

코로나19 팬데믹 상황에서 감성분석을 이용한 미국, 중국, 한국 여행자의 온라인 리뷰 비교 분석*

홍준우** · 홍태호***

A Comparative Analysis of Travelers' Online Reviews among
China, USA, and South Korea using Sentiment Analysis in the
Era of the COVID-19 Pandemic*

Junwoo Hong** · Taeho Hong***

■ Abstract ■

In this study, we performed a comparative analysis of the sentiment value for the tourists in USA, China, and Korea on the COVID19 pandemic era to explore and find out the features of the tourists by using online reviews. We collected a total of 243,826 online hotel reviews for metropolitan city and vacation spot in the three countries to compare the features between the business and the vacation trips. We collected the online reviews into the tow groups from Jan. 1, 2019 to Nov. 31, 2019 for before COVID19 pandemic and from Apr. 1, 2020 to Deb 28, 2021 for during COVID19. Online reviews were categorized into 6 dimensions using LDA model. Sentiment analysis were presented for 6 dimensions by utilizing a lexicon base. We proposed an approach to analyzing the importance of each attribute by applying 6-dimensional sentiment values to conjoint analysis. Our empirical analysis showed that the proposed approach could explore and find out the changed features of travelers during the COVID19 pandemic.

Keyword : COIVD-19 Pandemic, Online Review Analysis, Topic Modeling, Sentiment Analysis, Conjoint Analysis, Smart Tourism

1. 서 론

중국 우한시에서 최초로 발견된 코로나19(COVID-19)는 2019년 12월 31일 세계보건기구(WHO)에 처음 보고되어 확산되었으며 세계 각 국가를 비상사태로 만들었다. 이후 2020년 3월 11일 세계보건기구 사무총장은 중국 이외의 지역에서 확진자, 사망자, 영향을 받는 국가 수가 증가했으며 향후 더 증가할 것이라는 말과 함께 세계적인 팬데믹(pandemic)을 선포하였다(WHO, 2021).

한편 전 세계적으로 코로나19의 확산은 인명피해, 경제적 피해, 정서적 피해 등의 문제점을 일으키고 있다. 우선 인명피해로 2021년 7월 2일을 기준으로 한국의 코로나19 확진자 수는 158,549명, 사망자 수는 2,024명, 미국 확진자 수는 33,317,803명, 사망자 599,089명, 중국 확진자 수는 91,810명, 사망자 4,636명을 기록했다(중앙방역대책본부 2021년 7월 3일). 한국의 경우 코로나19 확진자 수가 초기에는 하루 1명씩 증가했지만 2020년 2월 19일 이후에는 하루 평균 320명씩 증가하는 모습을 보였다(김남순, 2020).

경제적 피해를 살펴보면 코로나19는 전 세계적으로 유례없는 경제위기를 발생시켰고 2020년에는 전세계 92.9%의 국가에 경제적 피해를 입혔다. 이는 1870년 천연두 팬데믹 시기의 40.9%, 20세기 대공황 시기의 83.3% 보다 큰 규모이다(홍태희, 2020). 세계관광기구(WNTO)는 각 국가별 여행제한정책으로 인해 관광 수입은 최소 9,100억 달러에서 최대 1조 1,700억 달러 감소하고 관광 부문 일자리 또한 최소 1억 개에서 최대 1억 2,000만 개 감소했다고 보고한다(정대영, 이수진, 2020). 또한 김동준 외(2020)는 코로나19가 항공, 숙박, 기타 서비스업에 직접적인 영향을 끼쳤으며 2020년 3월 기준으로 전년 동기대비 국제 여객 이용은 93.5%, 국제공항 이용은 91.6% 감소하여 2020년 상반기 국내 항공사의 매출 피해 전망은 6조 3,000억 원으로 전망하고 있다. 이는 항공사의 파산과 국제항공 네트워크의 붕괴를 야기할 수 있다.

정서적 피해로는 코로나19의 확산 방지를 위한 출입국 제한조치, 사회적 거리두기 일환인 집합금지, 외출

자제 등 방역조치 차원에서 외출활동이 위축되면서 사회적 우울 현상으로 코로나 블루가 발생하고 있다(손현일 외, 2020). 코로나19 확산으로 인해 전문의료진, 일반인 모두 불안, 우울증, 스트레스 등의 심리적 문제를 겪고 있다(Liu et al., 2020). Xiang et al.(2020)은 코로나19 의심환자는 잠재적인 감염 결과에 대한 두려움을, 격리된 환자는 지루함, 외로움, 분노 등의 감정을 경험하고 있다. 따라서 본 연구에서는 코로나19로 인해 급격하게 변해가는 시대적 상황을 이해하고 직접적인 대응을 할 수 있는 방안을 고찰하고자 한다.

한편, 정보통신 기술과 전자기기 및 인터넷의 보급으로 인터넷 접속이 용이해지고, 온라인을 통한 의견 공유가 빈번하게 이루어지고 있다. 온라인을 통한 의견 공유는 온라인 리뷰의 형태로 고객의 구매 및 사용 경험을 공유하는 것으로 다양한 전자상거래 플랫폼에서 발생하고 있고 이는 전자상거래 시장을 성장시켜 이용률을 증가시키는 원인이 되고 있다. 최근 온라인 리뷰는 지속적으로 증가하는 추세이며(Kim, 2020), 코로나19의 확산으로 더욱 가속화되고 있다(Al-maaitah et al., 2021). 또한 온라인 리뷰는 정보 탐색 및 구매에 영향을 미치는데(Yao et al., 2020), 고객은 상품의 탐색, 구매 등 구매에 대한 의사결정의 이점을 제공하며(Cheng and Ho, 2015), 기업은 상품에 대한 리뷰 분석을 통해 매출 변동에 대한 근거를 파악할 수 있어 온라인 리뷰를 적절하게 활용하는 것은 중요하다(Li et al., 2013).

한편 텍스트 마이닝 기법을 적용한 연구들이 일부 진행되고 있는데, Tong and Zhang(2016)은 적은 수의 데이터에 LDA를 적용하여 Wikipedia와 Twitter의 토픽을 추출하였으며 Putri and Kusumaningrum(2017)은 비정형 데이터에서 토픽을 추출하는 방법 중 하나로 LDA를 사용하여 온라인 리뷰 주제의 빈도를 분석하고 중요한 주제를 추출하여 의미를 분석하였다. Xie et al.(2021)은 LDA기법을 적용하여 토픽 주제와 해당 키워드를 추출하여 Weibo에서 시민의 반응을 이해하는 연구를 진행하였으며, Li et al.(2010)은 감성분석을 활용하여 긍정적 의견 혹은 부정적 의견이 표명되는지 여부를 정량화 하였다. 앞선 선행연구를 요약하면

텍스트 마이닝 기법으로 LDA, 감성분석 등을 이용하여 사회적 트렌드 변화를 파악할 수 있다. 온라인 리뷰가 갖는 특징을 파악하고 소비자들이 갖는 상대적 중요도를 확인하기 위해 컨조인트 분석을 사용한 연구들이 진행되고 있는데, 임정수(2013)는 인구사회학적 변인에 따른 중요도 분석에서 최근성, 생산국, 이용기기, 요금 4가지 속성을 성별, 연령대, 소득수준별로 컨조인트 분석을 실시하여 상대적 중요도를 확인했으며, 김다혜, 김주연(2019)은 공간마케팅 활용을 위한 중요도 분석을 위해 매대, 브랜드, 제품색채 3가지 속성과 속성별 3가지 속성수준을 활용하여 상대적 중요도를 확인하였다.

본 연구는 국가별 대표적 대도시 및 관광지에 대한 온라인 호텔 리뷰를 대상으로 토픽 추출 기법인 LDA와 감성 분석을 활용해 코로나19 이전과 이후의 호텔 산업에 대한 변화를 파악하고, 컨조인트 분석을 통해 호텔 리뷰의 6가지 속성별 중요도를 파악해 비즈니스적 함의를 고찰한다. 이를 통해 코로나19 팬데믹 상황에서 호텔 산업의 생존 및 성장을 위한 비즈니스 전략 및 생존전략에 기여할 것으로 기대됨과 동시에 잠재적인 고객들은 더 만족스러운 경험 및 의사결정을 할 수 있을 것으로 기대된다.

2. 이론적 배경

2.1 온라인 리뷰와 6가지 속성

트립어드바이저는 여행 플랫폼으로 여행에 관련된 숙박 시설, 음식점, 체험, 항공권, 크루즈 여행 등에 관한 가격 정보를 비교할 수 있다. 온라인 리뷰는 온라인상에서 상품에 대한 구매 경험을 바탕으로 소비자에 의해 작성된 상품정보로써, 유용성 및 중요성이 커지고 있으며(김근형, 오성열, 2009), 이를 의사결정 과정에 활용되어 중요한 역할을 한다(Gretzel et al., 2020). 소비자에 의해 작성된 온라인 리뷰는 평균점수, 여행 타입, 리뷰 제목 및 내용, 작성날짜 등의 형태로 구성되어 있다. 하지만 트립어드바이저의 숙박 시설에 대한 리뷰는 6가지 속성(가격, 객실,

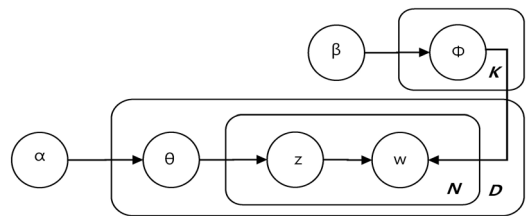
위치, 청결도, 서비스, 침대의 퀄리티) 관점을 추가하여, 선택적으로 리뷰를 작성할 수 있다. [그림 1]은 트립어드바이저에서 숙박 시설의 리뷰를 작성할 때, 고려할 수 있는 6가지 속성이다.



