

텍스트 마이닝 기법을 활용한 자율주행자동차 인식분석연구

Analysis of the Perception of Autonomous Vehicles Using Text Mining Technique

임 이 정* · 송 재 인** · 이 자 영*** · 황 기 연****

* 주저자 : 홍익대학교 도시계획과 박사수료
 ** 공저자 : 홍익대학교 도시계획과 박사과정
 *** 공저자 : 홍익대학교 도시계획과 석사과정
 **** 교신저자 : 홍익대학교 도시공학과 교수

I-Jeong, Im* · Jae-In, Song** · Ja-Young, Lee*** · Kee-Yeon, Hwang****

* Dept. of Urban Planning, Univ. of Hongik
 ** Dept. of Urban Planning, Univ. of Hongik
 *** Dept. of Urban Planning, Univ. of Hongik
 **** Dept. of Urban Planning, Univ. of Hongik

† Corresponding author : Kee-Yeon Hwang, keith@hongik.ac.kr

Vol.16 No.6(2017)
 December, 2017
 pp.231~243

ISSN 1738-0774(Print)
 ISSN 2384-1729(On-line)
<https://doi.org/10.12815/kits.2017.16.6.231>

Received 15 November 2017
 Revised 24 November 2017
 Accepted 8 December 2017

© 2017. The Korea Institute of
 Intelligent Transport Systems. All
 rights reserved.

요 약

자율주행자동차는 미래 교통수단으로써 주목받고 있으며, 전 세계적으로 관련 기술 개발 및 수용성 연구가 진행되고 있다. 그러나 자율주행자동차와 같은 신기술의 수용에 대한 인식 조사는 부족한 실정이다. 이에 본 연구에서는 자율주행자동차의 법제화와 상용화를 위한 기반 조성 작업의 일환으로 인터넷 기사와 댓글을 활용하여 텍스트 마이닝 기법 중 감성평가 기법을 적용하여 자율주행자동차에 대한 시민들의 인식분석연구를 수행하였다. 분석 결과, 자율주행자동차에 대한 긍정적인 시각도 있으나 부정적인 인식이 더 큰 것으로 나타났으며, 대부분 유보적인 판단을 내리는 것으로 나타났다. 이는 자율주행자동차 기술의 불확실성, 탑승 경험 부족으로 인한 것이라 판단되었으며, 사회적 수용성 향상을 위해 도입 이전 해소되어야 할 문제가 사료된다. 또한 지속적인 인식조사 및 설문조사를 통해 사회적 수용성을 확보할 방안을 강구해야할 필요가 있을 것으로 판단된다.

핵심어 : 자율주행자동차, 인식분석, 텍스트 마이닝, 감성분석, 사회적 수용성

ABSTRACT

The purpose of this study is to improve the social acceptance of AVs by analyzing the citizen's perception using an emotional analysis technique which belongs to a type of text mining. The source of the data is originated from 3 year accumulated internet articles and comments on AV from 164 newspapers and Naver. According to the study results, there exists a positive perception on AVs, although negative ones are more frequent than the positive. Also most of people take neutral position on AV due to the unfamiliarity and lack of experience on AVs And these problems needs to be responded before AV's commercialization through continuous analyses on the perception and social acceptance.

Key words : Autonomous Vehicle, Analysis of Perception, Text mining, Emotional Analysis, Social Acceptance

I. 서론

1. 개요

자율주행자동차는 현 도로교통 및 사회적 문제를 해결할 미래 교통수단으로 전 세계적으로 자율주행자동차 및 관련 기술 개발을 위한 국가 차원의 연구가 진행되고 있다. 우리나라의 경우, 2015년 ‘자율주행차량 상용화 지원 방안’을 토대로 2020년까지 레벨 3수준의 자율주행자동차를 상용화하려는 계획을 발표하였다.

자율주행자동차의 도입은 교통사고 및 혼잡 감소 등 교통 부분의 문제를 해결하고, ITS 및 정보통신 분야의 신기술이 접목된 미래 교통수단으로써 다양한 사회적 이익을 창출할 것으로 예상되고 있다. 그러나 국외 자율주행자동차 개발 선도 업체인 구글이 진행한 자율주행자동차 실험 및 테슬라의 오토파일럿 운행 중 자율주행자동차의 충돌사고가 발생하면서 이에 따른 사회적 수용성 문제가 대두되고 있다. 자율주행자동차의 도입으로 인해 발생 가능한 사회적 수용성 문제는 자율주행자동차의 기능과 운행 영향의 불확실성, 사고 발생 전 자율주행자동차의 윤리적 판단에 대한 수용 여부, 사고 발생 후 사고 책임소재 규명 등으로 정리할 수 있다.

스마트폰과 소셜 네트워크 매체(Social Network Service, SNS)의 보급은 온라인 상에서 다양한 사람들의 정보 교환을 편리하게 만들었다. 특히 인터넷 뉴스 기사와 소셜 네트워크 매체 간의 연계가 실현되면서 이용자들이 더 많은 정보를 공유하는 것이 용이해졌다. 자율주행자동차에 대한 국내외적인 관심도가 높아지면서 관련 기사와 동영상 등의 정보의 양이 날로 증대하고 있으며, 인터넷을 통해 어디서든 쉽게 이용자들이 관련 정보를 습득하고 공유할 수 있는 환경이 마련되어져 있다.

자율주행자동차의 도입에 앞서, 자율주행자동차와 같은 신기술에 대한 사회적 수용성을 향상시키기 위해서는 사전에 시민들의 인식조사가 필요한 상황이나 이와 관련된 연구는 미흡한 실정이다. 따라서 본 연구에서는 자율주행자동차에 대한 인식조사를 위하여 관련 인터넷 뉴스 기사와 해당 기사의 댓글을 활용한 텍스트 마이닝 분석을 통해 자율주행자동차에 대한 시민들의 인식을 분석하고자 한다.

