

アンサンブルとアテンションによる 遺伝的プログラミングの線形性の向上

加藤 慎二† 長尾 智晴‡

横浜国立大学大学院環境情報学府†

横浜国立大学大学院環境情報研究院‡

1 はじめに

進化計算の一つである遺伝的プログラミング (Genetic Programming; GP) [1]は、基礎関数の組み合わせを最適化することで、任意の合成関数を生成する手法である。GP で生成される関数は、基礎関数の組み合わせで表現されるため、解析性が高い。しかし、精度向上のために三角関数や指数関数などの非線形性の強い基礎関数を用いると解析性が悪化する。そこで本研究では、アンサンブル学習とアテンション機構を GP に用いることで、非線形性の強い基礎関数を用いなくても精度を維持できることを示す。

2 提案手法

2.1 解表現

提案手法の解表現は出力値を計算する N 個の出力木構造と、それに対応してマスクを計算する N 個のマスク木構造から構成される。マスク木構造の出力値はルートノードの値が 0.5 以上のとき True, それより小さいとき False を出力する。そして、マスク木構造の出力が True のときに対応する出力木構造の出力を、False のときは 0 を出力する。さらに、この処理を全ての木構造で行なった総和を解の出力値とする。図 1 に提案手法の解表現を示す。

2.2 遺伝子操作

提案手法では、交叉・突然変異の遺伝子操作を次のように定義する。

● 交叉

交叉では、2 つの解の間で各木構造同士の部分木交換交叉を行う。図 2 に提案手法の交叉方法を示す。このとき、各木構造に対して一定の確率で部分木交換が実行される。提案手法では各木構造の順序は常に同一であり、独立かつ同時に最適化される。

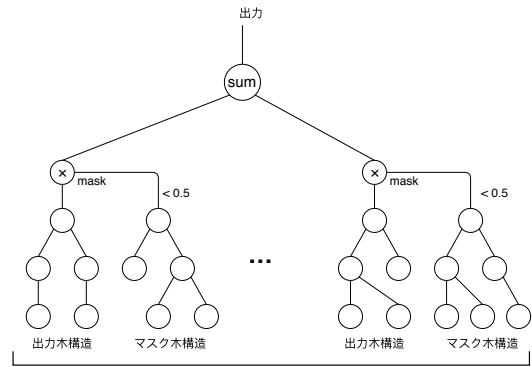


図 1 提案手法の解表現

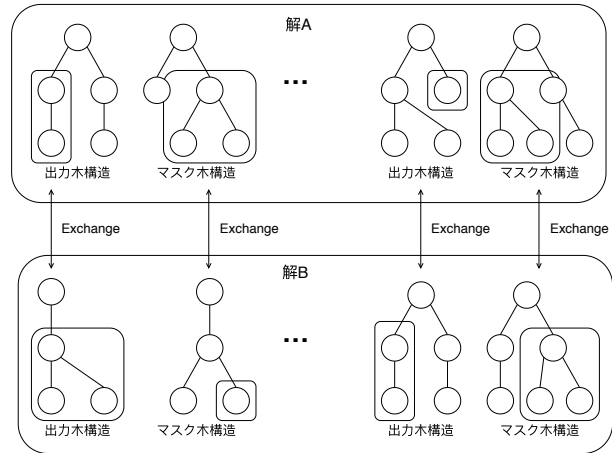


図 2 提案手法の交叉

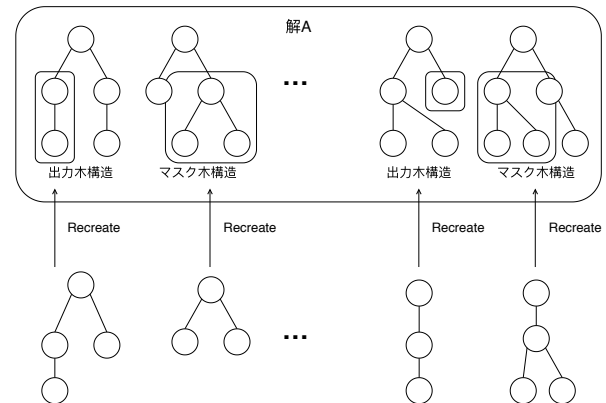


図 3 提案手法の突然変異

Ensemble and Attention Method for Improvement of Linearity in Genetic Programming

† Shinji Kato, Graduate School of Environment and Information Sciences, Yokohama National University

‡ Tomoharu Nagao, Faculty of Environment and Information Sciences, Yokohama National University

