

ニューラルネットワークの入出力間相互情報量を用いた データ自動分割アルゴリズム

東北学院大学 菊地 登志子
自治医科大学 岸 浩一郎
宇都宮大学 宮道 壽一

1. はじめに

知識発見の手がかりを抽出するための一手法として、著者らはデータ集合がもつ分布構造の特徴を可能な限り機械的に抽出し、その特徴に基づいた分割結果が自動的に得られる効率的データ自動分割アルゴリズムを提案した[1]。本稿では、このアルゴリズムの概略と、分布構造の特徴を最もよく捉えた分割数を推定する手法を解説し、財務データによる地方銀行の分類に適用することでその有用性を示す[2]。

2. 相互情報量を用いた効率的データ自動分割アルゴリズム

効率的データ自動分割アルゴリズムはニューラルネットワークの競合学習を基本としているため、ネットワークの構造は入力層と、入力データの分類を行う出力層（競合層）のみをもつ2層の階層型ネットワークである[3]。入力層は入力データの次元数と等しいユニット数、出力層は予め指定された分割数と等しいユニット数で構成される。出力層の各ユニットは入力層の全ユニットと完全結合している。

2.1 学習アルゴリズム

$w_{ij}(t)$ を時刻 t における入力層ユニット i から出力層ユニット j への結合荷重とし、ここではこの結合荷重をランダムに選出した入力データで初期化する。時刻 t における入力層ユニット i への入力データを $x_i(t)$ とし、各入力データから作られる入力ベクトル $\mathbf{x}(t) = (x_1(t), x_2(t),$

$\dots, x_n(t))$ と、各入力ユニットから出力ユニット j への結合荷重ベクトルとの距離 $d_j (j=1, \dots, K)$ を次式で計算する。

$$d_j = \sum_{i=1}^n (x_i(t) - w_{ij}(t))^2 \quad (2-1)$$

n は入力データの次元数（入力層ユニット数）、 K は分割数（出力層ユニット数）である。 d_j が最小となる出力ユニットを j^* （勝者）とし、結合荷重 $w_{ij}(t)$ を (2-2)、(2-3) 式に従って更新する。

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t) \quad (2-2)$$

$$\Delta w_{ij}(t) = \begin{cases} (1 - \mu(\tau)) g_w (x_i(t) - w_{ij}(t)) & (j = j^*) \\ (1 - \lambda(\tau)) g_l (x_i(t) - w_{ij}(t)) & (j \neq j^*) \end{cases} \quad (2-3)$$

g_w を勝者の学習係数、 g_l を敗者の学習係数と呼び、 $0 < g_w, g_l \leq 1$ 、 $g_l \ll g_w$ とする。全入力データをランダムな順序で1回ずつ入力層に提示するのを1サイクルとし、 τ はこのサイクル数を表している。1サイクルで、 t は入力データ数だけ増加する。 $\mu(\tau)$ 、 $\lambda(\tau)$ は、以下の2.2に示す r 値によりその形状が決まるシグモイド型増加関数である。

$\mu(\tau) = 0.0$ 、 $g_l = 0.0$ とすると競合学習のアルゴリズムになり、 $\mu(\tau) = 0.0$ 、 $\lambda(\tau) = 0.0$ とすると漏えい学習のアルゴリズムになる。効率的データ自動分割アルゴリズムは、入力データやネットワークに与えられた初期値に応じて $\mu(\tau)$ 、 $\lambda(\tau)$ を設定し、漏えい学習から競合学習へ自動的に移行していくアルゴリズムである。

2.2 r 値

次のような2つの確率事象系 A, B を考える .

$$A = \begin{pmatrix} a_1, & a_2, & \dots, & a_\alpha \\ p(a_1), & p(a_2), & \dots, & p(a_\alpha) \end{pmatrix}$$

$$B = \begin{pmatrix} b_1, & b_2, & \dots, & b_\beta \\ p(b_1), & p(b_2), & \dots, & p(b_\beta) \end{pmatrix}$$

