

토픽 모델링을 이용한 전기차 연구 동향 분석

천위안*, 조석수**

The Analysis of Research Trends in Electric Vehicle using Topic Modeling

Yuan Chen*, Seok-Swoo Cho**

요약 환경 문제와 에너지 효율성을 향상시키기 위하여 전기자동차를 도입함으로써 이에 따른 연구가 급증하고 있다. 그러나 전기 자동차 분야의 연구 동향을 전체적으로 파악기 위해서는 방대한 데이터를 체계적으로 분석할 필요가 있다. 본 연구에서는 SCIE 데이터베이스에서 수집한 36,519편의 전기 자동차 관련 논문을 바탕으로 LDA주제 모델링을 수행하여 전기 자동차 분야의 연구 동향을 체계적으로 분석하고 주요 연구 주제를 파악하였다. 데이터 분석 결과, 총 10개의 주요 주제가 도출되었으며, 이 중 3개 주제는 상승 추세를 보이는 핫 토픽으로 확인되었으며 그 분야는 Electric Vehicle Charging Infrastructure, Energy and Environmental Policy, Optimization and Algorithms이었다. 그러나, 5개 주제는 하락 추세를 보이는 콜드 토픽으로 분류되었으며 그 분야는 Battery Temperature and Cooling, Battery Materials and Chemistry, Motor and Mechanical Design, Control Strategies and Systems, Battery Components and Materials이었다. 본 연구에서는 전기 자동차 분야의 최신 연구 동향을 이해하는 데 중요한 기초 자료를 제공하였으며, 전기자동차 관련 연구자가 연구 주제 선정에 필요한 유용한 정보를 제공하였다.

Abstract To address environmental challenges and improve energy efficiency, the adoption of electric vehicles has led to a surge in related research. However, to comprehensively understand the research trends within the field of electric vehicles, it is necessary to systematically analyze vast amounts of data. This study systematically analyzed research trends in the field of electric vehicles and identified key research topics through LDA topic modeling, based on 36,519 papers related to electric vehicles collected from the SCIE database. The data analysis revealed a total of 10 major topics, of which three were identified as hot topics showing an upward trend: Electric Vehicle Charging Infrastructure, Energy and Environmental Policy, and Optimization and Algorithms. Conversely, five topics were identified as cold topics exhibiting a downward trend: Battery Temperature and Cooling, Battery Materials and Chemistry, Motor and Mechanical Design, Control Strategies and Systems, and Battery Components and Materials. This study provides basic data for understanding the current research trends in electric vehicles and offers valuable information for researchers in selecting research topics related to electric vehicles.

Key Words : Electric vehicle, LDA, Topic modeling, Research topic, Linear regression model

1. 서론

최근 글로벌 오염 수준의 심화와 화석 자원의 지속 가능성에 대한 우려로 인해 온실가스 배출과 같은 환

경 문제에 대한 관심이 커지고 있다. 2022년 데이터에 따르면 코로나19와 관련된 이동 제한 해제 이후 교통 분야의 이산화탄소 배출 비율이 상승세를 보이며 약 37%에 도달하고 있다[1]. 전통적인 내연기관(ICE) 차

*Department of Mechanical-Computer-Industrial Engineering, Graduate School, Kangwon National University

**Corresponding Author : Division of Mechanical Design Engineering, Kangwon National University (sscho394@kangwon.ac.kr)

Received August 13, 2024

Revised August 25, 2024

Accepted August 27, 2024

량의 이산화탄소 배출은 지구 온난화 문제를 악화시키고 있다. 차량 전동화는 교통 분야의 온실가스 배출을 줄이는 중요한 방법이다. 최근 통계에 따르면, 2022년 말까지 전 세계 전기차 판매량(배터리 전기차와 플러그인 하이브리드 전기차 포함)이 1,052만 대로 급증했으며 이는 전년 대비 55% 증가한 수치이다 [2]. 동시에, 전 세계 자동차 산업은 전기차에 대한 대규모 투자를 진행하고 있다. 유럽에서는 유럽연합 자동차 산업이 배출가스 이산화탄소 목표를 강화하였고, 영국 런던에는 초저배출구역(ULEZ)이 도입되었다[3]. 한국에서는 “2030 국가 온실가스 감축 목표(NDC) 상향안”에 따라 2030년까지 전체 등록 차량 약 2700만 대 중 전기차 및 수소차의 보급 대수가 450만 대(16.7%)에 이를 것이라고 예측하고 있다[4].

전기차에 대한 관심이 높아지면서 판매량이 계속 증가하고 있을 뿐만 아니라 전기차 분야에 대한 다양한 연구도 날로 증가하고 있으며 이로 인해 전기차에 관한 문헌이 많이 생성되고 동시에 몇 가지 문제가 발생되고 있다. 첫째는 전기차 분야 문헌의 급증으로 인해 연구자들이 방대한 문헌 중에서 자신의 연구가 얼마나 유용한지를 알기 어렵다. 둘째는 전기차 분야의 빠른 발전에 따라 현재의 연구 동향과 핫이슈를 파악하는 것이 점점 더 복잡해지고 있다. 따라서 연구자들은 미래 연구 방향을 정확히 이해하고 예측하는 데 어려움을 겪고 있다. 본 논문에서는 LDA (Latent Dirichlet Allocation) 토픽 모델링을 이용하여 전기차 분야 관련 문헌의 연구 토픽을 분류하고 분석하여 이를 통해 관련 연구자들과 이해관계자들에게 지난 20년간 전기차 분야의 연구 동향에 대한 명확한 통찰력을 제공하였으며 앞으로 고려해야 할 토픽들을 제시하였다. 본 논문의 나머지 부분은 다음과 같이 구성하였다. 제2장에서는 관련 연구를 논의하고 제3장에서는 채택된 연구 방법을 소개하고 제4장에서는 LDA 토픽 모델링 결과를 분석하였고 제5장에서는 본 연구의 의미와 결론을 설명하였다.