2. 연구의 방법

본 연구에서는 자율주행자동차 관련 인터넷 뉴스 기사와 해당 기사의 댓글 활용한 텍스트 마이닝 분석을 통해 자율주행자동차에 대한 시민들의 인식을 분석하고자 하였다. 첫 번째로 자율주행자동차의 정의와 인식 관련 선행연구와 본 연구에서 활용된 텍스트 마이닝 분석 중 감성 분석을 수행한 선행연구를 고찰하였다. 두 번째로 텍스트 마이닝 분석 틀을 설정하였으며, 마지막으로 텍스트 마이닝을 활용한 인식 분석을 수행하였으며 분석결과를 도출하였다.

II. 선행연구 고찰

1. 자율주행차량(Autonomous Vehicle, AV) 인식 연구

자율주행자동차(Autonomous Vehicle 또는 Automated Vehicle)는 「자동차관리법」 제2조 1의3에 법적으로 정의되어 있으며, 운전자 또는 승객의 조작 없이 자동차 스스로 운행이 가능한 자동차를 의미한다. 최근 자율주행 자동차의 개발수준이 높아짐에 따라 관련한 연구 또한 증가하고 있으나 기술개발, 운행 영향, 법규, 윤리 등의 연구에 비해 자율주행 자동차의 인식과 관련한 연구는 다소 부족하였다. 본 장에서 자율주행

자동차의 인식과 관련한 선행연구를 고찰하였다.

Lee et al.(2015)는 자율주행 자동차 도입이 교통부문에 미치는 영향에 대해 검토하고, 이용자 의식 및 선호도 조사를 통해 교통계획분야의 대응과제를 도출하였다. 이용자 의식조사 결과 자율주행 자동차를 도입할 때의 편익으로 운전 피로도 감소, 운전 이외의 활동, 교통안전의 순이었으며 시스템 오류 및 보안문제를 가장 우려하는 것으로 나타났다. 자율주행 자동차에 대한 지불 의사 금액은 자율주행 기술수준이 증가함에 따라 커졌으나, 증가액은 크지 않는 것으로 분석되었다. Kim et al.(2016)은 자율주행 자동차와 관련된 윤리와 운행 책임을 정리하고, 새로운 기술에 대한 운전자 수용성 분석 및 개선방안을 도출하였다. 자율주행 자동차의 윤리 실험과 집단별(전문가, 비전문가) 운전자 수용성 조사를 진행하였으며, 이용 요인분석을 위해 다중회귀모형을 활용한 자율주행자동차 이용 모형을 개발하였다. 수용성 조사결과 3, 4단계에서 응답자의 절반 정도가 이용 의사를 밝혀 현재까지 자율주행자동차의 수용성이 높지 않은 것으로 나타났으나, 회의적인 견해를 보이는 응답자가 15% 이내로 향후 정책 및 수용성 향상에 대한 노력에 따라 변동의 여지가 클 것으로 분석하였다.

2. 텍스트 마이닝(Text mining) 및 감성평가 연구

데이터 마이닝(Data mining)의 한 분야인 텍스트 마이닝(Text mining)은 비정형데이터를 처리하는 자연어 처리(Natural Language Processing)기반 기술로 국내에서는 1998년부터 관련 연구가 시작되었으며, 2011년 이후 빅데이터 기반의 연구가 활발해지면서 관련 연구가 지속적으로 증가하고 있는 추세이다(Kim et al., 2016).

교통분야에서도 텍스트 마이닝을 활용한 연구들이 있으며 Oh et al.(2016)은 경계분석이론과 텍스트마이닝을 통해 빅데이터의 정책이슈를 탐색하는 연구를 진행하였다. 연구 중 텍스트 마이닝 부분에서 LDA (Latent Dirichlet Allocation)모형과 비대칭-대칭 어휘소기반 LDA 응용을 통해 ITS 관련 뉴스, 문헌 등을 분석하였고 이를 통해 정책이슈를 탐색하는 기법을 정립하였다. Oh(2015)는 교통 ICT 분야의 연구동향을 정량적으로 분석하고, 교통분야의 IoT 활성화를 위해 필요한 연구영역을 도출하였다. 해당 연구에서는 기존 교통분야의 연구자료를 토대로 동시단어 출현기반 분석과 토픽모델링을 통해 텍스트 마이닝을 진행하였다. Jung et al. (2013)은 SNS 기반 교통안전 개선프로그램을 분석 및 운영함에 있어서 실제 주민이 느끼는 위험을 파악하기 위해 텍스트 마이닝 기법을 활용하였다. 설문조사의 결과를 온톨로지 기법과 워드클라우드를 통해 시각화 분석을 수행하여 주요 위험요소를 도출하였다. 이처럼 교통 분야의 텍스트 마이닝 관련 연구들은 주로 텍스트 노출 빈도에 따라 이슈를 도출하는 것에 그 목적이 있다고 볼 수 있다.

주요 텍스트 마이닝 기법으로는 감성평가(오피니언 마이닝(Opinion Mining) 또는 감성분석), 토픽모델링(Topic Modeling), 시맨틱 웹(Semantic Web) 및 온톨로지 기법(Ontology) 등이 활용되고 있으며, 이 중 감성평가 방법론이 가장 많이 활용되고 있다(Kim et al., 2013). 감성평가 방법론은 텍스트에서 긍·부정의 의견을 판단하여 사회 현상 분석을 위해 관련 결과를 활용하는 목적으로 사용된다(Jang et al., 2015). 본 연구에서 자율주행 자동차에 대한 긍·부정 의견 판단하기 위해 감성평가 방법론 선택하였으며 이에 대한 선행연구를 고찰하였다.

