

# 제품, 서비스, 융합제품서비스의 소비자 니즈 비교 분석 :아마존 온라인 리뷰를 중심으로

## Comparative Analysis of Consumer Needs for Products, Service, and Integrated Product Service : Focusing on Amazon Online Reviews

김성범

금오공대 IT융합학과

Sungbum Kim(it89@kumoh.ac.kr)

### 요약

이 연구는 텍스트 마이닝을 사용하여 하드웨어 제품에 대한 리뷰, 서비스 상품에 대한 리뷰, ICT분야의 하드웨어와 클라우드 서비스가 융합된 형태의 상품을 대상으로 소비자 리뷰를 분석한다. 분석을 위해 각 리뷰의 키워드를 도출하고 토픽 도출에 사용된 단어의 차별성을 찾는다. 마지막으로 전체 리뷰를 대상으로 군집분석을 실시하고 각각의 상품군의 리뷰가 어떤 군집에 속하는지를 검토한다. 이 연구를 통해서 각 상품의 유형별로 특화되어 사용된 핵심어를 도출하였고, 토픽모델링을 사용하여 제품과 서비스의 특성을 표현하는 주제를 도출하였다. 서비스 상품 리뷰에서는 공급자의 우수성을 의미하는 professional, technician과 같은 핵심어를 도출하였고, 융합제품서비스상품으로서 아마존 에코 리뷰에서는 favorite, fine, fun, nice, smart, unlimited, useful 등의 긍정적 의미의 형용사를 도출하였다. 군집분석을 사용하여 전체 리뷰를 분석하였고, 3개의 상품 유형별 리뷰가 배타적으로 서로 다른 각각의 군집에 속하는 결과를 발견하였다. 이 연구는 소비자의 니즈(needs)를 상품의 유형별로 온라인 리뷰를 이용하여 차이점을 분석하였고 실무적으로 상품 유형에 기반한 상품기획과 마케팅 프로모션 차별화의 필요성을 제시하였다.

■ 중심어 : | 아마존 | 온라인 리뷰 | 제품 | 서비스 | 융합제품서비스 | 텍스트마이닝

### Abstract

The study analyzes reviews of hardware products, customer service products, and products that take the form of a convergence of hardware and cloud services in ICT using text mining. We derive keywords of each review and find the differentiation of words that are used to derive topics. A cluster analysis is performed to categorize reviews into their respective clusters. Through this study, we observed which keywords are most often used for each product type and found topics that express the characteristics of products and services using topic modeling. We derived keywords such as "professional" and "technician" which are topics that suggest the excellence of the service provider in the review of service products. Further, we identified adjectives with positive connotations such as "favorite", "fine", "fun", "nice", "smart", "unlimited", and "useful" from Amazon Eco review, an integrated product and service. Using the cluster analysis, the entire review was clustered into three groups, and three product type reviews exclusively resulted in belonging to each different cluster. The study analyzed the differences whereby consumer needs are expressed differently in reviews depending on the type of product and suggested that it is necessary to differentiate product planning and marketing promotion according to the product type in practice.

■ keyword : | Amazon | Online Review | Product | Service | Integrated Product Service | Text Mining |

\* 이 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2018S1A5A8027897)

접수일자 : 2020년 05월 04일

심사완료일 : 2020년 06월 16일

수정일자 : 2020년 06월 16일

교신저자 : 김성범, e-mail : it89@kumoh.ac.kr

## I. 서론

전자상거래에 있어서 소비자의 리뷰(Review)는 소비자의 마음을 읽을 수 있는 소중한 마케팅 자원이며 미래에 제품의 개선 방향과 마케팅 전략의 방향성을 수립하는 분석의 기초로서 중요한 역할을 한다[1][2]. 제품의 리뷰는 추가적인 소비자의 유입을 일으키기도 하고 반대로 소비자가 유출되는 단서가 되기도 한다[3]. 소비자가 온라인에서 상품을 구매할 때 제품 리뷰에 높은 영향을 받는다는 것을 고려할 때[3] 제품에 대한 리뷰의 중요성은 점점 더 중요해 지고 있다.

이 연구는 텍스트 마이닝(Text mining)을 사용한 내용분석(Content Analysis)를 통해서 세 가지 상품 유형의 소비자 리뷰를 분석한다. 첫째는 하드웨어 형태의 제품에 대한 리뷰이고, 둘째는 물건의 형태가 아닌 사람의 용역을 대상으로 하는 서비스 상품에 대한 리뷰이며, 셋째는 ICT분야의 하드웨어 형태와 클라우드 서비스가 융합된 형태의 상품의 리뷰이다.

기존의 선행연구들은 리뷰 분석의 정확성을 높이기 위한 텍스트 마이닝 방법론을 제시하거나[4-7], 소비자의 리뷰와 구매와의 관련성을 탐구하기도 하였고[8][9], 내용상으로 리뷰의 감성 분석에 주목하기도 하였다[10-12]. 이 분야의 연구들은 특별히 리뷰가 발생하는 분야별로 호텔, 관광[3][13], 레스토랑에 대한 사용자 리뷰 분석을 통해[3][14] 특정 카테고리의 리뷰 특성을 도출하기 위한 연구로 확장되었다.

기존의 연구와 비교하여 이 연구의 차별화된 목적은 첫째, 분석의 대상을 제품(product), 서비스(service), ICT분야의 하드웨어 제품과 서비스가 융합된 상품으로 나누어 소비자의 리뷰 내용의 차이점을 비교 분석하는 점이다. 특별히 ICT분야의 하드웨어 제품과 서비스가 융합된 상품인 경우 이 상품에 대한 소비자 리뷰의 특성 분석은 미래에 출현하는 많은 ICT 제품서비스융합 상품의 기획과 개발, 마케팅을 위해 필수적인 작업이다. 이런 차별화된 연구목적 가지고 이 연구에서는 상품의 3개 유형을 구분하여 각각의 리뷰를 분석하기 위해 아래와 같이 연구 질문을 설정하였다.

(연구질문 1-1) 각각의 상품 카테고리별 (하드웨어 제품, 서비스, ICT 분야의 제품 서비스 융합 상품) 리뷰

의 주제는 무엇인가?

(연구질문 1-2) 각각의 상품 카테고리별 리뷰에 사용된 형용사와 동사 중에서 공통으로 사용된 단어와 차별적으로 사용된 단어는 무엇인가?

(연구질문 2) 세 가지 유형으로 구분된 카테고리별 상품 리뷰는 서로 차별적이어서 배타적으로 구분된다고 할 수 있는가?

리뷰 내용 중에서 공통된 키워드와 차별적인 키워드의 도출은 해당 상품을 기획하는 데 있어서 필수적인 기능을 탐색하는데 사용될 수 있으며, 소비자에게 무엇을 소구할 것인지, 시장에 어떻게 포지셔닝(positioning) 할지를 고민하는 마케팅 전략 수립에 실무적 시사점을 줄 수 있을 것이다.

이 연구를 진행하는 기본적인 접근 방법론은 텍스트로부터 (i) 각 리뷰의 키워드를 순위별로 도출하여 차별적인 단어의 존재를 탐구하고, (ii) 특정 제품군 리뷰에만 차별적으로 쓰여진 키워드를 도출하고, (iii) 토픽을 도출하여 토픽 또는 토픽모델링에 사용된 단어의 차별성을 탐색한다. (iv) 또한 전체 리뷰를 대상으로 군집화했을 때 각각의 상품군 리뷰가 어떤 방식으로 군집화되는지를 검토한다. 이를 통해 상품의 유형별 리뷰의 차별성을 검증한다.

이 연구는 (i) 분석의 대상, (ii) 분석 알고리즘, (iii) 소프트웨어 툴의 사용, 그리고 (iv) 해석까지 종합적이고 포괄적이라는 특징을 가지며 실무적으로는 텍스트 기반의 빅데이터를 분석할 때 제품의 유형별 리뷰 분석의 실무 지침을 제공하는 것을 목적으로 한다.

