

# 미래 자동차 분야 국가연구개발사업의 주요 연구 토픽과 투자 동향 분석: LDA 토픽모델링을 중심으로<sup>+</sup>

(Exploring Key Topics and Trends of Government-sponsored  
R&D Projects in Future Automotive Fields: LDA Topic  
Modeling Approach)

마 형 렬<sup>1)</sup>, 이 철 주<sup>2)\*</sup>

(Ma Hyoung Ryul and Lee Cheol-Ju)

**요 약** 글로벌 자동차 산업은 연결, 자율주행, 공유, 전동화 등의 주요 방향 아래 지속적으로 발전하고 있으며, 국내 자동차 산업 또한 기존의 전통적인 자동차 부품 제조로부터 미래 트렌드에 부합하는 전략적인 업의 전환을 꾀하고 있다. 본 연구에서는 2013년부터 2021년까지 산업통상자원부에서 지원한 미래 자동차 분야 연구개발 과제를 대상으로 토픽 모델링을 수행하였다. 해당 기간을 3개 기간으로 구분하여 주요 토픽의 변화를 분석하였다. 센서와 통신, 운전자 보조 기술, 배터리 및 전력 기술은 전 기간 동안 지속적인 주요 토픽으로 나타났으며, 고강도 경량 차체와 같은 주제는 1기에서만 관찰되었다. 한편, AI, 빅데이터, 수소 연료전지와 같은 주제는 2기와 3기에 점점 더 중요한 토픽으로 부상하였다. 또한, 토픽별 정부 투자액과 투자 증가율을 기준으로 각 기수별 집중 투자 분야를 분석하였다. 이러한 연구 결과는 향후 자동차 분야의 정책 수립 및 연구개발 전략 마련 시 기초 자료로 활용될 것으로 예상되며, 증거 기반의 정책 수립과 결정에 기여할 것으로 기대된다.

**핵심주제어:** 미래자동차, 전기자동차, 자율주행자동차, LDA, 토픽모델링, 연구개발 전략

**Abstract** The domestic automotive industry must consider a strategic shift from traditional automotive component manufacturing to align with future trends such as connectivity, autonomous driving, sharing, and electrification. This research conducted topic modeling on R&D projects in the future automotive sector funded by the Ministry of Trade, Industry, and Energy from 2013 to 2021. We found that topics such as sensors, communication, driver assistance technology, and battery and power technology remained consistently prominent throughout the entire period. Conversely, topics like high-strength lightweight chassis were observed only in the first period, while topics like AI, big data, and hydrogen fuel cells gained increasing importance in the second and third periods. Furthermore, this research analyzed the areas of concentrated investment for each period based on topic-specific government investment amounts and investment growth rates.

**Keywords:** Future cars, Electric Vehicle, Autonomous Vehicle, Latent Dirichlet Allocation (LDA), Topic Modeling, Research and Development Strategy

\* Corresponding Author: designme74@gmail.com

+ 본 연구는 2023년 한국산업정보학회 추계학술대회에서 발표된 논문을 수정 및 보완한 논문입니다.

Manuscript received November 29, 2023 / revised January

10, 2024 / accepted January 16, 2024

1) 산업통상자원 R&D전략기획단, 제1저자

2) 한국산업기술기획평가원, 교신저자

## 1. 서론

디지털 전환과 그린 전환으로 인해 전통적 산업 분야의 경계가 허물어지거나 융합되는 등 산업 패러다임 변화가 가속화되고 있다 (KIET, 2017). 대표적으로 자동차 산업은 전통 제조업에서 벗어나 서비스 중심의 모빌리티 산업으로 변화하면서 산업 내 경쟁구도 역시 더욱 복합적으로 변화하고 있다. 화석연료로 인한 기후변화를 대응하고 탄소중립을 위한 친환경 자동차로의 전환을 위해 기존의 가솔린 및 디젤엔진은 모터를 중심으로 하는 전동화(Electrification) 방향으로 변화되었으며, 디지털 기술의 발달과 함께 연결(Connectivity), 자율주행(Autonomous), 공유(Sharing)를 위한 첨단기술이 급격하게 발전하고 있다. 이와 같이 전기자동차 및 자율주행자동차 등 미래자동차 시장은 향후 전 세계적으로 큰 성장이 예상되면서, 기존 전통자동차 시장을 조속히 대체할 것으로 예상된다.

한편 국내 자동차 산업은 여전히 기존 내연기관 위주로 구성되어 있어 미래트렌드에 부합하는 산업 전환이 원활하게 이루어지지 않는다면 향후 새로운 모빌리티 산업에서 경쟁력을 유지하기가 어려울 것으로 보인다. 디지털 전환과 그린 전환으로 인한 거대한 산업의 전환은 개별 자동차 기업의 이슈뿐만 아니라 산업과 국가 경제에 중요한 부분일 수 있다. 자동차 산업은 국가 수출의 선봉 역할을 하는 국가경제에 중요한 산업분야 중 하나이다. 국내 자동차 산업에 속한 대다수 기업은 미래자동차 기술 개발, 인력, 예산 등의 어려움으로 인해 미래자동차 트렌드에 부합하는 유망 사업분야로 전환하기에 쉽지 않은 상황이다. 정부는 기업이 적극적으로 투자할 수 있도록, 유망분야로 성공적인 전환이 될 수 있도록 시장선도자의 역할과 지원자의 역할을 충분히 감당할 필요가 있다. 이를 위해서는 정부가 적극적으로 정책을 수립하고 연구개발 로드맵을 전략적으로 수립하는 것이 중요하다. 특히, 정부가 지원하는 R&D 투자의 성과가 전반적으로 미흡하다는 지적이 많은 것을 감안할 때, 좀 더 과학적이고 증거 중심의 연구개발 전략을 마련해야 한다.

따라서 본 연구에서는 미래 자동차분야의 국가연구개발사업을 대상으로 그 간의 주요 연구 토픽과 투자 동향을 분석하고자 한다. 먼저 산업통상자원부가 그간 지원한 국가연구개발과제 데이터를 모두 수집하였다. 수집된 각각의 과제에 대해 미래 자동차 분야의 과제인지를 판단하여 분석대상 데이터를 선별하고 전처리를 수행하였다. 분석대상 데이터를 총 세 개의 기간으로 분류하였는데, 각 기간별로 주요 키워드를 분석하고, 워드클라우드를 통해 시각화를 진행하였다. 또한 토픽모델링을 수행하기에 앞서 적합도를 측정한 후 자동차 분야 전문가 3인의 판단을 거쳐 최적의 토픽수를 결정하였다. 3개의 기간별로 각각 토픽모델링을 수행하고 전문가의 검토를 통해 토픽명을 명명하였으며, 기간에 걸친 토픽의 변화를 확인하였다. 마지막으로 토픽별 투자액을 산정하고 기간에 걸친 변화를 분석하였다.

기존의 연구동향을 분석한 논문들을 살펴보면 주로 특허나 논문 데이터를 사용한 경우가 많고, 일부 연구에서는 특정 분야의 국가연구개발사업 데이터를 활용한 경우가 있었다. 다만, 본 연구는 기존 연구에서 다루지 않았던 미래 자동차 분야를 특정하여 국가연구개발사업 데이터를 활용하였고, 실제 토픽별로 지원된 투자금액까지 분석하였다는 점에서 기존 연구에서 한발 더 나아갔다.

본 연구를 통해 미래자동차 연구분야에서 현재까지 꾸준히 투자되고 있는 주제, 중요도가 줄어든 주제, 점점 중요성이 커지는 핵심 주제, 각 기간별로 유망분야의 변화 등을 분석하였으며, 전 기간에 걸친 각 주제별 투자동향을 확인하였다. 본 연구결과는 향후 정책수립과 전략적인 연구개발 로드맵 수립, 과제 기획 등에 기초 자료로 활용할 수 있을 것으로 기대된다.

본 논문의 구성은 다음과 같다. 먼저 2장은 미래 자동차 선행연구, 기존 연구 동향 분석, 토픽모델링 개념에 고찰한다. 3장에서는 연구절차 및 데이터 수집 및 전처리 등 연구방법을 설명하고, 4장은 키워드 빈도, 주요 토픽 및 투자 동향에 대해 분석한다. 끝으로 5장에서는 연구결과를 요약하고 시사점을 제시한다.

## 2. 선행 연구 분석

### 2.1 미래 자동차 기술

산업통상자원부에서 발간한 산업기술 R&D 투자전략에 따르면 미래 자동차는 크게 전기자동차와 자율주행자동차로 구분될 수 있다(MOTIE, 2019). 먼저 전기구동차(xEV)는 이차전지 또는 연료전지로부터 전기에너지를 공급받아 동력원으로 사용하는 차량이다. 하이브리드자동차(HEV)를 포함하여 순수전기차(BEV), 수소전기차(FCEV), 플러그인 하이브리드차(PHEV) 등을 전기구동차로 분류한다. 최근에 발표된 산업기술수준조사에 의하면 전기구동차 분야에서 일본이 최고 기술국으로 한국의 기술수준은 최고기술국 대비 96%이며, 약 0.2년의 기술격차를 가지고 있다(KEIT, 2022). 전기자동차 국내 기술 수준은 꾸준히 상승하고 있으며, 글로벌 시장의 점유율 확보를 위해서는 지속적인 R&D 투자와 정부 지원이 필요하다. 전 세계적으로 다수의 국가가 R&D 투자를 확대하고 있으며, 중국은 특히 공격적으로 투자하고 있다.

