“PoET: design and implementation of collaborative machine learning”

Christopher Lo
Department of Economics

Chris Chia
Department of Mathematics

This research paper aims to address the issue of *information asymmetry* in the market for machine learning and proposes a novel solution in the form of the PoET (Proof of Ensemble Training) blockchain protocol.

Machine learning provides productivity gains to prediction heavy industries such as insurance and healthcare. Predictive performance depends both on the model and parameters selected, and access to large and diverse datasets to prevent overfitting. However, data sharing across firms to generate these more diverse datasets is limited due to privacy concerns, and collaboration between firms to select more representative machine learning models is also constrained due to the potential for free-riding. This leads to information asymmetry and centralisation of ML and data.

We propose a solution to the privacy and free-rider problem using an interdisciplinary approach, blending concepts from economics and computer science.

We first propose an economic surplus-sharing mechanism to solve the privacy and free-rider problem in a decentralised ML context. We then introduce a novel blockchain consensus protocol, PoET (Proof of Ensemble Training): every block in the network represents a collaborative attempt (ensemble) to improve performance on a specific machine learning problem by aggregating individual contributions. This protocol ensures that the blockchain network enforces both the rules of the proposed mechanism and the consistency of the information between firms. Lastly, we provide an interpretation of each network as a tradable financial asset: the price of a network is the present value of future cash flows generated by the machine learning performance improvements.