

온라인 전시 행사에서 개인 맞춤형 추천 시스템 적용 사례 연구 : 산학협력 EXPO를 중심으로

변 문 경¹ · 이 정 훈^{2*}

¹성균관대학교 인공지능융합학과 박사과정

²성균관대학교 소프트웨어학과 석사과정

Case study on application of personalized recommendation system in online exhibition event : Focusing on industry-academic cooperation EXPO

Moon-Kyoung Byun¹ · Jung-Hoon Lee^{2*}

¹Ph.D. Student, Department of Applied Artificial Intelligence Sungkyunkwan University

²*Master Student, College of Computing Sungkyunkwan University

[요 약]

코로나 19의 확산으로 인해 많은 전시·행사가 온라인 비대면 형태로 진행되고 있다. 온라인 전시·행사는 오프라인 행사와 비교하여 공간과 시간 제약 없이 언제 어디서든 관람할 수 있는 장점이 있다. 반면 온라인 전시·행사는 관람객이 몰입하는 시간이 상대적으로 적기 때문에 콘텐츠 전달 효과가 떨어진다. 이러한 단점에도 불구하고 코로나 19의 확산세가 장기화하며 온라인 전시·행사는 확대되고 있다. 게다가 MICE 전문가들은 코로나 19가 종식된 이후에도 오프라인 행사와 온라인 행사를 병행하게 될 것으로 전망한다. 이에 본 연구에서는 온라인 전시·행사에서 프로그램에 대한 몰입감을 높일 수 있는 전략으로 개인 맞춤형 인공지능 추천 알고리즘을 적용하여 관람객이 머무는 제한된 시간 동안 콘텐츠 전달 효과를 극대화할 방안을 제안하고자 한다.

[Abstract]

Due to the spread of Corona 19, many exhibitions and events are being conducted online, non-face-to-face. Compared to offline events, online exhibitions and events have the advantage of being able to watch anytime, anywhere without space and time constraints. On the other hand, in online exhibitions and events, content delivery is inferior because the amount of time for the audience to immerse is relatively small. Despite these shortcomings, the spread of Corona 19 is prolonged and online exhibitions and events are expanding. Also, MICE experts predict that even after the end of Corona 19, offline and online events will be held simultaneously. Therefore, in this study, as a strategy to increase the sense of immersion in the program in online exhibitions and events, we intend to propose a method that can maximize the effect of delivering content during a limited time of visitor's stay a personalized artificial intelligence recommendation algorithm.

색인어 : Matrix Factorization, TF-IDF, Sentence-BERT, 온라인 전시회, 추천 시스템

Key word : Matrix Factorization, TF-IDF, Sentence-BERT, Online exhibition, Recommendation system

<http://dx.doi.org/10.9728/dcs.2021.22.4.655>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 03 February 2021; Revised 29 March 2021

Accepted 29 March 2021

*Corresponding Author; Jung-Hoon Lee

Tel: 

E-mail: curiymoonlight@gmail.com

I. 서론

코로나 19의 확산으로 인해 오프라인으로 진행되던 전시회들이 온라인으로 진행되고 있다. 온라인 전시회는 거리와 공간의 제약이 없어 어디서든 관람할 수 있으며 시간적 제약도 없어 원하는 시간대에 관람할 수 있다[1].

그러나 온라인 전시회는 현장감이 떨어지며 관람객이 전시물을 세세하게 들여다보기 힘들다. 또한, 오프라인 전시회와 비교해 온라인 전시회는 관람객의 몰입감이 떨어지기 때문에 상대적으로 전시회에 체류하는 시간이 적다. 이러한 단점에도 불구하고 코로나 19의 확산세가 장기화하며 온라인 전시·행사는 확대되고 있다. 게다가 MICE 전문가들은 코로나 19가 종식된 이후에도 오프라인 행사와 온라인 행사를 병행하게 될 것으로 전망한다[2]. 온라인으로 한 번 구축된 콘텐츠는 서버와 호스팅이 유지되는 경우 행사 기간 외에도 관람이 가능한 지속가능성을 가지고 있기 때문이다[3].

