 추출하여 주제를 분석하거나 다양한 연구에서 핵심 주제를 도출 및 응용할

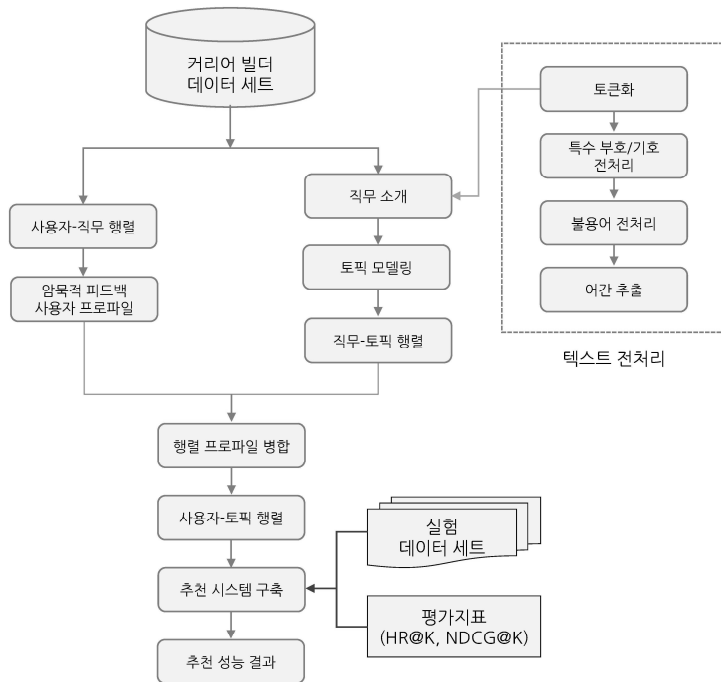
때 LDA 기법을 많이 사용하고 있다(Griffiths and Steyvers 2004; 정영진, 조운호, 2017).

최근에는 이와 같은 LDA 기법을 개인화 추천 서비스에 활용하는 다양한 연구가 수행되고 있다. Liu et al.(2011)는 iExpand 방법론을 통해 새로운 CF 기반 추천 시스템을 제안하였다. 이는 LDA 기법을 사용하여 아이템 및 사용자 관심사 간의 상호작용을 파악하여 추천 시스템에 반영하였다. McAuley and Leskovec(2013)는 LDA 기법을 적용하여 제품 리뷰에서 사용자의 관심사를 추출하고 이를 MF(Matrix Factorization) 알고리즘에 결합하여 사용자의 선호도를 예측하여 우수한 성능을 보여주었다. Wilson et al.(2014)는 평점과 텍스트를 동시에 고려하는 새로운 추천 방법론을 제안했다. 구체적으로 아이템 소개에 포함되는 텍스트를 LDA 기법을 적용하여 아이템의 잠재적 속성을 추론하고 이를 바탕으로 유사도를 계산하고 최적의 이웃 사용자를 선택하여 사용자의 선호도를 예측하였다. Zhao et al. (2015)는 사용자가 특정 아이템을 평가할 확률을 학습하기 위해 LDA 기법을 적용하고 이를 MF 기법과 결합하는 하이브리드 추천 방법론을 제안했다. MovieLens 및 EachMovie 데이터 세트를 사용한 실험에서 제안한 방법론의 우수한 추천 성능을 보여주었다. Leksin and Ostapets(2016)는 채용공고를 통해 사용자의 지원 여부를 예측하는 새로운 직무 추천 방법론을 제안했다. 해당 연구는 CF 알고리즘과 LDA 기법을 사용하여 구직자 정보와 직무 정보를 분류하여 구직자에게 적절한 직무를 추천하였다. Pavlinek and Podgorelec(2017)은 LDA 기법을 사용하여 텍스트 데이터를 분류하여 분류된 텍스트 데이터의 라벨링 값과 실제 텍스트 데이터의 라벨링 값을 비교하는 연구를 수행하였다. 실험 결과 연구에서 제안하는 방법론이 기존의 방법론보다 더 우수한 성과를 나타냈다. 또한, LDA 기법과 Word2Vec 기법을 적용하여 텍스트 데이터를 군집화하는 방법을 제안하였으며, LDA 기법으로 텍스트로 구성된 문서들을 군집화하는 연구를 제안했다(Li et al., 2018; Onan et al., 2017).

기존의 대부분 선행연구를 살펴보면 텍스트 데이터에 포함되는 주제를 파악하기 위해 LDA 기법을 적용하고 이를 전통적인 CF 기법에 결합하여 최종적으로 사용자의 선호도 평점을 예측하였다. 그러나 최근에는 선호도 평점 예측이 우수한 추천 성능을 나타내는 것이 아니라는 문제점이 제기되고 있다 (Zhou et al., 2016). 이에 따라 최근에 수행되는 많은 연구에서는 추천 시스템 관련 연구를 암묵적 피드백(Implicit feedback) 관련 문제로 정의하고 평점 예측이 아닌 추천 순위 예측에 중점을 두고 있다. 이에 본 연구는 사용자와 특정 직무 사이의 상호작용 가능성을 예측하는 문제로 정의하고자 한다. 이를 위해 먼저, 토픽 모델링 기법을 적용하여 직무 소개의 내용을 분석하고 이를 바탕으로 해당 직무를 가장 연관성이 높은 주제로 분류함으로써 사용자가 지원한 직무에 대한 주제를 파악한다. 이를 통해 사용자가 지원한 직무의 주제를 고려하여 CF 기반 추천 알고리즘을 구축하면 추천 성능이 제고될 수 있을 것을 기대한다.

### 3. 추천 방법론

본 연구에서는 사용자 맞춤형 직무 추천 시스템을 구축하기 위해 토픽 모델링 기법을 적용하여 직무 소개에 포함된 잠재적 주제를 추출하여 분석하였다. 이를 통해 직무 간의 유사성 및 이를 활용한 딥러닝 기반의 사용자 맞춤형 직무 추천 방법론을 제안하고자 한다. 본 연구에서 제안하는 추천 방법론은 [그림 1]과 같으며 총 3단계로 구성된다. 먼저, 실험을 위해 글로벌 채용 정보 사이트 커리어 빌더(Career Builder)에서 사용자 정보, 직무 정보 및 직무 소개를 포함하는 직무 지원 기록 데이터를 수집하였다. 수집한 데이터에는 사용자가 직무에 부여한 평점 정보가 없으므로 상호작용 여부를 예측하는 암묵적 피드백 문제로 정의하고 실험을 수행하였다. 다음으로 토픽 모델링 기법을 적용하여 직무 소개를 분석하여 사용자가 지원한 직무를 이와 가장 연관성이 있는 토픽으로 분류하고 최종적으로 사용자-토픽 행렬을 구축하였다. 마지막으로 기존 개인화 추천 서비스



[그림 1] 추천 방법론 프레임워크

관련 연구에서 우수한 추천 성능을 나타낸 NCF 프레임워크를 적용하여 추천 성능을 평가한다.

