

ILP システムと属性-値学習システムとの比較検討*

2C-2

溝口文雄

清水健一

大和田勇人†

東京理科大学 理工学部‡

1 はじめに

帰納学習は帰納論理プログラミング (ILP) の枠組と属性-値学習の枠組に大別することができ、それぞれ長所と短所を有している。一般に、複雑な問題対象を記述し学習を行なうためには、述語論理の特性や背景知識の利用などの長所から ILP が向いているといわれる。しかし、実際に ILP システムと属性-値学習システムとの実行能力を比較してみると、実行時間、処理できるデータの量などにおいて、属性-値学習システムの方が優れていることもある。そこで、ILP システムと属性-値学習システムの学習可能領域と実行能力を比較し吟味することが必要である。

本研究ではこの背景に基づき、ILP、属性-値学習システムの代表として Progol[1]、C4.5[2] を選び、これらについて実行能力の比較を行なう。また、学習対象の問題としては、代表的な論理プログラムの例題とフロアプランニングのデータを取り上げる。

2 帰納学習システム：C4.5 と Progol

現実の問題から学習を行なう場合、大量のデータの処理、探索の効率、数値データの処理などが重要な課題となる。既存の ILP、属性-値学習システムの中では、Progol、C4.5 が特にこの課題を満たすものであると考えられる。

C4.5 は、数値やノイズ処理が得意な属性-値学習システムである。決定木を生成することで高速に学習が行なえる。Progol は効率が良い探索を行なう ILP システムであり、確定節の利用が可能であり、数値の処理も GOLEM などの従来のシステムよりも強化されている。

3 実行能力の比較の手順

1. 前準備, 2. 学習の実行, 3. 比較, という手順で学習システムの実行能力の比較を行なう。

3.1 前準備

3.1.1 ILP の問題の属性-値への変換

Progol と C4.5 の実行能力の比較を行なうため、以下の点に注意しながら、ILP の問題を属性-値の形式に変換する。

分類クラス: 学習問題には、ある状態を学習する二値学習と、複数のクラスの分類規則を学習する多値学習がある。属性-値の形式に変換するには、二値学習の場合、真と偽のクラスとに分類するように記述する。(負事例が考えられない問題は学習不可能である。) 多値学習の場合、複数のクラスをそのまま学習するクラスとする。

学習ターゲットの分割: 属性-値学習では、複数の変数の学習が行なえない。例えば、`move(Peace, Position)` という述語でチェスの駒の特性を学習する場合、複数の Peace について一度に学習することができないので、Peace ごとに分割して学習を行なう。

属性: 学習に用いる属性は学習者が決定する。深さ 1 以上の関係も必要に応じて加える。(考えられる関係を全て枚挙することも可能ではある。)

属性値: 属性値は互いに排他的な関係にある。例えば、“方角” という属性に対して、{北, 南, 北東, ...} という属性値が考えられるとき、北東 < 北 という関係は属性-値学習では推論されない。このような場合、四方向と八方向の方角に関する属性を定義する。

3.1.2 予備学習

Progol ではノイズの比率や探索の深さ、C4.5 では枝刈り率、葉の最小数などに関して、最適と思われる条件を予備学習により決定する。

3.2 学習の実行

クロスバリデーション (CV) 法に基づき各学習システムを実行させる。CV 法は、与えられた事例を用いて未知の事例に対する予測力を測定するシミュレーション手法である。

3.3 学習能力の比較

比較の指標は、正解率と実行時間とする。2つのシステムの予測能力に差があるかを調べるためには、2つのシス

*Comparison between Inductive Logic Programming systems and Attribute-Value Learning Systems

†Fumio Mizoguchi, Kenichi Shimizu, Hayato Ohwada

‡Faculty of Sci. and Tech., Science Univ. of Tokyo

テムの判定が異なる事象の数に基づき、統計的に検定する。

4 論理プログラムの例題の学習

代表的な論理プログラムの問題(表1参照)に対し, Progol, C4.5で学習した結果を表2に示す. なお, chessの問題に対するC4.5の結果は5つの駒についての結果の合計である.

表1: $|E^+|$, $|E^-|$ は正・負事例数. $|B|$ は Progol の背景知識の数. $|Class|$ は分類するクラスの数. $|Att|$ は属性の数.