2. 관련 연구

이전 연구에서 Peng[5] 등은 공간적 및 통계적 모

델을 사용하여 전기차 시장 점유율에 영향을 미치는 요인을 조사하였고 미국과 유럽 간의 지역적 차이점과 유사점을 규명하였다. Cheng[6] 등은 전기차 배터리의 공급망을 연구하고 전기차 배터리의 재료 화학 성분과 전기차 공급망 간의 관계를 분석하였다.

Farinloyee[3] 등은 31명의 전기차 분야 전문가를 인터뷰하여 나이지리아에서 전기차 도입 과정에서 직면하는 도전과 기회를 분석하였다. 구[7] 등은 키워드 네트워크 분석 기술을 이용해 한국의 전기차 관련 특허를 분석하고 연도별로 관련 특허의 키워드를 정리 및 분석하여 전기차 특허 기술의 트렌드를 제공하였다. 최[8]은 토픽 모델링을 이용하여 한국과 중국의 전기차 특허를 분석하고 비교하였으며 한국과 중국의 전기차 분야 기술 경쟁력과 기술의 특징을 규명하였다.

토픽 모델링은 대규모 비구조적 텍스트 데이터에서 잠재 변수를 찾기 위한 인기 있는 통계 도구로서 LDA가 현재 널리 사용되는 토픽 모델링 중 하나이다. LDA 기반의 토픽 모델링은 다양한 연구에 널리 사용되고 있다. 예를 들어, 특허 개발 지도 생성[9], 소셜 미디어에서 사건을 자동으로 추출하여 범죄 예측[10], 환자의 전자 건강 기록에서 잠재적 치료 패턴 추출[11], 호텔 산업의 온라인 리뷰 데이터에서 고객 만족도 차원 추출[12], 중국과 미국의 온라인 의사 리뷰 비교[13] 등이다. LDA 기반의 토픽 모델링은 제품 계획 과정에서 사용되는 신제품 개발에도 사용되며[9], 제품 기술의 하위 기술 토픽과 경쟁정보계획을 결정하는 데에도 이용된다[14]. 이것으로부터 LDA 기반의 토픽 모델링이 대규모 말뭉치에서 잠재적 토픽을 식별하는 효과적인 기술임을 알 수 있다.

기존의 전기 자동차 연구 토픽은 주로 해당 분야의 특정 토픽에 집중되거나 특허 문헌을 토픽으로 한 연구가 많았다. 또한 이러한 연구는 주로 단일 국가나 지역에 국한되어 있었다. 전기 자동차 분야의 비특허 문헌이나 글로벌 범위의 전기 자동차 문헌에 대한 분석은 매우 부족한 실정이다. 따라서 본 연구에서는 LDA 토픽 모델링을 이용하여 SCIE 데이터베이스를 통해 전세계에서 출판된 전기 자동차 문헌의 토픽을 식별하고 연구 동향을 탐색하고자 한다.

표 1. 전기차 연구 검색 방법

Table 1. The search method for electric vehicle research

TOPIC (title-abstract-keywordplus-author keywords)	
Search Terms	"electric car*" (Topic) or "new energy car*" (Topic) or "alternative fuel car*" (Topic) or "Eco-Friendly car*" (Topic) or "Green car*" (Topic) or "fuel cell car*" (Topic) or "hybrid car*" (Topic) or "battery car*" (Topic) or "electric automobile*" (Topic) or "new energy automobile*" (Topic) or "Eco-Friendly automobile*" (Topic) or "Green automobile*" (Topic) or "fuel cell automobile*" (Topic) or "hybrid automobile*" (Topic) or "battery automobile*" (Topic) or "electric vehicle*" (Topic) or "new energy vehicle*" (Topic) or "Eco-Friendly vehicle*" (Topic) or "Green vehicle*" (Topic) or "fuel cell vehicle*" (Topic) or "hybrid vehicle*" (Topic) or "battery vehicle*" (Topic)
SICE Literature Categories	(1) Engineering (Engineering Electrical & Electronic, Engineering Environmental, Engineering Chemical, Engineering Mechanical, Engineering Multidisciplinary, Engineering Industrial, Engineering Manufacturing) (2) Electrical & Computer Science (Telecommunications, Automation Control Systems, Computer Science & Information Systems, Operations Research Management Science, Computer Science Theory & Methods, Computer Science Artificial Intelligence, Computer Science Interdisciplinary Applications, Computer Science Hardware Architecture, Computer Science Software Engineering, Computer Science Cybernetics) (3) Energy & Fuels (Energy and Fuels) (4) Environmental Studies (Environmental Studies, Environmental Sciences, and Green & Sustainable Science Technology) (5) Transportation (Transportation Science & Technology, and Transportation) (6) Materials Science & Chemistry (Materials Science Multidisciplinary, Materials Science Coating Films, Materials Science Characterization Testing, Chemistry Physical, Electrochemistry, Chemistry Multidisciplinary, Chemistry Analytical, Chemistry Applied, Chemistry Inorganic Nuclear, Chemistry Organic, Materials Science Composites, Materials Science Biomaterials, and Nanoscience Nanotechnology).

3. 연구방법

3.1 데이터 수집 및 전처리

본 연구에서는 SCIE 데이터베이스를 사용하여 전기차 관련 연구 문헌을 수집하였으며, 검색 기간은 2004년 1월 1일부터 2023년 12월 31일까지이다. 언어는 영어로 제한하였고, 문헌 유형은 "Article"로 제한하였다.

