

Symbolic Regression 振り返り

Yoshihiro Michishita(Riken CEMS / Proxima Technology)

Outline

- **SR with genetic algorithm/ SINDy**
- **AI-Feynmann / Deep Symbolic Regression**
- **Symbolic Physics Learner**


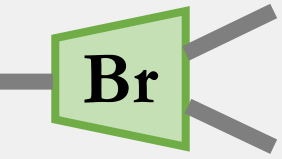
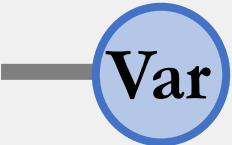
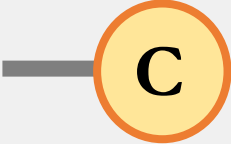
➤ Genetic Algorithm (genetic programming)

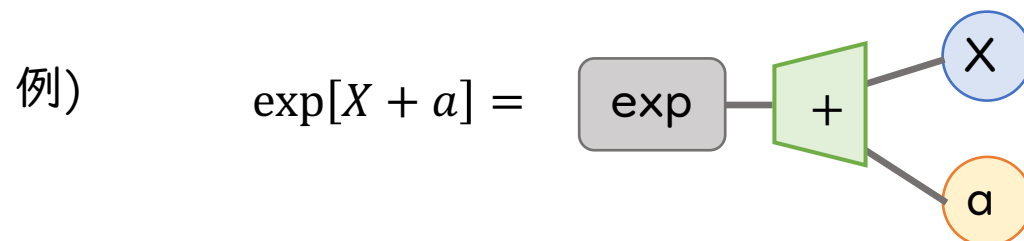
生物の進化をなぞったようなアルゴリズム

1. N個の個体をランダムに生成
2. 評価関数について上位 n 個体に対し、「複製」「交配」「突然変異」をN個になるまで実行
3. 2を繰り返し、世代を進める。G世代後もしくは評価値が一定以上の個体が出てきたら終了。

➤ Symbolic Regression

Nodeとして“branch(operator)”, “function”, “variable”, “constant”を用意し、tree graphで数式を表現。

Node type	Function	Branch	Variable	Constant
Graph				
Examples	$\exp, \int dt, \frac{\partial}{\partial x}$	$+, -, \times, -i[], \{, \}$	x, ψ_i, H	a, C



➤ Genetic Algorithm (genetic programming)

サンプルのダイナミクスを与え、それに近いダイナミクスを与える式を優秀な種(高得点)とする。

最初にランダムにtreeを生成し、以下を(ランダムに)実行。

- ・ 淘汰... 優秀な種(tree graphのみ一定数残し、後を消去)
- ・ 複製... 優秀な種を複製
- ・ 交配... 優秀な種同士の枝を交換する
- ・ 突然変異... 優秀な種のノードを一部別のものに変異させる

Discovering governing equations from data by sparse identification of nonlinear dynamical systems

