

온라인 리뷰의 텍스트 마이닝을 이용한 통합 감성 기반의 지능형 추천시스템

홍준우(부산대학교 경영학과)

홍태호(부산대학교 경영학과)

2022 경영정보관련 춘계통합학술대회 한국정보시스템학회:
초지능, 초연결, 초실감 시대의 가치창출 전략

목 차

1.서론

2.문헌연구

- 빅데이터 기반 추천시스템

3.연구 프레임워크

4. 실험 및 실험결과

- 1단계: 데이터 수집 및 전처리
- 2단계: 텍스트 마이닝을 통한
상품 속성 및 고객 속성 추출
- 3단계: 감정별 개인화 영화 추천

5. 결론

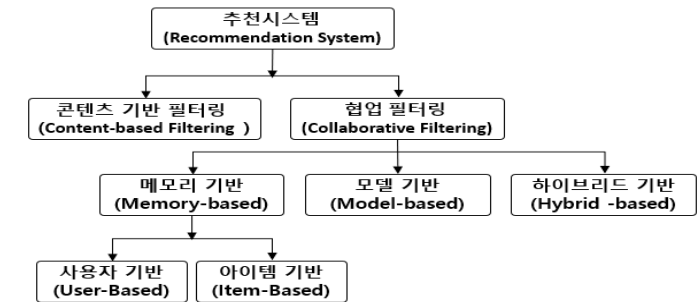
1. 서론

- 디지털 기술이 산업 전반의 전자상거래 시장에 융합되면서 시장의 활성화와 이용률을 증가시켰고, 구매 경험을 기반으로 작성한 온라인 리뷰는 고객의 정보 탐색 및 구매 의사결정에 영향을 미침(Yao et al., 2020).
- 전자상거래 시장의 성장은 고객들의 소비를 불러와 다양하고 방대한 데이터를 생성하고 있으며(Kim, 2020), 최근 코로나19 팬데믹의 영향으로 비대면 유동인구가 증가하면서 가속화되고 있는 추세임(AI maaitah et al., 2021).
- 고객 리뷰 데이터는 고객의 잠재적 니즈를 담고 있어 고객관리 및 확장을 위한 필수적 요소로 사용됨(Anderson & Sullivan., 1993)
 - 고객: 구매 의사결정 과정에서 상품 탐색과 구매 과정 등에 대한 이점(Cheng and Ho., 2015).
 - 기업: 매출 변동의 근거 탐색, 고객 패턴 파악 등에 대한 이점(Li et al., 2013).
- 정보홍수시대 지나친 상품의 다양화는 고객들의 직관적 선택과 효율적인 의사결정을 어렵게 만들고 있으며(김진화 등, 2008), 고객의 선호도에 맞는 상품을 제공해주는 개인화 추천시스템의 필요성이 대두되고 있음(Kim et al., 2010).
- 리뷰 데이터의 상품 속성(영화 속성 등)은 고객의 구매 의사결정에 영향을 미치는데(Vany and Walls., 1996), 기존 총평점을 이용한 추천시스템 연구는 상품의 세부적 선호도를 반영하지 못함(윤호민 & 최규완., 2020).
- 총평점에만 국한된 추천시스템은 정확도가 떨어진다는 한계점이 제기되고 있음(Jeon and Ahn., 2015). 반면 정성적 데이터인 리뷰 데이터는 추천시스템 구축에 중요한 요소로써 역할을 하고 있으며(Park & Chon, 2015), 일부 연구에서는 리뷰 데이터를 직접적으로 추천모형에 적용하여 기존의 총평점 기반 추천모형 보다 높은 정확도를 보이고 있음.
- 본 연구에서는 정성적, 정량적 데이터를 모두 활용한 통합 감성 기반의 지능형 추천모형을 제안함

2. 문헌연구: 빅데이터 기반의 추천시스템

추천시스템(Recommendation)

- 과거 고객 데이터를 기반으로 미래에 고객이 선호할 것으로 예상되는 상품을 추천해주는 서비스이며, 학계뿐만 아니라 산업계에서 마케팅 관점에서 많은 관심을 받고 있음(Kim & Park., 2018).
- 현재 가장 우수한 성능을 보이는 추천알고리즘은 협업필터링 기법이 있으며(Son et al., 2021), 이외에도 머신러닝, 딥러닝, 하이브리드 기반의 다양한 추천알고리즘 연구가 진행되고 있음.



