



2021년 5호

GTC BRIEF

GTC BRIEF는 기후기술과 관련하여 시의성 있는 현안 및 동향정보를 알기 쉽게 정리한 자료임



1. 국내외 녹색회복 추진현황과 시사점: 세계회복관측소 자료를 중심으로 01

_ 강한나 김민철 한민지 / 정책연구부

2. 2020년 국가연구개발과제의 기후기술 분류체계 기반 딥러닝 분류모델 적용 연구 13

_ 주경원 한수현 / 기술총괄부

2021년 5호

GTC BRIEF

2021
OCTOBER
Vol.2 No.5

1. 국내외 녹색회복 추진현황과 시사점: 세계회복관측소 자료를 중심으로 01

_ 강한나 김민철 한민지 / 정책연구부

2. 2020년 국가연구개발과제의 기후기술 분류체계 기반 딥러닝 분류모델 적용 연구 13

_ 주경원 한수현 / 기술총괄부

ISSUE
01

국내외 녹색회복 추진현황과 시사점: 세계회복관측소 자료를 중심으로

강한나, 김민철, 한민지 / 정책연구부 | hkang@gtc.re.kr, eco@gtc.re.kr, a2pmin@gtc.re.kr

하이라이트

- **(녹색회복 현황진단)** 세계회복관측소(Global Recovery Observatory) 자료에 따르면 '20년 한해 세계 경제규모 상위 50개국의 코로나 경기회복책 대비 환경에 긍정적 영향을 가지는 녹색회복책 비중은 2.5%에 그친 것으로 나타나, 향후 포스트코로나 시대 '더 나은 재건(building back better)'을 위해서는 국가 예산과 투자에서 녹색 우선순위 확대가 요구됨을 시사하고 있음
- **(녹색회복 중점분야)** △ (녹색에너지) 송전용량, 스마트그리드, 전력저장기술 확대, △ (녹색교통) 전기자동차(EV) 보조금, 전기 및 수소 기반 청정 교통수단 전환·확대, △ (녹색건물) 녹색건물·주택 개조(retrofit) 보조금 및 인프라
- **(녹색회복 정책방향)** 녹색지출 확대 및 균형 제고를 목적으로 △ 녹색예산 우선지출 이행방안 수립(온실가스감축인지 예·결산제, 기후변화영향평가 연계), △ 녹색 중소·벤처기업 대상 투자 확대(기후변화대응기금, 그린 뉴딜 분야 기금 활용), △ 녹색자원(재생에너지, 그린수소 등) 기반 산업구조로의 공정한 전환 지원

키워드

- 녹색회복, 포스트코로나, 경기부양책, 환경·기후 영향, 탈탄소사회 전환, 그린 뉴딜

왜 녹색회복인가

코로나 위기 대응을 위한 경기부양책 추진에 있어 환경적 지속가능성을 고려하는 녹색회복 추진

- 녹색회복(green recovery)은 경기부양책 측면에서 녹색부양책이라고도 표현되며, '단기적으로는 경제활동을 촉진하고, 장기적으로는 환경 및 천연자원의 질을 보존 및 보호하고 강화하기 위한 상생 정책' 등으로 정의됨¹⁾
 - 장기화된 코로나 팬데믹으로 인한 경제위기 상황에서 경제부양 정책이 추진되고 있으나, 환경이 고려되지 않은 근시안적 관점의 경기부양책은 기후변화 위기를 더욱 가속화시켜 예측 불가능한 위험을 야기하는 등 현재의 경제위기와는 비교할 수 없는 부정적 파급력을 가져올 수 있음
 - 최근 발표된 연구에서는 코로나 팬데믹으로 인한 각종 경기부양책이 탄소배출량을 증가시킬 수 있다고 경고하며 녹색부양책 대응의 중요성을 환기시킨 바 있음.²⁾
 - ※ '07년 세계경제위기 이후 글로벌 탄소배출량이 급격하게 감소하였으나 각 정부의 경기부양 정책에 환경적 고려가 충분하게 반영되지 않음에 따라 '10년 탄소배출량 최고치를 기록한 전례가 있음³⁾

- G20 에너지·환경장관회의*가 처음 개최된 지 1년 후인 '20. 9월, 환경 부문만을 다룬 첫 G20 환경장관회의에서 「환경장관 선언문」이 채택되어, '코로나-19 극복을 위한 환경적으로 지속가능하고 포용적인 회복 정책의 추진'에 대한 각국의 의지를 표명하며 녹색회복에 대한 국가적 합의 토대가 마련됨⁴⁾
 - * 지속가능발전을 위한 에너지전환 및 환경보호의 중요성 인식을 공유하고, 파리협약 이행, 청정에너지 기술 혁신 등을 촉진하는 합동선언문과 자원효율성 촉진, 해양폐기물 저감, 기후적응 기반시설(인프라) 확대를 위한 환경장관선언문 채택
- '21. 5월에는 이틀에 걸쳐 환경 분야 다자 정상회의인 제2차 P4G* 서울 녹색미래 정상회의가 '포용적인 녹색회복을 통한 탄소중립 비전실천'이라는 주제로 개최되었고, 문재인 대통령은 개회사를 통해 글로벌 기후·환경 위기극복을 위한 포용적 녹색회복 추진의 필요성을 강조함
 - * P4G(Partnering for Green Growth and the Global Goals 2030)는 녹색성장 및 글로벌 목표 2030을 위한 공공-민간 파트너십 강화 이니셔티브로 녹색경제 관련 5대 중점분야(식량·농업, 물, 에너지, 도시, 순환경제)에서 민관협력 촉진 노력
 - 코로나 팬데믹과 같은 전염병 대응이 기후변화 해결과 직결되어 있다는 인식의 확산⁵⁾과 맞물려 코로나 위기로부터의 녹색회복이 '탄소중립', '탈탄소사회 전환', '지속가능발전과 녹색경제' 등의 글로벌 목표 달성을 위해 고려되어야 할 중요한 변수로 떠오르고 있음
 - 이러한 의미에서 최근의 녹색회복은 탈탄소사회로의 전환 및 자원효율성 향상을 바탕으로 한 경쟁력 제고 등 장기적인 환경목표에 맞춘 경제회복 활동을 촉진하는 전략으로도 이해할 수 있음
 - 위와 같은 흐름에 맞춰 본 정상회의에서는 탄소중립사회로의 전환을 위한 녹색회복 추진을 목적으로 하는 「서울선언문」을 채택하였으며, 이를 통해 △ 지구온도 상승 1.5도 이하 억제, △ 재생에너지 발전 비중 확대를 통한 에너지 전환 촉진 및 탈석탄 지향, △ 해양플라스틱 대응 등 기후 및 환경목표에 대한 기후선도국과 개도국의 의지를 결집하였음⁶⁾
- 국제사회의 합의와 함께 녹색회복을 위한 우리나라의 정책방향은 대한민국 대전환 뉴딜 종합계획 중 '그린 뉴딜'로 대표될 수 있음
 - '20. 7월 발표된 「한국판 뉴딜 종합계획」은 코로나 팬데믹으로 인한 극심한 경제 침체 위기를 극복하고 구조적 대전환을 이루고자 마련된 종합적 국가발전 전략으로 디지털 뉴딜과 그린 뉴딜 두 축으로 추진되고 있음
 - 뉴딜 발표 후 당시 환경부 장관(조명래)은 유엔환경계획(UNEP: UN Environment Program) 사무총장과의 양자회담에서 한국 그린 뉴딜이 미래사회 청사진으로서 탄소중립 달성 등 전 세계 녹색회복의 본보기가 되겠다고 밝힌바 있음⁷⁾
 - ※ 뉴딜 종합계획에서 온실가스 저감 사업이 우선투자 대상으로 논의되고 있으며, 이는 코로나 경기침체 대응에서 '녹색경기부양'의 중요성이 인식되었음을 시사

