CS 289 Project S - Final Writeup

Chenhui Hao
ch_hao@berkeley.edu

Kaijing Ding
kaijing@berkeley.edu

Ke Liu
liuke126@berkeley.edu

Zishan Cheng
z.cheng@berkeley.edu

Abstract

This project is inspired by ancient Chinese astronomer Liu Hong's method of predicting positions of heavenly objects and moon phases. We reviewed and improved his works by building parametric models to acquire more accuracy in prediction of positions and lunar illumination. We also applied Neural Network and Random Forests models to discover patterns of lunar movements and make predictions on lunar illumination without any Astronomy knowledge beforehand, and found inspiring results in prediction accuracy compared with theory-based OLS model.
https://github.com/haochenhui97/AstronomyML

1 Introduction

Liu Hong (A.D. 135-210), an ancient Chinese mathematician and astronomer, is the first Chinese astronomer who has provided us with the earliest detailed and complete theory of the Sun and Moon. He created the Moon Departure Table, a numerical table of the difference between the actual and average degrees of the moon every other day after the moon passes through the perigee, to correct the uneven motion of the Moon.[1] We have collected Liu Hong's measurements to build our dataset as from our early project. In this project we are motivated by Liu Hong's method of precisely predicting moon's position, moon phase and eclipses, based on the Moon Departure Table.[2]

1.1 How Liu Hong predict lunar position

To begin with, Liu Hong employed a simple first order interpolation equation to calculate the rate of moon movement at any given time t:

$$f(x) = f(x_i)/24 + t/24 * [f(x_{i+1} - f(x_i)](0 < t < 24) \tag{1}$$

For example, if we want to calculate the rate of speed at 8 am on day (i+1), we need to first gather data of daily average speed for day i $f(x_i)$ and day (i+1) $f(x_{i+1})$ from Moon Departure Table (named as Parts of Lunar Daily Motion by Liu Hong). Then the desired speed will be

$$f(x_i)/24 + 8/24 * [f(x_{i+1}) - f(x_i)](unit : degree/hour). \tag{2}$$

To obtain lunar position, multiply the speed by 8 hours to get the degrees of moon movement at that time. Lastly, add it to previous location at 12am on that day.

As can be seen, Liu's method was quite straight forward but could still be improved. How he calculated the speed was actually based on hypothesis that moon movement speed is even throughout the day, which is not true. Thus a more precise model could be built here.

Preprint. Work in progress.

1.2 How Liu Hong predict moon phase and eclipses

Liu Hong figured out that moon movement from new moon to the next new moon had a time cycle of 29+773/1457 days (with error around 4 sec) and developed the Qianxiangli as lunar calendar accordingly. Thus moon phase can be told directly by the date in that month. For example, first day of a month always has a new moon, and full moon occurs on the 15th or 16th day.

As for eclipses, Liu Hong figured out the angle between ecliptic and lunar orbit was 6 du 1 fen (around 6 degrees). Though he did not distinguish solar and lunar eclipse, he recognized a rule for detecting the eclipses: when it's new moon or full moon, if sun is away from the node of ecliptic and lunar orbit more than 14°33 , then there must be no eclipse. He also developed other concepts that enabled him to predict eclipse to the precision of two hour. In this project we intend to predict both moon phase and percentage of illumination. We also predict solar and lunar eclipse separately by their relative position.

2 Method

2.1 Introduction on datasets

We utilized the dataset from our early project S, and add one more feature of percentage of illumination on moon surface. Also, since moon phase is also influenced by the solar position, we downloaded the same format of solar ephemerides data from NASA HORIZONS System[3]. Our dataset contains 9 variables: RA, DEC, time as year, month, day, hour, dRA, dDEC, percentage of illumination. We download 10-year data ranging from 2009-1-1 0:0 UT to 2019-1-1 0:0 UT with step size of one hour at Beijing. The total size of data contains 87648 observations. In developing our model we divided data into training, test and validation set.