〈그림 1〉. 추천시스템 종류

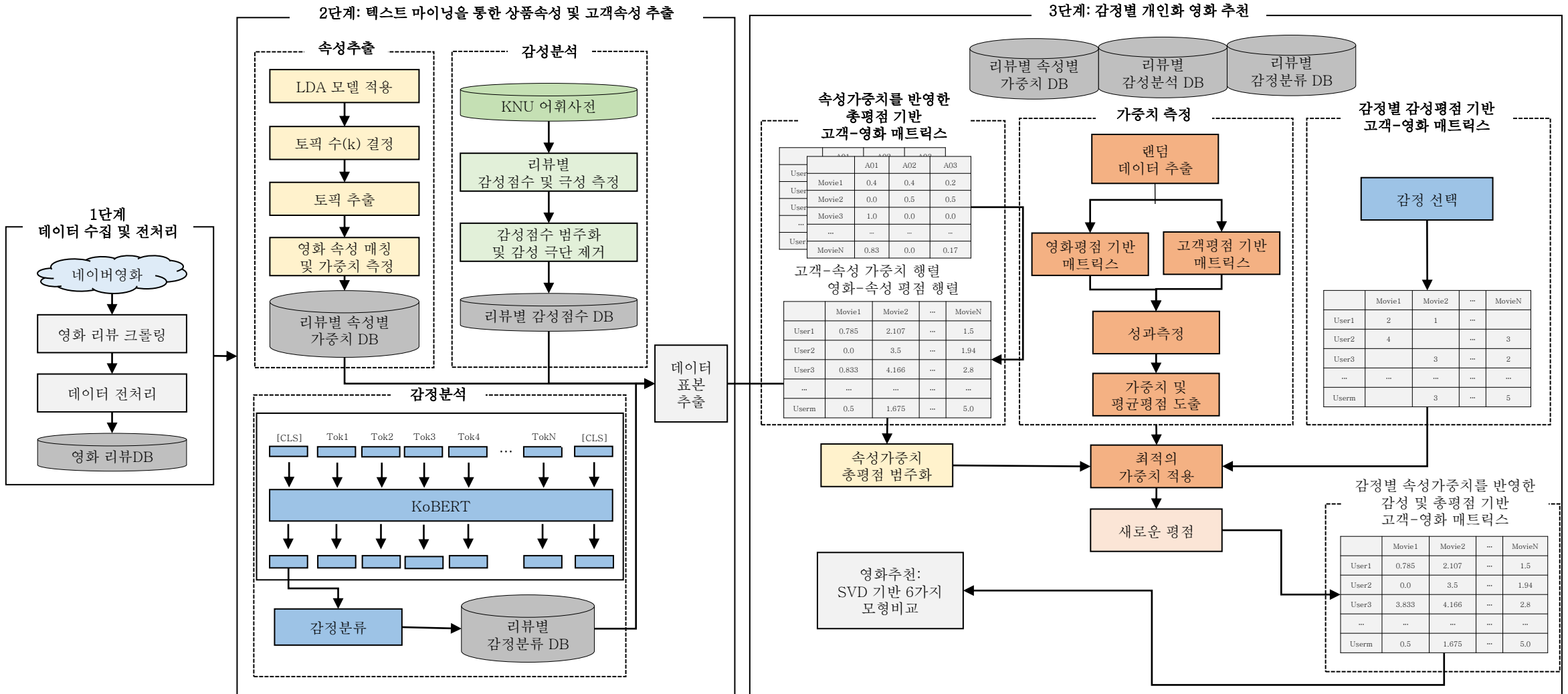
협업필터링(Collaborative Filtering, CF)

- 상품에 대한 선호도가 비슷한 고객들은 다른 아이템에 대한 선호도가 유사할 것이라는 가정을 기반으로 유사도 측정 및 선호도를 예측함.
- CF 기반 추천시스템은 메모리 기반(memory-based)과 모델 기반(model-based)으로 구분할 수 있는데 본 연구에서는 모델 기반 MF(Matrix Factorization) 방식의 대표적인 알고리즘인 SVD(Singular Vector Decomposition)를 사용함.
- SVD는 특이값 분해라고 부르며 행렬을 분해함으로써 암묵적 요인(Latent Factor)을 반영하여 분석의 정확성을 높이는 방법이며(Jeong & Kim., 2017), 적용방법으로 하나의 행렬을 U , Σ , V 으로 분해하고 U , V 을 직각행렬, Σ 을 대각행렬로 정의하여 특이값 분해 알고리즘을 수행함.
- 예시로 고객과 아이템에 대한 $m \times n$ 크기의 행렬 A 를 특이값 분해할 경우, $A = U\Sigma V^T$ 와 같이 나타낼 수 있으며 이때, $U_{m \times m}$ 은 고객 행렬, $V_{n \times n}^T$ 는 아이템 행렬을 의미하고, $\Sigma_{m \times n}$ 은 특이값에 대한 대각 항을 갖는 대각 행렬을 나타냄.