구동 및 전력변환시스템, 에너지 저장 및 충전시스템, 공조 및 열관리시스템, 연료전지시스템, 수소저장 및 공급시스템, 초고효율 하이브리드시스템, 탄소중립 합성연료 적용 시스템 등을 포함하여 전기구동차의 세부 기술분야에서 한국은 특히 배터리, 수소연료전지 분야에서 강점을 가지고 있다. 그 외 분야에서는 전 세계 수준에 뒤처지거나 아직 미흡한 상태이다. 향후 초기시장 창출을 위해 모터, 배터리, 충전, 연료전지, 수소저장, 공조, 경량화 등의 부품 기술을 중점적으로 개발하는 것이 중요하다.

자율주행자동차는 운전자의 개입 없이 스스로 주행을 제어하여 안전하게 목적지에 도달하는 차량을 말한다. 이를 위해 센서, 제어기 등의 핵심 부품이 고도화되고, AI 기술을 활용한 인지 성능도 향상되고 있다. 또한, 주변 차량, 보행자, 인프라와의 연결을 통한 커넥티드 자율주행 기술도 개발되고 있으며, 다양한 형태의 모빌리티 서비스와의 연계로 자율주행 기반 MaaS(Mobility as a Service) 산업도 발전하고 있다.

자율주행자동차는 크게 양산형과 서비스용으로 나뉘어진다. 양산형 자율주행차는 Ford, Toyota, 현대자동차 등이 레벨2 제품을 양산하고 있으며, 일부 완성차는 고속도로용 레벨3 제품을 출시하였다. 반면, 서비스용 자율주행차는 자율배송, 로봇택시, 자율셔틀 등의 서비스를 제공하기 위해 레벨4 기술을 기반으로 개발되고 있으며, Google Waymo, GM Cruise 등이 해당 분야에서 선도적인 위치를 차지하고 있다.

자율주행자동차 기술은 자율주행 부품, 자율주행 시스템, 자율주행 서비스뿐만 아니라 AICBM(AI, IoT, Cloud, BigData, Mobile)과 같은 공통 핵심 기술 및 연계산업 기술을 포함한다. 자율주행 부품에는 센서, 제어기, 통신 모듈, HVI, 정밀맵 모듈, 전동식 조향 및 제동 등이 포함되며, 주행환경 인지, 판단, 동작, 인프라 연계 기술은 자율주행 시스템 기술로 분류할 수 있다. 또한, 자율주행자동차를 활용한 대중교통, 택시, 차량 공유, 택배, 무인 식료품 판매 등의 서비스와 AICBM, 교통, 통신, 인터넷, 법제도 등과 관련된 연계 산업 기술도 모두 자율주행자동차 기술에 포함된다.

한국은 자율주행차 기술 분야에서 미국 대비 약 81.0%의 기술 수준을 가지고 있으며, 두 나라 간의 기술 격차는 약 1년으로 나타난다(KEIT, 2022). 특히, 자율이동서비스 분야와 주행환경 인지기술 분야에서 한국은 기술 수준이 상대적으로 낮은 것으로 나타나고 있다.

정부는 자율주행차 분야를 미래 핵심 산업으로 설정하고 국내 제조사의 기술 개발을 지원하여 자율주행차의 조기 상용화를 추진하고 있다. 관계 부처들이 합동으로 자율주행차 분야를 혁신성장BIG3 산업 중점 추진과제로 설정하고, 레벨4 자율주행 상용화를 목표로 정책적 지원을 하고 있다. 정부는 약1조 1,000억원의 규모로 자율주행 융합 신산업 육성과 국민 수용성 향상을 목표로 하는 자율주행 기술개발 혁신사업을 추진하고 있으며, 자율주행차 분야에 대한 투자 규모도 확대할 것으로 전망된다.

### 2.2 연구동향 분석 연구

시장 동향 및 고객과 소비자의 관심 주제를 분석하는 등 다양한 분야에 토픽모델링이 활용되고 있으나 (Kim et al., 2021; Lim et al., 2021; Hong et al., 2018; Choi et al., 2022; Chung et al., 2018) 여기에서는 연구동향 분석과 관련한 연구에 대해 중점적으로 살펴보고자 한다. 먼저 Kim(2021)은 자율주행 자동차의 연구개발 동향에 대해 분석하였다. 자율주행 자동차 연구와 관련된 1,870개의 영문 뉴스 기사를 수집하고 토픽모델링을 통해 20개의 주제를 추출하였다. 이러한 주제를 각각 투입, 활동, 산출, 성과라는 논리모형에 대입하여, 자율주행 자동차 연구개발 논리모형을 제시하기도 하였다. 분석 결과는 국내외 자율주행 자동차 연구개발 프로젝트의 진행상황을 전체적으로 파악하고 급변하는 기술 변화를 감지하기 목적으로 활용될 수 있다.

Nam et al.(2018)은 자동차 반도체 산업에서 연구개발 정책 수립과 기술 전략에 대한 통찰력을 얻기 위해 차량용 반도체 기술의 동향을 분석하였다. 특허를 수집한 후 토픽모델링을 적용한 결과, 전기차, 운전자 지원 시스템, 디지털 제조 등과 같은 주요 기술들을 포함하여 기술적 동향을 파악하였다.

또한 Jo et al.(2019)은 금융, 물류, 정부 서비스 등 다양한 분야에서의 블록체인 기술 연구 동향을 탐구하였다. LDA 기법을 사용하여 블록체인 분야의 연구 키워드를 추출하고 미국, 중국, 한국의 연구 동향을 분석한 결과, 미국에서는 경제와 금융에, 중국에서는 기술과 시스템에, 한국에서는 정책과 규제에 중점을 두는 등 국가간 연구 초점의 차이가 난다는 것을 발견하였다.

Park et al.(2019)은 스마트시티 정책, 산업 및 학문적 방향을 제시하기 연구를 수행하였다. Scopus DB와 Springer DB에서 스마트시티에 관련된 학술 논문을 수집하고, LDA 기반 주제 모델링을 적용하여 연구주제와 동향을 분석하였다. 분석 결과를 토대로 서비스 및 응용 분야, 기술 분야, 시민/사용자 관점으로 총 8개의 연구주제를 분류하였는데, 특히 데이터와 개인정보 보호와 관련된 주제가 연결중심성이 높게 나타나 다른 주제의 연구에서도 가장 많이 포함된

다는 것을 발견하였다.

Kim et al.(2017)은 정보시스템 분야의 연구 동향을 분석하기 위해 2002년부터 2016년까지의 주요 학술지에서의 초록을 수집하여 토픽모델링과 시계열 회귀 분석을 실시하였다. 추출된 주제에는 시스템 구축, 혁신 능력, 고객 충성도 등이 포함되었으며, 고객 충성도, 커뮤니케이션 혁신, 정보 보호 및 개인정보 보안에 대한 증가 추세와 시스템 구축 및 웹사이트의 감소 추세를 발견하였다.

Park et al.(2017)은 과학기술의 동향을 예측하기 위해 키워드 네트워크 분석을 수행하였다. 이 연구는 인공지능 특허에 초점을 맞추고 세 가지 시기로 나누어 키워드 네트워크 분석을 실시하였다. 시간이 지남에 따라 인공지능 적용과 관련된 키워드가 주목받는 것을 확인하였으며, 연구결과는 AI 연구에서의 응용 방법과 신규 분야에 대한 통찰력을 제공하고 있다.

상기의 여러 논문에서 토픽 모델링 방법을 통해 각 분야의 연구동향을 파악하고 관심을 끄는 새로운 주제를 식별하였으며, 이러한 연구들은 미래 연구개발 방향과 정책 결정을 위한 중요한 근거로 활용될 수 있다.

## 2.3 토픽모델링

토픽 모델링은 문서 집합에서 토픽을 찾아내는 텍스트마이닝 기법이다. 토픽은 문서에서 발견되는 단어들의 집합으로 정의할 수 있으며, 토픽 모델링은 다양한 분야에서 핵심주제를 찾아내기 위해 사용된다. 예를 들어, 검색 엔진에서는 토픽 모델링을 사용하여 문서의 주제를 추출하고, 사용자의 검색어와 관련성이 높은 문서를 추천할 수 있다. 고객 민원 시스템에서는 토픽 모델링을 사용하여 고객 민원의 주제를 추출하고, 고객의 민원을 효과적으로 분류하고 처리할 수 있으며, 마케팅 분석에서는 토픽 모델링을 사용하여 고객의 관심사를 파악하고, 고객에게 맞춤형 마케팅 캠페인을 진행할 수 있다.

