Can Explainable AI Replace AI?

A feature of statistical approaches to predictive algorithms, such as AI and Machine Learning (AIML) is that it can be difficult to explain the rationale behind any one particular prediction. Explainable AI refers to the collection a of tools that data scientists can use to dilute this "black box" nature of their models with a view towards improving user understanding of the models and thus increasing user adoption.

There is an additional benefit to Explainable AI which appears in a recent Nature publication (see also here) by celebrated Australian mathematician Geordie Williamson and his colleagues, where they have reported results arising from a collaboration between pure mathematicians and Machine Learning (ML) researchers at Googles DeepMind.

In this work, Williamson et. al. used Explainable AI to explore possible connections between different areas of mathematics and to then formulate several conjectures. Then armed with intuition garnered from working with the Explainable AI algorithms, they managed to prove these conjectures using purely rules-based, non-statistical mathematical methods. So, while the AI was crucial to the exploratory step where they synthesised their thinking, their final product could in principle have been published without any reference to AI at all!

This is an interesting workflow which is familiar from my work in industrial applications of AI. At Fujitsu Data & AI we often refer to the initial exploratory stages of a data science project as "searching for the signal". This typically involves first building simple ML models for predictions which are already understood perfectly well using existing methods. Then once we are satisfied that the data contains a signal, we move on to more ambitious models whose predictions are not currently performed by existing rules-based methods. Finally, using techniques from Explainable AI such as Shapley values we look to find the essence of the prediction algorithm, keeping an open mind that we may be able to find a far simpler rules-based model to replace the ML model.

For example, in an industrial preventative maintenance project, we may have thousands of sensor channels and using a complicated ML model, we may demonstrate that we can predict when the system is approaching an abnormal operating state. However, the Explainable AI algorithm may show us that in fact there are only two critical sensors which are needed for the prediction; we can subsequently dispense with the AI model and just monitor these two sensors with a rule's base algorithm.

While the AI inspired mathematical proofs of Williamson et al are a far cry from industrial data science, we think this work provides an important example where Explainable AI replaced the AI itself with rules based methods. It leads to a deeper understanding of the workflow of Explainable AI which in turn will lead to improved business value from AIML data science projects in a host of industries.

If your business needs help with AIML projects, please contact a Fujitsu Data & AI specialist now.

Contact

Fujitsu Data & AI

+61 3 9924 3000