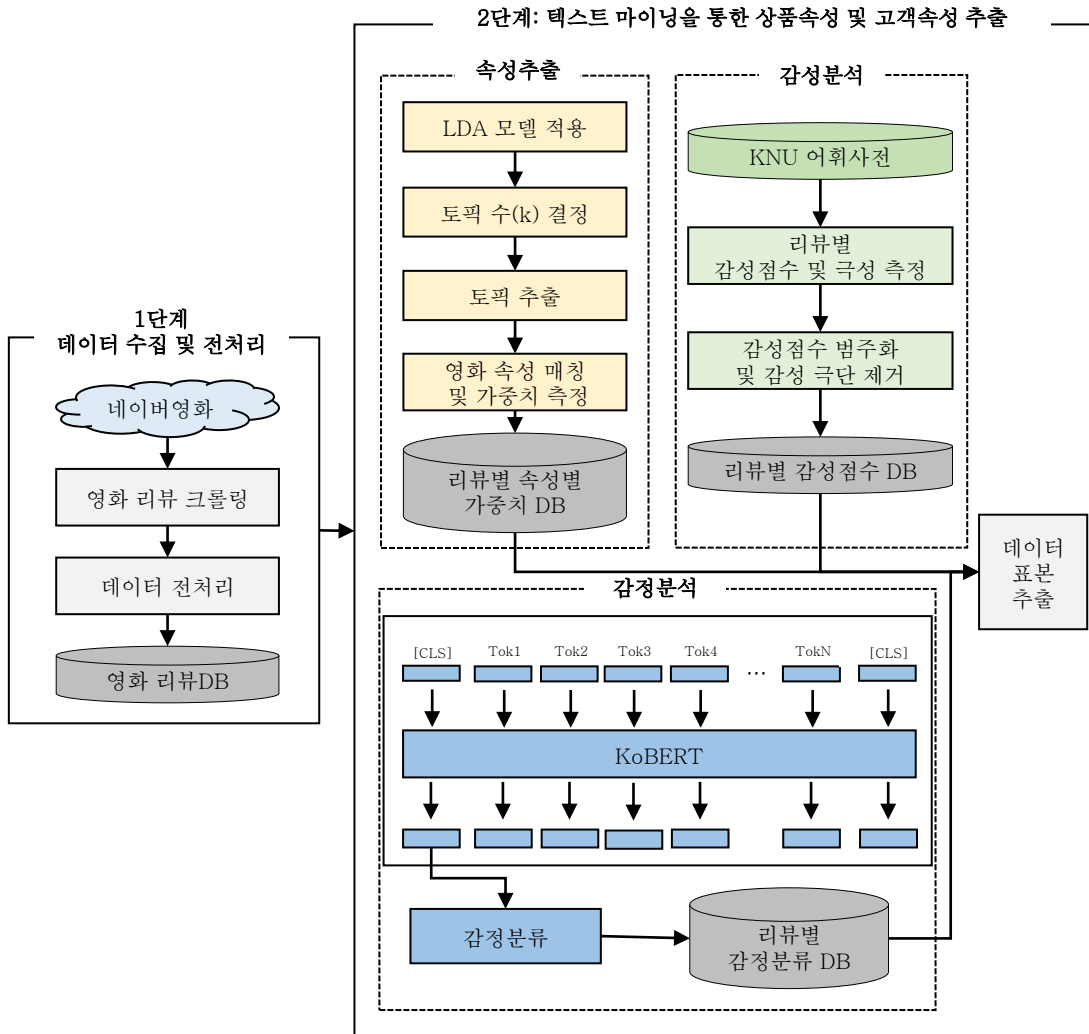
$$A = U \Sigma V^T$$

〈그림 2〉. SVD와 Latent Factor 모형

3. 연구 프레임워크



3. 연구 프레임워크: 1단계 & 2단계



[1단계: 데이터 수집 및 전처리]

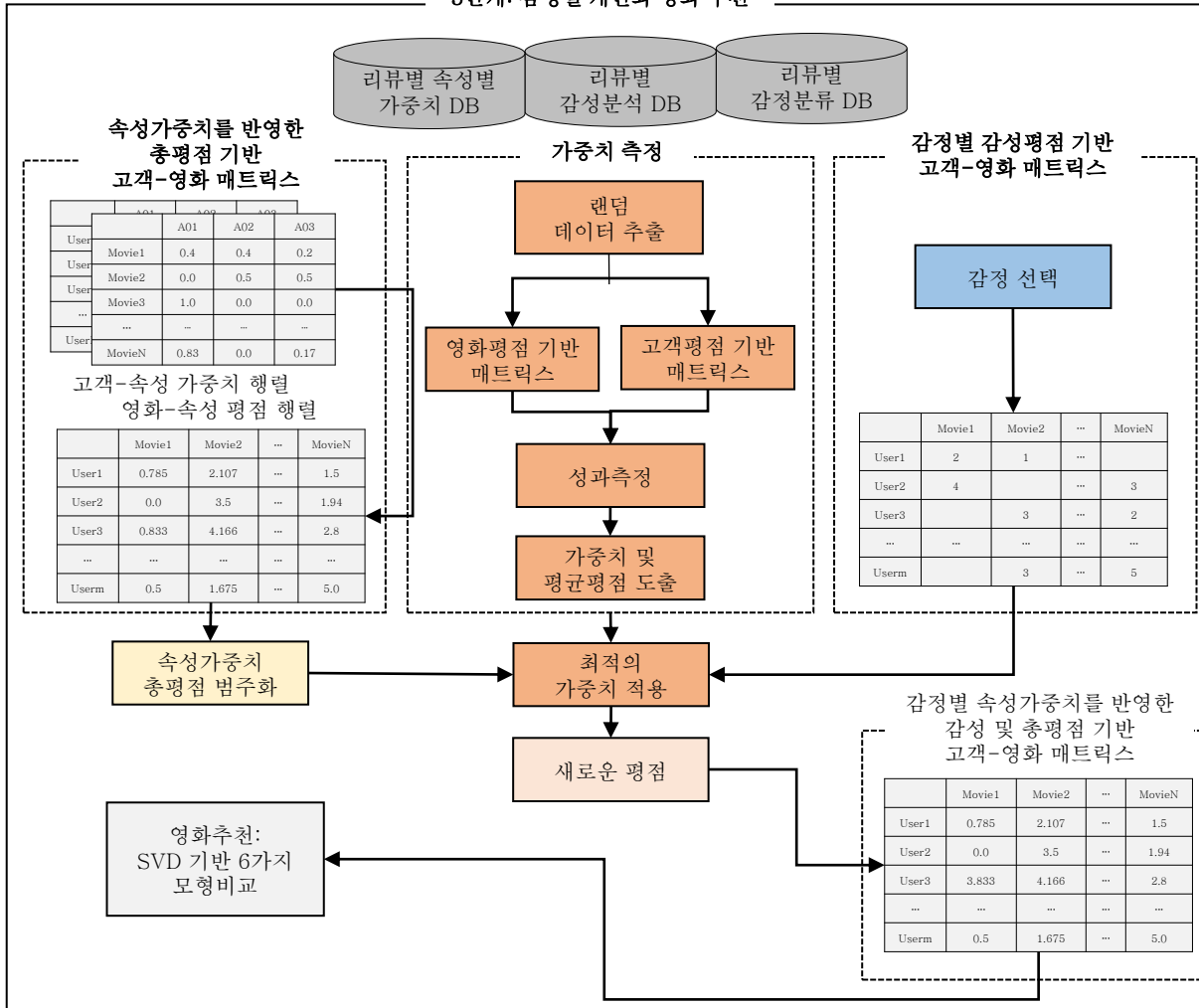
- ① 데이터 수집: “네이버영화” 에서 제공하는 영화 리뷰 데이터 수집
- ② 데이터 전처리: 불필요한 리뷰 데이터 제거 및 한국어 특성에 맞게 변환