● 突然変異

突然変異では、解の各木構造からランダムに選ばれた部分木を再生成する。図3に提案手法の交叉方法を示す。このとき、各木構造に対して一定の確率で突然変異は実行する。

3 実験内容

実験では、非線形基礎関数を用いた GP (Simple Genetic Programming; SGP) と、非線形基礎関数を用いない提案手法による GP (Ensemble and Attention Genetic Programming; EAGP) に対して、Boston と Diabetes の回帰問題を適用し、提案手法の解析性と回帰精度を検証した。

回帰問題の前処理は Minmax 正規化を行い、学習データとテストデータは 7:3 の割合で分割した。また、本実験で用いた GP の設定を表1に示す。

4 実験結果

Boston と Diabetes の回帰問題に対して SGP と EAGP を適用し、テストデータの回帰誤差を検証した。このとき、木の深さと EAGP の木の数が精度に与える影響を調査するために、各パラメータに対して 10 試行実験した。

図4に Boston 回帰問題における SGP と EAGP の回帰誤差を示す。SGP は深さ 9 が最も良い精度であり、EAGP は深さ 12、木の数 3 が最も良い精度である。

図5に Diabetes の回帰問題における SGP と EAGP の回帰誤差を示す。SGP は深さ 9 が最も良い精度であり、EAGP は深さ 3、木の数 4 が最も良い精度である。

実験の結果より、提案した EAGP は非線形性が強い基礎関数を用いていないにも関わらず、非線形性が強い基礎関数を用いた SGP と同程度の精度であることがわかる。とくに、Diabetes 問題では深さが 3 の小さな木構造で構成された EAGP は SGP よりも良い精度であった。

表1 GP の実験設定

name	parameters
population size	500
max generation	200
depth	3, 6, 9, 12, 15, 18
EAGP tree num	2, 3, 4, 5
crossover rate	0.8
mutation rate	0.2
selection	Batch Tournament[2]
SGP function nodes	$+$, $-$, $*$, $/$, \sin , \cos , \exp , \log x_i , ERC
EAGP function nodes	$+$, $-$, $*$ x_i , ERC
ERC range	$[-1, 1]$

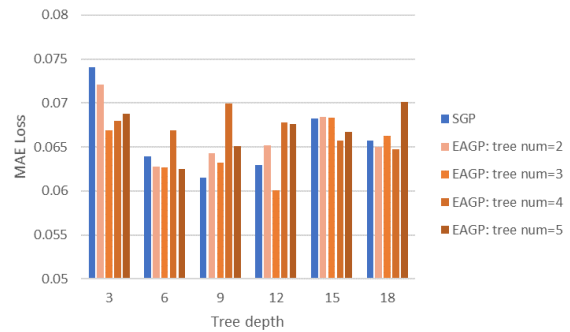


図4 Boston dataset の回帰誤差

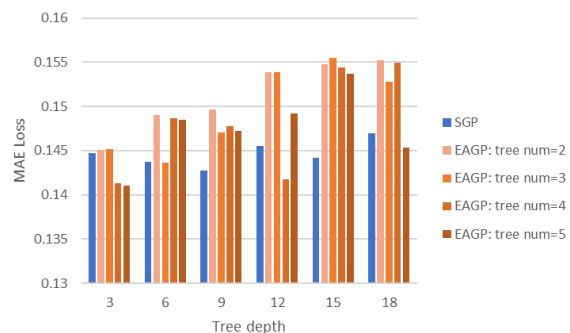


図5 Diabetes dataset の回帰誤差

5 おわりに

本研究ではアンサンブル学習とアテンション機構を用いることで線形性を向上した遺伝的プログラミングを提案した。提案手法による GP の木構造は小さな木構造と線形基礎関数によって構築されているため解析性に優れている。実験では非線形基礎関数を用いた通常の GP と比較し、同等の精度であることを確認した。これより提案手法は精度を維持したまま解析性を向上することができる。今後の課題としては、他の問題への適用による汎用性の検証や、Cartesian Genetic Programming[3]などを用いた簡潔な解表現方法の提案などが挙げられる。

参考文献

[1] Koza, John R. "Genetic programming: on the programming of computers by means of natural selection." Vol. 1. MIT press, 1992.

[2] Melo, Vinicius V., Danilo Vasconcellos Vargas, and Wolfgang Banzhaf. "Batch Tournament Selection for Genetic Programming." , arXiv:1904.08658. 2019).

[3] Miller, Julian Francis, and Simon L. Harding. "Cartesian genetic programming." Proceedings of the 10th annual conference companion on Genetic and evolutionary computation. ACM, 2008.