

국가R&D와 소셜 데이터를 활용한 수소연료전지 기술마이닝과 감성분석

이병희¹, 최정우², 김태현¹

¹한국과학기술정보연구원 NTIS센터

²과학기술연합대학원대학교 과학기술경영정책전공

bhlee@kisti.re.kr, zephy708@ust.ac.kr, heemang@kisti.re.kr

Technology Mining and Sentiment Analysis on Hydrogen Fuel Cell Using National R&D and Social Data

Byeong-Hee Lee¹, Jung-Woo Choi², Tae-Hyun Kim¹

¹NTIS Center, Korea Institute of Science and Technology Information

²Dept. of Science and Technology Management Policy, UST

요 약

온실가스 배출 문제가 세계적인 현안으로 부각되면서 수소를 에너지원으로 사용하는 수소 경제가 주목받고 있다. 수소연료전지는 수소경제의 구성요소 중 하나로, 수소를 활용해 열과 전기를 생산하며 에너지 변환 효율이 높이에 장점이 있다. 본 연구는 세계적인 온라인 커뮤니티인 레딧(Reddit)에서 수집한 수소연료전지와 관련된 소셜 데이터를 텍스트마이닝과 감성분석 기법으로 분석하였다. 분석 결과 9,211건의 댓글을 LDA(Latent Dirichlet Allocation)을 이용해 4개의 토픽 그룹으로 분류할 수 있었다. 이 중 수소연료전지와 관련이 높은 그룹을 선정해 STM(Structural Topic Model) 분석으로 10개 토픽을 추출하였고, 기후 환경, 수소 산업, 수소 차와 관련 있는 토픽 3개를 발견할 수 있었다. 이 연구 결과를 통해 수소연료전지의 세계적으로 실제적인 내용을 빠르고 효과적으로 파악하여 수소연료전지에 대한 예측하고, 우리나라의 수소연료전지 관련 국가R&D의 정책적 방향을 제시하고자 한다.

1. 서론

화석연료와 달리 공해물질을 배출하지 않는 친환경 연료를 사용한다는 점에서 수소에너지를 주축으로 한 수소경제가 전 세계적으로 주목받고 있다. 수소에너지는 수소의 형태로 에너지를 저장하고 사용할 수 있는 신에너지다. 수소는 자연 상태에서 독립적으로 존재하지 않아 천연가스 등 화석연료의 개질, 물의 전기 분해 과정 등을 필요로 하고, 이는 국가R&D를 통해 달성할 수 있다. 우리나라는 2019년 수소경제 활성화 로드맵 발표 이후 정부 차원의 대규모 투자를 실시하고 있다.

한편, 수소연료전지의 최종 수요처 중 하나인 수소연료전지 자동차의 경우 2022년 1월 기준 승용차는 19,270대, 수소버스는 129대로 각각 목표치의 33%, 6.5%에 머무르고 있다[1]. 이와 같은 상황에서 수소연료전지 자동차의 보급 확산을 통한 수소경제의 진전을 위해서는 최종 소비자인 대중의 수소 이용 및 활용에 대한 인식 제고가 필요하다.

레딧(Reddit)은 월간 활성화 유저가 약 5,000만 명에 달하며, 활성화된 게시판인 서브레딧은 약 14만 개에 달하는 등 주요 소셜미디어의 하나로 꼽을 수 있다. 레딧 리뷰 연구[2]에 따르면 레딧을 이용한 저널 연구 338건 중 과학기술 관련은 128건으로, 레딧은 과학기술에 대한 대중의 관심사를 파악할 때 활용할 수 있는 소셜 커뮤니티이다.

본 연구는 NTIS의 국가R&D 데이터를 통해 현재의 수소연료전지 관련 R&D 동향을 분석하고, 레딧의 소셜 데이터를 활용해 수소연료전지에 대한 대중의 인식을 분석하고자 한다. 이종의 데이터 분석을 통해 현재 연구의 트렌드가 대중의 인식과 어느 정도의 갭을 가지고 있는지 확인하고, 대중의 인식을 제고할 필요성을 논의하고자 한다.

