

LDA 토픽 모델링을 활용한 스마트폰 카메라 관련 뉴스기사 분석

박 천 호¹ · 김 태 경¹ · 최 정 일^{2*}¹송실대학교 대학원 IT정책경영학과 박사과정^{2*}송실대학교 경영학부 교수

Analysis of News Articles on Smartphone Camera Using LDA Topic Modeling

Cheon-Ho Park¹ · Tae-Gyong Kim¹ · Jeongil Choi^{2*}¹Doctoral Student, Graduate School of IT Policy and Management, Soongsil University, Seoul 06978, Korea^{2*}Professor, School of Business Administration, Soongsil University, Seoul 06978, Korea

[요 약]

본 연구는 스마트폰 카메라 관련 뉴스 빅데이터를 활용하여 스마트폰 카메라의 주요 키워드와 토픽을 도출하고 트렌드를 확인하여, 분석 결과를 기반으로 스마트폰 카메라에 대한 시사점을 제안하고자 수행되었다. 33,554건의 뉴스 기사를 분석한 결과, 스마트폰 카메라 기반 서비스 및 응용, 스마트폰 카메라 시장 및 부품 산업, 스마트폰 제조사 및 경쟁 제품 비교, 스마트폰 카메라 기능 및 디자인 등 4개의 토픽으로 분류되었다. 각 토픽의 트렌드로부터 스마트폰 카메라 기능에 대한 경쟁이 가속화되고 있으며, 카메라 기능 향상을 위한 주변 부품 산업 성장 및 기술 투자가 증가하고 있고, 소비자들의 기대치가 높아지고 있음을 확인하였다. 본 연구 결과를 바탕으로 미래 스마트폰 카메라 산업에 대한 변화 기술 발전, 소비자들의 관심 및 기대를 예측하고 대응할 수 있는 학술적 기반을 제공한 것에 연구의 의의를 갖는다.

[Abstract]

This study derived major keywords and topics of news big data on smartphone cameras, checked trends of topics, and suggested implications for smartphone cameras based on the results. After analyzing 33,554 news articles, they were classified into four topics: smartphone camera application and service, smartphone camera market and parts industry, smartphone manufacturers and competitive products comparison, and smartphone camera function and design. From the trend of each topic, competition for smartphone camera functions is accelerating, growth and technology investment in the peripheral parts industry to improve camera functions are increasing, and consumer expectations are increasing. The study is meaningful in that it provides an academic foundation to predict and respond to the development of change technologies and consumers' interest and expectations for future smartphone cameras.

색인어 : 스마트폰 카메라, 뉴스기사, 키워드, 토픽 모델링, 트렌드**Keyword** : Smartphone Camera, News Articles, Keyword, Topic Modeling, Trend<http://dx.doi.org/10.9728/dcs.2023.24.8.1753>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 02 July 2023; Revised 20 July 2023

Accepted 24 July 2023

***Corresponding Author; Jeongil Choi**

Tel: +82-2-820-0561

E-mail: jichoi@ssu.ac.kr

1. 서론

2007년 애플이 세계 최초로 출시한 스마트폰, 아이폰(iPhone)은 인터넷 이후에 또다시 인류의 삶에 혁명적인 변화를 가져왔으며, 현대인의 필수품이 되어 사회·문화·정치·경제적인 측면에서 매우 큰 영향을 주고 있다.

스마트폰 카메라의 발전은 전문가용 카메라의 수요와 시장에 큰 영향을 미치고 있다. 일반적으로 전문가용 카메라는 DSLR(Digital Single Lens Reflex) 카메라와 미러리스(Mirrorless) 카메라로 구분된다. 이러한 카메라는 고품질의 이미지 센서와 교환 가능한 렌즈를 통해 다양한 촬영 상황에 맞게 정밀하게 조절할 수 있는 장점이 있다. 하지만 비싼 가격, 큰 외형과 무게, 복잡하고 어려운 조작 방법 등의 단점이 있다.

반면 스마트폰 카메라는 저렴하고 휴대하기 편하며 간단하게 조작할 수 있는 장점이 있다. 또한 스마트폰 카메라는 인공지능, 컴퓨터 비전, 딥러닝 등의 기술을 활용하여 이미지 센서와 렌즈의 한계를 극복하고 있다.

스마트폰으로 사진을 찍는 것은 단순한 기록이나 기억보다는 자신의 정체성과 가치관을 표현하는 수단이 되고 있다. 사람들은 소셜 미디어나 메신저 등을 통해 자신이 찍거나 받은 사진을 공유하고 소통하면서 인정받기를 원한다. 이러한 과정에서 사진은 개인적인 감정이나 의미보다는 타인의 반응이나 평가에 영향을 받기도 한다.

스마트폰 카메라의 기술적 진화는 사진의 가치와 의미에 변화를 가져오고 있다. 스마트폰 카메라로 쉽고 빠르게 찍고 공유할 수 있는 사진은 기록적이고 소비적인 성격을 강화한다. 이러한 사진은 순간적인 감정이나 정보를 전달하는 용도로 사용되며, 장기적인 가치나 의미보다는 당시의 반응이나 평가에 중점을 두게 된다. 따라서 스마트폰 카메라로 찍은 사진은 대부분 잊혀지거나 삭제되기 쉽다.

스마트폰 카메라는 소비자들이 현실과 가상의 경계를 넘나드는 새로운 경험을 제공하여, AR/VR 등의 기술과 결합하여 소비자들이 현실과 다른 공간이나 시간, 상황에 들어가거나 현실에 가상의 요소를 추가하는 경험을 제공한다. 또한 스마트폰 카메라는 메타버스와 같은 디지털 세계와 연결되어 소비자들이 자신의 아바타를 만들고 다른 사람들과 만나거나 각종 활동을 하는 경험을 제공하기도 한다.

현대인의 삶에 큰 영향을 미치고 있는 스마트폰 카메라는 기술 개발과 응용 방법에 대하여 주로 연구되고 있으며, 한편으로 주사용자 경험, 마케팅 전략 수립, 특히 전략 수립 등과 같은 연구가 시도되고 있으나, 뉴스기사를 분석하여 스마트폰 카메라 시장 및 기술동향, 경쟁상황, 기능 비교, 토픽 트렌드 등에 대한 연구는 거의 없는 실정이다.