표 1에는 본 연구의 검색어와 검색 범위가 제공되어 있다. 검색어는 전기 자동차와 그 동의어를 선택하였으며, 검색 범위는 기존 연구[15]를 참고하여 전기 자동차 기술과의 연관성이 적은 연구 분야를 제외하고, 검색 범위를 (1) Engineering (2) Electrical & Computer Science (3) Energy & Fuels (4) Environmental Studies (5) Transportation (6) Materials Science & Chemistry의 6가지로 나누었으며 총 37,902편의 논문을 검색되었고, 전문 초록이

없거나 출판 연도가 부족하여 분석에 적합하지 않은 논문을 삭제했고 최종적으로 36,519개의 논문을 분석에 사용하였다. 36,519개의 논문에서 영문 초록만을 추출하여 LDA 토픽 모델링을 위한 전처리 과정을 거친다. 특수문자 제거, 토큰화, 품사 태깅 과정을 거쳐 명사만을 추출한 후 표제어로 변환시켰다. 불용어(Stop word)를 삭제하기 위해 Python NLTK(Natural Language Tool Kit) 패키지에서 제공하는 불용어를 기본으로 사용하였다.

3.2 특징 추출

전처리된 문서에서 특징을 추출하기 위해서는 자연어 처리를 위한 다양한 특징 추출 기법이 사용되고 있다. 본 연구에서는 문서를 벡터화하고 특징을 추출하기 위해 TF-IDF[16] (단어 빈도 - 역문서 빈도) 기법을 적용했다. TF-IDF는 정보 검색과 자연어 처리에서 자주 사용되는 통계적 기법으로, 특정 용어가 문서 집합

내에서 얼마나 중요한지를 결정한다. 이 방법은 문서의 단어를 중요도 숫자로 변환하는 텍스트 벡터화 방법이다.

TF-IDF는 주로 문서 내에서 단어의 중요도를 계산하는 데 사용된다. 단어의 중요도는 주로 문서에서의 출현 빈도에 의해 결정되며, 빈도가 높을수록 중요성이 높다고 평가되고, 빈도가 낮을수록 중요성이 낮다고 평가된다. 그러나 단어의 출현 순서를 고려하지 않는 Bag of Words (BOW) 기법은 대규모 코퍼스에서는 효과가 떨어질 수 있다 [17]. 이 문제를 해결하기 위해 TF-IDF 기법이 개발되었으며, 이를 통해 스케일링 또는 정규화 계수를 사용하여 계산의 정확성을 높일 수 있다. TF-IDF의 계산 방법은 다음식과 같다[17].

(a) TF(단어 빈도)

각 문서 d에 대해, 각 단어 t의 단어 빈도 TF(t,d)를 계산한다:

$$TF(t,d) = \frac{\text{문서d에서 단어t가 등장한 횟수}}{\text{문서d에 등장한 모든 단어의 수}} \quad (1)$$

(b) IDF(역문서 빈도)

각 단어 t의 역문서 빈도 IDF(t,D)를 계산한다:

$$IDF(t,D) = \log\left(\frac{\text{총문서의 개수}}{\text{단어t를 포함하는 문서의 수}}\right) \quad (2)$$

만약 어떤 단어가 모든 문서에 등장한다면, IDF(t,D)값은 낮게 나온다. 반대로, 특정 단어가 소수의 문서에만 등장한다면, IDF(t,D)값은 높아진다.

(c) TF-IDF 계산

각 문서 d에서 각 단어 t의 TF-IDF 값을 계산한다:

표 2. TF-IDF 코퍼스 예시

Table 2. TF-IDF corpus example

Content after removing stop words	TF-IDF corpus
['backdrop', 'concern', 'clim', 'ate', 'change', 'air', 'pollutio', 'n', 'contribution', 'petroleum', 'combustion', 'transport', 'sector', 'issue', 'china', 'ra', 'nge', 'policy', 'incentive', 'u	[(32, 0.1258), (56, 0.0371), (80, 0.04572), (82, 0.0131), (91, 0.1096), (95, 0.05409), (126, 0.0713), (143, 0.057), (218, 0.0497), (235, 0.1238), (256, 0.01725), (331,

ptake', 'vehicle', 'incentive', 'increase', 'stock', 'since', '2010s', 'trend', 'year', 'woul', 'turn', 'lead', 'ev', 'mark', 'et', 'objective', 'project', 'un', 'derstanding', 'effect', 'policy', 'incentive', 'ev', 'data', '4', '3', 'city', '2011', '2017', 'pol', 'cy', 'ev', 'driving', 'stage', '74', 'novelty', 'paper', 'policy', 'e', 'ffect', 'adoption', 'behavior', 'adoption', 'intention', 'study', 'result', 'abolishment', 'pu', 'rchasing', 'restriction', 'gove', 'rnnment', 'subsidy', 'tax', 'ex', 'emption', 'license', 'diffusio', 'n', 'effect', 'two', 'restriction', 'study', 'abolishment', 'pu', 'rchasing', 'restriction', 'ev', 'market', 'share', 'policy', 'c', 'onsumer', 'motivation']	(0.0410), (347, 0.0731), (669, 0.0772), (896, 0.0815), (1113, 0.0745), (1118, 0.1193), (1131, 0.0041), (1157, 0.05457), (1179, 0.2975), (1210, 0.0779), (1211, 0.0703), (1244, 0.0658), (1261, 0.1091), (1269, 0.0638), (1270, 0.0862), (1272, 0.0690), (1280, 0.0784), (1311, 0.0558), (1323, 0.1426), (1356, 0.0777), (1357, 0.2644), (1374, 0.0666), (1398, 0.0923), (1399, 0.1141), (1400, 0.0628), (1405, 0.0850), (1432, 0.1402), (1445, 0.1099), (1463, 0.1053), (1604, 0.0955), (1620, 0.0663), (1631, 0.0703), (1665, 0.0697), (1673, 0.0795), (1676, 0.0940), (1679, 0.0569), (1753, 0.083), (1774, 0.1193), (1784, 0.1210), (1869, 0.1042), (2456, 0.1898), (2716, 0.2806), (3625, 0.3347), (3785, 0.1198), (4013, 0.1043), (4618, 0.1514), (4812, 0.1096), (7674, 0.4109), (7675, 0.1872)]
--	--

$$TF-IDF(t,d,D) = TF(t,d) \times IDF(t,D) \quad (3)$$

이것은 단어 t가 문서 d에서 얼마나 중요한지를 나타내며, 단어가 해당 문서에서 얼마나 자주 등장하는지와 전체 코퍼스(corpus)에서 얼마나 희귀한지를 모두 고려한다.

