

과제계획요청서

PROJECT PROPOSAL REQUEST (PPR)

과학기술계학을 이용한 극단 기후·기상변화 예측 및 재난위험 맵핑 기술
(CLIMECAST: Climate Event Prediction Using AI and Science Technology)

PPR No. ASTRA02_2402PPR1

과제계획요청서(PPR) 공고일: 2024. 4. 4.

연구개발계획서 접수 마감일: 2024. 5. 9.

공모유형: 분야공모

보안과제여부: 일반



과학기술정보통신부



한국연구재단

한계도전전략센터

Advanced Science & Technology Research Agency (ASTRA)

1. 과제 개요

한국연구재단 한계도전전략센터에서는 극심한 변동성을 갖는 기후·기상 예측과 물리 지식 기반 인공지능 관련 연구개발 수준이 글로벌 선단에 이를 수 있는, 과학기계학습을 이용한 극단 기후·기상변화 예측 및 재난위험 맵핑 기술 (CLIMECAST¹⁾) 프로젝트를 추진합니다.

이 프로젝트는 “지구계 탄소 순환 불균형에 따른 극단적 기후·기상 변화 예측 모델을 개발하기 위해서는 유효 데이터 부족의 한계를 극복하는 새로운 기계학습(Machine Learning, ML) 기술이 필요하다”라는 문제 인식에서 출발합니다. 이 문제를 해결하기 위해 설명 가능한 인공지능 기술과 예측 데이터를 물리 지식의 확장에 활용하는 기술 개발을 시도합니다. 이런 도전적 시도와 기후 과학 기술과의 연계를 통해 지구 온난화 위기에 우리 사회가 선제적으로 대응할 수 있기를 기대합니다.

이 시도가 성공한다면 기후 변화 적응(Adaptation) 문제 해결 가능성과 의미를 보여주는 세계 선단의 기술이 될 것이며, 그 과정에서 얻는 새로운 지식을 바탕으로 유효 데이터가 부족해도 사회에 가치 있는 정보를 제공하는 과학기계 학습(Scientific Machine Learning, SciML) 발전에도 이바지할 것입니다.