본 연구에서는 뉴스기사 빅데이터를 활용하여, 스마트폰 카메라와 관련하여 어떤 내용의 뉴스기사들이 주로 언급되고 있는지 확인하고, 어떤 이슈가 나오고 있는지 살펴보고자 한다. 뉴스기사 빅데이터 텍스트 분석을 위하여 토픽 모델링 방

법 중에서 최근에 주로 사용되고 있는 LDA (Latent Dirichlet Allocation) 알고리즘을 사용하여, 뉴스기사 빅데이터에 나오는 스마트폰 카메라의 주요 키워드(Keyword) 및 토픽(Topic)을 도출하고 트렌드(Trend)를 확인하여, 분석결과 데이터를 기반으로 스마트폰 카메라에 대한 시사점을 제안하고자 한다.

본 연구를 위한 연구문제는 다음과 같다. 첫째, 스마트폰 카메라 뉴스기사 빅데이터에 나타난 주요 키워드는 무엇인가에 대해 연구하고자 한다. 둘째, 스마트폰 카메라 뉴스기사 빅데이터에 나타난 주요 토픽과 트렌드는 무엇인가에 대해 연구하고자 한다.

II. 이론적 배경

2-1 텍스트 마이닝

텍스트 마이닝(Text Mining)은 패턴과 트렌드를 인식하여 텍스트에서 정보를 추출하는 기술이다. 텍스트 마이닝 또는 텍스트 분석이라는 용어는 어휘 자원, 태깅 또는 주석과 연관성, 시각화 및 예측과 같은 기술을 통해 정보를 검색하는 프로세스를 말한다. 텍스트 마이닝은 자연어 처리(NLP; Natural Language Processing) 기법을 이용해 텍스트를 정형화된 데이터로 변환하고 머신러닝 기법을 적용해 관심이 있는 어떤 사건을 예측하고자 하는 방법론이라고 말할 수 있다[1].

1960년대에 기본적인 자연어 처리를 성공적으로 개발한 후에 차원 축소, 잠재 요인 식별 및 데이터베이스 텍스트 처리와 같은 다양한 기술의 채택은 새로운 정보 검색 시대의 번영에 기여했다. 또한, LSA(Latent Semantic Analysis) 및 LDA 같은 토픽 모델링(Topic Modeling)과 기계 학습 알고리즘은 1990년대 이후에 더 실질적인 기반을 제공한 것으로 보이며 저널, 소셜 네트워크 서비스 및 온라인 고객 리뷰 분석에 대한 성공적인 시사점은 이메일 필터링, 제품 제안, 사기 탐지, 검색 엔진 및 파산 예측과 함께 모든 측면에서 그 중요성을 높였다[2]-[5].

온라인에서 생성되는 텍스트, 이미지 등 비정형 데이터가 기하급수적으로 증가하고 있고 그중에서 텍스트 데이터 분석은 텍스트 내에 숨어있는 의미를 파악할 수 있으며, 이를 현상학적으로 접근하거나 분석할 수 있다는 측면에서 학문적·실무적으로 관심이 높아졌다[6].

2-2 토픽 모델링

토픽 모델링은 텍스트 마이닝 기법 중에서 가장 많이 활용되는 기법 중 하나로, 다양한 문서 집합에 내재한 토픽을 파악할 때에 쓰는 방법이며, 문서 분류가 텍스트의 내용을 파악

해서 무엇인가를 예측하는 것에 목적을 두었다면, 토픽 모델링은 예측보다는 내용의 분석 자체를 목적으로 하는 기법이라고 할 수 있다[1].

텍스트 데이터를 분석 대상으로 한 토픽 모델링은 구조화되지 않은 대량의 문헌에서 토픽을 찾아내기 위한 것으로 맥락과 관련된 단서의 벡터값을 활용하여 유사한 의미를 지닌 단어들을 그룹화하는 방식으로 토픽을 분석하거나 추론하는 방법이다[7]-[8].

토픽 모델링은 방대한 양의 문서 집합에서 주요 토픽을 추출하고 각 토픽에 대응되는 문서를 식별하여 제공한다. 하나의 문서가 하나의 토픽으로만 할당되는 군집화(Clustering) 기법과 달리 토픽 모델링은 하나의 문서가 여러 토픽에 동시에 대응될 수 있기 때문에 현실 세계의 모델링에 보다 적합한 기법으로 평가받고 있다[9].

2-3 LDA

LDA는 추출된 토픽의 단어 비중과 문서의 토픽 비중이라는 두 변수가 모두 양의 실수를 요소로 가지며, 모든 요소를 더한 값이 1이 되는 Dirichlet 분포를 따르는 알고리즘이다[10]. 특히 하나의 문서가 여러 토픽과 동시에 대응되는 알고리즘이기 때문에 현실의 현상에 대한 모델링에 적합한 기법으로 평가받고 있다[11].

그림 1은 LDA 모형의 구조를 나타낸다. 토픽 모델링은 전체 문서에 공통적으로 내재한 토픽들을 식별한다. 즉, 각 문서가 개별적으로 전혀 다른 토픽들로 구성되는 것이 아니고 전체 말뭉치를 관통하는 토픽들이 있으며, 문서는 이러한 공통적인 토픽들이 다양한 비중으로 결합된 것으로, 문서에 따른 토픽의 확률분포를 추정하는 것이 토픽 모델링의 첫째 목적이 된다[1]. 그림 1에서 D 는 말뭉치, N 은 문서, 그리고 z 는 토픽으로 단어는 토픽을 반영하며 토픽은 문서에 잠재되어 있다. θ 는 해당 문서의 토픽 분포를 말하고 Dirichlet 분포를 따르며, Dirichlet 분포의 매개변수인 α 에 의해 결정된다.

그런데, 적절한 토픽 수를 정하는 것은 토픽 모델링 결과의 타당성을 확보하는데 있어 매우 중요하다[12]. 토픽 수를 결정하는 것에 정답이 있는 것은 아니지만, 토픽 수가 적으면 포괄적인 이슈가 도출되고 토픽 수가 많으면 너무 세부적인 이슈가 도출되기 때문에 연구자가 최적의 토픽 수를 결정해야 한다. 토픽 수를 결정하기 위한 기준으로 주로 혼잡도(Perplexity)를 측정하고 있으며, 혼잡도가 낮으면 토픽 수가 안정적으로 구성되어 무질서도가 낮은 상태를 의미한다[12]. 즉, 가장 최적의 토픽 수는 가능하면 혼잡도가 낮은 경우의 토픽 수로 결정한다[13].

