知能と情報（日本知能情報ファジィ学会誌）Vol.17, No.4, pp.459-467(2005)

原著論文

カーネル関数を用いた階層的クラスタリング†

春山 秀幸[*1]・遠藤 靖典[*2]・大久保 貴義[*1]

従来のクラスタリングアルゴリズムにおいて，分類境界面が線形に分離不可能なデータに対しては，良好なクラスタリング結果を得ることは難しい．これは通常，各個体間の非類似度として，ユークリッド空間におけるノルムの自乗を用いているのでその結果，分類境界面が線形に分離し，非線形な境界線を持つようなデータ分布を扱うのが困難になるからである．そこで本論文では，カーネル関数を用いた，新たな階層的クラスタリングアルゴリズムを提案する．カーネル関数とは，カーネル関数と1対1に対応する写像により高次元の特徴空間に写像された2つのベクトルの内積を与えるものであり，データ分類の分野で有効な道具の1つである．その写像は陽に現れず，そのため，高次元空間での内積の計算を元の空間の次元で行えるという利点を持つ．これによって非線形な分類境界面を持つようなデータにも対応出来るようになる．また，いくつかの数値例を通して，本論文で提案したクラスタリングアルゴリズムの有効性を検討する．

キーワード：階層的クラスタリング，カーネル関数，パターン認識

1. まえがき

データ分類は，パターン認識・データ処理等における基礎的な技法の一つとして重要である．データ分類には，あらかじめ分類結果が分かっている教師データに基づいて学習を行い，分類・識別性能を上げていくものと，教師データなど外的基準なしで，分類すべき個体の間に定義された類似度や距離などに基づいて，分類を行う手法の二つに大別することが出来る．前者は教師付き分類，後者は教師なし分類またはクラスタリングと呼ばれる．

ここで，類似度とは，対象となる2つの個体(データ)が存在したとき，その対象間に付与される実数値であり，その値が大きければ大きい程，2つの個体はお互いに類似しているとする．反対に，非類似度とは，その値が小さければ小さい程，2つの個体は類似しているとする．一般に，ユークリッド距離の自乗を非類似度として分類が行なわれることが多い．

クラスタリングには大きく分けて，非類似度の小さい個体の組を1組ずつ結合して，小さなクラスタから大きなクラスタへ逐次結合していく階層的クラスタリングと，あらかじめクラスタ数を設定し，非類似度を評価基準としてクラスタを生成する方法の非階層的クラスタリングがある．さらに非階層的クラスタリングには，帰属度に曖昧さの概念を導入した，ファジィ理論を用いるファジィクラスタリングと，帰属度が0か1のどちらかに所属するクリスプクラスタリングがある．

クラスタリングアルゴリズムについては，様々な手法が研究・開発されている．代表的なアルゴリズムでは，Zadeh[1]が創始したファジィ理論を用いたクラスタリング技法や，Kohonen[2]による自己組織化マップに関連するクラスタリング技法，数理統計における最尤推定に基礎をおくEMアルゴリズム[3, 4]などがある．

これらを含め，従来のクラスタリング手法では，一般に非線形な境界を持つデータに対して適切な分類を行うことが難しい．これは通常，各個体間の非類似度として，ユークリッド空間におけるノルムの自乗を用いているので，その結果，分類境界面が線形に分離し，非線形な境界線を持つようなデータ分布を扱うのが困難になるからである．そこで最近，SVM (Support Vector Machine) [5]と呼ばれる，カーネル関数を利用して非線形な境界を求めることが出来る教師付き分類の一手法が，多くの研究者によって注目されている．

そこで，本論文では，SVMに利用されているカーネル関数に着目し，さらにデータ分布にとらわれずにクラスタリングを行うために，階層的クラスタリングにカーネル関数を導入した新たなアルゴリズムを提案する．カーネル関数を導入することにより，データをパターン空間よりはるかに次元の高い空間で扱うことが

† On Some Hierarchical Clustering Algorithms Using Kernel Functions
Hideyuki HARUYAMA, Yasunori ENDO and Takayoshi OKUBO
*1 筑波大学大学院システム情報工学研究科
Graduate School of Systems and Information Engineering, University of Tsukuba
*2 筑波大学大学院システム情報工学研究科リスク工学専攻
Department of Risk Engineering, Faculty of Systems and Information Engineering, University of Tsukuba

出来るようになり，より良い分類結果を得ることが期待される．さらに提案したアルゴリズムの有効性を，数値例を用いて，従来のクラスタリング手法と比較することにより検証する．

2．階層的クラスタリング

階層的クラスタリングにおいて，クラスタ形成のための手続きはまず，各データ間に非類似度を定義し，この非類似度に基づいてクラスタを逐次結合していく，という流れである．階層的クラスタリングには，階層併合的方法(agglomerative hierarchical method)と階層分割的方法(divisive hiearchical method)があるが，本論文では階層併合的方法を用いる．この階層併合的方法の手続きをAHC(agglomerative hierarchical clustering)と呼ぶこととする．

アルゴリズム1 (AHC) データの集合を $X=\{x_1,\ldots,x_n\}$，非類似度を $d(x,y)\,(x,y\in X)$ とする．

Step.1 初期設定として，n 個のデータを n 個のクラスタとする．

$$G_i := \{x_i\},$$
$$d(G_i,G_j) := d(x_i,x_j) \quad (i,j=1,\ldots,n)$$

Step.2 クラスタ間の非類似度行列 $D=[d_{ij}]$ を更新する．

Step.3 D を参照し，最も類似性の高い2つのクラスタを結合して1つのクラスタを生成する．

$$s(G_q,G_r) = \max_{i,j} s(G_i,G_j) \text{ または}$$
$$d(G_q,G_r) = \min_{i,j} d(G_i,G_j)$$

