

텍스트 마이닝 기법을 활용한 구글 플레이 스토어 영어 학습 앱 사용자 리뷰 분석

김기연¹ · 한수미^{2*}¹한림대학교 융합소프트웨어학과 석사과정^{2*}한림대학교 디지털인문예술전공/영어영문학과 조교수

User Review Analysis of English Learning Applications on Google Play Store Using Text-Mining

Gi-Youn Kim¹ · Sumi Han^{2*}¹Master's Course, Department of Convergence Software, Hallym University, Chuncheon 24252, Korea^{2*}Assistant Professor, Digital Humanities and Arts/Department of English Language and Literature, Hallym University, Chuncheon 24252, Korea

[요약]

모바일 앱을 활용한 에듀테크(EduTech) 시장이 커지면서, 사용자의 만족도를 높이기 위한 연구가 활발하다. 이에 본 연구는 텍스트 마이닝을 통해 리뷰 속에 담긴 영어 학습 앱 사용자의 만족, 불만족 요소를 분석하였다. 구글 플레이 스토어의 영어 학습 앱 225개에서 추출한 246,073개 리뷰의 감성 분석 결과, 긍정 리뷰 168,577개(68.5%)와 부정 리뷰 77,497개(31.5%)로 이루어져, 다수 사용자의 긍정적인 인식을 알 수 있었다. 토픽모델링 분석으로 도출된 긍정 토픽은 '단어 암기', '회화', '예문', '게임', 그리고 '강의'였고, 부정 토픽은 '콘텐츠', '오작동', '결제', '기능', 그리고 '가격'이었다. 단어 및 회화 등의 학습 내용에 대한 다소 긍정적인 평가에 비해, 영어 학습 앱 기능적인 결함에 대한 불만족이 있었다. 이러한 사용자 만족, 불만족 요소에 대한 이해는 향후 영어 학습 앱 분야에서 마케팅 전략 및 콘텐츠 개발 전략 수립에 활용될 수 있어, 실용적 의미가 크다.

[Abstract]

As the EduTech market of mobile applications grows, research to increase user satisfaction is active. In this regard, this study analyzed the satisfaction and dissatisfaction aspects of user reviews on English learning applications using text-mining. A total of 246,073 reviews from 225 English learning applications on Google Play Store were sentimentally analyzed into 168,577 positive reviews (68.5%) and 77,497 negative reviews (31.5%), indicating more positive user perception. Positive topics from topic modeling include 'word memorization', 'conversation', 'example sentence', 'game', and 'lecture', and negative topics include 'content', 'malfunction', 'payment', 'function', and 'price'. Despite positive evaluation of learning contents, functional defects of the applications were pointed out. These results are of great practical significance since they can be used in marketing and content development in the target domain.

색인어 : 영어 학습 앱, 사용자 리뷰, 텍스트 마이닝, 감성 분석, 토픽 모델링

Keyword : English learning application, User review, Text-mining, Sentiment analysis, Topic modeling

<http://dx.doi.org/10.9728/dcs.2022.23.10.1901>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 14 August 2022; **Revised** 05 October 2022

Accepted 11 October 2022

***Corresponding Author, Sumi Han**

Tel: +82-33-248-1532

E-mail: sumihan20@gmail.com

I. 서론

Pew Research(2019)는 대한민국의 스마트폰 보급률이 95%에 달한다고 밝혔다. 조사대상에서 선진국으로 분류된 나라 전체 평균인 76%를 상회하는 수치이다. 이러한 높은 스마트폰 보급률은 새로운 산업이 창출될 수 있는 기반이 되었다. 동시에 블루오션과도 같은 그 기반 위에서 시장을 선점하기 위한 다양한 앱이 출시되었고, 이에 따라 성장한 국내 모바일 콘텐츠 산업은 2016년부터 연평균 10.3%씩 증가하여 2019년에는 2018년 대비 12.5% 증가한 26조 28억 규모로 추산되고 있다[1]. 이는 생활에 밀접한 스마트폰의 특성상 다양한 산업 분야를 접목시킬 수 있고 동시에 카메라, 위성 항법 장치(GPS; global positioning system), 온도도 센서 등의 다양한 스마트폰의 기능을 이용해 창의적인 기능을 구현하는 것이 가능해 소비자들의 다양한 니즈를 충족시키기에 최적의 조건을 갖추고 있기 때문이다.

특히 신문물에 대한 적응력이 높은 젊은 연령층의 스마트폰 보급률은 매우 가파르게 상승하였다. 애플의 아이폰이 2009년 11월 28일에 국내 최초 출시된 후 약 4년만인 2012년 11월에 20대의 스마트폰 보급률은 95%, 30대의 스마트폰 보급률은 91%에 달할 정도로 젊은 연령층의 스마트폰 수요는 급증하였다[2]. 이러한 상황 속에 대한민국의 교육 시장 역시 스마트폰을 주목하였다. 한국모바일산업연합회가 2012년 9월에 출간한 글로벌 앱 정보 제공지 The Mobile는 영어 키키토리왕국이라는 일본의 영어 학습 앱의 성공비결을 다각도로 분석한 바 있고, 2013년 8월 호에는 스테디맥스(스피킹 맥스 서비스)의 성공 비결과 접근 방법을 다루며 분석하였다. 이러한 관심 속에 성장한 2019년 대한민국 모바일 교육 분야 매출액은 2018년 대비 13.8% 성장한 6,002억 원을 기록하며 가파른 성장을 보여주고 있다[1].

그러나 시장의 규모 대비 서비스 자체에 관한 연구는 거의 이루어지지 않은 것이 실정이다. 몇몇의 기존 연구는 주로 이를 이용한 학습자의 학업 성취도 향상에 집중하면서, 기존 학습 방법과 비교하여 모바일 교육 앱을 활용한 학습 방법의 향상된 학업 성취도에 주목했으며 동시에 대부분의 연구는 교육학적 관점에서 교육 앱의 효용성을 바라보고 있다[3]-[5]. 구체적으로 교육 앱의 어떤 요소가 사용자에게 긍정적인지 부정적인지에 대한 연구는 거의 없는 편이다. 최근, 교육 앱 시장은 계속 성장할 것이 예측되었기에 교육 앱의 효용성에 관한 연구뿐 아니라 교육 앱 자체의 기획, 개선, 마케팅을 위한 사용자 경험의 연구가 필요한 시점이다.