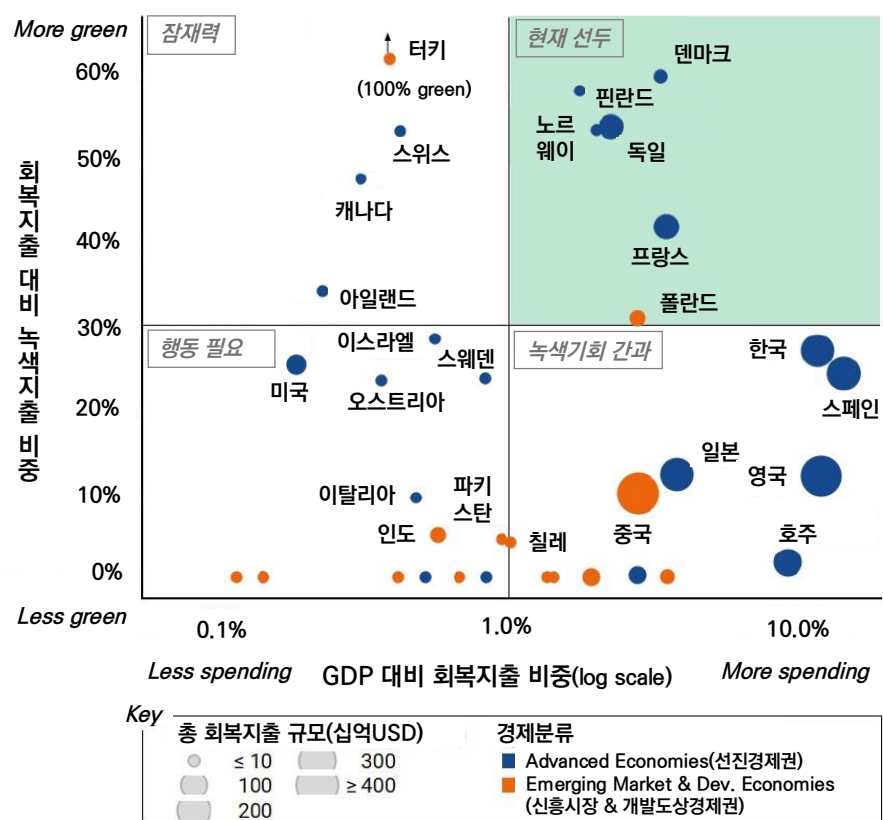
- '21년 '한국판 뉴딜 2.0 추진계획'에서도 미래 성장 동력으로서 탄소중립 전략이 강조되며 코로나 위기 회복을 위한 녹색대응 의지가 다시금 확인되었음
 - ※ '21. 8월, 2050 탄소중립위원회는 석탄발전, 전기수소차 비율, 건물에너지 관리, CCUS(탄소포집 및 활용·저장) 및 흡수원 확보량 등 핵심 감축수단을 달리 적용한 '2050 탄소중립 시나리오 초안'을 공개
- 이러한 녹색회복책은 단기적으로는 경기부양과 고용 창출, 중장기적으로는 환경적 지속 가능성 제고라는 편익을 가져올 것으로 예상됨⁸⁾
 - ※ 유럽의 경우, 750억~1,500억 유로 상당의 녹색회복 자금이 1,800억~3,500억 규모의 총부가가치를 생산하고 3백만 개 이상의 새로운 일자리를 창출할 것으로 예측⁹⁾
- (본 브리프의 목적) 국내외 코로나 위기 대응을 위한 경기부양책 추진에서 녹색회복 현황을 살펴보고, 우리나라 상위 녹색지출 분야를 중심으로 향후 녹색회복 정책 방향을 모색하고자 함
 - 본 브리프 내 녹색회복 현황은 세계회복관측소(Global Recovery Observatory) 자료를 토대로 작성('20년 한해 기준)

녹색회복 추진 현황 세계 경제규모 상위 50개국의 코로나-19 관련 회복지출 대비 녹색지출 비중은 2.5%로 낮은 수준('20년 기준)

- 옥스퍼드대학과 유엔환경계획(UNEP)의 '20년 코로나-19 관련 재정정책(fiscal spending policy)* 분석 보고서¹⁰⁾에 따르면 주요 경제국의 회복지출 대비 녹색지출 비중은 상당히 낮은 것으로 나타남
 - * 경기과열이나 침체가 일어날 경우 정부지출이나 조세를 변화시켜 총수요에 영향을 주고 이를 통해 경기를 조절하는 정책
- 분석대상은 세계회복관측소 데이터베이스¹¹⁾에 포함된 경제규모 상위 50개국 3,000개 이상의 재정정책('20년 한해 기준)
 - ※ 해당 데이터베이스는 각 국가가 발표한 재정정책 지출에 대한 계획으로 이행여부는 미반영되어 있으며, 계획 발표에 따라 주기적으로 갱신
- '20년 코로나-19 관련 재정지출 총 규모는 14.6조 달러로, 그 중 11.1조 달러는 단기 구제책(rescue measures), 1.9조 달러는 장기 회복책(recovery measures) (나머지 1.6조 달러는 분류가 불명확)
- 회복지출 중 녹색지출(환경적으로 긍정적 영향을 가져온 지출)은 '20년 3월 이후 꾸준히 증가하였지만 그 규모는 3,410억 달러로 회복지출 대비 18%에 그침
- 위 분석의 50개 국가별 녹색지출 현황을 보면 덴마크, 핀란드, 노르웨이 등 북유럽 국가들의 회복지출 대비 녹색지출 비중이 가장 높은 것으로 나타남[그림 1]
 - ※ 단, 녹색지출의 절대규모로는 한국, 스페인, 독일이 상위를 차지하나 그 이유는 회복지출 규모 자체가 크기 때문인 것으로 이해

- 한국, 스페인, 영국, 일본, 중국은 회복지출 규모와 GDP 대비 회복지출 비중 모두 높은 수준인 것에 비해, 회복지출 대비 녹색지출 비중은 30% 이상을 넘지 않음
- 반면, 터키, 캐나다, 스위스의 지출 규모는 작으나 녹색지출 비중이 크며, 그 중 터키*의 경우 회복지출 전체가 녹색지출로 구성됨
- * 회복지출 대비 녹색지출 비중이 100%인 것은 지출 규모 자체가 작기 때문일 수 있으나, '30년까지 온실가스 21% 감축을 목표로 재생에너지, 산림 등에 투자 확대 추진 중
- 50개 국가 중 13개는 녹색지출 비중이 0%로 유럽의 네덜란드와 포르투갈, 중남미의 멕시코와 아르헨티나, 아시아의 태국, 말레이시아 등이 여기에 속함
- 경제분류 기준(IMF)으로 살펴보면, 선진경제권(AE: Advanced Economies)에 속한 국가들이 신흥시장 및 개발도상경제권(EMDE: Emerging Market & Dev. Economies) 국가들에 비해 GDP 대비 회복지출 비중과 회복지출 대비 녹색지출 비중 모두가 높은 것으로 나타남
- ※ 유엔개발계획(UNDP) 인간개발지수(HDI: Human Development Index)가 높을수록 GDP 대비 회복지출 비중이 높은 것과도 비교 가능

[그림 1] 회복지출 및 녹색지출 기준에 따른 경제규모 상위 50개국 분포('20년)

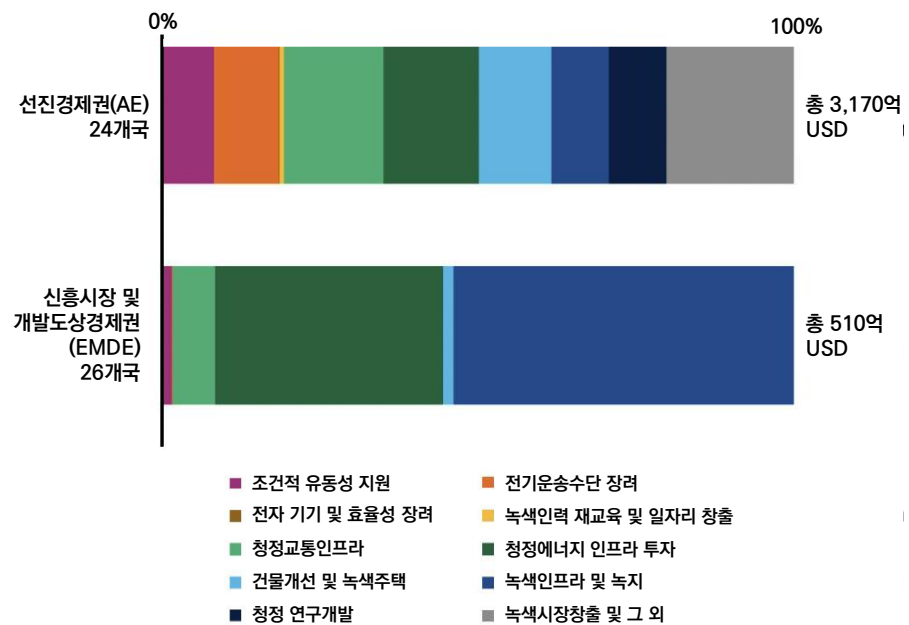


※ 출처: Global Recovery Observatory; O'Callaghan(2020), p. 12에서 재인용

- 50개 국가 중 선진경제권(AE)의 녹색지출은 비교적 다양한 정책으로 구성되어 있는 반면, 신흥시장 및 개발도상경제권(EMDE)의 녹색지출은 청정에너지나 청정교통에 한정되어 있음[그림 2]

- EMDE의 경우 높은 선진기술 진입장벽, 낮은 파생기술(enabling technology) 보급률 및 기술개발 흡수력, 자본부족 등이 녹색지출 확대의 걸림돌로 작용하고 있음
- 녹색지출의 정책부문별 구성으로 보면, 높은 실업률에도 불구하고 AE와 EMDE 모두 직업훈련에 대한 녹색지출이 거의 전무하며, 에너지, 농업, 교통, 건물 부문에서의 구조적 전환 대응 차원의 녹색인력 양성 노력도 매우 미비한 것으로 나타남

[그림 2] 경제규모 상위 50개국 녹색지출의 정책부문별 구성('20년)



