

# Generalized neural collaborative filtering

In Jun Hwang<sup>a</sup>, Hee Ju Kim<sup>a</sup>, Yu Jin Kim<sup>a</sup>, Yoon Dong Lee<sup>1,a</sup>

<sup>a</sup>Business School, Sogang University

## Abstract

In this study, we conduct an exploratory analysis of the MovieLens data, which is frequently used in many recommender system researches, to examine the detailed characteristics of the data. Also, we seek alternatives to improve the well-known neural collaborative filtering (NCF) method. NCF improved matrix factorization method by using deep neural networks in recommender systems. We devise, generalized NCF (G-NCF), a variant of NCF and test the performances. The G-NCF we propose shows superior characteristics on average performance across key evaluation metrics, compared to the NCF, but it also has a slightly larger variance in the evaluation metrics. Evaluation metrics such as MAP and nDCG were considered for comparison.

Keywords: recommender system, collaborative filtering, deep neural network, MovieLens Data

## 1. 서론

최근 급격한 산업 발전과 정보기술의 발달로 인하여 소비자들은 매우 많은 상품과 정보에 노출되고 있으며 개인이 선택해야 하는 서비스와 상품의 종류는 점차 무궁무진하게 증가하고 있다. 따라서 소비자 개인에게 맞춤형 정보를 제공하고 대상 상품을 쉽게 선택하고 만족도를 높이는 데 기여하는 ‘추천시스템’ (recommender system)에 대한 연구가 활발하게 진행되고 있다 (Kim, 2019). 특히 ‘협업필터링’ (collaborative filtering)은 각 ‘사용자’들이 각 대상 ‘상품’에 대하여 평가한 ‘평가기록’에서 사용자들의 상품에 대한 평가의 패턴을 파악함으로써 사용자들이 아직 소비하지 않은 상품 중 향후 선택 가능성이 높은 상품을 추천하는 방법으로 가장 관심을 받고 있는 추천방법 중 하나이다 (Kim, 2021; Seo, 2018).

협업필터링은 상품이나 소비자에 대한 별도의 특성 데이터를 활용하지 않고 소비자들의 평가기록만을 이용하여 평가패턴이 유사한 소비자들을 선정하고 상품을 추천하기 때문에, 추천대상이 되는 목표사용자에게 다양한 범위의 상품을 추천할 수 있는 장점이 있으며, 사용자들의 행동 데이터에서 다양한 ‘잠재 요인’을 추출하여 효과적인 추천이 가능하다. 그러나 적절한 추천을 위해 사용자들의 상품에 대한 피드백 데이터가 축적되고 관리되어야 하기 때문에 평가기록이 없거나 충분히 정보가 축적되지 못한 경우에는 적용이 불가능하거나 성능이 떨어지는 단점이 있다 (Zhang 등, 2021). 효과적인 협업필터링 방법으로 행렬분해법을 이용한 알고리즘이 개발되었고, 최근에는 이를 심층신경망을 이용하여 개선한 신경망 협업필터링 (neural collaborative filtering; NCF) 방법이 제안되었다.

무비렌즈 데이터는 추천시스템 연구를 위하여 공개된 대표적인 연구용 데이터이고, 많은 연구에서 예제 데이터로 활용되고 있다. 그럼에도 불구하고, 무비렌즈 데이터에 대한 통계학적 관점에서의 특성분석을 제공하고 있는 연구는 찾기 힘들다. 본 연구에서는 제 2절에서 무비렌즈 데이터에 탐색적 분석을 수행하여 데이터의 특성을 파악하고, 제 3절에서는 NCF 알고리즘을 개선한 새로운 대안적 방법을 제안하고, 이후 제 4절에서 대안적 방법의 성능을 비교 시험할 것이다.

<sup>1</sup>Corresponding author: Business School, Sogang University, PA 804, BaekBumRo 35, Mapo, Seoul 04107, Korea. E-mail: [widylee@sogang.ac.kr](mailto:widylee@sogang.ac.kr)

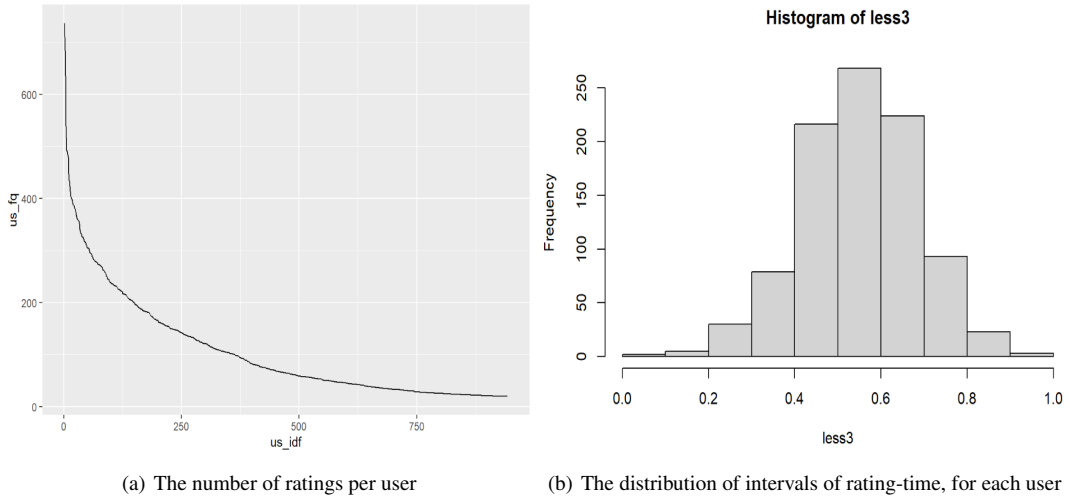


Figure 1: The number of ratings and the distribution of intervals of rating-time (ML-100K).