Step.4 収束条件は，クラスタ数が1なら終了．そうでないなら**Step.2**へ戻る．結果は樹形図によって出力される．

階層的クラスタリングでは，D の与え方が分類結果に大きな影響を与える．以下に D の代表的な方法をいくつか挙げる．

1. 最短距離法 (nearest neighbor method: NN)
2. 最長距離法 (furthest neighbor method: FN)
3. 群間平均法 (average linkage between the merged grourp: AHC-B)
4. 群内平均法 (average linkage within the merged group: AHC-I)
5. メディアン法 (median method)
6. 重心法 (centroid method: AHC-C)
7. Ward法 (Ward's method: AHC-W)

3．カーネル関数

近年，サポートベクトルマシン(SVM)[5,9]という新たなパターン分類技法が注目されている．この技法は非線形な教師付き分類を行なう方法で，非線形写像を用いて高次元空間へ写像しながら，その写像は陽には表れないため，それらの間の内積を元の空間の次元で行えるという特徴を持つ．この内積関数をカーネル関数と呼び，SVM以外にも，回帰分析やクラスタリング等，様々な手法に適用されている．カーネル関数の定義を以下に述べる．

入力空間であるパターン空間 R^p から無限次元空間を含む特徴空間 R^s への写像 $\phi: R_p\to R^s$ を考える．ただし，一般に $p\ll s$ とする．さらに，$x=(x^1,\ldots,x^p)\in R^p$ に対して，$\phi^i: R^p\to R\,(i=1,\ldots,s)$ を用いて，非線形関数を $\phi(x)=(\phi^1(x),\phi^2(x),\ldots\phi^s(x))$ と定義する．このとき，

ある関数 $K: R^s\times R^s\to R$ について，

$$K(x,y) = \langle\phi(x),\phi(y)\rangle \tag{1}$$

を満たすような ϕ が存在するとき，K をカーネル関数と呼ぶ．

カーネル関数を利用する最大の特徴は，ϕ は K に対し陽に現れないので，ϕ によって定義される関数 $K(x,y)$ を求める場合，内積形 $\langle\phi(x),\phi(y)\rangle$ さえ定義されていれば $\phi(x)$ を計算する必要がないということである．すなわち，$\phi(x)$ がどんなに大きくなろうとも，そのことは K の型に直接影響を及ぼさないので，低次元空間における計算で高次元空間のデータを扱うことが可能になる．言いかえると，写像 ϕ の具体的な型を知らなくても，カーネル関数を扱うことで高次元空間での分類が可能になる．このような，式(1)を満たす ϕ が存在するカーネル関数 K の具体例として，以下のようなカーネルが存在する．

$$K_P(x,y) = (1+\langle x,y\rangle)^p \tag{2}$$
$$K_G(x,y) = \exp(-\frac{\|x-y\|^2}{2\delta^2}) \tag{3}$$
$$K_T(x,y) = \tanh(a\langle x,y\rangle-\delta) \tag{4}$$
$$K_S(x,y) = \frac{1}{1+\exp(-\beta\langle x,y\rangle)} \tag{5}$$

式(2)は多項式型カーネル，式(3)はガウス型カーネルと呼ばれ，式(4)はパーセプトロンカーネル[6]，式(5)はシグモイドカーネルと呼ばれる．

カーネル関数には様々な形が数多く知られており[7,8]関数 $K(x,y)$ が半正定値であるとき，式(1)が成立することが知られている[9]．

4. カーネル関数を用いた階層的クラスタリング

ここでは，階層併合的手法にカーネル関数を適用した，カーネル関数を用いた階層的クラスタリングアルゴリズム K-AHC(Kernel-aggolomerative hierarchical clustering)を提案する．従来の階層的クラスタリングの方法では，非類似度としてパターン空間上の距離を用いているが，本論文で提案する手法では，パターン空間上の点(データ)をカーネル関数を用いて特徴空間に写像し，非類似度として特徴空間における距離の自乗を用いる．なお，データの集合を X とし，$x \in X$ と $y \in X$ の距離$\|x-y\|$が定義されているとする．

4.1 K-AHC クラスタリングアルゴリズム

パターン空間 R^pから特徴空間 R^sへの写像 ϕ：$R^p \to R^s$を用いて，カーネル関数 $K(x,y)=\langle\phi(x),\phi(y)\rangle$ が定義され，パターン空間 R^pにおけるデータ集合 $X=\{x^1,\cdots,x^p\}\in R^p$が与えられたとする．

なお，K-AHC における非類似度 $d_\phi(x,y)\equiv d(\phi(x),\phi(y))$は，特徴空間における距離の自乗$\|\phi(x)-\phi(y)\|^2$ $(x,y\in X)$とする．特徴空間内での x_iと x_jとの距離の自乗は，カーネル関数を用いることにより，ϕ の形が分からなくても計算する事が出来る．すなわち，

$$\begin{aligned} d_\phi(x,y) &= \|\phi(x)-\phi(y)\|^2 \\ &= \|\phi(x)\|^2 - 2\langle\phi(x),\phi(y)\rangle + \|\phi(y)\|^2 \\ &= K(x,x)-2K(x,y)+K(y,y) \end{aligned} \tag{6}$$