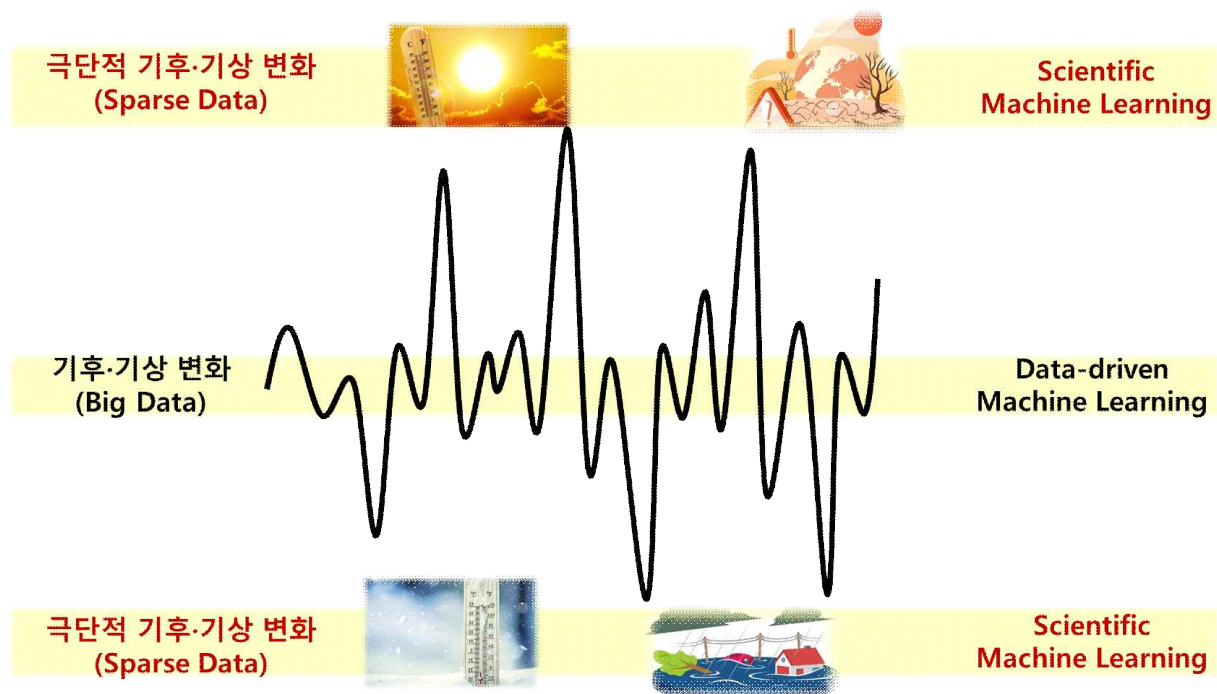


그림 1. 프로젝트(CLIMECAST)의 예측 모델 개발에 대한 설명도

1) CLIMECAST: Climate Event Prediction Using AI and Science Technology

2. 추진 배경

지구의 해양-육상-대기 사이에서 나타나는 탄소의 생화학적 순환과정에서, 대기 중의 탄소는 대부분 이산화탄소로 존재합니다. 그런데, 인구 증가, 문명 발달, 경제 규모 증가에 따라 막대한 양의 탄소자원(석탄, 석유 등)을 에너지원으로 사용하면서, 대기권의 이산화탄소 농도가 비정상적으로 증가하게 되었습니다. 이로 인해 자연적 탄소 순환 불균형이 발생하면서, 지구 온난화라는 심각한 문제를 초래하게 되었고, 과거에는 경험하지 못했던 극단적 기후·기상 변화(폭우, 폭염, 가뭄, 한파)가 자주 나타나면서 재앙 수준의 광범위한 재해를 겪고 있습니다.

지구 온난화와 기후·기상 변화의 인과 관계는 대기 물리학 및 기후 과학의 원리에 기반한 연구 결과²⁾³⁾⁴⁾와 각종 보고서⁵⁾를 통해 이해할 수 있습니다. 대기 중 이산화탄소의 증가로 인해 더 많은 에너지가 지구 시스템에 갇히게 되고, 이런 과잉 에너지는 기온 상승, 습도 증가, 기상이변에 사용될 수 있습니다. 그리고, 대기의 온도 상승은 제트 기류 및 해류 순환 패턴을 방해하여 날씨 변화가 심해지면서 폭염, 가뭄, 폭우, 폭풍과 같은 극한 현상의 빈도와 심각성을 증가시킵니다. 이상에서 살펴본 대기 중 에너지 증가와 대기 순환의 변화 이외에 바다, 호수, 지표면 수분 증발량 증가도 중요한 인과 관계 중 하나입니다. 대기 중 수분이 증가하면 강우와 폭설의 발생 조건이 충족되면서 더 강렬하고 빈번한 홍수 발생의 원인이 됩니다. 무엇보다 지구 온난화가 극단적 기상변화를 일으키는 다양한 피드백 메커니즘을 더욱 자극한다는 점이 중요합니다. 예를 들어 북극 해빙이 녹으면 지구 표면의 반사율이 감소하면서 태양 복사열이 더 많이 흡수되고 이로 인해 온난화가 더욱 가속화되는 것처럼 극한 기상 현상의 빈도와 강도에 영향을 미칠 수 있습니다.

태풍, 가뭄, 산불, 홍수와 같은 사건은 점점 흔해지고 있으며 심각한 사회·경제 문제를 일으키고 있습니다. UNDRR 2022년 보고서⁶⁾에 따르면 2015년 약 400회 수준의 재해는 2030년 560회에 이르면서 약 40% 증가가 예상되고, 예년에는 볼 수 없었던 극단적인 기상 온도가 나타나는 경우도 2001년

2) Proceedings of the National Academy of Sciences 109, E2415 (2012).