[2단계: 텍스트 마이닝을 통한 상품속성 및 고객속성 추출]

- ① 속성추출: 토픽모델링의 LDA를 통해 추출한 토픽을 3가지의 영화 속성에 매칭 및 대표단어 정의
영화 속성 및 대표단어를 기반으로 속성별 가중치 측정
- ② 감성분석: 어휘사전을 기반으로 감성점수 및 감성극성 측정
군집분석을 통한 범주화
- ③ 감성분석: BERT 기반 모델을 활용하여 7가지 감정으로 리뷰 데이터를 분류함.
- ④ 데이터 표본 추출: 데이터 희소성 검토, 감성점수의 극단치 제거

3. 연구 프레임워크: 3단계

3단계: 감정별 개인화 영화 추천



[3단계: 감정별 개인화 영화 추천]

■ 감정별 속성가중치를 반영한 감성 및 총평점 기반 고객-영화 매트릭스 생성:

① 속성가중치 반영한 총평점 기반 고객-영화 매트릭스:

■ 고객-속성 가중치 행렬 & 영화-속성 총평점 행렬:

두 개의 행렬을 전치행렬 후 연산을 하여 하나의 매트릭스를 생성

■ 범주화: 군집분석을 통해 5개로 범주화

② 감정별 감성평점 기반 고객-영화 매트릭스

③ 가중치 측정: 0.01 단위로 고객평점/영화평점 두 가지 경우에 대한 RMSE 수치를 측정함.

④ 새로운 평점: ①, ② 매트릭스에 최적의 가중치를 적용하여 하나의 평점을 생성

감정별 속성가중치를 반영한 감성 및 총평점 기반 고객-영화 매트릭스 구축

⑤ 감정별 개인화 영화 추천

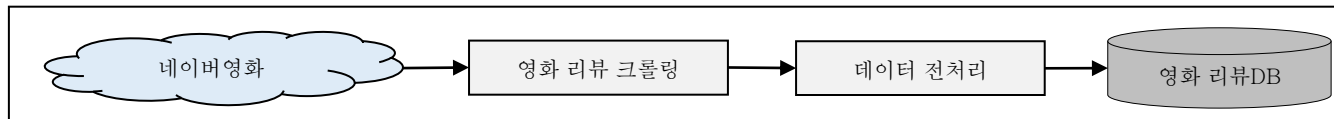
4. 실험 및 실험결과: 1단계 데이터 수집

■ 데이터 수집 및 전처리:

① 데이터 수집:

■ 네이버(NAVER):

- 국내 검색포털 시장에서 76.7%의 점유율을 갖고 있으며 PC뿐만 아니라 모바일 서비스 측면에서도 높은 우위를 선점하고 있음(심형섭 외, 2015).
- 2022년 4월 이용자는 27,103,788명으로 국내 최고 포털사이트로 인정받고 있음(Nielsen 2022년 4월).
- “네이버영화” 페이지에서 웹 크롤링을 통해 데이터(리뷰번호, 고객ID, 영화ID, 총평점, 리뷰 작성일자, 고객 리뷰) 수집



<그림 3>. 1단계: 데이터 수집 및 전처리

수집 기간	2022.02.14 ~ 2022.05.01(77일)				
리뷰 수	82,501				
영화 수	6,007				
고객 수	29,419				
총평점(1~10)	1(8,377)	2(5,514)	3(915)	4(3,447)	5(2,318)
	6(6,457)	7(5,546)	8(10,293)	9(5,181)	10(34,453)
총평점(1~5)	1(13,891)	2(4,362)	3(8,775)	4(15,839)	5(39,634)

<표 2>. 네이버영화: 영화 리뷰 데이터 수

18057042	닥터 스트레인지: 대혼돈의 멀티버스 ★★★★★ 1	lzs1****
	마블이 아니고 그냥 정성스럽게 만든 공포영화 수준. — 중간에 나가고싶었음. 신고	22.05.06
18057041	서유기 2 - 선리기연 ★★★★★ 10	homm****
	왜 인터스텔라 인셉션을 최고라고했을까 나 울것같아 이걸이제알았네 신고	22.05.06
① 18057040	③ 닥터 스트레인지: 대혼돈의 멀티버스 ★★★★★ 8 ④	youl**** ②
	기존 마블 히어로물이나 액션 영화같지 않고 호러영화에 가까웠던듯. 그리고 내용 전개도 기존 마블영화처럼 빠르게 전환되지 않고 초~중반엔 좀 질질끄는 느낌이 있는데다가 중간에 완다가 귀신. 살인마처럼 뒤에서 확 놀래켜줘서 지루함이 좀 덜한 느낌..? 그렇지만 보는 사람에 따라 약간 불쾌할수도 있을 것 같음. 완다라는 캐릭터를 모른다면 내용을 이해하는데 약간 어려움을 느낄수도 있을 것 같음. 그래도 마블영화답게 실망시키지 않는 듯. 후반부르 갈수록 내용이 다시 기존 마블영화물로 돌아오는 느낌이었음. CG 화려하고, 베네딕트의 연기는 우리를 실망시키지 않음. 연휴에 나름대로 재밌게 즐길 수 있는 영화인듯. 신고	22.05.06 ⑤

<그림 4>. “네이버영화” 웹 페이지의 고객 평점 리뷰 데이터 샘플

4. 실험 및 실험결과: 1단계 데이터 전처리

② 데이터 전처리:

