

국외 출장 결과보고서

결 재	담당자	실장	사업단장
	2월 18일	2월 19일	2월 19일
	오테진	김정한	박훈

1. 출장개요

출장목적	AGU Fall Meeting 2024 참석 및 발표		
출장일정	2024. 12. 8. ~2024. 12. 15. (6박 8일)		
성명		소속	직급(직위)
오테진		자료응용팀	책임급연구원(팀장)
방문기관 (또는 회의장소)	미국, 워싱턴 D.C., Walter E. Washington Convention Center		
출발지	인천	도착지	미국, 워싱턴 D.C.

2. 업무수행내용

가. 출장 배경 및 목적

- 미국 지구물리학연합(American Geophysics Union, 이하 AGU)은 매년 미국에서 열리는 전 세계에서 가장 큰 규모의 지구과학분야의 연합회로 이번에는 ‘What’s next for SCIENCE’라는 주제로 2024년 12월 9-13일에 미국 워싱턴 D.C.에서 개최했다. 올해 운영체계실에서의 AGU 2024 학회 출장 인원은 자료응용팀 1명 (오테진 팀장) 이다. 이번 학회에 참석함으로써 사업단에서 개발하고 있는 강수 예측 모델 개발 전략과 관측기반 모델, 예측장 기반 후처리 보정 모델 아키텍처에 대한 설명과 검증 체계로 나타난 결과 등의 연구 결과를 발표하고 해외 전문가들로부터 피드백을 받았다. 또한 해외 최신 연구 동향을 파악하여 향후 인공지능 개발 연구 계획에 활용하고자 한다.

나. 주요 업무 수행 사항

- 차세대수치예보모델개발사업단 인공지능 기반 강수예측 시스템 연구 개발 현황을 발표했다. 관측 및 후처리보정 모델 개발 전략과 개발한 모델들의 예측 성능에 대해 소개를 하였다. 사업단에서 개발한 관측기반 모델인 ClusterCast가 DGMR에 비해 더 좋은 성능을 낼 수 있었던 이유에 대한 질의가 많았다. 또한 후처리보정 모델들의 아키텍처에 따른 예측 성능 특성에 대한 논의가 있었다. 강한 강수에 대한 예측 성능을 CSI를 기준으로 훈련시키고 평가할 수 있지만 이는 대체로 강수장에 대한 과다모의를 초래할 가능성이 높다는 논의가 있었으며 이를 극복하기 위한 대안 지표가 필요하다는 공감대가 형성되었다.

- 관측기반 하이브리드 모델 개발: Swin-UNet 모델과 Diffusion 모델을 같이 사용한 예측 모델에 대한 소개가 있었다. Swin-UNet은 시계열 강수예측 성능을 높이는데 도움이 되지만 강수장이 예측 시간이 길어질수록 평활화 되는 경향이 있는 반면 Diffusion 모델은 강수 예측성의 시계열 특성을 잘 표현하지 못하고 모델 예측 결과의 일관성을 담보하기 어렵지만 평활화 문제는 피해갈 수 있는 특성이 있다. 두 모델을 하이브리드로 구성할 경우 각 모델이 갖고 있는 단점을 보완할 수 있다. 다만 Swin-UNet을 단독으로 구성할 경우 RMSE 성능이 하이브리드 모델에 비해 여전히 더 좋은데 이는 하이브리딩을 통한 고해상화 효과에 대한 부작용으로 보인다.
- AFNO 기반 지역모델 개발: FourCastNet과 같은 인공지능 기반 전지구 모델들은 통상 ERA5 재분석장으로 훈련되기에 해상도가 0.25도로 제한된다. 해상도를 증가시키기 위해 AFNO를 이용한 지역모델을 인공지능 전지구 모델에 네스팅 하고 지역 모델은 CERRA 자료로 훈련시킨 발표가 있었다. 아직 모델 개발 초기단계이지만 좋은 예측 성능을 보여줘 향후 주목할 만한 기술로 보여진다.
- 기계학습 기반 홍콩 강수 예측 시스템: 최근 많은 인공지능 기반 모델들이 쏟아지고 있는데 이들 모델들의 결과를 한데 모아서 이에 대한 하이브리드 인공지능 모델을 별도로 구성시 각각의 개별 모델들보다 더 좋은 성능을 낼 수 있다는 발표였다. 이는 앙상블을 구성하는 것보다 좀 더 광의의 개념으로 모델들의 예측 변수가 다를지라도 이를 개별 채널들로 묶어서 구성한 인공지능 시스템이다. 발표된 연구에서는 GraphCast, FourCastNet, FuXi 모델 결과를 이용하였다.

