人　工　知　能　101－4
（１９９５．　９．　１２）

クラスタリングに対する例からの学習

神嶌敏弘　　新田克己

mailto:kamisima@etl.go.jp
http://www.etl.go.jp:8080/etl/suiron/~kamisima/

電子技術総合研究所　推論研究室
〒305 茨城県つくば市梅園 1-1-4

本研究では，分類対象の集合を，『似ている』ものどうしを集めたクラスタなる部分集合に分割するクラスタリングを対象にした『クラスタリングに対する例からの学習』を取り上げ，この学習を行った．

クラスタリングは，画像処理の領域分割などの分野で利用されているが，クラスタリングの利用者が望む分割を獲得することが困難であることが多い．そこで，分類対象の集合とその集合の望ましい分割の組である学習事例の集合から，未知の分類対象の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を新たに考えた．従来の例からの学習と数値分類の手法を組み合わせてこの学習を行い，その結果を評価した．

Learning from Examples for Clustering

Toshihiro KAMISHIMA and Katsumi NITTA

Machine Inference Section, Electrotechnical Laboratory
1-1-4 Umezono, Tsukuba, Ibaraki, 305 Japan

We solve the novel machine learning problem: "Learning from Examples for Clustering," that handles the clustering, that is the method to divide the set of individuals to the subsets containing the "similar" individuals.

Though the clustering is the method used in image understanding and some other fields, clustering users offten gain the partitions which they do not wish. Therefore, we try an approach to get desirable partition for an unknown set of individuals from the set of examples, which are pairs of a set of individuals and the desirable partition for the set. We try this approach with the technique of the learning from examples and the numerical taxonomy, and propose criteria to evaluate results of this learning problem.

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の学習の手法を考案し，その結果を定量的に評価する．

クラスタリングは，分類対象の集合を，何らかの規準に基づいて『似ている』ものどうしを集めたクラスタに分割することである．このクラスタリングを分類対象の集合に対するある望ましい分割を獲得する目的で用いる場合，その望ましい分割を直観的でない数値に置き換えて表す必要があるため，そのような分割を得ることができないことが多い．そこで，分類対象の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，未知の分類対象の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問題を考え，これを『クラスタリングに対する例からの学習』と呼ぶことにする．

以後，2節では，このクラスタリングに対する例からの学習を定式化し，3節では，従来の例からの学習と数値分類のクラスタ分析とを組み合わせてこの学習を行う方法と，この学習結果の評価方法について述べる．4節では，2種類のデータを対象に実験し，その結果について考察する．5節では，まとめと今後の予定について述べる．

2 クラスタリングに対する例からの学習

『分類』全体の中で，クラスタリングに対する例からの学習がどのような位置にあるかについて述べ，この学習の定式化を行う．

分類を2種類：『クラス分け』と『クラスタリング』に分け，これらを対比して考える．

第一の分類：『クラス分け』を，ここでは，1個の分類対象を，事前に定めたクラスの中のいずれかに分類することとする．この分類に対する学習問題，すなわち，分類対象とそれが分類されるべきクラスの組である学習事例の集合から，未知の分類対象を望ましいクラスに分類するための規準を獲得する問題は，機械学習では例からの学習，数値分類では判別分析と呼ばれ多くの研究がある．本論文では，この学習を特に『クラス分けに対する例からの学習』と呼ぶ．

第二の分類：『クラスタリング』を，ここでは，分類対象の集合を，その中の分類対象どうしは似ていて，その外にある分類対象とは似ていないようなクラスタと呼ぶ部分集合に分割することとする．この分類は，機械学習では観察による学習や概念形成と呼ばれ，FisherのCOBWEB[2]などの研究があり，数値分類ではクラスタ分析と呼ばれ多くの研究がある．[1]

クラスタリングは，分析者の問題に対する専門的見地に基づく裏付けがあったうえで，分類対象の集合にそれを求めて，それらを要約する手段であって[4]，ある分類対象の集合に対する，望ましい分割があって，それを自動的に獲得する手法ではない．にもかかわらず，クラスタリングは，画像処理における領域分割など多くの分野で自動的に望ましい分割を獲得する手法として利用されているため，そのような分割を獲得できない場合が多い．

本研究では，自動的に望ましい分割を獲得できるようにするために，分類対象の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，未知の分類対象の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問題に取り組み，この問題を『クラスタリングに対する例からの学習』と名付けた．この学習を，分類対象の集合$V_j=\{v_j^1,v_j^2,\ldots,v_j^n\}$に対する記述$G_j$とその集合に対する望ましい分割$\pi_j$の組である学習事例を$K$個含む集合$\{(G_1,\pi_1),(G_2,\pi_2),\ldots,(G_K,\pi_K)\}$から，未知の分類対象の集合$G_{new}$に対する望ましい分割$\pi^*_{new}$を得るための規準を獲得することであると定式化する．ただし，分割πは分類対象の集合Vの部分集合であるクラスタ$\{C^1,C^2,\ldots,C^g\}$の集合であり，全ての分類対象は必ずいずれか一つのクラスタの要素でなければならない．