### 1단계 : 사용자 프로파일 정의

본 연구에서 제안한 추천 방법론의 성능을 평가하기 위해 글로벌 채용 정보 사이트 커리어 빌더에서 제공하는 직무 지원 기록 데이터를 수집하여 실험에 사용하였다. 수집한 데이터는 사용자 정보, 직무 정보, 직무 소개 등으로 구성되었으며, 기존 개인화 추천 서비스 연구에서 주로 사용되는 평점 정보는 존재하지 않는다. 따라서 본 연구에서는 수집한 데이터를 사용하여 사용자와 직무 간의 상호작용 여부를 예측하는 암묵적 피드백 관련 문제로 정의하고 사용자 프로파일을 구축했다. 이를 위해 식 (1)과 같이 사용자  $u$ 와 직무  $i$  사이에 상호작용이 존재할 때는 1, 아닌 경우에는 0으로 표시하였다.

$$R_{ui} = \begin{cases} 1, & \text{if interaction(user, job) is observed;} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

여기에서 사용자  $u$ 와 특정 직무  $i$  사이에 상호작용이 존재할 때는  $R_{u,i}$ 는 1을 나타내고 있다. 이때

사용자  $u$ 가 직무  $i$ 를 선호하는 것을 의미하지 않는다. 마찬가지로  $R_{u,i}$ 가 0일 때 사용자  $u$ 가 직무  $i$ 를 선호하지 않는다는 의미는 아니며 사용자가 직무를 인식하지 못할 수도 있다(He et al., 2017). 이와 같은 암묵적 피드백 데이터는 사용자의 선호도에 대한 노이즈 정보를 제공하므로 암묵적 데이터를 학습할 때 한계가 존재한다. 사용자와 상호작용이 존재한다면 적어도 직무에 대한 사용자의 선호도를 간접적으로 반영할 수 있지만 반대의 경우에는 사용자가 직무에 대한 내거티브 피드백(Negative Feedback)이 부족하므로 직무에 선호도 정보를 확인하기 어렵다(Lee et al., 2008; He et al., 2017). 따라서 본 연구에서는 기존의 암묵적 피드백 관련 연구의 전략을 참고하여 사용자와 직무 사이에 상호작용이 발생할 가능성을 예측하는 문제로 정의한다(Hu et al., 2008; He et al., 2017). 이에 따라, 수집한 데이터에서 사용자가 지원한 직무 기록이 <표 1>의 예시와 같으며, 여기에 식 (1)을 적용하여 <표 2>의 예시와 같은 사용자-직무 행렬  $M_{ui} \in R^{User \times Job}$ 을 구축할 수 있다.

<표 1> 사용자 지원 직무 기록 예시

User	Jobs	User	Jobs
$User_1$	{ $Job_1, Job_5$ }	$User_6$	{ $Job_2, Job_4, Job_5$ }
$User_2$	{ $Job_2, Job_3, Job_5, Job_7$ }	$User_7$	{ $Job_2, Job_6$ }
$User_3$	{ $Job_1, Job_4$ }	$User_8$	{ $Job_1, Job_4$ }
$User_4$	{ $Job_1, Job_2, Job_5, Job_7$ }	$User_9$	{ $Job_3, Job_6, Job_7$ }
$User_5$	{ $Job_1, Job_3, Job_6$ }	$User_{10}$	{ $Job_2, Job_3, Job_5$ }

<표 2> 사용자-직무 행렬 예시

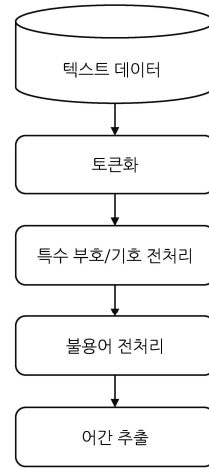
	$Job_1$	$Job_2$	$Job_3$	$Job_4$	$Job_5$	$Job_6$	$Job_7$
$User_1$	1	0	0	0	1	0	0
$User_2$	0	1	1	0	1	0	1
$User_3$	1	0	0	1	0	0	0
$User_4$	1	1	0	0	1	0	1
$User_5$	1	0	1	0	0	1	0
$User_6$	0	1	0	1	1	0	0
$User_7$	0	1	0	0	0	1	0
$User_8$	1	0	0	1	0	0	0
$User_9$	0	0	1	0	0	1	1
$User_{10}$	0	1	1	0	1	0	0

**2단계 : 텍스트 전처리 및 토픽 모델링**

두 번째 단계에서는 수집한 데이터에서 직무 소개의 주제를 분석하기 위해 텍스트 데이터에서 주제를 효과적으로 분석할 수 있는 토픽 모델링 기법을 적용하였다. 이를 위해, 먼저 수집한 직무 소개 텍스트 데이터를 [그림 2]와 같이 전처리 작업을 수행하였다. 직무 소개에는 태그, 특수 부호, 기호 등 실제 분석 과정에서 분류 성능을 저하할 수 있으므로 사전에 제거해야 한다. 또한, “I”, “my”, “she” 등 불용어는 텍스트에서 나타내는 빈도수는 많지만 큰 의미가 없으므로 제거하였다. 이처럼 실제 분석에 필요하지 않은 기호나 텍스트를 제외하고 나머지 텍스트는 모두 소문자로 변환하여 최종적으로 직무별로 토큰화 작업을 통해 키워드를 추출한다.

다음으로 Python 기반으로 배포된 Gensim 패키지를 사용하여 직무 소개의 주제를 분석하고 각 직무를 가장 연관성이 높은 토픽으로 분류하였다. LDA 기법을 적용하여 토픽 모델링 분석을 수행할 때 토픽의 개수가 설정되어 있지 않으면 통계적 분석 결과를 근거로 적절한 토픽 개수를 설정할 수 있다 (김지원 외, 2020). 본 연구에서는 LDA 관련 연구에서 널리 사용되고 있는 Coherence Score를 바탕으로 최적의 토픽 개수를 설정하였다(O’callaghan et al, 2017). <표 3>의 예시는 직무별로 각 토픽에 연관되는 비율을 보여주고 있다.

이와 같은 분석 결과를 바탕으로 본 연구에서는 기존 토픽 모델링 관련 연구 방법을 적용하여 <표 4>와 같이 직무-토픽 행렬  $N_{jt} \in R^{Job \times Topic}$ 을 구축하였다. 여기에서 직무  $i$ 와 토픽  $t$  사이의 연관되는 토픽



[그림 2] 텍스트 전처리 과정

비율이 가장 높을 때 해당 직무  $i$ 를 토픽  $t$ 에 분류하고 1로 표시하였으며, 나머지 토픽에는 0으로 표시하였다(박상현 외, 2017). 마지막으로 앞서 도출한 사용자-직무 행렬과 직무-토픽 행렬에 식 (2)을 적용하여 사용자와 토픽에 대응하는 행렬을 도출하였으며, 구체적인 도출 과정은 [그림 3]과 같다.