코로나 19와 언택트 문화의 정착으로 온라인 전시는 지속해서 확대될 전망이다. 이에 온라인 전시에서 관람객의 몰입감을 높일 방안은 관련 연구는 필요한 시점이다. 기존의 온라인 전시에서 몰입감을 높이는 연구는 주로 가상전시관 구축과 관련된 것이 대부분이었다 [4] [5]. 하지만 가상전시관 구축은 비용이 많이 들고, PC나 모바일 체험 환경에 따라서 관람객의 몰입감을 높일 수도 있고, 반대로 제한할 수도 있는 한계가 있었다. 본 연구에서는 추천 알고리즘을 사용해 사용자에게 최적화된 전시물을 추천해 관람객의 전시회 체류 시간과 만족감을 늘리고자 하였다.

개인 맞춤형 추천이란 특정 시점에 특정 고객이 필요로 하는 상품을 제공하는 것을 의미한다. 방대한 콘텐츠 중 추천 시스템 없이 사용자가 원하는 콘텐츠를 선택하는 것은 힘들다. 사람들은 영화가 보고 싶을 때 현재까지 개봉된 모든 영화 정보를 검색할 필요가 없다. 추천 알고리즘에서 제공해주는 추천 목록 중 보고 싶은 영화를 선택하면 된다. 추천 알고리즘은 영화, 음악, 책, 물건 등 다양한 분야에 적용될 수 있다.

본 논문에서는 2020 산학협력 EXPO 행사에 개인화 맞춤 추천과 유사 전시물 추천 시스템을 적용했다. 개인화 맞춤 추천은 관람객의 전시 로그를 기반으로 구현되었으며 개인마다 추천 목록이 다르게 제공된다. 유사 전시물 추천은 사용자가 특정 전시물을 관람하면 그와 내용이 유사한 전시물을 추천해준다. 전시물의 텍스트를 벡터로 변환한 후, 코사인 유사도를 적용해 유사 게시물을 판단한다.

본 논문의 구성은 다음과 같다. 2장에서는 2020 산학협력 EXPO 행사를 소개한다. 3장에서는 기존에 연구되었던 추천 알고리즘, Matrix Factorization, Word Representation 방법론에 대해 정리한다. 4장에서는 추천 시스템의 구현 흐름과 결과물을 설명한다. 5장에서는 본 논문의 한계점과 후속 연구로 마무리한다.

II. 2020 산학협력 EXPO

산학협력 EXPO의 목적은 학계 및 지역 산업계가 함께하는 EXPO 개최를 통해 산학협력 문화의 저변 확산 및 활성화하는 것이다.

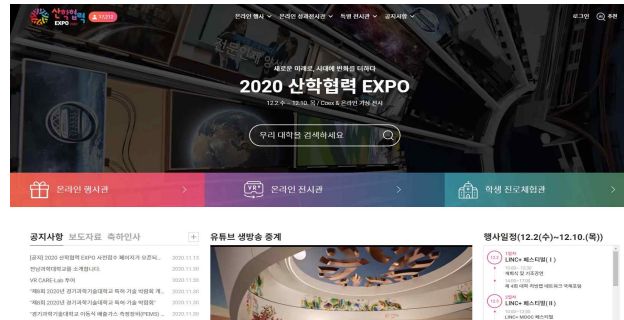


그림 1. 산학협력 EXPO 시작페이지
Fig. 1. EXPO Entrance Page

주최기관은 교육부와 한국 연구재단이며 행사 일정은 12월 2일(수)부터 10(목)일까지 9일간 진행됐다. 온라인 중심으로 행사를 진행했으며 그림1과 같이 산학협력 EXPO 홈페이지와 유튜브를 사용해 진행됐다. 총 300개 이상 기관(대학기업·유관 기관)이 참여했으며 한국산업기술진흥원, 대학교, 청년기업가재단, 일반대/전문대 LINC+사업협의회 등이 참여해 성과를 전시했다.

산학협력 EXPO 행사는 수백 개의 전시관이 있고 전시관 안에 수십 개의 전시물이 존재하는 구조이다. 전시관은 학교기업지원 사업 성과관(서울시립대학교), 일반대 LINC+ 4차 산업혁명 혁신 선도대학 성과관(서울시립대학교), 전문대 LINC+ 사회 맞춤형학과 중점형 성과 전시관(동주대학교), 조기취업형 계약학과 선도대학 육성사업 성과관(경일대학교), 대학 산학협력단지 조성지원 사업 성과관(연세대학교) 등 총 311개의 전시관이 존재한다.

