

텍스트 마이닝을 활용한 스마트 스피커 제품의 포지셔닝: 인공지능 속성을 중심으로¹

Positioning of Smart Speakers by Applying Text Mining to Consumer Reviews: Focusing on Artificial Intelligence Factors

이 정 현(Jung Hyeon Lee) 가톨릭대학교²

선 형 주 (Hyung Joo Seon) 연세대학교 경영대학³

이 흥 주 (Hong Joo Lee) 가톨릭대학교⁴

ABSTRACT

The smart speaker includes an AI assistant function in the existing portable speaker, which enables a person to give various commands using a voice and provides various offline services associated with control of a connected device. The speed of domestic distribution is also increasing, and the functions and linked services available through smart speakers are expanding to shopping and food orders. Through text mining-based customer review analysis, there have been many proposals for identifying the impact on customer attitudes, sentiment analysis, and product evaluation of product functions and attributes. Emotional investigation has been performed by extracting words corresponding to characteristics or features from product reviews and analyzing the impact on assessment. After obtaining the topic from the review, the effect on the evaluation was analyzed. And the market competition of similar products was visualized. Also, a study was conducted to analyze the reviews of smart speaker users through text mining and to identify the main attributes, emotional sensitivity analysis, and the effects of artificial intelligence attributes on product satisfaction. The purpose of this study is to collect blog posts about the user's experiences of smart speakers released in Korea and to analyze the attitudes of customers according to their attributes. Through this, customers' attitudes can be identified and visualized by each smart speaker product, and the positioning map of the product was derived based on customer recognition of smart speaker products by collecting the information identified by each property.

Keywords: Customer review, AI speaker, Artificial intelligence, Positioning map

1) 본 연구는 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구이며(NRF-2017S1A3A2066740), 2019년도 가톨릭대학교 교비연구비의 지원으로 이루어졌음.

논문접수일: 2020년 1월 20일; 1차 수정일: 2020년 3월 1일 게재확정일: 2020년 3월 8일

2) 제 1저자(95743@naver.com)

3) 제 2저자(linear_state@yonsei.ac.kr)

4) 교신저자(hongjoo@catholic.ac.kr)

1. 서론

스마트 스피커는 많은 제조기업에서 홈 엔터테인먼트의 허브로서 출시하고 있고, 다양한 형태로 수요가 늘고 있다. 스마트 스피커는 인공지능 비서 기능이 포함되어 사람이 목소리로 명령을 내릴 수 있으며, 기기의 제어와 연계된 오프라인 서비스를 제공한다. 스마트 스피커에 인공지능 비서 역할을 하는 애플 시리, 아마존 알렉사, 삼성 빅스비 등은 스마트 스피커 뿐만 아니라 다양한 기기에 탑재되어 인공지능 기능을 제공하고 있다. 스마트 스피커는 기본적으로 Wi-Fi나 Bluetooth 기반의 인터넷 연결성을 가진 포터블 스피커에 인공지능 속성이 포함되어 있는 기기이다. 음성 인식, 정보제공, 연결된 앱의 조절 및 제어, 연결된 기기의 조절 및 제어가 대표적인 인공지능 속성이며, 음성 인식을 통한 정보제공에서 출발하여 다른 앱 및 기기와의 연동성을 점차 증대 시키고 있다 (이홍주, 2019).

스마트 스피커가 대중적으로 활용되면서 사용되는 분야와 만족, 불만족 요인에 대한 연구들이 이루어 졌다 (이진명 외 2019; 이홍주, 2018a; 이홍주, 2019; 킨슈머 인사이트, 2018). 설문을 통한 만족 요인과 불만족 요인 파악 뿐만 아니라 고객 리뷰 분석을 통한 인공지능 속성 요인에 대한 고객 태도 또한 분석되었다. 설문을 통한 스마트 스피커의 수용 요인에 대한 연구는 특정 제품보다는 스마트 스피커 전체에 대한 수용 요인을 분석하였으며, 고객 리뷰를 분석한 연구들도 스마트 스피커 제품들을 비교하는 연구보다는 특정 스마트 스피커 제품에 대한 고객 태도를 분석(이홍주, 2019)하거나 스마트 스피커와 비 스마트 스피커를 비교하였다(이홍주, 2018a). 스마트 스피커 제품이 다양하게 출시되고 사용되는 상황에서 사용자들의 고객 리뷰를 분석하여 제품들에 대한 고객의 태도를 분석하여 제품 간에 비교할 필요가 증대하고 있다. 고객 리뷰를 활용하여 제품간 비교를 시각화하는 방안들은 주로 제품명의 동시

발생 정보나 제품명 주변의 단어나 제품명이 포함된 문장의 감성 값을 중심으로 제품간 비교를 시각화하고 있다(Chen et al., 2015; Lee and Bradlow, 2011; Netzer et al, 2012). 제품명이 아니라 제품이 가지고 있는 특정 속성을 중심으로 제품별로 속성에 대한 사용자의 의견을 파악하여 제품의 특징을 파악하고, 속성에 대한 고객 경험을 기반으로 제품에 대한 인식을 시각화하는 방안이 필요하다.

따라서, 본 연구는 고객 리뷰를 활용하여 스마트 스피커 제품들의 포지셔닝을 파악하는 방안을 제안한다. 스마트 스피커가 가진 다양한 속성 중에서 인공지능 속성에 집중하여 분석하였으며, 국내에 출시된 스마트 스피커에 대한 고객 경험 텍스트를 수집하여 분석하였다. 텍스트에 출현한 단어를 인공지능 속성 카테고리에 맵핑하고, 문장을 감성분석하여 속성에 대한 태도를 파악하였다. 단어를 확장하여 인공지능 속성에 할당하는 방안을 고안하였으며, 제품별 속성에 대한 태도를 측정하고 이를 기반으로 제품별 포지셔닝 맵을 도출하였다.

