

クラスタ例からの学習 — クラスタ属性の利用

Learning from Cluster Examples — Employing Attributes of Clusters

神嵐 敏弘

Toshihiro Kamishima

産業技術総合研究所

National Institute of Advanced Industrial Science and Technology (AIST)

mail@kamishima.net, http://www.kamishima.net/

赤穂 昭太郎

Shotaro Akaho

(同上)

s.akaho@aist.go.jp, http://staff.aist.go.jp/s.akaho/

元吉 文男

Fumio Motoyoshi

(同上)

f.motoyoshi@aist.go.jp

keywords: machine learning, clustering, learning from cluster examples, EM algorithm

Summary

Learning from cluster examples (LCE) 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 given examples. A general method for learning such partitioning rules is useful in any situation where explicit algorithms for deriving partitions are hard to formalize, while individual examples of correct partitions are easy to specify. In this paper, to improve estimation accuracy of LCE, we employ attributes of clusters and propose a method that can handle this type of attributes. We show improvements of performance by applying this method to artificial data.

1. はじめに

クラスタ例からの学習 (Learning from Cluster Examples; LCE) [神嵐 97] とは、クラスタリングに例からの学習の要素を加えた問題で、未知の分類対象集合の正しい分割を推定するための規則を学習事例集合から獲得する問題である。ただし、各学習事例は、特徴記述を付随させた分類対象集合である分類対象集合表現とその集合の正しい分割との対である。分割とは、分類対象集合を排他的で網羅的な部分集合に分けたものである。

文献 [Kamishima] では、分類対象集合の特徴の表現形式として多数の属性ベクトルを用いた形式を採用していた。これは、 k -means 法などのクラスタリング手法で用いられる表現形式、すなわち、集合中の各分類対象を、それらに関連づけた属性ベクトルで表現する形式を拡張したもので、分類対象だけでなく、分類対象対や分割全体にも属性ベクトルを関連付け、全部で三種類の属性ベクトルを用いていた。

本論文では、分類対象集合のより多様な特徴を考慮して分割を推定できるようにするために、新たに四種類目の属性ベクトルを導入してこの表現形式を拡張する。このベクトルは、各クラスタ (分割した分類対象集合の各部分集合) に関連付けたものである。本論文では、この新たな属性を利用する手法を提案し、これ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 LCE 問題の定式化について、3 節では分割の推定方法と学習方法について、4 節では実験の方法と結果について述べる。5 節では計算量と属性に関して考察し、最後の 6 節でまとめを述べる。

2. クラスタ例からの学習

クラスタリングは、全体のポテンシャルや分類対象の間の類似度のような、事前に定めた規準に基づき分類対象集合を分割する問題である。よって、このような明確な規準があって、それを満たす分割を求める場合に適用すべき手法である。しかし、目標とする分割は明確だが、その分割を導く規準を明示できないときに、目標分割を自動的に導く方法としてクラスタリングが用いられる場合がある。このような適用は、本来の利用法ではないので、目標分割を得るのは一般には困難である。だが、このようなときでも、目標分割そのものを示すことは容易である場合があるので、LCE 問題として解くことができれば有用である。以下、このような場合の一例である画像のセグメンテーションについて述べる。

セグメンテーションとは、画像を何らかの意味があると考えられる部分ごとに分割する処理である。例えば、線分の集合で対象を記述するベクトルデータ形式による論理回路の画像を処理する場合に、図 1(a) の元画像を、(b) のように点線で示した個所で分離することで、“AND

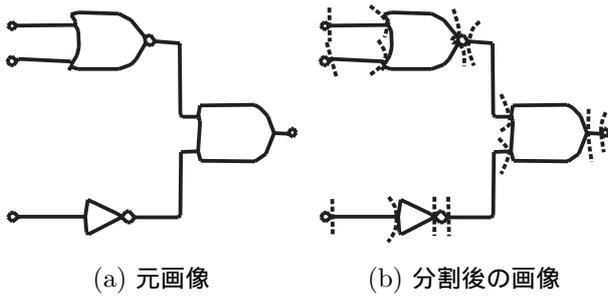


図 1 画像のセグメンテーションの例

ゲート”などの一つの部品を構成している線分の部分集合ごとに分割する処理である。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。この処理は、クラスタリングの手法を利用して行われることが多いが、人間が分割基準を明示することが困難であるという上記のような問題が生じる。図 1 の例では、クラスタリングを適用するためには、線分が同じクラスタに分類されるかどうかを判定するための規準を、ポテンシャル関数や類似度関数の形式で、あらかじめ明示しておく必要がある。しかし、線分の長さや傾きといった特徴量と、分割の目標である図面の部品といった抽象的な概念は大きく隔たっているため、人間が直観的に対応を見いだして、分割の規準を示すことは一般には困難である。ところが、このような場合でも、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 LCE によってセグメンテーションを行う規準が獲得できれば有用である。

2.1 クラスタ例からの学習の定式化

この節では、LCE の定式化について述べる。

図 2 は LCE 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、学習事例集合から分割用の規準を獲得し、推定段階では、獲得した規準を用いて、未知の分類対象集合表現に対する推定分割を求める。

分類対象集合表現は対 $(O_I, F(O_I))$ で、 O_I は分類対象集合、 $F(O_I)$ はこの集合 O_I の特徴である。分類対象集合 O_I は $\#O_I$ 個の分類対象 o^i の集合で、分類される対象であり、この集合の特徴が $F(O_I)$ である。k-means クラスタリング [Jain 88] の場合で例を示すと、分類対象集合 O の特徴 $F(O)$ は、各分類対象の特徴を記述した $\#O$ 個の属性ベクトルの集合に相当する。

事例は $\langle (O_I, F(O_I)), \pi_I^* \rangle$ で、事例集合 EX は $\#EX$ 個の事例からなる集合

$$EX = \{ \langle (O_1, F(O_1)), \pi_1^* \rangle, \dots, \langle (O_{\#EX}, F(O_{\#EX})), \pi_{\#EX}^* \rangle \}$$

である。ただし、 π_I^* は O_I の正しい分割で、分割 π とは分類対象集合の排他的な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。