3) Science 370, 1095 (2020).

4) Nature Climate Change 2, 491 (2012)

5) <https://www.ipcc.ch/report/ar6/wg2/>

6) GAR2022 Report, UNDRR(United Nations Office for Disaster Risk Reduction).

과 2030년 사이에 2배 증가할 것으로 예상합니다(그림 2, 3). 이러한 역사적 데이터는 지구의 기후가 전례 없는 변화를 겪고 있고, 향후 심각한 재난 문제가 계속 증가할 것이라는 주장을 뒷받침합니다.

지구 온난화가 일으키는 여러 문제에 대응하기 위해 전문가들은 온실가스를 감축하는 ‘완화(Mitigation)’ 방법과 기후 변화 ‘적응(Adaptation)’ 방법을 제안합니다. 온실가스 감축은 탄소 배출 에너지를 태양광 발전과 같은 무탄소 에너지로 대체하거나, 배출된 이산화탄소를 포집 저장하는 등 보다 적극적인 대응 방법이고, 기후 변화 적응은 현재 나타나고 있거나 미래에 나타날 기후 변화의 파급효과와 피해를 인위적 시스템의 개발과 활용을 통해 대비하는 것을 의미합니다.

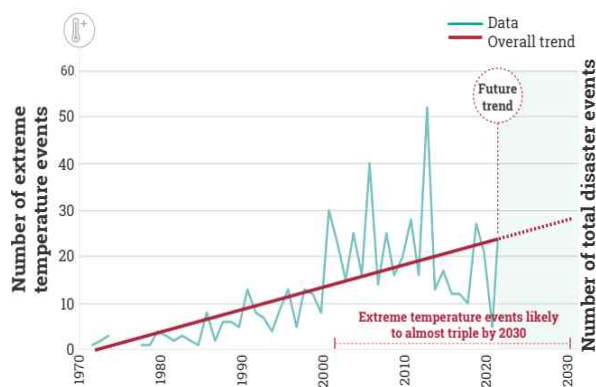


그림 2. 극단 기상 온도 발생 횟수 및 예측

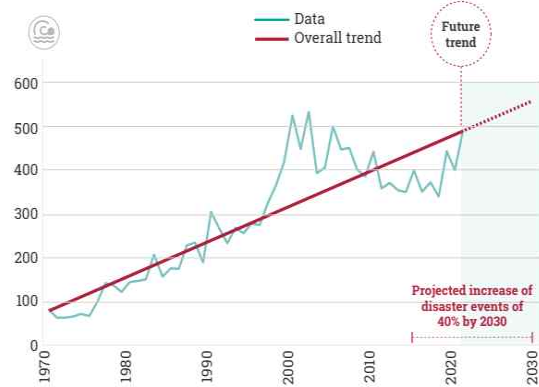


그림 3. 재난 발생 횟수 및 예측

탄소 에너지를 기반으로 지금까지 성장한 세계 경제의 향후 추세를 고려하면 지구 온난화 문제를 적응 기술 개발 관점에서 다루는 선제적 조치가 시급하며, 경험하지 못한 기상이변에 따른 재난 수준의 문제에 대응하기 위해서는 여러 기술이 융합된 접근 방식이 필요합니다. 왜냐하면, 기상이변이 이를 소화할 수 없는 인공환경 및 생활 편리성에 맞춰진 복잡한 도심 구조와 연결되면서, 속수무책의 상상할 수 없는 인명과 경제 손실이 나타날 수 있기 때문입니다.⁷⁾

기후·기상의 극단적 변동성은 인류의 지속적 성장에 중대한 영향을 미칩니다. 미국기상학회(American Meteorological Society, AMS)는 기후 변동성을 평균 상태를 중심으로 대기-해양 시스템의 시간적 변화, 또는 계절 이상의 긴 시간 범위에서 정의합니다.⁸⁾ 열대성 저기압 및 중규모 대류 시스템의 변

