

第2章 ファジィc-平均クラスタリング と距離空間

2.1 代表的なクラスタリング手法

クラスタリング (clustering) はクラスター分析 (cluster analysis) とも呼ばれ [2, 33], データ解析の諸技法のなかで, 外的基準なしに自動的に分類を行う方法, いいかえれば, データ以外にあらかじめ基準を設定することなく, データの集まりをいくつかのグループに分ける方法のことである. クラスタリングは, パターン認識の分野では, 教師なし分類 (unsupervised classification) と呼ばれる. 外的基準がないため, 分類すべき個体 (データ) の間に定義された類似性や距離にもとづいて, グループ分けを行う. クラスタリングの方法は次のように規定される [47].

“有限個の対象からなる集合が与えられ, かつ, 任意の二つの対象の対をとったとき, それらがどの程度の関連をもつかを示す数値が与えられていると仮定する. クラスタ分析, あるいはクラスタリングとは, この数値に基づいて, 与えられた集合をいくつかのクラス (部分集合) に分割し, 同一クラスの中の対象は互いに関連の度合いが大きく, 異なるクラスにはいる対象は関連が少ないようにする方法を意味する. このクラスをクラスターと呼ぶ. ここで, 分割するクラスの数, はあらかじめ与えられていることもあり, 与えられていないこともある. 後者の場合, 幾つかの異なる数のクラスの場合を想定して, 幾通りかの分割を行う.”

ここで, 任意の二つの対象の対をとったとき, それらがどの程度の関連をもつかを示す数値を類似性測度 (similarity measure) あるいは類似度と呼び, 個体を2つ与えたとき, その対に対して決まる実数である. その値が大きければ大きいほど, 2つの個体は互いに類似していると仮定される. 逆に, 非類似性測度 (dissimilarity measure) あるいは非類似度とは, 個体を2つ与えたとき, その対に対して決まる実数であり, その値が小さければ小さいほど, 2つの個体は互いに類似しているとされる. 非類似性を表す測度が与えられた場合は, 同じクラスにおいては非類似度が

小さく、異なるクラスについては非類似度が大きくなるようにしなければならない。一般には、非類似度という言葉の代りに我々が使いなれた距離 (distance) という用語を用いることも多い。

さて、クラスタリングは、階層的な手法と非階層的な手法に二分される。階層的な手法では、クラスタリングの結果、グループが形成される様子を示した樹形図と呼ばれる階層構造が出力され、階層構造におけるある階層を見ることにより、クラスター分割が得られる。観察する階層によって得られるクラスターの数が異なる。また、非階層クラスタリングでは、はじめにクラスターの数を与え、個体間の距離に基づいてクラスターが得られる。

階層的なクラスタリング手法

階層的な手法では、各々の個体間の類似度が与えられていると仮定する。この類似度を用いて、次のようなステップにより、クラスタリングを行う。

Step1 initialize: すべての個体を単一の要素からなる初期クラスターとみなす。

Step2 merge: 類似度の最も大きいクラスターの対を結合 (merge) してより大きいクラスターをつくる。

Step3 update: すべてのクラスターが結合され、全体が一つのクラスターになれば終り。そうでなければ Step2 で結合されたクラスターとほかのクラスター間の類似度を再計算 (update) して Step2 に戻る。

このアルゴリズムでは、Step1 において各々の個体からなるクラスターがまず生成され、Step2, Step3 において類似度の大きい対の結合を繰り返す。アルゴリズムを行った結果、クラスターとクラスターが結合した順番とその結合レベル (結合したときの類似度) が得られる。このクラスターが結合した順番とその結合レベルを表現するために樹形図が用いられる。樹形図では、クラスターが結合する段階が木構造として表現される。ある結合レベルで樹形図をみると、クラスター (分類) が得られる。階層クラスタリングでは、最初の段階ですべての個体間の類似度の計算が必要であること、個体数が多いと樹形図の表現が困難であることから大きいサイズのデータに適用するのは困難とされている。

