KIPS Trans. Softw. and Data Eng.
Vol.12, No.1 pp.19~30
ISSN: 2287-5905 (Print), ISSN: 2734-0503 (Online)

https://doi.org/10.3745/KTSDE.2023.12.1.19

Analysis of Social Trends for Electric Scooters Using Dynamic Topic Modeling and Sentiment Analysis

Kyoungok Kim† · Yerang Shin††

ABSTRACT

An electric scooter(e-scooter), one popularized micro-mobility vehicle has shown rapidly increasing use in many cities. In South Korea, the use of e-scooters has greatly increased, as some companies have launched e-scooter sharing services in a few large cities, starting with Seoul in 2018. However, the use of e-scooters is still controversial because of issues such as parking and safety. Since the perception toward the means of transportation affects the mode choice, it is necessary to track the trends for electric scooters to make the use of e-scooters more active. Hence, this study aimed to analyze the trends related to e-scooters. For this purpose, we analyzed news articles related to e-scooters published from 2014 to 2020 using dynamic topic modeling to extract issues and sentiment analysis to investigate how the degree of positive and negative opinions in news articles had changed. As a result of topic modeling, it was possible to extract three different topics related to micro-mobility technologies, shared e-scooter services, and regulations for micro-mobility, and the proportion of the topic for regulations for micro-mobility increased as shared e-scooter services increased in recent years. In addition, the top positive words included quick, enjoyable, and easy, whereas the top negative words included threat, complaint, and ilegal, which implies that people satisfied with the convenience of e-scooter or e-scooter sharing services, but safety and parking issues should be addressed for micro-mobility services to become more active. In conclusion, this study was able to understand how issues and social trends related to e-scooters have changed, and to determine the issues that need to be addressed. Moreover, it is expected that the research framework using dynamic topic modeling and sentiment analysis will be helpful in determining social trends on various areas.

Keywords : Personal Mobility Vehicle, Electric Scooter, Text Mining, Dynamic Topic Modeling, Sentiment Analysis

동적 토픽 모델링과 감성 분석을 활용한 전동킥보드에 대한 사회적 동향 분석

김 경 옥† · 신 예 랑††

요 약

마이크로 모빌리티 중 하나인 전동킥보드의 이용은 세계적으로 급격히 성장하고 있는 추세이다. 국내에서는 2018년 서울에서 서비스를 시작한 킥고잉을 비롯하여 서울을 포함한 일부 대도시에서 공유킥보드 서비스를 제공하는 업체가 생기면서 전동킥보드의 이용이 크게 증가했다. 하지만, 전동킥보드의 이용은 여전히 주차, 안전에 대한 문제로 인해 논란의 대상이 되고 있다. 이동수단에 대한 인식은 사용자들이 어떤 이동수단을 선택할지에도 영향을 끼치므로 전동킥보드 이용 및 공유킥보드 서비스 활성화를 위해서는 관련 이슈와 그에 대한 대중의 인식을 파악할 필요가 있다. 이에 본 연구에서는 전동킥보드 관련 이슈에 대한 사회적 동향을 파악하는 것을 목표로 시간에 따른 이슈의 변동성을 고려해 동적 토픽 모델링과 감성 분석을 활용하여 2014년에서 2020년까지의 전동킥보드 관련 뉴스 기사를 분석하였다. 토픽 모델링을 통해 마이크로 모빌리티 기술, 공유킥보드 서비스, 킥보드 관련 규제 관련 토픽을 도출하였으며, 공유킥보드 서비스 증가와 함께 안전에 대한 이슈가 크게 불거지면서 킥보드에 대한 규제 관련 토픽의 비중이 최근 들어 크게 증가함을 확인했다. 그뿐만 아니라 감성 분석을 통해 킥보드 관련 뉴스에 주로 등장하는 긍정어는 신속, 즐기다, 손쉽다, 편리 등이 있고 부정어는 위협, 불법, 침해 등으로 나타나 킥보드나 공유킥보드 서비스의 편의성에는 만족하지만, 마이크로 모빌리티 서비스에서 안전, 주차 등의 문제는 여전히 해결해야하는 이슈임을 알 수 있었다. 결론적으로, 본 연구를 통해 전동킥보드에 대한 이슈와 그에 대한 관심과 사회적 감성의 변화를 확인하고 어떤 이슈에 대한 대응이 필요한지 파악할 수 있었다. 이 연구의 분석의 틀은 향후 다양한 사회 현안에 대한 사회적 동향을 파악하고 그에 대한 대응 방안을 마련하는데 활용할 수 있을 것으로 기대된다.

키워드 : 개인형 이동 장치, 전동킥보드, 텍스트 마이닝, 동적 토픽 모델링, 감성 분석

※ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2020R1F1A1054496).
† 정 회 원 : 서울과학기술대학교 산업공학과 부교수
†† 비 회 원 : 서울과학기술대학교 데이터사이언스학과 석사수료

Manuscript Received : February 15, 2022
First Revision : April 5, 2022
Accepted : May 10, 2022
* Corresponding Author : Kyoungok Kim(kyoungok.kim@seoultech.ac.kr)

※ 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.

1. 서 론

1.1 연구의 배경 및 목적

인구가 밀집된 도시에서는 주택, 토지, 교통, 실업, 범죄 등 다양한 사회 문제가 발생한다. 그중 교통 분야에서는 승용차의 급증으로 인한 교통 체증, 소음, 공해와 같은 고질적인 문제의 해결방안을 찾는 연구가 지속해 이뤄지고 있다[1]. 도심에서의 늘어나는 교통량과 혼잡 문제를 해결하려면 대중교통을 활성화할 필요가 있는데 대중교통은 대중교통을 이용할 시 출발 구간과 도착 구간에서의 추가적인 이동수단이 필요할 수 있는 퍼스트-라스트 마일 문제(first-last mile problem)가 존재한다[2]. 최근에는 이 문제의 해결방안으로 전기와 같은 친환경 동력을 활용하는 소형 이동수단인 마이크로 모빌리티가 주목을 받고 있다[3].

대표적인 마이크로 모빌리티로는 자전거, 전기자전거, 전동킥보드 등이 있는데, 이런 개인형 이동수단은 등장한 지는 오래되었지만, 2010년 이후에 개인이 저렴한 가격으로 단기간에 이용할 수 있는 공유 서비스가 세계 많은 도시에 급속도로 도입되면서 이용이 급속히 증가하고 있다[4]. 마이크로 모빌리티 중 전동킥보드는 가장 뒤늦게 등장했는데, 스타트업 Bird가 Santa Monica를 기반으로 2017년에 최초의 전동킥보드 공유서비스를 시작했다[5]. 그 이후 공유킥보드 서비스 지역과 이용자 및 이용횟수는 빠르게 증가하고 있으며, 미국을 기준으로 2019년 공유킥보드 이용횟수는 8천6백만에 달했는데 이는 2018년에 비해 2배 이상 증가한 것이다[6].

국내에서는 2018년 9월 킥고잉이 최초의 공유킥보드 서비스가 선보였으며 뒤이어 라임, 빔, 씽씽 등 다양한 업체가 등장해서 현재 20여 개 공유킥보드 업체가 서비스를 운영 중이다[7]. 출시 초기에는 서울 내에서도 유동인구가 많은 도심지 중심으로 서비스가 제공되었으나, 점차 수도권, 부산, 대전 등 대도시를 중심으로 서비스 지역이 확대되고 있으며, 2018년 대비 2020년의 서울시 관내 공유킥보드는 연평균 119.5배 증가하였다[8].

이렇게 공유킥보드 서비스의 확대로 인해 전동킥보드의 이용이 국내에도 크게 늘었지만, 주차, 안전 등의 이슈에 대한 우려 또한 해소되지 못한 채 지속해서 제기되고 있다[9, 10]. 이동수단의 선택에는 선택에 대한 접근성, 편리성 등만이 아니라 이용자가 이동수단에 대해 갖는 인식도 큰 영향을 끼친다[11, 12]. 그러므로 전동킥보드의 이용을 더 활성화하기 위해서는 사회적으로 어떤 이슈들이 제기되고 있으며, 그것들이 긍정 혹은 부정적으로 인식되고 있는지 파악할 필요가 있다. 또한, 제기되는 이슈와 그에 대한 인식이 어떻게 변화하는지 파악해 대응 방안을 빠르게 수립하는 것 또한 전동킥보드가 유용한 이동수단의 하나로 안착하는 데 도움이 된다.

