

리뷰 텍스트 기반 감성 분석과 네트워크 분석에 관한 연구

Sentiment Analysis and Network Analysis based on Review Text

김 유 미 (Yumi Kim)*

허 고 은 (Go Eun Heo)**

목 차

- | | |
|-----------|-------------|
| 1. 서 론 | 4. 실험 결과 분석 |
| 2. 이론적 배경 | 5. 결 론 |
| 3. 연구 설계 | |

초 록

리뷰 텍스트는 이용자들의 경험과 의견이 구체적으로 담겨있어 이를 분석하면 리뷰 대상에 대한 많은 내용을 파악할 수 있다. 이에 따라 리뷰 텍스트에 대해 감성 분석을 진행하여 음식점의 각 요인에 대한 이용자의 평가 등을 파악하는 연구, 네트워크 분석을 통한 이용자들의 선호를 파악하는 연구들이 진행되어왔다. 본 연구에서는 음식점 리뷰 텍스트의 별점 기반 만족도가 높은 음식점과 낮은 음식점을 분석대상으로 선정하여 감성 분석과 네트워크 분석을 통합적으로 수행하였다. 서로 다른 두 집단의 리뷰 텍스트에서 나타나는 차이를 음식점의 특성을 파악하여 좋은 음식점의 기준과 음식점 만족도에 영향을 미치는 주요인을 확인하고자 하였다.

ABSTRACT

As review text contains the experience and opinions of the customers, analyzing review text helps to understand the subject. Existing studies either only used sentiment analysis on online restaurant reviews to identify the customers' assessment on different features of the restaurant or network analysis to figure out the customers' preference. In this study, we conducted both sentiment analysis and network analysis on the review text of the restaurants with high star ratings and those with low star ratings. We compared the review text of the two groups to distinguish the difference of the two and identify what makes great restaurants great.

키워드: 텍스트 마이닝, 리뷰 텍스트, 감성 분석, 네트워크 분석

Text Mining, Review Text, Sentiment Analysis, Network Analysis

* 연세대학교 영어영문학과 학사과정(ggsnim1917@yonsei.ac.kr / ISNI 0000 0005 0358 1186) (제1저자)

** 연세대학교 문헌정보학과 연구교수(goeun.heo@yonsei.ac.kr / ISNI 0000 0004 7707 1202) (교신저자)

논문접수일자: 2021년 7월 30일 최초심사일자: 2021년 8월 10일 게재확정일자: 2021년 8월 20일

한국문헌정보학회지, 55(3): 397-417, 2021. <http://dx.doi.org/10.4275/KSLIS.2021.55.3.397>

© Copyright © 2021 Korean Society for Library and Information Science

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

각종 리뷰 사이트 및 SNS가 활발해지면서 온라인상에 등재되는 이용자 리뷰는 제품, 서비스, 혹은 장소에 대한 이용자의 평가를 확인하는 주요한 방법으로 활용되고 있다. 이러한 이용자 리뷰는 다른 이용자의 구매에도 영향을 미칠 뿐만 아니라 판매자의 입장에서도 이를 피드백으로 활용할 수 있다.

대부분의 리뷰 사이트는 별점 혹은 숫자로 평가되는 등급과 언어로 표현되는 리뷰 형태가 사용되고 있다. 이상재, 최준연, 최진호(2014)에 따르면 이 중에서도 언어로 표현되는 리뷰는 이용자의 구체적인 감정, 경험을 파악하기 용이하므로 다른 이용자들에게 더 큰 영향을 미친다.

온라인 이용자 리뷰가 활발해지며 리뷰를 작성할 수 있는 서비스 또한 다양해지고, 이에 대한 연구 또한 활발하게 진행되고 있다. 국내에는 인스타그램, 페이스북, 네이버 블로그, 네이버 스마트플레이스, 카카오톡 등에서 해외에서는 Yelp, OpenTable, TripAdvisor, Google Map 등의 서비스에서 리뷰 텍스트를 작성, 확인할 수 있다.

지난 3월 네이버 스마트플레이스는 리뷰 서비스에서 별점 기능을 없애고, 방문객 리뷰 텍스트를 중심으로 핵심 키워드를 추출하여 제공하겠다고 밝혔다(문지영, 2021). 이는 업계에서도 별점보다 리뷰를 더 중점적으로 보고 중요하게 여기고 있다는 것이다. 또한, 네이버가 국내에서 영향력이 있는 서비스인 만큼 다른 서비스도 이와 비슷한 행보를 밟을 것으로 예상된다. 그러나 별점 없이 이러한 키워드 형태의 리뷰 서비스가 제공되는 것이 처음이므로 어떠한 키워드가 많

을수록 더 좋은 장소인지 판단하기는 힘들 것이다. 이런 시점에서 이용자의 만족도가 상대적으로 더 높은 장소가 어떠한 특성을 가지고 있는지, 이런 장소의 리뷰 텍스트에서는 어떠한 내용을 주요하게 확인할 수 있는지 알아보는 것은 중요한 과제일 것이다.

따라서 본 연구에서는 온라인 리뷰사이트에 등재된 이용자들의 리뷰 중 별점기능 외에도 주요한 정보로 활용될 수 있는 리뷰 텍스트를 확인하여 구체적인 내용을 파악해보고자 한다. 특히, 온라인 리뷰 텍스트 중에서도 이용층이 편중되지 않으며 리뷰 작성이 활발하게 이루어지는 요식업계를 대상으로 선정하였고 실방문자의 상세한 리뷰를 확인하기 위하여 음식점 예약 사이트인 OpenTable에 등록된 리뷰 텍스트를 활용하였다. 본 연구에서 설정한 연구 질문은 다음과 같다.

- 연구 질문 1. 만족도가 최상위인 음식점과 최하위인 음식점의 리뷰 감성분석 결과는 어떠한 차이를 보이는가?
- 연구 질문 2. 별점 기반 음식점에 대한 만족도와 리뷰 감성분석 결과가 통계적으로 유의한 인과성을 보이는가?
- 연구 질문 3. 만족도가 최상위인 음식점과 최하위인 음식점의 특성과 차이를 리뷰 키워드 기반 네트워크 분석을 통해 발견할 수 있는가?

음식점 리뷰를 활용하여 이용자들의 만족도가 높은 음식점과 만족도가 낮은 음식점을 비교함으로써 개별 음식점의 특성 파악을 넘어서서 어떠한 특성이 좋은 음식점을 만들며 사람

들의 음식점 만족도를 높여주는지 확인하고자 한다. 이를 통해 운영자의 관점에서는 마케팅 전략 및 음식점 운영에 전반적으로 도움이 되는 기초 자료를 수립할 수 있으며 고객의 관점에서도 만족도가 높은 음식점과 주요인을 파악하여 좋은 음식점을 선별할 수 있는 기준을 제공할 수 있으리라 기대한다.

2. 이론적 배경

2.1 감성 분석 연구

기존의 온라인 텍스트에 대한 감성 분석에 관해 진행된 연구를 살펴보면 크게 더 정확한 감성 사전 혹은 감성 분석 방법 구축에 관한 연구나 텍스트에서 나타나는 감성 분석의 결과의 활용에 관한 연구가 있다.

Zhang, Ye, Zhang, Li(2011)는 머신 러닝 감성 분석 기법을 활용하여 광둥어로 쓰인 음식점 리뷰를 분석하였다. 나이브 베이즈 알고리즘을 활용하여 광둥어를 감성 분석할 때 바이그램(bigram)을 활용하는 것이 가장 높은 정확성을 보인다는 것을 밝혔다. 또한 이를 통해 표준 중국어 및 다른 언어 사용자들이 광둥어로 작성된 리뷰의 내용을 편하게 파악할 수 있도록 하였다. Kang, Yoo, Han(2012)은 음식점 리뷰의 감성 분석 시 긍정어와 부정어의 정확도 차이를 줄이기 위해 나이브 베이즈 알고리즘을 개선하여 부정어 감성 분석의 정확도를 향상시켰다. Yu et al.(2017)은 Yelp에 등록된 어 있는 음식점 리뷰를 SVM 모델을 활용하여 리뷰에 등장하는 감성 단어의 출현 빈도를 통

해 각 리뷰의 감성 경향을 확인하였고 감성 단어에 극성값을 부여하였다. Shung et al.(2018)은 감성 분석에 있어서 기존 딥 러닝 모델의 한계점을 극복하기 위하여 Bidirectional LSTM을 활용하여 정보를 수집한 후 감성 분석을 진행하는 새로운 모델을 제시하였다. Hossain, Sharif, Hoque, Sarker(2020)는 딥 러닝 기법을 활용하여 소비자가 작성한 리뷰에서 나타나는 긍정·부정의 극성값을 최고 91.35%의 정확도로 확인하는 방법론을 제안하였다.