$p(a_l)$ は事象 a_l ($l=1, \dots, \alpha$) の出現確率, $p(b_m)$ は事象 b_m ($m=1, \dots, \beta$) の出現確率とする . このとき, 2つの事象系 A, B 間の相互情報量は (2-4) 式で定義される .

$$I(A; B) = -\sum_{l=1}^{\alpha} p(a_l) \log_2 p(a_l) + \sum_{l=1}^{\alpha} \sum_{m=1}^{\beta} p(a_l, b_m) \log_2 p(a_l | b_m) \quad (2-4)$$

$p(a_l, b_m)$ は事象 a_l , 事象 b_m の結合確率, $p(a_l | b_m)$ は, 事象 b_m のもとでの事象 a_l の条件付確率とする .

ここでは, 入力データの値の分布とニューラルネットワークによる分割結果のグループとの間の相互情報量を, 入力層の各ユニットごとに計算する . 入力層ユニット i に提示される全入力データの値 x_i^h ($h=1, \dots, N$) の取り得る範囲を M 等分したそれぞれの区間を事象 a_l ($l=1, \dots, M$) とし, その区間に入る入力データ x_i^h の頻度分布から $p(a_l)$ を計算する . N は入力データ数である . ニューラルネットワークの入力層に全入力データを提示して, その出力から各入力データの属するグループ b_m を求める . このグループを事象 b_m , b_m に属するデータ数の分布を求め $p(b_m)$ とする . 各グループ b_m ($m=1, \dots, K$) ごとに入力分布との結合確率 $p(a_l, b_m)$, 条件付確率 $p(a_l | b_m)$ を求め, (2-4) 式に基づいて各入力ユニットごとに相互情報量 I_i ($i=1, \dots, n$) を計算する . これらの相互情報量の最大値 I_{\max} と, 分割数 K の対数との比を (2-5) 式で計算し, これを r 値とよぶ .

$$r = \frac{I_{\max}}{\log_2 K} \quad (2-5)$$

相互情報量 I_i が取り得る値の最大値は $\log_2 K$ となるため, r 値は 0 と 1 の間の値をとる .

効率的データ自動分割アルゴリズムでは, 第 1 サイクルの学習終了後に r 値を計算し, この r 値の値を用いて第 2 サイクル以降では (2-3) 式の $\lambda(\tau)$ を以下のようにサイクル数 τ に応じて減少するよう設定する .

$$\lambda(\tau) = \frac{1}{1 + \phi(\tau)} \quad (2-6)$$

$$\phi(\tau) = \exp\left\{\frac{a(s(1-r) - \tau)}{s(1-r) - 1}\right\} \quad (2-7)$$

s および a は定数で, 以下の実験では $s = 40.0$, $a = 5.0$ を用いた . ただし, $r > 0.8$ では $\mu(\tau) = 0.0, \lambda(\tau) = 1.0$ として競合学習を行い, $0.6 < r \leq 0.8$ では $\mu(\tau) = 0.0$ として勝者の学習係数を一定にした漏えい学習, $r \leq 0.6$ では $\mu(\tau) = \lambda(\tau)$ として勝者の学習係数もサイクル数に応じて減少する漏えい学習を自動的に選択する . このように, 効率的データ自動分割アルゴリズムの学習係数は, r 値により入力データに応じて自動的に調整される . 詳細は文献 [1] を参照されたい .

2.3 分割結果の評価方法

データ分割における結果の良さの評価には, (2-8) 式で表される各データと属するグループと中心との距離の summed deviation を用いることとし, ここではいくつかの異なる初期条件から得られた分割結果の summed deviation の最小値が一致する場合, その最小値に対応する分割を最適解とする .

$$\text{summed deviation} = \frac{1}{N} \sum_{h=1}^N \sum_{j=1}^K \sum_{i=1}^n v_{hj} (x_i^h - \mu_{ij})^2 \quad (2-8)$$

ただし, v_{hj} は h 番目の入力データが出力層ユニットの j 番目のユニットのグループに属する場合 1, その他の場合 0 とする. μ_{ij} は j 番目のユニットのグループに属する入力データ群の重心ベクトルの第 i 成分である.

