

属性集合が可変な場合における決定木学習方式

6C-6

浅見徹 橋本和夫

国際電信電話株式会社 研究所

1. はじめに

従来のID3系の決定木学習方式^[1]では、インクリメンタルな学習へ拡張したID4^[2], ID5R^[3]も含め、学習期間中の属性集合に関しては一定であることが仮定されている。しかし、実際に専門家から知識を獲得する場合、対象を分類するに足る属性を当初から把握して、学習用の事例集合を集められないことがある。このとき、学習期間中に新たな属性の導入が必要になる場合があり、学習用事例集合を新しい属性集合を用いて、再収集しなければならなくなる。本稿では、属性集合が学習期間中に漸増する仮定の下に、収集済み事例集合を破棄せずに決定木を導出できるようID3系決定木学習理論の拡張を試みる。

2. 教師モデル

A, C , および S_l ($0 \leq l < N$, N はノードの総数)を、それぞれ m 個の属性の集合 $\{A_1, A_2, \dots, A_m\}$, p 個の分類クラスの集合 $\{C_1, C_2, \dots, C_p\}$, および決定木のあるノード N_l での学習用事例集合とし、ある属性 A_i の取り得る値の集合を $Range(A_i)$, また、集合 Set の要素の数を $|Set|$ で表すことにする。このとき、 $V_i \in Range(A_i), i = 1, \dots, m, C_k \in C, k = 1, \dots, p$ とすると、 $(m+1)$ -tuple の $E_n = \langle V_1, V_2, \dots, V_m, C_k \rangle$ で S_l 中の各事例を表すことができ、属性 A_i の値を知ることによって得られる相互情報量が最大になる A_i を用いて、事例集合 S_l の分割を行い、決定木を導出するのがID3系の決定木導出アルゴリズムである。以下の議論では、作成した決定木を T で、 T の終端ノードの集合を $TN(T)$ で表すことにする。

学習用の事例集合 S_0 を最初から全て与えて決定木を作成するID3の場合、前提となる教師の役割は、(1) 各事例 E_n の分類属性 C_k を指示するだけである。しかし、事例を追加するたびに決定木の更新を行うID5Rの場合には、事例 E_n を決定木 T で分類する際、ノード N_l での分類する事例 E_n の属性 A_i の値 V_i が A_i の値域に含まれていないと、 V_i に対応する枝を持つ新たな決定木を構成し、 C_k に E_n を分類する。このため、更に(2) 新たな属性・属性値対の指示が教師の役割として仮定されている。さて、インクリメンタルな学習を実際に応用した場合、全ての既存の属性集合 A を用いても、ある終端ノードでクラス C_k と C_k' が分類不可能になることがある。この場合、通常の知識獲得作業では、専門家の指示により、(3) 新しい属性 A_o が分類のために導入される。

ID5Rはインクリメンタルな学習が可能であるが、教

師の(3)の条件を満足させようとする、集合 A の増加以前に収集した事例を全て破棄してしまわなければならない。従って、学習効率を上げるためにはこのような事例の破棄を避けるアルゴリズムに拡張する必要がある。

3. 学習事例集合の操作アルゴリズム

まず、2つの疑似的な属性値 $*$ と $\neg x$ を導入し、属性 A_i の値域を拡張する。ここで、 $A_i = *$ の意味は A_i が属性値 $\forall V_i \in Range(A_i)$ に、 $A_i = \neg x$ は、 A_i が属性値 $\forall V_i \in \{V : V \in Range(A_i) \wedge V \neq x\}$ に等しいことを示す。このとき、以下の手続きで学習事例集合 S_0 を操作する。

