학사학위청구논문
2024 학년도

외계 행성 대기 분석을 위한 머신러닝 접근

Machine Learning Approaches
for Exoplanet Atmosphere Analysis

광운대학교
전자바이오물리학과
김 석 환

외계 행성 대기 분석을 위한 머신러닝 접근

Machine Learning Approaches
for Exoplanet Atmosphere Analysis

광운대학교
전자바이오물리학과
김 석 환

외계 행성 대기 분석을 위한 머신러닝 접근

Machine Learning Approaches
for Exoplanet Atmosphere Analysis

지도교수 이 현 규

이 논문을 이학 학사학위논문으로 제출함

2024년 11월 12일

광운대학교
전자바이오물리학과
김 석 환

이 논문을 김석환의 이학
학사학위 논문으로 인준함

지도 교수 _____ 인

학과장 _____ 인

광운대학교

2024년 11월 12일

국문요약

외계 행성 대기 분석을 위한 머신러닝 접근

Ariel 프로젝트는 유럽우주국(ESA)에서 진행하는 외계 행성 대기 연구 임무로, 2029년에 발사되어 약 1000개의 외계 행성 대기를 관측할 예정이다. 본 연구에서는 Airel 프로젝트를 지원하는 경연인 'Ariel Data Challenge 2024'에서 제공된 데이터를 기반으로 외계 행성 대기 분석을 위한 머신러닝 기반으로 토대로 다양한 모델을 적용하여 외계 행성 대기 성분을 분석하고자 하였습니다.

초기에는 MLP, CNN, RegNet 등의 복잡한 딥러닝 모델을 사용했으나, 과적합 문제와 데이터에 포함된 높은 수준의 노이즈로 인해 성능 개선에 한계가 있었고, 이러한 문제를 해결하기 위해 Mobilenet v3와 같은 경량화된 모델을 적용하여 모델의 복잡도를 줄이고 노이즈 처리를 강화하였습니다. 이 과정에서 과적합 방지를 위해 dropout을 비활성화하고 batch normalization을 제거하여 모델을 단순화하였습니다.

모델 성능 최적화를 위해 평균 예측과 비례상수 조정을 통한 접근을 시도하였고, in-transit과 out-transit 구간의 평균 값이 행성 대기 특성과 강한 상관관계가 있음을 발견하였습니다. 이를 통해 최적의 비례상수를 찾기 위한 여러 실험을 수행하여 최고 0.420의 점수를 기록할 수 있었습니다. 그러나 데이터의 복잡성과 노이즈를 완벽하게 해결하지는 못하였습니다.

결론적으로, 본 연구는 외계 행성 대기 분석을 위한 머신러닝 모델의 가능성을 확인하였으며, 후속 연구에서는 Gaussian Processing과 Bayesian inference 등 더 정교한 분석 기법을 통해 노이즈 처리를 강화하고, 주요 분자별 흡수 스펙트럼을 활용한 특성 엔지니어링을 통해 더 나은 예측 성능을 달성할 가능성을 제안하였습니다.

핵심 단어: 머신 러닝, 외계 행성, 대기 성분 분석

ABSTRACT

Machine Learning Approaches for Exoplanet Atmosphere Analysis

Kim, Seok Hwan

Dept. of Electrical and Biological Physics

Kwangwoon University

The Ariel project, conducted by the European Space Agency (ESA), is a mission focused on studying the atmospheres of exoplanets and is scheduled to launch in 2029, aiming to observe the atmospheres of approximately 1,000 exoplanets. This study leverages data provided in the "Ariel Data Challenge 2024," a competition supporting the Ariel project, to apply various machine learning models for analyzing exoplanetary atmospheric compositions.

Initially, complex deep learning models such as MLP, CNN, and RegNet were employed; however, challenges like overfitting and high levels of noise within the data limited performance improvement. To address these issues, we applied a lightweight model, Mobilenet v3, to reduce model complexity and enhance noise handling. In this process, dropout was deactivated, and batch normalization was removed to simplify the model and prevent overfitting.

