

集合間の距離に基づく重み付きカーネルk-平均法のパラメータ探索法

柳 信 一

北海道情報大学

A Search Algorithm for Gaussian Parameter of Weighted
Kernel k-means Method Based on the Distance between the Set

Shinichi YANAGI

Hokkaido Information University

平成28年 3 月

北海道情報大学紀要 第27巻 第 2 号別刷

〈論文〉

集合間の距離に基づく重み付きカーネル k -平均法のパラメータ探索法

柳 信一*

A Search Algorithm for Gaussian Parameter of Weighted Kernel k -means Method Based on the Distance between the Set

YANAGI SHINICHI*

要旨

本論文では、集合間の距離に基づき重み付きカーネル k -平均法のガウス関数のパラメータを逐次的に計算する手法を提案する。重み付きカーネル k -平均法は、代表的なクラスタリング手法である k -平均法に基づいており、クラスタ同士の境界が非線形であるデータの分類が可能である。しかし、計算結果は距離計算に用いるガウス関数のパラメータに強く依存し、パラメータ調整が難しいという問題点がある。提案手法は、集合間の距離に基づき初期クラスタを決定し、その初期クラスタを用いてパラメータの探索を行う。いくつかの人工データと実データを用いて計算機実験を行い、クラスタ間が十分な距離で分離しているデータに対する提案手法の有効性を示す。

Abstract

This paper proposes a weighted kernel k -means algorithm to compute a Gaussian parameter based on the distance between the set. The weighted kernel k -means algorithm is based on the k -means algorithm which is a well known clustering algorithm, and makes it possible to identify clusters that are non-linearly separable in input space. However, a conclusion identifying clusters with kernel k -means depends heavily on chosen Gaussian parameters which are used for computing distance between data and determining an optimal Gaussian parameter is difficult. The proposed algorithm identifies clusters with the original weighted kernel k -means algorithm after making a initial clusters based on the distance between the set and searching a Gaussian parameter. An experimental comparison in some artificial and real data set shows that a performance of the proposed algorithm is efficient for a data in which each cluster is separated with adequate gap.

キーワード

クラスタリング, k -平均法, カーネル法, ガウス関数, パラメータ

1. まえがき

クラスタリングとは、データ集合をある基準によって分類することである。 k -平均法 (MacQueen, 1967) は代表的なクラスタリング手法であり、アルゴリズムが単純で、事前に必要とするパラメータがクラスタ数のみであるという利点のため、クラスタリングを必要とする計

算過程において利用されることが多い。しかし、距離計算にユークリッド距離を使用するため、各クラスタの分離境界が非線形となるデータの分類に適していない。

カーネル k -平均法 (Girolami, 2002) は、 k -平均法に対してカーネル法 (Muller, Mika, Ratsch, Tsuda and Scholkopf, 2001) を適用した手法であり、非線形に分離されているクラスタの分類が可能である。しかし、計算結果が距離計算に使用するカーネル関数のパラメータに敏感であり、カーネル関数のパラメータの調

* 北海道情報大学経営情報学部システム情報学科講師,
Lecturer, Department of Business and Information Systems, HIU

整が困難であるという問題点がある。

一方で、与えられたデータ集合をデータ間の類似度を定義することでグラフとみなし、クラスタリングをグラフのカット問題に帰着するスペクトラルクラスタリング (Ng, Jordan and Weiss, 2002) がある。スペクトラルクラスタリングは計算過程で k -平均法を利用する手法であり、カーネル k -平均法同様、非線形に分離されているクラスタの分類が可能であるものの、類似度の設計が結果を大きく左右するため類似度のパラメータ依存度が高い。さらに、固有値問題を解く必要があるため、数値計算を必要とする。これに対して、Dhillon (Dhillon, Guan and Kulis, 2004) らは、カーネル k -平均法の距離計算に重み付き距離を適用した重み付きカーネル k -平均法 (Dhillon, Guan and Kulis, 2004) を提案している。重み付きカーネル k -平均法の最適化問題は数学的にスペクトラルクラスタリングと等価であり、数値計算を必要とせず距離計算のみで計算可能な手法として有効性を示している。

上記の手法で使用するカーネル関数、および類似度関数は、ほとんどの場合がガウス関数であり、パラメータはデータ間の距離に関するスケール調整の意味を持つ。適切なパラメータを選択する手法としては、Zelnik-Manor ら (Zelnik-Manor and Perona, 2004) が、スペクトラルクラスタリングにおいて各データの類似度に対して、同一のパラメータではなく、それぞれのデータの近傍距離を利用する手法を提案し有効性を示しているものの、近傍の度合いに関する根拠が十分に示されていない。また、Yu ら (Yu, Tranchevent, De Moor and Moreau, 2011) は、スペクトラルクラスタリングとパラメータの最適化を交互に行う手法を提案しているが、事前に解となるパラメータの有限列を用意する必要があり、依然としてパラメータ選択の困難さが残る。スペクトラルクラスタリング以外では、文献 (柳, 2014) において、Yu らの手法と同様、カーネル k -平均法とパラメータの更新を交互に行う手法を提案し、計算機実験によりいくつかのデータに対して有効性を示しているものの、クラスタ数やデータの分布形状によっては結果が不安定となる場合がある。

本論文では、文献 (柳, 2014) の手法において、まず集合間の距離を基に初期クラスタを作成し、初期クラスタを利用してカーネル関数のパ

ラメータを計算するように改良したアルゴリズムを提案する。いくつかの人工データと実データを用いて比較実験を行い、提案手法はクラスタ間が十分な距離で分離しているデータに対しては有効であることを示す。

以下、本論文は次のように構成されている。2章で従来の重み付きカーネル k -平均法を説明する。また、3章で重み付きカーネル k -平均法の従来のパラメータ決定に関する研究を述べる。4章では新たに重み付きカーネル k -平均法のパラメータ決定法を提案し、5章では従来手法と提案手法の比較実験を行い、提案手法の有効性を検討する。6章では本論文のまとめを行う。

2. 重み付きカーネル k -平均法

クラスタリングとは、対象とする n 個のデータからなるデータ集合 $X = \{x_i \mid i = 1, 2, \dots, n, \forall x_i \in R^d\}$ を何らかの基準により k 個の部分集合 $C_j = \{x \mid \forall x \in X\}$ に分類することであり、この部分集合 $C_j (j = 1, 2, \dots, k)$ をクラスタと呼ぶ。一般的にクラスタリングを必要とする場面では、データ集合のみが与えられ、クラスタ数 k は未知であることもあるが、本論文では k を既知のものとして取り扱う。

2-1 k -平均法

k -平均法は各データ $x_i \in X (i = 1, 2, \dots, n)$ と、クラスタの中心 $m_j \in R^d (j = 1, 2, \dots, k)$ との距離の総和を最小とする最適化問題として式 (1) で定式化される。

$$\min_{C_j} \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - m_j\|^2,$$

$$\text{s.t. } \bigcup_{j=1}^k C_j = X,$$

$$C_j \cap C_l = \emptyset, 1 \leq j, l \leq k, j \neq l. \quad (1)$$

式 (1) の $\|\cdot\|$ は L_2 ノルムである。式 (1) の解はクラスタ $C_j (j = 1, 2, \dots, k)$ であり、クラスタ中心 m_j は式 (1) より、 C_j に属するデータ数 $|C_j|$ を用いて式 (2) のように求まる。

$$m_j = \frac{1}{|C_j|} \sum_{x \in C_j} x, (j = 1, 2, \dots, k). \quad (2)$$

式 (1) の定式化は暗に各クラスタが凸形状となっていることを仮定しているため、線形分離不可能なクラスタからなるデータの分類に

適していない。

2-2 カーネル k-平均法

カーネル k-平均法 (Girolami, 2002) は k-平均法にカーネル法を適用した手法である。カーネル法は、無限次元まで含めた、現在の空間の次元数 d よりはるかに大きい $D (\gg d)$ 次元空間上に、写像 ϕ により $x \in R^d$ を非線形写像し、空間 R^D 上でデータの線形分離を試みる方法である。最大の特徴は、計算量削減のため実際の写像 $\phi(x)$ 自体の計算は行わず、式 (3) のカーネル関数を利用して $\phi(x)$ と $\phi(x')$ の内積のみを使用することで問題を解くことである。

$$K(x, x') = \phi(x) \cdot \phi(x'). \quad (3)$$

具体的なカーネル関数としては、カーネル関数が高次元空間における内積となっていることを保証するマースルカーネル (Máté, 1989) が望ましく、特に多くの研究では、代表的なマースルカーネルである式 (4) のガウスカーネルが使用されている。

$$K(x, x') = \exp\left(\frac{-\|x - x'\|^2}{\delta^2}\right). \quad (4)$$

式 (4) において、 δ はスケールパラメータを表す。カーネル k-平均法では、式 (1) の最適化問題は式 (5) のように表現される。

$$\begin{aligned} \min_{C_j} & \sum_{j=1}^k \sum_{x_i \in C_j} \|\phi(x_i) - M_j\|^2, \\ \text{s.t.} & \bigcup_{j=1}^k C_j = X, \\ & C_j \cap C_l = \emptyset, 1 \leq j, l \leq k, j \neq l. \end{aligned} \quad (5)$$

また、式 (2) のクラスタ中心は式 (6) のように表現される。

$$M_j = \frac{1}{|C_j|} \sum_{x \in C_j} \phi(x), M_j \in R^D (j = 1, 2, \dots, k). \quad (6)$$

写像 ϕ が未定義の場合、式 (6) のクラスタ中心を計算することはできないが、クラスタ中心 $M_j \in R^D$ とデータ $\phi(x) \in R^D$ の距離は式 (7) で計算することができる。

$$\begin{aligned} & \|M_j - \phi(x)\|^2 \\ &= \|M_j\|^2 - 2M_j \cdot \phi(x) + \phi(x) \cdot \phi(x) \\ &= \frac{1}{|C_j|^2} \sum_{x'', x' \in C_j} K(x'', x') - \frac{2}{|C_j|} \sum_{x' \in C_j} K(x, x') \\ & \quad + K(x, x), (j = 1, 2, \dots, k). \end{aligned} \quad (7)$$

2-3 スペクトラルクラスタリング

カーネル k-平均法は線形分離不可能なクラスタからなるデータの分類が可能であるものの、高次元空間上のデータ分布が球形状であることを仮定しているため分類能力には限界がある。一方で、スペクトラルクラスタリングはデータ $x \in X$ をグラフの節点とみなし、データ間の類似度を定義することで、クラスタリングをグラフのカット問題に帰着させる手法である。具体的には、目的関数を定義し、与えられたグラフの節点集合 V を目的関数が最小となるような部分グラフ V_1, V_2, \dots, V_k への分割を求める問題である。様々な目的関数が提案されており (Chan, Schlag and Zien, 1994; Shi and Malik, 2000; Yu and Shi, 2003), 式 (8) の Normalized Cut (Shi and Malik, 2000; Yu and Shi, 2003) が代表的である。

$$NCut(V, K) = \min_{V_1, V_2, \dots, V_k} \sum_{j=1}^k \frac{\text{links}(V_j, V - V_j)}{\text{links}(V_j, V)}. \quad (8)$$

ここで、 $K = [k_{ij}]$ は節点 v_i と v_j の類似度 k_{ij} を要素とする類似度行列である。類似度としては、式 (4) のガウス関数がよく使われる。そのため、パラメータ δ が分割に大きな影響を与える。類似度行列 K の要素をカーネル関数としたときはカーネル行列とも呼ばれる。また、 $\text{links}(A, B)$ は節点 $a \in A$ と節点 $b \in B$ の類似度の総和である。部分グラフの割り当ては離散最適化であるため、k-平均法同様 NP 困難であるが、Ncut は式 (9) の行列の固有値問題に緩和できる。

$$D^{-1/2}(D - K)D^{-1/2}. \quad (9)$$

ここで、式 (9) の行列 D は、節点 v_i の次数 $d_i = \sum_{j=1}^n k_{ij}$ を対角成分とする対角行列である。式 (9) の上位 k 個の固有値に対応する固有ベクトル $z_j (j = 1, 2, \dots, k)$ を求め、第 j 列を z_j とする行列 $Z = [z_1, z_2, \dots, z_k]$ の第 i 行をデータとみなして k-平均法を行いクラスタリングを達成する。

2-4 重み付きカーネル k -平均法

スペクトラルクラスタリングは、うまく類似度行列の設計ができていれば、最終的に分類を行う k -平均法における初期クラスタへの依存が低くなり、さらに、線形分離不可能なクラスタからなるデータの分類にも適している。一方で、固有値問題を解く必要があるため数値計算を必要とする。これに対して、Dhillon(Dhillon et al., 2007) らは、データ間の距離計算のみで問題を解くことができる重み付きカーネル k -平均法を提案した。重み付きカーネル k -平均法は式 (5) の最適化問題の目的関数を式 (10) のように、各データ x_i の重み w_i を考慮した形式に変換した手法である。

$$\begin{aligned} \min_{C_j} & \sum_{j=1}^k \sum_{x_i \in C_j} w_i \| \phi(x_i) - M_j \|^2, \\ \text{s.t.} & \bigcup_{j=1}^k C_j = X, \\ & C_j \cap C_l = \emptyset, 1 \leq j, l \leq k, j \neq l. \end{aligned} \quad (10)$$

したがって、式 (10) よりクラスタ中心は式 (11) となる。

$$M_j = \frac{\sum_{x_i \in C_j} w_i \phi(x_i)}{\sum_{x_i \in C_j} w_i}, M_j \in R^D (j = 1, 2, \dots, k). \quad (11)$$