<표 4> 직무-토픽 행렬 예시

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
$Job_1$	1	0	0	0	0
$Job_2$	0	0	1	0	0
$Job_3$	0	1	0	0	0
$Job_4$	0	0	0	1	0
$Job_5$	0	0	0	0	1
$Job_6$	1	0	0	0	0
$Job_7$	0	0	0	0	1

<표 3> 직무별 토픽 연관 비율 예시

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
$Job_1$	0.026001	0.011345	0.025443	0	0
$Job_2$	0.022201	0.021564	0.031582	0	0
$Job_3$	0.021027	0	0.014544	0	
$Job_4$	0	0.032516	0	0	0.021515
$Job_5$	0.010101	0	0	0	0.0345844
$Job_6$	0.021478	0	0	0.014442	0
$Job_7$	0.021451	0	0	0.021546	0.036441



[그림 3] 사용자-토픽 빈도 행렬 병합 과정 예시

$$P_{ut} = M_{ut} \cdot N_{jt} \quad (2)$$

먼저, [그림 3]과 같이 사용자-직무 행렬(a)과 직무-토픽 행렬(b) 간의 행렬 곱셈을 통해 사용자-토픽 빈도 행렬(c)을 도출한다(현윤진 외, 2015; 정영진, 조운호, 2017). 또한, 사용자와 토픽 간의 상호작용 발생 여부를 예측하는 암목적 피드백 문제로 정의하기 위해 사용자와 토픽 사이에 상호작용이 발생하였으면 1로 표시하고 발생하지 않으면 0으로 표시하여 최종적으로 <표 5>와 같이 사용자-토픽 행렬  $P_{ut} \in R^{User \times Topic}$ 을 도출하였다(Koren, 2008). 이를 통해 사용자가 각 토픽에 분류된 직무의 지원 여

부와 사용자가 해당 토픽에 대한 선호도를 간접적으로 파악할 수 있다.

### 3단계 : 딥러닝 기반 추천 시스템 구축

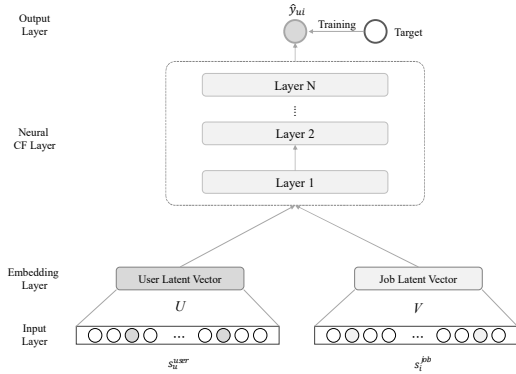
본 연구에서는 개인화 추천 서비스 관련 연구에서 널리 사용되는 CF 알고리즘에 DNN을 적용하여 우수한 성능을 보여준 NCF 프레임워크에 포함된 3가지를 알고리즘을 활용하여 본 연구에서 제안한 방법론의 추천 성능을 평가하였다(He et al., 2017). NCF 프레임워크는 정형화된 평점 선호도 데이터가 아닌 상호작용 가능성이 큰 제품을 추천하는 암목적 피드백 관련 연구를 목적으로 제안되었다. 즉, 사용자가 구매 가능성이 높은 제품이나 서비스의 추천 순위 점수를 예측하여 이를 바탕으로 맞춤형 추천 서비스를 제공한다.

먼저, MLP(Multi Layer Perceptron) 알고리즘은 사용자 잠재적 벡터(Latent Vector)와 직무 잠재적 벡터를 비선형 형태로 학습시키고 상호작용 여부를 예측하기 위해 DNN을 적용하였다. [그림 4]에서 입력 데이터  $s_u^{user}$ 와  $s_i^{job}$ 는 각각 사용자와 직무를 One-Hot Encoding 형태로 나타내고 있다. 그러나 입력 데이터의 차원의 크기가 커질수록 대부분의 값이 0으로 표현되는 데이터 희소성 문제점이

<표 5> 사용자-토픽 행렬 예제

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
User <sub>1</sub>	1	0	0	0	1
User <sub>2</sub>	1	0	1	1	1
User <sub>3</sub>	1	1	0	0	0
User <sub>4</sub>	1	0	1	1	1
User <sub>5</sub>	1	0	0	0	0
User <sub>6</sub>	0	1	1	0	1
User <sub>7</sub>	1	0	1	0	0
User <sub>8</sub>	1	1	0	0	0
User <sub>9</sub>	1	0	0	1	0
User <sub>10</sub>	1	0	1	0	1





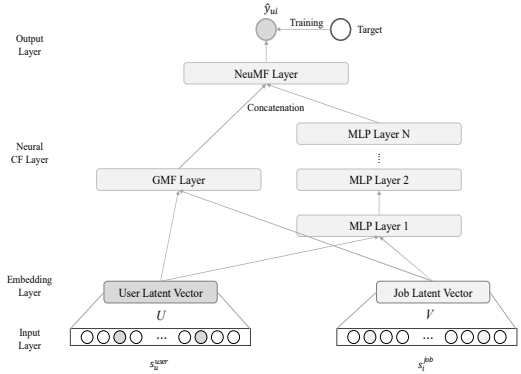
[그림 4] MLP 알고리즘

발생하게 된다. 따라서 임베딩 레이어를 통해 기존의 최소한 데이터 차원을 밀집 형태의 차원으로 표현되는 사용자 잠재적 벡터  $U$ 와 직무 잠재적 벡터  $V$ 로 변환한다. 이처럼 변환된 잠재적 벡터를 레이어 층을 통과시켜 상호작용 여부를 예측하는 과정이 MLP 알고리즘이며 식 (3)과 같이 정의할 수 있다. 최종적으로 Sigmoid 활성화 함수를 통해 사용자와 직무 간의 상호작용 여부를 예측하여 나타낸 것으로 0과 1 사이의 확률값으로 표현된다. 또한, 추천 목록의 크기  $K$ 에 따라 확률값이 가장 큰 상위  $K$ 개의 직무를 선택하여 사용자에게 추천 서비스를 제공한다.

$$\hat{y}_{ui} = a_L(W_L^T(a_{L-1}(\dots a_2(W_2^T \begin{bmatrix} p_u^M \\ q_i^M \end{bmatrix} + b_2) \dots) + b_L) \quad (3)$$

여기서  $p_u^M$ 는 사용자 임베딩 벡터를 나타내고  $q_i^M$ 는 직무 임베딩 벡터를 나타내고 있다. 또한,  $a_L$ ,  $W_L$ ,  $b_L$ 은 각각  $L$ 번째 레이어의 ReLU 활성화 함수, 가중치 그리고 편향을 나타내고 있다.