Lee et al.(2013)은 추가 정보 빅데이터를 활용하여 9가지 감성을 분류하고 이들이 갖고 있는 특징을 실증분석 하였다. 9가지 감정의 주성분 분석을 통해 긍정, 부정, 중립의 3가지로 분류하고, 이를 VAR(Vector Auto-Regression)분석으로 상호관계를 확인하였다. 분석결과 추가 정보를 담고 있는 감성들이 주식시장의 흐름과 상관관계가 있음을 확인하였다. Lee(2013)은 신문기사의 키워드와 트위터 데이터를 이용하여 감성분석을 진행하였으며, 이를 통해 최근 여론 동향을 파악하는 연구를 수행하였다. K-mean 알고리즘을 통해 키워드

를 군집화 하고, 기계학습(Machine learning)을 활용하여 긍·부정 판단 및 타당성을 검증하였다. 연구결과에 따르면, 신문기사 키워드의 감성 분석결과와 화제성의 상관관계가 65.97% 인 것으로 분석되었으며, 이는 단순 키워드 분석 결과로 문맥을 분석할 경우 상관성이 향상될 것으로 판단하였다. Yang et al.(2015)은 트위터를 통해 국내대학과 관련한 빅데이터를 수집하고 RHive를 통해 감성분석을 수행하여 대학별 인지도에 대한 긍·부정을 판단하였다. 해당 연구는 SNS 데이터를 활용한 대학분석을 통해 대학에 대해 일반인이 갖고 있는 생각을 분석하였다는 의의가 있다. Yoon(2016)은 수집된 트위터 데이터를 활용하여 한국어 감성분석을 위한 알고리즘 및 분석시스템을 제안하였다. 범용적으로 사용되는 SentiWordNet DB를 한국어로 번역하여 감성평가에 이용하였으며, 키워드의 극성계산 알고리즘을 통해 키워드의 긍·부정을 분석하였다. 해당 연구에서 제안된 감성분석의 정확도는 78.4%로 나타났으며 기존 알고리즘이 키워드 분석에 한정되어 있는 것에 비해 문장이 갖는 문맥상의 의미를 파악하였다는 것에 의의가 있다.

감성평가의 장점은 특정 상황에서 사용된 단어와 문장에 대한 감성 분석을 통해 긍정적인 요소와 부정적인 요소를 판단하여 이에 대한 분석을 수행할 수 있다는 점을 들 수 있다. 그러나 초기 감성평가의 경우, 국외에서 시작되었으며 영어를 중심으로 한 감성사전에 기반하여 한국어 매치를 통한 분석이 이뤄진다는 단점이 있다. 한국어의 경우, 단어만으로 감성평가를 하기 보다는 전체적인 문장의 문맥을 통해서 사용된 단어의 극성을 판단해야한다는 단점이 있다. 또한 한 문장 내에서도 다양한 감성(긍·부정)을 동시에 담을 확률이 높아 극성 판단이 어려운 판단이 빈번하여 중립의견 처리 비율이 높을 수 있다. 영어와 같이 단어가 가지고 있는 극성이 분명하지 않아, 긍정적인 단어이더라도 부정적으로, 부정적인 단어이더라도 긍정적으로 사용될 수 있기 때문에 중립으로 분석되는 경우가 많다. 따라서 감성평가를 수행하는데 있어서 단어의 문맥을 분석하여 극성을 분석하는 것이 중요하다고 할 수 있다.

3. 기존 연구와의 차별성

선행연구를 검토한 결과, 자율주행 자동차의 인식에 대한 연구와 교통 분야에서 텍스트 마이닝 기법을 활용한 연구는 부족한 것으로 나타났으며, 감성평가 방법론을 활용한 연구가 다양한 분야에서 진행되고 있는 것을 확인하였다. 따라서 본 연구에서는 감성평가 방법론을 활용하여 인터넷 뉴스 기사 및 기사의 댓글을 통해 자율주행자동차에 대한 시민들의 인식을 분석한 것에 차별성이 있다.

Ⅲ. 연구방법론

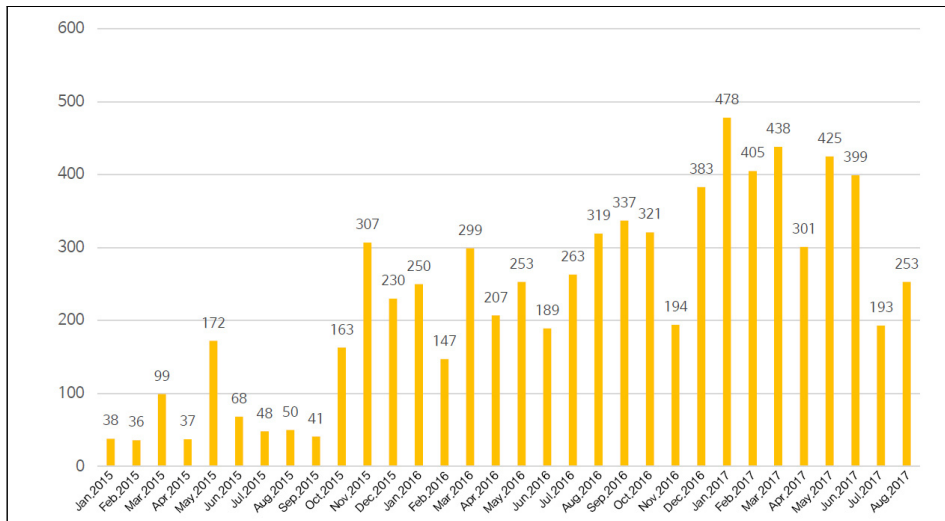
1. 분석의 틀

분석을 위해 데이터를 2015년 1월부터 2017년 8월까지 총 32개월간의 뉴스 기사 및 댓글 데이터를 수집하였다. 수집 범위는 라이브리 기능을 설치한 국내 언론사 164개 웹사이트의 기사 및 댓글과 네이버 뉴스 섹션의 기사 및 댓글로 한정하여 수집하였다. 수집 내용은 ‘자율주행’ 키워드를 중심으로 제목에 키워드가 포함된 기사 수, 해당 기사의 댓글, 해당 키워드가 언급된 댓글로 앞선 내용과는 중복되지 않도록 하였으며 <Table 1>과 같이 정리하였다.

