人工知能学会誌 Vol. 12 No. 2 (Mar. 1997) 〔別刷〕

論 文

Technical Papers

クラスタ例からの学習

Learning from Cluster Examples

神嶌 敏弘* 新田 克己*
Toshihiro Kamishima Katsumi Nitta

* 電子技術総合研究所知能情報部推論研究室
Machine Inference Sec., Machine Understanding Div., Electrotechnical Laboratory, Tsukuba 305, Japan.

1996年9月11日 受理

Keywords: machine learning, clustering, image understanding.

Summary

In this paper, novel machine learning problem that handles clustering and a solution for this problem are described.

The clustering is a method to divide a given set of individuals into clusters that are subsets having properties "internal cohesion" and "external isolation". The clustering is often used to get partition for a set of individuals desired to be fitted to some purpose. In such case, it is hard to define such desired partition by means of the properties internal cohesion and external isolation explicitly. But, it is usually easy to show desired partition itself. Therefore, by learning from examples that are pairs of a set of individuals and desired partition for the set, it is desirable to acquire a criterion that is used to get desired partition for any unknown set. In this paper, such a learning method is proposed. The method is different from ordinal "learning from examples". So we call it "Learning from Cluster Examples".

In our prior work, the learning method is applied to a problem to divide a logic diagram image. The experimental results show that the proposed method learns more desired partitions than the ones in our prior work. This method is extended to be so general that is applied to logic diagram understanding but also dot pattern clustering or problems in other fields.

1. はじめに

本研究では，クラスタリングを対象にした新たな学習問題を提起し，この問題に対する学習法を提案する．学習により獲得した知識に基づいたクラスタリングの結果を定量的に評価し，その結果について考察する．

クラスタリングとは，内的なまとまりと外的な分離が達成されるようなクラスタと呼ぶ部分集合に，分類対象集合を分割する操作である．この操作は分類対象集合に対するある望ましい分割を獲得する目的でよく利用される．しかし，その望ましい分割を，内的なまとまりや外的な分離といった規準によって表現することが困難であることが多い．そこで，分類対象集合とその集合に対する望ましい分割の組を学習事例として，未知の分類対象集合に対する望ましい分割を獲得するための規準を学習する問題を提起し，この問題を「クラスタ例からの学習」と名づける．

この学習は，[神嶌 95]でベクトルデータを対象にしていたものを，より多様な対象に適用できるように，一般的な問題として定式化しなおしたものである．より多様な対象に適用した結果をもとに[神嶌 95]の結果を改善し，また，実験結果の分析をもとに現在の学習方法の問題点などについて考察する．

以後，**2章**では，このクラスタ例からの学習の定義について，**3章**では，学習方法とその結果を利用したクラスタリングの方法について，**4章**では，獲得された分割の定量的な評価法について，**5章**では，実験とその結果に対する考察について，**6章**では，まとめと今後の予定について述べる．

2. クラスタ例からの学習

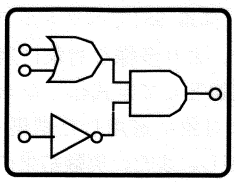
2・1 クラスタリングとその応用上の問題点

ここでは，クラスタリングについて述べ，次にクラスタリングを応用するうえでの問題点を示す．

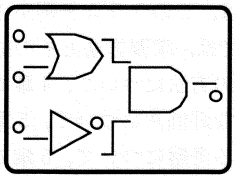
「クラスタリング」にはさまざまな定義があるが，ここでは，分類対象集合を，内的なまとまり(internal cohesion)と外的な分離(external isolation)が達成されるようなクラスタと呼ぶ部分集合に分割することとする[Everitt 93, p. 6]．このとき，どのような種類のクラスタがいくつ存在するかは事前にはわからないものとする．このクラスタリングは，機械学習の分野では観察による学習や概念形成と呼ばれ，FisherのCOBWEB[Fisher 87]などの研究が，数値分類ではクラスタ分析と呼ばれ多くの研究がある．これらの研究では，分類対象集合全体，もしくは，分類対象対について内的なまとまりや外的な分離を測る尺度を定義し，その尺度に基づき分割を獲得する．

このクラスタリングは，次のような場合に，分類対象集合に対するある望ましい分割を自動的に獲得する目的でよく利用される．

図1(a)のような論理回路の図面の画像から，この図面が表す命題論理式を出力するという画像認識の問題を考える．この図1(a)の画像は線分の集合によって対象を表したベクトルデータと呼ばれる画像であり，この画像は一般に次のような手順で認識される．最初に，図1(b)のようなANDゲートや接続線など何らかの意味を持った部分を構成していると推定される線分集合ごとに，この画像を分割する．次に，それらの線分集合にANDゲートといったラベルづけを行うことで記号的な表現を獲得し，この記号的表現から命題論理式を出力する．



