

遺伝的機械学習による多目的知識獲得



大阪府立大学 大学院工学研究科 准教授
能島 裕介

研究の背景

識別器設計やモデリングなど、数値データからの知識獲得は様々な分野で行われています。知識獲得において重要なことは、知識の精度とわかりやすさです。知識の精度は、その知識を信頼するために必要です。一方、知識のわかりやすさは、得られた知識を理解するために必要です。しかし、一般的には、高精度で単純な知識は存在せず、その両者にはトレードオフの関係があります(図1)。すなわち、高精度な知識ほど複雑な表現になり、わかりやすい知識ほど精度が落ち、必要な精度とわかりやすさは利用者に依存します。さらに、知識獲得では、大規模データの活用が重要です。ビッグデータ時代では、大規模データからの知識獲得は、多くの利用者にとって必要な技術になっています。

研究の成果

私は、遺伝的機械学習を用いた多目的知識獲得について研究しています。遺伝的機械学習とは、生物の進化を模した進化計算手法のひとつであり、個体表現として「If-thenルール」の形式を用います。これにより、数値データから言語的に理解可能な知識を抽出することができます。進化計算の欠点は、解候補の繰り返し評価が必要になることで、データが大きくなると深刻な問題になります。この問題に対して、遺伝的機械学習の並列分散実装を提案しました(図2)。

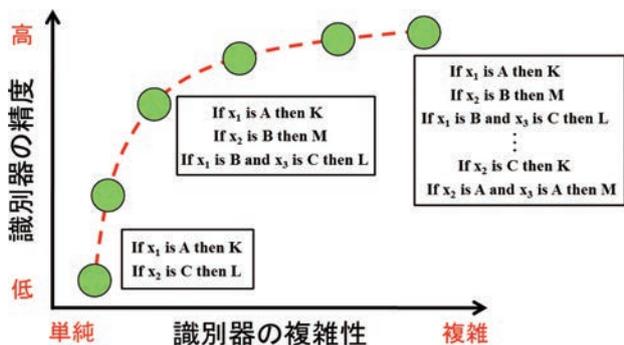


図1 知識の精度と複雑性のトレードオフ

個体群の分割とデータの分割を同時に行うと、計算に使用するCPUの個数の2乗倍の高速化が可能になり、さらに、未知データへの汎化性能の改善も可能になることを示しました。例えば、20CPUコアのワークステーションを用いると、単一CPUで1日かかった計算がわずか4分で終わります。さらに、進化型多目的最適化に拡張すると、精度と複雑性の異なる複数の知識を短時間で獲得することが可能になりました。

今後の展望

本研究で提案している並列分散実装を、大規模分散処理フレームワークである「Hadoopエコシステム」などに実装すれば、さらに大きなデータを高速に処理することが可能になると考えています。この高速化によって、オンデマンドで利用者に合わせた多目的知識獲得を大量のデータから行うことが可能になり、例えば、医療現場で、知識や経験の異なる医者と患者それぞれに適した診断情報を同時に提供することができるようになると期待しています。

関連する科研費

平成22-24年度 若手研究 (B) 「並列分散遺伝的知識獲得における効率的な個体群およびデータの分割」

平成25-27年度 基盤研究 (C) 「多目的遺伝的機械学習手法の並列分散実装」

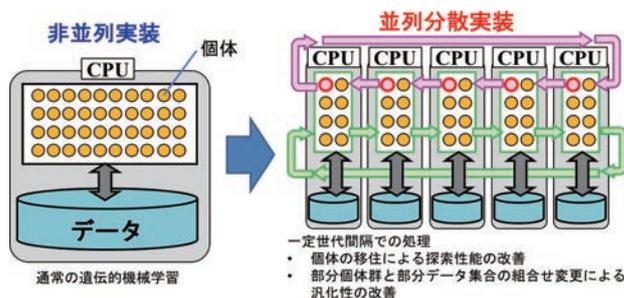


図2 遺伝的機械学習の並列分散実装