Qeios

Peer Review

Review of: "Kepler: The Pioneer of Data Science and AI"

Tony Hey[1]

1. Scientific Computing Department, Rutherford Appleton Laboratory, Didcot, United Kingdom

Review of Commentary by Ronaldo Mota

Kepler: The Pioneer of Data Science and AI

In February 2016, I attended the 'Galileo Symposium' at the University of Oklahoma and gave the keynote presentation. My title was 'Galileo, New Technology and Big Data' and included a discussion about Kepler's data analysis of Tycho Brahe's most accurate - pre-telescope – observations and measurements of the motions of the planets. It was the unprecedented accuracy of this data that forced Kepler to realise that the orbit of Mars was an ellipse and could not be fitted with Copernicus's circular orbits and epicycles. In his book describing his data analysis, *Astronomia Nova* (1609), Kepler writes:

' ... if I had believed we could ignore these eight [arc] minutes, I would have patched up my hypothesis accordingly. But since it was not permissible to ignore them, these eight [arc] minutes point the road to a complete reformation of astronomy ...'

It was the accuracy of Brahe's data that allowed Kepler to formulate his three Laws of Planetary Motion, which was the first time that planetary motion had been expressed in a precise mathematical form. Kepler eventually published Brahe's data, corrected to reflect Kepler's heliocentric view of the solar system, as the *Rudolphine Astronomical Tables* in 1627, some 26 years after Brahe's death. Together with Galileo's experiments on motion and acceleration, Kepler's work as a data scientist on planetary motions laid the foundation both for Newton's laws of motion and for the inverse-square law of gravity in his *Principia Mathematica* (1687).

My talk then went on to describe several 21st century experiments utilising 'New Technology and Big Data', including the ground-breaking Sloan Digital Sky Survey and its new generation of data scientists.

doi.org/10.32388/8OD7NM

The preamble given above is to explain why I was very excited to see the short paper by Ronaldo Mota titled *'Kepler: The Pioneer of Data Science and AI'*. Mota's paper summarises Kepler's work as a data scientist working on Tycho Brahe's astronomical data as above but then raises the question:

'Could an AI system, given Brahe's data, replicate Kepler's discoveries'

Mota answers this question in the affirmative and claims that this demonstrates that AI and machine learning can enable automated model discovery from vast datasets. He cites two papers successfully using machine learning methods to derive Kepler's Laws, 'Combining data and theory for derivable scientific discovery with AI-Descartes' [2] and 'From Kepler to Newton: Explainable AI for Science' [3]. The 'AI-Descartes' paper is interesting in that their system combines logical reasoning with symbolic regression to re-discover Kepler's third law. By contrast, 'The Explainable AI for Science' paper uses a neural network model as a black-box model for data fitting and augmentation. The resulting neural network function is then transformed into a human-understandable function using symbolic regression. The authors are then able to confirm that the orbit of Mars is an ellipse with the Sun at its focus - as in Kepler's first law.

The paper concludes with a look to the future. Mota believes that symbolic regression will be an increasingly important machine learning approach to extract interpretable mathematical expressions directly from datasets. He also highlights the integration of Kolmogorov-Arnold Networks (KANs) with symbolic regression as a particularly promising research area. As a final comment, Mota emphasises that human insight and experimental verification will remain a vital component of the AI for Science revolution.

All in all, this is a stimulating paper on AI for Science and one that pays homage to Kepler as an early visionary data scientist.

References

[1] Hey, Tony, invited talk at the 'Galileo Symposium' held at the University of Oklahoma, 25th February, 2016.

[2] Cornelio C, *et al*. 'Combining data and theory for derivable scientific discovery with AI-Descartes', Nature Communications 14:1477 (2023), doi:10.1038/s41467-023-37236-y

[3] Zelong Li, Jianchao Ji, Yongfeng Zhang, 'From Kepler to Newton: Explainable AI for Science', 2nd AI4Science Workshop at the 39th International Conference on Machine Learning (ICML), (2022), arXiv, doi:10.48550/arXiv.2111.12210

Declarations

Potential competing interests: No potential competing interests to declare.