

クラスタ例からの学習 — クラスタ属性の利用法の改良 (2)

Lerning from Cluster Examples — An Improvement of a method for handling Attributes of Clusters (2)

神鷹 敏弘*
 Toshihiro Kamishima

赤穂 昭太郎*
 Shoutarou Akaho

Abstract: Learning from cluster examples is a composite task of two common classification tasks: learning from examples and clustering. Learning from cluster examples involves an attempt to acquire a rule that can be used to partition an unseen object set from a given example set. Having an established method for a task is useful in situations where, though an appropriate partition for a given object set is clear, a way of deriving the partition is not known to users. Though we have been applied the method to problems in two domains, the estimated rules that cannot fully appropriate partitions was not learned. Formerly, we presented a method that considers attributes of clusters, such as arrangements of objects belonging to the same clusters. We show improvements of performance by applying it to artificial data.

キーワード: machine learning, clustering, learning from examples, learning from cluster examples

1 はじめに

クラスタ例からの学習 [4] とは、クラスタリングに例からの学習の要素を加えたもので、与えられた分類対象集合を適切に分割するための規則を、事例から獲得する学習問題である。

クラスタ例からの学習では、属性ベクトルによって分類対象集合を表現する。ただし、例からの学習のように一つの属性ベクトルで分類対象を表現するのではなく、多数のベクトルによって分類対象集合を表現する点異なる。文献 [2] では、三種類の属性ベクトルによって分類対象集合を表現した。文献 [3] では、各クラスタ (分類対象を分割したときの、分割された各部分集合) に関連付けた四種類目の属性ベクトルを新たに導入した。しかし、文献 [3] では、この種の属性を有効に利用できなかった。本論文では、この属性を利用する手法を改良する。この手法を人工的なデータに適用し、その有効性を示す。

以下、2 節ではクラスタ例からの学習について、3 節では学習方法について、4 節では実験の方法と結果について、そして、5 節ではまとめを述べる。

2 クラスタ例からの学習

図 1 はクラスタ例からの学習の概略を示したもので、左側には学習段階、右側には学習結果を利用した分割段階を表す。学習段階では、学習事例集合 EX から分割を獲得するための規則を学習し、推定段階では、学習した規則を用いて、未知の分類対象集合 O_U に対する推定分割 $\hat{\pi}_U$ を求める。

クラスタリングは、事前に定めた規準に基づき分類対象集合を演繹的に分割するものである。よって、この規準を明確に提示できない場合には、利用者にとって適切

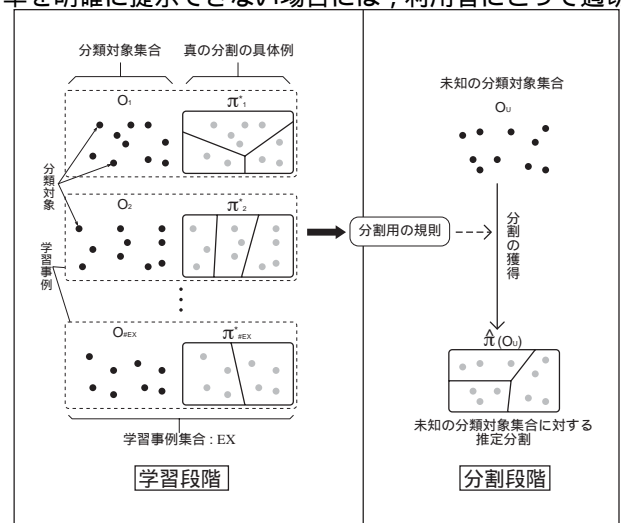


図 1: クラスタ例からの学習の概要

*産業技術総合研究所, 脳神経情報研究部門, 〒 305-8568
 茨城県つくば市梅園 1 丁目 1-1, 産総研つくば中央第 2,
 tel. 0298-61-5835, e-mail mail@kamishima.net, ホーム
 ページ <http://www.kamishima.net/>
 National Institute of Advanced Industrial Science and Technology, Neu-
 roscience Research Institute, AIST Tsukuba Central 2, 1-1, Umezono 1-
 chome, Tsukuba, Ibaraki, 305-8568 Japan