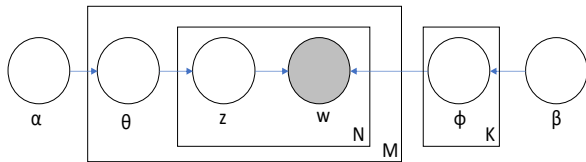
토픽 모델링의 방법론으로 초기에는 Latent Semantic Indexing (LSI)와 Probabilistic Latent Semantic Analysis (PLSA)가 주로 사용되었다

(Deerwester et al., 1990; Hofmann, 1999). LSI는 문서의 단어 빈도를 기반으로 토픽을 추출하는 방법이고, PLSA는 문서의 단어 빈도와 단어 사이의 연관성을 기반으로 토픽을 추출하는 방법이다, 그 후 2003년에 제안된 Latent Dirichlet Allocation (LDA)는 현재 가장 널리 사용되는 토픽 모델링 방법론 중 하나이다 (Blei et al., 2003). LDA는 LSI와 PLSA의 장점을 결합한 방법론으로, 문서 집합이 토픽들의 혼합으로 구성되어 있다고 가정한다. 각 토픽은 단어들의 분포로 표현되며, LDA는 문서에서 토픽을 추출하기 위해 토픽 할당 및 단어 할당을 수행하고 토픽과 단어 사이의 연관성을 학습한다.

Fig. 1에서 제시된 LDA 모델에서는 문서별 토픽의 분포( $\theta$ )와 토픽별 단어의 분포( $\phi$ )는 디리클레 분포로 이루어져 있다고 가정하고 있으며, 토픽 할당( $z$ )과 관측된 단어( $w$ )는 다항(multinomial) 분포를 이루고 있다. 모델의 작동 원리에 대해 구체적으로 살펴보면, 맨 왼쪽과 오른쪽에 보이는  $\alpha$ 와  $\beta$ 값은 디리클레 파라미터로서 연구자가 특정한 값을 설정해 줘야 한다.  $\alpha$ 값에 따라 각각의 문서 내에서의 토픽의 분포( $\theta$ )가 정해지며, 이로부터 특정 토픽 할당( $z$ )을 수행한다. 추출된 토픽에 해당하는 단어를 선택

하기 위해 토픽별 단어의 분포( $\phi$ )를 참고한다. 토픽별 단어의 분포( $\phi$ ) 역시  $\beta$ 값에 따라 결정되는 디리클레 분포를 따르고 있다. 단어가 선택되고 나면, 이렇게 추출된 단어가 실제 관측된 단어( $w$ )와 동일한 지를 지속적으로 비교하게 된다. 결국 LDA 방법론을 통해 추출된 단어가 실제 관측된 단어( $w$ )와 일치하는 확률을 최대화하는 특정 토픽 할당( $z$ ), 문서 내 토픽의 분포( $\theta$ ), 토픽별 단어의 분포( $\phi$ )를 찾아낼 수 있다. 이를 통해 각 문서가 어떤 토픽으로 구성되어 있는지, 각 토픽은 어떠한 단어로 구성되어 있는지 알 수 있게 된다.

이러한 LDA 방법론은 문서의 단어 빈도와 단어 사이의 연관성을 기반으로 토픽을 추출하기 때문에 다양한 분야에서 널리 활용될 수 있으며, 확률적 모델로서 새로운 문서에 대해서도 유연하게 적용할 수 있다. 다만 LDA 방법론에는 몇 가지 한계점이 존재한다. 첫째, 토픽 개수를 사전에 정해야 한다는 점에서 주관적인 결정이 필요하다. 너무 적은 토픽 수를 선택하면 토픽 간의 세부적 차이를 포착하지 못할 수 있고, 너무 많은 토픽 수를 선택하면 해석의 어려움이 생길 수 있다. 둘째, 추출된 토픽 간에 상관성이 존재할 수 있어 각 토픽을 독립적으로 해석하는데 어려움이 있다. 셋째, LDA는 유사어 처리에 대한 제약이 있을 수 있다. 예를 들어, 중요한 의미를 가진 유사한 단어가 서로 다른 토픽에 속할 경우, 토픽 모델링의 정확도에 영향을 줄 수 있다. 마지막으로, 분석자의 주관 개입이 필요할 수 있어 결과의 신뢰도에 영향을 줄 수 있다. 하지만 LDA 방법론은 텍스트 데이터에서 내재된 주제를 추출하는 유용한 텍스트 마이닝 기법으로 여전히 가장 널리 사용되고 있다.



K: number of topics  
 N: number of words in the documents  
 M: number of the documents

$\alpha$ : Dirichlet parameter of per-document topic distribution  
 $\beta$ : Dirichlet parameter of per-topic word distribution

$\phi$ : word distribution for topic K  
 $\theta$ : topic distribution for the document M

$z$ : topic assignment for  $w$   
 $w$ : observed word

Fig. 1 Graphical model representation of LDA

### 3. 연구 방법

#### 3.1 연구절차

본 연구의 목적은 미래 자동차분야의 국가연구개발사업을 대상으로 그 간의 주요 연구 토픽과 투자 동향을 분석하는 데 있다. 이를 위한

연구절차는 Fig. 2에 제시되어 있다. 구체적으로 살펴보면, 먼저 산업통상자원부가 지원된 국가 연구개발과제 데이터를 모두 수집하였다. 수집된 각각의 과제에 대해 미래 자동차 분야의 과제인지를 판단하여 분석대상 데이터로 확정하고 전처리를 수행하였다. 분석대상 데이터를 3개의 기간으로 분류하여, 각 기간별로 주요 키워드를 분석하고, 워드클라우드를 통해 시각화를 진행

하였다. 또한 토픽모델링을 수행하기 위해 사전적으로 정합도 측정 및 자동차 분야 전문가 3인의 판단을 통해 최적의 토픽수를 결정하였다. 3개의 기간별로 각각 토픽모델링을 수행하고 전문가의 검토를 통해 토픽명을 명명하였으며, 기간에 걸친 토픽의 변화를 확인하였다. 마지막으로 토픽별 투자액을 산정하고 기간에 걸친 변화를 제시하였다.

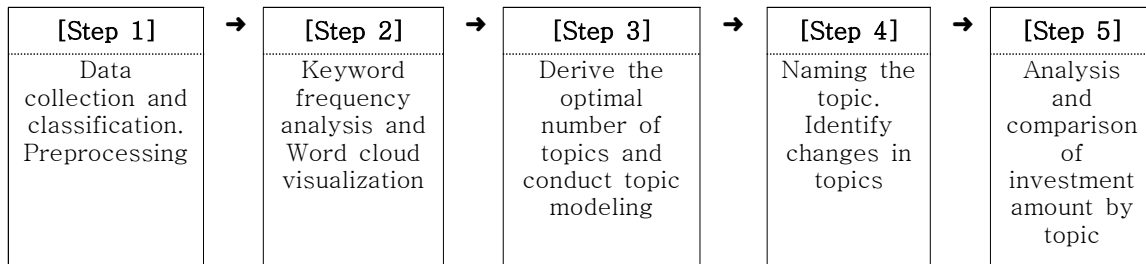


Fig. 2 The overall procedure of our analytical framework

### 3.2 데이터 수집 및 전처리

본 연구에서는 2013년부터 2021년까지 산업통상자원부에서 지원한 연구개발과제 데이터를 국가과학기술지식정보서비스(NTIS)를 통해 확보하였다. 산업통상자원부는 여러 부처 중에서도 특히 산업기술 개발을 중점적으로 지원하기에 본 연구에서는 산업통상자원부에서 지원한 미래 자동차 분야의 연구개발 과제에 초점을 맞추었다. 먼저 여러 산업 분야에 지원된 연구개발과제 중에서 미래 자동차 분야에 지원된 연구개발과제를 분류하는 작업을 수행하였다. 연구개발사업명을 기준으로 미래 자동차 분야에 지원된과제를 분류하는 방법도 있겠지만, 이는 상당한 오류를 포함할 여지가 크다. 예를 들어, 자동차 관련 연구개발사업이라 하더라도 해당 사업에서 지원된 과제에는 가솔린 또는 디젤자동차와 관련된 연구개발과제가 있을 수 있다. 또한 자동차 관련 연구개발사업이 아니더라도 미래 자동차와 연관된 연구개발과제가 지원되었을 수도 있다. 특히 소재부품관련 사업에서 자율주행에 필요한 센서 등을 개발하는 과제도 있을 수 있다. 따라서 본 연구에서는 미래 자동차관련 과제를 보다 정교하게 분류하기 위해 전문가의 판

단 및 머신러닝 기법을 함께 활용하였다. 일부 샘플을 대상으로 전문가의 판단을 통해 미래 자동차 과제를 사전에 분류한 후 이를 머신러닝을 통해 학습시킨 다음에 나머지 샘플에 대해 머신러닝을 통한 분류작업을 수행하였다. 이렇게 분류된 미래 자동차 연구개발과제를 전문가의 검토를 거쳐 확정하였다. 각 과제별로 여러 정보를 포함하고 있는데, 이 중에서 과제지원연도, 영문키워드, 정부출연금 등의 데이터를 주로 활용하였다.