※ 출처: Global Recovery Observatory; O'Callaghan(2020), p. 17에서 재인용

- 해당 보고서는 이러한 분석결과를 토대로, 현재 세계는 녹색 우선순위가 결여된 포스트코로나 대응을 펼쳐왔으며 그 결과 코로나위기 이후 '더 나은 재건(building back better)'을 이루고 있지 못하다는 평가를 내림
 - 녹색지출이 온실가스과 대기오염 감축 등에 기여한다는 점에서, '20년 한해 코로나 관련 재정정책의 2.5%, 회복지출의 18%만이 녹색지출이라는 결과는 지금까지 글로벌 차원의 기후위기 대응 노력이 충분하지 않은 것으로 이해될 수 있음
 - ※ 단, 세계회복관측소 데이터베이스 갱신 현황을 보면, '21.8월 현재 회복지출 대비 녹색지출 비중은 23%로, '20년에 비해 증가 추세인 것으로 확인
 - 회복지출 중 상당부분이 대기오염을 증가(16.4%)시키거나 도로교통, 국방사업 등 환경에 부정적 영향을 가져오는 자연자본으로 활용(17%)되었다는 점도 향후 회복지출 계획 시 고려가 필요한 사항임
 - 특히 높은 탄소집약도(GDP 대비 CO₂ 배출량)를 가졌음에도 녹색지출 비중이 적은 국가들이 다수 존재함에 따라 전 지구적 파리기후변화협정 목표*를 이루는데에 큰 우려로 인식되고 있음
 - * 지구의 평균 온도 상승 1.5도를 넘지 않도록 노력하겠다는 최초의 국제사회 기후 합의로 각 국가는 '30년까지의 온실가스 감축목표를 설정

녹색회복 중점분야 녹색에너지

- (중요성) 에너지는 글로벌 온실가스 배출의 73.2%¹²⁾를 차지하며, 녹색에너지 전환에 따른 경제적 파급효과와 민간투자 기회가 높아 경제전반에 밀접한 영향을 미치는 분야임¹³⁾
 - 재생에너지 사용을 위해서는 석탄에너지와는 전혀 다른 송전·배전방식이 요구되기 때문에 송전 인프라, 배전단계에서의 전력망(grid), 배터리 등 에너지 저장 기술에 대한 수요가 증가할 것임
 - 또한 녹색에너지(재생에너지, 수소 등)는 교통, 농업, 산업 등 다부문에 걸쳐 탈탄소 기술 개발·적용에 있어 필수적인 요소이며, 따라서 녹색에너지에 대한 녹색 민간 투자 연계가 용이하다는 특징이 있음
- (현황) '20년 기준 녹색에너지 분야 투자는 총 660억 달러이며, 세부적으로는 재생에너지, 바이오연료, 송전·배전, 수소, 배터리 및 저장 등으로 구성됨[그림 3]
 - (스페인 에너지전환) '정의롭고 포용적인 에너지전환'¹⁴⁾ 정책 일환으로 재생에너지 생산 및 생산부문 활용, 에너지저장 기술, 그린수소(72억 달러)
 - (독일 그린수소) '2020 국가수소전략'¹⁵⁾ 내 △ 수소무역과 유럽 수소사회 리더십 확보(24억 달러), △ 수소생산 및 산업·교통·난방 부문에서 활용(83억 달러)
 - (한국 재생에너지) '그린 뉴딜'¹⁶⁾ 전략 내 재생에너지, 수소, 스마트그리드*(133억 달러)
 - * '10년 스마트그리드 국가로드맵에서 '30년까지 스마트그리드(지능형전력망) 전국망 구축 비전을 제시하고, '22년까지 제2차 지능형전력망 기본계획을 수립하여 시행 중¹⁷⁾

[그림 3] 경제규모 상위 50개국 녹색에너지 분야 투자 세부구성('20년)



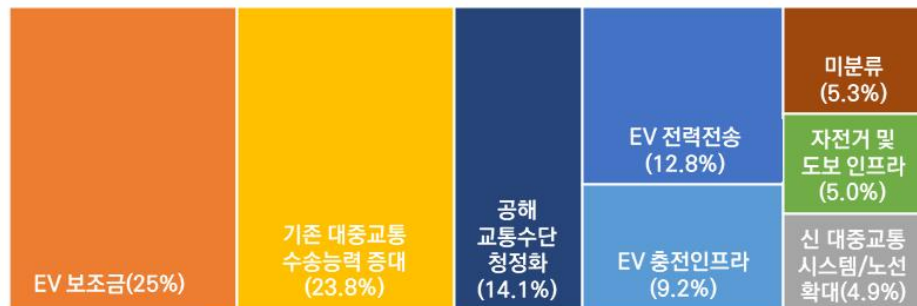
※ 출처: Global Recovery Observatory

- (기회 전망) 전 세계적으로 재생에너지 생산 비용이 크게 감소하고 있으며 송전용량, 스마트그리드, 전력저장기술 확대 등 녹색에너지 시장이 넓어지고 있음¹⁸⁾
 - 특히 수소는 탈탄소화를 위한 핵심 녹색에너지 자원으로 타 재생에너지와 비교하여도 점차 가격경쟁력을 확보할 것으로 예상되는 바, 기술개발 및 인프라 투자 기회가 지속적으로 증가할 것임

녹색교통

- (중요성) 전기자동차(EV: Electric Vehicle) 분야는 경제·환경·사회 부문에 긍정적 영향을 줄 수 있는 보편적 녹색교통 수단으로 재생에너지 보급이 높을수록 그 영향은 확장됨¹⁹⁾
 - 온실가스 배출과 대기오염 감소는 대중교통이 자가용을 대체할 때 중요도가 향상되며 에너지 효율성 측면에서 긍정적임
 - 또한 자전거 및 보행 인프라 투자로 건설 분야 일자리 확충 및 저소득층의 녹색교통에 대한 수요에 대응함
- (현황) '20년 기준 녹색교통 분야 투자는 총 861억 달러이며, 세부적으로는 EV 보조금 정책, 기존 대중교통 수송능력 증대, 교통수단 청정화 등으로 구성됨[그림 4]
 - (EV 관련 정책) EV정책 채택은 재생에너지 보급률이 높은 국가에서 온실가스 배출 감소와 밀접한 관련이 있으며, 내연기관 차량을 EV로 교체하여 전반적으로 건강 중심의 삶의 질 향상에 기여함으로써 대기오염 감소에 따른 편익이 증대됨²⁰⁾
 - ※ EV보조금 제도는 녹색교통 수단에 대한 접근성 확대 잠재력을 가진 분야일 뿐 아니라 EV구매 촉진을 위한 인센티브 조치 마련, 충전 인프라 확장을 통한 단기간 다수 일자리 창출에 이바지할 것으로 기대
 - (대중교통 수송능력 증대) 전기 및 수소 기반 대중교통 증대는 도시의 대기오염 감소에 기여하나 한편으로는 승객 안전문제와 관련한 각별한 고려 필요
 - (교통수단 청정화) 공해 교통수단 청정화는 탄소감축은 물론 에너지 효율성 제고 효과

[그림 4] 경제규모 상위 50개국 녹색교통 분야 투자 세부구성('20년)



(총액: 861억 달러)

※ 출처: Global Recovery Observatory

- (기회전망) EV 인프라에 대한 전략적 투자를 통하여 녹색교통의 중요성을 부각하고 있으며 대중교통에 대한 새로운 투자와 프로그램 확장은 일자리 기회의 접근성 향상, 공중보건 개선과 사회·경제적인 혜택을 가져올 것으로 전망됨
 - (폴란드 EV 정책) 폴란드 정부는 '20년 6월 EV 생산 및 보급 촉진을 포함하는 녹색 투자부양패키지(21억 달러)를 발표함²¹⁾
 - ※ 폴란드는 유럽 내 승용차 주요 생산국이나 '19년 기준 전기자동차 생산비중은 0.5% (스웨덴 12%, 중국 4.2%, 독일 4.0%)²²⁾

- (EV 생산 및 판매지원) EV 생산 투자는 많은 일자리를 빠르게 창출하는 효과는 물론 온실가스 배출 감축목표 달성에 기여하므로 정부 차원의 추가 조치가 필요함
※ EV 판매에 대한 공적 지원으로 큰 수혜를 입을 수 있는 국가는 인도, 일본과 같이 기존 자동차 산업이 큰 국가들이며 유가 지원의 단계적 폐지 및 EV 충전소와 같은 물리적 인프라에 대한 투자와 결합할 때 효과적일 것으로 예상²³⁾