2.2 Prediction of solar and lunar positions

2.2.1 Theoretical analysis

Since the movement of both sun and moon with respect to the earth are periodic, it is intuitive that we want to estimate the sun and moon positions using the periodic characteristics. In fact, the lunar motion seen from the earth composes of at least eight kinds of different motions[4]:

(1) Earth's rotation, which makes the Moon appears to rise from the east and fall in the west everyday;

(2) the non-negligible radius of the Earth in the Earth–Moon system, which explains the differences among lunar speed observed at different locations on Earth;

(3) Moon's revolution around the Earth, which makes the Moon rises 48 minutes everyday;

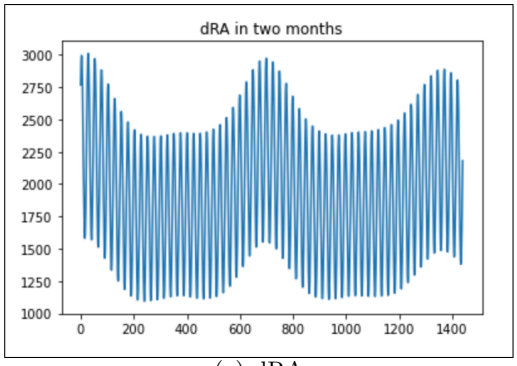
(4) inclination of the lunar orbit to the ecliptic plane, as a result of which the Moon seem to move in diagonals in the plane-of-sky;

(5) the elliptical orbit of the Moon's revolution, which makes the Moon moves faster near perigee and slower near apogee;

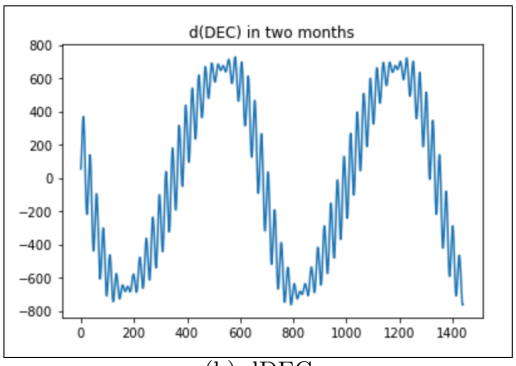
(6) lunar apsidal precession, the period of which is defined as the time it takes the major axis of the Moon's elliptic orbit to precess eastward by 360°, and is about 8.85 years[5];

(7) lunar nodal precession, the period of which is defined as the time it takes the ascending node to move through 360° relative to the vernal equinox (autumnal equinox in Southern Hemisphere), and is about 18.6 years[5];

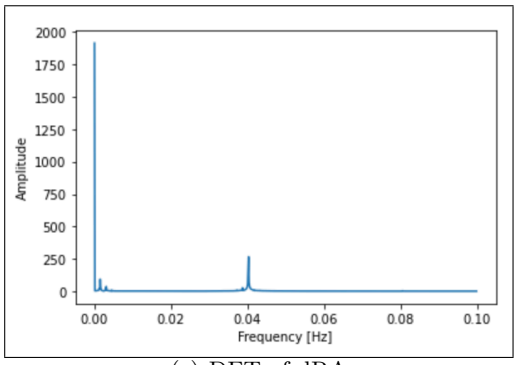
(8) the tidal effect of Moon, which slows the Earth's rotation and make the Moon escape the Earth gradually at the same time.



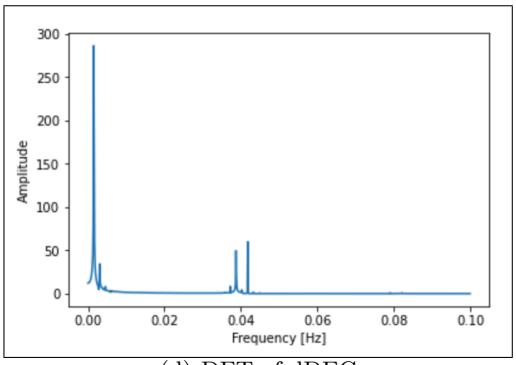
(a) dRA



(b) dDEC



(c) DFT of dRA



(d) DFT of dDEC

Figure 1: Theoretical analysis of solar and lunar positions

In summary, the lunar motion can be viewed as a combination of several periodic movements, which is similar to a periodic signal. This reminds us to try using Fourier transform to decompose the lunar location (which is a function of time) into its constituent frequencies.