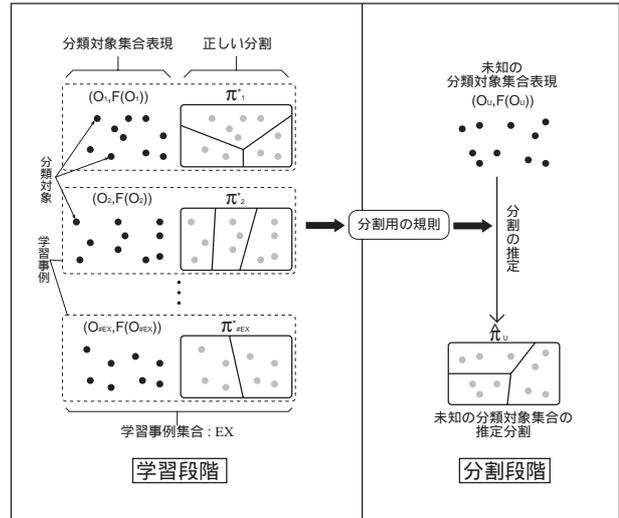


図 2 クラスタ例からの学習の概要

次に、分類対象集合の特徴 $F(O)$ の記述形式について述べる。文献 [Kamishima] では、より多様な記述を可能とするために k-means 法の記述形式を拡張し、以下のような三種類の属性ベクトルを用いた複合属性ベクトル形式で、特徴を記述していた。

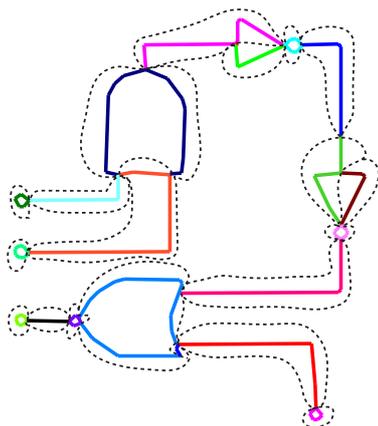
$A(o^i)$: 分類対象属性 k-means と同様に、各分類対象 o^i に関連付けたベクトルで、その分類対象の特徴を表現する。図 1 の例では、線分の長さや中点の位置の座標値といった属性が考えられる。

$A(p^{ij})$: 分類対象対属性 最短距離法によるクラスタリングでは、任意の分類対象対の間の非類似度を要素とする非類似度行列によって、集合の特徴が記述されている。分類対象対属性はこれを一般化したもので、分類対象 o^i と o^j の対である分類対象対 p^{ij} に属性ベクトルを関連付け、単にその対の間の非類似度だけでなく様々な特徴を記述できるようにしたものである。図 1 の例では、線分間の距離や、傾きの差といった属性が考えられる。

$A(\pi)$: 分割属性 これは厳密に言えば属性とは異なるものである。上記二つの属性は、分類対象集合だけの関数で、分類対象集合が与えられた時点で属性値が定まる。しかし、この属性の属性値は、任意の分割を決めたときに定まるもので、分類対象集合と分割の両方を入力とする関数である。この属性は、分割全体の特徴を表し、図 1 の例では、クラスタの数などの属性が考えられる。

これらの属性に加えて、本論文では、新たに以下の属性を導入し、複合属性ベクトル形式を拡張する。

$A(C^J)$: クラスタ属性 この属性は分割属性と同様に、分類対象集合に加えてクラスタをも入力とする関数である。すなわち、分類対象集合を与えられた時点では属性値は定まらず、任意の分割を定め、その中のクラスタを指定すると値が定まる。この属性は、分割



点線で囲んだ部分が一つのクラスタを表す

図 3 LCE の適用結果の例

を構成するクラスタの特徴を表すためのもので、図 1 の例では、部品の形状の特徴量（例：円形度）などが考えられる。

これら四種類の属性を用いた複合属性ベクトル表現では、分類対象集合の特徴 $F(O)$ は $A(\pi), \{A(C)\}, \{A(p)\}, \{A(o)\}$ となる。ただし、 $\{A(o)\}$ は、 O 中の各分類対象に関連付けられた属性ベクトル全体の集合を表す。 $\{A(p)\}$ や $\{A(C)\}$ も同様である。

ここで、クラスタ属性の導入の意図について述べる。図 3 は文献 [Kamishima] の標準的な結果で、手書き図面を取り込み、それをベクトルデータ画像に変換したものを処理している。この画像には、画像ごとに人間が介在してパラメータを調節するなどの処理をしていないため、小さなすきまやヒゲが存在している。この結果を観察すると、小さな端子などはよく抽出されているが、より大きな部品である AND ゲートなどはあまりよく抽出されていない。これは、すでに導入した分類対象属性や分類対象対属性は局所的な情報を表すことはできるが、大域的な情報を表せないためと考えた。そこで、より大域的な情報を扱うクラスタ属性を導入すれば、推定精度の向上が期待される。よって、本論文では、クラスタ属性を扱う手法を提案し、この属性が有効なことが確実な人工問題に適用することで、提案手法が動作することを実証することを目的とする。

本手法では、学習アルゴリズムの制限により属性の値域に次の制限がある。分類対象属性と分類対象対属性では、C4.5[Quinlan 93]などで一般に利用されている数値属性とカテゴリ属性が利用できる。一方、分割属性とクラスタ属性では、 $[0, 1]$ の範囲に変換しておく必要がある。

3. 分割用規則の学習と分割推定の方法

学習段階で分割用の規則を獲得し、分割段階ではその規則を用いて分割を推定するので、この順番で説明する

のが通常だが、説明の容易さのため分割の推定方法を先に述べる。

3.1 分割の推定方法

分割の推定とは、分割段階で、未知の分類対象集合表現 $(O_U, F(O_U))$ が与えられると、特徴 $F(O_U)$ に基づき、学習により獲得された規則を適用して、 O_U の推定分割 $\hat{\pi}_U$ を求めることである。