非階層的で中心を用いないクラスタリング手法

さて、本論文では非階層的なクラスタリング手法を用いている。非階層的なクラスタリング手法では、一般的にある評価関数を定義しその評価関数を最適化することにより、クラスタリングが行われる。非階層的なクラスタリング手法には、クラスター中心を用いる手法とクラスター中心を用いない手法がある。クラスター中心を用いない手法の1つとして次のような手法があげられる。

個体の集合 X が $X = \{x_1, x_2, \dots, x_n\}$ で与えられるとする。クラスターは、 X を分割した部分集合で与えられ、 C_1, C_2, \dots, C_K と表現する。

$$\bigcup_{k=1}^K C_k = X, \quad C_k \cap C_l = \emptyset, \quad (k \neq l).$$

個体 x_i と個体 x_j の間の非類似度 $d(x_i, x_j)$ が与えられていると仮定する。この中心を用いない最適クラスタリングでは、目的関数 $F(C_1, C_2, \dots, C_K)$ が、すべての可能な分割の組み合わせに関して最適化される。目的関数には以下のような関数を用いる。

$$F(C_1, C_2, \dots, C_K) = \sum_{k=1}^K \frac{1}{|C_k|^\alpha} \left\{ \sum_{\substack{i \neq j \\ x_i, x_j \in C_k}} d(x_i, x_j) \right\}.$$

ここで、 $|C_k|$ はクラスター C_k に含まれる個体数であり、 α は非負のパラメータである。クラスターをすべてまとめて、 $C = \{C_1, C_2, \dots, C_K\}$ のように C で表現することもある。

この目的関数は、クラスター分割の組み合わせを変数とする関数で、最適解を求めるには計算量が多く困難である。そこで、目的関数を最適化(最小化)するために、ローカルサーチ、多スタートローカルサーチ、遺伝アルゴリズム、タブーサーチ等の探索手法を用いて近似解を求める手法が用いられる。

このクラスタリング手法によるクラスタリング結果は、パラメータ α の値によって大きく異なる。目的関数に含まれるパラメータ α の好ましい値を求めるために、クラスタリングの結果得られた分割に対する評価規範が提案されている [58]。ここでは、得られたクラスタリング結果に対する新たな評価規範を提案し、パラメータ α の値と新たに導入した評価規範の値の関係から、パラメータ α の適切な値について考察している。

ここまでは、本論文では直接は扱っていない階層的クラスタリングや、中心を用いない非階層クラスタリング手法を紹介した。以降、本論文で扱う非階層でかつ中

心を用いるクラスタリング手法について説明する。

2.2 c-平均法

c-平均法 (c-means) は, K -平均法 (K -means) と呼ばれ, 非階層クラスタリングの代表的な手法の一つである. この方法は, MacQueen によって K -means と呼ばれていた. MacQueen [40] によれば,

K -means の手続きをインフォーマルに述べれば, 次のようになる. K 個ランダムに個体を選び, それぞれのグループの代表とする. ほかの個体を1つずつ選び, 最も近い平均値をもつグループに割り当てる. 割り当てられたグループについて, 平均値を更新する. 各ステージにおいて, K 個の平均値がグループを代表する. (したがって K -means と呼ぶ).

MacQueen はこれを K -means プロセスと呼んでいる. MacQueen のこのプロセスをもとにした K -means のクラスタリングのプログラムは, クラスタ数を変更するパラメータなどが含まれ, 以下で述べる c-平均法のアルゴリズムよりも複雑な手続きとなっている [40].

分類すべきデータ対象を p 次元ユークリッド空間上に存在するものと考え, データの各個体 (p 次元空間上の点) は $X = \{x_1, x_2, \dots, x_n\}$,

$$x_k = (x_k^1, x_k^2, \dots, x_k^p)^T$$

と表される. これらの点を c 個に分類する. 分類されたグループには中心点がそれぞれ存在し, $V = \{v_1, v_2, \dots, v_c\}$,

$$v_i = (v_i^1, v_i^2, \dots, v_i^p)^T$$

で表されるとする. つまり n 個の点からなる集合 X を, それぞれがクラスター中心 v_i をもつ c 個のクラスターに分類する.