녹색건물과 에너지효율

- (중요성) 건물은 에너지 관련 탄소배출의 28%를 차지하는 분야이며, 건물에서 사용되는 에너지의 절감을 위해서는 에너지 효율성 향상이 가장 핵심적인 요소임²⁴⁾
※ 녹색건물은 건축 주기 전반에서 에너지 절약 및 환경 보호를 고려하고 건물 사용자에게 건강한 사용공간을 제공하는 것을 목적으로 하며 에너지 소요량 최소화(패시브), 에너지 고효율화(액티브), 신재생에너지 활용으로 구성
- 녹색건물은 에너지 절약에 따른 탄소배출 및 폐기물 저감 효과와 더불어, 유지비 절감 등 장기적인 경제성 향상에 따른 에너지 빈곤층 지원에 기여할 수 있음
- (현황) '20년 기준 녹색건물 분야 투자는 총 352억 달러이며, 녹색 개조(retrofit) 및 옥상 태양광으로 구성됨[그림 5]
 - (영국 녹색주택 보조금) 주택개량을 통해 에너지 효율성을 향상시키는 '녹색주택 보조금사업'(30억 파운드, 그 중 5억 파운드는 저소득 가구 대상) 내 바우처(voucher)를 발급하여 주택의 전기 히트펌프, 단열재 설치 작업 비용을 일부 보조²⁵⁾
※ '21.3월 영국 정부는 행정력 미비로 인해 해당 사업을 종료하고, 대신 지방정부 주도 보조금 사업으로 3억 파운드를 마련
 - (네덜란드 제로에너지 주택) '에너지스프롱(Energiesprong)' 사업 내 낡은 주택의 녹색 개조를 통한 넷제로화, 생애 파이낸싱(whole-life financing) 기법*을 적용하여 주택조합 형태로 관리²⁶⁾
* 주택 임차인은 에너지 효율 증가에 따른 에너지 사용료 절감분을 주택조합에 지불하고 이를 주택개조 비용으로 충당
 - (캐나다 제로배출 건물) 기후위기 행동계획의 일환으로 '30년까지 상업용 건물 및 독립주택의 탄소배출을 절반으로 감소'²⁷⁾

[그림 5] 경제규모 상위 50개국 녹색건물 분야 투자 세부구성('20년)



※ 출처: Global Recovery Observatory

- (기회 전망) 녹색주택 개조는 배출 감소 효과뿐 아니라 에너지 빈곤 및 불평등 해결 측면의 사회적 비용 절감에 기여할 수 있음²⁸⁾
 - 기존의 에너지 취약소비자 대상 단기적인 사회정책(소득지원 및 보조금)에서 한 단계 나아가, 에너지 빈곤의 악순환을 근본적으로 해결할 대안으로서의 녹색주택 개조 사업이 활발히 추진되고 있음

우리나라 녹색회복 정책방향 제시

코로나 위기로 인한 탈탄소사회 달성의 지연 방지를 위해서는 녹색예산 우선투자 등을 통한 녹색지출 확대와 고탄소 산업 공정전환 지원을 통한 사회적 합의 도출 필요

- **(현상진단: 녹색지출 확대 및 균형 제고 필요)** '21.8월 현재 우리나라의 회복지출 대비 녹색지출 비중은 26.7%로 유럽국가에 비해 높지 않은 수준이며, 녹색지출 구성으로 보면 근로자 훈련, 일자리 창출 등 비(非)인프라 투자가 미비함
 - (지출 확대) 한국판 뉴딜 정책(~'25년)으로 인해 회복지출의 규모는 확대된 것에 반해 그 중 상당 지출이 기후·환경에 부정적인 영향을 가진다는 점에서 **녹색지출 비중 확대를 통해 그 영향을 최소화**하여야 함
 - ※ 주요국 회복지출 대비 녹색지출 비중: 벨기에(89.4%), 캐나다(74.5%), 덴마크(63.2%), 독일(47.3%)
 - (균형 제고) 녹색 인프라 투자(전기차 보조금, 재생에너지, 건물 개조 등)와 더불어 **사회·경제적 관점의 녹색기반 조성**(사회적 약자 포용, 녹색전환에 따른 일자리 창출 및 근로자 재교육) 노력은 녹색전환으로의 다각적 접근을 위해 필수적인 요소임
 - ※ 녹색지출에 일자리 창출 및 근로자 재교육이 포함된 경우는 스웨덴, 스페인, 프랑스 등 6개 국가(총 23.8억 달러)에 한정
- **(정책방향 1: 녹색예산 우선지출 이행방안 수립)** 국가 예산과 기금에 대하여 온실가스 감축인지 예·결산서 작성 의무화를 골자로 하는 「국가재정법」과 「국가회계법」 일부 개정법률안이 지난 5월 국회 본회의를 통과한 바 있으며, **온실가스감축인지 예·결산 제도 도입('23 회계연도)에 따른 구체적 운용계획 마련**이 시급함
 - (조직기반) △ 기획재정부와 환경부를 주부처로 하는 부처합동 온실가스감축 TF 구성·운영, △ 온실가스감축인지 예·결산안 심사·평가위원회 구축·운영, △ 온실가스감축인지 예·결산 정보시스템 구축·운영(디지털예산회계시스템(dBrain) 연계), △ 온실가스감축효과 검증·평가기구 구축·운영에 대한 추진계획 마련
 - (평가체계) △ 온실가스감축사업 유형화, △ 온실가스감축예산 방법론 개발* 및 이를 적용한 시범사업 운영, △ 예·결산서에 기후변화 영향 평가항목 구성하여 인센티브 제공
 - * 도입 시점까지의 짧은 시간을 고려, 프랑스 탄소인지예산에서 사용하는 단순한 태깅(긍정, 중립, 부정)을 우선 적용하고 향후 고도화하는 방안 고려 가능²⁹⁾

- **(정책방향 2: 녹색유망기업 투자 확대)** 우리나라 회복지출에서 가장 높은 비중을 차지하는 중소·벤처기업 유동성 지원이 환경에 부정적 영향을 미친 것으로 분석, **중소·벤처기업에 대한 재정지출을 녹색으로 전환·확대**시켜야 할 필요성을 제기함
 - 녹색기술 중소·벤처기업 대상의 투자 확대를 위해 출자사업 중 모태펀드*를 이용, 기존 '중진계정'에서 녹색기술 기업을 대상으로 하는 펀드를 마련하는 방안을 고려할 수 있음
 - * '05년 벤처투자자원 공급체계 마련을 위해 결성된 펀드로 '35년까지 운용
 - 기후변화대응기금(2조 5천억원), 한국판 뉴딜 종합계획 후속조치 일환인 '스마트 대한민국펀드' 내 그린 뉴딜 분야(1,300억원)를 녹색기술 중소·벤처기업 육성* 자금으로 일정 부분 배정(earmark)하는 방안을 추진할 수 있음
 - * 그린 뉴딜 5대 주요사업 내 '녹색산업 혁신성장 생태계 조성'과 관련
 - 녹색기술 관련 투자기업 선정은 신재생에너지, 탄소저감, 녹색교통, 녹색건물(주택) 등을 포함하도록 녹색기술인증제*와 연계하여 추진 가능함
 - * 온실가스 감축, 에너지 이용 효율화, 청정에너지생산 등 사회·경제활동 전 과정에 걸쳐 에너지와 자원을 절약하고 효율적으로 사용하여 온실가스 및 오염물질 배출을 최소화하는 기술에 대해 기술우수성, 녹색성 평가항목에 따라 인증하는 제도
- **(정책방향 3: 산업부문 공정한 전환 지원)** 산업(특히 고탄소배출 산업)은 탈탄소 경제로의 전환에 밀접하게 영향을 받는 분야로, **산업이 녹색자원(재생에너지, 그린수소 등) 기반 구조로 전환하는 과정의 사회·경제적 영향을 최소화하는 포용적 전략**이 필요함
 - ※ EU는 유럽그린딜(European Green Deal) 추진에 공정한 전환 메커니즘(Just transition mechanism)을 도입하고 경제구조 전환에 영향을 받는 지역을 대상으로 '27년까지 650억~750억 유로를 지원할 계획
 - (현재 여건) 우리나라 그린 뉴딜 예산 중 '신재생에너지 확산기반 구축 및 공정한 전환 지원'에 9조 2천억 원(~'25년)이 배정되어 있으나, 산업부문별 피해를 체계적으로 전망하고 이해관계자 의견수렴을 통해 공정전환 지원 수요를 파악하여 종합적 지원 정책을 수립하는 등의 시도는 부족함³⁰⁾
 - (종합적 지원체계 구축) 녹색사회로의 전환 과정에서 축소 또는 소멸이 불가피할 것으로 예상되는 에너지·제조 기업의 사업전환, 해당 근로자 실업 대비 및 녹색산업 전직훈련* 등 기업과 근로자 모두를 위한 선제적·체계적 지원방안 마련이 필요함
 - * 특히 근로자가 재직 중 직무전환 훈련을 받을 수 있는 환경을 조성하기 위해서는 장기 유급휴가 제공, 상시 훈련과정 개설, 훈련비 자부담 면제 등의 지원 필요
 - (지자체 지원계획 마련) 산업전환에 따른 지역 피해를 최소화하기 위해서 지역별 위기요소 파악 및 대응방안 도출을 통해 이에 상응하는 재원할당 및 지원계획을 마련하여야 함
 - ※ 고탄소 산업 집중 지역(예, 충남-석탄발전, 울산-자동차부품, 부산-조선)을 대상으로 우선적·집중적 지원 필요

