

Work done here

Google Research

RELAY[®] Curr THERAPEUTICS

Currently here

SyFES: Symbolic regression for finding DFT functionals

IPAM Workshop: Learning and Emergence in Molecular Systems

January 2023



Application of highly parallelizable symbolic regression system for finding complex expressions for functionals in Density Functional Theory (DFT) demonstrates useful power of symbolic regression on a real scientific problem.

- Problem of symbolic regression and our approach
 DFT and symbolic regression problem setup
 Results demonstrating effectiveness of the system
 How is the GAS22 functional different?
- **5. Reflections**

Ma H, Narayanaswamy A, Riley P, Li L. Evolving symbolic density functionals. Sci Adv. 2022;8: eabq0279.

Symbolic Regression

Problem of symbolic regression and our approach

What is Symbolic Regression?

$$t=rac{r_A+r_O}{\sqrt{2}(r_B+r_O)}$$
 .

VS.

Small models with few parameters are often represented as equations Symbolic Regression refers to a class of machine learning techniques to find these small compact expressions

7x7 conv, 64, /2 pool, /2 3x3 conv, 64 ¥ 3x3 conv, 64 3x3 conv, 64 ¥ 3x3 conv, 64 3x3 conv, 64 3x3 conv, 64 ****** 3x3 conv, 128, /2 3x3 conv, 128 3x3 conv, 128 3x3 conv, 128 3x3 conv, 128 ¥ 3x3 conv, 128 3x3 conv, 128 3x3 conv, 128 3x3 conv, 256, /2 3x3 conv, 256 3x3 conv, 512, /2 3x3 conv, 512 3x3 conv, 512 3x3 conv, 512 3x3 conv, 512 3x3 conv, 512





Mathematical Expressions as Computer Programs





Inspired by: Real E, Liang C, So DR, Le QV. AutoML-Zero: Evolving Machine Learning Algorithms From Scratch. arXiv: 2003.03384. 2020.

One key advantage compared to tree based representations is to have non-functional parts of the program that can be mutated later.

Allowed operations



Name	Symbolic form	Notes
Addition	$F = v_0 + v_1$	
Subtraction	$F = v_0 - v_1$	
Multiplication	$F = v_0 * v_1$	
Division	$F = v_0/v_1$	
AddMultiply	$F += v_o * v_1$	Common form in programming languages
Power	$F = v_0^n, n \in \{2, 3, 4, 6, \frac{1}{2}, \frac{1}{3}\}$	
Finite domain transform	$F = \gamma v_0 / (1 + \gamma v_0)$	γ is a parameter, this is inspired by forms in existing functionals



Operation	Example before	Example after
Insert new instruction	$F = c_0 + F$	$F = c_0 + F$
		$v_0 = c_1 * v_0$
Removal of instruction	$F = c_0 + F$	$F = c_0 + F$
	$v_0 = c_1 * v_0$	
Change operation	$v_1 = v_0 + c_0$	$v_1 = v_0/c_0$
	$v_1 = v_0 + c_0$	$v_1 = v_0^2$
Change argument	$F += c_2 * v_1$	$F += v_o * v_1$

Parameter Optimization





These are constant parameters; how to pick their values?

CMA-ES to minimize error on the training set 5-10 repeated attempts to minimize Initialization from a unit Gaussian Limit parameters to [-10, 10]





Illustration of CMA-ES on two 2D problems from Ha D. A Visual Guide to evolution strategies. blog otoro net. 2017. Available: <u>https://blog.otoro.net/2017/10/29/vis</u> ual-evolution-strategies/



Oldest, not least fit are removed

Random subset put into "tournament" to decide who reproduces





are numerically equivalent.