## II. 선행연구

### 1. 분석대상으로서의 상품의 종류

상품의 형태는 크게 제품과 서비스로 구분할 수 있다. 특별히 서비스 개념은 서비스 마케팅 분야에서 물리적 형태의 제품(physical product) 마케팅과의 차이점을 고민하면서 제품과 구별되는 차별적인 개념으로 인식되었다. 서비스는 무형성(intangibility), 이질성(heterogeneity), 분리불가능성(inseparability), 쉽게 사라지는 특성(perishability)을 통해 일반적인 제품과

구별되고 있다[15-17]. 무형성에 기인하여 소비자는 서비스 공급자의 평판(reputation)에 의존하게 되며 서비스 공급자는 서비스가 어떠한 것이라는 서비스의 내용을 소비자에게 설명해야 한다. 서비스는 공급 회사의 직원에 의하여 전달되는 것이므로 직원이 바뀔 때 따라 또는 서비스 공급 상황에 따라 다양한 모습을 가지게 된다(heterogeneity).

제품(product)과 서비스(service)의 차별성을 고민한 연구들은 과거 광고 분야에서 두 분야의 차이점을 제시하거나[18], 서비스 혁신의 주제로 제품혁신과의 유사점과 차이점을 분석하면서 서로 다른 특성과 경로를 제시하기도 하였으며[19][20], 2000년대 초반부터 제품서비스시스템(Product service system: PSS)이라는 이름으로 두 분야의 융합을 다루는 연구로 발전되어 진행되었다[21-24].

PSS연구에서 제품서비스시스템이란 기존 하드웨어 제품을 개발 판매하던 회사가 해당 제품의 서비스 기능을 부가적인 기능이 아닌 필수적인 기능으로 인식하면서 등장한 개념으로서 제품과 서비스를 동시에 고려하여 경제적이고 지속 가능한 방식으로 소비자의 니즈(needs)를 충족하고 이해관계자에게 보다 높은 가치를 전달하는 시스템으로 정의된다[23-26]. 대표적으로 제록스의 제품 판매와 서비스, 파커셀(Parkersell)의 조명 시스템, 유탄유 서비스 패키지, 토탈 에너지 소비관리 시스템, 금융서비스를 포함하는 효율적인 세탁 시스템 등을 성공한 PSS로 소개하고 있다[26]. PSS 연구 초기에는 지속가능성(sustainability) 개념과 연계하여 지속 가능한 자원관리의 개념으로 강조되기도 하였고 기업이 어떻게 PSS 비즈니스 모델을 구현하여 실행하는가의 문제[27], 제조업 분야의 서비스화(servitization)에 대한 가능성으로 확대 연구되었다. PSS 연구는 ICT와 같은 특정 분야의 모델로[28][29] 또는 궁극적으로 PSS 성과에 대한 평가를 어떻게 할 것인가의 주제로 다양화되었다. 이 연구들을 종합해 보면 PSS는 서비스 분야의 독자적인 개념 정립과 성과 창출, 그리고 혁신을 위한 차별화된 접근방법을 탐색하고 새로운 개념의 융합제품서비스 분야의 기획과 비즈니스 모델 정립에 기여하고 있다. 이 연구에서 사용하는 융합제품서비스 개념은 기본적으로는 PSS 개념에서 강조하는 서비스의 필수적

기능을 보유하는 것으로서, 연결된 서비스가 없는 경우 존재의 의미가 없는 상품을 의미한다.

## 2. 소비자 리뷰에 관한 연구

소비자 리뷰에 관한 기존의 연구는 몇 가지 기준을 사용하여 분류할 수 있다. 첫째, 판매 제품 또는 서비스 상품의 종류를 기준으로 분류할 수 있으며, 여행, 호텔, 레스토랑, 영화, 아마존에서 판매하는 상품 리뷰를 주된 연구의 대상으로 한다. 둘째, 리뷰가 발생하는 온라인 사이트의 성격을 기준으로 분류할 수 있으며 amazon [4][10-12][30][31], Airbnb[32] 등의 직접 판매와 예약이 이루어지는 곳의 리뷰, trip advisor[14][33][34], yelp[35] 와 같이 리뷰를 전문적으로 수집하는 사이트, Twitter[36][37], blog[38] 등의 SNS 매체에서의 소비자 리뷰 등으로 구분할 수 있다.

소비자 리뷰 연구는 여행 블로그에서 특정 단어의 출현을 조사하여 소비자의 기쁨(delight)을 발견하거나 [38], 온라인 입소문(WOM:word of mouth)과 리뷰가 여행에 미치는 영향을 탐구하는 등 여행을 핵심소재로 전개되기도 하였으며[37], 레스토랑을 대상으로 리뷰의 다차원 감성분석[39]을 시도하거나, 레스토랑 리뷰의 감성분석과 레스토랑 등급을 연결하는 연구가 시도되기도 하였다[40]. 특별히 하드웨어 제품의 기능을 텍스트 마이닝의 방법론을 사용하여 추출한 후 소비자의 구매와의 관련성을 파악하는 연구 등은 제품의 기능에 관한 소비자의 인식을 탐구하는 목적으로 제품의 리뷰를 사용하였다는 점에서 소비자 연구에 대한 리뷰의 기여도를 볼 수 있다[8][9][41].

소비자의 행태와 관련된 연구는 제품의 리뷰가 제품 구입을 고민하는 소비자에게 어떤 영향을 미치는지를 중점적으로 탐색하였으며 리뷰의 성격과 유용성(Usefulness), 유익함(helpfulness)의 점수와의 연관성을 분석하는 연구로 확장되었다[13][35][42][43].

소비자 리뷰 연구는 다양한 방법론을 적용하여 의미 있는 결과를 도출하는 방향으로도 심화하여 진행되었다. 유사하지만 다른 의미가 있는 문장을 판별하는 방법을 SVM(Support vector machine)을 이용하여 접근한 연구[5], 정확한 리뷰 분석을 위한 방법론을 고민하는 연구로서 잘못된 리뷰, 스팸 리뷰를 걸러내는 방

법론을 제시한 연구[4][6][7], 신경망 방법론을 이용하여 제품 리뷰의 점수에 대한 예측 모델링을 시도한 연구 등이 대표적이다[44]. 아마존 리뷰도 이 분야의 직접적인 대상이 되고 있다. 아이폰을 대상으로 아이폰의 기능(feature), 제품 자체, 아이폰에 관한 서비스를 연구자가 정립한 원칙(rule base)에 근거하여 구분한 후 제품등급(product rating)과 서비스등급(service rating)으로 분류한 연구[12]와, 아마존에서 판매하는 디지털 카메라와 캠코더를 대상으로 총 50개 이상의 제품과 15개월간의 자료를 기반으로 패널 데이터 세트를 구성하여 텍스트 마이닝과 소비자선택이론을 융합한 연구가 주목할 만하다. 특히 이 연구는 리뷰에서 등장하는 기능 관련 단어에 점수를 매기는 방법을 사용하여 다양한 제품의 기능에 대한 선호를 보여 주는 계량 모델을 구성하였다. 계량 모델을 구성한 후 판매순위와의 관계를 탐구하였다는 점에서 실무적으로 많은 시사점을 준다[8].

이상의 내용을 정리하면 소비자 리뷰를 대상으로 한 연구는 감성분석(sentiment analysis) 분야, 문장을 이해하고 분류함에 있어서 정확성이 높은 방법론을 발굴 및 적용하는 분야, 소비자의 리뷰를 점수화하여 구매에 어떤 영향을 미치는지 고민한 분야 등으로 구분할 수 있다.