本論文では、LCE-MAP (LCE-Maximum A Posteriori) アプローチによって分割を推定する。これは、事後確率最大 (MAP) 原理に基いた評価関数を学習段階で獲得し、未知の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アプローチである。

ある分割 π の評価関数は次のとおりである。最初に、分割 π のときの分割属性とクラスタ属性の属性値を計算する。MAP 原理に基づく評価関数は、分割 π が正しい分割と一致するという事象と、分割の表現を構成する属性値ベクトルとの次の結合確率である。

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

ただし、 $\pi=\pi^*$ は π が正しい分割であるという事象である。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。

$$\text{定数項: } \Pr[\{A(o)\}, \{A(p)\}] \quad (2)$$

$$\text{事前確率: } \Pr[\pi=\pi^*|\{A(o)\}, \{A(p)\}] \quad (3)$$

$$\text{分割属性密度: } \Pr[A(\pi)|\pi=\pi^*] \quad (4)$$

$$\text{クラスタ属性密度: } \Pr[\{A(C)\}|\pi=\pi^*] \quad (5)$$

以後、一般には成立しない独立性をいくつか仮定するが、できるだけ独立性のある属性を利用することでこの問題は回避できると考える。また、このような仮定に基づいていても、本手法が有効であることを実験により示す。

定数項(2) は、分割の状態には依存しないので最大化に関しては無視できる。事前確率(3) と分割属性密度(4) を求める手法は、詳しくは文献 [Kamishima] にあるが、その概要のみを次節で述べる。本稿では、最後のクラスタ属性密度(5) の計算方法の詳細を 3.3 節で述べる。

次に、学習段階で上記の評価関数が獲得されたとして、それを最大にする分割の探索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、分割属性密度(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 [Kamishima]) .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、そのときの分割を推定分割とする .

3.2 事前確率(3) と分割属性密度(4) の計算と学習の方法

事前確率(3) と分割属性密度(4) の計算と学習の方法について文献 [Kamishima] の内容を簡単にまとめる .

事前確率(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獲得しておく .

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

ただし、 $\text{in}(p, \pi)$ は、分類対象 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 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 $f_1(p^{ij}, \pi)$ 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、全ての分類対象対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、事前確率(3) の定数倍が次式で計算できる .

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

ただし、 P^+ は、分類対象対の集合の要素のうち、 $\text{in}(p, \pi) = 1$ 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 $f_1(p^{ij}, \pi)$ 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 .

もう一方の分割属性密度(4) の計算と学習の方法について述べる .

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属性の値を計算する .この属性値ベクトルは、分割属性密度(4) の分布から発生したものと見なせる .よって、学習段階では、これら属性値ベクトルから確率密度関数を推定する .分割段階では、この確率密度関数に、属性値を代入すれば分割属性密度(4) の値を計算できる .

3.3 クラスタ属性密度(5) の学習と計算の方法

ここでは、クラスタ属性値ベクトルの集合 $\{A(C)\} = \{A(C^1), A(C^2), \dots, A(C^{\#\pi})\}$ の生じる密度を表すクラスタ属性密度(5) の学習と計算の方法について述べる .

最初に、クラスタ属性密度(5) のモデルについて述べる .属性ベクトル $A(C)$ の各要素は独立に発生するモデ

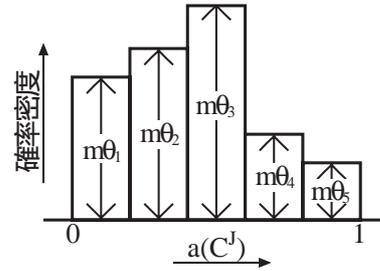


図 4 $\Pr[a(C^J) | \pi = \pi^*; \Theta^s]$ の分布

ルを考えると、クラスタ属性密度(5) は次式で表される .

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

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

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

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

これは、超パラメータ \mathbf{H}^s で指定されるパラメータ分布 $\Pr[\Theta^s; \mathbf{H}^s]$ に従ってパラメータ Θ^s が発生し、このパラメータで指定された属性値分布 $\Pr[a^s(C^J) | \pi = \pi^*; \Theta^s]$ 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 を省略する .

本論文では、属性値分布 $\Pr[a(C^J) | \pi = \pi^*; \Theta]$ に、図 4 の幅 $1/m$ のビンをもつヒストグラム型の分布を採用した .ただし、 m はヒストグラムのビンの数で、計算量を考慮して、実験では 5 とした .図のパラメータ $\Theta = (\theta_1, \dots, \theta_m)$ は $\sum_{k=1}^m \theta_k = 1, \theta_k > 0$ である .また、2 章で述べたように $a(C^J)$ の値域は $[0, 1]$ である .区間 $[0, 1]$ の値の分布では、ベータ分布が代表的だが、0 や 1 の点で密度が無限大になることがあり、数値計算時に特別な扱いを必要とする問題があるため採用しなかった .

このヒストグラム型の分布に従うときの $a(C^1), a(C^2), \dots, a(C^{\#\pi})$ の確率密度は以下のとおりである .属性値の入るヒストグラムのビンの番号を $k(a(C^J))$ 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、 $a(C^1), \dots, a(C^{\#\pi})$ のうち k 番目のビンに入る属性値の

数を n_k で表すと式(6) は次式になる .

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

一方のパラメータ分布 $\Pr[\Theta; \mathbf{H}]$ には, パラメータの総和が 1 になる制約から, このような値の代表的な分布である次の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, $\mathbf{H} = (\eta_1, \dots, \eta_m), \eta_k > 0$ は超パラメータであり, $B(\{\eta\})$ は次の多変量ベータ関数 (式中の $\Gamma(\cdot)$ はガンマ関数)