참고문헌

- 1) Strand, J. and Toman, M., Green stimulus: Economic recovery, and long-term sustainable development, Policy research working paper no. 5163, World Bank, 2010
- 2) Yuli, S., Ou, J., Wang, D. Zeng, Z., Zhang, S., Guan, D. and Hubacek, K., Impacts of COVID-19 and fiscal stimuli on global emissions and the Paris Agreement, Nature Climate Change, 11, pp. 200-206, 2021
- 3) Jaeger, J., Westphal, M. I., and Park, C., Lessons learned on green stimulus: Case studies from the global financial crisis, Working paper, World Resources Institute, 2020
- 4) 대한민국정책브리핑, G20 환경장관, 코로나19 극복을 위한 녹색회복 합의, 2020.9.17. Retrieved from <https://www.korea.kr/news/pressReleaseView.do?newsId=156411382> (2021.8.21. 검색)
- 5) 김상훈, 코로나19 근본 원인은 기후위기...우리 사회 저탄소 구조로 전환할 것, 나라경제 특집 9월호, KDI 경제정보센터, 2020, Retrieved from https://eiec.kdi.re.kr/publish/naraView.do?fcode=00002000040000100001&cidx=12980&sel_year=2020&sel_month=09 (2021.08.26. 검색)
- 6) 서울선언문 전문, Retrieved from <https://2021p4g-seoulsummit.kr/notice/detail.do> (2021.8.28. 검색)
- 7) 환경부 보도자료, 한국의 그린 뉴딜, 전 세계 녹색회복의 본보기가 되다, 2020.11.25., Retrieved from <http://me.go.kr/home/web/board/read.do?menuId=286&orgCd=&boardId=1413290&boardMasterId=1> (2021.8.28. 검색)
- 8) 김호석, 코로나19 경기 대응을 위한 환경 분야 재정지출 확대의 유효성: 그린 뉴딜의 경제학, 자원·환경경제연구, 29(2), pp. 293-312, 2020
- 9) McKinsey Sustainability, How a post-pandemic stimulus can both create jobs and help the climate, 2020.5.27., Retrieved from <https://www.mckinsey.com/business-functions/sustainability/our-insights/how-a-post-pandemic-stimulus-can-both-create-jobs-and-help-the-climate> (2021.8.2. 검색)
- 10) O'Callaghan, B. J. and Murdock, E., Are we building back better? Evidence from 2020 and pathways to inclusive green recovery spending, United Nations Environment Programme, 2021
- 11) Global Recovery Observatory, Retrieved from <https://recovery.smithschool.ox.ac.uk/tracking/> (2021.08.11. 검색)
- 12) World Resource Institute, Global GHG emissions by sector, 2020
- 13) International Renewable Energy Agency, REmap 2030: A renewable energy roadmap, 2014
- 14) Presidency of the Government of Spain, Recovery, transformation, and resilience plan, 2020
- 15) Federal Ministry for Economic Affairs and Energy of Germany, The national hydrogen strategy, 2020
- 16) 관계부처합동, 한국판 뉴딜 종합계획, 2020
- 17) 산업통상자원부, 제2차 지능형전력망 기본계획(2018-2022), 2018
- 18) International Renewable Energy Agency, Green hydrogen supply: A guide to policy making, 2021
- 19) Ke, W., Zhang, S., He, X., Wu, Y., and Hao, J. Well-to-wheels energy consumption and emissions of electric vehicles: Mid-term implications from real-world features and air pollution control progress. Applied Energy, 188, pp. 367-377, 2017
- 20) Buekers, J., van Holderbeke, M., Bierkens, J., and Int Panis, L., Health and environmental benefits related to electric vehicle introduction in EU countries, Transportaion Research Part D: Transport and Environment, 33, pp. 26-38, 2014
- 21) van Sonja, R. Poland sows the seeds of a post-Covid future beyond coal, Energy Monitor, Retrieved from <https://energymonitor.ai/policy/green-deals/poland-sows-the-seeds-of-a-post-covid-future-beyond-coal> (2021.08.26. 검색)
- 22) International Council on Clean Transportation, Global and U.S. electric vehicle trends, 2019
- 23) International Energy Agency, 2020 Global EV outlook, 2020
- 24) UNEP, 2020 Global status report for buildings and construction: Towards a zero-emission, efficient and resilient buildings and construction sector, 2020
- 25) Government of the United Kingdom, Green Homes Grant, Retrieved from <https://www.gov.uk/guidance/apply-for-the-green-homes-grant-scheme> (2021.08.22. 검색)

- 26) Energiesprong, The Netherlands, Retrieved from <https://energiesprong.org/?country=the-netherlands> (2021.08.22. 검색)
- 27) Government of Canada, Government of Canada confirms ambitious new greenhouse gas emissions reduction target, 2021.7.21. Retrieved from <https://www.canada.ca/en/environment-climate-change/news/2021/07/government-of-canada-confirms-a-ambitious-new-greenhouse-gas-emissions-reduction-target.html> (2021.08.22. 검색)
- 28) 김종우·박지용, 저소득층 에너지 효율사업 개선방안에 대한 연구, 에너지경제연구원, 2020
- 29) 허경선, 온실가스감축인지 예산제도(탄소인지예산제도)의 도입과 적용 방안, 한국조세재정연구원, 2021
- 30) 관계부처합동, 산업구조 변화에 대응한 공정한 노동전환 지원 방안, 2021

본 내용은 녹색기술센터(GTC)의 주요사업 [R2110101] 「탈탄소 사회 전환을 위한 정책 아젠다 발굴 및 산업 육성 전략 연구」 일환으로 분석 중인 연구 내용의 일부를 요약·정리한 것입니다.

ISSUE
022020년 국가연구개발과제의 기후기술
분류체계 기반 딥러닝 분류모델 적용 연구

주경원, 한수현 / 기술총괄부 | kwjoo@gtck.re.kr, sue@gtck.re.kr

하이라이트

- 딥러닝 기법을 활용한 시 문서분류 알고리즘의 발전에 따라 국가 기후기술 연구개발과제의 기후기술 분류체계 기반 분류 알고리즘 개발 및 적용성 검토
- 딥러닝 모델은 텍스트 전처리, 워드 임베딩 벡터, 합성곱신경망(CNN), 장단기메모리(LSTM) 알고리즘을 활용하여 구축하였으며, 국가 연구개발과제의 과제명을 통해 기후기술 분류체계의 대분류, 중분류, 소분류를 예측
- 학습 및 테스트 데이터로 기존 기후기술 분류체계에 맞춰 분류된 4만여건의 국가 연구개발과제 정보를 활용하였으며, 대·중·소분류 별로 각각 약 90%, 78%, 70%의 분류 정확도를 나타냄
- 향후 데이터 증강 및 사전학습 활용도 확장을 통해 딥러닝 모델을 추가적으로 개선하고 국가 연구개발의 기후기술 분류검토 전문가 자문 시 보조자료로 활용

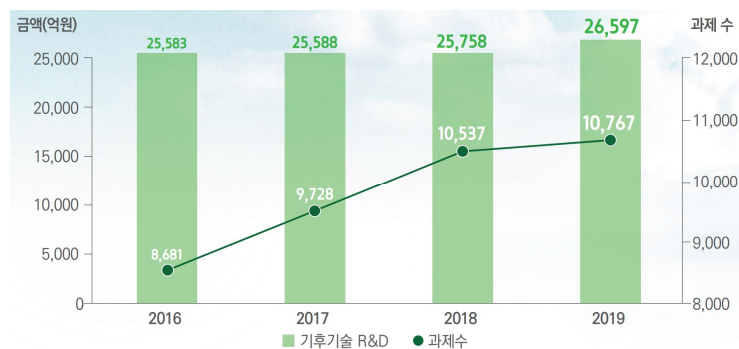
키워드

- 딥러닝, 다중분류, 합성곱신경망, 장단기메모리, 기후기술 분류체계

배경

기후기술 분류체계에 따른 국가 R&D 분류과정

- 기후기술 분류체계는 대분류-중분류-소분류 세 개의 계층적(hierarchical) 구조로 이루어져 있으며, 각각 3개, 14개, 45개의 기술분야로 정의됨
- 국가 연구개발과제 중 기후기술 R&D는 2016년도 8,681건에서 2019년 10,767건으로 매년 꾸준히 성장하고 있으며, 투자액은 2019년 기준 약 2.7조원으로 국가 전체 R&D 중 12.9%를 차지¹⁾

