

온라인 리뷰의 속성별 감성값을 이용한 호텔추천

홍준우(부산대학교 경영학과)

김은미(국민대학교 정보기술연구소)

김상희(부산대학교 BK21 Four 디지털금융 교육연구단)

홍태호(부산대학교 경영학과)

2021년 추계 학술대회 한국인터넷전자상거래학회:
메타버스 기반 사회와 경영혁신

목 차

1.서론

2.문헌연구

- 추천시스템: 협업필터링

3.연구 프레임워크

4. 실험 및 실험결과

- 데이터 수집 및 전처리
- 주제별 문장 추출
- 호텔 추천 및 결과

5. 결론

1. 서론

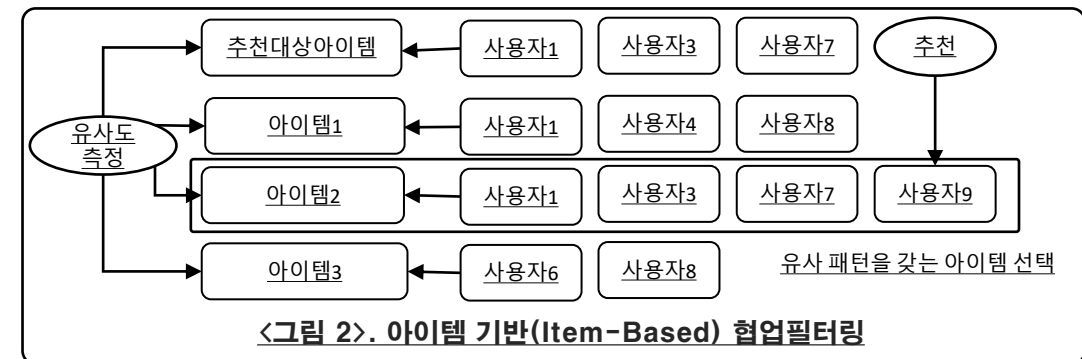
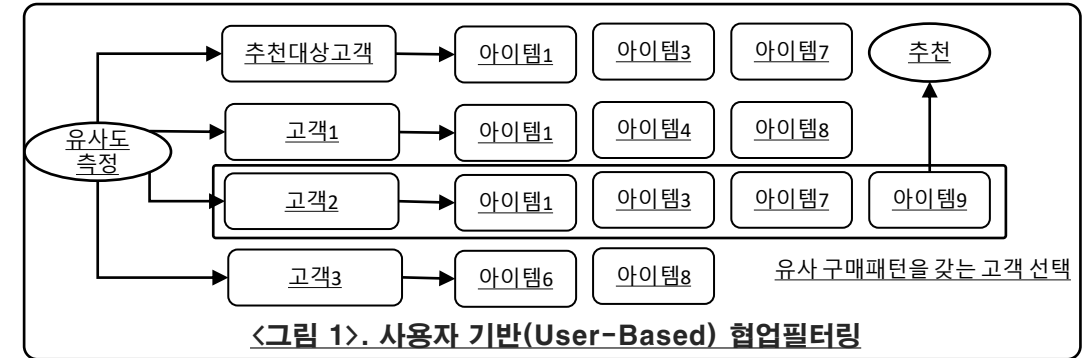
- 디지털 기술(소셜 미디어, 온라인 리뷰 등)이 다양한 산업에 융합되면서 패러다임의 변화를 가져와 소비자에게 많은 선택의 기회를 제공한다(구철모, 2017).
- 온라인 리뷰는 소비자의 구매 경험을 기반으로 작성되어, 입장에 따른 이점을 갖는다.
 - 고객: 상품 탐색 및 구매 과정에서 구매 의사결정에 대한 이점을 얻는다(Cheng and Ho., 2015)
 - 기업: 고객 분석을 통해 매출 증가 원인, 고객 패턴 파악 등에서 이점을 얻는다(Li et al., 2013).
- 전자상거래 시장의 성장으로 온라인 리뷰는 다양화되고 방대해 졌으며(유병준, 2019), 최근 코로나19 팬데믹의 영향으로 더욱 가속화되고 있다(Al-maaitah et al., 2021).
- 시장의 활성화에 따른 고객 유인으로 고객의 니즈에 맞는 상품이 생겨나고 있지만, 지나친 상품의 다양화는 직관적인 선택과 의사결정 능력을 저하시키고 있다(이진화 등, 2008).
- 상품 속성은 고객이 상품을 구매할 때 효과적인 의사결정 도와주는 기능을 한다(문혜선, 2015).
- 기존 상품추천모형은 상품 예측을 위해 전체 선호도를 나타내는 평점을 활용하는데, 이는 세부 항목에 대한 선호도를 반영하지 못한다(Zhanget al., 2014). 이에 효과적인 구매 의사결정을 위해 상품속성을 반영한 연구가 필요하다.
- 따라서, 본 연구에서는 효과적인 의사결정을 위해 **리뷰의 속성을 반영한 개인화 상품추천** 모형을 제안한다.