위에서 계산한 TF-IDF 값을 사용하여 TF-IDF 코퍼스를 구축한다. 표 2는 TF-IDF를 사용하여 문서를 벡터화하여 얻은 코퍼스 예시이다.

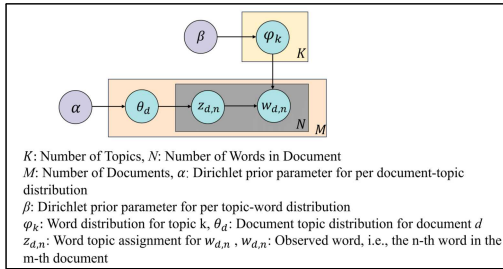


그림 1. LDA 그래픽 모델
 Fig. 1. LDA graphical model

3.3 LDA 토픽 모델링

본 논문에서는 수집된 데이터에서 TF-IDF 기법을 적용하여 문서의 특성을 벡터화하고, 이 벡터화된 데이터를 바탕으로 LDA 토픽 모델링을 이용해 토픽을 도출하였다.

그림 1은 LDA의 문서 생성 과정을 나타내었다[18]. 그림 1의 각 노드는 랜덤 변수를 나타내며, α 와 β 는 디리클레 분포(Dirichlet distribution)의 하이퍼파라미터(Hyper Parameter)인 ϕ_k 와 θ_d 의 하이퍼파라미터이다. α 는 문서-토픽 밀도를 나타낸다. α 값이 높을수록 문서가 더 많은 토픽으로 구성되며, 낮을수록 문서에 포함된 토픽이 적어진다. β 는 토픽-단어 밀도를 나타낸다. β 값이 높을수록 토픽이 말뭉치의 대부분의 단어로 구성되며, 낮을수록 토픽에 포함된 단어의 수가 적어진다. ϕ_k 는 k 번째 토픽에 속하는 단어의 분포를 나타내며, θ_d 는 d 번째 문서에 속하는 토픽의 분포를 나타낸다. 각 단어가 속하는 토픽 z 는 θ_d 에 의해 결정된다. w 는 θ_d 와 $z_{d,n}$ 을 기반으로 생성된 단어이다. 이 생성 과정은 M 개의 문서에 포함된 N 개의 단어에 대해 반복적으로 수행된다. LDA 알고리즘에서는 α , β , K 가 사용자가 결정해야 하며, ϕ_k 와 θ_d 는 LDA에 의해 학습된다.

본 연구에서는 토픽 모델링을 위해 Python의 gensim 패키지를 사용하였다. α 와 β 는 각각 0.1과 0.01로 설정되었다[19]. 최적의 토픽 개수 K 를 결정하기 위해 토픽의 개수를 1~30개까지 변화시키며 coherence 값을 측정하였다. Coherence 값은 토픽 모델링 결과를 평가하는 방법 중 하나로서 일반적으로

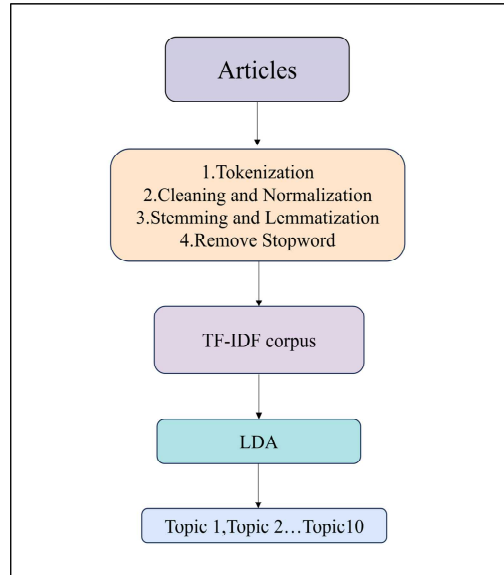


그림 2. 토픽 모델 구축 과정
 Fig. 2. Topic model building process

coherence 값이 높을수록 토픽의 일관성이 높다 [20].

Cv Coherence은 gensim 패키지에서 기본적으로 사용하는 일관성 측정 방법으로, 단어의 공기 정보와 맥락 정보를 결합하여 토픽의 일관성을 평가한다. 아래는 Cv Coherence을 사용하여 최적의 토픽 수를 결정하는 과정과 관련된 공식 및 설명이다 [20].

1. 일관성 측정 계산

주어진 토픽 수 K 에 대해, LDA 모델은 K 개의 토픽을 생성한다. 각 토픽 t 에 대해, 가장 중요한 N 개의 단어(토픽 단어)를 추출한다. Cv Coherence은 이러한 토픽 단어 간의 의미적 유사성을 계산하여 토픽의 일관성을 평가한다.