7) https://link.springer.com/chapter/10.1007/978-3-030-72196-1_17

화로 인해 변동성이 나타나는데, 연구개발 관점에서 해결 솔루션을 찾기 위해서는 해당 원인과 현상의 이해가 가장 중요합니다. 그런데, 1) 제한된 관측 기록, 2) 지구 시스템 모델의 세밀하지 못한 격자 간격(Grid Spacing), 3) 극한 현상을 유발하는 미세 규모의 불확실성이 원인과 현상의 이해를 어렵게 합니다.⁹⁾

지금까지 알려진 대부분의 기상 예측은 대기의 습도 및 기온 그리고 바람의 변화를 설명하는 유체역학 방정식을 풀어냄으로써, 현재와 미래의 대기 상태를 설명 및 예측하는 물리학적 방법론에 기반한 수치 예보 방식에 의해 이루어져 왔습니다. 대표적으로 영국기상청의 수치 예보 통합모델(United Model, UM)은 하나의 수치 방법론으로 모든 영역의 응용문제를 해결하고 있으며, 이러한 통합적 접근 방식은 미국 기상청과 유럽중기예보센터(European Centre for Medium-Range Weather Forecasts, ECMWF)를 비롯해서 여러 기술 선진국에서도 확인 가능합니다.

많은 연구자들은 기후 변화 연구를 위해 지구 시스템 모델(Earth System Model, ESM)을 주로 사용하고 있습니다. 수치 예보의 초기 발전 단계에서는 역학 코어나 물리 과정과 관련된 수치 모델을 응용한 예보기술이 발전하였고, 방대한 수준의 많은 모델을 이용한 사례도 꾸준히 알려져 왔습니다. 위험 기상의 대표 격인 집중호우, 태풍, 대설, 국지 강풍을 주로 다루는 연구가 많이 진행되고 있는데, 범용 중규모 모델이 세계적으로 보급되면서 이를 이용한 사례 연구도 많이 늘어나고 있습니다.¹⁰⁾ 그런데 이러한 방법은 실제 대기 상태에 대한 방대한 고해상도 관측 데이터와 복잡한 수학·물리학 방정식을 다뤄야 하는 만큼 엄청난 양의 컴퓨팅 자원을 필요로 합니다. 해당 방식의 계산 특성상 주로 대기와 같은 거대 움직임만을 예측할 수 있고, 구름과 같은 미시적인 운동까지 수식으로 해석하고 표현하는 것은 어렵다는 단점이 있습니다. 이를 극복하려는 다양한 시도가 있었고, 그중 하나는 위에서 언급한 물리 방정식을 인위적으로 단순화시키는 매개 변수화 방법을 통해 모델의 복잡성과 계산량을 줄이는 것입니다.

인공지능 및 기계학습은 4차 산업혁명의 가장 핵심적인 요소로 자리 잡고 있으며, 거의 모든 학문 영역의 전문가들이 자기 분야에서 기계학습의 새로

8) https://glossary.ametsoc.org/wiki/Climate_variability

9) <https://www.fao.org/3/a1247e/a1247e02.pdf>

10) Mesoscale Meteorological Modeling, RA Pielke Sr, (2013).

운 가능성을 시험해보고 있습니다. 일 예로 일부 전문가들은 인공지능과 기계학습 기술이 종전의 수치 예보 방식의 한계를 극복할 수 있다고 기대하고 있습니다.¹¹⁾¹²⁾ 이는 기상 예측 모델과 같이 복잡한 물리 방정식을 쓰지 않는 대신, 과거 날씨 데이터를 통해 패턴을 인식하면 계산량을 획기적으로 줄일 수 있기 때문입니다. 이와 관련된 2개의 의미있는 결과가 네이처(Nature) 저널에 중국 빅테크 기업 화웨이 테크놀로지스(Huawei Technologies)¹³⁾와 UC 버클리 대학¹⁴⁾에 의해 처음 공개되었습니다.