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

次に, 学習段階で, このモデルの超パラメータ Θ を獲得する方法について述べる . このような超パラメータを獲得する問題は, 文献 [Dempster 77] に EM アルゴリズムによる解法が述べられており, これを適用する . 学習事例集合中の各事例 $((O_I, F(O_I)), \pi_I^*)$ の O_I を π_I^* に分割したときのクラスタ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なので, この属性値ベクトルの集合について最尤推定で超パラメータを求める . 学習事例集合 EX の I 番目の事例から得た属性ベクトルの s 番目の属性値の集合を $\{a_I\} = \{a_I(C^1), \dots, a_I(C^{\#\pi_I^*})\}$ で表す . ここでも, s 番目の属性値のみについて考えるので, s は省略した . また, $\{a_I\}$ を発生させた分布のパラメータを $\Theta_I = (\theta_{I1}, \dots, \theta_{Im})$ で, そして, $\{a_I\}$ のうち k 番目のビンに入る属性値の数を n_{Ik} で表す . そして, 文献 [Dempster 77] の手法に従い, 逐次的に現在の超パラメータ \mathbf{H} を次式を最大にする \mathbf{H}' に更新することで超パラメータを求めた .

$$\begin{aligned} Q(\mathbf{H}'|\mathbf{H}) = & E_{\Theta_1, \dots, \Theta_{\#EX}} \left[\ln \prod_{I=1}^{\#EX} \Pr[\Theta_I|\mathbf{H}']\{\{a_I\}, \mathbf{H}\} \right] = \\ & \sum_{I=1}^{\#EX} \left\{ (\eta'_k - 1) \sum_{k=1}^m E_{\Theta_I} [\ln \theta_{Ik} | \{a_I\}, \mathbf{H}] \right\} \\ & - \#EX \ln B(\{\eta'\}) \end{aligned}$$

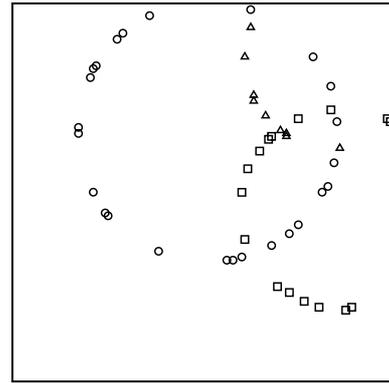


図 5 ドット・パターンの例

ただし,

$$\begin{aligned} \{\{a\}\} &= \{\{a_1\}, \{a_2\}, \dots, \{a_{\#EX}\}\} \\ E_{\Theta_I} [\ln \theta_{Ik} | \{a_I\}, \mathbf{H}] &= \int \frac{F_I(\{a_I\}, \Theta_I|\mathbf{H})}{G_I(\{a_I\}|\mathbf{H})} \ln(\theta_{Ik}) d\Theta_I \\ F_I(\{a_I\}, \Theta_I|\mathbf{H}) &= \prod_{k=1}^m (m\theta_{Ik})^{n_{Ik}} \frac{\prod_{k=1}^m \theta_{Ik}^{\eta_k - 1}}{B(\{\eta\})} \\ G_I(\{a_I\}|\mathbf{H}) &= \int F_I(\{a_I\}, \Theta_I|\mathbf{H}) d\Theta_I \end{aligned}$$

実験では \mathbf{H} の初期値を $\eta_k = 1.0, 1 \leq k \leq m$ とした . また, 関数 $Q(\mathbf{H}'|\mathbf{H})$ を最大化するパラメータ \mathbf{H}' の計算には文献 [Press 93] の simplex tabular 法を用いた .

学習段階で学習事例から超パラメータ \mathbf{H} が獲得されたので, これを用いて, 分割段階で, 未知の分類対象集合表現 $(O_U, F(O_U))$ のクラスタ属性密度(5) を計算する方法について述べる . まず, π_U に分割したときの, クラスタ属性の値 $A(C_U^1), A(C_U^2), \dots, A(C_U^{\#\pi_U})$ を計算する . この値の確率密度は次式で計算される .

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

ただし, n_{Uk}^s は, $a^s(C_U^1), a^s(C_U^2), \dots, a^s(C_U^{\#\pi_U})$ のうち k 番目のビンに入る要素の数である . Θ^s についての期待値をもってクラスタ属性密度(5) の値とするのは, 超パラメータの計算に EM アルゴリズムを用いたが, この手法はこの期待値について最尤な超パラメータを求める手法であるので, それに従い期待値を確率密度とした .

4. 実験と考察

4.1 実験対象

実験では図 5 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 .

ドットパターンは以下の手順で、 X と Y 軸の座標値がそれぞれ $0 \sim 127$ の間の整数となるよう生成した。ドットパターンは 50 個の点で構成されており、2 から 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 $1.3r^I$ の円周上に配置した。 r^I は各クラスタごとに与えたパラメータで、各クラスタの中心から r^I を半径とする円を描いたとき、一つ以上の他のクラスタのこの円とちょうど接触するような条件のもとでランダムに定めている。

このドットパターンの各点が分類対象 o であり、これらの集合の特徴 $F(O)$ のうち、クラスタ属性以外の属性には、文献 [Kamishima] と同じものを採用した。使用する属性として点の位置の X と Y 座標値のみを用いる実験も行ったが、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 5 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 ($4\pi \times \text{面積} / \text{外周長}^2$)
- (3) 推定中心の位置の標準偏差
- (4) 推定中心からのクラスタ中の各点までの距離の標準偏差
- (5) ドットパターン全体の点のうち、そのクラスタに含まれるものの割合。

これらの属性を、値 a が $[0, 1]$ の範囲になり、サンプル平均がほぼ 0.5 となるような変換をして利用した。具体的に、上記の (1) ~ (5) のそのままの値を a_{raw} とすると、 a_{raw} のサンプル平均に近い定数 \bar{a}_{raw} を用いて、属性値には $a = \exp(a_{raw} \times -\ln(2/\bar{a}_{raw}))$ の変換をした値 a を用いた。

これらの属性について簡単に説明をする。1 番目の属性の凸包とは、クラスタ内の点全てを内部に含む最小の凸多角形である。この属性は点が外周上に並んでいれば 1 となる。2 番目は代表的な形状特徴である円形度で、この値は外周が円に近ければ 1 となる。推定中心とは、三点を通る円の中心のことである。この推定中心を、クラスタ中の点について、三点の全ての組み合わせに対して求める。3 番目の属性は、これらの位置の X と Y 方向の分散の和の平方根で、4 番目の属性は、推定中心のセントロイドから各点までの距離の標準偏差である。ただし、推定中心の X か Y の座標値が $[0, 128]$ の範囲にない場合は、これらの点を属性の計算から除外した。