또한, 응집도(Coherence)는 자료 안에서 단어 간 연결의 정도를 평가하는 데 사용되는 지표로서, 주어진 단어 W 에 대한 응집도를 계산한다[14]-[15]. 응집도가 높다는 것은 구성된 단어 간의 의미적 연결이 잘 되었음을 의미한다[12].

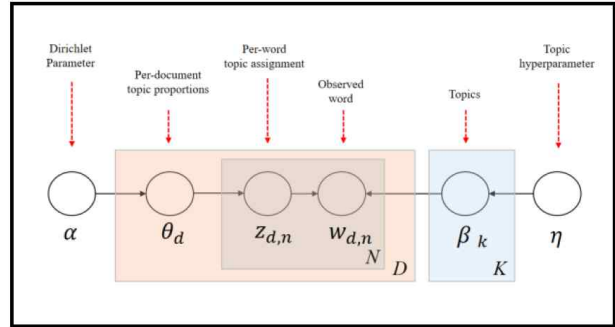


그림 1. LDA 모형 구조

Fig. 1. Principles of LDA

2-4 선행 연구

차윤정 등의 연구에서는 소셜 미디어 토픽 모델링으로 스마트폰 마케팅 전략 수립 지원 연구를 통하여, iPhone 6 & 6+ 및 Galaxy S5 & Note4 스마트폰에 대한 트위터 데이터를 활용하여 토픽 모델링 분석을 진행하였으며, 제품에 대한 온라인상의 사용자 의견을 실증적으로 분석하여 기업이 사용할 수 있는 바람직한 마케팅 전략 기법과 R&D 방향을 제시하였지만, 두 스마트폰 제조사의 대표적인 모델에 대하여 제한된 기간의 트위터 텍스트를 분석하였기 때문에 시간의 흐름에 따른 토픽 주제의 트렌드를 반영하지 못한 한계가 있다[16].

이연지와 설순옥의 연구에서는 스마트폰 카메라 분야 중심으로 설문조사에 기반한 특허 전략 수립에 대한 연구를 통해, 출원 수 중심의 정량적 분석 방법에 고객의 요구를 파악하는 기법인 설문조사 결과를 연계하는 방안을 제시하여 더욱 적극적인 특허 확보 전략을 수립할 수 있음을 보여주었으며, 특허 전략을 제시한 이후에 수년간 수행된 제품 개발 결과를 확인하여 제안된 전략의 유효성을 확인하였다[17].

김용철과 김승인의 연구에서는 스마트폰 카메라 애플리케이션에 대한 사용자 경험 연구를 통하여, 조작·편집·공유 3가지 측면에서 총 10가지의 태스크를 선정하여 20대 남녀 각각 4명을 대상으로 태스크 수행 시간을 측정하여 인터페이스별로 사용성의 차이가 있다는 것을 확인하였고, 피실험자 모두 다양한 기능과 조작의 편의성을 장점으로 꼽았음을 확인하였지만, 실험 집단의 수가 적어 실험 결과가 객관성을 띠고 하기에 어려움이 있을 수 있다[18].

강민정 등의 연구에서는 Z세대의 스마트폰 카메라 사용 행태 연구에서 3개의 그룹으로 분류하였는데, A그룹은 대학생 여성과 중고등학생으로 카메라 사용 행태에 있어 추억 기록과 편리성이 높게 나타났으며, B그룹은 주로 예체능 전공의 여성으로 편리성, 전문가적 마인드, 소통, 자연스러움, 타인 의식의 특성이 고루 나타났고, C그룹은 주로 사회초년생 이공계 남성으로 편리성, 자연스러움, 소통의 수치가 높음을 확인하였다[19].

김형지 등의 연구에서는 토픽 모델링의 LDA 기법을 적용

해 미디어에 보도된 스마트폰 중독 관련 뉴스기사의 주요 토픽을 분석하였으며, 가장 많이 언급된 단어는 스마트폰 중독 이외에 사용, 유아동, 게임, 부모 등으로 나타났으며, 토픽 모델링 분석을 통해 스마트폰 중독의 주요 토픽을 총 12개로 추출하였고, 이들 12개 토픽 중에서 지난 8년 동안 가장 많은 비중을 차지한 토픽은 중독 조사이었고, 중독 예방과 중독 관리 토픽도 상대적으로 높은 비중을 구성되었고, 중독 억제를 제외한 11개 토픽은 시기별로 구성 비율이 유의미한 차이를 보였다[20].

이혜선 등의 연구에서는 어린이 타자화와 출판 편향에 주목하여 어린이 스마트폰 이용에 관한 언론보도의 재현 양상을 논의하였으며, 어린이의 스마트폰 이용에 관한 언론보도에서 가장 많이 출현한 단어는 교육, 게임, 인터넷, 중독, 부모였으며 연결 중심성 지수가 가장 높은 단어는 중독이었고, 2011년~2016년에는 중독이라는 단어가 가장 많이 출현했으나, 2017년~2021년에는 교육이라는 단어가 가장 많이 출현하여, 어린이 스마트폰 이용에 관한 언론보도 의제가 중독에서 교육으로 변화하고 있음을 확인할 수 있었다[21].

III. 연구 방법

3-1 자료 수집

본 연구의 분석에서 활용된 신문기사는 한국언론진흥재단의 뉴스기사 빅데이터 시스템 ‘빅카인즈(BIGKinds)’에서 수집되었다. 빅카인즈는 1990년부터 오늘날까지 54개의 뉴스 미디어에 수록된 디지털 뉴스 콘텐츠 및 이에 대한 빅데이터 분석 결과를 제공하고 있는데, 저작권 문제로 인하여 기사 전문은 제공되지 않으며 기사에서 불용어가 제거되고 명사형태로 추출된 키워드와 형태소 분석 결과 가중치 상위 50개 키워드 등이 제공된다[22]. 본 연구에서는 불용어가 제거되고 명사형태로 추출된 키워드를 분석에 활용하였다.