국내에서도 한국어 감성 사전의 연구가 진행되고 있다. 송종석, 이수원(2011)은 제품 특징에 대한 서술어의 의미를 파악하는 감성 분석의 정확도를 높이기 위해 리뷰 대상의 제품 특징과 별점 점수를 고려하여 도메인 긍정/부정 사전을 구축하였다. 소진수, 신판섭(2020)은 음식점의 '음식', '가격', '서비스'에 대하여 새로운 한국어 감성 사전을 구축하였다. 이를 통해 상세한 내용을 파악하기 어려운 별점 점수와 별개로, 음식점에 대한 다양한 세부 평가 항목 점수를 예측할 수 있도록 하였다. Gang(2020)은 음식점 리뷰를 활용하여 소비자의 감정 경향과 온라인 리뷰 측정에 효과적인 새로운 머신 러닝 감성 분석 방법을 제안하였다.

이렇게 발전 중인 감성 분석 기법을 활용하여 텍스트의 내용을 심도 있게 확인하고자 하는 연구도 진행되었다. 이시환, 이훈영(2018)은 음식점 리뷰의 수와 감성 분석 결과를 기반으로 회귀 분석을 실시하여 리뷰의 극단적 감성 표현이 소비자들의 인지 유용성과 인지 즐거움에 영향을 미친다는 것을 발견하였다. Agüero-Torales, Cobo, Herrera-Viedma, López-Herrera(2019)는 미식에 관한 감성을 효과적으로 분석

하기 위해 TripAdvisor 별점 및 리뷰수 기준 그라나다 지방의 최고의 음식점 10개를 선정하였고 영문 리뷰에 대해 VADER 감성 사전을 활용하여 감성 분석을 진행하였다.

온라인 리뷰에 나타나는 리뷰 대상의 다양한 특성을 구분하여 각 특성에 대한 감성과 별점 점수의 관련성을 확인하는 연구도 진행되었다. 이렇게 각 특성을 개별적으로 확인하는 연구들에서는 주로 기 연구된 음식점의 특성 분류를 활용한다. Gan, Ferns, Yu, Jin(2017)은 음식, 서비스, 분위기, 가격, 사회적 맥락 등 음식점의 5가지 특성에 대한 감성 분석을 각각 진행하였고 수치적 계산을 통해 각 특성이 별점 점수에 얼마나 영향을 미치는지 연구하였다. Tian, Lu, McIntosh(2021)은 Yelp 리뷰를 통해 감성분석을 진행하여 이용자들이 리뷰 작성 시 부정 단어보다는 긍정 단어를 더 많이 사용한다는 것을 밝혔고, 음식에 대한 것보다 서비스에 대해 이야기할 때 감성 어휘를 더 많이 사용한다는 것을 발견했다. Mathayomchan, Taecharungroj(2020)은 구글맵에 등재되어있는 영국 내 음식점 중 5,010개의 음식점에 대한 리뷰 총 935,386개를 확인하였다. 이들은 음식점의 주요 특성 4개(음식, 서비스, 분위기, 가격)에 대해 VADER 감성 사전을 활용하여 감성 분석을 진행하였고, 이 결과와 별점 분석에 대한 회귀 분석을 진행하여 음식에 대한 가치가 별점에 가장 크게 영향을 미치고 가격에 대한 가치가 별점에 가장 부정적으로 영향을 미친다는 것을 밝혔다. Gao, Tang, Wang, Yin(2018)은 온라인 음식점 리뷰에 대해 비교 문장을 추출하여 각 비교 대상을 기준으로 감성 분석을 진행하였고, 네트워크 분석을 통해 어떤 음식점끼리 경쟁 구도를 보이

는지 파악하였다. 이를 통해 얻어내는 결과는 실제 시장에서 경쟁하기 위한 효율적인 전략 수립에 도움이 될 수 있을 것이다.

2.2 네트워크 분석 연구

네트워크 분석을 통해 텍스트에서 다루고 있는 대상 혹은 분야를 거시적으로 확인하는 연구가 진행되었다. Milojević, Sugimoto, Yan, Ding(2011)은 논문 제목을 동시출현단어 분석 및 계층적 클러스터링을 활용하여 문헌정보학 분야의 지적구조를 확인하였다. Kim, Yoon, Crowcroft(2012)는 네트워크 분석을 이용하여 정보 보안 분야의 학술지 논문과 프로시딩을 분석하였다. 이들은 한국인 연구자들과 저명한 연구자들의 출판물 비교 분석하여 두 그룹의 출판 경향 차이를 밝혔다. 허고은, 송민(2013)은 의료정보학 분야의 저널인 *IEEE ENG MED BIOL* 저널에 대해 저자동시인용분석 및 동시출현단어 분석을 진행하여 의학과 정보학의 융합학문인 의료정보학 학문의 지적구조를 확인하였다. 김지현(2016)은 광역도시 대표 트위터의 내용을 네트워크 분석을 통해 분석하여 연도별, 도시별로 시민에게 제공한 정보의 경향을 확인하였다. 정지나, 정힘찬, 김용(2017)은 재난정보와 관련한 국내 학위논문 및 학술지 논문에 대해 통계분석, 빈도분석 및 네트워크 분석을 진행하여 재난정보와 관련된 국내 연구 동향을 파악하였다.

텍스트에 대한 네트워크 분석을 진행하여 텍스트로부터 실제 현상에 대한 이해, 새로운 관계를 발견하는 연구도 진행되었다. 이러한 연구들은 소비, 질병 등 다양한 분야에서 진행되고 있다. 진설아 외(2013)는 아이폰 관련 트윗에

대하여 네트워크 분석, 키워드 빈도 분석, 토픽 모델링을 진행하여 해당 제품에 관한 토픽에 대한 변화 시점 및 패턴을 확인하였다. 이를 통해 대중의 반응을 온라인으로 빠르게 확인하여 기업이 신속한 대응을 할 수 있는 방안을 모색하였다. Song, Heo, Lee(2015)는 PubMed에서 알츠하이머에 관한 논문을 추출하여 생산성 분석, 네트워크 분석 및 내용 분석을 통해 알츠하이머 병 연구 동향을 파악하였다. Jo, Kim, Kim(2017)은 TripAdvisor의 리뷰를 활용하여 음식점 리뷰에서 가장 자주 사용되는 단어 72개를 추출하여 서비스의 질, 음식 메뉴, 즐거움, 지역 음식 총 4개의 그룹으로 나누었다. 이렇게 나눈 단어들의 연결 정도 중심성, 아이겐벡터 중심성을 분석하여 소비자의 선호를 파악하였다.