実験対象として、このドットパターンを選択した理由について補足する。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その特徴が明確になる。そのため、クラスタ属性の採用によって、もし推定精度が向上すれば、この属性の有効性を示せると考えた。また、上記の属性は、円形のクラスタとの関連がほぼ自明なものを集めた。これらクラスタの抽出への寄与が明らかな属性があったとしても、これらの属性の

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	0.860	3	0.829
1	0.840	4	0.863
2	0.889	5	0.870

規則性を、学習手法によって見いだせるかどうかは自明ではない。本研究は、属性の有効性ではなく、規則の学習方法の有効性を示すことが目的なので、これら関連が自明な属性を用いた。

4.2 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方法について述べる。ここでは、 $\#EX$ 分割の交叉確認である leave-one-out 法によって、各事例について推定分割と正しい分割の類似度を求め、それらの平均を評価尺度とした。類似度には次の情報損失量 (RIL; Ratio of Information Loss) を用いた。真の分割が生じる事象を Π^* 、推定分割が生じる事象を $\hat{\Pi}$ と記す。事象 Π^* の事前エントロピー $H(\Pi^*)$ と、推定分割を知ったあとの事後エントロピー $H(\Pi^*|\hat{\Pi})$ を用いて RIL は次式で表される。

$$RIL = \frac{H(\Pi^*|\hat{\Pi})}{H(\Pi^*)}$$

この値は、0 に近いほど真の分割に近い推定分割が獲得できたといえ、また、推定分割が真の分割と一致するときだけ 0 になる。具体的に、 $H(\Pi^*)$ と $H(\Pi^*|\hat{\Pi})$ は次式で表される。

$$H(\Pi^*) = \sum_{x=0}^1 \frac{N_x}{\#P} \log \frac{\#P}{N_x}$$

$$H(\Pi^*|\hat{\Pi}) = \sum_{x=0}^1 \sum_{y=0}^1 \frac{M_{xy}}{\#P} \log \frac{M_{0y} + M_{1y}}{M_{xy}}$$

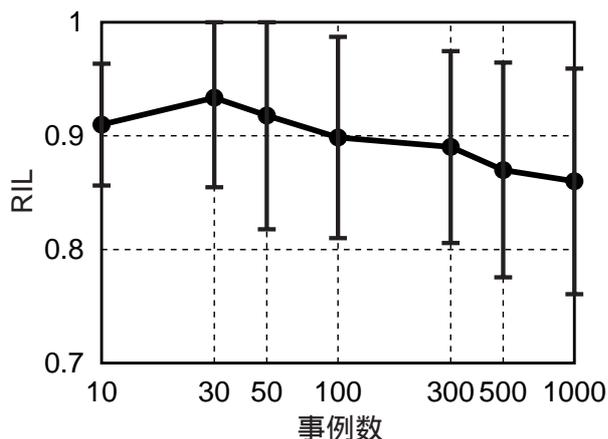
ただし、 N_x は分類対象対のうち $\text{in}(p, \pi^*) = x$ を満たすものの数、 M_{xy} は、 $\text{in}(p, \pi^*) = x$ かつ $\text{in}(p, \hat{\pi}) = y$ を満たすものの数である。

4.3 実験結果と考察

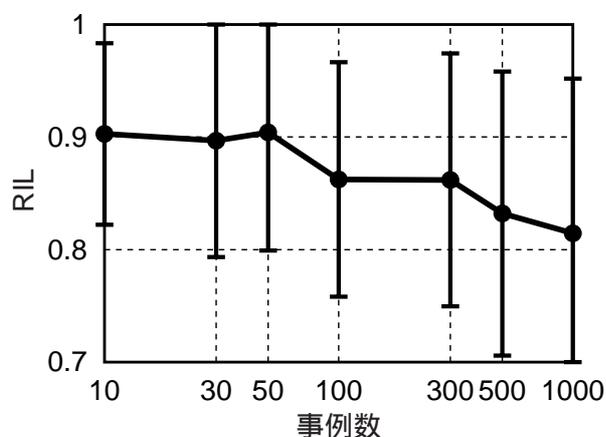
§1 クラスタ属性の選択

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。

分類対象集合の特徴として、前節の 5 種類のクラスタ属性全てを用いるのではなく、いずれか一つだけを用いて分割用の規則を獲得した。総事例数が 1000 個の場合の RIL の平均を表 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。クラスタ属性を全く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以降では採用することとした。



(a) クラスタ属性を用いない場合



(b) クラスタ属性を用いた場合

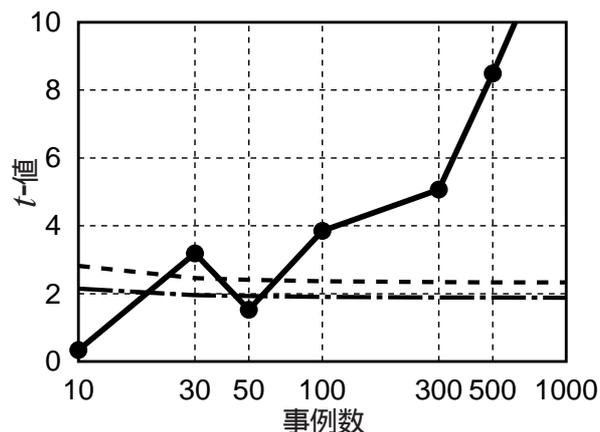
図 6 事例数を変えて実験した RIL の平均と標準偏差の変化

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クラスタリングでは不可能であった。これは、文献 [神鷹 97] の LCE 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。

§2 クラスタ属性の有効性の検証

前節で選択したクラスタ属性を用いて獲得した分割用の規則が、クラスタ属性を用いない場合よりも、より正しい推定分割を導くことを実験により示す。

クラスタ属性を用いないで推定した分割の RIL の平均を図 6(a) に、そのエラーバーで標準偏差を示し、図 6(b) にはクラスタ属性を用いた場合の RIL の平均と標準偏差を示した。総事例数が 10, 30, 50, 100, 300, 500, 1000 の場合について実験した。図には正確な数値は記していないが、1000 事例の値は、クラスタ属性を用いない場合は平均 0.860, 標準偏差 0.0993 で、用いた場合は平均 0.815, 標準偏差 0.1374 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

