

クラスタ例からの学習 — クラスタ属性の利用法の改良

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

神島 敏弘*

Toshihiro Kamishima

元吉 文男*

Fumio Motoyoshi

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. Unfortunately, this method did not successfully work, so we show the improved method and applied it to dot-pattern partitioning.

1 はじめに

クラスタ例からの学習 [3] とは、例からの学習にクラスタリングの要素を加えたものである。これは、未知の分類対象集合を与えられたとき、その集合を適切に分割するための規則を獲得する学習問題である。

例からの学習では、一般的に、属性ベクトルを用いて分類対象を表現する。クラスタ例からの学習でも同様に、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個の属性ベクトルで分類対象を表現するのではなく、数種類の多数のベクトルによって分類対象集合を表現する点が異なる。

文献 [2] では、次の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 [1] では、各クラスタ(分類対象を分割したとき、分割された部分集合)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 [1] では、この種の属性を有効に利用できなかったため、本論文では、その利用方法を改良する。この改良手法を人工的なデータに適用し、その効果を調査した。

以下、2 節ではクラスタ例からの学習について、3 節では新たな学習方法について、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、左側は学習段階、右側は学習結果を利用した分割の推定段階を表す。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、事前に定めた基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、この基準を明確に提示できない場合には、利用者にとって適切な分割を得ることができない。そこで、基準そのものを示す代わりに、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基準を推定する問題がクラスタ例からの学習である。このような、基準を示すことは困難

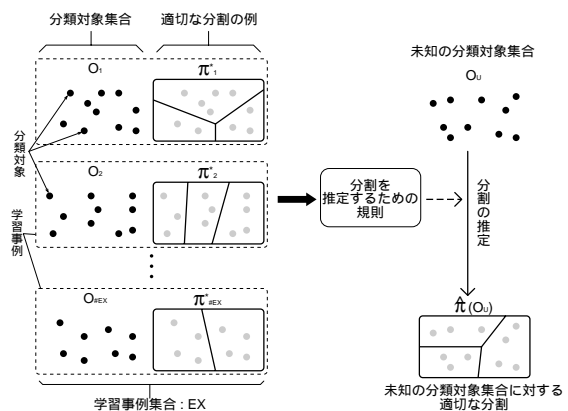
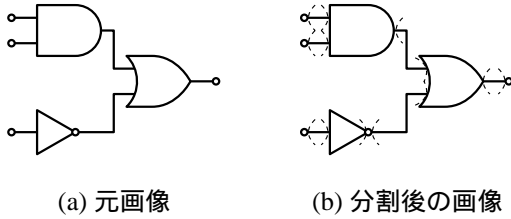


図 1: クラスタ例からの学習の概要

*電子技術総合研究所, 知能情報部, 〒 305-8568 茨城県つくば市梅園 1-1-4 tel. 0298-61-5835, e-mail kamisima@etl.go.jp, <http://www.etl.go.jp/~kamisima/>
Electrotechnical Lab., Machine Understanding Div., Umezono 1-1-4, Tsukuba, Ibaraki, 305-8568 JAPAN



(a) 元画像

(b) 分割後の画像

図 2: 画像のセグメンテーションの例だが、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タ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た。

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。分割に依存せず、分類対象集合のみの関数。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。 $A(o^i)$ と同様、分類対象集合のみの関数。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。分類対象集合 O と同時に分割 π の関数。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。 $A(\pi)$ と同様に、 π の関数。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。

まとめると、分類対象集合が与えられた時点で $\{A(O)\}$ と $\{A(P)\}$ (それぞれ、 $A(o)$ と $A(p)$ 全体の集合) の属性値が定まり、さらに、この分類対象集合を π に分割すると $A(\pi)$ と $\{A(C)\}$ ($A(C)$ 全体の集合) が計算できるようになる。適切な分割を導く規則は、これら四種類の属性を利用して記述される。