<Table 1> Summary of Data Collection

Collection	Description
Period	• 2015. 01 ~ 2017. 08 (32 months)
Scope	• Articles and comments on 164 web sites of domestic press which installed Libraries • Articles and comments on news section of Naver
Contents	• Data related with 'Autonomous' (keyword) 1) The number of Articles which include 'Autonomous' on title 2) The number of Comments which include 'Autonomous' on articles title 3) The number of Comments which include 'Autonomous' keyword (Except for overlapping above data.)
Analysis contents	• Analysis of positive and negative keyword of all articles and comments about 'Autonomous' vehicle

텍스트 마이닝 분석을 위해 데이터를 수집하여, '자율주행' 키워드를 포함한 자율주행자동차 관련 기사 발행 추이를 확인하였다. '자율주행' 키워드가 기사 제목에 포함되어 발생한 기사 건수는 총 7,343건으로 집계되었으며, 댓글 유무와 상관없이 해당 키워드를 포함한 총 기사의 발행 추이는 다음과 같다.



<Fig. 1> The number of articles published

기사 분석의 경우, 우선 라이브리 서비스를 사용하는 164개 언론사와 네이버 뉴스 섹션에 노출된 모든 언론사 별 기사 데이터를 수집하였다. 이후 수집된 데이터를 기사 제목, 언론사, 기사 작성일, 기사 URL로 구분하여 정리하였다. 수집한 기사 목록의 예시는 다음의 <Fig. 2>와 같다.

Article's title	Media	published date	Article's URL
화성에 자율주행차 실험도시... 속구장 40배 규모	YTN	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=052&aid=0001053116
세계 최대규모 K-CITY 자율주행차 상용화 앞당긴다!	TV조선	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=448&aid=0000220889
자율주행차 위한 3차원 디지털지도 구축한다	한국일보	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=469&aid=0000231182
정부 기업 손발 척척 맞는 일본 도로에 자율주행 난이도 매긴다	한국경제	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=104&oid=015&aid=0003817483
장애물 피해가며 요리조리 척척 자율주행차 직접 운전해보니	중앙일보	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=102&oid=025&aid=0002750189
포드 자율주행차량, 피자 배달 서비스 시험	조선일보	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=103&oid=366&aid=0000381470
국토지리정보원 도로소, 정밀지도 구축 손잡았다. 자율주행 지원 박차	이데일리	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=018&aid=0003912229
자율주행차량으로 피자 배달합니다	오마이뉴스	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=004&oid=047&aid=0002160868
자율주행용 위한 3차원 디지털지도 구축 속도낸다	아시아경제	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=277&aid=0004065907
김현미 장관 자율주행차 시승	서울신문	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=081&aid=0002849656
스마트시티 자율주행차 국토부, 4차 산업혁명에 박차	머니투데이	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=008&aid=0003927419
지연스, 자율 주행 보폭 강화 연도 그래픽 이어 타스 인수	디지털타임스	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=105&oid=029&aid=0002419836
LG이노텍, 자율주행차용 '2세대 통신 모듈' 세계 첫 출시	동아일보	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=020&aid=0003091309
자율주행차 시험장 K-City 확공, 2018년 말 완공	데일리안	2017.08.31.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=019&aid=0002207508
인천 남동인더스파크에 자율주행버스 생키나	해럴드경제	2017.08.30.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=016&aid=0001283580
LG이노텍, 자율주행차용 통신 시스템 '2세대 V2X 모듈' 첫 선	한국경제TV	2017.08.30.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=215&aid=0000570741
LG이노텍, 자율주행차용 2세대 V2X 모듈 세계 첫 개발	파이낸셜뉴스	2017.08.30.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=014&aid=0003866053
LG이노텍, 자율주행차 모듈 세계 첫 개발	파이낸셜뉴스	2017.08.30.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=101&oid=014&aid=0003866633
'속구장 50배' 자율주행차 도시 조성	채널A	2017.08.30.	http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=103&oid=449&aid=0000136756

(Fig. 2) The example of collected articles

수집한 ‘자율주행’ 키워드 관련 데이터는 다음과 같다. 첫 번째로 ‘자율주행’이 제목에 포함되어 발행된 기사 건수는 7,343건이었으며, 그 중에서도 댓글이 있는 기사는 1,291건이었다. 두 번째로 ‘자율주행’이 제목에 포함된 기사의 댓글 수와 댓글에서 ‘자율주행’을 언급한 댓글 수는 총 10,721개였으며 이때, 해당 키워드가 제목에 포함된 기사의 댓글 수는 9,652개, 댓글에서 해당 키워드를 언급한 댓글 수는 1,069개인 것으로 분석되었다.

