

4037

背景知識を用いたファジィID3
Fuzzy ID3 Using Background Knowledge

阪南大学 林 勲, ○川村 美智, 岡田 直
松下電器産業(株) 中央研究所 小澤 順

Isao Hayashi, Yoshitomo Kawamura, Naoshi Okada, Hannan University
Jun Ozawa, Central Research Laboratories, Matsushita Electric Industrial Co. Ltd.

Abstract. Algorithms of the Fuzzy ID3 family generate decision trees of fuzzy logic rules using mutual information. However, the algorithm is stopped if value of mutual information equals to zero. The algorithm presented in this paper tries to generate more continual and understandable decision trees by means of background knowledge.

1. はじめに

if-thenルールを構築するID3(Interactive Dichotomizer3)アルゴリズム¹⁾がある。平均相互情報量によりデータを分類し、テスト回数の期待値が最小になる決定木を構築する。また、ファジィルールによる決定木を構築するファジィID3²⁾がある。しかし、平均相互情報量が正でない場合にはアルゴリズムは停止し、分類は不可能となる。

一方、背景知識を用いて決定木を簡素化する手法³⁾が提案されている。背景知識とは、入力層の観測属性と上位層の抽象属性をもつ知識の階層表現である。各層はISA関係で結合する。本論文では、背景知識を用いた新たなファジィID3を提案する。平均相互情報量が正でない場合、背景知識を用いて初期のデータ集合の再構成を行い、データ分類を可能にする。背景知識を用いるので、ファジィルールは集約・細分化される。

2. 背景知識とは

背景知識では過去の経験から得た知識を階層構造で表現する。入力層には観測できる属性を定義し、上位層には観測属性の上位概念である抽象属性を定義する。各層のノードはISA関係で結合する。出力層の抽象属性が入力層の観測属性まで次々に具現化される。背景知識の一例を図1に示す。信号色は青系、黄系、赤系に具現化され、青系は青、緑の観測属性に具現化されている。黄系と赤系についても同様である。

3. 提案するファジィID3のアルゴリズム

いま、属性 $X_j, j=1, 2, \dots, n$ に対して、 N 個のデータ $x_{ij}, i=1, 2, \dots, N$ が観測されたとする。 x_{ij} は r 個のクラス $C_k, K=1, 2, \dots, r$ に分類されている。また、各 X_j に対して、 $m(j)$ 個のファジィ集合 $F_{tj}, t=1, 2, \dots, m(j)$ を定義する。

アルゴリズムでは、データ集合 D を C_k に分類する場

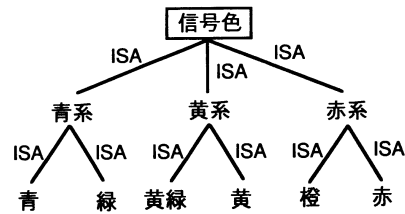


図1 信号色に関する背景知識

合、どの属性 X_j のファジィ集合 F_{tj} を用いるかを決定する。分類できない場合には、背景知識を用いてファジィ集合を再構成し、アルゴリズムを継続する。具体的なアルゴリズムを次に示す。

[手順1] データ集合 D におけるクラス C_k の出現度合 G_k を次式より求める。

$$G_k = \left(\frac{\sum_{i \in C_k} \mu^1(i)}{\sum_{i \in D} \mu^1(i)} \right), \quad 1=0, 1, 2, \dots \quad (1)$$

ただし、 $\mu^1(i)$ は決定木の第1層での i 番目のデータに対する重視度合いであり、 $\mu^0(i)=1$ とする。

[手順2] 第1層でのデータ集合 D のエントロピー $H(D)$ を計算する。 $H(D)$ は D の情報の不確かさを表す。

$$H(D) = - \sum_{k=1}^r G_k \log G_k \quad (2)$$

[手順3] 第1+1層において、属性 X_j の F_{tj} によりデータ集合 D 进行分类する。 X_j に関する条件付きエントロピー $H(D|j)$ を計算する。 $H(D|j)$ は属性 X_j の F_{tj} によって、 D 进行分类した場合の情報の不確かさを表す。

$$H(D|j) = \sum_{t=1}^{m(j)} (g_{tj} \times (-\sum G_k \log G_k)) \quad (3)$$

ただし、

$$g_{tj} = \left(\sum_{j=1}^n \mu_{F_{tj}}(x_{ij}) / \sum_{t=1}^{m(j)} \sum_{i=1}^N \mu_{F_{tj}}(x_{ij}) \right) \quad (4)$$

