

# 영화 리뷰의 상품 속성과 고객 속성을 통합한 지능형 추천시스템

홍태호

부산대학교 경영학과  
(hongth@pusan.ac.kr)

홍준우

부산대학교 경영학과  
(ehddml70@naver.com)

김은미

국민대학교 정보기술연구소  
(keunmi100@gmail.com)

김민수

부산대학교 경영학과  
(mskim\_growth@pusan.ac.kr)

디지털 기술이 산업 전반의 전자상거래 시장에 융합되면서 온라인 거래의 활성화와 이용률을 증가시켰으며, 이러한 시장의 흐름은 최근 코로나와 같은 감염병이 확산함에 따라 더욱 가속화되어 다양한 상품 정보를 온라인을 통해 고객들에게 제공할 수 있게 되었다. 다양한 정보의 제공은 고객들에게 다양한 선택의 기회를 제공하지만 의사결정에 어려움을 주기도 한다. 추천시스템은 고객의 의사결정에 도움을 줄 수 있으나 기존 추천시스템 연구는 정량적 데이터만에 국한되어 있으며, 상품 및 고객의 세부적인 요인을 반영하지 못하였다. 이에 본 연구에서는 온라인 리뷰를 기반으로 정성적 데이터를 텍스트 마이닝 기법을 적용하여 상품 및 고객의 속성을 정량화하고 기존의 객관적 지표인 총평점과 감성 및 감정을 통합한 지능형 추천시스템을 제안한다. 제안된 지능형 추천모형은 총평점 위주의 추천 모형보다 우수한 추천성적을 보여주었으며, 상품 및 고객의 세부적 요소를 반영한 추천결과를 통해 새로운 비즈니스 가치를 창출할 것으로 기대한다.

**주제어** : 추천시스템, 협업필터링, SVD, 텍스트마이닝, 영화 리뷰

논문접수일 : 2022년 5월 27일

논문수정일 : 2022년 6월 16일

게재확정일 : 2022년 6월 17일

원고유형 : Regular Track

교신저자 : 김은미

## 1. 서론

정보통신기술의 발전과 디지털 기기의 보급 및 대중화로 전 산업에서 디지털 대전환이 진행되면서, 고객의 경험을 기반으로 작성된 온라인 리뷰에 대해 빅데이터 분석을 통한 연구들이 진행되어 왔다. 온라인 리뷰는 고객의 잠재적 니즈를 담고 있어 고객관리 및 확장을 위한 필수적 요소로 사용되며(Anderson & Sullivan, 1993; Yao et al., 2020), 고객의 의견을 쉽게 파악할 수 있다는 이점을 갖는다. 고객은 온라인 리뷰를 통해 상품탐색 및 구매과정에 대한 정보와 지식에 접

근할 수 있으며(Cheng & Ho, 2015), 기업은 매출 증가 및 하락에 대한 원인 분석과 고객의 행동 패턴 파악 등을 리뷰 분석을 통해서 할 수 있다(Li et al., 2013). 특히 영화나 관광과 같은 경험재에서는 온라인 리뷰에서 제공하는 정보가 잠재고객들의 의사결정에 더욱 중요한 영향을 미친다. 이처럼 유용한 온라인 리뷰는 전자상거래 시장의 급격한 성장과 함께 방대하고 다양해지고 있으며, 최근 코로나19 팬데믹의 영향으로 비대면 활동이 급격히 증가하면서 데이터의 증가 및 다양화를 가속화시키고 있다(AI-maaitah et al., 2021). 전자상거래 시장을 이용하는 고객들

\* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A5A2A03055790)

이 증가함에 따라 고객들의 니즈 또한 새로운 생활방식에 맞추어 다양하게 진화하였다. 기업은 온라인 리뷰 정보를 이용하여 고객들의 니즈에 맞는 다양한 상품정보를 시장에 제공하였고, 고객은 개인의 선호도에 맞는 다양한 상품선택의 기회를 얻게 되었다. 하지만 지나친 상품의 다양화는 일부 고객의 직관적 선택과 의사결정을 어렵게 만들고 있으며(김진화 등, 2008), 기존 추천시스템에서 상품 속성 및 고객 속성에 대한 선호도를 반영하지 못하고 있어(윤호민, 최규완, 2020), 상품 및 고객의 특성을 고려한 개인화 추천시스템의 필요성이 제기되고 있다. 추천시스템 연구는 정량적 데이터를 적용한 연구가 진행되어 왔으며, 최근에는 정성적 데이터인 리뷰 데이터를 감성분석 및 가중치 측정을 통해 정량화하여 총평점 기반의 고객-상품 매트릭스와 결합한 추천모형을 제안하는 연구들이 발표되었다(Yun & Yoon, 2020). 정성적 데이터를 활용한 추천모형은 기존의 총평점 기반의 모형보다 우수한 정확성을 증명했다. 현지연 등(2019)은 감성분석을 통해 리뷰 데이터를 정량화하고 객관적 지표인 총평점과 결합하여 하나의 통합 평점을 생성하여 추천모형에 적용했으며, 객관적 지표인 총평점 기반의 단일 추천모형보다 우수한 성과를 증명했다. Lee et al. (2019)은 리뷰 데이터를 토픽모델링의 LDA를 통해 6가지 상품 속성과 대표단어를 분류한 후에, 고객의 세부속성을 반영한 추천모형으로 구현했다. 김민정 등(2021)은 리뷰 데이터를 텍스트 마이닝 기법인 워드클라우드, N-gram, TF-IDF, CONCOR 분석을 통해 다양한 도메인에 맞는 추천 결과 정보를 추출했다. 이처럼 정성적 데이터인 리뷰 데이터를 정량화하여 직접적으로 추천모형에 적용하면서 세부적 특성을 고려한 추천시스템을 구현했으며, 기

존의 총평점 기반 추천모형보다 우수한 성과를 증명했다. 이러한 연구결과는 최근 총평점에만 국한된 추천시스템이 정확도가 떨어진다는 한계점을 극복했다(전병국, 안현철, 2015). 따라서, 본 연구에서는 영화 리뷰 데이터를 대상으로 LDA를 이용하여 상품 속성과 감성 및 감정분석에 기반한 고객 속성을 정량적 지표인 총평점 데이터와 통합한 지능형 추천시스템을 제안한다. 영화 데이터는 추천시스템 연구에 많이 활용되어 왔으나 영화 리뷰 데이터의 활용은 미비하다. 영화 리뷰는 영화선택에 중요한 영향을 미치며 영화의 객관적인 정보와 더불어 관람후의 감정을 서로 공유할 수 있다. 제안된 방법론은 정성적인 데이터를 활용하여 온라인 시장을 이용하는 고객들에게 상품의 속성과 고객의 속성을 고려한 추천결과를 제시할 수 있으며, 동시에 고객만족도를 높임으로써 고객관계관리를 더욱 강화할 수 있을 것으로 기대한다.