(Table 2) Summary of Data

Classification	Description	
Article*	• The number of total articles : 7,343	• The number of comments of articles which marked with keyword in title 9,652
	• The number of articles which have comments : 1,291	
Comment**	• The number of total comments : 10,721	• The number of comments which include ‘Autonomous’ keyword : 1,069
	• The number of comments of articles which marked with keyword in title 9,652	

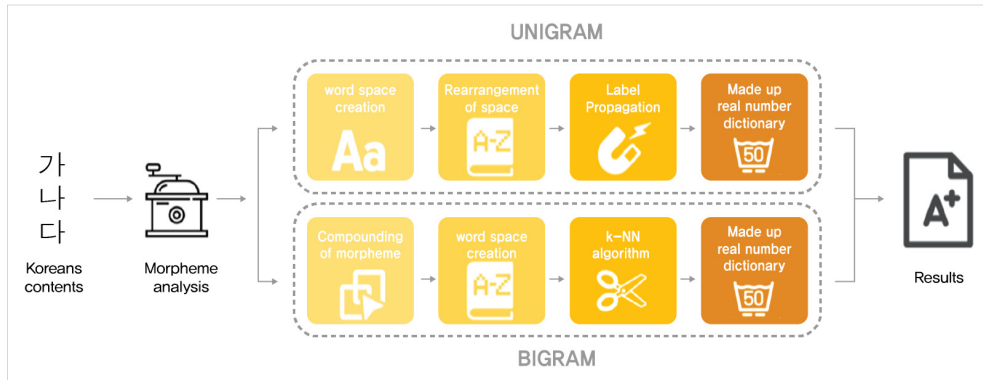
* This article means that article includes ‘Autonomous’ on title

** This comment means that comment includes ‘Autonomous’ on articles title and ‘Autonomous’ keyword

2. 댓글 키워드의 감성평가 틀

본 연구에서는 텍스트 마이닝 기법중 하나인 댓글 키워드의 감성을 분석하는 감성평가 방법론을 활용하였다. 감성평가 방법론이란 한글 콘텐츠의 각 음절과 어절을 벡터 좌표 상에 재배열하여 수학적 방법으로 풀이하고 그 값에 대한 의미를 분석하는 방법이다. 감성평가의 정확도를 높이기 위하여 형태소를 개별적으로 Label Propagation 기법을 활용하고 감성관계를 분석하고, 형태소간의 결합을 통해 단어의 문맥을 k-NN 기법을

사용하여 스코어를 도출하였다. 이때, 긍정·부정·중립의 점수를 산출하는 방식은 Uni그램과 Bi그램의 값을 각각 50%비중으로 구분하여 최종 결과를 산출하였다. Uni그램은 본문(댓글)의 내용을 수학적 풀이방법에 따라 형태소의 부정비율이 20% 이하일 때 긍정으로 구분하여 스코어링을 하고, Bi그램은 긍정 단어의 비율이 80% 이상일 때 긍정으로 스코어링 하였다. Uni그램과 bi그램 각각의 점수가 상이하거나, 긍·부정의 극성이 일정 비율에 미치지 않는 경우는 중립으로 구분하였다.



〈Fig. 3〉 Methodology of Sentiment Diffusion Analysis

1) Label Propagation 알고리즘

Label Propagation(레이블 전파, 이하 LP알고리즘)는 단어의 감성을 구하기 위해 활용한 준지도적 학습법으로 그래프 라플라시안(Graph Laplacian)을 이용하여 레이블링(Labeling) 결과의 일관성을 최대화 하는 알고리즘이다. 소수의 레이블을 포함한 데이터를 이용하여 레이블을 갖지 않는 데이터에 값을 할당해 주며, 이를 텍스트 마이닝에 적용할 경우 각 단어별 정점을 연결하고 알고리즘을 반복하여 인접한 다른 단어들의 감성으로부터 해당 단어의 감성을 유추한다(Heo et al., 2017).

가령, 감성을 갖고 있는 단어데이터를 $(x_1, y_1) \cdots (x_n, y_n)$, 감성을 갖지 않는 데이터를 x_{l+1}, \dots, x_{l+u} , y_i 는 C 개의 클래스 중 하나에 속한다고 할 때, 감성을 갖지 않는 데이터와 감성을 갖는 데이터의 유사성으로 판단을 하며, 거리기반으로 감성을 전파한다. 클래스 행렬인 Y 와 확률 전이 행렬인 T 를 행렬 곱하여 Y 를 갱신하며, Y 값이 수렴할 때 까지 반복하여 감성을 전파한다. 각 단어의 노드(Node)간 엣지(Edge)의 가중치는 노드(Node)의 유사성(거리)과 반비례하고 편차를 크게 만들기 위해 다음과 같은 식을 활용한다(Kim et al., 2017).

$$w_{ij} = \exp\left(-\frac{\sum_{d=1}^D (x_i^d - x_j^d)^2}{\sigma^2}\right)$$

여기서, w_{ij} : 노드 i, j 간 엣지의 가중치

D : 전체 데이터 차원의 개수

x_i^d : i 번째 데이터의 d 번째 차원

σ : 사용자 정의 변수 값

T 는 확률 전이 행렬을 나타내며 가중치를 행별 평균화하여 다음의 식과 같이 계산한다.

$$T_{ij} = \frac{w_{ij}}{\sum_{k=1}^{l+u} w_{kj}}$$

여기서, T_{ij} : 행렬 T 의 (i, j) 번째 값

l : 클래스가 있는 데이터 개수

u : 클래스가 없는 데이터 개수

Y 는 클래스 행렬로 개별 데이터마다 해당 클래스에 속할 확률의 의미하고 다음의 식과 같이 계산한다.

$$Y_{ic} = \begin{cases} 1 & \text{if } i < l \text{ and } C_i = c \\ 0 & \text{if } i < l \text{ and } C_i \neq c \end{cases}$$

여기서, Y_{ic} : i 번째 데이터의 클래스 c 에 대한 확률

C : 클래스 집합

C_i : i 번째 데이터의 클래스

2) k-NN 알고리즘

k-NN 분류방법은 특정 패턴에 대한 학습 데이터로 저장된 패턴 중에 거리가 최소인 k개의 패턴을 정하고 그 중 다수가 속한 범주를 패턴으로 분류하는 알고리즘이다. k-NN 알고리즘은 기존 학습 데이터(d_j)와 신규 데이터(d_x)의 감성 유사성을 판단하고, 유사도가 높은 상위 k개의 이웃 데이터들을 기존 학습데이터 집단에서 찾아내며 유사도를 계산하는 식은 다음과 같다(Eo D. S, 2014).