## 2. 무비렌즈 데이터

본 연구에 활용할 무비렌즈 데이터는 미국 미네소타대학의 GroupLens Research Project에서 연구를 목적으로 수집하여 공개하고 있는 데이터로서 1997년 처음 공개된 이후 지속적으로 다른 형태의 자료들이 제공되고 있다. 이 데이터는 공개 시점과 목적에 따라 여러 버전으로 구성되어 있으며 대체로 사용자의 아이디, 영화의 아이디, 사용자의 평가점수, 평가시각으로 구성되어 있다. 각 데이터는 포함된 평가 횟수에 따라 100K, 1M, 20M로 구별된다. 이 중 100K와 1M 데이터가 관련 연구에서 자주 이용되고 있다 (Harper와 Konstan, 2015).

본 논문에서는 추천시스템의 성능 비교를 위하여 무비렌즈 데이터 중 ‘MovieLens-100K’ (ML-100K) 데이터를 사용할 것이다. ML-100K 데이터는 무비렌즈 데이터셋 중 가장 작은 규모의 데이터로서 1997년 9월 19일부터 1998년 4월 22일까지의 기간 동안, 943명의 사용자들이 1,682편의 영화에 대하여 평가한 100,000건의 평가점수와, 사용자에 대한 정보, 영화에 대한 정보로 구성되어 있다. 평가점수는 사용자의 아이디(user-id)와 영화의 아이디(movie-id), 각 사용자의 평가점수(rating), 평가시각(timestamp)을 나타내는 네 개의 변수로 이루어져 있다. 무비렌즈 데이터의 평가점수는 1점부터 5점까지 1점 단위로 기록되어 있다. 사용자에 대한 정보는 사용자의 아이디, 성별(female/male), 연령대(age), 직업(occupation), 우편번호(zip)가 포함되어 있으며 영화에 대한 정보는 영화 아이디, 영화 제목, 영화 개봉일(release-date), 비디오 발매일(video-release-date), 영화 DB 인터넷 주소(imdb-url), 영화 장르로 구성되어 있다.

Figure 1은 ML-100K 데이터에서 943명의 사용자들을 평가 횟수별로 정렬하여 사용자별 평가 횟수에 대한 패턴과 사용자들이 영화에 평점을 부여한 시간 간격을 보여주고 있다.

Figure 1(a)에서 x-축 변수 us\_idf는 평가횟수가 가장 많은 사용자를 1 그 다음 많은 사용자를 2와 같이 평가횟수에 따라 사용자들의 순서를 새로 부여한 값이다. Figure 1(a)를 통하여, 각 사용자에 대한 평가횟수의 변화를 보여주고 있다. 평가횟수가 가장 높은 사용자는 737건 평가를 했다. 특정 사용자가 여러 영화를 평가하게 되므로, 동일 사용자들 사이에 평가의 시간간격을 구할 수 있다. 동일한 사용자에 의한 평가 시간간격의 특징을 알아보기 위하여, 각 사용자 별로 각자의 평가 시간간격 중에서 3초 이내에 평가한 비율(less3)을 구하여 이 값을 기준으로 히스토그램을 그리면 Figure 1(b)와 같다. Figure 1(b)에 나타난 바와 같이, 평가 시간간격 중 3초 이내 평가횟수의 비율이 50%인 경우가 가장 많으며, 이는 약 250명 정도인 것을 알 수 있다. 대체로 각 사용자가 매우 짧은 시간 간격으로 여러 영화에 대하여 평가한 것임을 알 수 있다.

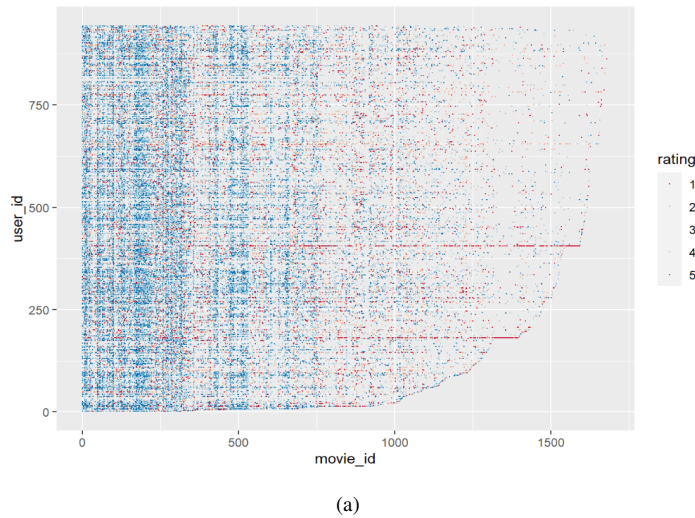


Figure 2: Ratings in (user, movie) plane.

Figure 2는 ML-100K 데이터에서 사용자(user\_id)와 영화 아이디(movie\_id) 두 개의 변수로 생성되는 평면에, 평가기록이 있는 (영화, 사용자) 순서쌍의 위치에 ‘평가점수’를 색깔로 나타낸 그림이다. 점수가 낮은 1점은 붉은색으로, 점수가 높은 5점은 파란색으로 표현했다(그림에서의 색깔은 종이 논문에서는 확인이 어려울 수 있으나, 인터넷에서 배포되는 논문 파일을 통하여 확인이 가능하다). 그림에서 볼 수 있듯이 일부 영화를 제외하고는 사용자들에 의해 평가가 부여된 경우가 적으며 전체적으로 데이터가 희박한 특징을 확인할 수 있다. 그림에서 살펴보면 영화 중 아이디가 앞에 속하는 영화들은 많은 사용자들이 좋은 평가를 부여한 인기 있는 영화들인데 반하여 뒤쪽 번호에 속하는 영화들은 사용자들에게 잘 선택되지 않고 점수도 낮게 평가된다는 것을 알 수 있다. 특히 영화 아이디 번호가 높고 사용자 아이디 번호가 낮은 경우에 해당하는 우측 하단에는 평가기록이 전혀 나타나지 않는다. 이는 큰 번호를 갖는 movie\_id에 해당하는 영화들이, 자료 수집 과정에서 나중에 추가된 영화들이고, 작은 번호를 갖는 user\_id에 해당하는 사용자들이 대체로 나중에 가입한 사용자일 가능성이 있음을 보여준다.