で与えられる．以下に，非類似度行列のオプションを示す．

1. Kernel-最短距離法(K-NN)

$$\begin{aligned} d(G_i,G_j) &= \min_{x\in G_i, y\in G_j} d_\phi(x,y) \\ &= \min_{x\in G_i, y\in G_j} \|\phi(x)-\phi(y)\|^2 \end{aligned} \tag{7}$$

2. Kernel-最長距離法(K-FN)

$$\begin{aligned} d(G_i,G_j) &= \max_{x\in G_i, y\in G_j} d_\phi(x,y) \\ &= \max_{x\in G_i, y\in G_j} \|\phi(x)-\phi(y)\|^2 \end{aligned} \tag{8}$$

3. Kernel-群間平均法(K-B)

$$\begin{aligned} d(G_i,G_j) &= \frac{1}{|G_i||G_j|}\sum_{x\in G_i y\in G_j} d_\phi(x,y) \\ &= \frac{1}{|G_i||G_j|}\sum_{x\in G_i y\in G_j} \|\phi(x)-\phi(y)\|^2 \end{aligned} \tag{9}$$

4. Kernel-群内平均法(K-I)

$$\begin{aligned} d(G_i,G_j) &= \frac{1}{{}_{|G_i\cup G_j|}C_2}\sum_{\substack{x,y\in G_i\cup G_j \\ x\neq y}} d_\phi(x,y) \\ &= \frac{1}{{}_{|G_i\cup G_j|}C_2}\sum_{\substack{x,y\in G_i\cup G_j \\ x\neq y}} \|\phi(x)-\phi(y)\|^2 \end{aligned} \tag{10}$$

ここで，

$${}_{|G_i\cup G_j|}C_2 = |G_i\cup G_j|(|G_i\cup G_j|-1)/2$$

5. Kernel-重心法(K-C)

まず，特徴空間内での x と y との距離の自乗は

$$\begin{aligned} &\|\phi(x)-\phi(y)\|^2 \\ &\quad = K(x,x)-2K(x,y)+K(y,y) \end{aligned} \tag{11}$$

で与えられる．さらに，特徴空間におけるクラスタ C の重心 $M_\phi(G)$は，

$$M_\phi(G) = \frac{1}{|G|}\sum_{x\in G}\phi(x) \tag{12}$$

で与えられる．式(12)は以下のように展開することが出来る．

$$\begin{aligned} \|\frac{1}{|G_i|}\sum_{x\in G_i}\phi(x)\|^2 &= \frac{1}{|G_i|^2}\|\sum_{x\in G_i}\phi(x)\|^2 \\ &= \frac{1}{|G_i|^2}\sum_{x\in G_i}\sum_{y\in G_i}\langle\phi(x),\phi(y)\rangle \\ &= \frac{1}{|G_i|^2}\sum_{x\in G_i}\sum_{y\in G_i}K(x,y) \end{aligned} \tag{13}$$

従来の重心法では，クラスタ G_iと G_jの間の非類似度はクラスタ重心間の距離の自乗で与えられるが，ここでは，特徴空間上でクラスタリングを行うため，特徴空間上における，クラスタ G_i と G_jの間の非類似度 $d(G_i,G_j)$を，

$$d(G_i,G_j) = \|M_\phi(G_i)-M_\phi(G_j)\|^2 \tag{14}$$

と定義する．すると，式(14)は以下のように変形することが出来る．

$$\begin{aligned} d(G_i,G_j) &= \|M_\phi(G_i)-M_\phi(G_j)\|^2 \\ &= \|M_\phi(G_i)\|^2 - 2\langle M_\phi(G_i),M_\phi(G_j)\rangle \\ &\quad +\|M_\phi(G_j)\|^2 \\ &= \frac{1}{|G_i|^2}\sum_{x\in G_i}\sum_{y\in G_i}K(x,y) \\ &\quad -2\langle M_\phi(G_i),M_\phi(G_j)\rangle \\ &\quad +\frac{1}{|G_j|^2}\sum_{x\in G_j}\sum_{y\in G_j}K(x,y) \end{aligned}$$

$$= \frac{1}{|G_i|^2}\sum_{x\in G_i}\sum_{y\in G_i}K(x,y) - \frac{2}{|G_i||G_j|}\sum_{x\in G_i}\sum_{y\in G_j}K(x,y) + \frac{1}{|G_j|^2}\sum_{x\in G_j}\sum_{y\in G_j}K(x,y) \quad (15)$$

6. Kernel-Ward 法(K-W)
この方法では，クラスタ内のデータと重心との距離平方和を損失と考え，結合の際に損失が最小とするようなクラスタを選択・結合する．そこで，平方距離和を

$$E_\phi(G) = \sum_{x\in G}\|\phi(x) - M_\phi(G)\|^2 \quad (16)$$

$$= \sum_{x\in G}K(x,x) - \frac{1}{|G|}\sum_{x\in G}\sum_{y\in G}K(x,y) \quad (17)$$

