

객체 간 관계 정보를 포함하는 지식 그래프 구축 기법 및 추천 시스템에서의 활용 방안

박성준, 배홍균, 채동규, 김상욱*
한양대학교 컴퓨터·소프트웨어학과
{tjdwns32, hongkyun, dongkyu, wook}@hanyang.ac.kr

An Approach to Constructing Knowledge Graph for Recommender Systems based on Object Relations

Sung-Jun Park, Hong-Kyun Bae, Dong-Kyu Chae, Sang-Wook Kim
Dept. of Computer Science, Hanyang University

요 약

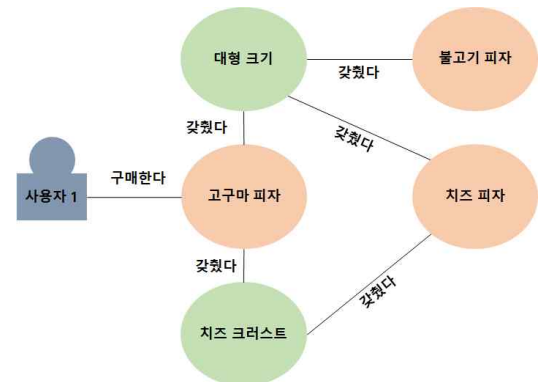
최근 사용자, 상품, 그리고 상품의 메타 정보 사이의 관계를 표현한 지식 그래프 (knowledge graph) 가 추천 시스템 분야에서 많은 관심을 받고 있으며 활발히 이용되고 있다. 하지만 기존의 지식 그래프는 각 노드 (사용자, 상품, 메타 정보 등) 사이의 단순한 사실 관계만을 표현하고 있으며, 이는 사용자의 선호도를 정확히 파악하는 데 한계가 있다. 본 논문에서는 지식 그래프의 정보 부족 문제를 보완하기 위해 각 상품에 남겨진 텍스트 리뷰를 감정 분석 (sentiment analysis) 하고, 이를 각 노드 간의 선호도 정보로 활용하여 지식 그래프를 구축하는 방법을 제안한다.

1. 서론

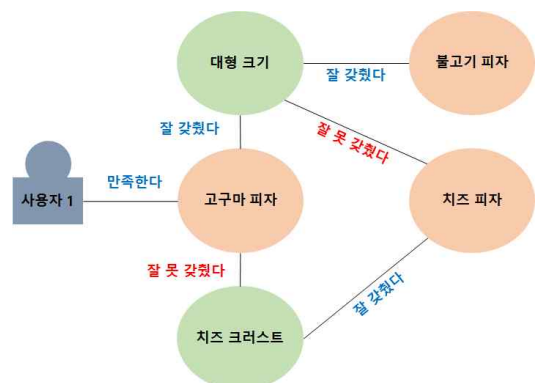
지식 그래프 (knowledge graph)는 최근 추천 시스템 분야에서 많이 사용되고 있는 데이터 저장 모델로 [1, 2], 두 객체 h, t 와 그 사이의 관계 r 을 (h, r, t) 트리플렛으로 표현한다. 추천 시스템에서 활용되는 지식 그래프의 경우 보통 사용자, 상품 그리고 상품의 메타 데이터와의 관계를 나타내서 사용자의 선호도를 파악하는 데 사용된다. 하지만 두 객체 간 단순 사실 관계만 표현한 지식 그래프는 추천해주고자 하는 타겟 사용자의 선호도를 정확히 파악하는 데 한계가 있다.

예를 들어, 그림 1에 표현된 지식 그래프를 활용해서 사용자 1에게 피자 메뉴를 추천할 때, 추천 시스템은 사용자 1이 구매한 고구마 피자가 갖춘 특징인 대형 크기와 치즈 크러스트를 갖추고 있는 치즈 피자를 추천할 가능성이 높다. 하지만 객체 간 선호 관계가 표현된 그림 2의 지식 그래프를 보면, 사용자 1이 고구마 피자를 구매하고 만족한 이유는 고구마 피자가 잘 갖춘 특징인 대형 크기 때문이고, 따라서 이 특징을 잘 갖춘 불고기 피자를 추천하는 것이 사용자 1의 선호도를 잘 파악한 추천이라고 볼 수 있다.

본 논문에서는 사용자가 상품에 부여한 텍스트 리뷰 (text review)를 활용해서 객체 간 선호 관계가 표현된 지식 그래프를 구축하는 방법을 제안한다.



(그림 1) 사실 관계만 표현한 지식 그래프 예시



(그림 2) 객체간 선호도를 표현한 지식 그래프 예시

* 교신 저자

2. 텍스트 리뷰를 통한 객체 간 선호 관계 파악

텍스트 리뷰는 사용자가 상품을 구매한 후 남긴 글로, 각 상품의 고유한 특징들과 그에 대한 사용자의 평가 등 사용자의 선호도 파악에 필요한 많은 정보를 담고 있다. 텍스트 리뷰에서 상품의 특징과 사용자, 상품 간의 선호 관계를 파악하기 위해서 우리는 참고 문헌 [3] 에서 사용한 sentiment analysis 기법을 사용했다. 선호 관계를 파악하는 단계는 다음과 같다. 우선, 텍스트 리뷰의 문법, 형태를 고려해서 상품의 특징이라고 할 수 있는 단어들의 집합 F (feature) 를 추출한다. 그 다음, 특징에 대한 의견 단어들의 집합 O (opinion) 를 만들어서 (F, O) 쌍을 만든다. 마지막으로 각 (F, O) 쌍이 텍스트 리뷰에서 발견되었을 때 긍정적인지 부정적인지를 나타내는 감정 S (sentiment) 를 결정해서 (F, O, S) 트리플렛을 완성한다.