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、これを適切な推定分割とする。この評価関数は、MAP 原理に基づくもので、未知の分類対象集合について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、この集合に関連付けられた属性値ベクトルとの結合確率密度である。この密度は、分類対象と分類対象対の集合 $\{A(O)\}$ と $\{A(P)\}$ が与えられた状態での、分類対象集合を π に分割したときの $\pi = \pi^*$ (この π が真に適切な分割であるという事象) と $A(\pi)$, $\{A(C)\}$ との結合確率密度で、次式で表される。

$$\Pr[\pi = \pi^*, A(\pi), \{A(C)\}; \{A(O)\}, \{A(P)\}] \quad (1)$$

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、この結合確率を次の三つの確率分布・密度に分解し、それぞれを簡略化する。

$$\Pr[\pi = \pi^*; \{A(O)\}, \{A(P)\}] \quad (2)$$

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

$$\Pr[A(C) | \pi = \pi^*, A(\pi); \{A(O)\}, \{A(P)\}] \quad (4)$$

式 (2) を求める手法は、詳しくは文献 [2] にあるが、その概要を 3.2 節に述べる。

属性 $A(\pi)$ は、分類対象集合と分割との関数であるため、その属性値が計算されるときに前提となる $\{A(O)\}$ と $\{A(P)\}$ の影響を暗黙のうちに受けていると見なせる。よって、前提条件 $\{A(O)\}$ 、 $\{A(P)\}$ を明示せず、式 (3) を次式で代用する。

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

この分布を求める手法も、詳しくは文献 [2] にあるが、その概要を 3.3 節に述べる。

式 (4) も、式 (3) と同様に、前提条件を省略し次式で代用する。

$$\Pr[A(C)|\pi=\pi^*, A(\pi)] \quad (6)$$

今回の研究では、この分布の計算方法を改良した。詳細は 3.4 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. 二個のクラスを一個のクラスに併合
2. あるクラスの要素一つを別のクラスに移動

ただし、式 (5) 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 分布 $\Pr[\pi=\pi^*; \{A(O)\}, \{A(P)\}]$ の計算と学習の方法

ここでは、式 (2) 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。

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。

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

ただし、 $\text{isc}(p, \pi)$ は、分類対象 p が分割 π の同じクラスの要素であるとき 1、そうでないとき 0 をとる関数である。ここで、 $\text{isc}(p^{ij}, \pi^*) = 1$ 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

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、この事象基本確率とみなす。分割の推定段階で、分類対象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 (2) 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、分類対象集合の要素のうち、 $\text{isc}(p, \pi) = 1$ 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。

3.3 分布 $\Pr[A(\pi)|\pi=\pi^*]$ の学習法

ここでは、式 (5) 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。

式 (5) の確率密度に関して、学習段階で、学習事例集合 EX の各要素についてその分割全体の属性の値を計算し、その属性値ベクトルの集合から確率密度関数を推定しておく。分割の推定段階では、この確率密度関数に、単に属性値を代入すれば式 (5) の値を求められる。

この確率密度関数を回帰木によって記述した。回帰木とは、よく知られた決定木とほぼ同じ物だが、木の葉ノードには、入力属性値ベクトルを分類すべきクラスの代わりに、入力ベクトルの確率密度が割り当てられている。決定木と同様に入力ベクトルの値に応じて左右の枝をたどり、たどり着いた葉ノードの確率密度を出力とすることで、確率密度関数を表現する。

3.4 分布 $\Pr[A(C)|\pi=\pi^*, A(\pi)]$ の学習法

文献 [1] では、以下の手順によりクラス属性の確率密度を求めた。最初に、式 (6) は、 $\{A(C)\}$ と $A(\pi)$ の相関は低いものと考え独立と仮定し、 $\Pr[A(C)|\pi=\pi^*]$ と置き換えた。次に、クラス属性の各要素は独立に発生すると仮定する、すなわち、属性ベクトル $A(C^J)$ の要素、 $a^1(C^J)$ 、 $a^2(C^J)$ 、 \dots 、 $a^{\#A(C)}(C^J)$ は独立とする。ただし、 $\#A(C)$ は属性ベクトルの要素数。この仮定を加えると、式 (6) は次式で表される。

$$\prod_{s=1}^{\#A(C)} \Pr[a^s(C^1), \dots, a^s(C^{\#\pi})|\pi=\pi^*] \quad (7)$$

