

ILPにおけるGAを用いた仮説探索の一手法

A method of the hypothesis search using GA to ILP

大橋 政彦^{*1} 松井 藤五郎^{*2} 大和田 勇人^{*2}
 Masahiko Ohashi Tohgoroh Matsui Hayato Ohwada

^{*1} 東京理科大学大学院理工学研究科経営工学専攻

Department of Industrial Administration, Graduate School of Science and Technology, Tokyo University of Science

^{*2} 東京理科大学理工学部経営工学科

Department of Industrial Administration, Faculty of Science and Technology, Tokyo University of Science

The inductive logic programming derives the hypothesis that is the generalization of the example from a lot of rules and the background knowledge. However, because the hypothesis space of the target for the search usually becomes infinity, the study speed is known to be slow as for ILP. The genetic algorithm is an algorithm that is a method of optimization, and it is effective to the problem of a vast space. The technique of this research proposes the application of GA to the technique of the search for the hypothesis of ILP. This thesis proposes the learning method of the gene expression by the horn clause coding in learning from the bottom clause.

1. はじめに

帰納論理プログラミング(Inductive Logic Programming:ILP)は、一階述語論理を表現言語とし、他の手法に比べて表現力が豊かであり現実的な問題を扱う際に強力な手法であるといえる。また、多くの事例とその関連情報(背景知識)から、事例の一般化である仮説(ルール)を導出する帰納推論を一階述語論理上で行うものである。しかし、ILPは探索の対象となる仮説空間が通常無限大になるため学習速度が遅いことが知られている。

遺伝的アルゴリズム(Genetic Algorithm:GA)は、生物の進化の過程を真似て作られたアルゴリズムで、最適化の一手法である。特徴としては、解空間が不明であり、決定的な解法が発見されておらず、また、全探索が不可能と考えられるほど広大な空間の問題に有効であることが挙げられている。

現在、ILPにGAを適用した研究は様々な手法が提案されており、有効性が示されている。[Tamaddoni-Nezhad]が提案している研究は最弱仮説をボトムとする探索空間で学習する方法を提案している。また、遺伝的アルゴリズムの適用の際、ビット列コーディングをせずホーン節のまま探索[市瀬]を提案している。

本論文は、最弱仮説を生成することで仮説空間を保持し、最弱仮説リテラルに基づくコーディングの学習を提案する。ボトム節からの探索によって、それぞれの最弱仮説におけるリテラル間でのランダム選択が容易に適用でき、遺伝的アルゴリズムにおける無意味な個体群の生成を制御できる。最弱仮説からの学習により、市瀬によるホーン節の連続性や引数の修正の必要がなくなる。また、GAには様々な手法が提案されているが、それらの手法はSGA(Simple Genetic Algorithm)を拡張したものである。ゆえに、今後改良や新たな手法提案をする上で基本的なモデルからの提案を考える。これにより、新たな学習戦略を提案する。

2. SGAによる仮説探索

ILPにおける仮説空間探索アルゴリズムは、一般的な仮説から探索を始めるトップダウン探索と特殊な仮説から探索を始めるボトムアップ探索がある。提案手法においては特殊な仮説(最弱仮説)を生成し、一般化していくボトムアップ探索を採用する。生成された最弱仮説は、逆伴意法により導出する。このボトムアップ探索における探索戦略に遺伝的アルゴリズムの一般的な手法であるSGA(Simple Genetic Algorithm)を用いる。初期個体の生成において、最弱仮説リテラルに基づくコーディングを適用する。SGAは最も基本的な遺伝的アルゴリズムの手法で、母集団の形成、評価、選択、突然変異により各世代で遺伝子操作を繰り返す。

2.1 最弱仮説リテラルに基づく遺伝子表現

一般的にSGAでは、遺伝子表現においてビットストリング表現で表す。ここで、最弱仮説リテラルに基づく遺伝子表現を行う。最弱仮説から探索することにより、ホーン節での連続性や引数の修正の必要がなくなる。

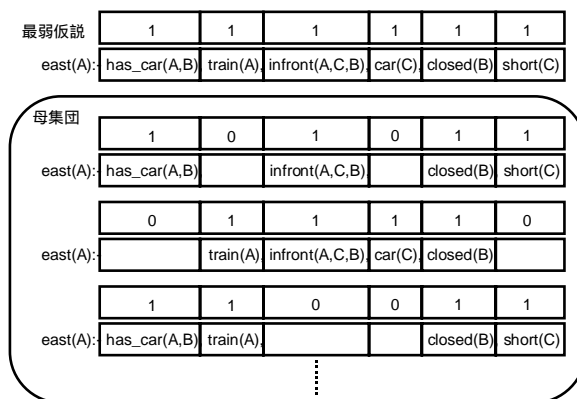


図1、最弱仮説リテラルに基づく遺伝子表現

連絡先: 大橋 政彦, 東京理科大学大学院理工学研究科経営工学専攻, E-mail: j7404620@ed.noda.tus.ac.jp

本研究では、最弱仮説のボディ部に含まれるリテラルをビットに対応させて遺伝子をコーディングする。最弱仮説を

$$h: -L_1, L_2, \dots, L_k.$$

とすると、遺伝子は k ビットのビット列によって表される。図1のように、個体 i の第 j ビットが 1 ならば仮説 h_i のボディ部にリテラル L_j が含まれ、0 ならば h_i のボディ部には L_j は含まれない。

