

Research on Constructing a Sentiment Lexicon for the F&B Sector based on the N-gram Framework

Yeryung Moon*, Gaeun Son*, Geonuk Nam*, Hanjin Lee**

*RA, School of ICT & GE, Handong Global University, Pohang, Korea

**Professor, School of Creative Convergence Education, Handong Global University, Pohang, Korea

[Abstract]

Online and mobile reviews strongly influence consumer behavior, especially in the service industry, and play a key role in determining customer retention and revisit rates. Systematically analyzing the information in these reviews can effectively assess how they directly influence customers' purchase decisions. In this study, we applied the existing KNU sentiment dictionary to food and beverage (F&B) review data to build a customized sentiment lexicon using N-grams based on about 10,000 reviews. Comparing its performance with the existing dictionary, we found that the sentiment lexicon generated using the 1-gram, 2-gram, and 3-gram models had the highest accuracy, precision, recall, and F1 scores. These results can serve as a powerful business support tool for SMEs in the F&B and grocery shopping sector, also be used to predict customer demand for technology and policy.

▶ **Key words:** Sentiment Analysis, Business Support, Social Impact, N-gram, Mobile Grocery Shopping

[요 약]

구매경험 후기는 온라인 및 모바일 서비스 산업에서 소비자 행동에 큰 영향을 미치며, 지속적인 이용여부를 결정짓는 중요한 요소이다. 이에 리뷰에서 제공되는 정보를 체계적으로 분석하면 고객의 구매결정에 어떻게 직접적으로 영향을 미치는지 효과적으로 평가할 수 있다. 본 연구에서는 국립국어원 기구축 KNU 감성사전을 식음료(F&B) 분야에 적용하여, N-그램 프레임워크 기반 약 10,000개의 리뷰 데이터 훈련 모델로 검증한 산업특화 감성사전을 구축하였다. 기존 사전과 성능을 비교한 결과, 1-그램, 2-그램, 3-그램 조합 기반 신규 생성된 감성사전이 가장 높은 정확도, 정밀도, 재현율, F1 점수를 나타냈다. 이 분석결과는 F&B 및 식품 부문 소상공인 관점에서 효과적인 비즈니스 지원 도구로도 활용할 수 있으며, 고객 수요예측에도 기술적, 정책적으로 활용할 수 있다.

▶ **주제어:** 감성분석, 비즈니스 지원도구, 소셜임팩트, N-그램, 모바일식품구매

- First Author: Yeryung Moon, Corresponding Author: Hanjin Lee
- *Yeryung Moon (22000247@handong.ac.kr), School of ICT & GE, Handong Global University
- *Gaeun Son (irreplaceable@handong.ac.kr), School of ICT & GE, Handong Global University
- *Geonuk Nam (nuuk@handong.ac.kr), School of ICT & GE, Handong Global University
- **Hanjin Lee (cus@handong.edu), School of Creative Convergence Education, Handong Global University
- Received: 2024. 07. 31, Revised: 2024. 09. 09, Accepted: 2024. 09. 25.

I. Introduction

모바일과 온라인 리뷰는 디지털 서비스 산업에서 소비자 행동에 큰 영향을 미치며 성공여부를 결정짓는 중요한 요소이다. F&B 산업은 레스토랑, 카페, 바와 같은 식음료 관련 서비스업을 포함하며, 소비자 경험과 리뷰가 사업 성패에 직접적인 영향을 미치는 산업으로, 많은 레스토랑과 소매점들은 고객의 반응을 신속하고 쉽게 이해하여 선호도와 만족도를 향상시키기 위해 노력하고 있다[1]. 그러나 시공간적 한계, 물적 자본과 인력 부족, 관리의 지속가능성 저하, 데이터 수집 및 분석의 어려움, 효과성 한계 등 여러 문제에 직면해 있다[2].

현재 많은 한국의 소상공인 매장들은 네이버 플레이스와 같은 플랫폼에서 자주 추출된 키워드에 의존하고 있으며, 이는 충분한 양의 리뷰가 있을 때만 효과적이다. 또한 인스타그램과 메타(구 페이스북)와 같은 소셜미디어 플랫폼에서 해시태그, 사진, 동영상상을 포함한 비구조화 데이터를 종합적으로 분석하는 것은 거의 불가능하여 최신 게시물의 영향에 의존하게 된다[3]. 이로 인해 많은 소상공인들은 고객의 반응을 실시간 이해하고, 고객이 원하는 것을 파악하며, 서비스를 개선하는 데 어려움을 겪고 있다[4].

디지털 리뷰는 고객접점 분야에서 소비자의 행동에 상당한 영향을 가져온다[4,5]. 리뷰의 다양한 측면(음식 품질, 서비스, 위생 기준 등)을 체계적으로 분석함으로써 비즈니스는 이 정보가 잠재고객의 구매결정에 어떻게 직접적으로 영향을 미치는지 효과적으로 평가할 수 있다[6]. 기존에도 음식점과 온라인 리뷰를 주제로 진행했던 선행 연구들이 있었지만, 대부분 평점 예측에 중점을 둔 소비자 리뷰와 평점 간의 상관관계 분석에 초점을 맞추거나, 평점이 음식점 순위에 어떤 영향을 미치는지 등의 연구들이 주로 진행되었다.[7,8] 그러나 현재 네이버 스마트플레이스의 경우 일부 소비자가 부당한 요구를 하거나, 개인적인 불만을 과도한 평점으로 표현하는 사례가 늘어나면서 사업자 보호와 공정한 평가를 위해 2021년 10월 26일 자로 별점 평가 시스템을 중단하고 키워드 리뷰로 대체하였다.

이에 오히려 평점 기반의 분석이 중단된 배경에서는 텍스트 기반의 감성분석이 더욱 중요해지고 있다. 그러나 현재 널리 사용되는 감성분석 도구와 최신 NLP 모델들은 일반적인 텍스트 분석에 뛰어난 성능을 보일 수 있지만, F&B 산업의 특수한 언어적 특성과 맥락을 충분히 반영하기 어렵다는 한계가 지적되었다. 또한 대규모 데이터 및 높은 연산 자원을 필요로 한다는 점에서 실용적인 제약도 클 수 있다.