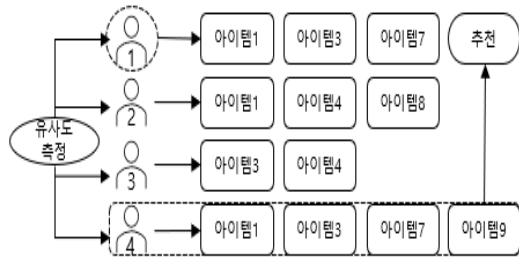
## 2. 문헌연구

### 2.1 빅데이터 기반 추천알고리즘

추천시스템은 과거의 고객 데이터를 기반으로 고객이 관심있을 것으로 예상되는 상품을 추천하는 기법이다. 협업필터링(Collaborative Filtering)은 Goldberg et al. (1992)에 의해 처음 소개되었으며 추천 기법 중 가장 우수한 성능을 보이고 있다(Son et al., 2015; Sarwar et al., 2000). 아이템에 대한 선호도가 비슷한 고객들은 다른 아이템에 대한 선호도 또한 유사할 것이라는 가정을 기반으로 고객 혹은 아이템에 대한 유사도를 측정하고 선호도를 예측하며, 메모리 기반과 모델 기

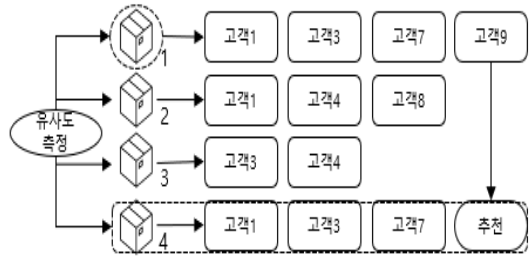
반 방식으로 구분할 수 있다(Breese et al., 1998; Son et al., 2015).

메모리 기반 방식은 이웃모형이라고도 하며, 매트릭스를 구축하고 매트릭스의 고객 혹은 아이템을 이용하여 유사한 고객 또는 아이템을 찾아 예상 평점을 추론하는 방식으로 이루어지며, 고객 기반 협업필터링과 아이템 기반 협업필터링으로 구분한다. 고객 기반 협업필터링은 <그림 1>과 같이 고객이 남긴 선호도 정보를 사용하며, 아이템 1, 3, 7을 선호하는 고객 1과 유사도가 가장 높은 고객 4를 기반으로 고객 4가 선호하는 아이템 중에 고객 1이 사용하지 않은 상품을 고객 1에게 추천한다. 음악, 친구 등의 추천에 많이 활용되고 있으며 페이스북, 인스타그램 등에서 사용되고 있다.



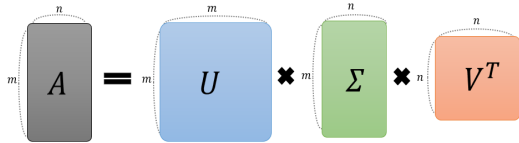
<그림 1> 고객 기반 협업필터링

아이템 기반 협업필터링은 Sarwar et al. (2001)에 의해 처음 소개되었으며 유튜브, 넷플릭스와 같은 영상 플랫폼에서 콘텐츠 추천 등에 사용된다. <그림 2>와 같이 상품목록에서 추천 대상인 상품 1은 고객 1, 3, 7, 9가 선호하는 상품이며, 유사도가 가장 높은 상품 4를 선택한다. 이후, 상품 4를 사용하지 않은 고객 9에게 상품을 추천한다.



<그림 2> 아이템 기반 협업필터링

모델 기반의 협업필터링은 메모리 기반 방식이 갖고 있는 데이터 희소성, 일관성 없는 사용자에 대한 문제점 등을 보완할 수 있다(Basilico, 2004; Son et al., 2015). 또한, 고객이 남긴 평점 데이터를 기반으로 모델링하여 평점을 제공하지 않아도 평점 예측이 가능하며 MF(Matrix Factorization)와 기계학습 방법이 있다. MF방식은 고객과 아이템 간의 잠재적 요인을 파악하여 평점을 예측하는 방식으로 SVD(Singular Vector Decomposition) 알고리즘이 대표적이다. SVD는 특이값 분해라고도 하며, 행렬을 분해함으로써 암묵적 요인(latent factor)을 반영하여 분석의 정확성을 높일 수 있다(Jeong & Kim, 2017). 이에 대한 적용 방법으로 하나의 행렬을  $U, \Sigma, V$ 로 분해하고  $U$ 와  $V$ 를 직각 행렬,  $\Sigma$ 는 대각행렬로 각각 정의하여 특이값 분해 알고리즘을 적용한다. 예를 들어 고객과 아이템에 대한  $m \times n$  크기의 행렬  $M$ 에 대해 특이값 분해를 할 경우,  $M = U\Sigma V^T$ 와 같이 나타낼 수 있다. 이때  $U_{m \times m}$ 은 고객 행렬을 나타내고,  $\Sigma_{m \times n}$ 은 특이값에 대한 대각행렬,  $V_{n \times n}^T$ 는 아이템 행렬을 나타내며(Son et al., 2015), <그림 3>과 같다.



〈그림 3〉 SVD와 암묵적 요인

기계학습 방식은 고객과 상품간의 비선형적인 복잡한 관계를 학습하여 관계를 추론하는 방식으로 데이터를 학습하여 새로운 지식을 얻을 수 있는 인공지능망, 의사결정나무, SVM 등의 기법이 있다.

## 2.2 온라인 리뷰 분석 기법과 응용

온라인 리뷰는 특정 재화나 서비스에 대한 소비자 개인의 평가와 경험 등을 의미한다(조신희, 이문용, 2014). 펜데믹 이후 디지털 트랜스포메이션이 가속화됨에 따라 어플리케이션 기반의 플랫폼을 중심으로 소비자들이 생성하는 온라인 리뷰의 수는 더욱 증가하고 있다. 이들 온라인 리뷰는 재화나 서비스에 대한 정보뿐 아니라 사용자의 사용경험과 추천여부의 성격도 포함하기 때문에 구매를 앞둔 개인에게 유용한 정보로서의 가치가 있다(Park et al., 2007). 소비자들은 제품 구매 의사결정 시 온라인 리뷰의 유용한 정보를 바탕으로 그 불확실성을 줄일 수 있다. 온라인 리뷰는 정보의 기능뿐 아니라 공유와 확장의 역할도 가지고 있어 소비자에 대한 마케팅, 판매와 같은 다양한 촉진활동에 영향을 주게 된다(Hu et al., 2008).

온라인 리뷰를 분석하기 위해 Moreo et al. (2012)은 온라인 뉴스 댓글의 어휘기반 (lexicon-based)의 감성분석 모델을 제안하였다. Darko and Liang (2022)은 LDA 기반의 통계적

모델을 제안하여 온라인 고객 리뷰의 만족도를 추출하는 방법론을 제안하였다. 기계학습 기반의 방법론으로도 Singh and Tucker (2017)은 머신러닝 기법을 이용해 고객 리뷰를 다섯 가지 제품 특성으로 분류가 가능함을 주장하고 이들 특성이 제품 평점과 관계가 있음을 주장하였다. Kiran et al. (2020)은 LSTM과 CNN 결합한 아키텍처를 구축하고 이를 이용해 온라인 리뷰의 감성지수를 분류하는 모델을 제안하였으며 나아가 어휘기반과 기계학습 기반의 방법론이 결합한 하이브리드(hybrid) 방법론도 연구되었다. Liu et al. (2021)은 LDA 기법을 이용하여 리뷰의 토픽을 추출하고 SVM을 이용해 토픽 예측 모델을 제안하였으며 박효연과 김경재(2021)는 기계학습의 단점을 보완하기 위해 BERT 기반의 감성 분석을 추천시스템에 활용하였다.