Statcounter(2021)에 따르면 국내에서 모바일 운영체제로 안드로이드를 사용 중인 인구가 약 70% 이상이므로 본 연구는 구글 플레이 스토어에 게시된 다양한 모바일 교육 앱 중 영어 학습 관련 앱을 대상으로 사용자 경험을 긍정과 부정으로 나누어 주요 사용자 경험 요소를 추출하고 분석하였다. 이를 위해서 텍스트 마이닝 기법인 딥러닝 모델을 활용한 감성분석과 LDA 토픽모델링을 활용하여 영어 앱 사용자 리뷰 분석을 시도하였다.

II. 문헌연구

2-1 앱 연구 동향

영어 학습 앱 관련 연구는 학습자 개인의 성취도 향상에 관한 연구가 대부분을 차지하고 있다. 동시에 영어 학습 앱이 가지는 교육학적 의미가 크기에 대부분의 연구가 교육학적 관점에서 진행되어 학습 만족도, 학습 경험, 스마트폰을 통한 협동학습 등의 내용이 주를 이뤘다. 예를 들어, 영어 학습 앱을 통한 학습이 학습자의 학업 성취도에 미치는 효과에 대한 연구들[3]-[4]과 영어 학습 중에서도 어휘 학습 분야에 영어 학습 앱을 도입했을 때 학습자의 어휘 습득에 미치는 영향에 대한 연구가 있다[5]. 또한 앞선 영어 학습 앱 관련 연구들은 설문조사와 같은 양적 조사를 이용하거나 면담과 같은 질적 조사의 연구 방법을 채택하기도 하였다. 이러한 연구들의 결론은 대체로 영어 학습 앱이 학습자의 학업 성취도를 향상시키는 데 효과가 있다는 것이었고 더 나아가 그 활용을 확대해야 한다는 제안이었다.

영어 학습 앱과 달리 타 분야 앱에 대한 연구는 대부분 해당 앱의 기획, 개선, 그리고 마케팅에 대한 인사이트를 도출하는 방향의 연구가 주를 이루고 있다. 이를 위해 대량의 사용자 리뷰를 수집하고 이를 텍스트 마이닝 기법을 통해 분석하여 유의미한 사용자 경험 지표를 도출하는 방식을 택한 연구가 대부분이다. 예를 들어, 모바일 증강현실 앱 리뷰를 분석하여 모바일 증강현실 앱의 개선 방향을 제시하거나 지역화폐 앱 사용자 리뷰 분석을 통해 지역화폐 앱의 마케팅 방향을 제시하고 영어 모바일 헬스 애플리케이션 사용자 리뷰 분석을 통해 사용자들의 니즈를 도출한 연구가 있다[6]-[8]. 더 나아가, 텍스트 마이닝 기법과 ARIMA(Auto-regressive Integrated Moving Average) 모델을 결합해 배달의 민족 앱 사용자 리뷰를 분석하고 이슈요인을 추출하여 서비스 개선의 필요를 제시한 연구도 있다[9]. 이러한 연구는 해당 서비스들의 분석을 통해 개선점 및 시사점을 제시했을 뿐 아니라 해당 산업에도 의미 있는 결과를 도출해냈다. 따라서 본 연구에서는 이와 같은 연구들의 방식을 채택해 양적 조사와 질적 조사 방법에서 더 나아가 영어 학습 앱 연구를 진행해 영어 학습 앱의 발전에 활용될 수 있는 다양한 공부정 요소를 도출해내고자 한다. 이러한 방식으로 도출된 공부정 요소는 관련 분야의 실무 환경에서 다각도로 활용이 가능할 것이다.

2-2 텍스트 마이닝과 리뷰 분석

사용자 경험을 분석해 제품 또는 서비스의 기획, 개선, 마케팅에 관한 통찰을 제시하는 연구는 꾸준히 이루어져 왔다. 이를 위해 기존의 사용자 경험 관련 연구들은 FGI(Focus Group Interview), 설문 조사 등의 전통적 양적·질적 연구방

법을 이용하였다. 이때 설문 조사는 사전에 정교하게 짜여진 설문 문항을 이용하기 때문에 사용자별로 다른 경험을 도출하기 어려웠다. 동시에 표본의 크기가 제한되거나, 측정 항목의 일관성이 보장되지 않아 편향된 데이터를 수집할 수 있으며 또한 많은 시간 소모와 정확도의 저해 가능성이 있다. 더불어 설문 조사에 응하는 사용자들은 연구 진행자를 의식하거나 주변 환경에 의해 편향된 응답을 할 수 있다. 마지막으로 앞선 연구에서 다루지 않은 주제는 고려되지 않을 수 있기에 새로운 주제의 발굴이 어렵다[10].

한편, 사용자가 생성한 콘텐츠 중 하나인 사용자 리뷰는 이러한 한계를 보완한다. 사용자 리뷰는 고객의 고객 경험 정보의 원천이 될 수 있기 때문이다[11]. 특히 온라인 사용자 리뷰는 고객이 자발적으로 통찰력 있는 의견을 작성한 것이고 무료 또는 아주 적은 비용으로 쉽게 접근 가능한 존재이다[10]. 동시에 불특정 다수가 작성한 방대한 양의 사용자 리뷰는 대중의 지혜(wisdom of crowds)를 만들어 내기도 한다.

하지만 소비자 리뷰는 비교적 짧은 데이터이며 동시에 비정형 데이터이기에 이를 정형화하고 분석이 가능한 형태의 데이터로 바꾸기 위한 자연어처리 분석과정이 필요하다[8]. 이를 위해 최신의 다수 연구들은 다양한 텍스트 마이닝 기법을 활용하는데 그 중 토픽모델링이 주로 이용되고 있다[12]. 토픽 모델링은 대량의 데이터를 분석할 수 있고 다양한 비정형 데이터에 적용하여 패턴을 찾을 수 있기 때문이다[13]. 앞선 앱 연구 동향에서 밝힌 다수의 기존 앱 연구에서도 토픽 모델링을 이용한 결론을 도출해내기도 하였다. 그렇기에 본 연구에서는 대량의 영어 학습 앱 리뷰의 긍정, 부정요인을 도출해내기 위해 토픽 모델링 기법을 사용하려 한다.