$$sim(d_x, d_j) = \frac{\sum_k t_{xk} \times t_{jk}}{\sqrt{\sum_k (t_{xk})^2} \times \sqrt{\sum_k (t_{jk})^2}}$$

t_{xk}, t_{jk} = d_x, d_j 에 나타난 가중치

유사도 식을 통해 학습 데이터를 추출 시 k개의 학습 데이터에 부여된 범주들은 새로운 데이터에 할당될 후보 범주 리스트가 된다. 이로부터 신규 데이터에 할당할 최적 감성을 찾기 위해 각 범주에 학습데이터가 분류된 범주별 적합성 점수인 $rel(C_k|d_x)$ 을 산정하여 점수에 따라 신규 데이터에 감성을 할당한다.

$$rel(C_k|d_x) \approx \sum_j sim(d_x, d_j) \times P(C_k|d_j)$$

3) 감성평가 기준

앞선 알고리즘을 통해 점수화 된 음절과 어절의 값에 대해 긍정, 부정으로 구분하고 긍정과 부정으로 구분할 수 없을 때 중립, 그 외 의미를 파악할 수 없는 경우를 기타로 구분하여 값을 도출한다. 본 연구에서

정한 분석 기준은 다음과 같다.

<Table 3> Criteria of analysis

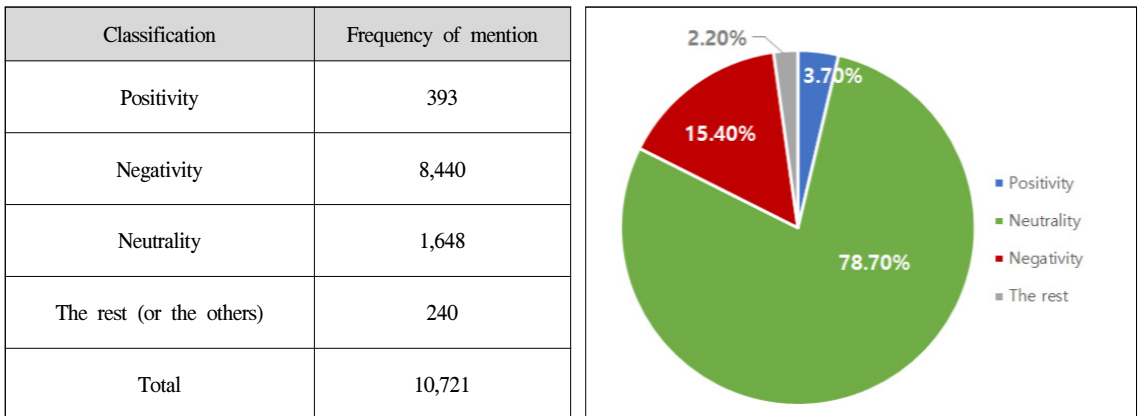
Classification	Description
Positivity	<ul style="list-style-type: none"> Classified to positivity if positive index of comment is high
Negativity	<ul style="list-style-type: none"> Classified to negativity if negative index of comment is high
Neutrality	<ul style="list-style-type: none"> Classified to neutrality if comment is included in following 3 cases <ul style="list-style-type: none"> If all contents don't have sensitivity(emotion) If positive and negative comments are similar If comment is within the scope of neutrality
The rest (or the others)	<ul style="list-style-type: none"> Classified to the rest if comment does not have emotion or comment is the advertisement

IV. 분석 결과

1. 텍스트 마이닝 분석 결과

앞서 수집된 ‘자율주행’이 제목에 포함된 기사의 댓글 수와 댓글에서 ‘자율주행’을 언급한 댓글에 대해 감성평가 방법을 사용하여 분석한 결과, 중립 의견이 가장 높고(8,440개) 다음으로 부정 의견(1,648개), 긍정 의견(393개), 기타 의견(240개)로 나타났다.

<Table 4> Result of comment analysis



1) 중립 의견

첫 번째로 중립 의견을 포함한 댓글 내용을 분석하였을 때, 주요 키워드는 <Table 4>의 Section 1과 같이 나타났다. ‘자율주행’에 대해 중립 의견을 보인 댓글 키워드 중 ‘운전’의 빈도가 가장 높았고 ‘사람’, ‘자율주행’, ‘기술’ 순으로 빈도가 높게 나타났다. 중립 키워드가 포함된 댓글에서는 자율주행 기술의 불확실성과 관련된 사회적 수용성 대한 의견과 함께 스마트폰과 같은 전자제품과의 연계 기술에 대한 의견이 대다수를 차

지하였으며, 중립 의견을 보인 댓글의 예시는 다음과 같다.

“현재 자동차도 일부분을 기계 의존적인 것이니 자율주행 자체를 신뢰하기 어려운 것이 아니라 아직은 자율주행이 대세가 되기엔 시기적으로 안심이 확산되지 못한 것이겠죠 점점 안정성이 증명되어 가고 제도와 인식도 그에 맞춰 정비되고 정리되어 가면 자율 주행이 지금의 자동차 문명처럼 확산되어 있겠지요. 기술이 만들어지자마자 대부분의 사람들로 부터 호응을 얻는 일은 어려운 일이라는 생각이 들게 되네요“

2) 부정 의견

두 번째로 부정 의견을 포함한 댓글 내용을 분석하였을 때, 주요 키워드는 <Table 4>의 Section 2와 같이 나타났다. ‘자율주행’과 관련하여 부정적인 의견을 보인 댓글의 키워드 중 ‘사고’의 언급 빈도가 가장 높았으며 다음으로는 ‘문제’, ‘김여사’, ‘급발진’ 순으로 빈도가 높게 나타났다. 부정 의견 대다수는 자율주행자동차 운행 시 특정 운전자 층(김여사)에 의한 사고를 염려하는 것으로 나타났다. 특히 부정 의견의 키워드와 댓글 내용의 경우, 최근 사회적 이슈인 급발진 사고와 함께 교통사고, 음주운전과 관련된 것으로 나타났다. 또한 신기술이 적용된 자율주행자동차의 해킹, 시스템 오류, 오작동에 대한 우려가 있는 의견이 대다수를 차지하였으며, 부정 의견을 보인 주요 댓글의 예시는 다음과 같다.