### III. 방법론

#### 1. 분석 프로세스

이 연구는 연구문제의 해결을 기준으로 자료 수집부터 분석까지 총 6단계의 프로세스를 구성하였다. 1단계에서는 분석의 대상을 설정한다. 이때의 대상은 온라인 상거래(commerce)사이트와 분석의 대상이 되는 상품을 의미한다. 온라인 상거래(commerce) 사이트 중에서 소비자의 리뷰를 제공하는 웹사이트를 선정한다. 대상이 되는 웹사이트는 제품(Hardware Product), 서비스(service), 융합제품서비스상품의 리뷰를 포함하는 사이트를 의미한다. 제품과 서비스, 제품서비스융합상품의 대상을 설정할 때 전문가 인터뷰와 기존의 선행연구를 참조하여 최적의 대상을 설정한다. 이 연구에서는

전문가 인터뷰를 통해서 ICT 분야의 융합상품으로 아마존 에코를 대상으로 하였다. 2단계는 텍스트를 추출하는 단계이다. 파이썬을 사용하여 크롤러를 개발하여 사용하였다. 파이썬에서 제공하는 urllib, requests, beautiful soup 모듈을 사용하여 크롤러를 제작하였다. 3단계는 데이터 전처리의 단계이다. 엑셀이나 csv 파일로 추출된 텍스트는 작성일자, 작성자(author), 본문(body text)으로 구분하고 해석이 불가능한 기호나 무의미한 이모티콘을 제거하고 불용어(stopwords)를 사용하여 텍스트를 정제하는 작업을 수행하였다. 4단계는 키워드를 도출하는 단계이다. 핵심어 도출은 일반적으로 빈도수를 기준으로 선정할 수 있으나 이 연구에서는 TF-IDF를 이용하여 키워드를 선정하였다(R and T-LAB plus 사용). 5단계는 주제를 도출하는 단계로 R과 T-LAB plus를 사용하여 토픽모델링(Topic modeling) 기법 중의 하나인 LDA를 사용하였다. 6단계는 군집분석의 단계이다. 텍스트 마이닝 프로그램의 하나인 T-LAB plus를 사용하였고, 군집분석 방법론은 이분화(Bisect) K-means 방법론을 사용하였다.

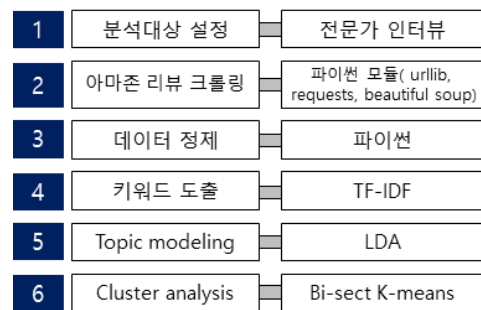


그림 1. 분석순서와 방법론

#### 2. 분석의 대상과 데이터

1차 분석대상으로서 소비자 리뷰가 제공되는 웹사이트는 (i) 하드웨어 중심의 모바일 단말기의 소비자 리뷰를 제공하는 GSMarena, (ii) 서비스 중심의 리뷰를 제공하는 웹사이트로서 Angie's List, (iii) 제품 및 서비스, 제품과 서비스가 융합된 상품 등 종합적인 제품이 소개되고 있는 Amazon, (iv) 기타 리뷰 사이트로서 Choice, Trustpilot, TextFreaks, Yelp를 검토하였

다.

각 사이트는 구조와 목표 등이 다르고 사이트의 특성으로 인한 차이점이 있다. 차이점에 기인하는 연구상의 오류 발생을 통제할 필요가 있으므로 이 연구에서는 연구문제에서 제기한 세 가지 유형의 상품군을 모두 포함하고 있는 아마존(Amazon)을 분석대상으로 하였다. 각 각의 상품 유형별로 리뷰수를 비슷하게 구성하여 각각 15,000개씩 총 4만 5천개의 리뷰를 수집한 후 리뷰의 단어 수가 일정 이하인 리뷰를 제외하고 총 24,899개를 분석대상으로 선정하였다.

융합제품서비스상품을 선정하기 위하여, 기업의 상품 기획 경력자 5명의 전문가 인터뷰를 통하여 의견을 수집하였다. 전문가들은 첫째, PSS의 개념인 제품과 서비스의 연결 고리로서 ICT기술 분야에서 제품 서비스융합상품을 선택해야 한다는 의견을 개진하였고[29], 주된 이유로 ICT 분야는 단말기, 네트워크, 콘텐츠로 구성된 제품과 융합된 서비스를 찾을 수 있는 최적의 산업이라는 주장이 설득력 있게 제시되었다. 둘째, ICT서비스의 대표적인 예로서 클라우드 서비스 사용이 분석의 대상이 되어야 한다는 의견이 개진되었으며 서비스를 구현하기 위해서 필수적으로 단말기가 필수적인 제품으로 구성되어야 한다는 점을 피력하였다.

이 연구에서는 클라우드(cloud) 서비스를 소비자가 단말기를 사용하여 이용하는 대표적인 상품으로 아마존 에코를 선정하였다. 융합제품서비스 상품을 하드웨어 제품 유형, 서비스 유형과 같이 복수의 상품을 구성할 것인지, 단일제품으로 접근할 것인지의 논쟁의 여지가 있을 수 있다. 제품서비스융합의 개념이 단일하게 정의되기 전 단계임을 고려할 때 복수의 상품을 분석할 경우 이질성을 통제하기 어려울 수 있다. 따라서 1차적인 탐색 연구의 의미가 강한 이번 연구에서는 ICT 분야의 단일제품으로 접근하는 것이 개념화에 적절할 것으로 판단하였다. 아마존 에코(Amazon echo)는 단말기(device)를 구입한 후 인터넷에 연결하여 클라우드 서비스를 이용하는 하드웨어 단말기와 서비스가 복합된 상품이다.

이 연구는 최종적으로 하드웨어 제품과 서비스, 하드웨어와 서비스가 결합된 상품에 대한 리뷰 분석을 대상으로 한다. 분석의 대상으로서 제품서비스융합상품은

단일제품이지만 하드웨어 제품과 서비스는 복수의 제품으로 구성되었으며 리뷰수가 많은 순서로 선정되었다. 일반적으로 1개의 제품에 대한 리뷰수가 하드웨어 제품이 서비스 상품에 비해 많기 때문에 동등한 수의 리뷰를 확보하기 위해서 서비스 상품의 종류가 하드웨어 제품에 비해 많이 선정되었다.

분석대상이 된 사이트는 영어로 구성되었으며 이 연구는 기본적으로 제품과 서비스에 대한 영어로 쓰인 소비자의 리뷰를 대상으로 한다.