[그림 1] 온라인 리뷰의 6가지 속성

2.2 토픽 모델링(Topic Modeling)

토픽 모델링의 대표적인 방법 중 하나인 LDA(Latent Dirichlet allocation)는 구조화되지 않은 비정형 텍스트 데이터를 분석하는데 널리 사용되고, 주어진 문서에서 단어들의 관련성을 토픽별로 분류하는 확률 기반 모형이다(Blei et al., 2003). LDA는 추출하는 토픽 수(k)를 지정할 수 있는데, 입력하는 토픽 수(k)에 따라 확률 분포 값이 달라지므로, 적절한 k값을 찾아 토픽 추출에 사용하는 것은 중요하다. 서예령 외(2021)는 적절한 토픽 수(k)를 추출하기 위해 일관성(coherence)을 측정하였으며, Hong et al.(2018)은 5겹 교차검증과 혼잡도를 사용하여 최적의 토픽 수를 결정하였다.



[그림 2] LDA 모형

- α, β, Φ : 하이퍼 파라미터 변수 값
- θ : 문서의 토픽의 디리클레 분포
- z : 단어의 토픽 번호
- w : 실제 관찰 가능한 값
- N : 문서에 있는 토픽의 수
- D : 문서의 수

[그림 2]는 LDA 모형으로 문서 생성 과정을 표현하였다. 비정형 텍스트 데이터에 존재하는 유의미한 정보를 추출하기 위해 디리클레 분포를 따르는 토픽의 단어 ϕ 와 문서의 토픽 분포 θ 를 생성하고 θ 에서 각 문서 내 단어에 속한 토픽 번호 z 를 추출하여 해당 토픽의 어휘 분포에서 단어 w 를 추출한다. 토픽모델링은 단어 간의 동시 발생 빈도에 따라 숨겨진 변수를 학습하여 문서의 토픽을 얻는 것이 특징이며 새로운 문서에 대한 경험적 접근없이 쉽게 확률을 적용할 수 있는 분석 기법이다(Blei et al., 2003). Poria et al.(2016)은 온라인 호텔 리뷰를 위치, 서비스, 가치 세 가지 관점을 기준으로 LDA기법을 적용하여 토픽을 추출하였으며, Priyantina and Sarno(2019)는 온라인 호텔 리뷰를 청결도, 편안함, 서비스, 음식, 위치 5가지 관점을 기준으로 LDA 기법을 적용하여 토픽을 추출하였다.

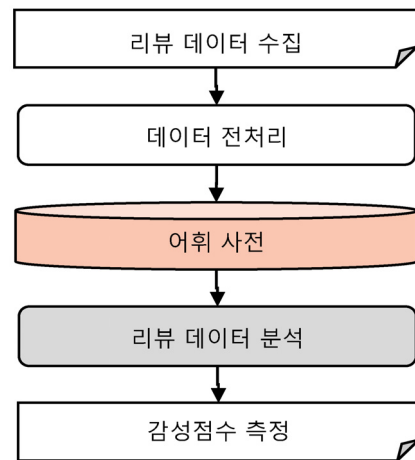
2.3 감성분석(Sentiment Analysis)

감성분석은 오피니언 마이닝이라고 부르며 컴퓨터를 통해 개인의 경험에서 만들어지는 사람의 의견, 태도, 정서를 추출하는 연구 분야이다(Liu, 2012). Medhat et al.(2014)은 온라인 리뷰가 개인의 감정과 특징을 나타내는 것을 증명하였으며 온라인 리뷰를 감성분석에 활용하였다.

한편, 감성분석을 위한 감성점수 측정 방법으로 어휘 기반, 기계학습 기반, 하이브리드 방법이 있다(Alamoodi et al., 2020). 일부 연구에서는 감성분석을 위해 어휘 기반의 방법을 사용하는데, 그 이유는 적용하기 쉽고 다른 연구에서도 널리 사용되기 때문이다(Ma et al., 2018).

[그림 3]은 감성 사전 기반 감성 분석 모형으로 온라인 리뷰 수집, 리뷰 전처리, 감성 사전, 리뷰 분석, 개체 및 관점별 감성 점수 측정 순으로 구성되어 있다. 감성 분석은 비정형 텍스트 데이터에 내포하고 있는 긍정적 혹은 부정적 정서를 추출하여 감정 요소를 분류해 최종적으로 감성 표현의 정도를 정량화한다. Shafqat and Byun(2020)은 관광 리뷰를 분석

하여 높은 평점과 긍정적인 감정 점수를 가진 관광지를 추천하는 추천 시스템을 개발하였으며, Yi Xiaowei(2021)는 리뷰의 평점 및 감성 점수의 차이를 통해 감성 점수를 계산하고, 점수를 사용하여 불일치 여부를 찾았다.



[그림 3] 감성 분석 모형

2.4 컨조인트 분석(Conjoint Analysis)

컨조인트 분석은 consider과 jointly를 결합한 합성어로서 특정 상품(서비스, 정책, 물품, 아이디어 등)이 갖는 속성의 선호도 및 반응을 측정하는 기법이다(Green and Srinivasan, 1990). 일반적으로 상품은 고유의 속성을 가지며 이를 사용하는 사용자에 따라 순위의 차이를 발생시킨다. 이를 컨조인트 분석을 통해 어떤 속성 조합이 더 중요한지를 파악한다. 컨조인트 분석의 목적은 상품이 갖는 속성의 조합이 사람들에게 미치는 영향에 파악하여 속성들 중에서 가장 큰 영향을 미치는 조합을 찾는 방법이다(신윤정 외, 2003). 장성순, 최재영(2018)은 생체인식기술 활용한 서비스에 대한 소비자 선호 분석 연구에서 6가지 속성과 속성별 3가지 속성수준을 활용하여 속성별 중요도를 비교하였으며, 이지선(2019)은 건강 기능식품에 대한 중요도를 5가지 속성과 속성별 2~3가지의 속성수준을 통해 시행하였다.

3. 연구 프레임워크

본 연구에서는 데이터 수집 및 전처리, 토픽모델링을 통한 문장 분류, 빈도 분석, 감성분석, 국가별 비교로 총 5단계로 진행하며 [그림 4]와 같다.

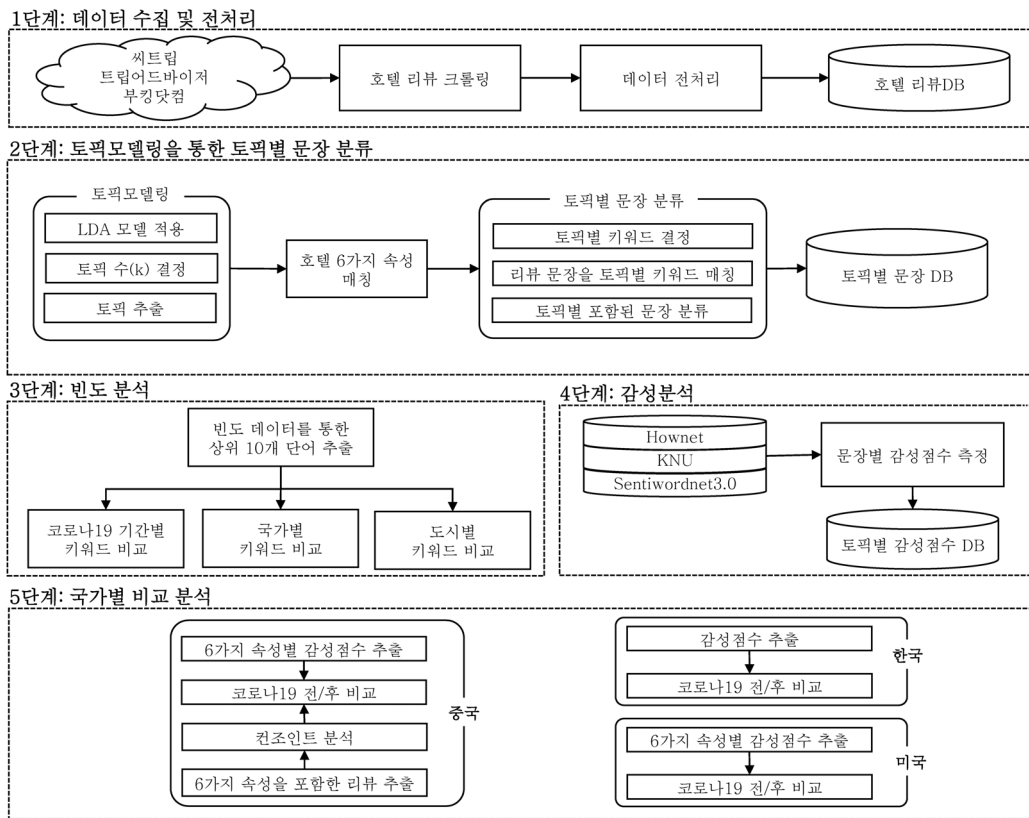
첫 번째 단계는 데이터 수집 및 전처리 과정이다. 데이터 수집은 국가별 대표적 대도시 및 관광지인 중국 베이징과 쑤야, 한국 서울과 제주도, 미국 뉴욕과 하와이를 대상으로 연구를 진행한다. 국가별 호텔 리뷰 데이터는 한국과 미국은 트립어드바이저(tripadvisor.com)와 부킹닷컴(booking.com), 중국은 씨트립(Ctrip.com)에 수집하며, 언어별 특성을 반영하기 위해 중국은 중국어, 한국은 한국어, 미국은 영어로 온라인 리뷰를 수집한다. 수집한 데이터를 언어별 특성에 맞게 리뷰 데이터에 존재하는 불용어