クリスプな (ファジィでない) クラスタリングにおいては, 各々の点は必ずどれか一つのグループに属さなければならない. これを言い換えると次のようになる. 個体 x_k がクラスター i に所属するか否かを 2 値変数 $U = \{u_{ik}\}$ によって

$$u_{ik} = \begin{cases} 1 & (x_k \text{ がクラスター } i \text{ に所属する}) \\ 0 & (x_k \text{ がクラスター } i \text{ に所属しない}) \end{cases}$$

と決める. この変数 u_{ik} はメンバシップと呼ばれる. ただしすべての個体は, ただ一つのクラスターに所属するので,

$$\sum_{i=1}^c u_{ik} = 1$$

が成立する.

このような条件の下で, 目的関数 J_1 を最小にすることにより, クラスタリングを行う.

$$J_1(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} \|x_k - v_i\|^2.$$

ここで $\|\cdot\|^2$ は, ユークリッド距離の2乗を表す. このような, ある関数を最適化することによりクラスタリングを行う方法を, 最適クラスタリングと呼ぶ.

個体 x_k がクラスター i に所属するか否かを表す2値変数 u_{ik} の許容集合 M_c についてまとめると次のよう表される.

$$M_c = \{(u_{ik}) \mid \sum_{i=1}^c u_{ik} = 1, u_{ik} \in \{0, 1\}, k = 1, 2, \dots, n\}.$$

上の目的関数 J_1 の最適化 (最小化) を (U, V) について一度に最適化するのは困難である. そこでメンバシップ U を固定した場合に目的関数 J_1 をクラスター中心 V について最小化するという段階と, クラスタ中心 V を固定して目的関数 J_1 をメンバシップ U について最小化するという2つの段階に分けて行うことにする. この手順を示した c -平均法と呼ばれる2段階アルゴリズムを紹介する. このアルゴリズムは Anderberg[2] によって Forgy のアルゴリズムとして述べられている [19].

c -平均法アルゴリズム

CM1. \bar{U} と \bar{V} の初期値を適当に決める.

CM2. $\min_{U \in M_c} J_1(U, \bar{V})$ の最適解 \hat{U} を求め, $\bar{U} = \hat{U}$ と置き換える.

CM3. $\min_V J_1(\bar{U}, V)$ の最適解 \hat{V} を求め, $\bar{V} = \hat{V}$ と置き換える.

CM4. 最適解 (\bar{U}, \bar{V}) が収束していれば終了, そうでなければステップ CM2 に戻る.

CM2 と CM3 の最適化は, 実際には次のようにして行われる. CM2 では,

$$u_{ik} = \begin{cases} 1 & \iff v_i = \arg \min_q \|x_k - v_q\|^2 \\ 0 & \iff \text{otherwise} \end{cases}$$

で最適化しており, $\|x_k - v_q\|^2$ が最小となるクラスターに対するメンバシップ値 u_{ik} を1とし, 他のクラスターに対するメンバシップ値を0にしている. いいかえると $\|x_k - v_q\|^2$ が最小となるクラスターに個体 x_k を割り当てている.

また, CM3 の最適解 \hat{V} を求めるステップでは, v_i に関する最適化を行っている. 目的関数 J_1 のユークリッド距離の2乗の部分をそれぞれの成分に分解して表示にすると

$$\|x_k - v_i\|^2 = \sum_{j=1}^p (x_k^j - v_i^j)^2$$

であるので, 目的関数は

$$J_1(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} \sum_{j=1}^p (x_k^j - v_i^j)^2$$

と書き表せる. 目的関数 J_1 をクラスター中心 v_i に関して最適化する. ところで, J_1 は $v_i^j (j = 1, \dots, p)$ に関する2次関数なので, 目的関数 J_1 のクラスター中心 v_i の j 次元成分 v_i^j に関する偏導関数を0とおくことにより最適解が求められる.

$$\begin{aligned} -\frac{1}{2} \frac{\partial J_1}{\partial v_i^j} &= \sum_{k=1}^n u_{ik} (x_k^j - v_i^j) \\ &= \sum_{k=1}^n u_{ik} x_k^j - v_i^j \sum_{k=1}^n u_{ik} = 0 \end{aligned}$$

より,

$$v_i^j = \frac{\sum_{k=1}^n u_{ik} x_k^j}{\sum_{k=1}^n u_{ik}}$$

となる.