각 대학의 전시관 내부에는 다양한 전시물 포스터가 존재한다. 각 전시관은 그림2와 같이 3D로 구현되어 있으며 전시물을 클릭하면 확대해서 볼 수 있다. 개인화 맞춤 추천은 사용자 전시 시청 로그에 Matrix Factorization을 적용해 311개의 전시관 중 9개의 전시관을 사용자에게 추천한다. 사용자 전시관 방문 로그 기반이기 때문에 특정 시간마다 추천 목록을 업데이트해 사용자에게 새로운 전시관을 추천한다.

산학협력 EXPO의 전시물은 각 기관에서 제출한 포스터와 영상 파일이다. 그림3은 포스터의 예시이며 이미지, 텍스트, 동영상 등이 혼합되어 있다. 전체 포스터의 개수는 1,585개이며 전시물 주제는 바이오, 의료서비스, 인공지능, 도시 인프라, 에너지 플랫폼 등 다양한 주제를 다루고 있다.

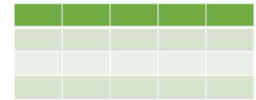
유사 전시물 추천은 전시물 포스터의 설명을 기준으로 의미상 유사한 전시물을 추천한다. 일반대 LINC+ 고도화형(계명대학교)의 포스터 설명은 다음과 같다. “코로나 19 극복 희망 나눔 캠페인 시즌 1,2. 기술개발성과(이미지 및 도표 활용). 대한민국의 코로나-19 확산으로 국민의 아픔과 고통 분담을 위한 대학의 지역사회 희망 나눔의 인식 확산 운동 ‘코로나 희망, 나눔 캠페인’을 통해 지역사회 재난 극복. 지역 문제 해결을 위해 대학의 사회공헌 촉매제 역할 수행을 통한 지역사회의 빠른 피해 복구 및 사회경제 회복 촉진”. 이와 같은 전시물 설명을 다른 전시물 설명과 비교해 1,585개의 전시물 중 가장 유사한 전시물 6개를 사용자에게 추천한다.



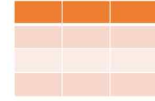
그림 2. 산학협력 EXPO 전시관
Fig. 2. Industry-Academic Cooperation EXPO exhibition

	Item A	Item B	Item C	Item D	Item E
User A	1	2	0	3	2
User B	1	1	1	0	2
User C	1	3	2	1	1

User - Exhibition Matrix



User Latent Matrix



Exhibition Latent Matrix

그림 4. 행렬 분해 벡터
Fig. 4. Matrix Factorization Vector

VR 영상상도론트랙

- 우수사례명: 3D스캐너를 활용한 실감콘텐츠 제작 교육 실험
- 내용:
 - 현장미러형 실습환경 활용 및 3D스캐너를 이용한 실감콘텐츠 Case study
 - 3D스캐너와 3D기술 반영에 따른 실감콘텐츠 변화 제시
 - 기업의 요구 기술 해결 및 산업체 공동 교육
- 우수성과
 - 다양한 3D스캐닝 촬영 기법을 통한 4차 산업혁명 시대 전문가 양성
 - 3D스캐닝의 전문 렌더링 기술을 접목한 우수 결과 도출
 - 협약기업과의 연계 운영 및 학생들의 자기주도적 실무역량강화 향상

IOT 융합서비스개발트랙

- 우수사례명: IOT 융합기술 단기 집중교육 운영
- 내용: 4차산업혁명시대 실무 능력을 갖춘 IOT 전문 인재 양성을 위한 PBL 기반의 SW 개발 과정
- 우수성과
 - 사전 설명회, 설문조사 및 진단평가를 통해 학생 수준을 고려하여 맞춤형 교육 운영
 - 교육과정 중 데이터베이스관리사 자격증 취득 지원: 자격증 취득률 80%(16명/20명)
 - 1인 1서버 사용, git를 통한 소스 형상관리 등 실무에서 SW를 개발 하는 환경과 비슷한 프로세스를 통한 미니프로젝트 진행
 - 미니프로젝트 과제를 캡스톤디자인으로 연계하고 확대·발전시켜 교내 PBL 교육성과 경진대회에서 "우수상" 수상