다음으로 GMF 알고리즘은 전통적인 MF 알고리즘을 일반화하여 Element-wise 연산을 수행하는 알고리즘이다. [그림 5]와 같이 입력 데이터에서 사용자 A와 직무 B는 임베딩 레이어를 통해 각각 사용자 잠재적 벡터  $U$ 와 직무 잠재적 벡터  $V$ 로 변환된다. 이는 앞서 방법과 같이 입력된 사용자와 직무가 모두 고차원의 최소한 벡터로 표현되기 때문에 효율적으



[그림 5] NeuMF 알고리즘

로 분석하기 위해 저차원의 벡터로 변환하는 과정이 필요하다. 최종적으로 GMF 알고리즘을 식 (4)와 같이 비선형 활성화 함수를 사용하여 사용자와 직무 간의 상호작용 관계를 정교하게 표현할 수 있다.

$$\hat{y}_{ui} = p_u^G \odot q_i^G \quad (4)$$

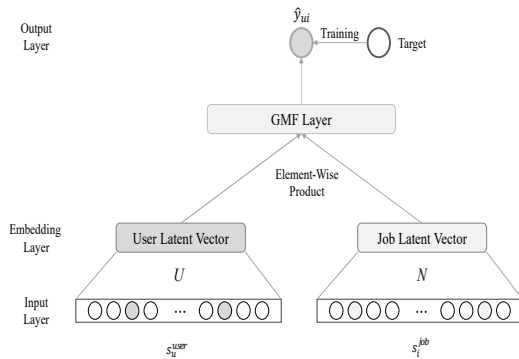
여기서  $p_u^G$ 와  $q_i^G$ 는 MLP 알고리즘에서 사용된 같은 사용자 임베딩 벡터와 직무 임베딩 벡터를 나타내고 있다. GMF 알고리즘에서는 사용자 잠재적 벡터  $U$ 와 직무 잠재적 벡터  $V$ 는 각 행렬의 원소를 곱하는 Element-wise 연산을 수행한다. 이에 따라 MLP 알고리즘보다 사용자와 직무의 상호작용 관계를 학습하는데 더 효율적인 구조로 되어 있다.

$$\phi^{GMF} = p_u^G \odot q_i^G$$

$$\phi^{MLP} = a_L(W_L^T(a_{L-1}(\dots a_2(W_2^T \begin{bmatrix} p_u^M \\ q_i^M \end{bmatrix} + b_2) \dots) + b_L) \quad (5)$$

$$\hat{y}_{ui} = \sigma(h^T \begin{bmatrix} \phi^{GMF} \\ \phi^{MLP} \end{bmatrix})$$

마지막으로, NeuMF 알고리즘은 [그림 6]과 같이 MLP 알고리즘과 GMF 알고리즘의 출력 결과를 결합하여 사용된다. 식 (5)와 같이 두 알고리즘의 출력 결과를 결합하고, 여기에 가중치를 곱하고 Sigmoid 활성화 함수를 적용하여 최종 점수  $\hat{y}_{ui}$ 를 예측한다. 여기서  $\hat{y}_{ui}$ 는 사용자  $u$ 와 직무  $i$ 간의 상호작용 가능성을 확률로 나타내고 있다.



[그림 6] GMF 알고리즘

## 4. 실험 데이터 및 결과

### 4.1 실험 데이터 및 성능평가

본 연구에서 제안하는 추천 방법론의 성능을 평가하기 위해 글로벌 채용 정보 사이트 커리어 빌더에서 제공하는 직무 기록 관련 데이터 651,638개를 사용하여 실험을 수행하였다. 데이터에는 156,258명 사용자와 27,198개 직무 및 직무 소개로 구성되었다. 또한, 이와 같은 암묵적 피드백 데이터를 사용하여 딥러닝 기반 추천 알고리즘을 학습시키기 위해 사용자가 실제 지원하지 않은 직무를 임의로 선택하여 학습 데이터에 포함하는 네거티브 샘플링(Negative Sampling) 전략이 필요하다(He et al., 2017). 본 연구에서 수집한 데이터는 사용자가 실제 지원한 직무 기록을 포함하고 있다. 그러나 이와 같은 지원 데이터만을 사용하여 지원 여부를 예측 변수로 설정하여 모델을 학습하기 어렵다. 또한, 사용자가 지원하지 않은 직무가 실제 지원한 직무보다 많으므로 지원하지 않은 직무를 모두 학습 데이터로 사용할 때는 학습하는 과정에서 편향이 발생할 수 있다. 따라서 본 연구에서 기존 관련 연구의 방법을 참고하여 실제 지원한 각 직무에 4개의 네거티브 샘플링을 설정하고 실험을 수행하였다(He et al., 2017).

추천 성능을 평가하기 위해 기존 순위 예측 관련 연구에서 널리 사용되는 Leave-One-Out 평가 방법을 적용하였다(Bayer et al., 2017; He et al., 2017).

이를 위해 각 사용자와 상호작용이 존재하는 직무에서 한 개의 직무는 테스트 데이터로 사용하여 나머지는 학습 데이터로 사용한다. 특히 모든 직무에 대하여 순위 점수를 예측하려면 많은 연산 시간과 비용이 소모된다. 이에 따라 사용자가 상호작용이 존재하지 않는 100개의 직무를 무작위로 샘플링하여 100개 직무 중 테스트 데이터의 순위를 평가하는 기존 연구의 평가 방법을 사용했다(He et al., 2017). 평가지표는 순위 예측 관련 연구에서 주로 사용되고 있는 Hit Ratio(HR)와 Normalized Discounted Cumulative Gain(NDCG)를 사용하여 Top-10 리스트의 추천 성능을 측정했다(He et al., 2015).

HR 평가지표는 추천 순위의 성능을 평가할 때, 각 사용자의 추천 목록 순위에 테스트 데이터가 존재하는지를 측정하는 평가지표로 식 (6)과 같이 계산할 수 있다.

$$HR@K = \frac{\text{Number of Hits}}{\text{Number of Hits} + \text{Number of Misses}} \quad (6)$$

식 (6)에서  $K$ 는 추천 목록의 크기를 나타내고 *Number of Hits*는 추천 목록에 테스트 데이터가 존재하는 사용자의 수를 나타내며, *Number of Misses*는 추천 목록에 테스트 데이터가 존재하지 않는 사용자의 수를 나타낸다.

