

토픽모형을 이용한 중국 전기차 산업의 기술 분석*

최 의 현**

· 목 차 ·

- | | |
|-------------------|-----------|
| I. 서론 | IV. 분석 결과 |
| II. 이론적 배경 | V. 결론 |
| III. 전기차 특허 추이 분석 | |

주제어 : 전기차, 특허 분석, 빅데이터, 토픽모형, 텍스트마이닝

I. 서론

본 연구는 중국 전기차 산업의 기술 수준을 텍스트마이닝 기법인 토픽모형을 통해 분석한다. 구체적인 방법론은 잠재 디리클레 할당(LDA, Latent Dirichlet Allocation) 모형이며, 분석대상은 중국이 출원한 국제특허(PCT) 최근 20년간 데이터이다.

중국 전기차 산업은 정부의 적극적인 정책에 힘입어 2015년 이후 세계 1위의 판매량을 기록하고 있으며, 다수의 기업이 시장 점유율을 놓고 치열하게 경쟁한다. 또한 중국의 전기차 배터리 산업은 CATL을 중심으로 세계적인 경쟁력을 갖고 있으며, 전기차 업체와 유기적인 협력관계를 유지하고 있다(이상운 2021, 조정원 2019).

국가, 산업, 또는 기업의 기술력을 분석하는 데 있어서 특허 통계는 매우 유용한 지표로 역할을 한다(Aghion et al. 2012). 오새새 외(2017)는 특허 통계를 통해 중국 전기차 산업의 기술력을 분석하였는데 그 결과 특허의 양은 증

* 본 연구는 2022년도 영남대학교 학술연구조성비 지원에 의하여 연구되었음.

** 영남대학교 경제금융학부 교수(yxcui@yumail.ac.kr)

가하고 있으나, 질적인 특허는 한국이나 일본과 비교하여 상당히 부족한 것으로 나타났다. 최의현(2020)은 특허 통계를 통해 전기차 특허를 다섯 개의 기술분야로 구분하고 중국기업의 혁신성과 정부 지원에 대해 분석하였다. 국제적으로 혁신성을 인정받는 수준 높은 특허는 민간기업을 중심으로 출원되었고, 정부와 관련된 기업, 연구소의 특허는 양적인 측면에서 증가하였음을 밝혔다. 정부의 적극적인 전기차 산업 정책이 기술혁신 측면에서 효율성이 떨어질 수 있다는 뜻이다.

이같이 중국 전기차 산업의 기술력을 분석하는 방법으로 정부정책과 기업실적에 관한 정성적 연구가 있고, 기술력의 지표로서 특허 통계를 활용하는 연구도 있다. 본 연구 역시 특허 통계를 대상으로 한다는 점에서는 기존 연구와 같지만, 특허를 단순하게 통계 수치로 분석하는 것이 아니라 특허의 문헌적 내용을 분석하였다는 점에서 큰 차이가 있다. 지난 20년간 중국이 출원한 전기차 특허 전체를 대상으로 모든 문서의 초록(abstracts)을 빅데이터 분석기법인 토픽모형으로 분석하여 특허의 기술을 주제(토픽)별로 구분하였다. 이런 방법을 사용하면 중국이 어느 기술 분야에서 상대적으로 의미 있는 발전을 했는지 그리고 현재 활발하게 진행되는 기술의 특징과 과거의 그것을 구분해 볼 수 있다.

본 연구의 구성은 다음과 같다. II장은 이론적 배경으로 특허 통계의 특성과 토픽 모형을 설명한다. III장은 중국 전기차 산업의 특허를 한국, 일본, 미국 등의 특허와 비교, 분석한다. IV장은 LDA 모델을 통해 도출한 토픽에 관해 설명하고, 전체 토픽을 핫(hot) 토픽과 콜드(cold) 토픽으로 구분하여 그 의미를 살펴본다. 마지막으로 V장에서 결론을 맺는다.

II. 이론적 배경

1. 분석대상

중국 전기차 산업의 기술 수준과 그 추이를 분석하기 위해 WIPO에 등재

된 전기차 특허 데이터를 사용하였다. 국제특허 분류 기준인 IPC 코드의 'B60L'은 전기차에 대한 특허를 집합적으로 보여준다. 이 기준에는 전기로 추진력을 얻는 차량(철도 차량 포함)의 설치와 배치에 관한 특허로 정의되고 있는데 자기부상열차나 고속철과 같은 궤도형 차량에 대한 특허도 포함하지만, 전반적인 전기차 산업의 기술력을 포괄적으로 나타낸다.¹⁾

특허는 어느 국가(또는 기관)에 등록하였는지에 따라 그 가치를 인정받는다. 미국, 유럽, 일본 등 선진국 특허기관에 모두 등록된 특허(소위 삼국특허)라면 그 기술이 세계적 수준이라고 할 수 있고, 미국이나 유럽 중 한 곳에 등록된 특허라고 해도 그 가치는 상당하다. 반면 중국특허청(국가지식산업국)에만 등록된 특허라면 기술적으로 중요한 의미를 부여하기 어렵다. 최근 중국에서는 정부주도 내지는 윗선의 재촉으로 정부측 연구기관이나 기업이 특허 출원을 남발하는 경향이 강하다. 또한 특허는 신청(출원)이후 등록까지 1년 반에서 2년 정도 걸리기 때문에 심사결과가 나오지 않은 특허에 기술적 의미를 부여하기에는 상당한 주의가 필요하다.

