12th Fuzzy System Symposium (Tokyo,June 4～7,1996)

THA2-2

学習型ファジィＩＤ３の高速化に関する一考察

A Study on High-Speed Algorithm of Fuzzy ID3 with Ability of Learning for AND/OR Connectives

林　勲*　　小澤　順**

Isao Hayashi　　Jun Ozawa

* 阪南大学　経営情報学部

** 松下電器産業(株)　中央研究所

* Dept. of Industrial Computer Science, Hannan University

** Central Research Laboratories, Matsushita Electric Industrial Co. Ltd.

Abstract. An ability of learning for AND/OR operators is discussed here to overcome a drawback of fuzzy ID3. The fuzzy ID3 is a powerful method to acquire fuzzy decision tree. However, it is nearly impossible to obtain the most suitable fuzzy rules for representing the data sets since the fuzzy ID3 has a couple of problems, i.e., a problem of a lack of representation and an adjusting problem. In our fuzzy ID3, AND/OR operators are formulated using t-norm and t-conorm connectives with parameters and each parameter is adjusted using golden section method, such that takes the maximal value of the average mutual information. By adjusting parameters, the proposed fuzzy ID3 gives more accurate fuzzy rules for representing the data sets. If t-conorm connective is selected as AND/OR operator, the decision tree has more flexible representation.

1．はじめに

if-thenルールを獲得する手法に，Quinlan[1)]が提案したID３(Interactive Dichotomizer 3)アルゴリズムがある．観測データとクラスからなるデータ集合が与えられると，平均相互情報量[2)]が最大となる入力属性を繰り返し抽出し，if-thenルールの決定木を構築する．また，馬野ら[3),4)]が提案したファジィID３がある．ファジィID３では，データ集合とファジィ集合が与えられた場合，メンバーシップ値による平均相互情報量を求め，最大となる入力属性を抽出し，ファジィルールの決定木を構築する．

本論文では，さらに，平均相互情報量が最大となるAND演算子とOR演算子も抽出するAND/OR演算子学習型ファジィID３[5)]を提案する．本手法では，パラメータ付きt-norm演算子とt-conorm演算子[6)]を用いて，AND演算子とOR演算子を定義し，黄金分割法[7)]により平均相互情報量が最大となるパラメータを探索する．黄金分割法は探索のための収束性にすぐれ，アルゴリズムが簡便であるという特徴がある．ただし，黄金分割法では，目的関数が単峰性を満足する必要がある．本論文では，平均相互情報量を目的関数とした場合に，この単峰性を満足する条件について議論する．

一方，馬野らはメンバーシップ関数とt-norm演算子を学習するチューニング型ファジィID３[4)]を提案している．しかし，本手法では，t-conorm演算子も学習するので，決定木の構築過程において，t-conorm演算子が入力属性間のAND/OR演算子として抽出された場合，新たなサブ決定木を発生する．通常のファジィID３[8)]では，過去に抽出された入力属性は必ずファジィルールに用いられ，データ集合を最適に表現するファジィルールが必ずしも得られるとは限らない．本手法では，抽出された入力属性に関与しない新たなファジィルールを構成できる．簡単な事例を用いて，その有用性を検討する．

2．学習型ファジィID３のアルゴリズム

関数$f:R^1\rightarrow R^1$が与えられた場合，最適化問題を解くアルゴリズムとして黄金分割法は有用である．いま，目的関数fが区間$L=[a,b]\subset R^1$内に最大点$x^*\in L$をもつものとする．このとき，fが区間Lで単峰ならば，任意のx_1, x_2, $x_1<x_2$に対して，次のように，区間Lの幅を縮小できる．

$$\left.\begin{array}{l} f(x_1)\geq f(X_2)\ \text{のとき，}\ x^*\in L^{new}=[a,\ x_2] \\ f(x_1)<f(X_2)\ \text{のとき，}\ x^*\in L^{new}=[x_1,\ b] \end{array}\right\} \qquad (1)$$

ただし,

$$x_1 = a + \frac{(3-\sqrt{5})}{2}(b-a) \qquad (2)$$

$$x_2 = a + \frac{(\sqrt{5}-1)}{2}(b-a) \qquad (3)$$

黄金分割法では，区間Lの縮小を繰り返すことにより，最大点x^*を探索できる．また，新たな区間L^{new}での点x_1とx_2は(2)式と(3)式のどちらか一方のみを計算すればよいので，収束時間の短縮化が図れる．