및 결측치를 제거하는 전처리 과정을 수행하고 정제된 리뷰 데이터를 데이터베이스에 저장한다.

두 번째 단계는 토픽모델링을 통한 토픽별 문장 분류 과정이다. 토픽모델링의 LDA기법을 호텔 리뷰 데이터에 적용하며, 토픽 수(k) 만큼의 토픽을 추출한다. 추출한 토픽을 트립어드바이저에서 제공하는 호텔 리뷰의 6가지 속성을 기준으로 매칭하고 토픽별 문장 분류를 위해 토픽별 키워드 결정한다.

그리고 리뷰 문장을 토픽별로 키워드로 매칭하여 토픽별 문장을 분류한다. 이후, 토픽별 데이터베이스에 저장한다.

세 번째 단계는 빈도 분석 과정이다. 전처리 과정을 통해 정제된 호텔 리뷰 데이터를 기간별, 국가별, 도시별로 비교할 수 있도록 자주 발생하는 상위 10개 단어로 추출한다.



[그림 4] 연구 프레임워크

네 번째 단계는 감성분석 과정이다. 온라인 리뷰에 대한 감성 분석은 리뷰어가 표현하는 감정을 텍스트로 구분하는 과정이다(Ganu et al., 2009). 감성 분석을 수행하기 위한 다양한 방법이 있지만 본 연구에서는 각 리뷰를 긍정, 중립, 부정으로 분류하기 위해 간단하고 널리 사용되는 사전 기반 방식인 어휘집을 사용하며, 언어별 특성을 고려하여 한국어 어휘집은 KNU, 영어 어휘집은 Sentiwordnet 3.0, 중국어 어휘집은 HowNet을 사용하여 감성점수를 측정한다. 이후 토픽별 감성점수 데이터 베이스에 감성점수 데이터를 저장한다.

다섯 번째 단계는 국가별 비교 분석 과정이며, 국가별로 코로나19 전/후를 비교 분석한다. 중국은 호텔 리뷰를 6가지 속성별 감성점수를 추출하여 속성별 코로나19 전/후를 비교 분석한다. 추가적으로 6가지 속성을 포함하는 리뷰를 대상으로 컨조인트 분석을 진행하며 속성별 상대적 중요도를 코로나19 전/후로 비교 분석한다. 미국은 중국과 동일하게 6가지 속성별 감성점수를 추출하여, 속성별 코로나19 전/후를 비교 분석하고, 한국은 속성별 감성점수가 부족하여 전체 리뷰 데이터에 대해서 감성점수를 측정하여 코로나19 전/후를 비교 분석한다.

4. 실험 및 결과

4.1 데이터 수집 및 전처리

본 연구에서는 경제규모가 가장 큰 미국과 중국의 대도시 및 관광도시를 대상으로 연구를 진행하고 추

가로 한국의 대도시와 관광지를 비교분석한다. 대표적인 각 나라의 수도와 관광도시를 선정하여 미국의 뉴욕과 하와이, 중국의 베이징과 쑤아, 한국의 서울과 제주도를 대상으로 관광 여행자의 온라인 리뷰를 분석하도록 한다. 2019.01.01~2021.03.31의 기간의 호텔 리뷰 데이터(‘고객ID’, ‘리뷰 날짜’, ‘숙박일’, ‘기여도’, ‘유용성’, ‘평점’, ‘리뷰 제목 및 내용’ 등)를 수집했으며 리뷰 수집 도구로서 Python(버전 3.7.1)의 웹 스크레이퍼를 사용했다. 한국과 미국은 트립어드바이어 부킹닷컴, 중국은 씨트립에서 각각 수집하여 중국은 222,019개, 한국은 5,682개, 미국은 16,125개의 리뷰를 확보했다. 이렇게 수집한 호텔 리뷰 수는 국가별 확진자 수, 코로나19 확산 방지를 위한 국가별 대응 및 전략 등으로 인해 국가별, 도시별 호텔 리뷰 수의 차이가 생겼다.

<표 1>은 국가별, 도시별 수집한 호텔 리뷰 수이다. 표를 통해 미국 도시별 리뷰 수가 코로나19 전/후를 비교했을 때 큰 폭으로 감소했고 한국은 중국과 미국에 비해 리뷰 수가 부족했다. 또한 중국 베이징과 미국 하와이의 평균점수는 코로나19 이전 시기보다 코로나19 이후 시기가 높게 나타났다. 다음 데이터의 전처리 과정으로 리뷰 데이터의 결과에 신뢰성을 낮추고 결과 해석을 어렵게 만드는 웹사이트 링크, 불필요한 용어, 특수문자, 숫자, 공백 등을 제거했으며(Aggarwal and Zhai, 2012; 이인섭, 박준, 2019), 결측치를 제거했다. 그리고 텍스트에 존재하는 불용어를 제거했으며, 미국과 중국은 R 라이브러리 ‘tm’, ‘stopwords’를 사용하여 제거했으며 한국은

<표 1> 국가별 호텔 리뷰 수

국가	중국				한국				미국			
	베이징		쑤아		서울		제주도		뉴욕		하와이	
도시	이전	이후	이전	이후	이전	이후	이전	이후	이전	이후	이전	이후
평균점수	4.74	4.76	4.71	4.68	4.74	4.68	4.57	4.5	4.58	4.34	4.19	4.31
리뷰 수	37,941	19,515	84,165	80,382	1,328	1,597	1,035	1,722	9,661	1,691	4,201	572
총 리뷰 수	57,456		164,563		2,925		2,757		11,352		4,773	

* 코로나19 이전: 2019.01.01~2019.11.31

* 코로나19 이후: 2020.04.01~2021.02.28

잘 알려진 불용어 사전이 없어 Python을 통해 텍스트 토큰화 하여 수작업으로 불용어를 선별하여 제거했다. 코로나19 전/후를 비교 분석하기 위해 팬데믹을 선포한 2021년 3월 11일 기준으로 전/후를 분류했으며, 코로나19 이전 시기를 2019년 1월부터 2019년 11월까지로 정의했고 코로나19 이후 시기를 2020년 4월부터 2021년 3월까지로 정의하여 데이터를 분류했다.

4.2 토픽모델링을 통한 토픽별 문장 분류

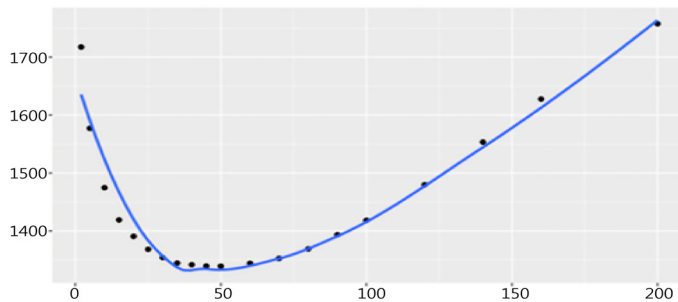
본 연구는 토픽모델링 종류인 LDA를 통해, 토픽을 추출했으며, 이를 호텔 리뷰의 6가지 속성에 매칭했다. 그리고 토픽별 키워드 결정 및 리뷰문장을 토픽별 키워드에 매칭하는 과정으로 토픽별 문장을 분류 및 데이터베이스에 저장했다. 우선 토픽모델링의 LDA를 통해 토픽을 추출하기 위해 최적의 토픽 수(k)를 결정해야한다. 따라서 본 과정에서는 5겹 교차 검증(5-fold cross validation)과 혼잡도(perplexity)를 통해 최적의 토픽 수(k)를 결정했다.

[그림 5]는 미국의 대도시 뉴욕을 대상으로 5겹교차 검증을 결합해 토픽 수(k)의 증가에 따른 혼잡도 변화를 나타냈다. 그림의 토픽 수는 2부터 200까지로 정의했으며, 기울기가 완만해지는 지점인 50을 최적의 토픽 수(k)로 지정했다.