(a) 정규화된 상호 정보(NPMI):

단어 w_i 와 w_j 의 공기 확률 $P(w_i, w_j)$ 는 슬라이딩 윈도우(Sliding Window) s 에서 함께 나타나는 빈도로 계산된다. NPMI 계산 공식은 다음과 같다 :

$$NPMI(w_i, w_j) = \frac{\log \frac{P(w_i, w_j) + \epsilon}{P(w_i) \cdot P(w_j)}}{-\log(P(w_i, w_j) + \epsilon)} \quad (4)$$

여기서, $P(w_i)$ 와 $P(w_j)$ 는 각각 w_i 와 w_j 의 출현 확률을 나타내며, ϵ 은 로그의 0을 방지하기 위해 사용하는 작은 상수이다.

(b) 단어 벡터와 토픽 벡터:
NPMI를 기반으로, 각 토픽 단어 w_i 에 대해 길이 N의 단어 벡터 \vec{v}_{w_i} 를 생성하며, 벡터의 각 요소는 다음과 같다:

$$\vec{v}_{w_i}[j] = \text{NPMI}(w_i, w_j), j = 1, 2, 3, \dots, N \quad (5)$$

각 토픽 t 에 대해, 모든 단어 벡터를 합산하여 해당 토픽의 토픽 벡터 \vec{v}_t 를 구한다:

$$\vec{v}_t = \sum_{i=1}^N \vec{v}_{w_i} \quad (6)$$

(c) C_v 일관성 점수:
 C_v 일관성 점수는 모든 토픽 단어와 해당 토픽 벡터 간의 코사인 유사도의 평균값으로 계산된다. 공식은 다음과 같다:

$$C_v(t) = \frac{1}{N} \sum_i S_{\cos}(\vec{v}_{w_i}, \vec{v}_t) \quad (7)$$

모델의 모든 토픽에 대해, 최종 C_v 일관성 점수는 모든 토픽 일관성 점수의 평균이다:

$$C_v(K) = \frac{1}{K} \sum_{t=1}^K C_v(t) \quad (8)$$

C_v 일관성 측정 방법은 문맥 벡터와 코사인 유사도를 사용하여 보다 유연하고 의미에 민감한 토픽 일관성 평가 방법을 제공한다. 이는 LDA 모델에서 토픽의 품질을 평가할 때 사람의 판단과 더 일치하는 경향이 있다.

LDA 모델의 최적 토픽을 C_v 일관성을 통해 결정한 후, 각 주제에서 TF-IDF 가중치가 가장 높은 단어를

주제어로 선택했다. 그림 2는 토픽 모델 구축 과정을 나타낸다.

3.4 토픽 추세 분석

LDA 결과를 기반으로, 본 연구는 시간이 지남에 따라 활발히 연구된 토픽(즉, 핫 토픽)과 관심을 끌지 못한 토픽(즉, 콜드 토픽)를 구분하였다 [21,22].

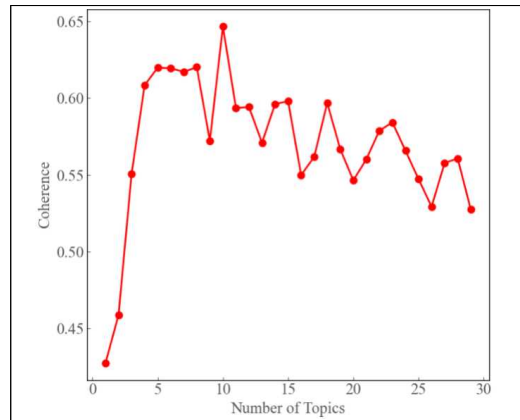


그림 3. 토픽 개수에 따른 coherence 값
Fig. 3. Coherence value by the number of topics

각 토픽의 가중치 변화를 관찰하여 그 차이를 명확히 했다. 각 토픽에 대해 선형 회귀 모델을 사용하여, 토픽 가중치(종속 변수)와 시간(독립 변수) 간의 관계를 분석했다. α 와 β 를 정의하기 위한 비용 함수를 최소화하기 위해, Ordinary Least Square 방법을 채택했다 (식 (1)). 추정된 선형 회귀 방정식(식 (2))에서 θ_{jt} 는 시간(연도) t 에서 토픽 j 의 평균 점유율을 의미한다. 이 연구에서 관심의 주요 요소는 β_j 계수이다. 이 값이 유의미하게 양수(음수)인 경우, 해당 토픽은 핫(콜드) 토픽이라고 간주된다.

$$\min \sum_{j=1}^n (\theta_{jt} - \alpha_j - \beta_j t)^2 \quad (9)$$

$$\theta_{jt} = \alpha_j + \beta_j t + \epsilon_j \quad (10)$$

4. 연구결과

4.1 토픽 분석

그림 3와 같이 토픽의 개수를 1~30개까지 변화시키며 coherence 값을 측정하였다. 그림 2에서 토픽의 개수가 10개 일 때, coherence 값이 0.647로 가장 높은 값을 나타내므로, 최적의 토픽의 개수를 10개로 결정했다.

표 3. LDA에서 얻어진 토픽

Table 3. Obtained topics from the LDA

No	Topic	Top 10 relevant words	share
1	Power Conversion and Circuits	converter, voltage, power, coil, charger, topology, inverter, output, wireless, circuit	0.0888
2	Electric Vehicle Charging Infrastructure	ev, station, network, power, distribution, load, grid, demand, energy, problem	0.2011
3	Battery State Estimation and Prediction	estimation, battery, soc, fault, state, model, accuracy, soh, prediction, method	0.0696
4	Energy and Environmental Policy	emission, policy, technology, fuel, energy, hydrogen, transport, market, impact, electricity	0.1775
5	Battery Temperature and Cooling	temperature, battery, cell, heat, pack, degree, cooling, module, lithiumion, rate	0.093
6	Battery Materials and Chemistry	material, cathode, electrode, carbon, capacity, anode, density, lithium, electrolyte, g1	0.0899
7	Optimization and Algorithms	mg, fcs, routing, bilevel, evcss, lca, mix, milp, acdc, swapping	0.0374
8	Motor and Mechanical Design	motor, machine, magnet, rotor, torque, vibration, stator, pm, flux, design	0.0542
9	Control Strategies and Systems	control, strategy, controller, fuel, motor, speed, consumption, energy, system, driving	0.1497
10	Battery Components and Materials	sei, cs, binder, flc, fourwheel, fem, handling, misalignment, solidstate, biomass	0.0389