学習型ファジィID３のアルゴリズムを次に示す．いま，属性x_j, $j=1,2,\cdots,n$に対してN個の観測データ$\mathbf{x}_i=(x_{i1},x_{i2},\cdots,x_{in})$, $i=1,2,\cdots,N$が得られたとする．各x_jに対して，m(j)個のファジィ集合F_{tj}, $t=1,2,\cdots,m(j)$を定義する．観測者は，この観測データ$\mathbf{x}_i$をr個のクラスC_k, $k=1,2,\cdots,r$に分類する．アルゴリズムの手順を次に示す．

［手順１］決定木の第L層目でのデータ集合DのクラスC_kに対する出現度合$G_k{}^L$を次式より求める．

$$G_k{}^L = \left(\sum_{i\in C_k} \mu^L(i) \Big/ \sum_{i=1}^{N} \mu^L(i)\right), \quad k=1,2,\cdots,r \qquad (4)$$

ただし，決定木の層は$L=0,1,2,\cdots$とする．また，$\mu^L(i)$は第L層目での観測データ$\mathbf{x}_i$に対する重視度合いであり，$\mu^0(i)=1$とする．

［手順２］第L層目でのデータ集合DのエントロピーH(D)を計算する．

$$H(D) = -\sum_{k=1}^{r} G_k{}^L \log_2 G_k{}^L \qquad (5)$$

［手順３］ファジィ集合F_{tj}を用いて，次層の属性$x_j{}^*$とAND/OR演算子を以下の手順により決定する．

［手順３－１］属性x_jに関する条件付きエントロピーH(D|j, p)を次のように計算する．

$$H(D|j,p) = \sum_{t=1}^{m(j)} \left(g_{tj} \times \left(-\sum_{k=1}^{r} G_k^{L+1} \log_2 G_k^{L+1}\right)\right),$$

$$j=1,2,\cdots,n, \quad p \geqq 0 \tag{6}$$

ただし，

$$g_{tj} = \left(\sum_{i=1}^{N} \mu_{F_{tj}}(x_{ij}) \Big/ \sum_{t=1}^{m(j)} \sum_{i=1}^{N} \mu_{F_{tj}}(x_{ij})\right) \tag{7}$$

$$\mu^{L+1}(i) = \begin{cases} \mu^{L}(i) \,Ⓣ\, \mu_{F_{tj}}(x_{ij}) & \text{；t-norm演算子の場合} \\ \mu^{L}(i) \,⊥\, \mu_{F_{tj}}(x_{ij}) & \text{；t-conorm演算子の場合} \end{cases} \tag{8}$$

である．ここで，Ⓣはパラメータp_1をもつt-norm演算子を表し，⊥はパラメータp_2をもつt-conorm演算子を表す．

［手順３－２］平均相互情報量I(j, p)を次式により求める．

$$I(j,p) = H(D) - H(D|j,p) \tag{9}$$

［手順３－３］I(j, p)が最大となるt-norm演算子のパラメータp_1^*とt-conorm演算子のパラメータp_2^*をそれぞれ，黄金分割法を用いて求める．

［手順３－４］次式を満足する属性x_jでの最適なパラメータp^*を求める．

$$I(j,p^*) = \max\{I(j,p_1^*),\ I(j,p_2^*)\} \tag{10}$$

［手順３－５］次式を満足する属性$x_j{}^*$を求める．

$$I(j^*,p^*) = \max_{j} I(j,p^*), \quad j=1,2,\cdots,n \tag{11}$$

［手順３－６］属性$x_j{}^*$を第L+1層目の属性として採用する．$p^*=p_1^*$の場合には，p_1^*のt-norm演算子をAND演算子として採用し，$p^*=p_2^*$の場合には，t-conorm演算子のOR演算子が採用されるので，属性$x_j{}^*$を用いた新たなサブ決定木を発生する．

［手順４］次の条件を満足する場合，第L+1層目ではデータ集合Dに関するファジィルールが獲得されたとして，それ以後の層でのアルゴリズムの手順を停止する．

$$G_k^{L+1} > \alpha \quad \text{または} \quad \sum_{i=1}^{N} \mu^{L+1}(i) < \beta \tag{12}$$

ただし，αはクラスC_kが全クラスの中で占める最大占有率を表すしきい値であり，βは観測データの最小度合いを表すしきい値である．

［手順５］手順４の停止則を満足しない場合には，L+1=Lとして手順１に戻り，アルゴリズムを継続する．

３．学習型ファジィID３における性質

黄金分割法を用いるには，パラメータp_1, p_2に関して，平均相互情報量I(j, p)が単峰性を満足する必要がある．ここでは，単峰性を満足するための条件について議論する．