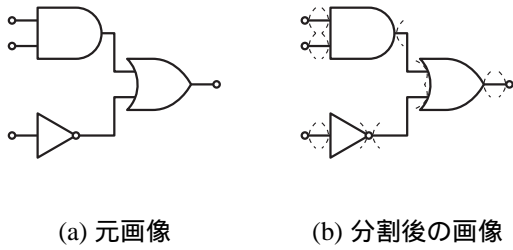


図 2: 画像のセグメンテーションの例

な分割を得ることができない。そこで、規準そのものを示す代りに、よい分割の具体例を提示し、その例から帰納的に分割のための規準を推定する問題がクラスタ例からの学習である。規準を示すことは困難だが、適切な分割そのものを示すことは容易である問題の一例として画像のセグメンテーション問題について述べる。

セグメンテーションとは、図 2 の (a) の原画像を、(b) のように意味があると考えられる部分ごとに切り分ける (点線で示した部分で分離する) 操作である。この図の場合、論理回路図の画像が与えられ、それを図面の部品ごとに切り分けている。このセグメンテーションは画像認識の過程でよく利用され、その結果の良否は認識結果に大きく影響する。

このセグメンテーションは、画像を構成する要素 (ピクセルや線分など) を属性ベクトルで表現し、一種のクラスタリングの技法を利用して行われることが多い。しかし、クラスタリングを適用するためには、分類対象が同じクラスに分類されるかどうかを判定するための、分割の規準をあらかじめ明示しておく必要があるが、これは一般には困難である。なぜならば、画像の構成要素は位置や明るさといった観測値で表されているのに対し、分割の目標は図面の部品といった抽象的な概念であり、両者は大きく隔たっていて、直観的に人間が対応を見いだすことが難しいからである。ところが、この場合でも、適切な分割の具体例を与えることは容易であり、この具体例をもとにクラスタ例からの学習によってセグメンテーションを行うことができれば有用であると考えられる。

以下、クラスタ例からの学習の定式化について述べる。 EX は $\#EX$ 個の事例からなる集合 $\{(O_1, \pi_1^*), (O_2, \pi_2^*), \dots, (O_{\#EX}, \pi_{\#EX}^*)\}$ であり、各学習事例は分類対象集合 O_I と、これに対する真に適切な分割の具体例 π_I^* の組である。分類対象集合 O は分類対象 o^i からなる集合であり、分割 π は、 O の排他的で網羅的な部分集合であるクラスタの組 $\{C^1, C^2, \dots, C^{\#\pi}\}$ である。

個々の分類対象集合 O_I は、属性ベクトルを分類対象集合の一部と関連付け、ベクトルにその部分の特徴を記述させることで表現する。これらの属性ベクトルは関連

- 付けられる部分によって以下の四種類に分けられる。
- $A(o^i)$: 分類対象属性 分類対象 o^i に関連付けられたベクトル。ドットパターンでは点の位置などを表すために用いる。
 - $A(p^{ij})$: 分類対象対属性 分類対象の対 p^{ij} に関連付けられたベクトル。ドットパターンでは点の間の距離などを表すために用いる。
 - $A(\pi)$: 分割全体の属性 分割全体に一つだけ関連付けられるベクトル。ドットパターンではクラスタの数などを表すために用いる。
 - $A(C^J)$: クラスタ属性 新しく導入する属性で、分割を構成するクラスタ C^J に関連付けられる。ドットパターンではクラスタ内のドット数などを表すために用いる。