To optimize model performance, we experimented with average predictions and proportional constant adjustments, discovering a strong correlation between the mean values in the in-transit and out-transit phases and the characteristics of the planetary atmosphere. Through numerous trials to identify the optimal proportional constant, we achieved a maximum score of 0.420. However, the complexity of the data and noise issues were not entirely resolved.

In conclusion, this study demonstrates the potential of machine learning models for analyzing

exoplanetary atmospheres. Future research could further enhance noise handling through more sophisticated methods such as Gaussian Processing and Bayesian inference, and improve prediction accuracy by implementing feature engineering based on the absorption spectra of key molecules.

Key words : Machine learning, exoplanet, atmosphere analysis

차 례

그 림 차 례

표 차 례

수 식 차 례

Ⅰ. 서론

최근 몇 년간 데이터 분석과 머신러닝은 천체물리학과 우주 탐사에서 중요한 도구로 자리 잡고 있다. 본 논문은 Kaggle의 'Ariel Data Challenge 2024'에서 제공된 데이터를 기반으로 행성 대기의 특성을 분석하는 것을 목표로 한다. Ariel(Atmospheric Remote-sensing Infrared Exoplanet Large-survey)은 ESA(유럽우주국)에서 진행하는 외계 행성 대기 연구 프로젝트로, 2029년에 발사되어 4년간 태양-지구 라그랑주 포인트2(L2)에서 운용되며, 약 1000개의 다양한 외계 행성의 대기를 관측하여 화학 조성, 온도 구조, 계절 변화를 분석하고자 한다. 이 논문은 머신러닝을 통해 추후 Ariel 프로젝트로 얻어진 데이터를 이용해 대기 분석을 할 수 있도록 유사한 데이터를 이용하여 다양한 머신러닝 기법과 데이터 분석을 시도하고 관측 데이터를 가장 잘 설명하는 모델을 찾아내고자 하는 것을 주목적으로 한다.

Ⅱ. 본론

2.1. 도메인 지식(Domain Knowledge)[5]

본 데이터 분석에는 보다 수월한 접근과 데이터 이해를 위해 Eclipse, Occultation, 빛 손실, Limb Darkening, Radial Velocity, Rossiter-McLaughlin 효과와 주 노이즈 원인을 이해할 필요가 있다.

Eclipse는 한 천체가 다른 천체에 의해 가려지는 현상으로 작은 천체가 큰 천체 앞을 지나는 것이고 Occultation은 작은 천체가 큰 천체 뒤로 숨는 것으로 secondary eclipse이라고도 한다. 외계 행성의 특성을 이해하기 위해 일식(Eclipse)과 월식(Occultation) 현상을 관측한다. 이러한 현상을 통해 행성의 궤도, 질량, 반지름, 온도, 대기 성분 등을 파악할 수 있다.

본 데이터에서는 Eclipse에서 발생한 transit data로만 구성되어 있다. 행성이 항성 앞을 지나갈 때, 항성의 일부 빛이 가려지며 이로 인해 발생하는 빛 손실을 통해 행성의 반지름과 대기 구성 정보를 얻을 수 있다. 또한 항성은 가장자리로 갈수록 온도와 불투명도의 차이로 인해 밝기가 감소하는데, 이를 Limb Darkening이라고 한다. 이 현상은 행성이 항성의 가장자리를 통과할 때 빛 손실의 양에 영향을 미친다. 이 현상을 고려해 더 정밀한 모델링이 가능하다.

Radial Velocity는 천체가 관측자를 향해 다가오거나 멀어지는 속도를 측정하여 행성의 질량과 궤도 특성을 추정하는 방법이다. 이 정보를 통해 행성의 질량을 추정할 수 있으며, 이는 Light Curve와 함께 사용되어 행성의 특성을 파악하는 데 중요하다.

Rossiter-McLaughlin 효과는 행성이 항성의 일부를 가릴 때 항성의 자전에 의해 발생하는 스펙트럼의 청색 편이 및 적색 편이 현상으로, 이 효과를 통해 행성의 공전축과 항성의 자전축 간의 정렬 관계를 파악할 수 있다.