이때 기존 연구들은 사용자 리뷰에 매겨진 별점을 이용해 사용자의 감정을 추측하였다[6][8]. 별점 1~2점은 부정, 별점 4~5점은 긍정으로 리뷰를 분류하여 대량의 리뷰를 분류하였는데 이는 평점과 리뷰가 불일치하는 경우를 대응할 수 없는 한계점이 있다[8]. 동시에 별점과 실제 리뷰 내용에 존재하는 편향을 발견하는 연구도 최근 진행된 바 있다[14]. 이는 텍스트만을 이용해 자동으로 긍부정 리뷰를 분류할 수 있는 딥러닝 기법을 활용해 보강될 수 있다[8].

토픽 모델링은 텍스트 마이닝 기법 중 통계적 추론을 이용한 방법으로 문장 속에 숨겨진 주제(Topic)를 추출하기 위해 만들어진 방법이다. 대표적으로 LDA(Latent Dirichlet Allocation) 알고리즘을 사용하는데 이 알고리즘은 문헌에서 단어의 출현이 곧 문서의 주제에 의해 결정된다고 가정한다. LDA 알고리즘을 활용한 토픽 모델링은 사전학습 과정 없이 숨겨진 주제의 구조를 찾을 수 있다는 것이다[15]. LDA 토픽 모델은 함께 자주 나타나는 단어의 그룹을 하나의 토픽으로 간주하며, 이때 토픽은 데이터 속 단어의 출현 비율에 따라 정해진다.

III. 연구 방법

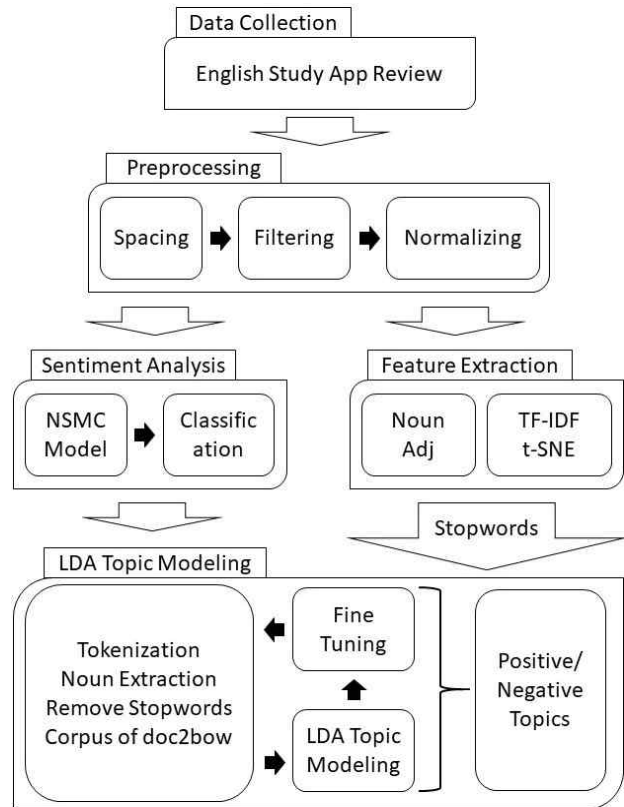


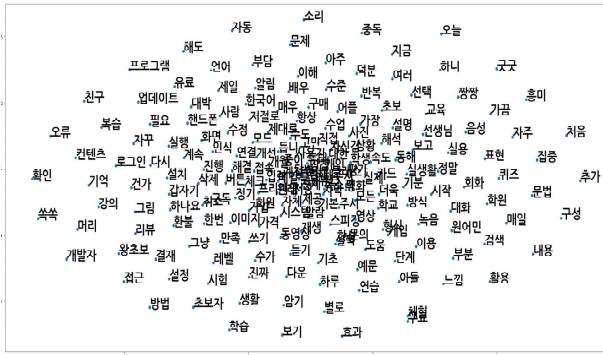
그림 1. 모델링 과정
Fig. 1. Modeling Procedure

<그림 1>은 본 연구의 데이터 수집 및 분석 절차를 나타낸 것이다. 우선, 연구 대상 앱 선정하고, 데이터 수집, 전처리, 그리고 특징을 추출 단계를 거쳤다. 이후 본 연구에서 사용한 텍스트 마이닝 기법인 감성분석과 토픽 모델링 분석을 거쳐서 사용자의 리뷰 속에 담긴 영어 학습 앱의 만족 그리고 불만족 요소를 파악하였다. 다음 서브섹션에서 단계별 구체적인 설명을 확인할 수 있다.

3-1 연구 대상 앱 선정

본 연구의 연구 대상인 구글 플레이 스토어 영어 학습 앱의 선정 기준은 구글 플레이 스토어에 “영어 공부”를 검색했을 때 도출된 앱이었다. 이는 사용자가 영어 학습 앱을 찾기 위해 사용할 수 있는 키워드 중 가장 간단한 키워드가 “영어 공부”이고, 동시에 가장 다양한 앱 검색 결과를 도출해 낼 수 있을 것이라는 연구자의 판단에 따른 것이다. 2021년 7월1일 기준으로, ‘영어 공부’라는 키워드 사용하여 총 250개의 앱 검색 결과를 얻을 수 있었고, 이 중 리뷰가 존재하지 않는 25개의 앱을 제외한 225개의 영어 학습 앱을 본 연구의 분석 앱으로 선정하였다. 본 연구에서 선정한 앱 목록과 기타 시각화 자료(그림)는 저자의 구글 드라이브에 공개되어 있다. (https://drive.google.com/drive/folders/1xv7HWf44GM4T2eCj9gmDup8Gw2dLwu_C?usp=sharing).

3-2 데이터 수집, 전처리, 그리고 특징 추출



* The distribution of the Korean keywords is shown instead of its English one as translation could change the output.