式 (11) より、クラスタ中心 $M_j \in R^D$ とデータ $\phi(x_i) \in R^D$ の距離は式 (12) で計算できる。

$$\begin{aligned} & \| M_j - \phi(x_i) \|^2 \\ &= \| M_j \|^2 - 2M_j \cdot \phi(x_i) + \phi(x_i) \cdot \phi(x_i) \\ &= \frac{\sum_{x_t, x_l \in C_j} w_t w_l K(x_t, x_l)}{(\sum_{x_i \in C_j} w_i)^2} - \frac{2 \sum_{x \in C_j} w_i K(x_i, x)}{\sum_{x_i \in C_j} w_i} \\ & \quad + K(x_i, x_i), (j = 1, 2, \dots, k). \end{aligned} \quad (12)$$

式 (10) の問題は、重み w_i と $K(x_i, x_j)$ を i 行 j 列の要素とするカーネル行列 K を適切に設定することにより、いくつかの種類の目的関数のスペクトラルクラスタリングと数学的に等価になることが証明されている(Dhillon et al., 2007)。本論文で扱う Ncut の場合は、 x_i の次数 $d_i = \sum_{j=1}^n K(x_i, x_j)$ を対角成分とする対角行列 D と式 (4) のガウス関数を要素とする類似度行列 A により、式 (12) の $K(x_i, x_j) (1 \leq i, j \leq n)$ を要素とするカーネル

行列 K を $K = D^{-1}AD^{-1}$ とし、各データの重みを対角成分とする対角行列 W を $W = D$ としなければならない。すなわち、 $w_i = d_i$ とする。ただし、実際には $D^{-1}AD^{-1}$ の正定値性を保証するため正の実数値 σ を用いた正則化項を加え $K = \sigma D^{-1} + D^{-1}AD^{-1}$ とする必要がある。 σ は原則として正の値であるがデータの分布形状によっては負の値にした方が良好な結果を得る場合もあり(Dhillon, Guan and Kulis, 2004) 調整が難しい。そこで本論文では、厳密には等価とされないとされているが(Dhillon, Guan and Kulis, 2007), $W = D^{-1}, K = A$ とした。

式 (10) の問題は、可能な C_j の選択に関して組合せ的な探索を必要とする離散最適化であるため NP 困難となる。したがって、局所解の出力を認める以下の発見的アルゴリズムを用いて計算する。

Algorithm 1 入力: $X, k, \delta, C_1^{(1)}, C_2^{(1)}, \dots, C_k^{(1)}$

- 1 繰返し番号 t を $t = 1$ で初期化する。
 - 2 各データに対して各クラスタ $C_j^{(t)}$ の中心との距離を式 (12) により計算し、最も近いクラスタ中心のクラスタに割り当てる。
 - 3 2の結果から式 (10) の目的関数を再計算し、変化が無ければ $C_j^{(t)}$ ($j = 1, 2, \dots, k$) を出力して計算を終了する。そうでなければ $t = t + 1$ として 2 へ戻る。
-

Algorithm 1 の入力 $C_j^{(1)}$ ($j = 1, 2, \dots, k$) は初期クラスタである。 k -平均法とカーネル k -平均法も基本的には Algorithm 1 で計算を行い、初期クラスタは、対象データ $x \in X$ をランダムに割り当てることが多い。しかし、重み付きカーネル k -平均法は初期クラスタ依存度が高く、初期クラスタとして、対象データ $x \in X$ をランダムに割り当ててもうまく分類できない。そこで、Dhillon らは、データをグラフの節点とみなし適切な枝でグラフを構成した後、グラフの粗大化(Karypis and Kumar, 1998) と局所探索(Dhillon, Guan and Kogan, 2002) により初期クラスタを決定している。

グラフの粗大化は、節点集合 V 、枝集合 E 、類似度行列 K からなるグラフ $G(V, E, K)$ が入力として与えられたとき、グラフ G の規模が小さくなるように粗く表現したグラフ

$G'(V', E', K')(|V'| > |V'|)$ を作成し、問題を解く手法である。データ集合 X を節点集合 V とみなし、類似度行列 K をカーネル行列とする。ランダムに選択した $x \in V$ に対して式 (13) を満たす探索済みではない節点 $y \in V$ を選択し、 x と y を一つの節点としてまとめる。

$$\max_{(x,y) \in E} \frac{e(x,y)}{w(x)} + \frac{e(x,y)}{w(y)} \quad (13)$$

ここで、式 (13) の $e(x,y)$ は x と y の類似度で、 x と y が複数個の節点からなるときはそれらの類似度の和となる。 $w(x_i)$ は重み w_i と等価であり、複数個の節点のときには、やはり重みの和となる。式 (13) を満たす y が見つければ探索済みとし、 x はいずれにしても探索済みとする。全ての節点が探索済みとなった時点で、粗大化した節点を新しい節点とし、同じ操作を段階的に繰り返す。ある段階における節点、つまりデータ集合の数が $5k$ 以下であれば、その節点集合をデータ集合とみなして重み付きカーネル k -平均法を実行する。次に、得られたクラスタリング結果に関して、粗大化を一段階解除する。節点の数は $5k$ よりおおきくなり、その節点集合をデータ集合、前段階のクラスタリング結果を初期クラスタとして、再び重み付きカーネル k -平均法を実行する。以下、粗大化が全て解除され、元のデータ集合が復元されまで繰り返す。

さらに、各段階で粗大化を解除し、重み付きカーネル k -平均法を実行した後、局所探索を行う。局所探索は、ある一つのデータを現在のクラスタから他の異なるクラスタに割り当て変更した場合の式 (10) の目的関数の変化を観察する手法である。全データに対して評価を行い、最終的に目的関数が最小となるデータのみを割り当て変更を行う。この処理を目的関数が減少しなくなるまで繰り返し行い、結果として得られたクラスタを、次の粗大化解除後の初期クラスタとする。

3. パラメータ選択に関する研究

3-1 Zelnik-Manor らの手法

Zelnik-Manor (Zelnik-Manor and Perona, 2004) らは、スペクトラルクラスタリングにおいて、ガウス関数を類似度としたときのパラメータを自動的に決定する方法を提案している。データの分布における局所的なスケールを考慮する方法であり、カーネル行列の各要素

$k_{ij}(1 \leq i, j \leq n)$ を式 (14) で決定する。

$$k_{ij} = \exp \left(-\frac{\|x_i - x_j\|^2}{\sigma_i(m)\sigma_j(m)} \right). \quad (14)$$

ここで、式 (14) の $\sigma_i(m)$ は x_i の m 近傍点までの距離である。Zelnik-Manor らは、 m は次元に考慮して決定する必要があるとしているものの、詳細な議論はしていない。しかし、実験では $m = 7$ と固定して良好な結果を得ている。

3-2 Yu らの手法

Yu ら (Yu, Tranchevent, De Moor and Moreau, 2011) は、スペクトラルクラスタリングにおけるカーネル行列を、一種類のパラメータを用いて一つ用意するのではなく、式 (15) のように複数のパラメータを用いて用意した各カーネル行列の線形和とすることで、パラメータ推定の困難さを緩和する手法を提案している。

$$\Omega = \sum_{r=1}^p \theta_r G_r, \forall \theta_r \geq 0, \sum_{r=1}^p \theta_r^\delta = 1. \quad (15)$$

ここで、式 (15) の G_r は $G_r = PK_r P$ であり、 K_r は r 番目のカーネルパラメータを用いて用意したカーネル行列、 P は $P = I - (1/n)zz^T$ によりデータの中心を原点に強制する行列で、 I は単位行列、 z は要素が全て 1 の列ベクトルである。また、 δ は問題調整のパラメータである。式 (15) より、事前に p 個のカーネルパラメータを用意した後、 $\theta_r (r = 1, 2, \dots, p)$ を推定する必要があるが、Yu らは、まずスペクトラルクラスタリングを行い、得られたクラスタリング結果を教師ラベルとして多クラスに対するカーネル判別分析 (Mika, Ratsch, Weston, Scholkopf and Mullers, 1999) を行うことで θ_r を推定している。任意の判別軸に対してクラス間の散布を最大にするよう、適当な判別軸から始めて、その判別軸に対する θ_r を推定し、得られた θ_r で構築しなおしたデータに対して最適となる新しい判別軸を推定することを、クラス間散布が変化しなくなるまで繰り返す。正則化項 $G_{p+1} = I$ も加え、連続量である任意の判別軸に対して有限個の $\theta_r (r = 1, 2, \dots, p+1)$ を決定する半無限計画問題として、式 (16) で定式化される。

$$\begin{aligned}
& \max_{\theta, u} u, \\
& \text{s.t. } \theta_r \geq 0, r = 1, 2, \dots, p+1, \\
& \sum_{r=1}^{p+1} \theta_r^\delta \leq 1, \delta \geq 1, \\
& \sum_{r=1}^{p+1} \theta_r f_r(\beta) \geq u, \forall \beta, \\
& f_r(\beta) = \sum_{q=1}^k \left(\frac{1}{2} \beta_q^T G_r \beta_q - \beta_q - \beta_q Y_q^{-1} b \right), \\
& r = 1, 2, \dots, p+1. \tag{16}
\end{aligned}$$

ここで、式(16)の β は判別軸を定める変数で判別分析における最適化問題の解である。また、 Y_q は q 番目とそれ以外のクラスを1と-1で分類する n 個の教師ラベルを対角成分とする対角行列である。 δ は問題を調整するパラメータで Y_u らは $\delta = 1$ として線形計画問題で定式化している。式(16)の解き方は、まず、固定した任意の β に対して式(16)の線形計画問題を解き θ を求める。次に、得られた θ を用いて1対他法で多クラスに拡張した最小二乗サポートベクターマシン(Suykens, Van Gestel, De Moor and Vandewalle, 2002)により β を求める。最小二乗サポートベクターマシンは各2組のクラスに対して式(17)の連立方程式を解くことと等価である。

$$\begin{bmatrix} 0 & z^T \\ z & \Omega \end{bmatrix} \begin{bmatrix} b \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ Y^{-1}z \end{bmatrix}. \tag{17}$$

以下、線形計画問題と連立方程式を繰り返し計算するが、任意の β に対して式(16)を最適化する θ を求める必要があるため、式(16)の4行目の β に関する制約式は1回の繰り返し毎に1つつ増えていくことになる。 Y_u らは式(16)の解 θ は10回程度の繰り返しで求まるとしている。求めた θ を用いて再びスペクトラルクラスタリングを行い、以下、式(16)の半無限計画問題とスペクトラルクラスタリングをクラスが変化しなくなるまで交互に繰り返す。

3-3 文献(柳, 2014)の手法

文献(柳, 2014)の手法は、初期パラメータ δ を入力として、2対のクラス間の距離が最大になるように強制的にクラスの変更を行い、その変更を満たすようにパラメータの更新を行う。そして、カーネル k -平均法を行い、以下目的関数の値が変化しなくなるまで、上記の3

過程を繰り返す方法である。クラスの変更は2対のクラス C_a, C_b に対して、移動データの集合を $X_a \subset C_a$ とすると、 $\forall x \in C_a - X_a$ と C_b との距離が $\forall x' \in X_a$ と C_b との距離より小さくならず、かつ、 $C_a - X_a$ と $C_b \cup X_a$ の距離が最大になるように決定する。さらに、 $\forall x \in C_a - X_a$ と $\forall x' \in C_b \cup X_a$ が互いに k 近傍とならないようにする。このとき、クラス $C_a^{\text{new}} = C_a - X_a$ と $C_b^{\text{new}} = C_b \cup X_a$ の各中心、および、 C_b^{new} の最近傍点 $x_a \in C_a^{\text{new}}$ 、 C_a^{new} の最近傍点 $x_b \in C_b^{\text{new}}$ との距離に関して以下の式が成立するように δ^{new} を決定する。

$$\|M_a^{\text{new}} - \phi(x_a)\|^2 \leq \|M_b^{\text{new}} - \phi(x_a)\|^2, \tag{18}$$

$$\|M_a^{\text{new}} - \phi(x_b)\|^2 \geq \|M_b^{\text{new}} - \phi(x_b)\|^2. \tag{19}$$

ここで、 M_a^{new} はクラス C_a^{new} の中心点であり、カーネル関数により M_a^{new} が関わる計算をするときは δ^{new} を用いる。パラメータ δ は式(20)、式(21)により逐次更新を行う。

$$\delta^{\text{new}} = \rho \delta. \tag{20}$$

$$\rho = \sqrt{\frac{1}{\left(1 + s \cdot \frac{\|M_a^{(\text{new})}, x_a\|^2 - \|M_b^{(\text{new})} - x_a\|^2}{\left(\frac{(|C_b^{\text{new}}| - 1)}{|C_b^{\text{new}}|} + 2 \frac{(|C_a^{\text{new}}| - 1)}{|C_a^{\text{new}}|} \right)} \right)}}. \tag{21}$$

式(21)の $M_a^{(\text{new})}$ はクラス C_a^{new} の中心点であり、カーネル関数により $M_a^{(\text{new})}$ が関わる計算をするときは δ^{new} ではなく δ を用いる。 s は $s \in \{-1, 1\}$ で探索方向が正方向なら-1、負方向なら+1をとる。正方向と負方向の切り替えには、データ X の二点間の距離の分布情報を利用する。クラス C_b の任意の二点間の距離、および、クラス C_a の任意の点と x_a の距離の最大値、最小値、中央値を、それぞれ、 $d_{b,\text{max}}, d_{b,\text{min}}, d_{b,\text{med}}$ 、および、 $d_{a,\text{max}}, d_{a,\text{min}}, d_{a,\text{med}}$ とする。もし、 $d_{b,\text{max}} \leq d_{a,\text{max}}$ なら δ の探索区間を $[d_{b,\text{min}}, d_{b,\text{max}}]$ とし、 $\delta \leq d_{b,\text{med}}$ なら正方向、そうでなければ負方向を探索する。また、 $d_{b,\text{max}} > d_{a,\text{max}}$ なら δ の探索区間を $[d_{a,\text{min}}, d_{a,\text{max}}]$ とし、 $\delta \leq d_{a,\text{med}}$ なら正方向、そうでなければ負方向を探索する。