이런 우주에서 벌어지는 사건을 관측하는 과정에서는 관측 방법에 따라 Photon Noise, Scintillation, Differential Extinction, Flat Fielding 등 다양한 노이즈가 존재하며 우주에서 관측하는 우주망원경의 경우 지구 대기와 관련된 Scintillation, Differential extinction의 경우 영향이 적지만 망원경 위치 포인팅에서 발생하는 Jitter noise 같은 noise가 지배적으로 작용한다.

2.2. 데이터 구성[1]

'Ariel Data Challenge 2024'에서 제공된 데이터는 외계 행성의 대기 스펙트럼 데이터로 구성되어 있으며, 두 개의 장비(FGS1, AIRS-CH0)를 이용하여 얻은 데이터로 이루어져 있다. Train data set의 경우 ARIEL의 시뮬레이션 모델을 사용하여 만들어진 가상의 데이터로 이루어져 있으며, Jitter noise, Photon noise, Flat fielding과 같은 다양한 노이즈가 포함되어 있다. 실제 평가에 이용되는 hidden test data set의 일부는 실제 관측 데이터로 구성되어 있다.

FGS1은 가시 스펙트럼(0.60~0.80 μm)에서 고정밀 광도 측정을 수행하며, AIRS-CH0은 적외선 분광기(1.95~3.90 μm)로 약 R=100의 해상도를 가지고 있다. 각 장비는 일정 시간 동안 축적된 전하를 측정하고 리셋하는 과정을 반복하여 각 행성별로 FGS1은 135,000 프레임, AIRS-CH0은 11,250 프레임의 이미지를 생성한다.

이 데이터는 uint16 형식으로 제공되며, 원래의 동적 범위를 복원하기 위해 gain과 offset 값을 각기 적용해야 한다. 이 데이터는 제한된 수의 photon으로부터 수집되었다는 조건 하에 상당한 양의 노이즈를 포함하고 있다.

관측된 신호 데이터에 해당하는 Signal File, 데이터 수집과정에서 발생한 노이즈를 제거하기 위한 Calibration Files, 데이터에 관한 정보가 들어 있는 Metadata Files로 구성되어 있으며 Signal File에는 각각 AIRS-CH0, FGS1 신호 데이터로(11250, 11392), (135000, 1024) 크기로 구성된 AIRS-CH0/FGS1_signal.parquet이 들어 있다.

Calibration Files에는 센서의 열 잡음과 바이어스 수준을 캡처한 데이터로, dark current를 제거하는 데 사용되는 AIRS-CH0/FGS1_calibration/dark.parquet, 빛에 반응하지 않는 dead pixel과 지속적으로 높은 신호를 생성하는 hot pixel 위치 데이터인 AIRS-CH0/FGS1_calibration/dead.parquet, 픽셀 간 감도 차이와 광학 시스템의 불규칙성을 보정하는 데 사용되는 데이터인 AIRS-CH0/FGS1_calibration/flat.parquet, 센서의 선형성 보정에 대한 정보로, 전자 포화 정도에 따른 픽셀의 비선형성을 보정하기 위한 데이터인 AIRS-CH0/FGS1_calibration/linear_corr.parquet, 센서의 판독 과정에서 발생하는 전자적 노이즈에 대한 데이터인 AIRS-CH0/FGS1_calibration/read.parquet으로 구성되어 있다.

Metadata Files로는 아날로그-디지털(ADC) 변환 파라미터(gain 및offset)를 포함하고 있어 데이터의 원래 동적 범위를 복원하는 데 사용된다. 또한 각 행성의 시뮬레이션에 사용된 항성 번호도 포함되어 있으며 train data set의 경우 0, 1 star에 속한 행성들로 구성되어 있고 test data set의 경우 0, 1 star 이외에도 추가의 star 1개에 속한 행성들로 구성되어 있는

[train/test]_adc_info.csv, 실제 스펙트럼(Ground truth spectra)이 들어 있는 train_labels.csv, 두 장비에 대한 축 정보가 들어 있는 axis_info.parquet, 데이터 셋의 각 실제 스펙트럼에 대한 파장 그리드인 wavelength.csv. 0.705 μm ~ 3.895036 μm 범위를 가진다.