$$\mu^{1+1}(i) = \mu^1(i) \times \mu_{F_{tj}}(x_{ij}) \quad (5)$$

[手順4] 平均相互情報量 $I(j)$ を次式により求める。

$I(j)$ は X_j の $F_{i,j}$ によって得られた情報量の確かさを表す。

$$I(j) = H(D) - H(D|j) \quad (6)$$

[手順5] $I(j)$ が最大となる X_j で D を分割する。

[手順6] 第1層において、 $I(j) \leq 0$ となる場合には、データ集合 D を再構成する。まず、

$$\{j, L, S \mid f^{L,s}_j \in D, s=1, 2, \dots, M\} \quad (7)$$

となる (j, L, S) を得る。ただし、 $f^{L,s}_j$ は背景知識での第 j 属性に関する第 L 層目の第 S 番目の抽象属性のファジィ集合である。 $f^{L,s}_j$ とISA関係にある $f^{L-1,s}_j$ と $f^{L+1,s}_j$ を D に加える。なお、 (j, L, S) が得られない場合には再構成をしない。

[手順7] 新たなデータ集合 D を用いてアルゴリズムを継続する。ただし、次の条件が満足される場合には、アルゴリズムを停止する。

$$G_k > \alpha \quad \text{または} \quad \sum_{i \in D} \mu^1(i) < \beta \quad (8)$$

ただし、 α はクラス C_k の最大占有率を表すしきい値であり、 β は最小データ数を表すしきい値である。

[手順8] 手順7の停止則が適用されない場合には、 $l+1$ として手順1に戻り、アルゴリズムを継続する。

4. 数値例

簡単な数値例を用いて、ファジィID3の有用性を検証する。いま、運転手が一直線の道路を走行し、前方の信号から、“加速”か“減速”、“定速”を判断する。観測データを表1に示す。与えられた各属性のファジィ集合と背景知識(図1も含む)から、走行の状況判断に対するファジィルールの決定木を構築した。

決定木の構築過程を図2に示す。距離の属性でデータ集合を分類後、ファジィ集合“普通”で $I(j)=0$ となり、分類不可能となった。背景知識を参照し、抽象属

表1 運転手の判断データ

データ	信号色(nm)	降水量(mm/h)	距離(m)	クラス
1	485.0	0.01	2.3	加速
2	490.0	0.01	2.3	加速
3	485.0	0.01	48.0	減速
4	490.0	0.01	48.0	減速
5	635.0	49.0	11.2	減速
6	560.0	16.0	1.9	加速
7	670.0	16.0	1.9	加速
8	730.0	49.0	39.5	減速
9	542.5	12.0	5.1	加速
10	490.0	9.5	37.0	減速
11	405.0	8.4	98.0	定速
12	740.0	7.1	120.0	定速

性のファジィ集合“青系”をデータ集合 D に加えて、アルゴリズムを継続し、最終結果を得た。

5. おわりに

本手法はファジィ集合を集約・細分化するので、ファジィルールの数の調整手法でもある。今後、より具体的な応用例に適用し有効性を検討する。なお、本研究の一部は、1994年度阪南大学産業経済研究所助成による研究の一環として行った。

参考文献

- 1) J.R. Quinlan : Discovering Rules by Induction from Large Collections of Examples, Expert Systems in the Micro Electronics Age, Edinburgh University Press (1979)
- 2) 馬野など : ID3に基づくファジィ決定木の油中ガス分析診断への適用について、第4回インテリジェントFAシンポジウム, pp. 201-204 (1993)
- 3) M. Nunez : The Use of Background Knowledge in Decision Tree Induction, Machine Learning, Vol. 6, pp. 231-250 (1991)

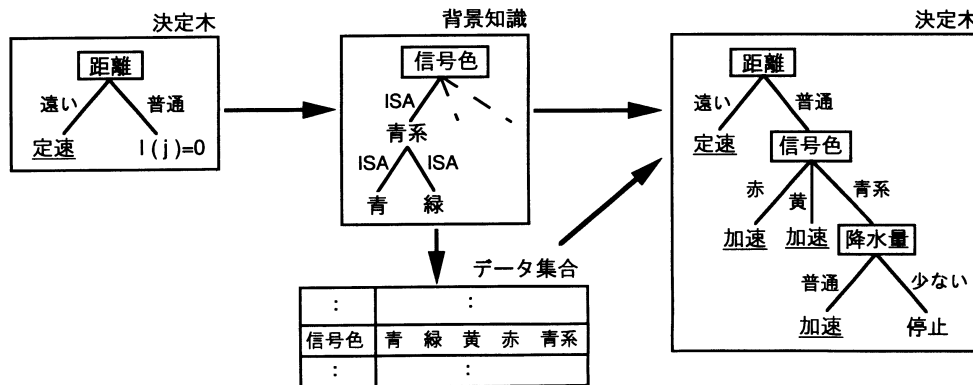


図2 決定木の構築過程