

# 딥러닝을 활용한 개인 성향 분석에 맞춘 여행 추천시스템

김솔비<sup>1</sup>, 조창석<sup>2</sup>

<sup>1</sup>한신대학교 정보통신학과 석사과정

<sup>2</sup>한신대학교 정보통신학과 교수

so7345@naver.com, cscho@hs.ac.kr

## A travel recommendation system tailored to personal tendency analysis using deep learning

Sol-Bi Kim<sup>1</sup>, Chang-Suk Cho<sup>2</sup>

<sup>1</sup>Dept. of Information and Telecommunications, Han-Shin University

<sup>2</sup>Dept. of Information and Telecommunications, Han-Shin University

### 요 약

본 연구에서는 기존 여행지 추천의 플랫폼에 있어 개인의 취향에 맞는 여행지 추천이 어렵다는 점을 해결하고자, 비선형적 관계를 해결할 수 있는 NCF 심층신경망 추천시스템을 이용하여 개인의 성향에 따라 여행지를 추천해 주는 시스템을 제안하고 이를 평가한 결과를 보고한다.

### 1. 서론

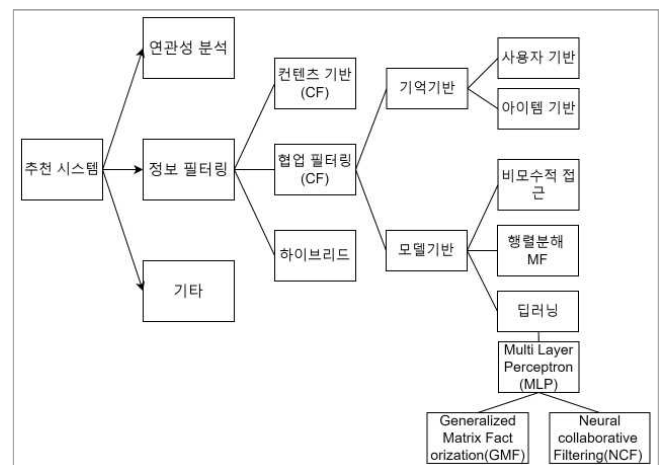
과거에는 사용자가 접하는 상품이 한정적이었지만, 산업이 발달하고 웹, 모바일 환경이 발전하면서 다양한 정보와 많은 양의 콘텐츠를 접하게 될 수 있게 되었다. 모든 정보를 알아볼 수 없고, 자신의 취향에 맞는 정보도 얻기 쉽지 않은 속에서 제공된 서비스가 바로 추천시스템이다. 추천시스템은 이미 다양한 산업(음악, 영화, 판매상품)에서 사용되고 있다. 그러나 추천시스템 서비스가 가장 필요하다면, 그것은 바로 관광 산업이다. 코로나 시기로 인해 주춤했던 관광 산업이 이제 전보다도 더 활발해지기 시작하면서 사용자들은 각종 블로그, 인스타그램, 페이스북을 통해서 많은 여행 정보들을 얻을 수 있게 되었다. 하지만 정보량이 워낙 많다 보니 막상 자신이 가고자 하는 여행 계획을 짤 때면 많은 시간과 노력이 필요하다.

현재 여행 플랫폼에서는 여행지, 맛집, 카페, 숙소 등 아주 많은 정보를 사용자에게 제공하고 있는데, Trip Advisor[1] 같은 경우 여행지를 추천할 때 기준은 인기 있는 곳을 추천해 주거나, 사용자가 우선순위로 설정한 카테고리(가성비, 요금, 아동)를 토대로 추천해 주는 정적인 부분이 많이 존재한다. 이렇게 되면 사용자가 앱을 사용할 때마다 설정해야 하는 불편함과 추천해 주는 여행지 또한 범위가 넓어 그 결과가 만족스럽지 못하다. 이러한 점을 고려

하여 본 연구에서는 기계 스스로 사용자의 데이터를 학습하여 여행 성향을 나누었고, 각 여행 성향에 해당하는 여행지들을 유사도로 계산하여 top-k 형식의 추천시스템 모델로 설계하도록 구축하였다.