이어서, <표 2>는 국가별, 도시별 최적의 토픽 수(k)를 나타낸 표이다. 이를 통해 중국(베이징 40, 상하이 50), 미국(뉴욕 45, 하와이 30)의 최적의 토픽 수(k)를 추출했다. 토픽별 문장 분류를 위해 LDA 통해 추출한 최적의 토픽 수만큼의 토픽을 호텔 리뷰의 6가지 속성에 매칭했으며, <표 3>과 같다. 매칭한 토픽을 통해 토픽별 키워드를 결정하고 리뷰 문장을 토픽별 키워드에 매칭하여 토픽별 문장을 분류 및 데이터베이스에 저장했다. <표 4>는 미국 뉴욕의 토픽별 키워드를 매칭한 결과이다.

<표 2> 도시별 최적의 토픽 수(k)

국가	중국		미국	
	베이징	상하이	뉴욕	하와이
토픽 수	40	50	45	30



[그림 5] 토픽 수(k)에 따른 혼잡도(미국-뉴욕)

<표 3> 토픽 추출 및 속성 매칭 결과(미국-뉴욕)

토픽 ID	속성	토픽	상위 10 단어
1	서비스	staff	location, staff, friendly, time, clean, comfortable, view, service, helpful, kind
2	서비스	dining	room, time, breakfast, bed, restaurant, made, helpful, clean, location, rooms
3	위치	location	location, place, view, small, clean, bar, room, rooms, friendly, square

49	위치	Facilities	central park, experience, square, clean, rooms, staff, friendly, bar, city
50	객실	rooms	room, time, friendly, clean, bed, service, floor, comfortable, night, view

〈표 4〉 토픽별 키워드 매칭 결과(미국-뉴욕)

뉴욕	가치	cheap, value, expensive, fee(s), worth, bar, parking fee, valet, price
	객실	room, desk, bed, floor, place, bathroom, view, small, large, big, air, conditioner
	위치	walking, location, time, night, square, central, park, times, front, close, city, entrance
	청결도	clean, dirty, covid, smell, trash, bathroom, detergent, area, poor, concern
	수면의 질	noise, noisy, quiet, sound, atmosphere, comfortable, deep, condition, experience, sleep
	서비스	staff, helpful, friendly, check, lobby, comfortable, time, desk, service, breakfast, bar, front, experience, food, menu, trip, service, wine, coffee, beverages, lobby, snack, furniture
하와이	가치	free, easy, worth, fee(s), value, price, pay, extra, rental
	객실	sunset(s), warm, view(s), room(s), small, space, pool(s), chair(s), facilities, outdated, suite, old, options, main, private, space, daily
	위치	time, spent, location, close, park, easy, access, oceanfront, rays, bay, sunsets, forward, construction, looking, seaside
	청결도	covid, clean, dirty, negative, bad, cleaning, evening, dirt, issue, full
	수면의 질	relax(ing), hear, loud, quiet, close, air, partial, noise, night
	서비스	lobby, reviews, buffet, bar, drinks, happy, restaurant(s), food, dinner, live, breakfast, spa, service(s), amenities, accommodations, options, front desk, people, help, care, housekeeping, shuttle, check, welcoming, professional, housekeeping, concierge, attentive, busy, enjoy, kids, children, families, available, taken, working, quick, job

4.3 빈도 분석

〈표 5〉는 빈도 분석 결과이다. 본 연구는 국가별 호텔 리뷰 데이터를 대상으로 도시별, 시기별 상위 10개의 단어를 추출했으며 코로나19의 시기적 관련성이 가장 높은 ‘청결도’와 관련된 단어를 기준으로 해석하여 비교 분석했다.

우선 코로나19 전/후의 미국 뉴욕과 하와이의 단어 빈도 결과를 비교했을 때 뉴욕은 2위까지, 하와이는 1위 단어가 동일한 결과임을 확인했다. 코로나19 이전 뉴욕은 ‘Clean’이 10위로 나타났으며 코로나19 이후 8위로 나타났다. 그리고 코로나19 이전 하와이는 ‘청결도’와 관련된 단어가 10위권 안에 들지 못했지만 코로나19 이후에는 ‘Clean’이 9위, ‘COVID’가 7위로 나타났다. 두 도시 모두 청결도에 대한 단어 빈도가 상승했으며, 코로나19 이후 청결도에 대한 관심이 증가한 것으로 해석된다.

코로나19 전/후의 중국 베이징과 쑤야의 단어 빈도 결과를 비교했을 때 두 도시 모두 상위 4위까지 단어가

동일한 결과임을 확인했다. 코로나19 이전 베이징은 ‘clean’이 9위로 나타났으며, 코로나19 이후 5위로 나타났다. 코로나19 이전 쑤야는 ‘clean’이 9위로 나타났지만, 코로나19 이후에는 10위권 안에 들지 못했다. 중국 베이징은 청결도에 대한 단어 빈도가 큰 폭으로 상승했고, 코로나19 이후 청결도에 대한 관심이 증가했음을 알 수 있다. 반면, 중국 쑤야는 청결도에 대한 단어 빈도가 하락했으며 코로나19 이후 청결도에 대한 관심이 감소했는데 지역적 특성이 오히려 코로나19와 거리가 매우 멀었기 때문으로 해석된다.

코로나19 전/후의 한국 서울과 제주도의 단어 빈도 결과를 비교했을 때 서울의 1위 단어를 제외하고 전반적으로 동일하지 않은 결과임을 확인했다. 그리고 코로나19 이전과 이후 시기를 비교했을 때 서울과 제주도는 청결도와 관련된 단어가 10위권 안에 들지 못했다. 이를 통해 두 도시 모두 청결도에 대한 단어 빈도를 확인하지 못했으며 전반적으로 ‘청결도’에 대해 관심이 없었거나 청결하기에 관광객의 관심을 많이 끌지 못한 것으로도 해석될 수 있다.

<표 5> 상위 10개 단어 추출 결과

도시	베이징		싼야		서울		제주도		뉴욕		하와이	
	이전	이후	이전	이후	이전	이후	이전	이후	이전	이후	이전	이후
코로나19												
1	hotel	hotel	hotel	hotel	kindness	kindness	travel	kindness	Room	Room	Room	Room
2	service	service	service	service	Hotel	employee	family	employee	Hotel	Hotel	Stay	Hotel
3	room	room	room	room	employee	service	facility	Hotel	Stay	Price	Night	Stay
4	breakfast	breakfast	breakfast	breakfast	service	Hotel	Hotel	breakfast	Staff	Staff	Hotel	Service
5	convenience	clean	convenience	check in	visit	breakfast	kindness	facility	Location	Location	Resort	View
6	location	convenience	location	convenience	satisfaction	use	employee	pool	Price	Stay	Hotel	Beach
7	stay	front desk	stay	stay	food	satisfaction	use	location	Night	Breakfast	Staff	COVID
8	surroundings	location	surroundings	surroundings	room	room	satisfaction	use	bed	Clean	Pool	Pool
9	clean	surroundings	clean	front desk	use	facility	child	satisfaction	Time	Sleep	Beach	Clean
10	check in	check in	check in	swimming pool	facility	check in	accommodation	travel	Clean	Night	View	Ocean

4.4 국가별 비교 분석

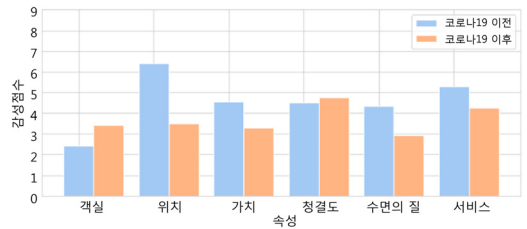
4.4.1 미국 감성분석

본 연구는 LDA를 통해 추출한 토픽을 호텔 리뷰의 6가지 속성별로 정렬했으며, 어휘기반 감성분석을 위해 Sentiwordnet3.0을 활용하여 데이터 세트에 적용했다. <표 6>은 호텔 리뷰의 6가지 속성 중 하나 이상에 포함된 리뷰 수를 코로나19 이전과 이후로 나누어 나타내는 표이다. 표의 리뷰 수는 코로나19 전/후로 10배의 차이를 확인했다.

<표 6> 코로나19 전/후의 감성분석 리뷰 수

지역	기간	감성분석 리뷰 수	
		코로나19 이전	코로나19 이후
뉴욕		1,231	126
하와이		1,293	129

[그림 6], <표 7>은 코로나19 전/후의 미국 뉴욕의 6가지 속성별 감성점수를 나타낸 그래프 및 표이다. 그림과 표를 보면, ‘객실’, ‘청결도’ 속성의 감성점수는 코로나19 이후 증가했으며 그 중 ‘객실’은 큰 폭으로 증가했다. 반면 ‘위치’, ‘가치’, ‘수면의 질’, ‘서비스’ 속성은 코로나19 이후 감소했으며 그 중에서 ‘위치’, ‘가치’는 큰 폭으로 감소했다. 따라서 뉴욕은 ‘객실’, ‘청결도’ 속성을 제외한 나머지 속성의 감성점수는 코로나19 이전에 비해 코로나19 이후 크게 감소했다.

