

# 공연 콘텐츠 추천을 위한 소셜 행위 기반 협업필터링 방법에 대한 연구

송재오<sup>0</sup>, 곽한경<sup>\*</sup>, 조정현<sup>\*</sup>, 이상문<sup>\*\*</sup>

<sup>0</sup>(주)제오시스 기업부설연구소

<sup>\*\*</sup>한국교통대학교 컴퓨터정보공학과

e-mail: jos@zeosis.com<sup>0</sup>, hkkwak@zeosis.com<sup>\*</sup>, jhcho@zeosis.com<sup>\*</sup>, smlee@ut.ac.kr<sup>\*\*</sup>

## A Study on Collaborative Filtering Method based on Social Behavior for Performance Contents Recommendation

Je-O Song<sup>0</sup>, Han-Kyeong Kwak<sup>\*</sup>, Jung-Hyun Cho<sup>\*</sup>, Sang-Moon Lee<sup>\*\*</sup>

<sup>0</sup>R&D Institute, ZEOSIS Co.,Ltd.

<sup>\*\*</sup>Dept. of Computer Sci. & Info. Eng., Korea Nat'l Univ. of Transportation

### ● 요약 ●

스마트폰을 중심으로 한 모바일 기기의 보급과 온라인 소셜 네트워크 서비스의 이용자들이 증가하면서 사용자들은 많은 콘텐츠를 소비하고 공유한다. 이는 콘텐츠 사용자들의 개별적 기호에 맞지 않거나 만족도가 떨어지는 콘텐츠를 소비하게 한다. 이와 같은 문제를 해결하기 위해 소셜 네트워크 사용자에게 적합한 콘텐츠를 추천하기 위한 기법에 대한 연구가 활발하게 진행되고 있다. 본 논문에서는 온라인 상에 존재하는 다양한 정보 중에서 공연과 관련한 콘텐츠들을 중심으로 사용자 성향별로 추천을 해 줄 수 있는 협업필터링 방법에 대하여 제안한다.

**키워드:** 공연(Performance), 콘텐츠(Contents), 소셜(Social), 필터링(Filtering), 추천(Recommendation)

### I. Introduction

4차 산업혁명 시대로 진입하면서, 일상 속에서도 인공지능 스피커 등을 통해 체험할 수 있는 추천 서비스가 구현되고 있다. 하지만, 대부분의 온라인 추천 서비스는 소셜 미디어를 중심으로 통계적인 정보에 기반하여 불특정 다수에게 제공되고 있다. 본 논문에서는 소셜 빅데이터를 이용하여 공연 콘텐츠를 중심으로 사용자 성향을 분석하고 유사한 사용자들의 정보를 분석하여 추천을 할 수 있는 협업필터링 방법을 제안한다.

### II. Implementation

전체적인 구조는 크게 사용자 성향 판별, 유사한 사용자 선별, 랭킹 알고리즘으로 나누어 개발한다. 사용자 성향 판별은 전처리 작업으로써 지속적으로 소셜 미디어 사용자들의 소셜 활동들을 수집하여 분석한다. 사용자 판별을 위해서 그림1과 같이, 사용자 행위에 대한 값을 정규화하고 행위의 비율 계산하여 행위에 대한 가중치를 도출한다.



Fig. 1. 콘텐츠 추천 절차

계산된 행위 가중치를 통해 최종적인 콘텐츠별 선호도 값을 계산한다. 콘텐츠별 선호도 값은 각 행위에 대한 실제 값과 위의 계산 과정을 통해 얻은 행위 가중치 값을 곱하여 모든 값을 합하게 되면 최종적인 선호도 값이 계산된다. 동일한 리뷰에 대해서 최신의 성향에 가중치를 더해주기 위해서 시간 가중치를 고려한 선호도 값을 계산한다.

$$TW_{(u,i)} = Pref_{(u,i,t)} * \frac{Pref_{(u,i,t)}}{Pref_{(u,i,t-1)}}$$

Pref은 이전 단계에서 계산한 선호도 값이며 t는 시간 값이다. 협업필터링을 위해 코사인 유사도를 이용하며, 두 대상의 데이터 집합이 얼마나 한 직선으로 표현되는지에 대한 측정값에 해당한다. 두 대상 사이의 유사도는 코사인 유사도 수식으로 계산하며, 성향 기중치와 시간 기중치가 적용된 아이템 선호도를 바탕으로 유사도를 계산한다. 코사인 유사도는 수식에서 u는 해당 사용자, e는 유사도 계산의 대상 사용자를 의미하며 i는 과거 평점을 준 콘텐츠 객체, T는 유사도 계산에 사용되는 콘텐츠의 집합을 의미 한다. T는 가장 최근에 평점을 준 콘텐츠 10개와 가장 높은 선호도를 가진 콘텐츠 10개를 결합하여 최저 1개에서 최고 20개까지의 콘텐츠를 선별한다.

$$S(u,e) = \frac{\sum_{i \in I} (r_i(u) - \bar{r}(u))(r_i(e) - \bar{r}(e))}{\sqrt{\sum_{i \in I} (r_i(u) - \bar{r}(u))^2} \sqrt{\sum_{i \in I} (r_i(e) - \bar{r}(e))^2}}$$

선별된 유사 사용자와 해당 사용자의 콘텐츠 이용 정보를 이용해서 협업필터링을 수행한다.