[그림 1] 국가 기후기술연구개발 투자금액 및 과제 수¹⁾

※ 출처: 녹색기술센터 (2020), 2019 기후기술 국가연구개발사업 조사·분석 보고서

- 위와 같이, 기후기술 분류체계는 국가 연구개발의 기후기술 분야 별 연구개발 투자금액의 조사분석, 성과분석 등의 통계자료를 산출하는데 활용
- 녹색기술센터에서는 매년 한국과학기술기획평가원(KISTEP) 및 국가과학기술지식정보시스템(NTIS)을 통해 국가연구개발 정보를 이관받아 일차적으로 분류하고, 30~40명의 각 분야별 기술전문가의 자문을 통해 기술분류를 확정하고 있음
- 하지만 현 기술분류과정에서는 분야별 전문가의 경험적인 판단에 의존하고 있어 아래와 같은 문제점이 제기될 수 있음
 - ‘녹색기술센터 검토 - 전문가 자문 - 자문결과의 취합 및 통계량 검토’의 과정은 적지 않은 시간과 자문료가 소요되며, 보통 두 번의 라운드를 거치며 최초 연구개발정보 취득 후 분류결과를 확정하기까지 최소 3달 가량의 기간이 소요됨
 - 동일한 연구과제에 대해 각 전문가의 분류의견이 다를 경우 의사결정이 어려우며, 같은 전문가로부터도 동일한 연구과제에 대해 매년 다른 분류결과로 검토받는 경우가 다수 존재함
 - 자문을 수행하는 전문가가 자신의 전문분야에 투자되는 연구비가 과도하게 산출되어 공표될 것을 우려하며 의도적으로 기후기술에서 배제할 가능성
- 본 연구의 목적은 딥러닝 기반 AI 문서분류 알고리즘의 적용성을 검토하여 시의성 있는 국가 기후기술 연구개발의 통계자료 산출에 기여하고, 견고(robust)한 분류모델을 구현하기 위한 프로토타입을 개발하는 것임

AI 기반의 문서분류 연구동향

- 인터넷이 발달함에 따라 데이터의 양이 급격히 증가하고 있으며, 이에 따라 축적되는 문서들을 분석하기 위해 AI 기반의 다양한 텍스트 분석 방법들이 개발되고 있음. 본 연구에서는 문서 분류(classification)에 관한 연구동향을 분석
- 기계학습은 크게 지도학습(Supervised Learning), 비지도학습(Unsupervised Learning), 강화학습(Reinforcement Learning) 등으로 구분
 - 지도학습은 데이터별로 라벨*이 준비되어 있는 자료를 모형에 제공하여 훈련시키는 방법으로 크게 분류(classification) 및 회귀(regression)로 나뉘며, 비지도학습은 라벨링이 되어있지 않은 데이터를 통해 데이터가 가지고 있는 특성을 추출하거나 구조를 파악
 - *각 데이터의 정답 (손글씨 이미지의 숫자, 사진 상 사물의 이름 및 위치, 뉴스의 카테고리 등)
 - 강화학습은 어떤 환경(Environment)이 주어졌을 경우 에이전트(Agent)가 현재의 상태(State)를 인식하고, 행동(Action)을 통해 보상(Reward)을 최대화하는 방향으로 학습하는 알고리즘으로 대표적인 예로 딥마인드(구글)의 알파고, 자동차의 자율주행 기술 등이 존재

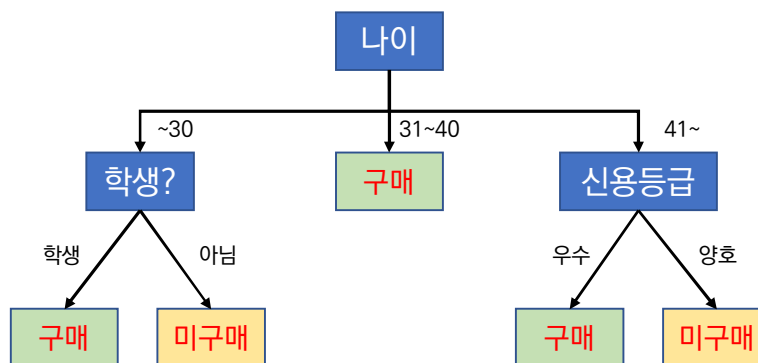
- 감성분석(sentiment analysis)은 대표적인 문서분류 모델이며 상품이나 영화 등의 리뷰 문장과 그와 함께 매겨진 점수(별점)를 훈련자료로 사용
 - 고객 피드백, 콜센터 메시지 등과 같은 데이터를 분석하며, 외부적으로는 기업과 관련된 뉴스나 SNS 홍보물 등에 달린 댓글의 긍정, 부정을 판단하는데에 사용
- 스팸메일분류의 자동화는 가장 오랫동안 사용되어 온 SI 모델 중 하나이며, 과거에는 특정 키워드, 발신자 정보 등을 통해 필터링하였으나, 현재는 다양한 기계학습 및 방법론이 적용되어 높은 분류 성능을 나타냄
- 일반적으로 실생활에서 쓰이는 문서의 경우 사전학습(pre-trained)된 모델이 지속적으로 배포되어 왔으며, 최근에는 기술문헌과 같이 구조가 어렵고, 희소한 단어들이 많이 포함된 문장들에 대한 분석도 다양하게 이루어짐²⁾³⁾

문서 분류를 위한 기계학습 방법론

의사결정트리 (Decision Tree)

- 의사결정트리를 이용한 기계학습 방법은 어떤 항목에 대한 관측값과 목표값을 연결시켜주는 예측 모델링 방법 중 하나이며, 분류와 회귀에서 모두 사용할 수 있기 때문에 CART(Classification And Regression Tree)라고도 함 (ex. 스무고개)
- 플로우 차트와 같이 분기점을 지날 때마다 특정 방향을 따라가는 구조를 가지고 있으며 각 잎(leaf) 노드는 클래스 레이블(결과)을 나타냄

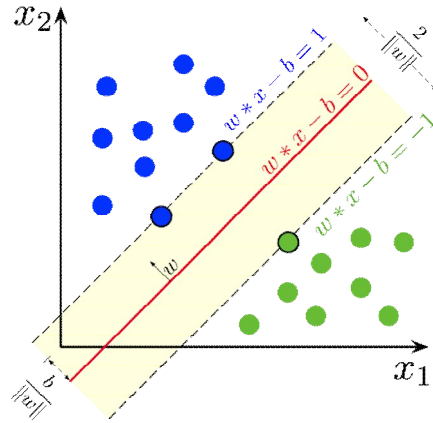
[그림 2] 고객의 정보를 바탕으로 구매여부를 예측하는 의사결정트리 예시



SVM (Support Vector Machine)

- SVM은 패턴 인식과 자료 분석을 위해 제안된 지도학습 모델이며, 두 개의 범주를 갖는 데이터의 집합에서 이진 선형 분류모델(경계)을 생성하여, 아래의 그림과 같이 가장 큰 폭을 갖는 경계선을 찾는 알고리즘임⁴⁾

[그림 3] SVM 경계 결정 알고리즘 모식도



※ 출처: https://en.wikipedia.org/wiki/Support-vector_machine (2021.09.06. 접근)

- 최근 신경망 기반의 딥러닝 방법이 개발되기 전에는 문서 분류작업에서 가장 우수한 성능을 나타내는 방법이었음⁵⁾

Naive Bayes

- Naive Bayes 자료들 사이의 독립(Independent)을 가정하는 베이즈 정리(Bayes' Theorem)를 적용한 확률이론에 기반한 분류 알고리즘으로 1950년대 이후 광범위하게 연구되어 적용되고 있으며 문서 분류에서는 스팸메일 필터링 알고리즘으로 주로 사용됨
- 예시로, 스팸메일(Spam)인지 정상 메일(Ham)인지 분류하는 Naive Bayes 추정식은 아래와 같으며 해당 계산결과에 부호에 따라 결과를 추정함

$$\log \left\{ \frac{P(Ham)}{P(Spam)} \right\} + \sum_{i=1}^m \log \left\{ \frac{P(w_i | Ham)}{P(w_i | Spam)} \right\}$$

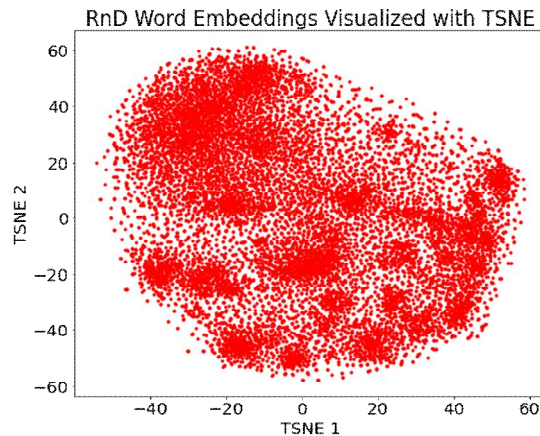
- 여기서 $P(Ham)$, $P(Spam)$ 은 메일들이 정상 또는 스팸메일일 확률이며, m 은 특정 메일 내 단어의 수, $P(w_i | Ham)$, $P(w_i | Spam)$ 은 단어 w_i 가 정상 또는 스팸메일에서 등장하는 조건부확률

워드 임베딩 (Word Embedding)

- 워드 임베딩은 컴퓨터가 자연어를 사람과 같이 이해할 수 있도록 단어를 벡터로 표현하는 방법으로 밀집 표현(dense representation)하는 과정을 통해 계산되며, 임베딩 벡터(embedding vector)라고도 부름
- 임베딩 벡터는 원-핫(one-hot) 인코딩 방법에 비해 낮은 메모리를 사용하며, 벡터의 값을 훈련 데이터로부터 학습하기 때문에 자연어 분석을 위한 딥러닝 모델링에서 첫 번째 레이어로 주로 사용
- 2016~2019년의 국가연구개발과제명의 워드 임베딩 벡터를 2차원으로 표현하면 아래의 그림과 같이 나타낼 수 있음