반현정, 김학선(2019a)은 소비자들의 레스토랑 인식을 조사하기 위해 TripAdvisor에 등재된 부산 지역 음식점의 808개의 영문 리뷰 중 58,800개의 단어 중 상위 100개 단어를 출현빈도 및 연관 단어 빈도 및 연결정도 중심성 분석을 진행하였다. CONCOR 군집 분석을 통해 각 음식, 서비스, 가격, 환경 요인에서 중요한 단어를 추출하였다. 그리고 이들은 위 4개요인 별 중요 단어의 빈도를 독립 변수로, 레스토랑 총점을 종속 변수로 다중 회귀 분석을 진행하였는데, 음식요인만이 음식점 리뷰에 유일하게 정(+)의 영향을 미친다는 결론을 도출하였다. 또한 반현정, 김학선(2019b)은 TripAdvisor에 등재된 국문리뷰를 통해 부산지역 음식점의 한국인 고객의 반응을 살펴보았다. 711개의 국문 리뷰를 살펴, 상위 100개 단어를 추출한 후 단어의 중심성을 분석하였다. 또한 웹사이트에서 제시하는 서비스 품질 평가요인(음식, 서비스,

가격, 환경)과 만족도의 관계를 선형 회귀 분석을 통해 살펴보았다. '음식', '서비스', '환경' 요인은 만족도에 정(+)의 영향을 미쳤으나 '가치'는 만족도에 유의한 영향을 미치지 않는 것을 밝혔다. 그러나 해당 연구들은 TripAdvisor 서비스에 치중되어 수집한 리뷰의 수가 부족하다는 한계점이 존재한다. 수집한 리뷰 개수의 부족 뿐 아니라 전반적인 TripAdvisor 내 음식점 리뷰 개수의 부족으로 음식점 리뷰를 확인하기에 TripAdvisor는 적합한 웹사이트가 아니다. 또한 국문리뷰의 경우, 수집된 텍스트가 명사로 이루어져 있어 단어의 추가적 의미 분석이 힘들다는 점에서 또 다른 한계가 있다.

텍스트에서 나타나는 만족도와 그에 대한 원인을 파악하고자 할 때, 감성 분석만을 활용한 연구는 선정한 요인 내 단어들의 연관성 혹은 각 요인 내 연관성을 파악하기 힘들다는 한계점이 있다. 네트워크 분석만을 활용한 연구는 리뷰 텍스트에서 나타나는 만족도 요인을 파악하나, 실제 텍스트에서 만족도가 얼마나 나타나고 있는지 확인하지 않았다는 한계점이 있다.

감성 분석과 네트워크 분석을 모두 적용하여 리뷰 텍스트를 확인한 연구로는 앞서 소개한 Gao, Tang, Wan, Yin(2018)의 연구가 있으나 이들은 리뷰 텍스트에서 확인할 수 있는 만족도 요인을 네트워크화하지 않고 존재하는 음식점 간의 관계를 네트워크화 하였다. 동일한 텍스트 데이터에 대해 감성 분석과 네트워크 분석을 적용하는 연구로는 최근 문헌정보학 영역에서 LDA토픽모델링, 네트워크 분석, 감성 분석 방법을 모두 적용하여 병명 개정 이후 조현병에 대한 사회적 인식 변화 파악을 시도한 김현지, 박서정, 송채민, 송민(2019)의 연구가 유

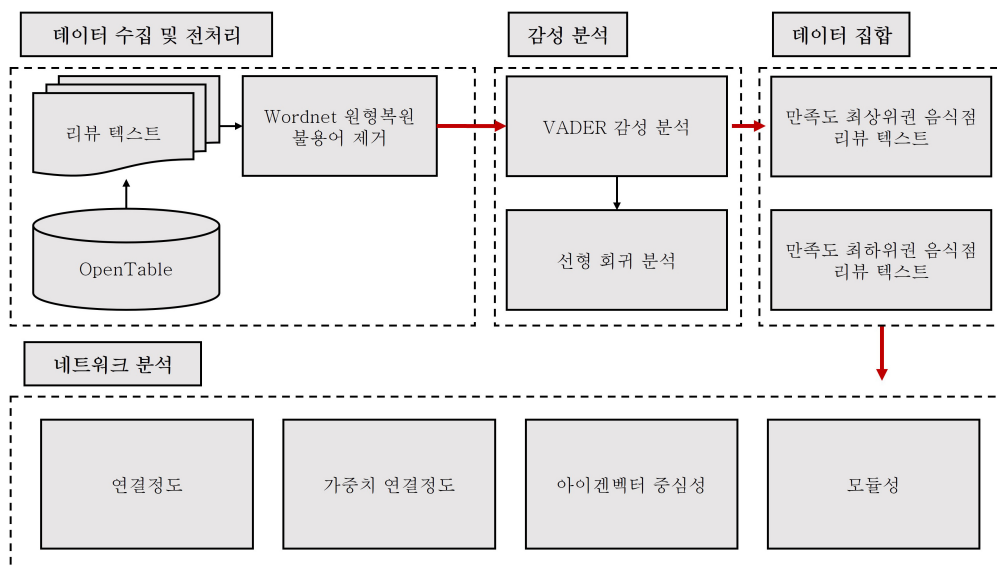
일하다. 이 연구는 다양한 텍스트 마이닝의 기법과 뉴스 프레임 유형을 활용하여 기사 내용을 분석했다는 점에서 의의가 있다. 그러나 네트워크 분석의 경우 연결정도 중심성만을 대상으로 네트워크 분석을 수행하였고 감성분석의 경우 병명개정 전 후의 평균 극성값, 부정 단어 출현 횟수만을 확인했다. 주제에 대한 긍정, 부정의 정도와 해당 원인을 파악하는데 분석 결과는 개별적으로 이루어졌고 분석 내용이 통합적으로 이루어지지 않았다.

따라서 본 연구에서는 리뷰 텍스트에서 나타나는 리뷰 대상의 특성을 파악하고자 만족도가 높은 음식점과 낮은 음식점의 리뷰 텍스트에 대한 비교 분석을 위해 감성 분석과 네트워크 분석을 통합적으로 진행하고자 한다. 개별 음식점의 만족도를 향상시키는 방법에서 더 나아가 어떤 음식점이 좋은 음식점인지 특성을 파악해보고자 하였으므로 이용자들의 만족도가

높은 음식점과 만족도가 낮은 음식점을 선정하여 감성 분석을 통해 텍스트에 나타난 사람들의 의견을 분석하여 실제 리뷰에서 긍정적, 부정적 평가가 얼마나 나타나는지 확인하고자 하였다. 더불어 네트워크 분석을 통해 리뷰 텍스트에서 파악할 수 있는 해당 음식점의 특성을 살펴봄으로써 고객들이 생각하는 좋은 음식점의 판단 기준과 음식점 만족도에 영향을 미치는 주요인을 종합적으로 확인하고자 한다.

3. 연구 설계

본 연구는 리뷰 텍스트를 통해 나타나는 긍정적 리뷰 대상의 특징을 파악하기 위해 <그림 1>과 같이 연구를 설계하였으며 데이터 수집, 전처리, 감성분석, 네트워크 분석으로 구성된다. 각 처리 과정은 다음 절부터 상세히 서술한다.



<그림 1> 연구 설계

3.1 데이터 수집

본 연구에서는 5개의 음식점을 표본으로 선정하여 진행하였다. 관광지는 관광객의 수요와 여행객의 특성에 따라 서로 다른 특색을 나타낼 수 있으므로, 관광지에 위치한 음식점이 아닌 사람들이 일상에서 방문하기 쉬운 음식점으로 선정하였다. OpenTable에 등재되어있는 지역 중 음식점 상권이 활성화 되어있는 비관광지 지역인 미국 코네티컷 주에 위치한 음식점을 조사했다. 그 중에서도 리뷰 1,000개 이상의 음식점 중 별점 점수 최상위 음식점 2개, 그리고 별점 점수 최하위 음식점 3개를 선정하였다. 별점 점수가 낮은 음식점 중에서는 리뷰가 많은 음식점이 부족하여 분석할 텍스트 리뷰 수를 늘리기 위하여 최상위 음식점보다 더 많은 음식점을 분석했다. 따라서 The Capital Grille, Union League Cafe, Lakeview Restaurant, Kokomo's Restaurant, Temple Grill 5개 음식점에 대하여 각 음식점이 OpenTable에 등재된 시점부터 2021년 6월 13일까지 업로드 된 총 4,541개의 리뷰 텍스트를 수집하였다.

3.2 데이터 전처리

효과적인 분석을 위해서 수집한 리뷰 텍스트 내 단어들을 WordNet(Princeton University, 2010)을 활용하여 원형복원을 진행하였다. WordNet은 영어 어휘 데이터베이스로 명사, 동사, 형용사, 부사가 신셋이라고 불리는 동의어 집합으로 이루어져 있다. 각 단어는 소문자로만 구성된, 문법적으로 활용되지 않은 원형(lemma)이 등재되어 있다. 이 기능을 사용한다면 단수, 복

수 형태, 동사의 시제, 소유격 등 문법적 요인으로 인한 동일한 단어의 활용형을 모두 같은 원형의 형태로 복원할 수 있다. 예를 들어, 주어와 수 일치에 위하여 변형된 동사 am, are, is는 모두 be로 복원된다. 본 연구에서는 다른 형태로 활용된 동일한 단어를 모두 하나로 인식하기 위하여 WordNet의 원형복원 기능을 사용하였다.