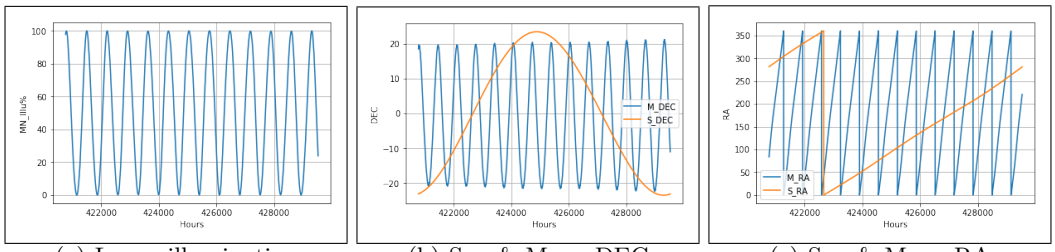
2.2.2 Fourier transform

In this section, we use the numpy.fft.fft funtion to decompose the Moon's motion in RA and DEC into their constituent frequencies.

Before directly estimating the lunar location, we first tried to estimate the lunar speed, since that's how Liu Hong performed the work. According to Fig 1(a) and 1(b), lunar speed in RA and DEC also seem to be a combination of several periodic movements. Then Fig 1(c) and 1(d) show the decomposition of lunar speed in RA and DEC respectively using Discrete Fourier Transform(DFT). As 1(c) illustrates, the amplitude at frequency of 0 reflects the trend that dRA always varies around 2000 seconds per hour. The other amplitude peaks reflect other latent frequencies.

However, we failed to use integration to deduce location from speed directly and resort to DFT to decompose location directly.

Since Fourier transform is based on time series, we didn't random shuffle the data and just split the data to be training set, validation set, and test set with the percentage of 60%, 20%, and 20%. Also, because DFT is employed to data of certain period, we decided the training data size and validation data size to be a multiple of 27.321582 days = 655.717968 hours, which is the length of a tropical month. The results are shown in section 3.1.



(a) Lunar illumination (b) Sun & Moon DEC (c) Sun & Moon RA

Figure 2: Periodical behaviors in variables

2.3 Prediction of moon phases

In this section, we provide two models to predict the moon phase. One model is based more on the theoretical analysis on the motion of the Moon, the Sun and the Earth to generate reliable feature and to adopt more straightforward model. The second method we directly use the raw variables from the post-processing datasets into the neural network.

2.3.1 Theoretical model

This subsection specifies prediction models for moon phased to be estimated in the project. We use the post-processing datasets from Early project, which includes the information of the Sun and Moon with time interval of one hour from 2018-Jan-01 00:00 to 2018-Dec-31 23:00, which indicates we have 8760 records in our model. The specific variables we selected from the post-processing datasets in the model specifications are listed in Table 1. Based on basic statistic analysis, we found that the lunar illumination percentage is a periodical function along the time, and there are also certain periodic behaviors in terms of the movements of the Sun and Moon as shown in Figure 2. We have adopted the cross validation to evaluate the performance of these variables and it shows that the combination of delta_DEC and delta_RA_R6 has the best performance.

When testing the prediction model and the performance of the features, 80% of samples is used as training set and 20% is for testing set. Two methods –ordinary least squares regression, and random forest are adopted to do the prediction.