Figure 3는 ML-100K 데이터에서 평가횟수를 기준으로 평가횟수가 높은 영화부터 번호를 부여하여 아이디를 만들고(mv\_idf), 마찬가지로 평가횟수가 많은 순서로 사용자 아이디(us\_idf)를 만들어, 앞서 Figure 2와 마찬가지로, 사용자 and 영화로 이루어진 평면에 평점을 색깔로 표시한 것이다(반대로 낮은 점수에 파란색 높은 점수에 빨간색을 주는 경우 본 논문에서와 같은 자료의 특징을 확인하기 힘들다). 그림에서 살펴보면 영화 중 아이디가 앞에 속하는 영화들은 많은 사용자들이 좋은 평점을 부여한 인기 있는 영화들인데 반하여 뒤쪽 번호에 속하는 영화들은 사용자들에게 잘 선택되지 않고 평점도 낮다는 것을 알 수 있다. 사용자 아이디가 앞쪽에 위치한 사용자는 많은 영화를 평가하였고, 평가횟수가 낮은 영화들에 대하여도 평가를 한 반면 매우 낮은 평점을 부여한 것을 볼 수 있다.

영화 평점 부여 시각(unix timestamp)을 횡축으로, 영화 아이디를 종축으로 배치하였을 때 Figure 4와 같이 몇 군데에서 붉은색 굵은 세로 줄이 나타나는 것을 볼 수 있다. 이는 일부 사용자가 매우 짧은 시간에 많은 영화들에 대하여 연이어 매우 낮은 평점을 부여하여 발생하는 현상이다. Figure 1에서의 평가 시간간격과 Figure 4가 보여주는 특징은, 무비렌즈 데이터가 개개의 영화를 감상하고, 감상한 영화에 대한 평가점수를 입력하는 방식으로 수집된 자료가 아니고, 단순히 사용자들이 자신들이 알고 있는 영화에 대한 평점을 한꺼번에

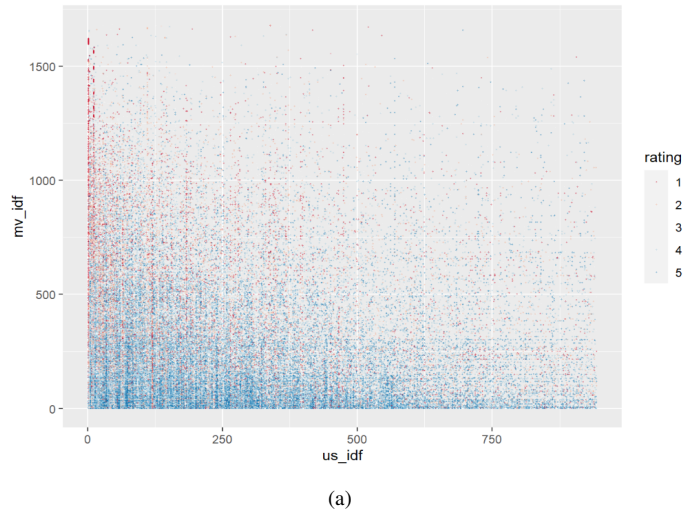


Figure 3: Ratings in (user, movie) plane, rearranged in their order of the number of ratings.

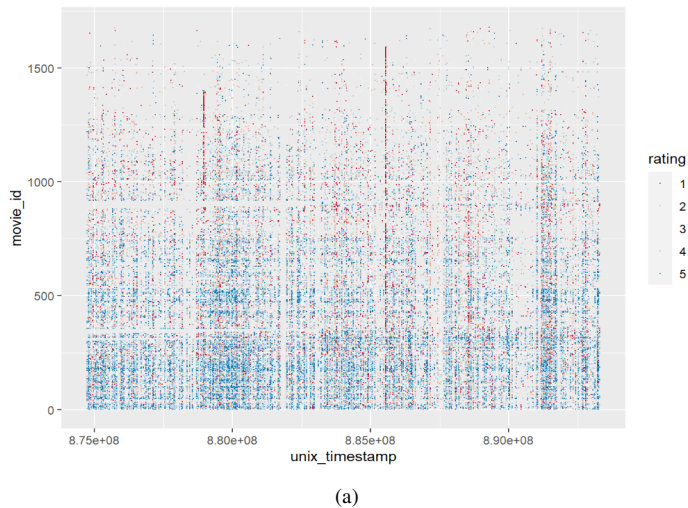


Figure 4: Ratings for movies at rating-time.

부여하는 방식으로 이루어진 것이라는 것을 시사한다. 또한 평가에 참여한 사용자들에게 평가횟수에 따른 인센티브가 제공되었을 가능성도 짐작할 수 있다.

### 3. NCF 알고리즘과 그 개선

추천시스템의 구현 및 테스트를 위하여 제공되는 Microsoft사의 Python Recommenders 패키지에는 다양한 추천시스템 알고리즘이 포함된다. Recommenders 패키지 포함된 여러 알고리즘들 중, 성능이 우수한 대표적인 방법으로 SAR, BPR, NCF을 들 수 있다.