3 クラスタリングに対する例からの学習の方法

本節では，クラス分けに対する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リング

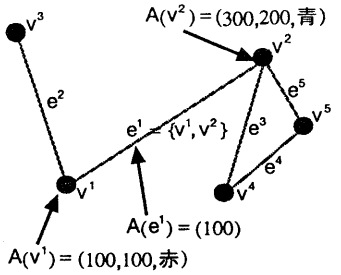

図1: 属性付グラフの例

に対する例からの学習を，分類対象の集合の記述の方法・学習の方法・学習結果の評価方法の三つの段階に分けて述べる．

3.1 属性付グラフ

本研究では，分類対象の集合の記述法として，以下に定める属性付グラフを用いた．

属性付グラフ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$\{v^1, v^2, \ldots, v^n\}$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$\{e^1, e^2, \ldots, e^n\}$であり，$e^k$は頂点の対$\{v^i, v^j\}, i \neq j$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$A(v^i) = (a^1(v^i), a^2(v^i), \ldots, a^p(v^i))$で，辺$e^i$の属性ベクトルを$A(e^i) = (a^1(e^i), a^2(e^i), \ldots, a^q(e^i))$と表記する．

点の集合をクラスタリングする場合の，説明のための簡単な属性付グラフの例を，図1に示す．このグラフは，各点が頂点で表されており，5個の頂点と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，及び，色であり，辺の属性$a^1(e)$は点の間の距離である．

3.2 学習の方法

本研究では学習を次の2段階で行った．第1段階では，クラス分けに対する例からの学習の手法を用いて分類対象の集合の事例から，1対の分類対象が同じクラスタに含まれるかどうかを判別する規則を獲得する．第2段階では，この規則とクラスタ分析とを用いて未知の分類対象の集合に対する分割を獲得する．

まず，第1段階での，規則の獲得について述べる．K個の事例$(G_1, \pi_1), (G_2, \pi_2), \ldots, (G_K, \pi_K)$の全ての辺$e = \{v^i, v^j\} \in E_k$, $k = 1, \ldots, K$について，次の3種類の属性からなる属性ベクトル$A(e, v^i, v^j)$を作る．

- 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方の値をとる属性と，大きい方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$r \times (r+1)$個の属性値をとる属性

図1の辺e^1に対する属性ベクトル$A(e^1, v^i, v^2)$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$a^1(v^2)$の小さい方と大きい方の値，4・5番目は2・3番目と同様，6番めの属性は$a^3(v^1)$と$a^3(v^2)$を組み合わせた値"赤－青"であり，まとめると$A(e^1, v^1, v^2)$は属性値$(100, 100, 300, 100, 200,$赤－青$)$となる．

クラス分けに対する例からの学習を用いて，この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，$ISC(e_{new}, \pi^*_{new})$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，$e_{new}$は未知の属性付グラフ$G_{new}$の辺であり，$\pi^*_{new}$は$G_{new}$に対する正しい分割である．また，$ISC(e, \pi)$は，辺$e = \{v^i, v^j\}$について$v^i$と$v^j$が共に，分割$\pi$中の同じクラスタ$C^x$の要素であるとき1，そうでない場合に0をとる関数である．

これらの学習事例を，文献[3]のMOKSHAを改良した，MOKSHA–3アルゴリズムに与え，規則$f(e; G)$を獲得した．

第2段階で，この規則を利用して，未知の属性付グラフG_{new}に対する推定分割$\hat{\pi}_{new}$を求める．分割の推定は，学習により獲得した規則$f(e; G)$

を利用して，G_{new}の頂点集合V_{new}中の任意の二つの頂点の間の非類似度を定め，非類似度行列を作る．頂点v^iとv^jの対の非類似度を次に示す．

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

ただし，cはクラスタ分析の手法に依存した定数．

この非類似度行列をもとに，クラスタ分析の代表的な三つの手法：最小距離法，最大距離法，及び，群平均法，によって分割$\hat{\pi}_{new}$を推定した．これら三つの手法では，非類似度行列を求めるときの定数cとクラスタの併合を停止する条件が異なるが，これらについて以下に記す．

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

3.3 推定分割$\hat{\pi}$の評価方法

クラスタリングの結果は，それを図示して観察するなどの方法で定性的に評価されてきた．このような評価も重要ではあるが，その結果を応用する場合，より良い結果を利用すべきであるし，また，結果を利用する場合に，結果の良さの目安があれば便利である．そこで，以下のような定量的な結果の評価を行った．