Density Functional Theory

DFT and symbolic regression problem setup





DFT is Everywhere



- Nvidia estimates that 15% of world's supercomputer time is devoted to DFT calculations.
- Over 40k papers published per year.
- A new family of near-room-temperature superconductors is predicted by DFT. (Phys. Rev. Lett. 125, 217001)
- Chemistry and material datasets:
 - Open Catalyst Project
 - Materials Project
 - QM9
 - •



The Nobel Prize in Chemistry 1998 Walter Kohn "for his development of the density-functional theory"



A Pribram-Jones, DA Gross, K Burke - Annual Review of Physical Chemistry, 2015

Meta-GGA Form of our Functional



14





Constructed to produce the ω B97M-V functional

84 datasets with "4986 data points (and requiring 5931 single- point calculations)"

Categorized into

- NCED (non-covalent dimers (easy))
- NCEC (non-covalent clusters (easy))
- NCD (non-covalent dimers (difficult))
- IE (isomerization energies (easy))
- ID (isomerization energies (difficult))
- TCE (thermochemistry (easy))
- TCD (thermochemistry (difficult))
- BH (barrier heights)

The authors carefully defined train/test splits and weighting among datasets

• We use their definitions

Mardirossian N, Head-Gordon M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. J Chem Phys. 2016;144: 214110.

Results

Demonstrating the effectiveness of the system





Discovering a New Functional: Improving ωB97M-V





 $F_{C-OS}^{\,\omega B97M-V}$ has $c_{00}, c_{10}, c_{20}, c_{60}, c_{21}, c_{61}$

Discovering a New Functional: Results





Discovering a New Functional: Results





Technical note: Symbolic regression step is *not* done with Self Consistent Field (SCF) calculations, but we validated the final functionals with SCF calculations and differences are minimal as you can see above

GAS22

How is the GAS22 functional different?



General form of
$$\omega$$
B97M-V
 $F^{\omega B97M-V} = \sum_{i} \sum_{j} c_{ij} w^{i} u^{j}$

 $F_{\rm x} = 0.862 + 0.937u + 0.318w$

Very similar to starting value

$$F_{\rm c-ss} = u - 4.108w - 5.242w^2 - 1.766u^6 + 7.538w^4u^6$$

Similar form, but different set of powers for the power series

$$F_{\rm c-os} = 0.805 + 7.989 \,w^2 - 7.548 \,w^6 + 2.001 \,w^6 \sqrt[3]{x^2} - 1.761 \,w^2 \sqrt[3]{x^2}$$

These terms are new, not seen in other functionals



Reflections



Fumans Define the datasets and objective function Define the evaluation of experimental results that would be convincing Constrained the part of the functional that would be modified Constrained the search space of that part of

Constrained the search space of that part of the functional (note the AddMultiply and FiniteDomainTransform instructions and parameter constraints)

Set up parameters of the evolutionary search algorithm



Machines

Proposed changes to existing equations

Weighted random selections towards functionals with good performance

Evolutionary algorithms in particular tend to be quite specialized to an application This is somewhat true with neural net applications as well



Deep Blue Much of value comes from brute force tree search



VS.

Current symbolic regression feels much more like this

AlphaGo Learned functions for evaluating board positions directs search procedure



70 hours

AlphaGo Zero plays at super-human level. The game is disciplined and involves multiple challenges across the board.

https://en.wikipedia.org/wiki/Deep_Blue_(chess_computer)#/media/File:Deep_Blue.jpg



- The symbolic regression search did not evaluate SCF convergence, but the GAS22 functional does converge.
 - Did we just get lucky? Would other similar searches still work in this way?
- Is symbolic regression actually a good proxy problem for "learning emergent concepts"?
 - Pure symbol manipulation with no meaning behind it. Is that enough?
- How important were the various tweaks to the available operators and evolutionary algorithm? Is this result robust with respect to other reasonable algorithmic choices?
 - Brief aside: This question of result robustness to other reasonable analytic choices is really important! Two papers from social science I recommend reading:
 - Silberzahn, R., et al. "Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results." Advances in Methods and Practices in Psychological Science, vol. 1, no. 3, Sept. 2018, pp. 337–56.
 - Schweinsberg, Martin, et al. "Same Data, Different Conclusions: Radical Dispersion in Empirical Results When Independent Analysts Operationalize and Test the Same Hypothesis." Organizational Behavior and Human Decision Processes, vol. 165, July 2021, pp. 228–49.

Questions?



Thank you to my co-authors who did all the hard work!



(Google)

He Ma



Li Li (now at Meta)



Arunachalam Narayanaswamy (now at Ghost)



