

論文の内容の要旨

論文題目 共生進化の帰納学習への適用に関する研究

氏 名 大谷紀子

人工知能の分野では、人間の知能と知的活動をコンピュータ上で実現することを目標として、さまざまな研究が行なわれてきた。人間の行なう知的活動のうち、とりわけ有用かつ高度といえる知的活動は学習であり、コンピュータに学習機能を持たせるべく行なわれているのが機械学習の研究である。

機械学習の代表的研究パラダイムの中で特に活発に研究されているのは、帰納推論によって複数の事例から一般的な概念を導出する帰納学習である。帰納学習の過程は、導出すべき概念の候補をすべて含む概念空間の探索として捉えることができる。与えられた事例だけではなく、未知の事例の存在も考慮しながら概念空間を探索することで、予測問題をも解決するような概念を獲得することができる。

一般に概念空間は広大であるため、目標とする概念を効率よく見つけ出す手法が必要とされている。さまざまな解探索手法のうち、遺伝的アルゴリズムは広大な解空間を探索できる多点探索手法であり、複数の局所解から大域的な最適解を発見したり、実用的な時間で評価の高い解を発見したりすることができる。また、組合せ最適化問題や関数最適化問題などさまざまな最適化問題に適用することができ、評

価関数に関する詳細な知識がなくても適用可能であるため、応用面においても有用な探索手法といえる。生物の進化のメカニズムを単純にモデル化した基本的なアルゴリズムのほか、目的に応じたさまざまな拡張アルゴリズムが提案されている。

Moriarty らは、動的で複雑な問題にも応用できるようなアルゴリズムとして共生進化を提案した。共生進化は、“そのままでは解決できないような複雑な問題を単純な副問題に分割し、副問題の解を統合して最終的な解を導き出す”という統治分割法に基づいた手法である。共生進化では、問題に対する解をいくつかの部分解の組合せにより表現する。部分解を個体とする部分解集団と、部分解の組合せを個体とする全体解集団を保持し、前者において最適解に含まれるべき部分解を、後者において最適解となる部分解の組合せを探索する。両集団は、互いに影響を及ぼし合いながら並行して進化する。共生進化の特長は、局所解への収束を回避するとともに最適解への収束を促進する点にあり、ニューラルネットワークやファジールールの学習に適用できることが報告されている。

共生進化は、ある目標を達成するために複数の人が協働する状況をモデル化したアルゴリズムといえる。複数のメンバーから構成されるグループの働きは、各メンバーの能力の高さだけではなく、メンバーの組合せにも依存する。能力の高いメンバーを育成しながら、目標達成に効果を発揮するメンバーの組合せをも見つけ出さなくてはならない。部分解集団の進化は個人の能力の向上に相当し、全体解集団の進化は目標を達成できるグループ編成の探索に相当する。

本論文では、共生進化の枠組みを明らかにするとともに、帰納学習全般への共生進化の適用可能性を確認することを目的とする。共生進化の基本的な性質を調べるため、OneMax 問題、Harik のだまし問題、Floyd 問題といった単純な問題を用いて、一般的な遺伝的アルゴリズムと共生進化との比較実験を行なった。両者で得られる解に加え、適応度の推移や集団の多様性の推移を比較することにより、共生進化の解探索手法としての有用性が示された。この結果を踏まえて、解構造が複雑な帰納学習の問題として帰納論理プログラミングと決定木生成を取り上げ、共生進化に基づく手法を提案する。以下、提案手法を実装したシステムの概要と動作結果を示す。

帰納論理プログラミングシステム

帰納論理プログラミングでは、与えられた正事例を被覆し、負事例を被覆しない

ような最適仮説の集合を導出する。Horn 節により表現される最適仮説の探索は、最適仮説に含まれるべきリテラルと各リテラルの一般化の度合いを決定することといえる。したがって、リテラルを部分解、リテラルの組合せで表現される Horn 節を全体解とする共生進化を最適仮説探索に適用することができる。予測正解率を向上させるため、アンサンブル学習を導入する。従来の探索では得られない解の多様性を未知事例への適合可能性と捉え、最適仮説集合の生成を繰り返して得られた複数の最適仮説集合の多数決により未知事例を分類する。帰納論理プログラミングシステムのベンチマークとして最もよく用いられている Mutagenesis のデータにより評価実験を行なったところ、4 種類の背景知識のいずれを用いた場合にも、提案手法を実装したシステムでは既存システムより高い予測正解率が得られた。また、一般的な遺伝的アルゴリズムよりも共生進化を適用する方が予測正解率が高く、多数決によるアンサンブル学習が予測正解率に向上に寄与することも示された。