각 과제별로 5개 내외의 영문키워드를 포함하고 있는데, 이에 대한 전처리를 수행하였다. 구체적으로 특수문자 제거, 소문자 변환, 불용어(stop word) 제거, 표제어 추출(lemmatization), 유사한 단어에 대한 표준변환 등의 순서로 진행하였다. 영문키워드에서 마침표, 느낌표 및 물음표 등 특수문자를 모두 제거하고, 모든 키워드를 소문자로 변환하였다. 또한, NLTK에서 정의된 불용어 리스트를 이용하여 불용어를 제거하였다. 복수형태의 단어를 모두 단수로 만드는 등 표제어 추출(lemmatization)도 수행하였다. 마지막으로 동일한 의미이지만 다르게 표현되는 키워드에 대해서는 표준형 키워드로 변환하였다. 예를 들어, 전기차의 경우에 'ev', 'electric

vehicle', 'electric car' 등으로 표현되고 있는 것을 'electric vehicle'로 통일하였다.

마지막으로 시간의 흐름에 따른 토픽의 변화를 파악하기 위해 연구개발과제 데이터를 대상으로 3개의 기간으로 구분하였다. 정부 정책 기조의 변화를 충분히 감안하기 위해 산업통상자원부에서 발간한 산업기술R&D전략의 계획기간과 동일한 기준으로 분석기간을 구분하였다. 산업기술R&D전략은 산업기술 R&D 예산을 산업정책에 맞추어 효율적으로 배분하고 사업·과제를 기획하기 위한 산업통상자원부의 종합적 R&D 투자전략이다. 산업기술R&D전략이 2013년에서 2015년, 2016년부터 2018년, 2019년에서 2021년까지를 대상으로 해당 시점의 기술과 시장 변화 등을 반영하여 각각 수립되고 발간된 것을 감안하면, 이와 동일하게 기간을 분류하는 것이 주요 토픽의 변화를 보다 잘 파악할 것으로 기대된다.

최종적으로 1기(2013년에서 2015년) 350개 과제, 2기(2016년부터 2018년) 631개 과제, 3기(2019년에서 2021년) 1,054개 과제 등 총 2,035개 과제를 대상으로 토픽모델링을 수행하였다. 구체적으로 1기 1,457개, 2기 2,758개, 3기 4,771개 등 총 8,986개의 영문 키워드가 분석에 사용되었다.

## 4. 분석 결과

### 4.1 키워드 빈도

토픽모델링을 수행하기에 앞서 토픽에 대한 전체적인 모습을 사전에 파악하기 위해 각 기수별 주요 키워드를 분석하고 워드 클라우드로 시각화를 진행하였다. 각각의 결과는 Table 1과 Fig. 3에 제시되어 있다. Table 1에서 보듯이 각 기수에 걸쳐 'electric vehicle'은 공통적으로 가장 빈번하게 나타나는 있다. 각 기수별로 자세하게 살펴보면, 1기에 'sensor fusion'이 두 번째 순위를 차지하고 있는데 본 키워드는 2기와 3기에도 관찰되고 있다. 자율주행자동차의 센서 또는 통신과 관련된 것으로 보이며, 1기에 나오

는 키워드 중 'wave' 또는 'integrated control', 2기에 나오는 'signal processing' 또는 'v2x' 등과 관련성이 높은 것으로 보인다. 또한, 'inverter'가 1기, 2기와 3기에 걸쳐 지속적으로 나타나고 있는데, 이는 전력변환기술과 관련된다. 전기자동차와 자율주행자동차의 안전성 이슈가 중요한데, 'safety' 키워드가 3개의 기간에 공통적으로 나타나고 있으며, 1기에는 차량의 안정성과 관련된 'aeb'가 주요키워드로 분석되었다. 'aeb'는 autonomous emergency braking의 약자로서 자동 긴급제동 시스템을 의미한다. 배터리는 전기차의 가장 중요한 구성요소 중의 하나인데, 3개의 기간에 걸쳐 관련된 키워드들이 여러 발견된다. 1기에는 이차전지(secondary battery), 리튬이온 배터리(lithium ion battery), 고전압(high voltage) 및 관련 시스템(energy storage system) 등의 키워드가 나타나며, 2기에는 배터리 팩(battery pack)과 배터리 관리시스템(battery management system)의 키워드가 추가로 나타나고 있다. 이런 키워드들은 3기에도 지속적으로 나타나고 있다. 고강도 차체와 관련하여 'high stiffness' 키워드를 1기에서 발견할 수 있는데, 2기와 3기에서는 더 이상 주요 키워드에서 나타나지 않는다. 다만, 경량차체와 관련해서는 'lightweigh'가 1기와 3기에서 발견된다. 운전자보조시스템인 'adas' (advanced driver assistance system)는 1기와 2기에 나타나는데, 1기에는 'acc', 'pedestrian detection', 2기에는 'collision avoidance', 3기에는 'object detection' 등의 키워드가 중점적으로 나타나고 있다. 자율주행과 관련된 키워드는 1기에는 보이지 않다가, 2기부터 눈에 띄게 나타나고 있다. 'autonomous driving', 'autonomous driving', 'deep learning' 등이 여기에 속한다. 특히, 딥러닝 등 인공지능 기술은 자율주행 기술개발에 필수적인 기반기술로 판단된다. 또한 2016년 3월 이세돌-알파고 대국을 기점으로 대한민국에 인공지능 열풍이 불었으며 정부R&D에서도 이러한 흐름을 반영한 결과인 것으로 보인다. 2기에는 자율주행기술에 필수적인 부품인 'lidar'가 상위 순위에 나타난 것을 볼 수 있다. 3기에 들어서 주목할 만한 키워드는 'automated driving

shuttle’, ‘micro mobility’이다. 자율주행자동차가 보다 확대되어 셔틀 또는 마이크로 모빌리티로 진화하고 있는 것을 보여준다. 또한 2기부터 나타난 인공지능 기술이 보다 확대되어 3기에는

‘big data’ 및 ‘platform’으로 확장되고 있음을 알 수 있다. 그리고 수소연료전지를 기반으로 하는 수소차 기술, 즉 ‘fuel cell’이 본격적으로 등장하고 있다.

Table 1 Keyword frequency analysis according to each period

Rank	Period 1 (2013~2015)		Period 2 (2016~2018)		Period 3 (2019~2021)	
	keyword	frequency	keyword	frequency	keyword	frequency
1	electric vehicle	49	electric vehicle	76	electric vehicle	120
2	sensor fusion	14	smart car	19	autonomous driving	60
3	ecu	8	autonomous driving	18	autonomous vehicle	39
4	green car	7	sensor fusion	18	battery	33
5	high voltage	7	secondary battery	16	platform	24
6	safety	7	autonomous vehicle	16	battery pack	23
7	motor	6	lidar	15	artificial intelligence	21
8	aeb	6	inverter	13	sensor fusion	20
9	secondary battery	6	safety	12	fuel cell	18
10	high stiffness	5	motor	12	big data	15
11	smart car	5	adas	12	fuel cell electric vehicle	13
12	adas	5	battery	11	lightweight	13
13	acc	5	battery management system	10	inverter	12
14	lithium ion battery	5	vehicle	10	ai	12
15	inverter	5	automotive	9	automated driving shuttle	12
16	ami	4	vehicle test	9	high efficiency	12
17	wave	4	high voltage	9	micro mobility	12
18	augmented reality	4	battery pack	8	automated vehicle	12
19	integrated control	4	signal processing	7	deep learning	12
20	fusion	4	reliability	7	battery management system	11
21	energy storage system	4	deep learning	7	secondary battery	11
22	high capacity	4	high efficiency	7	traction motor	10
23	pedestrian detection	4	fuel cell	7	lithium ion battery	10
24	battery	4	energy storage system	7	object detection	10
25	reliability	4	collision avoidance	7	safety	10
26	lightweight	4	intelligent vehicle	6	reliability	10
27	technology	3	ecu	6	automotive	9
28	small volume battery cooling fan	3	laser scanner	6	durability	9
29	vehicle simulation	3	v2x	6	standardization	9
30	surface treatment	3	surface treatment	6	high voltage	9