(a) もとの論理回路図



(b) 分割後の論理回路図

図 1 望ましい分割を獲得することが必要な場合の例

このような認識問題の過程で望ましい分割を獲得することはセグメンテーションと呼ばれ，画像認識における領域分割問題[三宅 91]や音声認識における連続音声の音素区間への分割など多くの応用分野が存在する．また，機械学習の一つである多戦略学習でも知識変形(knowledge transmutation)の一つagglomerationとして望ましい分割の獲得のためにクラスタリングが利用されている[Michalski 93]．

ところで，この望ましい分割を獲得するための規準は「小さくまとまった丸く見える部分はひとまとまり」といった，人間の観察によって作成されることが多い．このようにして作成された規準を利用する場合，次の二つの問題が生じる．

第1の問題は，クラスタ分析の手法を利用するために，人間の観察によって得た定性的な規準を，分類対象の集合全体がどれだけ望ましい分割を獲得しているかを表す定量的な尺度や，二つの分類対象の間の類似度に変換して表す必要があるが，この変換は一般に非常に複雑なものになることである．

第2の問題は，仮に，与えられた分類対象の集合に対して望ましい分割を導くような規準を獲得できたとしても，その規準によって，未知の集合に対しても望ましい分割を獲得できることは保証されないということである．

これら問題のうち，前者を解決するために**次節**で新たな学習問題を定義し，**3章**でその解法を示す．後者については**4章**で述べる学習結果の評価方法の工夫によって解決する．

2・2 クラスタ例からの学習

クラスタリングを望ましい分割を獲得する目的で利用する場合，そのような分割を導く規準を獲得することが一般に困難であることが問題であると述べた．しかし，このような場合でも，前述のベクトルデータの例の場合のように，望ましい分割そのものを示すことは容易であることが多い．そこで，分類対象集合とその集合に対する望ましい分割の組である学習事例の集合から，未知の分類対象集合に対する望ましい分割を獲得するための規準を獲得する学習問題を考え，この問題を「クラスタ例からの学習」と名づける．

このクラスタ例からの学習は，ID 3[Quinlan 86]に代表される一般の「例からの学習」(本論文では，この学習を特に「種類例からの学習」と呼ぶことにする)と類似しているが，次に述べるような点で異なっている．

例からの学習は，分類対象とそれが分類されるべきクラスの組である事例から，どのクラスに分類対象を分類すべきかを決定する規準を獲得することである．一方，クラスタ例からの学習は，分類対象がどのクラスタに分類されるかを決定するのではなく，互いの分類対象が同じクラスタに分類されるべきかどうかを決定する規準を獲得する．このように，これら二つの学習は獲得すべき規準が異なっている．

3. クラスタ例からの学習の方法

本章では，種類例からの学習とクラスタ分析とを組み合わせたクラスタ例からの学習を，分類対象集合の記述の方法，学習の方法，学習結果を用いた望ましい分割の獲得方法の三つの段階に分けて述べる．

3・1 属性つきグラフ

本研究では，次に述べる「属性つきグラフ」によって分類対象集合を記述する．属性つきグラフ G は $(V, E, A(V), A(E))$ なる四つ組である．V は分類対象の集合 $\{v^1, v^2, \cdots, v^n\}$ であり，v^i を頂点，V を頂点集合と呼ぶ．E は分類対象の対の集合 $\{e^1, e^2, \cdots, e^m\}$ であり，e^k は頂点対 $\{v^i, v^j\}$ $(i \neq j)$ で，この対を辺と呼び，E を辺集合と呼ぶ．$A(V), A(E)$ はそれぞれ，頂点と辺に対する属性ベクトルの集合を表す．頂点 v^i の属性ベクトルを $A(v^i)=(a^1(v^i), a^2(v^i), \cdots, a^p(v^i))$ で，辺 e^i の属性ベクトルを $A(e^i)=(a^1(e^i), a^2(e^i), \cdots, a^q(e^i))$ と表記する．

説明のために，ドットの集合をクラスタリングする場合を想定した簡単な属性つきグラフの例を図2に示す．このグラフは，一つのドットが一つの頂点で表されており，5個の辺を含む．さらに，3個の属性を含む頂点の属性ベクトルと1個の属性を含む辺の属性ベクトルを持ち，頂点の属性 $a^1(v), a^2(v), a^3(v)$ はそれぞれ点の X 座標，Y 座標，および，色であり，辺の属性 $a^1(e)$ は点の間の距離である．このように，頂点の属性によって各分類対象の性質，辺の属性によって分類対象の間の性質を表すことで分類対象集合を表現する．