- 총평점 범주화: 1부터 10사이의 총평점을 5개로 범주화
- 텍스트 제거: 웹사이트 링크, 특수문자, 이모티콘, 한국어 불용어, 공백, 숫자 및 영어
 - ※ 한국어 불용어: URL, 웹사이트, 텍스트 및 문서 등"에 대한 키워드 분석을 다루는 RANK NL에서 제공함.
- 텍스트 변환: 띄어쓰기, 맞춤법 검사기, 반복문자(이모티콘, 자음, 모음), 외래어

(1). 고객 리뷰 데이터: 원본

'넷플릭스에 있는 콘텐츠 하도 많이 봐서 볼게 없어서 ππππππππππ 중국영화 오랜만에 봐야지 했는데 정말 재밌어 보여서 보게 됐습니다 보는 내내 1. 지루할 틈이 없고 2. 스토리도 괜찮고 3. 액션신 또한 나이스라서 삼박자 모두 만족합니다 좋은 (Good) 영화 잘 보고 힐링 잘했네요 😊 좋은영화 추천합니다 신고

(2). 고객 리뷰 데이터: 리뷰어 제거

'넷플릭스에 있는 콘텐츠 하도 많이 봐서 볼게 없어서 ππππππππππ 중국영화 오랜만에 봐야지 했는데 정말 재밌어 보여서 보게 됐습니다 보는 내내 지루할 틈이 없고 스토리도 괜찮고 액션신 또한 나이스라서 삼박자 모두 만족합니다 좋은영화 잘 보고 힐링 잘했네요 좋은영화 추천합니다

(3). 고객 리뷰 데이터: 리뷰어 변환

넷플릭스에 있는 콘텐츠 하도 많이 봐서 볼 게 없어서 ππ 중국영화 오랜만에 봐야지 했는데 정말 재밌어 보여서 보게 됐습니다 보는 내내 지루할 틈이 없고 스토리도 괜찮고 액션 신 또한 나이스라서 삼박자 모두 만족합니다 좋은 영화 잘 보고 힐링 잘했네요 좋은 영화 추천합니다

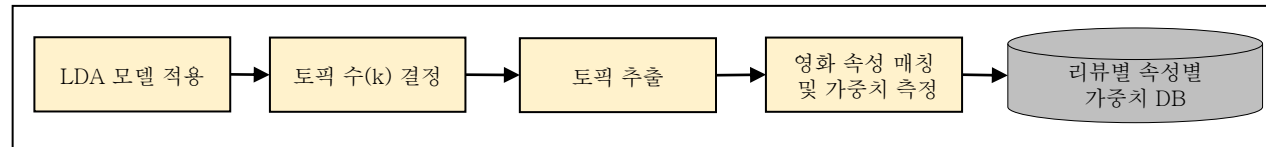
〈그림 5〉. 데이터 전처리: 텍스트 제거 및 변환 예제

4. 실험 및 실험결과: 2단계 속성추출

■ 텍스트 마이닝을 통한 상품속성 및 고객 속성 추출 :

① 속성추출:

- 토픽모델링의 LDA를 통해 토픽 수를 추출하는데 5겹 교차 검증(5-fold cross validation)과 혼잡도(Perplexity) 및 일관성(coherence) 성과지표를 통해 최적의 토픽 수(K) 50을 측정함.
- 온라인 영화 리뷰 데이터를 핵심적, 주변적, 정서적 속성으로 분류했으며(Neelamegham and Jain., 1999), 50개의 토픽을 추출하여 속성별 매칭 및 대표단어를 정의함.
- 영화 속성 및 속성별 대표단어”를 기반으로 리뷰별 3가지 영화 속성에 대한 각각의 가중치 측정.



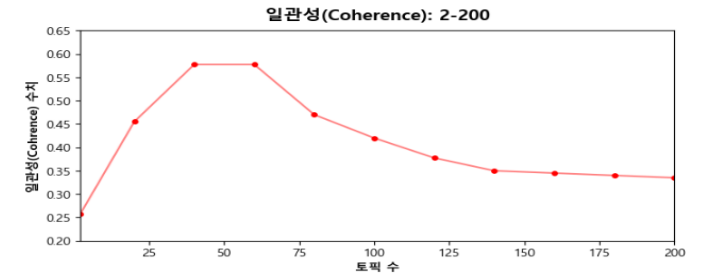
〈그림 6〉. 2단계: 속성추출: LDA(Latent Dirichlet Allocation)

구분	속성	키워드
영화	핵심적	한국영화, 중국, 일본, 한국, 액션, 애니메이션, 코믹, 연기, 리메이크, 시대, 후속, 스토리, 시리즈, 사랑이야기, 스토리, 마블, 아저씨, 여배우, 조폭, 연기, 작품, 멜로, 배우, 원작, 스토리, 아이, 여자, 스파이, 이야기, 스릴러, 캐릭터, 배우
	주변적	사운드, 러닝타임, 작품, 제작, 주제, 시간, 스케일, 시리즈, 연출, 자막, 시나리오, 관계, 주제, 상황, 메시지, 엔딩, 반전, 실화, 장면, 의도, 작가, 영상미, '노래', 분위기, 사람, 관객, 소재, 영상미, 포스터
	정서적	재미, 순수, 감동, 최고, 매력, 사랑, 리뷰, 느낌, 실망, 선동, 완벽, 낭비, 현실적