2.3. 평가 지표[1]

예측된 스펙트럼(μ_{user})과 해당 불확실성(σ_{user})을 실제 픽셀 수준 스펙트럼(y)과 비교하여 Gaussian Log-likelihood (GLL) 함수로 평가한다.

각 파장에서 계산된 GLL 값들은 모든 파장과 테스트 세트를 통틀어 합산하여 최종 GLL 값(L)을 산출한다. 이 최종 GLL 값은 다음 변환 함수에 의해 점수로 변환된다.

$$GLL = -\frac{1}{2}(\log(2\pi) + \log(\sigma_{user}^2) + \frac{(y-\mu_{user})^2}{\sigma_{user}^2})$$

수식 1. Gaussian Log-likelihood 함수

$$score = \frac{L - L_{ref}}{L_{ideal} - L_{ref}}$$

수식 2. score 계산 공식

여기서 L_ideal은 제출이 실제 값과 완벽하게 일치하며, 불확실성이 10 ppm(백만 분의 10)일 때를 의미하며, L_ref는 모든 데이터에 대해 학습 데이터 셋의 평균과 분산을 예측값으로 사용하는 경우로 정의한다.

점수는 0에서 1 사이의 실수로 반환되며, 높은 점수가 더 우수한 성능을 나타낸다. 0보다 낮은 점수는 0으로 처리된다.

2.4. 데이터 전처리[6]

모델링과 데이터 분석에 앞서 관측 과정에서 발생한 다양한 노이즈를 제거하기 위해 주어진 calibration file들을 사용하여 기본적인 전처리 과정을 수행하였다.

비선형 픽셀 반응을 보정하기 위해 AIRS-CH0/FGS1_calibration/linear_corr.parquet을 사용하여 각 픽셀에 대해 다항식을 적용하였고 AIRS-CH0/FGS1_calibration/dark.parquet을 사용하여 센서의 열 잡음과 바이어스 수준을 보정하고 시간에 따른 조정 계수 FGS dt = 4.5, AIRS-CH0 dt = 0.1을 적용하였다.

추가적으로 AIRS-CH0/FGS1_calibration/flat.parquet을 사용하여 픽셀 간 감도 차이와 광학
시스템의 불규칙성을 보정, AIRS-CH0/FGS1_calibration/dead.parquet을 사용해 죽은 픽셀

과 핫 픽셀의 값을 NaN으로 처리하였다.

이후 시간에 따른 변화를 줄이기 위해 FGS1의 경우 30*12, AIRS-CH0의 경우 30으로 나누어 Time Binning을 하였다. 이를 통해 FGS1과 AIRS-CH0의 시계열 데이터 사이즈를 일치시켜 병합하기 쉽게 하였고 노이즈를 감소시켜 모델 학습의 효율성을 높이고자 했다.

2.5. 모델링 방법(Modeling Approach)

단순한 jitter noise만으로도 200ppm 크기로 지구 행성의 경우 50ppm, 목성형 행성의 경우 200ppm으로 대기 성분에 따른 신호의 변동이 노이즈의 신호 크기보다 훨씬 작은 데이터 셋으로 단순한 model으로는 대기 성분에 의한 흡수 스펙트럼의 패턴을 알아내기 어려울 것으로 보여 이를 해결하기 위해 기본적인 DNN(Deep Neural Network) model인 MLP(Multi Layered Perceptron), 이미지에 특화된 기본적인 CNN(Convolution Neural Network), 그중 성능이 일반적으로 뛰어나다고 알려진 CNN model인 RegNet(Regular Network)과 같은 DNN(Deep Neural Network) model을 사용하여 depth와 구조, 하이퍼 파라미터를 달리하여 여러 시도를 하였다. train data set이 총 673개의 행성으로 이루어져 있어 각 행성을 Time Binning을 통해 187개의 시계열 데이터로 축소를 했더라도 이미지의 사이즈가 (283, 32)로 특성의 개수가 데이터의 개수보다 매우 커 DNN model의 경우 과적합의 위험성이 크다는 점, 너무 큰 노이즈로 뒤덮여 있다는 점에서 깊은 구조를 가지는 복잡한 DNN의 경우 적절한 model이 아니라 판단되었다.