전기자동차부품개발트랙

- 프로그램명: 전공역량강화프로그램
- 내용: Auto CAD 강좌, CATIA강좌를 통한 현장 맞춤형 부품설계 도면 작도를 통한 실무능력 향상
- 우수성과
 - 도면 해석 및 적도 능력을 보유한 전문인력 양성을 위한 프로그램 운영
 - 맞춤형 지도를 통한 기술력 향상과 실습 위주 교육으로 전기자동차 부품 개발 능력을 갖춘 창의적 현장 실무능력 및 취업역량 강화
 - 자격증 취득률: 국제자격 제도 설계 기술관리사 2급 100%(10명/10명), CATIA Part Design 80%(8/10)

그림 3. 전시물 예시
Fig. 3. Exhibit poster example

III. 관련 연구

3-1 추천 알고리즘

추천 알고리즘은 크게 Collaborative Filtering과 Contents Based Filtering으로 분류된다. Collaborative Filtering은 사용자의 과거 패턴을 분석해 유사한 성향의 사용자에게 아이템을 추천하는 방법이다. Collaborative Filtering은 사용자의 로그 기록과 과거 패턴 같은 데이터를 많이 가지고 있을수록 성능이 높아진다[6]. 하지만 신규 사용자의 경우에는 관찰된 데이터가 거의 없어서, 추천의 정확도가 급격히 떨어지는 Cold Start 문제가 발생한다[7][8].

Contents Based Filtering은 아이템 간의 유사도를 측정하여 사용자가 아이템을 조회했을 때 유사한 아이템을 추천해주는 방법이다[9][10]. Contents Based Filtering은 Collaborative Filtering의 문제점인 Cold Start 문제가 없으므로, 아이템에 대한 평가가 없는 신규 사용자에게도 적용할 수 있다. Contents Based Filtering은 사용자 간의 유사도를 기준으로 하는 User-Based와 아이템 간의 유사도를 기준으로 하는 Item-Based로 구분된다. User-Based는 나와 프로필이 비슷한 다른 사용자가 선호하는 전시물을 추천해주는 방식이다. Item-Based는 전시물의 이미지와 텍스트 같은 정보를 활용해 유사한 전시물을 추천하는 방식이다.

본 연구에서 개인화 맞춤 추천 시스템은 Collaborative Filtering을 사용해 구현하고 유사 전시물 추천 시스템은 Item Based Contents Filtering을 적용해 구현한다.

3-2 Matrix Factorization

Matrix Factorization은 3-1에서 설명한 Collaborative Filtering 기법의 하나이다[6][7]. Matrix Factorization은 사용자-전시물 행렬을 두 개의 낮은 차원의 직사각형 행렬의 곱으로 분해한다. 전시물 Latent Vector는 (차원 x 전시물 수)이며 사용자 Latent Vector는 (차원 x 사용자 수)이다. 차원은 개발자가 조정 가능한 파라미터이며 보통 50에서 200 사이로 설정한다. 각 Latent Vector의 초깃값은 랜덤하게 설정되며 연산을 통해 최적화된다.

최적화는 전시물 Latent Vector와 사용자 Latent Vector를 행렬 곱한다. 사용자 Latent Vector를 전치하면 (사용자 x 차원)이 되고 전시물 Latent Vector와 행렬 곱을 하면 (사용자 x 전시물 수) 형태가 된다. 예측 행렬은 사용자 전시물 행렬과 형태가 같 으며 Loss Function을 사용해 연산 후 사용자-전시물 행렬값과 같아지도록 최적화한다. Loss Function의 수식은 아래와 같다.

$$\min_{x^*, y^*} \sum_{u,i} (r_{ui} - x_u^T y_i)^2 + \lambda (\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2) \quad (1)$$

Matrix Factorization의 장점은 영화 평점, 구매 평점과 같은 선호도를 직접 나타내는 로그뿐만 아니라 구매 내역, 검색 기록, 검색 패턴과 같은 간접 로그를 사용해 추천 시스템을 구현할 수 있다는 것이다.

3-3 Word Representation

Word Representation이란 텍스트를 구성하는 단어 하나를 벡터로 표현하는 방법이다. Word Representation의 전통적인 방법은 빈도수 기반 단어 표현(Count Based Word Representation)이며, 대표적으로 Bag of Words와 TF-IDF가 있다. 빈도수 기반 단어 표현 방법은 문장에서 단어의 빈도수를 표현하는 방식으로 문장을 벡터로 변환한다. 빈도수 기반 방법의 단점은 문장을 표현하는데 불필요하게 큰 벡터와 필요하다는 것과 단어 사이의 관련성을 파악할 수 없다는 것이다.