이에 국외에서는 전동킥보드에 대한 사회적 감성을 분석한 몇몇 사례를 찾을 수 있다. Feng *et al.*[13]은 18개월 동안의 전동킥보드 관련 트윗을 수집하여 토픽 모델링을 이용해 전동킥보드에 대한 어떤 이슈들이 있는지 분석하였다. Aman *et al.*[14]은 전동킥보드 사용자의 전동킥보드에 대한 만족도와 만족도에 영향을 끼치는 요인을 파악하기 위해 12,026개의 공유 전동킥보드 앱의 리뷰를 분석하였다. 그 결과 사용자의 만족도에 크게 영향을 끼치는 주요 요인은 안전, 가격, 앱의 기능임을 밝혔다.

그렇지만 국내에서 진행된 전동킥보드에 대한 연구는 전동킥보드 제조 기술에 대한 주제를 제외하고는 이용 내역이나 설문조사를 기반으로 해서 전동킥보드 사용자의 이용 패턴을 분석하거나[15, 16], 전문가에 의한 동향 분석[17, 11]이 주를 이루고 있다.

이에 본 연구에서는 뉴스 데이터를 활용해 1) 토픽 모델링(Topic Modeling)을 통한 전동킥보드 관련 사회적 이슈를 추출하고 2) 감성 분석(Sentiment Analysis)을 통해 이슈별로 사회적 감성지수가 어떻게 되는지 파악하고자 한다.

1.2 연구의 범위 및 방법

본 연구의 시간적 범위는 국내에 전동킥보드가 소개되기 시작하는 2014년부터 2020년까지이며, 분석 기간 동안의 전동킥보드 관련 뉴스를 분석 대상으로 삼았다.

분석 방법으로는 이 연구는 다년간의 걸친 뉴스 데이터로부터 이슈의 시계열적인 변화 과정까지 고려해 이슈를 파악하기 위해서 기존 연구에서 토픽 모델링에 많이 활용하고 있는 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 대신에 동적 토픽 모델링(Dynamic Topic Modeling, DTM)을 활용하고 토픽 모델링 결과와 개별 뉴스의 감성 분석 결과를 통합하여 전동킥보드에 대한 사람들의 감성지수가 이슈에 따라서 어떻게 다르게 변화했는지 정량적으로 보여준다는 점에서 의의가 있다.

2. 이론적 배경

2.1 동적 토픽 모델링

토픽 모델링은 구조화되지 않은 비정형 데이터 중 하나인 텍스트로부터 의미 있는 정보를 추출하는 텍스트 마이닝의 한 기법으로, 문서 집합 내에서 추상적인 주제(토픽)를 찾는 통계적인 분석 방법이다. 여러 기법 중 LDA는 토픽 모델링의 대표적인 알고리즘으로, 문서들은 토픽들의 혼합으로 구성되어 있고, 토픽들은 서로 다른 단어들의 등장 확률 분포로 정의된다는 가정을 바탕으로 문서 집합으로부터 지정된 개수의 토픽을 추출한다[18]. 하지만, 이 방법은 개별 토픽이 시간에 따라 변화하지 가정하므로 서로 다른 토픽들이 서로 영향을 주며 변화하는 과정을 포착하지 못하는 문제가 있다[19].

이와 같은 단점을 보완하기 위해 등장한 기법이 시계열적인 문서 집합에 대해 시간에 따른 토픽의 변화까지 파악할 수

있는 DTM이다. DTM도 LDA와 유사하게 토픽을 단어들의 등장 확률 분포로 정의하지만, 문서 집합을 서로 다른 시간 구간으로 분할하고 한 구간에서의 토픽은 이전 구간에서의 토픽에서 진화한 것이라고 가정하는 점에서 LDA와 차이가 있다[20]. DTM은 동적으로 변화하는 시장의 상황을 관찰하거나, 산업과 기술의 융합 과정을 분석하는 등 다양한 분야의 시계열 문서 집합의 분석에 활용되고 있다[21, 22].

2.2 감성 분석

감성 분석은 토픽 모델링과 마찬가지로 다양한 분야에서 널리 활용되는 텍스트 마이닝의 한 기법으로, 텍스트에서 나타나는 사람들의 태도, 의견, 성향 등과 같은 주관적인 정보를 파악하는 것을 목적으로 한다. 감성 분석은 온라인상에서 손쉽게 얻을 수 있는 리뷰, 블로그, 뉴스 기사, SNS 데이터 등을 활용해서 기업들이 자사의 제품이나 서비스에 대한 소비자의 반응을 살피기 위해서나 특정 사회적 문제에 대한 대중들의 의견을 파악하는 등 다양한 분야에서 활용되고 있다[23-26].

감성 분석에는 다양한 기법들이 활용되고 있는데 그중 사전 기반 분석 방법은 미리 구축해 놓은 감성사전(sentiment lexicon)을 기반으로 문서에 사전에 존재하는 단어가 있으면 사전에 정의된 감성점수를 부여하고 이들을 종합하여 문서의 감정을 분류하는 방법이다[26]. 사전의 구축은 수동으로 이뤄지기도 하고 일부 시드(seed) 단어를 기반으로 자동으로 사전을 확장해나가는 방법을 사용하기도 한다. 일단, 감성사전 기반의 감성 분석은 감성사전이 구축되고 나면 쉽고 빠르게 문서의 감성을 파악할 수 있는 장점이 있다. 반면에, 개별 단어를 독립적으로 보기 때문에 문맥을 고려할 수 없고 구축된 사전이 잘못된 경우 부정확한 결과를 얻을 수 있다[27].

또 다른 감성 분석의 접근법은 기계학습 방법론을 활용하는 것이다. 이는 감성점수를 알고 있는 문서를 학습 데이터로 이용해서 감성점수를 예측하는 회귀 모델을 학습하거나, 긍정/부정 여부를 판별하는 분류 모델을 학습한다[27]. 감성점수를 예측하는 회귀 이 방법은 감성 사전 기반 방법보다 정확한 감성을 추출할 수 있지만, 감성점수를 미리 알고 있는 문서는 제한되어 있어 학습 데이터를 확보하는 데 어려움이 있다. 학습 데이터가 확보되면, 비정형 데이터인 텍스트를 정형 데이터에 적용할 수 있는 기계학습 알고리즘을 활용하기 위해 벡터화시키는 과정이 필요하다. 과거에는 개별 단어들의 하나의 차원으로 표현되는 벡터 스페이스 모델(vector space model)이 많이 활용되었다[28]. 하지만 이 방식은 문서 내 단어들의 등장 빈도와 단어가 등장한 문서의 수를 기반으로 하기 때문에 맥락을 반영하기 어렵다는 단점이 있다. 이 단점을 극복하기 위해 단어들 사이의 맥락을 반영할 수 있는 신경망 기반의 doc2vec과 같은 문서 임베딩(embedding) 기법이 개발되었다[29, 30]. 최근에는, Google이 개발한 양방향성 순환신경망 기반의 자연어 언어 모델인 BERT (Bidirectional Encoder Representations from Transformers)가 기존 순환신경망 모델보다 더 뛰어난 성능으로 각광받고 있으며[31, 32], 이 모델을 한국어에 적용한 KoBERT도 개발되었다[33, 34].

텍스트가 벡터화한 이후에는 선형 회귀, 의사결정나무, SVM, 신경망 등 다양한 기계학습 기법이 감성점수 예측 또는 감성 판별 모델 학습에 활용되고 있다[35]. 문서 벡터화 기법에 대한 연구와 마찬가지로 모델링 단계에서도 맥락을 고려한 감성 분석을 위해 CNN, RNN 등과 같은 신경망 기반의 알고리즘이 많이 개발되고 있다[36, 37].