Tamaddoni-Nezhad らは変数の束縛関係をビットに対応させてコーディングしているが、これに対し本手法のコーディングは非常にシンプルであるという利点を持つ。

2.2 MDLに基づく適応度

GA において次世代に残す母集団を選択する場合、適応度を求めて評価を行う。本提案手法においてホーン節による学習を行うことを考え、適応度関数を ILP の仮説評価基準として一般的な記述長最小原理 (minimum description length principle: MDL) を適用する。MDL の考えは、正例だけを説明して、負例を説明しない仮説が良い仮説ということである。仮説自身の記述長と、その仮説を用いて記述されるデータの記述長の総和を最小にする仮説が、最良のモデルであるということである。

候補仮説によって被覆される正例の数 p_i 、候補仮説によって被覆される負例の数 n_i 、仮説本体部のリテラル数 c_i をとり、以下を評価値とする。この評価関数は、記述長をリテラル数で近似することで、

$$fitness(i) = p_i - (n_i + c_i) \quad (1)$$

求めるべき最適解は、 $fitness(i)$ が非負かつ最大の値とする。

2.3 遺伝子操作

遺伝子操作において、親の節を二個体とするリテラルの一点交叉と節のリテラルへの突然変異を行う。遺伝子操作を行う上で、最弱仮説を生成していることで仮説空間を保持している。よって遺伝子操作が最弱仮説によって定義される探索空間を超えることはない。これにより、無意味な個体群が生成されたとしても適応度関数によって排除される。

突然変異においては、それぞれの生成された節に対して突然変異率 p でランダムに選択し、選ばれた節のリテラルに対して突然変異を行う。

生成された母集団は適応度関数で評価され、ルーレット選択によって次世代に残す。

ルーレット選択とは、各個体の適合度から選択確率を変え、確率に沿ってルーレットを回すような選択を行う方法である。利点として、適合度の高い個体ほど残りやすくなるが、適合度の低い個体でもそれなりの確率で生き残れるということである。各個体について求めた評価(適応度)を $fitness(i)$ とし、全個体の評価の総和を G とし、各個体の選択される確率 $fitness(i)/G$ を求める。そして、前の個体群からこの確率を用いて個体を M 個選択し、新しい母集団とする。

3. システム構成

本研究のシステム構成において、ILP の学習システムとして多くの有効性が示されている GKS をベースに提案する。以上で述べた提案を GKS 上の仮説探索上で改良する。

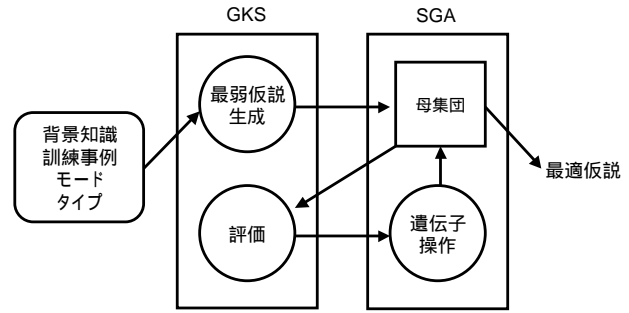


図2, システム構成

本論文で提案する方法は、最弱仮説をボトム、最も一般的な単位節をトップとする包摂束を仮説空間とし、仮説空間からの最適仮説探索をSGAによる探索戦略で行う。

図2は GKS を用いて最弱仮説を生成し、SGA のおいて母集団を形成する。形成された母集団は GKS で評価され、適応度を求める。適応度によって遺伝子操作され、次世代における母集団を生成する。世代間の繰り返しを行い、最適化探索を行う。世代間における繰り返しは、終了条件を設定し過学習を制限する。探索部で遺伝子操作を行うシステムの実装を試みる。生成される最弱仮説は逆伴意法により導出され生成される。それぞれの世代における母集団の形成を行い、評価と遺伝子操作を繰り返す。世代における繰り返しは、終了条件を設定し過学習を制限する。以上の学習過程のシステムにより最適化探索を行う。

4. 終わりに

本論文では、最弱仮説をボトムとする仮説空間においてSGAの最適化アルゴリズムを適用する。SGAの遺伝子表現において、最弱仮説リテラルに基づく遺伝子表現を提案した。これにより遺伝子操作が最弱仮説によって定義される探索空間を超えることはなく、また遺伝子におけるコーディングが非常にシンプルである。

本研究はまだ提案の段階であり、今後実装をする段階である。今後の課題として明確な手法の定義、実験と検証が必要である。この研究により、ILP における貪欲な探索より良い探索法の提案することが課題である。

参考文献

- [Divina 02] Federico Divina and Elena Marchiori, "Evolutionary Concept Learning", in Proc. The Genetic and Evolutionary Computation Conference, pp. 343—350, 2002.
- [Tamaddoni-Nezhad 02] A Tamaddoni-Nezhad and S.H. Muggleton, "A genetic algorithms approach to ILP", in Proc. The 12th International Conference on ILP, pp 285—300, 2002.
- [市瀬 99] 市瀬 龍太郎, 沼尾 正行, "帰納学習における帰納論理プログラミングと遺伝的プログラミングの統合", 人工知能学会誌, Vol. 14, No. 2, pp 307—314, 1999.
- [大谷 02] 大谷 紀子, 大和田 勇人, "共生進化に基づく帰納論理プログラミングの予測精度の向上", 人工知能学会誌, Vol. 17, No. 4, pp 431—438, 2002.