で定義することが出来る．そして，クラスタ G_i と G_j の間の非類似度 $d(G_i, G_j)$ を

$$d(G_i,G_j) = E_\phi(G_i\cup G_j) - E_\phi(G_i) - E_\phi(G_j) \quad (18)$$

と定義する．

4.2 K-AHC における非類似度の再計算と計算量

上記に述べた，非類似度の式は原理であり，実際に非類似度の更新は，新たに作られたクラスタについてのみ行われる．その際に用いられる更新式は，カーネルを用いない従来のアルゴリズムと同じものを使用することが出来る．階層的クラスタリングは，比較的良い分類結果が得られることが知られているが，計算量は最大 $O(N^3)$ を必要とする．一般に，カーネル関数を導入した場合，アルゴリズムの計算量は増大するという欠点があるが[10, 11]，本論文で提案した方法に関しては，計算量の増大はなく，従来と同様の計算量で計算結果を得ることが出来るのが特徴の1つである．各手法の更新式を以下に示す．ここで，$G_{qr} := G_q \cup G_r$ である．

4.2.1 Kernel-重心法(K-C)

$\alpha = |G_q|/(|G_r|+|G_q|)$ とすると，式(12)より，

$$d(G_i, G_{qr}) = \|M_\phi(G_i) - M_\phi(G_{qr})\|^2 = \alpha d(G_i,G_q) + (1-\alpha)d(G_i,G_r) - \alpha(1-\alpha)d(G_q,G_r) \quad (19)$$

4.2.2 Kernel-Ward 法(K-W)}

特徴空間上におけるクラスタ G 内のデータと重心 $M(G)$ の間の距離平方和は式(16)より，

$$E_\phi(G) = \sum_{x\in G}\|\phi(x) - M_\phi(G)\|^2 = \sum_{x\in G}\|\phi(x)\|^2 - |G|\|M_\phi(G)\|^2$$

で与えられる．よって，

$$\begin{aligned}
d(G_i,G_j) &= E_\phi(G_i\cup G_j) - E_\phi(G_i) - E_\phi(G_j)\\
&= |G_i|\|M_\phi(G_i)\|^2 + |G_j|\|M_\phi(G_j)\|^2 - |G_i\cup G_j|\|M_\phi(G_i\cup G_j)\|^2\\
&= |G_i|\|M_\phi(G_i)\|^2 + |G_j|\|M_\phi(G_j)\|^2 - |G_i\cup G_j|(\alpha^2\|M_\phi(G_i)\|^2 + (1-\alpha)^2\\
&\quad \cdot\|M_\phi(G_j)\|^2 + \alpha(1-\alpha)(\|M_\phi(G_i)\|^2 + \|M_\phi(G_j)\|^2 - \|M_\phi(G_i) - M_\phi(G_j)\|^2))\\
&= |G_i|\|M_\phi(G_i)\|^2 + |G_j|\|M_\phi(G_j)\|^2\\
&\quad -(|G_i|+|G_j|)(\frac{|G_i|}{|G_i|+|G_j|})^2\|M_\phi(G_i)\|^2\\
&\quad -(|G_i|+|G_j|)(\frac{|G_j|}{|G_i|+|G_j|})^2\|M_\phi(G_j)\|^2\\
&\quad -(|G_i|+|G_j|)\frac{|G_i||G_j|}{(|G_i|+|G_j|)^2}\|M_\phi(G_i)\|^2\\
&\quad -(|G_i|+|G_j|)\cdot\frac{|G_i||G_j|}{|G_i|+|G_j|}\|M_\phi(G_j)\|^2\\
&\quad +(|G_i|+|G_j|)\frac{|G_i||G_j|}{(|G_i|+|G_j|)^2}(\|M_\phi(G_i) - M_\phi(G_j)\|^2)\\
&= \frac{|G_i||G_j|}{|G_i|+|G_j|}\|M_\phi(G_i) - M_\phi(G_j)\|^2 \qquad (20)
\end{aligned}$$

と与えられ，式(19)と式(20)より，

$$\begin{aligned}
&d(G_i, G_{qr})\\
&= \frac{|G_i||G_{qr}|}{|G_i|+|G_{qr}|}\|M_\phi(G_i) - M_\phi(G_{qr})\|^2\\
&= \frac{|G_{qr}||G_i|}{|G_{qr}+|G_i|}\left[\frac{|G_q|}{|G_{qr}|}\|M_\phi(G_q) - M_\phi(G_i)\|^2 + \frac{|G_r|}{|G_{qr}|}\|M_\phi(G_q) - M_\phi(G_i)\|^2 - \frac{|G_q||G_r|}{|G_{qr}|^2}\|M_\phi(G_q) - M_\phi(G_r)\|^2\right]\\
&= \frac{(|G_q|+|G_r|)|G_i|}{|G_q|+|G_r|+|G_i|}\left[\frac{|G_q|+|G_i|}{(|G_q|+|G_r|)|G_i|}\cdot d(G_i,G_q) + \frac{|G_r|+|G_i|}{(|G_q|+|G_r|)|G_i|}d(G_i,G_r) - \frac{1}{|G_q|+|G_r|}d(G_r,G_q)\right]\\
&= \frac{|G_i|+|G_q|}{|G_i|+|G_q|+|G_r|}d(G_i,G_q)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{|G_i|+|G_r|}{|G_i|+|G_q|+|G_r|}d(G_i,G_r)\\
&-\frac{|G_i|}{|G_i|+|G_q|+|G_r|}d(G_q,G_r) \qquad (21)
\end{aligned}
$$