2. 문헌연구: 협업필터링

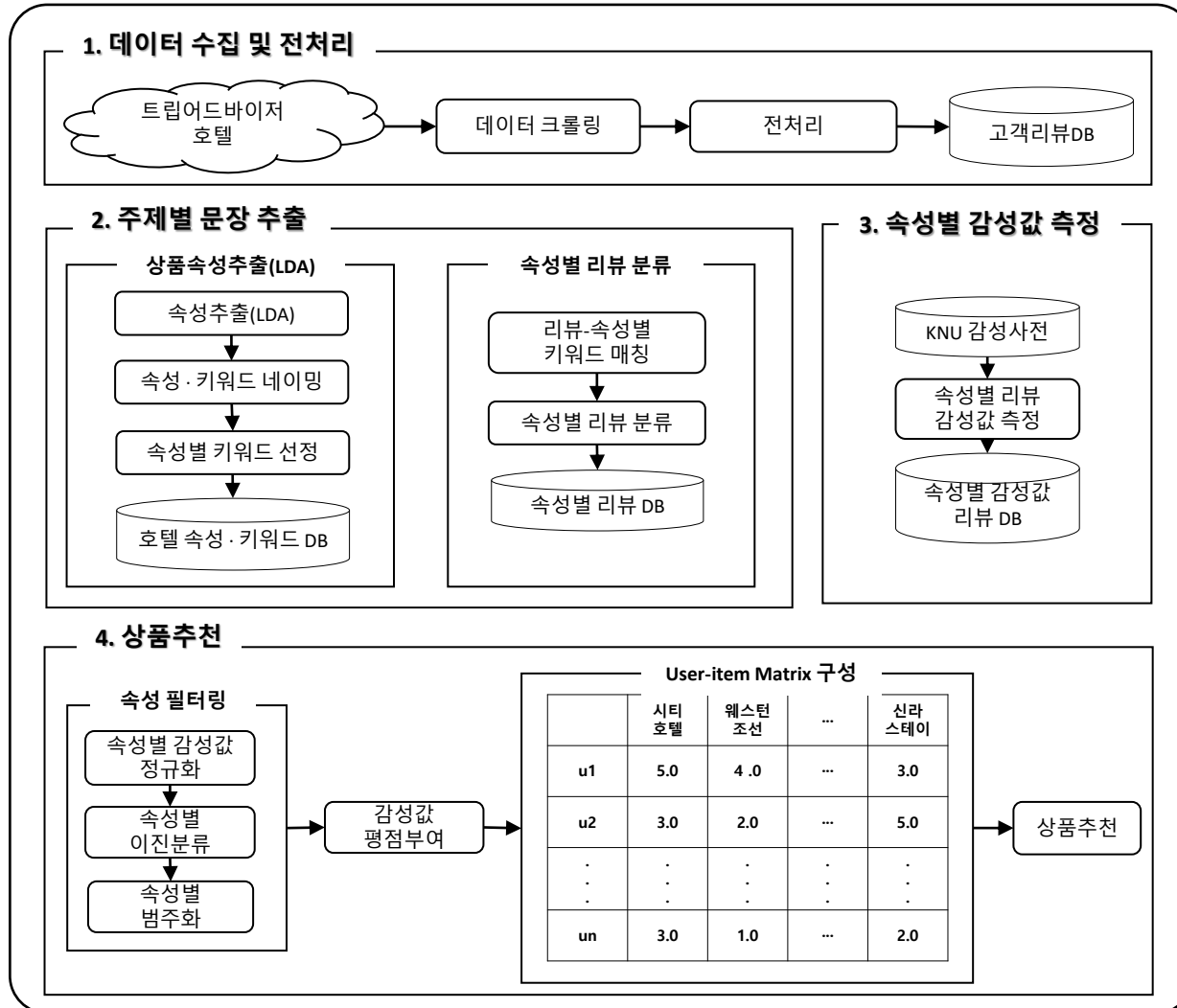
- 추천 시스템이란, 고객에게 상품정보를 제공하여 효과적인 타겟 마케팅 역량을 제공한다(Linden et al.,2003).