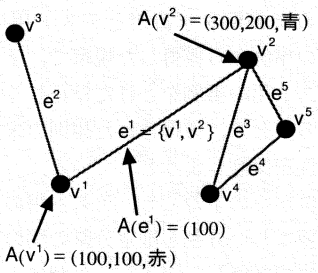


図 2 簡単な属性つきグラフの例

この属性つきグラフを用いてクラスタ例からの学習は次のように表すことができる．

> 分類対象集合を表現した属性つきグラフ G_j と，そのグラフの頂点の集合に対する望ましい分割 π_j の組である学習事例を K 個含む集合 $\{(G_1, \pi_1), (G_2, \pi_2), \cdots, (G_K, \pi_K)\}$ から，未知の属性つきグラフ G_{new} の頂点の集合に対する望ましい分割 π^*_{new} を求める規準を獲得する．ただし，分割 π は頂点の集合 V の部分集合であるクラスタ $\{C^1, C^2, \cdots, C^q\}$ の集合であり，すべての分類対象は必ずいずれか一つのクラスタの要素であり，かつ，複数のクラスタの要素であってはならない．

3・2 クラスタ例からの学習の方法

ここでは，望ましい分割の獲得のために分析すべき分類対象集合の性質について論じ，次に，クラスタ例からの学習の方法について述べる．

本研究では，内的なまとまりと外的な分離を二つの分類対象の間の類似度によって表すが，この分類対象AとBの間の類似度を定めるために，Michalskiらは[Michalski 88, p. 59]で，次のような分類対象集合の性質が重要であると述べている．

1. 分類対象AやBの性質
2. 分類対象AとBをともに含む集合の性質
3. クラスタ全体のゲシュタルト的性質

属性つきグラフでは，1の性質は頂点の属性によって，2の性質は辺の属性によって表すことができる．しかし，以下に提案する方法では，3のようなゲシュタルト的な性質の特徴を直接的には学習できない．このことは，以下に提案する方法の問題点であるが，属性つきグラフの頂点や辺の属性に対して**5・1節**で述べるような工夫によって，ある程度この問題は解消できる．

次に，クラスタ例からの学習の方法，すなわち，分類対象間の類似度を求める規則の学習方法について述べる．この規則は，1対の分類対象が同じクラスタの要素になるかどうか判別する規則として，以下の2段階の手続きで求める．

〔1〕 **属性ベクトルの作成**

最初に，K 個の事例 $(G_1, \pi_1), (G_2, \pi_2), \cdots, (G_K, \pi_K)$ のすべての辺 $e=\{v^i, v^j\} \in E_k$ $(k=1, \cdots, K)$ について，次の3種類の属性からなる属性ベクトル $A(e, v^i,$

v^j) をつくる．

- 辺 e の属性 $a^x(e)$.
- 属性 $a^x(v^i)$ と $a^x(v^j)$ が連続値属性のとき，これら二つのうち小さいほうの値をとる属性と，大きいほうの値をとる属性の二つの属性.
- 属性 $a^x(v^i)$ と $a^x(v^j)$ が r 個の属性値をとる離散値属性のとき，これらの属性値を2個組み合わせた r^2 個の値と，二つの属性のうちいずれか一方の属性値が不明である場合の r 個の値とを合わせた値をとる属性.

図2の辺 e^1 に対する属性ベクトル $A(e^1, v^1, v^2)$ の例を示す．1番目の属性は $a^1(e^1)$，2と3番目の属性は，それぞれ $a^1(v^1)$ と $a^1(v^2)$ の小さいほうと大きいほうの値，4・5番目は2・3番目と同様，6番目の属性は $a^3(v^1)$ と $a^3(v^2)$ を組み合わせた値“赤-青”であり，まとめると属性ベクトル $A(e^1, v^1, v^2)$ は(100,100,300,100,200,赤-青)となる．

〔2〕 **類似度を求める規則の学習**

辺 $e=\{v^i, v^j\}$ について v^i と v^j がともに，分割 π 中の同じクラスタの要素であるときに1，そうでない場合に0をとる関数を $ISC(e, \pi)$ とし，前段階で生成した属性ベクトルと関数 $ISC(e, \pi)$ の値の組($A(e, v^i, v^j)$, $ISC(e, \pi)$) を学習事例として，種類例からの学習によって $ISC(e_{new}, \pi^*_{new})$ が1となる確率を推定する規則 $f(e;G)$ を獲得する．ただし，e_{new} は未知の属性つきグラフ G_{new} の辺であり，π^*_{new} は G_{new} に対する望ましい分割である．これらの学習事例を **MOKSHA-3** アルゴリズムに与え規則 $f(e;G)$ を獲得した．この **MOKSHA-3** は，[神嶌 95]の **MOKSHA** に，規則の記述長の符号化の方法と符号長の短い規則を探索するときの評価関数に改良を加えたものであり，Rissanen の MDL 基準[Rissanen 83]に基づいて，ある対象の属性ベクトルから，その対象が特定のクラスに分類される確率を出力する規則を求める．