이러한 온라인 리뷰 분석기법을 비즈니스 활동에 활용하기 위한 연구도 지속적으로 진행되어 왔다. Hou et al. (2019)는 제품 디자인적 관점에서 온라인 고객 리뷰를 제품 특징, 제품의 행동 유도성, 감정, 제품 수용도, 제품 유용성 등 다섯 가지 기준으로 요약하는 모델을 제안하고 이들 정보가 제품 디자인 과정에 유용함을 주장하였다. Xu (2021)은 온디맨드 플랫폼 형태의 레스토랑 리뷰를 분석하고, 고객의 만족도에 영향을 미치는 서비스 특성을 연구하였다. Dhar and Bose (2022)는 어플리케이션 리뷰를 텍스트와 이모티콘을 이용하여 감정 및 감성 분석을 진행하고, 해당 감성 값이 어플리케이션 평점과 긍정적 관련이 있음을 밝혀냈다.

## 2.3 상품과 고객 속성 기반추천

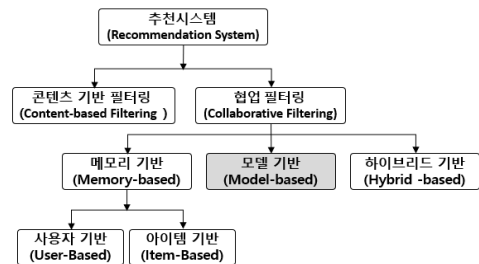
온라인 리뷰는 고객의 경험을 기반으로 생성

되어 만족 및 불만족, 긍정 및 부정 등에 대한 의견을 표현하며, 고객관리를 위한 필수적 요소로 인식되어 활용되고 있다(Anderson & Sullivan, 1993; Yao et al., 2020). 특히 영화와 같은 경험제 기반 산업에서 온라인 리뷰는 의견을 표현하고 사전정보 및 평가정보를 제공하여 영화를 결정하는 과정에 중요한 영향을 미친다. 다시 말해, 영화를 시청하기 전 고객들은 영화의 속성 및 주변정보에 의존하여 영화를 선택하며(Vany & Walls, 1996), 영화를 관람하고 작성된 고객의 리뷰는 관객의 영화 선택에 영향을 미친다(Vapnik, 1999). Neelamegham and Jain (1999)은 영화의 속성을 핵심적, 주변적, 정서적의 세가지 속성으로 분류하였다. 핵심적 속성은 배우, 스토리, 장르 등과 같이 영화 내용과 직접적으로 연관된 요소이며, 주변적 속성은 세트, 의상, 배경음악, 특수효과 등과 같이 영화의 전반적인 흐름과 관련된 요소이다. 정서적 속성은 재미있는, 호소력 있는, 흥미있는 등과 같이 영화 시청을 통해 고객이 느끼는 감정과 관련된 요소이다. 김광수(2000)는 영화선택과 관련된 속성을 핵심적 속성, 주변적 속성, 그리고 감정적 속성을 즐거움과 각성으로 나누어서 분류하였다. 본 연구에서는 영화리뷰에서 나타나는 영화의 속성을 Neelamegham and Jain (1999)과 같이 3개의 속성으로 분류하여 분류하고자 한다.

### 3. 연구 프레임워크

본 연구에서는 온라인 리뷰를 통해 상품과 고객의 세부 특성을 추출하여 추천시스템에 적용하고자 지능형 추천시스템을 제안한다. 기존의 총평점 기반의 추천시스템의 한계를 극복하고

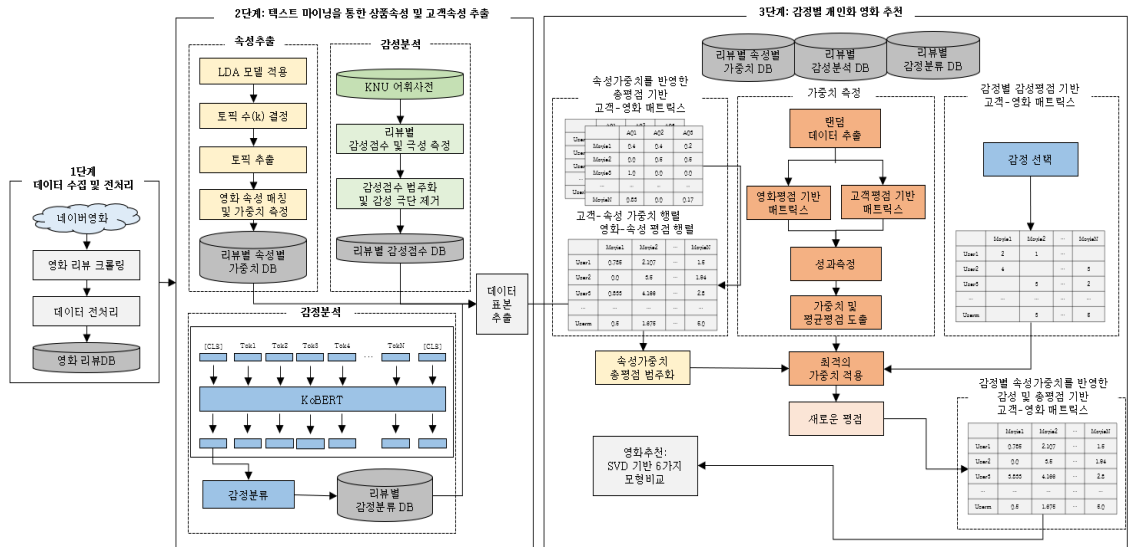
데이터 희소성의 문제를 해결하기 위해 리뷰 데이터를 활용하였으며 영화 리뷰의 평점, 영화의 속성, 고객의 감성 및 감정 결과를 통합하여 적용한다. 추천시스템은 <그림 4>와 같이 분류되며 협업필터링은 메모리 기반, 모델 기반, 하이브리드 기반으로 분류할 수 있다. 본 연구에서는 협업필터링의 모델기반에서 대표적인 알고리즘인 SVD를 적용함으로써 기존 추천시스템의 데이터 희소성에 대한 문제를 해결하고자 한다.



<그림 4> 추천시스템의 분류

연구 프레임워크는 <그림 5>와 같으며, 총 3 단계로 이루어진다. 1단계는 데이터 수집 및 전처리 과정이다. 데이터 수집은 국외 및 국내 영화를 대상으로 하며, 네이버 영화에서 한국어로 작성된 온라인 영화 리뷰 데이터를 수집한다. 한국어 특성을 반영하여 불필요한 텍스트 제거 및 변환을 통해 데이터 전처리 후 데이터베이스에 저장한다.

2단계는 텍스트 마이닝을 통한 상품속성 및 고객속성 추출 과정이다. 상품속성을 추출하기 위해 토픽모델링의 LDA를 적용하여 추출된 토픽과 영화의 3가지 속성인 핵심적, 주변적, 정서적 속성을 매칭시킨다. 속성별로 매칭된 대표 단어를 분류하여 리뷰별 속성 가중치를 측정하고 데이터베이스에 저장한다. 감성분석은 온라인



〈그림 5〉 연구 프레임워크

리뷰를 작성한 리뷰어가 표현하는 감성을 구분하는 과정이며, 어휘기반은 긍정, 중립, 부정으로 분류가 용이하다는 장점이 있다(Ma et al., 2018). 이에 감정사전을 활용하여 어휘기반의 감정분석을 수행한다. 감정분석을 위해서는 BERT 종류의 기계번역 모델로서 한국어 번역 성능이 우수한 KoBERT를 사용하여 기쁨, 슬픔, 놀람, 분노, 공포, 혐오, 중립 등의 7가지 감정으로 분류한다. 다음으로 추천시스템이 갖는 고질적 문제인 데이터 희소성 문제를 예방하기 위해 최소한의 리뷰 수를 갖는 고객을 대상으로 데이터를 추출하여 활용한다.