[협업필터링(Collaborative Filtering, CF)]

- 협업필터링은 메모리 기반(Memory based)과 모델 기반(Model-based)으로 구분하며, 메모리 기반의 경우, 사용자 기반(User-Based)과 아이템 기반(Item-Based)로 분류할 수 있다 (손지은 등, 2015).
- 사용자 기반(User-Based) 협업필터링이란, 추천 대상 고객과 다른 고객과의 동일 구매 이력을 기반으로 유사도를 측정하여 높은 유사도를 갖는 고객을 선택하는 방법.
 - 사용자 기반 K-NN 협업필터링을 통해 장르별 협업필터링을 제안하여, 기존 협업필터링의 한계와 성능을 향상시켰다(이재식 & 박석두, 2007).
 - 기존 협업필터링에서 산업·수준을 고려한 최근접 이웃 협업필터링을 제안했다(권병일 & 문남미, 2010)
- 아이템 기반(Item-Based) 협업필터링이란, 추천 대상 아이템을 기준으로 유사한 아이템을 선택한 뒤, 선정 아이템을 구매하였으나, 추천 대상 아이템을 구매하지 않은 다른 고객에게 아이템을 추천하는 방법.
 - 음악 콘텐츠 추천에서 아이템 간 협업필터링을 통해 사용자 기반 방식 보다 폭 넓은 수준의 콘텐츠를 추천제안을 할 수 있음을 제시했다(유영석 등, 2017).



3. 연구 프레임워크



[데이터 수집 및 전처리]

- 호텔 및 음식점 등의 정보를 제공하는 트립어드바이저에서 데이터 수집
- 전처리: 불필요 데이터 제거 및 한국어 텍스트 변환

[주제별 문장 추출]

- 상품속성추출: LDA를 통해 추출한 토픽을 6가지 속성으로 네이밍 그리고 속성별 키워드를 선정
- 속성별 리뷰 분류: 속성별 키워드를 통해 속성별 리뷰를 분류

[속성별 감성값 측정]

- 감성분석: KNU감성사전을 기반으로 6가지 속성별 감성값 측정

[상품추천]

- 속성 필터링: 대상의 관심속성을 반영하기 위한 과정(정규화, 이진분류, 범주화)
- 감성값 평점: 속성별 감성값을 5점 척도로 변환
- 상품추천: User-Item Matrix 구성 및 KNN기반 상품추천

<그림 3>. 연구 프레임워크

4. 실험 및 실험결과(1)

4.1 데이터 수집 및 전처리

[데이터 설명]

- 데이터 출처: 트립어드바이저(Tripadvisor)
- 데이터 기간: 2010.01 ~ 2021.09
- 호텔 수: 31개
- 데이터 수: 17,653

[데이터 전처리]

- 제거: 웹사이트 링크, 불용어, 특수문자, 숫자, 공백
- 변환: 띄어쓰기, 맞춤법, 반복문자(이모티콘, 자음, 모음), 외래어

항목	국가				
국가/도시	대한민국/서울				
총 기간	2010.01 ~ 2021.09				
리뷰 수	17,653				
호텔 수	31				
속성평점 구분	1(156)	2(3,248)	3(10,156)	4(3,979)	5(114)
총 평점 구분	1(173)	2(213)	3(844)	4(4,060)	5(12,363)