3. データの分布構造の特徴を捉えた分割数の推定

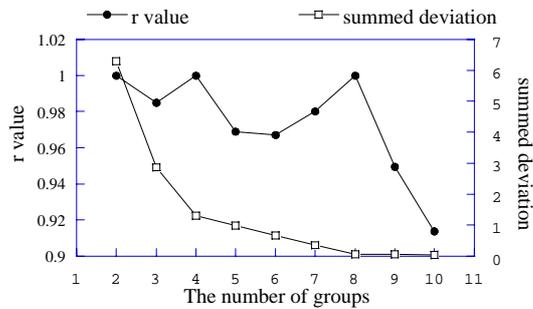
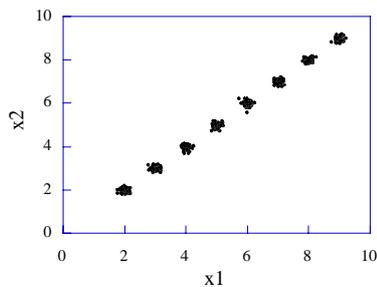
r 値は入力データの各成分の分布と分割結果の相関を表す相互情報量から計算され, 効率的データ自動分割アルゴリズムでは第 1 サイクルの学習終了後に求められる. 様々な分割数を指定してデータ分割を行った場合, 入力データの

分布の特徴を捉えた分割数では, 学習の早い段階で最適な分割結果に近い出力が得られる. 言い換えれば, この分割結果は他の分割数の第 1 サイクル終了後の結果より入力データと高い相関をもち, r 値は他の分割数より大きな値を示す. そこで, r 値によりデータの分布構造の特徴を捉えた分割数の推定がどの程度可能かをサンプルデータを用いて検討した[2].

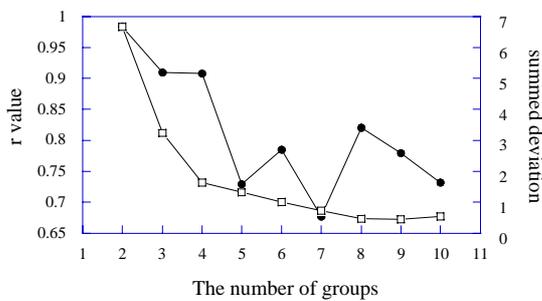
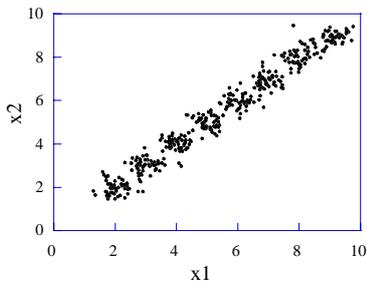
3.1 データの分布構造を捉えるための指標としての r 値

各成分が 8 つの正規分布 (それぞれの平均値 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 標

(a) standard deviation 0.1



(b) standard deviation 0.3



(c) standard deviation 0.5

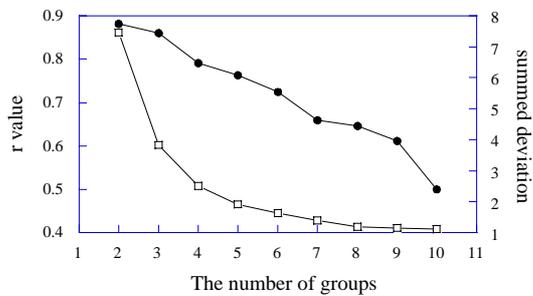
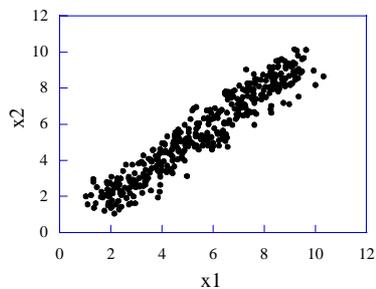