説明を簡単にするために，パラメータpをもつt-norm演算子のみ議論する．パラメータpの微小変化に対する平均相互情報量I(j, p)への影響dI(j, p)/dpから，I(j, p)の単峰性に関する性質を導く．dI(j, p)/dpは次のようになる．

$$\frac{dI(j,p)}{dp} = \frac{dH(D)}{dp} - \frac{dH(D|j,p)}{dp} = -\frac{dH(D|j,p)}{dp} \tag{13}$$

ここで，

$$\frac{dH(D|j,p)}{dp} = -\sum_{t=1}^{m(j)} \left(g_{tj} \times \frac{d}{dp} \sum_{k=1}^{r} G_k^{L+1} \log_2 G_k^{L+1}\right) \tag{14}$$

である．いま，

$$M_k(p) = \sum_{i \in C_k} \mu^{L+1}(i) = \sum_{i \in C_k} \left(\mu^{L}(i) \,Ⓣ\, \mu_{F_{tj}}(x_{ij})\right) \tag{15}$$

$$M_0(p) = \sum_{i=1}^{N} \mu^{L+1}(i) = \sum_{i=1}^{N} \left(\mu^{L}(i) \,Ⓣ\, \mu_{F_{tj}}(x_{ij})\right) \tag{16}$$

とすると，(14)式のパラメータpの微小変化に対する第L+1層のエントロピーH(D)への影響Eは次のようになる．

$$\begin{aligned} E &= \frac{d}{dp} \sum_{k=1}^{r} G_k^{L+1} \log G_k^{L+1} \\ &= \sum_{k=1}^{r} \frac{d}{dp}\left(\frac{M_k(p)}{M_0(p)} \log_2 \frac{M_k(p)}{M_0(p)}\right) \\ &= \sum_{k=1}^{r}\left(\frac{d}{dp}\left(\frac{M_k(p)}{M_0(p)}\right) \log_2 \frac{M_k(p)}{M_0(p)} + \frac{M_k(p)}{M_0(p)}\left(\frac{d}{dp}\left(\log_2 \frac{M_k(p)}{M_0(p)}\right)\right)\right) \\ &= \sum_{k=1}^{r}\left(\frac{1}{(M_0(p))^2}\left(\log_2 \frac{M_k(p)}{M_0(p)} + \frac{1}{\log_e 2}\right) \times \left(\frac{dM_k(p)}{dp} M_0(p) - \frac{dM_0(p)}{dp} M_k(p)\right)\right) \end{aligned} \tag{17}$$

ここで，

$$E_1 = \log_2 \frac{M_k(p)}{M_0(p)} + \frac{1}{\log_e 2} \tag{18}$$

$$E_2 = \frac{dM_k(p)}{dp} M_0(p) - \frac{dM_0(p)}{dp} M_k(p) \tag{19}$$

とすると，E_1, E_2の符号の正負によって，Eの符号が決定し，dI(j, p)/dpの符号が決定する．

E_1, E_2の符号の正負によるI(j, p)の単峰性について議論しよう．いま，t-norm演算子として，次のSchweizerのt-norm[6]を用いる．

$$T(x_1,x_2) = 1-\left((1-x_1)^p+(1-x_2)^p-(1-x_1)^p(1-x_2)^p\right)^{1/p}$$

$$p \geqq 0 \tag{20}$$

ただし，パラメータpを$p \geqq 0$の範囲で変化させることにより，激烈積($p \to 0$)から論理積($p \to \infty$)までを表現できる．

E_1において，$M_k(p)/M_0(p)$は，第k番目のクラスの出現度合を表している．$E_1=0$とすると，$M_k(p)/M_0(p)=0.368$となる．一般に，データ集合では多くのクラス数を設定するので，$k \geqq 3$ならば，$E_1<0$となる．

一方，(20)式のt-norm演算子は単調関数であるので，E_2はpのほとんどの定義域で単調関数となっている．この意味から，解析的に，$E_2=0$となるpが存在する．

Eは$E_1 \times E_2$をすべてのクラスに対して合計するので，Eはpに関して単調関数となり，E=0となるpが存在する．したがって，dI(j, p)/dpは単調関数となり，dI(j, p)/dp=0となるpが存在する．これより，I(j, p)は単峰性を満足しているといえる．t-conorm演算子の場合にも同様なことがいえる．