2장에 관련연구를 정리하였으며, 3장에서 본 논문이 활용한 자료에 대해 기술하였다. 4장에서 분석방안과 스마트 스피커의 포지셔닝 결과를 제시하며, 5장에서 본 연구의 결론과 시사점을 논의하였다.

2. 관련연구

Berger et al. (2020)은 텍스트를 활용한 연구를 텍스트 생산자와 수용자인 고객, 기업, 투자자, 사회로 구분하고, 텍스트 생산자와 수용자별로 수행된 연구들을 구분하였다. 텍스트 분석을 수행하기 위한 데이터 전처리, 분석 도구, 측정 방안, 타당성 검증 방안을 정리하였다. 고객 리뷰 분석을 통해 제품이나 속성에 대한 고객의 태도 분석 연구는 크게 감성 점수를 파악하거나 제품간의 연관관계를 분석하는 연구로 나누어 볼 수 있다.

첫번째 연구들은 고객 리뷰 분석을 통해 제품의 기능이나 속성에 대한 고객 태도, 제품에 대한 감성 분석 및 감성이 제품 평가에 미치는 영향을 파악하여 왔다 (김정훈 등 2015; 박명석 등 2018; Cao et al. 2011; Humphreys and Wang, 2017). 텍스트 마이닝을 적용하여 고객 리뷰에서 속성이나 기능에 해당하는 단어를 추출하여 감성분석을 수행하거나 평가에 미치는 영향을 분석하여 왔다 (이홍주, 2018b; Cao et al. 2011; Ghose et al. 2012). 고객 리뷰에서 토픽을 추출한 후 평가에 미치는 영향을 분석하거나 텍스트 데이터에서 정보를 추출하고, 추출된 정보들을 텍스트 데이터의 다른 속성과 연관관계를 측정하였다 (Büschken and Allenby, 2016). 스마트 스피커 사용자들의 고객 리뷰를 분석하는 연구도 수행되었으며, 주요하게 언급되는 속성 파악, 속성에 대한 감성분석 및 인공지능 속성이 제품 만족에 미치는 영향 분석(이홍주 2018a)과 제품 출시 이후 시간이 흐르면서 스마트 속성에 대한 태도 변화를 분석하였다 (이홍주 2019). 위 연구들은 대체로 한 제품이나 제품의 속성에 대한 고객의 태도를 분석하거나 동태적인 변화를 추적하는 방안을 제안하였다.

두번째 연구들은 주로 제품 커뮤니티나 다양한 제품이 판매되는 쇼핑몰에서 리뷰를 수집하여 다양한 제품을 비교하는 방안을 제안하였다(Chen et al. 2015; Lee and Bradlow, 2011; Netzer et al. 2012). Netzer et al. (2012)는 자동차 포럼의 게시물을 수집하여 자동차 브랜드, 제품모델, 수식어를 추출하여, 자동차 모델 간의 동시발생 네트워크를 그렸다. 동시발생 네트워크에 다차원 척도(Multidimensional Scaling, MDS)를 적용하여 제품 모델의 시장 구조를 파악하였다. 게시물을 통해 파악한 시장 구조를 고객 설문을 통해 관심제품 집합 데이터와 차량 교체 데이터를 통해 파악한 시장 구조와 비교하였다. 또한, 제품 모델과 가장 많이 언급되는 단어들을 시각화하여 제품에 대한 사용자들의 의견을 파악하는 방안도 제시하였다. Lee and

Bradlow (2011)은 디지털 카메라에 대한 고객 리뷰를 수집한 후에, 텍스트에 K-평균 클러스터링을 적용하여 공통 속성에 대한 기술들을 하나 혹은 여러 개의 클러스터로 구성하였다. 제품 속성 계층을 구성하여 속성에 대한 연관 단어들을 파악하였으며, 이를 활용하여 제품간의 관계를 표현하는 시각화 방안을 제안하였다. 고객 리뷰를 활용하여 제품의 랭킹을 측정하는 방안들도 연구되었다 (Chen et al. 2015; Ghose et al. 2012). Chen et al. (2015)는 스마트폰과 피쳐 폰 고객 리뷰에 토픽 모델링을 적용하여 중요 토픽과 토픽의 분포를 파악하고, TOPSIS 방안을 활용하여 제품의 랭킹을 매겼다. 토픽 분포와 제품 랭킹을 고려하여 MDS를 이용한 시장 구조 맵을 제시하였다. Ghose et al. (2012)는 브랜드 선택과 제품 속성 수요 모형에 기반하여 제품의 랭킹을 매겼으며, 입력변수로 고객 리뷰에서 얻을 수 있는 리뷰 수, 평점과 리뷰의 독해 용이성에 관한 리뷰 글자수, 리뷰에 포함된 오타 자 수, 문장 길이 등의 변수를 사용하였다.

본 연구는 고객 리뷰 분석에 대한 두 가지 연구흐름을 혼합하여 제품 속성에 대한 고객 태도를 분석하여 제품 포지셔닝 맵을 구성하는 방안을 제안한다. 스마트 스피커 제품이 기존의 포터블 블루투스 스피커에 인공지능 속성을 결합한 제품으로 볼 수 있기에 인공지능 속성에 대한 고객 태도를 중점으로 제품간 비교를 수행하는 방안을 제안하였다.