- 그림의 점은 국가 연구개발과제 단어사전의 각 단어들을 의미하며, 64차원으로 학습된 각 단어의 벡터공간을 t-SNE 방법을 통해 2차원으로 축소하여 X(TSNE1), Y(TSNE2)축으로 도시

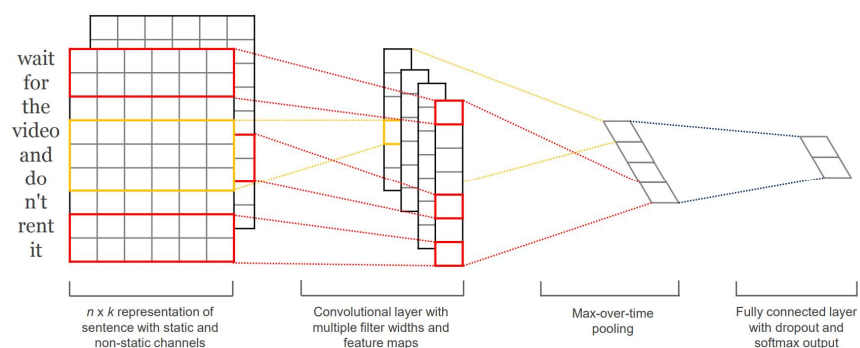
[그림 4] 국가 연구개발과제 워드 임베딩 벡터



합성곱신경망 (CNN, Convolution Neural Network)

- CNN은 일반적으로 이미지 분류에 필수적인 딥러닝 방법으로 이미지의 영역 특성을 추출하는 데에 탁월하여 영상 및 동영상 인식, 추천 시스템, 영상 분류, 의료 영상 분석 등에 사용
- 텍스트를 워드 임베딩으로 표현할 시 이미지와 유사하게 한 축은 문장, 한 축은 임베딩 벡터로 표현할 수 있으므로 이미지 분류와 유사하게 적용할 수 있으며, 의미있는 특성들을 추출할 수 있음⁵⁾

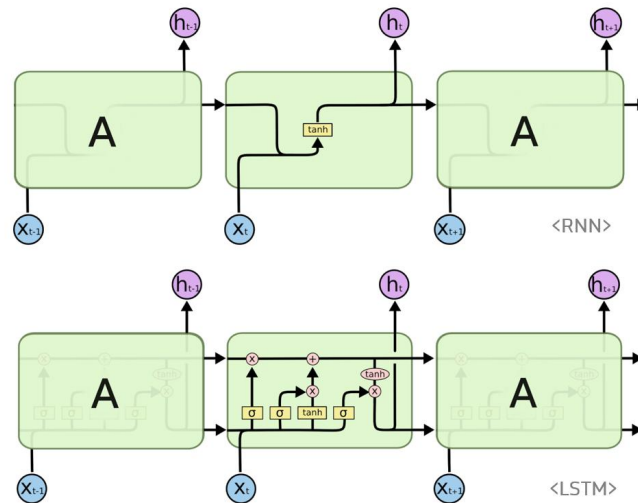
[그림 5] 문장 시퀀스에 대한 합성곱신경망 적용 예시⁵⁾



장단기메모리 (LSTM, Long Short-Term Memory)

- LSTM은 순환신경망(RNN)의 일종으로 입력 시퀀스가 길어질 경우 그라디언트 소실 (vanishing gradient)를 해결하기 위해 고안된 방법으로 추가적인 은닉 층을 구성하여 의미있는 그라디언트가 유지될 수 있도록 개선된 방법임

[그림 6] 바닐라 RNN과 LSTM 구조 모식도



※ 출처: <https://towardsdatascience.com/long-short-term-memory-networks-are-dying-whats-replacing-it-5ff3a99399fe> (2021.09.06. 접근)

- 문서분류를 위한 순환신경망 구조는 여러개의 입력으로부터 하나의 출력을 가지는 Many-to-One 입출력 구조를 가지게 되며 마지막 토큰이 들어왔을 때 이를 완전연결층(fully-connected layer)로 넘겨 소프트맥스를 활용하여 어떤 범주에 속할 지 예측하는 순서로 계산

모델 적용 및 결과

모델 구조 및 데이터

- 국가연구개발의 기후기술 분류체계 기반 분류를 위한 기계학습 모형은 기존에 전문가 그룹을 통해 분류(라벨링)된 데이터를 활용하여 학습하는 모형으로 지도학습 방법에 속하며, 대·중·소분류에 따라 각각 3, 14, 45개의 분야를 예측하는 모델로 다중분류(Multi-class Classification)모형에 속함
- 2016~2019년 국가연구개발과제 중 기후기술로 분류된 39,713개의 데이터를 대상으로 적용
- 일반적으로 텍스트 전처리(preprocessing)시에는 한글 및 영문만을 대상으로 추출 하지만, 연구개발과제 정보는 전문적인 기술문헌으로 단위(MW, Hz 등) 및 각 기술분야에서 통용되는 약어 등이 포함되어 있음
 - 따라서, 형태소 분석 후 제거하는 특수문자 등의 불용어(stopword)를 최소화하고, 출현 빈도가 낮은 한글, 영문의 고유명사도 워드 벡터에 포함함
- 훈련데이터와 테스트(검증)데이터는 8:2로 나누어 수행하였으며, 대·중·소분류 별로 데이터가 편중되지 않도록 계층적으로(stratified) 분할
 - ※ 기계학습 모델 훈련 시 일반적으로 훈련/테스트 비율은 8:2로 분할하며 데이터의 수가 100,000개를 넘어가는 경우에는 훈련 데이터 비율을 상향함 (본 연구에서는 39,713개가 사용됨)

- 연구개발과제와 같은 기술문헌의 경우 고유한 단어가 많고 희소한 단어를 제외하기 어려워 입력계층의 차원이 큰 점을 고려하여 모델 구조는 임베딩 벡터-CNN(-LSTM)의 구조를 선택하였으며, 최적화 방법은 Adagrad의 문제점을 보완한 RMSProp을 사용함
- 각 모델별 하이퍼 파라미터를 설정하고 그에 따른 테스트 데이터를 기준으로 대분류, 중분류, 소분류별 정확도를 계산

분석결과

- 임베딩 벡터와 CNN을 활용한 모델 1~3번의 결과는 아래의 표와 같음
 - 1~3번 모델 별로 CNN필터의 수는 각각 16, 32, 64개를 적용하였으며, 3번 모델에는 2개의 레이어를 적용하고, 과적합을 방지하기 위해 50%의 dropout을 설정함

[표 1] 워드 임베딩 벡터와 CNN을 활용한 딥러닝 모델 구조 및 분류 결과 정확도

		Model		
		#1	#2	#3
Embedding Vector		64	64	64
CNN	Layers	1	1	2
	Filters	16	32	64
	Kernel Size	5	5	5
	Activation	relu	relu	relu
	MaxPooling	4	4	4
	Dropout	0.0	0.0	0.5
Batch		64	32	64
ReduceLROnPlateau		사용	사용	사용
Early stopping		5	5	5
정확도(%)	대분류	90.1	89.9	87.5
	중분류	77.4	77.6	78.1
	소분류	66.4	67.1	69.3

- 임베딩 벡터와 CNN을 활용한 모델 중 대분류 기준으로는 1번 모델(90.1%)이 가장 정확하며, 중분류와 소분류 기준으로는 3번 모델(78.1%, 69.3%)이 가장 높은 정확도를 나타냄
- 배치의 크기나 CNN의 레이어와 필터의 수는 정확도에 큰 영향을 미치지 않음
- 임베딩 벡터, CNN, LSTM을 활용한 모델 4~6번의 결과는 아래의 표와 같음
 - CNN 구조에서 4번 모델에는 3개의 레이어와 20%의 dropout을 적용하였으며, 5번 모델에는 40%의 dropout을 적용함
 - LSTM 구조에서 4번모델에는 16개의 레이어와 20%의 dropout을 적용하였으며 5, 6번 모델에는 32개의 레이어를 적용함