### 2. 고객의 선호를 추천하는 기존 추천시스템 분석

추천시스템은, 사용자가 관심을 가질 만한 콘텐츠를 추천하는 것으로, 사용자의 선호도 및 과거 행동을 토대로 사용자에게 적합한 콘텐츠를 제공한다.



(그림 1) 추천시스템 종류.

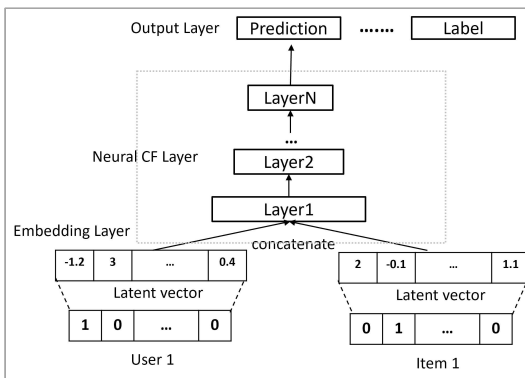
추천시스템은 크게 콘텐츠기반 필터링[3]과 협업 필터링[3]으로 나뉜다. 콘텐츠기반 필터링은 사용자가 구매한 상품 중 좋아하는 상품들과 유사한 상품들을

추천하고, 협업필터링 방식은 특정 집단에서 발생하는 유사한 사용 행동을 파악하여 비슷한 성향의 아이템을 추천하는 기술이다. 협업필터링 방식의 주요 점은 모델 기반의 행렬분해 MF (Matrix Facton) 알고리즘이다. 이 방식은 추천시스템에서 가장 대중적으로 사용하고 있는 기술이고, Netflix Prize[4] 통해 알려지게 되었다. MF는 사용자-아이템 평점 매트릭스 속에 숨어있는 잠재요인을 추출해 예측할 수 있게 하는 기법이다. 그러나 과거 경험과 행동 방식에 의존하다 보니 새로운 사용자가 나타났을 때 기존 사용자-아이템의 매트릭스 정보가 존재하지 않아 추천하지 못하는 콜드 스타트(Cold-start) 문제가 발생하게 된다. 또한 MF의 모델링은 사용자와 아이템의 잠재특성(latent feature)을 선형적으로 곱하는 단순한 방법으로 매우 효율적이지만, 이는 동시에 벡터들로 이루어지는 선형적인 관계만 내적으로 수치화할 수 있다는 의미이다.

그래서 본 연구에서는 이러한 비선형적인 관계를 해결할 수 있는 심층신경망 (Deep Neural Network)을 이용하였고, 또한 사용자를 개인단위로 분석하기보다 비슷한 여행 성향을 지닌 사용자끼리 그룹으로 묶어 신규 사용자가 들어왔을지라도 간단한 프로필을 토대로 그에 해당하는 여행 성향 그룹을 분류할 수 있게 구축하였다.

3. 딥러닝을 이용한 추천시스템의 제안

본 연구에서는 심층신경망 모델 중 우수한 추천 성능을 보인 NCF (NCF: Neural Collaborative Filtering, 신경망 협업 필터링) 모델을 사용하였다.



(그림 2) 본 논문에서 사용한 Neural Collaborative Filtering의 구조.

Coving-ton[5]은 비디오 추천을 위한 NCF 신경망 기반의 추천 알고리즘을 제안하여 youtube에 적용한 결과 동영상 클릭의 60%가 알고리즘에 의해 추천된 동영상이라는 성과를 창출한 바 있다.