4. 提案法

従来法である文献(柳, 2014)の手法はクラスタ間の距離が十分に離れていて, 2つのクラスタからなるデータに対しては有効であるものの, クラスタ間の距離が十分離れていてもクラスタ数が増えるとクラスタリング結果の精度が落ち, また, クラスタ間の距離が十分離れていないときはクラスタが含むデータ数が極端に偏るなどの問題点がある。本論文では, 上記の問題点の改良を検討する。基本的には, 従来法と同様の考え方により, ガウスカーネルのパラメータを探索するが, 従来法のようにパラメータ探索とカーネル k -平均法を交互に繰り返すのではない。従来法は, ランダムに決めた初期クラスタに対してカーネル k -平均法を行い, 得られた結果から距離を基準にクラスタ内の部分集合を入れ換えるという処理をしているため, クラスタ数が増えると計算結果が不安定になる。そこで, 本論文では, 以下の各手順を1回ずつ計算する。

- (1) 集合間の距離を基準に初期クラスタを作成する。
- (2) 初期クラスタを分離するカーネルパラメータを探索する。
- (3) 重み付きカーネル k -平均法によりクラスタリングを行う。

上記の(3)にあるように, スペクトラルクラスタリングとの比較も行うため, クラスタリング手法として重み付きカーネル k -平均法を採用する。

4-1 初期クラスタの決定

まず, 任意の有限集合 $A = \{a \mid a \in R^d\}$ と $B = \{b \mid b \in R^d\}$ の集合間の距離 $d(A, B)$ を式(22)により定式化する。

$$d(A, B) = \min_{\forall a \in A, \forall b \in B} \|a - b\|^2. \quad (22)$$

式(22)で要素が一つの集合, つまり, $A = \{a\}, B = \{b\}$ のときは, 通常の L_2 ノルム $d(A, B) = d(\{a\}, \{b\}) = d(a, b) = \|a - b\|^2$ となる。また, 特殊な場合として, 空集合との距離を0, つまり, $d(A, \emptyset) = 0$ と定義する。式(22)より, R^d 上の集合 A, B, C, D , に関して, 式(23)の性質が成り立つ。

$$\begin{aligned} d(A \cup B, C \cup D) = \\ \min\{d(A, C), d(A, D), d(B, C), d(B, D)\}. \end{aligned} \quad (23)$$

次に, 本論文で使用する, クラスタ間の距離が十分に離れているデータ集合という基準を定義1により定量化する。

定義1. 式(24)を満たすデータ集合 X の $k(\geq 2)$ 個の部分集合を要素とする集合を $disjoint(X, k)$ と定義する。

$$\begin{aligned} disjoint(X, k) = \{X_i \mid \cup_{i=1}^k X_i = X, \\ X_i \cap X_j = \emptyset, \\ \max_{\forall A \subseteq X_i} d(A, X_i - A) \leq d(X_j, X_i), \\ 1 \leq i, j, t \leq k, j \neq t\}. \end{aligned} \quad (24)$$

式(24)は $disjoint(X, k)$ の要素を任意に2分割した部分集合同士の距離が, $disjoint(X, k)$ の要素同士の距離より小さいことを意味する。まず, 集合を2分割したときの部分集合同士の距離の性質を示す。

定理1. $|X| \geq 2$ となる有限のデータ集合 $X = \{x \mid x \in R^d\}$ の部分集合 Y と Z が式(25)を満たすとき, 式(26)と式(27)を満たす少なくとも1組の Y, Z が存在する。

$$\begin{aligned} Y = \arg \max_{A \subseteq X} d(A, X - A), \\ Z = X - Y. \end{aligned} \quad (25)$$

$$\max_{\forall B \subseteq Y} d(B, Y - B) \leq d(Y, Z). \quad (26)$$

$$\max_{\forall C \subseteq Z} d(C, Z - C) \leq d(Y, Z). \quad (27)$$

証明. まず, X の任意の二つの要素の距離 $d(x, x')(\forall x, x' \in X)$ が全て異なる場合を考え, 式(26)が成立することを証明する。仮に式(28)を満たす Y_1, Y_2 が存在するとする。

$$\begin{aligned} d(Y_1, Y_2) > d(Y, Z), \\ Y_1 \cup Y_2 = Y, Y_1 \cap Y_2 = \emptyset. \end{aligned} \quad (28)$$

X の任意の二つの要素の距離が全て異なるという仮定より, 一般性を失わず $d(Y, Z) = d(Y_1, Z) < d(Y_2, Z)$ と仮定することで, 式(23)より, 式(29)が成立する。

$$d(Y_1 \cup Z, Y_2) = \min\{d(Y_1, Y_2), d(Z, Y_2)\}. \quad (29)$$

式(29)において, もし, $d(Y_1 \cup Z, Y_2) = d(Y_1, Y_2)$ ならば式(25)に矛盾する。また,

$d(Y_1 \cup Z, Y_2) = d(Z, Y_2)$ ならば, $d(Y_1 \cup Z, Y_2) = d(Z, Y_2) > d(Y, Z)$ となり, やはり, 式 (25) に矛盾する。したがって, 式 (28) を満たす Y_1, Y_2 は存在せず, 式 (26) が成立する。式 (27) についても, 同様に証明できる。

次に, X の任意の二つの要素の距離に同じものがある場合を考える。式 (25) を満たす任意の Y, Z において, もし, 式 (26) を満たさない Y の部分集合 Y_1, Y_2 が存在する場合, すなわち, $d(Y_1, Y_2) > d(Y, Z)$ を満たす Y_1, Y_2 が存在する場合, 式 (30) を満たす必要がある。

$$\begin{aligned} d(Y_1, Z) &= d(Y, Z), \\ d(Y_2, Z) &= d(Y, Z). \end{aligned} \quad (30)$$

なぜなら, 新たに $Y = Y_1, Z = Z \cup Y_2$ としても, 式 (31) より式 (25) が成立するためである。

$$\begin{aligned} d(Y_1, Z \cup Y_2) &= \min\{d(Y_1, Z), d(Y_1, Y_2)\} \\ &= d(Y, Z). \end{aligned} \quad (31)$$

式 (26) を満たさない Y_1, Y_2 に関して, 式 (30) の性質は再帰的に成立するため, 以下 Y が式 (26) を満たすまで, Y の分割と Z と Y の部分集合の結合を繰り返すことで, 式 (25) と式 (26) を満たす Y, Z を見つけることができる。

上記で見つけた Y, Z に関して, 式 (27) を満たさない Z の部分集合 Z_1, Z_2 が存在する場合, すなわち, $d(Z_1, Z_2) > d(Y, Z)$ を満たす Z_1, Z_2 が存在する場合, 式 (32) を満たす必要がある。

$$\begin{aligned} d(Z_1, Y) &= d(Y, Z), \\ d(Z_2, Y) &= d(Y, Z). \end{aligned} \quad (32)$$

さらに, 式 (33) を満たす Z_1 の部分集合 ZZ_1, ZZ_2 が存在すると仮定する。

$$\begin{aligned} d(ZZ_1, ZZ_2) &> d(Y, Z), \\ ZZ_1 \cup ZZ_2 &= Z_1, \\ ZZ_1 \cap ZZ_2 &= \emptyset. \end{aligned} \quad (33)$$

この場合, ZZ_1, ZZ_2 は式 (34) を満たす必要がある。

$$\begin{aligned} d(ZZ_1, Y) &= d(Z_1, Y), \\ d(ZZ_2, Y) &= d(Z_1, Y). \end{aligned} \quad (34)$$

何故なら, 仮に $d(ZZ_2, Y) > d(Z_1, Y)$ であれば, 式 (35) より, 式 (25) に矛盾するためである。

$$\begin{aligned} d(Y \cup ZZ_1 \cup Z_2, ZZ_2) \\ &= \min\{d(Y, ZZ_2), d(ZZ_1, ZZ_2), d(Z_2, ZZ_2)\} \\ &> d(Y, Z). \end{aligned} \quad (35)$$

同様に, $d(ZZ_1, Y) > d(Z_1, Y)$ ならば, $d(Y \cup ZZ_2 \cup Z_2, ZZ_1) > d(Y, Z)$ となり, やはり, 式 (25) に矛盾する。したがって, Z_1 部分の任意の 2 分割, および, Y の任意の 2 分割 $Y_1, Y_2 (Y_1 \cup Y_2 = Y, Y_1 \cap Y_2 = \emptyset)$ に関わる距離に関して式 (36) が成立する。

$$\begin{aligned} d(Y_1 \cup ZZ_1, Y_2 \cup ZZ_2) &= \\ &\min\{d(Y_1, Y_2), d(Y_1, ZZ_2), d(ZZ_1, Y_2), \\ &\quad d(ZZ_1, ZZ_2)\} \leq d(Y, Z), \\ d(Y, ZZ_1 \cup ZZ_2) &= \\ &\min\{d(Y, ZZ_1), d(Y, ZZ_2)\} \leq d(Y, Z). \end{aligned} \quad (36)$$

式 (36) より, Y, Z を $Y = Y \cup Z_1, Z = Z_2$ と置き換えても Y に関して式 (26) が成立しており, 式 (27) を満たさない Z に関する式 (32) の性質, および, その Z の部分集合が式 (33) を満たす場合の式 (34) の性質は再帰的に成立する。したがって, 以下, Z が式 (27) を満たすまで, Z の分割と Y と Z の部分集合の結合を繰り返すことで, 式 (25), 式 (26), 式 (27) を満たす Y, Z を見つけることができる。□

次に, 定理 1 を用いて, $\text{disjoint}(X, k)$ の存在可能性を示す。

定理 2. 有限のデータ集合 $X = \{x \mid x \in R^d\}$ と整数 k が与えられ, $|X| \geq k$ を満たすとき, 少なくとも一つの $\text{disjoint}(X, k)$ が存在する。

証明. まず, $k = 2$ のときは, 定理 1 より, $\text{disjoint}(X, 2)$ が存在する。次に, $\text{disjoint}(X, k-1)$ が存在すると仮定して, $\text{disjoint}(X, k)$ の存在可能性を考える。 $\text{disjoint}(X, k-1) = \{X_1, X_2, \dots, X_{k-1}\}$ とし, 各 $X_i (1 \leq i \leq k-1)$ の $\text{disjoint}(X_i, 2) = \{A_i, B_i\}$ に対して, $\arg \max_i d(A_i, B_i) = k-1$ とする。このとき, $A_{k-1} = S \cup T, S \cap T = \emptyset$ を満たす任意の S, T , および, $B_{k-1} = U \cup V, U \cap V = \emptyset$ を満たす任意の U, V は式 (37) を満たす。

$$\begin{aligned} d(S, T) &\leq d(A_{k-1}, B_{k-1}) \leq \min_{1 \leq i, j \leq k-1, i \neq j} d(X_i, X_j), \\ d(U, V) &\leq d(A_{k-1}, B_{k-1}) \leq \min_{1 \leq i, j \leq k-1, i \neq j} d(X_i, X_j). \end{aligned} \quad (37)$$

また, $X_t (1 \leq t \leq k-2)$ に関して, 式 (38) が成立する。

$$\begin{aligned} d(X_{k-1}, X_t) &\leq d(A_{k-1}, X_t), \\ d(X_{k-1}, X_t) &\leq d(B_{k-1}, X_t). \end{aligned} \quad (38)$$

従って $\{X_1, X_2, \dots, X_{k-2}, A_{k-1}, B_{k-1}\} = \text{disjoint}(X, k)$ となる。 \square

ここで、集合を要素とする集合 $Z = \{X_1, X_2, \dots, X_n\}$ があたえられたとき、 $\forall Y \in Z$ の最近傍要素 $NN(Y)$ を式 (39) で定義する。

$$NN(Y) = S = \arg \min_{\forall S \in Z - \{Y\}} d(Y, S). \quad (39)$$

定理 2 の証明の考え方より系 1 が得られる。

系 1. $\text{disjoint}(X, k) = \{X_1, X_2, \dots, X_k\}$ において $\arg \min_i d(X_i, NN(X_i)) = t$ のとき式 (40) が成立する。

$$\begin{aligned} \text{disjoint}(X, k-1) &= (\{X_1, X_2, \dots, X_k\} \\ &\quad - \{X_t, NN(X_t)\}) \\ &\quad \cup \{X_t \cup NN(X_t)\}. \end{aligned} \quad (40)$$

証明. 一般性を失わず $t = k, NN(X_k) = X_{k-1}$ とする。このとき、 $\text{disjoint}(X, k) = \{X_1, X_2, \dots, X_k\}$ の仮定より $\forall A \subseteq X_k, \forall B \subseteq X_{k-1}$ に対して式 (41) が成立する。

$$\begin{aligned} d(A, X_k - A) &\leq d(X_{k-1}, X_k), \\ d(B, X_{k-1} - B) &\leq d(X_{k-1}, X_k), \\ d(A, B) &\geq d(X_{k-1}, X_k), \\ d(X_k - A, X_{k-1} - B) &\geq d(X_{k-1}, X_k). \end{aligned} \quad (41)$$

また、 $S \cup T = X_k \cup X_{k-1}, S \cap T = \emptyset$ を満たす任意の S, T を $S = A \cup (X_{k-1} - B), T = B \cup (X_k - A)$ と表現できる。式 (41) より、 $d(S, T)$ に関して式 (42) が成立する。

$$\begin{aligned} d(S, T) &= \min\{d(A, B), d(A, X_k - A), \\ &\quad d(B, X_{k-1} - B), d(X_k - A, X_{k-1} - B)\} \\ &\leq d(X_{k-1}, X_k). \end{aligned} \quad (42)$$