3. 연구 방법

3.1 데이터 수집

최근에는 본 연구에서는 국내 전동킥보드 관련 이슈에 대한 분석을 위해서 전동킥보드를 주제로 다루고 있는 뉴스 데이터를 활용한다. 블로그나 SNS에서는 특정 킥보드 제품이나 공유킥보드 서비스에 대한 사용자의 평가 및 후기가 주를 이루고 있으므로 전동킥보드 관련 현안을 파악하는 데는 적절하지 않다고 판단하였다.

뉴스 데이터는 국내 포털 사이트 네이버와 다음을 통해 수집하였다. 사용된 뉴스는 각각의 포털 사이트 뉴스 플랫폼 내에서 제공되는 기사로 한정하였다. 포털 사이트에서 검색되는 이미지를 기준으로 총 3개의 검색어 '전동킥보드', '공유킥보드', '공유스쿠터'를 선정하였고, 연관성이 낮은 뉴스를 제외하기 위해 제목에 선정된 키워드가 포함되는 뉴스만 수집하였다. 공유킥보드 서비스가 등장한 것은 2018년이지만, 이전에도 전동킥보드 제품은 국내 시장에 도입되었기 때문에 전동킥보드와 관련된 뉴스가 등장하기 시작하는 시점부터 뉴스를 수집하였다. 최종적으로 수집된 뉴스의 시간적 범위는 2014년 1월 1일부터 2020년 12월 31일까지이며 중복된 뉴스는 제거하고 총 7,310개의 뉴스 기사를 확보하였다.

추가로 뉴스 데이터 이외에 국내의 전동킥보드 관련한 사람들의 관심 정도가 어떻게 변화해왔는지 확인하기 위해 네이버 트렌드를 이용해 전동킥보드 관련 키워드의 검색량 추이를 살펴보았다. 네이버 트렌드는 구글 트렌드와 동일한 서비스로 특정 기간 내에 사용자가 입력한 키워드 집합에 대한 검색량을 기간 내 최대치를 100으로 하여 상대적으로 보여준다. 이 연구에서는 뉴스를 추출하기 위해 사용한 키워드 집합(Set 1) 외에도 공유킥보드 업체들(킥고잉, 지쿠터, 씽씽, 빔, 라임, 디어)에 대한 키워드 집합(Set 2)에 대해서도 네이버 트렌드 값을 추출하였다. 네이버 트렌드 값은 2016년부터 얻을 수 있어 전동킥보드 키워드에 대해서는 2016년에서 2020년까지 트렌드 값을 얻었고, 검색어로 사용한 공유킥보드 업체 중 가장 먼저 국내 서비스를 시작한 킥고잉이 2018년 9월에

서비스를 시작해서 공유킥보드 업체 키워드에 대해서는 2018년 8월부터 2020년 12월까지 트렌드 값을 얻었다.

감성 분석을 위해서는 한국어 감정 및 의견 분석 코퍼스인 KOSAC(KOrean Sentiment Analysis Corpus)[36]을 활용하였다. 이 사전은 1-gram에서 3-gram까지 총 16,362개의 표제어에 대해서 극성(polarity)과 강도(intensity) 등의 정보를 제공하고 있다.

3.2 데이터 전처리

토픽 모델링을 위해서는 특수문자를 제거한 후 형태소 분석을 거쳐 명사만 추출하였다. 또한, '전동킥보드', '킥보드', '스쿠터', '전동', '스쿠터', '공유킥보드', '공유스쿠터' 등과 같이 모든 뉴스 기사에 빈번하게 등장하는 단어는 토픽을 정의하는데 쓸모가 없으므로 이런 단어를 제거하기 하기 위해서 전체 뉴스 기사의 80% 이상에 등장하는 단어는 제거하였다. 마찬가지로 효율적인 분석을 위해서 전체 뉴스 기사의 5% 미만에 등장하는 단어도 제거하였다.

KOSAC을 활용해서 뉴스별 감성점수를 계산하기 위해서 형태소 분석을 진행한 후에 1-gram, 2-gram, 3-gram을 추출하였다. 그 이후에 전동킥보드와 관련된 단어들이 포함된 감성사전을 구축하기 위해서 1-gram 중에서 명사, 동사, 형용사, 부사만을 따로 추출하였다.

3.3 분석 방법

토픽 모델링 분석에서는 우선 적절한 토픽 개수를 결정하기 위해 뉴스 날짜를 고려하지 않고 토픽 수를 바꿔가며 LDA를 적용하고 주제 일관성 지표를 비교하였다. 최적의 토픽 개수가 결정된 이후에는 연도에 따라 뉴스 기사를 분할하고 DTM을 적용하였다. DTM에서 동적 변화의 정도를 조절하는 파라미터인 체인 파라미터(chain parameter)는 0.005로 설정하였다.

뉴스 데이터의 감성 분석은 두 단계에 걸쳐 진행한다. 먼저, 문장 단위로 추출한 n-gram과 KOSAC 사전을 대조하여 사전에 존재하는 단어를 파악한다. 이때 3-gram부터 차례로 대조하여, 3-gram이 존재하면 해당 3-gram으로부터 얻어지는 2-gram과 1-gram 조합은 중복해서 감성점수가 부여되지 않도록 했다. 매칭되는 n-gram의 감성점수는 기존 연구[37]를 참조하여, n-gram의 극성뿐 아니라 강도까지 고려하여 산출했다. 강도는 High, Medium, Low 중 하나로 제시되는데 강도에 따라서 극성점수를 각각 3배, 2배, 1배 하였다. 개별 n-gram의 점수는 문서 단위로 평균값을 계산했으며 문서의 최종 감성점수는 문서의 긍정점수에서 부정점수를 뺀 값으로 정의했다.

문서의 감성점수를 얻은 다음에는 수집한 뉴스 데이터에 특화된 감성 사전을 구축하기 위해서 감성점수가 높은 상위 30% 뉴스(긍정 뉴스)와 점수가 낮은 하위 30% 뉴스(부정 뉴스)를 추출하고 긍/부정을 판별하는 분류 모델을 elastic net 정규화 로지스틱 회귀를 통해서 학습하였다. 이때 Bag-of-Words(BoW) 방식으로 문서-단어 행렬을 구성하는데 명사, 동사, 형용사, 부사만 선택해서 설명변수로 이용하였다. 단, 문서-단어 행렬을 얻을 때 단어의 실제 등장 빈도를 그대로 사용하게 되면 뉴스 길에 따라 단어의 등장 빈도의 편차가 커지는 문제가 발생하므로 문서-단어 행렬에서 각 문서를 구성하는 단어들의 빈도를 Equation (1)을 이용해 정규화하여 단어-빈도 행렬을 구성하였다. Equation (1)에서 $f_{w,d}$는 d번째 뉴스에서 단어 w에 대한 빈도를 의미한다.

$$f'_{w,d} = \sqrt{\frac{f_{w,d}^2}{\sum_w f_{w,d}^2}} \tag{1}$$

elastic net 정규화는 계수의 절댓값에 제약(l_1 제약)을 주는 방식과 계수의 제곱합에 제약(l_2 제약)을 주는 방식을 혼합해서 사용한다. 제약 조건에는 전체 제약 조건의 강도를 조절하는 하이퍼 파라미터(hyper parameter) α와, l_1제약의 가중치를 조절하는 하이퍼 파라미터 l이 존재하는데 최적의 하이퍼 파라미터는 5-겹 교차 검증을 통해 결정하였다. 최적의 하이퍼 파라미터를 결정한 후에는 추출한 전체 데이터를 이용해서 로지스틱 회귀 모형을 구축해서 모형의 계수를 추정하였다. 이때 계수의 부호는 극성의 방향을 계수의 절대값은 극성의 정도로 해석할 수 있으므로 이를 이용해 감성사전을 구축하였다[40]. 이렇게 구축된 감성사전을 바탕으로 뉴스의 최종 감성점수를 산출하였다.