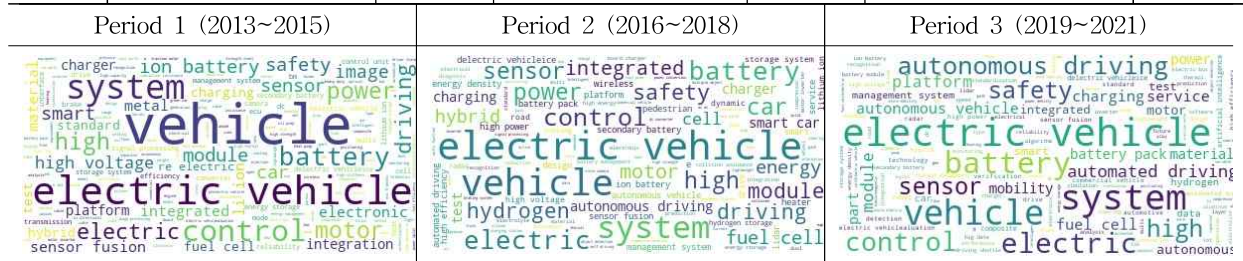


Fig. 3 Wordcloud on keywords in each period



## 4.2 토픽 모델링

토픽모델링을 수행하는 데 필요한 적절한 토픽의 개수를 결정하기 위해 기존 논문과 전문가 검토를 거친 후 일관성 점수를 참고하였다. 자동차분야 연구동향 또는 인공지능 분야 국가연구개발사업 연구주제를 분석한 기존 연구를 보면 대부분 5개에서 10개 내의 토픽을 선정한 것으로 확인된다. 더욱이 토픽 개수를 너무 크게 잡는 경우에는 결과를 해석하기에 용이하지 않을 가능성이 크다. 따라서, 5개에서 10개의 범위 내에서 일관성 점수가 가장 높게 나타나는 토픽 개수를 사용하고자 한다. Fig. 4에서 볼 수 있듯이 1기(2013년에서 2015년)의 데이터를 기준으로 일관성 점수를 계산하였는데, 토픽이 6개인 경우에 상대적으로 가장 높은 점수를 보였다. 또한 시기에 따른 토픽의 변화 등을 일관성 있게 분석하고 해석의 용이성을 위해 본 연구에서는 1기(2013년에서 2015년), 2기(2016년부터 2018년), 3기(2019년에서 2021년)에 모두 동일하게 6개의 토픽 개수를 사용하였다. 또한 디리클레 파라미터인  $\alpha$ 와  $\beta$ 값은 각각 0.1, 0.01로 설정하였다.

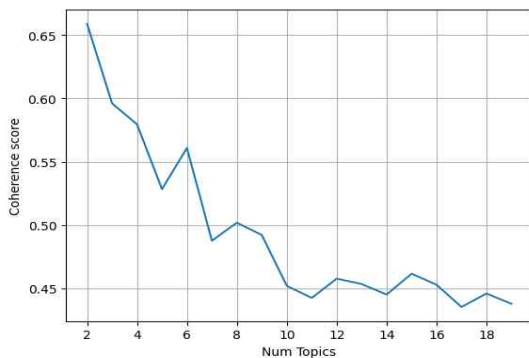


Fig. 4 Coherence score

각 기수별로 추출된 토픽에 적절한 이름을 부여하기 위해, 먼저 산업통상자원부에서 발간한 산업기술R&D전략 (MOTIE, 2019) 중, 전기·수소자동차 및 자율주행차 분야의 기술분류체계를 참고한 후, 전문가 검토를 통해 각 토픽별 차별성을 고려하여 최종 확정하였다.

먼저 Table 2에서 나타난 제1기의 토픽을 살

펴보면, 토픽 11은 센서, 통신, 신호처리와 관련성이 높다. 주요 키워드에서 sensor fusion, optical sensor, integrated control 등 센서 연구 개발과 관련 것들이 보이며, wave 또는 signal processing 등 통신과 신호처리 관련 키워드도 나타난다. 토픽 12는 주요 키워드로 ACC (Adaptive Cruise Control), ADAS (Advanced Driver Assistance Systems) 등이 발견되며 운전자의 편의를 증대시키는 운전자 보조기술 개발과 관련되어 있다. 토픽 13은 고강도 경량차체에 대한 주제로 보이며, high stiffness, lightweight, autobody 등을 통해 유추해 볼 수 있다.

토픽 14는 여러 주제의 키워드가 혼재되어 있는데, 특히 secondary battery, safety, reliability 등을 통해 배터리, 안전과 신뢰성 주제와 관련이 높아 보인다. 토픽 15에는 다른 토픽에서 발견되지 않은 inverter, high voltage, high power 등이 관찰된다. 이는 전기차에서 모터를 구동하기 위한 전력변환 기술과 관련되어 있다. 토픽 16에는 integrated energy management system, light battery pack, re-electric vehicle, electric energy storage system 등의 키워드가 나타나며, 차량내 에너지 관리, 경량 배터리 팩, 주행거리 향상, 전기에너지 저장시스템 등 에너지 효율향상 기술개발을 전반적으로 포함하고 있는 것으로 보인다.

Table 3은 제2기의 토픽을 보여주고 있다. 토픽 21에 여러 주제의 키워드가 관찰되지만 핵심 키워드는 deep learning으로 판단된다. 2016년 3월 이세돌-알파고 대국을 기점으로 정부R&D에서 인공지능 관련 연구가 강조된 결과로 보이며, lidar 등을 통한 객체 인식 기술에서도 deep learning을 활용한 기술개발이 진행되었음을 알 수 있다. 토픽 22는 자율주행 실증과 관련된 주제로서, 이와 관련한 키워드가 많이 등장하고 있다. Autonomous driving, autonomous vehicle, electric test, v2x communication 등 자율주행 기술을 개발하고 테스트하고, 또 통신기술을 융합하는 기술개발을 진행하고 있는 것으로 판단된다. 토픽 23에서는 안전 및 에너지관련 기술 개발 주제를 포함하고 있다. Safety, rolling, high strength steel 등 주요 키워드로 부터 자

동차의 선회시 톨링을 저감하는 안전 기술 또는 고강도 스틸을 통해 자동차의 안전을 증가시키는 기술을 유추해 볼 수 있다. 또한 regulator 또는 energy storage system을 통해 자동차 내 전압을 안정적으로 공급하여 배터리의 수명을 연장시킬 수 있는 레귤레이터 기술이나 에너지 저장시스템과 관련된 기술개발이 진행되는 것으로 파악된다.

토픽 24는 전력변환 및 배터리 기술개발 주제를 다수 포함하고 있는 것으로 보인다. 특히 inverter, motor 및 high power는 전력변환 기술과 관련이 높고, battery management system 또는 efficiency는 배터리 기술개발과 연관성이 높다. 토픽 25에는 신호처리, 센서 및 소프트웨어 플랫폼 등이 복합적으로 포함되어 있다. Signal processing, sensor fusion, autosar 등이 주목할 만한 키워드이다. 여기에서 autosar는 AUTomotive Open System Architecture의 약자로서, 개방형 자동차 표준 소프트웨어 구조를 의미한다. 토픽 26에는 여러 다양한 키워드가 나타나지만, 운전자 지원 또는 고효율 기술과 관련되어 있다. 주요 키워드로는 ADAS, SOC,

high efficiency 등을 들 수 있다. ADAS는 운전자의 편의를 증대시키는 운전자 보조기술 개발과 관련되어 있고, SOC는 Systems On a Chip 약자로서 다양한 기능과 서브시스템을 하나의 칩에 통합하는 기술을 의미하며, 자동차의 효율을 향상시키는 기술의 일부로서 볼 수 있다.

Table 4의 토픽 31에서는 이전에는 나오지 않았던 새로운 키워드들이 나타나고 있다. Fuel cell, system, international standard, bipolar plate 등이 여기에 해당하는데, 연료전지 시스템에 대한 연구개발이 중점적으로 진행되고 있음을 알 수 있다. Bipolar plate는 수소연료전지 자동차용 분리판으로서 막-전극접합체와 기체확산층과 함께 스택의 핵심부품이다. 수소, 산소 및 냉각수를 막-전극접합체에 공급하고 전기화학반응으로 생성된 전류를 수집하여 전달하는 역할을 한다. 또한 연료전지 기술은 초기기술로서 향후 글로벌 시장 선점을 위해서는 국제 표준을 확보하는 것이 중요한데, 이러한 맥락에서 국제표준과 관련된 연구개발 과제가 지원된 것으로 판단된다. 토픽 32에는 자율주행 실증관련 기술개발 주제로 판단되며, 주요 키워드로

Table 2 Topic modeling results in period 1 (2013-2015)

Topic	Naming	Keywords
Topic 11	Sensor & communication	sensor fusion, integrated control, wave, optical sensor, reliability critical vehicle component, rse, obe, signal processing, pedestrian identification, utis
Topic 12	Driver assistance	acc, electric vehicle, adas, motor, high voltage, pedestrian detection, ecu, das, battery pack, 4wd
Topic 13	High-strength lightweight body	electric vehicle, high stiffness, green car, lightweight, platform, vacuum, autobody, traction module, electric-traction module, e-4wd module
Topic 14	Battery, safety & reliability	electric vehicle, secondary battery, safety, fusion, augmented reality, reliability, ecu, infrastructure, eps, multi-pore structure
Topic 15	Power conversion	electric vehicle, sensor fusion, inverter, high voltage, lithium ion battery, electrolyte additive li ion battery, noise, high power, assisted parking system, separator
Topic 16	Energy efficiency	electric vehicle, smart car, aeb, sensor fusion, magnet, integrated energy management system, integrated underbody package, light battery pack, re-electric vehicle(range extender electric vehicle), electric energy storage system

Table 3 Topic modeling results in period 2 (2016-2018)