따라서 본 연구는 F&B 산업의 고유한 어휘와 표현을 보다 직접적으로 다룰 수 있는 산업 특화적 감성 사전을 개발함으로써, 특정 산업에 최적화된 분석을 더 간단하고 효율적인 방식으로 제공하고자 한다.

이를 위해, N-그램(N-gram) 프레임워크를 사용하여 F&B 산업에 특화된 감성 사전을 구축하는 것을 목표로 한다. N-그램 기법은 텍스트에서 자주 사용되는 단어나 구를 분석하여 특정 산업의 맥락에 적합한 감성 분석을 가능하게 한다. 이를 통해 F&B 산업에서 자주 사용되는 고유한 어휘와 표현을 정확하게 분석하고, 리뷰 데이터를 보다 세밀하게 해석할 수 있는 도구를 제공하고자 한다.

요컨대 본 연구는 N-그램 기법을 활용한 맞춤형 감성 분석 도구를 통해, 소상공인들이 고객 피드백을 더 효과적으로 이해하고 대응할 수 있도록 돕고자 하며, 이를 통해 실제 비즈니스 환경에서의 문제를 해결하고, 감성 분석의 정확성과 효용성을 높이는 데 기여할 것이다.

II. Preliminaries

1. Problem Definition

현재 사용가능한 감성분석 도구들은 주로 일반 언어 사용에 중점을 두고 개발되었다. 이러한 도구들은 특정 산업 맥락에서 중요한 전문 용어와 맥락적 미묘함을 정확히 포착하지 못하는 경우가 많아 감성분석의 정확성이 크게 제한된다. 예를 들어, "맛이 없을 수가 없다"라는 문구는 일반 감성분석에서 "없다"라는 단어의 반복으로 인해 매우 부정적인 진술로 분류될 수 있지만, 실제로는 음식이 매우 맛있다는 고객의 긍정적인 경험을 나타내는 중요지표이다.

Table 1. Comparison of Sentiment Scores Using General and Customized Dictionaries

General Dictionary	Customized Dictionary
익명1 ... 샌드는 맛이 없을 수가 없죠. +1	익명1 ... 샌드는 맛이 없을 수가 없죠. -2
익명2 ... 박스 안에 쿠키들이 맛별로 있어요. 0	익명2 ... 박스 안에 쿠키들이 맛별로 있어요. -1
익명3 ... 맛을 고를 수 있으면 좋을 텐데요. -1	익명3 ... 맛을 고를 수 있으면 좋은 텐데요. +1

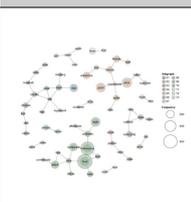
이러한 분석 부정확성은 여러 문제를 초래할 수 있다. 첫째, 고객 피드백의 오해는 실제 고객 만족도 또는 불만족도에 대한 부정확한 이해로 이어진다. 둘째, 잘못된 감성분석 결과는 비즈니스 전략 수립에서 오류를 초래하여 서비스 개선이나 마케팅 전략의 효과를 감소시킨다. 셋째, 고객 감성을 정확하게 이해하지 못하면 고객과의 효과적인 소통이 어려워져 고객 만족도가 감소할 수 있다.

산업별 맞춤 감성사전의 부족과 특정 산업 리뷰 분석의 어려움을 해결하기 위해, 이에 본 연구에서는 식음료 데이터를 활용하여 F&B 업계에 특화된 감성사전을 개발 및 평가하고자 한다. 이는 업계 특유의 어휘와 표현을 통합하여 감성분석의 정확성을 향상시킬 뿐만 아니라 리뷰 데이터에서 더 미묘한 감성을 식별할 수 있게 할 것이다.

2. Related works

이 연구의 문제정의와 직접적으로 관련된 선행연구는 두 가지 주요 방향에서 검토되었다. 첫 번째 방향은 데이터 전처리 및 분석의 정확성을 향상시키는 연구들을 검토했다. 두 번째 방향은 특정 주제의 언어적 특성과 맥락적 미묘함을 정확하게 해석하는 선행결과들을 포함한다.

Table 2. Related works review [9-11]

Korean Stopword	Co-occurrence network	Classified online reviews
		

2.1 Data Pre-processing

한국어 사전 연구는 텍스트 데이터에서 불필요한 정보를 제거하여 데이터 정리를 통해 감성분석의 정확성과 효율성을 높이는 중요한 기초를 제공한다[9]. 특히 한국어 리뷰 데이터에서 관련 없는 어휘를 식별하고 제거함으로써 더 정확하고 신뢰할 수 있는 감성 점수를 도출할 수 있다. 이 데이터 전처리 연구는 감성분석 도구의 개발 및 정교화에 필수적이며 분석의 정확성을 높이고 리뷰의 감성가치를 정확히 평가하는 데 기여한다.

2.2 Topic Modeling

기존 연구에서 살펴본 결과, 감성분석과 토픽 모델링은 리뷰 데이터에서 주제의 중요성을 분류하고 구조화하는데 필수적인 도구이다. 토픽 모델링은 대량의 텍스트 데이터에서 반복되는 패턴이나 주제를 자동으로 감지하는 기계학습 기반 알고리즘을 사용하여 리뷰 내용의 주요 주제나 트렌드를 식별한다. 예를 들어 레스토랑 리뷰에서는 "서비스", "가격", "맛", "위생", "음식"과 같은 주제가 중요한 평가기준으로 나타난다. 각 주제가 전체 리뷰에 얼마나 기여하는지 분석함으로써 소비자 만족도와 기대를 더 정확히 이해할 수 있다. 이전 연구에서는 KH-Coder 도구를 사용하여 단어 간의 공유된 네트워크를 시각화하고 이를 13개의 클러스터로 나누어 분석했다. 다시 이를 평가하여 공출된 네트워크를 최종 5개의 클러스터로 시각화하여 중요한 통찰력을 제공했다[10].