3. 자료

국내에 출시된 스마트 스피커 중에서 KAKAO의 KAKAO MINI, NAVER의 클로바, SK텔레콤의 NUGU, KT의 Giga Genie, LG thinkQ hub의 리뷰 데이터를 수집하였다. 스마트 스피커를 판매하는 채널의 고객 리뷰 데이터는 대부분 짧게 작성되어 있으며, 스마

트 스피커 사용경험보다는 구매경험에 대한 내용이 많기 때문에 제외하였다. 스마트 스피커 제품 사용 경험에 대해 비교적 길게 작성되어 있으며 다양한 측면을 언급하고 있는 블로그 포스ティング을 수집하기로 하였으며, 네이버 API를 사용하여 네이버 블로그 포스ティング을 가져왔다.

블로그 포스ティング 내용 수집은 API로 가져온 URL을 통해 BeautifulSoup을 활용해서 HTML구조를 가져온 후에 블로그 내용을 크롤링(Crawling)하여 수집하였다. 국내 첫 스마트 스피커 출시일인 2016년 9월 1일부터 2018년 9월 7일까지 포스ティング된 블로그 내용을 수집하였다. 네이버 API 검색 키워드로는 '브랜드명(카카오, 네이버 등)', '제품명(미니, 클로바 등)', '인공지능 스피커', 'AI 스피커' 등을 조합하여 <표 1>과 같이 총 112개를 사용했다.

숫자, 특수기호 등을 제거하는 등의 전처리를 수행하였다. 가져온 블로그 포스트에서 스마트 스피커 관련

글인지 확인한 후에 블로그 포스트를 각각의 브랜드별로 분류하기 위해서 규칙기반 분류 방안을 활용하였다. <표 2>의 규칙 1에 기술된 것처럼 '인공지능', '비서', '인공', '지능', '인공지능', '자연어', '음성인식', '음성', '인식', 'ai', '스마트'가 속해 있는 포스트만을 1차로 추출하였으며 총 18,241개 포스트가 수집되었다.

1차로 추출된 블로그 포스트에 규칙 2를 적용하여 블로그 포스트에 기술된 내용이 어떤 브랜드에 대한 것인지를 분류하였다. '카카오', '네이버', 'SK', 'KT', 'LG' 관련 단어들을 조건문에 넣어서 조건을 만족하는 경우에 블로그 포스트를 해당 브랜드로 분류하였다. 규칙 2를 적용하여 브랜드가 분류된 블로그 포스트는 총 3,491개였다. 수집된 데이터를 탐색하여 기사 스크랩 되어진 포스트이거나, 광고, 유사 주제의 포스트 등 449개의 포스트를 제거하였으며, 최종적으로 3,042개의 포스트를 수집하였다.

다음 <그림 1>은 분석에 사용된 전체 블로그 포스트

<표 1> Examples of Keywords

인공지능 스피커	카카오 스마트 스피커	네이버 스마트 스피커	sk 스마트 스피커	KT 스마트 스피커	LG 스마트 스피커
AI 스피커	카카오 미니	네이버 클로바	sk 누구	KT 기가지니	LG 씽큐
스마트 스피커	헤이 카카오	스마트 스피커 클로바	스마트 스피커 누구	스마트 스피커 기가지니	스마트 스피커 씽큐

<표 2> Classification Rule

Rule		분류된 게시물 (건)
Rule1	필요조건	18,241
Rule2	KAKAO	1,317
	NAVER	301
	SK	1,226
	KT	477
	LG	170

를 활용하여 작성한 워드 클라우드이다. 데이터 수집을 위해 사용한 검색어는 워드 클라우드 표시에서는 제외하였다. 스마트 스피커 제조 업체의 브랜드에 해당하는 단어가 많이 등장하였다. 이런 단어를 제외하고는 ‘음악’, ‘기능’, ‘날씨’, ‘서비스’, ‘사용’과 같은 단어가 많이 등장하였다. 스마트 스피커가 제공하는 기능에 해당하는 단어들이다.

<표 3>은 제품별로 가장 빈번하게 등장하는 단어 20개를 선정해 보았다. <그림 1>과 마찬가지로 데이터 수집에 활용된 검색어들은 리스트에서 제외하였다. 스마트 스피커 별로 빈번히 등장한 단어와 등장 횟수를 테이블에 표시하였다.

SK 기가지니 제품에 대해서 가장 빈번하게 등장한 단어는 ‘사용’, ‘멜론’, ‘서비스’, ‘음성’, ‘알람’ 등이고 ‘무드등’, ‘캔들’, ‘일정’은 다른 제품에 잘 등장하지 않는 단어들이다. KT 제품에 대해서 가장 빈번하게 등장한