“앞으로 자율주행시대가 온다고 해도 문제다. 자율주행차랑 인간운전차가 나란히 달리면 엄청난 사고와 혼란이 생길텐데 거기다 헬조선 운전자들의 매너와 실력이 보통급을 넘어서는건 사실이잖아?”

3) 긍정 의견

세 번째로 긍정 의견을 포함한 댓글 내용을 분석하였을 때, 주요 키워드는 다음 <Table 4>의 Section 3과 같이 나타났다. ‘자율주행’과 관련하여 긍정적인 의견을 보인 댓글의 키워드 중 ‘안전’이 가장 높은 빈도를 보였으며 다음으로는 ‘스마트’, ‘해결’, ‘완벽’ 순으로 빈도가 높은 것으로 나타났다. 긍정 키워드가 포함된 댓글에서는 자율주행자동차의 운전에 따른 안전을 높게 평가하는 것으로 나타났으며, 해당 기술에 따라 사회 전반적인 문제를 해결할 수 있을 것이라는 의견을 보였다. 긍정 의견을 보인 댓글의 예시는 다음과 같다.

“당연하다 사람이 운전하는 것보다 자율주행차가 운전하는게 훨씬 안전하다 왜냐면 사람이 운전을 하면 시시각각 감정이 변해서 분노 기타 등등으로 사고가 발생한다. ...그러기 때문에 자율주행차가 훨씬 더 안전하고 사고도 일어나지 않는다.“

〈Table 5〉 Summary of frequency of keyword mention

Section 1(Neutrality)			Section 2(Negativity)			Section 3(Positivity)		
Position	Keyword	Frequency of mention	Position	Keyword	Frequency of mention	Position	Keyword	Frequency of mention
1	Driving	2,823	1	Accident ¹⁾	1,617	1	Safety	411
2	People	2,439	2	Problem	453	2	Smart	144
3	Autonomous driving	1,257	3	Mrs. Kim	336	3	Solution	140
4	Technology	964	4	Sudden Acceleration	255	4	Perfect	110
5	Road	796	5	Deadly weapon	238	5	Acknowledgment	100
6	Driver	497	6	Hacking	210	6	The best	88
7	Human	480	7	Error	189	7	Completion	82
8	Taxi	454	8	Danger	147	8	Success	80
9	Our Country (Korea)	408	9	Malfunction	143	9	Revolution	80
10	Google	397	10	Traffic accident ²⁾	130	10	Innovation	67
11	Samsung	362	11	Concern	123	11	Sympathy	67
12	Bus	295	12	Drunk driving	115	12	Freedom	62
13	Sensor	280	13	Anxiety	107	13	Convenience	48
14	Artificial intelligence	217	14	Negligence	94	14	Permission	47
15	China	209	15	Defect	84	15	Growth	43

2. 소결

텍스트 마이닝의 감성평가 분석에 따른 결과는 다음과 같이 정리할 수 있다. 첫 번째로, ‘자율주행’ 키워드를 포함한 기사의 발행 추이를 통해 자율주행자동차에 대한 관심이 점차적으로 증가하고 있음을 확인할 수 있었다. 본 연구에서 수집한 기사 데이터를 통해 수집 기간 동안 자율주행자동차 관련 기사 수가 급진적으로 증가함을 확인하였다.

두 번째로 감성평가 분석을 수행한 결과, 자율주행자동차에 대한 긍정적인 시각도 있으나 우려의 시각이 더 큰 것으로 나타났으며 대부분 유보적인 판단을 내리는 것으로 나타났다. 대다수 유보적인 판단을 내리는 것은 아직까지 자율주행자동차가 도입되지 않았고 관련된 경험이 부족하기 때문에 이와 같은 판단을 내린 것으로 사료된다.

부정적인 의견 중 큰 비중을 차지하는 의견으로는 자율주행자동차의 윤리문제를 포함한 교통사고와 관련된

1), 2) 해당 ‘Accident(사고)’ 키워드는 자율주행자동차의 차량 결함(오작동, 시스템 오류 등)으로 인하여 자동차 운행시 자체적으로 발생한 사고를 의미하며, 10번의 ‘Traffic accident’의 경우, 차대차 또는 차대사람 간 발생하는 교통사고를 의미함

의견인 것으로 분석되었다. 특히 최근 발생 빈도가 높아진 차량의 급발진 문제와 자율주행자동차 시스템 등의 차체 결함에 따른 우려가 있음을 확인하였다. 또한 일반자동차와 자율주행자동차의 혼재 상황과 혼재 상황에서 보복운전, 끼어들기 등 운전자 에티켓과 관련된 부정적인 인식이 주를 차지하는 것으로 나타났다.

긍정적인 의견의 경우, 안전과 관련된 의견이 주요 의견으로 분석되었으며 현재의 운행 환경이 안전해질 것으로 기대하는 의견과 신기술에 대한 기대하는 의견이 대다수인 것으로 분석되었다.