HR 평가지표는 추천 목록에 포함되는지를 확인하는 지표로 추천 순위에 따른 정확도를 반영하지 못하고 있다. 이 문제를 해결하기 위해 본 연구에서는 NDCG 평가지표를 사용하여 추천 목록에서 상위 순위의 직무에 더 높은 가중치를 부여하여 평가하였으며 이는 식 (7)과 같이 나타낼 수 있다.

$$NDCG@K = \sum_{j=1}^K \frac{2^{r_j} - 1}{\log_2(j+1)} \quad (7)$$

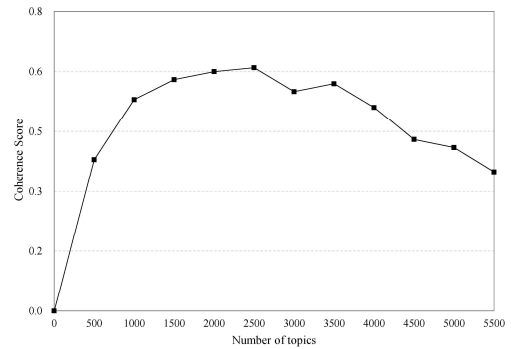
식 (7)에서  $K$ 는 추천 목록의 크기를 나타내고,  $j$ 는 추천된 직무를 나타내고,  $r_j$ 는 추천 순위의  $j$ 번째 위치에 있는 직무의 관련성을 나타낸다. 테스트 데이터에서  $j$ 번째 직무를 지원하면  $r_j$ 는 1로, 지원하지 않으면 0으로 간주하였다.

본 연구에서 제안하는 방법론의 추천 성능을 평가하기 위해 적용한 NCF 프레임워크의 각 알고리즘 파라미터는 다음과 같이 설정하였다. 최적화 알고리즘(Optimizer)은 관련 연구에서 우수한 성능을 보여준 Adam을 적용하였다(Kingma and Ba, 2014). 본 연구는 사용자와 토픽 사이에 상호작용 여부를 측정하는 이진 분류의 문제이므로 손실 함수는 Binary Cross-Entropy를 적용하고, Epoch는 알고리즘별로 가장 높은 성능을 나타내는 횟수로 설정하였다. Embedding 크기는 16으로 설정하고 Batch 크기는 64개로 설정하였다. 본 연구의 실험은 CPU Intel Core i5-9400F, 16GB RAM, GeForce RTX 2060 환경에서 수행되었다.

#### 4.2 실험결과

기존 추천 방법론에서는 직무의 유사성을 고려하지 않고 전체 직무를 사용하여 추천 서비스를 제공하였다. 그러나 이와 같은 방법론은 직무 개수가 증가할수록 데이터 희소성 문제가 발생할 수 있다. 즉, 직무 제목은 다르지만 직무 내용은 유사한 경우가 많으므로 유사한 직무를 분류하여 추천 시스템에 반영할 필요성이 있다. 따라서 본 연구에서 기존 연구의 방법을 참고하여 직무 간의 유사성을 분석하여 비슷한 주제를 나타내는 직무를 분류하였다(현윤진 외, 2015; 정영진, 조운호, 2017). 먼저 토픽 모델링 연구에서 대표적으로 사용하는 LDA 기법을 적용하여 최적의 토픽 개수를 선택하였다.

이를 위해 LDA 기법 연구에서 널리 사용되고 있는 Coherence Score를 바탕으로 최적의 토픽 개수를 도출하였다. 여기에서 Coherence Score가 높은 경우에는 같은 토픽으로 분류되어 텍스트 간의 유사도가 높다는 것을 나타내므로 Coherence Score가 가장 높은 지점을 최종 토픽 개수로 선택하였다. 본 연구에서는 토픽의 개수를 0부터 5500까지 적용하여 최적의 토픽 개수를 도출하였다. [그림 7]은 토픽 개수 별 Coherence Score를 나타내고 있으며, 토픽의 개수가 2,500개 일 때에 Coherence Score가 다른 토픽의 개수보다 높게 나타내고 있으므로 최종 토픽의 개수로 선택하였다. <표 7>은 직무 소개에 토픽 모델링 분석을 수행한 결과를 부분적으로 보여준다. 이처럼 직무 간의 유사성을 파악하여 비슷한 주제를 나타내는 직무들을 분류하고 이를 추천 시스템에 반영하여 기존의 전체 직무를 포함하는 방법론과 비교하였다.



[그림 7] 토픽 개수에 따른 Coherence Score 결과

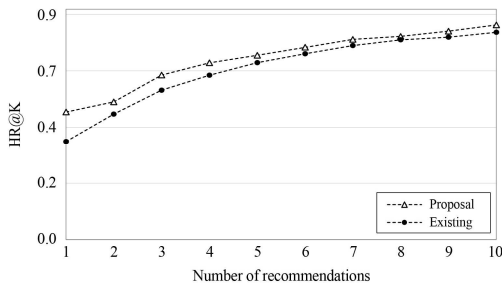
<표 7> 토픽 모델링 분석 결과 예시

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Account	Administrative	Electronics Assembler	Insurance	Health
Accounting Assistant	Administrative Opportunities	Electronics Assembly	Insurance Compliance	Health Service Agent
...	...	...	...	...
Account Executive	Administrative Support	Electronics Repair	Insurance Pre-Authorization	Health Care
Accounting Clerk	Administrative Staff	Electronics Equipment	Insurance Verification	Health Care Position

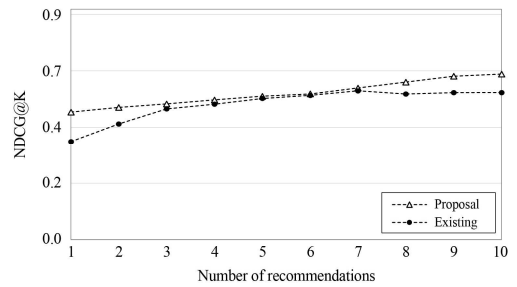
본 연구에서 제안한 방법론의 성능을 측정하기 위해 다양한 크기의 추천 목록에서 HR 성능과 NDCG 성능을 평가하였다. 첫 번째, HR 성능을 평가하기 위해 [그림 8]과 같이 추천 목록의 크기를 1부터 10까지 설정하고 실험을 수행하였다. 여기서 “Existing”은 모든 직무를 포함하는 기존의 추천 방법론을 나타내고 “Proposal”은 본 연구에서 제안한 유사한 주제를 나타내는 직무를 분류하여 반영하는 추천 방법론을 나타낸다. 본 연구에서 제안한 추천 방법론의 경우 기존 추천 방법론과 비교했을 때, 추천 목록의 크기에 상관없이 전체적으로 우수한 추천 성능을 나타내고 있다. 결과적으로 본 연구에서 제안한 추천 방법론이 기존 추천 방법론과 비교

했을 때 각각 평균적으로 7%(MLP), 11%(GMF), 14%(NeuMF)만큼 HR 성능이 개선되었다.