NCF의 구조를 살펴보면 먼저 user와 item에 대한 벡터를 one-hot encoding으로 생성해내고, 그다음 Embedding Layer 단계에서 부족한 input vector를 밀집하게 바꿔준다. 임베딩이 완료된 user와 item latent vector는 여러 층의 신경망을 거쳐 학습되면서 실제값과 예측값의 오차를 구하여 정답 값으로도 출하게 된다.

4. 사용자 여행 성향 분류

실제로 여행 분류모델을 구축하기에 앞서 비슷한 여행 성향을 지닌 사람끼리 나눌 기준을 정하기 위하여 Kim[6]의 논문을 참고하여 총 5가지의 여행 성향을 분류하였다.

- (1)패력및소비지향형 <-> 여흥추구-레저
- (2)자기개발성향 <-> 일상탈피-자연경관, 휴양/모험심-레저
- (3)문화예술체험지향형 <-> 이벤트참여-문화유적지
- (4)사회교류지향형 <-> 대인관계-레저
- (5)기타여행성향 <-> 추억추구-자연경관,휴양

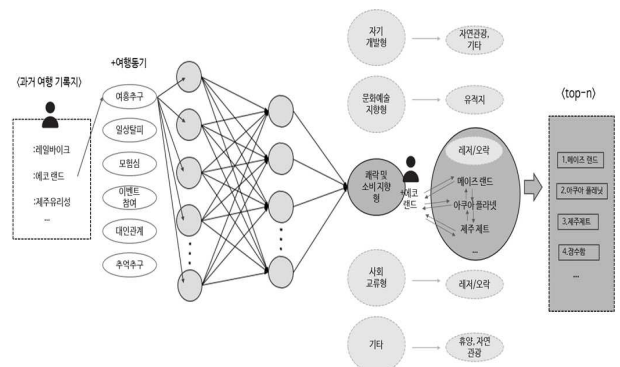
(그림 3) 여행 성향 분류

Kim[6]은 논문에서 관광객을 선택할 때 그 결정에 영향을 미치는 동기요인은 무엇인지, 이들 간의 관계를 구체적인 여행지 선택행위와 관련지어 파악하고자 하였다. (그림3) 과같이 총 6개의 여행동기 변수에 따라 여행 성향이 5가지로 분류되고, 이것은 유의적인 차이가 있다는 것을 증명했다.

그리하여 본 연구에서는 사용자의 여행동기를 포함해 5가지의 여행 성향을 분류하는 모델을 만들었다.

5. 심층신경망을 이용한 여행 추천시스템 모델

본 연구에서는 심층신경망 모형에 기반한 행렬분해의 NCF 모델을 이용하여 여행 추천시스템을 제안한다. (그림 4)는 여행 추천시스템의 전체적인 시스템 구성도이다.

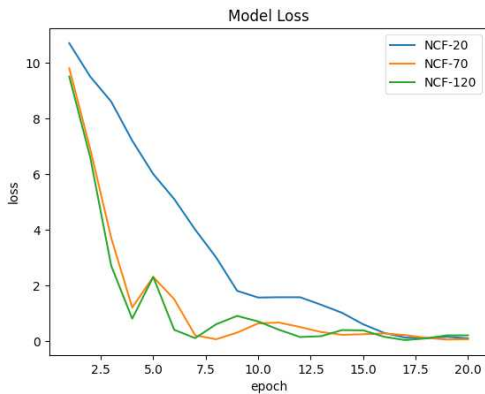


(그림 4) 제안하는 여행 추천시스템의 전체 구성

총 100명의 사용자와 120개의 제주도 여행지를 구성해 각 사용자의 여행기록과 동기를 입력데이터로, 1차 분류는 여행 성향으로, 2차 성향은 여행지 분류를 최종 정답 데이터로 설정해 학습시켰다. 1차 여행 성향 분류를 위해 사용자-여행지 데이터의 행렬을 분해하였고 처음 부족한 형태의 데이터를 임베딩 레이어를 거쳐 잠재벡터(Latent vector) 형태로 만든 후에 신경망 모델을 거쳐 가중치를 학습하여 최종 사용자의 여행 성향 다섯 가지 중 한 개의 결과값을 얻게 된다. 2차 여행지 분류는 여행 성향의 결과값의 해당하는 여행지를 토대로 사용자의 여행지와 유사도를 구한다. 유사도는 각 여행지에 대한 설명 데이터를 통해 TF-IDF[7]의 알고리즘을 이용하여 여행지끼리의 유사도 구한 후, 사용자와 가장 관련이 높은 순서대로 여행지를 추천한다.