3단계는 감정별 개인화 영화추천을 위해 감정별 속성가중치를 반영한 감정 및 총평점을 만드는 과정이다. 먼저 고객-속성 가중치 행렬과 영화-속성 평점 행렬을 전치행렬 연산하여 속성가중치를 반영한 총평점을 생성한다. 다음은 감정별 감정평점을 생성하여, 두 평점에 최적의 가중

치 지점을 탐색하여 감정별 속성가중치를 반영한 감정 및 총평점을 만들어 고객-영화 매트릭스를 생성한다. 이후, 모델 기반의 Matrix Factorization의 대표 알고리즘인 SVD를 사용하여 다양한 평점 기반의 모형과 성과를 비교한다.

## 4. 실험 및 실험결과

### 4.1 데이터 수집 및 전처리

본 연구에서는 감정별 개인화 추천시스템을 위해 네이버 영화에서 영화 리뷰 데이터를 Python의 웹 스크래퍼를 사용하여 수집하였으며, 2022.02.14.~2022.05.01.까지 82,501개의 리뷰를 수집하였다. 6,007편의 영화에 대해 29,419명의 고객이 작성한 리뷰로 총평점은 1점에서 10점까지로 분류된다. 평점별 데이터 부족현상과 편의성을 위해 <표 1>와 같이 기준 평점을 5

〈표 1〉 영화 리뷰의 평점 범주화

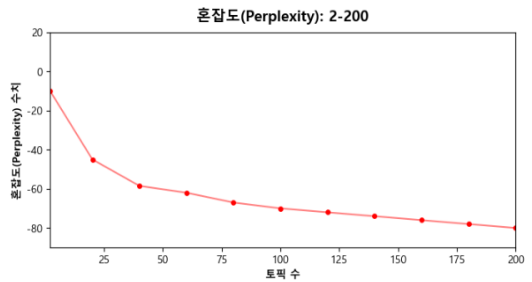
기존 평점	1	2	3	4	5	6	7	8	9	10
	8,377	5,514	915	3,447	2,318	6,457	5,546	10,293	5,181	34,453
조정된 평점	1		2		3		4		5	
	13,891		4,362		8,775		15,839		39,634	

점으로 범주화하여 조정하였다. 한국어는 1개 이상의 형태소들이 결합되어 단어를 이루고, 대부분의 자연어 처리 알고리즘이 외국 언어에 맞게 만들어졌기 때문에 영어와 같은 언어에 비해 자연어 처리 부분에서 분석 성능 및 적절성에 대한 문제가 제기되고있다(박대민, 2016). 본 연구에서는 리뷰 데이터의 신뢰성을 높이고 결과해석의 용이성을 위해 웹사이트 링크, 특수문자, 이모티콘, 공백, 숫자 및 영어, 결측치, 불용어를 제거했다(Aggarwal & Zhai, 2012; 길호현, 2018). 불용어 처리는 웹 페이지 분석 업체인 RNAK NL에서 제공하는 한국어 불용어 사전을 활용하였으며 한국어 특성을 반영하여 띄어쓰기, 맞춤법 검사기, 반복문자, 외래어를 검사하여 텍스트 데이터를 정제하여 한국어 리뷰 분석에 대한 문제점을 개선하였다.

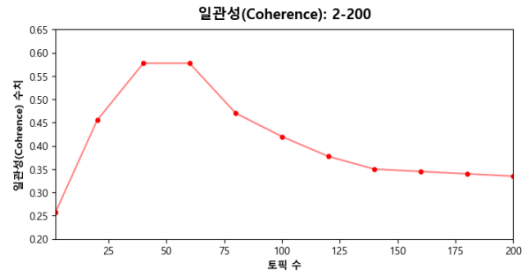
#### 4.2 영화 속성추출

영화의 속성추출을 위해 토픽모델링의 LDA를 적용하여 영화의 3가지 속성에 매칭하고 대표 키워드를 결정하였다. 최적의 토픽 수 k를 결정하기 위해 5겹 교차 검증(5-fold cross validation)으로 혼잡도(perplexity)와 일관성(coherence)을 함께 고려하였다. 일관성은 토픽별 단어가 의미론적으로 유사한지를 확인할 수 있으며 혼잡도는 특정 확률 모델이 실제로 관측되는 값을 얼마

나 잘 예측하는지 확인할 수 있다.



〈그림 6〉 토픽 수 k에 따른 혼잡도



〈그림 7〉 토픽 수 k에 따른 일관성

토픽 수 k는 2부터 200까지 설정하였으며 혼잡도 및 일관성은 <그림 6>, <그림 7>과 같다. 일관성 지표는 수치가 높을수록 유의미하며, 40부터 60까지의 지점의 크기가 동일하게 높게 나타내고 있으며, 혼잡도 지수는 낮을수록 유의미하며 점진적으로 감소하는 40과 60의 중간지점인 50을 최적의 토픽 수 k로 정했다. 최적의 토픽 수에 대한 상위 10개의 키워드를 3가지 영화 속

성에 매칭하였으며 영화 속성별 대표단어는 <표 2>와 같다. 영화 속성 및 대표 단어를 통해 리뷰별 3가지 속성의 가중치를 측정했다.

<표 2> 영화 속성별 대표단어

영화속성	단어
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, 스틸러, 캐릭터, 배우
주변적	사운드, 러닝타임, 작품, 제작, 주제, 시간, 스케일, 시리즈, 연출, 자막, 시나리오, 관계, 주제, 상황, 메세지, 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

### 4.3 감성 및 감정분석

본 연구에서는 어휘기반 감성분석을 위해 KNU사전 활용했으며, 리뷰어를 토큰화시켜 명사, 형용사, 동사만을 추출하여 리뷰별 감성점수 및 감성극성을 측정했다. 측정한 감성점수의 점수별 데이터 수가 10개 미만인 데이터를 추출하여 38개의 양극단치를 제거했다. 이러한 과정을 통해 기존의 -12부터 +15까지 분포되어 있던 감성점수를 -8부터 +8까지로 새롭게 범위를 조정했으며, 군집분석을 통해 -8~2을 1점, -1~0을 2점, 1을 3점, 2~3을 4점, 4~8을 5점으로 범주화하여 감성평점으로 활용했다.

리뷰별 감정분류를 위해서는 BERT 종류 중에서 우수한 한국어 번역 성과를 보이는 한국어 전용 번역 모델인 KoBERT를 활용하였다. 모델학

습을 위해 AI-HuB에서 제공하는 7가지 감정(기쁨, 슬픔, 놀람, 분노, 공포, 혐오, 중립)으로 분류된 38,594개의 ‘한국어 단발성 대화 데이터’를 사용했다(이세화, 2020). 감정분석 결과 16,040개의 온라인 영화 리뷰 데이터를 6,980개 행복, 997개 놀람, 3,784개 중립, 2,094개 혐오, 1,389개 슬픔, 423개 분노, 373개 공포로 분류했다.