3・3 学習結果を利用したクラスタリングの方法

前節の方法で獲得した規則 $f(e;G)$ を利用して，未知の分類対象集合に対する推定分割 $\hat{\pi}_{new}$ を求める方法について述べる．本研究では，非類似度行列を入力とするクラスタ分析の代表的なクラスタリング手法：最小距離法，最大距離法，および，群平均法を利用して推定分割を求める．非類似度行列とは，n 個の分類対象がある場合には $n\times n$ の行列であり，その要素 d_{ij} は，分類対象 v_i と v_j の非類似度，すなわち，より似ているものほど小さく，また，$d_{ij}=d_{ji}$ を満たす値である．G_{new} に対するこの行列の各要素は，学習により獲得した規則 $f(e;G)$ を利用して次式で求める．

$$d_{ij}=\begin{cases}1-f(\{v^i, v^j\};G_{new}) & if\ \{v^i, v^j\}\in E_{new}\\ c & others\end{cases}$$

ただし，c はクラスタ分析の手法に依存した定数で，行列の対角要素は0とする．

この非類似度行列をもとに，推定分割 $\hat{\pi}_{new}$ を求める手続きを最小距離法を用いた場合について述べる．最小距離法は，一つの分類対象が一つのクラスタに相当する状態から始めて，二つのクラスタ間の非類似度の最も小さなクラスタの対を併合することを繰り返す．そして，クラスタ C_a と C_b を併合して C_n にしたとき，これと他のクラスタ C_x との間の非類似度を，分類対象 $v^\alpha\in C_n$ と $v^\beta\in C_x$ との間の非類似度 $d_{\alpha\beta}$ のうち最小のものにする．

他の二つの方法も，最大距離法では新たな非類似度を $d_{\alpha\beta}$ のうち最大のものに，また，群平均法ではその平均にする点だけが異なる．

ところで，すべての分類対象が一つのクラスタに含まれてしまう前に，適当な条件で併合を停止する必要がある．最後に，この条件と前述の非類似度行列を求めるときの定数 c とを，各クラスタリング手法ごとに分けて述べる．

最小距離法 定数 c は1.0，クラスタ間の非類似度が0.5未満になったときに併合を停止.

最大距離法 定数 c は0.0，クラスタ間の非類似度が0.5未満になったときに併合を停止.

群平均法 定数 c は0.5，学習用事例(G_x, π_x)の分割に含まれるクラスタ数の平均よりもクラスタ数が小さくなった場合に併合を停止.

4. 未知の分類対象集合に対する推定分割の評価

学習事例から獲得した規準では未知の分類対象集合に対する望ましい分割を獲得できることが保証されないことについて**2・1節**で述べた．種類例からの学習では，与えられた事例を学習用とテスト用の事例に分け，学習用の事例から獲得した規準で，学習用の事例ではなくテスト用の事例をどれだけ正しく分類できるかを評価することで，獲得された規則の未知の事例に対する有効性を保証しようとしている．

クラスタ例からの学習でも，テスト用の分類対象集合に対する望ましい分割と推定分割の間の類似性を次に述べる方法で評価し，望ましい分割により類似した分割を獲得することを学習の目標とした．K 個の事

例から，最初の事例(G_1, π_1^*)を取り除き，残りの$K-1$個の事例を学習事例としてグラフG_1の分割$\hat{\pi}_1$を推定し，後に述べる情報損失度を求める．この情報損失度を，残り$2\sim K$番目の事例についても求め，その平均によって学習の結果を評価した．

分割の類似性の定量的な尺度には[Rand 71]のものや[Jain 88, p. 18]のJaccard係数を利用したものなどがある．しかし，これらの尺度は0から1の間の値をとるが，実際にはほとんどの値がより限定された範囲に集中してしまうため実用上不便である．そこで，獲得すべき情報量のうちどれだけの割合の情報量を獲得できなかったかを表す情報損失度を用いた．n個の分類対象をもとにした$N=n(n-1)/2$個のすべての分類対象対のうち，望ましい分割π^*の同じクラスタの要素であり，かつ，推定分割$\hat{\pi}$でも同じクラスタの要素である分類対象対の数をn_{11}，π^*では同じだが$\hat{\pi}$では異なるクラスタの要素である分類対象対の数をn_{10}とする．n_{00}とn_{01}も同様に定める．

$$\text{情報損失度}=\frac{\sum_{i=0}^{1}\sum_{j=0}^{1} n_{ij}\log((n_{0j}+n_{1j})/n_{ij})}{\sum_{i=0}^{1}(n_{i0}+n_{i1})\log(N/(n_{i0}+n_{i1}))}$$

情報損失度はπ^*と$\hat{\pi}$が一致したときに限り0となり，1以下である．

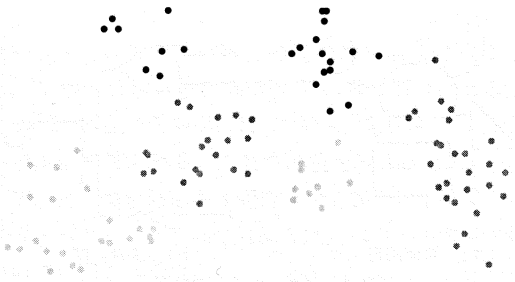
この式は分割そのものではなく分類対象対に関する情報損失度を表している．すなわち，分類対象A，BとCがあった場合にABとACは同じクラスタの要素だが，BCが異なるクラスタの要素になることはあり得ないといったような制約を無視している．しかし，二つの分割の類似性を比較する目的には十分利用可能である．