2. 연구방법

본 연구는 우리나라의 국가R&D에서 수소연료전지 연구 동향을 분석하기 위해 NTIS 데이터를 활용해 수소연료전지 관련 국가R&D과제 기술마이닝을

수행하였다. 국가과학기술지식정보서비스(NTIS)에서 키워드 수소연료전지로 검색한 결과인 5,023건의 과제를 다운로드하여 수소연료전지의 국내 연구개발(R&D) 과제 현황을 파악하였다.

수소연료전지에 대한 대중의 인식을 조사하기 위해 소셜 커뮤니티인 레딧을 분석했다. 레딧의 소셜 데이터를 수집하기 위해 R 언어의 해당 패키지를 이용해 레딧의 전체 서브레딧에서 ‘Hydrogen Fuel Cell’ 키워드가 포함된 댓글을 9,211건을 다운로드하였다. 레딧 데이터는 날짜 처리 및 다양한 텍스트, 전처리를 활용해 연관 규칙, Bi-그램, N-그램, 동시 발생 등의 분석하였다. 이후 토픽 모델링을 실시하여 수소연료전지와 관련성이 떨어지는 댓글을 제외할 수 있었다. 관련성이 높은 1,544건의 댓글을 구조적 토픽 모델링(STM; Structural Topic Modeling)으로 분석했다.

STM은 디리클레 할당 방식을 사용하는 LDA(LDA; Latent Dirichlet Allocation)의 확장된 프레임워크로써 문서 안에 존재하는 단어의 빈도수 뿐만 아니라 문서에 메타데이터를 활용하여 토픽을 구성하는 단어들의 분포를 결정한다. STM은 메타데이터와 문서 내 존재하는 토픽들의 상관관계에 대해 추정할 수 있고, 각 토픽들의 관계를 구분하여 해석할 수 있는 장점을 가지고 있다[3].

3. 연구결과

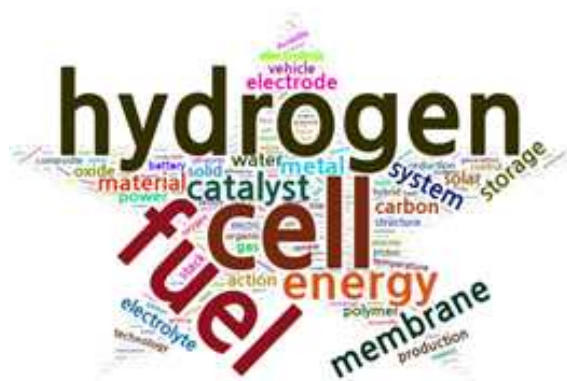
국가과학기술지식정보서비스(NTIS)에서 수집한 수소연료전지 관련 국가R&D과제의 2012년부터 2021년까지 최근 10년간 추이를 살펴보았다. 과제 연구비의 경우 2018년부터 2021년까지 150억 원대에서 400억 원 대로 급격하게 증가하였다. 과제수 또한 연구비와 유사한 경향임을 확인할 수 있었다. 수소연료전지 관련 연구비와 과제가 2018년부터 증가한 원인으로는 문재인 정부의 수소경제 정책의 영향인 것으로 추정된다.

2018년 이후 부처별 수소연료전지 관련 국가 R&D과제를 확인한 결과 과학기술정보통신부, 산업통상자원부, 교육부, 중소벤처기업부 순으로 나타났다. 2018년 이후 수소연료전지 관련 국가R&D과제를 많이 수행한 기관은 한국에너지기술연구원(4.24%), 한국과학기술원(4.09%), 한국과학기술연구원(3.42%), 울산과학기술원(3.01%), 고려대학교(2.14%) 등으로 나타났다.

2018년 이후 연구개발단계는 기초연구가 52.7%

로 가장 많은 비중을 차지하고 있으며, 개발연구(27.7%), 응용연구(14.1%) 기타(5.5%) 순으로 나타났다.