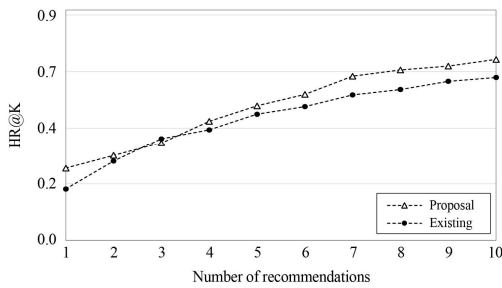
두 번째, NDCG 성능을 측정하기 위해 추천 목록의 크기를 1부터 10까지 설정하고 [그림 9]와 같이 제안한 추천 방법론의 NDCG 성능을 평가하는 실험을 수행하였다. 본 연구에서 제안한 추천 방법론의 경우 기존 추천 방법론과 비교했을 때, HR 성능과 마찬가지로 추천 목록의 크기에 상관없이 전체적으로 우수한 예측 성능을 나타내고 있다. 결과적으로 기존 추천 방법론과 비교했을 때 본 연구에서 제안한 추천 방법론의 NDCG 성능은 각각 평균적으로 8%(MLP), 16%(GMF), 26%(NeuMF)만큼 개선되었다.



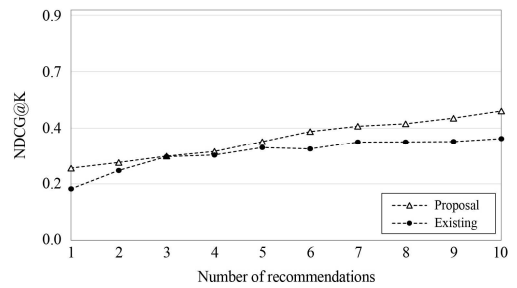
(a) MLP-HR@K



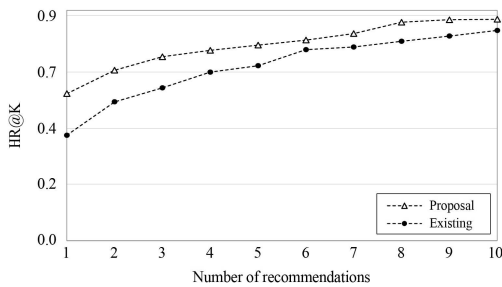
(a) MLP-NDCG@K



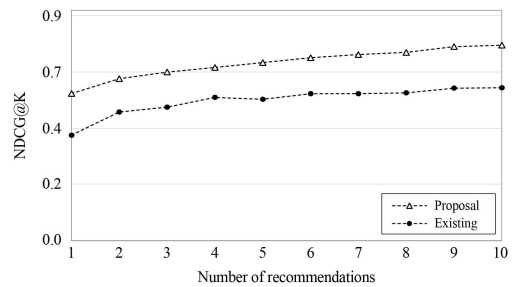
(b) GMF-HR@K



(b) GMF-NDCG@K



(c) NeuMF-HR@K



(c) NeuMF-NDCG@K

[그림 8] 추천 목록 크기에 따른 HR 결과

[그림 9] 추천 목록 크기에 따른 NDCG 결과

## 5. 결 론

본 연구에서는 직무 자체만을 고려하는 기존 직무 추천 서비스의 한계를 개선하기 위해 직무에 대한 구체적인 소개 등 정성적인 선호도 정보를 정교하게 반영하여 추천 성능을 향상하는 새로운 추천 방법론을 제안하였다. 이를 위해 본 연구에서는 사용자 프로파일의 직무 소개에 대해 토픽 모델링 분석을 수행하여 직무 간의 유사성을 파악하여 비슷한 주제를 나타내는 직무를 분류하여 새로운 사용자-토픽 행렬을 구축하고 이를 바탕으로 사용자 맞춤형 추천 서비스를 제공하였다. 본 연구의 실험을 위해 글로벌 채용 정보 사이트 커리어 빌더에서 제공하는 직무 지원 기록 관련 데이터를 수집하여 사용하였다. 실험 결과 모든 직무를 포함하는 기존 추천 방법론과 비교했을 때 유사한 주제를 나타내는 직무를 분류하고 이를 추천 시스템에 반영하는 본 연구에서 제안하는 방법론이 더 우수한 추천 성능을 보여주었다.

본 연구의 학술적 및 실무적 시사점은 다음과 같다. 첫째, 최근에는 실업 문제가 사회적 해결과제로 등장하고 구직에 대한 수요가 높아지면서 온라인 취업 사이트가 활성화되고 있다. 특히 직무를 선택할 때 직무 내용 등 세부적인 정보는 중요한 역할을 하고 있다. 따라서 사용자 맞춤형 직무 추천 서비스를 제공할 때 직무에 대한 세부적인 정보를 반영하는 것이 중요하다. 하지만 기존에 직무 추천 서비스는 광고 형태의 추천, 선호도 정보, 지원 여부 등 정량적 정보만을 사용하여 직무를 추천하였다. 본 연구에서는 직무 내용에 대하여 토픽 모델링 분석을 수행하여 주제를 탐색하고 사용자의 정성적 선호도를 정교하게 추출하여 새로운 사용자 프로파일을 구축하는 새로운 추천 방법론을 제안하였다. 이는 기존 정량적 정보만을 고려하는 연구에 국한되지 않고 텍스트 데이터 등 정성적 선호도를 동시에 고려하여 추천 성능을 향상하게 시킴으로써 사용자 맞춤형 직무 추천 서비스 연구의 확장에 기여한다. 둘째, 영화 소개, 상품 소개 등 제품에 대한 소개와 사용자

상호작용 정보를 포함하는 다양한 분야의 연구에 적용할 수 있다. 최근에 온라인 플랫폼에서 사용자 선호도를 명시적으로 나타내는 정량적인 평점 정보를 수집하기 어려운 실정이다. 따라서 구매 여부, 클릭 여부 등 암묵적 피드백 관련 데이터를 수집하는 연구가 활발히 진행되고 있다. 본 연구에서는 직무 추천 분야에서 유사한 직무를 분류하고 이를 추천 시스템에 반영하여 추천 성능을 평가하였으나 제품에 대한 소개 정보를 제공하고 사용자와 제품 간의 상호작용 여부를 확인할 수 있는 다른 분야에도 적용할 수 있다. 셋째, 온라인 취업 사이트 실무자에게 다음과 같은 시사점을 제공할 수 있다. 기존 온라인 취업 사이트에서 직무를 추천할 때 주로 광고 형태의 추천, 선호도 정보, 지원 여부 등 정량적 정보만을 반영하여 직무나 서비스를 제공했다. 이는 사용자의 선호도 정보 혹은 행동 패턴 정보를 통해 실제 사용자가 특정 직무를 선호할 거라고 생각하기 때문이다. 하지만 이와 같은 정량적 정보들은 실제 선호도를 나타내는 데 한계가 존재할 수 있으며 오히려 추천을 제공할 때 성능을 하락시키는 요인을 작용하고 있다. 본 연구에서는 이와 같은 문제를 개선하기 위해 새로운 직무 추천 방법론을 제안하여 전통적인 방법론에 비해 우수한 추천 성능을 확인함으로써 기존 온라인 취업 사이트 추천 서비스 전략에 대하여 재고의 여지가 있음을 확인할 수 있다.