SAR (simple algorithm for recommendation) 알고리즘은 비교적 간단한 방법으로서 사용자(user)의 상품(item)에 대한 상호작용(평가, 구매, 검색, 클릭 등)을 사용자 및 상품 간 유사도를 기반으로 평가함으로써 사용자가 기존에 상호작용한 상품과 가장 유사한 상품을 추천하는 방법이다. BPR (Bayesian personalized ranking) 방법은 기존의 추천 시스템이 사용자 및 상품간 상호작용 데이터가 결측치로 인한 희소성 때문에 정교한 추천에 한계가 있다는 점에 착안하여, 동일 사용자에 의하여 평가된 영화들 사이에 선호도의 우선순위가 존재하는지의 여부를 모형화하는 방법을 이용하였다. BPR에서는, 사용자의 선호도가 높은 상품을 추천하기 위한 방법으로, 모형 구성과정에서 릿지(ridge) 벌칙항을 부과하여, 기존에 획득된 자료에 대한 적합성에 매몰되지 않고, 사용자가 선택할 가능성이 높은 상품을 잘 ‘예측’하여 사용자에게 추천하는 방법을 사용하였다 (Rendle 등, 2012). Lee 등 (2020)은 여러 추천시스템 알고리즘에 대하여 그 성능을 비교 평가하였다.

본 논문에서는 Recommenders 패키지에서 제공되는 이들 알고리즘 중, NCF (neural collaborative filtering) 방법을 중심으로 그 특징을 살펴보고, 개선 대안을 모색한다. 먼저 NCF 알고리즘의 작동 방식을 살펴보고, 문제점과 개선 대안을 살펴보자.

### 3.1. NCF 알고리즘

NCF 알고리즘은 He 등 (2017)이 제안한 심층신경망 기반 알고리즘으로, 다양한 데이터에 대하여 높은 성능을 보여주는 것으로 알려져 있다. NCF 알고리즘은 기존의 행렬분해 방식에 기반한 협업 필터링 알고리즘을, 심층신경망을 이용하여 일반화한 방법이다. NCF 알고리즘은, 특정 사용자와 특정 영화 사이에 평가 기록이 이루어질 확률을 모형화 하는 방법으로 추천 우선순위를 도출한다. 사용자  $u$ 의 상품  $i$ 에 대한 평가기록이 있는 경우  $y_{u,i} = 1$ 이라 하고, 그렇지 않은 경우를  $y_{u,i} = 0$ 이라 하자. 평가기록이 있는  $(u, i)$ 쌍들의 집합을  $R^+$ 라 하고(즉,  $y_{u,i} = 1$ 인 경우), 평가기록이 없는 쌍들의 집합을  $R^-$ 라 하자(즉,  $y_{u,i} = 0$ 인 경우).

NCF 알고리즘은 0과 1로 이루어진 이진값을 갖는 관측값  $y_{u,i}$ 을 모형화하기 위한 방법으로, 심층신경망을 이용하여 행렬분해법을 일반화한 generalized matrix factorization (GMF) 망과 multilayer perceptron (MLP) 망을 결합하여 얻는 함수를 이용한다. 신경망에 입력으로 사용하는 값은, 사용자( $u$ )와 상품( $i$ )에 대한 정보이고, 이를 임베딩하여 사용한다. 사용자와 상품에 대한 정보는 사용자번호와 상품번호를 one-hot-encoding 방법으로 가변수 형태로 변환하고, 이를 동일한 차원의 (행)벡터인  $\mathbf{p}_u$ 와  $\mathbf{q}_i$ 로 임베딩한다. 사용자와 상품에 대한 임베딩벡터  $\mathbf{p}_u$ ,  $\mathbf{q}_i$ 와 상수항(bias)을 결합하여 얻게 되는 (행)벡터를  $\mathbf{r}_{(u,i)} = (b_0, \mathbf{p}_u, \mathbf{q}_i)$ 라 하자.

MLP 망을 통한 출력값은 다음과 같이 표현된다.

$$\Phi_{(u,i)}^{MLP} = [[\mathbf{r}_{(u,i)} W_1] W_2] \cdots W_K$$

와 여기서 대괄호  $[ ]$ 는 매번 각기 적절하게 선택되어 적용되는 활성화함수를 의미하고, 행렬  $W_1, \dots, W_K$ 는 각 층에서 적용되는 가중치행렬이다. 또 GMF 망을 통한 출력값은 다음과 같이 표현된다.

$$\Phi_{(u,i)}^{GMF} = [(\mathbf{p}_u \odot \mathbf{q}_i) \mathbf{w}], \quad (3.1)$$

여기서 대괄호  $\odot$ 는 원소별 곱을 의미하고,  $\mathbf{w}$ 는 각 좌표에 대한 가중치를 부여하기 위하여 고려하는 (열)벡터이다. 임베딩벡터의 차원을  $d$ 라고 하면,  $\mathbf{p}_u = (p_1, \dots, p_d)$  이고,  $\mathbf{q}_i = (q_1, \dots, q_d)$  일 때,  $\mathbf{p}_u \odot \mathbf{q}_i = (p_1 q_1, \dots, p_d q_d)$  이다. NCF 알고리즘은,  $y_{u,i}$ 에 대한 예측값으로 다음과 같이 구한  $\hat{y}_{u,i}$ 을 사용한다.

$$\hat{y}_{u,i} = \left[ \left( \Phi_{(u,i)}^{GMF}, \Phi_{(u,i)}^{MLP} \right) \mathbf{h} \right],$$

여기서  $\mathbf{h}$ 는 가중치벡터이다.