5. 実　　験

5・1 実 験 対 象

クラスタ例からの学習を，数値分類の分野での実験で一般的に用いられるドットパターンと，画像処理への応用であるベクトルデータの2種類のデータに対して適用した．

第1のデータ「ドットパターン」とは，図3のような2次元空間上のドットで各分類対象を表現したものである．各クラスタは，表1に示した円状の分布に従って発生したドットの集合である．ただし，表中のRは各クラスタの勢力範囲の半径で，この勢力範囲は隣接するクラスタの勢力範囲と接触するように定めた．クラスタリングは，均一とガウスではガウスのほうが，クラスタの分布の範囲に関しては重複，接触，分離の順に容易である．このドットパターンを，各ドットが頂点に相当し，頂点のすべての対に辺が存在する属性つきグラフに変換した．これらグラフはすべて，50個の頂点と1 225個の辺を含み，頂点には4種類，辺には8種類の属性を付加した．



（a） 均一・接触　　（b） ガウス・接触

図 3 ドットパターンの例

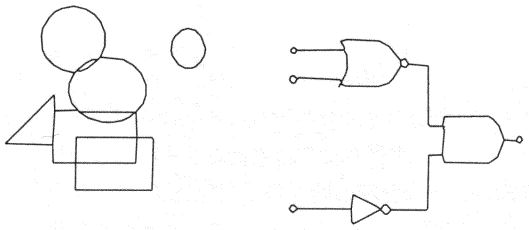
表 1 ドットパターン中のクラスタのドットの分布

ガウス・重複	標準偏差$R/2.0$のガウス分布
ガウス・接触	標準偏差$R/2.5$のガウス分布
ガウス・分離	標準偏差$R/3.0$のガウス分布
均一・重複	半径$1.3R$の円内に均一分布
均一・接触	半径$1.0R$の円内に均一分布
均一・分離	半径$0.7R$の円内に均一分布

ところで，**3・2節**で，ゲシュタルト的な性質をある程度反映した分割について学習することが属性の工夫により可能であると述べたが，このような工夫を行った属性の一つについて述べる．まず，ドットパターンに対して最小距離(Minimum Spanning Tree)を考える．最小距離木とは，与えられたドットすべてを結ぶ木のなかで，枝の長さの合計が最小のものである．辺の両端の頂点に対応するドットを結ぶ最小距離木の上のパスを見つけ，そのパスに含まれる頂点の数を属性として用いた．この量は，二つのドットの間の領域のドット密度といったゲシュタルト的性質を反映することがグラフ理論的クラスタリングの[Zahn 71]に報告されている．このような性質を備えた属性を利用することにより，単に二つのドットの関係に基づいたクラスタではなく，ゲシュタルト的な性質を反映したクラスタを獲得するための規準を学習することができる．

第2のデータ「ベクトルデータ」とは線分の集合によって対象を表した画像で幾何図形と論理回路と名づけた2種類の画像に対して実験を行った．

「幾何図形」とは，楕円，長方形，および三角形の線図形をランダムに4〜6個描いた図4（a）のようなビットマップ画像をベクトル化して生成した画像である．この画像に対して，ひとまとまりの図形，すなわち，三角形や長方形1個分を表す線分の集合を一つの



（a） 幾何図形　　（b） 論理回路

図 4 ベクトルデータの例

表 2 ベクトルデータの属性つきグラフの特徴

事例集合名	クラスタ数	頂点数	辺数
論理回路・ランダム	16.7	102.9	2776
論理回路・近傍	16.7	102.9	152
幾何図形・ランダム	4.99	55.5	833
幾何図形・近傍	4.99	55.5	86

クラスタとする分割の獲得を試みる．

「論理回路」とは，図 4（b）のような，AND ゲートなどの 5 種類の部品から構成された，手書きの図面をイメージスキャナで計算機に入力したものをベクトルデータに変換したもので，著者が[神嶌 95]で用いたものである．この画像から，一つの部品を表す線分の集合を一つのクラスタとする分割の獲得を試みる．

これらのベクトルデータを，各線分が一つの頂点に相当する属性つきグラフに変換した．頂点の対から適当な条件を満たすものを選択して辺とするのだが，その条件の異なる「ランダム」と「近傍」の 2 種類のグラフを作成した．ランダムは頂点のすべての対のうち無作為に半分を選択したもので，近傍は線分の最も近い端点の間の距離が画像の 1 辺の長さの 1％以下の対を選択したものである．また，頂点には 8 種類，辺には 7 種類の属性を付加した．これらの属性つきグラフの頂点，辺，およびクラスタの平均数を表 2 に示す．

