

論文の内容の要旨

論文題目 ノンパラメトリックな確率モデルを用いたプログラム進化

氏名 丹治 信

本研究では、ノンパラメトリックな確率モデルを用いた遺伝的プログラミング(Genetic Programming, GP)に関する研究について述べる。

GPは、生物進化を模倣した進化計算の一手法であり、木構造データを扱うことでプログラムそのものや関数を生成することが可能な手法である。ロボットの動作生成や複雑な数式など、人間が直接作ることが難しいような問題に対して、機械に自動的に解を獲得させることが可能である。

GPや遺伝的アルゴリズム(Genetic Algorithm, GA)などの進化計算では、交叉、突然変異と呼ばれる有性生殖を模倣した遺伝的オペレータを用いて解の組み換えを行ってきた。近年、確率モデルを用いた進化計算の枠組みが盛んに研究されている。確率モデルを用いた進化計算はGAで初めて提案され、1次元で固定長のパラメータを最適化させる問題で有効であることが示されてきた。特に最適化する対象の変数間に依存関係があるような問題は従来のGAでは、ベイジアンネットワークや変数のグループへと分解するようなモデルを使うことで従来のGAよりも正確に解けることが報告されている。

現在、GPに対しても、確率モデルを使用することで、複雑な問題や騙し問題などを解くアプローチが研究されており、確率モデルGPと呼ばれる。これらの確率モデルを明示的に扱うアプローチでは、PPT(Probabilistic Prototype Tree)表現や確率文法が用いられる。しかし、可変長な木構造を遺伝子として持つGPでは、ノード間の依存関係を表現することが難しい。またGPでは使用するシンボルの数が多く、パラメータ数や十分な推定に必要な個体数が増加し、結果計算量が増大してしまう問題がある。本研究で示す実験では、PPT表現を用いるタイプの確率モデルGPの比較を行い、モデルの複雑性と解くことができる問題の難しさに相関があることを示す。

本論文では、上記の問題に対し、PORTS(Program Optimization by Random Tree Sampling)とPERCE(Program Evolution using Related Clique Extraction)の二つの手法を提案する。これらの手法は、ネットワークの確率テーブルや条件付確率表などの明示的なパラメータを持たない、ノンパラメトリックな確率モデルであると言える。そのため、少ない計算量、少ないメモリで実行できるメリットがある。

PORTS では、優良個体からの木構造断片と呼ばれる単位の部分構造を組み合わせることで新たな個体を生成する。木構造断片のサイズは幾何分布に従うようにサンプリングされ、適応的なスケジューリングにより、この分布のサイズを変化させることで、効率的に解の探索を行う。この特徴から、GP-Easy な問題や Regression 問題に対して有効に働くことが示されている。

PERCE(Program Evolution using Related Clique Extraction)は、PPT 表現を、互いに重複を許すクリークに分割することで有用な部分構造を保存する手法である。従来の確率モデル GP では、部分構造の周辺分布を保持する必要があったが、優良個体からデータをサンプリングすることでメモリを節約と、大きなサイズの部分構造を保存可能である。従来の GP では解きづらかった騙し問題に対して、従来の GP や、確率モデル GP より高速に解けることを示す。

最後に、本研究で得られた結果についてまとめ、GP と確率モデル GP の利用方法について考察する。また確率モデル GP の研究の課題と将来の方向性について述べる。