となる．

4.2.3 Kernel-最短距離法(K-NN)

$$
\begin{aligned}
d(G_i,G_{qr}) &= \min_{x\in G_i, y\in G_{qr}} d_\phi(x,y)\\
&= \min\{d(G_i,G_q), d(G_i,G_r)\} \qquad (22)
\end{aligned}
$$

4.2.4 Kernel-最長距離法(K-FN)

$$
\begin{aligned}
d(G_i,G_{qr}) &= \max_{x\in G_i, y\in G_{qr}} d_\phi(x,y)\\
&= \max\{d(G_i,G_q), d(G_i,G_r)\} \qquad (23)
\end{aligned}
$$

4.2.5 Kernel-群内平均法(K-B)

$$
d(G_i,G_{qr}) = \frac{|G_q|d(G_i,G_q)+|G_r|d(G_i,G_r)}{|G_q|+|G_r|} \qquad (24)
$$

4.2.6 Kernel-群内平均法(K-I)

$$
d(G_i,G_{qr}) = \frac{I(G_i)+I(G_{qr})+d'(G_i,G_{qr})}{{}_{|G_i\cup G_{qr}|}C_2} \qquad (25)
$$

ここで，$I(G_i)$, $d'(G_i, G_{qr})$の2つを以下のように定義する．

1. 初期クラスタ $G_i=\{x_i\}$に対して，
   $I(G_i)=0$, $d'(G_i, G_{qr})=d_\phi(x_i, x_{qr})$とする．
2. $I(G_{qr})=I(G_q)+I(G_r)+d'(G_q, G_r)$
   $d'(G_i, G_q)=d'(G_i, G_q)+d'(G_i, G_r)$

5. 数値例によるクラスタリングアルゴリズムの分類結果

ここでは，幾つかの人工データ及びIrisデータを用いて，従来のクラスタリングアルゴリズムと本論文で提案したK-AHCでクラスタリングを行なった．

5.1 サンプルデータによるクラスタリング

今回使用したデータは，データ1～5の5通りで，各データの特徴は次の通りである．

- データ1(個体数300)…各々のクラスタが入れ子状になっている．
- データ2(個体数270)…1つのクラスタが他方のクラスタを包含する形状となっている．
- データ3(個体数312)…2つのクラスタの個体数が大きく異なる．
- データ4(個体数306)…1つのクラスタは線形，他方のクラスタは球状である．
- データ5(個体数241)…3つのクラスタからなり，一般には分類不能である．

これらのデータは，すべてのクラスタリングアルゴリズムに共通する，いわゆる「クラスタリングしにくい特徴」を人工的に構成したものとなっている．この

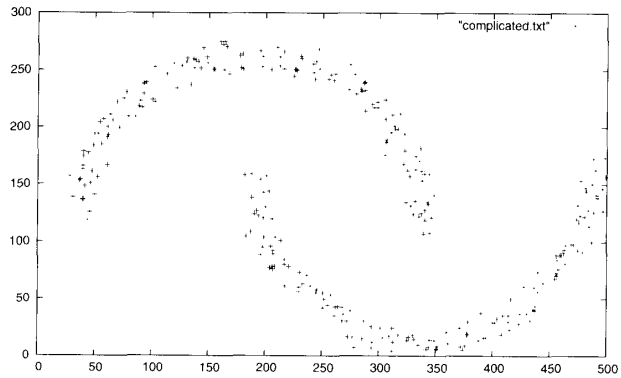


図1 データ1(個体数300)

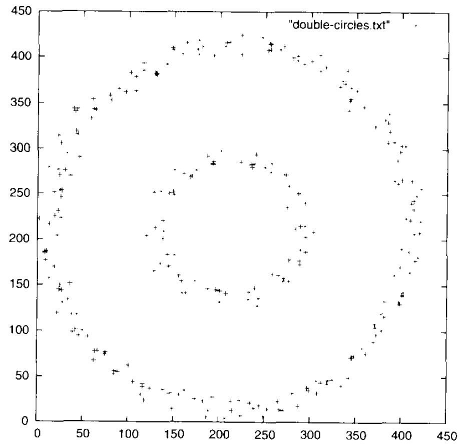


図2 データ2(個体数270)

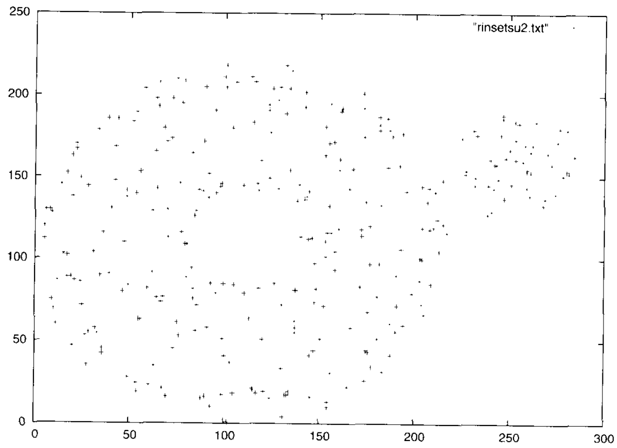


図3 データ3(個体数312)