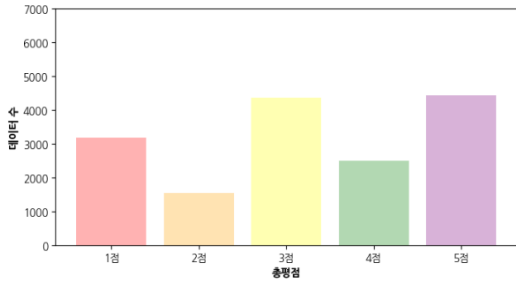
### 4.4 데이터 표본 추출

데이터 희소성은 고객이 작성한 총평점 및 리뷰를 전체 영화목록 중에서 평점 및 리뷰를 작성한 영화보다 작성하지 않은 영화 수가 많아 발생하는 추천시스템의 대표적 문제점이다. 이는 고객과 아이템간 관계의 유사도를 구하는 과정을 방해하여 잘못된 결과가 도출될 수 있다(Son et al., 2015). 본 연구에서는 데이터 희소성 문제를 줄이기 위해 고객별 10개 이상의 총평점 및 리뷰를 작성한 고객을 선택하였다. 또한, 총평점 및 감성평점의 극단치를 제거하기위해 5점 척도의 평점인 두 변수에 중립을 제외하고 긍정과 부정의 리뷰 의미가 서로 다른 데이터를 제거하여 <표 3>과 같이 16,040개의 데이터를 분석에 사용하였다. <그림 8>과 <그림 9>는 데이터 필터링 후 총평점, 감성평점의 평점별 데이터 수이다.

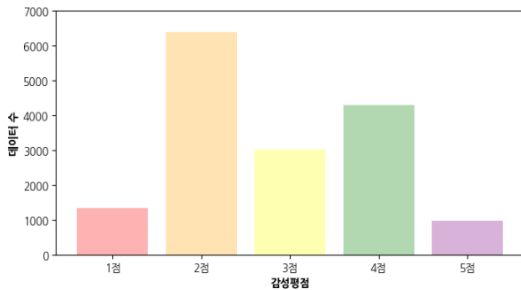
<표 3> 데이터 필터링 전/후

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





<그림 8> 총평점: 평점별 데이터 수



<그림 9> 감성평점: 평점별 데이터 수

#### 4.5 감정별 개인화 추천시스템

##### 1) 속성가중치를 반영한 총평점 기반 고객-영화 매트릭스

상품은 다양한 속성을 가지며 각 속성에 대한 고객의 중요도는 상이하하다. 또한, 모든 상품은 상대적 강점과 약점이 존재하며 각 특징에 대한 고객들의 평점은 이를 반영한다. 미래에 영화를 시청하고자 하는 고객들은 상품(영화)의 속성 및 기타 정보를 반영하여 영화를 선택하는 경향이 있다(Vany & Walls, 1996). 본 연구에서는 고객-속성 가중치 행렬과 상품-속성 가중치 행렬의 전치행렬 연산을 통해 고객의 속성가중치를 반영한 총평점 기반의 고객-영화 매트릭스를 구현했으며, <표 4>와 같다.

<표 4> 속성가중치를 반영한 총평점 기반 고객-영화 매트릭스

	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

고객-속성 가중치는 고객이 상품을 평가할 때 우선적으로 고려하는 속성에 대한 우선순위라고 할 수 있다. 특정 고객이 리뷰에서 빈번하게 언급하고 있는 속성일수록 고객의 가중치는 증가한다고 가정하며, 고객-속성 가중치 점수는 식(1)과 같다.

$$X_{ij} = \frac{8}{1 + e^{z_{ij}}} - 3 \quad \text{식(1)}$$

$z_{ij}$  : 고객  $u_i$ 가 feature  $F_j$ 에 대한 언급 횟수

식(1)에서  $u_i$ 는 고객  $u_1, u_2, \dots, u_M$  중  $i$ 번째 고객을 의미한다. 상품의 여러가지 속성들  $F_1, F_2, \dots, F_l$  중  $j$ 번째 속성인  $F_j$ 에 대하여 사용자  $u_i$ 가 언급한 횟수를  $z_{ij}$ 로 나타냈으며, 고객-속성 가중치 점수를 식(2)의  $X'_{ij}$ 에 대한 식으로 구할 수 있다. 고객별 속성 가중치의 합이 1이 되도록하며 고객-속성 가중치 행렬식은 식(2)와 같으며, <표 5>는 식을 통해 계산된 행렬이다.

$$X'_{ij} = \frac{X_{ij}}{\sum_{j=1}^3 X_{ij}} \quad \text{식(2)}$$

〈표 5〉 고객-속성 가중치 행렬

$X'_{ij}$	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

상품-속성 총평점 행렬 Y는 세부 속성에 대해 평균적으로 몇점으로 평가되고 있는지를 나타내며, 상품-속성 총평점은 식(3)과 같다.

$$Y_{ij} = \frac{2}{1 + e^{-k \cdot s_i}} + 1 \quad \text{식(3)}$$

$k$  : 상품  $p_i$ 의 속성들 중에서  $F_j$ 에 대한 리뷰의 수

$s_i$  :  $F_j$ 에 대한 평균 총평점

식(3)에서  $p_i$ 는 전체 N개 상품  $p_1, p_2, \dots, p_N$  중  $i$ 번째 상품을 의미한다. 상품  $p_i$ 의 속성 중에서  $F_j$ 에 대한 리뷰의 수를  $k$ 라고 정의하며,  $F_j$ 에 대한 평균 총평점을  $s_i$ 로 나타낸다. 그리고 상품-속성 총평점 점수  $Y_{ij}$ 를 각 행의 합이 영화의 평균 총평점( $u_i$ )이 되도록하며, 식(4)를 통해, 상품-속성 총평점 행렬을 <표 6>과 같이 구성한다.

$$Y'_{ij} = u_i \frac{Y_{ij}}{\sum_{j=1}^3 Y_{ij}} \quad \text{식(4)}$$

〈표 6〉 상품-속성 가중치 행렬

$Y'_{ij}$	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

## 2) 감정별 감성평점 기반 고객-영화 매트릭스

평점만을 활용한 기존 추천시스템은 정확도를 떨어뜨린다는 문제점이 제기되고 있으며(전병국, 안현철, 2015), 텍스트 리뷰 데이터를 추천시스템에 활용하여 문제점을 해결하고자 한다. 본 연구에서는 5점 척도의 감성평점 변수와 중립감정을 제외한 6가지 감정변수를 활용하여 감정별 감성평점 기반 고객-영화 매트릭스를 생성했으며, <표 7>과 같다.

〈표 7〉 감정별 감성평점 기반 고객-영화 매트릭스

	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

## 3) 가중치 측정

<표 4>와 <표 7>의 매트릭스 평점을 하나의 평점으로 만들기 위해 각 평점별 최적의 가중치 지점을 찾는다. 전체 데이터에서 70%를 랜덤으로 추출하여 속성가중치 총평점과 감정별 감성평점 변수를 불러와 가중치의 합이 1이 되도록

0.01단위로 차등적으로 변수에 적용하여 총 99번의 RMSE를 측정했다. 이를 통해 최적의 가중치 지점으로 감정평점 0.99, 속성가중치 총평점 0.01을 확인했다. 그리고 영화평점 기반의 RMSE 평균이 비교적 고객평점 기반 보다 낮게 나타나 매트릭스의 결측치를 영화평균으로 대체했다.