불용어(Stop Words)는 다양한 텍스트에서 공통적으로 자주 등장하지만 문헌의 주제와 관련 없는 단어이다. 즉, 영어의 관사, 전치사와 같이 자주 사용되지만 주요한 의미를 지니지 않은 단어를 의미한다. 불용어는 서로 다른 문헌의 차이를 파악하는데 도움이 되지 않는 정보이다. 따라서 본 연구에서는 파이썬 NLTK 패키지에서 제공하는 영문 불용어 코퍼스를 활용하여 불용어 제거 후 분석하였다.

3.3 감성 분석

감성 분석은 자연어에서 나타나는 주관성, 태도, 의견 등을 확인하고 정량화하여 분석하는 기법이다. 데이터의 주관성과 객관성을 구분한 후 사전 혹은 머신러닝을 활용하여 데이터에 나타나고 있는 극성을 파악한다. 감성 사전은 각 단어와 해당 단어의 긍정 부정의 정도를 나타내는 극성값 등으로 구성되어 있다. 사전 기반으로 극성을 파악할 때, 문헌에 있는 각 단어의 극성값을 사전에서 파악하고, 이를 활용하여 문헌 전체의 극성값을 계산한다.

본 연구에서는 감성 분석을 통해 텍스트에 나타난 사람들의 의견을 분석하여 리뷰 텍스트에서 리뷰 대상에 대한 긍정적, 부정적 평가가 얼마나 나타나는지 확인하고, 별점 기반 최상위의

만족도를 보이는 음식점과 최하위의 만족도를 보이는 음식점의 리뷰에서는 이용자들의 의견이 어떻게, 얼마나 상이하게 나타나는지 알아보고자 감성 분석을 적용하였다.

이를 위해 다양한 감성 사전 중 소셜 미디어에서 나타나는 짧은 형식으로 된 글의 감성 분석에 효과적인 VADER(Hutto & Gilbert, 2014)를 활용하였다. VADER는 질적, 양적 기법, 기본 문법 규칙 등을 복합적으로 활용하여 개발된 감성 사전이다. VADER 감성 사전은 각 단어에 긍정, 부정, 중립을 라벨링하고, 단어의 긍정과 부정을 -4(가장 부정적)와 4(가장 긍정적) 사이의 극성값으로 분류한다. 또한 느낌표 등의 구두점, 대문자 사용 등을 복합적으로 확인하여 문헌의 극성값을 계산한다. 감성 분석의 평가는 실제 사람의 분석 결과와 얼마나 비슷한가를 기준으로 행해지는데, Hutto, Gilbert (2014)에 따르면 VADER의 소셜미디어 텍스트, 영화 리뷰, 제품 리뷰, 뉴욕 타임즈 사설에 대한 감성 분석은 인간의 감성 분석만큼 또는 그 이상 정확하다.

감성 분석의 결과 통해 앞서 선정한 만족도가 최상위권인 음식점과 상대적으로 만족도가 낮은 음식점의 별점의 차이가 리뷰 텍스트에서도 유의미하게 나타나고 있는지 단순 회귀 분석을 통해 검증하였다. 단순 회귀 분석은 두 변수의 값을 비교하는 통계기법으로, 독립 변수와 종속 변수 간 인과관계를 확인할 때 사용된다.

3.4 네트워크 분석

단어 쌍을 네트워크화하여 분석하는 네트워크 분석 기법은 두 개의 단어가 문헌 내에서 함께 출현하는 횟수를 기반으로 단어의 연관성과

분석하고자 하는 텍스트 데이터의 특성을 파악하는 분석 기법이다. 이 기법은 해당 문헌으로부터 주제어와 그들 간의 관계를 파악하는데 용이하게 사용된다.

본 연구에서는 이용자의 만족도 정도에 따라 리뷰 텍스트에서 중점적으로 다루어지는 내용의 차이를 확인하기 위해 네트워크 분석을 적용하고자 한다. 최상위의 만족도를 보이는 음식점과 최하위의 만족도를 보이는 음식점의 리뷰 텍스트에서 나타나는 이용자들의 인식을 구체적으로 파악하여 어떠한 특성이 이용자의 만족도에 영향을 미치는지 확인하고자 하였다.

네트워크 분석에서 네트워크의 전체 구조와 결합성을 분석하기 위해서 다양한 지표를 확인한다. 본 연구에서는 연결정도(degree), 가중치 연결정도(weighed degree), 아이겐벡터 중심성(Eigenvector centrality)을 이용했다. 연결정도는 특정 노드에 연결되어 있는 다른 노드의 수로 문헌 내에서 얼마나 많은 다른 단어와 함께 사용되었는지 나타낸다. 특정 노드와 연결된 노드 쌍의 관계 강도를 고려하기 위하여 가중치 연결정도도 함께 확인하였다. 중심성이란 네트워크 전체 구조 내에서 중요하고 영향력 있는 노드를 파악하기 위한 지표이다. 본 연구에서는 연결된 다른 노드의 중심성도 고려하기 위하여 아이겐벡터 중심성을 확인하였다.

네트워크 내부에서 비슷한 특성을 가진 노드끼리 군집을 이루기도 한다. 군집을 살펴보는 방법으로 Blondel, Guillaume, Lambiotte, Lefebvre (2008)의 모듈성(modularity) 알고리즘이 있다. 이를 활용하면 큰 네트워크 내부의 높은 모듈성을 찾아 네트워크 내부에 있는 군집들을 파악할 수 있다.

본 연구에서는 각 음식점의 특성을 파악하는 것이 아니라 만족도가 높은 음식점과 만족도가 낮은 음식점의 특성을 비교하고자 별점이 최상위권인 음식점인 The Capital Grille, Union League Cafe 두 개를 묶어 하나의 집합으로 분석을 진행하였고, 별점이 최하위권인 음식점인 Lakeview Restaurant, Kokomo's Restaurant, Temple Grill Restaurant를 묶어 하나의 '만족도 최하위권 음식점' 집합으로 분석하였다. 리뷰 텍스트로부터 단어 쌍을 추출한 후 시각화를 위한 입력 파일로 .csv 파일을 만들었다. 시각화 툴 Gephi(Bastian, Heymann, & Jacomy, 2009)를 활용하여 네트워크 분석을 통해 연결정도, 가중치, 아이겐벡터 중심성 계산을 통해 계산한 위세 중심성, 모듈성 등을 확인하여 리뷰 텍스트에서 나타나는 음식점의 특성을 파악하였고, 이를 통해 이용자들이 어떠한 특성을 가진 음식점들이 좋은 음식점으로 여기는지, 혹은 나쁜 음식점으로 여기는지 확인해보았다.

4. 실험 결과 분석

4.1 감성 분석

VADER 감성 분석 사전을 활용하여, 선정된

5개 음식점의 리뷰 텍스트에서 나타난 사람들의 의견을 분석하여 해당 텍스트 정보에 담긴 감정을 positive(긍정), negative(부정), neutral(중립)으로 나누어 각각의 비율로 표현해주었다. 이를 통해, 실제 리뷰에서 긍정적, 부정적 평가가 얼마나 나타나는지 확인했다.

〈표 1〉은 표본으로 설정한 5개 음식점의 리뷰에 대한 감성 분석 결과를 나타낸다. 〈표 1〉의 좌측 2개의 음식점이 별점 점수 최상위권 음식점, 우측 3개의 음식점이 별점 점수 최하위권 음식점이다. 모든 음식점의 리뷰 텍스트에서 부정 점수보다 긍정 점수가 높게 나타났다. 이는 이용자들이 리뷰 작성 시 부정 언어보다 긍정 언어를 많이 사용한다는 Tian, Lu, McIntosh (2021)의 연구 결과와 일치한다. 별점 점수가 높은 음식점의 경우 긍정 점수가 모두 0.35 이상이었지만 별점 점수가 낮은 음식점은 0.25에 미치지 못하였다. 부정 점수는 모든 음식점의 리뷰 텍스트에서 가장 낮게 나타났으나 별점 점수 최상위권 음식점에 비해 별점 점수 최하위권 음식점의 부정 점수의 차이는 대부분 두 배 이상으로 확실하게 나타났다.