본 연구는 WIPO(세계지식재산권기구)의 국제특허(PCT)를 사용하였다. PCT 특허는 기술에 대한 기술적 검토가 없고, 법적 효력을 갖추지 않았기에 출원인이 특허 출원을 남발할 수는 있지만, 질적인 측면에서 개별 선진국 특허와 중국내 특허 사이에 있다고 볼 수 있다. 이런 이유로 PCT 통계를 사용할 때는 그 기술의 질적 수준을 분별하기보다 기술의 추세적 변화에 초점을 맞출 필요가 있다.

본 연구의 분석 연도는 2001년부터 2020년까지 20년 기간이다. 중국의 첫 전기차 PCT 특허가 1999년 6월이었고, 2000년에도 2건의 출원이 있었지만, 20년이라는 기간을 맞추기 위해 2001년 특허부터 조사하였다. 또한 통상적으로 데이터 추출 시점 대비 1년 이내에서는 특허 통계가 자주 바뀌는 경우가 많아 2020년 말까지로 기간을 정했다. 분석 기간에 중국인이 출원한 전기차 PCT 특허는 총 566건이고, 이 특허 문서에서 초록(abstract)을 추출했다.

1) 전기차 기술의 핵심중 하나인 배터리 제조에 대한 특허는 B60L에 포함되지 않는다. 배터리 코드 H01M으로 차량용, 휴대폰용 등 모든 종류의 배터리 관련 특허가 여기에 있다.

따라서 본 연구가 분석할 데이터는 특히 제목, 특히 초록, 출원인, 출원 연도 등으로 구성되어 있다.

2. 방법론

특히 초록이라는 텍스트로 구성된 비정형(unstructured) 데이터를 분석하기 위해 본 연구는 R 프로그램을 이용하여 단어 전처리 과정을 진행하고,²⁾ 그 후 문서와 단어 간의 빈도수를 값으로 갖는 행렬인 document-term matrix(DTM)를 구축했다. 최종적으로 566개 문서에서 3,505개 단어로 구성된 DTM으로 토픽모형 분석을 시도하였다.

토픽모형은 문서와 단어로 구성된 행렬을 기반으로 문서에 잠재되어 있다고 가정된 주제(토픽)의 등장확률을 추정하는 자연어 처리기법이다(백영민 2020:243). 본 연구는 토픽모형 가운데 광범위하게 활용되는 잠재적 디리클레 할당(LDA) 모형을 사용한다. LDA 모형에 따르면 본 연구의 데이터는 566개의 문서(document)가 행(column)으로, 3,505개의 단어가 열(row)로 배치된 말뭉치(corpus)이다. LDA 알고리즘은 문서내 잠재된 주제들을 찾아내는 생성모델(generative model)이다.³⁾ 문서내 단어들을 대상으로 문서 구조와 보이지 않는 변수를 추론하고, 문서 집합내 단어들에 대한 디리클레 분포를 통하여 지정한 수만큼의 토픽을 찾아낸다(박주섭, 홍순구, 김종원 2017, 임선규 2021).

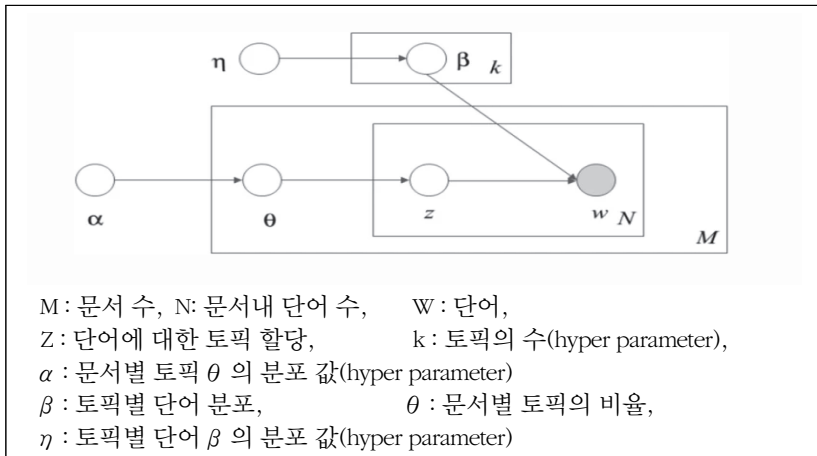
LDA 모형을 시각화하면 [그림 1]과 같다.⁴⁾ 여기에서 관찰된 값은 단어(W_N)이고, 이것으로부터 Θ , Z , β 를 추론하게 된다. β 는 hidden parameter, Θ 와

2) 대문자를 소문자로 전환(lowcase), 특수문자, 조사, 관사 등 불용어(stopword) 처리, 단어의 어미변화를 조정하는 표준단어 변환(stemming) 등의 작업을 수행하였다.

3) 머신러닝에서 데이터 분류를 위한 모델은 생성모델과 판별모델로 나뉜다. LDA는 생성모델 중 데이터의 확률 분포를 모르는 상태에서 샘플링으로 추정하는 암묵적 모델에 속한다.

4) LDA 모형의 구조를 시각화한 설명은 Blei, Ng, Jordan(2003)이 제시한 [그림 1]이 모범적이다. 모형에 대한 자세한 설명은 텍스트마이닝 이론서인 백영민(2020)을 참고하기 바란다.