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量を求める．この情報損失量を，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．

情報損失量は，音声認識の評価などで利用される量で，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．取り除いたグラフG_tの頂点の対$e_t=\{v_t^i,v_t^j\},v_t^i,v_t^j\in V_t,i\neq j$に関して，$ISC(e_t,\pi_t^*)$が0である事象を$a_0$，1である事象を$a_1$，同様に$ISC(e_t,\hat{\pi}_t)$に関して事象$b_0$と$b_1$を定めたとき，情報損失量は次式で与えられる．

$$RIL=\frac{\sum_i\sum_j -P(a_i,b_j)\log P(a_i|b_j)}{\sum_i -P(a_i)\log P(a_i)}$$

この量はπ_t^*と$\hat{\pi}_t$が一致したときに限り0となり，1以下である．また，この定義では，辺相互の制約条件などを無視しているため，厳密に情報量の損失の割合を示してはいないが，二つの分割を比較する目的には十分利用可能であると考える．

4 実験結果

4.1 実験内容

クラスタリングに対する例からの学習を，実験的な人工問題であるドット・パターンと，画像処理への応用である文献[3]のベクトル・データの2種類のデータについて行った．

ドット・パターンのデータとして，各クラスタ中のドットの分布が均一分布のものと，ガウス分布のものの2種類，クラスタの領域が，十分に分離しているもの，接触しているもの，わずかに重複しているものの3種類，合計6種類の事例集合を用意した．クラスタリングは，ガウス分布の方が均一分布よりも容易であり，クラスタの領域が重複しているものよりも，十分に分離していいるものの方が容易である．

ベクトル・データでは『論理回路』と『幾何図形』と名付けた2種類の画像を用意した．これらの画像それぞれについて，辺の存在する条件が異なる『ランダム』と『近傍』と呼ぶ属性付グラフを作成した．クラスタリングは，『幾何図形』の方が『論理回路』よりも容易であり，『近傍』の方が『ランダム』よりも容易である．

4.2 実験結果

合計10種類の属性付グラフをそれぞれ100個集めて事例集合を作成した．これらの属性付グラフについて，ドット・パターンもしくは，ベクトル・データのものであれば共通した特徴を表1に，各グラフごとに異なる特徴を表2にまとめた．ただし，表2のgは学習事例として与えた分

分類集合名	頂点の属性数	辺の属性数
ドット・パターン	4	8
ベクトル・データ	8	7

表 1: 同種の属性付グラフに共通の特徴

名称	g	n	m	Acc	RIL
ガウス重複	2.92	50	1225	.878	.513
ガウス接触	2.96	50	1225	.949	.252
ガウス分離	3.02	50	1225	.981	.106
均一重複	2.97	50	1225	.763	.775
均一接触	2.99	50	1225	.840	.591
均一分離	3.02	50	1225	.991	.044

(a) ドット・パターン

名称	g	n	m	Acc	RIL
論理回路ランダム	16.7	102.9	2776	.960	.543
論理回路近傍	16.7	102.9	152	.889	.541
幾何図形ランダム	4.99	55.5	833	.893	.588
幾何図形近傍	4.99	55.5	86	.897	.464

(b) ベクトル・データ

表 2: 各属性付グラフごとに異なる特徴

割πに含まれるクラスタ数の平均，nとmは，それぞれ，属性付グラフの頂点と辺の平均数である．また，Accは，属性付グラフに含まれる辺eの$ISC(e,\pi^*)$の値が，クラスタリングを行う前に正しく判別されていることに関する正解率であり，RILはこのことに関する情報損失量である．

これらの属性付グラフについて，クラスタリングに対する例からの学習を行い，その結果に対して3節の情報損失量を求めた．これらの量を，ドット・パターンについては図2に，ベクトル・データについては図3にそれぞれ示す．