그림 2. 특징 추출 시각화

Fig. 2. Feature Extract Visualization

영어 학습 앱 사용자 리뷰를 수집하기 위해 오픈 소스 프로그램 Python 3.7.9와 웹 크롤링에 대표적으로 사용되는 라이브러리 중 하나인 Selenium을 사용하였다. 구글 플레이 스토어(<https://play.google.com/store/apps>)는 사용자의 검색 결과에 따른 앱 목록을 반환해준다. 사용자 리뷰 수집은 2021년 07월 14일 부터 시작하여 2021년 07월 17일에 종료되었고 총 225개 앱, 246,073개의 사용자 리뷰를 수집하였다. 사용자 리뷰는 대표적인 비정형 데이터 중 하나로 이를 연구하기 위해서는 전처리 과정(preprocessing)이 필수적이다. 전처리 과정은 다음과 같다. 우선, 잘못된 띄어쓰기에 따른 부정확한 텍스트 전처리의 가능성을 줄이기 위해 한국어 전처리 패키지 중 하나인 PyKoSpacing(<https://github.com/haven-jeon/PyKoSpacing>)을 이용해 사용자 리뷰의 띄어쓰기 교정을 실시하였다. 이후 정규식을 이용해 텍스트 교정이 가능한 re.sub을 이용해 특수문자와 같은 특정 문자열을 제거했고 마지막으로 Konlpy의 Okt.normalize 기능을 사용해 더 자세한 정규화를 시행하였다. 이를 통해 문법에 맞지 않는 단어나 오타를 포함하고 있는 단어를 정규화 하였다. 그 뒤 명사 상위 500개, 형용사 상위 500개를 추출하였다. 이후 전체 리뷰의 키워드 분포를 간단히 확인하기 위해 문서의 유사도를 구하는 작업에 많이 쓰이는 TF-IDF(Term Frequency-Inverse Document Frequency)를 특징추출을 위해 구축하였고 t-SNE(t-Stochastic Neighbor Embedding)를 이용하여 <그림 2>와 같이 시각화하였다.

3-3 감성분석 모델을 통한 사용자 리뷰 분류

본 연구에서는 대표적 한국어 감성 말뭉치인 네이버 영화 리뷰 감성 말뭉치(<https://github.com/e9t/nsmc>)를 이용해 감성 분석 모델을 구축하였다. 네이버 영화 리뷰 감성 말뭉치는 총 20만개의 사용자 리뷰로 이루어져 있고 긍정 리뷰는 1,

부정 리뷰는 0으로 각각의 리뷰에 태그가 되어 있다. 이 말뭉치를 이용, 약 85%의 정확도를 가진 LSTM(Long Short Term Memory)기반 감성 분석 모델을 구축하였다. LSTM은 RNN(Recurrent Neural Network)의 장기 의존성 문제를 개선하여 자연어 처리에서 주로 활용되는 모델이다. 이후 해당 감성 분석 모델을 이용해 앞서 정규화가 진행된 약 25만개 사용자 리뷰에 대해 긍정과 부정 사용자 리뷰로 분류하였다.

3-4 LDA 토픽 모델링

영어 학습 앱 사용자 경험에서 도출된 긍정과 부정의 주요 요인을 추출하기 위해 LDA(잠재 디리클레 할당) 알고리즘을 이용한 LDA 토픽 모델링 분석을 시행하였다. 이를 위해 자연어 처리 패키지 Mecab과 토픽 모델링에 유용한 Gensim을 사용하였다. 이때 실제 주제(Topic)를 내포하고 있는 명사만을 LDA 토픽 모델링에 이용했으며 이를 위해 mecab.nouns()를 활용하였다. 이와 함께 불필요한 단어를 제거하기 위해 불용어(stopword) 사전을 구축하여 활용하였다. 이때, 분석 대상이 영어 학습 앱이기 때문에 기존에 활용되던 불용어 사전을 적용하기는 어려웠다. 이를 해결하기 위해 앞서 추출한 명사 상위 500개와 형용사 상위 500개를 참고하여 “구글”, “안드로이드”, “갤럭시”와 같은 단어를 불용어 사전에 추가하였다. LDA분석에서 토픽 수(K)는 사용자가 지정하는 하이퍼 파라미터이다. 최적의 토픽 수는 LDA 확률 모델이 얼마나 잘 예측하고 있는지 알려주는 혼란도(Perplexity) 점수를 참고하거나[16] 각 토픽 내에 얼마나 일관성 있는 단어가 모여 있는지 알 수 있는 토픽 일관성(Topic Coherence) 점수를 참고하는 방법이 있고 데이터에 특징에 따라 연구자가 주관적으로 토픽수를 정할 수도 있다. 본 연구에서는 혼란도 점수와 토픽 일관성 점수를 모두 만족하는 토픽 수를 도출 하기위해 LDA 모델의 최적 학습 횟수(Passes)를 먼저 구하고 이후 최적 토픽수를 구한 뒤 도출된 LDA 결과에서 중복으로 출현하는 키워드들을 불용어 사전에 추가하여 필터링하여 결과 도출 및 재구성의 과정을 반복하였다. 이를 통해 도출된 긍정 사용자 리뷰의 최적 학습 횟수는 30회, 토픽 수는 6개로 도출되었고 부정 사용자 리뷰의 최적 학습 횟수는 10회, 최적 토픽 수는 6개로 도출되었다.

IV. 연구결과

4-1 감성분석결과

네이버 영화 리뷰 감성 말뭉치를 이용해 만들어진 감성 분석 모델로 텍스트를 분류한 결과 전체 246,073개 리뷰 중 긍정 리뷰 168,577개(68.5%)와 부정 리뷰 77,497개(31.5%)의 결과를 도출할 수 있었다. 긍정 리뷰의 경우 ‘핵심을 잘 집어준다.’, ‘출근길에 사용하기 좋다.’, ‘초보자에게도 큰 도움

이 된다.’는 내용이 포함되었고 부정 리뷰의 경우 ‘학습에 흥미가 생기지 않는다.’, ‘가격이 너무 비싸다.’, 그리고 ‘학습 기능에 제약이 너무 많다.’와 같은 내용이 포함되었음을 확인할 수 있었다. 과반수 이상의 리뷰가 긍정을 표하고 있는 것으로 보아 현 시점의 영어 학습 앱 사용자 경험은 다수의 사용자들에게 긍정적인 것으로 보인다.

4-2 토픽 모델링 결과

LDA 토픽 모델링분석을 통해 최종적으로 도출된 토픽은 총 12개로, 긍정 토픽과 부정 토픽 각각 6개씩의 토픽이 도출되었다. 이때 긍정 토픽의 경우 단어 암기, 회화, 예문, 게임, 강의 등의 내용으로 이루어져 있었고 부정 토픽은 콘텐츠, 오작동, 결제, 기능, 가격이라는 내용을 포함하였다.