Z 는 hidden variable이다. α 와 η 는 설정된 값인 hyper parameter다.⁵⁾ Z 는 문서별 토픽 비율인 θ 로부터 생성되고, θ 는 디리클레 분포를 따르는 값으로 α 값에 의해, 또한 토픽별 단어 분포인 β 는 η 값에 의해 결정되며 디리클레 분포를 따른다. 각 단어의 토픽을 나타내는 값인 Z 와 토픽별 단어 비율인 β 값에 따라 단어(w_N)가 결정된다.⁶⁾



자료: Blei, Ng, Jordan(2003), Park and Liu(2020)

<그림 1> LDA 모형의 구성

3. 기존 연구

LDA 모형은 Blei, Ng, Jordan(2003)에 의해 처음으로 등장하였다. 이들 개발자는 LDA 모형을 생성적 확률모형, 3층 위계적 베이지안 모형 등의 용어로

5) 하이퍼 파라미터는 머신러닝에서 사용자가 직접 정한 값을 말한다. 임의로 정한다고는 하지만, 토픽모델링에서는 경험적 또는 방법론적으로 통용되는 법칙이 있으며 이는 차후의 결과분석에서 설명한다.

6) 백영민(2020, 244-246), 김태경 외(2016), Park, and Liu(2020) 등의 설명을 참고하였다.

정의한다. 사회과학 분야에서 LDA 모델은 텍스트 분류나 분석에 자주 활용되는데, 본 연구와 같이 특허 내용을 분석하는 연구는 사회과학 분야보다 기술 추이를 분석하는 산업공학분야에서 자주 활용되었다.

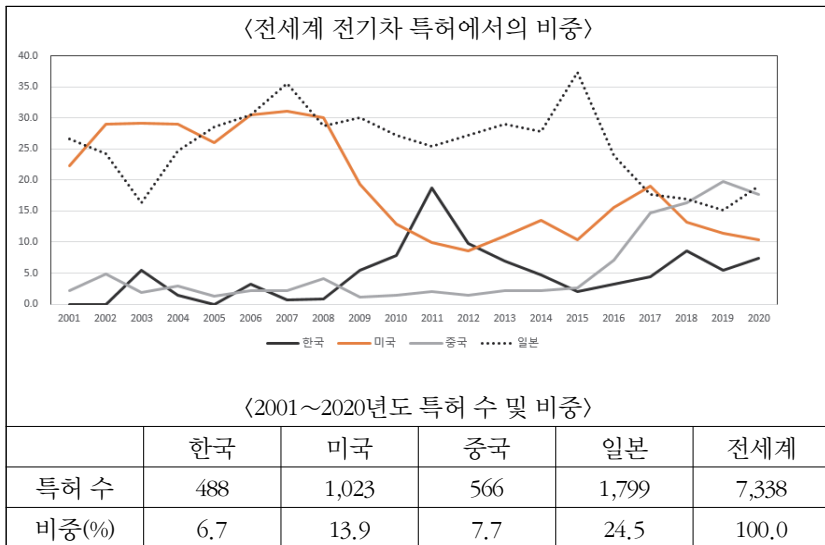
박주섭, 홍순구, 김중원(2017)은 미국특허청의 인공지능특허 초록을 대상으로 LDA 분석을 시도하였다. 20개의 토픽을 정하고, 핵심기술의 연도별 추이와 비중 분석을 통해 핫토픽과 콜드토픽을 구분하였다. 김태경, 최화련, 이흥철(2016)은 핀테크 특허를 대상으로 유사한 분석을 하였다. Wang et al.(2021)은 중국내 블록체인 특허를 LDA 기법으로 분석하였다. 분석대상은 다르지만, 그 방법론에서는 큰 차이가 없었다. 그 외 유사한 연구로 특허 대신 논문 초록을 이용하여 기술의 방향을 추적하는 연구가 있는데 대상만 바뀌었을 뿐 방법론은 유사하다.

기존 연구들은 산업공학(또는 경영공학) 측면에서 방법론의 적용 및 확대, 그리고 자신들의 연구분야에 어울리는 기술적인(또는 엔지니어링적인) 차원에서 결과를 해석한다. 이에 비해 본 연구는 중국 전기차 산업에 대한 경제학적 결과를 도출하고, 기존 연구에서의 결과를 보완 내지는 지지하기 위해 LDA 모델을 방법론으로 활용한다. 기존 전기차 연구에서는 첫째, 중국 전기차 산업에 대한 정책이나 기업에 대한 정성적 문헌 분석, 둘째, 생산량, 무역액 등 2차 통계자료에 근거한 분석, 셋째, 특허 통계와 같은 1차 통계자료에 대한 분석 등의 방법론을 사용하였다. 기존의 연구방법론에 빅데이터 분석 기법인 텍스트마이닝 방법론을 추가하면 중국 전기차 산업의 기술력에 대해 좀 더 세밀한 분석이 가능하다고 본다.

Ⅲ. 전기차 특허 추이 분석

이 장에서는 중국의 전기차 기술을 특허 비교를 통해 분석한다. 먼저 [그림 2]는 중국의 PCT 특허를 한국, 미국, 일본과 비교한다. PCT 특허는 신청의 개념이어서 기술의 질적 수준을 보장하지는 않는다. 특허 수량에서 일본이 가장 많고, 미국, 중국, 한국의 순이었다. 일본은 2015~16년도에 가장 많은

특허를 출원했고, 미국은 2010년 이후 특허 출원이 급감했다. 한국은 2011년 이 피크였고, 중국은 2019~2020년에 특허 출원이 많았다. 이러한 추세에 대해 중국의 기술추적이 빠르게 진행 중이라고 해석할 수도 있지만, 대부분의 핵심기술이 2010년을 전후로 이미 개발되어서 그 뒤로는 후속 기술개발이 더디었다고도 볼 수 있다.