図 1 合成された正規分布 5 次元データの分布と r 値

標準偏差はすべて 0.1) を合成した分布である 5 次元のデータ (データ数は 400) をもとに, 8 つの正規分布の標準偏差を 0.3, 0.5 とした 3 種類のデータを用意する. 標準偏差 0.1 としたデータでは, 完全に分離した 8 つのグループからなり (図 1 (a), 左側), これらの 8 つのグループが得られる 8 分割がデータの分布構造の特徴を捉えた分割数となる. 標準偏差 0.5 ではグループにかなりのオーバーラップが生じ (図 1 (c), 左側) 散布図からも分布の特徴を捉えた分割数は判断できない.

このデータの 2 分割から 10 分割までを, 結合荷重に異なる初期値を与えた 10 種類のネットワークに, 効率的データ自動分割アルゴリズムを適用し実行した. それぞれの分割数ごとに, 10 種類のネットワークから得られた summed deviation が最も小さな値をもつネットワークを選択し, その r 値と summed deviation の値を分割数ごとに示す (図 1, 右側). 8 つの正規分布の標準偏差が 0.1 のデータでは, 2 分割, 4 分割, 8 分割に (図 1 (a), 右側), 標準偏差が 0.3 のデータでは, 6 分割, 8 分割に r 値の増加が見られる (図 1 (b), 右側). 何れの場合も 8 分割の r 値は増加し, 8 分割という特徴を r 値で捉えることが十分可能であることを示してい

る. 一方, 標準偏差 0.5 としたデータではもはや 8 分割の特徴は存在せず, r 値はほぼ単調に減少している (図 1 (c), 右側).

標準偏差が 0.1 のデータは全くオーバーラップのない 8 つのグループからなるため, 2 分割, 4 分割の結果からもデータのもつ大まかな特徴は捉えられる. しかし, 8 分割の結果からは最も詳細な特徴を得ることが可能で, r 値が大きく増加する分割数が複数個ある場合には, 分布構造の特徴を最もよく捉えた分割数として, 分割数の多いほうを選択するのが適しているといえる.

4. 実データへの応用

効率的データ自動分割アルゴリズムを日本の地方銀行の財務データによる分類に適用し, 実データに対しても r 値によりデータの分布構造の特徴を捉えた分割数の推定が十分可能であることを示す. 更に, 地方銀行のデータに対し主成分分析を行い, 提案するシステムから得られた分割結果を分割数ごとに主成分平面上に表示して, どのようなグループが抽出されたとき r 値が増加したかを調べる.