2.3 Categorization

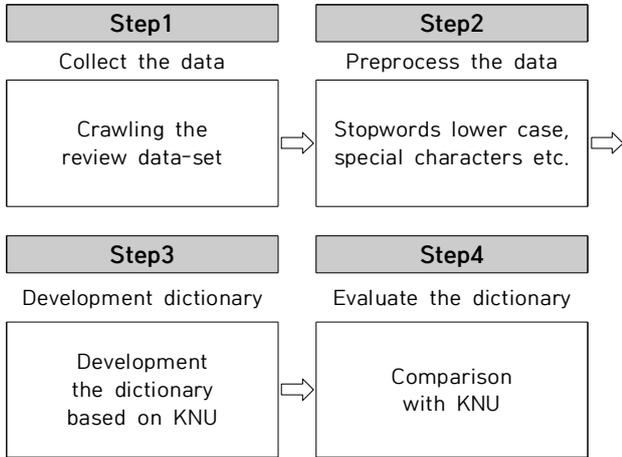
감성분석은 F&B 외식산업에서 고객 리뷰를 통해 소비자 만족도와 경험을 이해하는 데 중요한 역할을 한다. 특히, 레스토랑 산업에 맞춤화된 감성사전을 사용하여 리뷰의 다양한 측면을 더 정확하게 분석하고 분류하는 것이 필수적이다. 이러한 맥락에서 본 연구는 온라인 리뷰에서 얻은 식음료매장 관련 데이터의 감성분석을 미세 조정하기 위해 개발된 맞춤형 감성사전의 적용을 제안한다[11].

본 논문에서 제시된 산업특화 감성사전은 음식 맛, 서비스 품질, 가격, 분위기 등의 주요 평가항목에 따라 리뷰 문장을 분류하며, 해당 산업 특유의 어휘와 표현을 포함하고자 한다. 이러한 접근법은 각 평가항목이 소비자의 구매결정과 레스토랑 성과에 미치는 영향을 더 정확하게 평가할 수 있게 한다. 이에 신규 제안된 분류 방법은 특정 감성 어휘에 따라 문장을 분류하고 리뷰에서 표현된 전체 감성을 결정할 수 있을 것으로 예상된다.

이전 연구 검토는 본 분석에서 직면한 문제를 해결하기 위한 방법론적 및 기술적 통찰력을 제공한다. 이러한 통찰력에 기반하여 본 연구는 리뷰 데이터의 감성 가치를 더 정확하게 분석하여 레스토랑의 비즈니스 성과와 직접 관련된 실질적인 결과를 도출하는 것을 목표로 한다. 이 과정은 식음료 매장의 시장 경쟁력을 향상시키고 매출과 직접적으로 관련된 소상공인의 운영성과를 극대화하는 데 기여할 것으로 기대된다.

III. The Proposed Scheme

Table 3. System Overview



본 연구에서 제안하는 시스템은 리뷰 데이터를 수집하고, 전처리하며 감성사전을 개발 및 평가하는 일련의 과정을 포함한다. <표3>은 시스템의 전체적인 개요를 나타낸다.

먼저, 리뷰 데이터를 수집하는 단계에서는 웹 크롤링 기술을 활용하여 다양한 음식점 리뷰 데이터를 수집한다. 이 과정에서 가능한 많은 리뷰 데이터를 확보하는 것이 중요하다. 수집된 데이터는 전처리 과정을 거치게 된다. 전처리 단계에서는 불용어(stopwords)를 제거하고, 모든 텍스트를 소문자로 변환하며, 특수 문자를 제거하는 작업이 포함된다. 이러한 전처리 작업은 데이터의 일관성을 높이고, 분석의 정확성을 향상시키기 위해 필수적이다.

전처리가 완료된 데이터는 감성사전 개발에 사용된다. 본 연구에서는 KNU에서 제공하는 감성사전을 기반으로 산업특화 감성사전을 개발한다. 여기서 중요한 점은 음식점 리뷰에 특화된 어휘와 표현을 포함하도록 사전을 확장 및 수정하는 것이다.

마지막으로, 개발된 감성사전의 성능을 평가하기 위해 기존의 KNU 감성사전과 비교하는 과정을 거친다. 이를 통해 새로운 감성사전이 얼마나 정확하게 감성을 분석하고 분류할 수 있는지 검증한다. 비교 결과를 통해 산업특화 감성사전의 유효성을 확인하고, 필요에 따라 사전을 수정 및 보완한다.

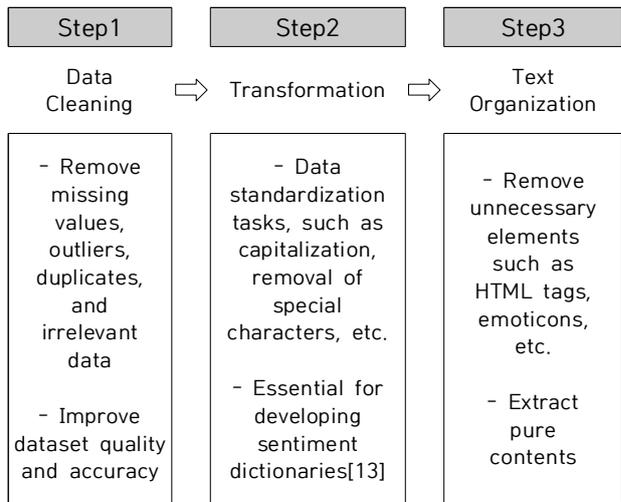
이러한 일련의 단계를 통해 본 연구는 음식점 리뷰 데이터에 특화된 감성분석 시스템을 개발하고, 그 정확성과 효과성을 입증하고자 한다.