さらに、任意の $X_t (1 \leq t \leq k-2)$ に関して、式 (43) が成立する。

$$\begin{aligned} d(X_t, X_{k-1} \cup X_k) &= \min\{d(X_t, X_{k-1}), d(X_t, X_k)\} \\ &\geq d(X_{k-1}, X_k). \end{aligned} \quad (43)$$

式 (42) と式 (43) より、 $X_{k-1} \cup X_k$ は $\{X_1, X_2, \dots, X_{k-2}\}$ に対して $\text{disjoint}(X, k-1)$ の条件を満たしてい

る。また、 $S \cup T = X_t, S \cap T = \emptyset$ を満たす任意の S, T に関して式 (44) が成立する。

$$\begin{aligned} d(S, T) &\leq d(X_t, X_{k-1}), \\ d(S, T) &\leq d(X_t, X_k). \end{aligned} \quad (44)$$

式 (44) より式 (45) が成立する。

$$\begin{aligned} d(S, T) &\leq \min\{d(X_t, X_{k-1}), d(X_t, X_k)\} \\ &\leq d(X_t, X_{k-1} \cup X_k). \end{aligned} \quad (45)$$

式 (45) より、 $\{X_1, X_2, \dots, X_{k-2}\}$ は $X_{k-1} \cup X_k$ に対して $\text{disjoint}(X, k-1)$ の条件を満たしている。さらに、 $\text{disjoint}(X, k) = \{X_1, X_2, \dots, X_{k-2}, X_{k-1}, X_k\}$ より、 X_1, X_2, \dots, X_{k-2} に関しては相互に条件を満たしている。したがって、 $\text{disjoint}(X, k-1) = \{X_1, X_2, \dots, X_{k-1} \cup X_k\}$ となる。 \square

定理 3. データ集合 X の任意の二つの要素 x, x' の距離が全て異なるとき、 $\text{disjoint}(X, k)$ は一意に定まる。

証明. まず、 $k=2$ のとき一意に定まらないと仮定し、2組の $\text{disjoint}(X, 2)$ が存在するものとする。2組を $\{A, B\}, \{Y, Z\}$ とすると式 (46) を満たす。

$$\begin{aligned} X &= A \cup B, A \cap B = \emptyset, \\ X &= Y \cup Z, Y \cap Z = \emptyset, \\ \{A, B\} &\neq \{Y, Z\}. \end{aligned} \quad (46)$$

また、任意の A_1, A_2, B_1, B_2 を用意し、 $\{A, B\}, \{Y, Z\}$ 間の関係を式 (47) とする。

$$\begin{aligned} A &= A_1 \cup A_2, A_1 \cap A_2 = \emptyset, \\ B &= B_1 \cup B_2, B_1 \cap B_2 = \emptyset, \\ A_1 \cup B_1 &= Y, \\ A_2 \cup B_2 &= Z. \end{aligned} \quad (47)$$

このとき、 $d(A, B) = d(Y, Z)$ と仮定すると、 $\text{disjoint}(X, 2) = \{A, B\}$ および X の任意の 2 点間の距離が全て異なる仮定より、 $d(A, B) < d(A_2, B_2)$ または $d(A, B) < d(A_1, B_1)$ を満たさなければならない。仮に $d(A, B) < d(A_2, B_2)$ とすると、一方で $A_2 \cup B_2 = Z$ かつ $\text{disjoint}(X, 2) = \{Y, Z\}$ の仮定より $d(A, B) = d(Y, Z) \geq d(A_2, B_2)$ を満たさなければならない矛盾する。また、 $d(A, B) \neq d(Y, Z)$ かつ $d(A, B) < d(Y, Z)$ と仮定すると $d(Y, Z)$ は式

(48)~式(51)のいずれかを満たさなければならぬ。

$$d(Y, Z) = d(A_1, A_2). \quad (48)$$

$$d(Y, Z) = d(B_1, B_2). \quad (49)$$

$$d(Y, Z) = d(A_1, B_2). \quad (50)$$

$$d(Y, Z) = d(B_1, A_2). \quad (51)$$

式(48)は $d(A_1, A_2) \leq d(A, B) < d(Y, Z)$ より矛盾する。式(49)も同様である。式(50)は $d(B_1, B_2) < d(A, B) < d(A_1, B_2) = d(Y, Z)$ を満たさなければならず、一方で $B_1 \subseteq Y, B_2 \subseteq Z$ より $d(B_1, B_2) \geq d(Y, Z)$ も満たす必要があり、やはり矛盾する。式(51)も同様である。したがって、異なる $\text{disjoint}(X, 2)$ は存在せず一意に定まる。

次に、 $\text{disjoint}(X, k-1)$ が一意に定まると仮定する。 $\text{disjoint}(X, k)$ が一意に定まらなると仮定し、2組の $\{A_1, A_2, \dots, A_k\}, \{B_1, B_2, \dots, B_k\}$ が共に $\text{disjoint}(X, k)$ であるとする。一般性を失わず $\arg \min_{1 \leq i \leq k} d(A_i, NN(A_i)) = k$, $\arg \min_{1 \leq i \leq k} d(B_i, NN(B_i)) = k$ とすると系1より式(52)が成立する。

$$\begin{aligned} \text{disjoint}(X, k-1) &= \{A_1, A_2, \dots, A_{k-1} \cup A_k\} \\ &= \{B_1, B_2, \dots, B_{k-1} \cup B_k\}. \end{aligned} \quad (52)$$

$A_{k-1} \cup A_k \neq B_{k-1} \cup B_k$ とすると式(52)より、 $A_{k-1} \cup A_k = B_t (1 \leq t < k-2), B_{k-1} \cup B_k = A_u (1 \leq u < k-2), t \neq u$ を満たす t, u が存在し、 X の要素の二点間の距離が全て異なるという仮定より、 $d(A_{k-1}, A_k) \neq d(B_{k-1}, B_k)$ となる。一般性を失わず、 $d(A_{k-1}, A_k) > d(B_{k-1}, B_k)$ とすると、 $S \cup T = B_t, S \cap T = \emptyset$ を満たすある S, T に対して $d(S, T) = d(A_{k-1}, A_k) > d(B_{k-1}, B_k)$ が成立するため $\{B_1, B_2, \dots, B_k\} = \text{disjoint}(X, k)$ に矛盾する。また、 $A_{k-1} \cup A_k = B_{k-1} \cup B_k$ かつ $\{A_{k-1}, A_k\} \neq \{B_{k-1}, B_k\}$ とすると $\text{disjoint}(A_{k-1} \cup A_k, 2) \neq \text{disjoint}(B_{k-1} \cup B_k, 2)$ となり、 $k=2$ で $\text{disjoint}(X, 2)$ が一意に定まることに反する。したがって、 $\{A_1, A_2, \dots, A_k\} = \{B_1, B_2, \dots, B_k\}$ であり、 $\text{disjoint}(X, k)$ も一意に定まる。□

次に、提案手法の初期クラスタを定義するための準備をする。 X の部分集合 $T \subseteq X$ からなる

集合 $N = \{T \mid T \subseteq X, T \cap T' = \emptyset, \forall T' \in N\}$ が与えられたとき、 N の部分集合からなる集合 S を式(53)のように定義する。

$$\begin{aligned} S &= \{T \mid T \in N, \\ &(\exists NN(T) \in S) \vee (\exists T' \in S, NN(T') = T)\}. \end{aligned} \quad (53)$$

一般的に、 $NN(S)$ は必ずしも対称性を満たすとは限らない、つまり、 $S = NN(S')$ かつ $NN(NN(S')) = S$ とは限らない。式(53)は S の要素となる条件が、他の S の要素の最近傍要素であるか、または、自身の最近傍要素が S の要素であることを示す。つまり、ある $T \in N$ が $T \in S$ となるならば、 T と S は式(54)または式(55)を満たす必要がある。

$$\exists NN(T) \in S. \quad (54)$$

$$\exists T' \in S, NN(T') = T. \quad (55)$$

したがって、 N から適当な要素 T を一つ選択し、 $S = \{T\}$ とした後、 S の全要素と N の全要素に対して式(54)と式(55)を繰り返し評価し、いずれかを満たした T' に対して、 $S = S \cup \{T'\}, N = N - \{T'\}$ とすることで、 T を含む S を一つ作成できる。また、この操作を $N = \emptyset$ となるまで繰り返すことで、 N の部分集合の集合 S を要素とする集合を作成できる。そのような集合 $D(N)$ を式(56)で定義する。

$$\begin{aligned} D(N) &= \{S \mid S = \{T \mid T \in N, \\ &(\exists NN(T) \in S) \vee (\exists T' \in S, NN(T') = T)\}, \\ &\cup_{S \in D(N)} S = N\}. \end{aligned} \quad (56)$$

S の性質より、明らかに $\forall S, S' \in D(N), S \cap S' = \emptyset$ が成立する。また、 N の各要素の最近傍要素に関して距離が大きい上位 k 個の要素からなる集合を式(57)のように $NN_k(N)$ と定義する。

$$\begin{aligned} NN_k(N) &= \{T \mid |NN_k(N)| = k, \\ &d(T, NN(T)) \geq d(T', NN(T')), \\ &\forall T' \in N - NN_k(N)\}. \end{aligned} \quad (57)$$

式(57)を用いて、式(56)を最近傍要素との距離が大きい上位 k 個の $T \in N$ に式(55)を適用しないように拡張する。すなわち、 X の部分集合 $T \subseteq X$ からなる集合 $N = \{T \mid T \subseteq X, T \cap T' = \emptyset, \forall T' \in N\}$ が与えられたとき、 N の部分集合の S を要素とする集合 $D_k(N)$ を式(58)で定

義する。

$$\begin{aligned}
 D_k(N) = \{S \mid S = \{T \mid T \in N, \\
 (NN(T) \in S) \vee \\
 ((\exists T' \in S) \wedge (T' \notin NN_k(N)), \\
 NN(T') = T)\}, \\
 \cup_{\forall S \in D_k(N)} S = N\}. \quad (58)
 \end{aligned}$$

式 (58) は $T \in S$ が $NN_k(N)$ の要素のときは、 $NN(T)$ を S の要素としないことを意味する。

式 (58) を用いて X の部分集合を要素とする集合 $D_{i,k}$ を式 (59) のように再帰的に定義する。

$$D_{i,k} = \begin{cases} D_k(X) & (i = 1) \\ D_k(D_{i-1,k}) & (i > 1) \end{cases} \quad (59)$$

最終的に、 $|D_{i,k}| = k$ を満たすとき、 $D_{i,k} = D_k$ と定義し、提案手法における初期クラスタとする。 D_k の性質を示すため、 $D_{i,k}$ の各要素とその最近傍点との距離からなる集合 $Dist_i$ を式 (60) で定義する。

$$Dist_i = \{d(D, NN(D)) \mid \forall D \in D_{i,k}\}. \quad (60)$$

また、 $D_{i,k}$ の各要素とその最近傍点との距離の大きい上位 k 個の集合を $Dist_{i,k} \subseteq Dist_i$ と定義し、 $Dist_{i,k}$ の最小値を $\min(Dist_{i,k})$ と定義する。このとき、以下の定理が成立する。

定理 4. データ集合 X の任意の二つの要素 x, x' の距離が全て異なるとき、 $NN_k(D_{i-1,k})$ の要素 D の中で $\min(Dist_{i-1,k}) = d(D, NN(D))$ を満たす要素 D は $\{D\} \in D_{i,k}$ 、または、 $\{D \cup NN(D)\} \in D_{i,k}$ となる。さらに、 $NN_k(D_{i-1,k})$ の D 以外の要素集合 $NN_k(D_{i-1,k}) - \{D\}$ は $NN_k(D_{i-1,k}) - \{D\} \subseteq D_{i,k}$ となる。

証明. $NN_k(D_{i-1,k}) = \{D_1, D_2, \dots, D_k\}$ として、一般性を失わず $\min(Dist_{i-1,k}) = d(D_1, NN(D_1))$ とする。まず、 $D_j (2 \leq j \leq k)$ が $\{D_j\} \in D_{i,k}$ となることを示す。これは、 $D_j (2 \leq j \leq k)$ が式 (54) と式 (55) を満たさないことを示せばよいが、 $D_{i,k}$ の定義より、 $D_j (2 \leq j \leq k)$ に対して式 (55) の評価は行わない。つまり、 $S = \{D_j\}$ に関して、 $S = S \cup \{NN(D_j)\}$ とはしないため、式 (54) を満たさないことを示せばよい。したがって、 $D_j \notin S$ となる S に対して、 $S = S \cup \{D_j\}$ という評価が発生しないよう、 $NN(D) = D_j$ を満たす $D \in D_{i-1,k} - NN_k(D_{i-1,k})$ が存在しないことを示せばよい。仮に、そのような D が存

在するならば $d(D, NN(D)) \leq \min(Dist_{i-1,k})$ となるため $2 \leq j \leq k$ に対して式 (61) を満たさなければならない。

$$\begin{aligned}
 d(D, NN(D)) &\leq \min(Dist_{i-1,k}) \\
 &< d(D_j, NN(D_j)) \\
 &\leq d(D_j, D). \quad (61)
 \end{aligned}$$