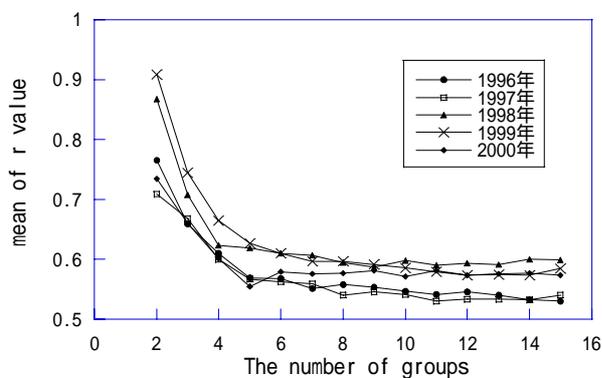


図2 地方銀行の分割における r 値

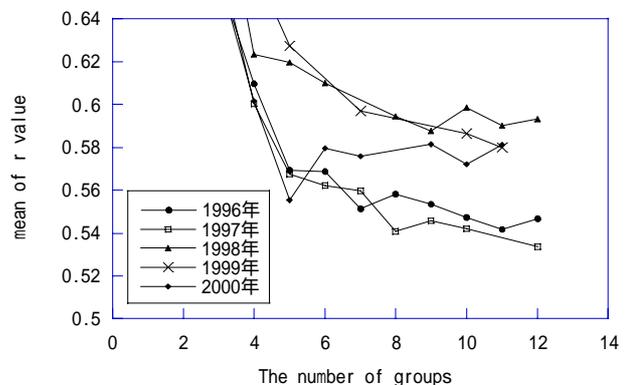


図3 地方銀行の5分割以上の r 値

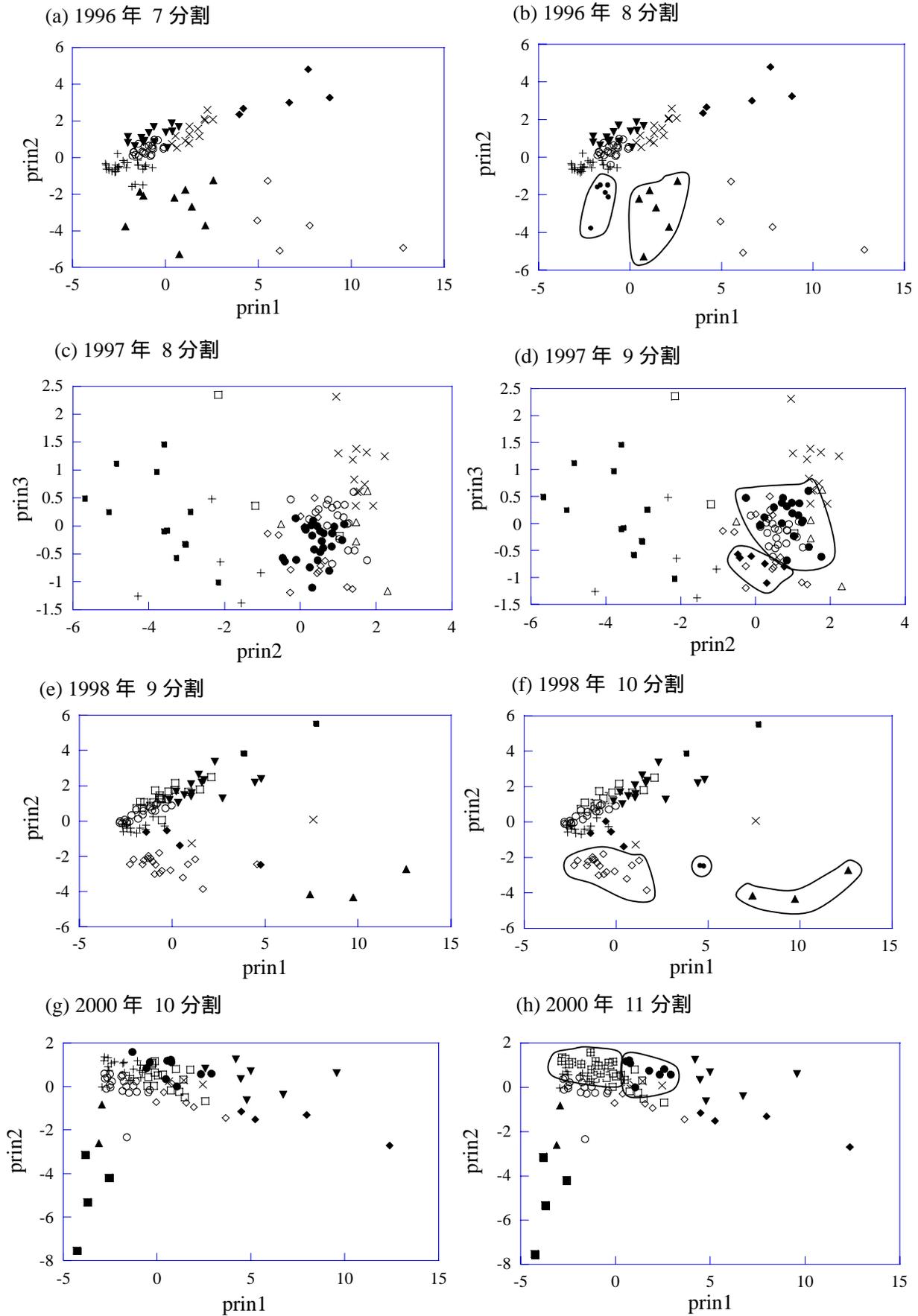


図4 地方銀行の分割結果と主成分得点

4.1 入力データ

日本の地方銀行の1996年から2000年までの3月期決算における財務データを入力データに使用した。用いた財務データは、資本金、総資産、株主資本、株主資本比率(%)、外部負債、資金量、ROE(%), 含み損益、最高経常利益、経常収益、業務純益、経常利益、利益の13項目の数値である。

4.2 地方銀行の分類における r 値

4.1のデータをもとに結合荷重に異なる初期値を与えた20種類のネットワークに効率的データ自動分割アルゴリズムを適用し、2分割から15分割までを行った。各分割数ごとに、20種類のネットワークから得られたr値の平均である平均r値を計算した(図2)。すべての年で、2分割から5分割までの平均r値は減少したため5分割以上を対象に、最適解が得られた分割数のみの平均r値を図3に表示した。図3から、データの分布構造の特徴を捉えた分割数は、1996年では8分割、1997年9分割、1998年10分割、2000年11分割と判断することができる。ただし、1999年には平均r値の増加する分割数は存在しなかったため、分布構造の特徴を捉えた分割数を推定することはできなかった。

次に、どのようなグループが抽出されたとき平均r値が増加したかを調べ、推定された分割数の妥当性を検討した。ただし、13次元データの分布構造の特徴は視覚化することが困難なため、ここでは地方銀行のデータに対し主成分分析を行い、主成分平面上に分割結果から得られるグループを表示することにした(図4)。分布構造の特徴を捉えた分割数として推定された分割結果を右列(図4(b)(d)(f)(h))に、その分割数よりグループ数が1つ少ない分割結果を左列(図4(a)(c)(e)(g))に表示する。何れも分割数が