1. Collect and Pre-process The Data

F&B 산업 특화 감성사전 제작에 앞서, 리뷰 데이터를 모으기 위해 탐색과정을 거쳤다. 먹거리가 풍부한 경북 K 지역 내 100개 이상 네이버 리뷰를 보유한 식음료 매장 중 무작위로 지역 내 6개의 레스토랑에서 리뷰 데이터를 크롤링하였다. 매장 A, B, C의 리뷰 데이터는 감성사전 개발에 사용되었으며, 나머지 매장 D, E, F의 리뷰 데이터는 감성사전 모델의 최종 선택 및 평가에 사용되었다[12].

특히, F&B 산업에 특화된 감성사전을 구축하기 위해 다양한 한국어 F&B 리뷰 데이터를 수집하였다. 주요 데이터 출처는 네이버 리뷰를 중심으로, 카카오톡 지도 리뷰, 구글 지도 리뷰, 캐치테이블 리뷰 등 5곳 채널이 포함되었다. 수집된 데이터는 리뷰 텍스트, 작성자 정보, 작성 날짜 등의 23종 메타데이터로 구성되었다. 수집된 데이터는 형식과 품질이 다양하므로 일관된 분석을 위해 철저한 전처리 과정을 거쳤다. 전처리 단계는 다음의 <표4>와 같다.

Table 4. Step of Pre-processing



2. Developing Sentiment Lexicon(Dictionary)

본 연구에서는 KNU 감성사전을 탐색하고, 리뷰 데이터를 분석하여 F&B 산업에 특화된 새로운 감성사전을 개발하였다. 먼저, 기존의 KNU 감성사전을 검토하여 그 구조와 내용을 분석하고, 사전의 범위와 한계를 이해하였다. 이를 통해 현재 감성사전이 포함하고 있는 단어들의 감성적 의미와 사용빈도를 파악하였다.

이후, 선택된 음식점의 리뷰 데이터를 분석하여 감성사전에 포함될 적합한 단어들을 식별하였다. 이 과정에서는 자주 등장하는 단어와 감성적으로 중요한 단어들에 중점을 두었다. 특히, F&B 산업 특유의 어휘와 표현을 파악하기 위해 빈도 분석 및 감성적 관련성을 집중적으로 검토하였다[14].

마지막으로, KNU 감성사전의 단어들과 리뷰 데이터에서 맞춤 선택된 단어들을 결합하여 새로운 감성사전을 개발하였다. 이 사전은 F&B 산업에 특화된 것으로, 해당 산업에서 자주 사용되는 어휘와 표현들을 포함하도록 구성되었다[14, 15, 16]. 이러한 과정을 통해 개발된 감성사전은 기존 사전의 한계를 보완하고, F&B 산업의 맥락에서 더욱 정확하게 감정을 분석할 수 있도록 설계되었다.

특히 N-gram과 휴먼코더 기법으로 F&B 산업 특화 감성사전을 개발하였다.

2.1 N-gram

본 연구에서는 F&B 산업에 특화된 감성사전을 개발하기 위해 N-gram(엔-그램) 기법을 활용하였다. N-gram 분석은 자연어 처리(NLP) 분야에서 널리 활용되는 기법으로, 특히 감성분석 및 텍스트 마이닝에서 중요한 역할을 한다. 이 기법은 단어들 간의 관계와 패턴을 식별할 수 있게 하여 더욱 정확한 감성분석을 가능하게 하고, 이러한 N-gram의 분석 효과는 수많은 연구에서 입증되었다[17, 18, 19].

N-gram은 N개의 연속된 단어들의 시퀀스를 의미하며, 텍스트 내 단어 빈도와 연관성을 분석하는 데 유용하다. 예를 들어, "The food delivery is really delicious"라는 문장에서 2-그램(bigram)은 ["The food", "food delivery", "delivery is", "is really", "really delicious"]와 같이 구성되며, 3-그램(trigram)은 ["The food delivery", "food delivery is", "delivery is really", "is really delicious"]와 같이 구성된다.

데이터 수집 및 전처리 과정을 거친 후, 약 10,000개의 리뷰 데이터를 N-gram 분석을 통해 <표 5>와 같은 단계를 수행하였다. 그러나, 4gram 또는 그 이상의 N-gram 모델을 사용하지 않은 이유는 다음과 같은 몇 가지 이유에 기인한다.

- 데이터 스파스 문제: N-gram 모델에서 N이 증가할수록, 특히 4gram, 5gram과 같이 더 높은 N값을 사용할 경우, 데이터의 희소성이 크게 증가하는 문제가 발생한다. 그 결과 학습 데이터에서 나타나는 표현의 수가 급격히 줄어들게 된다. 따라서, 4gram 이상의 모델에서는 학습에 충분한 데이터가 확보되지 않아 오히려 성능이 저하될 가능성이 높다.
- 계산 비용 증가: N-gram의 수가 늘어날수록, 계산 복잡성과 저장 비용이 기하급수적으로 증가하게 된다. 4gram, 5gram 모델을 사용하면, 데이터의 크기가 커지고, 이를 처리하기 위한 메모리와 연산 자원이 크게 증가할 수 있다. 이는 특히 소상공인과 같이 제한된 자원을 가진 사용자에게는 실용적이지 않을 수 있다.

따라서 본 연구에서는 1-gram, 2-gram, 및 3-gram 모델을 결합하여 F&B 산업에 최적화된 감성 사전을 구축하고자 하였다. 데이터 수집 및 전처리 과정을 거친 후, 약 10,000 개의 리뷰 데이터를 N-gram 분석을 통해 <표5>와 같은 단계를 수행하였다.

Table 5. N-gram Analysis (Step1)

1-gram Analysis	- Analyze each word individually to calculate frequency. - Extract the top 1,000 most frequent words.
2-gram Analysis	- Analyze two consecutive pairs of words to calculate frequency. - Extract the top 1,000 most frequent words.
3-gram Analysis	- Analyze three consecutive sets of words to calculate frequency. - Extract the top 1,000 most frequent words.