<표 1>은 긍정 리뷰 토픽 모델링 결과이다. 긍정 리뷰의 첫 번째 토픽인 토픽 1은 ‘단어장’, ‘뜻’, ‘영상’, ‘화면’, ‘검색’ 등의 단어를 포함하고 있어 ‘단어 암기’로 정의하였다. 토픽 2는 ‘학습’, ‘회화’, ‘표현’, ‘연습’, ‘대화’ 등의 단어로 구성되어 ‘일상 회화’로 정의하였다. ‘예문 활용’으로 정의된 토픽 3은 ‘도움’, ‘기초’, ‘활용’, ‘예문’, ‘생활’ 등의 단어를 포함하고 있었다. 토픽 4는 ‘학원 강의’로 정의하였고 토픽 5는 ‘게임’, ‘초보’, ‘흥미’, ‘친구’, ‘문제’ 등의 단어를 포함하고 있어 ‘개인 게임 성취 모형’이라 정의하였다. 마지막 토픽 6은 ‘무료’, ‘유료’, ‘구매’, ‘처음’ 등의 단어를 포함하고 있어 ‘과금 구조’로 정의하였다.

<표 2>는 부정 리뷰 토픽 모델링 결과로, 토픽 1은 ‘실력’, ‘콘텐츠’, ‘수업’ 등을 포함한 ‘콘텐츠’로, 토픽 2는 ‘반복’, ‘오류’, ‘문제’ 등을 포함한 ‘오작동’으로 정의하였다.

표 1. 긍정 리뷰 LDA 토픽 모델링 결과
Table 1. Result of LDA Topic Modeling (Positive Review)

Topic Number	Definition	Keywords
Topic 1	Memorization	Vocabulary Lists, Addition, Meaning, Video, Part, Advantage, Screen, Context, Search, Choice
Topic 2	Daily Conversation	Learning, Repeat, Conversation, Useful, Expression, Effective, Practice, Memory, Talk, Variety
Topic 3	Sentence Usage	Help, Needs, Satisfied, Basic, Utilization, Understanding, Example Sentence, Living, Review, Composition
Topic 4	Academical Lecture	Memorization, Start, A day, Explanation, Class, Teacher, Burden, Lecture, Download, Academy
Topic 5	Game Achievement Model	Game, Beginner, Interests, Language, Friend, Voice, Question, Head, In Favor, Card
Topic 6	Business Model	Free, Paid, Ability, Use, Purchase, First, Awesome, Convenience, Review, Developer

표 2. 부정 리뷰 LDA 토픽 모델링 결과
Table 2. Result of LDA Topic Modeling (Negative Review)

Topic Number	Definition	Keywords
Topic 1	Contents	Ability, Game, Update, Start, Feeling, Head, Contents, Class, Irritation, Comprehension
Topic 2	Error	Repeat, Learning, Practice, Day, Error, Memorization, Meaning, Sentence, Problem, Video
Topic 3	Purchase Issue	Cancel, Refund, Automation, Login, Subscription, Purchase, Sign up, Confirmation, Approval, Install
Topic 4	Entry Barriers	Game, Beginner, Sound, First, Person, Conversation, Login, Setting, Test, Recoding
Topic 5	Feature	Useful, Help, Screen, Inconvenience, Problem, Vocabulary Lists, Addition, Use, Video, Play
Topic 6	Price	Free, Paid, Recognize, Step, Use, Academy, Voice, Interests, Standard, Card

토픽 3은 ‘환불’, ‘구매’, ‘결제’ 등을 포함하여 ‘결제 문제’로, 토픽 4는 ‘초보자’, ‘테스트’, ‘로그인’ 등을 포함하여 ‘진입 장벽’으로, 그리고, 토픽 5는 ‘화면’, ‘불편’ 등을 포함하여 ‘기능 관련’으로 정의하였다. 토픽 6은 ‘무료’, ‘유료’, ‘이용’, ‘학원’ 등을 포함하여 ‘가격’으로 정하였다.

4-3 토픽모델링 시각화

LDavis는[17] LDA 모델의 학습 결과를 시각적으로 표현하는 라이브러리이다. 토픽의 크기에 따라 전체 문서에서 해당 토픽의 비중을 알 수 있고 위치를 통해 각 토픽들의 독립성 또한 알 수 있다. 각 토픽별 거리가 가까우면 의미상 유사한 토픽이고 거리가 멀면 의미상 유사하지 않은 토픽이다.

<그림 3>은 긍정 리뷰 토픽 모델링 결과를 시각화한 것이다. 이때 토픽 1 ‘단어 암기’와 토픽 5 ‘개인 게임 성취 모형’의 거리가 가까운 것을 보아 ‘개인 게임 성취 모형’은 ‘단어 암기’와 밀접한 관련성을 띄고 있음을 알 수 있고 토픽 3 ‘예문 활용’과 토픽 4 ‘학원 강의’의 거리가 가까우므로 두 토픽의 관련성이 높은 것을 알 수 있다. 토픽 2와 토픽 6은 같은 사분면에 위치하지만 거리가 앞선 토픽들에 비해 먼 것을 알 수 있는데 ‘일상 회화’로 정의된 토픽 2와 ‘과금 구조’로 정의된 토픽 6에 큰 관련성이 존재하지는 않음을 알 수 있다.