これらの属性は、その実現値の計算の方法によって、前者二つと後者二つに分けられる。前者では、属性値は分類対象集合が与えられた時点で定まっている。一方、後者では、ある分割を定め、その分割の状態と分類対象集合の両方に依存して属性値が計算される。実装では、属性値が必要なときに、分類対象と分類対象対の属性値、及び、分割の状態をもとに分割全体やクラスタ属性の値を計算している。

未知の分類対象集合のこれら四種類の属性を参照して、その集合の真の分割を求めることのできる規則を獲得するのが学習の目標である。

3 学習と分割推定の方法

3.1 分割推定の方法

本論文では、分割の適切さを表す評価関数を定め、与えられた分類対象集合に対する可能な分割の中から、この評価関数を最大にするものを探索し、それを推定分割とする。この評価関数は、MAP 原理に基づくもので、ある分割を定めたとき、その分割が適切な分割であるという事象と、この集合の属性値ベクトルとの次式の結合確率で表される。

$$\Pr[\pi = \pi^*, A(\pi), \{A(C)\}, \{A(o)\}, \{A(p)\}] \quad (1)$$

ただし、 $\pi = \pi^*$ は π が真の分割と一致する事象で、 $\{A(O)\}$ と $\{A(P)\}$ はそれぞれ、 O と分類対象対全体の集合 P の要素に関連付けられた属性値ベクトル全体の集合、 $\{A(C)\}$ は分割中のクラスタに関連付けられた属性値ベクトル全体の集合である。この結合確率は、非常に多くの属性値に依存し、また、属性値の数も変動するため非常に計算が困難である。そこで、種類の異なる属性は独立であるとの仮定を用いて、次の項の積に分解し、そ

れぞれを簡略化する．

$$\Pr[\{A(O)\}, \{A(P)\}] \quad (2)$$

$$\Pr[\pi=\pi^*|\{A(O)\}, \{A(P)\}] \quad (3)$$

$$\Pr[A(\pi)|\pi=\pi^*] \quad (4)$$

$$\Pr[\{A(C)\}|\pi=\pi^*] \quad (5)$$

式 (2) は，分割の状態には依存しないので最大化に關しては無視できる．式 (3) と式 (4) を求める手法は，詳しくは文献 [2] にあるが，その概要のみを次節で述べる．本稿では，最後の式 (5) の計算方法の詳細を 3.3 節で述べる．

次に，上記の評価関数を最大にするような分割を探索する方法について述べる．分類対象集合に対する可能な分割の総数は非常に多く，最適な分割を実際には求めることができないので，次の方法で準最適な分割を探索する．

最初に，各クラスタに分類対象が一個ずつだけ含まれる初期分割を作成する．この分割に，次の二種類の操作を行い評価関数を最大にする分割を次の時点の分割とする．

1. 二個のクラスタを一個のクラスタに併合

2. あるクラスタの要素一つを別のクラスタに移動

ただし，式 (4) が 0 の場合を考慮するため若干の工夫を行った（詳細は文献 [2]）．そして，これらの操作では評価関数を大きくする分割を発見できなかったときに探索を終了し，そのときの分割を最終的な推定分割とする．

3.2 式 (3) と式 (4) の計算と学習の方法

式 (3) と式 (4) の計算と学習の方法について文献 [2] の内容を簡単にまとめる．

最初に式 (3) の計算のために，学習段階で学習事例集合から次の確率分布関数を推定しておく．

$$f_1(p^{ij}) = \Pr[\text{in}(p^{ij}, \pi^*)=1; A(p^{ij}), A(o^i), A(o^j)]$$

ただし， $\text{in}(p, \pi)$ は，分類対象対 p が分割 π の同じクラスタの要素であるとき 1，そうでないとき 0 をとる関数である．ここで， $\text{in}(p^{ij}, \pi^*) = 1$ となる事象は，あらゆる可能な分割のうち分類対象対 p^{ij} が同じクラスタ要素となっている分割のいずれかが生じる事象と等価であることに注目する．そして，確率 $f_1(p^{ij})$ を， $A(p^{ij})$ ， $A(o^i)$ ， $A(o^j)$ 証拠とした，Dempster&Shafer 理論での，この事象の基本確率とみなす．分割の推定段階で，分類対象対集合の全ての要素についてこの確率を求め，それらを Dempster&Shafer 理論の確率結合則を用いて統合す