뉴스별로 감성점수가 산출되면 토픽 모델링 결과와 결합하여 토픽별 감정점수를 Equation (2)를 이용해서 계산한다. Equation (2)에서 s_d는 d번째 뉴스의 감성점수이고 $p_{d,t}$는 d번째 뉴스의 t번째 토픽에 대한 비중을 뜻하고 이를 이용해서 계산한 t번째 토픽의 감성점수가 S_t이다.

$$S_t = \frac{\sum_d s_d p_{d,t}}{\sum_d p_{d,t}} \tag{2}$$

4. 분석 결과

4.1 탐색적 데이터 분석 결과

Fig. 1은 두 키워드 집합에 대한 네이버 트렌드 결과를 보여주고 있다. 업체 관련 키워드는 킥보드 키워드에 비해 검색량이 적어 개별 키워드 집합에 대해서 따로 결과를 얻었다. 2019년 기준으로 두 키워드 집합에 대해 동시에 트렌드 값을 추출했을 때와 업체 키워드만 단독으로 트렌드 값을 추출했을 때를 비교한 결과 업체 관련 키워드의 검색량은 킥보드 키워드에 비해 검색량이 1/12 정도 되는 것으로 나타났다. 킥보드 관련 키워드의 검색량은 산발적으로 검색량이 늘어나는

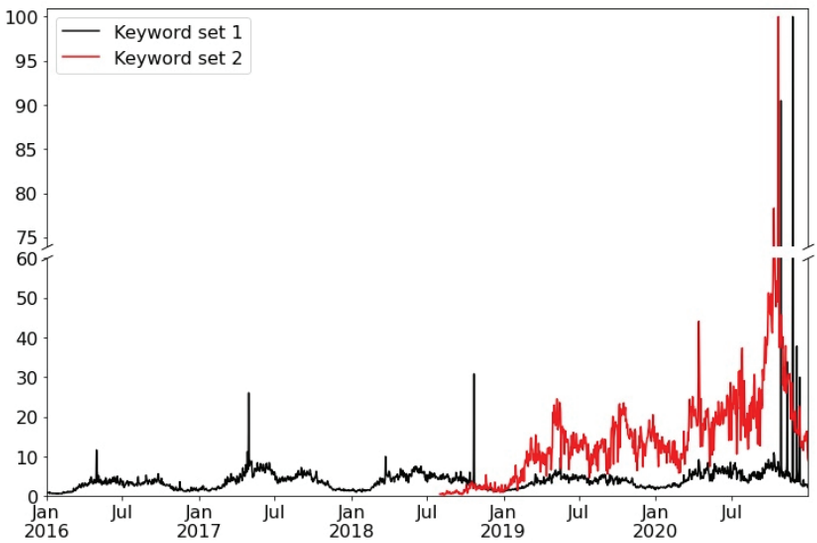
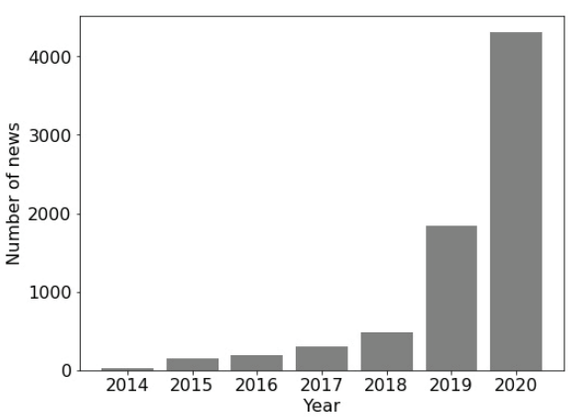
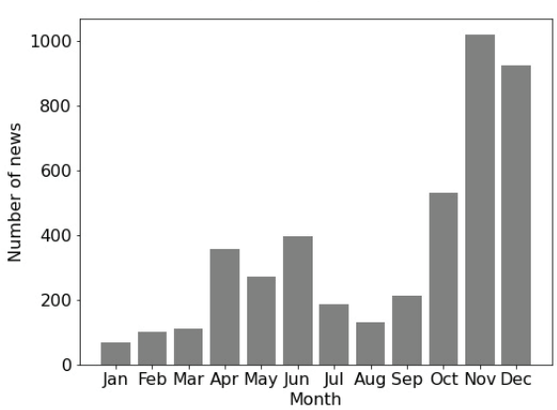


Fig. 1. Naver Trend Scores

날이 있긴 하지만 2016년 이후 큰 변동 없이 유지되다가 2020년 11월 이후에 검색량이 폭증한 날이 다수 나타났다. 이는 전동킥보드에 대한 법규가 추가된 2020년 6월 9일에 개정된 도로교통법이 2020년 12월 10일부터 적용되면서 언론에서 관련 기사를 다수 내보낸 것의 영향으로 보인다.



(a) Yearly



(b) Monthly

Fig. 2. The Number of News Articles

업체 관련 키워드는 꾸준히 증가하는 추세를 보이며, 개정된 도로교통법 시행 전에 급격히 검색량이 급증하는 현상이 나타났다. 도로교통법 개정 전에는 전동킥보드는 원동기장치자전거로 분류돼 면허가 필요하고 차도로만 주행해야 했다. 그러나 개정된 도로교통법에는 시속 25km, 30kg 미만의 이동수단은 개인형 이동 장치로 분류해, 자전거와 유사한 안전의무를 부가하고 만 13살 이상이면 면허 없이 이용할 수 있게 했다. 이와 같은 법 개정은 공유킥보드 서비스 업체에 큰 영향을 끼치므로 업체 관련 검색량이 증가한 것으로 보인다.

Fig. 2는 수집한 뉴스 기사의 연도별 분포와 2020년의 월별 분포를 보여주고 있다. 관련 뉴스는 2018년까지는 서서히 증가하다 2019년과 2020년에 크게 증가했다. 2020년의 뉴스 기사의 분포를 보면 도로교통법이 개정된 전후인 4~6월에 기사가 이전보다 증가했으며, 개정된 법규가 적용되는 12월 전에 다시 뉴스가 급증한 것을 볼 수 있다.

4.2 동적 토픽 모델링 결과

DTM 모형의 최적 토픽 수를 결정하기 위해서 토픽의 수를 변경하여 LDA 모델을 학습하고 모델의 주제 일관성(topic coherence)를 C_v 값을 계산하였다. C_v를 선택한 것은 여러 연구와 실험을 통해서 이 지표가 토픽의 일관성을 가장 잘 측정한다는 결과가 있었기 때문이다[40]. Fig. 3은 토픽의 개수를 2에서 10으로 변화시켰을 때의 C_v값을 보여주고 있는데, 3일 때 C_v값이 가장 높아 토픽 개수를 3개로 결정하였다. 이후 뉴스를 연도별로 구분해 DTM을 적용하여 세 개의 토픽에 대한 연도별 단어 확률 분포를 얻었다.

각 토픽별로 등장 확률이 높은 단어들을 살펴본 결과, 첫

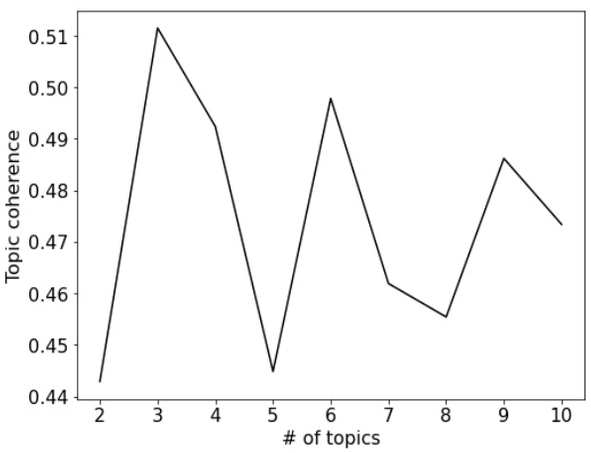


Fig. 3. Topic Coherecne by the Number of Topic

번째 토픽(Topic 1)은 마이크로모빌리티 기술과 관련이 많았으며, 두 번째 토픽(Topic 2)은 공유킥보드 서비스와 연관되어 있었으며, 세 번째 토픽(Topic 3)은 킥보드 관련 규제에 대한 토픽으로 정의 내릴 수 있었다. Table 1, Table 2, Table 3에 각 토픽별로 평균 확률이 높은 20개 단어에 대해서 연도별로 확률이 어떻게 변화했는지를 정리하였다.