- (1) 新しい事例 E_n を S_0 に追加する際、まず決定木 T で分類を試みる。
- (2) E_n の分類の際、 T の終端ノード N_l へ到達した場合
 - i) E_n のクラスが N_l の示すクラスに等しいとき、 $S_0 \leftarrow S_0 \cup \{E_n\}$.
 - ii) そうでないとき、 N_0 から N_l に至るパスで未使用の属性 A_o を E_n の分類属性として、導入する。
 - a) $A_o \notin A$ ならば、 $A \leftarrow A \cup \{A_o\}$ とし、 E_n に対しては A_o の観測値 V_o , S_0 中の既存事例に対しては $\neg V_o$, $S_0 - S_l$ 中の既観測事例については $*$ を A_o の値として与える。
 - b) $A_o \in A$ ならば、 E_n の A_o の値に対しては A_o の観測値 $V_{o,n}$ を与え、 S_l 中の既存事例の A_o の値は $*$ を $\neg V_{o,n}$ で置換する。
 - c) $S_0 \leftarrow S_0 \cup \{E_n\}$.
- (3) T の終端ノード N_l へ到達以前に、ある非終端ノード(検査属性を A_i とする)の枝で、 $A_i = V_i$ に対応する枝がない場合
 - i) $\neg V$ に対応する枝があり、 $V \neq V_i$ のときはその枝の先に分類を進める。
 - ii) そうでなければ、 $S_0 \leftarrow S_0 \cup \{E_n\}$.

4. 相互情報量の計算方法

ID3系決定木導出に使われるGain(相互情報量)計算の基礎は、学習事例集合 S_l の中でクラス C_k に属し、属性 $A_i = V_{i,j}$ となる事例の数 $N_{k,i,j}$ にあり、以下の手続きで求める。ここで、3.節で導入した2つの疑似的な属性

値“*”と“¬x”を考慮し、 $V_{i,j} \in Range(A_i)$ の最初の2つのこれらの値に割り当てているものとする。すなわち、 $V_{i,1} = *, V_{i,2} = \neg x$ とする。

- (1) 疑似的な属性値“*”以外の値 $V_{i,j} (j > 1)$ に対して事例集合 S_l の中で各クラス C_k に属し属性 $A_i = V_{i,j}$ となる事例の数 $\nu_{k,i,j}$ を求め、 $\mathcal{N}_{k,i,j} \leftarrow \nu_{k,i,j}$ と置く。
- (2) $A_i = *$ なる事例 E_n (クラス C_k , 重み $W_{l,n}$) があれば $j > 1$ に対して、

$$\mathcal{N}_{k,i,j} \leftarrow \mathcal{N}_{k,i,j} + W_{l,n} \frac{\nu_{k,i,j}}{\sum_{\beta>1} \nu_{k,i,\beta}}$$

と置く。また、 E_n の A_i による分割後、 $A_i = V_{i,j}$ に対応するノード N_p に属す E_n の重み W_p を $W_{l,n}P(i,j)$ とする。ここで、

$$P(i,j) = \frac{\sum_k \nu_{k,i,j}}{\sum_{\beta>1} \sum_k \nu_{k,i,\beta}}$$

であり、 S_0 に対しては $W_{0,n} = 1$ と置く。

すなわち、ここでは I.Kononenko 等が ASSISTANT^[1]において提案したベイズの判定規則によって未観測属性の値を比例配分する方法を疑似属性の処理に用いている。

5. 考察

3節のステップ(2)ii)の付加属性の選択は、通常、専門家の指示によるが、図1の1から8に示す Quinlan の学習事例^[3]を用いて逐次学習させた場合の事例集合の変化を求めて見ることにする。ここで、“+”、“-”はクラスを、height, hair, eyesは属性を、short, tall, blond等は属性値を表わす。事例の追加は1から8の番号順とし、

- 1 (-, height=short, hair=blond, eyes=brown)
- 2 (-, height=tall, hair=dark, eyes=brown)
- 3 (+, height=tall, hair=blond, eyes=blue)
- 4 (-, height=tall, hair=dark, eyes=blue)
- 5 (-, height=short, hair=dark, eyes=blue)
- 6 (+, height=tall, hair=red, eyes=blue)
- 7 (-, height=tall, hair=blond, eyes=brown)
- 8 (+, height=short, hair=blond, eyes=blue)

図1: Quinlan の学習事例^[3]

新しい属性の追加は height, hair, eyes の順としたときの事例集合の増加を図2に示す。ここで、-blond, -bluc の代わりに Xblond, Xbluc を用いて表している。