화웨이 테크놀로지스(Huawei Technologies)가 보고한 연구는 기온과 풍속 등의 다양한 날씨 변수를 최대 일주일 전에 예측하는 ‘판구웨더(Pangu-Weather)’ 모델인데, 이 모델은 유럽중기기상센터(ECMWF) 예측 모델보다 최대 10,000배 빠르면서 동일한 정확도를 보였고, 열대성 저기압의 경로까지 추적할 수 있다고 알려줍니다. ‘판구웨더(Pangu-Weather)’ 모델 개발을 위해 1979년부터 2017년까지의 40년에 걸친 전 세계 기상 데이터를 학습시켰고, AI가 처리해야 할 데이터를 여러 계층으로 나눔으로서 방대한 데이터를 한 번에 처리했을 때 발생할 수 있는 오류의 가능성을 낮춘 것이 특징입니다. UC 버클리 대학의 마이클 조던 교수가 발표한 나우캐스트넷(NowcastNet)은 단기간 내 발생하는 강우량에 대한 예측을 전문적으로 수행하는 딥러닝 기반 AI 알고리즘입니다. 이 모델은 2048kmX2048km 규모의 지역 내 강수량을 최대 3시간 전에 높은 정확도로 예측할 수 있다고 알려줍니다. 전 지구 예측 모델을 인공지능/기계학습으로 구현하는 이러한 작업은 많은 인력과 컴퓨팅 자원이 필요한데, 실질적인 예측성 향상에 초점을 맞추고 있어서 예측 불확실성 원인은 알 수 없습니다.

인공지능 및 기계학습 전문가들은 예측 모델의 불확실성 문제를 해결하기 위해 기존 데이터 집합을 보강하여 규모와 다양성을 늘리는 데이터 증강(Data Augmentation), 한 데이터에서 학습된 모델을 레이블이 제한된 다른 관련 데이터에서도 잘 작동하도록 조정하는 전이 학습(Transfer Learning), 레이블링 과정의 효율성이 최대화될 수 있도록 정보 함축적 데이터를 지능적으로 선택하는 능동 학습(Active Learning), 데이터 사이의 연관 관계를 기억하는 뉴럴 네트워크(Neural Network) 등 다양한 기법을 제안하고 있습니다.

11) <https://doi.org/10.1371/journal.pone.0194889>

12) Nature 597, 672 (2021).

13) Nature 619, 533 (2023).

14) Nature 619, 526 (2023).

다. 콜로라도 주립대(Colorado State University) Elizabeth A. Barnes 연구팀은 뉴럴 네트워크 기법을 이용한 기후 변화 예측 결과를 발표했습니다.¹⁵⁾ 해당 인공지능 모형은 종전과 같은 프로세스에 기반한 기후 예측 장기 시뮬레이션만을 통해 학습되었기 때문에 실제 관측 정보는 전혀 모르는 상태로 학습이 됨에도 불구하고, 특정한 기후 발현이 예상되는 연도 수준의 정보를 제공합니다.

그러나 첨단 AI 기후·기상 예측 모델도 지구 온난화 적응 기술 관점에서 평균에서 크게 벗어나는 예측은 수행하지 못할 가능성이 큼니다. 왜냐하면, AI 시스템은 정확한 기후 예측을 위해 과거의 유효하면서도 많은 기상 데이터에 의존할 수밖에 없는데, 최근 세계 각지에서 지구 온난화 현상에 의해 발생하는 폭염, 폭우, 한파, 가뭄과 같은 이변은 너무 극단적이며 과거의 데이터가 없기 때문입니다. 또, 앞선 예에서 보였듯, 프로세스 기반 모형의 시뮬레이션을 통해 샘플 수를 증가시키려 해도 이마저도 현재 수준에서 평균에서 크게 벗어나는 현상에는 적용하기 어렵다는 한계가 있습니다. 글로벌 선단 연구그룹은 막대한 컴퓨팅 자원을 이용하여 이를 극복하려 시도합니다. 그러나, 이 또한 불확실성이 있는 예측 데이터 중 하나이고, 이렇게 큰 비용으로 얻은 데이터가 갖는 확장성도 매우 제한적입니다.