事例数が 1000 個のときの t -値は 13.46図 7 各事例数での t 値と危険率 1%での有意水準

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の検定を行った。検定には、対応のある場合の t 検定を用い、クラスタ属性を用いた場合とそうでない場合の間の t 値を求めた。この t 値は、自由度 $\#EX - 1$ 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、RIL の平均に有意な差があるといえ、クラスタ属性の有効性が検証できる。それぞれの事例数での t 値を図 7 に示す。実線は t 値を表し、一点鎖線は危険率 3% での有意水準、破線は危険率 1% での有意水準を表す。 t -値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、用い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 10 個と 50 の場合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 -値は大きくなる傾向にあり 100 個以上では明らかに有意な差が見られる。この結果から、本手法によってクラスタ属性を取り扱うことの有効性が示されている。

この結果は自明のようにも見えるが、3 章では、“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”などの幾つかの独立性の仮定を導入しているため、これらの独立性が大きく損なわれる場合には、本論文の手法は有効ではなくなる。しかし、上記の結果は、利用する属性によっては、これらの仮定があっても、本手法が有効であることを示している。

§3 Hough 変換との比較

次に、点集合から関数で表された曲線を抽出する代表的な方法である Hough 変換 [Suetens 92] と、本手法の比較を行う。本実験の目的は、本手法と Hough 変換のどちらがより優れているかを示すことではないことに注意されたい。Hough 変換は、与えられたドットパターンを後に述べる事前に定めた規則に基づき分割する。一方、クラスタ例からの学習はそのような規則そのものを、学習事例集合から獲得する問題である。このように、全く

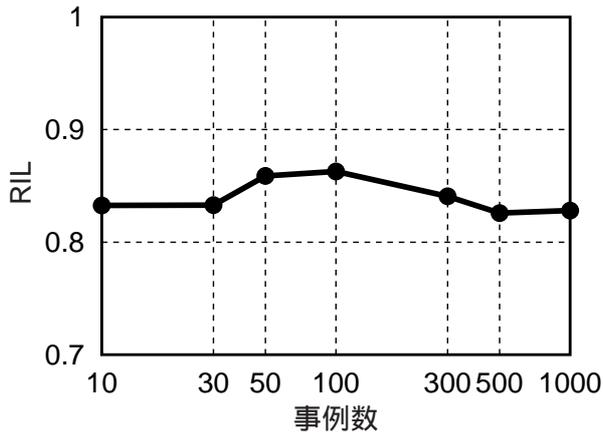


図 8 Hough 変換を用いた場合の RIL の平均

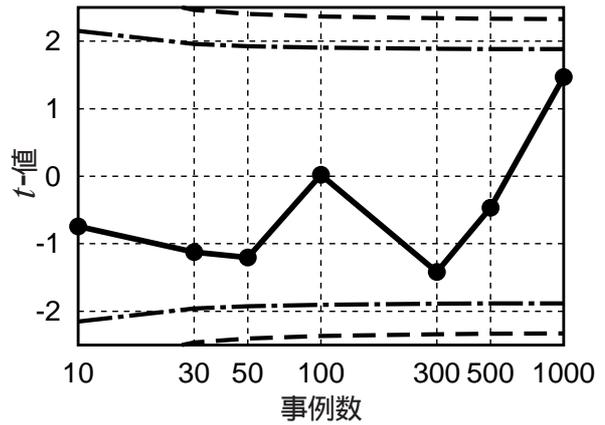


図 9 各事例数での t 値と危険率 1%での有意水準

異なる問題を扱う手法であるので、本質的に優劣を論じることは無意味である。この実験では、Hough 変換を適用するにあたり、ドットパターン中にある曲線は円である、そのパラメータの範囲、そして、円の正しい数などの分割の推定に必要な情報を用いている。ここで、もし Hough 変換と同等の正しさの分割を推定する規則を、学習事例から獲得できれば、分割用の規則には分割推定に必要な情報が十分に含まれていると結論づけられる。このことを検証するのが本実験の目的である。

まず、Hough 変換について簡単に説明する。Hough 変換とは、点の集まりの中から、曲線を表す関数のパラメータを求める手法である。当初は直線を抽出する方法として提案されたが、その後、円や楕円なども抽出できるよう拡張された。図 5 のようなドットパターンから円を抽出する手順は以下のとおりである。まず、中心 (c_x, c_y) と半径 R で構成されるパラメータ空間を考え、各パラメータを離散化して空間をセルに分割する。ここでは、 c_x と c_y は 0 から 127 の 128 個に、 R は 0 から 255 までの 256 個のセルに分割した。このセルには、頻度情報をつぎのように格納する。まず、全てのセルの頻度を 0 とし、ある点 (x_i, y_i) について、セルの表すパラメータ値が、円を表す式

$$(c_x - x_i)^2 + (c_y - y_i)^2 = R^2$$

を満たすならば、そのセルの頻度に 1 を加える。これを、全てのセルと全ての点について行う。そして、頻度が上位 n までのセルを選び出すことで、それらのセルののパラメータで表される円が、ドットパターンの中から抽出できる。実験では、抽出される円の数 n には、正しいクラスタ数 $\# \pi^*$ を与えた。そして、ドットパターン中の各点を、最も近い円に割り当てることで、推定分割を求めた。

図 6 の実験と同じ事例集合について、各分類対象集合表現を上記の方法で分割したものと、事例中の正しい分割との間の RIL を求めた。その平均を図 8 に示す。この実験では、標準偏差は 0.244 ~ 0.298 もあり、エラーバー

分類対象数	30	50	100
クラスタ属性有	0.930	0.862	0.821
クラスタ属性無	0.940	0.899	0.843
(t -値)	1.5065	3.8480	2.7788
Hough 変換	0.822	0.863	0.821
(t -値)	-3.5541	0.0221	-0.0062