3. 출장성과·시사점 및 향후 업무 활용 계획

가. 출장성과·시사점

- 사업단에서의 인공지능 연구 개발 활동을 전세계 다양한 전문가들에게 홍보하였으며 강한 강수 예측성을 개선하기 위해 전세계 선진 기관들이 어떻게 인공지능을 활용하고 있으며 어떤 성과를 이루어 내고 있는지를 파악할 수 있었다.
- 인공지능 기반 전지구 예측시스템들이 어떻게 활용되고 있는지 확인할 수 있었다. 현재 GraphCast가 가장 많이 알려졌고 이외 Pangu, FourCastNet, FuXi등이 있다. 이들은 각기 다른 설계 방식으로 구성되었는데 특히 동아시아 강수 예측 성능의 경우 FuXi가 월등히 좋은 것이 눈에 띄었다. 이는 전지구 인공지능 모델의 설계에 따라 예측 성능 특성이 크게 달라짐을 시사한다.
- 최근 각광받고 있는 NeuralGCM 모델의 경우 새로운 업데이트를 통해서 기존의 떨어지던 강수

예측성능을 끌어올리기 위해 ERA5 재분석장 강수 데이터가 아닌 IMERG 관측 데이터셋을 훈련 자료로 사용하여 강수 성능을 대폭 향상 시킨 것을 확인할 수 있었다. 이는 인공지능 기반 강수 모델을 구성할 때 재분석장의 다른 변수들은 사용하더라도 강수량 자료는 관측자료를 사용하는 것이 바람직함을 시사한다.

- 관측기반 강수 예측 모델을 구성할때 가장 중요한 자료로 꼽히는 레이더 자료는 그 자체로 많은 불확실성이 내포되어 있다. 반사도 자료를 강수량으로 변환시 ZR 관계식에서 이용되는 계수에 대한 검증 및 최적화를 통해 지상 게이지 관측자료와의 상관관계를 개선할 수 있는지 검토해봐야 할 것으로 보인다.
- 중동 사막지역에 최근 강한 강수 사례가 생겨 이를 인공지능 모델을 통해서 예측하는 것이 가능한가에 대한 연구가 있었다. 인공지능 모델은 훈련 분포 범위 내의 사례들에 대해서는 (데이터 내삽) 예측을 잘 하는 반면 극히 희귀한 사례의 경우, 즉 훈련 데이터에 잘 보이지 않는 분포는 제대로 예측하지 못하는 특성이 있다, 중동 사막지역에서의 강한 강수는 중동에서는 매우 희귀한 사례이지만 훈련 데이터의 지역 범위를 넓힐 경우 강한 강수 사례들이 증가하므로 이를 통해 목표 중동지역에서의 강수 예측성을 높일 수 있었다고 한다. 따라서 한반도 강수 예측의 경우에도 한반도 데이터 뿐만 아니라 보다 넓은 지역의 데이터를 활용하여 지역 강수 예측성능을 올릴 여지가 있을 것으로 보인다.

나. 향후 업무 활용 계획

- 팀에서 관측장 기반 모델과 수치예보 예측장 후처리 보정 모델을 별도로 개발하여 향후 하이브리딩을 하려 하였으나 좀 더 유연한 개발 전략으로 후보정 모델에도 일부 관측 자료를 활용하는 방안을 고려할 수 있겠다. 이를 실험함으로 초기 예측성능을 어느정도로 개선할 수 있는지를 확인해보는 작업을 검토해보려 한다.
- 이번 출장을 통해서 강수 모델들에 사용되는 다양한 손실함수 형태에 대해서 살펴볼 수 있었다. 희귀사례 훈련에 적합한 focal loss는 분류 문제에 많이 활용되는 손실함수이지만 이를 희귀 문제에 적용 가능하도록 변형시키 적용해보는 실험을 하고자 한다.
- 모델 아키텍처도 U-Net, ECA (Efficient Channel Attention), DA (Deformable Attention) 등이 눈에 띄었는데 이중 일부의 적용을 검토해보려 한다.