#### 4.6 분석 결과

감정별 개인화 영화추천을 위해 <표 4>와 <표 7>을 통해 도출한 최적의 가중치를 적용하여 새로운 하나의 평점을 생성했다. 본 연구에서 제안한 지능형 추천시스템 모형의 성과를 확인하기 위해 총 6가지의 모형을 구축하였다. 총평점 기반(R\_SVD), 감정평점 기반(S\_SVD), 감정 및 총평점(SR\_SVD), 속성가중치 기반 총평점(Weight R\_SVD), 속성가중치 기반 감정 및 총평점

(Weight SR\_SVD), 감정별 속성가중치 기반 감정 및 총평점(E\_Weight SR\_SVD)에 대한 성과모형을 비교 및 분석했다.

본 연구에서는 협업필터링의 모델 기반 방법인 SVD를 적용했으며, 결측치는 각 연구모형별 가중치 측정을 통해 영화평점의 평균으로 결측치를 대체했다. 그리고 최적의 잠재요인 k를 측정하기 위해 10부터 160까지의 잠재요인을 설정하여 최적의 잠재요인 k를 도출했으며, R\_SVD는 90, S\_SVD는 10, SR\_SVD는 10, Weight R\_SVD는 20, Weight SR\_SVD는 20, E\_Weight SR\_SVD는 62로 각각 선정했다. 성과를 측정하기 위해서는 RMSE(Root Mean Square Error), MAE(Mean Absolute Error), MAPE(Mean Absolute Percent Error) 지표는 사용하였다.

6가지 종류의 평점 모형을 실험하여 성과를

<표 8> 비교모형별 성능지표

평가 지표 구분		MAE	MAPE	RMSE	대체 결측치 잠재요인	결측치	가중치
비교 모형	R_SVD	1.076	53.46%	1.208	영화, 90	99.71%	
	S_SVD	0.951	42.75%	1.050	영화, 10	99.71%	
	SR_SVD	0.859	36.56%	0.962	영화, 10	99.71%	총평점(0.12) 감성평점(0.88)
	Weight_R_SVD	0.211	11.11%	0.271	영화, 20	40.35%	
	Weight_SR_SVD	0.124	9.06%	0.209	영화, 20	40.01%	감성평점(0.97) 속성가중치(0.03)
	E_Weight_SR_SVD	0.020	1.33%	0.074	영화, 62	40.00%	감성평점(0.99) 속성가중치(0.01)

<표 9> 감정별 성능지표 및 파라미터

평가 지표 구분		MAE	MAPE	RMSE	대체 결측치 잠재요인	결측치	가중치
감정 분류 (중립 제외)	행복(6,980 개)	0.018	1.29%	0.093	영화, 90	39.64%	감성평점(0.99) 속성가중치(0.01)
	놀람(997 개)	0.019	1.31%	0.069	영화, 10	39.98%	
	슬픔(1,389 개)	0.015	0.90%	0.065	영화, 110	41.23%	
	혐오(2,094 개)	0.013	0.90%	0.060	영화, 70	39.62%	
	분노(423 개)	0.027	1.81%	0.076	영화, 50	38.12%	
	공포(373 개)	0.028	1.77%	0.079	영화, 40	41.18%	
평균 성과		0.020	1.33%	0.074	영화, 62	40.00%	

측정한 결과 <표 8>과 같으며 <표 9>은 감정별 성과지표를 계산한 결과이다. 본 연구에서 제안한 감정별 속성가중치 기반의 E\_Weight\_SR\_SVD 모형의 RMSE가 0.074로 가장 우수한 성과를 보였다. 두 번째로 우수한 모형인 Weight\_SR\_SVD의 RMSE는 0.209로 나타나 감정을 반영함으로써 성능이 향상됨을 확인할 수 있었다. 또한, 속성가중치를 반영한 총평점을 적용한 모형인 E\_Weight\_SR\_SVD, Weight\_SR\_SVD, Weight\_R\_SVD 등의 결측치가 약 40%로 약 99% 결측치를 갖는 다른 모형에 비해 데이터 희소성 문제를 개선시켰음을 확인하였다.

## 5. 결과 분석 및 토론

본 연구에서는 온라인 리뷰의 상품 속성과 고객 속성을 통합한 지능형 추천시스템을 제안하였으며 추천시스템의 대표적인 문제점인 데이터 희소성을 해결하고 추천시스템의 성능을 향상시키고자 하였다. 데이터 희소성의 문제를 해결하기 위해 속성의 가중치를 반영하여 모형을 구축함으로써 90% 이상의 결측치를 42% 미만으로 줄일 수 있었으며 온라인 영화리뷰에 토픽모델링, 감성 및 감정분석을 적용하여 본 연구에서 제안하는 지능형 추천시스템의 우수한 성과를 확인하였다. 기존의 평점을 중심으로 한 모형(R\_SVD)의 MAPE는 53.46%로 나타났으나, 감정 및 감성과 속성에 대한 가중치를 적용한 지능형 추천시스템에서는 50%이상 개선하여 MAPE가 1.33%로 나타났다. 또한, 한국어로 작성된 리뷰 데이터를 분석하기 위해 한국어의 특성을 반영하여 데이터 전처리를 수행함으로써 한국어 리뷰분석에 대한 문제점을 개선하고 데이터에 대

한 신뢰성을 확보하였다.

본 연구의 학술적 시사점은 다음과 같다.

첫째, 메모리 기반의 추천시스템은 데이터 희소성에 대한 문제가 존재하였으나 본 연구에서는 모델 기반 모형의 SVD 알고리즘을 적용하였으며 정량적 데이터와 정성적 데이터를 결합하여 데이터 희소성에 대한 문제를 해결하였다. 기존의 추천시스템은 대다수가 정량적인 데이터에만 의존하였지만, 본 연구에서는 정성적 데이터인 리뷰 데이터를 활용하여 정량적 데이터와 정성적 데이터를 결합한 개인화된 추천시스템을 구축하였다. 리뷰 데이터에서 고객의 속성, 감성 및 감정 정보를 추출하였으며 이를 정량적 데이터로 변환하여 추천시스템에 적용함으로써 대량의 데이터 손실을 예방할 수 있었으며 추천시스템의 데이터 희소성에 대한 문제를 해결하였다. 둘째, 속성가중치를 반영한 총평점 기반의 매트릭스와 감정별 감성평점 기반의 매트릭스를 하나의 고객-영화 평점 매트릭스로 생성하기 위해 평점별 최적의 가중치 지점을 측정하는 프로세스를 구축하였다. 가중치는 모형의 성능을 좌우하는 중요한 요인으로 감정평점과 속성가중치 총평점에 대한 최적의 가중치를 탐색하기 위한 방법을 제안하였다. 셋째, 온라인 리뷰가 갖는 정량적 데이터인 총평점과 정성적 데이터인 영화속성, 감성점수, 감정분류를 활용하여 하나의 통합모형을 제안하였다. 영화의 속성을 추출하기 위해 토픽모델링 기법인 LDA를 적용하였으며 고객 속성을 추출하기 위해 KNU 한국어 감성사전과 KoBERT를 적용하여 이들을 통합한 지능형 추천시스템을 제안하였으며 추천시스템을 위해 영화의 속성과 고객의 속성의 통합이 처음으로 시도되었다.