<그림 1> Word cloud of all posts

<표 3> Frequent words of each smart speaker

제품	SK		KT		Kakao		Naver		LG	
	단어	빈도	단어	빈도	단어	빈도	단어	빈도	단어	빈도
1	사용	28,090	가지	14,951	멜론	31,296	웨이브	18,200	음성	3,767
2	멜론	24,157	사용	12,232	사용	27,174	프렌즈	11,219	전자	3,379
3	서비스	22,892	뮤직	11,1889	네이버	26,272	카카오	10,713	기능	3,133
4	음성	21,194	음성	9,650	뉴스	25,374	음악	7,780	수	2,633
5	알람	18,404	블루투스	9,416	노래	25,082	기능	6,431	가전	2,607
6	버튼	17,962	버튼	8,080	알람	23,946	뮤직	6,344	네이버	2,534
7	오늘	17,382	홈	7,889	누구	23,384	사용	6,127	제품	2,450
8	노래	17,330	올레	7,519	수	22,204	등	5,710	구글	2,298
9	미니	15,900	이용	7,336	블루투스	21,578	바	5,586	사용	2,120
10	무드등	15,894	날씨	6,563	프렌즈	19,473	서비스	5,241	아마존	1,996
11	블루투스	15,879	정보	6,421	음성	19,270	블루투스	4,908	날씨	1,958
12	음성인식	13,181	생활	6,069	버튼	17,980	미니	4,451	제어	1,935
13	뉴스	13,133	연결	5,900	바	17,519	클로버	4,430	서비스	1,895
14	연결	12,915	기기	5,486	연결	16,342	검색	4,242	음악	1,861
15	제품	12,540	네이버	5,221	서비스	13,842	수	4,076	정보	1,746
16	캔들	12,398	연동	4,988	라디오	13,617	날씨	3,908	확인	1,716
17	시간	11,763	비서	4,893	스마트	13,207	출시	3,758	플랫폼	1,683
18	전원	11,670	제공	4,380	톡	13,113	음성	3,620	블루투스	1,662
19	말	11,498	노래	4,308	오늘	12,707	라인	3,439	바	1,632
20	일정	11,307	인터넷	4,285	정보	11,778	노래	3,438	음성인식	1,628



<그림 2> Word clouds of each product

단어는 ‘가지’, ‘사용’, ‘뮤직’, ‘음성’, ‘블루투스’이고, KT 제품에는 등장하지만 다른 제품에는 잘 등장하지 않는 단어들은 ‘생활’, ‘비서’와 같은 단어다.

카카오 제품에 대해서는 ‘멜론’, ‘사용’, ‘네이버’, ‘뉴스’, ‘노래’라는 단어들이 가장 빈번히 등장하였으며,

‘프렌즈’, ‘라디오’, ‘톡’이라는 단어가 비교적 많이 등장했다. 카카오 제품에 대한 설명에 비교를 위해서인지 네이버라는 단어가 많이 출현했고, 네이버 제품에도 ‘프렌즈’, ‘카카오’와 같은 카카오 제품에 대한 단어가 상대적으로 많이 출현했다. 네이버 제품에는 ‘웨이브’,

‘프렌즈’, ‘카카오’, ‘음악’, ‘기능’이라는 단어가 가장 빈번하게 등장했다. 네이버 서비스의 한 종류인 ‘라인’이라는 단어는 네이버 제품에 많이 등장하였다.

LG 제품에는 ‘음성’, ‘전자’, ‘기능’, ‘수’, ‘가전’이라는 단어가 빈번히 등장하였으며, ‘네이버’, ‘구글’, ‘아마존’ 같은 타 제품 제조사 명이 비교적 많이 등장하였다.

<그림 2>는 제품별 워드클라우드이다.

4. 분석 및 결과

블로그 포스트의 내용으로부터 스마트 스피커 제품에 대한 인식을 파악하기 위하여 다음과 같은 과정을 진행하였다. 첫번째 단계로 블로그 포스트 내용이 제품의 어떤 속성에 대한 것인지 파악하기 위해서, 포스트에 등장하는 단어가 속하는 제품 속성을 파악하였다. 두번째 단계는 블로그 포스트를 문장 단위로 나눈 후에 문장이 어떤 속성에 대한 내용인지를 첫번째 단계에서 마련한 단어-속성 매칭 결과를 활용하여 파악하였다. 세번째 단계는 문장이 속성에 대해 긍정적인지 부정적인지를 파악하기 위한 감성분석을 수행하였다. 네번째 단계는 세번째 단계에서 파악한 문장 별 감성점수

와 두번째 단계에서 평가한 문장 별 속성을 활용하여 제품별로 속성에 대한 평균 감성점수를 구하였으며, 이를 이용하여 제품들의 포지셔닝 맵을 작성하였다.

블로그 포스트에 등장하는 단어를 제품 속성에 할당하기 위해서 먼저 스마트 스피커의 제품 속성을 파악하여야 한다. 스마트 스피커의 속성을 정리 하기위해 스마트 스피커의 속성에 대해 연구한 논문을 참조하였으며 (강민영 2017; 신훈철 2016; 이홍주 2018a), 제시된 속성은 <표 4>와 같다.

스마트 스피커에 대한 기존문헌은 제품에 대한 속성을 스피커에 대한 일반속성과 스마트 속성으로 구분하였다. 일반속성은 음악 재생, 고객 서비스 경험 등에 대한 일반적인 내용이며, 스마트 속성이 정보 제공 및 연동된 기기 제어나 서비스 연결에 대한 속성이다. 기존 문헌에서 제시된 내용과 블로그 포스트에 많이 언급되는 단어들을 기반으로 일반 속성과 스마트 속성을 파악하였다. 일반 속성은 가격, 휴대성, 디자인, 사운드로 정의하고 스마트 속성으로는 음성인식, 음악 정보제공, 앱-기기연동, 기타편의기능으로 정의했다.

