

2-G-1-3

複数手法を組み合わせた遺伝的プログラミングによる 医療データの学習

新美 礼彦¹⁾ 田崎 栄一郎¹⁾
桐蔭横浜大学 工学部 制御システム工学科¹⁾

Combined Learning Methods of Genetic Programming for Medical Databases

Ayahiko Niimi¹⁾ Eiichiro Tazaki¹⁾

Department of Control & Systems Engineering, Toin University of Yokohama¹⁾

Abstract: There are many techniques about diagnostic system construction methods using medical treatment data. Genetic programming (GP) is a learning method by which the expression of the system is comparatively flexible and structural learning can be done at the same time, when the system is constructed. It is thought that the system which easily contains the representation of higher-order knowledge can be constructed if this can be used to construct the medical treatment diagnostic systems. However, if the node has the continuous value attributes in GP, the learning settling slows, and the decrease in the accuracy of the diagnostic system is invited occasionally. The technique by which the learning efficiency of GP is improved to the continuous value attribute is proposed by building in two or more techniques usual GP for such a problem. It is expected that more efficient GP learning can be done by the interaction of the plural techniques built in. To verify the validity of the proposed method, we developed a medical diagnostic system for the occurrence of hypertension, and compared the proposed method with prior methods.

Keywords: Genetic Programming, medical databases, decision tree, pruning redundant patterns, fitting random continuous nodes

1. はじめに

医療データからの診断支援システム構築に関する手法は、数多く提案されている。遺伝的プログラミング (GP) は、システムの表現が比較的柔軟で、かつシステム構築時に構造学習を同時に行うことができる学習法である。これを医療診断支援システムの構築に利用できれば、簡単に高次の知識表現を含んだシステムの構築が可能であると考えられる。しかしながら、GP ではノードが連続値属性を持つと学習収束が遅くなることや、診断システムの精度の低下を招くことがある。

このような問題に対し、通常の GP に複数の手法を組み込むことにより、連続値属性に対する GP の学習効率の改善をはかる手法を提案する。組み込む複数手法の相互作用により、より効率的な GP 学習が行えるものと期待される。組み込む手法は、異なる学習に適したものを使用するほうが、相互作用を期待できると考えられる。ここでは、素早い構造学習を行う手法、効率的に探索範囲を狭める手法、連続値属性を効果的に学習する手法、の 3 つの学習法を含ませることにした。また、組み込む手法の条件として、比較的簡単なアルゴリズムで計算量の少ないものが望ましい。これは、組み込み操作による計算量の増加を防ぐ意味

がある。

2. 遺伝的プログラミング

遺伝的プログラミング (Genetic Programming: GP) は、生物進化論の考えに基づいた学習法であり、そのアルゴリズムの流れは遺伝的アルゴリズム (Genetic Algorithm: GA) と同様である。^[1] その特徴は染色体表現が GA と異なり、構造表現ができるように拡張してあることである。今回は、決定木を表現するためにツリー構造を用いた。

3. GP と複数手法の組み合わせ

複数手法の組み合わせの相互作用による学習効率の改善は、連続値属性を多く含むデータベースや属性数が多い大規模データベースへの GP 学習の適用の可能性を開くことができると考えられる。

3.1 C4.5 による決定木構築手法

決定木による分類学習は短時間の学習ですむが、その訓練データにノイズが含まれていた場合、その分類能力は急速に劣化する。一方、GP は構造学習が可能であるため、高い分類能力を選られるが、学習の自由度が増すことにより、より多くの学習時間が必要となる。

この問題に対し、2つのアルゴリズムを組み合わせることで学習を行う。まず、C4.5^[2]を用いて決定木を構築し、GPの初期集団に取り込み、学習を行う。^[3]これにより、GPの初期集団内に有効と思われるスキーマを含ませることが可能となり、学習速度と分類精度の改善が期待される。

3.2 無効ノード削除操作

無効ノード削除を行なうことにより、GP個体内の無駄な構造を省き、木構造の縮小と計算量の削減が期待される。^[4]無効ノード削除はランダムに決められたサブツリーに対して、無効ノードパターンを探し、見つかったノードパターンを有効な終端ノードと置き換えることにより行なう。なお、無効ノードパターンについては問題に依存するので、問題ごとに設定する必要がある。

今回は、GPの木構造をノードオブジェクト配列に変換してから、無効ノードパターンのパターンマッチングを行なった。木構造を配列に変換するため、木の深さを気にすることなく、目的の無効ノードパターンを探すことが可能となる。また、オブジェクト配列として捉えることにより、ノード内容を気にすることなく、処理を行なうことが可能となる。

3.3 連続値属性適応操作

あるサブツリー内のすべてのノードに対して、定義されたノードの組み合わせすべてと構造が変わらないよう置き換えて、その中でもっとも適応度が高かった組み合わせを選択するというアルゴリズムを考える。このアルゴリズムでは、ある構造のときの局所最適解を探ことができ、強力な学習促進が期待される。しかしながら、計算量が増加する可能性がある。今回は、この考え方を連続値属性を多く含んだGPの学習に適応した。^[4]

決定木構築では、終端ノードに訓練データの連続値属性を出力するノードを定義することがある。決定木でこのノードと比較されるノードを乱数によって決めるとき、このような連続値属性の範囲が広いと、なかなか有効な定数が選択されない可能性がある。そこで、このようなノードに対して、訓練データをもとにより適応度が高くなるような定数と置き換えるアルゴリズムを提案する。

4. 医療診断支援システムへの適用

GPへの組み込みにはC4.5による決定木構築法、無効ノード削除操作、連続値属性適応操作の3つを使用した。また、提案した手法の有効性を検証するため、高血圧発症の医療データベースを用いた。このデータベースは多くの属性が連続値属性であるため、今回の検証に向いていると考えられる。実験の結果、通常のGPのみの場合に比べて、短時間に高精度の推論システムを構築することができた。

5. おわりに

本研究では、GPに複数手法を組み合わせることで学習を行うルール生成システムの構築法を提案した。また、提案した手法の有効性を検証するために、医療データベースを用いてシステム構築を行い、単独で使用した場合との比較を行った。

その結果、提案した手法に学習効率の改善が認められた。このことより、提案した手法は、ルール生成システムの学習効率の改善に有効な手法であるといえる。

今後は、提案した複数手法の組み合わせの相互作用による学習効率の改善が有効であると思われる、連続値属性を多く含むデータベースや属性数が多い大規模データベースへのGP学習の適用を試みる予定である。

参考文献

- [1] J. R. Koza: Genetic Programming, MIT Press, 1992.
- [2] J. R. Quinlan: C4.5 Programs for Machine Learning, Morgan Kaufmann Publishers, 1995.
- [3] 新美礼彦, 田崎栄一郎: オブジェクト指向を用いた決定木とADF GPの組み合わせによる学習, 第42回知識ベースシステム研究会, 人工知能学会, pp.19-24, 1999.
- [4] 新美礼彦, 田崎栄一郎: 無効ノード削除と連続値属性の適応操作を加えた遺伝的プログラミング, 第13回人工知能学会全国大会論文集, pp.257-258, 1999.