ることで，式 (3) の定数倍が次式で計算できる．

$$\prod_{p \in P^+} f_1(p) \times \prod_{p \in P^-} (1 - f_1(p))$$

ただし， P^+ は，分類対象対集合の要素のうち， $\text{in}(p, \pi) = 1$ となるものからなる集合で， P^- はそれ以外の要素からなる集合である．

関数 $f_1(p^{ij})$ は， EX を変換した学習事例から，通常の例からの学習のアルゴリズムを適用して解くことができる．

次に，式 (4) の計算と学習の方法についてまとめる．最初に，学習事例集合 EX の各要素について，分類対象集合 O_I を π_I^* に分割したときの分割全体の属性の属性値ベクトルを計算する．この属性値ベクトルは，式 (4) の分布から発生したものと見なせる．よって，これら属性値ベクトルから確率密度関数が推定できる．分割段階では，この確率密度関数に，属性値を単に代入すれば式 (4) の値を計算できる．

3.3 式 (5) の学習と計算の方法

ここでは，式 (5) のクラスタ属性の集合 $\{A(C)\} = \{A(C^1), A(C^2), \dots, A(C^{\#\pi})\}$ の確率密度の学習と計算の方法について述べる．クラスタ数 $\#\pi$ は分割全体の属性の中に含まれているため，この確率密度を計算するときには $\#\pi$ は与えられているものとする．さらに，属性ベクトルの集合の各要素は独立に発生するモデルを考えると，式 (5) は次式で表される

$$\Pr[\{A(C)\}|\pi=\pi^*] \approx \prod_{s=1}^{\#\pi} \left(\Pr[\{a^s(C)\}|\pi=\pi^*] \right)$$

ただし， $\#\pi$ は属性ベクトルの要素数で， $\{a^s(C)\}$ は属性値ベクトル $A(C^1), A(C^2), \dots, A(C^{\#\pi})$ の s 番目の要素を取り出した集合である．

この式中の結合確率 $\Pr[\{a^s(C)\}|\pi=\pi^*]$ に次のモデルを採用した．

$$\Pr[\{a^s(C)\}|\pi=\pi^*] \approx \left(\prod_{J=1}^{\#\pi} \Pr[a^s(C^J)|\pi=\pi^*; \Theta^s] \right) \Pr[\Theta^s; \mathbf{H}^s] \quad (6)$$

これは，パラメータ分布 $\Pr[\Theta^s; \mathbf{H}^s]$ は \mathbf{H}^s を超パラメータとし，この分布に従って発生したパラメータ Θ^s で指定される属性値分布 $\Pr[a^s(C^J)|\pi=\pi^*; \Theta^s]$ に従って $\#\pi$ 個の属性値が独立に発生するモデルである．以下， s 番目の属性値について論じるものとし，添字の s は省略する．

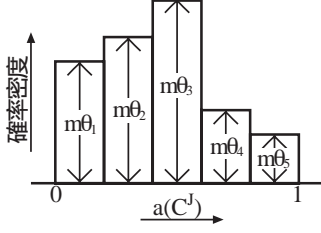


図 3: $\Pr[a(C^J)|\pi=\pi^*; \Theta^s]$ の分布

文献 [3] では、パラメータ分布には対数正規分布、属性値分布にはベータ分布を用いていたが、今回は、属性値分布として図 3 のヒストグラム型の分布を採用した。この図は分割数 $m = 5$ の場合を示しており、計算量の点を考慮して、実験ではこの値を採用した。パラメータは $\Theta = (\theta_1, \dots, \theta_{m-1})$, $\theta_m = 1 - \sum_{k=1}^{m-1} \theta_k$, かつ, $\theta_k > 0, k = 1 \dots m$ である。属性値 $a(C^J)$ の値域は $[0, 1]$ であるとする。このとき確率密度の制約