一方で、もし、 $d(D, D_j) = d(D, NN(D)) \notin Dist_{i-1,k} (2 \leq j \leq k)$ ならば式 (62) を満たす必要があり式 (61) に矛盾する。

$$d(D, D_j) = d(D, NN(D)) < \min(Dist_{i-1,k}). \quad (62)$$

次に、 D_1 が $\{D_1\} \in D_{i,k}$ 、または、 $\{D_1 \cup NN(D_1)\} \in D_{i,k}$ となることを示す。 $\{D_1\} \in D_{i,k}$ に関しては、 $d(D_1, NN(D_1)) \notin Dist_{i-1,k} - Dist_{i-1,k}$ 、つまり、 $NN(D) = D_1$ を満たす $D \in D_{i-1,k} - NN_k(D_{i-1,k})$ が存在しない場合であり、 $D_j (2 \leq j \leq k)$ の場合と同じように証明できる。 $\{D_1 \cup NN(D_1)\} \in D_{i,k}$ に関しては、 $NN(D) = D_1$ を満たす $D \in D_{i-1,k} - NN_k(D_{i-1,k})$ が存在する場合である。このような D に対しては式 (55) が一回だけ評価される。つまり、 $D \in S$ となる $S \in D_{i,k}$ が存在するならば、 $S = S \cup \{D_1\}$ は評価されるが、 $D_1 \in S$ に対しては式 (55) は評価されない。したがって、 $\{D \cup D_1\} \in D_{i,k}$ であるためには、式 (54) が評価されないよう、 $D = NN(A), D_1 = NN(B)$ を満たす $A, B \in D_{i-1,k} - NN_k(D_{i-1,k})$ が存在しないことを示せばよい。仮に、そのような A, B が存在するのであれば、 $NN(D_1) = D$ 、かつ、 $NN(D) = D_1$ より式 (63) が成立する。

$$\begin{aligned}
 d(A, NN(A)) &= d(A, D) > d(NN(D), D) \\
 &= d(D_1, D), \\
 d(B, NN(B)) &= d(B, D_1) > d(NN(D_1), D_1) \\
 &= d(D, D_1). \quad (63)
 \end{aligned}$$

式 (63) より、もし、 A, B が存在するのであれば $d(A, NN(A)), d(B, NN(B)) \in Dist_{i-1,k}$ とならなければならないが、 $A, B \in D_{i-1,k} - NN_k(D_{i-1,k})$ に矛盾する。□

定理 5. データ集合 X の任意の二つの要素 x, x' の距離が全て異なるとき、式 (64) の関係が成立する。

$$\min(Dist_{i-1,k}) \leq \min(Dist_{i,k}), (i \geq 2). \quad (64)$$

証明. まず, 任意の $D \in NN_k(D_{i-1,k})$ と $s(\geq 1)$ 個の $\{A_1, A_2, \dots, A_s\} \subseteq D_{i-1,k}$ に関して式 (65) が成立する。

$$d(D, A_1 \cup A_2 \cup \dots \cup A_s) \geq d(D, NN(D)). \quad (65)$$

また, $NN_k(D_{i-1,k}) = \{D_1, D_2, \dots, D_k\}$ として, 一般性を失わず $\min(Dist_{i-1,k}) = d(D_1, NN(D_1))$ とすると, 定理4より, $D_j \in D_{i,k}$ ($2 \leq j \leq k$) となるので, $NN(D_j) \in D_{i-1,k}$ を含む $D_{i,k}$ の要素を $S \in D_{i,k}$ とすると, 式 (65) より, 式 (66) が成立する。

$$\begin{aligned} d(D_j, S) &= d(D_j, NN(D_j)) \\ &\leq \min_{S' \in D_{i,k} - \{S\}} d(D_j, S'). \end{aligned} \quad (66)$$

したがって, S は $D_{i,k}$ における D_j の最近傍要素 $S = NN(D_j)$ となり, $Dist_{i-1,k} - \{\min(Dist_{i-1,k})\} \subseteq Dist_i$ を満たす。次に, D_1 に関して, D_1 が $D_{i,k}$ の要素となった場合は, D_j ($2 \leq j \leq k$) と同様, $NN(D_1) \in D_{i-1,k}$ に対して $d(D_1, S) = d(D_1, NN(D_1))$ を満たす $S \in D_{i,k}$ が存在し, S は $D_{i,k}$ における D_1 の最近傍要素 $S = NN(D_1)$ となり, $d(D_1, NN(D_1)) \in Dist_i$ となる。一方で, $\{D_1 \cup NN(D_1)\}$ が $D_{i,k}$ の要素となった場合, $NN(NN(D_1)) = D_1$ が成立するため, 式 (65) より, $\{D_1 \cup NN(D_1)\}$ の最近傍要素 S に関して式 (67) が成立する。

$$\begin{aligned} d(D_1 \cup NN(D_1), S) &= \min\{d(D_1, S), d(NN(D_1), S)\} \\ &\geq \min\{d(D_1, NN(D_1)), \\ &\quad d(NN(D_1), NN(NN(D_1)))\} \\ &= d(D_1, NN(D_1)). \end{aligned} \quad (67)$$

したがって, 式 (67) より, $D_1 \cup NN(D_1)$ とその最近傍要素 S との距離 d' は $d' \in Dist_i$ かつ $d' \geq d(D_1, NN(D_1))$ となる。もし, $d < \min(Dist_{i-1,k})$ を満たす d が $d \in Dist_{i,k}$ ならば上記の k 個が $Dist_{i,k}$ に含まれなければならないため $|Dist_{i,k}| > k$ となり矛盾する。□

最終的に D_k に関する以下の定理が得られる。
定理6. データ集合 X の任意の二つの要素 x, x' の距離が全て異なるとき, $D_k = disjoint(X, k)$ である。

証明. D_k の定義より, $\forall D, D' \in D_k$ に関して $D \cap D' = \emptyset$ となるので, $A \cup B = D, A \cap B =$

\emptyset を満たす任意の A, B の距離 $d(A, B)$ の最大値 $d_{D,max}$ が式 (68) を満たすことを示せばよい。

$$\forall D, d_{D,max} < \min_{\forall D' \in D_k} d(D', NN(D')). \quad (68)$$

D_k が n 回の式 (59) の操作で得られる, すなわち, $D_k = D_{n,k}$ とすると, $D = A \cup B$ ならば, ある $x \in A$ と $x' \in B$ の距離 $d(x, x')$ が $d(x, x') \in Dist_{t,k}$ ($1 \leq t \leq n-1$) を満たす必要がある。したがって, 定理5と合わせることで式 (69) が得られ, 式 (68) が成立する。

$$\begin{aligned} d_{D,max} &= d(A, B) \leq \min_{x \in A, x' \in B} d(x, x') \\ &\leq \min_{\exists t, 1 \leq t \leq n-1} \{Dist_{t,k}\} \leq \min\{Dist_{n,k}\} \\ &= \min_{\forall D' \in D_k} d(D', NN(D')). \end{aligned} \quad (69)$$

□

もし X の任意の2点間の距離が全て異なるのであれば, 定理6より D_k は一意に定まる。したがって, D_k を初期クラスタとしてガウスカーネルのパラメータを推定することで, $disjoint(X, k) = \{C_1, C_2, \dots, C_k\}$ を満たす δ が存在すれば, クラスタリング結果も C_1, C_2, \dots, C_k となり一意に定まる。一方で X が外れ値を含んでいたり, クラスタ内の分布が極端に異なるようなデータ集合の場合, $disjoint(X, k)$ がクラスタリングの意味で適切な分割とはならない可能性がある。そこで, 上記のような状況に対応できるよう, 整数値 r を用いて $D_{i,k}$ を式 (70) のように $D_{i,k}^r$ に緩和する。

$$D_{i,k}^r = \begin{cases} D(X) & (i=1, r>1) \\ D_k(X) & (i=1, r=1) \\ D(D_{i-1,k}^r) & (i < r) \\ D_k(D_{i-1,k}^r) & (i \geq r) \end{cases}. \quad (70)$$

式 (70) の $D_{i,k}^r$ が $|D_{i,k}^r| = k$ を満たすとき, $D_{i,k}^r = D_k^r$ と定義する。 $r=1$ のときは, $D_k^r = D_k$ となる。また, $r=2$ のときは, 一つのデータが D_k^r の要素となること, すなわち, $\exists \{x\} \in D_k^r$ となることを避けることができるため, 外れ値対策となる。 $D \in D_{r-1,k}^r$ ならば $D \in D_{r,k}^r$ とはならないため, 以下 r が増加するにつれて距離が大きい集合間の結合が速くなると考えられる。本論文では $r=3$ とする。 D_k^r を求めるアルゴリズムを Algorithm 2 に示す。

Algorithm 2 入力: $r, k, X = \{x_1, x_2, \dots, x_n\}$

- 1 $j \leftarrow 1$ とする。
- 2 一つのデータを初期クラスタとしてクラスタ集合 $D_k = \{C_i \mid C_i = \{x_i\}, i = 1, 2, \dots, n\}$ を作成する。
- 3 $|D_k| = k$ なら D_k を出力して終了する。
- 4 $NN_k \leftarrow \emptyset$ とする。
- 5 $T \leftarrow \emptyset$ とする。
- 6 $\forall C \in D_k$ に対して $NN(C)$ を求める。
- 7 $j \geq r$ ならば, $d(C, NN(C))$ が大きい上位 k 個の C を NN_k の要素として加え, $D_k \leftarrow D_k - NN_k$ とする。
- 8 $D \leftarrow \emptyset$ とする。
- 9 D_k の中から C を一つ選択し以下の更新を行う。

$$\begin{aligned} D_k &\leftarrow D_k - \{C\} \\ D &\leftarrow D \cup \{C\} \end{aligned}$$

- 10 $NN(C) \notin D_k \cup NN_k$ ならば 12 へ進む。
- 11 以下の更新を行い 10 へ戻る。

$$\begin{aligned} D &\leftarrow D \cup \{NN(C)\} \\ D_k &\leftarrow D_k - \{NN(C)\} \\ NN_k &\leftarrow NN_k - \{NN(C)\} \\ C &\leftarrow NN(C) \end{aligned}$$

- 12 $NNTMP \leftarrow D, D \leftarrow \emptyset$ とする。
- 13 $NNTMP$ の中から C を一つ選択し以下の更新を行う。

$$\begin{aligned} NNTMP &\leftarrow NNTMP - \{C\} \\ D &\leftarrow D \cup \{C\} \end{aligned}$$

- 14 $\forall C' \in D_k$ に対して $NN(C') = C$ ならば以下を行う。

$$\begin{aligned} NNTMP &\leftarrow NNTMP \cup \{C'\} \\ D_k &\leftarrow D_k - \{C'\} \end{aligned}$$

- 15 $NNTMP \neq \emptyset$ ならば 13 へ戻る。
 - 16 もし $D_k \neq \emptyset$ ならば $T \leftarrow T \cup \{D\}$ とし 8 へ戻る。
 - 17 $D_k \leftarrow NN_k \cup T, j \leftarrow j + 1$ とし 3 へ戻る。
-

4-2 パラメータの推定

D_k^r の各 2 対に対して, 文献 (柳, 2014) の方

法で δ を推定する。ただし, 文献 (柳, 2014) の方法とは異なり, k 個の初期クラスタを一度決定してしまえば変更することはなく, パラメータ探索と重み付きカーネル k -平均法の計算は各一回ずつ行う。そのため, パラメータの探索方向は正方向のみに限定することで計算回数 of 安定化を計る。さらに, 式 (21) を重み付きカーネル k -平均法の距離計算に適した式に変更する必要がある。まず, パラメータの探索区間について考える。初期クラスタに対して式 (18) を満たす δ が存在しない可能性もあるため, 事前に探索区間 $[d_{min}, d_{max}]$ を決めておき有限回の計算により δ を推定する。 d_{min} に関しては, 正方向のみの探索であるため, 十分小さい任意の正の実数 $\epsilon > 0$ を採用する。 d_{max} に関しては, Zelnik-Manor ら (Zelnik-Manor and Perona, 2004) が計算対象データの m 近傍点との距離を採用していること, および, 文献 (柳, 2014) の実験結果より, 2 対のクラスタ $C_i, C_j, 1 \leq i, j \leq k, i \neq j$ を覆う超球の半径の最大値をとれば十分であると推測する。したがって, 式 (71) のように区間 $[d_{min}, d_{max}]$ を決定する。

$$\begin{aligned} d_{min} &= \epsilon (> 0), \\ d_{max} &= \max_{\forall x, x' \in C_i \cup C_j, 1 \leq i, j \leq k, i \neq j} \frac{\sqrt{d(x, x')}}{2}. \quad (71) \end{aligned}$$

次に, パラメータの更新比率 ρ を文献 (柳, 2014) と同じ考え方に基づき式 (18) から導出すると式 (72) が得られる。

$$\rho_q = \sqrt{\frac{1}{1 - \frac{\|M_a^{(q-1)} - x_a\|^2 - \|M_b^{(q-1)} - x_a\|^2}{3}}}. \quad (72)$$

式 (72) における $M_a^{(q-1)}$ と $M_b^{(q-1)}$ は, $q-1$ 番目のカーネルパラメータ $\delta_{q-1} = \rho_{q-2} \delta_{q-2}$ を用いて M_a および M_b のカーネル関数を計算することを意味する。また, 式 (72) は探索幅の比率を定めるパラメータであり, 多少の誤差は許容できるため重み w_i の変化は無視した。以上より, δ を更新するためのアルゴリズムを Algorithm 3 に示す。

4-3 提案手法のアルゴリズム

前節より, Algorithm 2 で決定した初期クラスタの各 2 対に対して, まず式 (71) により探索区間 $[d_{min}, d_{max}]$ を決定した後, Algorithm

3によりパラメータ δ を推定する。

Algorithm 3 入力: $C_a, C_b, d_{\min}, d_{\max}$

- 1 繰返し番号を $q = 1$ とし, $\delta_q = d_{\min}$ とする。
 - 2 C_b に対して式 (12) を最小とする $x_a \in C_a$ を求める。
 - 3 C_a, C_b, x_a, δ_q より式 (72) で ρ_q を計算する。
 - 4 以下のいずれかであれば, δ_q を出力し計算を終了する。
 - 式 (18) と式 (19) を同時に満たす
 - $\rho_q = 1$
 - $\delta_q \geq d_{\max}$
 - 5 $\delta_q = \rho_q \delta_q, q = q + 1$ とし, 2へ戻る。
-