별점 점수의 차이가 리뷰 텍스트에서 나타나는 이용자의 만족도 차이와 유사하게 나타나는지, 감성 분석의 결과로 얻는 긍정 및 부정 점수와 별점 점수 사이에 유의미한 연관이 있는지

〈표 1〉 각 음식점의 별점, 리뷰 개수 및 감성 분석결과

	Capital Grille	Union League	Lakeview	Kokomo's	Temple Grill
ratings	4.8	4.8	3.8	3.8	3.2
total reviews	1988	2193	160	183	17
negative	0.024	0.003	0.069	0.068	0.04
neutral	0.605	0.609	0.698	0.701	0.776
positive	0.371	0.388	0.233	0.232	0.184

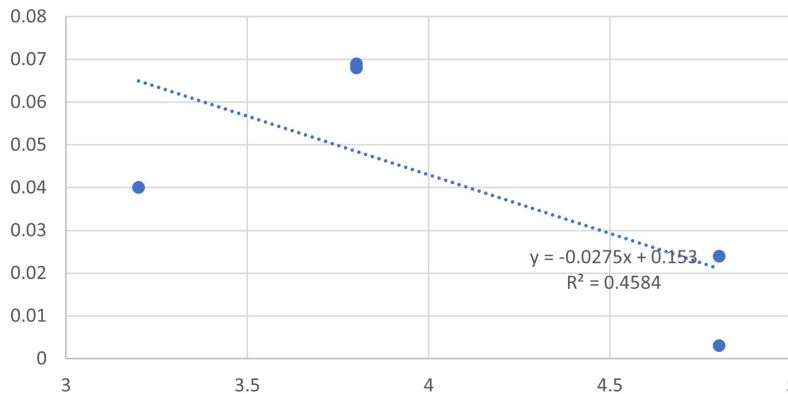
확인하기 위하여 회귀 분석으로 검증하였다. 이를 위하여, 독립 변수는 음식점의 별점 점수, 종속 변수는 음식점의 긍정 혹은 부정 점수로 설정하였다.

〈그림 2〉는 별점 점수와 리뷰 텍스트의 감성 분석 부정 점수 간의 회귀 분석 결과를, 〈표 2〉는 두 변수 사이의 분산분석 결과를 나타낸다. 별점 점수와 부정 점수, 두 변수 사이 회귀 분석 결과를 보면 결정계수가 0.4584이다. 그러나 두 95%의 신뢰 수준으로 F-test를 진행하였을 때, 유의미한 F의 값이 0.209315로 나타났는데, 이는 0.05보다 크므로 해당 회귀식이 유의하지 않다. 즉, 별점 점수와 리뷰 텍스트에서 확인되는 부정적 감정은 인과관계가 없다.

〈그림 3〉은 별점 점수와 리뷰 텍스트의 감성

분석 긍정 점수 간의 회귀 분석 결과이고 〈표 3〉은 두 변수 사이의 분산분석 결과를 나타낸다. 두 변수 간의 결정계수는 0.9767으로 표본의 좌표가 거의 일직선상에 위치하고 있다. 95%의 신뢰수준으로 F-test를 진행하였을 때, 유의미한 F의 값이 0.0015로 0.05보다 작게 나타나므로 두 변수 사이의 회귀식이 유의하다. 즉, 별점 점수가 높을수록 리뷰 텍스트 내에서 긍정적인 내용이 더 많이 나타난다. 따라서 별점 점수가 높은 음식점을 만족도가 높은 음식점, 별점 점수가 낮은 음식점을 만족도가 낮은 음식점으로 바라볼 수 있으며, 별점 점수 최상위권 음식점의 리뷰 텍스트와 별점 점수 최하위권 음식점의 리뷰 텍스트를 비교하는 것은 만족도 최상위권 음식점과 만족도가 최하위권

ratings - negative 회귀 분석

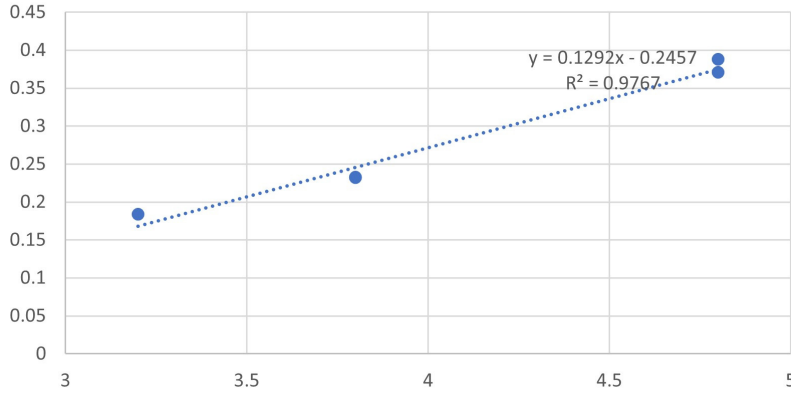


〈그림 2〉 별점 점수와 리뷰 내 부정 점수의 회귀 분석 그래프

〈표 2〉 별점 점수와 리뷰 내 부정 점수의 분산분석

	자유도	제공합	제공 평균	F 비	유의한 F
회귀	1	0.001488	0.001488	2.539039	0.209315
잔차	3	0.001759	0.000586		
계	4	0.003247			

ratings - positive 회귀 분석



<그림 3> 별점 점수와 리뷰 내 긍정 점수의 회귀 분석

<표 3> 별점 점수와 리뷰 내 긍정 점수의 분산분석

	자유도	제곱합	제곱 평균	F 비	유의한 F
회귀	1	0.032876	0.032876	125.529	0.001524
잔차	3	0.000786	0.000262		
계	4	0.033661			

음식점의 차이를 파악하는데 적합하다.

4.2 네트워크 분석

음식점에서의 만족도 차이가 나타나는 이유를 확인하기 위하여 만족도 최상위권 음식점과 최하위권 음식점의 리뷰 텍스트에서 나타나는 내용의 차이를 확인하였다. 이를 위하여 시각화툴 Gephi(Bastian, Heymann, & Jacomy, 2009)를 사용하여 네트워크 분석을 진행한 결과는 다음과 같다.

<표 4>는 만족도가 최상위권인 음식점 리뷰에 사용된 단어들 중 위세 중심성이 높은 상위 30개 단어에 대한 연결정도, 가중치 순위 및 수치, 위세 중심성 순위 및 수치를 분석한 결과이

다. 상위 30개 단어 중 긍정적인 형용사 혹은 부사는 노란색으로 표시한 단어들로 great, nice, delicious, perfect, perfectly, excellent, happy 총 7개 포함되었다. 음식 관련 단어는 초록색으로 표시한 단어들로 food, dessert, meal, side, main, course, salad 총 7개 포함되었다. 이 외에도 파란색으로 표시한 서비스에 관련된 3개의 waiter, service, staff 단어도 각 19위, 20위, 24위로 포함되었다. 특히 food와 great은 연결정도, 가중치, 위세 중심성에서 모두 1, 2위를 차지하였다. 이는 해당 두 단어가 만족도 최상위권 음식점의 리뷰 텍스트에서 다른 단어들과 가장 자주 사용되었고 가장 높은 영향력을 가지고 있다는 것을 의미한다. <표 4>에서 dessert가 위세 중심성 4위, 가중치빈도 12위인 것 또한 주목할 만하