스마트폰이 시장에서 큰 관심을 불러일으키고 본격적으로 판매가 되는 2010년부터 현재 분석 시점인 2023년 5월9일까지 ‘스마트폰 카메라’로 검색하였으며, 뉴스기사 내용 중복을 피하고 검색결과 데이터가 너무 방대해지는 것을 막기 위하여 조선일보, 중앙일보, 동아일보, 매일경제, 한국경제, 디지털타임스, 전자신문 언론사를 대상으로 경제·사회·문화·IT_과학 분야의 뉴스기사 총 33,554건의 데이터가 분석에 사용되었다.

그림 2와 같이 ‘스마트폰 카메라’ 관련 뉴스기사는 2010년 1,900건, 2016년 2,995건, 2022년 1,569건의 뉴스기사가 보도되었으며, 2021년부터 뉴스기사 건수가 줄어드는 현상을 보이고 있는데, 2023년 6월 1일에 코로나19 위기경보가 심각에서 경계로 하향되어 병·의원/약국의 마스크 의무착용도 해제되면서 다시 일상의 삶으로 돌아오게 되고 국내외 여

행이 자유롭게 되어 ‘스마트폰 카메라’ 관련 뉴스기사는 다시 증가할 것으로 예상된다.

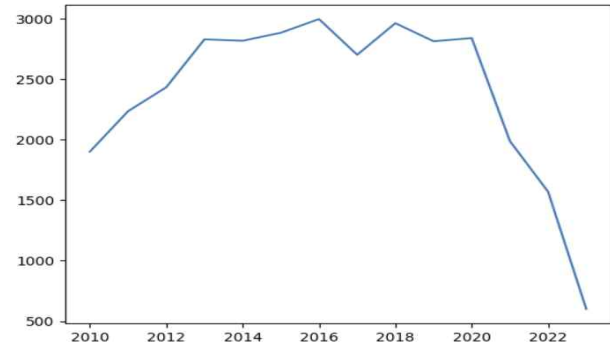


그림 2. 스마트폰 카메라 관련 뉴스기사 보도 추이
Fig. 2. News articles trend on smartphone camera

3-2 분석 방법

본 연구의 목적은 스마트폰 카메라 관련 뉴스기사 빅데이터에 나타난 주요 키워드를 도출하고 토픽을 분석하여 트렌드를 파악하는 것이므로, 대량의 텍스트에서 키워드를 추출하고 주요 트렌드를 도출하는 데에 유용한 토픽 모델링 분석 방법을 활용하였다. 토픽 모델링 분석 방법은 비정형 데이터를 대상으로 머신러닝 알고리즘을 활용하여 확률분포를 바탕으로 키워드와 토픽을 자동으로 추출하는 방법으로서, 대량의 텍스트 안에 잠재되어 있는 핵심 키워드와 주요 트렌드를 파악하는 데에 유용하다[23].

데이터 분석을 위하여 구글 코랩 환경을 활용하였으며 KoNLpy, LDA, pyLDAvis, Gensim 패키지를 사용했다. 본 연구에서는 적절한 토픽 개수를 확인하기 위해, 토픽 개수를 2부터 15까지 설정하여 토픽 모델링 분석을 수행하여 토픽별로 키워드가 중복되는 정도와 분류의 적절성을 비교하였다. 분석 결과로서 키워드의 중복 정도가 낮고, 토픽별로 키워드가 의미 있게 분류되는 최적의 토픽 개수는 4개로 나타났다.

본 연구의 분석 과정은 4 단계로 구분될 수 있는데 첫째는 분석에 필요한 데이터 수집 과정, 둘째는 수집된 텍스트에 대한 전처리 과정, 셋째는 키워드 도출 및 토픽 분석 과정, 넷째는 추출된 토픽의 시계열적 변화 분석이다.

데이터 수집 과정은 수집 대상 데이터 목록을 선정하고 수집을 위한 세부적 수집계획을 작성하여 적합한 데이터를 불러오는 것이다. 본 연구에서는 빅카인즈에서 ‘스마트폰 카메라’로 검색하여 관련 뉴스기사의 키워드를 분석 대상으로 하였다.

데이터 전처리 과정은 일반적인 양적연구에서 변수 리코딩 및 데이터 클리닝 과정에 비유될 수 있는데 한마디로 비정형화된 텍스트 자료를 텍스트 마이닝에 적합한 정형화된 자료로 변형하는 것을 의미한다[24]. 본 연구에서는 정규표현식을 사용하여 부호, 공란 등은 제거하고 한글, 영문, 숫자 데이

터만 활용하였으며 한 글자로 만들어진 단어와 ‘스마트폰’ 및 ‘카메라’는 제외하고 분석을 수행하였다.

키워드 도출 및 토픽 분석 과정은 전처리 과정을 통하여 얻은 텍스트 데이터를 기반으로 빈번하게 사용되고 있는 키워드를 찾아내어 관련성이 높은 단어들을 군집화하는 작업이다. 본 연구에서는 LDA 모형을 사용하여 키워드를 도출하고 토픽 응집도가 높고 토픽별 구분이 확실하게 되는 토픽 개수를 결정하여 토픽명을 부여하여 분석을 수행하였다.

토픽의 시계열적 변화 분석은 각 문서를 특정 토픽에 할당 한 후에 토픽의 시계열적 트렌드를 확인해보는 것이다. 본 연구에서는 해당 연도의 뉴스기사에 대하여 각 토픽에 해당하는 뉴스기사의 확률분포를 계산하여 연도별로 어떻게 변하는지를 확인하였다.

IV. 연구 결과

4-1 키워드 빈도분석

위의 자료 수집의 조건에서 ‘스마트폰 카메라’로 검색을 통하여 명사형태로 추출된 키워드는 총 48,463개이지만, 검색어 ‘스마트폰’과 ‘카메라’를 제외하고 분석에 불필요한 단어를 제거한 총 48,407개의 명사로 키워드 분석 및 토픽 모델링을 수행하였다. 분석 결과, 출현빈도가 높은 상위 10개의 단어는 삼성전자(54,488번), 시장(53,799번), 제품(53,213번), 출시(47,968번), 갤럭시(35,083번), 서비스(30,726번), 애플(27,227번), 사진(26,833번), 화면(25,043번), 촬영(23,145번)이다. 텍스트 분석에서 가장 단순하고 효과적인 방법은 여러 개의 문서에서 가장 많이 사용된 단어를 파악하는 것이며, 이것만으로도 많은 정보를 얻을 수 있다. 키워드의 빈도분석 결과를 한눈에 확인해볼 수 있는 워드 클라우드(Word Cloud)는 그림 3과 같다.