표 1. 분석대상

구분	분석대상
하드웨어 제품	Headphone, Honeywell air conditioner, LG air conditioner, POT Vornado, Washing machine
서비스 상품	Cleaning (Carpet, Deep, spring, Move In out, Window) Assembly(bed, Desk, Elliptical, Bike, Furniture, Grill, Office chair, Treadmill Setup(Business Laptop, home audio) Installation (Garbage Disposal, Window air conditioner, sound Bar Handyman Services, Repair, Wall hanging
하드웨어서비스 융합 상품	Amazon Echo

### 3. 키워드 선정:TF-IDF

텍스트에서 핵심 키워드를 선정하는 방법으로는 빈도수를 기준으로 하는 방법과 이를 보완하는 방법으로서 TF-IDF 방법이 있다. TF-IDF(Term Frequency - Inverse Document Frequency)는 TF에 DF의 역수를 곱한 값으로 구하며 문서에서 중요 단어를 추출하는 방법의 하나다. 선정된 핵심어는 단어가 문서 내에서 얼마나 중요한지를 표시해 준다. 단순 빈도수에 기반해서 키워드를 추출하는 방법론을 수학적 알고리즘을 바탕으로 수정해 주는 지표이다. TF는 문서 내의 특정 단어의 빈도수이며 DF는 여러 문서 안의 특정 단어 빈도수이다. IDF는 DF의 역수이며 DF가 커질수록 TF-IDF는 작아진다[45][46]. 단순 빈도수와 비교하여 너무 많은 문서에 등장하는 단어를 중요도에 있어서 패널티의 형식으로 보정하는 방법이다. 이 방법을 사용하여 리뷰에서 핵심어 추출의 정확도를 높일 것으로 판단한다.

$$TF-IDF = w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$  =  $j$ 문서에서  $i$ 단어의 출현빈도  
 $N$  = 문서의 수  
 $df_i$  = 단어  $i$ 가 발생하는 문서의 수

#### 4. 토픽모델링 : LDA

토픽모델링은 텍스트를 분석하여 주제를 도출하는 방법이다. 토픽모델링을 설명하는 구성요소는 문서와 주제와 단어이다. 분석의 단위는 문서이며 문서가 모인 문서 집합은 코퍼스(corpus)가 된다. 문서 안의 토픽을 분석하며 도출된 토픽은 단어로 구성된다[47]. 토픽모델링을 이용하여 주제를 도출할 수 있으며, 도출된 주제의 비중을 알 수 있고, 토픽을 설명하는 가장 확률 높은 단어군을 도출한다. 토픽모델링 방법론으로서 LDA (Latent Dirichlet Allocation)는 사전확률과 사후확률로 설명되는 베이스 추론에 바탕을 두고 문서에 존재하는 토픽을 확률적으로 판단하는 방법론이다[48]. 각 문서에 어떤 토픽이 존재하는지 확률적으로 추론한다. 사전확률로서 문서의 주제분포와 주제의 단어분포가 존재하며 이 두 분포 모두 디리클레 분포를 따르는 것으로 가정한다. 즉 LDA는 디리클레 분포를 사용하는 토픽모델링 방법의 하나이다[49]. 이 방법론은 복잡한 확률식을 추정해야 하는데, 이 연구에서는 깃스샘플링 기법(Gibbs Sampling Method)을 사용하여 은닉된 잠재변인을 추론한다[50]. 이 연구에서는 깃스샘플링 방법을 사용하고, 반복 1000회, 주요 파라미터인 알파는 0.01을 기준으로 LDA를 실행하여 분석한다. 토픽모델링은 대량 문서에서 주제를 도출하는 방법론으로서 이 연구에서 리뷰 유형별 차이를 분석하는데 사용한다.

#### 5. 군집분석

군집분석은 개체들을 여러 가지 변수를 기준으로 다 차원 공간에서 유사 특성을 가진 개체끼리 묶는 방법이다. 통계적으로 개체의 유사성에 기반하여 군집을 찾고 자료를 요약하는 탐색적인 자료 분석 방법의 하나다. 개체들의 여러 가지 특성들을 유사성 거리로 환산하는 작업을 거쳐서 거리가 상대적으로 가까운 개체들을 동질적인 집단으로 군집화한다. 핵심어 군집화의 비지도

학습 방법론으로서 리뷰를 분석하는데 적합한 방법론이다. 군집분석은 비교 시 사용할 변수 선정의 문제, 유사성 거리를 측정하는 방법, 어떻게 동질적인 집단으로 군집화할지에 대한 방법을 설정해야 한다. 군집분석의 방법론 중에 K-means 군집분석은 각각의 점(point) 다른 중심보다 해당 군집의 중심에 가까운 경우 특정 군집에 속해 있는 것으로 간주한다. 이분화(Bisecting) K-means는 K-means 군집분석을 변형한 것으로 문서 처리에 있어서 좀 더 효율적인 것으로 알려져 있다. 차선책을 희생하면서 계산 요구를 줄이는 방법론이다 [51].

### IV. 결과

#### 1. 기술적 통계

카테고리별로 유형1은 8,311개, 유형2는 8,128개, 유형3은 8,459개로 비슷한 리뷰 개수를 설정하였다. 리뷰 개수는 유사한 크기로 설계하였지만 각 유형별 사용된 단어의 수는 차이를 보인다. 하드웨어 제품에 대한 리뷰 단어 수가 다른 유형에 비해 월등이 많으며 이것은 대부분 하드웨어 제품의 리뷰가 길게 작성된 것에 기인한다.

표 2. 상품유형별 분석리뷰 개요

구분	리뷰수	단어수	평균단어수
유형 1 (복수: Hardware)	8,311	1,128,358	136
유형 2 (복수: Service)	8,128	432,291	53
유형 3 (단일: Amazon Echo)	8,459	546,697	65
총 리뷰수	24,898	2,107,346	85

각 제품유형별로 TF-IDF 기준으로 키워드를 20위까지 선별하였다. [표 3]에서처럼 서비스(Service)에 특화된 단어로 service, job, professional, recommend, technician, arrive, hour, install, excellent 등이 도출되었고, Amazon Echo에서는 play, device, ask, home, question, listen이 차별적으로 도출되었다. amazon echo의 특성이 ask, home, question 등에 반영되었다.

표 3. 각 상품 유형별 키워드 빈도수(Top 20)

	HW분야	Freq.	Service분야	Freq.	Amazon Echo	Freq.
1	good	4340	service	4726	sound	4747
2	work	4163	job	3559	love	3435
3	buy	3656	time	3536	good	2972
4	great	3601	clean	3140	play	2856
5	unit	3243	work	2969	great	2571
6	sound	3156	great	2512	device	1937
7	time	3087	professional	2455	ask	1898
8	pressure	2985	recommend	1893	work	1739
9	love	2746	technician	1711	dot	1631
10	easy	2567	arrive	1597	buy	1611
11	air	2549	hour	1292	quality	1547
12	instant	2272	good	1209	home	1325
13	room	2269	install	1207	time	1249
14	need	2114	need	1184	question	1220
15	cool	2106	excellent	1155	listen	1124
16	day	1956	highly	1023	product	1117
17	purchase	1926	definitely	1019	need	1044
18	quality	1875	complete	951	know	999
19	product	1818	call	942	generation	982
20	cancel	1781	home	926	easy	956

2. 토픽모델링

토픽모델링은 T-lab plus에 내장된 기능을 사용하였다. LDA와 김스샘플링을 사용하여 분석되었다. 토픽의 수인 K의 수를 변화하면서 분석을 진행하였고 비교적 토픽이 중복되는 경우가 많아서 적은 수의 토픽으로 결정하였다(k=10). 토픽모델링 결과는 세 유형 모두 상품의 종류 또는 서비스의 특징별로 토픽이 도출되는 특징을 보였다. 하드웨어 제품의 경우 요리, 부엌의 팬, 헤드폰, 방의 온도, 조명, 세탁기 관련된 제품별로 토픽이 도출되었으며 제품의 종류와 관계없이 추천, 가족, 고객서비스에 관련된 토픽이 도출되었다.

서비스 상품의 경우 상품 종류 측면에서 카펫, 집, 설치와 관련된 토픽이 도출되었으며 공급자와 숙련공 고객, 시간, 서비스를 공급하는 사람의 우수성을 의미하는 토픽으로서 professional, technician이 도출되었다.

아마존 에코로 분석한 세 번째 유형은 아마존 에코의 상품이 보유하는 특성으로 app, 질문과 응답, 아침, 와이파이, 방, 오디오, 선물, 음악, 스피커 등의 토픽이 도출되었다.

표 4. 토픽모델링 결과

토픽	하드웨어 제품	서비스	하드웨어 서비스 융합 상품
1	cook	provider customer	app
2	fan	carpet	question and answer
3	headphone	time	morning
4	recommend	house	WIFI
5	family	installation	room
6	light	great	audio
7	buy-spend	professional	Google Apple
8	customer service	satisfy	gift
9	cool-hot	technician	music
10	washer	handyman	speaker

세 유형 상품의 토픽모델링은 개별 상품의 특성을 묘사하는 단어 중심으로 도출되었으며 중첩되는 측면이 강했다.

