인스타그램 포스트 데이터를 이용한 협업 필터링 기반 맛집 추천 시스템

정한조*, 송은수*, 최현승*, 박원정O
OO목포대학교 융합소프트웨어학과,
*목포대학교 융합소프트웨어학과
e-mail: {hanjojeong, dmstn0331, gustmd96, fpahsckck}@mokpo.ac.krO*

A Collaborative Filtering-based Restaurant Recommendation System using Instagram-Post Data

Hanjo Jeong*, Eunsu Song*, Hyun-Seung Choi*, Won-Jeong ParkO
ODepartment of Software Convergence Engineering, Mokpo National University,
*Department of Software Convergence Engineering, Mokpo National University

● 요 약 ●

최근 소셜 미디어로 이름을 알린 이색 카페와 맛집을 찾아다니는 문화가 확산되는 추세이다. 블로그 포털 검색을 통해 찾아본 맛집은 광고성 게시물이 많아서 신뢰도가 떨어지고, 맛집 관련 게시물 수가 많아서 모든 게시물들을 수동으로 읽기는 불가능하다. 본 논문에서는 사용자들이 선호해서 자발적으로 공유하는 신뢰도 높은 인스타그램의 맛집 포스트 데이터를 이용하여 아이템 기반의 협업 필터링(Item-based Collaborative Filtering) 기법을 통해 사용자의 취향에 맞고 선호할 만한 맛집을 자동으로 추천해주는 알고리즘 및 시스템을 소개한다.

키워드: 추천 시스템(recommendation system), 협업 필터링(collaborative filtering), 인스타그램(Instagram)

I. Introduction

최근 소셜 미디어로 이름을 알린 이색 카페와 맛집을 찾아다니는 문화가 확산되는 추세이다. 블로그 포털 검색을 통해 찾아본 맛집은 광고성 게시물이 많아서 신뢰도가 떨어지고, 맛집 관련 게시물 수가 많아서 모든 게시물들을 수동으로 읽기는 불가능하다. 이러한 문제점을 해결하기 위해 본 논문에서는 인스타크램(Instagram)의 맛집 포스트 데이터를 이용하였다. 인스타그램은 사용자수가 2013년 1억명을 기록한 후, 매년 증가하여 2018년 10억명을 초과하였다. 인스타그램의 맛집 포스트 데이터의 경우, 대부분 사용자들이 직접 방문하여 주문한 메뉴의 사진과 함께 포스트를 업로드하여 자발적으로 공유하기 때문에 신뢰성 높은 데이터로 볼 수 있다.

추가적으로, 인스타그램의 사용자들은 대부분 선호하거나 만족한 경우, 해당 맛집의 메뉴 포스트를 고유한다. 즉, 일반적인 협업 필터링을 이용하는 영화 추천 [1] 데이터와 같이 평점이 없는 반면, Amazon.com의 아이템 기반 협업 필터링(Item-to-Item Collaborative Filtering) [2]에 사용되는 구매 데이터와 같이 사용자가 포스트로 공유한 맛집은 사용자가 선호하는 맛집이라고 가정할 수 있다.

이에 따라 본 논문에서는 아마존의 아이템 기반 협업 필터링 기법과 유사하게 맛집 포스트를 공통으로 게시한 사용자들을 기준으로 각 맛집 간의 코사인 유사도를 구해서 유사도가 가장 높은 Top N 맛집을 자동으로 추천해주는 추천 시스템 관련 방법론 및 알고리즘을 소개한다.

II. Method

1. Data

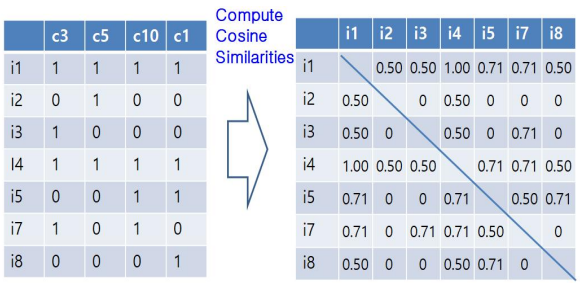
웹 크롤링(Web-crawling)을 통해 인스타그램에 업로드 된 전라남도 목포시 소재의 맛집 약 700개와 관련된 약 8천 개의 포스트 데이터를 수집했다. 크롤링하기 위해 각각의 맛집명을 해시태그로 검색한 후, 검색된 포스트 데이터의 사용자 아이디를 추출하여 (사용자 아이디, 맛집명) 컬럼으로 구성된 테이블을 생성하였다.

