

한국지능정보시스템학회 [2021 추계학술대회] 논문 투고

논문 전문(Full Paper) 제출 희망 여부 (√표시)	예 [<input type="checkbox"/>]	아니오 [<input checked="" type="checkbox"/>]
심사용 전문 제출 용도 (√표시) - 복수 선택 가능	우수논문 심사 [<input type="checkbox"/>] 학술지 Fast Track 심사 [<input type="checkbox"/>]	해당사항 없음

도서카테고리와구매 및 평점 히스토리 기반의 설명 가능한 추천시스템 개발

윤태준 (한양대학교 공과대학 컴퓨터소프트웨어학부 4 학년 학부생, 공동주저자 tj1616@hanyang.ac.kr)
 이강민 (한양대학교 공과대학 컴퓨터소프트웨어학부 4 학년 학부생, 공동주저자 kangs7434@gmail.com)
 정재용 (한양대학교 공과대학 컴퓨터소프트웨어학부 4 학년 학부생, 공동주저자 wodydthk7@gmail.com)
 채동규 (한양대학교 공과대학 컴퓨터소프트웨어학부 조교수, 교신저자 dongkyu@hanyang.ac.kr)

[저자 연락처]

◎ 윤태준(010-2434-2254)

주소 : 서울시 성동구 왕십리로 222 한양대학교 공과대학 컴퓨터소프트웨어학부

◎ 이강민(010-4445-2906)

주소 : 서울시 성동구 왕십리로 222 한양대학교 공과대학 컴퓨터소프트웨어학부

◎ 정재용 (010-4804-1813)

주소 : 서울시 성동구 왕십리로 222 한양대학교 공과대학 컴퓨터소프트웨어학부

◎ 채동규 (010-4951-1373)

주소 : 서울시 성동구 왕십리로 222 한양대학교 공과대학 컴퓨터소프트웨어학부

도서 카테고리화 구매 및 평점 히스토리 기반의 설명 가능한 추천시스템 개발

윤태준¹

한양대학교

컴퓨터소프트웨어학부

tj1616@hanyang.ac.kr

이강민¹

한양대학교

컴퓨터소프트웨어학부

kangs7434@gmail.com

정재용¹

한양대학교

컴퓨터소프트웨어학부

wodydthk7@gmail.com

채동규

한양대학교

컴퓨터소프트웨어학부

dongkyu@hanyang.ac.kr

Abstract - 본 논문에서는 사용자의 리뷰로부터 감정분석을 수행하여 중요한 키워드를 추출하고, 이를 기반으로 지식 그래프를 생성하는 방안을 제안한다. 생성된 지식 그래프는 추후 설명가능한 추천시스템을 위한 도구로써 사용된다. Amazon Books 의 평점-리뷰 데이터를 이용해서 실험한 결과 사용자들에게 추천 결과와 함께 의미있는 설명이 도출된 것을 확인하였다.

Key Terms - 협업필터링, 설명 가능한 추천시스템, 감정분석, 지식그래프, 인공지능

본 연구는 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원 (No.2020-0-01373, 인공지능대학원지원(한양대학교))과 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학지원사업의 연구결과로 수행되었음 (2016-0-00023)

1. 서론

추천시스템은 오늘날 산업 전반에서 사용자에게 개인화된 제품이나 콘텐츠를 제공하기위해 필수적으로 이용되고 있다 (He et al., 2017)(Wang et al., 2019). 최근에는 단순히 개인에게 추천결과만을 보여주는 것이 아니라, 해당 추천 결과가 만들어지기까지의 과정을

설명할 수 있는 설명 가능한 추천시스템에 대한 관심이 커지고 있다. 이를 구현하기 위해 많은 접근방법이 있으며, 그 중 본 논문에서는 지식그래프 데이터 구조를 기반으로 하는 설명가능한 추천시스템에 대해서 다룬다.

지식그래프는 서로 연관이 있는 객체들을 간선과 노드로 표현한 데이터 구조이다. 일반적인 상품 추천 시스템에서 활용되는 지식그래프의 경우 사용자, 상품, 상품에 대한 메타데이터 등이 노드를 구성하며 간선은 노드들 간의 관계를 나타낸다. 예를 들면 넷플릭스의 사용자들, 콘텐츠들, 그리고 장르, 배우, 감독 등 콘텐츠의 메타데이터 등이 노드를 구성하며, 어떤 사용자가 어떤 콘텐츠를 봤는지, 그리고 어떤 장르와 어떤 배우가 출연하는 것을 봤는지 등의 히스토리를 기반으로 지식 그래프가 생성된다. 그 후 어떤 알고리즘이 A 사용자에게 B 상품을 추천한다면, 지식 그래프 상에서 노드 A 에서 노드 B 까지의 경로가 곧 추천에 대한 설명이 된다.

그러나 기존의 추천시스템에서의 지식그래프에는 상품에 대한 키워드 정보가 고려되지 않았다. 키워드는 사용자가 작성하는 리뷰들에서 추출할 수 있으며, 다양한 정보를 나타낼 수 있다. 예를 들면 사용자의 상품에 대한 감정정보를 나타낼 수 있으며, 상품 고유한 특성이 될 수도 있다. 이를 지식그래프에 반영한다면 더 의미있는 설명을 제공할 수 있을 것으로 기대한다.