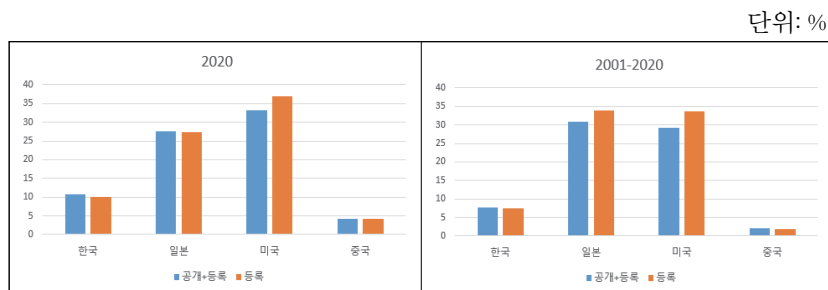
자료: wipson DB(wipson.com)

<그림 2> 주요국의 전기차 특허 추이

다음으로 미국특허청(USPTO)에 출원한 특허를 대상으로 4개국의 전기차 기술을 양적으로 비교해 보았다. 미국특허청은 유럽특허청과 함께 가장 권위 있는 특허청으로 해당 기술의 질적 수준을 담보한다. [그림 3]은 특허 공개와 출원, 두 가지 통계를 보여주는데 특허 공개는 특허 신청과 같은 개념이고, 특허 등록은 신청과 심사를 거쳐 법적 효력을 인정받은 특허이다. 등록된 특허가 많을수록 해당국의 기술 수준이 높다고 하겠다. 2001년부터 20년간의 통계를 보면 일본과 미국의 특허 수가 한국이나 중국을 압도하였다.

중국과 중국학 (제46호)

PCT 특허에서 양적 성장을 보여주었던 중국은 미국특허에서는 여전히 낮은 비중을 기록하였다. 중국의 전기차 기술이 새로운 발명으로 인정받기에는 부족하다는 의미이다.



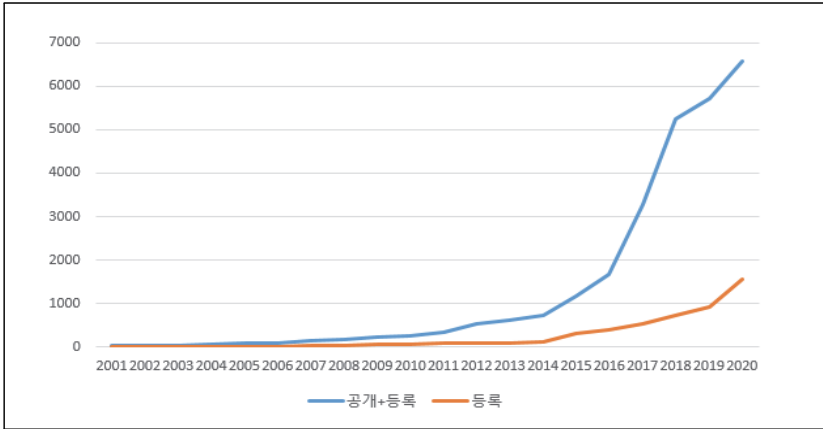
자료: wipson DB(wipson.com)

<그림 3> 주요국의 미국특허 출원 비중 비교

중국지식산업국(CNIPA) 특허 가운데 중국 국적자가 출원한 전기차 특허는 [그림 4]에서 확인할 수 있다. 2001년 이후 20년간 중국지식산업국의 전기차 특허는 약 3만 6천 건이고, 이중 중국인이 출원한 특허는 2만 7천 건 정도다. 2015년 이후 특허 출원이 폭발적으로 증가하였는데 이는 PCT 특허에서의 추이와 유사하다. 2015년은 ‘중국제조 2025’가 발표된 해였다.⁷⁾ 중앙정부의 강력한 기술개발 의지가 특허 출원에 반영된 것이라고 추론할 수 있다.⁸⁾

7) 중국의 전기차 정책은 조정원(2019)에 정리되었다.

8) 지난 20년간 한국특허청에 출원한 한국 국적의 전기차 특허는 약 9,500건이고, 한국의 PCT 특허는 488건이었다. 중국의 경우 두 통계는 각각 2만 7천 건과 566 건이다. 중국내 특허 신청이 남발되었을 가능성이 큼을 방증한다.



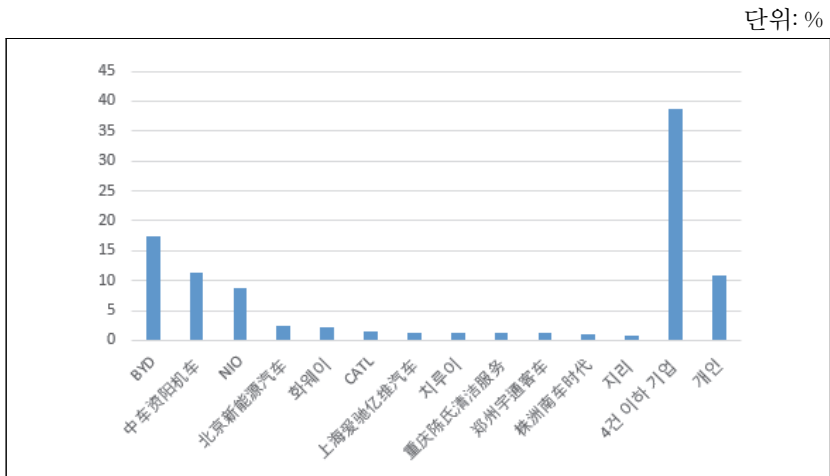
자료: wipson DB(wipson.com)