그럼에도 패턴을 파악하기 위해서는 단순한 선형 모델로는 부족해 보여 그 중간 단계에 있는 모델로 메모리의 사용이 제한적인 상황에서 사용하기 위해 고안된 MobileNet v3를 사용하여 학습을 시도하기로 하였다. 기존 MobileNet v3 model에서 과적합을 피하기 위한 추가적인 조치로 dropout을 0으로 하여 사용하지 않았고 batch normalization과 같은 정규화 층을 identity block으로 바꾸어 훈련을 진행하였다.

Try	Preds constant	Sigma constant	Score
1	0.2	0.2	0.246
2	0.2	0.2	0.246
3	0.217	0.2	0.254
4	0.217	0.02	0.106
5	0.2	0.02	0.000
6	0.217	0.215	0.244
7	0.217	0.185	0.264
8	0.217	0.180	0.268
9	0.217	0.1	0.339
10	0.217	0.0	0.026
11	0.217	−0.005	0.000
12	0.217	0.1201	0.319
13	0.217	0.05	0.381

14	0.217	0.005	0.000
15	0.217	0.063	0.377
16	0.217	0.017	0.000
17	0.217	0.012	0.000

표 1. MobileNet v3 Try Table Results

또한 quantile을 설정해 과대 예측과 과소예측을 의도적으로 하여 sigma 또한 정보를 기반으로 예측하기 위한 quantile loss function을 사용하였다. 물리적 이론 법칙을 기반으로 한 PIMM을 사용해보고자 시범적으로 Rayleigh scattering 공식을 loss function에 추가하여 시도도 해보았으나 오히려 결과가 안 좋아 quantile loss function만을 사용하여 prediction과 각 sigma에 상수를 달리하여 더해 총 17번의 시도를 하였고 이 방법의 최대치는 0.381이었다.

이 모델의 한계가 명확해 좀 더 EDA를 진행하면서 방법을 찾아보던 중 MCMC를 사용하여 우선 행성별 정확한 노이즈 사이즈를 알아내고 이를 이용해 노이즈를 제거하는 개념으로 시도해 보았으나 별 효용이 드러나지 않았고 현 데이터의 노이즈 특성상 여러 노이즈가 각기 파장과 시간에 따른 선형, 비선형 노이즈가 합성된 노이즈로 구성되어 있어서 더 복잡한 방법이 필요성을 느끼던 중 target의 값 변동이 평균치에서 최대 0.0001 이내의 변동을 보이는 걸 발견하고 평가 지표에 따라 가장 정확한 결과의 sigma가 0.00001이므로 우선적인 과제가 평균을 정확히 예측하는 것으로 생각하여 선형 모델을 이용하기로 하였다.