블로그 포스트에 포함된 단어를 관련된 속성 카테고리에 할당하였다. 우선 단어빈도수(TF)와 TF-IDF 방안 (Baeza-Yates & Ribeiro-Neto, 2011)을 통해서 TF값

<표 4> Features of smart speakers

	이홍주(2018a)	강민영(2017)	신훈철(2016)	도출 속성	
일반 속성	Customer Service Experience	날씨, 교통정보		가격	
	Music	음악재생		디자인	
		타이머/스케줄 관리		사운드	
스마트 속성	Connectivity	IPTV 연동 서비스	Strengthen linking	휴대성	
	Control Apps	라디오/뉴스/팟캐스트	User Recognizing	음성인식	
	Sound Quality	인터넷 정보검색	Integrating Function	정보제공	
	Upgrade	가정용 사물인터넷		앱-기기연동	
	Smart Features	쇼핑 및 음식주문			기타 편의기능
		일상대화			

<표 5> Keyword Example - Category mapping

일반속성				스마트 속성				
가격 (12개)	휴대성 (4개)	디자인 (33개)	사운드 (8개)	기타 편의기능 (32개)	음성인식 (27개)	음악 (12개)	정보제공 (29개)	앱-기기연동 (42개)
가격	배터리	디자인	소리	기능	음성인식	노래	알리	연결
구매	충전	버튼	음질	가지	누구	음악	날씨	스마트폰
무료	케이블	조명	사운드	서비스	기가지니	멜론	데이트	블루투스
판매	충전기	무드등	들리	가능하다	인공지능	이용권	시간	연동

과 TF-IDF 값이 큰 단어 1000개를 각각 선정하였다. 이 중에서 중복으로 선정된 단어들은 제거하였으며, 띄어쓰기가 제대로 되지 않은 단어도 제외하였다. 또한, 스마트 스피커와 무관한 ‘포스팅’, ‘리뷰’ 등의 단어를 제외하였으며 분류가 모호한 단어들도 제외하였다. 추가로 같은 뜻이지만 다르게 표현되는 이형동의어도 하나의 단어로 변환하였다. 이를 통해 총 199개의 키워드를 추출하였으며, 추출된 키워드를 <표 4>에서 도출된 속성에 할당하였다. 스마트 스피커를 사용중인 학생 5명이 199개의 키워드를 검토하였으며, 키워드가 속할 속성 카테고리를 논의하여 할당하였다. <표 5>가 속성과 속성에 할당된 키워드 리스트의 예이다.

가장 많은 단어가 할당된 속성은 앱-기기 연동이며, 디자인, 기타 편의기능, 정보제공 속성에도 많은 단어들 이 할당되었다. 휴대성과 사운드, 가격, 음악은 상대적으로 많은 단어가 할당되지는 않았다.

스마트 스피커 속성에 할당된 단어의 수가 199개로 수집한 블로그 포스트가 가지고 있는 전체 단어의 수에 비해 많지 않다. 이로 인해 많은 문장을 특정 속성에 할당하는 것이 어렵게 된다.

이를 개선하기 위해 속성에 할당된 키워드와 유사한 단어들 을 찾아서, 유사 단어를 해당 속성의 키워드로 추가하는 방법을 활용하였다. 유사한 단어의 선정은 3,042건의 블로그 포스트에 포함된 단어들을

Word2vec을 사용하여 임베딩 하였다.

Word2vec 임베딩은 제품별로 단어가 가진 의미를 다룰 수 있기에, 제품별로 임베딩을 수행하였으며, 단어 별 임베딩 벡터를 활용하여 가장 유사한 단어 30개를 선정하였다.

워드 임베딩을 활용하여 유사한 단어를 선정하는 방안은 문장 분류를 위한 방안에 활용되어 왔다(서덕성 등 2017; 이민식 등 2019). 유사한 단어를 선정하는 방안은 단어별로 가장 유사한 N개의 단어(Top-N)를 선택하는 방안과 코사인 유사도 값(Similarity)에 따라 선정하는 방안이 있을 수 있다. <표 6>은 두 가지 방식에 따라 유사 단어를 선정했을 때의 중복 단어 수를 보여 주고 있으며, 하나의 제품에 대한 블로그 포스트를 임베딩하여 사용한 경우이다. Top-30을 사용한 경우는 많은 단어들 이 유사 단어로 선정되기에 속성 카테고리에 중복할당되는 경우가 많지만 속성 카테고리에 해당하는 많은 단어를 확보할 수 있다. Top-30에 비해 Top-20, Top-10은 속성 고유의 단어만을 활용할 수 있다. 유사도 랭킹이 아니고 유사도 값을 활용하여 유사 단어를 선정하는 경우에는 더 고유한 단어를 활용할 수 있는 것으로 분석되었다. 유사도 0.4를 기준으로 하는 경우에도 총 단어의 수가 448개로 작은 단어만 선정된다. 따라서 많은 단어를 선정하여 속성 카테고리에 해당하는 단어를 많이 확보하는 방안과 적은 단어지만 속성에