求められた最適解 v_i^j に注目する. u_{ik} は $\{0, 1\}$ の2値変数であり, $u_{ik} = 1$ とは, 個体 x_k がクラスター v_i に所属していることを表す. このことに注意して求められた最適解をみると, 分母はクラスター i に所属する個体の数を表し, 分子はクラスター i に所属する個体ベクトルの総和である. このことから, 最適解 v_i^j はクラスター中心を v_i とするクラスターの重心と一致することがわかる.

2.3 ファジィc-平均法

c-平均法では1つの個体は1つのグループにしか属することができない。1つの個体が複数のグループに所属できるように制約を緩める。帰属度 u_{ik} に0から1までの区間の値を連続的にとることを許し、その値を個体 x_k のクラスター i への帰属度(メンバシップ)とする。ここに、帰属度 u_{ik} にファジィネス (fuzziness-ファジィである性質のこと) と呼ばれるあいまいさが導入される。ファジィ理論 [65] の基本的考え方は、あいまいさを0から1までの区間の値で表すという点にある。クラスタリング技法にファジィの考えを導入したファジィクラスタリングは、研究が盛んに行われてきた [3, 6, 11, 22, 37, 51]。その中でもファジィc-平均法は、多くの変形がなされ応用も多い。

本論文では、不確定性を含むデータに対するファジィc-平均法を考察しているが、まず、不確定性を含まない従来のファジィc-平均法について述べる。従来のファジィc-平均法では不確定性を含まないデータを対象とする。

分類すべきデータを p 次元ユークリッド空間上の個体の集合 $X = \{x_1, x_2, \dots, x_n\}$ と表し、データの各個体を

$$x_k = (x_k^1, x_k^2, \dots, x_k^p)^T \quad (2.1)$$

と表す。これらの点を c 個に分類する。クラスター中心

$$v_i = (v_i^1, v_i^2, \dots, v_i^p)^T \quad (2.2)$$

はまとめて $V = \{v_1, v_2, \dots, v_c\}$ と表す。また、 v_i を中心とするクラスター i に x_k が所属する度合いを表すメンバシップを u_{ik} で表し、行列 $U = (u_{ik})$ を定義する。 u_{ik} は次の集合 M に属すると仮定される。

$$M = \{(u_{ik}) \mid u_{ik} \in [0, 1], \sum_{i=1}^c u_{ik} = 1, k = 1, 2, \dots, n\}. \quad (2.3)$$

ただしここでも

$$\sum_{i=1}^c u_{ik} = 1$$

が成立するように u_{ik} を決める。つまり1つの点は、複数のグループに所属することができるけれども、所属する度合いをすべて加え合わせると1.0になるという正則化の条件を満たす。

このように帰属度 u_{ik} をファジィ化して、c-平均法における u_{ik} の許容集合 M_c のかわりに \mathcal{M} を用いてみるとc-平均法がファジィ化されるだろうか。

許容集合 \mathcal{M} のもとで評価関数 J_1 を最適化してみるとCM2における最適解は一般にクリस्प解となる。つまり帰属度 u_{ik} の最適解は0もしくは1となる。なぜなら、最適化すべき目的関数

$$J_1(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} \|x_k - v_i\|^2$$

は u_{ik} に関して線形であり制約条件 \mathcal{M} も線形であるので、この最適化問題は線形計画画となり、最適解は制約条件の端点となる。よって、ただ単に許容集合をファジィ化しただけでは0から1の間の値をとるようなファジィな解は得られない、つまりc-平均法をファジィ化したことにはならない。