본 연구의 실무적 시사점은 다음과 같다.

첫째, 정량적 데이터인 총평점 기반의 기존 추천모형에서 온라인 고객 리뷰 데이터에 텍스트 마이닝의 토픽 모델링, 감성 및 감정분석을 적용한 방법론을 제시했다. 온라인 리뷰는 실시간으로 제공되며 고객의 솔직한 반응을 즉각적으로 파악할 수 있다. 이러한 고객의 반응이 추천모형에 즉각적으로 반영된다면 보다 현실적인 추천이 이루어질 수 있을 것이다. 이는 온라인 리뷰를 활용하는 전자상거래 시장의 다양한 분야에 적용이 가능하여 호텔 추천, 레스토랑 추천, 관광지 추천 등에 적용할 수 있을 것이다. 둘째, 다양한 자료원천을 대상으로 추천시스템을 개발할 수 있는 방법론을 제시하였다. 3,343편의 영화와 1,407명의 고객에 대해 총 82,501개의 리뷰 데이터를 활용하여 제안된 지능형 추천시스템 모형에 대해 검증하였으며 우수한 성능을 확인하였다. 셋째, 추천시스템에 많이 활용되고 있는 협업필터링의 메모리 기반 모형의 한계점을 극복하기 위해 모델 기반 모형의 알고리즘을 활용한 추천모형을 제안했다. 과거의 데이터가 충분하지 않은 신규고객에게도 적용이 가능하며 새로운 영화에 대해서도 지능형 추천시스템을 활용할 수 있을 것이다. 넷째, 한국어로 작성된 리뷰 데이터를 활용한 한국어 자연어 처리를 통해 한국어 기반의 텍스트 마이닝 연구의 발전에 기여하였다. 한국어의 특성을 반영하여 데이터 전처리를 수행하여 한국어로 작성된 리뷰에 대한 신뢰도를 확보하였다.

## 6. 결론 및 향후 연구방향

다양한 상품정보는 고객에게 다양한 선택의 기회를 제공하지만 지나친 다양성으로 인해 고

객은 상품선택의 어려움을 겪고 있으며 추천시스템은 고객의 상품선택에 도움을 줄 수 있다. 본 연구에서는 영화 리뷰 데이터를 기반으로 정량적 데이터인 총평점과 정성적 데이터인 리뷰를 토픽모델링, 감성 및 감정분석을 함께 적용하여 하나로 통합한 지능형 추천시스템을 제안하였다. 기존의 추천시스템 연구에서 영화 데이터를 많이 활용해 왔으나 영화 리뷰 데이터에 대한 사용은 미비하다. 이에 본 연구에서는 영화의 리뷰 데이터를 활용하여 제안된 방법론의 성과를 검증하였다. 제안된 모형은 기존 추천시스템의 데이터 희소성에 대한 문제점을 개선하였으며 우수한 성능을 확인하였다.

본 연구의 한계 및 향후 연구방향은 다음과 같다.

첫째, 본 연구에서 82,501개의 리뷰 데이터를 수집하였으나 고객별 10개 이상의 총평점 및 리뷰를 작성한 고객만 추출하여 16,040개의 리뷰 데이터만 추천시스템 구축에 사용되었다. 신규고객일 경우 10개 이상의 데이터를 제공하지 않을 수 있으며 10개 미만 작성한 고객이라도 중요한 정보를 제공할 수 있다. 향후 연구에서는 보다 다양한 고객의 의견을 반영할 수 있도록 하며 고객의 성향에 따른 과소 또는 과대평가에 대한 적절한 방법도 모색할 필요가 있다. 둘째, 어휘 기반의 감성분석을 위해 특수문자와 이모티콘 등을 제거하고 감성점수 및 극성을 측정하였으며 명사, 형용사, 동사만을 추출하였으나 향후 연구에서는 대명사, 부사 등 다양한 품사를 활용하고 특수문자와 이모티콘을 추천모형에 반영해 볼 수 있을 것이다. 셋째, 협업필터링 모델기반의 SVD 알고리즘으로 추천시스템을 구축하였으나 딥러닝 기반의 추천알고리즘을 적용하여 보다 우수한 성과를 기대할 수 있을 것이다.

## 참고문헌(References)

### [국내 문헌]

- 길호현. (2018). 텍스트마이닝을 위한 한국어 불용어 목록 연구. *우리말글*, 78, 1-25.
- 김광수. (2000). 영화 선택 및 평가에 관한 연구. *Korean Association for Advertising and Public Relations*, 48, 139-164.
- 김민정, 김수현, 오지혜, 엄지윤, 강주영. (2021). SNS 텍스트 마이닝 기반 포스트 코로나 트렌드 차박 여행 지도 제작 및 차박지 추천에 관한 연구. *한국IT서비스학회지*, 20, 11-28.
- 김진화, 변현수, 이승훈. (2008). 온라인 리뷰와 미니멀리즘. *한국전자거래학회 심포지움 및 기타간행물*, 235-252.
- 박대민. (2016). 뉴스 기사의 자연어처리:< 뉴스 소스 베타> 를 중심으로. *커뮤니케이션 이론*, 12(1), 4-52.
- 박호연, 김경재. (2021). BERT 기반 감성분석을 이용한 추천시스템. *지능정보연구*, 27(2), 1-15.
- 윤호민, 최규완. (2020). 사용자 선호기반 개인화 음식메뉴 추천 기법 연구. *호텔경영학연구*, 29(1), 83-100.
- 이세화. (2020). KETI 지능정보 플래그십 R&D 데이터. (주)아크릴, <https://aihub.or.kr/opedata/keti-data/recognition-laguage/KETI-02-009>
- 전병국, 안현철. (2015). 사용자 리뷰 마이닝을 결합한 협업 필터링 시스템: 스마트폰 앱 추천에의 응용. *지능정보연구*, 21(2), 1-18.
- 조신희, 이문용. (2014). 온라인 제품 리뷰의 유용성 결정 요인 분석을 통한 리뷰 활용 방안 도출. *Entrue Journal of Information Technology*, 13(1), 29-40.

현지연, 유상이, 이상용. (2019). 평점과 리뷰 텍스트 감성분석을 결합한 추천시스템 향상 방안 연구. *지능정보연구*, 25(1), 219-239.