Topic 1에서는 모빌리티, 서비스, 배터리 등의 단어는 확률이 증가하는 반면에 제품, 전기차, 전기자전거 등의 단어는 확률이 감소하는 추세를 보인다. 이는 전동킥보드가 국내에 도입된 초기에는 전동킥보드에 국한된 이슈보다는 전기를 동력으로 삼는 이동수단에 대한 전반적인 이슈가 기사로 많이 다뤄지다가 이후에 국내에서도 전동킥보드 시장이 확대되고, 공유킥보드 서비스가 늘어나면서 이들을 중점으로 다룬 기사가 늘어났기 때문으로 보인다.

Topic 2에서는 업체, 추진, 주차, 이용자 등의 단어는 확률이 증가하는 반면, 이용, 설치, 규제, 운행 등의 단어는 확률이 감소하는 경향을 보인다. 이 토픽에서는 상위 20개 단어에는 포함되어 있지는 않지만, 2018년 이후부터는 공유, 관리, 확산, 방치, 킥고잉, 씽씽 등과 같이 국내 공유킥보드 서비스와 연관된 단어들의 비중이 증가하고 있다. 즉, 여러 공유킥보드 업체가 국내 서비스를 시작하면서 그와 연관된 다양한 이슈들이 뉴스에서 다뤄진다는 것을 유추할 수 있다. 특히 서울이라는 단어는 꾸준히 높은 비율로 등장하고 2020년에 비율이 크게 증가했다.

Topic 3에서는 도로교통법, 면허, 사고, 이상, 착용 등 단어의 등장 확률은 해가 갈수록 증가하는데, 이는 전동킥보드 사용이 증가하면서 이와 함께 안전에 대한 이슈도 증가했기 때문으로 보인다. Topic 2에서도 규제, 허용 등과 같은 단어들이 높은 빈도로 등장하지만, Topic 3에서는 전동킥보드 관련 문제 중 특히 안전과 관련된, 안전모, 교통사고, 충돌, 위험 등의 단어가 높은 빈도로 등장하는 점에서 다른 토픽과 차별화된다.

Table 1. Probability of the Top 20 Terms of Topic 1

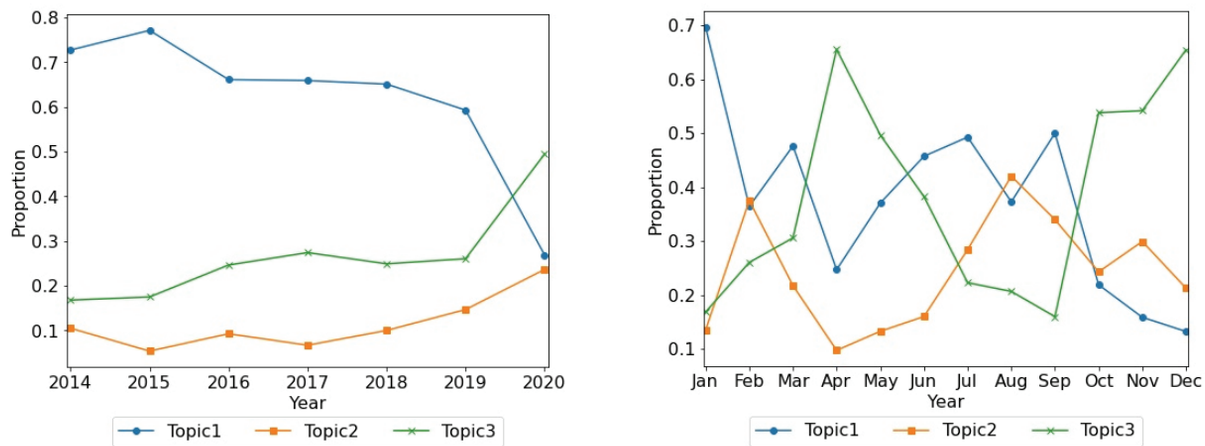
Term	2014	2015	2016	2017	2018	2019	2020	Avg	CAGR
Product	0.0490	0.0475	0.0430	0.0406	0.0326	0.0178	0.0112	0.0345	-0.2180
Mobility	0.0203	0.0227	0.0256	0.0243	0.0246	0.0369	0.0267	0.0259	0.0467
Service	0.0067	0.0068	0.0080	0.0117	0.0274	0.0626	0.0561	0.0256	0.4263
Electric car	0.0205	0.0229	0.0240	0.0354	0.0205	0.0078	0.0070	0.0197	-0.1649
Electric bicycle	0.0184	0.0177	0.0229	0.0289	0.0188	0.0159	0.0074	0.0186	-0.1409
Market	0.0163	0.0166	0.0179	0.0202	0.0198	0.0191	0.0142	0.0177	-0.0226
Domestic	0.0153	0.0166	0.0205	0.0208	0.0164	0.0162	0.0178	0.0177	0.0263
Charging	0.0220	0.0205	0.0181	0.0171	0.0165	0.0163	0.0113	0.0174	-0.1052
Battery	0.0152	0.0154	0.0160	0.0146	0.0163	0.0178	0.0185	0.0163	0.0331
Electricity	0.0130	0.0132	0.0146	0.0177	0.0251	0.0091	0.0048	0.0139	-0.1544
Possibility	0.0197	0.0195	0.0163	0.0129	0.0096	0.0075	0.0099	0.0136	-0.1080
Launch	0.0160	0.0163	0.0154	0.0140	0.0105	0.0080	0.0079	0.0126	-0.1105
Sale	0.0157	0.0159	0.0160	0.0159	0.0122	0.0050	0.0040	0.0121	-0.2038
Use	0.0072	0.0074	0.0081	0.0092	0.0118	0.0161	0.0233	0.0119	0.2164
Offer	0.0084	0.0084	0.0087	0.0090	0.0114	0.0159	0.0178	0.0114	0.1336
Development	0.0169	0.0151	0.0113	0.0089	0.0077	0.0076	0.0092	0.0110	-0.0962
Driving	0.0142	0.0147	0.0125	0.0106	0.0095	0.0069	0.0070	0.0108	-0.1105
Car	0.0149	0.0137	0.0119	0.0109	0.0100	0.0068	0.0049	0.0104	-0.1697
Smart	0.0130	0.0140	0.0161	0.0107	0.0074	0.0039	0.0041	0.0099	0.0009
Personal	0.0103	0.0110	0.0113	0.0106	0.0082	0.0069	0.0089	0.0096	-0.1750

Table 2. Probability of the Top 20 Terms of Topic 2

Term	2014	2015	2016	2017	2018	2019	2020	Avg	CAGR
Use	0.0329	0.0330	0.0333	0.0346	0.0351	0.0324	0.0287	0.0328	-0.0224
Preparation	0.0250	0.0249	0.0244	0.0233	0.0216	0.0199	0.0188	0.0226	-0.0469
Installation	0.0208	0.0205	0.0197	0.0186	0.0167	0.0149	0.0136	0.0178	-0.0684
Seoul	0.0159	0.0160	0.0156	0.0148	0.0136	0.0143	0.0189	0.0156	0.0288
Parking	0.0106	0.0108	0.0113	0.0128	0.0154	0.0205	0.0277	0.0156	0.1742
Government	0.0189	0.0182	0.0174	0.0168	0.0166	0.0124	0.0074	0.0154	-0.1435
Regulation	0.0125	0.0129	0.0141	0.0169	0.0220	0.0175	0.0057	0.0145	-0.1238
Operation	0.0174	0.0177	0.0179	0.0155	0.0129	0.0106	0.0076	0.0142	-0.1285
Vehicle	0.0105	0.0108	0.0115	0.0129	0.0149	0.0181	0.0208	0.0142	0.1217
Personal	0.0094	0.0095	0.0100	0.0111	0.0130	0.0169	0.0217	0.0131	0.1501
Promote	0.0133	0.0133	0.0131	0.0122	0.0112	0.0109	0.0138	0.0125	0.0061
Plan	0.0127	0.0126	0.0126	0.0122	0.0119	0.0105	0.0104	0.0118	-0.0318
Company	0.0092	0.0094	0.0096	0.0104	0.0115	0.0130	0.0162	0.0113	0.0990
User	0.0088	0.0089	0.0091	0.0097	0.0104	0.0123	0.0186	0.0111	0.1331
Service	0.0098	0.0098	0.0099	0.0106	0.0128	0.0135	0.0069	0.0105	-0.0568
Problem	0.0098	0.0099	0.0101	0.0103	0.0103	0.0095	0.0108	0.0101	0.0174
Schedule	0.0111	0.0111	0.0108	0.0096	0.0084	0.0081	0.0085	0.0097	-0.0436
Area	0.0095	0.0096	0.0099	0.0096	0.0094	0.0091	0.0068	0.0092	-0.0549
Permission	0.0090	0.0092	0.0097	0.0100	0.0112	0.0099	0.0045	0.0091	-0.1115
Business	0.0091	0.0092	0.0094	0.0095	0.0104	0.0099	0.0058	0.0090	-0.0716