クラスタリングは，幾何図形と論理回路とでは，事例当りの分類対象の数，クラスタ数が多く，さらに，手書き画像であるため図形の形状の散らばりの大きい論理回路のほうが困難である．近傍とランダムでは，近傍にある線分が，図面中の図形の形状を表現するのに重要であるというヒューリスティックを利用しているため近傍のほうが容易である．

5・2 実験内容と結果

本節では 4 種類の結果を示し，**次節**以降でこれらの結果について考察する．ただし，実験は 100 個の事例を含む事例集合を対象に行った．

第 1 に，望ましい分割を獲得できたかを評価する前に，分類対象の間の類似度を求める規則を評価する．

評価は，推定分割の評価に用いた **4 章**の情報損失度と，種類例からの学習では一般的な正解率とによって行った．ただし，正解率は，同じ事例集合に対する異なる学習方法の結果を比較する目的には適しているが，異なる事例集合に対する同じ学習方法の結果を比較する目的には，正規化などの問題から不向きなため参考にとどめ，**次節**では情報損失度を中心に議論する．

4 章では，分類対象集合のすべての対のうちどれだけの対が正しく判別できたかを情報損失度によって評

表 3 分類対象対の類似度を求める規則に対する実験結果

事例集合名	正解率	情報損失度
ガウス・重複	.878	.513
ガウス・接触	.949	.252
ガウス・分離	.981	.106
均一・重複	.763	.775
均一・接触	.840	.591
均一・分離	.991	.044
論理回路・ランダム	.960	.543
論理回路・近傍	.889	.541
幾何図形・ランダム	.893	.588
幾何図形・近傍	.897	.464

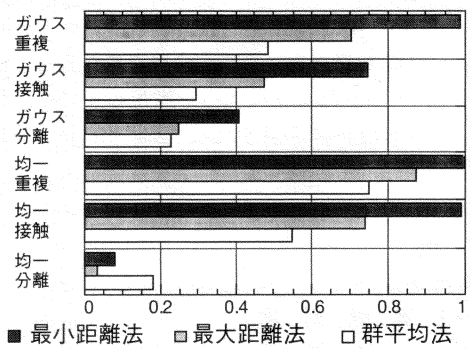


図 5 ドットパターンに対する推定分割の情報損失度の平均

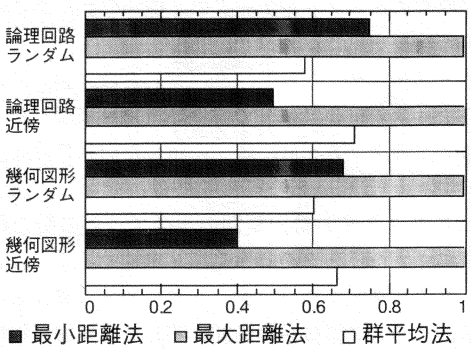


図 6 ベクトルデータに対する推定分割の情報損失度の平均

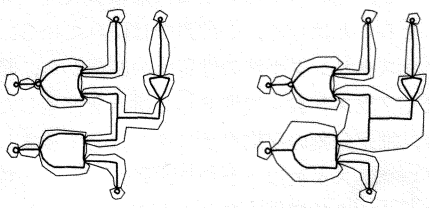
表 4 推定分割のクラスタ数から望ましい分割のクラスタ数を引いた値の平均

(a) ドットパターン

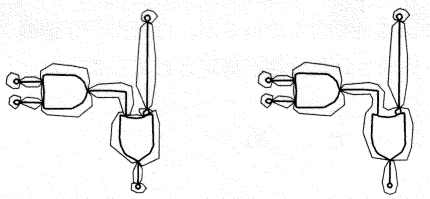
	最小距離法	最大距離法	群平均法
ガウス・重複	−1.90	3.75	0.08
ガウス・接触	−1.40	2.29	0.04
ガウス・分離	−0.74	0.96	−0.02
均一・重複	−1.97	4.77	0.03
均一・接触	−1.94	3.59	0.01
均一・分離	−0.13	0.16	−0.17

(b) ベクトルデータ

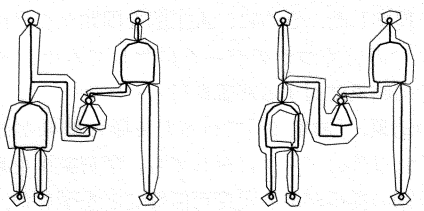
	最小距離法	最大距離法	群平均法
論理回路・ランダム	0.65	3.73	−0.25
論理回路・近傍	−2.26	−12.68	−5.63
幾何図形・ランダム	6.58	7.57	−0.01
幾何図形・近傍	0.52	−0.91	−1.18



(a) 情報損失度が最小であった事例(情報損失度：0.113)



(b) 情報損失度が30番目に小さかった事例(情報損失度：0.410)



(c) 情報損失度が50番目に小さかった事例(情報損失度：0.482)

