

2005

AND/OR演算子を学習するファジィID3  
Fuzzy ID3 with Ability of Learning for AND/OR Operators

阪南大学 林 勲, ○柴崎 真二

松下電器産業(株) 中央研究所 小澤 順

Isao Hayashi, Shinji Shibasaki, Hannan University

Jun Ozawa, Central Research Laboratories, Matsushita Electric Industrial Co. Ltd.

**Abstract.** The Fuzzy ID3 is a powerful method to generate decision trees of fuzzy logic rules. However, results don't always satisfy user requests since AND operators and OR connections are already fixed. The algorithm proposed in this paper has the ability of learning for AND/OR operators.

1. はじめに

ID3(Interactive Dichotomizer3)アルゴリズム<sup>1)</sup>はif-thenルールの決定木を生成する。平均相互情報量により、テスト回数の期待値が最小になるようにデータを分類する。また、ID3を拡張したものとしてファジィID3<sup>2)</sup>がある。ファジィ集合による決定木を生成する。しかし、ファジィルールのAND演算子が代数積<sup>3)</sup>であるので、満足する結果が得られるとは限らない。また、生成された決定木と同じ表現がある場合でも、ファジィルールを再構成できない。

本論文では、AND演算子をパラメータ付きのt-norm<sup>3)</sup>で定義し、演算子を学習する新たなファジィID3を提案する。本手法では、最大の平均相互情報量のパラメータをもつt-normを演算子に採用し、背景知識<sup>4)</sup>を用いて生成された決定木の集約・細分化を行う。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。経験的知識に基づいた決定木の調整が可能である。ファジィルールはOR結合されているので、本手法はOR結合の調整手法ともいえる。

2. t-normとは

t-norm  $\textcircled{T}$ とは、 $T(x_1, x_2) : [0, 1] \times [0, 1] \rightarrow [0, 1]$ なる境界条件、単調性、交換性、結合性を満足する関数である。代表的なt-normとして、次の演算子がある。

$$\text{激烈積} : x_1 \wedge x_2 = \begin{cases} x_1 & (x_2=0) \\ x_2 & (x_1=0) \\ 1 & (x_1, x_2 > 0) \end{cases} \quad (1)$$

$$\text{限界積} : x_1 \odot x_2 = 0 \wedge (x_1 + x_2 - 1) \quad (2)$$

$$\text{代数積} : x_1 \cdot x_2 = x_1 x_2 \quad (3)$$

$$\text{論理積} : x_1 \wedge x_2 = \min\{x_1, x_2\} \quad (4)$$

t-norm演算子には次のような大小関係がある。

$$\wedge \leq \odot \leq \cdot \leq \wedge \quad (5)$$

また、パラメータ付きのt-normが提案されている。例えば、Schweizerのt-norm<sup>3)</sup>では次式を用いる。

$$T(x_1, x_2) = 1 - p \sqrt[p]{(1-x_1)^p + (1-x_2)^p - (1-x_1)^p (1-x_2)^p} \quad (6)$$

$0 \leq p \leq \infty$

ただし、pはパラメータである。0 ≤ p ≤ ∞の範囲でpを変えることにより、激烈積から論理積まで表現できる。

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 $x_i, i=1, 2, \dots, N$ が観測されたとする。 $x_i$ 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 $F_{t,j}, t=1, 2, \dots, m(j)$ を定義する。具体的な手順を示す。

[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。

$$G_k = \left( \frac{\sum_{i \in C_k} \mu^1(i)}{\sum_{i \in D} \mu^1(i)} \right), \quad 1=0, 1, 2, \dots \quad (7)$$

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わす。

$$H(D) = - \sum_{k=1}^r G_k \log G_k \quad (8)$$

[手順3] 第1+1層において、属性 $X_j$ の $F_{t,j}$ によりデータ集合Dを分類する。 $X_j$ 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 $F_{t,j}$ によって、Dを分類した場合の情報の不確かさを表す。

$$H(D|j) = \sum_{t=1}^{m(j)} (g_{t,j} \times (-\sum G_k \log G_k)) \quad (9)$$

ただし、

$$g_{t,j} = \left( \frac{\sum_{j=1}^n \mu_{F_{t,j}}(x_{i,j})}{\sum_{t=1}^{m(j)} \sum_{i=1}^n \mu_{F_{t,j}}(x_{i,j})} \right) \quad (10)$$