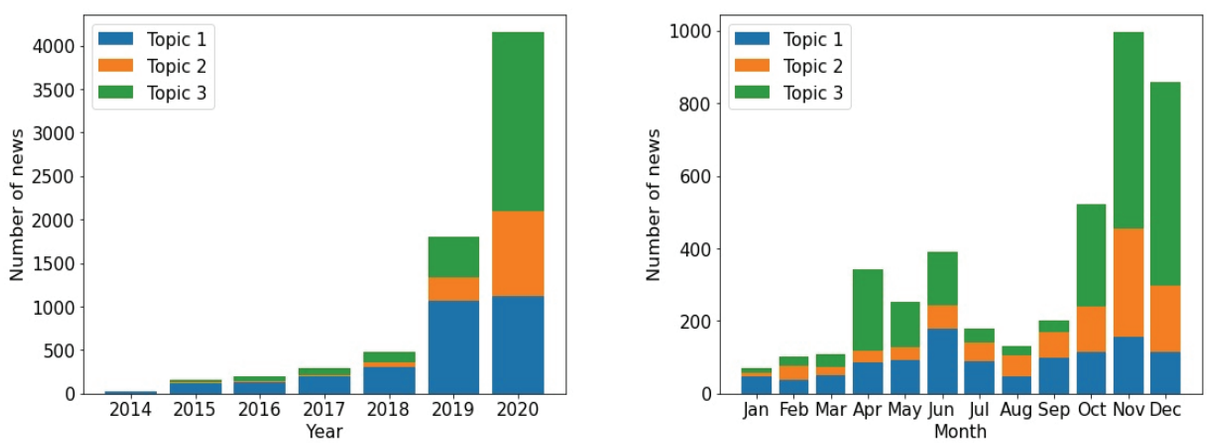
Table 3. Probability of the Top 20 Terms of Topic 3

Term	2014	2015	2016	2017	2018	2019	2020	Avg	CAGR
Wheel	0.0590	0.0665	0.0633	0.0254	0.0112	0.0047	0.0021	0.0332	-0.4249
Transportation	0.0192	0.0194	0.0228	0.0333	0.0241	0.0199	0.0172	0.0223	-0.0176
Use	0.0239	0.0236	0.0236	0.0208	0.0177	0.0181	0.0277	0.0222	0.0253
Footway	0.0208	0.0203	0.0190	0.0192	0.0195	0.0174	0.0114	0.0182	-0.0950
Road traffic law	0.0145	0.0143	0.0142	0.0146	0.0151	0.0166	0.0220	0.0159	0.0722
Personal	0.0119	0.0125	0.0140	0.0164	0.0167	0.0168	0.0189	0.0153	0.0800
Road	0.0155	0.0158	0.0161	0.0157	0.0164	0.0163	0.0106	0.0152	-0.0612
Driving	0.0123	0.0125	0.0134	0.0154	0.0190	0.0186	0.0113	0.0146	-0.0130
License	0.0155	0.0151	0.0147	0.0141	0.0133	0.0133	0.0159	0.0145	0.0040
Police	0.0149	0.0146	0.0142	0.0140	0.0136	0.0137	0.0136	0.0141	-0.0145
People	0.0160	0.0159	0.0158	0.0155	0.0142	0.0105	0.0070	0.0136	-0.1290
Operation	0.0123	0.0125	0.0138	0.0149	0.0159	0.0137	0.0114	0.0135	-0.0126
User	0.0116	0.0116	0.0117	0.0122	0.0135	0.0197	0.0119	0.0132	0.0040
Accident	0.0112	0.0114	0.0121	0.0136	0.0149	0.0150	0.0127	0.0130	0.0217
Occur	0.0120	0.0119	0.0121	0.0131	0.0141	0.0146	0.0118	0.0128	-0.0027
Abnormality	0.0127	0.0121	0.0111	0.0099	0.0090	0.0106	0.0229	0.0126	0.1038
Crackdown	0.0138	0.0137	0.0130	0.0124	0.0105	0.0085	0.0087	0.0115	-0.0736
Insurance	0.0145	0.0136	0.0119	0.0103	0.0094	0.0094	0.0103	0.0114	-0.0550
Wear	0.0087	0.0087	0.0089	0.0098	0.0113	0.0131	0.0125	0.0104	0.0624
Driveway	0.0111	0.0112	0.0112	0.0116	0.0121	0.0096	0.0061	0.0104	-0.0941



(a) Yearly (b) Monthly: 2020

Fig. 4. The Proportion of Each Topic



(a) Yearly (b) Monthly: 2020

Fig. 5. The Number of News by Each Topic

Fig. 4 - (a)는 연도별로 각 토픽의 비중을 보여주고 있다. Topic 1은 2015년 이후로 서서히 감소하기는 하지만 2019년까지는 가장 높은 비중을 차지했다. 그렇지만 2020년에는 그 비중이 급격히 감소했다. 반면에 Topic 2는 분석 기간동안 계속 가장 낮은 비중을 차지하지만, 증가하는 추세를 보이며, Topic 3은 2019년까지는 두 번째로 높은 비중을 차지하다 2020년에는 가장 높은 비중을 차지했다. 이는 2020년에 개정된 도로교통법이 개인형 이동장치에 규정이 포함되면서, 개인형 이동장치에 해당하는 전동킥보드 관련 기사가 급증했기 때문으로 추측된다. 이는 Fig. 4-(b)에서 볼 수 있듯이 2020년에 Topic 3의 비중이 갑자기 증가한 시점이 도로교통법이 개정이 이뤄진 시기와 개정된 법규가 적용되는 시점과 일치하는 것으로 뒷받침된다. 다만, 2020년에 Topic 1의 비중이 크게 감소한 것은 이와 관련된 뉴스 기사 수가 감소했기 때문이라기보다는 Fig. 5-(a)에서 알 수 있듯이 Topic 2와 Topic 3에 대한 기사가 2020년에 급증했기 때문이다. 그리고 Fig. 5-(b)를 보면, Topic 3의 비중은 도로교통법 개정 때가 약간 더 높지만 뉴스 기사의 수는 개정된 법규가 시행되는 시점에 더 많은 것을 알 수 있다.

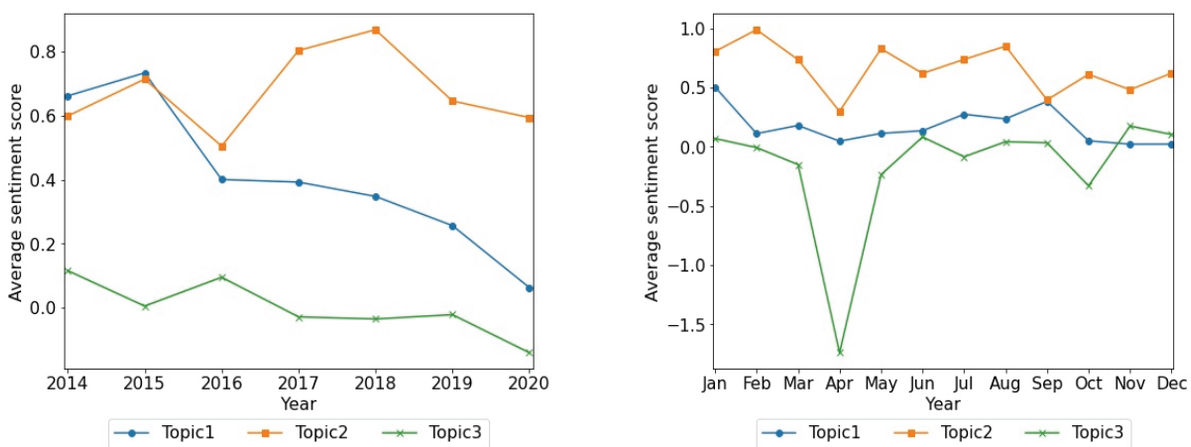
4.3 감성 분석 결과

5-겹 교차 검증으로 얻어진 최적의 하이퍼 파라미터는 는 0.01, 의 비중은 0.5로 얻어졌다. 그리고 이때 5-겹 교차 검증 정확도는 88.45%로 나타났다. 이 정도 성능은 감성사전을 구축하는 데 큰 문제가 되지 않으리라 판단하여 최적의 하이퍼 파라미터를 이용해 추출된 전체 데이터를 이용해 다시 학습시켜 최종 모델의 계수를 단어의 감성점수로 활용하여 감성사전을 구축했다.