図 7 論理回路・近傍に対する望ましい分割と推定分割の例

価したが，分類対象の間の類似度は，テスト用事例の属性つきグラフの m 個の辺を対象に求めるので，これらの辺がどれだけ正しく判別できたかを情報損失度や正解率によって評価する．

また，ここで用いる情報損失度は，**4章**の情報損失度の式の n_{ij} が，$i=1$ のとき辺 e の両端の頂点が望ましい分割の同じクラスタの要素である場合の辺の数を表し，$j=1$ のとき獲得した規則を用いて $f(e;G)$ の値が0.5よりも大きい場合の辺の数を表している点と，N 個の分類対象が m 個の辺の数に置き換わる点とが異なる．正解率は，この n_{ij} を用いて，$(n_{00}+n_{11})/m$ で表される．これらの値を表3に示す．

第2に，推定分割を**4章**の方法で評価した結果を，ドットパターンについては図5に，ベクトルデータについては図6に示す．

第3の推定分割が含むクラスタ数に対する実験結果について述べる．この実験では，望ましい分割におけるクラスタ数の推定を検証する目的で，推定分割のクラスタ数から望ましい分割のクラスタ数を引いた値の平均を求め，その値を表4に示す．

最後に，獲得された推定分割の具体例を示す．図7は論理回路・近傍に最小距離法を適用して推定された分割(右側)とその事例に対する望ましい分割(左側)である．ただし，図ではクラスタを構成する線分集合を細線で囲んで表示している．

5・3 [神嶌 95]の結果との比較

[神嶌 95]の実験も今回も，分類対象対の類似度をもとに分割を推定している点で基本的に同じであるが，次のような部分的な改良により類似度を求める規則を改善した．

- ゲシュタルト的な性質を反映した属性の採用
- 学習アルゴリズムの改良
- 属性つきグラフを学習事例に変換する方法
- 属性つきグラフの生成時に分類対象対から辺を選択する規準の改良
- 事例数を50から100に増やした

類似度を判別する規則の情報損失度は，[神嶌 95]の全体・1次属性実験の場合0.964，近傍・1次属性実験の場合0.865であったが，表3の論理回路の情報損失度は0.54程度になり大幅な改善が見られた．推定分割に対する情報損失度も，[神嶌 95]では0.68程度もあったが，図6のように事例集合ランダムでは群平均法を用いて0.581，近傍で最小距離法を用いて0.495に減少させることができた．

さらに，結果を定性的な面から評価すると，図7(a)では誤って二つのクラスタを一つに併合した誤りが2か所，図(b)では三つのクラスタを一つに併合した誤りが1か所存在する程度の誤りであり，これらの推定分割に望ましい分割と視覚的な類似性を十分に見い出せ，画像認識でも利用可能であると考える．しかし，図7の(c)の推定分割は望ましい分割とはかなり異なり，このような分割の利用は難しいと考える．こ

れらのことから，情報損失度が0.45程度以下の推定分割が利用可能であると考えると，[神嶌 95]ではこのような推定分割はほぼ皆無であったが，今回は全体の30〜40%の事例に対してこのような分割を推定できた．このことにより，クラスタ例からの学習の結果を実用的な問題で有効に利用できる可能性を示せたと考える．

5・4 論理回路以外のデータの結果に対する考察

図6の結果では，推定分割の情報損失度は，論理回路とほぼ同等もしくはそれ以下であり，クラスタ例からの学習が[神嶌 95]以外の問題にも利用できることが示されている．また，同じベクトルデータであれば，属性の種類などの学習の条件は同じものを用いて問題がないことも示されている．

次にドットパターンに対する結果について，最も情報損失度の小さな手法の結果を基準に考察する．これは，同じ事例集合でもクラスタリング手法によって結果が異なるためである．

図5のドットパターンで，重複，接触，および分離の三つを比較した場合，クラスタがよく分離されている順に推定分割の情報損失量が減少している．どのようなクラスタリング手法でも分割の獲得が困難な程度にクラスタを不明確にした重複では情報損失度が大きい．それでも，ガウス分布では30〜40%の事例に対して利用可能と考えられる分割が獲得できている．分割抽出が容易なように作成した分離では，ほとんどの事例に対して利用可能な分割を獲得でき，均一分布では83%の事例に対して望ましい分割そのものを推定することができ，非常に良好な結果が得られた．これらの中間的性質を持つ接触は，何らかのクラスタリング手法を用いれば分割の抽出がある程度は可能なように作成したが，均一分布ではそれほど情報損失度が小さくならなかった．これは，ゲシュタルト的な性質でも，点の密度やクラスタの分離の具合といった性質は反映できるようにしたが，クラスタの形状といった性質を反映する属性を考案できなかったことなどが原因であったと考える．

5・5 クラスタリング手法に対する考察

図5や図6に見られるように，類似度を推定する規則は同じであるにもかかわらず，クラスタリング手法の違いによって推定される分割は大きく変わり，安定して分割を推定できない問題がある．この問題は，各クラスタリング手法のどのような性質によって生じているのかについて考察する．