추출된 상위 1,000개의 N-그램은 휴먼 코더(human coder)를 통해 감성 점수를 평가하였다. 평가된 감성 점수는 KNU 감성사전에 추가되었으며, 이를 통해 감성분석을 수행하였다. 높은 정확성을 보장하기 위해 이 과정을 각 N-그램 카테고리에서 세 번 반복하였다.

Table 6. N-gram Analysis (Step2)

KNU, 1,2-gram	6,000 words added to the KNU dictionary with 1-gram and 2-gram sentiment scores.
KNU, 1,3 gram	6,000 words added to the KNU dictionary with 1-gram and 3-gram sentiment scores.
KNU, 2,3 gram	6,000 words added to the KNU dictionary with 2-gram and 3-gram sentiment scores.
KNU, 1,2,3 gram	6,000 words added to the KNU dictionary with 1-gram, 2-gram and 3-gram sentiment scores.

단일 N-그램 감성사전을 결합하여 보다 복잡한 감성사전을 구성하였다. 이는 보다 다양한 맥락에서 감성분석을 가능하게 한다. 예를 들어, KNU + 1-그램 + 2-그램 감성사전은 1-그램 및 2-그램 감성 점수를 포함하는 사전으로, KNU 사전에 6,000개의 단어를 추가한 것이다. KNU + 1-그램 + 3-그램 감성사전은 1-그램 및 3-그램 감성 점수를 포함하는 사전으로, KNU 사전에 6,000개의 단어를 추가한 것이다. KNU + 2-그램 + 3-그램 감성사전은 2-그램 및 3-그램 감성 점수를 포함하는 사전으로, KNU 사전에 6,000개의 단어를 추가한 것이다. KNU + 1-그램 + 2-그램 + 3-그램 감성사전은 1-그램, 2-그램, 3-그램 감성 점수를 모두 포함하는 사전으로, KNU 사전에 9,000개의 단어를 추가한 것이다.

이러한 과정을 통해 개발된 감성사전은 F&B 산업의 맥락에서 더욱 정확하게 감성을 분석할 수 있도록 설계되었으며, n-그램 분석 기법을 통해 기존의 감성사전을 보완하고, 비정형화된 리뷰 데이터의 특성을 효과적으로 반영하였다.

결과적으로 [KNU], [1gram], [2gram], [3gram], [1gram + 2gram], [2gram + 3gram], [1gram + 3gram], [1gram + 2gram + 3gram] 조합을 만들어, 총 7개의 사전을 구축하였다.

2.2 Human Coder

상위 1,000개의 추출된 N-gram(1-gram, 2-gram, 3-gram)에 대해 사람 평가(휴먼 코딩)를 통해 감성 점수를 매기는 과정을 수행하였다. 각 N-gram은 세 명의 휴먼코더에 의해 평가되어 감성 점수의 신뢰성을 보장하였다. 또한 휴먼코더 간의 일관성을 높이기 위해 코딩 가이드를 구축하여 사전 교육을 제공하였다.

휴먼코더를 평가할 때는 코헨의 카파계수(Cohen's Kappa Coefficient) 개념을 이용하였다. 이는 복수의 코더 간의 일치도를 측정하는 방법이며, 유사성이 0.7 이상인 경우 해당 평가가 신뢰할 만한 것으로 간주된다[20].

3명의 휴먼코더가 각각의 점수를 매긴 후, 휴먼코더 1과 휴먼코더2, 휴먼코더 2와 휴먼코더3, 휴먼코더 1과 휴먼코더3이 매긴 점수를 비교하였다. 같은 점수를 매겼을 때는 1, 다른 점수를 매겼을 때는 0으로 두고 계산한 결과, 평균 점수가 0.7을 상회하여 논리적으로 믿을 만하다고 판단하였다.

Reviews	Human Coder1	Human Coder2	Correctness
솔직히 여기는 맛이 없을 수가 없어요	1	2	0
웨이팅이 없어서 좋았고, 맛도 최고입니	2	2	1
고막비빔밥이 죽여줍니다.	2	2	1
이 가격에 이 퀄리티가 말이 안됩니다.	2	2	1
여길 왜 이제 알았을까요, 전 바보입니	1	1	1
둘이 먹다 하나 죽어도 몰라요.	1	1	1
폼 미쳤어요.	2	1	0
경주에 오게 된다면 다른 곳을 갈 것 같	-1	-2	0
놀랄 정도로 별로입니다.	-1	-1	1
서비스가 가장 최악입니다.	-2	-2	1
			0.7

Fig. 1. Evaluation of Human Coder (Code Book)

Table 7. Evaluation of Dictionary

	KNU	1gram	2gram	3gram	1-2gram	2-3gram	1-3gram	1-2-3gram	Performance Improvement
Accuracy	22.88%	57.51%	25.44%	22.88%	60.61%	62.28%	61.51%	63.52%	△178% p
Precision	24.64%	63.88%	24.41%	24.64%	64.54%	61.09%	63.88%	67.76%	△175% p
Recall	22.59%	61.41%	22.45%	22.59%	62.14%	60.41%	61.41%	62.41%	△176% p
F1 Score	19.98%	58.52%	20.07%	19.98%	60.42%	60.89%	61.52%	65.31%	△227% p

3. Experimental Results

3.1 Selection of the Optimal Sentiment Dictionary[21]

본 연구에서는 N-gram 기법을 사용하여 생성된 7개의 감성사전 모델(1-gram, 2-gram, 3-gram, 1-gram+2-gram, 2-gram + 3-gram, 1-gram + 3-gram, 1-gram + 2-gram + 3-gram)과 기존 KNU 모델을 비교 분석하고자 했다. 이를 위하여 크롤링한 음식점 D, E, F의 리뷰 데이터를 KNU 감성사전과 N-gram 기법을 사용하여 새로 구축한 감성사전 7개를 토대로 감성점수를 계산하였다. 각 레스토랑의 리뷰 중 무작위로 500개씩 샘플링하여 총 1,500개의 리뷰를 확보하고, 이를 휴먼코더들이 평가하여 감성점수를 매겼다. 이 점수를 기준으로 총 8개의 감성사전에 대한 성능 평가를 진행하였다.