$$\int_0^1 \Pr[a(C^J)|\pi=\pi^*; \Theta] d\Theta = 1$$

が成立することは容易に検証できる。属性値の入るピンの番号を $k(a(C^J))$ で表すと、属性値 $a(C^J)$ の確率密度は $m\theta_{k(a(C^J))}$ となる。ここで、 $a(C^J), \dots, a(C^{\#\pi})$ のうち k 番目のピンに入る属性値の数を n_k で表すと属性値分布は次式になる。

$$\begin{aligned} \Pr[\{a^s(C)\}|\pi=\pi^*; \Theta] \\ &\approx \left(\prod_{J=1}^{\#\pi} (m\theta_k)^{k(a(C^J))} \right) \\ &= \left(\prod_{k=1}^m (m\theta_k)^{n_k} \right) \end{aligned}$$

一方のパラメータの分布には、 Θ の制約を考慮し、 $H = (\eta_1, \dots, \eta_m)$ を超パラメータとする多変量ベータ分布を採用した。

$$\Pr[\Theta; \mathbf{H}] = \frac{\prod_{k=1}^m \theta_k^{\eta_k - 1}}{B(\{\eta\})}$$

ただし、 $B(\{\eta\})$ は次の多変量ベータ関数。

$$B(\{\eta\}) = \frac{\prod_{k=1}^m \Gamma(\eta_k)}{\Gamma(\sum_{k=1}^m \eta_k)}$$

未知の分類対象集合 O_U を π_U に分割にしたとき、式 (5) は次式となる。

$$\begin{aligned} \Pr[\{A(C)\}|\pi=\pi^*] \\ &= E_{\Theta} \left[\left(\prod_{k=1}^m (m\theta_k)^{n_k} \right) \frac{\prod_{k=1}^m \theta_k^{\eta_k - 1}}{B(\{\eta\})} \right] \\ &= m^{\#\pi} \frac{\prod_{k=1}^m \Gamma(n_k + \eta_k)}{\Gamma(\#\pi + \sum_{k=1}^m \eta_k)} \times \frac{1}{B(\{\eta\})} \end{aligned}$$

次に、学習事例集合から超パラメータ Θ を EM アルゴリズム [1] を用いて学習する方法について述べる。学習事例 (O_J, π_J^*) について、 O_J を π_J^* に分割したときのクラス属性の属性値ベクトルの集合は、 $\Pr[\{A(C)\}|\pi=\pi^*]$ に従って発生したものであるため、これを学習事例とする。ここでも、属性値ベクトルの s 番目の属性値のみについて考えて s は省略し、 EX の I 番目の事例から得た属性値ベクトルの集合を $\{a_I\} = \{a_I(C^1), \dots, a_I(C^{\#\pi_I})\}$ で、 $\{a_I\}$ のうち k 番目のピンに入る属性値の数を n_{Ik} で表す。この集合を発生させた分布のパラメータは $\Theta_I = (\theta_{I1}, \dots, \theta_{I,m-1})$ で表記し、以下の関数を定める。

$$\begin{aligned} F_I(\{a_I\}, \Theta_I | \mathbf{H}) &= \prod_{k=1}^m (m\theta_{Ik})^{n_{Ik}} \frac{\prod_{k=1}^m \theta_{Ik}^{\eta_k - 1}}{B(\{\eta\})} \\ G_I(\{a_I\} | \mathbf{H}) &= \int F_I(\{a_I\}, \Theta_I | \mathbf{H}) d\Theta_I \\ E_{\Theta_I}[\log \theta_{Ik} | \{a_I\}, \mathbf{H}] &= \frac{\int F_I(\{a_I\}, \Theta_I | \mathbf{H}) \log(\theta_{Ik}) d\Theta_I}{G_I(\{a_I\} | \mathbf{H})} \end{aligned}$$