이러한 문제를 해결하기 위해 단어를 밀집 벡터(Dense Vector)로 표현하는 단어 임베딩(Word Embedding) 방법이 개발됐다. 단어 임베딩은 단어를 밀집 벡터로 표현하기 때문에 문장을 표현하는데 큰 벡터가 필요하지 않고 단어 사이의 연관성을 알 수 있는 장점이 있다. 본 연구에서는 다양한 단어 임베딩 방법 중 Sentence BERT를 사용해 전시물의 텍스트를 밀집 벡터로 표현했다[13][14].

1) TF-IDF

TF-IDF(Term Frequency - Inverse Document Frequency)는 특정한 단어가 문서에서 얼마나 자주 등장하는지를 나타내는 값이다. TF-IDF는 단어의 빈도와 역문서 빈도를 계산해 단어에 가중치를 부여한다[15].

TF(단어 빈도, term frequency)는 특정 문서 d에서 특정 단어 t의 등장 회수를 의미한다. 예를 들어 "과자"라는 단어가 문서3에서 3번 등장하면 TF는 3이 된다. TF는 특정한 단어가 문서 내에 얼마나 자주 등장하는지를 나타내는 값으로, TF 값이 클수록 문서에서 중요한 단어라고 생각할 수 있다.

DF(문서 빈도, Document Frequency)는 특정 단어 t가 등장한 문서의 수이다. 예를 들어 "과자"라는 단어가 문서2와 문서3에서 언급되었다면 DF는 2이다.

IDF(역문서 빈도, Inverse Document Frequency)는 DF(t)에 반비례하는 수를 의미한다. 단어 자체가 문서 집합 내에서 자주 사용되는 경우, 이것은 그 단어가 흔하게 등장한다는 것을 의미한다. 예를 들면 여러 문서에서 “은”, “는”, “이”, “가”와 같은 조사가 많이 등장하지만, 그 문서를 대표하는 중요한 단어는 아니다. 반면, “손흥민”이라는 단어는 일반적인 문서 사이에서 잘 나오지 않고 그 문서의 핵심인 경우가 많다. TF-IDF는 TF와 IDF를 곱하며 수식은 아래와 같다.

$$W_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \quad (2)$$

표 1. TF-IDF 예시

Table 1. TF-IDF example

	Word 1	Word2	...	Word n
Doc 1	0.574	0.212	...	0.213
Doc 2	0	2.311	...	0
...
Doc n-1	0	6.52	...	0.733
Doc n	0.231	0	...	0

최종적으로 TF-IDF는 단어 하나하나를 숫자로 표현할 수 있다. 예를 들어 “과자”라는 단어는 2.3이라는 수치로 표현될 수 있고 “사과”라는 단어는 0.598로 표현될 수 있다. 이러한 숫자 값을 문장별로 뭉치면 표1과 같이 문장 집합을 2차원 행렬로 표현할 수 있다.

2) Sentence BERT

Sentence BERT는 사전 훈련된 BERT 구조를 Siamese 네트워크 구조로 수정하여 Semantic Textual Similarity(STS) 및 Natural Language Inference(NLI) 데이터로 학습한 문장 임베딩 모델이다[16]. 이런 네트워크 구조는 문장의 의미를 임베딩에 효과적으로 표현할 수 있게 해주며, 코사인 유사도를 통해 쉽게 유사도를 계산할 수 있다.

Sentence BERT는 딥러닝 구조이며 768개의 은닉층과 110,000,000개의 파라미터로 구성되어 있다. Sentence BERT의 Batch size는 16이며 Optimizer는 Adam을 사용한다. Sentence BERT는 104개 언어를 지원하는 사전 학습된 BERT (BERT-Base, Multilingual Cased) 모델을 사용했다.

BERT는 Google에서 만든 단어 임베딩 기법으로 NLP 11개 Task에 SOTA(State of the Arts)를 기록했으며, SQuAD v1.1에서는 인간보다 더 높은 정확도를 보여 주목을 받고 있다[17]. BERT 개발자들의 접근방식은 “(1) 범용 솔루션을 (2) 스케일러블한 형태로 구현해서 (3) 많은 GPU와 TPU로 학습해서 성능을 높인다”이다. BERT는 Attention 기반의 Transformer 모델의 Encoder 부분만 사용하여 구성되었다.