이 연구의 연구 질문에서 제기된 각각 제품유형별 리뷰의 차이점을 도출하기 위해 추가적으로 각각의 유형 리뷰에 사용된 차별화된 형용사와 동사를 토픽모델링을 구성하는 단어를 대상으로 추가적인 분석을 시행하였다. 세 유형의 상품별로 완전히 배타적으로 1개의 상품유형에서만 출현하는 단어와 두 분야에서 동시에 출현하는 단어, 세 분야에서 동시에 출현하는 단어를 정리하면 [표 5]와 같이 차별적 사용예를 보여 준다. 하드

웨어 제품에 특화된 형용사로 huge, portable, 동사로 charge, enjoy 등의 단어가 나타났다. 서비스 상품의 경우 동사 중에서는 expect, exceed, explain, perform, repair, estimate가 사람에 의한 서비스의 성격을 묘사하고 있으며, 형용사 중에는 사람에 의한 서비스의 능력치와 전문성, 태도를 보여주는 단어로써 competent, efficient, friendly, courteous, knowledgeable, neat, outstanding, polite, professional, super, timely, young이 특징적인 단어로 도출되었다. 아마존 에코에 해당하는 3유형에서는 사용하는 장면묘사를 중심으로 command, listen, respond와 같은 동사가 도출되었고, ICT융합분야의 새로운 서비스로서 difficult의 부정적 의미의 단어도 보이지만, favorite, fine,

fun, nice, smart, unlimited, useful등의 긍정적 의미의 형용사가 많이 도출되었다.

도출된 단어 중에서 서비스 상품에서 차별적으로 등장하는 professional에 대해서는 서비스 분야의 주목할 단어로서 특별히 맥락의 전개를 이해하기 위해 이 단어와 동시 출현하는 단어를 추가적으로 분석하였다 [그림 2]. professional에 선행하는 키워드는 time, friendly, polite이, 후행하는 단어로는 courteous와 friendly가 빈도수 측면에서 가까운 핵심어로 도출되었다. 선행과 후행에 따른 차이점은 크게 발견되지 않았으며 서비스 공급자의 행태와 관련된 긍정적 의미의 단어와 동시출현하고 있음을 보여 준다.

표 5. 토픽모델링에 사용된 각 상품 유형별 형용사, 동사 비교

구분	HW	Service	amazon echo	동사	형용사
Type1	√			allow, charge, cover, decide, end up, enjoy, forget, hang, hear, mean, order, pay, replace, spin, surprise, think, wait, watch, wish	at first, cold, comfortable, dry, good, great, hot, huge, instant, outside, portable, powerful, pretty, quiet, real, simple, slow
Type2		√		able to, arrive, assemble, check, clean up, confirm, estimate, exceed, expect, explain, hire, inform, meet, notify, perform, repair, review, satisfy, sure, thank	clean, competent, courteous, deep, dirty, early, efficient, entire, friendly, knowledgeable, neat, necessary, outstanding, personable, pleasant, polite, professional, punctual, reasonable, super, thorough, timely, young
Type3			√	change, command, continue, control, figure out, improve, list, listen, play, question, reminder, respond, talk, upgrade	bad, basic, clear, decent, difficult, favorite, fine, fun, large, multiple, nice, original, prime, short, smart, specific, trying, unlimited, useful
Type4	√	√		amaze, buy, install, recommend, return, stay	easy, extremely, perfect
Type5		√	√	answer, help, know, purchase, start, tell	excellent, fast, glad, happy, helpful, long, small, wonderful, worth
Type6	√		√	call, feel, recommend, support, update	cool, low, wrong
Type7	√	√	√	impress, receive, send, set, spend	awesome, definitely, fantastic, high, old, quick

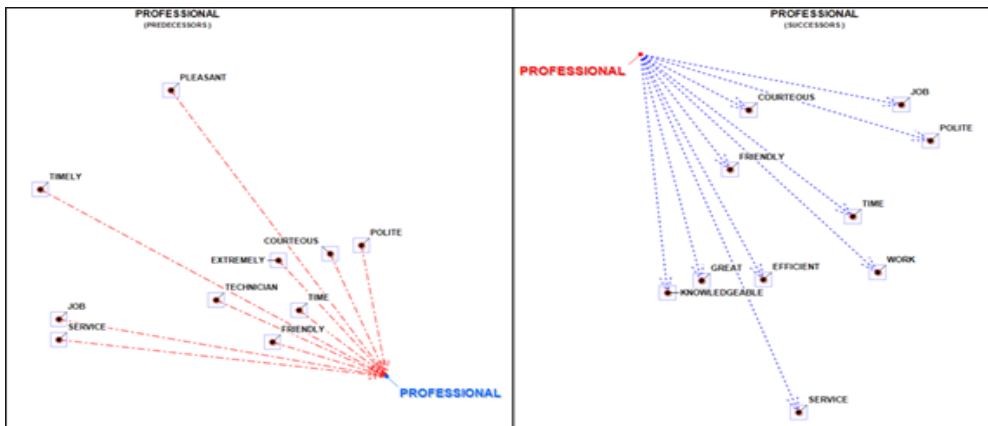


그림 2. Professional의 선행형 단어 분석



### 3. 군집분석

총 24,898의 리뷰를 동일한 수의 문서(documents)로 인식하여 비지도학습의 하나인 이분화(bisecting) K-means의 방법으로 군집화를 실시하였다. 동시 출현의 최소값은 2로 하였고 클러스터의 최대 상한은 10개를 넘지 않도록 모델링 하였다. 군집을 이루게 한 잠재요인(latent factor)은 2개로 도출되었으며 2개의 잠재요인에 의해 3개의 군집이 형성되었다. 하드웨어 상품은 군집1(53.97%)에 서비스 상품은 군집2(98.85%)에, ICT 제품서비스융합상품으로 간주한 amazon echo는 군집3(67.01%)에 대부분 포함되었다[표 6][그림 3]. Service분야의 리뷰가 군집2에 속한 비중이 HW가 군집1에, Amazon echo가 군집3에 속한 비중보다 월등하게 높다. 군집1은 전체의 31.58%를 차지하고 있으며 군집을 이루는 핵심어 역시 하드웨어 상품

중심의 cook, pot headphone 등의 제품과 이에 대한 특성인 sound, pressure, noise, quality 등의 핵심어로 구성되었다. 군집2는 전체의 35.18%를 차지하였으며 service와 관련된 리뷰 단어 중심으로 구성되었고, 서비스의 주제를 암시하는 clean, install 등의 단어와 서비스 수행의 행태와 관련된 professional, technician, arrived의 단어로 구성되었다. 군집3은 전체의 33.23%를 차지하였고 Amazon echo 서비스의 내용으로 추정되는 단어로서 music, weather등의 단어와 하드웨어 제품 중의 Vorando제품과 관련된 air, fan 등의 단어가 같이 등장하였다[표 7].

통계적 유의성에 의한 판단이 아닌 정성적인 해석이지만 연구문제로 제기한 각 상품 분야별로 차별적인 리뷰로 구성될 것이라는 가설은 클러스터링을 통해서 지지되고 있다.