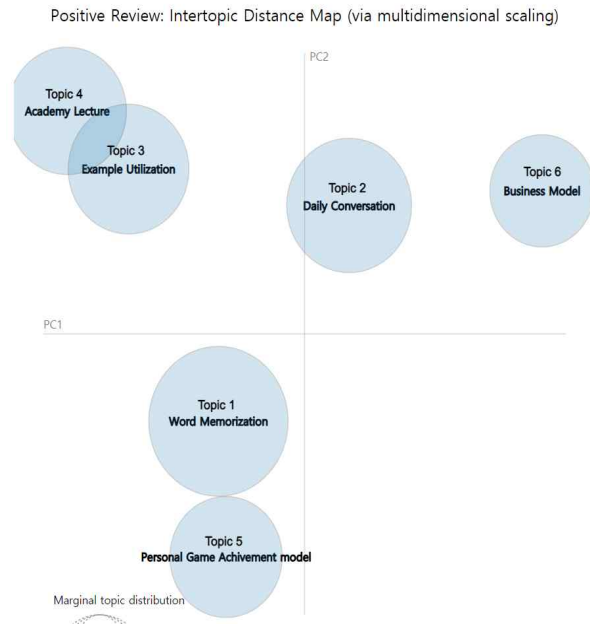


그림 3. 긍정 토픽 모델링 시각화
Fig. 3. Positive Topic Modeling Visualization

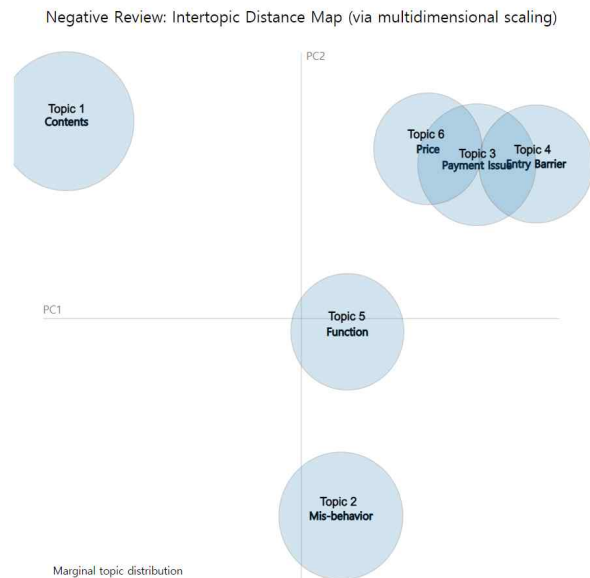


그림 4. 부정 토픽 모델링 시각화
Fig. 4. Negative Topic Modeling Visualization

<그림 4>는 부정 리뷰 토픽 모델링 결과를 시각화한 것이다. 이때 앞서 본 긍정 리뷰 토픽의 경우 두 개의 토픽이 각각 쌍을 이루고 있었던 것과 달리 세 개의 토픽이 매우 밀접한 연관성을 보이고 나머지 세 개의 토픽은 어느 정도 독립성을 보이고 있음을 알 수 있다. 거리가 가깝게 뭉쳐있는 토픽 3, 4, 6은 각각 ‘결제 문제’, ‘진입 장벽’, ‘가격’으로 정의가 되었는데, 앱 과금 구조 관련하여 밀접한 관련성을 가지고 있음을 알 수 있다. 같은 사분면 안에 위치한 토픽 2와 토픽 5는 각각 ‘오작동’, ‘기능 관련’으로 정의 되었는데, 앱의 원활한 작동과

관련성이 있음을 알 수 있다. 그리고 토픽 1의 경우 ‘콘텐츠’으로 위치상 떨어져 있어, 앞선 토픽들과 크게 독립된 토픽임을 알 수 있다.

V. 결 론

5-1 연구 요약 및 시사점

본 연구는 교육 분야의 구글 플레이 스토어 영어 학습 앱의 사용자 리뷰를 분석하여 만족, 불만족 요인을 파악하고, 향후 사용자 경험을 바탕으로 한 ‘앱 개선, 마케팅’ 등에 기여하고자 하였다. 구체적으로, 영어 학습 앱의 사용자 리뷰를 분석하기 위해서는 텍스트 마이닝 기법들 중 딥러닝 모델을 활용한 감성 분석을 진행했고 분류된 리뷰 데이터에 LDA 토픽 모델링 방법을 적용하였다.

연구 결과, 긍정 리뷰에서 첫 번째 토픽에 해당하는 단어 암기, 두 번째 토픽에 해당하는 일상 회화에 관한 긍정적 평가가 두드러졌음을 알 수 있었다. 동시에 예문 활용이라 정의한 세 번째 토픽과 개인 게임 성취 모형이라 정의한 다섯 번째 토픽에 대한 긍정적 평가 역시 이루어졌음을 알 수 있었다. 이러한 토픽들은 현시점의 영어 학습 앱이 가진 강점들이며 특히 영어 학습과 직결되는 단어 암기와 일상 회화에 대한 부분은 더욱 발전시켜 영어 학습 앱의 특징 점으로 삼을 수 있을 것으로 전망된다.

반면에, 부정 리뷰에서는 첫 번째 토픽에 해당하는 콘텐츠, 두 번째 토픽에 해당하는 오작동에 대한 부정적 평가가 두드러졌음을 알 수 있다. 동시에 기능 관련이라 정의한 다섯 번째 토픽을 통해 영어 학습 앱의 각종 기능에 대해 사용자들이 부정적 평가를 내비쳤음을 알 수 있다. 콘텐츠는 사용자들의 흥미를 지속적으로 유지하고 동시에 더 많은 사용자를 불러 모을 수 있는 요소이기에 영어 학습 앱 개발자들의 빠른 대응이 필요할 것으로 보인다. 동시에 오작동 관련한 토픽은 영어 학습 앱 사용자 경험에 있어 치명적일 수 있는 내용이기에 적극적인 대응 방안을 마련해야 할 것이다. 긍정 및 부정 리뷰의 여섯 번째 토픽은 각각 과금 구조와 가격이었는데 유사한 토픽이 긍정 및 부정 리뷰에서 동시에 도출된 것을 통해 사용자들의 평가가 중립적임을 알 수 있었다. 사용자의 결제는 곧 수익으로 직결되는 중요한 요소이지만 실제 사용자들은 영어 학습 앱에 결제하는 행위에 대해 중립적인 모습을 보여 사용자들이 과금을 통해 더 긍정적인 경험을 할 수 있는 방안을 마련해야 할 것이다.

본 연구의 시사점은 아래와 같다. 첫째, 영어 학습 앱에 대한 사용자 경험을 분석해 영어 학습 앱에 대한 개선점 및 마케팅 시사점을 도출해냈다는 점에서 의의가 있다. 기존 영어 학습 앱과 관련된 연구들은 대부분 영어 학습 앱을 통한 개인의 성취도 향상에 주목하였다. 하지만 다른 분야의 앱 관련 연구는 해당 분야 앱의 개선점 및 마케팅 시사점에 주목하여

앱의 개선을 주된 목적으로 삼았다. 이와 같이 본 연구는 영어 학습 앱 자체의 개선을 목적으로 하는 사용자 리뷰 연구를 진행하였다는 점에서 학문적 의의가 있다.