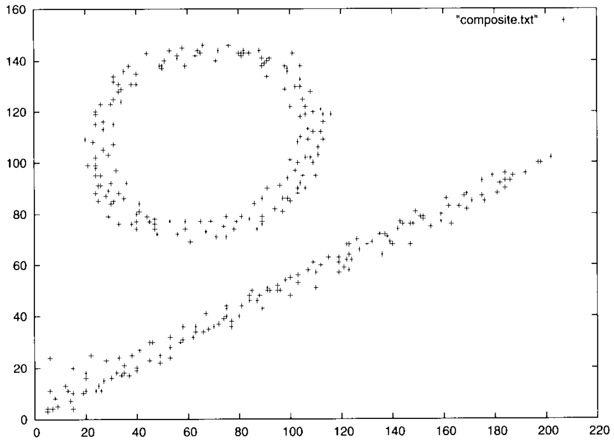


図 4　データ 4（個体数306）

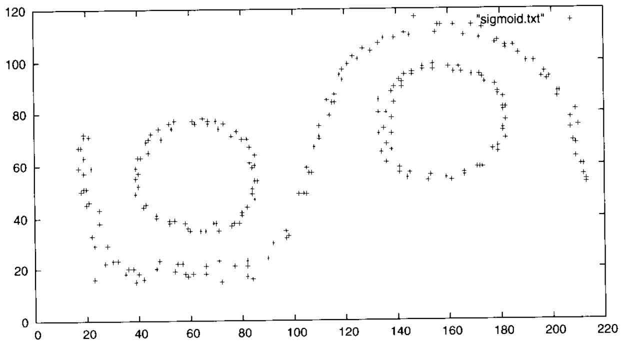


図 5　データ 5（個体数241）

5 つのデータに対して，K-AHC を用いてクラスタリングを行った．また，比較のために，従来の階層的クラスタリングの各手法，さらに，非階層的クラスタリングからは代表的な手法であるファジィ c-平均法(FCM)についても同様にクラスタリングを行った．クラスタ数はデータ 5 が 3，それ以外のデータに関しては 2 とした．

なお，今回使用したカーネル関数は，多項式型カーネル，ガウス型カーネル，パーセプトロン型カーネル，シグモイド型カーネルを用いた．

5.2　考察

表 1 は各アルゴリズムのクラスタリング結果であり，表 2 はカーネル関数を用いた階層的クラスタリング　によるクラスタリング結果である．表記しているアルファベットは，対応するカーネル関数を用いることによって良好な分類結果が得られることを示している．

従来の手法では，分類が困難なデータに対して，カーネル関数を用いた階層的クラスタリングでは，適切に分類されていることが分かる．さらに，表 2 を見ても分かるように，手法の違いによって分類結果に差が

表 1　従来のアルゴリズムによるクラスタリング結果

	FCM	NN	FN	AHC-I	AHC-B	AHC-C	AHC-W
データ 1	×	○	×	×	×	×	×
データ 2	×	○	×	×	×	×	×
データ 3	×	×	×	×	×	×	×
データ 4	×	○	×	×	×	×	×
データ 5	×	×	×	×	×	×	×

表 2　K-AHC における各クラスタリング結果

	K-NN	K-FN	K-B	K-I	K-C	K-W
データ 1	P, G, T, S	×	G	×	×	G
データ 2	P, G, T, S	×	G	×	×	G
データ 3	P, T	×	P, T	P, T	P, G, T	P, G
データ 4	S	S	G, S	×	G	S
データ 5	P, G, T, S	×	G	×	×	G

P：多項式型カーネル
G：ガウス型カーネル
T：パーセプトロン型カーネル
S：シグモイド型カーネル

生じていることが分かる．特に，K-NN，K-B，K-W の 3 つの手法については，どのデータに対しても比較的良好な分類結果が得られた．カーネル法を用いない階層的クラスタリングにおいても，NN, AHC-B, AHC-W の各手法が他の手法より性能が良いことを考え合わせると，妥当な結果と思われる．

次に，各手法に対するカーネルとの相性について考察する．K-NN のようにデータ 3，データ 4 を除いて，どのカーネルを用いても分類結果が良いものや，K-B，K-W のように，ある特定のカーネルを用いることによって良い分類結果が得られるものもある．これは，K-NN がデータ間の類似度のみで行うのに対し，K-B や K-W では，クラスタ間やクラスタ内の類似度の総和も考慮するからであると考えられる．ガウス型以外のカーネルを用いる際，$x \to x+\delta$，$y \to y+\delta$ とすると，ガ

ウス型では計算にノルムを用いているために $K_G(x+\delta, y+\delta)=K_G(x,y)$ が得られるが，他のカーネルでは計算に内積を用いているため，この関係を満たさず，原点の取り方によってクラスタリング結果が変化することがあり，このことは他の数値例からも確認している．これは，例えば多項式カーネルの場合，一般に，$\|\delta_1\|<\|\delta_2\|$ ならば $\|\phi(x+\delta_1)-\phi(y+\delta_1)\|<\|\phi(x+\delta_2)-\phi(y+\delta_2)\|$ となり，パターン空間における原点の移動が特徴空間におけるある種の引き延ばしの作用をもたらす結果であると考えられ，多項式型やシグモイド型等，原点からの距離に鋭敏に反応するカーネルを用いて，クラスタ間やクラスタ内の類似度の総和を計算するような場合，空間の伸長が計算に蓄積され，データ間の類似度を計算するだけの場合よりも計算結果に大きな影響を与えるようになると考えられるためである．この問題に関する詳細な計算は，ϕ が陽に表れないために非常に困難であり，今後の課題の1つと考えられる．