4. 数値例

ここでは，本手法の有用性を示すため，AND/OR演算子を学習する事例を示す．

4.1 AND演算子のパラメータを学習する事例

黄金分割法における性質を説明するため，ここでは，学生が就職先を決定する場合に，資本金，従業員数，売上高から企業を評価する事例を取り上げる．学生が判断したデータ集合を表1に示す．また，資本金と従業員数，売上高に関するファジィ集合を図1から図3に示す．$\alpha=0.8$, $\beta=1.0$として，本手法のアルゴリズムにより，企業分類に関するファジィルールを構築した．

表1　企業を分類するデータ集合

資本金（億円）	従業員数（千人）	売上高（億円）	クラス
290	12	2,000	発展企業
300	21	1,300	発展企業
1,000	160	3,300	安定企業
1,670	103	3,700	安定企業
1,010	97	4,200	安定企業
1,600	70	10,700	有望企業
1,630	90	11,000	有望企業
1,560	78	8,800	有望企業

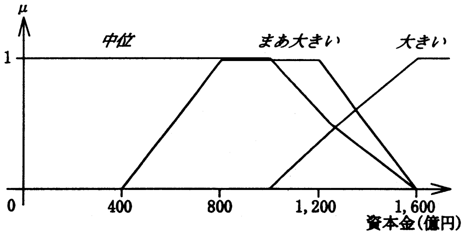

図1　資本金に関するファジィ集合

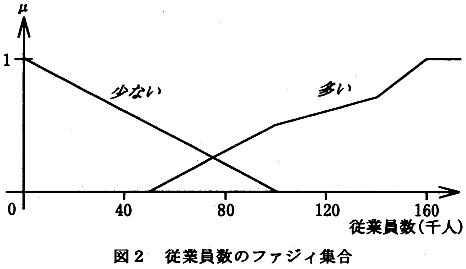

図2　従業員数のファジィ集合

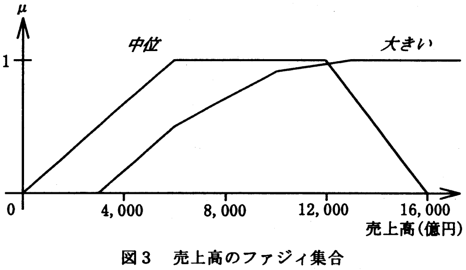

図3　売上高のファジィ集合

得られた決定木を図4に示す．結果として，$p^*=0.001$と$p^*=0.703$のt-norm演算子が得られた．一般に，資本金か売上高が満足できる企業は有望企業と考えられるので，平均演算子に近い$p^*=0.703$のAND演算子が得られたのは，我々の直感と一致している．パラメータpの値に対する$I(j,p)$の変化を図5に示す．また，(13)式から(19)式を用いて得られた$dI(j,p)/dp$の変化を図6に示す．図5から，$I(j,p)$は$p^*=0.703$において最大値$I(j,p^*)=0.2935$をとる単峰性関数であることがわかる．また，図6より，$p^*=0.703$において，$dI(j,p)/dp=0$となっていることがわかる．

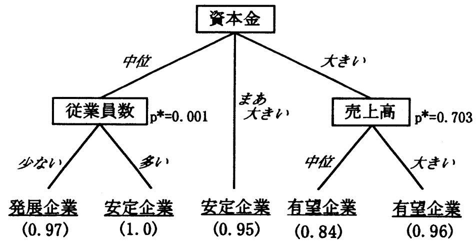

図4　獲得された決定木

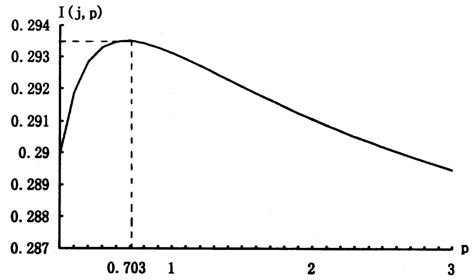

図5　パラメータpに対する$I(j,p)$の変化

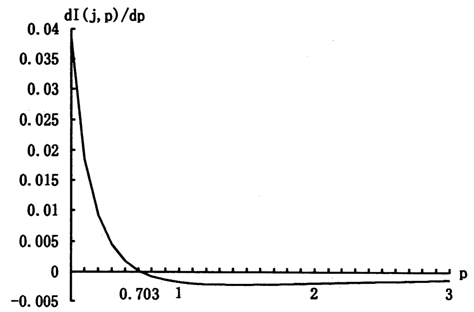

図6　パラメータpに対する$dI(j,p)/dp$の変化

次に，新たに作成した30種類の観測データに対して，全探索，最急降下法，黄金分割法によるパラメータpの平均収束時間の比較を行った．結果を表2に示す．黄金分割法は，全探索の1/10，最急降下法の1/3の時間で最適解を探索できる．