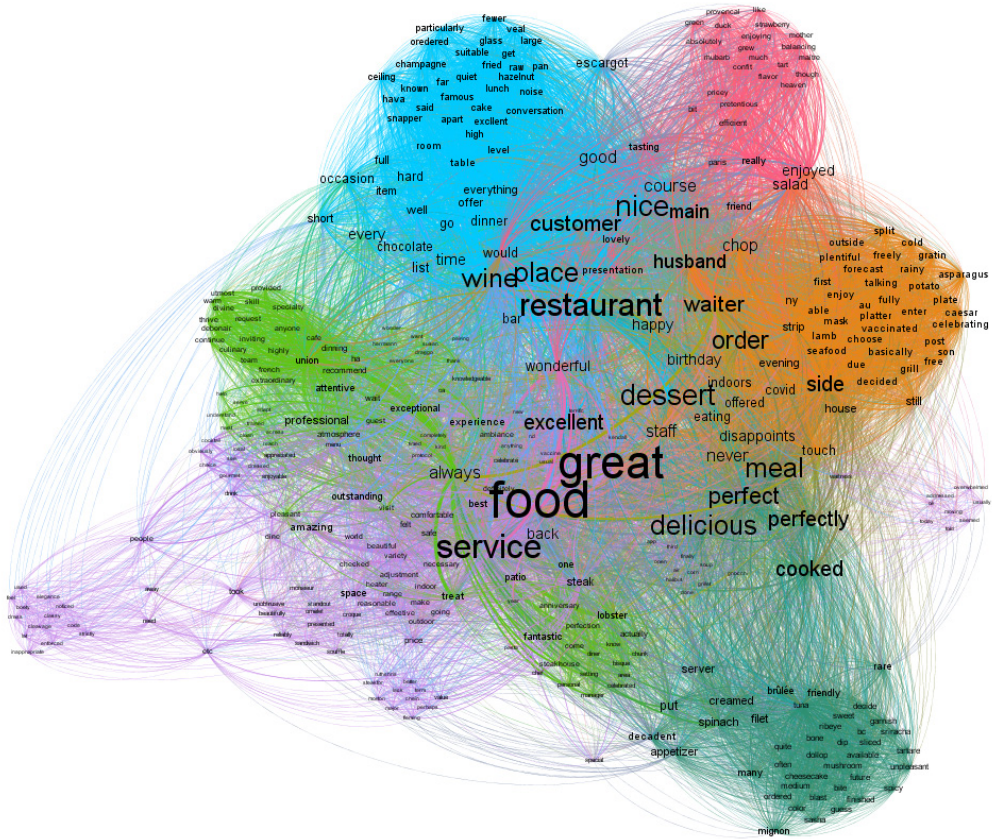
〈표 4〉 만족도 최상위권 음식점 리뷰의 연결정도, 가중치, 위세 중심성

단어	연결정도	가중치		위세 중심성	
		순위	수치	순위	수치
food	610	1	1060	1	1
great	566	2	1020	2	0.961185
restaurant	412	3	802	3	0.956405
dessert	372	12	478	4	0.906868
order	318	9	552	5	0.869843
nice	388	5	754	6	0.859263
place	362	7	626	7	0.836261
meal	358	10	526	8	0.817211
delicious	354	13	452	9	0.795193
customer	296	17	378	10	0.789627
perfect	316	15	430	11	0.751127
husband	262	4	766	12	0.748118
side	286	11	502	13	0.737239
main	256	23	320	14	0.734898
course	248	25	318	15	0.72776
perfectly	282	20	348	16	0.725567
chop	236	14	446	17	0.716559
cooked	282	19	354	18	0.711524
waiter	296	16	386	19	0.693203
service	444	6	668	20	0.664059
wine	350	8	612	21	0.650574
excellent	288	18	370	22	0.628196
never	248	23	320	23	0.617837
staff	238	26	294	24	0.589405
enjoyed	202	21	328	25	0.576171
salad	202	21	328	25	0.576171
eating	194	29	236	27	0.575468
disappoints	200	27	246	28	0.574399
happy	200	28	240	29	0.570719
indoors	192	30	222	30	0.569553

다. 우리나라는 후식을 별도로 먹는 문화인 반면 한 음식점에서 식사를 마친 후 후식을 주문하는 문화권에서는 식사 혹은 음식점에서의 경험이 불만족스러울 경우 해당 음식점에서 디저트를 주문하지 않고 떠나게 될 확률이 높다. 네트워크 분석 결과 dessert에 대한 내용이 음식점의 리뷰

에서 상위 단어로 확인된다는 것은 해당 음식점이 만족스러웠다는 점을 의미한다.

〈그림 4〉는 Gephi를 활용하여 만족도가 높은 음식점의 리뷰 텍스트를 시각화한 네트워크 분석을 나타낸다. 각 노드 레이블의 크기는 리뷰 텍스트에 나타나는 각 단어의 빈도를, 엣지



<그림 4> 만족도 최상위권 음식점 리뷰의 네트워크 시각화

는 각 노드의 연결을 의미한다. 동일한 색상의 엣지로 연결된 단어는 동일한 모듈성 커뮤니티에 속하는 단어들로 리뷰 텍스트 내에서 주로 함께 출현하였음을 확인할 수 있다.

<표 4>에서도 확인할 수 있었듯이 great, food는 단어 빈도수도 높고 네트워크의 중심에 나타나며, 여러 색상의 선과 연결되어, 다양한 모듈성 커뮤니티에 속해 있는 것을 확인할 수 있었다. 이는 이들이 특정 리뷰에만 사용된 것이 아니라 다수의 리뷰에 대부분 포함되어 있었음을 나타낸다. 이 외에도 service, dessert, delicious의 사용빈도가 높고, 네트워크의 중심부에 위

치하고 있으며, excellent, wonderful은 다양한 모듈성 커뮤니티에 속해있다는 점을 확인할 수 있다.

<표 5>는 만족도가 낮은 음식점 리뷰에 사용된 단어 중 위세 중심성이 높은 상위 30개 단어에 대한 연결정도, 가중치 순위 및 수치, 위세 중심성 순위 및 수치를 분석한 결과이다. 이 중 긍정적인 형용사 혹은 부사는 노란색으로 표시한 2개의 great, good 단어가 포함되었다. 음식과 관련된 단어는 초록색으로 표시한 단어로 food, meal, appetizer, roll 총 4개 포함되었다. 서비스에 관한 단어는 파란색으로 표현한 두

〈표 5〉 만족도 최하위권 음식점 리뷰의 연결정도, 가중치, 위세 중심성

단어	연결정도	가중치		위세 중심성	
		순위	수치	순위	수치
drink	702	2	2238	1	1
came	618	1	2382	2	0.941604
table	604	7	1560	3	0.918653
time	628	6	1564	4	0.912256
food	704	3	2198	5	0.860263
hour	522	17	1224	6	0.847106
great	578	5	2008	7	0.827081
order	486	13	1288	8	0.821768
one	498	9	1524	9	0.821013
sat	458	23	954	10	0.803995
would	522	15	1248	11	0.793335
never	456	11	1378	12	0.792784
good	514	4	2188	13	0.773661
server	496	22	1026	14	0.769396
first	424	14	1284	15	0.758347
minute	410	10	1404	16	0.750949
service	520	18	1166	17	0.746859
without	394	16	1244	18	0.736725
meal	456	21	1078	19	0.733617
plate	382	24	938	20	0.721483
took	368	28	722	21	0.705791
around	352	26	874	22	0.697916
back	448	25	900	23	0.692738
get	452	19	1148	24	0.677283
restaurant	440	12	1364	25	0.669291
appetizer	388	29	676	26	0.664899
place	434	8	1558	27	0.658007
even	418	20	1146	28	0.654135
roll	316	30	556	29	0.647396
dining	318	27	822	30	0.645520

단어로 server, service가 위세 중심성 상위 30개 단어 안에 포함되었다. 서비스와 관련된 단어가 food를 제외한 음식에 관련된 단어보다 위세 중심성 수치 상위에 나타났다. 이로 인하여, 음식보다 서비스에 대한 내용이 리뷰에 더 중요하게 등장했다는 것을 확인할 수 있다.

만족도가 높은 음식점의 리뷰와 마찬가지로

만족도가 낮은 음식점의 리뷰 텍스트에 등장한 단어들의 네트워크 분석 결과를 〈그림 5〉로 표현하였다. 각 노드 레이블의 크기로 리뷰 텍스트에 나타나는 각 단어의 빈도를 시각화하였고, 엣지는 각 노드의 연결을 표현하였다. 또한 엣지의 색상으로 각 단어가 속한 모듈성 커뮤니티를 확인할 수 있다.