新しい属性が追加されるごとに事例集合を作り直すすると、図2のステップ3および7で既存事例を全て消去する必要があり、ステップ8を終了した時点で事例番号7, 8だけが3個の属性を持って S_0 に残るだけである。一方、本方式ではステップ6で事例番号2と6が分類不可能となるため、新しい属性 eyes が導入され、最終的には事例番号6, 7, 8が3個の属性を持って残り、事例集合の増加速度が大きい。従って学習中の決定木の分類誤り率も小さくなるのが利点である。

- ステップ1
1 (-, height=short)
- ステップ2
1 (-, height=short)
2 (-, height=tall)
- ステップ3
1 (-, height=short, hair=*)
2 (-, height=tall, hair=Xblond)
3 (+, height=tall, hair=blond)
- ステップ4
1 (-, height=short, hair=*)
2 (-, height=tall, hair=Xblond)
3 (+, height=tall, hair=blond)
4 (-, height=tall, hair=dark)
- ステップ5
1 (-, height=short, hair=*)
2 (-, height=tall, hair=Xblond)
3 (+, height=tall, hair=blond)
4 (-, height=tall, hair=dark)
5 (-, height=short, hair=dark)
- ステップ6
1 (-, height=short, hair=*, eyes=*)
2 (-, height=tall, hair=Xblond, eyes=Xblue)
3 (+, height=tall, hair=blond, eyes=*)
4 (-, height=tall, hair=dark, eyes=*)
5 (-, height=short, hair=dark, eyes=*)
6 (+, height=tall, hair=red, eyes=blue)
- ステップ7
1 (-, height=short, hair=*, eyes=*)
2 (-, height=tall, hair=Xblond, eyes=Xblue)
3 (+, height=tall, hair=blond, eyes=*)
4 (-, height=tall, hair=dark, eyes=*)
5 (-, height=short, hair=dark, eyes=*)
6 (+, height=tall, hair=red, eyes=blue)
7 (-, height=tall, hair=blond, eyes=brown)
- ステップ8
1 (-, height=short, hair=*, eyes=*)
2 (-, height=tall, hair=Xblond, eyes=Xblue)
3 (+, height=tall, hair=blond, eyes=*)
4 (-, height=tall, hair=dark, eyes=*)
5 (-, height=short, hair=dark, eyes=*)
6 (+, height=tall, hair=red, eyes=blue)
7 (-, height=tall, hair=blond, eyes=brown)
8 (+, height=short, hair=blond, eyes=blue)

図2: Quinlan の学習事例^[3]による学習プロセス

6. まとめ

本稿では、実際に専門家から知識を獲得する場合に近い属性集合が学習期間中に漸増する条件下に、ID3系決定木導出アルゴリズムを効率よく応用するため、収集済み学習集合を破棄せずに決定木を導出する学習用事例集合の操作方法について述べた。本方式では、 $A_i = *$ を持つ事例は、全ての値域に分散されるため、決定木にはノイズの混入に相当し、決定木の複雑さを増すが、反面では属性集合の増加を促す効果がある。また、 $A_i = \neg x$ を持つ属性は相互情報量の大きな値を持ちやすいため、初期の分類属性として選ばれやすい。従って、効率良い決定木を得るためには、あるステップ以降では疑似属性値を持つ事例を削除する手順、及びその統計的な有効性を実証することが今後の課題である。

最後に日頃御指導頂く KDD 研究所小野所長、浦野次長、山本 AI 応用グループリーダ、並びに通信網支援ソフトウェアグループと AI 応用グループの各位に感謝する。

参考文献

- [1] J.R. Quinlan, "Induction of Decision Trees", Machine Learning, Vol.1, pp.81-106, Kluwe Academic Publishers1986.
- [2] J.C. Schlimmer and D. Fisher, "A case study of incremental concept induction", in AAAI-86, pp.496-501, Morgan Kaufmann, 1986.
- [3] P.E. Utgoff, "Incremental Induction of Decision Trees", Machine Learning, Vol.4, pp.161-186, Kluwe Academic Publishers1989.