このとき、EM アルゴリズムでは、逐次的に現在の超パラメータ H を次式を最大にする H' に更新する。

$$\begin{aligned} Q(\mathbf{H}' | \mathbf{H}) &= \\ &\sum_{k=1}^m \left\{ (\eta'_k - 1) \frac{1}{\#\text{EX}} \sum_{I=1}^{\#\text{EX}} E_{\Theta_I}[\log \theta_{Ik} | \{a_I\}, \mathbf{H}] \right\} \\ &\quad - \log B(\{\eta'\}) \end{aligned}$$

実験では \mathbf{H} の初期値を $\eta_k = 1.0, 1 \leq k \leq m$ とした。

4 実験と考察

4.1 評価方法

ここでは、学習により獲得された規則を用いて、どれだけ真の分割に近い分割が推定されたかを評価する文献 [4] の方法について概略を述べる。クラスタ例からの学習では、 $\#\text{EX}$ 分割の交叉確認 ($\#\text{EX}$ は総事例数) である leave-one-out 法によって分割を評価した。真の分割と推定分割との類似性の評価には情報損失量 (RIL; Ratio of Information Loss) を用いた。これは、真の分割を推定するのに必要な情報量のうち、推定分割はどれだけの情報量を得ていないかという割合であり、0 から 1 の間の値をとり、0 のとき推定分割は真に適切な分割と一致する。交叉確認で求めた各事例の真の分割と推定分割の間の RIL の平均値によってどれだけ適切に推定できているかを評価した。

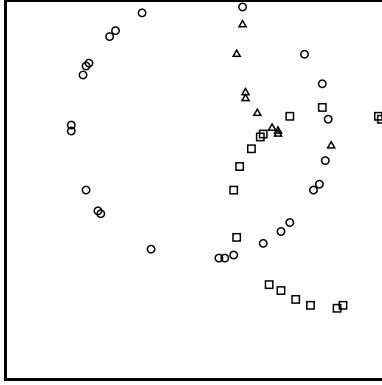


図 4: ドット・パターンの例

4.2 実験対象

実験では図 4 のような、同じクラスタの点が円周上に並んだドットパターンを対象にした。この分類対象集合には、局所的な特徴はなく、クラスタ中の要素全体が集まってはじめてクラスタとしての特徴が明確になる。よって、クラスタ属性を採用すれば、真の分割の推定の精度が向上することが予測される。もしこの分類対象集合について、クラスタ属性を用いた場合とそうでない場合を比較し、用いた場合に推定結果が改善されれば、本手法の有効性を示すことができると考える。

ドットパターンは以下の手順で生成した。各分類対象集合は 50 個の点で構成されており、3~4 個のクラスタに分割されている。各クラスタ内の点は半径 $1.3r$ の円周上に、中心方向に若干のガウスノイズを加えた位置に配置した。ただし、 r は各クラスタごとに与えたパラメータで、文献 [2] のドットパターンと同じものである。これは各クラスタを中心にしてこの r を半径とする領域が、他のクラスタのこの領域と接触し、かつ、重ならない条件のもとでランダムに定めている。

クラスタ属性以外の属性には、文献 [2] と同じものを採用した。使用する属性の組み合わせを変えた実験も行ったが、有意な差はみられなかった。

クラスタ属性には、上記の分類対象集合の特徴を考慮して、以下の 7 種類の属性について実験した。

1. クラスタ中の点のうち凸包の頂点に含まれるものの割合
2. 凸包の円形度 ($4\pi \times \text{面積} \times \text{外周長}^2$)
3. 推定中心の位置の標準偏差
4. 推定中心からの各点までの距離の標準偏差
5. クラスタの要素数の全体に対する割合
6. 重心までの距離の標準偏差
7. 重心からの距離でソートしたとき、近い方から 75% の順位にある点までの距離