表 2 分類対象数に伴う平均 RIL の変化と RIL 間の t -値

をグラフ中に表示できないため、図中には標準偏差を示さなかった。図 6(b) と図 8 を比較すると、平均には明確な差は見られない。Hough 変換は、画像中から図形を抽出する優れた方法として知られているが、ここでは、全体的にあまりよい分割は得られていない。これは、画像に Hough 変換を適用する場合は、少なくとも数万の画素を対象とするが、ここでは 50 個程度しか点がなく、パラメータ空間のセルの頻度に差が小さい。そのため、円を表さない部分の頻度が、円である部分の頻度を上回る場合が生じやすいからである。

これらの RIL の平均の間に差がないことをより厳密に示すために t -検定を行った。図 9 に、図 6(b) と図 8 の RIL の間の対応のある場合の t -値を示す。実線は t -値、破線は 1% の有意水準、そして、一点鎖線は 3% の有意水準である。 t -値が、下の有意水準を下回れば、Hough 変換を用いた方が有意に RIL が小さく、逆に、上の有意水準を上回れば、本手法による推定分割の RIL が有意に小さいことを示している。そして、 t -値がこれら二つの有意水準の間であれば、二組の RIL の平均に差があるとは結論づけられない。

図 9 の実験結果から、いずれの事例数でも、有意水準をたとえ 3% まで緩めても有意な差があるとはいえない。よって、本手法によって獲得した分割用の規則は、Hough 変換を用いた場合と同等の分割を推定することができたといえ、正しい分割の推定に必要な情報を、学習事例集合から獲得できたといえる。

§4 分類対象数の変化の影響

分類対象集合中の分類対象数の変化の影響について調べた。分類対象数が 30, 50, 及び, 100 である三種類の事例 100 個に対する LCE と Hough 変換による RIL の平均を表 2 に示す。クラスタ属性を用いた場合, 用いない場合, 及び Hough 変換の三つの場合の結果を示している。また, 表中に示した t -値は, クラスタ属性がある場合に対するもので, 正であればクラスタ属性を用いた LCE が優れていることを示し, $t_{0.99} = 2.3646$ 以上の差があれば, 危険率 1% で有意な差がある。

全般に, 分類対象数の増加に伴い RIL が減少し, より正確な分割が獲得されている。これは, 分類対象数の増加に伴う情報量の増加の影響と考えられる。クラスタ属性を用いた場合と用いない場合を比較すると, 前述の実験同様にクラスタ属性を用いる方がより適切な推定がなされており, 分類対象数が 50 個以上ではその差は統計的に有意である。Hough 変換と LCE を比較すると, 分類対象数が 30 個の場合は有意に LCE が悪いが, 50 個以上の場合には有意な差はなく同等の分割が獲得されている。30 個のように分類対象数が少ない場合には, 情報の不足により, 精度の高い規則が獲得できていないものと考えられる。

§5 ノイズに対する影響

頑健性について調べるために, ドットパターンの円の半径が, ドットごとに, 本来の半径を中心に正規分布に従って変化する事例集合について実験を行った。正規分布の標準偏差が, 円の本来の半径に対して 0% ~ 10% の場合について実験した結果を表 3 に示す。その他の条件は前節の実験と同じである。

一般的に, 変動幅の増加にともない RIL が増加し, 特に 10% ではどの手法もほとんど適切な分割が得られない。クラスタ属性を用いた場合とそうでない場合を比較すると, 変動の増大にともない, t -値が小さくなり, クラスタ属性が有効でなくなっている。これは, クラスタとしての特徴が曖昧になっていくためであると考えられる。次に, Hough 変換の結果をみると全体としてノイズの増加にともない RIL も増加する傾向がある。 t -値を見ると, LCE の方が若干良い分割を得ている。これは, LCE は, クラスタ属性以外の属性も利用してが, Hough 変換はクラスタの属性のみに依存しているため, クラスタの特徴が明確でない状況では, LCE がより適切にクラスタを推定できていると考える。

5. 考 察

5.1 計算量に関する考察

学習段階と分割段階の計算量について議論する。

学習段階では EM アルゴリズムを用いて超パラメータを計算する。EM アルゴリズムの計算量は関数 $Q(\mathbf{H}|\mathbf{H})$ の計算量を, 反復回数だけ掛けたものになるが, この

回数は定数オーダーなので, 計算量のオーダーは関数 Q についてのみ議論すればよい。関数 Q の計算には, 期待値 $E_{\Theta_I}[\ln \theta_{Ik} | \{a_I\}, \mathbf{H}]$ を $\#EX$ 回評価する必要がある。この期待値の計算量はクラスタ数に比例するため, 学習事例中の平均クラスタ数を $\#\bar{\pi}^*$ で表すと, 全体の計算量は $O(\#EX \#\bar{\pi}^*)$ となる。ところで, 関数 $f_1(p, \pi)$ の計算のために, 平均分類対象数を $\#\bar{O}$ で表すと, $O(\#EX(\#\bar{O})^2)$ の学習事例を参照するため, 必要な計算量はそれ以上になる。 $\#\bar{O} \gg \#\bar{\pi}$ であるので, クラスタ属性の学習に必要な計算量は, 関数 f_1 の学習より小さく, 学習全体の計算量に影響しない。

分割段階では, 収束までに, 少なくとも $\#O_U - \#\hat{\pi}$ 回 ($\#\hat{\pi}$ は最終クラスタ数) 併合操作が必要である。移動操作の回数は不定だが, 経験的には, 併合操作の数よりはるかに少ない。 $\#O_U \gg \#\hat{\pi}$ なので, 反復回数は $O(\#O_U)$ となる。クラスタ数が $\#\pi$ の状態の反復について考慮する。併合操作のために, 全てのクラスタ対を調査する必要があるが, 前回の反復で変化したクラスタと他のクラスタとを併合した場合だけを再計算すればよいので, 計算量は $O(\#\pi)$ となる。移動操作の場合も, 同様に前回から変化したクラスタについてのみ計算すればよい。変化したクラスタには平均 $\#O_U / \#\pi$ 個の要素があり, それらを残り $\#\pi - 1$ 個のクラスタへ移動させる場合を考えればよいので計算量は $O(\#O_U)$ となる。まとめると, 全体の計算量は $O(\#O_U^2)$ となり, これは最短距離法などの階層的アルゴリズムと同等の計算量である。さらに, 分割の途中でクラスタ属性を再計算する必要があるため, その計算量だけオーダーが加わる。