$$\mu^{1+1}(i) = \mu^1(i) \textcircled{T} \mu_{F_{t,j}}(x_{i,j}) \quad (11)$$

ここで、 $\textcircled{T}$ は(8)式のt-normである。0 ≤ p ≤ 100として、pの値を変える。

[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 $F_{t,j}$ によって得られた情報量の確かさを表す。

$$I(j) = H(D) - H(D|j) \quad (12)$$

[手順5]  $I(j)$ が最大となるパラメータ $p$ の値を求め、 $I(j)$ が最大となる属性 $X_j$ でデータ集合 $D$ を分割する。  
 [手順6] 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。

$$G_k > \alpha \quad \text{または} \quad \sum_{i \in D} \mu^1(i) < \beta \quad (15)$$

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

決定木の生成後、 $X_j$ の第1層より下位層がまったく同じ複数の部分集合が存在する場合には、集約・細分化を行う。 $X_j$ の上位層の第 $l-1$ 層目の $F_{i,j}$ について、

$$\{j, L, S \mid F_{i,j} = f^{L,j}, f^{L,j} \in BK\} \quad (16)$$

となる $(j, L, S)$ が存在する場合、 $F_{i,j}$ の代わりに、 $f^{L,j}$ とISA関係にある $f^{L-1,j}$ と $f^{L+1,j}$ を採用して決定木を変更する。ただし、 $f^{L,j}$ は背景知識BKに含まれる第 $j$ 属性に関する第 $L$ 層目の第 $S$ 番目のファジィ集合を表す。なお、(16)式が成立しない場合には、決定木を変更しない。

[手順7] 手順6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

提案したアルゴリズムの有用性を検証するため、資本金、従業員数、売上高で企業を評価する事例を考えた。表1にデータを示す。売上高は多少少ないが資本金と従業員数が大きい企業を安定企業とし、資本金と売上高の大きい企業を望ましい企業とした。また、それ以外を発展途上企業に分類した。資本金に関する背景知識を図1に示す。データ集合と背景知識から、企業分類に対するファジィルールを構成した。

集約以前の決定木の結果を図2に示す。また、背景

表1 企業を分類するデータ

資本金(億円)	従業員数(人)	売上高(億円)	クラス
290	12,000	2,000	発展途上企業
300	21,000	1,300	発展途上企業
1,000	160,000	3,300	安定企業
1,670	103,000	3,700	安定企業
1,010	97,000	4,200	安定企業
1,600	70,000	10,700	望ましい企業
1,630	90,000	11,000	望ましい企業
1,560	78,000	8,800	望ましい企業



図1 資本金に関する背景知識

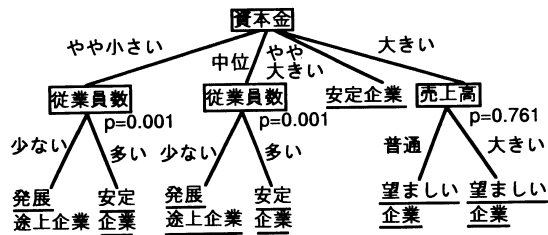


図2 集約前の決定木

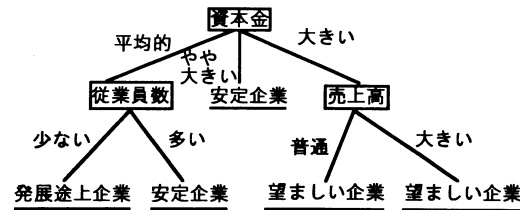


図3 集約後の決定木

知識により、資本金が“平均的”に集約した決定木を図3に示す。ただし、 $p=0.761$ などはt-normのパラメータの値を表す。一般に、資本金が売上高が満足できる企業は望ましい企業と考えられるので、 $p=0.761$ のAND演算子がOR的であるのは我々の直感と一致している。

#### 5. おわりに

AND/OR演算子を学習するファジィID3により、観測者の好みに対応した分類が可能となった。今後、より具体的な応用例に適用し、本手法の有効性を検討する必要がある。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) 水本 : ファイ集合とファジィ推論, 第3回ファジィシステムシンポジウム, pp. 37-48 (1987)
- 4) M. Nunez : The Use of Background Knowledge in Decision Tree Induction, Machine Learning, Vol. 6, pp. 231-250 (1991)