Transforming classical artificial intelligence with modern machine learning approaches: The case of the BACON system for equation discovery from scientific data

Jonah Miller [1], Soumya Banerjee [1]&

1 University of Cambridge, Cambridge, United Kingdom

& Corresponding author sb2333@cam.ac.uk

Abstract

`BACON` is a heuristic-based computational scientific discovery system (designed by Pat Langley), which aims to find invariants in multivariable systems [1]. We rebuilt `BACON` in a modern computing language, and we improve the noise-resilience of `BACON`. We demonstrate how such classical AI systems can be understandable, yet powerful.

Our `BACON` also outperformed `PySR` - a modern method utilising symbolic regression on a neural network [2] - conclusively in specific environments on small datasets.

`BACON` was also successfully overlain on a Monte Carlo Tree Search framework.

In 1997 the first neural network designed to solve this exact problem [3] was created. We suggest that there is potential in these forgotten approaches that modern deep learning systems can learn from. Integrative approaches that combine heuristic approaches like `BACON` with modern deep learning can be very helpful. We suggest integrating modern deep learning approaches with heuristic-based approaches as a way to analyse large scientific datasets.

Summary and conclusions

For smaller datasets, `BACON` *consistently thrives whilst* `PySR` *struggles.* `BACON` is made for this environment, whereas `PySR` is made to overfit on complicated, large datasets whilst applying their biases towards simplicity. When approaching this threshold, it is `PySR` that thrives whilst `BACON` suffers (see experiments in the Appendix on Black's Law). Here is an - albeit niche - situation where classical methods outperform modern techniques. It amplifies the need to reproduce, study and understand these seemingly anachronistic mechanisms and see what lessons can be taken going forward.

We applied this framework to a number of exemplar problems in physics and mathematics. Our results suggest that `BACON` is good at reducing noise and inferring the correct equation in smaller datasets, whereas `PySR` is significantly more successful on larger, noisier, datasets [4].

We release open-source code which will aid further development: `https://github.com/JonahMiller/BACON`.

The broad goal of this research project was to combine modern approaches to AI with the classical. Both have strengths which, when efficiently combined, could lead to refined systems, able to analyse large datasets effectively. We suggest that in the future, combining large-language models with classical AI approaches such as those presented here may help solve more complex scientific and mathematical problems.

References

1. Langley P (1977) BACON: A production system that discovers empirical laws. In: Proceedings of the Fifth International Joint Conference on Artificial Intelligence.

2. Cranmer M (2023) Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl. In: arXiv:2305.01582.

3. Saito K, Nakano R (1997) Law discovery using neural networks. In: Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence.

4. Miller J, soumya banerjee (2024) The bacon system for equation discovery from scientific data: Transforming classical artificial intelligence with modern machine learning approaches. In: AAAI 2024 Fall Symposium Series, Integrated Approaches to Computational Scientific Discovery. doi: 10.31219/OSF.IO/Z8KQV. URL `https://osf.io/z8kqv`.