IV. 제안 방법

4-1 개인화 맞춤 추천

개인화 맞춤 추천은 관람객의 로그에 Matrix Factorization을 적용해 사용자가 선호를 반영해 전시물을 추천한다. 본 시스템의 전체 흐름은 다음과 같다.

첫째, 사용자 로그를 수집한 후, 각 사용자가 관람한 전시물을 정리한다. 사용자 로그는 표2와 같은 형태로 저장되어 있으며 {2:[1-269, 1-270], 10:[1-270], 36:[1-270, 1-269, 1-272]}과 같은 python dictionary 자료형으로 변환한다. dictionary의 key는 회원코드가 되고 value는 전시관 코드 리스트가 된다.

표 2. 사용자 로그

Table 2. User log

User id	Exhibition id
2	1-269
2	1-270
2	7-72
...	
10	1-270
36	1-270
36	1-269

첫째, 사용자 로그를 수집한 후, 각 사용자가 관람한 전시물을 정리한다. 사용자 로그는 표2와 같은 형태로 저장되어 있으며 {2:[1-269, 1-270], 10:[1-270], 36:[1-270, 1-269, 1-272]} 과 같은 python dictionary 자료형으로 변환한다. dictionary의 key는 회원코드가 되고 value는 전시관 코드 리스트가 된다.

둘째, 전시관 코드 리스트를 사용자-전시물 벡터로 변환한다. 사용자-전시물 벡터는 2차원 행렬이며 행은 회원코드를 의미하고 열은 전시관 코드가 된다. 그 후, Matrix Factorization을 적용하기 위해 사용자 Latent Vector와 전시물 Latent Vector를 생성한다. 각 Latent Vector의 차원은 200으로 설정했다. 전체 사용자가 5535명이고 차원을 200으로 설정했기 때문에 사용자 Latent Vector는 (5535 x 200) 로 설정된다. 마찬가지로 전시관의 개수는 311개이고 차원은 200으로 설정했기 때문에 전시물 Latent Vector는 (311 x 200) 로 설정한다.

셋째, 사용자와 전시물 Latent Vector를 행렬 곱해 예측 행렬을 생성한 후, 손실 함수를 사용해 최적화한다. 전시물 Latent Vector를 전치 시켜 (200 x 311)로 만든 후, 사용자 Latent Vector와 행렬 곱셈을 한다. 사용자와 전치 시킨 전시물 Latent Factor를 행렬 곱셈하면 (5535 x 311) 형태가 되며 이는 사용자-전시물 벡터 차원과 같다. 그 후 앞에서 설명한 Loss Function을 15번 반복해 예측 행렬 안 요소가 사용자-전시물 Vector 값이 되도록 최적화한다.

표 3. 사용자-전시물 벡터

Table 3. User-Exhibit Vector

	1-269	1-270	...	7-72	7-74
2	1	1	...	0	0
10	1	0	...	1	1
...
32	0	0	...	0	1
36	2	3	...	1	0

표 4. 예측 행렬

Table 4. User-Exhibit Vector

	1-269	1-270	...	7-72	7-74
2	0.2783	3.211	...	0.72	0.213
10	1.231	0.83	...	6.11	0.811
...
32	0.13	0.31	...	0.31	1.21
36	3.212	8.32	...	1.4356	0.43



그림 5. 개인화 맞춤 추천 예시
Fig. 5. Exhibit poster example

마지막으로 예측 행렬을 반복해서 최적화하면 표4와 같이 0이 없는 밀집 행렬이 생성된다. 행은 사용자를 의미하며 열은 각 전시관을 의미한다. 행렬의 요소는 사용자가 해당 전시물을 선호하는지를 반영한 값이며 높을수록 사용자가 선호할 확률이 높다. 해당 행렬에서 사용자가 보지 않은 전시관 중 점수가 높은 9개를 추출해 관람객에게 추천한다.

그림5는 실제 2020 산학협력 EXPO에 반영된 결과물이며 사용자에게 최적화된 9개의 전시관을 추천해준다.

4-2 유사 전시물 추천

유사 전시물 추천은 전시물의 텍스트를 사용해 의미상으로 가까운 전시물을 추천해주는 시스템이다. TF-IDF와 Sentence BERT를 사용해 텍스트를 벡터로 변환시킨 후, 각각의 벡터들에 대해 코사인 유사도를 적용해 유사한 게시물을 찾는다. TF-IDF를 적용해 찾아낸 유사한 전시물 3개와 Sentence BERT를 적용해 찾아낸 전시물 3개를 관련 전시물을 사용자에게 추천한다. 유사 전시물 추천은 개인화 맞춤 추천과 다르게 모든 사용자에게 똑같이 추천된다.