<표 1>. 호텔 리뷰 데이터 수

4.2 주제별 문장 추출(1)

[호텔 속성 추출]

- LDA의 최적의 토픽 수(k)를 추출하기 위해 혼잡도(perplexity)를 구했으며, 20을 도출함.
- 트립어드바이저의 호텔 리뷰의 6가지 속성을 기준으로 토픽을 추출하여 네이밍.

[호텔 속성 추출]

- 호텔의 6가지 속성별 키워드 선정
- 속성별 키워드를 기반으로 온라인 리뷰 분류

토픽 수	토픽	단어	상위 키워드
1	서비스	체크	체크, 예약, 직원, 요청, 서비스, 체크인아웃, 사항
2	서비스	카페	카페, 야외, 좋아여, 많아, 괜찮고, 친절하셨고
3	가격	호텔	호텔, 시설, 가격, 객실, 위치
...
20	서비스	맛있어요	맛있어요, 쌀국수, 다방, 오믈렛, 웨스턴, 계란

<표 2>. 토픽 및 단어추출 및 속성별 네이밍

4. 실험 및 실험결과(2)

4.2 주제별 문장 추출(2)

도시	속성	키워드
서울	서비스	체크, 예약, 직원, 요청, 서비스, 체크아웃, 사항, 카페, 야외, 좋았어, 많아, 괜찮고, 친절하셨고, 라운지, 이용, 조식, 음식, 클럽, 수영장, 뷔페, 식사, 저녁, 사우나, 저렴하고, 친절하네요, 만족한, 밤늦게, 좋아했습니다, 좋았습니다, 좋았어요, 깔끔하고, 깨끗하고, 좋고, 친절하고, 레스토랑, 아침, 직원, 직원, 체크, 방문, 투숙, 친절하게, 프런트, 오물렛, 웨스턴, 계란, 맛있어요, 쌀국수, 다방
	장소	명동, 서울, 주변, 근처, 쇼핑, 교통, 광화문, 롯데, 백화점, 타워, 마트, 잠실, 석촌호수, 공항, 명동, 거리, 쇼핑, 걸어서
	가격	고풍, 생일, 세트, 파티, 위치, 시설, 가격, 객실, 호텔
	청결도	화장실, 샤워, 냄새, 욕실, 청소, 욕조
	객실	온도, 에어컨, 조절, 난방, 충전기, 교체, 발생, 쾌적한, 욕상, 세탁기, 테이블, 요리, 다리미, 인테리어, 느낌, 분위기, 깔끔하게, 모던, 럭셔리, 구성, 고급스러운, 깔끔한
	수면의 질	침대, 디럭스, 사이즈, 공간, 넓고, 베드, 패밀리

〈표 3〉. 6가지 속성별 키워드

호텔 속성	서비스	장소	가격	청결도	객실	수면의 질
리뷰 수	16,371	9,100	14,189	2,378	4,145	3,034

〈표 4〉. 호텔 속성별 온라인 리뷰 분류 수

4.3 호텔추천 및 결과(1)

[속성별 감성값 기반 필터링]

- 정규화: 정규화 범위를 “-1 ~ 1” 로 정의
- 이진분류: (정규화 감성값 X)
 - “ -0.5 >= x, “-0.38 <= x” → “1” 변환
 - 이외 “0” 변환
- 범주화: 속성별 데이터를 분류

속성	데이터 수
서비스, 장소, 가격	4,330
서비스, 가격	3,208
서비스	1,360
서비스, 장소, 가격, 객실	1,124
서비스, 가격, 객실	1,042
서비스, 장소, 가격, 수면의 질	768
서비스, 장소	594
서비스, 가격, 수면의 질	456
서비스, 장소, 가격, 청결도	436
서비스, 객실	381

〈표 5〉. 속성별 데이터 범주화 수(상위 10개)