Table 4는 최종적으로 얻어진 감성사전에서 감성점수의 크기가 상위 10위 안에 드는 긍정어와 부정어를 보여주고 있다. 긍정 단어에는 도움, 탑재, 신속 등과 같은 단어가 포함되어 있으며 부정 단어에는 위협, 불만, 혐의 등의 단어가 포함

Table 4. Top 10 Positive and Negative Words

Positive	Score	Negative	Score
help	4.0310	threat	-4.6996
equipped	3.1271	complaint	-4.3933
quick	2.9671	suspect	-3.7638
enjoy	2.6735	controversy	-3.4702
feature	2.4770	lead	-3.4225
thanks to	2.4027	illegal	-3.2866
design	2.3748	violation	-3.2362
prevention	2.3575	anxiety	-3.2078
ease	2.3336	loss	-2.9650
prospect	2.3182	critical	-2.9541



(a) Yearly (b) Monthly: 2020

Fig. 6. Sentiment Scores of Each Topic

되어 있다. 상위 10위 안에 들지 않는 단어까지 고려해서 살펴보면 신속, 즐기다, 손쉽다, 편리와 같은 긍정어와 느리다, 불편 등과 같은 부정어가 있어 이동의 편의성, 전동킥보드 사용성과 관련된 요소가 뉴스의 감정에 큰 영향을 끼치는 것으로 보인다. 또한, 부정어에는 위협, 불법, 침해, 불안, 혼란 등과 같이 안전, 주차 등과 연관된 요소와 관련된 단어가 포함되어 있으며 이 요소는 주로 부정적인 감정과 더 밀접한 관련이 있는 것으로 보인다.

Fig. 6은 Equation (2)로 구한 연도별 토픽별 감성점수와 2020년 월별 토픽별 감성점수를 보여주고 있다. 값이 양수로 클수록 해당 토픽에 대한 사회적 감성이 긍정적이고, 반대로 값이 음수로 크면 해당 토픽에 대한 사회적 감성이 부정적임을 뜻한다. 그리고 0에 가까운 값을 가지면 해당 토픽에 대한 감성이 중립에 가깝다는 뜻이다. Fig. 6 - (a)를 보게 되면 Topic 1과 Topic 3은 전반적으로 감성점수가 감소하는 추세인 반면, Topic 2는 공유킥보드 서비스가 국내에 처음 등장한 2018년에 가장 높은 감성점수를 보이다 다시 하락하는 걸 볼 수 있다. 다만, Topic 1은 감성점수가 높았다가 중립 상태로 바뀐 것이지만, Topic 3의 감성점수는 2014년에도 높지 않았는데 2017년 이후에는 부정적으로 바뀌었다. 정리하면, 전동킥보드가 국내에 출시된 초기에는 전동킥보드 기술과 관련된 긍정적인 뉴스 기사가 많았지만 전동킥보드가 활성화된 후에는 그런 기사가 점차 줄어들었고, 안전에 대한 이슈는 전동킥보드 이용이 활성화된 이후에 불거져 그와 관련된 부정적인 기사가 많아진 것으로 추측할 수 있다.

Fig. 6 - (b)를 보게 되면 도로교통법이 개정되기 직전인 4월에 Topic 3에 대한 감성점수가 크게 낮은 것을 확인할 수 있다. 마찬가지로 같은 시기에 Topic 2에 대한 감성점수도 하락했다. Topic 3에 대한 감성점수는 개정된 도로교통법이 시행되기 전인 10월에 다시 낮아졌다가 11~12월에는 10월보다 높은 감성점수를 보이는데 개정된 도로교통법에 대한 논란이 가중되면서 12월에 개인형 이동장치에 대한 주의의무를 강화하는 방향으로 다시 도로교통법이 재개정되면서 규정에 대한 부정적인 의견이 줄어든 것으로 보인다.

5. 결 론

본 연구는 뉴스 기사를 대상으로 토픽 모델링과 감성 분석을 결합하여 전동킥보드와 관련된 이슈의 동향을 분석하였다. 일반적으로 토픽 모델링을 활용한 동향 분석에서는 특정 분야에서의 확인할 수 있는 토픽을 추출하는데 그치는 경우가 다수인데, 이 연구에서는 토픽 모델링 이후에 감성 분석까지 진행하여 개별 이슈에 대한 사회적 감정이 어떻게 되는지까지 분석하였다. 또한, 평점이 존재하는 리뷰와는 달리 뉴스 기사는 긍/부정 여부를 판별할 수 없으므로 KOSAC 감성사전을 이용해서 확실하게 긍정 또는 부정인 기사를 추출하고 elastic net 정규화 로지스틱 회귀를 이용해서 전동킥보드에 특화된 감성사전을 구축하여 뉴스의 감성점수를 계산했다.

토픽 모델링 결과 전동킥보드에 대한 기술, 킥보드 업체 및 공유킥보드 서비스, 안전 및 규제 토픽을 얻을 수 있었다. 전동킥보드 도입 초기에는 뉴스에서 기술에 대한 비중이 높았으나 공유킥보드 서비스가 등장하면서 이에 대한 비중과 함께 안전과 규제에 대한 비중이 급격히 증가했다. 그리고 감성 분석 결과 2014~2015년에는 전동킥보드 관련 기술과 업체에 대해서는 긍정적인 기사가 많았다. 기술과 관련된 토픽에 대해서는 그 이후에는 서서히 감성점수가 감소하였나, 킥보드 업체 및 공유킥보드 서비스에 대해서는 이후에도 대체로 긍정적인 반응을 보이고 있다. 안전 및 규제에 대해서는 나머지 두 토픽에 비해 낮은 감성점수를 보이는데 그마저도 해가 갈수록 낮아지고 있다.

이를 통해 전동킥보드 기술이나 관련 사업, 전동킥보드로 인한 이동의 편의성에서는 대체로 긍정적인 반응이 많지만, 전동킥보드 이용이 논란을 야기하지 않고 활성화되려면 주차, 안전과 같은 문제를 해결할 수 있는 방안과 적절한 규제 마련이 필요하다는 것을 알 수 있다. 특히 2020년 5월의 도로교통법개정에서는 부정적 감정이 두드러져, 사회 전반적으로는 안전을 위해 규제 강화를 원하는 것을 알 수 있다.

추가로, 본 연구에서 사용한 동적 토픽 모델링과 감성 분석을 결합합 분석 프레임워크는 다양한 사회적 문제에 대해 시간에 따라 바뀌는 이슈 상황을 파악하는데 적용할 수 있어 여러 분야에서 활용할 수 있다.

본 연구는 뉴스의 감성점수를 얻기 위해 이미 구축되어 있는 감성사전을 사용하였으며, 전동킥보드에 특화된 감성사전 또한 KOSAC 감성사전을 바탕으로 얻어진 감성점수를 활용하였다는 한계점이 있다. 추후 전동킥보드 제품에 대한 리뷰나 공유킥보드 서비스에 대한 리뷰 등을 확보하여 추가 분석을 수행한다면 보다 정확한 감성사전을 구축할 수 있을 것이다. 또한, 이 연구에서는 사회적 감성이 전동킥보드나 공유킥보드 이용에 끼치는 영향을 직접적으로 확인하지는 못했다. 킥보드 이용 데이터를 확보한다면, 사회적 감성이 공유킥보드 이용에 끼치는 영향력 또한 밝힐 수 있을 것이다.