- 모델 컴파일 시 6번 모델에는 학습률을 조정하는 ReduceLROnPlateau를 적용하였으며 4, 5번 모델에는 학습률을 고정하여 학습함

[표 2] 워드 임베딩 벡터, CNN, LSTM을 활용한 딥러닝 모델 구조 및 분류 결과 정확도

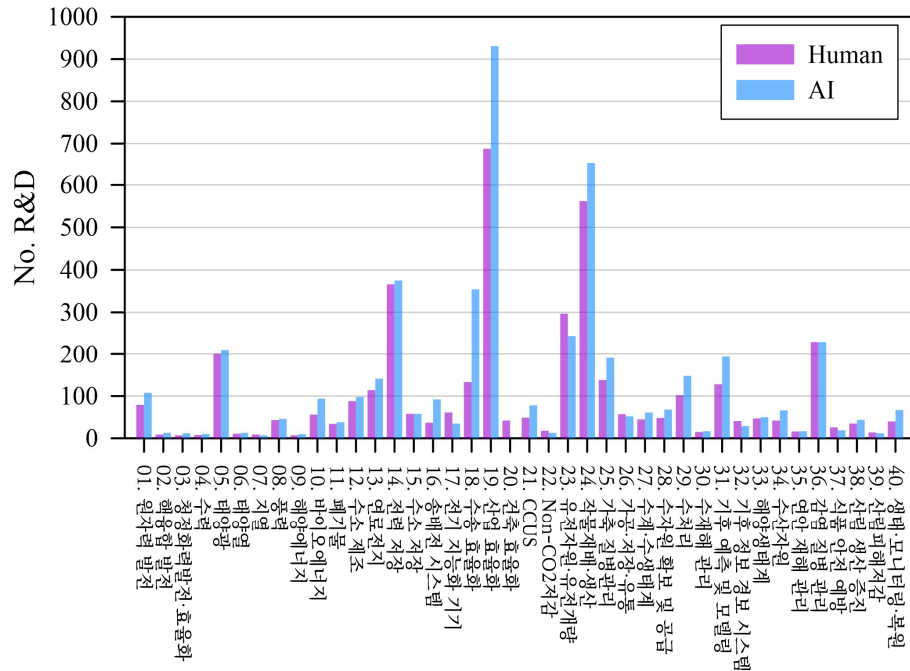
		Model		
		#4	#5	#6
Embedding Vector		64	64	32
CNN	Layers	3	1	1
	Filter	32	32	32
	Kernel Size	5	5	5
	Activation	relu	relu	relu
	MaxPooling	4	4	4
	Dropout	0.2	0.4	0.0
LSTM	Layers	16	32	32
	Dropout	0.2	0.0	0.0
Batch		64	64	64
ReduceLROnPlateau		미사용	미사용	사용
Early stopping		5	3	5
정확도(%)	대분류	89.0	89.2	90.0
	중분류	77.9	77.3	78.3
	소분류	70.9	69.7	70.4

- 대분류, 중분류 정확도 기준으로는 모델 6번이 가장 우수(90.0%, 78.3%)하며, 소분류 기준으로는 모델 4번이 가장 우수(70.9%)하나 모델 간 큰 차이를 나타내지 않음
- 본 모델은 기존에 기후기술로 분류된 연구개발과제에 한해 적용한 결과이며, 전체 국가 연구개발과제수인 연간 약 7만여개에 해당하는 데이터셋에 적용할 경우에는 소분류 기준 93.8%의 정확도를 나타냄
- 불균형(imbalance) 데이터의 성능 측정에 사용되는 f1-score(macro)의 경우 전 모델에서 0.70 내외의 수치를 나타내었으며, 2020년 신규 연구개발과제에 대한 AI 예측결과를 기술분류 전문가 자문 요청 시 보조자료로 제시함
 - 보조자료로 제공한 AI 모델의 예측결과는 기술분류 자문 시 긍정적인 피드백을 받았으며, 전문가가 분류한 약 70%의 연구개발과제는 전문가 검토결과와 동일하게 분류됨

2020년 신규과제 분석결과

- 본 모델의 적용성을 검토하기 위해 정확도 결과를 바탕으로 4번 모델과 6번 모델의 결과를 앙상블하여 훈련/테스트 과정에서 사용되지 않은 2020년 4,898개의 신규과제를 적용하여 소분류를 예측하였으며, 적응 및 감축 분야의 전문가 자문결과와 비교함
 - ※ 그림 7의 AI 예측결과는 적용성 검토를 위해 개발된 모델의 예측결과로, 향후 사전학습 모델 활용과 앙상블 추론을 통한 최종 예측결과는 변경될 수 있음

[그림 기] 2020년 국가연구개발과제(신규)의 전문가그룹 및 딥러닝 모델 예측결과



- 전반적으로 전문가 분류결과와 유사한 예측결과를 나타내는 것을 확인하였으며, 유의하여 살펴봐야 할 점으로는 다음과 같은 특징이 있음
 - 전문가 그룹에 비해 AI는 더 많은 연구과제를 기후기술 분류체계 기반의 연구과제로 분류함
 - 19.산업효율화는 전문가 집단이 예측한 688개에 비해 높은 931개로 분류하였으며, 기존에 산업효율화로 분류된 연구개발과제의 단어 범위가 포괄적인 점이 영향을 미친 것으로 보임
 - 16.송배전시스템 또한 전문가 집단이 분류한 37개에 비해 높은 92개로 분류하였으며, 이는 송배전시스템 분야가 다루는 단어의 범위가 넓어 훈련과정에 영향을 미친 것으로 보임
 - 20.건축효율화의 경우 전문가 집단은 40개의 과제를 분류했지만 딥러닝 모델에서는 대부분 19.산업효율화로 분류하였는데, 산업효율화로 분류된 연구개발과제들이 건축분야의 단어 특성을 포함하고 있어 이와 같은 결과가 나타남

결론

- 본 연구에서는 AI, 딥러닝 분야에서 활발히 연구 및 활용되고 있는 문서분류 딥러닝 알고리즘을 활용하여 기후기술 국가연구개발과제의 한국어 말뭉치에 적용하였으며, 기후기술 분류체계 기반으로 정확도를 검토함
- 훈련 및 테스트 자료로 2016~2019년도의 국가 연구개발과제 중 기후기술로 분류된 39,713개의 연구개발과제를 활용하였으며, 이 중 80%를 훈련자료로 사용함
- 딥러닝 모델은 워드임베딩벡터, CNN, LSTM의 구조의 사용하였으며, 대분류, 중분류, 소분류 별 정확도는 각각 약 90%, 78%, 70%의 정확도를 나타냄

- 2020년 국가연구개발과제 중 신규과제에 대해 본 연구에서 학습시킨 모델 중 우수한 2개의 결과를 앙상블하여 예측한 결과와 전문가 집단의 분류결과와 비교한 결과 75.3%의 정확도를 나타내며, 아래와 같은 사항을 개선하여 추가적으로 모델의 성능을 제고할 필요가 있음
 - 소분류별 데이터 불균형을 고려하여 macro f1-score를 목적함수로 추가할 필요가 있으며, 각 소분류별 f1-score를 고려하여 여러 모델의 경험적 앙상블 추론 구현
 - 개선 방안으로 ①부족한 소분류에 대한 데이터 증강(augmentation), ②기술문헌을 해석하기 위한 사용자 사전 추가, ③텍스트 전처리 방법의 고도화, ④워드 임베딩 벡터의 사전학습 모델활용(koBert, koelectra 등)이 있으며, 향후 추가적인 연구를 통해 모델을 개선시킬 수 있을 것으로 기대함
- 본 모델을 통해 예측한 국가 연구개발과제의 기후기술 분야를 바탕으로 ①전문가 자문 과정에 소요되는 시간을 단축하고, ②예측 확률(softmax)이 높은 과제는 내부적으로 검토하여 전문가 자문에 소요되는 예산을 절감하고, ③전문가의 의견이 상충될 시 보조 지표로 사용할 수 있도록 활용 예정

참고문헌

- 1) 녹색기술센터 (2020), 2019 기후기술 국가연구개발사업 조사·분석 보고서
- 2) 황상흠, 김도현 (2020), 한국어 기술문서 분석을 위한 BERT 기반의 분류모델, 한국전자거래학회, Vol.25, No.1
- 3) KISTEP (2019), 기계학습 기반 바이오의료분야 과학기술정보데이터 분석활용 모형 고도화
- 4) Thorsten Joachims (1998), Text categorization with Support Vector Machines: Learning with many relevant features
- 5) Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2015), Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. Expert Systems with Applications, 42(1), 306-324
- 6) Yoon Kim (2014), Convolutional Neural Networks for Sentence Classification (2014), arXiv:1408.5882 [cs.CL]
- 7) <https://towardsdatascience.com/long-short-term-memory-networks-are-dying-whats-replacing-it-5ff3a99399fe> (2021.09.06. 접근)
- 8) https://en.wikipedia.org/wiki/Support-vector_machine (2021.09.06. 접근)

본 내용은 녹색기술센터(GTC)의 주요사업(한수원, 안세진, 우아미, 주경원, 「기후기술 분류체계 기반 통계생산 및 국제확산」)으로 수행한 내용을 요약·정리한 것입니다.

2021년 5호 2021 OCTOBER Vol.2 No.5

GTC BRIEF

발 행 인 정병기

발 행 일 2021년 10월 19일

발 행 처 녹색기술센터

주 소 04554 서울특별시 중구 퇴계로173
남산스퀘어 빌딩 17층

전 화 02.3393.3961

팩 스 02.3393.3919~20

홈페이지 <http://www.gtck.re.kr>

I S S N 2733-9696(온라인)
2733-9572(인쇄본)

디 자 인 리드릭 02.2269.1919



04554 서울특별시 중구 퇴계로173
남산스퀘어 빌딩 17층
Tel. 02.3393.3900
Fax. 02.3393.3919~20
www.gtck.re.kr

* 본 GTC BRIEF의 내용은 필자의 개인적 견해이며, 센터의 공식적인 의견이 아님을 알려드립니다.