그림 3. 워드 클라우드 분석 결과
Fig. 3. Word cloud results

4-2 잠재토픽 추출

적절한 잠재토픽의 개수를 결정하기 위해 혼잡도와 응집도 값이 토픽 개수에 따라서 어떻게 변화하는지 분석하였다. 혼잡도는 특정한 확률 모델이 실제로 관측되는 값을 얼마나 잘 예측하는지 평가하는 지표로서 혼잡도 값이 작을수록 토픽 모델이 잘 학습되었다고 볼 수 있다. 응집도는 토픽이 얼마나 의미적으로 일관성이 높은지 평가하는 지표로서 응집도 값이 높을수록 토픽 모델이 잘 학습되었다고 볼 수 있다. 따라서, 혼잡도와 응집도 값을 참고하여 토픽의 개수를 정해야 하는데, 그림 4의 혼잡도 값 추이와 그림 5의 응집도 값 추이를 참고하여 4개의 토픽으로 분류하여 결과를 분석하였다.

파이썬에서 제공하는 LDA 시각화 라이브러리 pyLDAvis 2.1.2 버전을 사용하여 4개의 토픽에 대한 2차원 위치를 확인하였으며, 그림 6에서와 같이 각각의 토픽은 사시분면에 각각 1개의 토픽이 위치하고 있는 것으로 보아, 토픽 분류는 잘 된 것으로 보인다.

첫 번째 토픽은 서비스, 정보, 활용, 사람, 로봇, 영상, 스마트 등의 핵심어를 포함하고 있는데, 이는 ‘스마트폰 카메라 기반 서비스 및 응용’ 주제를 표현하고 있음을 알 수 있다. 스마트폰 카메라가 발전함에 따라 그 기술과 성능을 활용하는 다양한 서비스 및 응용 기술들이 등장하고 있으며, 이러한 기술들은 스마트폰 사용자들에게 사람들의 인식 능력을 향상시키

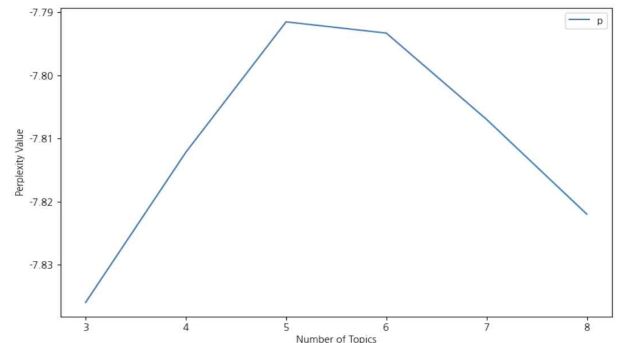


그림 4. 토픽 개수에 따른 혼잡도 분석 결과
Fig. 4. Perplexity values according to number of topics

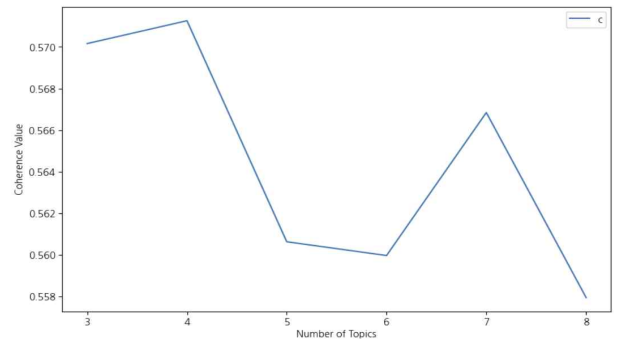


그림 5. 토픽 개수에 따른 응집도 분석 결과
Fig. 5. Coherence values according to number of topics

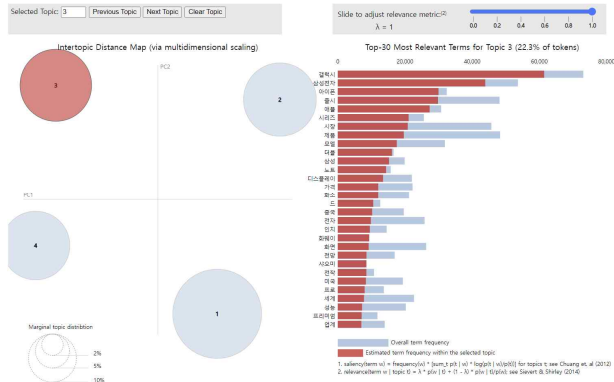


그림 6. LDA 토픽 모델링 시각화 결과
 Fig. 6. Visualization of LDA topic modeling

는 기능, 데이터 이용, 분석 등의 다양한 영역에서 혁신을 가져올 것으로 기대된다.

두 번째 토픽의 핵심어는 시장, 사업, 반도체, 부품, 투자, 매출, 주행 등이며, 이는 ‘스마트폰 카메라 시장 및 부품 산업’ 성격을 가진 주제임을 알 수 있다. 스마트폰 카메라의 성능 향상은 부품 및 세부 구성요소의 진화와 혁신에 기반을 두고 있으며, 이를 통해 스마트폰 카메라 관련 시장 및 기업들의 기회와 동시에 고도화된 생산 라인과 투자를 요구하는 도전이 있다는 것을 알 수 있다.

세 번째 토픽은 갤럭시, 삼성전자, 아이폰, 출시, 애플, 시리즈, 시장 등의 핵심어를 포함하고 있는데, 이는 ‘스마트폰 제조사 및 경쟁 제품 비교’ 주제를 표현하고 있음을 알 수 있다. 주요 스마트폰 제조사들은 스마트폰 카메라 기능과 관련하여 각기 다른 전략과 제품을 선보이고 있으며, 경쟁력을 높이기 위하여 카메라 기능 이외에도 다양한 혁신이 이루어지고 있음을 의미한다.