본 연구과제는 big data를 활용하는 data-driven machine learning 기법의 한계 극복을 통해 구현되는 기후·기상 변화 예측 모델 개발을 목표로 합니다. 그리고 차별화된 도전적 목표는

- a) 문제 정의에 기반한 데이터를 부족한 모집단에서 추려내고, 이후 유효 데이터 부족의 한계를 극복하는 차별화된 연구 접근법 개발,
- b) 설명 가능한 인공지능(eXplainable AI, XAI) 모델 개발,
- c) 그리고, 데이터 활용 확장성을 위한 예측 결과의 지식화 연구를 통해 달성 가능합니다.

15) <https://doi.org/10.1175/AIES-D-22-0001.1>

3. 과제 목표 및 범위

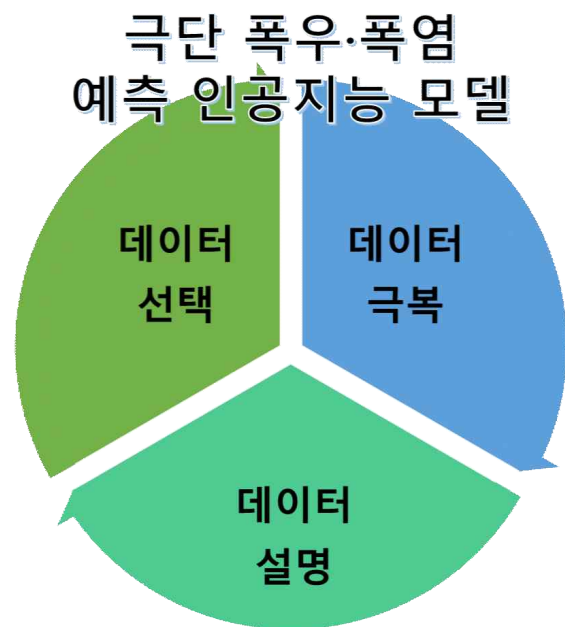
(최종 목표 및 연구내용)

본 과제계획요청서가 제시하는 중요한 연구목표는 “big data를 활용하는 data-driven machine learning 기법의 한계 극복을 통해 구현되는 한반도 지역의 극단적 폭우·폭염 예측 모델 개발”입니다.

목표 달성을 위해 제안하는 개념 및 검증 방법은 기계학습과 기후·기상 예측 분야에서 현재 글로벌 연구 커뮤니티 선단 연구그룹과는 다른 새로운 접근 방식, 가설, 검증이어야만 합니다.

연구 접근법의 다양성에 기반한 협력 연구 체계 구축을 위해, 다음과 같은 프로세스를 통한 연구개발 계획서를 제출합니다.

- a) 폭우와 폭염 중 어떤 종류의 모델을 대상으로 예측 불확실성 극복 문제를 다룰 것인가?
- b) (데이터 선택) 예측 불확실성 극복을 위해 어떤 문제 정의에 기반하여 데이터를 선택하고 마련할 것인가?
- c) (데이터 극복) 기계학습을 위한 데이터 부족의 한계를 어떤 방법으로 극복할 것인가?
- d) (데이터 설명) 극단 폭우·폭염 예측 결과의 정확성과 추적성 제고를 어떻게 담보할 수 있을까?



이 중(b~d) 어떤 것에 더 초점을 둘 것인지는 각 연구자의 연구 배경과 전문성에 따라 다를 것입니다. 이러한 연구 접근법의 다양성에 기반한 책임PM 중심 협력 연구 체계는 연구과제 수준을 넘는 프로젝트 목표 달성에 이바지할 것입니다.