そこで, まず各 $\delta_{i,j} (1 \leq i, j \leq k, i \neq j)$ の統合に関して考える。全てのクラスタ C_1, C_2, \dots, C_k に対して, 式 (73) を満たす δ の集合 Δ が存在するならば, 各2対のクラスタ $C_i, C_j, 1 \leq i, j \leq k, i \neq j$ に対して式 (73) を満たす $\delta_{i,j}$ の集合 $\Delta_{i,j}$ が存在し $\Delta \subseteq \Delta_{i,j}$ を満たす。

$$\forall x \in C_i, i = \arg \min_{1 \leq i \leq k} \|M_i - x\|^2. \quad (73)$$

したがって, $\delta \in \Delta$ の最小値を δ_{\min} , 最大値を δ_{\max} , $\delta_{i,j} \in \Delta_{i,j}$ の最小値を $\delta_{i,j,\min}$, 最大値を $\delta_{i,j,\max}$ とすると, 式 (74) より, 式 (75) が成立する。

$$\Delta \subseteq \bigcap_{1 \leq i, j \leq k, i \neq j} \Delta_{i,j}. \quad (74)$$

$$\begin{aligned} \max_{1 \leq i, j \leq k, i \neq j} \delta_{i,j,\min} &\leq \delta_{\min}, \\ \delta_{\max} &\leq \min_{1 \leq i, j \leq k, i \neq j} \delta_{i,j,\max}. \end{aligned} \quad (75)$$

式 (75) より, 本論文では, Algorithm 2 により計算した各2対の初期クラスタ $C_i, C_j, 1 \leq i, j \leq k, i \neq j$ に対して, $\delta_{i,j} \in \Delta_{i,j}$ を計算し, その最大値を δ と決定する。ただし, Algorithm 2 で決定した初期クラスタに対して $\Delta = \emptyset$ の場合もあり, 必ずしも式 (18) を満たす δ が存在する保証はないため, その場合 $\delta = d_{\max}$ となる。特に, クラスタが分離していないようなデータに関しては, この傾向が強くなると考えられる。一般的にガウス関数のパラメータは距離のスケールを調整する役割があり, 式 (4) より, δ が小さいときは距離の変化に敏感で, δ の増加に伴い距離の変化に対して鈍感になっていく。つまり, δ の増加にしたがって

カーネル k -平均法は通常の k -平均法の性質に近くなっていくと考えられる。仮に, クラスタが分離していないデータであれば, 提案手法は初期クラスタに分割する手がかりがないため, むしろ, 通常の k -平均法の分割基準に委ねるのが妥当であると考ええる。一方で, 全てのクラスタが分離していないデータであれば, そのような基準は妥当であると考ええるが, クラスタの一部が分離していて, それ以外が分離していないような場合は, $k(k-1)/2$ 個の δ の一部が $d_{\min} < \delta < d_{\max}$ を満たし, それ以外が $\delta = d_{\max}$ となる可能性がある。この場合, 分離可能と思われる $d_{\min} < \delta < d_{\max}$ を満たしたもののの中から最大値を選択し δ の推定値とするのか, または, $\delta = d_{\max}$ とするかの判断が難しい。もし, $k(k-1)/2$ 個の δ のうちの $k-1$ 個以上が d_{\max} となっている場合は, 一つのクラスタに他のクラスタが結合してしまっている可能性が高いと考えられるので, 本論文では $k(k-1)/2$ 個の δ のうち $\delta \geq d_{\max}$ となる個数を n_δ とし, n_δ の値によって式 (76) の基準で最終的な δ を推定する。

$$\delta = \begin{cases} d_{\max} & (n_\delta \geq k-1) \\ \max_{1 \leq i, j \leq k, i \neq j} \delta_{i,j} & (\text{otherwise}) \end{cases}. \quad (76)$$

最終的な提案法のアルゴリズムを Algorithm 4 に示す。

Algorithm 4 入力: X, k, r

- 1 X, k, r を入力とし, Algorithm 2 により初期クラスタ C_1, C_2, \dots, C_k を求める。
 - 2 式 (71) より δ の探索区間 $[d_{\min}, d_{\max}]$ を決定する。
 - 3 全ての2対の初期クラスタ $C_i, C_j, (1 \leq i, j \leq k, i \neq j)$ に対して, $C_i, C_j, [d_{\min}, d_{\max}]$ を入力として, Algorithm 3 により $\delta_{i,j}$ を求める。
 - 4 $\delta_{i,j} (1 \leq i, j \leq k, i \neq j)$ と $[d_{\min}, d_{\max}]$ を用いて, 式 (76) より δ を決定する。
 - 5 $X, k, C_1, C_2, \dots, C_k, \delta$ を入力として Algorithm 1 を実行し, クラスタ C_1, C_2, \dots, C_k を再計算する。
 - 6 C_1, C_2, \dots, C_k を出力して計算を終了する。
-

4-4 計算量

Algorithm 1 の重み付きカーネル k -平均法の

計算量のオーダーは Algorithm 1 の繰返し数 t , クラスタ数 k , データ数 n , データの次元数 d を用いて $\mathcal{O}(tkn^3d)$ と表すことができる。Algorithm 2 の計算量のオーダーは $\mathcal{O}(dn^3 \log n)$, Algorithm 3 の計算量のオーダーは Algorithm 3 の繰返し数 q を q_p とすると, そのオーダーは $\mathcal{O}(q_p n^2 d)$ となる。したがって, Algorithm 4 の計算量のオーダーは $\mathcal{O}(dn^3 \log n + q_p n^2 d + tkn^3 d)$ となり, $\log n > tk$ とならない限りは重み付きカーネル k -平均法の計算オーダーとなる。一方で, 従来法の計算量のオーダーは, カーネル k -平均法を複数回実行する必要があるため, その繰返し数を l とすると $\mathcal{O}(lk^2 q_p n^2 d + lk^2 n^2 d(1 + ndk \log k) + ltkn^3 d)$ となり (柳, 2014), $l > \log n$ となる状況において, 提案法は従来法に比較して効率化できていると考えられる。

5. 実験

実験は, 教師ラベル付きの人工データと実データに対して行う。実験環境は, Intel Xeon 2.00 GHz 4 コア \times 2 のデュアルプロセッサ, および, 24GB RAM のワークステーションで, OS は 64 ビット版 Linux を使用した。Matlab v 8.5.0 上で提案手法, 文献 (柳, 2014) の手法, スペクトラルクラスタリング (以下スペクトラル法), そして, 重み付きカーネル k -平均法を実装した。また, Yu らの手法 (以下 OKKC と表記) に関しては著者らが提供している Matlab 用のスクリプト (Yu, Tranchevent, De Moor and Moreau, 2011) をそのまま利用した。

5-1 クラスタリング結果の評価方法

教師ラベル付きのデータのクラスタリング結果に対して評価を行う代表的な指標として, Adjusted Rand Index (ARI) (Hubert and Arabie, 1985) や Normalized Mutual Information (NMI) (Strehl and Ghosh, 2003) がある。本論文では NMI を評価指標として用いる。あらかじめ教師ラベルにより決定されている k 個のクラスタ集合を $C^t = \{C_1^t, C_2^t, \dots, C_k^t\}$, 実際にクラスタリングすることにより得られる k 個のクラスタ集合を $C = \{C_1, C_2, \dots, C_k\}$, そして, $n = |C_1^t \cup C_2^t \cup \dots \cup C_k^t| = |C_1 \cup C_2 \cup \dots \cup C_k|$ をデータ数とするとき, C^t と C のエントロピーは式 (77) で定義される。

$$\begin{aligned}
 H(C^t) &= - \sum_{j=1}^k \frac{|C_j^t|}{n} \log \frac{|C_j^t|}{n}, \\
 H(C) &= - \sum_{i=1}^k \frac{|C_i|}{n} \log \frac{|C_i|}{n}. \quad (77)
 \end{aligned}$$

以降, 本論文では C^t を正解クラスタ集合と呼ぶことにする。また, C^t と C の結合エントロピーは式 (78) で定義される。

$$H(C^t, C) = - \sum_{i=1}^k \sum_{j=1}^k \frac{|C_i \cap C_j^t|}{n} \log \frac{|C_i \cap C_j^t|}{n}. \quad (78)$$

NMI の計算式は式 (77) と式 (78) を用いて, 式 (79) により定義される。

$$\text{NMI} = \frac{H(C^t) + H(C) - H(C^t, C)}{\sqrt{H(C^t)H(C)}}. \quad (79)$$

NMI は 0 と 1 の間の値をとり, 正解クラスタと実際に得られたクラスタが完全に一致するとき, $H(C^t) = H(C) = H(C^t, C)$ より, 最大値の 1 をとる。

5-2 実験条件

まず, 提案手法と文献 (柳, 2014) のパラメータ推定における最大繰返し数を 300 とし 300 を越えても計算を終了しない場合は強制的に計算を終了するようにした。ガウスパラメータに関して, 提案手法の探索区間の左端点と, 従来手法の初期値は文献 (柳, 2014) 同様, ガウス関数の値が計算可能な有効桁以上となるように, Matlab における倍精度浮動小数型変数の正の最小値である 2.25074×10^{-308} , 計算対象データのデータ間の最大距離 D_{\max} を用いて式 (80) とした。

$$d_{\min} = \frac{D_{\max}}{\sqrt{308 \log_e 10 - 1}}. \quad (80)$$

提案手法の D_k^r のパラメータは $r = 3$ とした。文献 (柳, 2014) の従来手法に関しては, 2 つの新しいクラスタの任意のデータ同士が互いに k 近傍とはならないという条件があるが, $k > 1$ に関しては有効性を確認出来なかったため (柳, 2014), $k = 1$ とした。また, 全体の繰返し数 l の最大繰返し回数を 20 回とし, 20 回の繰返しを越えても計算を終了しない場合は強制的に計算を終了するようにした。スペクトラル法のカーネルパラメータに関しては, Zelnik-Manor らの手法を採用し, 近傍のレベ

ル m を $m = 7$ とした。重み付きカーネル k -平均法に関しては、パラメータを自動設定できないため、スペクトラル法と同様の方法を採用した。OKKC に関しては、事前に複数のカーネルパラメータを用意する必要があり、Yu らの実験と同様 c をデータの共分散行列の要素の平均とし、 $\{\frac{1}{4}c, \frac{1}{2}c, c, 2c, 3c, 4c, 5c, 6c, 7c, 8c\}$ を用いてカーネル行列を用意した。各手法における k -平均法、カーネル k -平均法、重み付きカーネル k -平均法の最大繰返し数を 100 回とし、100 回の繰返しを越えても計算を終了しない場合は強制的に計算を終了するようにした。

5-3 使用データの詳細

人工データは全て 2 次元とし、Kuncheva ら (Kuncheva and Vetrov, 2006) の実験を参考に分類精度の評価に適していると思われるものを 19 種類用意した。表 1 に各データの詳細を示し、図 1 ~ 19 に各データの分布状態を示す。

表 1 人工データの詳細

データ名	図番号	サンプル数	クラスタ数
データ1	図 1	400	2
データ2	図 2	300	3
データ3	図 3	400	2
データ4	図 4	600	3
データ5	図 5	600	3
データ6	図 6	601	2
データ7	図 7	800	4
データ8	図 8	600	2
データ9	図 9	400	2
データ10	図 10	400	2
データ11	図 11	600	3
データ12	図 12	400	2
データ13	図 13	400	2
データ14	図 14	600	3
データ15	図 15	400	4
データ16	図 16	900	9
データ17	図 17	900	9
データ18	図 18	500	5
データ19	図 19	500	5

実データに関しては UCI machine learning repository (Frank and Asuncion, 2010) より、iris, wine, ionosphere, glass, soybeans-small, satimage, pen digit の 7 つのデータを使用した。satimage と pen digit に関しては、元のデータのデータ数が非常に多いため、各クラスタからランダムに 80 個を選択した。各

データの詳細を表 2 に示す。

表 2 実データの詳細

データ名	次元数	サンプル数	クラスタ数
iris	4	150	3
wine	13	178	3
ionosphere	35	351	2
glass	9	214	6
soybean-small	35	47	4
satimage	36	480	6
pen digit	16	800	10

5-4 5 つの手法のクラスタリング精度の比較

5 つの手法を用いて各データに対して 30 回の試行を行い、計算時間の平均と最小と最大、および、NMI 値を調べた。表 3 に NMI 値の実験結果を示し、表 4 に計算時間の実験結果を示す。表 3 において、提案手法は 30 回の試行全てにおいて同じ値となったため、1 つの値のみ示している。これは、初期クラスタが全て同じ値となったためであり、Algorithm 2 が一意の結果を出力することを実験的に確認できた。また、重み付きカーネル k -平均法とスペクトラル法は、ガウスパラメータを固定しており、式 (10) および式 (8) の目的関数を最小とする結果を採用することができるため、30 回の試行において目的関数の最小値に対応する NMI 値、NMI 値の最小値と最大値、そして、目的関数が最小となった回数を調べた。表 3 では対応する手法の列において、括弧内が (最小, 最大) を示し、最後の列が回数を示している。従来法と OKKC に関しては、NMI 値の平均と最小と最大を調べた。表 3 の対応する手法の列において、括弧内が (最小, 最大) を示している。

データ 1, 3, 4, 5, 6, 7, 8, 10, 11, 12 は対応する図 1 ~ 図 12 の形状より、クラスタ数 k に関して $disjoint(X, k)$ と推測できる。これらのデータは表 3 より、提案法の NMI が全て 1 となっており、データ集合 X の分布形状やクラスタ数が異なっても $disjoint(X, k)$ を満たしていれば、提案法は従来法と比較して安定して正解クラスタを見つけることができると考えられる。上記のデータに対しては重み付きカーネル k -平均法とスペクトラル法はデータ 3, 5, 6, 7, 12 の結果がやや悪い。データ 3, 5, 7, 12 は初期クラスタへの依存が大きい形状であるためと考えられる。データ 6 に関しては、二つのクラスタ