Topic	Naming	Keywords
Topic 21	Deep learning	electric vehicle, deep learning, lidar, battery, anode, high voltage, commercialization, automotive parts, secondary battery, nano
Topic 22	Autonomous driving demonstration	electric vehicle, autonomous driving, autonomous vehicle, vehicle test, fuel cell, sensor fusion, smart car, v2x communication, laser scanner, vehicle
Topic 23	Safety & energy	safety, electric vehicle, hydrogen vehicle, vehicle, rolling, high strength steel, integrated system, regulator, ecu, energy storage system
Topic 24	Power conversion & battery	inverter, motor, smart car, electric vehicle, reliability, battery management system, sensor fusion, heat pump, high power, efficiency)
Topic 25	Signal processing, sensor & S/W platform	electric vehicle, signal processing, sensor fusion, autosar, sic, battery pack, high voltage, stereo camera, hydrogen storage system, cathode
Topic 26	Driver assistance & high efficiency	electric vehicle, adas, secondary battery, autonomous driving, intelligent vehicle, mosfet, soc, automotive, smart car, high efficiency

Table 4 Topic modeling results in period 3 (2019-2021)

Topic	Naming	Keywords
Topic 31	Fuel cell system	smart car, fuel cell, electric bus, micro, system, international standard, crash, electric mobility, certification, bipolar plate
Topic 32	Autonomous driving demonstration	autonomous vehicle, lightweight, autonomous driving, automated driving shuttle, simulation, integrated system, battery management system, intelligent vehicle, sensor fusion, verification autonomous driving
Topic 33	Autonomous driving & high efficiency	electric vehicle, autonomous driving, traction motor, high efficiency, reliability, battery pack, automated driving, automotive, micro electric vehicle, airless tire
Topic 34	Battery & power conversion	electric vehicle, autonomous driving, battery, platform, lithium ion battery, v2x, high voltage, battery pack, big data, internet things
Topic 35	AI & bigdata	electric vehicle, artificial intelligence, secondary battery, soc, big data, deep learning, automated vehicle, battery, ai, automated bus requirement
Topic 36	Sensor, etc	electric vehicle, sensor fusion, battery, light weight, electric vehicle management, fuel cell, fuel cell electric vehicle, inverter, artificial intelligence, battery pack

autonomous vehicle, autonomous driving, automated driving shuttle, simulation, verification autonomous driving 등이 관찰된다. 토픽 33은 자율주행 기술과 함께 효율향상 등 다른 주제들이 혼재되어 있는 것으로 보인다. Autonomous driving 및 automated driving이 나타나며, high efficiency 및 airless tire 등 타이어 경량화로 인해 연비를 향상시키는 기술들이 포함되어 있다.

토픽 34는 배터리 기술이 많이 포함되어 있으며 전력변환 기술도 일부 포함되어 있다. 주요 키워드로 battery, lithium ion battery, battery

pack 등이 있고, high voltage 등 전력관련 기술도 나타난다. 토픽 35는 AI와 빅데이터 관련 주제에 집중되어 있다. 기존의 딥러닝 기술에서 AI와 빅데이터 전반으로 기술이 확대 발전되고 있는 것으로 보인다. 주요 키워드로는 artificial intelligence, big data, deep learning, ai 등이 나타나고 있다. 토픽 36은 센서를 포함한 복합 주제를 담고 있다. 주요 키워드로 Sensor fusion 과 함께 light weight, battery, inverter 등이 곁고루 나타나고 있다.

Table 5 Topic change over the entire periods

Category	Period 1 (2013-2015)	Period 2 (2016-2018)	Period 3 (2019-2021)
A	(T11) Sensor & communication	(T25) Signal processing, sensor & S/W platform	(T36) Sensor, etc
B	(T12) Driver assistance	(T26) Driver assistance & high efficiency	(T33) Autonomous driving & high efficiency
C	(T14) Battery, safety & reliability (T15) Power conversion (T16) Energy efficiency	(T24) Power conversion & battery (T23) Safety & energy	(T34) Battery & power conversion
D	(T13) high-strength lightweight body		
E		(T21) Deep learning	(T35) AI & bigdata
F		(T22) Autonomous driving demonstration	(T32) Autonomous driving demonstration
G			(T31) Fuel cell system

3기에 걸친 토픽의 전반적인 변화는 Table 5에 정리되어 있다. 카테고리 A, B와 C에 대한 연구개발은 3기에 걸쳐 지속적으로 투자되고 있음을 알 수 있다. 카테고리 A는 센서, 통신, 소프트웨어에 대한 주제로서 자율주행 기술개발을 위해 필수적이며 지속적인 개선이 필요한 분야이다. 카테고리 B는 운전자 보조기술로서 편의성 증대를 위해 1기와 2기에 지속적으로 연구개발이 진행되었으며, 3기에 와서는 자율주행 기술이라는 큰 주제로 발전되고 있다. 카테고리 C는 배터리, 전력변환, 에너지 효율향상과 관련된 주제가 모두 포함되어 있다. 전기자동차에 있어 배터리 성능이 가장 중요하기에 3기에 오면서

배터리 기술에 집중적으로 연구개발이 진행되는 것으로 판단된다. 카테고리 D는 고강도 경량차체와 관련한 기술개발로서 1기에는 핵심 주제로 나타났으나, 2기와 3기로 가면서 다른 주제에 비해 비중이 상대적으로 줄어들어 핵심 주제에서 사라진 것으로 보인다. 이와는 반대로 카테고리 E에서 나타난 바와 같이 딥러닝, AI와 빅데이터는 2기와 3기로 오면서 점점 중요성이 커져 핵심 주제로 부상함을 알 수 있다. 또는 카테고리 F와 G를 통해, 자율주행 기술의 사업화를 위한 실증연구가 2기부터 활발히 진행되어 오고 있으며, 3기에 들어서서는 수소연료전지 기술개발에 집중적인 투자가 이루어지고 있는

것으로 판단된다.

### 4.3 토픽별 투자 동향 분석

토픽별 투자 동향을 살펴보기 위해 LDA 분석을 통해 확보한 문서별 토픽의 분포( $\theta$ )를 활용하였다. 즉, 이것은 각 연구개발과제별 토픽의 구성 비중을 의미한다. 각 연구개발과제에 투입된 정부출연금을 토픽의 구성 비중에 따라 각 토픽에 투입되는 금액으로 모두 세분화하였다. 이렇게 산출한 금액을 토픽을 기준으로 다시 집계하여, 기수별, 연도별로 토픽에 얼마만큼의 정부출연금이 지원되었는지를 파악하였다. 각 기수는 3개년으로 구성되어 있는데, 해당 기수안에서 각 토픽의 3년간의 상대적 증가율과 상대적 지원비중을 기준으로 유망 토픽을 확인하였다. Table 6은 각 기수에서 토픽별로 지원된 정부출연금 및 기간 내 증가율을 보여주고 있다. 또한 이 절대값에서 평균값을 차감하여 상대적인 정부출연금 및 증가율을 계산하였다. 이러한 두 종류의 상대값을 기준으로 유망 토픽을 파악하기 위한 매트릭스를 Fig. 5에서 제시하였다. 1사분면에 위치한 토픽은 투자증가율이 상대적으로 높고 지원금액도 상대적으로 많으므로 유망하다고 인식되는 분야로 볼 수 있다.

1기 (2013 ~ 2015)를 살펴보면 토픽 12 (Driver assistance)와 토픽 15 (Power conversion)가 1사분면에 위치해 있다. 당시에는 운전자 지원 기술과 전력변환 기술이 상당히 강조되었음을 짐작할 수 있다. 즉 전기차의 전력변환 기술, 자율주행을 위한 기초단계로서 스마트카를 위한 운전자 편의기술의 대폭적인 수준향상에 상대한 무게를 두고 있었던 것으로 보인다. 반면 토픽 11 (Sensor & communication)과 토픽 13 (High-strength lightweight body)에는 상대적으로 지원이 축소되는 것으로 관찰된다. 센서 등 특정 부품 단위의 기술개발보다는 센서를 활용한 운전자 지원기술 등 통합 기술개발로 방향이 전환되는 것으로 보이며, 고강도 차체와 같이 전통 자동차에서도 강조해 왔던 분야는 상대적으로 주목을 덜 받는 것으로 판단된다.

2기에는 토픽 22 (Autonomous driving

demonstration)가 유망 토픽으로 부상하였다. 1기에는 주요 토픽으로 나타나지 않는 자율주행 기술인데, 이 시기에 본격적으로 지원되고 있다. 자율주행 기술은 기술개발과 실증이 동시에 진행되어야 하는 분야로서 경쟁국과의 기술격차를 좁히기 위해 집중적으로 정부가 지원하고 있는 것으로 보인다. 1사분면 안에는 위치하고 있지는 않지만 1사분면과 가까운 곳에 토픽 21 (Deep learning)이 나타나고 있다. 아직 충분한 지원은 되지 않고 있으나, 증가율이 상대적으로 상당히 높다. 동 시기에 딥 러닝 기술의 급속한 진전과 함께 이를 자동차 분야에 응용하려는 적극적 시도가 이루어진 것으로 판단된다.