사용자가 작성한 텍스트 리뷰에 감정 분석 기법을 적용하면, 사용자가 상품의 어떤 특징을 고려하여 상품을 구매했는지, 그리고 이 특징에 대하여 어떤 평가를 내렸는지 등에 관한 정보를 알 수 있다.

3. 실제 데이터셋에 적용 결과

<표 1> Amazon beauty dataset

객체 타입	사용자	상품	특징	브랜드	카테고리
객체 수	22,363	12,101	1,892	2,077	248

우리는 선호 관계 정보가 반영된 지식 그래프 구축을 위해 Amazon Dataset¹의 beauty 카테고리의 데이터셋을 사용했다. 표 1 에는 데이터셋에 대한 정보가 표시되어 있다. 브랜드, 카테고리는 상품의 메타 데이터고, 특징은 텍스트 리뷰에서 추출된 상품의 특징이다.

기존 메타 데이터와 텍스트 리뷰만 사용하면, 사용자와 상품, 사용자와 특징, 그리고 상품과 특징 사이를 각각 ‘구매한다’, ‘언급한다’, 그리고 ‘갖추다’와 같은 단순 사실 정보로만 표현할 수 있다.

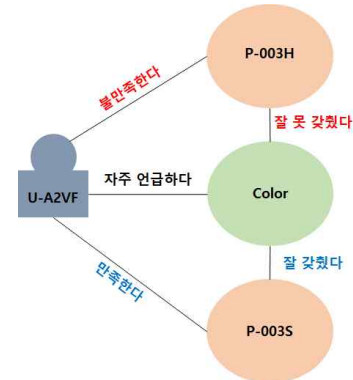
<표 2> 감정 분석 기술 적용 전후의 관계 변화

감정 분석 기술 적용 전	감정 분석 기술 적용 후
구매한다 (사용자 → 상품)	불만족한다 보통이다 만족한다
언급한다 (사용자 → 특징)	가끔 언급한다 자주 언급한다 매우 자주 언급한다
갖추다 (상품 → 특징)	잘 못 갖추다 갖추다 잘 갖추다

우리는 2 장에서 구한 (F, O, S) 트리플렛을 사용하여 일반적인 사실관계를 선호 관계에 따라 나눴다. 이때, 두 객체 간 관계를 항상 긍정, 부정 중 하나로 확정 짓기에는 위험이 있어서 기준을 세워서 3 가지

단계로 관계를 구분했고, 표 2 에 정리하였다.

그림 3 은 우리 방법으로 구축한 beauty dataset 의 지식 그래프 일부를 보여준다. 사용자 ‘U-A2VF’ 가 자주 언급하는 특징인 ‘color’ 를 잘 갖춘 상품인 ‘P-003S’ 는 사용자에게서 만족한다는 평가를 받았지만, 그렇지 못한 상품 ‘P-003H’ 은 불만족한다는 평가를 받았다. 우리는 이것을 보고 사용자가 상품 ‘P-003S’ 를 좋아하는 이유 중 하나로 color 가 맘에 들기 때문이라고 예상할 수 있다.



(그림 3) Beauty dataset 지식 그래프 구축 결과 예시

4. 향후 연구 계획

우리는 실험을 통해 객체 간 선호 관계가 잘 반영된 지식 그래프를 성공적으로 구축하였다. 이 그래프는 크게 두 가지 측면에서 활용될 수 있는데, 첫 번째는 사용자의 정확한 선호도를 파악하고, 이에 기반해 더 정확한 추천을 해줄 수 있는 추천 모델을 개발하는 것이다. 두 번째는 이 그래프 상에서의 사용자로부터 추천된 상품까지의 경로 (path) 를 이용해서 추천의 이유를 설득력 있게 설명해주는 모듈을 개발하는 것이다.

감사의 글

이 논문은 (1) 2019 년도 정부(교육부)의 재원으로 한국연구재단의 지원과 (No. 2019R1I1A1A01061588), (2) 2018 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원과 (No. 2018R1A5A7059549), (3) 2017 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단-차세대정보·컴퓨팅기술개발사업의 지원을 받아 수행된 연구임 (No. NRF-2017M3C4A7069440).

참고문헌

- [1] H. Wang et al. "DKN: Deep knowledge-aware network for news recommendation.", *WWW, Lyon, France, 2018*, pp. 1835-1844.
- [2] X. Wang et al. "Explainable reasoning over knowledge graphs for recommendation.", *AAAI, Honolulu, USA, 2019*, pp. 5329-5336.
- [3] Y. Zhang et al. "Explicit factor models for explainable recommendation based on phrase-level sentiment analysis.", *ACM SIGIR, Gold Coast, Australia, 2014*, pp. 83-92

¹ <http://jmcauley.ucsd.edu/data/amazon/>