연구개발 계획서는 상기 연구목표 달성을 위한 연구 내용으로써

- a) 예측 모델의 대상으로 폭우, 폭염 중 선택하게 된 배경,

- b) (데이터 선택) 연구에 활용할 데이터의 종류와 이를 선택하게 된 문제 정의,
- c) (데이터 극복) 부족한 데이터 수의 한계를 극복하기 위해 제안하는 고유한 해결 방법에 대한 설명과 과학적 가치,
- d) (데이터 설명) 극단 폭우·폭염 예측 데이터의 신뢰성과 추적성 제고를 위한 알고리즘 제안과 수행 방안,
- e) (데이터 지식화) 예측 결과를 물리 지식의 확장을 위해 활용할 수 있는 방법론,
- f) 연구 결과의 파급효과 확대를 위한 계획¹⁶⁾

을 포함해야 하는데, b)~e) 항목은 연구자 전문성에 따라서 연구 수행 가중치를 별도 표기하고, 이에 관련된 연구내용도 구분(데이터 선택/극복/설명/지식화)하여 제시합니다.

(연구 결과 산출물 및 성능 지표)

연구 결과 산출물 계획에는 폭우·폭염 변화 예측 인공지능 모델과 이것이 부족한 데이터 문제를 해결한 모델임을 알 수 있는 연구 결과가 포함되어야 합니다. 그리고, 모델의 예측 우수성 판단을 위한 성능 지표*¹⁷⁾를 함께 제시해야 합니다.

* 성능 지표 (예)

- 각종 최근 신기술(state-of-the-arts) 또는 기상청 현업 서비스 대비 성능 향상 수준
- 과거 10년 기간 동안 나타난 이상 폭우·폭염 데이터 대비 모델의 결과 데이터 비교 및 정확도 수준

상기 연구 결과 산출물 이외, 데이터 선택/극복/설명/지식화 연구 각 단계의 우수성을 나타내는 별도의 지표, 성능, 성과물 계획도 자유롭게 제시할 수 있습니다.

16) 예) 결과물에 쉽게 접근할 수 있는 프로그램 코드 모듈화와 소스 코드 공개 계획 등

17) 열거된 성능 지표는 전체 목록이 아닐 수 있음. 연구 제안자는 새롭게 조정된 지표를 제안할 수 있으며, 이 경우 해당 지표가 인공지능 모델의 우수성과 관련된 의미를 설명해야 함.

(제외 대상)

본 프로젝트에서는 아래에 해당하는 기술 개발을 지원 대상으로 고려하지 않습니다.

- 선행 기후·기상 변화 예측 연구를 통해 알려진 방법론을 활용하여 연구 범위를 확장하는 연구
- 유사 연구의 중복, 후속 또는 이것을 확장하는 연구
- 기존 기술을 활용한 점진적 성능 향상 연구
- 데이터 부족의 우려가 없는 기후·기상 현상을 예측하는 data-driven machine learning 연구

4. 추진 일정

(프로그램 구조, 과제 구성)

한계도전전략센터에서는 1인 연구 책임자 중심, 3개 내외의 과제(선정 과제 별 연간 5억 원 내외¹⁸⁾, 연구 기간 최대 3년 7개월) 선정을 계획하고 있습니다. 연구개발 계획서를 제출하는 전문가는 연구내용과 목표를 고려하여 센터가 계획하는 자원과 기간 범위의 계획서를 제출해야 합니다. 연구과제는 단계별 (19개월/12개월/12개월) 평가를 통해, 다음 단계 연구 수행에 대한 진행 여부(Go/No-Go)가 결정되고, 연구비는 책임PM의 연구 수행내용 검토를 통해 가감 가능할 것입니다.



그림 4. CLIMECAST 프로젝트의 주요 일정

본 연구과제는 big data를 활용하는 data-driven machine learning 기법의 한계 극복을 통해 구현되는 폭우·폭염 변화 예측 모델 개발을 목표로 합니다. 그리고, 이러한 도전적 시도를 통해 설명할 수 있고(explanative), 인과 관계가 있는(causal) 인공지능 모델이 개발될 수 있기를 기대합니다.