표 3은 이 10개 토픽의 목록을 보여주며, 각 토픽에 대해 상위 10개의 단어와 전체 말뭉치에서의 토픽 비율이 포함되어 있다. 토픽은 LDA가 도출한 순서에 따라 배열되고 번호가 매겨졌다. LDA는 알고리즘에 의해 토픽을 나타내지만, 각 토픽에 대해 구체적인 이름을 제공하지는 않았다. 본 연구는 각 토픽의 상위 10개 단어를 분석하여 토픽을 명명했다.

토픽 1은 토픽에 기여한 상위 단어들이 전압, 전력, 변환기, 충전기, 무선 기술 및 회로 등과 관련된 용어를 포함하고 있고, 이 토픽은 전력 변환 장치 및 전기 회로와 관련된 내용인 것으로 생각된다.

토픽 2는 토픽에 기여한 상위 단어들이 충전소, 네트워크, 전력망, 전력 분배, 에너지 수요 등과 관련된 용어를 포함하고 있고, 이 토픽은 전기차 충전을 위한 인프라와 관련된 내용인 것으로 생각된다.

토픽 3은 토픽에 기여한 상위 단어들이 배터리의 상태 추정(SOC), 결함 감지, 예측 방법, 모델 정확성 등과 관련된 용어를 포함하고 있고, 이 토픽은 배터리 상태 추정 및 예측과 관련된 내용인 것으로 생각된다.

토픽 4는 토픽에 기여한 상위 단어들이 배출, 에너지 정책, 기술, 연료, 수소, 교통 등과 관련된 용어를 포함하고 있고, 이 토픽은 에너지 정책의 영향과 환경 기술의 시장 동향과 관련된 내용인 것으로 생각된다.

토픽 5는 토픽에 기여한 상위 단어들이 배터리의 온도, 열 관리, 냉각 모듈 등과 관련된 용어를 포함하고 있고, 이 토픽은 배터리의 열 성능 및 냉각 솔루션과 관련된 내용인 것으로 생각된다.

토픽 6는 토픽에 기여한 상위 단어들이 음극, 양극, 전극, 탄소 재료, 전해질 등과 관련된 용어를 포함하고 있고, 이 토픽은 배터리 재료 및 화학 성분과 관련된 내용인 것으로 생각된다.

토픽 7는 토픽에 기여한 상위 단어들이 라우팅 문제, 혼합 정수 선형 계획법(MILP) 등과 관련된 용어를 포함하고 있고, 이 토픽은 최적화 및 알고리즘과 관련된 내용인 것으로 생각된다.

토픽 8는 토픽에 기여한 상위 단어들이 모터, 자석, 회전자, 토크, 진동, 고정자 등과 관련된 용어를 포함하고 있어 이 토픽은 모터와 기계 시스템의 설계 및 성능 특성과 관련된 내용인 것으로 생각된다.

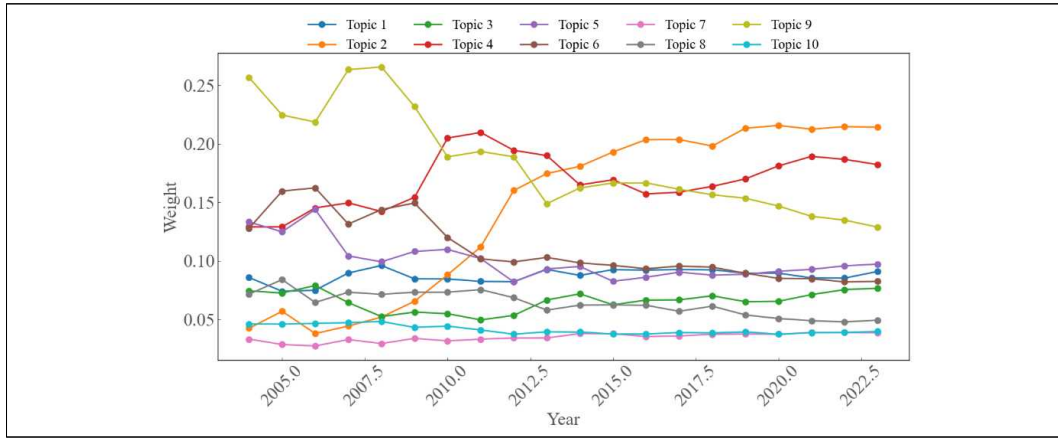


그림 4. 시간에 따른 토픽 동향
Fig. 4. Topic trends over time

토픽 9는 토픽에 기여한 상위 단어들인 제어기, 전략, 연료 관리, 모터 속도, 에너지 소비 등과 관련된 용어를 포함하고 있고, 이 토픽은 제어 전략 및 시스템과 관련된 내용인 것으로 생각된다.

토픽 10는 토픽에 기여한 상위 단어들인 고체 전해질 계면(SEI), 바인더, 사륜구동 시스템, 유한 요소법(FEM), 불일치 문제, 고체 상태 전지 및 바이오매스 등과 관련된 용어를 포함하고 있고, 이 토픽은 배터리 구성 요소와 기술적 문제와 관련된 내용인 것으로 생각된다.

또한 표 3은 각 토픽이 전체 논문에서 차지하는 비율을 나타내고 있다. 10개의 토픽 중에서 가장 큰 비율을 차지하는 토픽은 “토픽 2. Electric Vehicle Charging Infrastructure”로, 이 범주의 비율은 0.2011이다. 10개의 토픽 중에서 두 번째로 큰 비율을 차지하는 것으로 “토픽 4. Energy and Environmental Policy”와 “토픽 9. Control Strategies and Systems”이 각각 0.1775와 0.1497의 비율을 나타내고 있다.