Steven L. Brunton^{a,1}, Joshua L. Proctor^b, and J. Nathan Kutz^c

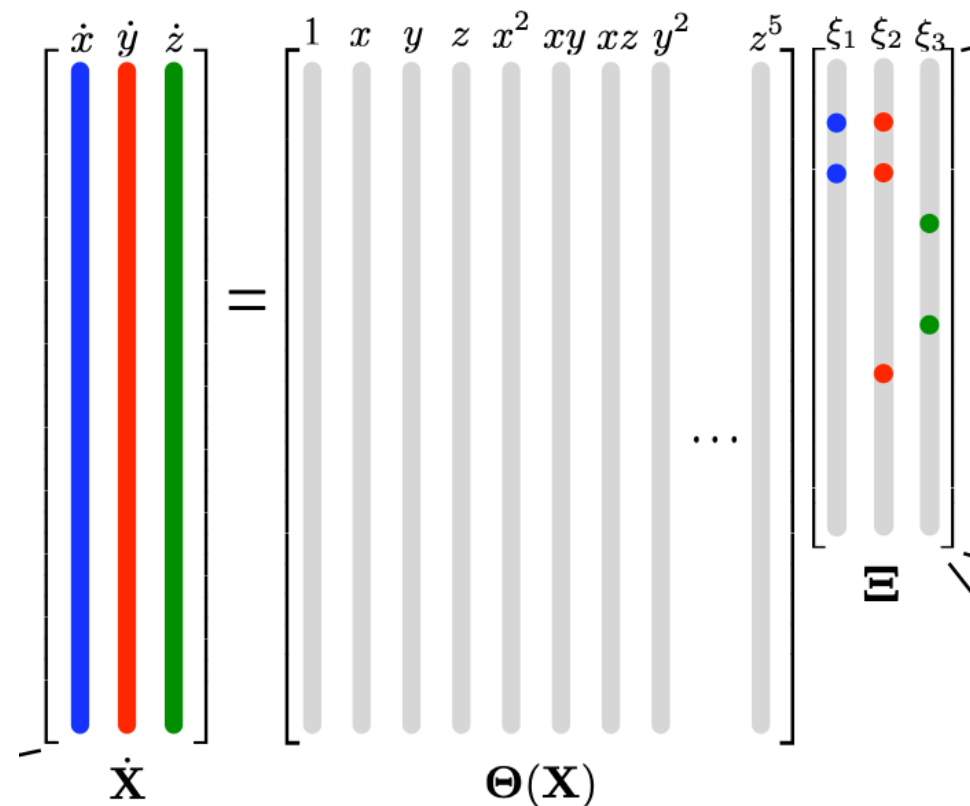
^aDepartment of Mechanical Engineering, University of Washington, Seattle, WA 98195; ^bInstitute for Disease Modeling, Bellevue, WA 98005; and ^cDepartment of Applied Mathematics, University of Washington, Seattle, WA 98195

PNAS : 10.1073/ 1517384113(2016)

与えられた力学系ダイナミクス

と変数の関数基底で線形回帰

(LASSOでSparseな解を同定)



	SR with GA	SINDy
良い点	簡単	簡単
悪い点	収束の保証なし 解が安定しない	関数基底による 変数が増えると大変 綺麗にsparseになるの？



SR with Reinforcement learning

➤ 問題の分割手法

1. 何かしらの方法で方程式を生成
2. 物理定数が合うように要請・変数を無次元化
(ver2では変数の微分の性質を使ってグラフを分割して簡略化)
3. それぞれ変数・定数をフィッティング

関数をNNで近似したのちに、
変数の微分の性質から小さいグラフに
分割して、総当たりで計算(本当に出来る?)
(ソースは公開されているが動かない。。)