3기에 들어서면 토픽 34 (Battery & power conversion)와 토픽 33 (Autonomous driving & high efficiency)이 1사분면에 위치해 있다. 이 시기에는 미래자동차의 가장 필수적인 기술인 배터리 기술에 다시 집중하고 있는 것으로 보인다. 결국 기존의 화석연료기반의 전통 자동차에서 미래 자동차로 성공적으로 전환되기 위해서는 배터리의 충전시간 단축이나 주행거리 향상은 가장 핵심적인 요소이다. 또한 자율주행과 함께 고효율, 고신뢰성도 지속적으로 관심을 받고 있다.

1기부터 3기에 걸쳐 전체적으로 바라보면, 정부가 지원해온 톱 유망분야는 1기의 운전자 지원기술에서 2기의 자율주행 실증으로, 3기의 배터리 및 전력변환 기술로 변화해 오고 있다.

마지막으로 Table 5에서 구분한 토픽별 카테고리 기준을 기준으로 하여 3기간에 걸친 총 정부출연금의 규모를 정리하였다. 본 결과는 Table 7과 Fig. 6에서 볼 수 있다. 지난 3기에 걸쳐 산업통상자원부가 미래자동차 분야의 기술개발에 투자한 금액은 총 1조 6000억원에 달한다. 1기에 투자한 총 금액이 약 3,300억원, 2기에는 약 4,500억원, 3기에는 무려 약 8,400억원의 예산이 투입되었다. 미래 자동차가 여러 산업 중에서 유망한 분야이기도 하지만, 2020년에 기획재정부의 혁신성장을 위해 발표한 빅3(BIG3) 정책의 영향이 컸던 것으로 보인다. 당시 정부는 미래차, 시스템 반도체, 바이오헬스에 대한 대대적인 산업 육성정책을 발표한 바 있다. 카테고리별로

Exploring Key Topics and Trends of Government-sponsored R&D Projects in Future Automotive Fields: LDA Topic Modeling Approach

구체적으로 살펴보면, 배터리와 전력변환 기술과 관련된 카테고리 C영역에 가장 많은 5,000억 원 규모의 투자가 이루어졌다. 그 다음으로 운전자 지원기술을 다루는 카테고리 B와 센서 및 통신분야인 카테고리 B순으로 투자가 진행된 것을 알 수 있다.

Table 6 Government funding amount and ratio of increase according to topics in each periods

Period 1 (2013 ~ 2015)	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15	Topic 16
Government fund (100 million won)	394	570	480	724	567	640
Ratio of increase (within each period)	-14%	142%	-21%	-44%	49%	-23%
Government fund (relative)	-168	7	-83	161	4	77
Ratio of increase (relative)	-29%	127%	-36%	-59%	34%	-38%
Period 2 (2016 ~ 2018)	Topic 21	Topic 22	Topic 23	Topic 24	Topic 25	Topic 26
Government fund (100 million won)	740	931	717	827	789	582
Ratio of increase (within each period)	143%	151%	44%	48%	0%	59%
Government fund (relative)	-24	167	-47	63	25	-183
Ratio of increase (relative)	69%	77%	-30%	-26%	-74%	-15%
Period 3 (2019 ~ 2021)	Topic 31	Topic 32	Topic 33	Topic 34	Topic 35	Topic 36
Government fund (100 million won)	1181	1222	1487	1620	1552	1344
Ratio of increase (within each period)	26%	74%	74%	79%	29%	53%
Government fund (relative)	-220	-179	86	219	151	-57
Ratio of increase (relative)	-30%	18%	18%	23%	-27%	-3%

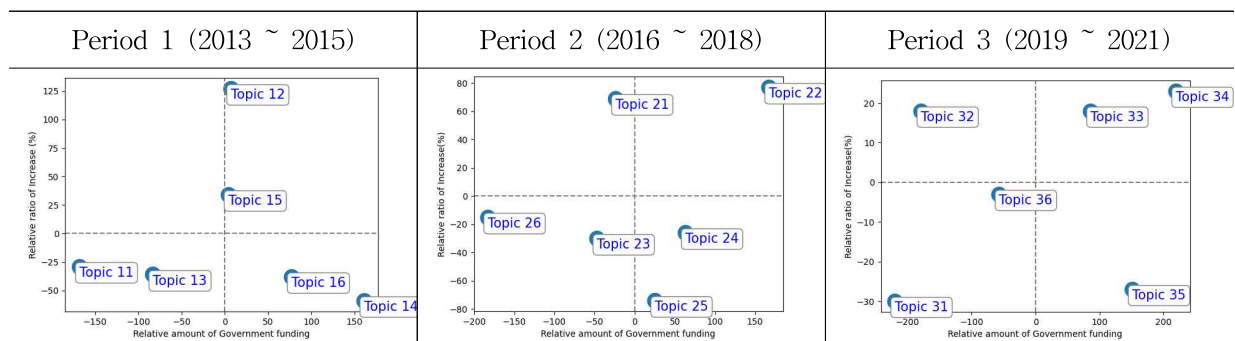


Fig. 5 Matrix diagram for identifying promising topics in each periods

Table 7 Government funding amount according to categories in each periods (unit: 100M won)

Category	Topics	Period 1 (2013-2015)	Period 2 (2016-2018)	Period 3 (2019-2021)	Total (2013-2021)
A	(T11) Sensor & communication (T25) Signal processing, sensor & S/W platform (T36) Sensor, etc	394	789	1344	2527
B	(T12) Driver assistance (T26) Driver assistance & high efficiency (T33) Autonomous driving & high efficiency	570	582	1487	2639
C	(T14) Battery, safety & reliability (T15) Power conversion (T16) Energy efficiency (T24) Power conversion & battery (T23) Safety & energy (T34) Battery & power conversion	1931	1544	1620	5095
D	(T13) high-strength lightweight body	480			480
E	(T21) Deep learning (T35) AI & bigdata		740	1552	2292
F	(T22) Autonomous driving demonstration (T32) Autonomous driving demonstration		931	1222	2153
G	(T31) Fuel cell system			1181	1181
		3,375	4,586	8,406	16,367

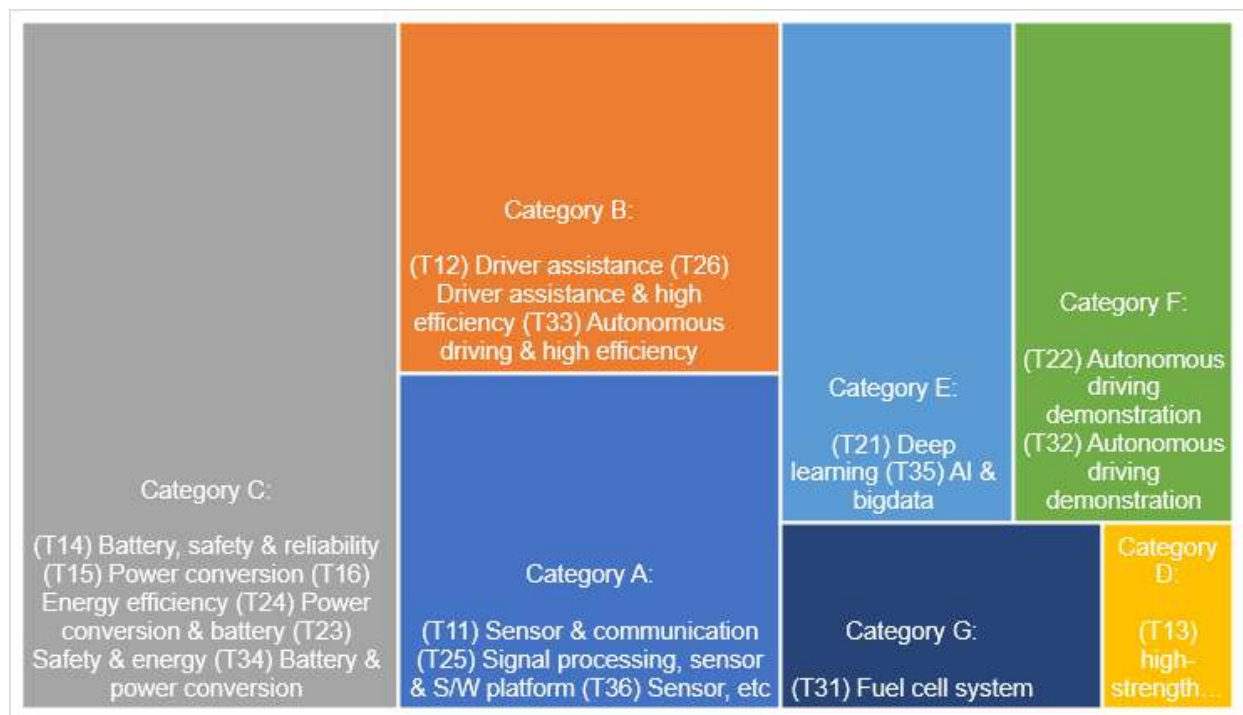


Fig. 6 Treemap chart for total government funding on each category over the three periods