4.2 토픽의 추세

그림 4은 4.1절에서 추출한 10개 토픽의 시간에 따른 가중치 변화를 나타내고 있다. “토픽 2. Electric Vehicle Charging Infrastructure”의 가중치는 2006년 이후로 계속 증가하고 있으며, 2014년부터 2023년까지 계속 1위를 차지하고 있다. “토픽 9. Control Strategies and Systems”의 가중치는 2010

년 이전에 모든 토픽 중 1위였지만, 2010년 이후로 매년 감소하는 추세를 보이고 있다. “토픽 4. Energy and Environmental Policy”는 초기부터 높은 가중치를 유지했으며, 2010년부터 2013년까지 모든 토픽 중 가중치 순위 1위를 차지했다. 그 이후에도 계속 높은 가중치를 유지하며 안정적인 수준을 유지하고 있다. “토픽 6. Battery Materials and Chemistry”는 초기에는 높은 가중치를 보였으나, 2009년 이후로 점차 감소하여 현재 모든 토픽 중 중간 수준을 차지하고 있다.

표 4는 10개의 토픽의 기울기(Slope, β)와 p 값(p-value), 그리고 이들의 유형을 제공하며, 이는 각 토픽의 가중치 변화 추세를 설명하는 데 사용된다. β 값에 1000을 곱해 더 명확한 수치 변화를 나타내고, p 값은 통계적 유의성을 판단하는 데 사용된다. p 값이 0.05보다 작은 경우 결과는 통계적으로 유의하다. “토픽 1. Power Conversion and Circuits”와 “토픽 3. Battery State Estimation and Prediction” p 값이 0.05보다 높다. 따라서, 통계적으로 유의하지 않아 여기서는 논의하지 않았다. 즉, 토픽 1 과 3은 전기 자동차 연구 분야의 일반적인 주제로 생각된다. “토픽 2. Electric Vehicle Charging Infrastructure”, “토픽 4. Energy and Environmental Policy”, 그리고 “토픽 7. Optimization and Algorithms” p 값이 모두 0.05보다 낮다, 통계적으로 유의하다. 그리고 β 값이 모두 양수이다. 따라서, 상승 추세가 뚜렷한 토픽

표 4. 토픽 유형 및 회귀 분석 결과

Table 4. Topic types and regression results

No	Topic	Slope β_j (1000 ^a)	p-value	Type
1	Power Conversion and Circuits	0.431	0.0528	-
2	Electric Vehicle Charging Infrastructure	11.380	7.670e ⁻¹⁰	hot
3	Battery State Estimation and Prediction	0.343	0.317	-
4	Energy and Environmental Policy	2.879	0.0175	hot
5	Battery Temperature and Cooling	-1.977	5.438e ⁻⁷	cold
6	Battery Materials and Chemistry	-3.960	2.096e ⁻⁷	cold
7	Optimization and Algorithms	0.514	5.621e ⁻⁷	hot
8	Motor and Mechanical Design	-1.548	1.199e ⁻⁷	cold
9	Control Strategies and Systems	-6.745	4.394e ⁻⁸	cold
10	Battery Components and Materials	-0.518	1.422e ⁻⁵	cold

^a p < .05, Slope is multiplied by 1000.

으로 간주될 수 있다. 반면에 “토픽 5. Battery Temperature and Cooling”, “토픽 6. Battery Materials and Chemistry”, “토픽 8. Motor and Mechanical Design”, “토픽 9. Control Strategies and Systems”, 그리고 “토픽 10. Battery Components and Materials”의 p 값이 모두 0.05보다 낮아 통계적으로 유의미하다. 또한 β_j 값이 모두 음수이므로 하락 추세가 뚜렷한 콜드 토픽으로 생각된다.

5. 결론

본 논문에서는 SCIE 데이터베이스에서 수집한 2004년부터 2023년까지의 전기 자동차 관련 36,519편의 논문을 대상으로 한 LDA 주제 모델링 결과를 소개하여 연구자와 이해관계자들이 전기 자동차 분야의 전반적인 연구 동향을 이해하는 데 도움을 주고자 한다. 전기자동차분야는 Power Conversion and Circuits, Electric

Vehicle Charging Infrastructure, Battery State Estimation and Prediction, Energy and Environmental Policy, Battery Temperature and Cooling, Battery Materials and Chemistry, Optimization and Algorithms, Motor and Mechanical Design, Control Strategies and Systems, Battery Components and Materials.의 총 10개의 주제가 확인되었다. 학술 연구자들이 관심 있는 연구 분야의 주제를 쉽게 파악할 수 있도록 본 연구는 선형 회귀 모델을 통해 10개의 주제에 대한 연구 동향 분석을 수행하였고 3개의 핫 토픽과 5개의 콜드 토픽을 확인하였다. 상승 추세를 보이는 핫 토픽은 Electric Vehicle Charging Infrastructure, Energy and Environmental Policy, Optimization and Algorithms의 3개 분야이다. 하락 추세를 보이는 콜드 토픽은 Battery Temperature and Cooling, Battery Materials and Chemistry, Motor and Mechanical Design, Control Strategies and Systems, Battery Components and Materials의 5개 분야이다.

향후에는 본 연구에서 도출된 핫 토픽과 콜드 토픽에 대한 추가 세분화 분석결과를 특허, 뉴스 보도 등과 결합하여 전기 자동차 관련 연구 주제에 영향을 미치는 요인과 그 관련성을 연구할 예정이다. 이러한 미래 연구는 연구자들이 다양한 각도에서 전기 자동차 관련 연구 주제의 동향을 이해하고 파악하는 데 도움을 주어, 연구 주제를 찾는 데 더 많은 도움을 제공할 것이다.