データ	$ E^+ $	$ E^- $	$ B $	$ Class $	$ Att $
animals	16	42	105	4	7
chess	27	12	34	2	5
krki	342	658	51	2	12
train	5	5	257	2	31

表2: 注) C4.5の実行時間は $c4.5 + c4.5rules$

データ	Progol		C4.5	
	実行時間	正解率	実行時間	正解率
animals	0.25s	100.0%	0.06s	99.5%
chess	3.65s	61.5%	0.18s	51.0%
krki	18.85s	99.4%	0.73s	99.6%
train	0.371s	100.0%	0.05s	50.0%

ProgolとC4.5の正解率を比較すると, trainのデータにおいてのみ予測力に差があると検定された. これは, Progolで得られた $eastbound(A):-has_car(A,B),closed(B),short(B)$ という規則がC4.5では学習されなかったことによる. これは $\forall x \exists y (P(x,y) \wedge Q(y))$ という関係を属性-値学習では推論できないことを示している. この関係を属性を加えて再学習すると, C4.5でも正解率100%の予測が行なえた.

また, C4.5の実行時間は Progol の $\frac{1}{4} \sim \frac{1}{20}$ であった.

5 学習の応用例

現実問題から学習を行なう場合, 複雑な構造の問題から重要な関係を導出するというタスクが多い. そのような問題の例としてフロアプランニングの問題がある. この問題では, 実際の設計図面からプランニングを行なう上で有効な規則を獲得することが目的となる.

5.1 フロアプランのデータからの学習

入力として敷地の形状・面積と玄関の位置が与えられたとき, 四隅に配置される部屋や壁に接する部屋など図面において一般的に成り立つ関係を求める. 学習においては, 面積や長さなどの数値情報を処理する必要がある.

本稿では, 部屋のタイプの学習を Progol, C4.5で行なった結果を示す. なお, 入力となる図面数が50,100個の場合, 背景知識 B が, B_1 :数値情報を含まない, B_2 :数値情報を含む, という各々の条件下で学習を行なった.

学習される規則の一例を以下に示す.

```
room(A,bath) :- adjacent(A,washroom),
has_wall(A,east), space(A,S), S < 200.
```

5.2 学習結果・考察

Progol, C4.5による学習結果を表3に示す.

表3: 注) - は学習未完了を示す.

データ	B	Progol		C4.5	
		実行時間	正解率	実行時間	正解率
50	B_1	2669.39s	68.3%	6.42s	67.3%
	B_2	10634.42s	73.9%	4.08s	78.8%
100	B_1	24010.72s	72.2%	48.91s	75.0%
	B_2	-s	-%	14.80s	89.2%

実行時間は, C4.5の方が圧倒的に短かった. 正解率は, 背景知識に数値情報を加えた場合のみC4.5の方が良かった. また, 背景知識, 事例数が多くなると正解率が高くなる傾向は明らかである. しかし, Progolの最後のケースでは学習が完了しなかったことから, 大量のデータに対する実行能力に限界があることが分かる. 特に数値のリテラルを扱う場合探索空間が莫大になるため, 数値情報の処理が課題となる.

このように, 正解率と実行時間だけから見ると, C4.5の優位は明らかであった. しかし, 属性として明示されている以外の関係を新たに発見することはできない. それゆえ, 何が重要な関係であるかが全く未知の問題に対しては属性-値学習システムは適さないといえる. また, 属性の追加や変更も困難であり, 試行錯誤的な学習にも向かない. これに対してILPにおいては, この属性-値学習の短所が長所となる.

6 おわりに

未知の問題から重要な関係を導くという学習の本質的なタスクに対しては, ILPの推論能力が適している. しかし, 大量のデータや数値の処理能力だけに限れば, 属性-値学習システムの方が優れている点も多いことが, 本研究で行なった比較実験によって示された. そこで, 現実の問題に対し学習システムを適用する場合, 最初にILPシステムで学習を行なった後に, 学習の精度を上げる目的でC4.5を用いるという方針が立てられる.

参考文献

- [1] S.Muggleton. Inverse Entailment and Progol. *New Generation Computing*, Vol.13, pp.7-71, 1995.
- [2] J.Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.