4. 실험 및 실험결과(3)

4.3 호텔추천 및 결과(2)

[호텔추천]

- 속성별 감성값 평점을 5점척도로 변환
- User-item Matrix 구성:
 - (1). 속성별 평점 기반
 - (2). 코사인 유사도 측정

5점 척도	속성별 감성값	감성척도 (추천여부)
1	-8 ~ -3	부정(추천)
2	-2 ~ 0	부정(추천)
3	1 ~ 5	긍정(선택)
4	6 ~ 13	긍정(추천)
5	14 ~ 21	긍정(추천)

〈표 6〉. 속성별 세부평점 정의

ID	호텔	롯데호텔월드	밀레니엄 힐튼	신라스테이 (광화문점)	코리아나	...	서울웨스턴 조선	호텔 페이토
Andrew		4.0	0.0	0.0	0.0	...	3.0	0.0
Daniel		3.0	3.0	0.0	0.0	...	3.0	0.0
David		3.0	3.0	3.0	0.0	...	0.0	3.0
Eric		0.0	0.0	0.0	3.0	...	0.0	0.0
...	
Michael		3.0	4.0	0.0	3.0	...	4.0	3.0
Tim		0.0	0.0	0.0	0.0	...	3.0	0.0

〈표 7〉. 호텔별 사용자 감성값 매트릭스

고객ID	코사인 유사도
Andrew	0.65
Tim	0.55
Daniel	0.54
David	0.48
Kim	0.32

〈표 8〉. Michael와 유사한 고객(Top5)

호텔이름	속성평점	추천여부
티마크 그랜드(명동점)	4	추천
오크우드 프리미어 코엑스 센터	4	추천
신라스테이(광화문점)	3	선택
임피리얼 팰리스 서울	3	선택
서머셋팰리스서울	3	선택

〈표 9〉. 호텔 추천 결과(속성평점 기준(오른쪽), 총 평점 기준(왼쪽))

[추천 결과]

- 추천대상: Michael
- 고려속성: 서비스, 장소, 가격
- 속성평점 및 총 평점 기반 추천결과에서 전반적으로 유사한 결과를 도출했다.
- MAE, RMSE를 통한 성능평가에서 속성평점을 기반으로 한 모델이 더 정확도가 높은 것을 확인했다.

기준	속성별 평점	총 평점
MAE	0.28	0.46
RMSE	0.42	0.59

〈표 10〉. 평가지표(MAE, RMSE)

호텔이름	평점	추천여부
오크우드 프리미어 코엑스 센터	5	추천
신라스테이(광화문점)	5	추천
롯데시티호텔(마포점)	5	추천
이비스 엠베서더 서울 (명동점)	5	추천
호텔 페이토 (강남점)	5	추천

5. 결론

[학술적 시사점]

- 총 평점 기반 기존 추천모형 보다 속성별 감성값 평점을 이용한 추천 모형의 예측성도가 더 우수함을 확인했다.
- 토픽모델링을 적용한 연구 모형에서 호텔 리뷰의 6가지 속성을 도출하여 감성값 기반 추천 모형을 제안했다.

[실무적 시사점]

- 전자상거래 시장의 성장과 고객의 니즈가 다양해지는 상황에서
다양한 속성을 반영한 개인화 추천시스템은 서비스의 질적 수준을 향상시켜 고객 만족도를 향상시킬 수 있을 것으로 예상된다.

[한계점 및 향후 연구방향]

- 호텔추천을 위해 온라인 호텔 리뷰를 수집하여 연구를 진행했지만, 충분한 데이터를 확보하지 못해 추가적인 데이터 확보가 필요하다.
- 본 연구에서는 MAE, RMSE를 통해 성능평가를 했지만, 향후 연구에서는 다른 지표를 기반으로 성능평가를 할 필요가 있다.
- 상품추천을 위해 머신러닝 기법인 최근접 이웃 협업필터링을 사용했지만,
향후 연구에서는 딥러닝 기반, 하이브리드 기반 등의 새로운 연구방법을 제안할 필요가 있다.



감사합니다