## 6. 성능 평가

본 제안 모델의 성능을 평가하기 위해 평균제곱오차(Mean Squared Error)와 정확도(Precision)을 이용하였다



(그림 5) 제안 모델의 임베딩 차원에 따른 정답 값과 추천 값의 Training Loss 비교

(그림 5)의 각각의 그래프는 학습 횟수에 따른 손실함수의 값을 비교한 결과이고, 하이퍼 파라미터인 임베딩의 차원의 크기에 따른 각각의 손실함수 결과값이다. 임베딩 차원은 사용자와 여행지를 나타내는 벡터의 표현력에 직접적인 영향을 미친다. 그 결과는 20차원일 때 최종 손실 값은 0.092, 70차원일 때 최종 손실 값은 0.05, 120차원일 때 최종 손실 값은 0.24로 나왔다. 임베딩의 적정한 차원값을 넣어야 사용자-여행지 관계를 잘 나타내는 것을 볼 수 있었다. 가장 낮은 손실의 임베딩 차원의 모델을 가지고, 각 타입별 정확도를 계산한 결과, 자기개발형은

0.83, 문화예술형은 0.54, 쾌락소비지향형은 0.84. 사회교류형은 0.66. 기타는 0.58의 결과 값이 나왔다.

## 7. 결론

본 논문에서는 여행 추천시스템을 활용한 기술과 기존 플랫폼에 대한 한계에 대해 언급하며, 기계 스스로 사용자의 취향에 맞게 여행지를 추천하지 못하는 정적인 문제를 해결하고자 하였다.

개인이 취향에 맞는 여행지를 받는 동적 추천방식을 받을 수 있도록 심층신경망 모델을 선택하였고, 그중 NCF 방식 선택해 모델을 구축해 나갔다. 사용자들의 과거 여행기록과 동기를 기반으로 총 5가지의 여행 성향 판단하였고, 그 결과에 따른 top-k의 여행지 추천을 제공하도록 설계하였다. 그 결과, 전체적으로 낮은 오류율과 높은 정확도의 결과 값을 구할 수 있었다.

현재 심층신경망을 이용한 여행 추천시스템에 관한 연구 발표들이 아직은 수가 적고 효과적이지도 않다는 점에서, 본 제안의 여행 추천시스템은 여행지 추천의 새로운 방향 제시 면에서 의의가 있다.

## 참고문헌

- [1] <https://www.tripadvisor.co.kr/>
- [2] <https://triple.guide/>
- [3] 손지은, et al. "추천시스템 기법 연구동향 분석." 대한산업공학회지 41.2 (2015): 185-208.
- [4] Koren, Yehuda, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." Computer 42.8 (2009): 30-37.
- [5] Covington, Paul, Jay Adams, and Emre Sargin. "Deep neural networks for youtube recommendations." Proceedings of the 10th ACM conference on recommender systems. 2016.
- [6] Kim, K. H., U. Y. Chung, and B. H. Kwon. "An Effect of Tour Motivation Factors on Tour Site Choice and Satisfaction; Focus on college students residing in Seoul." Journal of Hotel and Tourism 6.1 (2004): 50-73.
- [7] 오유리, et al. "워드임베딩 기반 TF-IDF 를 이용한 특허 문서의 유사 청구항 도출 기법 비교 분석." 한국정보과학회 학술발표논문집 (2017): 1001-1003.