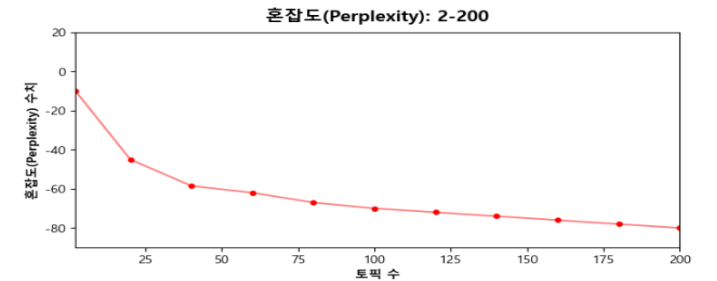
〈표 3〉. 영화 속성 및 속성별 대표단어

영화 속성	내용
핵심적	영화 내용과 직접적으로 연관된 키워드와 관련된 속성 ex. 배우, 스토리, 장르 등
주변적	영화의 전반적 흐름과 관련된 속성. ex. 관심지표, 배경, 연출, 음악, 특수효과 등
정서적	영화 시청에 대해 고객이 느끼는 감정과 관련된 속성 ex. 감동, 슬픔, 재미 등

〈그림 4〉. 토픽 수(k)에 따른 일관성(coherence)



〈그림 7〉. 토픽 수(k)에 따른 일관성(coherence)

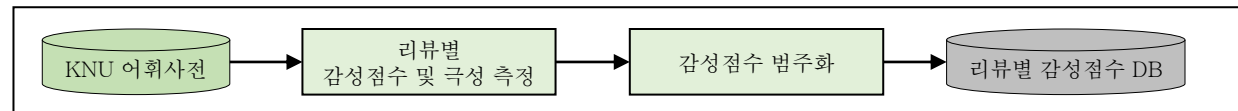


〈그림 8〉. 토픽 수(k)에 따른 혼잡도(Perplexity)

4. 실험 및 실험결과: 2단계 감성분석

② 감성분석:

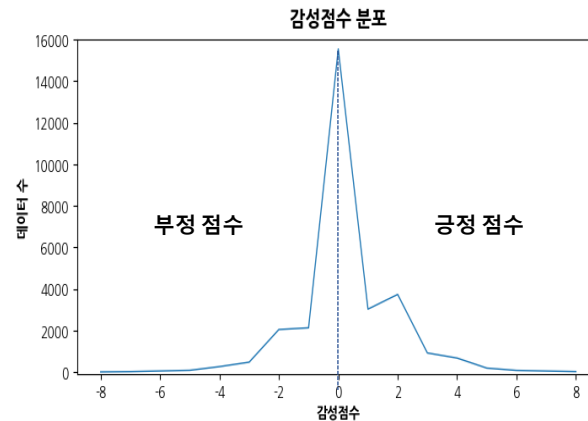
- 감성점수 및 감성극성 추출: 한국어 기반의 KNU감성 사전 사용했으며 명사, 형용사, 동사를 추출하여 측정
- 감성점수 범주화:
 - 초기 감성점수: 감성점수는 -12부터 +15로 분포하며, 시각화를 통해 감정점수별 10개 미만의 데이터 수를 갖는 양극단치 38개 제거
 - 감성점수 범주화: 감성점수는 -8부터 +8로 조정했으며 군집분석을 통해 5개로 범주화



<그림 9>. 2단계: 감성분석(Sentiment Analysis)

-12	1
-11	3
-10	6
-9	7
-8	13
-7	26
-6	58
-5	89
-4	266
-3	480
-2	2051
-1	2133
0	15549
1	3032
2	3740
3	923
4	679
5	195
6	84
7	56
8	28
9	9
10	4
11	3
12	3
13	1
15	1

<그림 10>. 감정점수: 감성점수의 양극단 값 제거 전/후



감성평점	감성점수 분포
1	-8 ~ -2
2	-1 ~ 0
3	1
4	2 ~ 3
5	4 ~ 8

<표 5>. 감성점수 범주화: 감성점수 분포

4. 실험 및 실험결과: 2단계 감정분석

③ 감정분석:

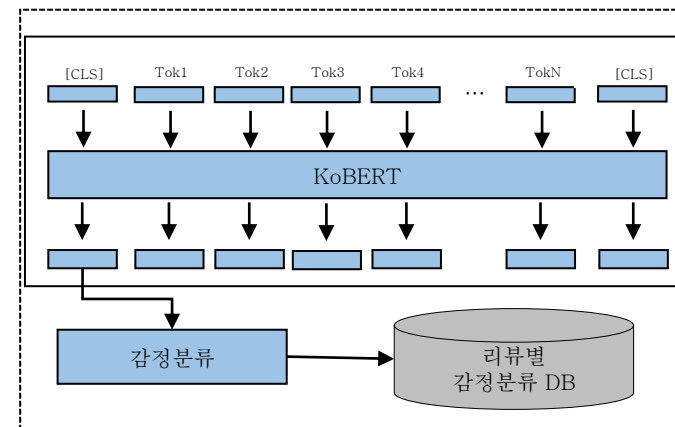
- BERT 종류의 기계번역 모델로써 한국어 성능의 한계를 극복하고자 SKT Brain에서 개발함.
이는 기존 BERT 모델에서 한국어 데이터를 추가로 학습시켜 한국어 데이터에 대해 높은 정확도를 낼 수 있는 모델임
※ 한국어 위키에서 5백만개의 문장과 54백만개의 단어를 학습시킨 모델
- KoBERT를 통한 다중 감정분류: 7가지(행복, 놀람, 중립, 공포, 분노, 슬픔, 혐오)
- 학습 데이터: AI Hub에서 제공하는 “한국어 단발성 대화 데이터”

총평점	행복(6,980)	놀람(997)	중립(3784)		
	혐오(2,094)	슬픔(1,389)	분노(423)	공포(373)	