이때 추천(학습)대상이 되는 모수는 바이어스  $b_0$ 와 가중치 행렬  $W_1, \dots, W_K$ , 가중치벡터  $\mathbf{w}$ ,  $\mathbf{h}$ 이다. He 등 (2017)은 이들 모수에 대한 추정을 위하여, 다음과 같이 정의되는 binary cross-entropy를 최소화 하는 방법을

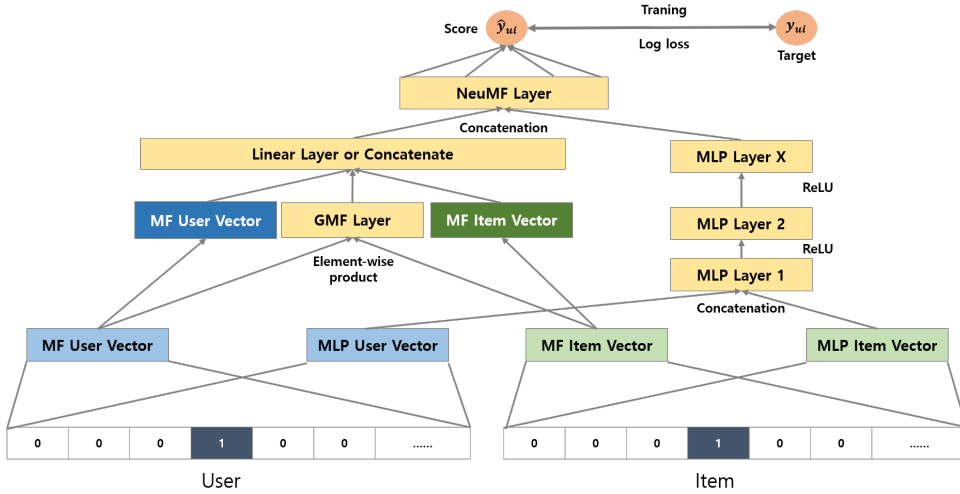


Figure 5: The structure of G-NCF.

제안하였다.

$$L = - \sum_{(u,i) \in R^+ \cup R^-} \{y_{u,i} \log \hat{y}_{u,i} + (1 - y_{u,i}) \log (1 - \hat{y}_{u,i})\}.$$

### 3.2. 일반화 NCF 알고리즘

NCF 알고리즘은 사용자와 해당 상품 간 상호작용이 발생할 가능성을 예측하는 것을 목적으로 사용자 · 상품 간에 잘 어울리는 내제된 차원의 변수가 존재할 수 있다는 점을 고려하여 이를 모형화 하기 위한 방법으로 GMF 망을 고려하고, 여기에 MLP 망을 결합시킨 것이다. GMF 망에서는 임베딩벡터로 축약된 사용자 정보와 상품 정보를 입력으로 받아, 임베딩벡터의 원소별 곱을 구하여, (3.1)과 같이 변환하여 그 출력을 얻는다. 이 과정에서 두 가지 개선 가능성을 찾아볼 수 있다.

하나는, 사용자 정보와 상품 정보를 축약한 임베딩벡터는 통계학에서의 잠재변수(latent variable)로서의 역할을 하므로, 이 과정에 층별정규화(layer normalization) 방법을 적용할 수 있다는 점이다. Layer normalization은 Ba 등 (2016)에 의하여 제안된 방법으로, 하나의 관측단위에 해당하는 여러 개의 잠재변수들이 갖는 값들을, 마치 하나의 변수가 여러 개의 관측값들로 이루어진 경우인 것처럼 평균과 표준편차를 구하고 표준화하여 이용하는 것을 말하고, 이 경우 심층신경망의 훈련 속도를 증가시킬 수 있고, 더 우수한 결과를 얻게된다고 알려져 있다. 보통 layer normalization은 batch normalization과 대비되는 개념으로 자주 언급된다. Batch normalization은, 여러 개의 관측단위가 묶여 하나의 배치로 처리되는 과정에서, 각 배치를 마치 전체 데이터인 것처럼 생각하고, 포함된 각 변수들을 배치단위로 표준화하여 적용하는 방법을 말한다. Batch normalization은 배치단위로 진행되는 심층신경망의 최적화(학습)과정에서, 각 배치의 동일성을 증가시켜서 신경망의 최적화 속도를 높이는 방법이다.

다른 하나는 (3.1)에 의하여 표현되는 함수에는, 사용자 정보와 상품 정보가 곱셈이라는 대칭적 구조를 통하여 모형에 개입된다는 제약이 내포되어 있고, 이를 일반화하여 대칭성 제약을 갖지 않는 방법을 고려할 필요가 있다. 이를 위하여 다음과 같은 방법으로, 사용자 임베딩벡터와 상품 임베딩 벡터를 원소별로 곱하여 얻게 되는,  $\tilde{p}_u \odot \tilde{q}_i$  항 외에, 사용자 임베딩벡터와 상품 임베딩벡터에 대한 정보를 별도로 포함하여 일반화하는

Table 1: Comparisons between search system and recommender system

	Search system	Recommender system
Usages	Response to queries	Item recommendation to users
Results	Search list of documnets	Recommendation list of items
Relevance function	0, 1, 2, ...	0, 1
Assumption	The same query can be issued infinitely many times	The same user selects items only in finite times

방법을 고려할 수 있다.

$$\Phi_{(u,i)}^{GMF} = [(\tilde{p}_u, \tilde{p}_u \odot \tilde{q}_i, \tilde{q}_i) \mathbf{w}], \quad (3.2)$$

여기서  $\mathbf{w}$ 는 적절한 차원을 갖는 가중치벡터이고,  $\tilde{p}_u$ 와  $\tilde{q}_i$ 는 각각 사용자 임베딩벡터  $p_u$ 와 상품 임베딩벡터  $q_i$ 이거나, 혹은 각각을 층별 정규화한 벡터이다. Figure 5은 G-NCF의 전체 구조를 그림으로 나타낸 것이다. 사용자 임베딩벡터와 상품 임베딩벡터를 처리하는 MLP와 별도로 GMP가 구성된다. GMP에는 사용자 임베딩벡터와 상품 임베딩벡터 각각과 두 임베딩벡터의 곱으로 이루어지는 벡터가 입력으로 사용된다.