<표7>에서 제시한 바처럼 각각의 감성사전의 Accuracy(정확도), Precision(정밀도), Recall(재현율), F1 Score를 계산해 본 결과 1-gram+2gram+3gram Dictionary가 모든 지표에서 가장 성능이 좋았다. 특히 KNU 감성사전의 성능과 비교해봤을 때, Accuracy는 178%p, Precision은 175%p, Recall은 176%p, F1 Score는 227%p 더 높은 것을 확인할 수 있다. 따라서 1gram + 2gram + 3gram 으로 구성된 감성사전을 최종 F&B 산업 특화 감성사전으로 선택하였다.

3.2 Comparison of the Top50 Words in Restaurants Reviews

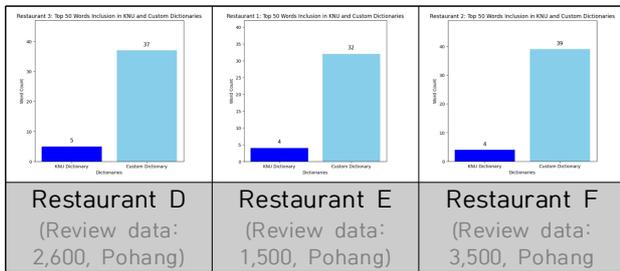
제작한 감성사전의 포괄성을 평가하기 위해 레스토랑 D, E, F의 리뷰 데이터에서 각각 빈도수가 높은 단어 50개를 리스트업했다. 그 후 이 상위 50개의 단어들이 과연 KNU 감성사전과 우리가 제작한 F&B 산업 특화 감성사전에 각각 몇 개씩 포함되어 있는지 확인하였다.

그 결과 <표8>의 결과와 같이, 레스토랑 D의 리뷰 데이터에는 KNU 사전에 포함되어 있는 단어가 단 4개, F&B 산업 특화 감성사전에 포함되어 있는 단어가 32개였다. 레스토랑 E의 리뷰 데이터에는 KNU 사전에 포함되어 있는

단어도 단 4개, F&B 산업 특화 감성사전에 포함되어 있는 단어가 39개였다. 마지막으로 레스토랑 F의 리뷰 데이터에는 KNU 사전에 포함되어 있는 단어가 단 5개, F&B 산업 특화 감성사전에 포함되어 있는 단어가 37개였다.

평균적으로 KNU에는 4~5개, F&B 산업 특화 감성사전에에는 평균 36개 정도 포함되어 있어 9배 높았다. 연구진이 제작한 사전이 다른 매장과 업장들에게도 좋은 결과를 보여 확장성이 있음을 파악할 수 있었다.

Table 8. Comparison of the Top50 Words in Restaurant Reviews



3.3 Statistical Testing

마지막으로 우리가 제작한 F&B 산업 특화 감성사전의 통계적 검정도 진행하였다[22]. <그림2>에 보이는 것처럼, 레스토랑 A의 리뷰 데이터에서 랜덤 샘플링한 리뷰 500개를 기준으로 KNU사전과 우리가 최종 산업특화 감성사전으로 점수를 매겨본 후 통계적 검정을 진행하였다.

먼저 정규성 테스트(Shapiro-Wilk)를 진행하였다. Shapiro-Wilk 테스트는 데이터 샘플이 정규분포를 따르는지 여부를 결정하는 통계적 방법이다. 이 테스트는 특히 작은 샘플 크기에 적합하다. p-value가 0.05보다 작으면 데이터가 정규분포를 따르지 않는 것으로 간주된다.

SPSS 26.0 버전으로 정규성 검정을 진행한 결과, <표 8>과 같이 KNU, Custom 사전 모두 p-value가 0.001을 넘지 않아 두 그룹 모두 정규성을 만족하지 않는다고 판단 가능했다.

No	Review Data	KNU	1gram-2gram-3gram
1	11시오븐이라 10시 40분에 도착해 캐지타이를 찍어팅달아왔는데 15분짜리 5분쯤 입장했어요 엄청너무도 끔찍양념이 자극적이기 없어서 어쩔건구들이 먹기 힘들게걸었어요 고요이 진짜진 짜맛이요 육회롤회도 시원하게먹기싫었어요 직원분들 작장 배쓰고 친절해요	0	2
2	경주 가죽여탕 중 우연히 들리던 불루리본 맛집 꼬막비빔밥 해물파전 소를고기 모두 정말 맛있었어요 다음번 방문 때는 육회롤회와 비빔밥도 먹어보고 싶어요 분위기와 맛 다잡은 경주 황리단길 맛집이네요추천드려요	2	3
3	육회비빔밥 먹을곳 찾다가 다름식당 리뷰에 여기 추천하는말들이 있길래 와봤습니다 꼬막비빔밥이 주력인것 같아서 꼬막비빔밥이랑 경주왔으니 육회롤회도 시켰어요 꼬막비빔밥 진짜 맛있었어요 꼬막 양도 엄청 많고 육회롤회는 알몸이 녹을수록 더 맛있어지더라구요	3	4
4	물요일 점심이라 510분내로 바로 입장했고 여기오길 걸렸다는 생각했습니다.맛있게 잘 먹고 왔어요	3	5

Fig. 2. Statistical Testing Results

Table 9. Result of Normality Test

	W	p-value
KNU	= 0.703	= 1.40e-45 < 0.001
Custom	= 0.781	= 5.47e-41 < 0.001

이어서 분산 동질성 테스트(Levene's)도 진행하였다. Levene 테스트는 두 그룹의 분산이 동일하지 여부를 결정하는 통계적 방법이다. p-value가 0.05보다 작으면 두 그룹 간의 분산이 동일하지 않음을 나타낸다.