협약 전 후보 과제 워킹과 전문가 컨설팅에 따른 연구개발 계획서 변경, 협력 과제 추가 기획 등 적극적인 피보팅(pivoting)을 통해 각 과제가 갖는 약점도 보완하고, 그림5와 같은 데이터 흐름 체계를 참고하여 여러 연구 배경(기후 과학, 수리 과학, 컴퓨터 과학 등)에서 도전적인 연구를 시도하는 과제를 선정하고, 책임PM 중심의 협력·융합 연구를 추진합니다.

연구개발 계획서를 제출하는 전문가는 수행 기간 전체에 걸쳐 연구개발 로드맵과 마일스톤 목표를 구체적으로 제시해야 합니다. 연구개발 로드맵에는 중간 단계에 적용해야 하는 특정 작업 및 이의 수준을 가늠할 수 있는 성과지표를 제공해야 합니다. 그리고, 과제 추진 일정을 준수하고 목표, 지표, 중간 및 최종 성과물을 충분히 달성할 수 있는 계획을 제시해야 합니다.

¹⁸⁾ 단계별 예산 규모는 공고문을 참고.

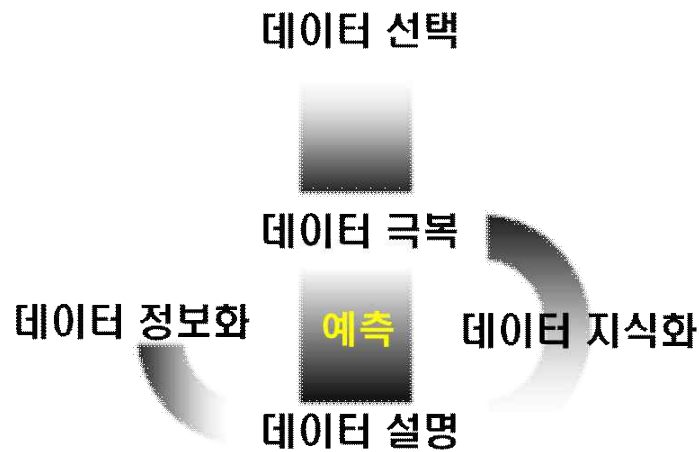


그림 5. 프로젝트의 데이터 흐름 체계

(기술 로드맵, 점검 평가 일정)

책임PM은 연구 과정의 지식화와 진화적 Risk 관리를 위해 전문가 리뷰 미팅을 계획하고 있습니다. 단계 점검을 통해 연구자가 설정한 목표가 달성되었는지 확인함으로써 다음 단계 연구과제의 진행 여부를 결정할 수 있습니다. 책임PM은 정기 리뷰 미팅에서 토론하는 실험실 데이터를 바탕으로 마일스톤 점검 보고서를 작성 및 연구진과 공유하고, 센터에 제출한 점검 보고서는 연차(단계) 보고서로 활용될 수 있습니다.

5. 제공 성과물

연구 책임자는 최소한 다음과 같은 연구 결과물을 제공해야 합니다.

- 책임PM과의 회의(현장 방문, 전문가 리뷰)에서 연구 수행의 진척 상황을 파악할 수 있는 실험 데이터 및 결과 요약 자료
- 단계 점검일 10일 전까지 제출되어야 하는 단계 완료 보고서
- 성과의 수준을 판단할 수 있는 연구 결과 산출물
- 연구 수행 과정에서 연구자의 목표 달성을 위한 노력을 설명하는 각종 자료 등

6. 기타

(협업 관련)

한계도전전략센터는 서로 다른 접근 방법으로 제안된 다수의 과제를 선정할 수 있습니다. 그리고, 모든 연구 수행자가 프로젝트의 목표 달성을 위해 협업할 것을 기대하고, 접근 방식이 다른 타 전문성에 기반한 상호 협력이 이루어지도록 운영할 것입니다. 지적 재산권 혹은 이해 상충의 문제가 없는 조건에서 투명성이 보장된 협업을 추진할 수 있습니다.