2022 Ariel Data Challenge Solution: Light Track

Team Name: The Gators

November 9 2022

Below we answer the specific questions from the organizers.

1 How did you use the training data?

We formatted the provided spectral information, auxiliary parameters, FM-parameters, and retrieved quantile data for each planet as several Numpy data-files, which we have included in our submission, since our code reads them as inputs. If you prefer to read the original files, please uncomment the corresponding block of code, and comment out the code which imports the Numpy arrays.

2 Did you perform any data preprocessing step?

1. For planets with spectral values higher than 0.1, we replaced these anomalously high values with a value constructed from the other (lower than 0.1) spectral values.

2. In order to focus on the effect of the atmosphere, we subtract the contribution of the planet itself:

$$M'_\lambda = M_\lambda - \frac{R^2_{\text{planet}}}{R^2_{\text{star}}} \tag{1}$$

3. We used the analytical solution of the thermal equilibrium between the star and planet to estimate the temperature T_p of the planet from the auxiliary parameters alone.

4. We constructed the additional features $\frac{R_P}{R_S}$, $\frac{D}{H}$, and $\frac{R_S}{H} \max_\lambda (M'_\lambda)$, and concatenated them with the auxiliary parameters.

5. We standardized the auxiliary features and normalized M'_λ and the noise data by dividing by their respective maximum values for each planet.

3 What kind of model did you go for?

We use several fully connected neural networks some of which use concatenations or products of the outputs of previous modules as inputs.

4 What is the input/output of the model?

The model takes in the 52 wavelength bins of the flux modulation data, the 52 corresponding uncertainties (noise), and the nine provided auxiliary parameters plus the three additional features we constructed. The model takes these three vectors as input, and has a total of 116 input neurons.

The model outputs 15 values which are the central and two quantile values for each of the five components. For the temperature we use the value of T_p computed in step 2 above and then a linear fit to the target data for Q_1 and Q_3 to predict $\{Q_1, Q_2, Q_3\} = \{0.9865\,T_p, T_p, 1.016\,T_p\}$.

5 Did you do any post-processing to the output?

- We sort the quantile values for each concentration. (For example, it would not make sense to predict that the 86th-percentile value is less than the median value, etc.)
- We constrain the output to be within the interval $[-12, -1]$ for each of the absorber concentrations.

6 Did you perform any sampling step? If so please describe.

There is no sampling.

7 Did you use any external library and/or forward model?

We did not use any forward model. We used standard libraries such as pytorch, scipy, pandas, and numpy.