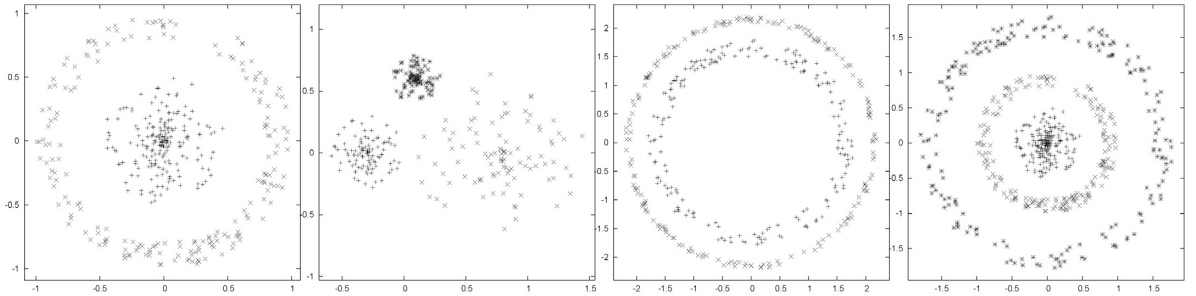


図 1 データ 1

図 2 データ 2

図 3 データ 3

図 4 データ 4

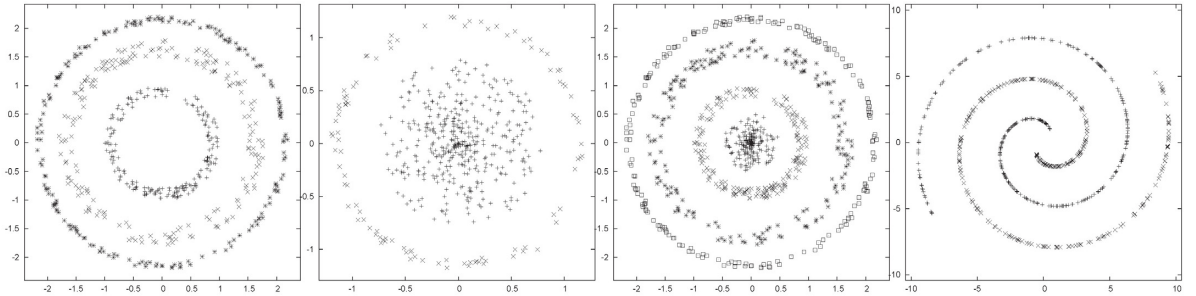


図 5 データ 5

図 6 データ 6

図 7 データ 7

図 8 データ 8

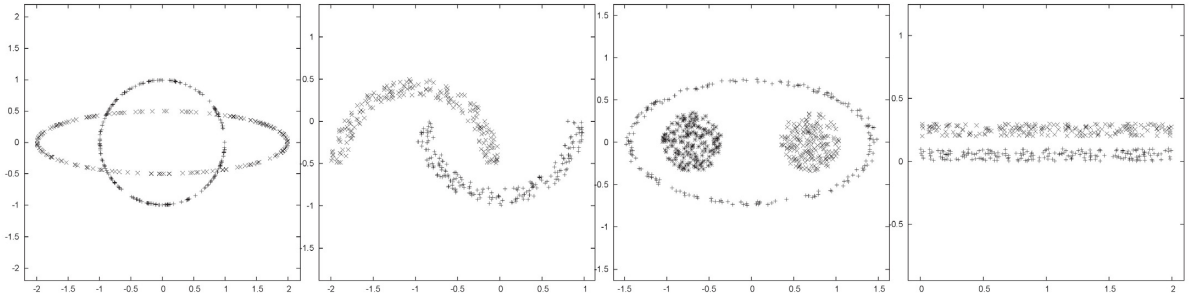


図 9 データ 9

図 10 データ 10

図 11 データ 11

図 12 データ 12

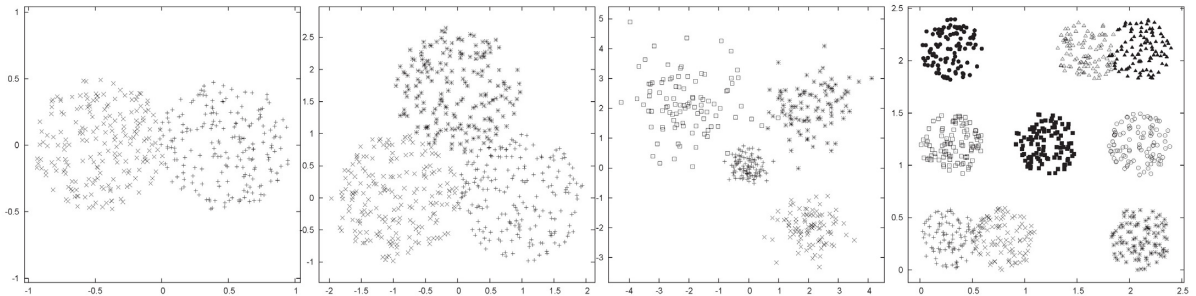


図 13 データ 13

図 14 データ 14

図 15 データ 15

図 16 データ 16

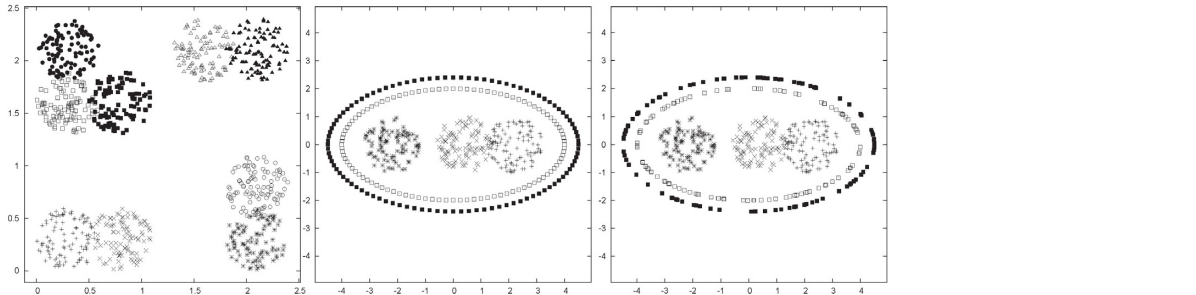


図 17 データ 17

図 18 データ 18

図 19 データ 19

数が異なり、ガウスパラメータを固定しているため、うまく分離できなかったと考えられる。また、計算時間については表 4 より、若干の入れかわりはあるものの、概ねスペクト

ラル法, 提案法, 従来法, 重み付きカーネル k -平均法, OKKC の順になっている。スペクトラル法はデータのサイズ n に対して 3 乗の計算オーダーとなるものの、MATLAB 上の行列

表3 各データに対する5つの手法の比較

データ	提案法	従来法	OKKC	重み付き	スペクトラル
1	1.000	1.000 (1.000, 1.000)	0.250 (0.013, 0.684)	1.000 (0.430, 1.000)	27 1.000 (1.000, 1.000) 30
2	1.000	0.979 (0.653, 1.000)	1.000 (1.000, 1.000)	1.000 (0.846, 1.000)	27 1.000 (1.000, 1.000) 30
3	1.000	1.000 (1.000, 1.000)	2.1e-3 (1.8e-5, 4.1e-3)	0.000 (0.000, 0.016)	1 8.9e-4 (8.9e-4, 8.9e-4) 30
4	1.000	0.916 (0.221, 1.000)	0.504 (0.361, 0.597)	1.000 (0.768, 1.000)	29 1.000 (1.000, 1.000) 30
5	1.000	0.620 (0.353, 1.000)	0.067 (7.4e-3, 0.289)	0.579 (0.341, 0.590)	3 0.653 (0.653, 0.653) 30
6	1.000	0.053 (0.053, 0.053)	0.079 (1.4e-3, 0.218)	0.505 (0.032, 1.000)	9 0.454 (0.454, 0.454) 30
7	1.000	0.661 (0.359, 1.000)	0.357 (0.147, 0.533)	0.804 (0.605, 0.804)	1 0.804 (0.630, 0.804) 26
8	1.000	1.000 (1.000, 1.000)	0.032 (9.8e-4, 0.065)	1.000 (0.350, 1.000)	28 1.000 (1.000, 1.000) 30
9	2.2e-3	0.024 (0.022, 0.024)	0.030 (7.3e-5, 0.089)	1.2e-3 (1.8e-5, 0.262)	11 8.8e-4 (8.8e-4, 8.8e-4) 30
10	1.000	1.000 (1.000, 1.000)	0.158 (0.014, 0.371)	1.000 (1.000, 1.000)	30 1.000 (1.000, 1.000) 30
11	1.000	1.000 (1.000, 1.000)	0.463 (0.169, 0.567)	1.000 (1.000, 1.000)	30 1.000 (1.000, 1.000) 30
12	1.000	1.000 (1.000, 1.000)	4.3e-4 (2.2e-16, 1.7e-3)	0.162 (0.011, 1.000)	15 2.9e-4 (2.9e-4, 2.9e-4) 30
13	0.915	0.016 (0.016, 0.016)	0.427 (8.9e-4, 0.902)	0.567 (0.426, 0.919)	9 0.902 (0.902, 0.902) 30
14	0.944	0.506 (0.140, 0.861)	0.488 (0.406, 0.879)	0.845 (0.471, 0.899)	4 0.902 (0.902, 0.902) 30
15	0.951	0.743 (0.489, 0.951)	0.823 (0.721, 0.951)	0.951 (0.852, 0.951)	28 0.958 (0.958, 0.958) 30
16	0.971	0.912 (0.842, 0.972)	0.736 (0.442, 0.961)	0.961 (0.933, 0.968)	18 0.974 (0.914, 0.974) 9
17	0.888	0.857 (0.774, 0.941)	0.679 (0.373, 0.921)	0.901 (0.837, 0.912)	8 0.927 (0.872, 0.927) 22
18	0.953	0.477 (0.360, 0.830)	0.487 (0.373, 0.620)	0.776 (0.737, 0.787)	10 0.646 (0.646, 0.696) 26
19	0.686	0.677 (0.412, 0.807)	0.457 (0.361, 0.578)	0.778 (0.620, 0.798)	5 0.664 (0.560, 0.706) 26
sat.	0.518	0.139 (0.050, 0.365)	0.579 (0.520, 0.628)	0.569 (0.543, 0.663)	1 0.624 (0.524, 0.624) 7
pen.	0.777	0.216 (0.066, 0.461)	0.682 (0.641, 0.724)	0.794 (0.726, 0.852)	1 0.765 (0.738, 0.788) 12
iris	0.798	0.662 (0.050, 0.758)	0.474 (0.222, 0.798)	0.798 (0.722, 0.885)	18 0.786 (0.786, 0.786) 30
wine	0.432	0.091 (0.091, 0.091)	0.394 (0.210, 0.430)	0.395 (0.232, 0.429)	15 0.420 (0.420, 0.420) 30
glass	0.418	0.344 (0.120, 0.443)	0.116 (0.103, 0.123)	0.422 (0.275, 0.428)	1 0.404 (0.303, 0.414) 2
iono.	0.018	0.026 (0.026, 0.026)	0.125 (0.081, 0.198)	0.136 (0.103, 0.182)	5 0.221 (0.221, 0.221) 30
soy.	0.848	0.652 (0.543, 0.848)	0.301 (0.206, 0.365)	0.716 (0.716, 1.000)	12 0.723 (0.723, 0.723) 30

演算が高速であることが大きな理由と考えられる。また、重み付きカーネル k -平均法はC言語で実装された Graclus が公開されており (Dhillon, Guan and Kulis, 2007), 入力である行列の疎行列化が必要ではあるものの、かなり高速であるため、全てC言語で実装した場合は必ずしもこの順位にはならない可能性がある。C言語またはC++言語での実装による提案手法と Graclus およびスペクトラル法との計算時間に関する厳密な比較は今後の課題としたい。OKKC に関しては表3のデータのほとんどの結果が他の手法より悪く、また、表4より、計算時間も突出して長くなっている。計算時間に関してはOKKCにおける半無限計画問題の計算の不安定さに原因があると考えられる。実際に半無限計画問題の計算過程において収束せずに計算を破棄する事が多々あったため、実験では半無限計画問題の計算で収束しなかったときは、そこで全体の計算を破棄するのではなく、収束しなかった計算過程のみ再計算して次の計算過程へ進むようにした。

各データの再計算数は多くても4~5回程度であった。実際に計算する際に結果を得るためには再計算を考慮しなくてはならないことから、再計算の時間も比較対象とすることは公平性の意味で妥当であると考えられる。

データ9は二つの環状のクラスタが重なっている土星型の分布のデータであり、全ての手法のNMI値が低い。OKKC以外の手法の平均または目的関数を最小とするNMI値は $10^{-3} \sim 10^{-4}$ のオーダーであり、ほぼ0とみなせる。OKKCのみ平均が、かろうじて0.03と相対的に高くなっているものの、最大値は0.1程度であり良好な結果とはいえない。重み付きカーネル k -平均法およびスペクトラル法で正解クラスタを得るのは困難であると考えられる。

データ2とデータ15は各クラスタの分散が異なり、特にデータ2は図2における右下のクラスタの右上部分にあるデータ間の距離が最大となるため、そのデータがいわゆる外れ値となり、正解クラスタが $disjoint(X, 3)$ とはな