본 연구가 가지는 한계점과 추후 연구계획은 다음과 같다. 첫째, 본 연구에서는 직무 기록 관련 데이터를 사용하여 사용자가 해당 직무를 지원할 가능성을 예측하는 암묵적 피드백 문제로 정의하였다. 그러나 연구에서 고려한 사용자, 직무 속성이 제한적이므로 추후 연구를 통해 다양한 속성을 고려할 필요가 있다. 예를 들어, 사용자 특성, 직무 카테고리 등 추가적인 정보들을 고려하면 추천 성능일 높일 수 있을 것으로 기대한다. 둘째, 텍스트 전처리 과정에서 오픈 소스(Open Source) 라이브러리 패키지를 사용하여 불용어 제거, 토큰화 등 작업을 진행했다, 하지만 분야에 따라 사용되는 불용어, 단어 등이 다르므로 추가 연구를 통해 특정

분야에서 적합한 사전을 구축하여 텍스트를 정교하게 정제하는 방법을 연구해야 한다. 셋째, 본 연구에서는 토픽 모델링을 수행할 때 기존 연구에서 대표적으로 사용되는 LDA 기법을 적용했다. 최근에는 토픽을 추출할 때 딥러닝 사용하는 다양한 기법들이 우수한 성능을 나타내고 있다. 추후 연구에서는 이와 같은 딥러닝 알고리즘을 적용하여 토픽을 정교하게 추출하면 본 연구에서 제안하는 추천 방법론의 추천 성능이 더욱 향상될 수 있을 것으로 기대한다.

## 참고문헌

- 김지원, 박상민, 박성호, 정하림, 윤일수, “한국도로공사 VOC 데이터를 이용한 토픽 모형 적용 방안”, *한국IT서비스학회지*, 제19권, 제6호, 2020, 1-13.
- 이병현, 최일영, 이정재, 김재경, 강현모, “온라인 뉴스를 이용한 기업평판 구성요인 탐색 및 지수 개발 연구”, *한국IT서비스학회지*, 제19권, 제6호, 2020, 145-159.
- 박상현, 문현실, 김재경, “토픽 모델링에 기반한 온라인 상품 평점 예측을 위한 온라인 사용 후기 분석”, *한국IT서비스학회지*, 제16권, 제3호, 2017, 113- 125
- 박수상, “협업적 필터링을 활용한 추천 채용 시스템의 설계와 구현”, 서울대학교 석사학위 논문, 2016.
- 정영진, 조운호, “온라인 구매 행태를 고려한 토픽 모델링 기반 도서 추천”, *지식경영연구*, 제18권, 제4호, 2017, 97-118.
- 현윤진, 김남규, 조운호, “토픽 분석을 활용한 관심 기반 고객 세분화 방법론”, *Journal of Information Technology Applications and Management*, 제22권, 제1호, 2015, 77-93.
- Al-Otaibi, S.T. and M. Ykhlef, “A survey of job recommender systems”, *International Journal of Physical Sciences*, Vol.7, No.29, 2012, 5127-5142.
- Bayer, I., X. He, B. Kanagal, and S. Rendle, “A generic coordinate descent framework for learning from implicit feedback”, *Proceedings of the 26<sup>th</sup> International Conference on World Wide Web*, 2017, 1341-1350.
- Bennett, J. and S. Lanning, “The netflix prize”, *Proceedings of KDD Cup and Workshop*, 2007, 35.
- Blei, D.M, “Probabilistic topic models”, *Communications of the ACM*, Vol.55, No.4, 2012, 77-84.
- Blei, D., L. Carin, and D. Dunson, “Probabilistic topic models”, *IEEE Signal Processing Magazine*, Vol.27, No.6, 2010, 55-65.
- Cheng, Z., Y. Ding, L. Zhu, and M. Kankanhalli, “Aspect-aware latent factor model : Rating prediction with ratings and reviews”, *Proceedings of the 2018 World Wide Web Conference*, 2018, 639-648.
- Das, A.S., M. Datar, A. Garg, and S. Rajaram, “Google news personalization : scalable online collaborative filtering”, *Proceedings of the 16<sup>th</sup> International Conference on World Wide Web*, 2007, 271-280.
- Griffiths, T.L. and M. Steyvers, “Finding scientific topics”, *Proceedings of the National academy of Sciences*, Vol.101, 2004, 5228-5235.
- Herlocker, J.L., J.A. Konstan, and J. Riedl, “Explaining collaborative filtering recommendations”, *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, 2000, 241-250.
- He, X., T. Chen, M.Y. Kan, and X. Chen, “Tri-rank : Review-aware explainable recommendation by modeling aspects”, *Proceedings of the 24<sup>th</sup> ACM International on Conference*

- on *Information and Knowledge Management*, 2015, 1661–1670.
- He, X., L. Liao, H. Zhang, L. Nie, X. Hu, and T.S. Chua, “Neural collaborative filtering”, *Proceedings of the 26<sup>th</sup> International Conference on World Wide Web*, 2017, 173–182.
- Hu, Y., Y. Koren, and C. Volinsky, “Collaborative filtering for implicit feedback datasets”, In *2008 Eighth IEEE International Conference on Data Mining*, 2008, 263–272.
- Hutterer, M., *Enhancing a Job Recommender with Implicit User Feedback*, Ph.D. thesis, University of Vienna, Vienna, Austria, 2011.
- Jelodar, H., Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao, “Latent Dirichlet allocation(LDA) and topic modeling : models, applications, a survey”, *Multimedia Tools and Applications*, Vol.78, No.11, 2019, 15169–15211.
- Kingma, D.P. and J. Ba, “Adam : A method for stochastic optimization”, *arXiv preprint arXiv:1412.6980*, 2014.
- Koren, Y., “Factorization meets the neighborhood : a multifaceted collaborative filtering model”, *Proceedings of the 14<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008, 426–434.
- Lee, T.Q., Y. Park, and Y.T. Park, “A time-based approach to effective recommender systems using implicit feedback”, *Expert Systems with Applications*, Vol.34, No.4, 2008, 3055–3062.
- Lei, X., X. Qian, and G. Zhao, “Rating prediction based on social sentiment from textual reviews”, *IEEE Transactions on Multimedia*, Vol.18, No.9, 2016, 1910–1921.
- Leksin, V. and A. Ostapets, “Job recommendation based on factorization machine and topic modelling”, *Proceedings of the Recommender Systems Challenge*, 2016, 1–4.
- Li et al., “LDA meets Word2Vec : a novel model for academic abstract clustering”, *Companion Proceedings of the the Web Conference*, 2018, 1699–1706.
- Linden, G., B. Smith, and J. York, “Amazon.com recommendations : Item-to-item collaborative filtering”, *IEEE Internet Computing*, Vol.7, No.1, 2003, 76–80.
- Liu, Q., E. Chen, H. Xiong, C.H. Ding, and J. Chen, “Enhancing collaborative filtering by user interest expansion via personalized ranking”, *IEEE Transactions on Systems, Man, and Cybernetics, Part B(Cybernetics)*, Vol. 42, No.1, 2011, 218–233.
- Lu, Y., Q. Mei, and C. Zhai, “Investigating task performance of probabilistic topic models : an empirical study of PLSA and LDA”, *Information Retrieval*, Vol.14, No.2, 2011, 178–203.
- McAuley, J. and J. Leskovec, “Hidden factors and hidden topics : understanding rating dimensions with review text”, *Proceedings of the 7<sup>th</sup> ACM Conference on Recommender Systems*, 2013, 165–172.
- O’callaghan, D., D. Greene, J. Carthy, and P. Cunningham, “An analysis of the coherence of descriptors in topic modeling”, *Expert Systems with Applications*, Vol.42, No.13, 2015, 5645–5657.
- Onan, A., H. Bulut, and S. Korukoglu, “An improved ant algorithm with LDA-based representation for text document clustering”, *Journal of Information Science*, Vol.43, No.2, 2017, 275–292.
- Park, D.H., H.K. Kim, I.Y. Choi, and J.K. Kim,