<그림 4> 중국내 중국인의 특허 출원 추이

마지막으로 어느 중국기업이 전기차 출원을 많이 했는지 PCT 특허를 기준으로 살펴보면 [그림 5]와 같다. 중국 전기차 산업에서 선구자 역할을 한 BYD가 가장 많은 PCT 특허(20년간 98건)를 보유했다. 다음으로 기관차 생산 회사인 中國中車와 최근 판매량이 급증하는 NIO의 특허가 많았다.⁹⁾ CATL은 배터리 전문회사로 배터리 특허의 출원은 많았지만, 전기차 관련 특허는 그렇지 않았다. 특이한 점은 20년간 4건 이하, 주로 1건의 특허만 출원한 기업들의 특허 비중과 개인의 비중이 매우 높다는 것이다. 이런 특허는 기술개발의 연속성이 떨어지고 보여주기식 특허일 가능성이 크다.¹⁰⁾

9) NIO에는 NextEV의 특허도 포함하였다.

10) PCT 특허가 심사 없이 등록만으로 가능하다는 점을 인식할 필요가 있다.



주: 4건 이하 기업은 20년간 특허 출원이 4건이 안 되는 기업으로 대부분 1건임. 개인은 사업체가 아닌 개인 명의로 출원한 것으로 5명만이 2건을 출원했고, 나머지는 모두 1건임.

자료: wipson DB(wipson.com)

<그림 5> 중국 PCT 특허의 출원인별 전기차 특허 비교

IV. 분석 결과

1. 토픽 모델링

전기차의 세부 기술을 추출하기 위해 전처리 과정을 거친 특허 초록 정보에 LDA 토픽모델링 프로그램을 사용하여 토픽을 추출하였다. 토픽 모델에서는 토픽 수의 사전적 설정이 중요하다. 잠재토픽의 개수를 선정하기 위해 흔히 R 프로그램의 topicmodels 패키지의 perplexity 함수를 사용한다. 하지만 본 연구는 백영민(2020: 409~416)에 따라 Griffiths2004, Deveaud2014, CaoJuan2009, Arun2010 등 4개의 모형적합도 지수를 동시에 고려하여 본 모형에 적합한 토픽 개수를 정하였다.¹¹⁾ 파라미터 추정을 위해 샘플링 반복(Gibbs

sampling)을 시도하였다. 다수의 연구에서 반복 횟수(iteration)를 1천 회 이상을 권고한다. 본 연구에서는 반복 횟수를 250회부터 5,000회까지 시도해보았는데 토픽별 단어빈도의 차이는 거의 없었다. R의 topicmodels 패키지에서는 $K=16$ 이면, 파라미터 α 는 $3(=50/K)$ 에 근사한다.¹²⁾ α 값이 LDA 모델에 미치는 효과를 검증한 결과, $\alpha=3$ 에서 거의 모든 문서가 하나의 토픽 값을 가졌다. 즉, 현재의 토픽 수에서 특정 문서에서 특정 토픽이 주도적으로 드러나는 결과를 얻었다.

〈표 1〉은 전기차 기술에서의 토픽과 토픽별로 빈도가 높은 단어를 보여준다. 토픽 명칭은 주어진 단어와 그 단어가 주로 사용되는 특허 분류를 종합적으로 고려하여 작성하였다. 전기차 특허인 IPC 중분류 코드 B60L에는 12개의 세분류 코드, 그리고 그 밑에 63개의 세세분류 코드가 있다.

그런데 전기차 세부 기술을 복잡한 모형을 이용하여 16개의 토픽으로 정할 필요 없이 비슷한 숫자인 12개의 IPC 코드를 이용하여 기술을 구분하는 방법도 생각해 볼 수 있다. 이러한 방법 역시 기존 연구에서 활용되었던 방법이지만 다음과 같은 문제가 존재한다. 첫째, 특허를 세분화할수록 경제적 의미가 있는 설명보다는 공학적 메커니즘에 의존하게 된다. 둘째, 하나의 특허가 복수의 특허 코드를 갖는 경우 특허 코드만으로 기술을 구분하기 어렵다. 실제로 다수의 특허가 B60L 이하 분류에서는 복수의 IPC 코드를 갖고 있었다.¹³⁾

전기차의 핵심부품은 구동모터 시스템, 차체 및 샤시 플랫폼, 배터리 관련 시스템, 공조 및 전장 공통부품 등 4개 영역으로 구성된다(이봉현 외 2011). 구동모터 시스템 영역에 속하는 토픽으로는 토픽 5, 6, 9, 11, 12가 해당된다. 차체와 관련된 토픽은 토픽 2, 기타 1 및 2이다. 배터리 관련 시스템은 토픽 1, 8, 10 그리고 공조 및 전장 공통부품의 영역에 속하는 토픽은 T3, 13, 15 등

11) 네 개의 모형적합도 지수를 비교한 결과 토픽의 개수가 16개인 경우와 37~40개인 경우에서 신뢰성이 있었다. 하지만 토픽의 수가 너무 늘어나면 기술의 특성을 집합적으로 분석하는 본 연구의 의도에서 벗어날 수 있기에 상대적으로 적은 수의 토픽으로 정하였다.