표 6. 상품형태별 리뷰의 군집비중

	군집1	군집2	군집3	Total
HW	53.97%	16.96%	29.07%	100%
Service	0.57%	98.85%	0.58%	100%
Amazon Echo	24.21%	8.78%	67.01%	100%
Total	31.58%	35.18%	33.23%	100%

표 7. 군집별 핵심 키워드 비중

	Lemma	X <sup>2</sup>	Lemma	X <sup>2</sup>	Lemma	X <sup>2</sup>
	Cluster 1		Cluster 2		Cluster 3	
1	cook	6723	service	11759	ALEXA	4586
2	pot	6339	job	7098	play	3998
3	pressure	4679	professional	5560	music	3200
4	cooker	4400	clean	4812	room	3011
5	headphone	4042	technician	3931	air	2957
6	sound	3833	arrive	3332	cool	2638
7	instant	3193	recommend	2842	fan	2582
8	noise	22612	time	2127	Echo	1995
9	quality	2490	install	2081	weather	1414
10	cancel	2384	highly	2050	hose	3390

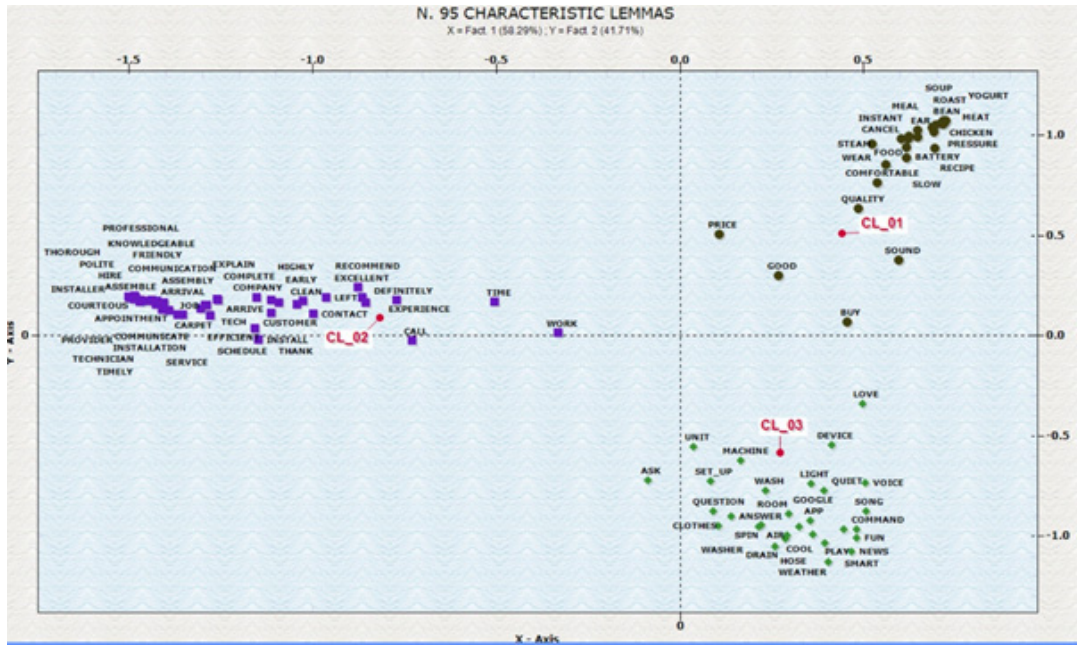


그림 3. 군집분석 결과(3군집)

## V. 논의와 결론

이 연구는 소비자 리뷰는 세 가지 상품유형별로 서로 다른 주제와 단어를 사용할 것이라는 연구 질문에서 시작되었다. 이와 같은 차이점은 텍스트로 작성된 리뷰를 상품을 기획하거나 개선할 때 또는 마케팅 방향성을 설정할 때 실무자들이 고려해야 할 필수적인 사항이라는 점이다. TF-IDF를 통해서 도출한 핵심어를 살펴보면 서비스 분야에서 하드웨어 제품 리뷰와 비교하여 차별적으로 사용된 단어들 나타났다. professional, technician, recommend 와 같은 단어들은 서비스 상품의 리뷰에서만 등장했다. 또한, 서비스 상품 분야의 리뷰에서 excellent, definitely 등과 같이 좀 더 강한 수식어가 사용되었다. 하드웨어 상품보다 서비스 분야에 대한 소비자의 만족도와 선호가 좀 더 극단적으로 아주 나쁘거나 아주 좋다고 평가되어 표현되고 있음을 보여준다.

토픽모델링을 통해서 도출한 토픽들은 제품과 서비스의 특성을 의미하는 주제들로 구성되었다. 서비스 상품의 리뷰에서 도출된 공급자, 숙련공, 고객, 시간, 서

스를 공급하는 사람의 우수성을 의미하는 토픽으로서 professional, technician와 같은 핵심어가 도출되었다. 이와 같은 토픽은 특별히 하드웨어 상품의 리뷰와 차별성을 보인다. 서비스 상품 리뷰에서 등장한 사람의 능력과 전문성과 태도를 보여주는 단어는 서비스 제공업체가 주목해야 할 핵심적인 내용이다.

토픽모델링을 구성하는 형용사와 동사의 사용현황에서도 상품기획자들이 주목해야 할 몇 가지 시사점이 존재한다. favorite, fine, fun, nice, smart, unlimited, useful 등의 긍정적 의미의 형용사는 특별히 아마존 에코 리뷰에서 등장하여 토픽을 구성한다. 특별히 기능적으로 unlimited, smart한 서비스와 기능과 동시에 fun한 제품의 속성은 Amazon echo의 후속 제품을 기획할 때 반드시 고려해야 할 핵심어들이다.

마지막으로 군집분석을 통해서 전체 리뷰가 3개의 집단으로 군집화되었으며 3개의 상품유형별 리뷰가 비교적 서로 다른 각각의 군집에 속하는 결과를 보였다. 특히 서비스 상품에 대한 리뷰는 그 댓글이 군집2에 대부분 속함으로써 다른 리뷰에 비교하여 구별되는 차별성을 가진다고 평가할 수 있다.

이 연구를 시작하게 된 상품유형별로 세 가지로 구분된 카테고리의 댓글은 차별적으로 구분된다고 할 수 있는가? 라는 연구 질문은 통계적인 가설 검정의 수준은 아니지만, (i) TF-IDF에 의한 핵심어, (ii) 토픽모델링의 주제, (iii) 토픽모델링에 배타적으로 사용된 동사와 형용사, (iv) 군집분석을 통해서 지지 된다고 할 수 있다.

이 연구의 결과는 경영학의 서비스 마케팅 분야에서 하드웨어 제품과 구별되는 서비스 사용자의 니즈 탐구에 차별적인 접근방법을 제공할 것이다. 또한, 비즈니스 모델을 연구하는 분야에서 융합제품서비스의 비즈니스 모델 설계와 분석을 위해 사용될 수 있을 것이다.

위와 같은 시사점에도 불구하고 이 연구는 몇 가지 한계를 가지고 있다. 첫 번째 연구 대상이 된 아마존 상품 리뷰가 전체 소비자 리뷰를 대표하지 못하는 한계와 세 가지 분류를 대표하는 상품 리뷰가 가지는 대표성의 한계를 가지고 있다는 점이다. 세 가지 유형의 리뷰를 선정함에 있어서 기존 선행연구와 참여 전문가의 의견에 바탕을 두고 있지만, 연구의 대상이 된 리뷰가 완전히 모집단을 대표하지는 못할 수 있다. 특히 세 번째 유형인 ICT 제품 서비스 융합 형태의 경우 이 연구에서는 아마존 에코를 대상으로 하였지만 이런 형태를 ICT분야의 제품서비스융합상품이라고 말하기에는 아직 연구자의 일치된 의견이 존재하지 않는다. 이 연구에서는 실무자들이 지적하는, 시사점이 가장 필요한 유형을 선정의 첫 번째 기준으로 채택하였고, ICT분야로 한정하여 하드웨어 서비스융합 상품을 선택하였을 뿐이다. 두 번째의 한계는 이 연구의 데이터 전처리의 문제에서 비롯된다. 이 연구에서는 이모티콘과 이와 유사한 의성어 의태어 등의 새로운 신조어는 제거하여 처리하였으나 이러한 단어들도 일정 부분 의미를 지니고 있다. 최근에는 이러한 기호에 해당하는 단어들의 의미를 밝히는 노력이 시도되고 있으나 우리의 연구에서는 이러한 부분은 적용되지 못하였다. 세 번째의 한계는 이 연구의 연구 질문을 다양한 텍스트 마이닝 방법론과 정성적 해석으로 탐구하려고 하였으나 가설검정 수준으로 통계적 유의성으로 증명하는 수준은 아니라는 점이다.