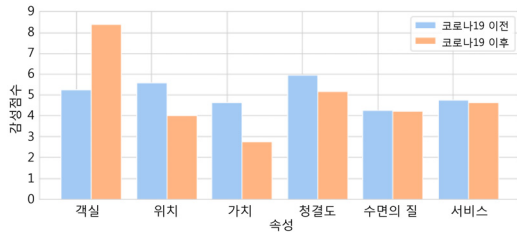


[그림 6] 6가지 속성별 감성점수(미국-뉴욕)

<표 7> 6가지 속성별 감성점수(미국-뉴욕)

속성	기간	
	코로나19 이전	코로나19 이후
객실	2.41	3.44
위치	6.4	3.5
가치	4.55	3.32
청결도	4.52	4.75
수면의 질	4.34	2.93
서비스	5.28	4.25

이어서, [그림 7], <표 8>은 미국 하와이의 코로나19 전/후의 6가지 속성별 감성점수를 나타낸 그래프 및 표이다. 그림과 표를 보면 ‘객실’ 속성의 감성점수는 코로나19 이후 큰 폭으로 증가했다. 반면 ‘위치’, ‘가치’, ‘청결도’, ‘수면의 질’, ‘서비스’ 속성은 코로나19 이후 감소했으며 그 중에서 ‘위치’, ‘가치’는 비교적 큰 폭으로 감소했다. 따라서 하와이는 ‘객실’ 속성을 제외한 나머지 속성은 감성점수가 코로나19 이후 감소했다.



[그림 7] 6가지 속성별 감성점수(미국-하와이)

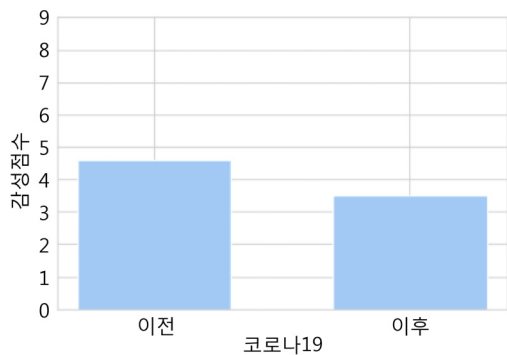
<표 8> 6가지 속성별 감성점수(미국-하와이)

속성 \ 기간	코로나19 이전	코로나19 이후
객실	5.26	8.4
위치	5.56	4.0
가치	4.62	2.78
청결도	5.96	5.17
수면의 질	4.27	4.20
서비스	4.74	4.64

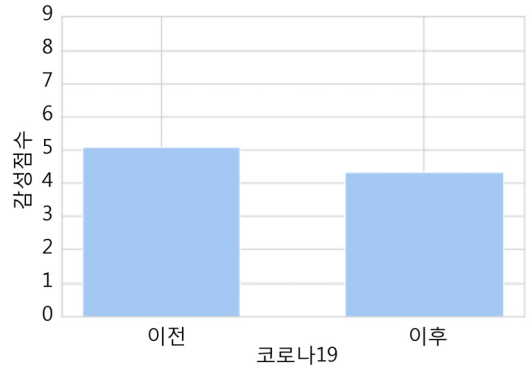
<표 9>, [그림 8], [그림 9]는 코로나19 전/후 뉴욕과 하와이의 리뷰에 대한 감성점수를 나타낸 그래프 및 표이다. 그림과 표를 보면 미국의 두 도시는 코로나19 이후 감성점수가 감소했다.

<표 9> 코로나19 전/후 도시별 감성점수(미국)

도시 \ 기간	코로나19 이전	코로나19 이후
뉴욕	4.58	3.52
하와이	5.07	4.34



[그림 8] 코로나19 전/후 감성점수(미국-뉴욕)



[그림 9] 코로나19 전/후 감성점수(미국-하와이)

한편, 본 연구에서 도출한 미국 뉴욕과 하와이의 ‘가치’ 속성의 감성점수를 보면 코로나19 이후 낮아지는 경향을 확인했다. 그리고 미국에서는 코로나19 이후 다양한 프로모션을 통해 가격 할인 등을 진행하고 있다(Le and Phi, 2021). 다양한 프로모션을 통한 가격 할인은 고객의 감성점수에 큰 영향을 미치지 못함을 확인했다. 그리고 하와이의 ‘청결도’ 속성의 감성점수를 보면 코로나19 전/후 시기에 동일한 서비스를 제공하더라도 고객들은 코로나19 이후에 더 불편함을 느끼는 것을 알 수 있다. 또한 코로나19 전/후의 도시별 전체 감성점수를 비교했을 때 전반적으로 감소했다. 결과적으로 코로나19 이후 뉴욕과 하와이의 숙박 시설을 이용하는 고객들의 만족도가 감소했음을 알 수 있다.

4.4.2 중국 감성분석

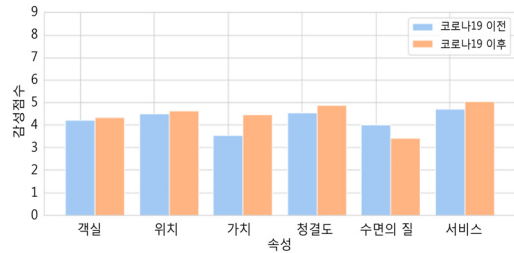
본 연구에서는 LDA를 통해 추출한 토픽을 호텔 리뷰의 6가지 속성별로 정렬했으며, 이를 어휘기반 감성분석을 위해 HowNet을 활용하여 데이터 세트에 적용했다. <표 10>은 호텔 리뷰의 6가지 속성 중 하나 이상에 포함된 리뷰 수를 코로나19 이전과

<표 10> 코로나19 전/후의 감성분석 리뷰 수

지역 \ 기간	코로나19 이전	코로나19 이후
베이징	20,046	9,303
싼야	16,666	12,683

이후로 나누어 나타내는 표이다. 표의 리뷰 수는 코로나19 전/후로 베이징은 약 2배, 쑤야는 약 1.3배 차이를 확인했다.

[그림 10], <표 11>은 코로나19 전/후 중국 베이징의 6가지 속성별 감성 점수를 나타낸 그래프 및 표이다. 그래프와 표를 보면, '수면의 질' 속성의 감성점수는 코로나19 이후 감소했다. 반면, '객실', '청결도', '위치', '가치', '서비스' 속성의 감성점수는 증가했다. 따라서 베이징은 '수면의 질' 속성을 제외한 나머지 속성 모두는 코로나19 이후 증가했다.

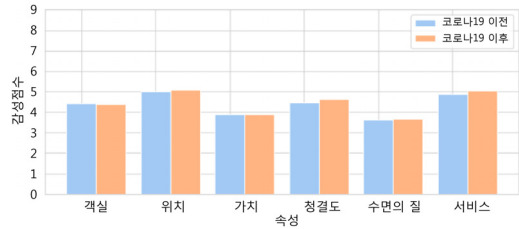


[그림 10] 6가지 속성별 감성점수(중국-베이징)

<표 11> 6가지 속성별 감성점수(중국-베이징)

속성 \ 기간	코로나19 이전	코로나19 이후
객실	4.21	4.32
위치	4.52	4.63
가치	3.57	4.48
청결도	4.56	4.87
수면의 질	4.01	3.45
서비스	4.73	5.04

[그림 11], <표 12>는 코로나19 전/후 중국 쑤야의 6가지 속성별 감성점수를 나타낸 그래프와 표이다. 그래프와 표를 통해 코로나19 이후 '위치', '가치', '청결도', '수면의 질', '서비스' 속성은 미세하게 증가했으며 '객실' 속성은 미세하게 감소했다. 따라서 쑤야는 '객실' 속성을 제외하고 나머지 속성 모두 미세하게 증가했다. 하지만 전반적으로 큰 변동이 없음을 확인했다.



[그림 11] 6가지 속성별 감성점수(중국-쑤야)

<표 12> 6가지 속성별 감성점수(중국-쑤야)

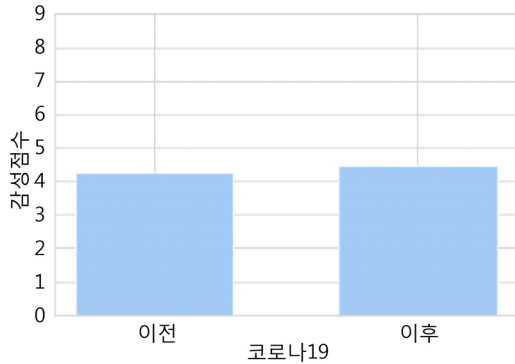
속성 \ 기간	코로나19 이전	코로나19 이후
객실	4.4	4.38
위치	4.99	5.08
가치	3.88	3.89
청결도	4.47	4.62
수면의 질	3.65	3.67
서비스	4.86	5.04

[그림 10], [그림 11]의 연구에서 도출한 중국 베이징과 쑤야의 6가지 속성별 감성점수를 보면 베이징의 '수면의 질'과 쑤야의 '객실' 속성을 제외하고는 코로나19 이전 보다 코로나19 이후의 감성점수가 더 높게 나타났다. 이러한 연구 결과는 국가별 여행 제한 정책으로 해외 관광이 어려워진 상황에서 예방접종을 맞았거나 여행을 할 수 있는 관광객들에 의해 국내 여행으로 만족감이 상승하는 것으로 해석된다 (정대영, 이수진, 2020).