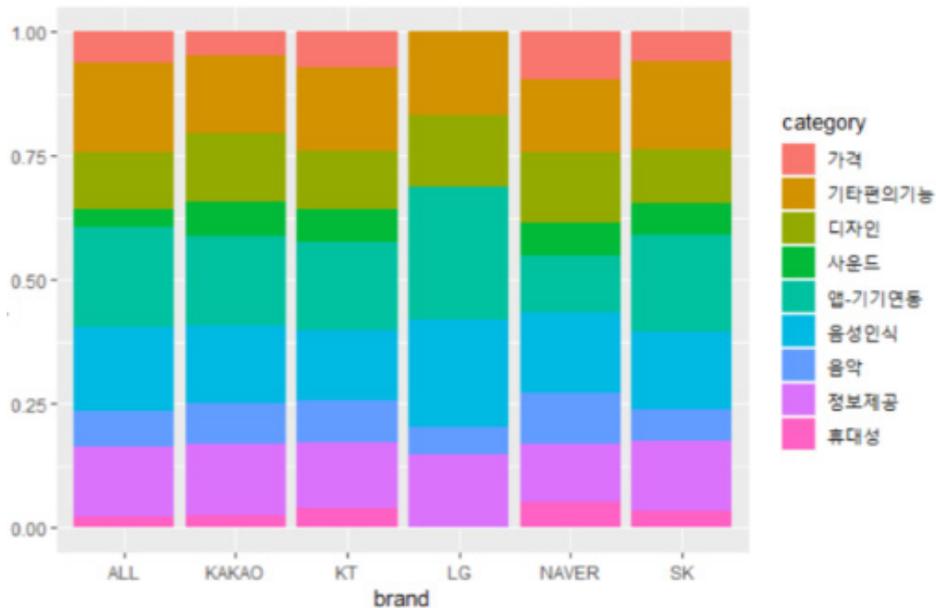
<표 6> Similar word selection

	Top-N			Similarity			
	30	20	10	> 0.7	>=0.6	>=0.5	>=0.4
중복 없는 단어	4181	3130	1862	120	152	272	150
2개 카테고리 중복	228	250	273	0	6	23	150
3개 카테고리 중복	212	192	153	0	0	2	37
4개 카테고리 중복	178	171	112	0	0	1	11
5개 카테고리 중복	179	142	58	0	0	0	0
6개 카테고리 중복	135	95	37	0	0	0	0
7개 카테고리 중복	123	66	12	0	0	0	0
8개 카테고리 중복	63	17	2	0	0	0	0
9개(모든) 카테고리 중복	10	2	0	0	0	0	0

대한 고유한 단어를 선정하는 방안이 있을 수 있다. 이는 텍스트 데이터를 가지고 <표 6>과 같은 분석을 수행한 후에 적절한 방안을 선택하는 것이 바람직하다. 본 연구에서는 되도록 문장을 속성 카테고리에 할당하기 위하여 많은 단어를 확보할 수 있는 Top-30 방안을 적용하였다. 5개 제품을 합쳐 총 18,230개의 단어가 선정되었으며, 제품별로는 평균 3,648개 단어가 사용되었다. 중복되는 단어를 제거하였기에 199개의 단어가

평균적으로 20개의 유사한 단어로 확장된 경우이다.

포스트의 문장이 어떤 속성에 대해 언급하고 있는 것인지를 파악하기 위하여, 블로그 포스트를 문장단위로 나누었으며 총 429,045개의 문장을 추출하였다. R의 KWIC 함수(Scott & Bondi, 2010)를 이용하여 문장에 포함된 단어들이 가장 많이 포함되어 있는 속성을 파악하여, 문장을 해당 속성에 할당하였다. <그림 3>이 전체 문장과 제품별 문장의 속성 분포이다. 전체



<그림 3> Feature distributions

<표 7> Scores for Smart Speakers

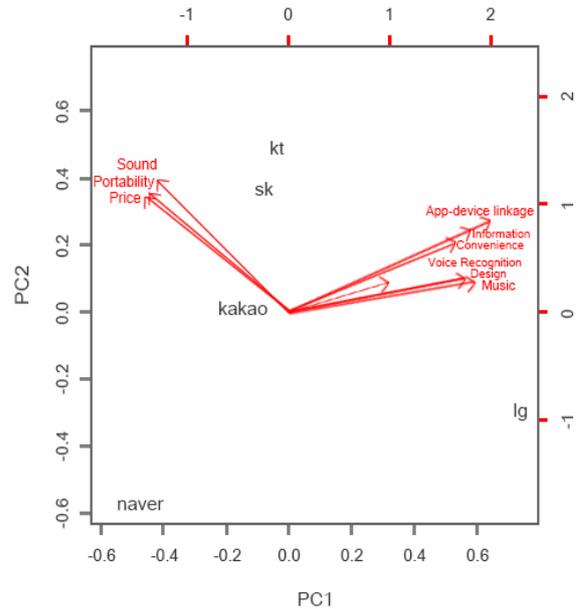
	SK	NAVER	LG	KT	KAKAO
디자인	0.495841	0.368778	0.642412	0.503263	0.461356
가격	0.549718	0.412791	NaN	0.563762	0.493343
사운드	0.484794	0.313043	NaN	0.506127	0.430045
휴대성	0.479233	0.360215	NaN	0.500531	0.321945
음성인식	0.521475	0.347552	0.636905	0.536781	NaN
앱-기기연동	0.481672	0.330769	0.616185	0.525256	0.583298
음악	0.524	0.380573	0.701408	0.532591	0.516645
정보제공	0.534306	0.35414	0.650108	0.543436	0.525242
기타편의제공	0.525793	0.374868	0.635185	0.541343	0.487525

적으로 앱-기기연동, 음성인식, 정보제공에 대한 문장이 많이 분포하고 있으며, 가격, 휴대성, 음악, 사운드에 대한 문장은 상대적으로 적게 분포하였다.

문장의 감성분석은 오픈한글 API(안정국과 김희웅 2015)를 이용하여 수행하였으며, 감성 분석 결과는 긍정, 부정, 중립으로 분류하였다. 중립으로 분류된 문장을 제외하고 긍정 문장에는 1점, 부정 문장에는 -1점을 부여하였으며, 브랜드별로 속성에 대한 평균 감성점수 값을 파악하였다. <표 7>이 제품별 속성의 평균 감성점수이다. NaN으로 표시된 경우는 해당 속성에 해당하는 긍/부정 문장이 없는 경우이며, 제품별로 평균 감성점수 값이 큰 세개의 속성을 음영으로 표현하였다.