V. 결 론

자율주행자동차는 다양한 사회문제를 해결하고 안전성을 확보할 차세대 교통수단으로 각광받고 있다. 그러나 도입 및 상용화를 위해서는 실제 이를 이용할 시민들의 인식 및 신기술 수용 여부에 대한 조사가 선행되어야 한다고 판단된다. 이에 따라 본 연구에서는 상용화 이전에 새로운 기술이 접목된 자율주행 자동차에 대한 시민들의 인식에 대한 분석이 선행되어야 함을 인지하고, 이를 위해 자율주행자동차에 대한 기사와 댓글 데이터를 수집하여 텍스트 마이닝을 활용한 자율주행자동차 인식분석연구를 수행하였다.

본 연구의 결과를 요약하면 다음과 같다. 첫 번째로 분석 결과에 따르면 자율주행자동차의 도입과 운행에 대해 중립적 입장이 과반수를 차지하는 것을 확인하였다. 또한 긍정적인 시각보다는 신기술에 대한 우려에 따른 부정적 입장인 것을 확인하였다. 특히 자율주행자동차에 대한 불확실성, 차체 결함, 사고 책임 등의 문제에 대해 부정적인 의견이 높은 것으로 나타났다. 이는 도입 이전에 해소되어야 할 문제이며 자율주행 자동차에 대한 사회적 수용성과의 향상을 위해 우선적으로 고려되어야 할 중요한 사안이라 판단된다.

연구의 한계로는 본 연구가 특정 기간 동안 발행된 인터넷 기사의 댓글을 활용하여 인식조사를 수행하였기 때문에 의견 수렴에 제한이 있다고 사료된다. 또한 댓글의 특성상 별도의 개인정보를 취득하는 것이 어렵기 때문에 특정 연령대나 성별 등(개인특성데이터)을 통해 의견 차이를 도출하는 것이 어렵다는 한계가 있다. 또한 현재까지 자율주행자동차의 도입 및 상용화와 관련하여 언론 및 시민들의 인식이 명확하게 형성되기에는 관련 정보가 제한적이기 때문에 한계가 있을 것으로 판단되며, 이에 따라 본 연구의 의견을 일반화하기는 어렵다고 판단된다. 향후 연구과제로는 앞서 언급한 연구의 한계를 보완하기 위하여 일반인을 대상으로 한 자율주행자동차 및 차량 운행으로 발생 가능한 문제 상황에 대한 설문조사를 수행하고 지속적인 텍스트 마이닝 분석을 통한 인식조사를 수행할 필요가 있을 것으로 사료된다.

ACKNOWLEDGEMENTS

본 논문은 국토교통부 교통물류연구사업의 연구비지원(과제번호 17TLRP-B131486-01)에 의해 수행되었습니다. 또한 2017년 한국ITS학회 추계학술대회에서 발표한 자료를 수정·보완하여 작성하였습니다.

REFERENCES

- Eo D. S.(2015), *Comparison of Learning Methods in Text Mining with Big Data*, Department of Data Science Graduate School, Inje University.
- Heo C. and Ohn S. Y.(2017), "A Novel Method for Constructing Sentiment Dictionaries using Word2vec and Label Propagation," *The Journal of Korean Institute of Next Generation Computing*, vol. 13, no. 2, pp.93-101.

- Jang K. A., Park S. H. and Kim W. J.(2015), “Automatic Construction of a Negative/positive Corpus and Emotional Classification using the Internet Emotional Sign,” *Journal of KIISE*, vol. 42, no. 4 pp.511-521.
- Jung Y. S., Um K. J. and Won M. S.(2013), *An Innovative Approach for Traffic Safety Improvement Based on Public Involvement via Social Network Services*, KOTI, 2013-08.
- Kim A. R. and Cho S. B.(2017), “A Fusion Method of Co-training and Label Propagation for Prediction of Bank Telemarketing,” *Journal of KIISE*, vol. 44, no. 7, pp.686-691.
- Kim K. O., Moon Y. J., Lee J. D. and Cho S. A.(2016), *The Korea Transport Institute*, A Fundamental Research on Public Perceptions on Ethics, Legal, and Social Acceptance of Autonomous Vehicles(AV).
- Kim S. G., Cho H. K. and Kang J. Y.(2016), “The Status of Using Text Mining in Academic Research and Analysis Methods,” *Journal of Information Technology and Architecture*, vol. 13. no. 2, pp.317-329.
- Lee B. J., Kim K. H. and Park J. I.(2015), *Korea Research Institute for Human Settlements*, Advanced Infrastructure Technologies and National Territorial Development -Focusing on Autonomous Vehicles-.
- Lee D. H., Kang H. G., Kim S. H. and Lee C. M.(2013), “Autocorrelation Analysis of the Sentiment with Stock Information Appearing on Big-Data,” *The Korean Journal of Financial Engineering*, vol. 12, no. 2, pp.79-96.
- Lee G. H. and Lee K. J.(2013), “Twitter Sentiment Analysis for the Recent Trend Extracted from the Newspaper Article,” *Korea Information Processing Society*, vol. 2, no. 10, pp.731-738.
- Oh C. S., Lee Y. T. and Ko M. S.(2016), “Establishment of ITS Policy Issues Investigation Method in the Road Section applied Text mining,” *Korea Institute of Intelligent Transport Systems*, vol. 15, no. 6, pp.10-23.
- Oh J. S.(2015), “Identifying Research Opportunities in the Convergence of Transportation and ICT Using Text Mining Techniques,” *Journal of Transportation Research*, vol. 22, no. 4, pp.93-110.
- Yang M. H., Jung I. S., Kim Y. T. and Cho W. S.(2014), “An Awareness Identification and Preference Analysis for Domestic University Using SNS Data,” *The Korea Big Data Service Society*, vol. 1, no. 1, pp.1-13.
- Yoon H. J.(2016), *Implementation of the Sentiment Analysis Algorithm with Korean Twitter Data*, Graduate School of Seoul National University of Science and Technology.