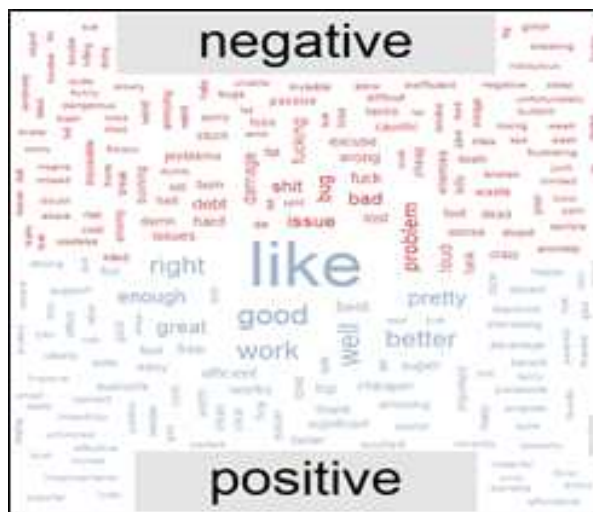
수소연료전지 관련 국가R&D과제 정보에서 영문 키워드를 추출하여 워드클라우드를 표현하면 [그림 1]과 같다. “cell”, “fuel”, “hydrogen” 외에도 “energy“, “catalyst“, “membrane“ 등의 순서로 키워드가 나타났다.



(그림 1) 국가R&D과제의 영문 키워드 워드클라우드.

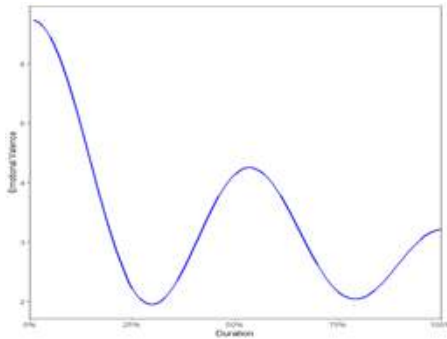
수소연료전지에 대한 대중의 인식을 조사하기 위해 소셜 커뮤니티인 레딧에서 받은 텍스트 데이터로 감성분석을 실시한 결과는 다음과 같다.

[그림2]는 긍정어와 부정어의 분포를 자주 등장하는 단어 빈도를 크기별로 보여준다. 긍정어로는 “Like”, “good” 등이 많이 나타나고 있고, 부정어로는 “problem”, “issue” 등이 많이 나타나고 있다.



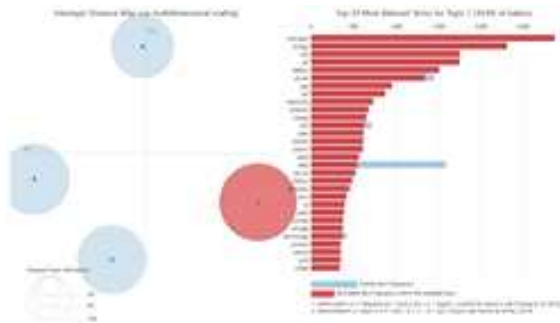
(그림 2) 레딧의 수소연료전지 관련 긍정어와 부정어

또한 연도별로 레딧에서 수소연료전지 관련 긍정과 부정의 상태를 보면 [그림3]과 같이 표현된다. 최초에는 큰 긍정 기대감으로 상승했다가 부정 기대감으로 하강하기를 반복하다가 최근으로 오면서 점차 안정적 상태로 진행됨을 볼 수 있다.



(그림 3) 레딧의 연도별 수소연료전지 긍정과 부정 트렌드

9,211개의 댓글을 활용한 토픽모델링 분석 결과는 다음과 같다. 우선 구어체 문장인 텍스트에서 의미가 없는 댓글을 제외하기 위해 전처리 과정으로 LDA(Latent Dirichlet allocation) 토픽 모델링을 통해 [그림4]과 같이 4개로 분류 후 의미가 있는 토픽 그룹을 추출하여 댓글 1,544건으로 줄였다.