다음 절에서는, 먼저 추천시스템의 성능 평가를 위한 척도들에 대하여 검토하고, 이어서 (3.1)을 이용하는 NCF 알고리즘과, (3.2)를 이용하는, 일반화 NCF (Generalized NCF) 방법의 성능을 비교한다. 시뮬레이션 실험결과에 의하면 일반화 NCF (G-NCF) 방법이 다소간의 장점이 있는 것으로 나타난다.

## 4. 성능의 비교와 평가

### 4.1. 평가 지표

추천시스템의 성능 평가는 추천된 상품이 실제로 사용자들에게 얼마나 채택되었는지를 정량화하여 비교, 평가한다. 추천된 상품목록과 실제 선택된 상품목록을 비교하고 정량화 하기 위하여 다양한 평가지표가 사용된다. 본 논문에서는 Precision 및 Recall, nDCG, MAP 값을 중심으로 추천시스템의 성능을 평가하였다. MAP과 nDCG는 순서학습(learn to rank) 분야에서 연구가 축적된 평가방법으로, 관련된 여러 연구가 있다(ref., Zhang 등, 2012; Chaudhury 등, 2017).

추천시스템은 각 사용자별로 상품 추천목록을 작성하고, 사용자가 선택한 상품 중 몇 개가 추천목록에 포함되었는지에 따라 그 성능을 평가할 수 있다. 추천시스템의 성능을 평가할 때 영향을 미치는 기본적인 요소는 추천목록의 크기(길이)  $K$ 와, 사용자가 선택한 전체 상품의 개수  $L$ 이다. 사용자가 선택한 상품 중 추천목록에 있는 상품의 개수가  $s$ 라면, 정확도(precision)는  $s/K$ 이고, 재현율(recall)은  $s/L$ 이다.

추천시스템은 사용자들에 대하여 추천하는 상품의 추천목록을 작성하는 것인데 비하여, 검색시스템은 질의(query)에 대하여 문서의 검색목록을 작성하여 제시한다는 점에서 추천시스템과 검색시스템은 유사점이 많다. 추천목록과 검색목록은 결과로 제시된 순서에 의미를 부여하여 해석할 수 있다. 검색시스템에서 검색 목록에 나타나는 각 문서와, 추천시스템에서 추천목록에 나타나는 각 상품의 적절성(relevance)이 성능평가의 기준이 된다. 함수  $\text{rel}(k)$ 은 검색목록과 추천목록에 위치한  $k$  번째 문서 혹은 상품에 대한 적절성을 나타내는 함수다. 검색시스템에서  $\text{rel}(k)$ 는 0, 1, 2, ...와 같이 다양한 값을 갖는데 비하여, 추천시스템에서  $\text{rel}(k)$ 는 비선택(0)과 선택(1) 두 가지 값을 갖는 것으로 단순화 되지만, 대신 각 사용자가 상품을 선택하는 횟수는 각기 다르고 유한한 횟수만큼 반복하여 상품을 선택할 수 있다고 본다. Table 1에는 검색시스템과 추천시스템의 특징이 비교되어 있다.

nDCG (normalized discounted cummulative gain)는 검색시스템 성능 평가를 위하여 개발된 지표이고, 추천시스템에도 적용될 수 있다. 추천시스템이나 검색시스템에서 nDCG는 DCG를 IDCG (Ideal DCG)로 나누어

구한다. 즉,  $nDCG = DCG/IDCG$  이다.  $IDCG$ 는 동일한 검색목록 혹은 추천목록에서 순위를 바꾸어 구할 수 있는  $DCG$ 의 최대값이다. 목록에서의 순위를  $rel(k)$  값이 높은 순서로 배치했을 때 얻게되는  $DCG$  값이  $IDCG$  이다. 추천시스템에서  $rel(k)$ 는 0 혹은 1인 값을 갖게 되므로,  $DCG$ 와 함께,  $IDCG$ 는 다음과 같이 구한다.

$$DCG = \sum_{k=1}^K rel(k) / \log_2(1+k),$$

$$IDCG = \sum_{k=1}^s 1 / \log_2(1+k).$$

MAP (mean average precision)는 추천시스템의 성능 평가를 위하여 개발된 대표적인 지표이다. 추천시스템이 특정 사용자( $u$ )에 대하여 작성한 추천목록의 크기가  $K_u$ 이고, 사용자가 선택한  $L_u$  개의 상품 중,  $s_u$ 개의 상품이 추천목록에 포함되어 있었다면, 정확도는  $s_u/K_u$ 이다. 이때 정확도를  $Precision@K_u$ 와 같이 나타내거나 그냥  $Precision@K$  라고 나타낸다. AP (average precision)는 추천목록에서의 추천순위를 고려하여 개발된 지표이다. 크기  $K$ 인 추천목록의 순위  $k$  이내에  $s_k$ 개의 상품이 선택되었다면, 추천목록 순위  $k$ 이내에서의 정확도는  $s_k/k$ 이다. 이로부터  $AP@K$ 는 다음과 같이 구한다.

$$AP@K = \frac{1}{s} \sum_{k=1}^K \left\{ \left( \frac{s_k}{k} \right) rel(k) \right\}$$

$$= \frac{1}{s} \sum_{k=1}^K \{ c_k \cdot rel(k) \}, \quad (4.1)$$

여기서  $c_k$ 는, 추천시스템에서의 적절성함수  $rel(k)$ 의 단순성을 보충하는 기능을 하는 항으로 다음과 같다.