最小距離法には，たった1対の誤って判別された分類対象対の影響によって，過剰な併合が生じる性質が，最大距離法には過剰な分割が生じる性質がある．この性質の影響のため，表4の結果では，最小距離法では過剰な併合のためクラスタ数の差が負に，逆に最大距離法では過剰な分割のためこの差が正になっている．また，図5では最小距離法を用いたいくつかの場合，図6では最大距離法を用いた場合に，分類対象対のわずかな判別誤りの影響を受けて，情報損失度が非常に大きくなっている．

一方，群平均法は，[鷲尾 89, p.247]にも限定つきではあるが広い意味で良い結果を与えることが述べられている．他の方法のように，情報損失度が非常に大きくなることはなく，この点では優れている．しかし，クラスタリング手法の性質上の制限から，類似度ではなくクラスタ数を規準として併合を停止した．このため，推定分割の平均クラスタ数は必ず望ましい分割の平均クラスタ数以下となり，また，多くの場合この平均クラスタ数個のクラスタからなる分割が推定される．実験でも，ドットパターンの学習事例は2〜4個のクラスタを含んでいるが，ほとんど3個のクラスタから構成された分割が推定されてしまった．

6. ま と め

本研究では，クラスタ例からの学習なる新たな問題の学習について，その学習法とその結果の定量的評価方法を提案した．さらに，人工的な問題から実用的な問題までのいくつかのデータに対して本手法を適用し，その結果に対する評価を行った．

その結果，[神嶌 95]で行った実験の結果が改善され，それ以外のデータに対しても，学習結果をもとにした分割の獲得がある程度可能なことを示した．さらに，より好ましい分割を獲得できるようにするために，各クラスタリング手法を本学習問題に適用する際の問題点について述べた．

今後は，ゲシュタルト的な性質を直接的に扱える枠組みと，今回明らかにした問題点に配慮したクラスタリング手法について研究したい．

◇参 考 文 献◇

[Everitt 93] Everitt, B. S.: *Cluster Analysis*, p. 6, 3rd edition, Edward Arnold (1993).

[Fisher 87] Fisher, D. H.: Knowledge Acquisition via Incremental Conceptual Clustering, *Machine Learning*, Vol. 2, pp. 139-172 (1987).

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

[神嶌 95] 神嶌敏弘, 美濃導彦, 池田克夫：非納学習を用いた図面部品の抽出と分類のための規則の形成, 情処学論, Vol. 36, No. 3, pp. 614-626 (1995).

[Michalski 88] Michalski, R. S., 電総研人工知能グループ訳：概念クラスタリング構造をもつ対象のゴール指向分類, Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.), 発見的学習, 第2章, pp. 56-84, 共立出版(1988).

[Michalski 93] Michalski, R. S.: Inferential Theory of Learning as a Conceptual Basis for Multistrategy Learning, *Machine Learning*, Vol. 11, pp. 111-151 (1993).

[三宅 91] 三宅 誠ほか：視覚と画像工学—見る・見せる—, 信学誌, Vol. 74, No. 4, pp. 309-408 (1991).

[Quinlan 86] Quinlan, J. R.: Induction of Decision Trees, *Machine Learning*, Vol. 1, pp. 81-106 (1986).

[Rand 71] Rand, W. M.: Objective Criteria for the Evaluation of Clustering Methods, *J. of the American Statistical Association*, Vol. 66, pp. 846-850 (1971).

[Rissanen 83] Rissanen, J.: A Universal Prior for Integers and Estimation by Minimum Description Length, *The Annals of Statistics*, Vol. 11, No. 2, pp. 416-431 (1983).

[鷲尾 89] 鷲尾泰俊, 大橋靖雄：多次元データの解析, シリーズ 入門統計的方法3, 岩波書店(1989).

[Zahn 71] Zahn, C. T.: Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters, *IEEE Trans. on Computers*, Vol. C-20, No. 1, pp. 68-86 (1971).

〔査読者：阿曽弘具〕

著 者 紹 介

神嶌 敏弘(正会員)

1992年京都大学工学部情報工学科卒業. 1994年同大学院修士課程修了. 同年, 電子技術総合研究所入所. 機械学習とその応用の研究に従事. 情報処理学会会員.

新田 克己(正会員)

1975年東京工業大学工学部電子工学科卒業. 1977年同大学院修士課程電子物理工学専攻修了. 1980年同大学院博士課程電子物理工学専攻修了. 工学博士. 同年, 電子技術総合研究所に入所. 1989～95年(財)新世代コンピュータ技術開発機構に出向. 1995年より電子技術総合研究所知能情報部推論研究室室長. 1996年より東京工業大学大学院総合理工学研究科教授を併任. 知識情報処理技術の応用システムの開発に興味を持つ. 情報処理学会会員.