EECS 189 - Project S Early Deadline

Abstract

Our goal is to create a dataset of similar quality to the data available to Greek astronomers during their time. By connecting the ancient Ptolemic model of the universe to our modern understanding of Fourier series, we've created a dataset that, given modern machine learning methods, can be used to learn planetary motion models and moon phases. In an attempt to model the uncertainty of Greek instrumentation, we've introduced noise to many of our "measurements," and taken the data from geographical regions similar to where the Greeks were.

Source code can be found here: `https://github.com/pnelson6679/projectS_early_deadline_teamSKEP`

1 Introduction

Ancient Greek astronomy was one of the earliest explorations of the natural world and some may even describe it as the first science. Astronomy, nonetheless, created a foundation for modern scientific thought. Despite being false in some aspects as with the concept of Geo-centrism, the Greeks were able to make incredibly accurate predictions regarding the movement of celestial bodies like the planets and the moon. Our group's aim is to provide a data set to emulate the observations and resources Greek astronomers had when making their predictions. We also want to represent this data in a way that a machine learning model would be able to ultimately use. With this goal in mind, we created a data set that can be used to train a linear regression model with Fourier features, inspired by Ptolemy's model which follows a similar approach.

The data set we created contains recurring coordinate data for the five planets observed by the Greeks - Mercury, Venus, Mars, Jupiter and Saturn - in addition to the sun and moon. The data includes ecliptic coordinate system coordinates relative to the Medicina Radio Observatory in Italy for all celestial bodies, and the Cartesian coordinates for all celestial bodies in intervals of every five days starting from 1850. It also includes moon phases in the same time intervals. Gaussian noise has been added to any data (i.e. coordinate data) that would have required Greek tools to measure to account for lower levels of instrument accuracy for tools like the Astrolabe during that time period.

2 Methods

2.1 Planetary Motion

As mentioned in earlier sections, for our planetary position dataset we took inspiration from the Ptolemic interpretation of the solar system; this includes the particulars of the coordinate system he used (i.e. the raw data) as well as the fundamental conceptualization of planetary movements (i.e. the featurizations).

The coordinate system that many ancient astronomers used was the **geocentric ecliptic** coordinate system. This coordinate system specifies the position of planets (and local stars) relative to two quantities:

1. the ecliptic – this is the plane of Earth's orbit around the sun or in the geocentric case, the Suns orbit around the Earth.

2. the primary direction – this is the direction that is the "start" angle; convention is to use the direction from the Earth to the Sun at the vernal equinox (with some nuance to account for procession) and all angles follow given the right-hand convention.

Given the plane and the direction, we can specify each planets relative position using 3 parameters:

1. the ecliptic longitude λ – this is the angle between the primary direction and the planetary object (i.e. the "horizontal" angle)
2. the ecliptic latitude β – this measures the angle from the planetary body to the ecliptic (positive towards the north and negative towards the south)
3. the distance Δ – the direct distance to the planetary body (units dependent on use case)

Now given planetary positions in this format, we can collect a ton of raw data using a library called **Astropy**. The library allows to generate the geocentric eclipticalcoordinates for any planetary body at any point in time; as detailed in an earlier section we collected for the last 170 years (from 1850 till today). The details of this generation are detailed in the source code for Astropy. Now that we have our raw data, we can think about the featurization aspect of our dataset. To this we turn to Ptolemic model of the universe.

While we won't go into the weeds here, the Ptolemic conceptualization of the universe treats planetary motion as the combination of a deferent and epicycle (as pictured above). What do we notice about the movement and setup of this model? It looks exactly like a **Fourier series**! This allows us to leverage some of the concepts we've learned in class to featurize our raw data.

To put in more formally, the Fourier series perspective of Ptolemy's model states that given two **constant** angular velocities k_0 and k_1 representing the speed of the deferent and the epicycle respectively and given two coefficients a_0 and a_1 we can compute the position of the planet as:

$$z(t) = a_0 e^{ik_0 t} + a_1 e^{ik_1 t} \tag{1}$$

Notice that $k_i = \frac{2\pi}{T}$ for some period T. This implies that to solve for the planetary position, we need to find the right coefficients a_0 and a_1 for some Fourier feature with right fraction of 2π. This means that if we have enough Fourier **features**, we can approximately find these true values with certain learning algorithms.

More concretely, given a data point for planet p, we provide the raw parameters from above, but we also create the Fourier features:

$$\phi_t = \left\{ e^{2\pi i \frac{k}{N}} \right\}_{k=1}^{N}$$

Notice that the true signals exists in this conceptualization for a large enough N (i.e. fine-grained enough angular velocities), which implies that a well selected learning algorithm should be able to find the true Fourier features representing the epicycle and the deferent (given that there doesn't exist a better representation for the planet position than Ptolemy's) by setting all other coefficients to 0 and computing a_0 and a_1!

Thus we've created our data matrix for planet p, which includes for the tth time step (i.e. row): the raw features λ, β, Δ as well as the Fourier features ϕ_t. The true labels y_t is simply the position $z(t)$.

2.2 Moon Phases

AstroPy was able to generate phase angles of the moon for each time as well. The phase angle is a signed degree between 0° and 180°. The new moon, the phase where the moon is not visible, is at -180° and +180° and the full moon, when the moon is at its brightest, is at 0°. This is to account for whether each phase is waxing or waning. Accordingly, the first quarter (when the moon is waning) is at a -90° phase angle and the third quarter is at a +90° phase angle. [1]

Although the ancient Greeks did not have access to phase angles, this level of precision was what made the most sense for our dataset because we gave the positions of all the celestial bodies at a specific time. It would not make sense to give the categorical phase of the moon at that time because it would not align with the rest of our data.

3 Results

Our completed dataset has 46 columns, describing the time of each data point along with the coordinates for each planet in the geocentric ecliptic coordinate system according to the λ, β, δ as described above and the appropriate Cartesian transformation of the coordinates given by the xyz plane.

The columns for angles are given in degrees, with the exception of the moon phase, which is specified in The angles for distance are specified in astronomical units (AU), such that $1\text{AU} = 1.496 \times 10^8\text{km}$, or the average Sun-Earth distance.

We were also able to generate the appropriate phase angles for the moon at the same times for the other celestial bodies.

4 Conclusion

Through this study, we sought to act as the torchbearers of Ptolemy's legacy by providing the tools for present and future scholars to practice their understanding of the heavens through machine learning. Ptolemy's model of planetary motion serves as a bridge connecting past Greek astronomers' interpretation of the universe with present practitioners' comprehension of Fourier series. By providing this dataset, we hope to increase people's exposure to the historical context of modern astronomy. This way, we can honor the accomplishments of past scientists.

For users planning to utilize this data set, please be aware that the data has not been divided into training/validation/test/other splits of any kind. We leave split designation up to user discretion. The size is approximately 9 MB for users concerned with size.

Bibliography

[1] P. Kenneth Seidelmann. *Explanatory Supplement to the Astronomical Almanac: A Revision to the Explanatory Supplement to the Astronomical Ephemeris and the American Ephemeris and Nautical Almanac*. University Science Books, 2005.

[2] Ecliptic Coordinate System. Ecliptic coordinate system — Wikipedia, the free encyclopedia, 2020. [Online; accessed 27-October-2020].

[3] Eccentrics, deferents, epicycles, and equants. `https://www.mathpages.com/home/kmath639/kmath639.htm`.