$$c_k = \sum_{j=k}^K \frac{1}{j} rel(j),$$

즉, 추천목록에 있는 상품에 대한 적절성은 사용자의 선택 여부에 따라 0 혹은 1 두 가지로만 구별되지만, 사용자가 선택한 상품 중 추천목록에 포함된 상품의 개수가 많고, 추천목록에 있는 상품들 중에서도 추천순위가 높은 상품일수록 ‘실질적 적절성’이 높다고 판단할 수 있다. 이런 점을 반영하여 적절성 함수  $rel(k)$ 에  $c_k$ 를 곱하여 ‘실질적 적절성’이 반영되도록 한 것이다.  $c_k$  값은 추천목록에서의 순위  $k$  번째 이후 순위로 추천되고 사용자에게 의하여 선택된 상품들에 대하여, 추천순위의 역수를 더하여 계산한다. 식 (4.1) 에서  $s$ 로 나누는 항 (즉,  $1/s$ )은  $nDCG$ 를 구하는 과정에서  $DCG$ 를 그 최대값인  $IDCG$ 로 나누어 normalize 하는 것과 동일한 역할을 한다. 이때 MAP는, 다음과 같이, 전체 사용자에 대하여  $AP@K$ 의 평균을 구한 것이다.

$$MAP@K = \frac{1}{U} \sum_{u=1}^U AP@K_u,$$

여기서  $U$ 는 추천목록이 제시된 전체 사용자 수이다.

## 4.2. 성능 평가 방법과 결과

앞서 세부적으로 살펴본 무비렌즈 ML-100K 데이터를 학습을 위한 훈련 데이터와 평가를 위한 테스트 데이터로 분할하되 의미있는 성능 평가를 위하여 학습 데이터와 테스트 데이터는 시간순으로 정렬하여 이용한다. 추천시스템의 성능 평가는 시간적으로 앞선 데이터를 학습한 이후 추천 영화를 도출하고, 시간적으로 뒤에 발생하는 테스트 데이터와 비교하여 평가하는 방식으로 진행하게 된다. Recommenders 패키지에서 제공하는



Table 2: Comparisons of NCF and G-NCF

	Dimension	MAP@10	nDCG@10	Precision@10	Recall@10
NCF	8	0.0524 (0.001)	0.2099 (0.003)	0.1890 (0.003)	0.1064 (0.002)
G-NCF	8	0.0522 (0.002)	0.2095 (0.005)	0.1870 (0.004)	0.1057 (0.003)
NCF	16	0.0535 (0.001)	0.2108 (0.002)	0.1889 (0.002)	0.1086 (0.002)
G-NCF	16	0.0543 (0.002)	0.2148 (0.005)	0.1917 (0.004)	0.1087 (0.002)
NCF	32	0.0516 (0.001)	0.2025 (0.003)	0.1815 (0.003)	0.1072 (0.002)
G-NCF	32	0.0534 (0.001)	0.2111 (0.004)	0.1884 (0.003)	0.1082 (0.002)
NCF	64	0.0489 (0.001)	0.1926 (0.003)	0.1722 (0.003)	0.1035 (0.002)
G-NCF	64	0.0525 (0.002)	0.2051 (0.004)	0.1827 (0.004)	0.1072 (0.002)

\*Standard deviations, in parentheses

chrono split 함수를 이용하여 이러한 시간순서 분할(chronological split) 작업을 손쉽게 수행할 수 있다. 본 연구에서는 학습 및 테스트 데이터 비율을 75 : 25 로 하였으며 평가에 활용되는 데이터 분할은 동일 시드(seed)를 부여하여 동일 환경에서 비교 평가를 할 수 있도록 하였다.

성능 평가는 NCF와 G-NCF 알고리즘을 조율 모수를 변화시켜 가며 각 30회씩 시뮬레이션을 통해 결과를 도출하였다. 조율모수는 선행연구인 He 등 (2017)과 Lee 등 (2020)에서 시험된 경우를 참고하여 선정하였다. NCF와 G-NCF는 MLP layer는 동일하고 GMF 부분에서 차이가 있기 때문에 MLP layer의 차원 수는 [32, 16, 8]로 고정하고 GMF 및 사용자 임베딩벡터와, 상품 임베딩벡터의 차원을 8, 16, 32, 64로 변경하면서 테스트를 진행하였다. He 등 (2017)의 연구에 의하면, 임베딩벡터의 차원(dimension)이 높아질수록 예측 품질이 올라가지만 너무 높을 경우 과적합의 문제가 발생할 가능성이 있다고 언급하고 있다. 본 연구에서는 64차원을 최대로 고려하였고, 그 외 학습에 소요되는 시간과 최적화 하기까지의 안정성을 고려하였다.

알고리즘별 각 조율모수에 따라 도출된 평가지표는 Table 2와 같다. 추천목록의 크기  $K$ 는 10으로 하였고, 네 가지 척도(MAP, nDCG, Precision, Recall)를 활용하여 평가하였다. Table 2에서의 값들은 신경망 학습과정에서 에폭(epoch)을 30으로 설정하여 시험한 경우이다. 에폭이 50인 경우도 시험하였으나, 전반적으로 30인 경우와 비슷하지만 30인 경우에 비하여 성능이 약간 떨어지는 것으로 나타나 30인 경우만을 수록하였다. 학습률은 0.001, 배치의 크기(batch size)는 1024로 동일하게 진행하였다.

Table 2를 살펴보면, 임베딩벡터의 차원(dimension)이 16인 경우가 상대적으로 성능이 좋은 것으로 나타났다, G-NCF가 NCF에 비하여 미세하게 성능이 우수하지만, 반면 시뮬레이션 결과 전체적으로 성능의 변동성 또한 약간 더 큰 것으로 나타났다. 이는 Figure 6를 통하여 잘 살펴볼 수 있다. Figure 6에서 G-NCF는 여러 경우들에 있어서 NCF에 비하여 약간 우수한 성질을 보이고 있지만, 상대적으로 그 값의 변동이 더 크다. 그 이유는 앞서 (3.2)에서, 임베딩벡터를 통한 사용자 정보와 상품정보가 가공되지 않은 상태로 다음 단계의 입력으로 주어지는 데서 오는 효과로 보인다.