少ないうちは、あまり類似していないデータも一つのグループとしてまとめられていたのが、分割数の増加に伴い他と分離度の高いグループが抽出されているのが分かる。このようなより密度の高い(データの集中した)グループが抽出できれば、グループに属するデータがもつ特徴をより鮮明に捉えることが可能になる。これらのグループが抽出された分割数でr値が増大することは、分布構造の特徴を自動的に捉えるための手法として提案手法が有用であることを示している。

5. まとめ

著者らの提案によるニューラルネットワークの入出力間で求めた相互情報量に基づくr値を用いた効率的データ自動分割アルゴリズムを紹介した。更に、r値をデータの分布構造の特徴を捉えた分割数を推定する指標として用いることを検討し、その有用性を地方銀行の実データを用いて示した。

文献

- [1] 菊地登志子, 岸浩一郎, 宮道壽一, "相互情報量により学習パラメータを調整した効率的データ自動分割アルゴリズム," 電子情報通信学会論文誌(D-), vol.J82-D- , no.4, pp.660-668, 1999.
- [2] 菊地登志子, 岸浩一郎, 宮道壽一, "ニューラルネットワークの入出力間相互情報量に基づくデータ自動分割システム," 日本社会情報学会学会誌, 第12巻1号, pp.43-56, 2000.
- [3] D.E.Rumelhart, J.L.McClelland, and The PDP Research Group, "Parallel Distributed Processing," MIT Press, Cambridge, Massachusetts, 1986.

