

# 遺伝的プログラミングを用いた 説明性を有する予測モデルの構築

阿部 有希<sup>†</sup>

<sup>†</sup> 横浜国立大学 理工学部

長尾 智晴<sup>††</sup>

<sup>††</sup> 横浜国立大学 大学院環境情報研究院

## 1. はじめに

近年、人工知能研究の中でもニューラルネットワーク (Neural Network; NN) を利用した機械学習手法の発展は目覚ましい。特に、ディープラーニングは一般的にコンピュータには高難度である画像認識のタスクでも人間に匹敵する精度を記録している。しかし、NN を用いたモデルはブラックボックスであり、実運用段階や意思決定に責任を伴う現場では用途が限定されていることが現状である。そこで、判断根拠の説明が可能な「説明できる AI」の研究が盛んに行われている。本論文では、判断根拠を説明しやすい簡易なモデルを利用し遺伝的アルゴリズム (Genetic Algorithm; GA) [1] を用いて最適化することで、ニューラルネットワークの出力を近似する手法を提案する。

## 2. 提案手法

本稿では、二分木を用いた回帰モデルによって、ニューラルネットワークの出力を近似する手法を提案する。二分木の最下層である葉ノードを入力変数、最上層である根ノードを出力として、葉ノードの二入力に対して簡単な演算 (加算, 減算, 乗算, 平均) を行い、その演算結果を根ノードの入力とする。これを繰り返すことで、最終的な出力を計算する。最下層の入力にどの入力変数を用いるか、また二入力に対してどの演算を行うかを遺伝子型として定義し、GA を用いて最適化させることで、解探索を行う。さらに、演算の中に右部分木, 左部分木の入力のみを演算結果とする演算を入れることで、モデルの構造単純化を図る。具体的な表現型, 遺伝子型の例を図 1 に示す。図中の「R」で表された演算は、左部分木の入力のみを演算結果とする演算である。

## 3. 実験

### 3.1. 実験設定

データセットは、Kaggle Kernel の Titanic (タイタニック号) の乗客データを用いた生存者予測) を利用し、NN を用いた機械学習モデル, および提案手法である説明モデルを用いて学習させた。NN の隠れ層の個数は 3 とした。入力変数の個数は 5 個 (Pclass, Sex, Age, Fare, Embarked) とし、これらを用いて乗客の生存確率を推定した。その後、推定した数値を提案手法である説明モデルを用いて近似し、GA を用いて最適化した。GA の個体数は 1000, 世代数は 5000 とした。

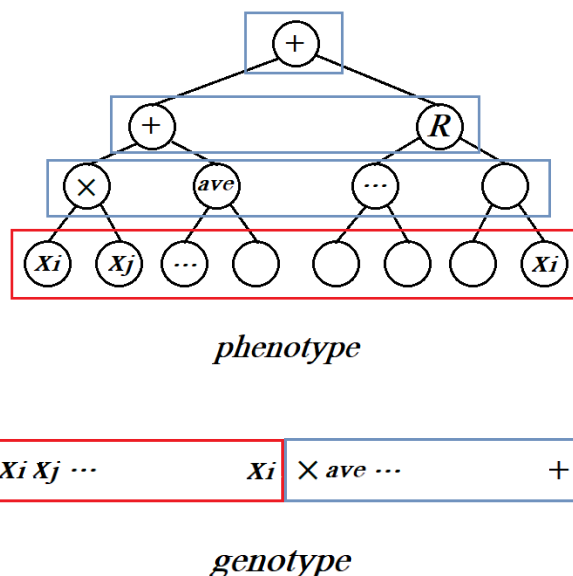


図1. モデルの表現型および遺伝子型

手法	スコア (10 試行平均)
NN(説明対象のモデル)	0.73256
説明モデル(提案手法)	<b>0.76555</b>

図2. 実験結果

### 3.2. 実験結果

実験結果を図2に示す。NN を用いた機械学習モデルでの正答率は 0.733, 説明モデルでの正答率は 0.766 となった。正答率は 10 試行平均で出したが、説明モデルは NN の正答率に近いだけでなく、高い正答率を実現できていることが確認できる。また、説明モデルの出力値の分布が NN の出力値の分布に近いことから、説明モデルは元のモデルをよく近似できていることが確認できた。

## 4. まとめ

本稿では二分木を用いて、NN の機械学習モデルを近似する手法を提案した。今後は、元のモデル内部の構造も考慮した、さらなる説明性の高いモデルを構築する必要がある。

### 参考文献

[1] Davis, L : “The Handbook of Genetic Algorithm”, Van Nostrand Reinhold, New York, 1990.