4.3 考察

最初に，従来から行われてきたクラスタリングの結果に対する定性的な評価と，3節の情報損失量による定量的との対応について述べる．

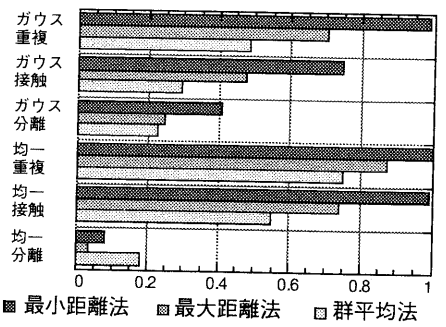

図 2: ドット・パターンに対する情報損失量

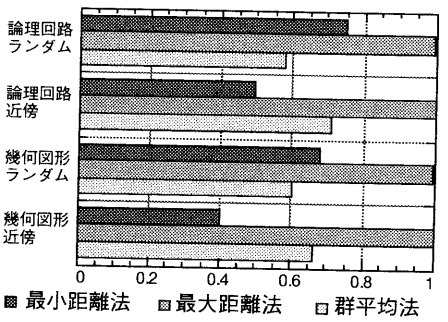

図 3: ベクトル・データに対する情報損失量

著者の観察によると，情報損失量が約0.4より大きくなると正しい分割π^*と推定分割πの間に視覚的な類似性を見い出せなかった．

図2のドット・パターンの場合，最小距離法では均一・分離のみ，最大距離法ではガウス・分離と均一・分離の二つ，そして，群平均法ではガウスのものと均一・分離の四つの事例集合で，多くの推定分割に対し正しい分割との視覚的類似性を見い出すことができた．

図3のベクトル・データの場合，文献[3]では情報損失量の平均が0.7程度もあり，ほとんどの分類対象の集合について視覚的な関連を見い出せなかった．それに対し，今回の実験では，属性付グラフの属性の選択の工夫と学習方法の改良によって，最小距離法を用いた論理回路・近傍において40%，幾何図形・近傍において50%程度の分類対象の集合に対して視覚的に関連を見い出せ

る分割を獲得できるようになった.

次に，表2のクラスタリング前の情報損失量と，図2と図3のクラスタリング後の情報損失量との関連について考察する．ドット・パターンの結果では，どの場合でも，クラスタリング前の情報損失量が小さいほど分割をより正しく推定できている．しかし，ベクトル・データでは，単純にこのような結果にはなっていない．ランダムと近傍の2種類の結果について比較すると，最小距離法の場合にはクラスタリング前の情報損失量が小さい方がより正しく分割を推定できているが，群平均法では逆に正しく分割できなくなっている．最大距離法では，クラスタリング前の結果に関わりなく正しい分割を獲得できていない．また，幾何図形・近傍の最小距離法の結果などは，クラスタリング前の情報損失量よりもより小さな情報損失量を実現する分割を獲得できでいる．

実験を行う以前は，3節で述べた学習方法の第一段階のクラス分けに対する例からの学習の結果を改善しさえすれば，クラスタリングに対する例からの学習の結果もそれに従い改善されると考えられるため，この学習は従来の学習と本質的に差はないとも考えられた．しかし，上記の結果から，たとえクラスタリング前の結果が改善されても，正しい分割が求められるようにならない場合や，逆に，クラスタリング前の結果がそれほど良くない場合でも，属性付グラフの辺の選択やクラスタリング手法に工夫によって，より正しいクラスタが求めることができる場合があることが分かる．このように，クラスタリングに対する例からの学習には固有の問題があり，この問題が本質的に従来の学習問題とは異なっていることが示されている．

最後に，ベクトル・データに対して最小距離法を用いた場合，属性付グラフの辺の選択規準の違いによってクラスタリング後の情報損失量が変わる理由について，最小距離法のもつ特徴をふまえて考察する．

最小距離法には，正しくは異なるクラスタに分類されるべきところが誤って同じクラスタに分類されると判断されたれ分類対象の対，すなわち，$ISC(e,\pi^*)=0$ となるべきが，誤って $ISC(e,\hat{\pi})=1$ と推定された辺 e が，クラスタの間にたった一つでも存在すると，それら二つのクラスタは併合されるため，この種の誤りに非常に弱い欠点がある．実際に，この誤り率は，論理回路の場合，ランダムでは32.5%，近傍では7.6%であり，最小距離法の結果のすぐれている近傍の方が小さい．一般に，この誤りを0にすることは困難であるため，この方法によって正しい分割を推定することにも限界が生じる．このことが，最小距離法をクラスタリングに対する例からの学習に利用する場合には問題となる．

同様に，最大距離法や群平均法にも長所・短所があり，従来のクラスタ分析の手法を用いる限り，どのようなデータに対しても，クラスタリング前の推定結果に応じた正しい分割を獲得することは難しい．よって，この学習に適した新たな手法が必要であると考える．

5 まとめ

本研究では，クラスタリングに対する例からの学習なる新たな問題の学習法を提案し，その結果を定量的に評価した．

実験の結果，この問題が従来の学習と本質的に異なるものであること，既存の手法によってこの問題を解くには問題があることが分かった．

今後は，新たなクラスタリング手法を中心に研究を進めたい．

参 考 文 献

1) Everitt, B. S.: *Cluster Analysis*, Edward Arnold, third edition (1993).

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

3) 神嶌敏弘, 美濃導彦, 池田克夫：帰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, 情処論, Vol. 36, No. 3, pp. 614–626 (1995).

4) 大隅昇：クラスター分析はどう使われるか, 数理科学, No. 190, pp. 26–34 (1979).