属性値 a が $[0, 1]$ の範囲になり、また、サンプル平均 \bar{a}

表 1: クラスタ属性を 1 種類だけ用いた場合の RIL の平均

属性	RIL	属性	RIL
なし	.860	4	.869
1	.840	5	.870
2	.889	6	.874
3	.852	7	.878

がほぼ 0.5 となるように、計算時には $\exp(a \times -(\ln 2/\bar{a}))$ の変換を行った。ここで、3 と 4 番目の属性について補足しておく。推定中心とは、クラスタ内の任意の組み合わせの 3 点を通る円の中心を考え、それらの重心のことである。3 番目の属性は、円の各中心の位置の X と Y 方向の分散の和の平方根で、4 番目の属性は、推定中心から各点までの距離の標準偏差である。

4.3 実験結果と考察

4.3.1 クラスタ属性の選択

最初に、有効なクラスタ属性を選択するための以下の予備実験を行った。

前節の 7 種類のクラスタ属性のうちいずれか一つだけと、文献 [2] の属性を用いて分割を推定した。総事例数を 1000 個にし、leave-one-out テストを行い、RIL の平均を求めた結果を表 1 に示す。ただし、属性の項目の“なし”とは、クラスタ属性を全く用いない場合を示す。何もクラスタ属性を用いない場合と比較して、1 番と 3 番の属性を用いた場合に RIL が減少したので、これらの 2 種類の属性を次の実験では採用する。

ここに示したような属性選択の手法は例からの学習では一般的である。しかし、クラスタリングの結果の良否を定量的に評価できなため、このような属性選択手法は通常のクラスタリングでは不可能であった。これは、文献 [4] のクラスタ例からの学習の枠組みを採用し、推定結果を定量的に評価する方法を確立したことによってはじめて可能となったことに注意されたい。

4.3.2 クラスタ属性の有効性の検証

1 と 3 番目のクラスタ属性を用いて推定した分割の RIL の平均を図 5 に、そのエラーバーで標準偏差を示し、図 6 にはクラスタ属性を用いない場合の RIL の平均と標準偏差を示した。ただし、総事例数が 10, 30, 50, 100, 300, 500, 1000 の場合を示している。図には正確な数値は記してないが、1000 事例の値は、クラスタ属性を用いた場合が平均 0.839 で、標準偏差 0.1130、用いない場合は 0.860 と 0.0993 である。いずれのサンプル数においても、RIL の平均はクラスタ属性を用いた方が小さ