REFERENCES

- [1] Tracking Transport. Available online: <https://wa.org/energy-system/transport>(accessed on 3 July 2024).
- [2] A. K. Karmaker, K. Prakash, M. N. I. Siddique, M. A. Hossain, H. Pota, "Electric vehicle hosting capacity analysis: Challenges and solutions", *Renewable and Sustainable Energy Reviews*, vol.189, 113916, Jan., 2024.
- [3] T. Farinloye, O. Oluwatobi, O. Ugboma, O. F. Dickson, C. Uzundu, E. Mogaji, "Driving the electric vehicle agenda in Nigeria: The

- challenges, prospects and opportunities", *Transp. Res. Part D: Transp. Environ.*, vol. 130, 104182, May, 2024.
- [4] Directions for Improving Environmentally Friendly Vehicle Distribution Policies. Available online: https://www.kdi.re.kr/research/focusView?pub_no=18382 (accessed on 3 July 2024).
- [5] R. Peng, J. H. C. G. Tang, X. Yang, M. Meng, J. Zhang, C. Zhuge, "Investigating the factors influencing the electric vehicle market share: A comparative study of the European Union and United States", *Appl. Energy*, vol. 355, 122327, Feb., 2024.
- [6] A. L. Cheng, E. R. Fuchs, V. J. Karplus, J. J. Michalek, "Electric vehicle battery chemistry affects supply chain disruption vulnerabilities", *Nat. Commun.*, vol. 15, no. 1, 2143, Mar., 2024.
- [7] M. Koo, H. Hwang, J. Jeon, "Analysis of patent trends of electric vehicle in Korea using network analysis", *J. Korea Acad.-Ind. Coop. Soc.*, vol. 24, no. 3, pp. 13-23, Mar. 2023.
- [8] E.H. Choi, "A study on the technological competitiveness of Korea and China in the EV industry using topic modeling", *J. Northeast Asian Econ. Stud.*, vol. 35, no. 1, pp. 101-124, Apr. 2023.
- [9] M. Kim, Y. Park, J. Yoon, "Generating patent development maps for technology monitoring using semantic patent-topic analysis", *Comput. Ind. Eng.*, vol. 98, pp. 289-299, Aug. 2016.
- [10] X. Wang, M. S. Gerber, D. E. Brown, "Automatic crime prediction using events extracted from Twitter posts", *Proc. Int. Conf. Social Comput.*, vol. 7227, pp. 231-238, Apr. 2012.
- [11] Z. Huang, X. Lu, H. Duan, "Latent treatment pattern discovery for clinical processes", *J. Med. Syst.*, vol. 37, no. 1, pp. 1-10, Feb. 2013.
- [12] Y. Guo, S. J. Barnes, Q. Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation", *Tourism Manag.*, vol. 59, pp. 467-483, Apr. 2017.
- [13] H. Hao, K. Zhang, W. Wang, G. Gao, "A tale of two countries: International comparison of online doctor reviews between China and the United States", *Int. J. Med. Inform.*, vol. 99, pp. 37-44, Mar. 2017.
- [14] J. Yoon, W. Seo, B. Y. Coh, I. Song, J. M. Lee, "Identifying product opportunities using collaborative filtering-based patent analysis", *Comput. Ind. Eng.*, vol. 107, pp. 376-387, May. 2017.
- [15] M. Haghani, F. Sprei, K. Kazemzadeh, Z. Shahhoseini, and J. Aghaei, "Trends in electric vehicles research", *Transp. Res. Part D: Transp. Environ.*, vol. 123, p. 103881, Oct. 2023.
- [16] L. George P. Sumathy, "An integrated clustering and BERT framework for improved topic modeling", *Int. J. Inf. Technol.*, vol. 15, pp. 2187-2195, May, 2023.
- [17] G. Salton and C. T. Yu, "On the Construction of Effective Vocabularies for Information Retrieval", *ACM SIGPLAN Notices*, vol. 10, pp. 48-60, Nov. 1973.
- [18] D. M. Blei, A. Y. Ng, M. I. Jordan, "Latent Dirichlet allocation", *J. Mach. Learn. Res.*, vol. 3, pp. 993-1022, Jan. 2003.
- [19] Best Practices for Topic Modeling. Available online: <https://msaxton.github.io/topic-model-best-practices/>
- [20] M. Röder, A. Both, A. Hinneburg, "Exploring the space of topic coherence measures", *Proc. 8th ACM Int. Conf. Web Search Data Min.*, pp. 399-408, Feb. 2015.
- [21] T. L. Griffiths, M. Steyvers, "Finding scientific topics", *Proc. Nat. Acad. Sci.*, vol. 101, pp. 5228-5235, Apr. 2004.
- [22] J. S. Park, S. G. Hong, J. W. Kim, "A study on science technology trend and prediction using topic modeling", *J. Korea Ind. Inf. Syst. Res.*, vol. 22, no. 4, pp. 19-28, Aug. 2017.

저자약력

천 위 안 (Yuan Chen)

[정회원]



- 2020년 2월 : 강원대학교 일반대학원 자동차공학과(공학석사).
- 2020년 3월~현재 : 강원대학교 일반대학원 기계-컴퓨터-산업경영 공학과(박사과정).

〈관심분야〉 재료강도학, 기술예측, 머신러닝, 딥러닝

조 석 수 (Seok-Swoo Cho)

[정회원]



- 1997년 2월 : 동아대학교 대학원 기계공학과 (공학박사)
- 1997년 8월~현재 : 강원대학교 기계시스템공학부

〈관심분야〉 재료강도학, 기술예측, 머신러닝, 딥러닝