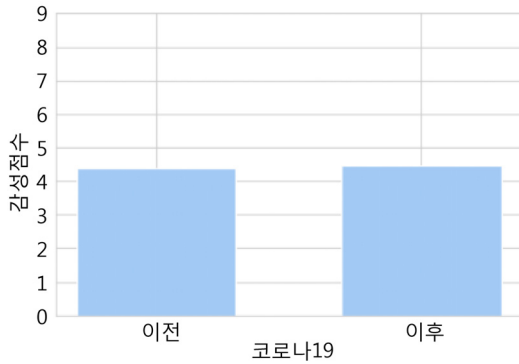
<표 13>, [그림 12], [그림 13]은 코로나19 전/후 베이징과 쑤야의 리뷰에 대한 감성점수를 나타낸 그래프 및 표이다. 그림과 표를 통해 중국의 두 도시는 코로나19 이후 감성점수가 증가했으며 이는 코로나19 이후 중국 베이징과 쑤야의 숙박 시설을 이용하는 고객들이 전반적으로 만족도가 증가했음을 알 수 있다.

<표 13> 코로나19 전/후 도시별 감성점수(중국)

도시 \ 기간	코로나19 이전	코로나19 이후
베이징	4.26	4.45
쑤야	4.38	4.45



[그림 12] 코로나19 전/후 감성점수(중국-베이징)



[그림 13] 코로나19 전/후 감성점수(중국-싼야)

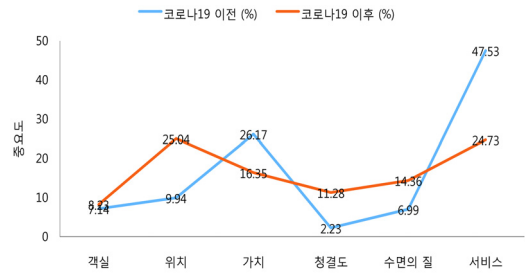
4.4.2.1 온라인 리뷰의 컨조인트 분석 결과

본 연구는 컨조인트 분석을 통해 중국 베이징과 싘냐의 호텔을 이용하는 고객을 대상으로 6가지 속성별 상대적 중요도를 코로나19 전/후로 비교 분석했다. 컨조인트 분석을 위해 5성급 호텔의 리뷰 중 6가지 속성에 모두 포함되는 호텔 리뷰 데이터를 대상으로 연구를 진행했으며 데이터 전처리 과정에서의 방법과 동일하게 코로나19 전/후를 구분하여 중국 베이징 코로나19 이전을 609개, 코로나19 이후를 426개로 분류했다. 이후 컨조인트 분석을 통해 6가지 속성에 대한 상대적 중요도 정도를 확률로 도출했으며 <표 14>, [그림 14]와 같다.

그림과 표를 통해 코로나19 이전 시기에는 ‘서비스’ 47.53%, ‘가치’ 26.17%, ‘위치’ 9.94%, ‘객실’ 7.14%, ‘수면의 질’ 6.99%, ‘청결도’ 2.23% 순으로 상대적 중요도를 측정했으며 코로나19 이후 시기에

는 ‘위치’ 25.04%, ‘서비스’ 24.73%, ‘가치’ 16.35%, ‘수면의 질’ 14.36%, ‘청결도’ 11.28%, ‘객실’ 8.23% 순으로 상대적 중요도를 측정했다.

[그림 14]를 통해 코로나19 전/후의 6가지 속성에 대한 상대적 중요도를 비교 분석했으며 코로나19 이전에는 ‘서비스’ 속성의 중요도가 큰 폭으로 높았다. 하지만 코로나19 이후에는 ‘위치’ 속성의 중요도가 가장 높았으며 전반적으로 6가지 속성 모두 25%를 초과하지 않는 선에서 비교적 균등하게 분포되었다.



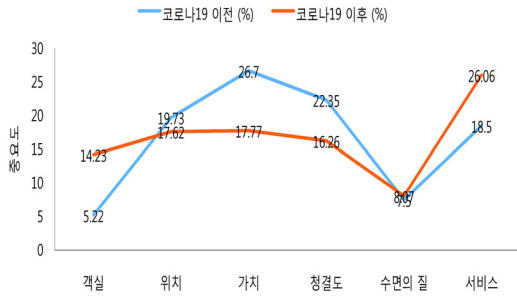
[그림 14] 코로나19 전/후 속성별 중요도(중국-베이징)

<표 14> 코로나19 전/후 속성별 중요도(중국-베이징)

속성 시기	객실	위치	가치	청결도	수면의 질	서비스	수량
코로나19 이전	7.14	9.94	26.17	2.23	6.99	47.53	609
코로나19 이후	8.23	25.04	16.35	11.28	14.36	24.73	426

다음, 중국 싘냐는 코로나19 이전을 514개, 코로나19 이후를 509개로 분류하여 컨조인트 분석을 통해 6가지 속성에 대한 상대적 중요도를 도출했으며 <표 15>, [그림 15]와 같다. 그림과 표를 통해 코로나19 이전 시기에는 ‘가치’ 26.7%, ‘청결도’ 22.35%, ‘위치’ 19.73%, ‘서비스’ 18.5%, ‘수면의 질’ 7.5%, ‘객실’ 5.22% 순으로 중요도를 측정했지만 코로나19 이후 시기에는 ‘서비스’ 26.06%, ‘가치’ 17.77%, ‘위치’ 17.62%, ‘청결도’ 16.26%, ‘객실’ 14.23%, ‘수면의 질’ 8.07% 순으로 중요도를 측정했다. 또한 [그림 15]를 통해 코로나19 전/후의 6가지 속성에 대한 상대적 중요도를 비교 분석했으며 코로나19 이전에는 ‘가치’

속성의 중요도가 높았고 코로나19 이후에는 ‘서비스’ 속성의 중요도가 높았다. 따라서 도시별 6가지 속성에 대한 상대적 중요도를 확률로 나타냈으며 이를 통해 중국 베이징의 호텔을 이용하는 고객들은 코로나19 이전에는 ‘서비스’ 측면을 큰 확률로 중요하게 여긴 반면, 코로나19 이후에는 ‘위치’, ‘가치’, ‘청결도’, ‘수면의 질’ 등 여러가지 측면을 중요하게 여겼다. 이러한 연구 결과는 여행자들이 서비스 측면만을 고려한 과거에 비해서 코로나19 팬데믹으로 인한 영향을 고려하고 있음을 알 수 있었다.



[그림 15] 코로나19 전/후 속성별 중요도(중국-싼야)

<표 15> 코로나19 전/후 속성별 중요도(중국-싼야)

속성 시기	객실	위치	가치	청결도	수면의 질	서비스	수량
코로나19 이전	5.22	19.73	26.7	22.35	7.5	18.5	514
코로나19 이후	14.23	17.62	17.77	16.26	8.07	26.06	509

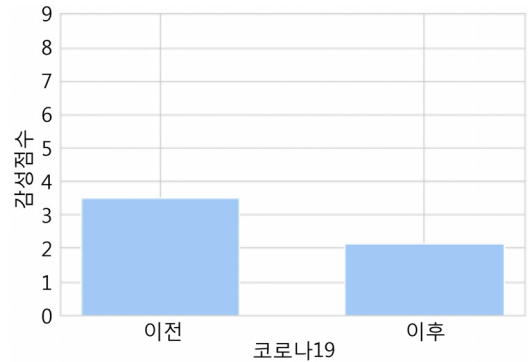
중국 싘야의 호텔을 이용한 고객들은 코로나19 이전에는 ‘가치’ 측면을 중요하게 여긴 반면, 코로나19 이후에는 ‘서비스’ 측면을 중요하게 여겼다. 또한, 코로나19 전/후를 비교했을 때 코로나19 이후 ‘청결도’, ‘위치’ 측면의 중요도 감소한 점과 ‘수면의 질’ 측면의 중요도 수치의 변화가 없는 것으로 볼 때 섬 지역인 싘야는 도심 지역의 베이징 보다 코로나19 팬데믹의 영향을 비교적 덜 받는 것을 알 수 있다. 예를 들어 코로나19 선포 후 집단감염 및 확진자 발생이 거의 없다.

4.4.3 한국 감성분석

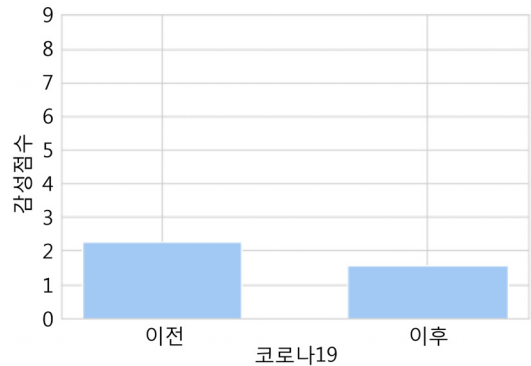
본 연구는 어휘기반 감성분석을 위해 KNU를 활용하여 데이터 세트에 적용했다.