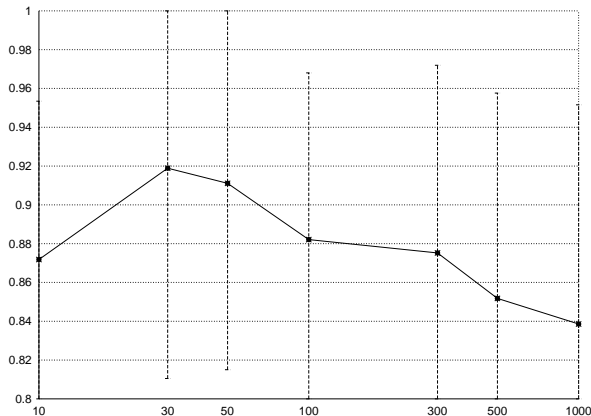


図 5: 各事例数でのクラスタ属性を用いた場合の RIL とその標準偏差

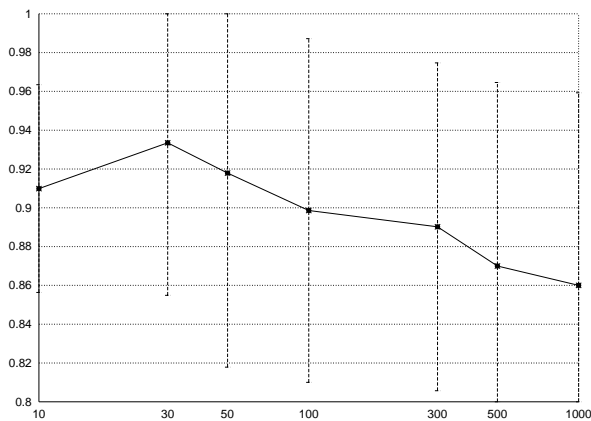


図 6: 各事例数でのクラスタ属性を用いない場合の RIL とその標準偏差

く、クラスタ属性を採用した方が良い。

さらに、クラスタ属性の有効性を厳密に確認するために t 検定を行った。事例数が n のとき、 i 番目の RIL の差を d_i で、 d_i の平均を \bar{d} であらわすと、 t 値は次式で表される。

$$t = \frac{\bar{d}}{\sqrt{\frac{1}{n} \sum (d_i - \bar{d})^2 / (n-1)}}$$

この t 値は、自由度 $n-1$ の t 分布に従うため、もしこの値が t 分布の 99% 点より大きければ、危険率を 1% で有意な差があるといえる。この t 値を、図 7 に示す。実線は t 値を表し、点線は危険率 1% での有意水準を表す。 t が有意水準より大きければ、クラスタ属性を用いた場合の方が、そうでない場合よりも有意に結果が良いことを示している。図では、事例数が 100 個の場合までは有意な差が見られない。しかし、事例数の増加にともない、 t は大きくなり 300 個以上では有意な差が見られる。

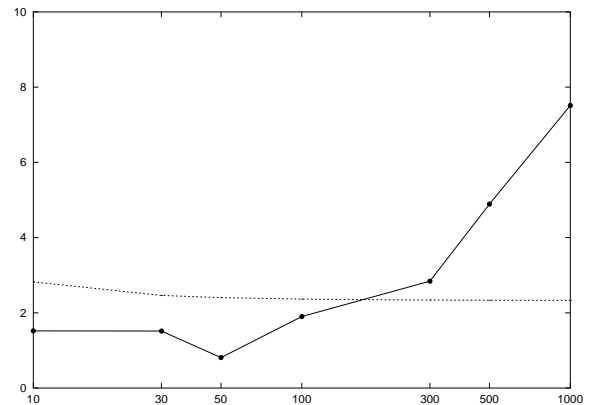


図 7: 各事例数での t 値と危険率 1% での有意水準

本手法では、“クラスタ属性の属性値ベクトル中の各要素は独立である”とか、“各属性値ベクトルは独立に発生する”などの仮定を導入している。しかし、これらの仮定にも関わらず、実験ではクラスタ属性を有効に利用できる結果がえられており、本手法が有効であることが示されている。

5 まとめ

本論文では、より適切な分割の推定を可能とするため、クラスタ属性についても考慮できるように、クラスタ例からの学習の手法を改良した。この手法を人工データであるドットパターンの分割問題に適用し、その有効性を示した。

今後は、文献 [2] にあるベクトルデータ画像の分割問題など、より実際の問題に近い対象に適用する予定である。

参考文献

- [1] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society (B)*, Vol. 39, No. 39, pp. 1–38, 1977.
- [2] 神鷹敏弘. *Learning from Cluster Examples*. 博士論文 (情報学), 京都大学, 2001.
- [3] 神鷹敏弘, 元吉文男. クラスタ例からの学習 — クラスタ属性の利用法の改良. 第 3 回情報論的学習理論ワークショップ (IBIS2000), pp. 81–86, 2000.
- [4] 神鷹敏弘, 新田克己. クラスタ例からの学習. *人工知能学会誌*, Vol. 12, No. 2, pp. 276–284, 1997.