둘째, 본 연구는 감성 분석 과정을 텍스트 마이닝 기법 중 하나인 딥러닝 모델을 통한 감성 분석으로 진행하였다는 점에서 의의가 있다. 기존의 연구들은 사용자 리뷰의 감성을 별점으로 예측하는 방식을 사용했는데 최근 연구에 따르면 실제 리뷰 내용의 감성과 별점을 통한 감성은 편향이 있을 수 있다고 한다[14]. 본 연구는 실제 리뷰 내용에 기반을 두어 사용자의 감성을 예측 및 분류할 수 있는 딥러닝 모델을 활용한 감성 분석을 진행하여 리뷰 내용에 기반을 둔 분류를 진행하였다는 점에서 학문적 의의가 있다.

셋째, 본 연구에서 도출된 영어 학습 앱 리뷰의 긍정 및 부정 요소는 실무적으로 활용이 가능하다. 먼저, 영어 학습 앱 제공자는 본 연구에 사용된 리뷰의 긍정 요소를 마케팅 전략 수립과 미래 콘텐츠 개발 전략 수립에 활용할 수 있다. 예를 들어, 단어 암기와 일상 회화 학습 기능은 기존 영어 학습 앱 사용자들이 긍정적으로 평가한 요소이다. 이를 강점으로 삼아 해당 니즈를 가진 잠재적 소비자 군에 특화된 마케팅을 전개해 기업은 소비자의 구매력을 자극할 수 있는 기회를 얻을 수 있다. 다섯 번째 토픽에 해당하는 개인 게임 성취 모형은 게임, 흥미, 친구 등의 단어를 포함하고 있는데 이는 곧 개인들의 경쟁을 통해 성취도를 높이는 방식이라 정의할 수 있다. 실무자는 타인과의 경쟁을 통한 자신의 성취를 목표로 하는 온라인 게임에 열광하는 사용자층에게 이를 어필해 소비자의 추가 유입을 기대할 수 있고 추가적인 화제성을 기대할 수 있다.

또한, 리뷰에 나타난 부정 요소는 영어 학습 앱을 개선하는데 활용할 수 있다. 예를 들어, 부정 리뷰의 첫 번째 토픽인 콘텐츠의 경우 업데이트, 수업 등의 단어를 가지고 있는 바, 기존 사용자들이 영어 학습 앱의 콘텐츠 업데이트에 관한 불만사항을 가지고 있음을 알 수 있다. 이는 사용자들에게 제공되는 콘텐츠가 적기 때문일 수도 있지만, 동시에 사용자들의 콘텐츠 소모 속도가 빠르기 때문일 수도 있다. 이를 통해, 콘텐츠의 공급량을 점검하고 콘텐츠의 공급 또는 소모 방식의 개선을 모색할 수 있다. 이는 곧 서비스의 지속성 측면에서 개선으로 이어질 수 있다. 동시에 오작동이라 정의된 두 번째 토픽을 통해 기존 사용자들이 영어 학습 앱이 오작동하여 정상적인 사용이 불가능한 것에 대해 불만이 있음을 알 수 있었다. 영어 학습 앱 제작자는 이와 같은 문제점을 신속히 파악하고 사용자들의 불만 요인에 즉각적으로 대응하여 사용자들의 신뢰 회복에 힘써야 하고, 기존 사용자 유출을 막아야 한다. 이미 구매력을 한번 행사한 사용자는 미래에 다시 한 번 구매하기 쉬워서 충성 고객이 될 수 있는 대상이기 때문이다. 따라서, 영어 학습 앱 제작자들은 본 연구에서 도출된 사용자 리뷰 부정 요소를 바탕으로 앱을 개선할 수 있을 것이다.

5-2 연구의 한계 및 향후 연구 방향

본 연구가 지닌 한계점은 다음과 같다. 첫째, LDA 토픽 모델링 기법의 경우 모수 통계라는 특성 때문에 데이터가 가지는 토픽의 수(K)를 연구자가 직접 설정해야 하는 한계점이 있다. 이때 데이터가 몇 개의 주제를 가지고 있는지 연구자가 미리 알기는 어렵기 때문에 혼란도를 뜻하는 Perplexity 값을 기준으로 K 값을 선정하거나 여러 K 값을 도입한 반복적인 실험을 통해 K 값을 선정하는 절차를 거친다. 이러한 과정에서 실제 데이터가 가진 토픽의 수보다 많거나 적은 K 값을 설정할 수도 있기에 결과가 부정확할 수도 있다. 향후 연구에서는 이를 개선한 기법을 적용할 수 있을 것이다.

둘째, 딥러닝 감성 분석 모델의 기반 학습 데이터에 따라 감성 분석 결과에 차이가 생기는 한계점이 있다. 딥러닝 감성 분석 모델은 학습 데이터로 제공된 말뭉치 속의 문서와 그 문서에 태깅된 감성 분류를 활용해 학습하고 이를 바탕으로 미래에 입력된 문서에 대한 감성을 예측해 이진 분류하는 방식으로 만들어진다. 이때 학습 데이터가 입력된 문서와 정 반대의 주제를 가지고 있다면 딥러닝 모델의 정확도가 떨어질 가능성이 존재한다. 본 연구에서는 네이버 영화 리뷰 코퍼스를 사용하여 학습시킨 딥러닝 모델을 영어 학습 앱 리뷰 감성 분석에 활용하였다. 향후 연구에서는 앱 리뷰와 관련된 코퍼스를 활용한 딥러닝 모델을 통해 이를 개선할 수 있을 것이다.

셋째, 긍정 리뷰와 부정 리뷰로 나누어 진행한 토픽 모델링 분석 결과에서 서로 상반된 의미의 키워드가 검출되었다. 이는 긍정과 부정 의견을 함께 제시한 사용자 리뷰에 따른 것으로 추정되며 향후 더욱 고도화된 감성 분석 기법을 적용하여 이를 개선할 수 있을 것이다.