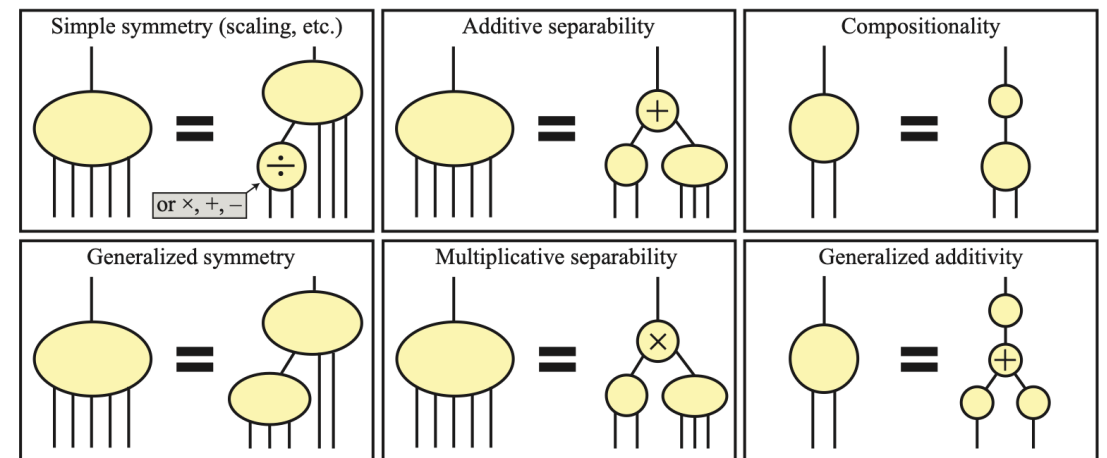
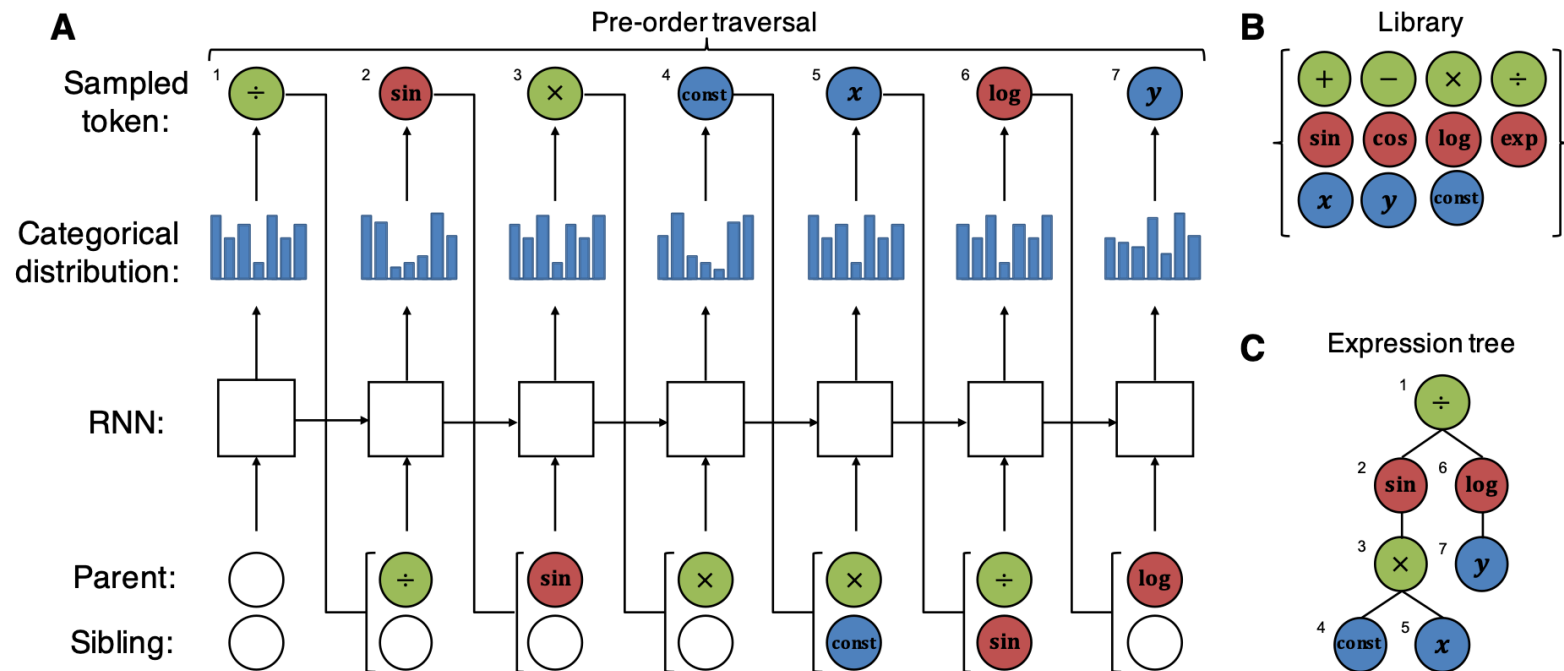


Figure 3: Examples of graph modularity that our algorithm can auto-discover. Lines denote real-valued variables and ovals denote functions, with larger ones being more complex.