この式中の結合確率 $\Pr[a^s(C^1), \dots, a^s(C^{\#\pi})|\pi=\pi^*]$ を表すために、パラメータ $\exp(\alpha^s)$ と $\exp(\beta^s)$ で指定されるベータ分布に従って $a^s(C^J)$ は独立に発生し、超パラメータ ($\Theta^s = (\alpha^s, \beta^s)$) の各要素は正規分布に従って

発生するというモデルを採用した．すなわち，式 (7) は次式で表される．

$$\prod_{s=1}^{\#A(C)} \left[\prod_{J=1}^{\#\pi} \Pr[a^s(C^J) | \pi = \pi^*] \right] \Pr[\Theta^s] \quad (8)$$

パラメータ分布の超パラメータを学習段階で推定しておく，与えられたクラスタ属性の属性値ベクトルに対する式 (8) の最大値をクラスタ属性の確率密度，すなわち，式 (6) の値とした．

この超パラメータは，学習事例 EX より以下の方法で求めた．学習事例の各要素について， O_I を π_I^* に分割したときの，各クラスタ属性値ベクトルの第 s 要素を抽出した集合， $\{a^s(C_I)\} = \{a^s(C_I^1), \dots, a^s(C_I^{\#\pi_I^*})\}$ (以下，添字 s は省略)，を求める．ベータ分布の平均と分散が次式であることを用い，

$$\begin{aligned} \text{平均} &= \frac{\alpha_I}{\alpha_I + \beta_I} \\ \text{分散} &= \frac{\alpha_I \beta_I}{(\alpha_I + \beta_I)^2 (\alpha_I + \beta_I + 1)} \end{aligned}$$

$\{a^s(C_I)\}$ の平均と分散がこの式に一致するとしてパラメータ Θ_I を求めた．パラメータの分布に正規分布を用いているので， $\alpha_1, \dots, \alpha_{\#EX}$ の平均と標準偏差を超パラメータとした．

しかし，上記の手法で新たな属性を採用したにもかかわらず，この手法では文献 [2] と比べてより適切な分割は推定されなかった．その原因は以下の点にあると考えた．このモデルでは，クラスタ属性の属性値の確率密度はクラスタ数に依存しないものとしている．これが原因となり，式 (8) はクラスタ数が多いほど大きな値になる傾向が生じていた．具体的には，確率密度 $\Pr[a^s(C^J) | \pi = \pi^*]$ が 1 より大きい場合には，クラスタ数が多いほど式 (8) も大きくなっていった．適切な分割中のクラスタ数は，分割全体の属性 $A(\pi)$ によって表されることを想定していたが，この属性の影響よりも上記の影響の方が強くなり，訓練事例から想定される数よりも多いクラスタの分割を導く評価関数を学習していた．

今回はこの問題に対処するため，クラスタ数を表す属性を $A(\pi)$ の要素として必ず採用することとし，加えて， $\{A(C)\}$ と $A(\pi)$ の間の依存性を無視しないようにした．具体的には，超パラメータ Θ は常に一定であったが， $A(\pi)$ の属性値に依存して変化するようにした．これは，クラスタ数の確率密度分布の定義域を適当に分割し，各領域ごとに個別に $\Pr[\Theta^s]$ を推定することで実現した．定義域の分割には，分割全体の属性の確率密度を記述するのに利用した回帰木の特徴を流用した．回帰木は，式 (5) の値がほぼ等しい事例が同じ葉ノードに分類

されるように作成されるが，これと同時に，分割全体の属性値が互いに類似した事例が同じノードに集まる．このことを利用し，学習段階では， EX 中の各事例を式 (5) を記述した回帰木で分類し，各葉ノードごとにそのノードに分類された事例だけを用いて推定することで，分割全体の属性に依存した超パラメータを求めた．そして，推定段階では，分割全体の属性値ベクトルを式 (5) を表す回帰木に入力し，その到達した葉ノードにある超パラメータを用いてクラスタ属性の確率密度分布を計算した．