네 번째 토픽의 핵심어는 촬영, 사진, 화면, 출시, 제품, 갤럭시, 성능 디자인 등이며, 이는 ‘스마트폰 카메라 기능 및 디자인’ 성격을 가진 주제임을 알 수 있다. 스마트폰 사용자들은 카메라 성능뿐만 아니라 이를 편리하게 사용할 수 있는 디자인에도 관심을 가지며, 이에 따라 스마트폰 제조사들은 카메라 기능과 디자인에 관한 지속적인 연구개발 및 혁신을 통해 고객의 요구를 충족시키려고 노력하고 있음을 알 수 있다.

4-3 각 토픽의 트렌드 분석

본 연구에서는 토픽 모델링을 통해 ‘스마트폰 카메라’ 관련 뉴스기사의 4가지 토픽에 대한 시계열적 변화를 확인해보았다. 그림 7에서와 같이 첫 번째 토픽은 2016년 정점을 찍고 점차 상대적 비중이 낮아지고 있으며, 두 번째 토픽은 2015년 저점을 찍고 다시 상대적 비중이 높아지고 있다. 그림 8에서 세 번째 토픽은 변동이 있지만 전체적으로 상대적 비중이 높아지는 추세이며, 네 번째 토픽은 전체적으로 상대적 비중이 낮아지고 있다.

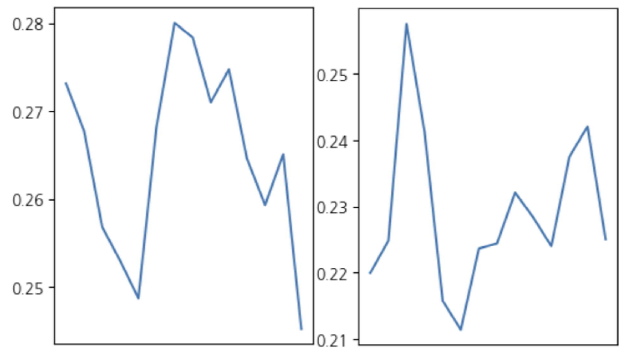


그림 7. 첫 번째 및 두 번째 토픽의 연도별 트렌드 추이
 Fig. 7. Trends about first & second topic

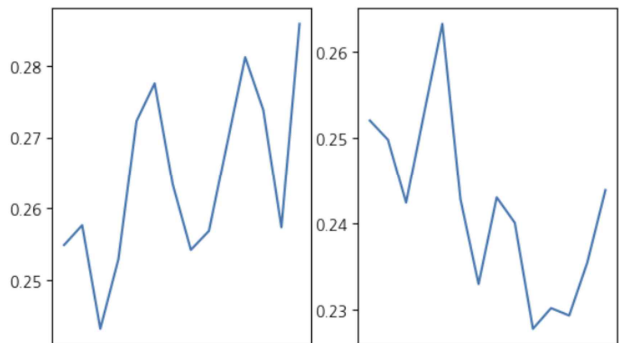


그림 8. 세 번째 및 네 번째 토픽의 연도별 트렌드 추이
 Fig. 8. Trends about second & fourth topic

V. 결론 및 시사점

본 연구는 LDA 토픽 모델링 분석방법을 통해 ‘스마트폰 카메라’ 관련 뉴스기사의 빅데이터에 나타난 주요 키워드와 토픽을 도출하고 트렌드를 파악하여, 분석결과를 기반으로 스마트폰 카메라에 대한 시사점을 제안하기 위하여 수행되었다. 스마트폰 카메라 관련 뉴스기사 빅데이터에 대한 토픽 모델링 분석을 수행한 결과, 4개의 토픽이 도출되었다. 첫 번째는 스마트폰 카메라 기반 서비스 및 응용이며, 두 번째는 스마트폰 카메라 시장 및 부품 산업이고, 세 번째는 스마트폰 제조사 및 경쟁 제품 비교이며, 네 번째는 스마트폰 카메라 기능 및 디자인이다.

첫 번째 토픽은 2016년 정점을 찍고 점차 상대적 비중이 낮아지고 있는데, 이는 시장 출시 초기에 스마트폰 카메라 기반 서비스 및 응용 기술의 확산으로 인해 많은 관심을 받았으나, 시간이 지남에 따라 다른 분야와 합쳐진 기존의 기술 또는 보다 통합적인 서비스들로 발전되었을 가능성을 보여준다.

두 번째 토픽이 2015년 저점을 찍고 상대적 비중이 높아지는 것은 스마트폰 카메라의 기능 향상에 따른 주변 부품 산업의 성장 및 카메라 기술에 대한 투자와 관심이 증가하고 있음을 보여준다. 이는 스마트폰 카메라에 대한 기대치 증가와 산업 성장세와 연관될 가능성이 있다.

세 번째 토픽은 상대적 비중이 전체적으로 높아지고 있는데, 이는 스마트폰 제조사 간의 카메라 기능 경쟁이 격화되고 있음을 나타낸다. 소비자들의 카메라에 대한 기대치가 높아지고 스마트폰 제조사들도 시장을 선점하기 위해 비교분석이 활발하게 일어나고 있음을 의미한다고 볼 수 있다.

네 번째 토픽은 전체적으로 상대적 비중이 낮아지고 있는데, 스마트폰 카메라에 대한 관심이 최근에는 배터리 수명이나 보안 등의 다른 기능에 주안점을 두고 있을 수 있다는 가능성을 보여준다고 할 수 있겠다.

스마트폰 카메라 기반 서비스 및 응용 분야는 초기에 큰 관심을 받았으나, 시간이 흐르면서 다른 분야와의 통합 및 기존 기술로 인한 상대적 비중 감소가 있다고 보여진다. 따라서, 기술 발전에 따른 새로운 서비스와 응용 기술의 출현 및 변화를 지속적으로 모니터링할 필요가 있다.

스마트폰 카메라 시장 및 부품 산업은 카메라 기능 향상에 따른 주변 부품 산업 성장 및 기술 투자 증가를 보여주고 있는데, 이는 카메라 기술 발전에 대한 지속적인 지원과 협업 노력이 필요하다는 것을 보여준다.

스마트폰 제조사 및 경쟁 제품 비교 분야에서는 카메라 기능 경쟁이 지속적으로 활발해지고 있으며, 이를 통해 소비자들의 카메라에 대한 기대치가 높아지고 제조사들도 이에 부응하여 혁신을 추구해야 함을 확인하였다.