¹ 공동제 1 저자

이에 따라 본 논문에서는 사용자가 상품을 구매한 후 작성한 리뷰 데이터와 평점 데이터를 기반으로 추출한 키워드, 그리고 사용자의 감정을 분석한 감정기반의 설명가능한 추천시스템을 제안한다. 또한 이를 위한 키워드가 포함된 지식그래프의 구현과 더욱 설득력 있는 추천에 대해서 논의한다.

2. 리뷰데이터 분석을 통한 감정기반 지식그래프

먼저 키워드를 고려하기 위해 본 논문은 상품의 특징을 나타내는 단어 집합을 추가로 사용한다. 이러한 집합은 사용자가 남긴 리뷰 데이터로부터 얻을 수 있다. 리뷰 데이터는 사용자가 상품을 구매한 후 남긴 텍스트 데이터이며, 상품의 특징을 나타내는 키워드와 상품에 대한 사용자의 감정을 나타내는 단어들을 포함하고 있어 감정 분석에 용이하게 사용될 수 있다.

리뷰 데이터에서 키워드들의 집합을 얻어내는 과정은 다음과 같다.

1. 모든 리뷰 데이터를 음절 단위로 나눈다.
2. 단어들 중 빈도수가 6000 이상인 것들과 TF-IDF 값이 0.1 이하인 것들을 제거한다.
3. 명사와 형용사가 아닌 단어들은 제거한다.

위의 2 번과 3 번 절차는 단어들 중 의미 없는 (예를 들어 a, the) 단어들은 키워드로 간주하지 않기 위함이다.

이후에는 앞서 구성한 노드들 간의 관계에서 긍정과 부정을 분석하고 간선으로 연결한다. 먼저 사용자, 상품 간의 관계는 사용자가 상품을 구매하고 남긴 평점 데이터를 이용해서 나타낼 수 있다. 평점 데이터는 1 에서 5 까지의 정수로 이루어져 있으며, 3 점을 기준으로 1, 2 점은 불만족, 3 점은 보통, 4, 5 점은 만족으로 나타내어 사용자의 감정을 반영할 수 있다.

사용자, 카테고리 간의 관계는 사용자가 해당 카테고리에 포함되는 상품을 하나 이상 구매했을 때 긍정으로 나타내며, (해당 카테고리에 포함되는 상품 구매 횟수)/(전체 상품 구매 횟수)를 간선의 가중치로 사용한다. 사용자와 키워드 집합 간의 관계 또한 감정정보를 모델링 할 수 있다. 본 논문에서는 해당 사용자가 작성한 리뷰 데이터에서의 키워드 빈도수를 이용해서 파악한다. 즉, 키워드들 중 사용자의 리뷰 데이터 집합에서의 빈도수가 높은 키워드들은 사용자가 자주 언급한 것이므로 긍정, 빈도수가 낮은 키워드들은 부정으로 간주한다. 끝으로 상품과 키워드 간의 감정 관계는 상품 별로 키워드가 언급된 횟수를 계산하고 언급 횟수 상위 20% 는 긍정, 그 외의 간선은 부정으로 간주한다.

3. 실험 및 결론

<표 1> Amazon Books 데이터 정보.

노드 유형	사용자	도서	키워드	카테고리
노드 개수	369,493	43,391	66,109	449

위에서 기술한 감정 기반 지식그래프를 구현하기 위해 우리는 Amazon 의 Book 데이터셋을 사용했다. 표 1 은 해당 데이터의 정보를 나타낸다. 또한 위 데이터에서 추출된 키워드의 수도 표시하였다.

타겟 사용자 A 에게 추천되는 상품 B 를 임의의 협업 필터링 알고리즘으로 도출했다고 가정하자. 그 후, 구현된 감정 기반 지식그래프에서 사용자 A 에게 추천되는 상품 B 의 추천 경로를 얻기 위해서 다음과 같은 절차를 거친다.

1. 일단 모든 간선의 가중치를 고려하지 않은 채로 사용자 A 에서 상품 B 까지의 경로 중 짧은 순서대로 100 개를 추출한다.
2. 100 개의 경로들 각각의 가중치 합을 구한다.
3. 가중치가 높은 상위 몇 개의 경로를 설명으로써 사용자에게 제공한다.

위 과정을 거쳤을 때, 예를 들어 타겟 사용자 ‘A26PRNQYA3Y662’ (사용자 ID 는 익명화되었음)에게 추천된 책 “Notorious Nineteen: A Stephanie Plum Novel” 의 경우 도출된 경로 중 하나로 ‘A26PRNQYA3Y662’ => ‘United States’ => ‘A12122R63P29R2’ => “Notorious Nineteen: A Stephanie Plum Novel” 을 얻을 수 있다. 이 경로를 해석하면, 사용자 ‘A26PRNQYA3Y662’ 가 ‘United States’ 카테고리의 책을 구매하고, 사용자 ‘A12122R63P29R2’ 또한 해당 카테고리의 책을 구매했을 때, 사용자 ‘A12122R63P29R2’ 가 구매한 또 다른 상품인 “Notorious Nineteen: A Stephanie Plum Novel” 을 결과적으로 추천해준다는 의미이다.

이와 같이 사용자에게 어떤 상품을 추천해줄 때 추천 경로를 통해 더욱 설득력 있는 추천이 가능함을 볼 수 있다.

4. 참고문헌

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017, April). Neural collaborative filtering. In Proceedings of the 26th international conference on world wide web (pp. 173-182).

Wang, X., He, X., Wang, M., Feng, F., & Chua, T. S. (2019, July). Neural graph collaborative filtering. In Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval (pp. 165-174).