긍정적인 단어는 별점 점수에 따라 빈도와 중요성이 크게 차이나지만 부정적인 단어의 경우, 주목할 만한 큰 차이는 나타나지 않았다. 또한, 별점 점수 최상위권 음식점의 리뷰에서는 음식에 관한 이야기가 주를 이루었고 서비스에 관한 이야기의 위세 중심성이 그보다 낮았다. 반면에, 별점 점수가 최하위권 음식점의 리뷰에서는 음식에 관한 단어보다 서비스에 관한 단어의 위세 중심성이 더 높게 나타났다. 즉, 사람들의 만족도가 더 높은 음식점은 음식에 대한 사람들의 만족도가 높고, 서비스 또한 우수하다는 것을 확인할 수 있다.

5. 결론

본 연구는 감성 분석과 네트워크 분석을 통합적으로 활용하여 온라인 음식점 리뷰를 분석하였다. 감성 분석을 활용하여 실제 텍스트에서 만족도가 얼마나 나타나고 있는지를 확인하였고, 이러한 만족도가 어떠한 요인에 의해 나타나는지 네트워크 분석을 통해 파악하였다. 주요 단어들 기반의 분석은 연구의 결과를 객관적으로 입증할 수 있도록 네트워크 분석의 특정 척도가 아닌 주요 척도인 연결정도, 가중치, 위세 중심성을 공통적으로 살펴보았다. 이를 통해 만족도 최상위권의 음식점과 만족도 최하위권의 음식점의 리뷰 텍스트에서 나타나는 차이를 확인할 수 있었다. 본 연구의 연구 질문에 대한 결과는 다음과 같다.

- 연구 질문 1. 감성 분석 결과 만족도 최상위권의 음식점 리뷰 텍스트의 긍정 점수는

0.3 이상, 부정 점수는 0.03 미만으로 확인되었다. 반면, 만족도 최하위권 음식점 리뷰 텍스트는 긍정 점수 0.3 미만, 부정 점수는 0.03 이상으로 나타났다. 만족도에 관계 없이 리뷰에서 나타나는 내용은 중립적인 내용, 긍정적인 내용, 부정적인 내용 순으로 많이 나타났으나, 그 구성 비율에서 차이가 확연하게 나타났다. 만족도 최상위권 음식점에 비해 만족도 최하위권 음식점의 부정 점수가 대부분 두 배 이상 높게 나타났다.

- 연구 질문 2. 별점 기반 음식점에 대한 만족도와 리뷰의 감성 분석 결과는 긍정 점수에 있어 통계적으로 유의한 인과성을 보였다. 그러나 리뷰의 별점 점수와 부정 점수의 관계는 통계적으로 유의한 인과성을 보이지 않았다. 별점 점수가 높은 음식점은 만족도가 높은 음식점이며 별점 점수가 낮은 음식점은 만족도가 낮은 음식점으로 파악할 수 있으며 리뷰 텍스트 기반으로 두 집단을 구분하여 비교하는 것은 만족도의 차이를 파악하는데 유의미하다는 점을 확인했다.
- 연구 질문 3. 연결정도 상위 단어의 차이는 리뷰 텍스트에서 주로 나타나는 토픽의 차이를, 위세 중심성 상위 단어의 차이는 각 리뷰 텍스트에서 중요하게 다루고 있는 토픽의 차이를 보여준다. 만족도 최상위권 음식점의 리뷰 텍스트에서는 음식 자체에 대한 이야기가 주로 나타났고, 만족도 최하위권 음식점은 음식보다는 서비스에 대한 내용이 많았다. 특히, 만족도 최상위권 음식의 리뷰 텍스트에서는 ‘음식’

이라는 단어에 연결 정도, 가중치, 위세 중심성이 모두 가장 높았다. 또한 만족도 최상위권 음식점의 리뷰 텍스트에서 긍정적인 형용사와 부사가 더 자주 중요하게 사용되었다. 만족도가 더 높은 음식점의 리뷰 텍스트에서는 음식에 대한 사람들의 만족도가 더 중요하게 나타나고 있었지만 만족도 최하위권 음식점의 리뷰 텍스트의 경우 음식 관련 내용에 중심성이 떨어졌다. 만족도가 낮은 음식점의 경우 음식에 관한 단어보다 서비스에 관한 단어의 위세 중심성이 더 높게 나타나면서 리뷰 텍스트에서 확인할 수 있는 이용자들의 주 인식이 서비스에 대한 내용임을 확인할 수 있었다.

본 연구에서는 만족도 최상위권 음식점의 리뷰 텍스트와 만족도 최하위권 음식점의 리뷰 텍스트에서 나타나는 차이점을 확인하였다. 도출된 결과는 음식점 마케팅을 위하여 리뷰 사이트에 리뷰를 등록할 때 다른 장점보다 음식에 대한 만족도를 더욱 부각시키는 등 마케팅 전략 및 음식점 운영에 전반적으로 도움이 되는 기초 자료로 활용 가능하다. 또한 음식점을 방문하는 이용자들도 음식점을 방문할 때 음식점과 리뷰 내용의 특성을 기반으로 만족도에 영향을 미치는 주요 요인을 파악하여 음식점을

방문한다면 음식점 선별의 질적 기준이 높아질 수 있고 결과적으로 음식점 이용 경험의 만족도가 향상될 수 있을 것이다. 특히 리뷰 서비스에서 별점 점수, 수치화의 부재로 만족도를 확인하기 어려울 때, 만족도 최상위권의 음식점은 리뷰 텍스트의 중심이 음식점의 다른 요인에 분산되는 것보다 음식에 집중된다는 점을 확인한다면 음식점 선별에 도움이 될 수 있을 것이다. 더불어 정보학적인 측면에서 방대한 양의 음식점 리뷰 텍스트에서 나타나는 만족도와 만족도 요인을 자동으로 분석함으로써 텍스트에 내재된 가치를 다양한 측면으로 활용할 수 있는 가능성을 제시하였다는 점에서 연구의 시사점을 지닌다.

본 연구에서 활용한 리뷰 등재 사이트인 Open Table이 음식점 예약 사이트라는 특성상 별점 점수가 3.0/5.0 이하로 낮은 음식점은 없었고 상대적으로 별점 점수가 낮은 음식점의 리뷰 또한 별점 점수가 높은 음식점의 리뷰에 비해 많이 부족했다. 따라서 향후 연구에서는 음식점의 제약이 적은 SNS의 리뷰를 활용하여 분석한다면 더욱 폭넓은 결론을 얻을 수 있을 것이다. 또한 본 연구에서는 하나의 지역에 있는 5개의 음식점만 표본으로 활용하여 비교하였지만 추후 연구에서는 타 지역에 위치한 음식점까지 포함하여 음식점 리스트를 확장하면 더욱 정확한 결과를 얻을 수 있을 것이다.

참 고 문 헌

- [1] 김지현 (2016). 언어네트워크 분석방법을 활용한 공공정보 내용분석: 광역도시 대표 트위터 내용을 중심으로. 한국비블리아학회지, 27(3), 151-171.