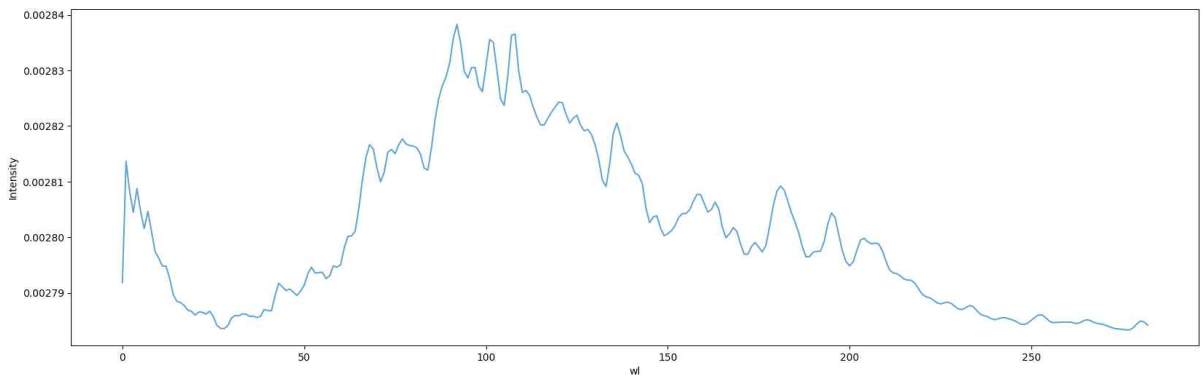

그림 1. ID.785834 행성의 각 파장별 실제 intensity

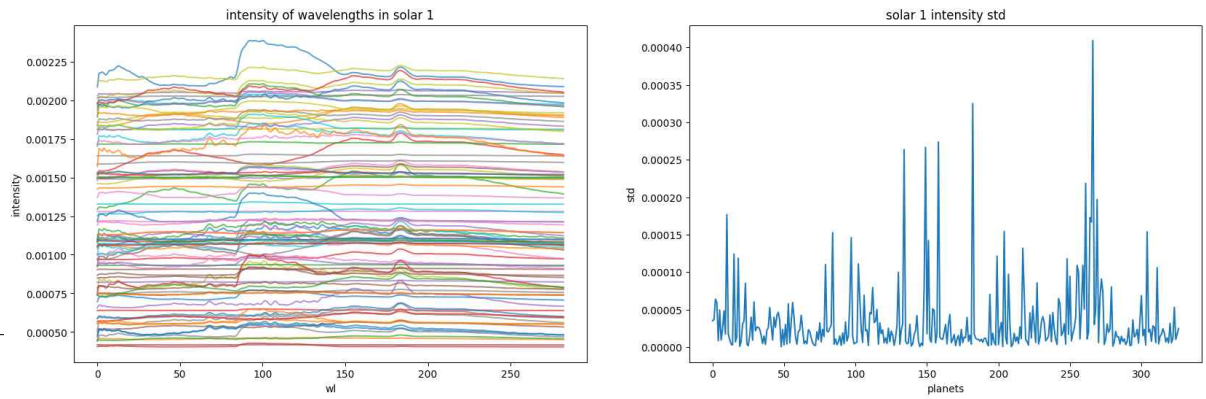

그림 2. Star 1에 속하는 100개 행성들의 각 파장별 실제 intensity(왼쪽), Star 1에 속하는 전체 행성들의 각 파장별 실제 intensity의 총 표준편차(오른쪽)

transit 구간을 행성이 별과 지구 사이를 지나가는 부분(in-transit)과 아닌 부분(out-transit)으로 나누어 파장 세기를 통합하여 시간 평균을 내어 in-transit과 out-transit의 평균과 label의 평균값이 강한 상관관계를 나타내는 것을 발견했고 이후 비례상수를 찾는 과정을 거쳤다.