SKTBrain / KoBERT, <https://github.com/SKTBRIN/KOBERT>

### [국외 문헌]

- Aggarwal, C. C., & Zhai, C. (2012). *A survey of text classification algorithms*. In *Mining text data* (pp. 163-222). Springer, Boston, MA.
- Al-maaitah, T. A., Majali, T. E., Alsoud, M., & Al-maaitah, D. A. (2021). The Impact of COVID-19 on the Electronic Commerce Users Behavior. *Journal of Contemporary Issues in Business and Government*, 27(1), 784-793.
- Anderson, E. W., & Sullivan, M. W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing science*, 12(2), 125-143.
- Basilico, J., & Hofmann, T. (2004, July). Unifying collaborative and content-based filtering. *In Proceedings of the twenty-first international conference on Machine learning* (p. 9).
- Breese, J. Heckerman, D., & Kadie, C. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. *In Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI.
- Cheng, Y. H., & Ho, H. Y. (2015). Social influence's impact on reader perceptions of online reviews. *Journal of Business Research*, 68(4), 883-887.
- Darko, A. P., & Liang, D. (2022). Modeling customer satisfaction through online reviews: A FlowSort group decision model under

- probabilistic linguistic settings. *Expert Systems with Applications*, 195, 116649.
- Dhar, S., & Bose, I. (2022). Walking on air or hopping mad? Understanding the impact of emotions, sentiments and reactions on ratings in online customer reviews of mobile apps. *Decision Support Systems*, 113769.
- Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61-70.
- Hou, T., Yannou, B., Leroy, Y., & Poirson, E. (2019). Mining customer product reviews for product development: A summarization process. *Expert Systems with Applications*, 132, 141-150.
- Hu, N., Liu, L., & Zhang, J. J. (2008). Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology and management*, 9(3), 201-214.
- Jeong, S. Y., & Kim, H. J. (2017). A recommender system using factorization machine. *Journal of Digital Contents Society*, 18(4), 707-712.
- Kiran, R., Kumar, P., & Bhasker, B. (2020). Oslcfit (organic simultaneous LSTM and CNN Fit): A novel deep learning based solution for sentiment polarity classification of reviews. *Expert Systems with Applications*, 157, 113488.
- Lee, R. K., Chung, N., & Hong, T. (2019). Developing the online reviews based recommender models for multi-attributes using deep learning. *The Journal of Information Systems*, 28(1), 97-114.
- Li, M., Huang, L., Tan, C. H., & Wei, K. K. (2013). Helpfulness of online product reviews as seen by consumers: Source and content features. *International Journal of Electronic Commerce*, 17(4), 101-136.
- Liu, F., Lai, K.-H., Wu, J., & Duan, W. (2021). Listening to online reviews: A mixed-methods investigation of customer experience in the sharing economy. *Decision Support Systems*, 149, 113609.
- Ma, E., Cheng, M., & Hsiao, A. (2018). Sentiment analysis—a review and agenda for future research in hospitality contexts. *International Journal of Contemporary Hospitality Management*, 30(11), 3287-3308.
- Moreo, A., Romero, M., Castro, J. L., & Zurita, J. M. (2012). Lexicon-based Comments-oriented News Sentiment Analyzer system. *Expert Systems with Applications*, 39(10), 9166-9180.
- Neelamegham, R., & Jain, D. (1999). Consumer choice process for experience goods: An econometric model and analysis. *Journal of marketing research*, 36(3), 373-386.
- Park, D. H., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International journal of electronic commerce*, 11(4), 125-148.
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001, April). Item-based collaborative filtering recommendation algorithms. *In Proceedings of the 10th international conference on World Wide Web* (pp. 285-295). arXiv:1301.7363.
- Singh, A., & Tucker, C. S. (2017). A machine learning approach to product review disambiguation based on function, form and behavior classification. *Decision Support Systems*, 97, 81-91.
- Son, J., Kim, S. B., Kim, H., & Cho, S. (2015).

- Review and analysis of recommender systems. *Journal of Korean institute of industrial engineers*, 41(2), 185-208.
- Vany, A. D., & Walls, W. D. (1996). Bose-Einstein dynamics and adaptive contracting in the motion picture industry. *The Economic Journal*, 106(439), 1493-1514.
- Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.
- Xu, X. (2021). What are customers commenting on, and how is their satisfaction affected? Examining online reviews in the on-demand food service context. *Decision Support Systems*, 142, 113467.
- Yao, Z. Y., Park, Y. K., and Hong, T. H. (2020). A study on the Effect of Reviewer Attributes on the Usefulness of Online Reviews. *Information Systems Journal*, 29(2), 173-195.
- Yun, S. Y., & Yoon, S. D. (2020). Item-Based Collaborative Filtering Recommendation Technique Using Product Review Sentiment Analysis. *Journal of the Korea Institute of Information and Communication Engineering*, 24(8), 970-977.



Abstract

## An Intelligent Recommendation System by Integrating the Attributes of Product and Customer in the Movie Reviews

Taeho Hong\* · Junwoo Hong\*\* · Eunmi Kim\*\*\* · Minsu Kim\*\*\*\*

As digital technology converges into the e-commerce market across industries, online transactions have activated, and the use of online has increased. With the recent spread of infectious diseases such as COVID-19, this market flow is accelerating, and various product information can be provided to customers online. Providing a variety of information provides customers with various opportunities but causes difficulties in decision-making. The recommendation system can help customers to make a decision more effectively. However, the previous research on recommendation systems is limited to only quantitative data and does not reflect detailed factors of products and customers. In this study, we propose an intelligent recommendation system that quantifies the attributes of products and customers by applying text mining techniques to qualitative data based on online reviews and integrates the existing objective indicators of total star rating, sentiment, and emotion. The proposed integrated recommendation model showed superior performance to the overall rating-oriented recommendation model. It expects the new business value to be created through the recommendation result reflecting detailed factors of products and customers.

**Key Words** : Recommendation system, Collaborative filtering, SVD, Text mining, Movie review

Received : May 27, 2022 Revised : June 16, 2022 Accepted : June 17, 2022

Corresponding Author : Eunmi Kim

---

\* College of Business Administration, Pusan National University

\*\* College of Business Administration, Pusan National University

\*\*\* Corresponding Author: Eunmi Kim

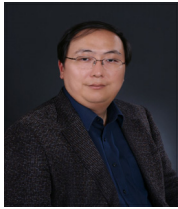
Kookmin Information Technology Research Institute, Kookmin University

77 Jeongneung-ro, Seongbuk-gu, Seoul 02707, Korea

Tel: +82-51-510-2531, E-mail: keunmi100@gmail.com

\*\*\*\* College of Business Administration, Pusan National University

## 저 자 소개



### 홍 태 호

현재 부산대학교 경영학과 교수로 재직 중이다. KAIST에서 경영정보시스템을 전공하여 공학석사와 공학박사를 취득하였다. 주요 관심분야는 비즈니스 애널리틱스, 딥러닝, 오피니언 마이닝, CRM 등이다. 주요 논문을 Expert Systems, Expert Systems with Applications, Information Processing & Management, Asia Pacific Journal of Information Systems, 지능정보연구, 정보시스템연구 등에 게재하였다.



### 홍 준 우

울산대학교 경영학과에서 학사학위를 취득하였다. 현재 부산대학교 경영학과 경영정보 전공 석사과정에 재학 중이다. 주요 관심분야는 비즈니스 애널리틱스, 텍스트 마이닝, 자연어 처리, 추천시스템 등이다.



### 김 은 미

부산대학교 경영학과에서 석사학위와 박사학위를 취득하였다. 현재 국민대학교 정보기술연구소 연구원으로 있으며 주요 관심분야는 비즈니스 애널리틱스, 딥러닝, 소셜미디어, CRM 등이다. 주요 논문을 Expert Systems with Applications, 인터넷전자상거래연구, 정보시스템연구, 지식경영연구 등에 게재하였다.



### 김 민 수

부산대학교 경영학과에서 학사학위를 취득하였다. 현재 부산대학교 경영학과 경영정보 전공 석사과정에 재학 중이다. 주요 관심분야는 비즈니스 애널리틱스, 신용평가, 딥러닝 등이다.