Table 1: Description of initial variables from Early Project phase

Variable	Description
Y variable	
MN_Illu%	Lunar illumination percentage (float, percent)
X variables	
Time	UTC time for observation YYYY-Mon-Day
Hours	Counted hours starting from 2018-Jan-01 00:00
S_RA	Astrometric right ascension of the Sun center in arc degrees
S_DEC	Astrometric declination of the Sun center in arc degrees
S_dRA*cosD	The angular rate of change in apparent RA of the Sun
S_d(DEC)	The angular rate of change in apparent DEC of the Moon
M_RA	Astrometric right ascension of the Moon center in arc degrees
M_DEC	Astrometric declination of the Moon center in arc degrees
M_dRA*cosD	The angular rate of change in apparent RA of the Moon
M_d(DEC)	The angular rate of change in apparent DEC of the Moon

Theoretically, the lunar illumination, or say moon phase is highly correlated with the relative position of the Sun and the Moon with respect to the Earth. We have studied about the RA and DEC to capture the potential relationship, and we found that the difference between the RA of the Moon and the Sun plays the major role while the difference between the

Table 2: Feature engineering

Variable	Formula and calculation
delta_DEC	= S_DEC − M_DEC
delta_RA	= S_RA − M_RA
delta_-RA	= M_RA − S_RA
delta_RA_R1	= (S_RA − M_RA) mod 180
delta_RA_R2	= abs(S_RA − M_RA) mod 180
delta_RA_R3	= 360−abs(S_RA − M_RA)
delta_RA_R4	= min(S_RA − M_RA,M_RA−S_RA)
delta_RA_R5	= 180-abs(S_RA − M_RA)
delta_RA_R6	= abs(180-abs(S_RA − M_RA))

DEC of the the Sun and the Moon also have some impacts. As RA and DEC are measured related to the earth and zero meridian in RA is labeled 0 degree which intersects the celestial equator at a point called the vernal equinox. It is the time when the Sun crosses the celestial equator in late March of each year, as shown in the Figure 2 (c). So we need to convert the relative position of the Sun and the Moon into absolute positions. Instead of directly adopting the raw variables from the post-processing datasets, we carried out some feature engineering to construct several features based on the raw variables to better capture the periodical performance of lunar illumination. The calculation for the feature engineering in show in Table 2. The detailed results are shown in section 3.2.

2.3.2 Neural network

First, we split the dataset into three components: Training set, validation set, and test set. Then, we construct a 3-layer neural network, whose structure is shown in Figure 3(a). There are eight moon phases in total. To be more precise, we use the percentage of the illumination of the moon as the output in our model. The input layer contains: *Hours, Month, Date, Time, RA, DEC, dRA*cosD, d(DEC)*, and the output of the network is *Percentage of the illumination of the moon*. The hyper-parameters in this model are: the learning rate, the number of neurons of each layer, the number of layers. We used cross-validation to set up proper hyper-parameters. We used ReLU as the nonlinear activation function. We tried SGD and Adam as the optimizers.

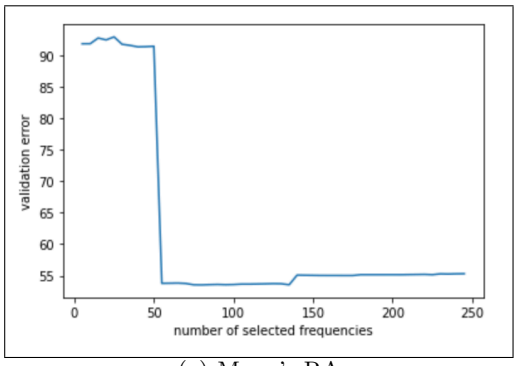
3 Results

In this section, the results of prediction on moon/sun position, moon phase and moon/sun eclipses are presented. We present the findings from model with special focus on the accuracy.