References

[1] S. Sultana, D. Salon, and M. Kuby, “Transportation sustainability in the urban context: A comprehensive review,” *Urban Geography*, Vol.40, No.3, pp.279-308, 2019.

[2] S. Jäppinen, T. Toivonen, and M. Salonen, “Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach,” *Applied Geography*, Vol.43, pp.13-24, 2013.

[3] B. Madapur, S. Madangopal, and M. N. Chandrashekar, “Micro-Mobility Infrastructure for Redefining Urban Mobility,” *European Journal of Engineering Science and Technology*, Vol.3, No.1, pp.71-85, 2020.

[4] G. McKenzie, “Urban mobility in the sharing economy: A spatiotemporal comparison of shared mobility services,” *Computers, Environment and Urban Systems*, Vol.79, pp.101418, 2020.

[5] J. K. Mathew, M. Liu, and D. M. Bullock, “Impact of Weather on Shared Electric Scooter Utilization,” *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp.4512-4516, 2019.

[6] NACTO, “Shared Micromobility in the U.S.: 2019,” 2020.

[7] Gyeonggi Research Institute. “Smart mobility services: Issue and policy implications,” 2020.

[8] M. J, Park, “A controversy over whether the Scooter-sharing Service is an Innovation or not: 2,000 Complaints Have Been Filed due to Abandoned Scooters,” Korea Joong-Ang Daily Economy, p.1, 2020.

[9] H. H. Jo, H. S. Noh, H. C. Yoo, J. Kang, J. Jung, H. S. Kim, “A study on the use behavior and safety of electric scooters - Focused on the survey of e-scooter owners,” *Geographical Journal of Korea*, Vol.55, No.1, pp.43-55, 2021.

[10] S. Y. Ko, “A study on the change of movement environment according to increase of personal mobility focusing on the road environment of highway safety and drivers using secure walkers,” *Korea Institute of Design Research Society*, Vol.2, No.3, pp.9-17, 2017.

[11] T. Eccarius and C. C. Lu, “Adoption intentions for micro-mobility-Insights from electric scooter sharing in Taiwan,” *Transportation Research Part D: Transport And Environment*, Vol.84, pp.102327, 2020.

[12] H. Fitt and A. Curl, “The early days of shared micromobility: A social practices approach,” *Journal of Transport Geography*, Vol.86, pp.102779, 2020.

[13] J. J. Aman, J. Smith-Colin, and W. Zhang, “Listen to E-scooter riders: Mining rider satisfaction factors from app store reviews,” *Transportation Research Part D: Transport and Environment*, Vol.95, pp.102856, 2021.

[14] Y. Feng, et al., "Micromobility in smart cities: A closer look at shared dockless E-Scooters via big social data," *ICC 2021-IEEE International Conference on Communications*, IEEE, 2021.

[15] H. S. Lee, K. H. Baek, J. H. Jeong, and J. H. Kim, "User's behaviors of smart personal mobility sharing services: Emperical evidence from electric scooter sharing service," *Proceedings of the KOR-KST Conference*, pp.462-463, 2019.

[16] S. J. Kim, S. Choo, and S. H. Kim, "A study on the usage pattern of E-Scooter sharing service," *Proceedings of the KOR-KST Conference*, pp.344-345, 2020.

[17] KISO Planning Team, "Domestic trends and expected effects of 'Shared Electric Kickboards'," 2019. [Internet], https://journal.kiso.or.kr/?p=9850.

[18] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, Vol.3, pp.993-1022, 2003.

[19] S. A. Jin, G. E. Heo, Y. K. Jeong, and M. Song, "Topic-network based topic shift detection on twitter," *Journal of the Korean Society for information Management*, Vol.30, No.1, pp.285-302, 2013.

[20] D. M. Blei and J. D. Lafferty, "Dynamic topic models," *Proceedings of the 23rd International Conference on Machine Learning*, 2006.

[21] A. Daud, "Using time topic modeling for semantics-based dynamic research interest finding," *Knowledge-Based Systems*, Vol.26, pp.154-163, 2012.

[22] H. Zhang, G. Kim, and E. P. Xing, "Dynamic topic modeling for monitoring market competition from online text and image data," *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015.

[23] T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," *2018 IEEE International Conference on Innovative Research and Development (ICIRD)*, IEEE, 2018.

[24] L. Jiang and Y. Suzuki, "Detecting hate speech from tweets for sentiment analysis," *2019 6th International Conference on Systems and Informatics (ICSAI)*. IEEE, 2019.

[25] D. Tumitan and K. Becker, "Sentiment-based features for predicting election polls: A case study on the brazilian scenario," *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. 2. IEEE, 2014.

[26] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational linguistics*, Vol.37, No.2, pp.267-307, 2011.

[27] M. Z. Asghar, A. Khan, S. Ahmad, M. Qasim, and I. A. Khan, "Lexicon-enhanced sentiment analysis framework using rule-based classification scheme," *PloS one*, Vol.12, No.2, pp.e0171649, 2007.

[28] K. Erk, "Vector space models of word meaning and phrase meaning: A survey," *Language and Linguistics Compass*, Vol.6, No.10, pp.635-653, 2012.

[29] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," *Proceedings of the 31st International Conference on Machine Learning*, Vol.32, pp. 1188-1196, 2014.

[30] G. Rao, W. Huang, Z. Feng, and Q. Cong, "LSTM with sentence representations for document-level sentiment classification," *Neurocomputing*, Vol.308, pp.49-57, 2018.

[31] Z. Gao, A. Feng, X. Song, and X. Wu, "Target-dependent sentiment classification with BERT," *IEEE Access*, Vol.7, pp.154290-154299, 2019.

[32] M. Munikar, S. Shakya, and A. Shrestha, "Fine-grained Sentiment Classification using BERT," *Proceedings of the 2019 Artificial Intelligence for Transforming Business and Society (AITB)*, Vol.1, pp.1-5, 2019.

[33] S. Lee, H. Jang, Y. Baik, S. Park, and H. Shin, "Kr-bert: A small-scale korean-specific language model," *arXiv Prepr. arXiv2008.03979*, 2020.

[34] D. Hyun, J. Cho, and H. Yu, "Building large-scale english and korean datasets for aspect-level sentiment analysis in automotive domain," *Proceedings of the 28th International Conference on Computational Linguistics,* 2020.

[35] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, Vol.5, No.4, pp.1093-1113, 2014.

[36] T. Ito, K. Tsubouchi, H. Sakaji, T. Yamashita, and K. Izumi, "Contextual sentiment neural network for document sentiment analysis," *Data Science and Engineering*, Vol.5, No.2, pp.180-192, 2020.

[37] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis," *Future Generation Computer Systems*, Vol.115, pp.279-294, 2021.

[38] H. Jang, M. Kim, and H. Shin, "KOSAC: A full-fledged Korean sentiment analysis corpus," *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation*, pp.366-373, 2013.

[39] H. P. Shin, M. H. Kim, and S. Z. Park, "Modality-based sentiment analysis through the utilization of the Korean sentiment analysis corpus," *Journal of the Linguistic Society of Korea*, Vol.74, pp.93-114, 2016.

[40] S. B. Kim, S. J. Kwon, and J. T. Kim, "Building sentiment dictionary and polarity classification of blog review By using elastic net," *Proceedings of Korean Institute of Information Scientists and Engineers*, pp.639-641, 2015.

김 경 옥

https://orcid.org/0000-0002-0196-3832
e-mail : kyoungok.kim@seoultech.ac.kr
2008년 POSTECH 신소재공학과(학사)
2013년 POSTECH 산업경영공학과
(Ph.D., 석·박사통합)
2013년 ~ 2015년 삼성경제연구소
산업전략1실 선임, 책임연구원
2015년 ~ 현 재 서울과학기술대학교 산업공학과 부교수
관심분야 : Transport Data Analysis, Machine Learning

신 예 랑

https://orcid.org/0000-0001-9118-9934
e-mail : fnel99@ds.seoultech.ac.kr
2019년 충남대학교 영어영문학과(학사)
2021년 서울과학기술대학교
데이터사이언스학과 석사수료
관심분야 : Data Mining, Text Mining