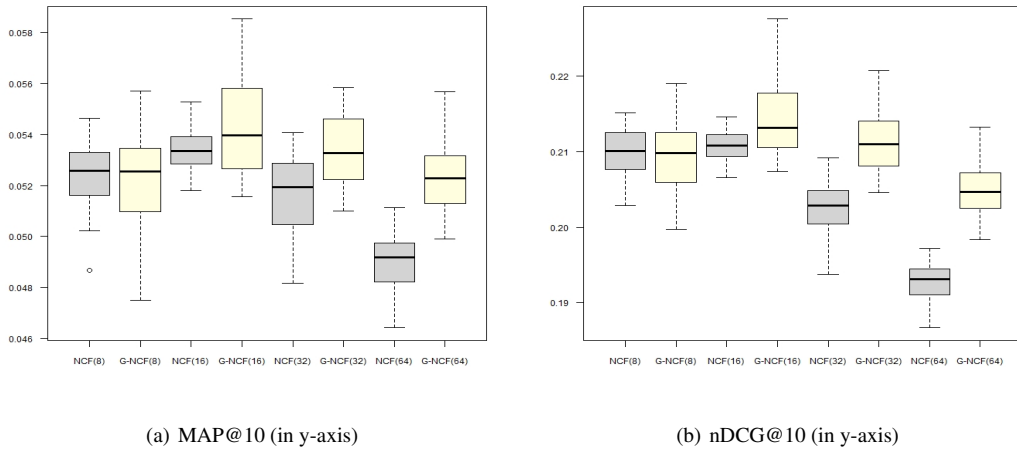


Figure 6: Comparisons of NCF and G-NCF.

## 5. 결론

본 연구에서는 추천시스템 연구에서 널리 이용되는 무비렌즈 데이터(ML-100K)에 대한 탐색적 데이터분석(EDA)를 통하여 데이터의 특징을 살펴보고, 협업필터링 방법 중, 알고리즘이 상대적으로 단순하면서도 우수한 성능을 보이는 NCF 방법의 특징과 그 개선 가능성을 시험해 보았다. 본연구에서 제안한 G-NCF 방법이 기존의 NCF 방법에 비하여 평균적인 성능지표에서는 우수한 점이 있었지만, 그 지표값의 편차가 다소 커진다는 한계를 보였다. 한계를 극복할 만한 더 좋은 알고리즘 개발을 위한 꾸준한 연구가 계속될 필요가 있다고 보인다. 본 연구를 통하여 살펴본 사항들이 고려되어 이후 더 나은 알고리즘 개발에 기여하게 되고, 소비자들이 보다 편리하게 각자 자신에게 맞는 상품을 선택하는데 도움이 되길 바란다.

## References

- Ba JL, Kiros JR, and Hinton GE (2016). Layer normalization, Available from: *pre-print*: arXiv:1607.06450
- Chaudhury S and Tewari A (2017). Online learning to rank with top-k feedback, *Journal of Machine Learning Research*, **18**, 1–50.
- Harper FM and Konstan JA (2015). The MovieLens datasets: History and context, *ACM Transactions on Interactive Intelligent Systems (TiiS)*, **5**, 1–19.
- He X, Liao L, Zhang H, Nie L, Hu X, and Chua TS (2017). Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, Perth, Australia, 173–182.
- Kim HJ (2019). A study on collaborative filtering based on clustering method (Master’s thesis), Sogang University, Seoul.
- Kim SE (2021). A study on collaborative filtering using deep neural network (Master’s thesis), Sogang University, Seoul.
- Lee S, Cho Y, Lee JS, and Yu D (2020). Comparative study of recommender systems using movie rating data, *Journal of the Korean Data & Information Science Society*, **31**, 975–991.

- Rendle S, Freudenthaler C, Gantner Z, and Schmidt-Thieme L (2012). BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, Montreal, QC, 452–461.
- Seo HJ (2018). Comparative study of heuristic approaches for collaborative filtering (Master’s thesis), Sogang University, Seoul.
- Zhang Y, Li Q, Choi I, and Kim JK (2021). *Information Systems Review*, **23**, 155–172.
- Zhang P and Su W (2012). Statistical inference on recall, precision and average precision under random selection. In *Proceedings of 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2012)*, Chongqing, 1348–1352.

*Received February 15, 2024; Revised March 7, 2024; Accepted March 9, 2024*

## 일반화 신경망 협업필터링

황인준<sup>a</sup>, 김희주<sup>a</sup>, 김유진<sup>a</sup>, 이운동<sup>1,a</sup>

<sup>a</sup>서강대학교 경영학부

---

### 요약

본 연구에서는, 추천시스템 연구에 자주 활용되는 무비렌즈 데이터에 대한 탐색적 분석을 통하여 무비렌즈 데이터의 자세한 특성을 살펴보고, 추천시스템에서 심층신경망을 이용한 협업필터링 (NCF) 방법으로 잘 알려진 신경망행렬분해법을 개선하기 위한 대안을 모색한다. 본 연구에서, 제안한 일반화 NCF (G-NCF) 방법은 기존의 NCF 방법에 비하여 주요 평가 지표에서 평균적으로 우수한 특성을 보이지만, 평가지표의 산포가 다소 커지는 단점도 함께 가진다. 성능 비교를 위한 평가 지표로 MAP와 nDCG 등을 이용하였다.

주요용어: 추천시스템, 협업필터링, 심층신경망, 무비렌즈 데이터

---

<sup>1</sup>교신저자: (04107) 서울시 마포구 백범로 35, 바오로관 804호, 서강대학교 경영학부. E-mail: widylee@sogang.ac.kr