- 출처: <https://doi.org/10.14699/KBIBLIA.2016.27.3.151>
- [2] 김현지, 박서정, 송채민, 송민 (2019). 조현병과 정신분열병에 대한 뉴스 프레임 분석을 통해 본 사회적 인식의 변화. *한국문헌정보학회지*, 53(4), 285-307.
출처: <https://doi.org/10.4275/KSLIS.2019.53.4.285>
- [3] 문지영 (2021. 3. 17.). 네이버, 식당·카페 “별점” 없앤다. “중소상공인 고충 줄이려”. YTN.
출처: https://www.ytn.co.kr/_ln/0102_202103171050020731
- [4] 반현정, 김학선 (2019a). 트립어드바이저 리뷰 분석을 통한 부산지역 레스토랑 인식에 관한 연구 1: 영문 리뷰를 중심으로. *한국조리학회지*, 25(4), 1-11.
- [5] 반현정, 김학선 (2019b). 온라인 리뷰 분석을 통한 부산지역 레스토랑 인식에 관한 연구 2: 국문 리뷰를 중심으로. *산업혁신연구*, 35(2), 185-207.
- [6] 소진수, 신판섭 (2020). 음식점 리뷰 감성분석을 통한 세부 평가항목별 평점 예측. *한국컴퓨터정보학회논문지*, 25(6), 81-89.
- [7] 송중석, 이수원 (2011). 상품평 극성 분류를 위한 특색별 서술어 긍정/부정 사전 자동 구축. *정보과학회논문지: 소프트웨어 및 응용*, 38(3), 157-168.
- [8] 이상재, 최준연, 최진호 (2014). 온라인 리뷰의 경제적 효과, 유용성과 유용성 투표수에 영향을 주는 결정요인. *한국IT서비스학회지*, 13(1), 43-55.
출처: <https://doi.org/10.9716/KITS.2014.13.1.043>
- [9] 이시환, 이훈영 (2018). 레스토랑 리뷰의 감성표현이 정보의 인지된 유용성과 인지된 즐거움에 미치는 영향. *한국관광학회 국제학술발표대회집*, 83(0), 113-118.
- [10] 정지나, 정희찬, 김용 (2017). 언어네트워크 분석을 통한 재난안전정보와 관련한 국내 연구동향 분석. *한국비블리아학회지*, 28(3), 67-93.
출처: <https://doi.org/10.14699/KBIBLIA.2017.28.3.067>
- [11] 진설아, 허고은, 정유경, 송민 (2013). 트위터 데이터를 이용한 네트워크 기반 토픽 변화 추적 연구. *정보관리학회지*, 30(1), 285-302.
- [12] 허고은, 송민 (2013). 저자동시인용 분석과 동시출현단어 분석을 이용한 의료정보학 저널의 지적 구조 분석. *정보관리학회지*, 30(2), 207-225.
출처: <https://doi.org/10.3743/KOSIM.2013.30.2.207>
- [13] Agüero-Torales, M., Cobo, M., Herrera-Viedma, E., & López-Herrera, A. (2019). A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor. *Procedia Computer Science*, 162, 392-399. Available: <https://doi.org/10.1016/j.procs.2019.12.002>
- [14] Bastian M., Heymann S., & Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*.
- [15] Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of

- communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.
- [16] Gang, Xu (2020). Customer Sentiment Analysis: Take Restaurant Online Reviews as an Example. *Asia-pacific Journal of Convergent Research Interchange*, 6(6), 25-33.
- [17] Gao, S., Tang, O., Wang, H., & Yin, P. (2018). Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *International Journal of Hospitality Management*, 71, 19-32. Available: <https://doi.org/10.1016/j.ijhm.2017.09.004>
- [18] Hossain, E., Sharif, O., Hoque, M. M., & Sarker, I. H. (2020). SentiLSTM: A Deep Learning Approach for Sentiment Analysis of Restaurant Reviews. arXiv:2011.09684
- [19] Hutto, C. & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225.
- [20] Jo, A-Ra, Kim, Tae-Ho, & Kim, Hak-Seon (2017). Exploratory Study of Restaurant Online Reviews Using Semantic Network Analysis. *한국관광레저학회 학술발표대회*, 199-210.
- [21] Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, 39(5), 6000-6010. Available: <http://doi.org.ssl.access.yonsei.ac.kr10.1016/j.eswa.2011.11.107>
- [22] Kim, H., Yoon, J. W., & Crowcroft, J. (2012). Network analysis of temporal trends in scholarly research productivity. *Journal of Informetrics*, 6(1), 97-110. Available: <https://doi.org/10.1016/j.joi.2011.05.006>
- [23] Mathayomchan, B. & Taecharungroj, V. (2020). "How was your meal?" Examining customer experience using Google maps reviews. *International Journal of Hospitality Management*, 90, 102641. Available: <https://doi.org/10.1016/j.ijhm.2020.102641>
- [24] Milojevic, S., Sugimoto, C., Yan, E., & Ding, Y. (2011). The cognitive structure of Library and Information Science: Analysis of article title words. *J. Assoc. Inf. Sci. Technol*, 62, 1933-1953.
- [25] Princeton University (2010). About WordNet. WordNet, Princeton University.
- [26] Qiwei Gan, Bo H. Ferns, Yang Yu, & Lei Jin (2017). A Text Mining and Multidimensional Sentiment Analysis of Online Restaurant Reviews. *Journal of Quality Assurance in Hospitality & Tourism*, 18(4), 465-492. Available: DOI: 10.1080/1528008X.2016.1250243
- [27] Shuang, K., Zhang, Z., Guo, H., & Loo, J. (2018). A sentiment information Collector-Extractor architecture based neural network for sentiment analysis. *Information Sciences*, 467, 549-558. Available: <https://doi.org/10.1016/j.ins.2018.08.026>

- [28] Song, M., Heo, G. E., & Lee, D. (2015). Identifying the landscape of alzheimer's disease research with network and content analysis. *Scientometrics*, 102(1), 905-927.
Available: <https://doi.org/10.1007/s11192-014-1372-x>
- [29] Tian, G., Lu, L., & McIntosh, C. (2021). What factors affect consumers' dining sentiments and their ratings: Evidence from restaurant online review data. *Food Quality and Preference*, 88, 104060. Available: <https://doi.org/10.1016/j.foodqual.2020.104060>
- [30] Yu Boya, Zhou Jiaxu, Zhang Yi, & Cao Yunong (2017). Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews. *arXiv:1709.08698*
- [31] Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, 38(6), 7674-7682.
Available: <https://doi.org/10.1016/j.eswa.2010.12.147>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Kim, Ji-Hyun (2016). A content analysis of public information using language network analysis: focused on contents of twitters of metropolitans. *Journal of the Korean BIBLIA Society for library and Information Science*, 27(3), 151-171.
- [2] Kim, Hyunji, Park, Seojeong, Song, Chaemin, & Song, Min (2019). Text mining driven content analysis of social perception on schizophrenia before and after the revision of the terminology. *Journal of the Korean Society for Library and Information Science*, 53(4), 285-307.
- [3] Moon, Ji Young (2021. 3. 17.). Naver, Getting rid of "Star Ratings" on Restaurants and Cafe. "To ease small business owners' troubles". YTN.
https://www.ytn.co.kr/_ln/0102_202103171050020731
- [4] Ban, Hyun-Jeong & Kim, Hak-Seon (2019a). A study on the tripadvisor review analysis of restaurant recognition in Busan 1: especially concerning english reviews. *Culinary Science & Hospitality Research*, 25(4), 1-11.
- [5] Ban, Hyun-Jeong & Kim, Hak-Seon (2019b). A study on the online review analysis of restaurant recognition in Busan 2: especially concerning korean reviews. *The Journal of Business and Economics*, 35(2), 185-207.
- [6] So, Jin-Soo & Shin, Pan-Seop (2020). Rating prediction by evaluation item through sentiment analysis of restaurant review. *Journal of the Korea Society of Computer and Information*,

- 25(6), 81-89.
- [7] Song, Jong Seok & Lee, Soo Won (2001). Automatic construction of positive/negative feature-predicate dictionary for polarity classification of product reviews. *Journal of KIISS: Software and Applications*, 38(3), 157-168.
- [8] Lee, Sangjae, Choeh, Joon Yeon, & Choi, Jinho (2014). The determinant factors affecting economic impact, helpfulness, and helpfulness votes of online. *Journal of Information Technology Services*, 13(1), 43-55
- [9] Lee, Si Hwan & Lee, Hoon Young (2018). The effects of sentiment expressed in restaurant reviews on perceived usefulness and enjoyment of information. *The Tourism Sciences Society of Korea International Conference*, 83(0), 113-118.
- [10] Jeong, Ji-Na, Jeong, Him-Chan, & Kim, Yong (2017). Analysis of trends on disaster safety information based on language network analysis methods. *Journal of the Korean BIBLIA Society for library and Information Science*, 28(3), 67-93.
- [11] Jin, Seol A, Heo, Go Eun, Jeong, Yoo Kyung, & Song, Min (2013). Topic-network based topic shift detection on twitter. *Journal of the Korean Society for Information Management*, 30(1), 285-302.
- [12] Heo, Go Eun & Song, Min (2013). Examining the intellectual structure of a medical informatics journal with author co-citation analysis and co-word analysis. *Journal of the Korean Society for Information Management*, 30(2), 207-225.