表4 各データに対する5つの手法の計算時間(秒)の比較

	提案法	従来法	OKKC	重み付き	スペクトラル
1	0.55 (0.53, 0.59)	0.48 (0.45, 0.51)	183.8 (33.81, 537.8)	2.65 (2.59, 2.73)	0.37 (0.36, 0.38)
2	0.39 (0.37, 0.47)	0.45 (0.28, 0.71)	132.9 (37.45, 469.8)	1.81 (1.71, 2.07)	0.22 (0.22, 0.23)
3	0.86 (0.84, 0.89)	0.66 (0.54, 0.77)	70.24 (29.20, 151.2)	2.97 (2.71, 3.41)	0.38 (0.37, 0.39)
4	1.41 (1.34, 1.46)	5.41 (1.18, 19.66)	502.2 (64.38, 1268.4)	5.43 (5.27, 5.56)	0.85 (0.84, 0.90)
5	1.85 (1.78, 1.91)	13.06 (2.50, 16.99)	364.6 (92.62, 1021.9)	5.78 (5.42, 6.59)	0.86 (0.85, 0.90)
6	0.79 (0.76, 0.82)	5.88 (5.70, 6.15)	206.1 (56.25, 580.4)	4.20 (3.93, 5.69)	0.61 (0.59, 0.64)
7	3.34 (3.06, 3.44)	24.39 (3.43, 38.09)	1226.8 (210.8, 2166.3)	9.74 (9.27, 10.53)	1.56 (1.54, 1.57)
8	1.81 (1.78, 1.86)	1.59 (1.07, 3.28)	208.3 (101.2, 391.7)	5.34 (5.21, 5.51)	0.85 (0.84, 0.87)
9	0.69 (0.67, 0.73)	2.16 (2.11, 2.20)	106.6 (52.39, 202.1)	2.84 (2.59, 3.24)	0.37 (0.36, 0.39)
10	0.74 (0.71, 0.78)	0.49 (0.47, 0.54)	96.56 (27.23, 208.2)	2.67 (2.55, 2.76)	0.37 (0.36, 0.38)
11	2.00 (1.91, 2.09)	1.59 (1.04, 2.71)	135.0 (63.01, 297.6)	5.35 (5.23, 5.46)	0.87 (0.85, 0.89)
12	0.84 (0.82, 0.86)	0.56 (0.55, 0.58)	128.9 (32.70, 416.5)	2.68 (2.59, 2.80)	0.37 (0.37, 0.38)
13	1.20 (1.14, 1.27)	0.42 (0.41, 0.44)	72.88 (32.52, 280.8)	3.29 (2.71, 4.04)	0.37 (0.36, 0.39)
14	5.28 (5.15, 5.42)	5.26 (1.16, 8.37)	148.8 (65.29, 430.9)	7.25 (5.98, 9.48)	0.85 (0.83, 0.88)
15	1.26 (1.19, 1.38)	2.46 (0.61, 4.76)	157.8 (48.49, 467.9)	3.15 (2.74, 4.47)	0.37 (0.36, 0.38)
16	15.61 (14.79, 17.64)	52.91 (10.89, 59.05)	3136.7 (246.2, 8242.6)	15.36 (14.50, 16.87)	2.05 (2.03, 2.12)
17	28.09 (21.91, 37.40)	59.68 (55.68, 63.63)	2376.2 (354.8, 4358.3)	17.07 (14.77, 18.92)	2.06 (2.04, 2.08)
18	1.39 (1.23, 1.57)	7.60 (1.79, 17.61)	432.8 (49.70, 1914.1)	4.77 (4.27, 5.54)	0.52 (0.47, 0.60)
19	1.43 (1.37, 1.52)	11.72 (4.05, 14.33)	607.0 (71.92, 1303.4)	4.65 (4.19, 5.44)	0.51 (0.42, 0.68)
sat.	7.01 (6.72, 7.38)	2.34 (1.37, 6.46)	232.3 (59.75, 409.3)	12.82 (10.35, 15.92)	0.75 (0.73, 0.84)
pen.	6.90 (5.82, 7.90)	10.76 (5.10, 20.95)	160.8 (71.20, 310.3)	18.77 (15.92, 23.84)	1.83 (1.80, 1.87)
iris	0.17 (0.15, 0.19)	0.42 (0.10, 0.78)	22.07 (3.26, 99.3)	0.95 (0.76, 1.14)	0.07 (0.07, 0.08)
wine	0.24 (0.22, 0.28)	1.22 (1.17, 1.25)	101.6 (24.04, 221.0)	0.93 (0.88, 1.08)	0.10 (0.10, 0.10)
glass	0.86 (0.61, 0.99)	0.64 (0.35, 1.13)	5.05 (1.45, 18.15)	3.33 (2.27, 4.24)	0.14 (0.13, 0.16)
iono.	1.43 (1.27, 1.59)	0.43 (0.39, 0.50)	54.93 (17.13, 158.4)	4.19 (2.93, 5.85)	0.39 (0.38, 0.42)
soy.	0.05 (0.05, 0.05)	0.07 (0.05, 0.11)	0.52 (0.31, 0.98)	0.44 (0.34, 0.70)	0.02 (0.02, 0.02)

らなないと考えられる。提案手法は D_k を D_k^r に緩和した効果により正解クラスタを得ている。データ15はクラスタ同士の分布が重なっているため提案手法で完全に分離する初期クラスタを得ることは困難であるが、 δ の範囲の決め方とk-平均法における球状のクラスタの境界は中心点を二等分する超平面となる特性により、スペクトラル法の結果には若干劣るものの良好な結果を得ている。

データ13とデータ14は各クラスタの分布が完全に重なっており、完全に分離はできない。データ13のNMI値に関しては提案手法の結果が最も良く、データ14に関しては一番高評価であったスペクトラル法の結果にはやや劣るもののほぼ同等の結果が得られている。ただし、表4より、提案手法の計算時間がスペクトラル法と比較して長くなっている。これは、データ13とデータ14の初期クラスタに関して各中心点を垂直2等分するような境界で分離できず、結果としてAlgorithm 3では、初期クラスタ間のパラメータを最大値まで探

索することになったため、その計算時間に依存していると考えられる。

データ16とデータ17はクラスタ数が共に9であり、クラスタ全てではなく、部分的に分布が重なっているデータ形状である。図16より、データ16は2対のクラスタが2種類、合計4つのクラスタの分布が重なっている。分布が重なっているクラスタ数が半数以下であるため、NMI値に関してはスペクトラル法には劣るもののほぼ同等の値で良好な結果となっている。計算時間はスペクトラル法に比較してかなり長いですが、一方でスペクトラル法において目的関数が最小となった回数が1/3以下となっているため、精度と計算時間を総合して考慮すると著しく劣っているわけではないと考える。また、図17より、データ17は全てのクラスタがいずれかのクラスタと分布が重なっており、4つのデータ集合の塊に分離しているように見える形状である。提案手法のNMI値はスペクトラル法と重み付きカーネルk-平均法より小さな値となっており、提案手法が苦手とする

データ形状であると考えられる。表4より、計算時間に関しても、データ14と同じ理由により、スペクトラル法および重み付きカーネル k -平均法と比較してかなり長くなっている。

データ18とデータ19は、図18および図19より、球状と環状のクラスタが混合しているデータ分布であり、内側の3つのクラスタのうち2つの分布が重なっている。また、2つのデータ集合は、外側の環状のクラスタの密度が異なっており、データ18が密集しているのに対して、データ19は、やや散在した分布となっている。データ18に関しては提案手法のNMI値が最も大きかったのに対して、データ19では重み付きカーネル k -平均法よりもかなり低い値となっている。また、重み付きカーネル k -平均法とスペクトラル法のデータ18とデータ19におけるNMI値が僅差であるのに対して、提案手法はかなり離れている。これは、外側の2つのクラスタの分布の密度に原因があると考えられる。データ18では外側のクラスタのデータ同士が互いに最近傍点となっており、Algorithm 2において、比較的早めにクラスタを形成した後、各クラスタとの距離が十分離れているため、そのまま結合することなく最終的なクラスタとなったのに対して、データ19は外側の2つのクラスタの結合速度が遅く、外側の2つのクラスタの部分集合同士が互いに結合してしまったと考えられる。正解クラスタ内の任意の2点間の距離が他のクラスタとの距離より大きいデータ集合の分布は、データ17と同様、提案手法が苦手とする形状であると考えられる。

実データのNMI値に関しては、表3より、satimageは2番目に低い値であったが極端な差はなかった。pen digitとglassに関しては、重み付きカーネル k -平均法の結果に次いで2位であったが、重み付きカーネル k -平均法において目的関数が最小となった回数が30回中1回であることを考慮するとNMI値の値がそれほど離れていないことからほぼ同等と見なせる。irisに関しては重み付きカーネル k -平均法と同じ値で共に最も高評価であった。wineとsoybeansに関しては、提案手法の評価が最も高かった。ionosphereに関しては最も結果が芳しくなく、下から2番目に低い評価結果であった。初期クラスタの構成が正解クラスタと大きく異なっていたためと考えられるが、最も評価が高かったスペクトラル法のNMI値

が0.22程度であることから、スペクトラル法の分類の妥当性にもやや疑問が残る。総合的にみて、提案手法は、ionosphere以外は概ね良好な結果であり、 $disjoint(X, k)$ を基準にパラメータを決定しているものの、 $disjoint(X, k)$ を満たしていないであろう実データにも適用可能であると考えられる。

6. むすび

本論文では、文献(柳, 2014)の手法を基に、データ集合間の距離を基準に初期クラスタを構成して、重み付きカーネル k -平均法のガウスカーネルの適切なパラメータを探索するアルゴリズムを提案した。いくつかの人工データと実データを用いて、4つの手法との比較実験を行い、クラスタ間が十分な距離で分離しているデータに関してはクラスタ数や形状に依存せず提案手法が有効であることを示した。一方で、クラスタが部分的に分離していたり内部が密集していないクラスタが混在しているデータに関しては有効性を確認できなかった。今後の課題として、上記の分布形状を持つデータに対する基準の検討がある。

参考文献

- [1] Chan, Pak K, Martine DF Schlag, and Jason Y Zien (1994) "Spectral k-way ratio-cut partitioning and clustering," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, Vol. 13, No. 9, pp. 1088–1096.
- [2] Dhillon, Inderjit S, Yuqiang Guan, and Jacob Kogan (2002) "Iterative clustering of high dimensional text data augmented by local search," in *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pp. 131–138, IEEE.
- [3] Dhillon, Inderjit S, Yuqiang Guan, and Brian Kulis (2004) "A unified view of kernel k-means, spectral clustering and graph cuts," *University of Texas at Austin, Tech. Rep.*, Vol. TR-04-25.
- [4] Dhillon, I.S., Y. Guan, and B. Kulis (2004) "Kernel k-means: spectral clus-

- tering and normalized cuts,” in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 551–556, ACM.
- [5] Dhillon, I.S., Y. Guan, and B. Kulis (2007) “Weighted graph cuts without eigenvectors a multilevel approach,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 29, No. 11, pp. 1944–1957.
- [6] Frank, A. and A. Asuncion (2010) “UCI Machine Learning Repository,” <http://archive.ics.uci.edu/ml>.
- [7] Girolami, M. (2002) “Mercer kernel-based clustering in feature space,” *Neural Networks, IEEE Transactions on*, Vol. 13, No. 3, pp. 780–784.
- [8] Hubert, L. and P. Arabie (1985) “Comparing partitions,” *Journal of classification*, Vol. 2, No. 1, pp. 193–218.
- [9] Karypis, George and Vipin Kumar (1998) “A fast and high quality multilevel scheme for partitioning irregular graphs,” *SIAM Journal on scientific Computing*, Vol. 20, No. 1, pp. 359–392.
- [10] Kuncheva, L.I. and D.P. Vetrov (2006) “Evaluation of stability of k -means cluster ensembles with respect to random initialization,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 28, No. 11, pp. 1798–1808.
- [11] MacQueen, J. (1967) “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.
- [12] Máté, L. (1989) *Hilbert space methods in science and engineering*: Hilger.
- [13] Mika, S., G. Ratsch, J. Weston, B. Scholkopf, and KR Mullers (1999) “Fisher discriminant analysis with kernels,” in *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, pp. 41–48, IEEE.
- [14] Muller, K.R., S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf (2001) “An introduction to kernel-based learning algorithms,” *Neural Networks, IEEE Transactions on*, Vol. 12, No. 2, pp. 181–201.
- [15] Ng, A.Y., M.I. Jordan, and Y. Weiss (2002) “On spectral clustering: Analysis and an algorithm,” *Advances in neural information processing systems*, Vol. 2, pp. 849–856.
- [16] Shi, J. and J. Malik (2000) “Normalized cuts and image segmentation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 22, No. 8, pp. 888–905.
- [17] Strehl, A. and J. Ghosh (2003) “Cluster ensembles—a knowledge reuse framework for combining multiple partitions,” *The Journal of Machine Learning Research*, Vol. 3, pp. 583–617.
- [18] Suykens, Johan AK, Tony Van Gestel, Bart De Moor, and Joos Vandewalle (2002) “Basic Methods of Least Squares Support Vector Machines,” in *Least Squares Support Vector Machines*: World Scientific.
- [19] Yu, S., L.C. Tranchevent, B. De Moor, and Y. Moreau (2011) “Optimized data fusion for kernel k -means clustering,” *Kernel-based Data Fusion for Machine Learning*, pp. 89–107.
- [20] Yu, Stella X and Jianbo Shi (2003) “Multiclass spectral clustering,” in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 313–319, IEEE.
- [21] Zelnik-Manor, Lihi and Pietro Perona (2004) “Self-tuning spectral clustering,” in *Advances in neural information processing systems*, pp. 1601–1608.
- [22] 柳信一 (2014) 「距離の分布に基づくカーネル k -平均法のパラメータ探索法」, 『北海道情報大学紀要』, 第26巻, 第1号, pp.87–107, 11月。