12) Griffiths and Steyvers(2004) 및 백영민(2020:248)을 참고

13) 하나의 특허가 여러 세부적인 특허 분야에 속해 있다는 뜻이다.

으로 구분할 수 있다.¹⁴⁾

<표 1> 중국 전기차 특허에서의 토픽과 토픽별 키워드

토픽1. 배터리 충전제어	토픽2. 차체, 프레임	토픽3. 공조장치	토픽4. 기타 1
control time controlling method mode charger vehiclemounted	body frame air structure front magnetic mounted	module comprises provided device terminal comprising connector	connected train comprises connection plurality means bus
토픽5. 전동식 브레이크 시스템	토픽6. 전기차의 출력 제어	토픽7. 철로 집전장치	토픽8. 배터리 관리
braking brake method equalization target feedback comprises	vehicle method state torque current speed condition	end rail current mechanism member contact device	battery pack charge management heating balancing packs
토픽9. 전기 모터 제어	토픽10. 전기에너지 저장 장치	토픽11. 전력 제어	토픽12. 차량컨트롤 유닛(VCU)
motor controller driving drive generator output range	energy storage electric device automobile fuel solar	power supply dc/dc converter means line distribution	control unit signal detection management safety switch

14) 토픽 14번은 충전소와 관련된 기술이고, 토픽 4와 16번은 철도와 전장부품이 혼재되어 있어 기타로 분류하였다.

토픽13. 차량용 발전기	토픽14. 충전기, 충전소	토픽15. 고전압 부품	토픽16. 기타 2
vehicle electric electrical electricity generation disclosed mileage	charging device information station position pile charged	circuit voltage current connected traction end direct	system high present operation invention low vehicles

지난 20년간 중국 전기차 특허에서 토픽 등장확률을 계산하면 <표 2>와 같다. 16개 토픽의 등장확률이 큰 차이가 없는 것은 어느 한 기술 분야에 특허가 집중되지는 않았다는 뜻이다. 이 가운데 토픽 7(철로 집전장치), 14(충전기·충전소), 6(전기차의 출력 제어)에 상대적으로 기술이 집중되었고, 토픽 5(전동식 브레이크 시스템), 4(기타 1), 1(배터리 충전제어)은 상대적으로 약하였다.

<표 2> 중국 전기차 특허에서의 토픽 등장확률

토픽1	토픽2	토픽3	토픽4	토픽5	토픽6	토픽7	토픽8
5.8%	6.0%	5.9%	5.6%	5.4%	7.0%	7.1%	6.5%
14위	10위	12위	15위	16위	3위	1위	5위
토픽9	토픽10	토픽11	토픽12	토픽13	토픽14	토픽15	토픽16
6.3%	6.0%	6.5%	5.9%	6.3%	7.0%	6.8%	5.8%
7위	9위	6위	11위	8위	2위	4위	13위

주: 전체 토픽의 등장확률 총합은 100%임.

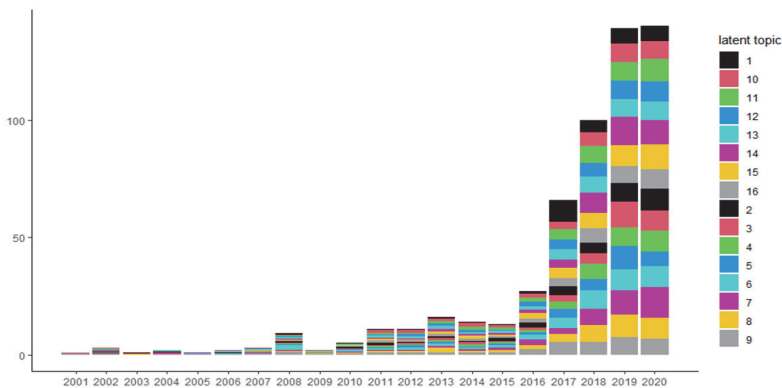
다음으로 연도별 토픽의 등장 추이를 분석하였다. [그림 6]을 보면 2015년까지는 특허의 수가 정체되어 있었으나 2016~2019년 기간에 급증하였음을 알 수 있었다. 2020년에는 토픽 7, 14, 15, 11에 해당하는 특허가 늘어났다. 어느 한 기술 분야에서 특허가 집중적으로 늘어나지는 않았고, 대체로 비슷한 규모로 증가하는 현상을 보였다. 중국의 전기차 기술이 상향 평준화된다고

도 해석할 수 있지만, 특정 분야에서의 기술력이 ‘치고 올라가는 모습’은 아니었다. 중국의 전기차 기술이 비교적 최근에서 발전한다고 할 수 있는데 선두 주자 역할을 하는 기술이 보이지 않는 것은 부정적인 측면에서도 해석할 수 있다.

2. 핫 토픽 및 콜드 토픽 분석

다음으로 중국의 전기차 기술에 관한 연구가 점점 확대되고 있는지 감소하고 있는지를 핫 토픽과 콜드 토픽으로 구분하여 분석한다. 연도별로 토픽이 상대적으로 증가 추세에 있다면 중국내 해당 기술에 관한 연구가 활발하게 진행중이고, 그 반대라면 연구가 줄어들고 있다고 해석할 수 있다. 그런데 콜드 토픽이라고 해서 부정적으로만 볼 수 없는 것이 과거에 이미 연구가 많이 행해진 분야에서는 신규 특허가 적게 등장할 수 있기 때문이다.