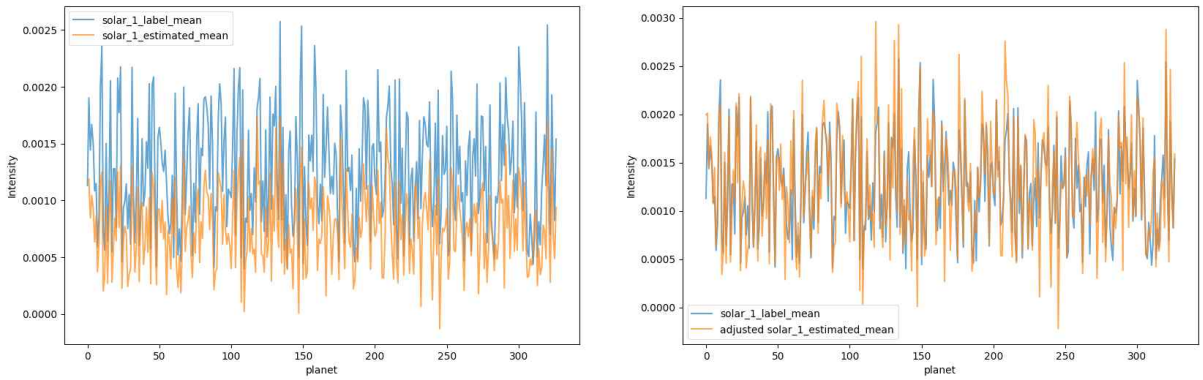

그림 3. Star 1에 속하는 각 행성의 파장별 실제 intensity의 평균과 In-transit mean, Out-transit mean의 평균값(왼쪽), Star 1에 속하는 각 행성의 파장별 실제 intensity의 평균과 (In-transit mean, Out-transit mean의 평균값) / 0.59(오른쪽)

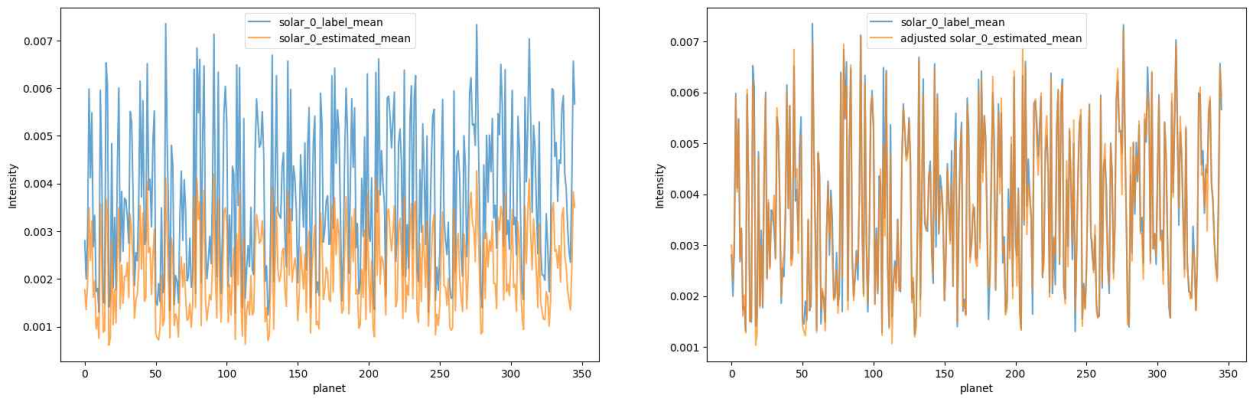

그림 4. Star 0에 속하는 각 행성의 파장별 실제 intensity의 평균과 In-transit mean, Out-transit mean의 평균값(왼쪽), Star 1에 속하는 각 행성의 파장별 실제 intensity의 평균과 (In-transit mean, Out-transit mean의 평균값) / 0.59(오른쪽)

Try	Preds ratio	Sigmas	Score

1	0.59(total)	0.000037	0.000
2	0.57(total)	0.000249	0.374
3	0.58(total)	0.000249	0.406
4	0.61(total)	0.000249	0.401
5	0.598(total)	0.000249	0.420
6	0.580(total)	0.00013	0.154
7	0.598(total)	0.00013	0.204
8	0.590(total)	0.00013	0.205
9	0.598(total)	0.00015	0.289
10	0.598(star 0), 0.621(star 1), 0.609(star 2)	0.000243, 0.000282, 0.0005	0.400
11	0.621(star 0), 0.598(star 1), 0.609(star 2)	0.000282, 0.000243, 0.0005	0.399
12	0.589(total)	0.00029	0.3166

표 2. Solar system 별 label 평균 추정 비례상수 탐색 결과

9번째 시도까지 solar system에 관계없이 동일한 비례 상수를 사용해서 시도한 결과 평균값으로는 약 0.00025 정도의 오차의 정확도로 예측하여 최고 0.420의 점수를 받아낼 수 있었다.

이후 train label 기준으로 내부적으로 동일한 평가 지표를 사용해 solar system 별 최적의 비례 계수를 찾았고 이를 이용해 시그마를 다르게 하여 여러 시도를 했으나 급격한 점수 상승은 기대할 수 없었다.