이와 같은 연구의 한계에도 불구하고 이 연구는 몇 가지 실무적 시사점을 가지고 있다. 시장에서 성공하는 상품은 상품기획 진행 시 소비자 니즈의 발굴과 니즈의

상품에의 반영이 이루어진 상품들이다. 소비자 니즈는 소비자가 말하는 것(saying)과 행동하는 것(doing)으로 부터 발굴하며 전자는 각종 인터뷰 기법으로 후자는 관찰(observation)을 통하여 진행한다. 전자의 경우 인터뷰의 산출물은 텍스트 스크립트(script)가 되며 이것을 정성적으로 분석하여 소비자의 니즈를 제품에 반영한다. 텍스트마이닝을 통한 소비자 니즈의 발굴은 정성 분석을 통해 소비자의 니즈를 발굴하던 기존의 연구에 정량화를 통한 정교한 발굴의 방법론을 제시한다. 또한, 웹이나 앱에서 소비자의 텍스트로 표현된 니즈는 그 양의 반대함으로 인하여 정성적 분석으로 접근하는 것이 불가능하다. 이 경우에는 필수적으로 텍스트의 정량화를 통한 소비자 니즈 탐구가 필요하다. 이 연구는 각 상품의 형태별로, 특별히 서비스가 융합된 상품을 포함하여 각 상품에 대한 소비자의 니즈가 어떤 텍스트로 표현되는지 분석함으로써 텍스트 기반의 제품 니즈 연구의 기초적인 자료를 제공할 수 있다.

이 연구는 다양한 제품의 리뷰가 존재하는 모든 텍스트로 분석대상을 확장할 수 있다. 하드웨어 제품과 단순 서비스뿐만 아니라 영화나 음악 등의 콘텐츠 분야, 여행과 레스토랑 등의 서비스 등으로 확장함으로써 카테고리별 해석의 준거 틀을 종합적으로 마련할 수 있다. 그리고 선행연구에서 사용하는 소비자 행동 분석 틀을 모든 제품과 서비스에 그대로 변형 없이 사용하는 것이 아니라 각각의 서비스의 형태에 따라 변형하여 사용하여야 함을 제시한다.

빅데이터 누적은 분석을 통해서 가치를 발휘한다. 데이터로서의 텍스트는 무엇을 대상으로 하고 있는지, 작성된 시점이 언제 인지에 따라서 다양한 해석이 가능하다. 이 연구에서는 하드웨어 제품, 서비스, ICT분야의 하드웨어와 서비스의 결합 상품의 형태에 따라서 소비자 리뷰의 모습을 정량적으로 분석하였다. 이를 통해서 빅데이터 분석 시에 어떤 기준을 가지고 평가하고 해석할 것인지에 대한 평가와 해석의 원형(prototype)을 제시하였다. 빅데이터의 대상이 되는 소비자 리뷰는 각종 리뷰 사이트, 블로그를 포함한 다양한 SNS(facebook, twitter, instagram 등)로부터 수집될 것이며 후속들은 다양한 대상별 (SNS의 종류)로 차이 나는 언어의 표현을 분석하여 정리한다면 매체별로 특화된 다양한 시

사점을 도출할 수 있을 것이다. 제품의 형태와 기간별 표현의 차이, 특별히 리뷰를 작성하는 소비자의 컨텍스트(context)가 반영된다면 좀 더 정확하고 정교한 소비자 니즈의 예측이 가능할 것이다. 최근의 소비자행동에 관한 연구는 소비전, 소비중, 소비후로 이어지는 시간적 흐름을 중시하고 있다. 그러나 연구자가 이러한 자료를 획득해서 연구 성과로 창출하기까지는 많은 어려움이 존재한다. 만약 리뷰를 작성한 시점과 어떠한 컨텍스트(context) 상황 인지가 표현된 데이터 확보가 가능하다면 연구의 정확성과 실효성을 확보하는 중요한 계기가 될 것이다.

#### 참 고 문 헌

- [1] Y. Chen and J. Xie, "Online consumer review: Word-of-mouth as a new element of marketing communication mix," *Management science*, Vol.54, No.3, pp.477-491, 2008.
- [2] Y. Chen, S. Fay, and Q. Wang, "The role of marketing in social media: How online consumer reviews evolve," *Journal of interactive marketing*, Vol.25, No.2, pp.85-94, 2011.
- [3] R. Filieri and F. McLeay, "E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews," *Journal of Travel Research*, Vol.53, No.1, pp.44-57, 2014.
- [4] E. I. Elmurngi and A. Gherbi, "Unfair reviews detection on Amazon reviews using sentiment analysis with supervised learning techniques," *Journal of Computer Science*, Vol.14, No.5, pp.714-726, 2018.
- [5] N. Jindal and B. Liu, Identifying comparative sentences in text documents. in *Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. August, pp.244-251, 2006.
- [6] R. Y. Lau, S. Y. Liao, R. C. Kwok, W. Xu, K. Y. Xia, and Y. Li, "Text mining and probabilistic language modeling for online review spam detection," *ACM Transactions on Management Information Systems*, Vol.2, No.4, pp.1-30, 2011.
- [7] C. L. Lai, K. Q. Xu, R. Y. Lau, Y. Li, and D. Song, High-order concept associations mining and inferential language modeling for online review spam detection. in *Proceedings - IEEE International Conference on Data Mining, ICDM*. 2010.
- [8] A. Nikolay, G. Anindya, and G. I. Panagiotis, "Deriving the pricing power of product features by mining consumer reviews," *Management Science*, Vol.57, No.8, pp.1485-1509, 2011.
- [9] N. Archak, A. Ghose, and P. G. Ipeirotis, Show me the money!: Deriving the pricing power of product features by mining consumer reviews, in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.59-65, 2007.
- [10] 류송희, 김용희, 김응모, "Amazon영화 리뷰를 활용한 감성분석 및 영화추천기법에 관한 연구," *한국정보과학회 학술발표논문집*, 2016년 6월, pp.1834-1836, 2016.
- [11] E. Sygkounas, G. Rizzo, and R. Troncy, "Sentiment polarity detection from Amazon reviews: An experimental study, In *Semantic Web Evaluation Challenge*," Springer, Cham. pp.108-120, 2016.
- [12] A. Bhatt, A. Patel, H. Chheda, and K. Gawande, "Amazon review classification and sentiment analysis," *International Journal of Computer Science and Information Technologies*, Vol.6, No.6, pp.5107-5110, 2015.
- [13] M. Lee, M. Jeong, and J. Lee, "Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website: A text mining approach," *International Journal of Contemporary Hospitality Management*, Vol.29, No.2, pp.762-783, 2017.
- [14] A. Ganzaroli, I. De Noni, and P. van Baalen,