추천의 정확성을 향상시키기 위해서 콘텐츠의 신뢰도를 평가한다. 신뢰도는 서비스 내의 인기도와 SNS에서의 언급량을 기준으로 계산한다. 서비스 내의 인기도는 찜하기, 클릭 수와 평균 평가값을 사용한다. SNS의 언급량은 불리언 기반의 TF-IDF 수식을 통해 계산한다. Like는 i 상품의 찜하기 횟수이며 Stay는 클릭 후 일정시간 머무른 횟수이다. 본 연구에서는 시간을 10초 이상에서 5분까지를 Stay 횟수에 반영하였다. AvgRating은 평균 평점이며 TF-IDF는 SNS에서 상품의 언급량을 계산한 수치이다.

$$Rel_i = (\alpha)((Like_i + Stay_i) * AvgRating_i) + (1 - \alpha)(TF-IDF_i)$$

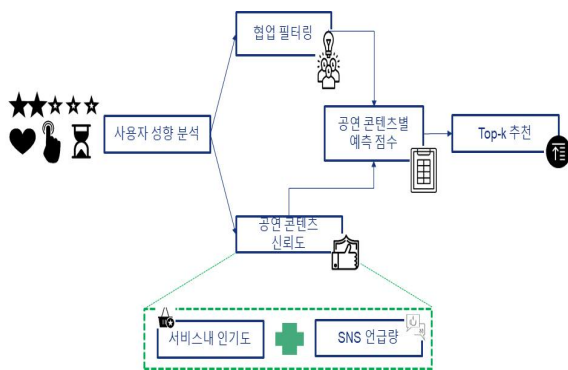


Fig. 2. 사용자 성향 기반 콘텐츠 추천 프로세스

그림2는 사용자 성향에 기반한 개인화 추천에 대한 전체적인 프로세스이다.

그림3은 해당 기술을 이용하여 구현한 공연 콘텐츠 추천 기반의 결과물로 랭킹 알고리즘에 의하여 추천 순서별로 콘텐츠를 나열하고 있다.

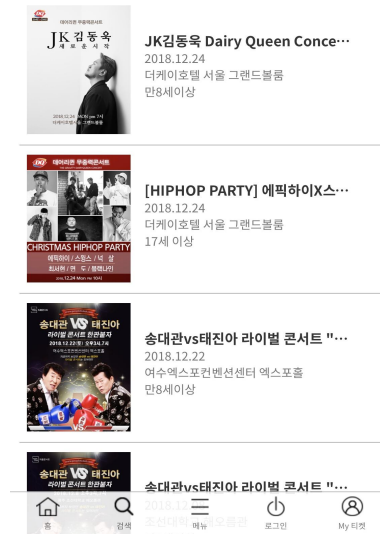


Fig. 3. 결과물 예시

### III. Conclusions

본 논문에서 제안하는 내용은 빅데이터 수집과 처리를 위해 Kafka, Spark, MongoDB와 같은 빅데이터 플랫폼을 활용하였다. 4차 산업혁명 시대의 일상 생활에도 스마트TV, 스마트 스피커 등을 통해 이미 많은 서비스들이 제공되고 있으며, 사용자의 편리를 위한 미디어 콘텐츠 제공 기술에 기반하고 있다. 이는 중장기적으로 다양한 콘텐츠 속에서 미디어 콘텐츠를 사용자 중심으로 추천해주는 기술에 대한 수요를 가져올 것이다. 이에 본 논문에서의 제안 내용은 향후 다양한 분야에서 활용될 수 있다.

### REFERENCES

- [1] Mohammad Julashokria, Mohammad Fathiana, Mohammad Reza Gholamiana, Ahmad Mehrbod, "Improving electronic customers' profile in recommender systems using data mining techniques", Management Science Letters of Growing Science, pp.449-456, 2011.
- [2] O. Shafiq, R. Alhadj and John G. Rokne, "On personalizing Web search using social network analysis," International Journal of the Information Sciences, Vol.314, pp.55-76. 2015.
- [3] Eun-Ju Lee, Je-O Song, Ina Kim\*, Jae-Soo Yoo, "Big-data Analysis based Mobile Services using Individual Skin-type and Genes for Cosmetic Recommendation", Proceedings of the KOCON Conference, Vol.16, No.1, pp.495-496, 2018.