決定木生成システム

決定木は分類規則の木構造による表現技法である。分類規則には解釈の容易さが求められることが多いため、予測正解率に加えて木の簡素さも決定木の重要な評価指標といえる。分類規則は複数の属性に関する分岐の組合せと見なすことができる。すなわち、決定木は高さ 1 の部分木の組合せであり、高さ 1 の部分木を部分解、高さ 1 の部分木を結合した決定木を全体解とする共生進化により決定木を生成することができる。過学習を回避するための指標として、終端ノードにおける正解事例数の散らばりを示す正解局在率を定義し、正解率と正解局在率により算出した適応度に従って、部分解集団と全体解集団を並行して進化させる。UCI 機械学習リポジトリのデータを用いて評価実験を行なった結果、提案手法を実装したシステムでは、代表的な決定木生成システム C5.0 よりも予測正解率が平均 2% 高く、ノード数が 70% 程度の決定木が生成された。また、共生進化では一般的な遺伝的アルゴリズムよりもノード数の少ない決定木を生成できることが示された。さらに、正解局在率による決定木の評価は、MDL 基準よりも簡素な決定木の生成に有用であることが確認された。

ポリマー判別システム

プラスチックのリサイクルにはポリマーの種類判別が必要であり、安価かつ簡便な判別手法として近赤外スペクトルデータの利用が検討されている。処理の透明

性、およびポリマーの特徴の明示という点から、決定木での判別規則の表現が求められている。近赤外スペクトルデータは 2401 個の波長と吸光度の組からなり、決定木の評価に用いられるベンチマークデータと比較すると、属性数が非常に多い。また、クラス間の事例数の偏りも大きく、異なるクラスに類似した事例が存在したり、同一クラスに特徴の異なる事例が存在したりしている。これらの特徴を踏まえ、多属性データを扱える遺伝子表現とクラス間の事例数の偏りを考慮した適応度関数による共生進化で決定木を生成するとともに、代表的な事例とは異なる少数派の事例の特徴をも学習するための 2 段階判別法を導入する。評価実験の結果、提案手法の有効性が示され、実験に用いたすべての種類のポリマーについて、正解率 100% の決定木が生成された。本研究で得られた決定木は、プラスチックのリサイクル処理において実用化が可能と考えられ、資源有効活用の一助となることが期待される。

本研究における評価実験では、一般的な遺伝的アルゴリズムよりも共生進化が解探索手法として有効であるという結果が得られたことから、既に遺伝的アルゴリズムが適用されている問題に共生進化を適用することで、さらに目的に適合する解が得られるものと考えられる。共生進化の適用に際しては、部分解と全体解の表現を検討する必要があるが、本研究を通して以下のような設計指針が得られた。

- 意味づけの明確な構造を部分解とする。
- 部分解の単純な組合せにより全体解を表現する。

この設計指針に従うことによって、共生進化を種々の問題に容易に適用することができる。

ポリマー判別のための決定木生成については、ベンチマークデータではなく、実データにより有効性を検証した。実データを用いたシステムの検証はしばしば望まれるところであり、本研究におけるポリマー判別システムが実データで他システムよりも良好な解を導出したことで、提案手法の有効性が明確に示されたものと考えられる。また、最終成果として得られた決定木により、化学の分野で取り上げられている問題が解決できた点も大きな成果といえる。

本論文では、共生進化が帰納論理プログラミングの最適仮説探索と決定木生成に有効であることを示した。また、Horn 節や決定木という構造の複雑な解を持つ問題を対象として良好な結果が得られており、実データにおける有用性も確認されていることから、多様な問題への適用可能性は十分に高いと考える。