- "Vicious advice: Analyzing the impact of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice," *Tourism Management*, Vol.61, pp.501-510, 2017.
- [15] R. G. Wyckham, P. T. Fitzroy, and G. D. Mandry, "Marketing of services an evaluation of the theory," *European Journal of Marketing*, Vol.9, No.1, pp.59-67, 1975.
- [16] S. L. Vargo and R. F. Lusch, "The four service marketing myths: remnants of a goods-based, manufacturing model," *Journal of service research*, Vol.6, No.4, pp.324-335, 2004.
- [17] P. J. Danaher, "Customer heterogeneity in service management," *Journal of Service Research*, Vol.1, No.2, pp.129-139, 1998.
- [18] M. Abernethy Avery, "Promoting customer contact people: a key difference in service and product advertising," *Journal of Services Marketing*, Vol.7, No.1, pp.4-12, 1993.
- [19] A. Hidalgo and L. D'Alvano, "Service innovation: Inward and outward related activities and cooperation mode," *Journal of Business Research*, Vol.67, No.5, pp.698-703, 2014.
- [20] E. J. Nijssen, B. Hillebrand, P. A. Vermeulen, and R. G. Kemp, "Exploring product and service innovation similarities and differences," *International Journal of Research in Marketing*, Vol.23, No.3, pp.241-251, 2006.
- [21] S. Goto, "Technology epiphany and an integrated product and service," *Journal of Technology Management and Innovation*, Vol.12, No.2, pp.34-44, 2017.
- [22] J. Z. Raja, D. Bourne, K. Goffin, M. Çakkol, and V. Martinez, "Achieving customer satisfaction through integrated products and services: An exploratory study," *Journal of Product Innovation Management*, Vol.30, No.6, pp.1128-1144, 2013.
- [23] F. H. Beuren, M. G. Gomes Ferreira, and P. A. Cauchick Miguel, "Product-service systems: A literature review on integrated products and services," *Journal of Cleaner Production*, Vol.47, pp.222-231, 2013.
- [24] M. J. Goedkoop, C. J. Van Halen, H. R. Te Riele, and P. J. Rommens, "Product service systems, ecological and economic basics," Report for Dutch Ministries of environment (VROM) and economic affairs (EZ), Vol.36, No.1, pp.1-122, 1999.
- [25] A. Annarelli, C. Battistella, and F. Nonino, "Product service system: A conceptual framework from a systematic review," *Journal of Cleaner Production*, Vol.139, pp.1011-1032, 2016.
- [26] M. J. Goedkoop, C. J. Van Halen, H. R. Te Riele, and P. J. Rommens, "State-of-the-art in product-service systems. Proceedings of the Institution of Mechanical Engineers," Part B: *Journal of Engineering Manufacture*, Vol.221, No.10, pp.1543-1552, 2007.
- [27] G. Vasantha, R. Roy, and J. Corney, "Challenges and opportunities in transforming laser system industry to deliver integrated product and service offers," In Working conference on Virtual Enterprises, 2014 October, pp.127-134, 2014.
- [28] Y. Suh and J. Jeon, "Do integrated products and services increase customer satisfaction? The case of the mobile industry in Korea," *Total Quality Management and Business Excellence*, Vol.27, No.11, pp.1261-1276, 2016.
- [29] A. Lelah and D. Brissaud, "Clarifying frontiers of PSS and information and communication technologies, in The Philosopher's Stone for Sustainability," Springer, pp.441-446, 2013.
- [30] S. O. Orimaye, S. M. Alhashmi, and E. G. Siew, "Buy it - Don't buy it: Sentiment classification on amazon reviews using sentence polarity shift," In Pacific Rim International Conference on Artificial Intelligence, 2012 September, pp.386-399, 2012.
- [31] A. Shafae, H. Issa, S. Agne, S. Baumann, and A. Dengel, "Aspect-based sentiment analysis of amazon reviews for fitness tracking devices,"

- In Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2014 May, pp.50-61, 2014.
- [32] J. Zhang, "What's yours is mine: exploring customer voice on Airbnb using text-mining approaches," *Journal of Consumer Marketing*, Vol.36, No.5, pp.655-665, 2019.
- [33] S. Okazaki, L. Andreu, and S. Campo, "Knowledge Sharing Among Tourists via Social Media: A Comparison Between Facebook and TripAdvisor," *International Journal of Tourism Research*, Vol.19, No.1, pp.107-119, 2017.
- [34] L. Garay Tamajón and G. Cánoves Valiente, "Barcelona seen through the eyes of TripAdvisor: actors, typologies and components of destination image in social media platforms," *Current Issues in Tourism*, Vol.20, No.1, pp.33-37, 2017.
- [35] L. Yao, "Mining helpfulness products reviews for C2C E-commerce," *International Journal of Advancements in Computing Technology*, Vol.4, No.14, pp.123-132, 2012.
- [36] Kim and K. Yun, "Sentiment Analysis on Monthly House Rent on Twitter," *Journal of the Korean Cadastre Information Association*, Vol.19, No.1, pp.37-55, 2017.
- [37] M. D. Sotiriadis and C. van Zyl, "Electronic word-of-mouth and online reviews in tourism services: The use of twitter by tourists," *Electronic Commerce Research*, Vol.13, No.1, pp.103-124, 2013.
- [38] V. P. Magnini, J. C. Crofts, and A. Zehrer, "Understanding Customer Delight," *Journal of Travel Research*, Vol.50, No.5, pp.535-545, 2011.
- [39] Q. Gan, B. H. Ferns, Y. Yu, and L. Jin, "A Text Mining and Multidimensional Sentiment Analysis of Online Restaurant Reviews," *Journal of Quality Assurance in Hospitality and Tourism*, Vol.18, No.4, pp.465-492, 2017.
- [40] Q. Gan and Y. Yu, "Restaurant rating: Industrial standard and word-of-mouth - A text mining and multi-dimensional sentiment analysis," In 2015 48th Hawaii International Conference on System Sciences, pp.1332-1340, 2015.
- [41] C. P. Wei, Y. M. Chen, C. S. Yang, and C. C. Yang, "Understanding what concerns consumers: A semantic approach to product feature extraction from consumer reviews," *Information Systems and e-Business Management*, Vol.8, No.2, pp.149-167, 2010.
- [42] A. Ghose and P. G. Ipeirotis, "Designing novel review ranking systems: Predicting the usefulness and impact of reviews," In Proceedings of the ninth international conference on Electronic commerce, pp.303-310, 2007.
- [43] W. Duan, Q. Cao, and Q. Gan, "Investigating determinants of voting for the 'helpfulness' of online consumer reviews: A text mining approach," in 16th Americas Conference on Information Systems, AMCIS, p.497, 2010.
- [44] J. Li, G. S. Wu, J. Y. Qi, and Y. Su, "A product review score prediction model based on neural network. in Information, Computer and Application Engineering - Proceedings of the International Conference on Information Technology and Computer Application Engineering, ITCAE 2014. 2015.
- [45] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, Vol.24, No.5, pp.513-523, 1988.
- [46] X. Yu, M. Li, K. A. Kim, L. Chung, and K. H. Ryu, "Emerging Pattern-Based Clustering of Web Users Utilizing a Simple Page-Linked Graph," *Sustainability*, Vol.8, No.3, p.239, 2016.
- [47] H. M. Wallach, "Topic modeling: beyond bag-of-words. in Proceedings of the 23rd international conference on Machine learning, 2006. ACM.
- [48] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, Vol.3, No.1, pp.993-1022,

2003.

- [49] M. Hoffman, F. R. Bach, and D. M. Blei, "Online learning for latent dirichlet allocation," in advances in neural information processing systems, 2010.
- [50] I. Porteous, et al. "Fast collapsed gibbs sampling for latent dirichlet allocation," in Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.567-577, 2008. ACM.
- [51] S. M. Savaresi and D. L. Boley, "On the performance of bisecting K-means and PDDP," in Proceedings of the 2001 SIAM International Conference on Data Mining, pp.1-14, 2001.

#### 저 자 소 개

김 성 범(Sungbum Kim)

정회원



- 2013년 2월 : KAIST 기술경영(공학박사)
- 2013년 9월 ~ 현재 : 금오공대 IT 융합학과 부교수

〈관심분야〉 : ICT 전략, 데이터사이언스, 기업가정신, 신제품/신사업 개발