<표 16> 코로나19 전/후 도시별 감성점수(한국)

도시	기간	코로나19 이전	코로나19 이후
서울	이전	3.51	2.13
	이후	2.28	1.56



[그림 16] 코로나19 전/후 감성점수(한국-서울)



[그림 17] 코로나19 전/후 감성점수(한국-제주도)

<표 16>, [그림 16], [그림 17]은 코로나19 전/후 서울과 제주도의 리뷰에 대한 감성점수를 나타낸 그래프 및 표이다. 그림과 표를 통해 한국의 두 도시는 코로나19 이후 감성점수가 감소했으며 이를 통해 코로나19 이후 서울과 제주도의 숙박 시설을 이용하는 고객들의 전반적으로 만족도가 감소했음을 알 수 있다.

5. 결론 및 향후 연구

본 연구는 코로나19 팬데믹 시기 중국, 한국, 미국의 온라인 호텔 리뷰를 통해 빈도분석, 감성분석, 키포인트 분석을 각각 적용해 코로나19 전/후를 비교 분석했다. 본 연구에서 사용된 데이터는 Python의 웹 스크레이퍼를 통해 수집했으며, 미국과 한국 데이터는 트립닷컴, 부킹닷컴, 중국 데이터는 씨트립에서 국가별 언어에 대한 온라인 호텔 리뷰를 수집했다. 수집한 호텔 리뷰 데이터를 언어별 특성에 맞게 불용어 처리 및 결측치 제거 등의 전처리 과정을 수행했으며, 이후 빈도분석을 통해 온라인 리뷰에서 가장 많이 사용되는 상위 10개의 단어를 코로나19 전/후로 국가별 및 도시별로 비교 분석했다. 그리고 감성분석을 이용하여 코로나19 전/후의 국가별, 도시별 6가지 속성 및 전체 리뷰의 만족도를 비교 분석하였다. 온라인 리뷰의 6가지 속성에 대해서 키포인트 분석을 수행하여 6가지 속성에 대한 상대적 중요도를 분석했다. 최종적으로 본 연구는 코로나19 팬데믹 시대 호텔업계에 대한 사고방식 및 정서적 변화를 이해하는 방법에 대해 제안했다.

다음은 도시별 연구결과를 요약한 내용이다.

첫 번째 미국 뉴욕과 하와이를 코로나19 이전과 이후로 비교 분석했을 때 뉴욕은 코로나19 이후 ‘청결도’와 관련된 단어 빈도가 증가했고 6가지 속성 중 ‘청결도’ 및 ‘객실’의 속성이 코로나19 이전 보다 증가했다. 또한 코로나19 전/후의 전체 데이터의 감성점수는 감소했다. 이를 통해, 뉴욕 지역의 호텔을 이용하는 고객들이 객실을 이용함에 있어 냄새, 환경, 상태 등 ‘청결도’ 및 ‘객실’ 측면에 만족스러움을 느끼지만, ‘위치’, ‘가치’, ‘수면의 질’, ‘서비스’의 측면에서는 불만족스러움을 느낀다는 사실을 확인했다. 따라서 뉴욕 지역의 호텔들은 호텔의 다양한 서비스, 숙소 환경 등의 측면에서 고객의 만족도를 높일 수 있는 방안을 모색해야한다. 하와이는 코로나19 이후 ‘청결도’와 관련된 단어 빈도가 증가했고, 6가지 속성 중 ‘객실’ 속성을 제외한 다른 속성은 코로나19 이후 감소했다. 또한 코로나19 전/후의 전체 데이터의 감성

점수는 감소했다. 이는 하와이 지역의 호텔을 이용하는 고객들이 객실 측면에서 만족도를 느끼지만 그 외의 측면에서는 불만족스러움을 느낀다는 사실을 확인했다. 따라서 앞의 뉴욕의 결과와 유사하게 호텔의 다양한 서비스, 숙소 환경 등의 측면에서 고객의 만족도를 높일 수 있는 방안을 모색해야한다.

두 번째 중국 베이징과 쑤야를 코로나19 이전과 이후로 비교 분석했을 때 베이징은 코로나19 이후 ‘청결도’와 관련된 단어 빈도가 증가했고 6가지 속성 중 ‘수면의 질’ 속성을 제외한 나머지 모든 속성은 증가했다. 또한, 코로나19 전/후의 전체 데이터의 감성점수는 증가했다. 이를 통해 베이징 지역의 호텔을 이용하는 고객들이 코로나19 이후 객실을 이용함에 있어 객실 소음, 수면 환경 등의 ‘수면의 질’ 측면에서 불만족스러움을 느끼지만 헬스장, 영화관, 수영장, 유흥시설 등의 ‘서비스’, ‘청결도’, ‘객실’, ‘위치’, ‘가치’ 측면에서 만족스러움을 느낀다는 사실을 확인했다. 따라서, 베이징 호텔들은 호텔의 ‘수면의 질’ 측면에서 고객의 만족도를 채울 수 있는 방안을 모색해야한다. 쑤야는 코로나19 이후 ‘청결도’와 관련된 단어 빈도가 감소했고 6가지 속성 모두 미세하게 변했으며 ‘객실’ 속성을 제외하고는 증가했다. 그리고 코로나19 전/후의 전체 데이터의 감성점수는 증가했다. 이는 쑤야 지역의 호텔을 이용하는 고객들이 ‘객실’ 측면을 제외한 5가지 측면에 대해 만족스러움을 느낀다는 것을 확인했다. 또한, 코로나19 전/후 ‘객실’ 측면의 감성점수의 차이가 미미하고, 전체 데이터의 감성점수가 코로나19 이전 보다 이후가 긍정적이라는 사실을 통해 쑤야 지역의 호텔들은 섬이라는 지역적 특성을 참고하여 관광객의 만족도를 극대화할 수 있는 방안을 모색해야 한다.

세 번째 한국 서울과 제주도를 코로나19 이전과 이후로 비교 분석했을 때 서울과 제주도는 코로나19 이후 ‘청결도’와 관련된 단어 빈도가 존재하지 않았고 코로나19 전/후의 전체 데이터의 감성점수는 감소했다. 그 중에서 서울은 큰 폭으로 감소했다. 이는 서울과 제주도의 호텔을 이용하는 고객들이 전반적으로 불만족스러움을 느끼는 것을 알 수 있다.

본 연구의 도시별 결과를 바탕으로 정리하면 다음과 같다.

첫 번째 각 국가의 최대 관광명소인 하와이와 썬야는 대부분의 속성이 전반적으로 코로나19 이전보다 코로나19 이후 감성점수가 더 높게 나타났다. 다국적 호텔기업은 국가별 차별화 전략에 활용할 수 있으며, 예를 들어 하와이와 썬야의 경우 주로 관광객들이 찾는 지역적 특성을 고려하여 청결도뿐만 아니라 호텔의 유희(카지노, 수영장, 헬스장 등), 가성비 등을 고려하여 시장 우위를 점할 수 있다. 반면, 인구의 분포가 높음 대도시인 베이징과 뉴욕은 객실 관리 및 소독 시스템을 갖는 시장이 우위를 점할 수 있다. 이처럼 지역적 특성에 맞게 의사결정 과정을 고려하여 유용하게 사용할 수 있다.

두 번째 빅 데이터를 기반으로 온라인 리뷰 동향을 이해하고, 코로나19 팬데믹 상황에서 적절한 생존 전략을 세우는데 도움이 될 것이며, 분석 기법을 적용하여 얻는 키워드는 현재 호텔의 제품 및 기능을 확장하는데 도움이 될 것이다.

본 연구가 갖는 한계점 및 향후 연구방향은 다음과 같다.

첫 번째 한국과 미국은 '사회적거리두기', '집합금지', '도시 봉쇄', '여행 제한' 등의 정부 정책으로 인해, 온라인 리뷰 수가 거의 없어, 서로 다른 웹사이트에서 데이터를 수집하여 불균형한 데이터를 사용하여 연구를 진행했으며, 이는 편향된 연구 결과로 이어질 수 있다. 향후 연구에서는 웹사이트의 종류와 데이터 수를 통일시켜 연구 과정의 통일성을 향상시킬 것이다.

두 번째 한국의 온라인 리뷰를 감성분석에 적용할 때, 신뢰할 수 있는 불용어 사전이 거의 없어서 불용어 필터를 만들어서 분석했다. 향후 연구에서 한국어 불용어를 효과적으로 처리하고 신뢰할 수 있는 풍성한 실험 결과를 도출할 것이다.

세 번째 감성분석을 할 때, 국가별 어휘기반 감성 점수를 측정했다. 어휘기반 감성분석의 경우, 특정 영역을 반영하지 못하는 단점이 있으며(Ma et al., 2018), 향후 연구에서는 인공지능, 기계학습과 딥러

닝 등의 더욱 발전된 방법을 적용할 것이다.