さらに，カーネル別に見てみると，特定なデータに対して有効なカーネルがある．例えば，多項式型カーネルではデータ3に対して良好な結果を得ることができ，シグモイド型カーネルでは，データ4に対して有効に動作することが分かる．また，ガウス型に関しては全般的に良い分類結果が得られている．特に，K-B，K-Wについてはそれぞれ，データ3，データ4以外は誤分類が無い結果が得られている．パーセプトロン型カーネルについては，他のカーネルを導入した時と比べて分類結果が若干悪い．これは，パラメータが複数存在するので，パラメータの設定がより困難になるためである．以上のように，各手法・各カーネルの両方面からクラスタリング結果に影響があることが数値例から分かる．

5.3 Irisデータ

ここでは，実際のデータとして，クラスタリングにおいて有名なIrisデータを使用し，クラスタリングを行なった．Irisデータとは，Iris-setosa, Iris-versicolor, Iris-virginicaと呼ばれる3つのアヤメ科の品種それぞれ50個体に関して，ガクの長さ，ガクの幅，花弁の長さ，花弁の幅の4通りのデータをまとめたものである．このデータは，データ分類の例としてよく用いられている．

5.4 考察(Irisデータ)

Irisデータのクラスタリング結果を表3～7に示した．本実験では，カーネル関数を用いた階層的クラスタリングとして，性能がよいと考えられるK-Wを用いた．表を見て分かるように，どのデータの組み合わせに対しても従来のクラスタリング手法より，カーネル関数を用いた階層的クラスタリング(K-W)の方が誤分類数

表3 Irisデータによるクラスタリング結果(花弁の長さ-花弁の幅)

	FCM	NN	AHC-W	K-W
Iris-setosa	50	50	25	50
Iris-versicolor	44	99	50	46
Iris-virginica	56	1	75	54
誤分類数	6	49	50	4

表4 Irisデータによるクラスタリング結果(ガクの長さ-ガクの幅)

	FCM	NN	AHC-W	K-W
Iris-setosa	45	49	49	48
Iris-versicolor	52	1	31	76
Iris-virginica	53	100	70	26
誤分類数	31	51	33	26

表5 Irisデータによるクラスタリング結果(ガクの長さ-花弁の長さ)

	FCM	NN	AHC-W	K-W
Iris-setosa	51	50	50	50
Iris-versicolor	41	99	88	68
Iris-virginica	58	1	12	32
誤分類数	78	51	38	18

表6 Irisデータによるクラスタリング結果(ガクの長さ-花弁の幅)

	FCM	NN	AHC-W	K-W
Iris-setosa	56	48	50	50
Iris-versicolor	40	1	74	63
Iris-virginica	54	1	26	37
誤分類数	26	99	24	13

表7 Irisデータによるクラスタリング結果(ガクの幅-花弁の幅)

	FCM	NN	AHC-W	K-W
Iris-setosa	49	49	50	50
Iris-versicolor	53	1	74	60
Iris-virginica	48	100	26	40
誤分類数	11	51	24	10

が少なくなり，分類結果が良いことが分かる．これは，AHC-Wなどの従来の手法で，パターン空間でクラス間同士の分布状態が曖昧なものは，クラスタリングを行っても分類が困難なことが明らかであるのに対し，K-Wでは，カーネルを用いて無限次元を含む高次元特徴空間にデータを写像し，特徴空間でのクラス間同士の分布が線形に分離しやすい状態になっているからであると思われる．ただし，ϕ が陽に表れないため，高次元空間でどのようにデータが分類しているかを把握することは容易ではない．これらの分布状況の把握と解析は今後の課題の一つである．

6. おわりに

本論文では，階層的クラスタリングにカーネル関数を導入することによって，高次元特徴空間への写像を利用した新たなクラスタリングアルゴリズムを提案した．これにより，分類困難であるような非線形な分類境界を持つデータに対し，適切な分類が行えることを示し，その分類性能を検証した．本提案手法の計算量は，従来の階層的クラスタリングと同等であり，カーネル関数を導入することによる計算量の増大はない．また従来のアルゴリズムがほぼそのまま使えるため，ハード・ソフト的なリソースがそのまま利用可能であり，既存の資産でクラスタリング結果の大きな改善が可能である．

今後の課題としては，カーネルやパラメータによって，どのような特性(データ分布)の非線形境界を，分離出来るのかを明らかにすることについて検討していく必要がある．そのためには，特徴空間でのデータの分布状態を把握する手法の開発が必要となる．

次に，カーネル関数の生成・選択及び，パラメータ自動設定の決定等について検討していきたい．$\phi(x)$ の振る舞いを議論出来れば，データが与えられた際に，そのデータに対して有効なカーネル関数をあらかじめ選択してクラスタリングが行えることが出来る．さらに，データが入力された時に既存のカーネル関数だけを使用するのではなく，データに対して，新たなカーネル関数を生成することも必要であると考えられる．

謝辞

本研究を進めるにあたって有益なご助言を頂きました，筑波大学大学院宮本定明教授に感謝致します．なお本研究の一部は，日本学術振興会科学研究費補助金基盤研究(B)課題番号16300065「ソフトコンピューティングによるクラスター分析に関する国際拠点形成(ミニCOE)」の補助を受けた．