(그림 4) 레딧 댓글 전처리를 위한 LDA 토픽 모델링

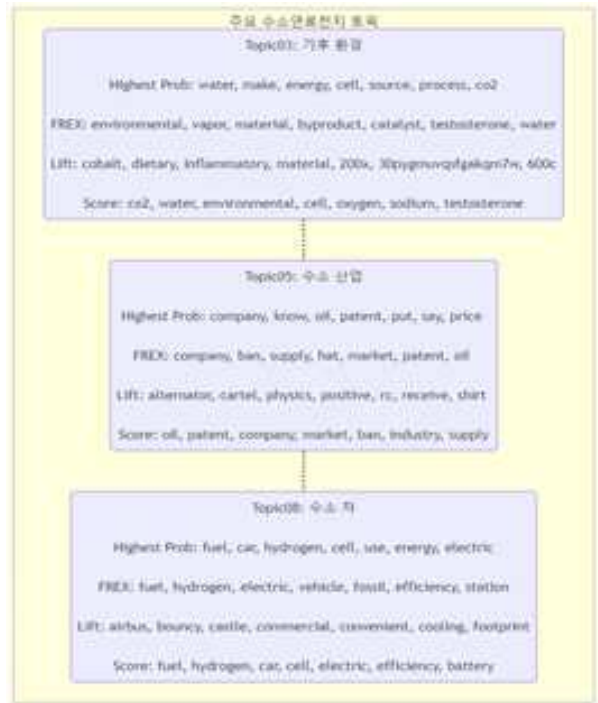
토픽 수를 10개로 지정하고 STM을 실시하였다. 그 결과 7개의 토픽은 수소연료전지와 관련성이 없는 것으로 나타났고, [그림6]과 같이 “기후 환경”, “수소 산업”, “수소 차” 등 3개 토픽을 추출할 수 있었다.

4. 결론

수소연료전지 관련 국가R&D과제 정보에서 영문 키워드를 추출하여 워드클라우드에 표현한 결과 “cell”, “fuel”, “hydrogen”, “energy”, “catalyst”, “membrane” 등의 순서로 키워드가 나타났다. 기초 연구가 52.7%로 가장 많은 비중을 차지한 것을 고려하면, 수소 경제를 달성하기 위한 연료전지, 촉매, 멤브레인 기술 개발에 중점을 둔 국가R&D가 실시되고 있는 것을 확인할 수 있다.

레딧의 주요 수소연료전지 관련 토픽은 “기후 환경”, “수소 산업”, “수소 차”로 추출할 수 있었다. 소

설 미디어를 통해 나타난 대중의 관심은 수소경제의 등장 배경과 이를 활용한 산업, 그리고 최종 수요처인 수소 차에 집중되고 있는 것으로 파악된다.



(그림 5) 레딧 주요 수소연료전지 토픽

두 결과를 종합하면 국가R&D는 기초단계에 머물러 있으나 대중의 관심은 수소경제의 등장 배경부터 최종 수요처까지 폭넓은 인식을 갖고 있는 것으로 파악된다. 연구개발의 촉진을 통해 대중의 관심에 부합하는 최종 수요처 개발까지 신속한 상용화를 기대해 본다.

사사

본 연구는 2022년도 한국과학기술정보연구원(KISTI) 주요사업 과제로 수행한 것입니다(NTIS 과제고유번호 1711173845).

참고문헌

[1] 에너지원별 발전량 현황(연, 2009~2020), 통계청
 [2] 추진기, “소셜 빅데이터 분석에 의한 신 소비시장 트렌드 연구 - ‘나홀로 소비’ 연관어를 중심으로”, 디지털융복합연구, 18(2), 367-376. 2020.
 [3] 박한샘, 김동현, 장성주. “구조적 토픽 모델링 기반 스마트 시티 연구 동향 분석”. 한국디지털콘텐츠학회 논문지, 20(9), 1839-1846, 2019