表２　各種法の比較

	全探索	最急降下法	黄金分割法
収束時間	4.39　秒	1.30　秒	0.43　秒

注1）全探索では，pのきざみ幅を0.1とし，0≦p≦50の範囲で探索した．
注2）最急降下法では，学習係数を0.3とした．

４．２　OR演算子のパラメータを学習する事例

OR演算子を学習する事例を次に示す．ここでも，学生が就職先を決定する事例を用いた．ただし，表３にデータ集合を示し，ファジィ集合を図７から図９に示す．α=0.8，β=1.0として，企業分類に関するファジィルールを構築した．

表３　企業を分類するデータ集合

資本金（億円）	従業員数（千人）	売上高（億円）	クラス
398	4	2,995	発展企業
530	21	1,300	発展企業
1,695	118	5,300	安定企業
1,800	150	13,700	安定企業
1,700	71	6,005	安定企業
1,770	50	12,400	有望企業
750	60	10,000	有望企業
1,405	120	11,800	有望企業

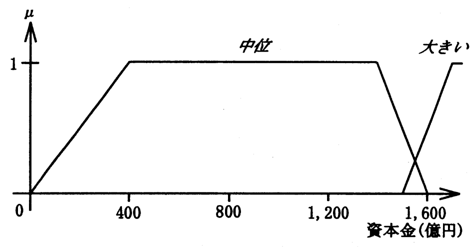

図７　資本金のファジィ集合

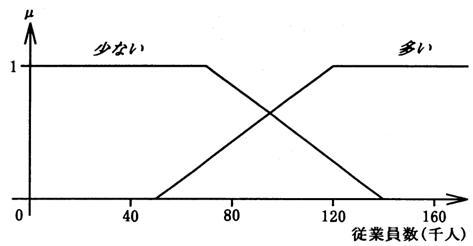

図８　従業員数のファジィ集合

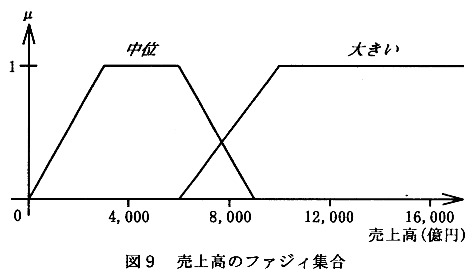

図９　売上高のファジィ集合

得られた決定木を図10に示す．AND/OR演算子として，t-conorm演算子を採用したので，サブ決定木が発生し，売上高に関与しない２個のファジィルールが構成されている．

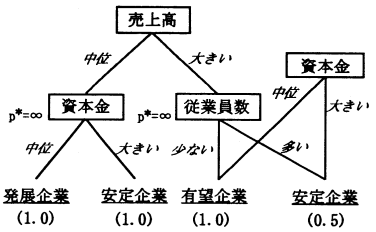

図10　獲得された決定木

５．おわりに

本手法では，データ集合をより表現するファジィルールの構築が可能となった．しかし，ファジィ集合の種類や形状に依存しており，平均演算子に関する取り扱い方法も議論していない．今後，これらの課題を解決する必要がある．なお，本研究の一部は，1995年度阪南大学産業経済研究所助成研究「ファジィ演算子を自動獲得するファジィID３の提案」の成果報告である．

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) 菊池豊彦：情報科学Ｉ，コロナ社（1990）
3) 馬野元秀，岡本宏隆，嶋野逸生，田村坦之，河内二三夫，梅津祐久，木下淳一：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）
4) 馬野元秀，北辻佳憲，嶋野逸生，田村坦之：チューニング型ファジィID３によるファジィ決定木の生成，日本ファジィ学会ファジィ・コンピューティング研究部会第７回ワークショップ（1995）
5) 林勲，小澤順：背景知識を用いたAND/OR演算子学習型ファジィID３，第11回ファジィシステムシンポジウム，pp. 137-140（1995）
6) 水本晴雅：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）
7) 今野浩，山下浩：非線形計画法，日科技連（1978）
8) C. Z. Janikow : Fuzzy Processing in Decision Trees, Proceedings of the International Symposium on Artificial Intelligence, pp.360-367 (1993)

［問い合わせ先］
580　大阪府松原市天美東５－４－３３
阪南大学 経営情報学部
林　勲
TEL: 0723-32-1224 (ext. 8412)
FAX: 0723-36-2633
Internet: ihaya@hannan-u.ac.jp
HomePage: http://www.hannan-u.ac.jp/~ihaya