높은 순위에 자주 들어간 스마트 속성은 정보제공, 음악, 기타 편의 제공이며, 일반 속성은 가격이였다. SK는 가격, 정보제공, 기타편의제공이 가장 긍정적이었으며, NAVER는 가격, 음악, 기타편의제공이 가장 긍정적이었다. LG는 음악, 정보제공, 디자인이며, KT는 가격, 정보제공, 기타편의제공, KAKAO는 앱-기기연동, 정보제공, 음악이었다.

제품별 속성에 대한 평균 감성스코어를 활용하여 PCA 차원분석을 통해 제품별 포지셔닝을 파악하였으며, <그림 4>이 PCA 차원분석의 결과이다.



<그림 4> Product positioning

KT와 SK는 소리와 가격, 휴대성과 같은 일반적인 속성에 큰 강점을 보였고, KAKAO는 대체적으로 모든 측면에서 이점을 보였다. LG는 음성인식과 음악과 같은 스마트 속성에서 큰 강점을 보였으며 NAVER는 전반적으로 속성에 대한 감성 스코어가 비슷하게 낮은 값을 보였다. 분석에 사용한 블로그 포스트 중에서 LG에 대한 내용이 매우 적었으며, 가격, 사운드, 휴대성에 대

해서는 감성 값이 측정되지 않았다. 따라서, 도출된 포지셔닝 맵에서 LG 제품이 인공지능 속성 측면에서 좋은 평가를 얻은 것으로 나온 것은 이러한 부분을 감안하여 해석하여야 한다.

5. 토의 및 결론

본 논문에서는 제품에 대한 고객 경험 리뷰를 활용하여 제품의 속성에 대한 사용자들의 태도에 근거한 제품 포지셔닝 맵 도출방안을 제안하였다. 본 연구는 감성 분석에 대한 연구와 제품 포지셔닝 연구를 연계하여 속성에 대한 감성 점수를 고려한 포지셔닝 맵을 제시하였다. 스마트 스피커가 가진 일반 속성과 스마트 속성을 구분하였고, 속성 카테고리에 속하는 단어의 확장을 위해 Word2Vec을 활용하였다. 이를 통해 제품별 혹은 속성별 감성 수치를 도출하는 것이 가능하며, 언급되는 텍스트 자체에 기반한 제품 포지셔닝이 아니고 속성에 대한 감성 수치를 활용한 포지셔닝 방안을 제안하였다. 스마트 속성에 대해 매우 긍정적인 제품과 일반 속성에 대해 매우 긍정적인 제품을 찾을 수 있었으며, 대부분의 속성에서 긍정적이지 못한 제품 또한 파악되었다. 본 연구의 학문적 의의는 다음과 같다.

첫째로, 리뷰 토픽이나 문장이 특정 속성에 해당하는지를 파악하여 속성에 대한 감성을 파악하는 연구와 제품간의 비교를 위한 연구를 연계하였다. 속성에 대한 태도를 분석하여 제품간의 비교를 수행하는 방안을 제안하였다. 각 제품에 대한 고객 태도 파악도 가능하며 제품간 비교도 수행할 수 있는 방안이다.

둘째로, 기존 문헌에 근거하여 스마트 스피커 속성 카테고리를 파악하였으며 빈번히 등장하는 단어를 카테고리화 매칭하여 문장이 언급하는 내용을 속성에 할당하기 위한 초기 단어 집합을 구성하였다. 스마트 스피커에 대한 텍스트 분석 문헌이 많지 않기에 카테고리

리 분류에 참조하고, 사람이 유의미하게 구분할 수 있는 수준의 단어들을 카테고리에 매칭하였다. 모호하거나 중복되는 단어들을 제외하고 명확한 단어들을 할당할 수 있는 방안을 제안하였다.

마지막으로 단어와 속성 카테고리에 관한 매칭을 연구자들이 일부 단어에만 수행하고, 워드 임베딩을 통해 다른 단어로 확장하는 방안을 제안하였다. 많은 단어들이 고객 리뷰에서 다양한 의미로 사용되므로 핵심적인 단어만 수동으로 분류된다면 이를 활용하여 유사한 다른 단어로 확장하는 것이 가능하다. 또한, 하나의 단어가 제품별 로 다양한 의미를 표현할 수 있기에 제품별 리뷰만을 활용한 텍스트를 활용하여 유사 단어를 찾을 수 있다. 일반적인 상황에서의 유사성 기준을 적용하지 않고 제품 리뷰에서의 유사성을 사용하여 단어를 확장할 수 있는 방안이다.

본 연구가 고객의 경험 또는 인식을 통한 제품 포지셔닝 맵 도출을 위한 방안을 제시하고 있지만 다음과 같은 한계가 있다. 본 연구에서 제품에 대해 언급한 블로그 포스트 수의 차이는 고려되지 않았다. 특정 제품은 훨씬 많은 고객 리뷰가 작성되었고 다른 제품은 그렇지 못하였다. 본 방안에서는 이러한 차이가 고려되지 않았으며, 속성별로는 고려되었지만 고객 리뷰 전체의 감성 수치는 고려되지 않았다. 감성 분석 연구에서는 전체 리뷰의 감성과 등장하는 횟수를 고려하는 방안들이 제시되고 있기에 이러한 측면을 반영하는 것이 필요하다. 마찬가지로 속성별 문장의 수도 상이하기에 이에 대한 고려도 필요하다. 속성별 평균 감성 수치가 활용되고 있기에 이러한 차이를 고려하지 않아도 속성별 비교는 타당하다. 두번째 한계는 유사도 단어 확장 방안의 기준 설정에 따라 다른 결과가 나올 수 있다는 것이다. 본 연구에서는 적절한 기준 설정 방안을 도출하지는 않았다. 본 문에 기술된 것처럼 텍스트 데이터에 따라 상이한 기준이 적용되어야 하며, 선택을 위한 기준은 제시 되어있다.