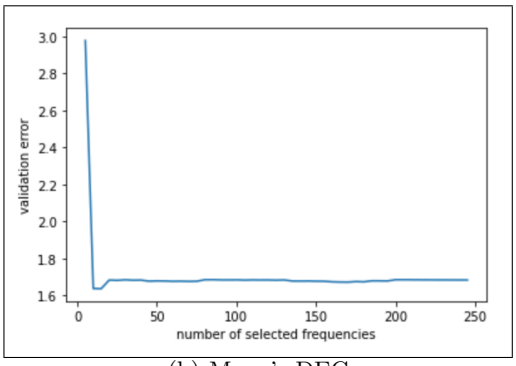
3.1 Sun and moon position

Employing the DFT method, we decomposed the location of Moon and Sun in RA and DEC into their constituent frequencies. However, the number of frequencies that DFT outputs equal the number of input data points, which, if all used in the final estimate model, will certainly lead to overfit. To prevent that, we tuned the number of selected frequencies for each of the four models(Moon's RA, Moon's DEC, Sun's RA, and Sun's DEC) with the validation set, as shown in Fig 5; and decided the number of selected frequencies to be 80, 15, 1, and 4 respectively.

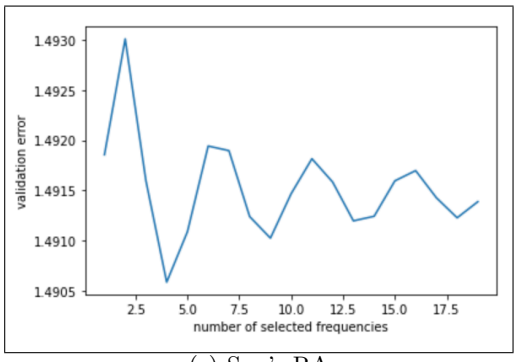
Finally, we used the models to predict the lunar and solar positions for the time period of test set and calculated the loss. The mean squared error for the test set are 387.23, 13.74, 3.49, and 1.87 for the four models respectively. Sun's errors are much smaller than moon's, probably due to the reason that solar motion (or Earth's revolution in fact) is much simpler than lunar motion, which is influenced by at least 8 factors, as discussed in previous section. Also, the period of lunar motion is actually even longer than the time period covered by the training set (e.g. the Saros cycle is approximately 18 years 11 days 8 hours), which indicates



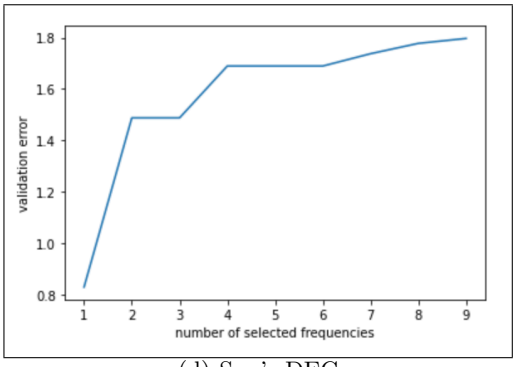
(a) Moon's RA



(b) Moon's DEC



(c) Sun's RA



(d) Sun's DEC

Figure 3: Tuning hyper-parameter

that our model loses some period information. Future study may use data of longer time span to increase the accuracy of lunar location prediction.

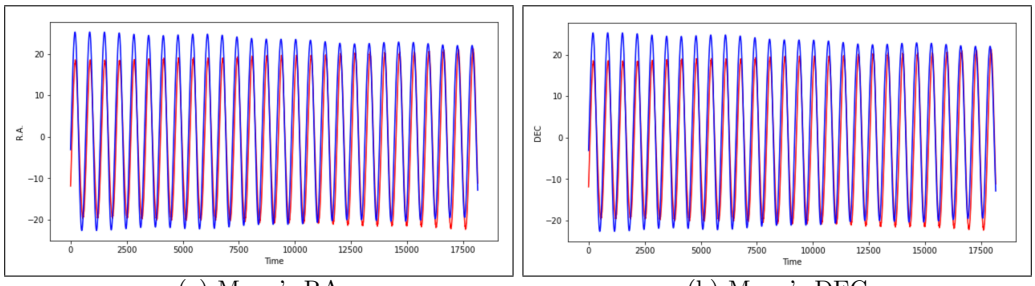
3.2 Moon phase

This subsection first presents the estimation results of the theoretical model on moon phase specified in Section 2.3.1. As shown in Table 2, we have carried out the feature engineering based on the RA and DEC of the Sun and the Moon. The DEC-related variable is delta_DEC, while there are 8 versions of RA-related variables for the variable selection. Cross validation has been adopted to select the best-performance variance. Because of the space limitation, here we only put the results from delta_RA_R5 and delta_RA_R6.