1) TF-IDF

TF-IDF를 이용한 유사 전시물 추천 시스템의 구현 방법은 그림6과 같다. 첫째로, 전시물 텍스트에 형태소 분석을 적용해 문장을 형태소라는 토큰으로 분리한다. 형태소는 뜻을 가진 가장 작은 말의 단위를 의미한다. 예를 들어, "아버지가 방에 들어가신다"라는 문장에 형태소 분석을 적용하면 "아버지/NNG", "가/JKS", "방/NNG", "에/JKB", "들어가/VV", "신다/EP+EC"로 분리가 된다. 문장을 형태소로 분리한 후, 조사와 동사 같은 불필요한 품사를 제외한다. 본 시스템에서는 Konlpy 형태소 분석기를 사용하며 여러 모듈 중 Komoran 모듈을 사용한다.

둘째로 형태소로 분리된 단어 집합에 TF-IDF를 적용해 2차원 행렬로 구성된 벡터를 만든다. 행렬의 행은 전시물을 의미하며 열은 단어들을 의미한다. 2차원 행렬에 코사인 유사도를 적용하면 각 전시물 중 가장 유사한 전시물을 알 수 있다.

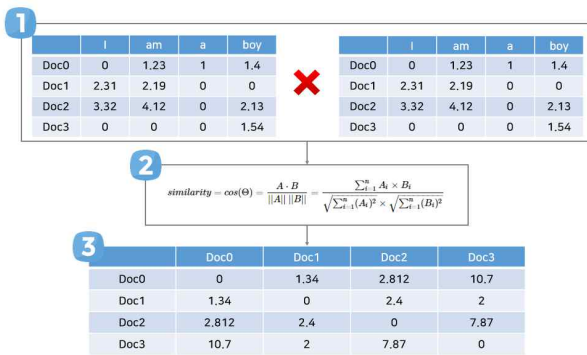


그림 6. 전시물 예시
Fig. 6. Exhibit poster example

셋째로, 2차원 행렬에 코사인 유사도를 적용해 문서 집합의 상관관계를 담은 행렬을 구성한다. 각 전시물 당 유사도 점수가 가장 높은 전시물 3개를 추출한다.

2) Sentence BERT

Sentence BERT를 이용한 유사 전시물 추천 시스템의 구현 방법은 다음과 같다. 첫째로 전시물 텍스트를 Sentence BERT에 입력해 1차원 행렬로 구성된 벡터를 생성한다. 전시물 설명을 Sentence BERT에 입력하면 (1 x 764) 벡터 형태로 변환된다. 해당 벡터는 전시물의 정보가 압축되어 숫자로 표현된 형태이다. 각각의 전시물 벡터를 합치면 (N x 764) 차원의 2차원 벡터가 된다.

둘째로, 해당 벡터에 코사인 유사도를 적용하면 TF-IDF와 마찬가지로 전시물 간의 유사도 점수를 가진 2차원 벡터가 생성된다. 전시물 중 가장 점수가 높은 3개를 유사한 전시물을 사용자에게 추천한다.

V. 결 론

코로나 19와 언택트 문화의 정착으로 온라인 전시는 지속적으로 확대될 전망이다. 온라인 전시에서 관람객의 몰입감을 높일 방안에 대한 연구로 추천 시스템을 사용해 실제 2020 산학협력 EXPO에 참여한 5,535명의 관람객에게 추천 목록을 제공했다. 그 결과 본 논문에서는 온라인 전시회에 최적화된 개인화 맞춤 추천과 유사 전시물 추천 방법을 제안해 보았다.

개인화 맞춤 추천 시스템은 사용자의 관람 기록에 Matrix Factorization을 적용하여 관람객에게 최적화된 9개의 전시관을 추천한다. 유사 전시물 추천은 전시물의 설명을 기반으로 유사한 내용의 전시물을 사용자에게 추천한다. TF-IDF와 Sentence BERT를 사용해 총 6개의 전시물을 사용자에게 추천한다.