참고 문헌

[국내 문헌]

1. 강민영 2017. “인공지능(AI) 가전제품 문제점 및 개선방안(음성인식 스피커를 중심으로)”, *한국소비자원 조사보고서*.
2. 김정훈, 송영은, 진운선, 권오병 2015. “텍스트마이닝을 통한 댓글의 공감도 및 비공감도에 영향을 미치는 댓글의 특성 연구”, *한국IT서비스학회지* (14:2), pp. 159-176.
3. 박명석, 권영진, 이상용 2018. “댓글이 음원 판매량에 미치는 차별적 영향에 관한 텍스트마이닝 분석”, *지식경영연구* (19:2), pp. 91-108.
4. 서덕성, 모경현, 박재선, 이기창, 강필성 2017. “워드 임베딩과 그래프 기반 준지도학습을 통한 한국어 어휘 감성 점수 산출”, *대한산업공학회지* (43:5), pp. 330-340.
5. 신희철, 김종학, 박영택 2016. “카노모델(Kano Model)을 이용한 스마트 오디오 컨셉 기능의 고객 만족에 관한 연구”, *한국품질경영학회* (44:4), pp. 951-963.
6. 이민석, 양석우, 이홍주 2019. “문장 분류를 위한 정보 이득 및 유사도에 따른 단어 제거와 선택적 단어 임베딩 방안”, *지능정보연구* (25:4), pp. 105-122.
7. 이진명, 정민지, 이주래, 김예은, 안치연 2019. “인공지능 스피커에 대한 소비자 인식과 수용의도: 비수용자를 중심으로”, *소비자학연구* (30:2), pp. 193-213.
8. 안정국, 김희웅 2015. “집단지성을 이용한 한글 감성어 사전 구축”, *지능정보연구* (21:2), pp. 49-67.
9. 이홍주 2018a. “A Ghost in the Shell? 고객 리뷰를 통한 스마트 스피커의 인공지능 속성이 평가에

미치는 영향 연구”, *한국IT서비스학회지* (17:2), pp. 191-205.

10. 이홍주 2018b. “헬스케어 서비스 리뷰를 활용한 서비스 품질 차원 별 중요 단어 파악 방안”, *지식경영연구* (19:4), pp. 171-185.
11. 이홍주 2019. “인공지능 속성에 대한 고객 태도 변화: AI 스피커 고객 리뷰 분석을 통한 탐색적 연구”, *지식경영연구* (20:2), pp. 25-42.
12. 컨슈머인사이트 2018. “뜨거운 AI스피커 시장, 차가운 소비자 평가”, 이동통신 리포트, https://www.consumerinsight.co.kr/voc_view.aspx?no=2924&id=ins02_list&PageNo=1&schFlag=0

[국외 문헌]

1. Baeza-Yates, R., B. Ribeiro-Neto 2011, *Modern Information Retrieval: The Concepts and Technology behind Search* (2nd Edition), Addison-Wesley Professional.
2. Berger, J., A. Humphreys, S. Ludwig, W. W. Moe, O. Netzer, D. A. Schweidel 2020. “Uniting the Tribes: Using Text for Marketing Insight”, *Journal of Marketing* (84:1), pp. 1-25
3. Cao, Q., W. Duan and Q. Gan 2011. “Exploring determinants of voting for the ‘helpfulness’ of online user reviews: A text mining approach”, *Decision Support Systems* (50:2), pp. 511-521.
4. Chen, K., G. Kou, J. Shang and Y. Chen 2015. “Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches”, *Electronic Commerce Research and Applications* (14:1), pp. 58-74.

5. Ghose, A., P. G. Ipeirotis and B. Li 2012. "Design Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content," *Marketing Science* (31:3), pp. 493-520.
6. Humphreys, A. and R. J. Wang 2018. "Automated Text Analysis for Consumer Research," *Journal of Consumer Research* (44: 6), pp. 1274-1306.
7. Lee, T. Y. and E. T. Bradlow 2011. "Automated marketing research using online customer reviews," *Journal of Marketing Research* (48:5), pp. 881-894.
8. Netzer, O., R. Feldman, J. Goldenberg and M. Fresko 2012. "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science* (31:3), pp. 521-543.
9. Scott, M., and Bondi, M. 2010. *Keyness in Texts*. (pp. 21-42). Amsterdam, Philadelphia: John Benjamins.

● 저 자 소 개 ●



이정현 (Jung Hyeon Lee)

현재 가톨릭대학교 정보통신전자공학부 학부생이다. 주요 관심분야는 NLP(자연어 처리), Machine learning 등을 통한 경영학적 문제 해결이다.



선형주 (Hyung Joo Seon)

현재 연세대학교에서 빅데이터 분석 전공 석사과정을 진행 중이다. 가톨릭대학교에서 심리학 학사를 취득하였다. 주요 연구 및 관심분야는 텍스트 마이닝, 네트워크분석, 딥 뉴럴 네트워크, 머신러닝 등을 활용한 비즈니스 문제 해결이다.



이홍주 (Hong Joo Lee)

현재 가톨릭대학교 경영학부 교수로 재직 중이다. KAIST 테크노경영대학원에서 경영학 박사 학위를 취득하였다. 주요 관심분야는 데이터 분석, 지능형 정보시스템, 온라인 사용자들의 상호작용 등이다. 지금까지 Journal of Electronic Commerce Research, Cyberpsychology, Behavior, and Social Network, Technological Forecasting & Social Change, International Journal of Electronic Commerce 등 주요 학술지에 논문을 발표하였다.