분산 동질성 테스트를 진행한 결과, <표10>과 같이 p값이 0.05보다 작았다. 즉, 두 그룹의 분산이 동일하지 않음을 나타내어, 그룹 간 분산 차이가 유의미했다. 테스트의 결과를 종합하면 두 그룹 모두 정규성을 만족하지 않고 분산이 동일하지 않으므로 독립 t검정을 사용할 수 없다.

Table 10. Result of Statistical Analysis

Results	Homogeneity of Variance Test	Non-parametric Statistics Test
Statistic	= 348.082	= 786383.523
p-value	= 1.28e-73 < 0.01	= 6.88e-60 < 0.001

따라서 마지막으로 비모수 통계 분석(Mann-Whitney U 테스트)을 진행하였다[23]. Mann-Whitney U 테스트는 종속 변수가 서열형 또는 연속형이지만 정규분포를 따르지 않는 경우 두 독립 그룹 간의 차이를 비교하는 비모수 테스트이다. p-value가 0.05보다 작으면 두 그룹 간에 유의미한 차이가 있음을 결론지을 수 있다.

<표10> 우측 내용과 같이 비모수 통계분석 결과 p값이 0.001보다 작아, 그룹의 중앙값이 동일하다는 귀무가설을 기각할 수 있게 한다. 이는 KNU_score와 Customized_score 간에 통계적으로 유의미한 차이가 있음을 나타낸다.

결과적으로, 본 연구에서 수행된 통계 분석은 기존 KNU 감성사전과 연구진이 제안한 F&B 산업특화 감성사전 간의 유의미한 차이를 나타낸다. Shapiro-Wilk 및 Levene 테스트는 두 그룹 모두 정규성을 만족하지 않고 분산이 동일하지 않음을 확인하였다. Mann-Whitney U 테스트는 그룹 간 중앙값의 유의미한 차이를 나타내었다. 이러한 결과는 산업특화 F&B 산업 감성사전이 해당 산업의 연구에 더 적합하다는 결론을 논리적으로 뒷받침한다.

IV. Conclusions

산업특화 감성사전이 기존 KNU 감성사전에 비해 리뷰 데이터를 분류하는 데 뛰어난 성능을 입증하였지만, 연구를 더욱 발전시키고 산업에 실질적인 기여를 하기 위해 해결해야 할 한계점들이 존재한다. 먼저 필연적으로 언어모델은 학습 데이터의 편향을 반영할 수 있다. 이는 특정 그룹이나 주제에 대한 왜곡된 표현을 초래할 수 있고, 모델의 성능 자체가 주어진 데이터에 크게 의존된다. 이에 모델의 출력에 대한 편향성을 분석할 수 있는 도구를 활용하여 지속적으로 모니터링하는 점이 필요할 것이다.

나아가 디지털 기술과 AI의 발전으로 시대의 흐름이 빠르게 변화하여 그에 따라 외래어, 신조어 등도 더 빠르게 변동되고 있다. 이는 모델의 유효성 검증 및 갱신에 대한 주기와 범위가 넓고 빠르다는 점을 의미한다. 후속연구로서 분석체계를 지속적으로 발전시켜가도록 제언한다.

다음으로, 감성사전의 적용을 한국의 식음료는 물론 다양한 산업으로 확장하고 검토해야 한다. 다양한 음식 유형 및 서비스 산업에서 산업특화 감성사전의 효과를 평가함으로써 그 적용 가능성을 확인할 수 있다. 셋째, 대규모 데이터 분석을 통한 정량적 검증을 강화해야 한다. 정교화된 통계모델을 적용하고 비교분석을 수행함으로써 감성사전의 신뢰성과 타당성을 더욱 높일 수 있다.

이러한 과제를 해결함으로써 본 연구는 식품구매, F&B 산업뿐만 아니라 다양한 서비스 산업에서 소비자만족과 고객경험을 정확하게 분석하고 개선하는 데 강력한 도구가 될 수 있다. 궁극적으로 이는 산업 전반의 서비스 품질과 고객만족도를 향상하는데 기여할 것이다.

본 연구는 레스토랑 리뷰 데이터를 사용한 산업특화 감성사전을 통한 감성분석의 잠재력을 입증하였다. 보다 정확하고 신뢰할 수 있는 고객 피드백 분석을 통해 비즈니스 성과를 극대화할 수 있음을 보여주었다. 향후 연구는 이러한 과제를 해결하고 더 발전된 감성분석 도구를 개발하며 실질적인 산업 환경에 적용하는 방법을 탐구할 것이다.