네 번째 국가별 연구결과 도출에서 통계변수를 적용하지 않고, 단순히 6가지 속성 기반으로 감성분석을 하여 분류했다. 이는 경제사회의 구체적 상황변수를 고려하지 못한 것이므로 향후 연구에서는 호텔의 등급, 여행 형태, 고객 특성 등에 대한 변수를 반영하여 접근할 것이다.

다섯 번째 컨조인트 분석에서 5성급 호텔에 대한 속성별 중요도를 비교 분석할 때, 6가지 속성별로 어느 정도의 차이를 발견할 수 있었으며 그중에서 위치의 차이가 가장 큰 것으로 분석되었다. 따라서 향후 연구에서는 동일한 5성급 호텔을 대상으로 동일한 속성을 통해 연구를 진행할 필요가 있다.

참고문헌

- 김근형, 오성열, "온라인 고객리뷰 분석을 통한 시장세분화에 텍스트마이닝 기술을 적용하기 위한 방법론", *한국콘텐츠학회논문지*, 제9권, 제8호, 2009, 272-284.
- 김남순, "코로나바이러스감염증-19 현황과 과제. 보건·복지", *Issue & Focus*, 제373호, 2020, 1-13.
- 김다혜, 김주연, "공간마케팅 활용을 위한 컨조인트 분석", *한국실내디자인학회 학술대회논문집*, 제21권, 제3호, 2019, 201-204.
- 김동준, 최현준, 조환기, 김광용, "코로나팬데믹 이후 관광산업 발전방안 연구", *호텔경영학연구*, 제29권, 제4호, 2020, 129-144.
- 김진유, "포스트코로나시대 도시계획의 과제", *Urban planners*, 제7권, 제3호, 2020, 11-15.
- 백창재, "금융세계화와 미국: 국가, 패권, 자본", *국가전략*, 제12권, 제3호, 2006, 37-72.
- 서예령, 고금석, 이재우, "빅데이터 LDA 토픽 모델링을 활용한 국내 코로나 19 대유행 기간 마스크 관련 언론 보도 및 태도 변화 분석", *한국정보통신학회논문지*, 제25권, 제5호, 2021, 731-740.

- 손헌일, 김기욱, 허종배, 박충훈, “코로나 19 극복을 위한 부산시 정책 대응”, *BDI 정책포커스*, 2020, 1-12.
- 송수연, 김한경, “언택트 시대의 대학교육: 코로나 19에 따른 비대면 강의 만족도와 수강지속 의사에 영향을 미치는 요인에 관한 연구”, *아시아교육연구*, 제21권, 제4호, 2020, 1099-1126.
- 신윤정, 김부용, 현용진, “컨조인트 분석 방법론에 의한 담배 포장의 금연 인식 유발 효과 분석”, *보건사회연구*, 제27권, 제1호, 2007, 27-51.
- 이승영, “코로나19와 전 지구적 대응 전략 비교 분석”, *시큐리티연구*, 특별호, 2020, 121-150.
- 이지선, “컨조인트 분석을 이용한 마늘 건강기능식품 선택속성의 상대적 중요도 분석”, *동아시아식생활학회지*, 제29권, 제4호, 2019, 336-343.
- 임정수, “최신 영화 VOD 이용자의 선호도에 대한 컨조인트 분석”, *한국콘텐츠학회논문지*, 제13권, 제5호, 2013, 191-198.
- 장성순, 최재영, “컨조인트 분석을 이용한 생체인식기술활용 서비스에 대한 소비자 선호 분석 연구”, *한국혁신학회지*, 제13권, 제4호, 2020, 181-206.
- 정대영, 이수진, “코로나 19, 여행의 미래를 바꾸다”, *이슈 & 진단*, 2020, 1-25.
- 조혜경, “세계경제 질서의 다극화와 미국의 경제 패권”, *동향과 전망*, 제84호, 2012, 11-37.
- 통계청, “코로나 19발생현황 (해외)”, 2021. A available at https://kosis.kr/covid/covid_index.do (Accessed July 02, 2021).
- 홍태희, “코로나 19 팬데믹 경제위기의 특성과 원인 그리고 전망”, *경상논총*, 제38권, 제3호, 2020, 79-99.
- Al-maaitah, T.A., T.E. Majali, M. Alsold, and D.A. Al- maaitah, “The Impact of COVID-19 on the Electronic Commerce Users Behavior”, *Journal of Contemporary Issues in Business and Government*, (27:1), pp. 784-793, 2021.
- Blei, D.M., A.Y. Ng, and M.I. Jordan, “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*, Vol.3, 2003, pp. 993-1022.
- Cheng, Y. H. and Ho, H. Y., “Social influence’s impact on reader perceptions of online reviews”, *Journal of Business Research*, Vol.68, No.4, 2015, pp. 883-887.
- Green, P.E. and V. Srinivasan, “Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice”, *Journal of Marketing*, Vol.54, No.4, 1990, 3-19.
- Gretzel, U., M. Fuchs, R. Baggio, W. Hoepken, R. Law, J. Neidhardt, and Z. Xiang, “e-Tourism Beyond COVID-19: A Call for Transformative Research”, *Information Technology & Tourism*, Vol.22, 2020, pp. 187-203.
- Hong, T.H., H. Niu, G. Ren, and J.Y. Park, “Multi-Topic Sentiment Analysis using LDA for Online Review”, *The Journal of Information Systems*, Vol.27, No.1, 2018, pp. 89-110.
- Kim, R.Y., “The Impact of COVID-19 on Consumers: Preparing for Digital Sales”, *IEEE Engineering Management Review*, Vol.48, No.3, 2020, 212-218.
- Li, F., M. Huang, and X. Zhu, “Sentiment Analysis with Global Topics and Local Dependency”, In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol.24, No.1, 2020.
- Li, M., L. Huang, C.H. Tan, and K.K. Wei, “Helpfulness of online product reviews as seen by consumers: Source and content features”, *International Journal of Electronic Commerce*, Vol.17, No.4, 2013, pp. 101-136.
- Liu, S., L. Yang, C. Zhang, Y.T. Xiang, Z. Liu, S. Hu, and B. Zhang, “Online Mental Health Services in China during the COVID-19 outbreak”, *The Lancet Psychiatry*, Vol.7,

- No.4, 2020, e17-e18.
- Ma, E., M. Cheng, and A. Hsiao, "Sentiment Analysis: A Review and Agenda for Future Research in Hospitality Contexts", *International Journal of Contemporary Hospitality Management*, Vol.30, No.11, 2018, 3287-3308.
- Medhat, W., A. Hassan, and H. Korashy, "Sentiment Analysis Algorithms and Applications: A Survey", *Ain Shams Engineering Journal*, Vol.5, No.4, 2014, 1093-1113.
- Priyantina, R.A. and R. Sarno, "Sentiment analysis of hotel reviews using Latent Dirichlet Allocation, semantic similarity and LSTM", *Int. J. Intell. Eng. Syst*, Vol.12, No.4, 2019, 142-155.
- Putri, I.R. and R. Kusumaningrum, "Latent Dirichlet allocation (LDA) for sentiment analysis toward tourism review in Indonesia", *Journal of Physics: Conference Series*, Vol.801, No.1, 2017, 012073.
- Shafqat, W. and Y.C. Byun, "A Recommendation Mechanism for Under-Emphasized Tourist Spots Using topic modeling and Sentiment Analysis", *Sustainability*, Vol.12, No.1, 2020, 320.
- Tong, Z. and H. Zhang, "A text mining research based on LDA topic modelling", In *International Conference on Computer Science, Engineering and Information Technology*, 2016, 201-210.
- WHO, "Timeline: WHO's COVID-19 response", 2021. A available at <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline#event-72> (Accessed July 02, 2021).
- Xiang, Y.T., Y. Yang, W. Li, L. Zhang, Q. Zhang, T. Cheung, and C.H. Ng, "Timely Mental Health Care for the 2019 Novel Coronavirus Outbreak is Urgently Needed", *The Lancet Psychiatry*, Vol.7, No.3, 2020, 228-229.
- Xie, R., S.K.W. Chu, D.K.W. Chiu, and Y. Wang, "Exploring Public Response to COVID-19 on Weibo with LDA Topic Modeling and Sentiment Analysis", *Data and Information Management*, Vol.5, No.1, 2021, 86-99.
- Yao, Z.Y., Y.K. Park, and T. H. Hong, "A study on the Effect of Reviewer Attributes on the Usefulness of Online Reviews", *Information Systems Journal*, Vol.29, No.2, 2020, 173-195.

◆ About the Authors ◆

**홍 준 우 (ehddml70@naver.com)**

울산대학교 경영학과에서 학사학위를 취득하였다. 현재 부산대학교 경영학과 경영정보전공 석사과정에 재학 중이다. 주요 관심분야는 소셜 네트워크, 오피니언 마이닝, 추천시스템 등이다.

**홍 태 호 (hongth@pusan.ac.kr)**

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, 정보시스템연구 등에 게재하였다.