각 토픽의 연도별 추이를 상승 또는 하락으로 판별하는 기준으로 선형회귀분석의 회귀계수 값을 이용하고, 잔차 분석을 통해 회귀모형의 적합도를 검토하였다. 독립변수는 연도, 종속변수는 전기차 기술(토픽)의 연도별 비중 평균값을 사용하여 선형회귀 분석을 하였다. 유의수준 0.05에서 유의한 확률을 가지는 토픽을 대상으로 회귀계수 값이 양이면 핫 토픽, 음이면 콜드



<그림 6> 연도별 전기차 특허의 토픽별 분포

토픽으로 구분하였다.

먼저 2001년부터 20년간의 토픽들의 연도별 추세를 분석한 결과 핫 토픽으로 철로집전장치, 전력제어, 충전기·충전소, 고전압 부품 등 4개 토픽이, 콜드 토픽으로 차체·프레임이 꼽혔다. 나머지 11개의 토픽은 각각 증가 또는 감소하였지만, 신뢰할만한 유의수준을 갖지는 못했다. 핫 토픽 가운데 철로집전장치는 전기차와 무관한 기술이고, 충전기·충전소도 전기차 개발의 핵심과는 거리가 먼 분야다. 차체·프레임은 전기차 전용 차체와 관련된 분야로 주행 기본성능을 결정하는 중요한 부분인데 콜드 토픽으로 나타났다.

다음으로 분석 기간을 2011년 이후 10년으로 줄여서 분석하였다. 대부분의 전기차 기술이 최근 연도에 개발되었으므로 최근의 변화를 좀 더 파악하기 위해서이다. 핫 토픽으로는 철로 집전장치와 충전기·충전소 등 두 개가 선정되었다. 콜드 토픽에서는 앞선 분석과 상당한 변화가 나타났다. 전기 모터제어, 전기에너지 저장장치, 차량용 발전기 등이다. 최근 중국의 전기차 판매량이 늘어나면서 충전시설 부족에 대한 이슈가 크게 제기되고 있어 충전기·충전소 기술은 충분히 핫 토픽이 될 수 있다.¹⁵⁾ 하지만, 전기차 개발과 관련된 핵심적인 기술은 아니며, 오히려 콜드 토픽의 기술들이 전기차와 직접 관련된 기술이다.

중국 전기차에 대한 핫 토픽과 콜드 토픽 분석은 산업정책 차원에서 매우 흥미로운 결과를 보여준다. 특히로 살펴본 중국의 전기차 기술은 폭발적으로 증가했다. 이 기술을 각 주제별로 분석해 보니 실제로 기술 발전이 두드러진 분야는 철도와 충전소 등 주변부적인 기술이고, 전기차의 성능과 관련된 핵심기술은 추세적으로 약화되었다. 앞에서 언급하였듯이 콜드 토픽은 이미 성숙된 기술에 속할 수 있지만, 중국 전기차 산업이 과거에 이미 개발해 놓은 기술은 없다고 봐야 한다.

15) <https://www.mk.co.kr/news/world/view/2021/10/944605/>

<표 3> 중국 전기차의 핫 토픽과 콜드 토픽 구분

기간: 2001~2020년			
	토픽명	계수	P-value
핫 토픽	7. 철로집전장치	0.002	0.036
	11. 전력제어	0.001	0.096
	14. 충전기 · 충전소	0.002	0.003
	15. 고전압부품	0.002	0.008
콜드 토픽	2. 차체, 프레임	-0.004	0.059
중립	11개		
기간: 2011~2020년			
	토픽명	계수	P-value
핫 토픽	7. 철로집전장치	0.006	0.022
	14. 충전기 · 충전소	0.005	0.005
콜드 토픽	9. 전기 모터제어	-0.004	0.011
	10. 전기에너지 저장장치	-0.007	0.004
	13. 차량용 발전기	-0.004	0.006
	16. 기타2	-0.002	0.076
중립	10개		

V. 결론

본 연구는 중국 전기차 산업의 기술력을 분석하기 위해 전기차 특허를 수집하고, 방법론적으로 토픽 모형을 적용하였다. 특허 통계는 2001년 이후 20년간 중국인이 출원한 국제특허(PCT) 566건 전수를 대상으로 하였다. 이 데이터에 토픽모델링을 적용하여 전기차에 대한 토픽을 추출하고 세부 기술을 정의하였다. 또한 추출한 기술에 대해 회귀분석을 통해 유의미한 기술에 대해 상승추세를 보이는 핫 토픽과 하락추세를 보이는 콜드 토픽을 구분하여 분석하였다.

그 결과 16개의 토픽을 추출하였고, 이들 토픽의 기술적 추세를 분석할 수 있었다. 대다수 토픽에서 기술력이 향상되었지만, 특정 분야에서의 기술력 향상이 추세적으로 두드러지게 나타나지는 않았다. 좀 더 세밀한 분석을 위

해 핫 토픽과 콜드 토픽을 구분한 결과 실제로 기술 발전이 두드러진 분야는 철도와 충전소 등 주변부적인 기술이었고, 전기차의 성능과 관련된 핵심기술인 전기 모터제어, 전기에너지 저장장치, 차량용 발전기 등은 추세적으로 약화되었다.