さらに，より精密に計算するために，文献 [1] の方法で求めた超パラメータを初期値とし，EM アルゴリズムにより精密な予測を行った．その方法について簡単に，パラメータ α と β ，それぞれに，超パラメータの平均 μ_α と標準偏差 σ_α があるが，ほぼ同じ手順で求められるため α についてのみ述べる．EM アルゴリズムでは，現在の超パラメータ $\mu_\alpha, \sigma_\alpha$ は次式で $\mu'_\alpha, \sigma'_\alpha$ に更新する．

$$\begin{aligned} \mu'_\alpha &= \frac{1}{\#EX} \sum_{I=1}^{\#EX} \int_{-\infty}^{\infty} \Phi(I) \alpha_I d\alpha_I d\beta_I \\ \sigma'^2_\alpha &= \frac{1}{\#EX} \sum_{I=1}^{\#EX} \int_{-\infty}^{\infty} \Phi(I) \alpha_I^2 - \mu'^2_\alpha d\alpha_I d\beta_I \\ \Phi(I) &= \frac{F(I)}{\int_{-\infty}^{\infty} F(I) d\alpha_I d\beta_I} \\ F(I) &= \left\{ \prod_{J=1}^{\#\pi_I^*} B(a(C_I^J); \exp(\alpha_I), \exp(\beta_I)) \right\} \times \\ &\quad N(\alpha_I; \mu(\alpha), \sigma(\alpha)) N(\beta_I; \mu(\beta), \sigma(\beta)) \end{aligned}$$

ただし， $B(x; a, b)$ はベータ分布での x の確率密度， $N(x; \mu, \sigma)$ は正規分布での確率密度を表し，同じ葉ノードに分類された事例の集合を EX とする．

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，どれだけ真に適切な分割に近い分割が推定されたかを評価する文献 [3] 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，次のような leave-one-out 法によって推定された分割を評価する．与えられた事例集合 EX の最初の事例 (O_1, π_1^*) を取り出し，これをテスト用事例，残りを学習用事例とする．学習用事例から獲得された規則を利用して，分類対象集合 O_1 に対する推定分割 $\hat{\pi}_1$ を求める．この推定分割と真に適切な分割 π_1^* との間の情報損失量を求める．ただし，情報損失量 (以下，RIL(Ratio of Information Loss) と略す) とは，真に適切な分割を推

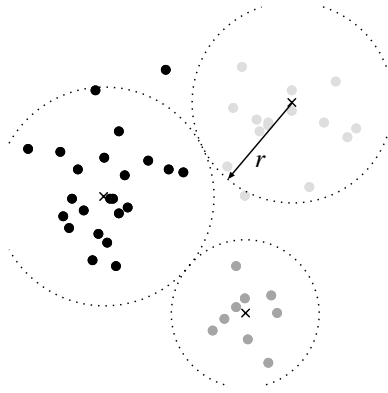


図 3: ドット・パターンの例
 定するのに必要な情報量のうち、推定分割はどれだけの情報量を得ていないかという割合であり、0 から 1 の間の値をとり、0 のとき推定分割は真に適切な分割と一致する。これら一連の操作を残りの 2 から $\#EX$ 番目までの事例に対して行い、 $\#EX$ 個の RIL の平均を求める。この平均値によってどれだけ適切な分割が推定できているかを評価した。