➤ RNNでgraphを生成し、強化学習手法(REINFORCE)で学習

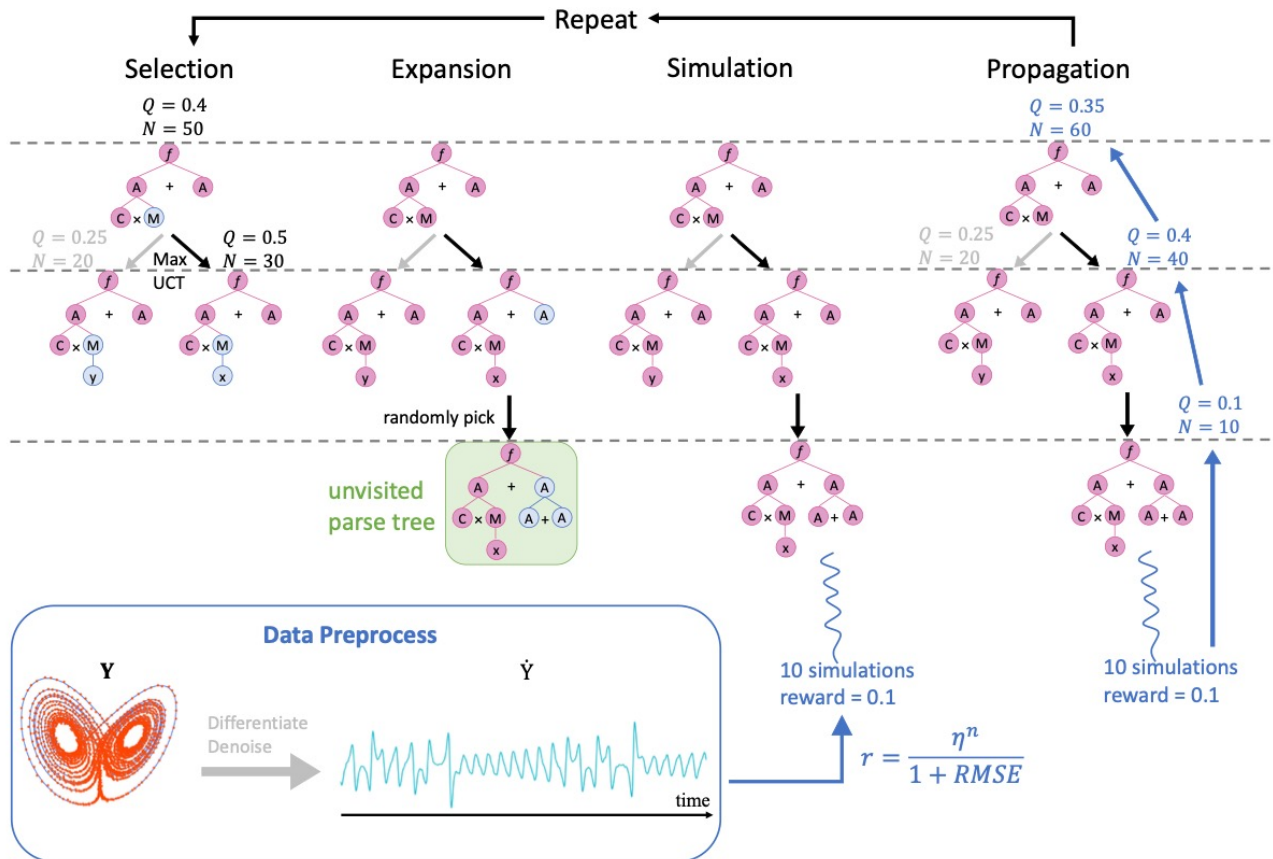
ダイナミクスの一貫性をスコアとして計算・学習



*RNNなので学習安定しなさそう

➤ MCTSで記号回帰の探索

arXiv:2205.13134



*GAの名残か植え替え(交配)的な操作を残している
ので、「環境が時間変化する」事に相当する。
ある種UCB(UCT)の保証をぶち壊す操作をしている。
(摂動解など似たようなパターンが複数回現れる
場合は強そう。ソースは公開されていない)

●まとめ

- これまでのSRは基本的に「未知の非線形ダイナミクスの方程式」を当てる
 - ・理論屋の道具という感じではない
 - ・アルゴリズムが少し古い(REINFORCE, GA, MCTS(+GA))
 - ・探索空間の関数系や変数に困るかも
- 理論屋として面白いのは「有効理論の構築」にSRを使う
 - ・オリジナルの模型に出てくる演算子等を使えば良い
 - ・式が分かれば、理論屋なら解釈出来るはず。(人間の仕事が残っている)
 - ・新しい物理的描像・物性探索。