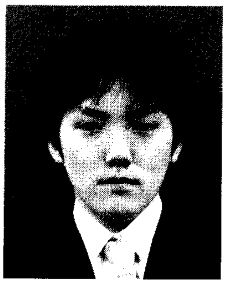
参考文献

[1] L.A. Zadeh : "Fuzzy sets, Information and Control", Vol.8, pp338-353 (1965).
[2] T. Kohonen : 'Self-Organizing Maps, 2nd Ed.', Springer, Berlin (1997).
[3] A. P. Dempster, L. M. Laird, D. B. Rubin : "Maximum likelihood from incomplete data via the EM algorithm", J. of the royal Statistical Society, B., Vol.39, pp.1-38 (1997).
[4] R. A. Redner, H. F. Walker : "Mixture densities, maximum likelihood and the EM algorithm", SIAM Review, Vol.26, No.2, pp.195-239 (1984).
[5] 前田英作："痛快！サポートベクトルマシン", 情報処理, Vol.42, No.7, pp.676-683 (2001).
[6] C. Burges : "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, Vol.2, pp.1-47 (1998).
[7] 津田宏治："カーネル設計の技術", 第5回情報論理学習論ワークショップ(IBIS2002), pp.1-10 (2002).
[8] 津田宏治："カーネル設計の方法", 日本神経回路学会誌, Vol.9, No.3, pp.190-195 (2002).
[9] V. Vapnik : 'Statistical Learning Theory', John Wiley & Sons, Inc (1998).
[10] 中山洋一, 宮本定明："高次元空間への写像を利用したクリスプ c-平均法", 第18回ファジィシステムシンポジウム講演論文集, pp.119-122 (2002).
[11] 水津大輔, 宮本定明："高次元空間への写像を利用したファジィ c-平均法", 第18回ファジィシステムシンポジウム講演論文集, pp.123-126 (2002).
[12] 宮本定明：'クラスター分析入門', 森北出版 (1999).
[13] Y. Endo, H. Haruyama, T. Okubo : "On Some Hierarchical Clustering Algorithms Using Kernel Functions", IEEE International Conference on Fuzzy Systems (2004).
[14] 春山秀幸, 遠藤靖典："カーネル関数を用いた階層的クラスタリングについて", 第19回ファジィシステムシンポジウム講演論文集, pp.909-912 (2003).
[15] 春山秀幸, 遠藤靖典："カーネル関数を用いた階層的クラスタリング", 電気情報通信学会総合大会, D-12-19 (2003).

(2004年10月4日　受付)
(2005年3月12日　採録)

[問い合わせ先]
〒305-8573 茨城県つくば市天王台1-1-1
筑波大学大学院システム情報工学研究科 遠藤研究室
春山 秀幸
TEL：029-853-5250
FAX：029-853-5135
E-mail：haruyama@esys.tsukuba.ac.jp

著者紹介



春山 秀幸（はるやま ひでゆき）［学生会員］

平成13年 東海大学工学部通信工学科卒，平成14年 筑波大学大学院システム情報工学研究科リスク工学専攻入学，現在に至る．クラスタリングに関する研究に従事．日本知能情報ファジィ学会学生会員．

大久保 貴義（おおくぼ たかよし）［非会員］

平成14年 東海大学工学部通信工学科卒，平成15年 筑波大学大学院システム情報工学研究科入学，平成17年3月 同大大学院修士(工学)取得後退学，同年4月 NTTデータフォース株式会社入社，現在に至る．クラスタリングに関する研究に従事．

遠藤 靖典（えんどう やすのり）［正会員］

平成2年 早稲田大学理工学部電子通信学科卒，平成7年 同大大学院理工学研究科博士後期課程了．博士(工学)．平成6年 早稲田大学理工学部助手，平成9年 東海大学工学部講師，平成13年 筑波大学機能工学系講師，平成16年 筑波大学大学院システム情報工学研究科助教授となり，現在に至る．クラスタリングに関する研究，鉄道車両のファジィABS制御に関する研究，ファジィ理論を用いたシステムの変動解析に関する研究に従事．電子情報通信学会平成5年度論文賞，同学会同年度米澤ファウンダーズ・メダル受賞記念特別賞，同学会平成6年度学術奨励賞，日本ファジィ学会平成8年度奨励賞各受賞．日本知能情報ファジィ学会，電子情報通信学会各会員．

On Some Hierarchical Clustering Algorithms Using Kernel Functions

by

Hideyuki HARUYAMA, Yasunori ENDO and Takayoshi OKUBO

Abstract :

By the former clustering algorithms, it is difficult to obtain a good clustering results for the data which is not able to be separated on linear classification boundary. The square of norms between data is used as dissimilarity so that those algorithms regard the classification boundary as linear. In this paper, the new hierarchical algorithm is proposed using kernel functions. The kernel functions are useful into the field of classification and give the value of the inner product of two vectors in a high-dimensional feature space by a mapping which one-to-one corresponds to the kernel function. The mapping is not explicit so that the value of the inner product of the feature space can be easily calculated on dimension of the original pattern space. By introducing the kernel functions, the proposed algorithm can separate the data on nonlinear classification boundary. Moreover, the availability of proposed algorithm is discussed through some numerical examples.

Keywords : hierarchical clustering, kernel function, pattern recognition

Contact Address : **Hideyuki HARUYAMA**

Graduate School of Systems and Information Engineering, University of Tsukuba

1-1-1, Tennoudai, Tsukuba, Ibaraki, 305-8573 Japan

TEL : 029-853-5250

FAX : 029-853-5135

E-mail : haruyama@esys.tsukuba.ac.jp