스마트폰 카메라 기능 및 디자인 분야는 다른 기능이나 기술 특성과의 상대적 과편화로 인해 상대적 비중 감소 추세를 보이고 있는데, 이는 제품 디자인에만 집중하는 것이 아니라 다양한 기능 및 기술 특성을 고려해 종합적인 스마트폰 경쟁력을 강화해야 함을 보여준다.

이러한 결과를 바탕으로 미래 스마트폰 카메라 산업에 대한 변화, 기술 발전, 소비자들의 관심 및 기대를 예측하고 대응할 수 있는 학술적 기반을 제공한 것에 연구의 의의를 갖는다. 이러한 연구 결과는 스마트폰 카메라 기술 개발 및 제품 전략을 구축하는데 도움이 될 것으로 기대된다.

미래에는 스마트폰 카메라의 활용 범위가 더욱 확대될 것으로 보인다. 더욱 높은 해상도와 함께 다양한 센서 기술의 개발로 스마트폰 카메라는 즉각적인 정보 수집 및 분석, 교육 및 학습 지원, 건강 관리 및 의료 지원, 그리고 안전한 여행 등에 기여함으로써 개인의 삶을 더욱 자유롭고 풍요롭게 만들어 줄 것으로 예상된다.

본 연구의 학술적 시사점은 스마트폰 카메라 관련 연구에서 머신러닝, 특히 토픽 모델링의 사용 가능성을 확인하였으며 향후에 더욱 고도화된 머신러닝 기법을 통하여 더 많은 인사이트를 도출하는 데 기여할 것이다. 또한, 이 연구 결과를 바탕으로 스마트폰 카메라 산업 뿐만 아니라 다른 산업 영역에서의 후속 연구를 수행할 수 있는 토대를 마련하였으며, 학계의 관련 연구자들이 이 연구 결과를 기반으로 새로운 가설을 설정하거나 다양한 분석 방법을 도입할 수 있는 계기를 마련하였다고 할 수 있겠다.

본 연구의 실무적 시사점은 스마트폰 카메라 산업의 관련

기업들은 이 연구 결과를 활용하여 제품 개발 및 마케팅 전략을 보다 명확하게 설정할 수 있으며, 특히 본 연구에서 도출된 인사이트를 기반으로 시장 트렌드 예측이나 제품 포트폴리오 전략에 참고할 수 있을 것이다. 소비자들의 관심과 기대를 충족시키는 스마트폰 카메라 기능 및 디자인 개선을 고려하는 것은 기업들의 경쟁력을 강화하는 데 도움이 되며, 기업들은 본 연구에서 도출된 결과를 바탕으로 카메라 성능 이외에도 다양한 기능 및 기술 특성 개선에 지속적인 연구 개발을 진행해야 할 것이다. 기술 혁신의 확산과 통합을 통해 스마트폰 카메라 기반 서비스 및 응용 기술의 발전을 꾀할 수 있을 것이며, 이를 위해서는 스마트폰 카메라 산업 종사자들은 본 연구 결과를 활용하여 혁신적인 서비스 개발에 힘쓰며, 서비스 제공 업체들과의 협업을 통한 전략적 파트너십을 구축해야 할 것이다.

본 연구의 한계점은 스마트폰 카메라 관련 뉴스 기사를 데이터로 사용하였는데, 실제 사용자의 리뷰, 소셜 미디어, 전문가 의견 등의 데이터를 추가적으로 분석하게 되면 좀더 신뢰할 수 있는 인사이트를 얻을 수 있을 것이다. 그리고, 본 연구에서 사용한 머신러닝 기법인 토픽 모델링도 유용한 도구이지만 네트워크 분석, 감성분석 등과 같은 다양한 분석 방법을 적용하여 분석하게 되면 결과의 신뢰성을 높일 수 있을 것이다. 또한, 본 연구에서는 2010년부터 2023년 5월까지 장기간에 걸쳐 분석을 진행하였는데, IT 기술의 급격한 발전으로 스마트폰 카메라의 이용 행태와 관련 이슈도 다양하게 변화하고 있으므로, 향후에는 적절한 시기를 구분하여 분석하는 연구가 필요할 것으로 생각한다.

참고문헌

- [1] S. U. Park, J. Y. Kang, and S. C. Jung, *Python Text Mining Perfect Guide*, Paju: Wikibooks, 2022.
- [2] M. Anandarajan, C. Hill, and T. Nolan, *Practical Text Analytics: Maximizing the Value of Text Data (Advances in Analytics and Data Science 2)*, Cham, Switzerland: Springer, 2019.
- [3] R. Chen and W. Xu, "The Determinants of Online Customer Ratings: A Combined Domain Ontology and Topic Text Analytics Approach," *Electronic Commerce Research*, Vol. 17, No. 1, pp. 31-50, March 2017. <https://doi.org/10.1007/s10660-016-9243-6>
- [4] Y.-J. Cho, P.-W. Fu, and C.-C. Wu, "Popular Research Topics in Marketing Journals, 1995-2014," *Journal of Interactive Marketing*, Vol. 40, No. 1, pp. 52-72, November 2017. <https://doi.org/10.1016/j.intmar.2017.06.003>
- [5] Z. Qiao, X. Zhang, M. Zhou, A. Wang, and W. Fan, "A Domain Oriented LDA Model for Mining Product Defects from Online Customer Reviews," in *Proceedings of the 50th*