Ⅲ 결론

본 연구에서는 'Ariel Data Challenge 2024'에서 제공된 외계 행성 대기 스펙트럼 데이터를 바탕으로 다양한 머신러닝 모델을 적용하여 대기 성분을 분석하고자 했다. 초기에는 MLP, CNN, RegNet 등의 복잡한 딥러닝 모델을 사용하였으나, 데이터와 노이즈의 특성상 과적합 문제와 노이즈의 영향을 극복하기 어려웠다. 이러한 한계를 해결하고자 MobileNet v3와 같은 경량화된 모델을 사용하여 모델의 복잡도를 줄이고 노이즈 처리에 집중하는 접근을 시도하였다.

MobileNet v3를 사용한 모델은 과적합 방지를 위해 dropout을 비활성화하고 batch normalization을 제거하였고 복잡한 딥러닝 모델들과 비교하여 훨씬 좋은 성능(0.381)을 보여주었지만, 여전히 데이터의 노이즈 특성으로 인해 높은 예측 성능을 내기 어려웠다. Label의 분포를 보았을 때 std가 0.001 scale에 위치하고 있어 우선적으로 비례상수를

조정하고 각 sigma에 대해 다양한 실험을 수행하여 최적의 평균값을 예측하고자 했다. 결과적으로 in-transit mean과 out-transit mean의 평균값과 target 값의 전체 평균이 비례한다는 점을 발견하였고 이후 최적의 비례상수를 찾기 위한 여러 번의 시도를 통해 최고 0.420의 점수를 얻을 수 있었으나, 결국 데이터의 복잡성과 노이즈를 완전히 해결하지 못한 한계가 있었다.

본 연구에서는 보다 전문화된 노이즈 제거와 같은 전처리 시도와 데이터 포인트 별 예측을 하지 못하였고 전체 평균 예측을 1차 목표로 하여 접근하였으나 단순한 평균과 비례상수를 통한 예측을 하였고 보다 유연한 행성별 평균값 예측을 하지 못하였다.

따라서 후속 연구에서는 다항식 fitting, Rigde regression과 같은 선형 모델을 통해 보다 세밀한 영역 내에서 fitting된 여러 예측을 한 뒤 앙상블하여 보다 정확한 예측을 할 수 있을 것이다. 또한, 충분한 시도를 하지 못했지만 Gaussian Process[7] 또는 이런 Airel data와 같이 사전에 정확한 확률 분포를 알기 어려운 경우 데이터를 이용해 사후 확률을 업데이트해 나가면서 정확한 예측이 가능한 Bayesian inference을 이용하고 중요한 몇몇 분자들(h2o, n2, co2, etc.)의 분자별 흡수 스펙트럼 추세를 활용한 feature engineering을 통해 더 정확한 예측이 가능할 것으로 기대된다.

참고문헌

1. Kaggle. (2024). Ariel Data Challenge 2024. Retrieved from https://www.kaggle.com/competitions/ariel-data-challenge-2024
2. Ariel Mission. (n.d.). Retrieved from https://arielmission.space/
3. ARIEL (Atmospheric Remote-sensing Infrared Exoplanet Large-survey). Retrieved from https://www.eoportal.org/satellite-missions/ariel#development-status
4. European Space Agency (ESA). (n.d.). Ariel factsheet. Retrieved from https://www.esa.int/Science_Exploration/Space_Science/Ariel_factsheet
5. Winn, J. N. (2014). Transits and Occultations. Massachusetts Institute of Technology. Retrieved from http://arxiv.org/abs/1001.2010v5
6. Sergei Fironov. (2024). Ariel_only_correlation. Retrieved from https://www.kaggle.com/code/sergeifironov/ariel-only-correlation
7. Fortune, M., Gibson, N. P., Foreman-Mackey, D., Evans-Soma, T. M., Maguire, C., & Ramkumar, S. (2024). How do wavelength correlations affect transmission spectra? Application of a new fast and flexible 2D Gaussian process framework to transiting exoplanet spectroscopy. Astronomy & Astrophysics