〈표 6〉. 감정분류 데이터 수

구분	SNS 글 및 온라인 댓글
성격	7개 감정: 기쁨, 슬픔, 놀람, 분노, 공포, 혐오, 중립
출처	AI Hub, 한국어 단발성 대화 데이터
특징	기존 한국어 텍스트 데이터는 이진분류(감성) 수준을 크게 벗어나지 못하지만 본 데이터는 다중분류(감정)에 대한 데이터 특성을 가짐
데이터 수	38,594

〈표 7〉. KoBERT 학습 데이터



〈그림 11〉. 2단계: 감정분석(Emotion Analysis)

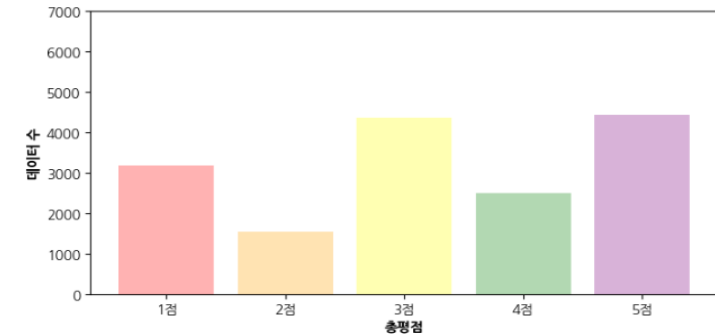
4. 실험 및 실험결과: 2단계 데이터 표본 추출

④ 데이터 표본 추출

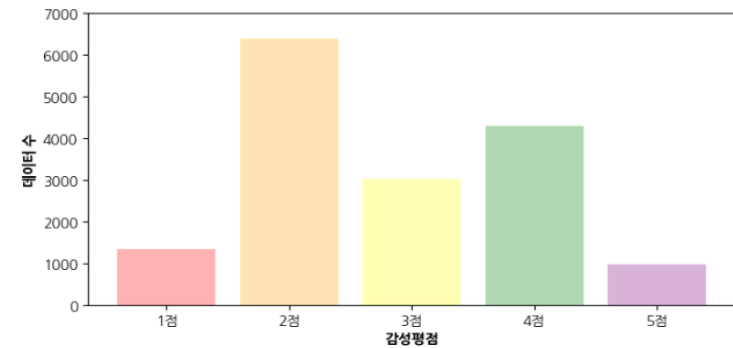
- 데이터 필터링: 유저별 10개 이상 평점 및 리뷰를 작성한 대상자를 선별함
- 극단치 제거: 총평점과 감성평점의 극단치 13,362개 제거
- 3,343편의 영화에 대해 1,407명의 고객이 작성한 16,040개의 데이터를 활용함.

구분	데이터 필터링 및 극단치 제거	
	이전	이후
영화	6,007	3,343
고객	29,419	1,407
데이터 수	82,501	16,040

〈표 8〉. 영화 리뷰 데이터 수: 데이터 필터링



〈그림 12〉. 총평점별 데이터 수



〈그림 13〉. 감성평점별 데이터 수

4. 실험 및 실험결과: 3단계 감정별 영화 추천

■ 감정별 개인화 영화 추천 :

① 속성가중치를 반영한 총평점 기반 고객-영화 매트릭스: - <표11>

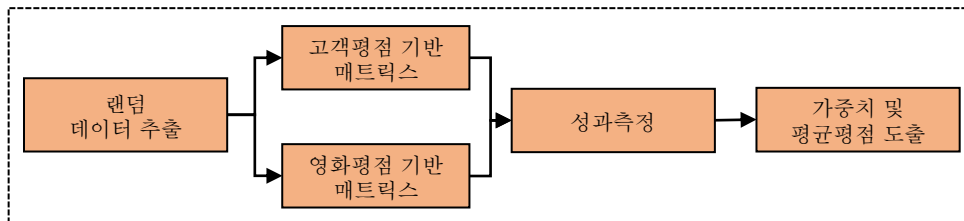
- 행렬01: 고객-속성 가중치 행렬: 고객별 속성 가중치 합이 10이 되도록 행렬 구성 - <표9>
- 행렬02: 영화-속성 총평점 행렬: 영화별 속성 가중치 합이 영화의 총평점이 되도록 행렬 구성 - <표10>
- 행렬01과 행렬02을 전치행렬 후 연산하여 “속성가중치를 반영한 총평점 기반 고객-영화 매트릭스 생성”

② 감정별 감성평점 기반 고객-영화 매트릭스: - <표12>

- 중립 감정을 제외한 6가지 감정별 “감성평점 기반 고객-영화 매트릭스 생성” ※ 6가지 감정: 행복, 놀람, 공포, 분노, 슬픔, 혐오

③ 가중치 측정:

- 두 가지 매트릭스 평점을 하나의 평점으로 만드는 과정으로 평점별 최적의 가중치 지점을 탐색.
- 전체 데이터에서 랜덤으로 70% 데이터를 추출하여 “속성가중치 총평점”과 “감정별 감성평점” 변수를 불러와 가중치 합이 10이 되도록 0.01단위 차등적으로 변수에 적용하여 고객/영화 평점 기반의 RMSE를 99회 측정함.
- 최적의 가중치 지점으로 영화평점 기반의 속성가중치 총평점 0.01, 감정별 감성평점 0.99를 확인함.