4.2 実験対象

文献 [3] のドットパターンを対象に実験を行ったので、これについて述べる。ドットパターンとは、図 3 のように平面にドットが分布している分類対象集合で、クラスタリング手法の実験的検証によく利用される。各分類対象集合は 50 個のドットから構成されており、3~5 個のクラスタに分割されている。各クラスタ内ではドットは円形のガウス分布に従って分布しており、その標準偏差は $r/2.5$ である。ただし、 r は各クラスタごとに与えたパラメータで、図 3 の点線のように、各クラスタを中心にこのパラメータ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たものである。分割の推定には、文献 [2] の 4 種類の分類対象の属性と 8 種類の分類対象対の属性の属性を利用した。分割全体の属性には、文献 [2] のもののうち、クラスタ数を分類対象数で割った値のみを用いた。クラスタ属性は、次の 4 通りの場合でそれぞれ実験を行った。

- 属性 1. クラスタ内のドットの寄与率
- 属性 2. クラスタ内の X と Y 軸方向の歪度
- 属性 3. クラスタ内の X と Y 軸方向の尖度
- 属性 4. クラスタ内のドット数

ただし、これらの値は適宜 (0, 1) の範囲となるよう正規化した。これらの属性の選択にあたっては、適切な分割の推定に有効であると著者が考えた属性を幾つか実験し、正しい分割での属性値と、ランダムに分割した場合

表 1: ドットパターンの事例集合に対する RIL の平均

| クラスタ属性 | 実験 1 | 実験 2 | 実験 3 |
|------------|------|------|------|
| 属性 1 (寄与率) | .166 | .173 | .166 |
| 属性 2 (歪度) | .181 | .182 | .179 |
| 属性 3 (尖度) | .191 | .196 | .185 |
| 属性 4 (要素数) | .161 | .161 | .161 |

実験 1: EM あり, 依存性考慮
 実験 2: EM あり, 依存性無視
 実験 3: EM なし, 依存性考慮

の属性値の平均および標準偏差の差が大きいものを選択した。これは、正しい分割での属性値の分布が、ランダムに分割した場合の属性値の分布と異なっている属性ほどより有効であると考えたからである。

4.3 実験結果と考察

3 節の手法をドットパターンに適用したときの RIL の平均を表 1 に示す。今回は、3.1 節で述べたように、文献 [1] の手法に対し

1. クラスタ属性の分布関数で、この属性と分割全体の属性の依存性を考慮
2. EM アルゴリズムにより厳密にパラメータを推定

の 2 点の改良を試みた。表 1 には、両方の改良を採用した場合 (実験 1)、依存性を考慮しなかった場合 (実験 2)、および、EM アルゴリズムを採用しなかった場合 (実験 3) の三通りの実験について、それぞれ、前節の四種類のクラスタ属性の集合を用いた結果を示した。前節でも述べたように、RIL は 0 に近い方がより適切な分割が予測されていることを示しているが、クラスタ属性を用いない場合の値 0.161 より、クラスタ属性を採用した方が若干 RIL が大きくなり、文献 [1] 同様、クラスタ属性の効果は見られていない。また、依存性を考慮しなかった場合や、EM アルゴリズムを採用しなかった場合でも結果は同様であった。

この原因を調査するために、式 (1), (2), (5), (6) の各項の値を、真に適切な分割と推定した分割それぞれについて求めた。各式の値の $-\log_2$ をとった値の平均を表 2 に示す。表 1 の三種類の実験にクラスタ属性を用いない場合を加えた。各欄の左側が真に適切な分割に対する値で、右が推定された分割に対する値であり、値の大小を間の不等号・等号で表示した。この値が小さければより適切な分割と判断されるので、真に適切な分割でこの値が最小となるのが望ましい。しかし、表 1 の式 (1) の値は、真の分割の方が大きい。よって、推定分割の探索の手法に問題があるのではなく、式 (1) の推定の精度