計算量のオーダーは多項式ではあるが, その次数が大きい。そのため, 分類対象数 $\#O$ を大きくすることは難しい。そこで, 文献 [Kamishima] の画像のセグメンテーションへの LCE の適用では, 幾つかの画素を併合して線分で表現したベクトル画像に変換する処理をしている。この併合は, ユーザの意図を反映したものではなく, 単に類似した画素を併合する処理である点が重要な点である。このような併合操作の意図の違いは Pavlidis によっても

It seems that no matter how sophisticated a surface fitting or step finding algorithm we use, we cannot improve the results of object outlining for a very large class of images [Pavlidis 92].

のように指摘されている。このように, ユーザの意図に合わせないような併合操作で分類対象の数を減し, その後, 具体例から学習した規則によって, ユーザの意図を反映するようにそれらの分類対象を併合をするようにすれば, 実用的な時間で計算が可能である。

ノイズの標準偏差	0%	1%	3%	5%	10%
クラスタ属性有	0.862	0.827	0.897	0.914	0.950
クラスタ属性無	0.899	0.840	0.899	0.916	0.948
(<i>t</i> -値)	3.8480	1.3168	0.3172	0.5448	-0.5206
Hough 変換	0.863	0.829	0.945	0.928	0.975
(<i>t</i> -値)	0.0221	0.0517	3.0738	0.8372	3.3473

表 3 ノイズの大きさに伴う平均 RIL の変化と RIL 間の *t*-値

5.2 属性に関する考察

クラス分類の学習問題において属性が分類器の性能にとって重要であるのと同様に、推定分割の精度向上には、学習手法の改良と共に、分類対象集合表現の属性も重要である。図 3 のような対象を分割するには、より高度な属性が必要になり、その開発にも困難が予測される。しかし、LCE の枠組は、次のように、この属性の開発に寄与すると考える。一つは、4.3.1 節のように、クラス分類で開発された属性選択手法を用いて、既存の属性から有用なものを定量的な基準に従って選択できること。さらに、複合属性ベクトル表現に、多様な属性を導入すれば、それらを組み合わせることで規則を生成することもできる。もちろん、任意の分類対象集合の推定分割を求めるのは不可能だが、与えられた事例と同じ発生源の分類対象集合であれば、その推定分割を求める規則を記述する属性ならば、開発が可能であると考えている。

6. ま と め

本論文では、未知の分類対象集合表現の分割用の規則を、学習事例集合から求める LCE の手法の改良について述べた。

従来の手法では、分類対象集合の特徴を三種類の属性ベクトルを用いた複合属性ベクトル形式で記述していた。本論文では、分類対象集合のより多様な特徴を考慮して適切な分割を推定できるようにするために、新たにクラスタ属性を導入してこの表現形式拡張した。

本論文では、このクラスタ属性を利用する手法を提案し、これを人工的なドットパターンデータに適用した。実験では、クラスタ属性を用いた場合と用いない場合の実験結果を比較し、クラスタ属性の利用によって統計的に有意に、より正しい推定ができることを示した。また、Hough 変換を用いて求めた分割と、本手法で獲得した規則を適用した分割を比較する実験も行った。これらの分割に差が無いことから、正しい分割を求めるために必要な情報を本手法で獲得できていることも示した。

今回は、人工データを用いて手法の有効性の検証を行ったが、今後は、文献 [Kamishima] 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 ◇

[Dempster 77] Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum Likelihood from Incomplete Data via The EM Algorithm, *Journal of the Royal Statistical Society (B)*, Vol. 39, No. 1, pp. 1-38 (1977)

[Jain 88] Jain, A. K. and Dubes, R. C.: *Algorithms for Clustering Data*, Prentice Hall (1988)

[Kamishima] Kamishima, T. and Motoyoshi, F.: Learning From Cluster Examples, *Machine Learning* (to appear)

[神鷹 97] 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)

[Pavlidis 92] Pavlidis, T.: Why Progress in Machine Vision is So Slow, *Pattern Recognition Letters*, Vol. 13, pp. 221-225 (1992)

[Press 93] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge Univ Press, second edition (1993)

[Quinlan 93] Quinlan, J. R.: *C4.5 Programs for Machine Learning*, Morgan Kaufmann (1993)

[Suetens 92] Suetens, P., Fua, P., and Hanson, A. J.: Computational Strategies for Object Recognition, *ACM Computing Surveys*, Vol. 24, No. 1, pp. 5-61 (1992)

[担当委員: 沼尾正行]

2002 年 5 月 2 日 受理

著 者 紹 介



神鷹 敏弘 (正会員)

1968 年生。1992 年京都大学情報工学科卒業。1994 年同大学院修士課程終了。同年電子技術総合研究所入所。2001 年博士 (情報学)。同年電子技術総合研究所は産業技術総合研究所へ再編。機械学習とその応用の研究に従事。ACM, 情報処理学会会員。



赤穂 昭太郎

1988 年東大・工・計数卒。1990 年同大学院工学系研究科計数工学専攻。修士課程修了。同年通商産業省電子技術総合研究所入所。2001 年より産業技術総合研究所 脳神経情報研究部門情報数理研究グループ主任研究員。博士 (工学)。統計的学習の理論や応用に関する研究に従事。電子情報通信学会, 日本神経回路学会会員。



元吉 文男

1976 年東京大学大学院理学系研究科物理学専門課程修士課程終了。同年, 通産省電子技術総合研究所入所。現在, 独立行政法人産業技術総合研究所所属。研究テーマは数式処理, プログラミング言語。情報処理学会, 日本ソフトウェア科学会, 日本応用数理学会, 日本数式処理学会各会員。