<그림 14>. 3단계: 가중치 측정

	Movie 1	Movie 2	...	Movie N
User 1	2	1	...	NaN
User 2	4	NaN	...	3
User 3	NaN	3	...	2
...
User M	NaN	3	...	5

<표 12>. 감정별 감성평점 기반
고객-영화 매트릭스

	핵심적 속성	주변적 속성	감성적 속성
User 1	0.75	0.00	0.25
User 2	0.52	0.08	0.40
User 3	0.62	0.00	0.38
...
User M	0.32	0.04	0.64

<표 9>. 행렬01: 고객-속성 가중치 행렬

	핵심적 속성	주변적 속성	감성적 속성
Movie 1	2.50	0.00	2.50
Movie 2	2.26	0.10	1.72
Movie 3	3.35	0.00	1.65
...
Movie N	0.50	0.00	0.50

<표 10>. 행렬02: 영화-속성 총평점 행렬

	Movie 1	Movie 2	...	Movie N
User 1	0.79	2.11	...	1.50
User 2	0.00	3.50	...	1.94
User 3	0.83	4.17	...	2.80
...
User M	0.50	1.68	...	5.00

<표 11>. 속성가중치를 반영한 총평점 기반
고객-영화 매트릭스

4. 실험 및 실험결과: 성과측정

연구모형별 성능비교:

1). 비교모형별 성능지표

평가 지표 구분		MAE	MAPE	RMSE	대체 결측치 잠재요인	결측치	가중치
비교 모형	StarR_SVD	1.076	53.46%	1.208	영화, 90	99.71%	
	SentiR_SVD	0.951	42.75%	1.050	영화, 10	99.71%	
	SentiStarR_SVD	0.859	36.56%	0.962	영화, 10	99.71%	총평점(0.12) 감성평점(0.88)
	AWStarR_SVD	0.211	11.11%	0.271	영화, 20	40.35%	
	AWSentiStarR_SVD	0.124	9.06%	0.2094	영화, 20	40.01%	감성평점(0.97) 속성가중치평점(0.03)
	E_AWSentiStarR_SVD	0.020	1.33%	0.074	영화, 62	40.00%	감성평점(0.99) 속성가중치평점(0.01)

- StarR_SVD: 총평점
- SentiR_SVD: 감성평점
- SentiStarR_SVD: 감성 및 총평점
- AWStarR_SVD: 속성가중치 총평점
- AWSentiStarR_SVD: 속성가중치 감성 및 총평점
- E_AWSentiStarR_SVD: 감정별 속성가중치 감성 및 총평점(★)

- 제안 연구 모형(E_AWSentiStarR_SVD)은 RMSE가 0.074으로 가장 우수한 성과를 증명했으며 MAE, MAPE 또한 우수한 성과를 증명함

- 속성가중치를 반영한 추천모형(AWStarR_SVD, AWSentiStarR_SVD, E_AWSentiStarR_SVD)은 결측치가 약 40%로 99%의 결측치를 갖는 다른 추천모형에 비해 낮음을 확인하여 데이터 희소성 문제가 개선함.

2). 6가지 감정별 세부 성능지표

평가 지표 구분		MAE	MAPE	RMSE	대체 결측치 잠재요인	결측치	가중치
감정분류 (중립제외)	행복(6,980개)	0.018	1.29%	0.093	영화, 90	39.64%	감성평점(0.99) 속성가중치평점(0.01)
	놀람(997개)	0.019	1.31%	0.069	영화, 10	39.98%	
	슬픔(1,389개)	0.015	0.90%	0.065	영화, 110	41.23%	
	혐오(2,094개)	0.013	0.90%	0.060	영화, 70	39.62%	
	분노(423개)	0.027	1.81%	0.076	영화, 50	38.12%	
	공포(373개)	0.028	1.77%	0.079	영화, 40	41.18%	
평균 성과		0.020	1.33%	0.074	영화, 61.67	40.00%	

5. 결론

▪ [학문적 기여도]

- ① 메모리 기반 추천시스템은 데이터 희소성 문제를 갖고 있어
본 연구에서는 모델 기반 추천시스템인 SVD 알고리즘을 사용하여 정량적 데이터와 정성적 데이터를 결합하여 데이터 희소성 문제를 해결함.
- ② 두 가지 평점 데이터를 기반으로 하나의 평점 데이터를 생성하기 위한 최적의 가중치 지점을 측정하는 프로세스를 구축함.
- ③ 온라인 리뷰가 갖는 정량적 데이터인 총평점과 정성적 데이터인 영화속성, 감성점수, 감정결과를 활용하여 하나의 통합 추천모형을 제안함.

▪ [실무적 기여도]

- ① 기존 정량적 데이터인 총평점 기반의 추천연구에서
정성적 데이터인 온라인 리뷰 데이터에 텍스트 마이닝의 토픽모델링, 감성 및 감정분석을 적용한 방법론을 제시함.
- ② 다양한 자료원천을 대상으로 추천시스템을 개발할 수 있는 방법론을 제시함.
- ③ 한국어 텍스트 마이닝의 자연어처리와 다중 감정분류를 통해 정성적 데이터를 활용한 추천시스템 연구의 발전에 기여함.

▪ [한계점 및 향후 연구방향]

- ① 본 연구에서 77일 동안 82,501개 리뷰 데이터를 수집하여 표본 추출과정을 통해 16,040개 리뷰 데이터를 추출하여 활용함.
향후 연구에서는 최소 한 분기(4개월) 이상의 기간동안 데이터를 수집하여 더 많은 리뷰 데이터를 활용할 필요가 있음.
- ② 어휘 기반의 감성분석을 명사, 형용사, 동사만을 기반으로 감성점수 및 극성을 측정함.
향후 연구에서는 대명사, 부사 등 다양한 품사를 활용한 감성분석을 진행할 필요가 있음.
- ③ 모델기반 추천알고리즘인 SVD를 활용하여 추천모형을 구축 및 제안함.
향후 연구에서는 최근 활발하게 연구되는 딥러닝 기반의 추천알고리즘을 적용할 필요가 있음.



감사합니다