## 5. 결론

본 연구에서는 미래 자동차분야의 국가연구개발사업을 대상으로 그 간의 주요 연구 토픽과 투자 동향을 분석하였다. 2013년부터 2021년까지 산업통상자원부가 지원된 국가연구개발과제 데이터를 모두 수집한 후, 기술분야 전문가의 검토 및 머신러닝 기법을 활용하여 미래자동차 대상과제를 확정하였다. 총 1,054개 과제를 대상으로 3개의 기간으로 구분하여 토픽모델링을 수행하고 각 기수별 주요 연구토픽 및 투자동향을 파악하였다. 본 연구의 주요 결과를 요약하면 다음과 같다. 첫째, 센서, 통신, 소프트웨어 등 카테고리 A 영역, 운전자 보조기술 등 카테고리 B 영역, 그리고 배터리, 전력변환, 에너지 효율 향상 등 카테고리 C 영역에는 전체 기간 (2013 ~ 2021)에 걸쳐 꾸준히 투자되고 있음을 알 수 있다. 둘째, 고강도 경량차체와 관련한 기술개발로서 1기에는 핵심 주제로 나타났으나, 2기와 3기로 가면서 다른 주제에 비해 비중이 상대적으로 줄어들어 핵심 주제에서 사라졌다. 셋째, 딥러닝, AI와 빅데이터, 자율주행 기술의 사업화를 위한 실증연구는 2기로 오면서 점점 중요성이 커져 핵심 주제로 부상하였으며, 3기에 들어서서는 수소연료전지 기술개발에 집중적인 투자가 이루어지고 있다. 넷째, 각 기수별로 톱 유망분야는 1기의 운전자 지원기술에서 2기의 자율주행 실증으로, 3기의 배터리 및 전력변환 기술로 변화하고 있다. 다섯째, 카테고리별로 전 기간에 걸친 투자금액을 살펴보면, 배터리와 전력변환 기술과 관련된 카테고리 C 영역에 가장 많은 5,000억원 규모의 투자가 이루어졌다. 그 다음으로 운전자 지원기술을 다루는 카테고리 B 영역과 센서 및 통신분야인 카테고리 B 영역순으로 투자가 진행되었다. 마지막으로, 미래자동차 분야에 대한 정부의 투자금액은 특히 3기 (2019 ~ 2021)에 대폭 증가하였다. 1기에 투자한 총 금액이 약 3,300억원, 2기에는 약 4,500억원, 3기에는 무려 약 8,400억원의 예산이 투입되었다. 미래 자동차가 여러 산업 중에서 유망한 분야이기도 하지만, 2020년에 기획재정부가 혁신성장을 위해 발표한 빅3(BIG3) 정책의 영향이 컸던

것으로 보인다.

본 연구결과를 통해 지난 10년간 미래자동차 분야에 대한 정부의 연구동향과 투자동향을 상세히 알 수 있었다. 향후 미래자동차 분야 정책 수립이나 연구개발 로드맵을 수립하는 과정에서 본 연구결과가 기초자료로 검토되고 활용되어 증거기반의 정책 및 전략 수립에 기여할 것으로 예상된다. 다만, 본 연구에서는 산업통상자원부 R&D과제만을 대상으로 분석하였기에, 국가연구개발 전체의 동향을 완전히 파악하는 데는 한계가 존재한다. 또한 본 연구에서는 정부가 투자한 연구주제와 투자금액 등 투입관점에 초점을 맞추었다. 향후 연구에서는 투입관점과 함께 산출 관점까지 동시에 고려한다면, 보다 의미 있는 결과를 도출할 수 있을 것으로 판단된다. 최근에 연구개발 패러독스 등 정부 연구개발 투자의 성과에 대한 전반적인 우려가 점증하고 있다 (Ma et al., 2022). 미래 자동차 분야의 연구개발 과제에서 산출된 논문이나 특허를 활용하여 성과가 나오는 토픽까지 동시에 분석한다면, 투자가 집중되는 토픽과 성과가 산출되는 토픽 간의 상호 비교를 통해 정부 연구개발 투자의 효과성을 제시할 수 있을 것으로 기대된다.

## References

- Blei, D., Ng, A. and Jordan, M. (2003). Latent Dirichlet Allocation, *The Journal of Machine Learning Research*, Vol. 3, pp. 993-1022.
- Choi, D., Song, B., Park, D. and Lee, S. (2022). Keyword trends analysis related to the aviation industry during the Covid-19 period using text mining, *Journal of the Korea Industrial Information Systems Research*, Vol. 27, No. 2, pp. 115-128.
- Chung, M. and Lee, J. (2018). Systemic Analysis of Research Activities and Trends Related to Artificial Intelligence(A.I.) Technology Based on Latent Dirichlet Allocation (LDA) Model, *Journal of the Korea Industrial Information Systems Research*, Vol. 23, No.



- 3, pp. 87-95.
- Deerwester, S., Dumais, S., Furnas, G., Landauer, T. and Harshman, R. (1990). Indexing by latent semantic analysis, *Journal of the American Society for Information Science*, Vol. 41, No. 6, pp. 391-407.
- Hofmann T. (1999). Probabilistic latent semantic indexing, *In Proceedings of the 22nd annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, pp 50 - 57.
- Hong, T., Niu, H., Ren, G. and Park, J. (2018). Multi-Topic Sentiment Analysis using LDA for Online Review, *The Journal of Information Systems*, Vol. 27, No. 1, pp. 89-110.
- Jo, H., Kim, J. and Lee, B. (2019). A Study on Research Trends of Blockchain Using LDA Topic Modeling : Focusing on United States, China, and South Korea, *Journal of Digital Contents Society*, Vol. 20, No. 7, pp. 1453-1460.
- Kim, C., Choi, S. and Kwahk, K. (2017). Investigation of Research Trends in Information Systems Domain Using Topic Modeling and Time Series Regression Analysis, *Journal of Digital Contents Society*, Vol. 18, No. 6, pp. 1143-1150.
- Kim, K. (2021). A Study on the Analysis of R&D Trends and the Development of Logic Models for Autonomous Vehicles, *Journal of Digital Convergence*, Vol. 19, No. 5, pp. 31-39.
- Kim, S., Park, D. and Kim, B. (2021). Analysis of Airbnb customers' reviews using Topic Modeling: focusing on the difference in topics before and after the COVID-19 Pandemic, *The Journal of Internet Electronic Commerce Research*, Vol. 21, No. 4, pp. 115-130.
- KIET. (2017). *The Influence and Challenges of the Fourth Industrial Revolution on Korean Major Industries*, Korea Institute for Industrial Economics and Trade.
- KEIT. (2022). *2021 Industrial technology level survey report*, Korea Planning & Evaluation of Industrial Technology.
- Lim, C., Ha, J., Cho, K., and Ha, H. (2021). Analyzing Customers' Perception of Service and Market Trends in Fresh Food E-commerce: Application of Text Mining Techniques, *The Journal of Internet Electronic Commerce Research*, Vol. 21, No. 4, pp. 169-182.
- Ma, H., Lee, C. and Oh, D. (2022). Identification of R&D Paradox after the Global Financial Crisis: Korean Case Focusing on the Smart Convergence and Conventional Industries, *Science, Technology and Society*, 27(2), 213-232.
- MOTIE. (2019). *2019 - 2021 Industrial Technology R&D Investment Strategy*, Ministry of Trade, Industry and Energy.
- Nam, D. and Choi, G. (2018). Technology Trend Analysis in the Automotive Semiconductor Industry using Topic Model and Patent Analysis, *Journal of Korea Technology Innovation Society*, Vol. 21, No. 3, pp. 1155-1178.
- Park, J., Hong, S. and Kim, J. (2017). A Study on Science Technology Trend and Prediction Using Topic Modeling, *Journal of the Korea Industrial Information Systems Research*, Vol. 22, No. 4, pp. 19-28.
- Park, K. and Lee, C. (2019). A Study on the Research Trends for Smart City using Topic Modeling, *Journal of Internet Computing and Services*, Vol. 20, No. 3, pp. 119 - 128.



**마 형 렬 (Ma Hyoung Ryul)**

- 회원
- 연세대학교 기계설계 학사
- 연세대학교 MBA 경영석사
- 뉴욕주립대(스토니브룩) 기술 혁신정책박사
- 산업통상자원 R&D 전략기획단

수석전문위원

- 한국뉴욕주립대 기술경영학과 겸임교수
- 관심분야: 기술혁신시스템, 개방형 혁신, R&D 투자효율성, 혁신역량 측정, 데이터 분석



**이 철 주 (Lee Cheol-Ju)**

- 서울대학교 섬유고분자 학사
- 서울대학교 재료공학 석사
- 성균관대학교 기술경영학과 공학 박사
- Virginia Polytechnic Institute and State University 방문학자

- 한국산업기술기획평가원 산업기술혁신본부 수석 연구원
- 관심분야 : 국가연구개발, 기술경영, 기술정책