본 논문의 개인화 맞춤추천은 사용자 로그 기록만 사용하기 때문에 콘텐츠의 특성이 반영되지 않는다. 또한, 유사 전시물 추천은 전시물의 영상과 이미지 정보를 사용하지 않는다. 추후 연구로 개인화 맞춤추천에 콘텐츠 정보와 사용자 로그를 섞어 사

용자에게 추천 목록을 제공하고자 한다. 또한, 유사 전시물 추천에 이미지와 텍스트를 벡터로 변환해 전시물을 추천하고자 한다.

참고문헌

[1] S. H. Jang, "The influence of usage intentions of social network services in MICE industry: Focused on the gender differences," The Journal of the Korea Contents Association, 20(3), 506-514, 2020.

[2] C. W. Park and T. Lim, "Changes in Exhibition and Convention in the era of Untact and Online Exhibition Case Study," International Journal of Trade Fairs and Exhibition Studies, 15(5), 57-82, 2020.

[3] H. J. Cho, "A Study of the Development Plan for Local Festival as a Contents Platform", Storynimage, 19, 299-323, 2020.

[4] S. E. Park and J. S. Kim, "Converting 2-D Artworks into 2.5-D Content for More Immersive HMD VR Exhibition Experience," Journal of Digital Contents Society, Vol. 21, No. 4, pp. 653-661, 2020.

[5] H. M. Lee and M. S. Kim, "Case Study of Digital Contents for Online Education in College - Focused on Virtual Museum," Korean J General Edu, 14(4):81-96, 14(4), 81-96, 2020.

[6] Koren, Y and Bell, R. "Advances in collaborative filtering," In Recommender systems handbook, pp. 77-118. Springer, Boston, MA, 2015.

[7] Herlocker, J. L., Konstan, J. A. and Riedl, J. "Explaining collaborative filtering recommendations." In Proceedings of the 2000 ACM conference on Computer supported cooperative work, pp. 241-250, December 2000.

[8] Breese, J. S., Heckerman, D. and Kadie, C. "Empirical analysis of predictive algorithms for collaborative filtering," arXiv preprint arXiv:1301.7363, 2013.

[9] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. "Item-based collaborative filtering recommendation algorithms." In Proceedings of the 10th international conference on World Wide Web, pp. 285-295, April 2001.

[10] Van Meteren, R. and Van Someren, M. "Using content-based filtering for recommendation." In Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop, Vol. 30, pp. 47-56, May 2000.

[11] Ma, H., Yang, H., Lyu, M. R. and King, I, "Sorec: social recommendation using probabilistic matrix factorization," In Proceedings of the 17th ACM conference on Information and knowledge management, pp. 931-940, October 2008.

- [12] He, X., Zhang, H., Kan, M. Y. and Chua, T. S., "Fast matrix factorization for online recommendation with implicit feedback," In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, pp. 549-558, July 2016.
- [13] Mikolov, T., Chen, K., Corrado, G. and Dean, J., "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.
- [14] Pennington, J., Socher, R., & Manning, C. D., "Glove: Global vectors for word representation," In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543). October 2014.
- [15] Tata, S. and Patel, J. M., "Estimating the selectivity of tf-idf based cosine similarity predicates," ACM Sigmod Record, 36(2), 7-12, 2007.
- [16] Reimers, N. and Gurevych, I. "Sentence-bert: Sentence embeddings using siamese bert-networks," arXiv preprint arXiv:1908.10084. 2019.
- [17] Devlin, J., Chang, M. W., Lee, K., and Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805.a, 2018.



변문경(Byun-Moon Kyoung)

2002년 : 성균관대학교 화학, 국어국문학 전공 (이학사, 문학사)
2015년 : 성균관대학교 대학원 교육학석사
2018년 : 성균관대학교 대학원 교육공학 박사
2020년~현 재 : 성균관대학교 대학원 인공지능 융합학과 박사과정

※ 관심분야 : 관심분야: STEM 교육, 인공지능 융합교육, 인공지능 콘텐츠 생성 API 개발, 테크놀로지 통합교육, 스토리텔링
E-mail: curiomoonlight@gmail.com



이정훈(Jung-Hoon Lee)

2019년 : 경기대학교 응용통계학과, 컴퓨터학과 (공학사)
2019년~현 재 : 성균관대학교 소프트웨어학과 석사과정

※ 관심분야 : 자연어처리(Natural Language Processing), 감정 분석(Sentiment Analysis), 대화형 시스템(Dialogue system)
E-mail: vhrefdl@gmail.com