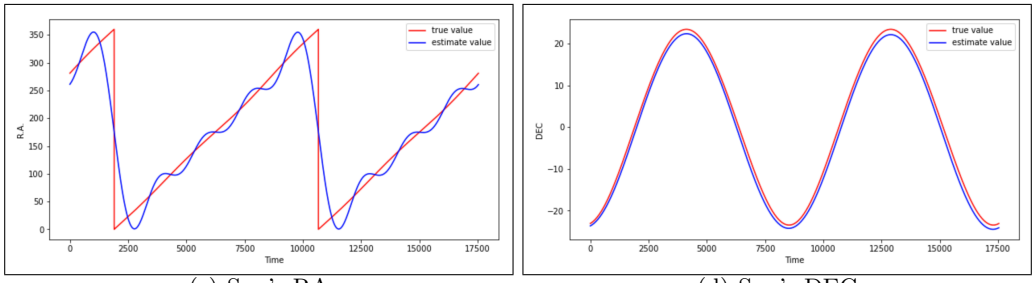
First, for OLS linear regression, we input delta_DEC with one of the RA-related variable and the model with delta_RA_R6 generates the best performance with R-squared of 0.983 and Table 3 shows the R-squared of different input feature. Therefore, model delta_DEC and delta_RA_R6 is the best. And Figure * shows the performance of the model for training and testing set, while the performance of model with delta_RA_R5 is also plotted here for comparison.

Table 3: Cross validation for feature selection in OLS model

RA-related	delta_RA	delta_-RA	delta_RA_R1	delta_RA_R2
R-squared	0.004	0.004	0.001	0.267
RA-related	delta_RA_R3	delta_RA_R4	delta_RA_R5	delta_RA_R6
R-squared	0.352	0.370	0.360	0.983