2. Item-based Collaborative Filtering using Cosine Similarity

Amazon.com의 아이템 기반 협업 필터링과 유사한 방법으로 추천 대상 사용자가 포스트한 맛집을 다른 사용자 집단을 실시간으로 찾은 후, 해당 사용자 집단에서 가장 많이 포스트한 다른 맛집을 추천 대상 사용자에게 추천한다.

이러한 맛집 포스트를 공통으로 공유한 사용자들을 중심으로 추천할 다른 맛집을 찾기위해서 코사인 유사도(Cosine Similarity) 함수를 이용하였다. 즉, 각각의 맛집 포스트를 공유한 사용자이이디 각각을 차원으로하여 포스트한 경우 1을 할당, 포스트하지 않은 경우 0을 할당하는 방식으로 각각의 맛집 벡터를 생성한다.

각각의 맛집 벡터 상호간의 코사인 유사도를 계산하여 Fig. 1과 같은 유사도 행렬을 생성한다. 유사도 행렬은 각 원소가 각각의 맛집, i_k 에 대한 상호간의 코사인 유사도를 나타내므로 대칭(Symmetric) 행렬이다.



	c3	c5	c10	c1
i1	1	1	1	1
i2	0	1	0	0
i3	1	0	0	0
I4	1	1	1	1
i5	0	0	1	1
i7	1	0	1	0
i8	0	0	0	1

	i1	i2	i3	i4	i5	i7	i8
i1		0.50	0.50	1.00	0.71	0.71	0.50
i2	0.50		0	0.50	0	0	0
i3	0.50	0		0.50	0	0.71	0
i4	1.00	0.50	0.50		0.71	0.71	0.50
i5	0.71	0	0	0.71		0.50	0.71
i7	0.71	0	0.71	0.71	0.50		0
i8	0.50	0	0	0.50	0.71	0	

Fig. 1. Cosine Similarity Matrix for Each Restaurant Vector

최종적으로, Fig.1의 유사도 행렬을 이용하여 사용자가 최근에 포스트한 맛집과 가장 유사도가 높은 5개의 맛집을 찾아서 추천한다. Table 1은 상호명이 '구라파소년'인 맛집을 검색했을 때, 코사인 유사도가 높은 순으로 5개의 다른 맛집을 추천한 결과이다.

Table 1. Top 5 Recommended Restaurant for a Restaurant '구라파소년'

상호명	코사인 유사도
이태리반점	0.500
비스트로로지	0.456
선셋테이블	0.456
곰집갈비	0.354
기찻길315	0.354

III. Conclusions

본 연구에서는 인스타그램 포스트 데이터를 이용하여 아이템 기반 협업 필터링 기반의 추천 시스템을 제안하였다. 인스타그램 포스트 데이터는 영화 리뷰 데이터와 달리 평점을 부여하지 않기 때문에, 일반적인 협업 필터링 기법을 사용할 수 없다. 또한, 인스타그램 사용자들은 본인들이 선호하는 맛집만 포스트하는 경향이 있기 때문에, 평점이 없어도 포스트한 맛집은 Amazon.com의 아이템기반 협업필터링의 '구매 제품=선호 제품' 가정처럼 사용자가 선호하는 맛집이라고 가정할 수 있다. 이러한 이유로 본 연구에서 제안한 추천 기법은 인스타그램 및 다른 대부분의 맛집 블로그와 같이 평점이 없는 데이터에도 효과적으로 적용될 수 있다.

ACKNOWLEDGEMENT

본 논문은 2020학년도 MNU혁신사업 지원에 의하여 연구되었음. 또한, 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2019-0744).

REFERENCES

[1] J. Hyun, S. Ryu, and S. Y. T. Lee, "How to improve the accuracy of recommendation systems: Combining ratings and review texts sentiment scores," Journal of Intelligence and Information Systems, Vol. 25, No. 1, pp. 219-239, 2019.

[2] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," IEEE Internet computing, Vol. 7, No. 1, pp. 76-80. 2003.