表 2: 評価関数の値の平均値 ($-\log_2$ をとった値)

| | | 式 (1) | 式 (2) | 式 (5) | 式 (6) |
|------|------|---------------|---------------|---------------|-----------|
| なし | | 254.2 > 152.8 | 266.3 > 164.8 | -12.1 = -12.1 | 0.0 = 0.0 |
| 実験 1 | 属性 1 | 258.3 > 164.1 | 266.3 > 171.3 | -12.1 < -12.0 | 4.0 < 4.9 |
| | 属性 2 | 257.7 > 170.4 | 266.3 > 178.3 | -12.1 < -12.0 | 3.5 < 4.1 |
| | 属性 3 | 256.7 > 179.5 | 266.3 > 188.4 | -12.1 < -12.0 | 2.5 < 3.1 |
| | 属性 4 | 255.8 > 154.3 | 266.3 > 164.3 | -12.1 = -12.1 | 1.6 = 1.6 |
| 実験 2 | 属性 1 | 258.3 > 164.2 | 266.3 > 171.7 | -12.1 < -12.0 | 4.1 < 4.5 |
| | 属性 2 | 257.6 > 175.2 | 266.3 > 182.4 | -12.1 < -12.0 | 3.4 < 4.8 |
| | 属性 3 | 256.4 > 191.8 | 266.3 > 200.5 | -12.1 < -12.0 | 2.2 < 3.4 |
| | 属性 4 | 255.4 > 153.8 | 266.3 > 164.8 | -12.1 = -12.1 | 1.1 > 1.0 |
| 実験 3 | 属性 1 | 258.3 > 164.1 | 266.3 > 171.3 | -12.1 < -12.0 | 4.0 < 4.9 |
| | 属性 2 | 257.7 > 170.1 | 266.3 > 178.2 | -12.1 < -12.0 | 3.5 < 3.9 |
| | 属性 3 | 256.7 > 176.9 | 266.3 > 186.0 | -12.1 < -12.0 | 2.5 < 2.9 |
| | 属性 4 | 255.8 > 154.3 | 266.3 > 164.8 | -12.1 = -12.1 | 1.6 = 1.6 |

に問題があることが分かる．さらに，式 (1) を構成する式 (2), (5), (6) の各項を調査する．式 (2) は真の分割でも大きな値になり，この項が大きく影響していることが分かる．これは，式 (2) は分類対象対という非常に局所的な部分の情報を基に計算されているため，予測の誤差が大きいためである．そこで，式 (2) の項を分割の詳細を決定するための項として利用し，加えて分割全体の大域的な特徴を表すために式 (5) や (6) の項を導入していた．これら新たな項の真に適切な分割での値とそうでない分割の差が大きければ，これらの項は有効に機能するが，その差が小さいことが問題であると考えられる．クラスタ属性としては，真の分割とそうでない場合に分布に大きな差があるものを選んで採用しており，分布関数を十分に表現できれば適切な分割とそうでない分割との間で確率密度に大きな差がでるはずであるが，全体的にあまり差はでていない．この分布関数が表現できなかった原因は，分布の属にベータ分布を選択したことはないかと推察している．これは，ベータ分布に従う人工データに対しては EM アルゴリズムによる精密な推定が効果があったが，ここでの実験では効果が無かったことから確かめられる．少数のパラメータで多様な分布を表現できるためベータ分布を採用していたが，他の分布，特に，パルス状の分布を的確に表現できる分布を採用して実験を行いたい．

あと，まだ十分な証拠を見つけてはいないが，分割全体とクラスタ属性と依存性を考慮しても結果が改善されなかった点については，分類対象集合の分割により，サンプル数が減少しパラメータの推定精度が下がった影

響のために，効果が打ち消されたのではないかと考えている．

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，この手法が実験的なドメインの問題 (ドットパターン) で有効であることの検証を試みた．

しかし，今回の方法では，より適切な分割を推定できなかった．実験の結果，クラスタ属性の分布の族にベータ分布を採用した点に問題があると推定されたので，他の分布の族を採用した場合について実験する予定である．

参考文献

- [1] 神鷹敏弘, 元吉文男. クラスタ例からの学習 — クラスタに関する属性の利用 —. 人工知能学会研究会資料, Vol. SIG-FAI-9804-4, pp. 23–28, 1999.
- [2] 神鷹敏弘, 元吉文男. クラスタ例からの学習 — 分類対象集合全体の属性の利用. 情報論, Vol. 40, No. 9, pp. 3345–3357, 1999.
- [3] 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.