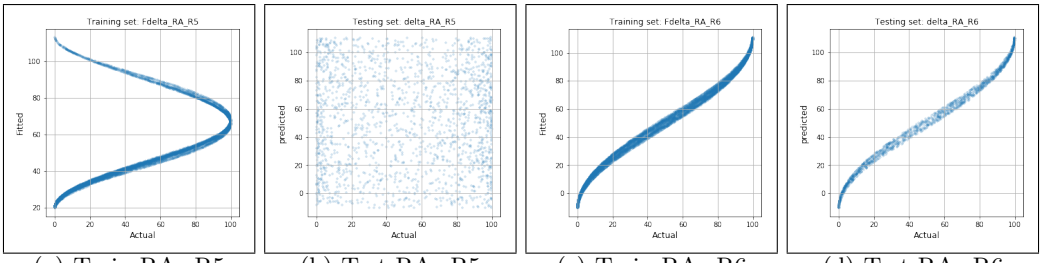


(a) Moon's RA (b) Moon's DEC



(c) Sun's RA (d) Sun's DEC

Figure 4: Estimate result



(a) Train RA_R5 (b) Test RA_R5 (c) Train RA_R6 (d) Test RA_R6

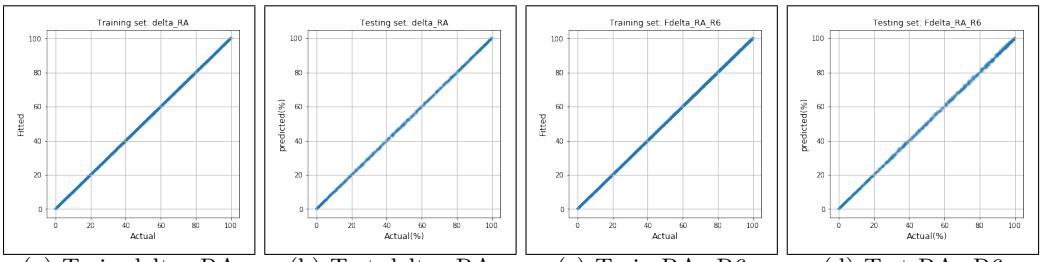
Figure 5: Performance of OLS model with delta_RA_R6

Second, we also adopted random forest to test our variables. The random forest generates much better model performance for different combination of variable input compared to OLS model as shown in Table 4. Figure * shows the performance of the model for training and testing set using variable delta_DEC and delta_RA_R6, while the performance of model with delta_RA is also plotted here for comparison.

Table 4: Cross validation for feature selection in Random Forest

RA-related	delta_RA	delta_-RA	delta_RA_R1	delta_RA_R2
R-squared	0.999	0.999	0.968	0.979
RA-related	delta_RA_R3	delta_RA_R4	delta_RA_R5	delta_RA_R6
R-squared	1.000	1.000	1.000	1.000

The result from Neural Network model showed that Adam ($loss : 3.4027e-05$) performed much better than SGD ($loss : 0.08$).The training and validation error plots are shown in Figure 7(b).

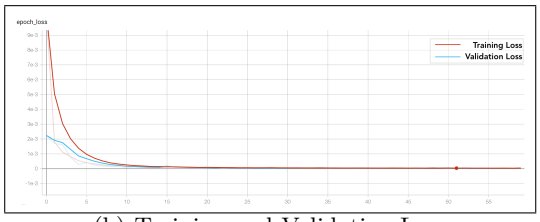


(a) Train delta_RA (b) Test delta_RA (c) Train RA_R6 (d) Test RA_R6

Figure 6: Performance of Random forest with delta_RA_R6

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 256)               2048
_________________________________________________________________
dense_1 (Dense)              (None, 128)               32896
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 129
=================================================================
Total params: 35,073
Trainable params: 35,073
Non-trainable params: 0
_________________________________________________________________
```

(a) Neural Network Structure



(b) Training and Validation Loss

Figure 7: Neural Network Structure and Training/Validation Loss

4 Conclusion

In predicting solar and lunar positions, we utilized fourier transform to build model on movement speed and obtain position with higher accuracy than Liu Hong. Solar position has higher accuracy than lunar, which might due to simpler movement pattern.

In predicting lunar illumination, Random Forest has much better performance than OLS model based on astronomy theory. Compared with Neural Network, Random Forest need much less amount of data to acquire similar precision. We brought the prediction to a higher level than ancient astronomers could ever do.

For further study, extra estimation on sun/moon eclipses could be applied. As stated before in introduction, Liu Hong has already developed method to predict eclipses. Modern astronomy defines that if sun is away from the node of ecliptic and lunar orbit more than 18°31 , then there must be no solar eclipse; 12°51 for lunar eclipse. Though due to lack of astronomy knowledge we did not perform precise predictions here, but once we know the relative positions of sun, earth and moon, we can calculate the eclipses.

Reference

[1] Christopher Cullen. (2002). The First Complete Chinese Theory of the Moon: The Innovations of LIU Hong c. A.D. 200

[2] Liu Hong. https://factpedia.org/index.php?title=%E5%88%98%E6%B4%AAvariant=zh

[3] HORIZONS Web-Interface. NASA JPL. https://ssd.jpl.nasa.gov/horizons.cgi#top

[4] 我们如何预测日食与月食? —— 从沙罗周期到精确计算. Mars Riu. https://zhuanlan.zhihu.com/p/142847801

[5] Lunar precession. Wikipedia. https://en.wikipedia.org/wiki/Lunar_precession