- Hawaii International Conference on System Sciences*, Waikoloa Village: HI, pp. 1821-1830, January 2017. <https://doi.org/10.24251/HICSS.2017.222>
- [6] B.-J. Suh and S.-Y. Shin, “A Study on the Research Trends on Domestic Platform Government Using Topic Modeling,” *Informatization Policy*, Vol. 24, No. 3, pp. 3-26, September 2017. <https://doi.org/10.22693/NIAIP.2017.24.3.003>
- [7] M. Steyvers and T. Griffiths, Probabilistic Topic Models, in *Handbook of Latent Semantic Analysis*, Mahwah, NJ: Lawrence Erlbaum Associates, Inc., pp. 420-440, 2007.
- [8] C.-H. Nahm, “An Illustrative Application of Topic Modeling Method to a Farmer’s Diary,” *Cross-Cultural Studies*, Vol. 22, No. 1, pp. 89-135, January 2016.
- [9] N. Kim, D. Lee, H. Choi, and W. X. S. Wong, “Investigations on Techniques and Applications of Text Analytics,” *The Journal of Korean Institute of Communications and Information Science*, Vol. 42, No. 2, pp. 471-492, February 2017. <https://doi.org/10.7840/kics.2017.42.2.471>
- [10] D. M. Blei, “Probabilistic Topic Models,” *Communications of the ACM*, Vol. 55, No. 4, pp. 77-84, April 2012. <https://doi.org/10.1145/2133806.2133826>
- [11] W.-R. Yang and H.-C. Yang, “Topic Modeling Analysis of Social Media Marketing Using BERTopic and LDA,” *Journal of Industrial Distribution & Business*, Vol. 13, No. 9, pp. 37-50, September 2022. <http://dx.doi.org/10.13106/jidb.2022.vol13.no9.37>
- [12] D. Y. Lee and H. S. Yi, “Exploring Methods for Determining the Appropriate Number of Topics in LDA: Focusing on Perplexity and Harmonic Mean Method,” *Journal of Educational Evaluation*, Vol. 34, No. 1, pp. 1-30, March 2021. <http://dx.doi.org/10.31158/JEEV.2021.34.1.1>
- [13] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, 2003.
- [14] S. Arora, R. Ge, Y. Halpern, D. Mimno, A. Moitra, D. Sontag, ... and M. Zhu, “A Practical Algorithm for Topic Modeling with Provable Guarantees,” in *Proceedings of the 30th International Conference on Machine Learning (ICML '13)*, Atlanta: GA, pp. 280-288, June 2013.
- [15] F. Rosner, A. Hinneburg, M. Röder, M. Nettling, and A. Both, “Evaluating Topic Coherence Measures,” in *Proceedings of Neural Information Processing Systems Foundation (NIPS 2013) - Topic Models Workshop*, Stateline: NV, December 2013. <https://doi.org/10.48550/arXiv.1403.6397>
- [16] Y.-J. Cha, J.-H. Lee, J.-E. Choi, and H.-W. Kim, “A Topic Modeling Approach to Marketing Strategies for Smartphone Companies,” *Knowledge Management Review*, Vol. 16, No. 4, pp. 69-87, December 2015. <http://dx.doi.org/10.15813/kmr.2015.16.4.005>
- [17] Y. Lee and S. Seol, “A Study on Developing Patent Strategy Based on Survey: Focusing on Smartphone Camera Field,” *Asia-Pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, Vol. 7, No. 7, pp. 913-923, July 2017. <http://doi.org/10.35873/ajmahs.2017.7.7.086>
- [18] E.-C. Kim and S.-I. Kim, “A Study on the User Experience of Smartphone Camera Application,” *Journal of the Korea Convergence Society*, Vol. 8, No. 12, pp. 221-226, December 2017. <https://doi.org/10.15207/JKCS.2017.8.12.221>
- [19] M. Kang, C. Lee, D. Lee, and Y. Lee, “Identifying Characteristics and Types of Generation Z according to the Behavior of Smartphone Camera Use,” *Archives of Design Research*, Vol. 33, No. 3, pp. 155-174, August 2020. <http://dx.doi.org/10.15187/adr.2020.08.33.3.155>
- [20] H.-J. Kim, S. Kim, and S.-T. Kim, “Topic Modeling of Media Reports on Smartphone Addiction: A Study on the Comparison of Government Policies between 2010 and 2018,” *Korean Journal of Broadcasting & Telecommunications Research*, No. 104, pp. 38-69, October 2018.
- [21] H.-S. Lee, G. Lee, and J. Cho, “What Words Have the Press Used to Describe Children’s Smartphone Use?: A Semantic Network Analysis and Topic Modeling of the Press Article between 2011 and 2021,” *Korean Journal of Communication & Information*, Vol. 114, pp. 204-246, August 2022. <https://doi.org/10.46407/kjci.2022.08.114.204>
- [22] Korea Press Foundation. Big Kinds [Internet]. Available: <https://www.bigkinds.or.kr>.
- [23] T. Kim, B. Ahn, W. Lee, and H. Kang, “Analysis of Metaverse Trends using News Big Data,” *Journal of Digital Contents Society*, Vol. 23, No. 2, pp. 203-216, February 2022. <http://dx.doi.org/10.9728/dcs.2022.23.2.203>
- [24] S.-C. Lee, “Topic Modeling of Newspaper Articles on Government ‘Senior Job Program’ via Latent Dirichlet Allocation,” *Journal of Digital Convergence*, Vol. 18, No. 10, pp. 537-546, October 2020. <https://doi.org/10.14400/JDC.2020.18.10.537>



박천호(Cheon-Ho Park)

1991년 : 서강대학교 물리학과 (이학사)
2008년 : 아주대학교 경영학과 (경영학 석사)
2023년 : 숭실대학교 대학원 IT정책경영학과 박사과정 재학 중

1991년~1998년: 삼성테크윈(주) 선임연구원
2001년~2009년: 삼성전기(주) 수석연구원
2009년~2012년: 삼성엘이디(주) 수석연구원
2012년~현 재: 삼성전자(주) 수석연구원
※ 관심분야 : 빅데이터 분석, 가상세계(Virtual World), 비전 & 디스플레이 시스템, 품질관리, 경영혁신 등



김태경(Tae-Gyong Kim)

1993년 : 건국대학교 화학과(이학사)
2023년 : 숭실대학교 대학원 IT정책경영학과 (공학석사)
2023년 : 숭실대학교 대학원 IT정책경영학과 박사과정 재학 중

1993년~2003년: 삼성생명(주) 과장
2003년~2014년: 티맥스소프트(주) 상무
2015년~현 재: 포에버아이씨티 전무이사
※ 관심분야 : 빅데이터, 딥러닝, 클라우드, 블록체인 등



최정일 (Jeongil Choi)

2004년 : 미국 University of Nebraska-Lincoln (경영학 박사)

2022년~현 재: 한국품질경영학회 회장
2007년~현 재: 숭실대학교 경영학부 교수
※ 관심분야 : 온라인 비즈니스 모델, IT기반의 서비스 혁신, 서비스 운영 및 품질평가 등