- “A literature review and classification of recommender systems research”, *Expert Systems with Applications*, Vol.39, No.11, 2012, 10059-10072.
- Pavlinek, M. and V. Podgorelec, “Text classification method based on self-training and LDA topic models”, *Expert Systems with Applications*, Vol.80, 2017, 83-93.
- Qin, C., H. Zhu, T. Xu, C. Zhu, C. Ma, E. Chen, and H. Xiong, “An enhanced neural network approach to person-job fit in talent recruitment”, *ACM Transactions on Information Systems*, Vol.38, No.2, 2020, 1-33.
- Reusens, M., W. Lemahieu, B. Baesens, and L. Sels, “A note on explicit versus implicit information for job recommendation”, *Decision Support Systems*, Vol.98, 2017, 26-35.
- Seo, S., J. Huang, H. Yang, and Y. Liu, “Interpretable Convolutional Neural Networks with Dual Local and Global Attention for Review Rating Prediction”, *Proceedings of the Eleventh ACM Conference on Recommender Systems*, 2017, 297-305.
- Su, X. and T.M. Khoshgoftaar, “A survey of collaborative filtering techniques”, *Advances in Artificial Intelligence*, 2009, 1-19.
- Tay, Y., L. Anh Tuan, and S.C. Hui, “Latent relational metric learning via memory-based attention for collaborative ranking”, *Proceedings of the 2018 World Wide Web Conference*, 2018, 729-739.
- Wilson, J., S. Chaudhury, and B. Lall, “Improving collaborative filtering based recommenders using topic modelling”, *IEEE/WIC/ACM International Joint Conferences on Web Intelligence(WI) and Intelligent Agent Technologies(IAT)*, Vol.1, 340-346.
- Yang, S., M. Korayem, K. AlJadda, T. Grainger, and S. Natarajan, “Combining content-based and collaborative filtering for job recommendation system : A cost-sensitive Statistical Relational Learning approach”, *Knowledge-Based Systems*, Vol.136, 2017, 37-45.
- Zhang, S., L. Yao, A. Sun, and Y. Tay, “Deep learning based recommender system : A survey and new perspectives”, *ACM Computing Surveys(CSUR)*, Vol.52, No.1, 2019, 1-38.
- Zhang, Z., D. Zhang, and J. Lai, “urCF : user review enhanced collaborative filtering”, *20<sup>th</sup> Americas Conference on Information Systems*, 2014.
- Zhao, X., Z. Niu, W. Chen, C. Shi, K. Niu, and D. Liu, “A hybrid approach of topic model and matrix factorization based on two-step recommendation framework”, *Journal of Intelligent Information Systems*, Vol.44, No.3, 2015, 335-353.
- Zheng, L., V. Noroozi, and S. Yu, “Joint deep modeling of users and items using reviews for recommendation”, *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 2017, 425-434.
- Zhou, X. and S. Wu, “Rating LDA model for collaborative filtering”, *Knowledge-Based Systems*, Vol.110, 2016, 135-143.



## ◆ About the Authors ◆



**이 청 용 (leecy@khu.ac.kr)**

경희대학교 경영학과에서 경영학 학사학위를 취득하고, 동 대학원 빅데이터 응용학과에서 공학 석사학위를 취득하였다. 현재 빅데이터응용학과 박사과정에 재학 중이며, 주요 관심분야로는 개인화 추천 서비스, 자연어 처리, 데이터 마이닝, 빅데이터 분석, 딥러닝 등이다.



**전 상 흥 (ksdp@khu.ac.kr)**

경희대학교 일반대학원 빅데이터응용학과에서 공학 석사학위를 취득하였다. 주요 관심분야는 Recommender Systems, Big Data Analytics, Marketing 등이다.



**이 창 재 (l2nobak@daum.net)**

경희대학교 경영대학원 경영학 석사학위를 취득하였다. 현재 경희대학교 경영학과 박사과정을 수료하였으며, (주) 피엠피아이, (주) 리서치퍼스트 대표이사이다. 수상 경력으로 (사) 한국전문기자협회의 '2019 대한민국 리더대상'과 2019년 경희대학교 총장 우수상 등이 있다. 주요 관심분야로는 개인화 추천 서비스, 자연어 처리, 빅데이터 분석, 통계분석기법연구 및 머신러닝 등이다.



**김 재 경 (jaek@khu.ac.kr)**

서울대학교에서 산업공학학사, 한국과학기술원에서 경영정보시스템 전공으로 석사 및 박사학위를 취득하였으며 현재 경희대학교 경영대학 및 빅데이터 응용학과 교수로 재직하고 있다. 미국 미네소타 주립대학교 그리고 텍사스 주립대학교(달라스)에서 교환교수를 역임하였다. 주요 관심분야로는 개인화 서비스, 추천시스템, 빅데이터, 및 딥러닝 등이다. IEEE Transaction on Services Computing, IEEE Transaction on SMC-A, International Journal of Human Computer Studies, International Journal of Information Management, Information and Management, Expert Systems with Applications, Applied Artificial Intelligence, 등 다수의 학술지에 논문을 게재하였다. 현재 4단계 BK21사업 연구단장(빅데이터 분야) 및 AI 비즈니스 연구센터 센터장을 맡고 있다.