REFERENCES

- [1] H.-J. Eom, and H.-J. Jin. (2024). "Analysis of the Effect of Online Reviews on Restaurant Rankings," *Korean Journal of Hospitality & Tourism*, 33(1), 57-77. <https://doi.org/10.24992/KJHT.2024.2.33.01.57>
- [2] J.Y. Choi, H.A. Kim, and Y.B. Kim. (2020). "The Impact of Online Review Volume, Rating, and Sentiment Score on Sales: Focusing on the Moderating Effect of Brand Reputation," *Journal of Channel and Retailing*, 25(3), 1-21. <https://doi.org/10.17657/jcr.2020.07.31.1>
- [3] J.-P. Park, and H. Chun. (2022). "Effects of Electronic Word-of-Mouth Quality on Information Satisfaction and Purchase Intention among Consumer Reviews on Food Delivery Apps," *Journal of Tourism and Leisure Research*, 34(5), 347-364. <https://doi.org/10.31336/jtlr.2022.5.34.5.347>
- [4] P. Mehta, and S. Pandya. (2020). "A review on sentiment analysis methodologies, practices and applications," *International Journal of Scientific and Technology Research*, 9(2), 601-609.
- [5] H.-M. Kim, and K. Park. (2019). "Sentiment analysis of online food product review using ensemble technique," *Journal of Digital Convergence*, 17(4), 115-122. <https://doi.org/10.14400/JDC.2019.17.4.115>
- [6] J. Kim, W. Hou, and H. Kahn. (2020). "The Power of Online Review: Consumer Evaluation Based on Online Review Types," *Journal of Product Research*, 38(4), 21-30. <https://doi.org/10.36345/kacst.2020.38.4.004>
- [7] J.-S. So, and P.-S. Shin. (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. <https://doi.org/10.9708/jksoci.2020.25.06.081>
- [8] S.-H. Park, Y.-E. Lee, and H.-J. Lee, "Research on Enhancing Customer Experience through AI-Supported Review Generation," *The transactions of The Korean Institute of Electrical Engineers*, 73(2), 334-342. <https://doi.org/10.5370/kiee.2024.73.2.334>
- [9] H. H. Gil. (2018). "A study on Korean unused word list for text mining," *The Korean Language and Literature*, 78, 1-25. <https://doi.org/10.18628/urimal.78.201809.1>
- [10] M. K. Cho, and B. J. Lee. (2021). "Comparison of service quality of full service carriers in Korea using topic modeling: based on reviews from TripAdvisor," *Journal of Hospitality and Tourism Studies*, 23(1), 152-165, <https://doi.org/10.31667/jhts.2021.2.86.152>
- [11] B.L. Chua, S. Karim, S. Lee, and H. Han. (2020). "Customer Restaurant Choice: An Empirical Analysis of Restaurant Types and Eating-out Occasions," *Int'l Journal of Environmental Research and Public Health*, 17(17), 6276. <https://doi.org/10.3390/ijerph17176276>
- [12] S. Li, S. Yun, and Y.W. Woo. (2019). "Crawling Methods for Web Data of Various Formats using Python," *Proceedings of General Conference of the Korean Institute of Information and Communication Sciences*, 343-346.
- [13] J.-H. Jung, H.-I. Chung, and Z.-K. Lee. (2021). "An Analysis of Mobile Food Delivery App Baemin by Using Text Mining and ARIMA Model," *Journal of Digital Contents Society*, 22(2), 291-299, <https://doi.org/10.9728/dcs.2021.22.2.291>
- [14] S.-W. Kang, and W.-G. Seo. (2023, October 26). "Generate a

Domain-specific Sentiment Lexicon for Stock Price Prediction," Proceedings of the General Conference of the Korean Institute of Information and Communication Sciences, Gyeonggi.

- [15] J. Kim, D. Lee, E. Lee, Y. Park, Y. Shin, and O. Kwon. (2023, May 31). "Development of a Sophisticated Sentiment Lexicon for an Automated Food Safety Inspection System," Proceedings of the Spring Joint Conference of the Korean Institute of Industrial Engineers, Jeju.
- [16] H. Kang, and H. Song. (2021). "Constructing Sentiment Lexicon for Subject-Specific Sentiment Analysis," *Korean Linguistics*, 93, 83-110, <https://doi.org/10.20405/kl.2021.11.93.83>
- [17] Z. Mengjuan, H. Changping, and S. Yu. (2022). "Sentiment Lexicon Construction for Chinese Book Reviews Based on Ultrashort Reviews," *The Electronic Library*, 40(3), 221-236. <https://doi.org/10.1108/EL-07-2021-0147>
- [18] A. Dey, M. Jenamani, and J.J. Thakkar. (2018). "Senti-N-Gram: An n-Gram Lexicon for Sentiment Analysis," *Expert Systems with Applications*, 103, 92-105. <https://doi.org/10.1016/j.eswa.2018.03.004>.
- [19] F. Yin, Y. Wang, J. Liu, and L. Lin. (2020). "The Construction of Sentiment Lexicon Based on Context-Dependent Part-of-Speech Chunks for Semantic Disambiguation," *IEEE Access*, 8, 63359-63367.
- [20] A. B. Cantor. (1996). "Sample-size Calculations for Cohen's Kappa," *Psychological Methods*, 1(2), 150-153. <https://doi.org/10.1037/1082-989X.1.2.15>
- [21] M. Lerasle. (2012). "Optimal Model Selection in Density Estimation," *Annales de l'I.H.P. Probabilités et Statistiques*, 48(3), 884-908. <https://doi.org/10.1214/11-AIHP425>.
- [22] E. Melilli, and P. Veronese. (2024). "Confidence Distributions and Hypothesis Testing," *Stat. Papers*, 65, 3789-3820. <https://doi.org/10.1007/s00362-024-01542-4>
- [23] S-W. Kim, and J-M. Song. (2021). "Statistical Hypothesis Testing using Deep Learning: Focusing on Two Sample T-Test," *Journal of the Korean Data And Information Science Society*, 32(1), 25-35. <https://doi.org/10.7465/jkdi.2021.32.1.25>

Authors



Yeryung Moon is currently double majoring in ICT Convergence and Business Management at Handong Global University, Korea. She is a research assistant (RA, research funded by IITP) at the School of

ICT & Global Entrepreneurship, HGU since 2023. Her main interests are the Google for Startups, Digital tech-based proptech, foundation for a youth startup ecosystem.



Gaeun Son is currently double majoring in ICT Convergence and Business Management at Handong Global University, Korea. As a data-driven marketing strategist and the head of CX-Lab, she focuses on achieving results

by integrating various academic fields. Her main interests are the AD Tech, Growth Marketing and Consulting as a manager of the Performance by TBWA Corp.



Geonuk Nam is currently double majoring in ICT Convergence and Business Management at Handong Global University, Korea. He have been interned at MYSC, an accelerating and impact investment company, in 2024,

focusing on sustainable innovation. He is interested in Social Impact, Financial Inclusion, ESG Evaluation & Ratings, and Data Analysis/Visualization using Python.



Hanjin Lee received the B.S. in Sociology, M.S. degree in Communication from Yonsei University, Korea in 2006 and 2011. Furthermore he received the Ph.D. degrees in Digital Management, Korea University, 2021.

Dr. Lee's research interests are focused on AI biz innovation, as well as the intersection between user behavior and digitalization. He has gained 16 years experience in the e-Commerce industry (Naver, eBay, and Coupang Corp.)