적극적인 과학기술 개발정책은 양적인 면에서 큰 성과를 거뒀음에도 불구하고, 핵심적인 부분에서의 성과는 상대적으로 퇴보한 것 같다는 본 연구 결과는 중국정부의 정책 효율성에 대해 강한 의문을 제기한다. 물론 최근 중국 전기차산업에 있어서 충전소 문제가 소비자 선택에 핵심적인 이슈인 것은 사실이고, 전기차 생산 기술만큼 중요한 것이 배터리 생산기술인 점도 타당하다. 중국 전기차가 세계적으로 성능 좋은 배터리를 탑재하였다면 그 자체만으로 경쟁력을 가질 수도 있다. 그럼에도 불구하고 늘어나는 양적 성과 중에 하필이면 중요한 기술에서만 성과가 상대적으로 줄어드는지에 대한 해석이 필요하다. 전기차는 상용화되었지만, 아직 기술적으로 보완해야 할 부분이 상당히 많다는 것을 시사한다.

참 고 문 헌

- 백영민. 2020. 『R를 이용한 텍스트 마이닝』. 한울.
- 오새새, 최의현, 최병헌(2017). 「중국 전기자동차 산업의 발전과 기술경쟁력 비교 분석: 특허 통계를 중심으로」, 『중소연구』, 41-3.
- 이상윤. 2021. 「중국 전기차 배터리산업 정책과 발전에 관한 연구」. 『비교경제연구』, 28-1.
- 이봉현, 김찬중, 김규식. 2011. 「전기차 기술개발 동향 및 기반구축 현황」. 『조명전기설비』, 25-4.
- 임선규. 2021. 「생성모델 연구 동향 및 금융에서의 활용」. 『주간기술동향』, 2021.12.8. 정보통신기획평가원.
- 조정원. 2019. 「중국의 사회·기술시스템 전환은 왜 어려운가?: 신에너지자동차 정책을 중심으로」, 『현대중국연구』, 20-4.
- 최의현. 2020. 「중국의 산업정책과 기술혁신: 차세대자동차 개발을 중심으로」, 『동북아경제연구』, 32권 2호, pp.
- Aghion, Philippe, Antoine Dechezlepretre, David Hemous, Ralf martin, John V. Reeneen. 2012. “Carbon Taxes, Path Dependency and Directed Technical Change: Evidence from the Auto Industry”. *NBER Working Paper*, no.18596.
- Blei, David. M., Andrew Y. Ng, Michael I. Jordan. 2003. “Latent Dirichlet Allocation”. *Journal of Machine Learning Research*, 3.
- Griffiths, Thomas., Mark Steyvers. 2004. “Finding Scientific Topics”. *PNAS*, Apr 6.
- Park, Seong-Taek, Chang Liu. 2020. “A study on topic models using LDA and Word2Vec in travel route recommendation: focus on convergence travel and tours reviews”, *Personal and Ubiquitous*

Computing. (<https://doi.org/10.1007/s00779-020-01476-2>)

Wang Jinli, Yong Fan, Hui Zhang and Libo Feng. 2021. “Technology Hotspot Tracking: Topic Discovery and Evolution of China’s Blockchain Patents Based on a Dynamic LDA Model”. *Symmetry*, 2021–13. <https://doi.org/10.3390/sym13030415>

〈국문요약〉

토픽모형을 이용한 중국 전기차 산업의 기술 분석

본 연구는 중국 전기차 산업의 기술 수준을 텍스트마이닝 기법인 토픽모형을 통해 분석한다. 구체적인 방법론은 잠재 디리클레 할당(LDA, Latent Dirichlet Allocation) 모형이며, 분석대상은 중국이 출원한 국제특허(PCT) 최근 20년간 데이터이다.

분석결과 16개의 토픽을 추출하였고, 이들 토픽의 기술적 추세를 분석할 수 있었다. 대부분의 토픽에서 기술력이 향상되었지만, 특정 분야에서의 기술력 향상이 추세적으로 두드러지게 나타나지는 않았다. 좀 더 세밀한 분석을 위해 토픽을 핫 토픽과 콜드 토픽을 구분하였다. 그 결과 실제로 기술 발전이 두드러진 분야는 철도와 충전소 등 주변부적인 기술이었고, 전기차의 성능과 관련된 핵심기술인 전기 모터제어, 전기에너지 저장장치, 차량용 발전기 등은 추세적으로 약화되었다.

현재 중국정부는 적극적인 산업정책을 통해 전기차의 기술력을 높이려 한다. 전반적인 차원에서의 기술력은 향상되었다고 할 수 있지만, 핵심 기술분야에서 특허가 상대적으로 줄어드는 것으로 나타났다.

<Abstract>

A Study on the Research Trends in Chinese EV Industry using Topic Modeling

Eui-Hyun Choi

This study analyzes the technological level of the Chinese electric vehicle industry through the topic modeling in text mining methods. The specific methodology is the Latent Dirichlet Allocation (LDA) model, and the analysis objects are data from the recent 20 years of international patents (PCT) applied by China.

Because of the analysis, 16 topics were extracted and technical trends of these topics could be analyzed. Although technology improved in most of the topics, technological improvement in a specific field has not been revealed as a trend. For in-depth analysis, the analysis results were divided into hot topics and cold topics. As a result, the topics in which technological development actually stood out were non-core technologies such as railways and charging stations, and the main technologies related to the performance of electric vehicles, such as electric motor control, electric energy storage, and vehicle generators, gradually weakened.

Currently, the Chinese government is trying to improve the technology of electric vehicles through active industrial policies. Although it can be said that the overall level of technology has improved, the number of patents in the core technology fields has been relatively decreased.

중국과 중국학 (제46호)

Key words: EV, Patent analysis, Text mining, Topic Modeling, Big data

• 투고: 2022년 04월 03일 / • 심사: 2022년 04월 24일 / • 게재확정 : 2022년 05월 20일