넷째, 본 연구의 앱 사용자 리뷰 분석 결과를 모든 영어 학습 앱에 적용하기는 어려움이 있다. 본 연구에서는 구글 플레이 스토어에서 ‘영어 공부’를 검색하였을 때 도출된 검색 결과 앱 250개에서 리뷰가 존재하는 225개 앱의 리뷰 246,073개를 분석하였다. 영어 학습 앱의 경우 그 분야가 넓기에 향후 연구에서는 더욱 다양한 검색 키워드를 활용해 더 다양한 영어 학습 앱의 리뷰를 수집하거나, 애플의 앱 스토어의 영어 학습 앱을 대상으로 사용자 리뷰 분석을 확대할 수 있다. 이렇게 본 연구의 한계는 후속 연구를 통해서 개선 및 보완될 수 있을 것이다. 본 연구가 미래의 영어 학습 앱을 통한 다양한 산업분야의 선행 연구로서 학문적 발판이 될 수 있기를 기대한다.

참고문헌

- [1] MOIBA, 2019, Status of Mobile Contents Industry in South Korea, MOIBA, Seoul, Mobile Content Industry Report 2019, 2019.
- [2] Gallup Korea, Smartphone Usage & Brand 2012-2020, Gallup Korea, Seoul, Gallup Report G20200812, 2020.
- [3] J. H. Joe, “The Effects of Various Learning Methods and Activities in Learning English on Learners’ Satisfaction and

Perceptions”, The Journal of Humanities and Social science, Vol. 9, No. 3, pp. 1059-1074, June 2017.
<http://dx.doi.org/10.17053/jcc.2021.24.1.004>

[4] E. H. Song, “Preference and Usefulness of Smartphone English Apps for English Study of EFL University Learners”, The Journal of Foreign Studies, Vol., No. 29, pp. 95-124, Sep 2014. <http://doi.org/10.15755/jfs.2014..29.95>

[5] I. S. Kil, “The Relationship between Smartphone-Based Learning and TOEIC Scores”, The Journal of Linguistics Science, Vol. 95, No., pp 1-20, Dec 2020.
<http://dx.doi.org/10.21296/jls.2020.12.95.1>

[6] J. L. Hong and M. R. Yu and B. R. Choi, “An Analysis of Mobile Augmented Reality App Reviews Using Topic Modeling”, Journal of Digital Contents Society, Vo.1 20, No. 7, pp 1417-1427, Jul 2019.
<http://dx.doi.org/10.9728/dcs.2019.20.7.1417>

[7] S. M. Lee and T. W. Lee, “Establish Marketing Strategy Using Analysis of Local Currency App User Reviews – Focused on ‘Dongbackjeon’ and ‘Incheonum’”, The Journal of the Korea Contents Association, Vol. 21, No. 4, pp 111-122, Apr 2021.
<https://doi.org/10.5392/JKCA.2021.21.04.111>

[8] G. B. Park and S. W. Lee, “Examining user needs for mHealth apps based on text mining of user reviews”, in Proceedings of 2020 Spring Conference of ESK, Jeju-Island: Jeju-Island, pp 85-89, June 2020.
<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NOD E09412385>

[9] J. H. Jung and H. I. Chung and Z. K. Lee, “An Analysis of Mobile Food Delivery App ”Baemin“ by Using Text Mining and ARIMA Model”, Journal of Digital Contents Society, Vol. 22, No. 2, pp 291-299, Feb 2021.
<http://dx.doi.org/10.9728/dcs.2021.22.2.291>

[10] Y. Guo, S. J. Barnes, and Q. Jia, “Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation,” Tourism Management, Vol. 59, No., pp.467-483, Apr 2017.
<http://dx.doi.org/10.1016/j.tourman.2016.09.009>

[11] M. Chau and J. Xu, “Business intelligence in blogs: Understanding consumer interactions and communities,” MIS quarterly, Vol. 36, No. 4, pp.1189-1216, Dec 2012.
<http://dx.doi.org/10.2307/41703504>

[12] S. J. Hwang, Y. K. Lee, and Y. S. Kim, “Hallym Systematic Analyzer of Korean (H-SAK) ver.1.5”, in Proceedings of Korea Computer Congress 2021, Jeju-Island:Jeju-Island, pp 316-318, June 2021.
<https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NO DE10582917>

[13] M. Song, Text Mining, Chunglam Publishing Co., pp. 112-117, 2017.

[14] S. Sadiq and M. Umer and S. Ullah and S. Mirjalili and V. Rupapara and M. Nappi, “Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning”, Expert Systems with Applications, Vol. 181, Article 115111, Nov 2021.
<https://doi.org/10.1016/j.eswa.2021.115111>

[15] D. M. Blei, “Probabilistic topic models,” Communications of the ACM, Vol. 55, No. 4, pp. 77-84, Apr 2012.
<https://doi.org/10.1145/2133806.2133826>

[16] T. L. Griffiths, and M. Steyvers, “Finding scientific topics”, Proceedings of the National Academy of Sciences, Vol. 101, No. 1, pp. 5228-5235, Apr 2004.
<https://doi.org/10.1073/pnas.0307752101>

[17] C. Sievert & K. E. Shirley, “LDAvis: A method for visualizing and interpreting topics”, Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Vol., pp. 63-70, June 2014.
<http://dx.doi.org/10.3115/v1/W14-3110>



김기연(Gi-Youn Kim)

2022년 : 한림대학교 인문대학
 영어영문학과/미래융합스쿨
 디지털인문예술전공 (학부)

2022년~현 재: 한림대학교 융합소프트웨어학과 석사과정
 ※ 관심분야 : 자연언어처리(Natural Language Processing)



한수미(Sumi Han)

2004년 : 서울대학교 인문대학
 영어영문학과(학부)
 2009년 : 서울대학교 인문대학
 영어영문학과 대학원
 (문학석사-영어학)
 2015년 : 노던애리조나대학, 미국
 (응용언어학 박사-
 코퍼스언어학, 언어평가)

2016년 : 서울대학교 대학영어프로그램 강사
 2017년~현 재: 한림대학교 미래융합스쿨 디지털인문예술전공/
 인문대학 영어영문학과 조교수
 ※ 관심분야 : 코퍼스언어학(Corpus Linguistics), 언어평가
 (Language Testing), 디지털인문학(Digital Humanities)