2.3.1 標準的なファジィc-平均法

c-平均法をファジィ化するために、c-平均法の目的関数を変更したファジィc-平均法 [4, 12, 13] (ここでは、標準的なファジィc-平均法と呼んでいる) の目的関数が提案された。

$$J_m(U, V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m \|x_k - v_i\|^2.$$

ここで、 m は、 $m > 1$ を満たすようなファジィ化パラメータである。($m = 1$ とおくと、c-平均法の目的関数 J_1 と一致することに注意する。)

標準的なファジィc-平均法では、c-平均法の目的関数 J_1 のかわりに J_m を最適化し、制約条件は、メンバシップが0もしくは1をとるc-平均法の制約条件 M_c のかわりに、0から1の値をとることを許容した制約条件 \mathcal{M} を使用する。

この最適化でもさきほどのc-平均法と同様に、目的関数には2つのパラメータ U と V が含まれるので一度に最適化することはできず、2段階の最適化アルゴリズムを用いて最適化する。そのアルゴリズムFCMは次のようなものである。

ファジィc-平均法のアルゴリズム

FCM1. \bar{U} と \bar{V} の初期値を適当に決める。

FCM2. $\min_{U \in \mathcal{M}} J_m(U, \bar{V})$ の最適解 \hat{U} を求め、 $\bar{U} = \hat{U}$ と置き換える。

FCM3. $\min_V J_m(\bar{U}, V)$ の最適解 \hat{V} を求め, $\bar{V} = \hat{V}$ と置き換える.

FCM4. 最適解 (\bar{U}, \bar{V}) が収束していれば終了,
そうでなければステップFCM2に戻る.

このファジィc-平均法のアルゴリズムでは, FCM2とFCM3においてそれぞれメンバシップとクラスター中心に関する最適化が行われるが, その最適解は次のようになる.

FCM2のメンバシップに関する最適解は, すべての $v_i (i = 1, \dots, c)$ に対して $x_k \neq v_i$ であるような x_k については

$$u_{ik} = \left[\sum_{j=1}^c \left(\frac{\|x_k - \bar{v}_i\|^2}{\|x_k - \bar{v}_j\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}, \text{ for } x_k \neq v_i, i = 1, \dots, c \quad (2.4)$$

となり, ある v_i について $x_k = v_i$ となる x_k については,

$$u_{ik} = 1; u_{jk} = 0 (j \neq i)$$

となる. このように, 個体と一致する中心がある場合を除いて, 帰属度 u_{ik} はファジィな解, すなわち0から1までの間の値をとることがわかる.

FCM3のクラスター中心に関する最適解は

$$v_i = \frac{\sum_{k=1}^n (\bar{u}_{ik})^m x_k}{\sum_{k=1}^n (\bar{u}_{ik})^m} \quad (2.5)$$

となる.

ここで, これら2つの最適解の導出過程をしてみることにしよう. FCM2におけるメンバシップ U の最適解について見てみる. いま, 制約条件 $M(2.3)$ をさらに緩めて, 制約条件を $\sum_{i=1}^c u_{ik} = 1$ のみとし,

$$u_{ik} > 0, 1 \leq i \leq c, 1 \leq k \leq n \quad (2.6)$$

の条件を取り去ってみる. (2.6) の条件なしで得られた最適解が仮に M にはいっていただけると, 緩い条件のもとでの最適解であるのだから, 制約 M のもとでの最

適解でもあるはずである。目的関数 J_m を緩い制約条件 $\sum_{i=1}^c u_{ik} = 1$ のもとで U に関して最小化する問題は、ラグランジュ乗数法を用いて解くことができる。

ラグランジュ乗数を $\lambda_k (k = 1, \dots, n)$ とすると、ラグランジュ関数は、

$$\begin{aligned} L &= J_m + \sum_{k=1}^n \lambda_k \left(\sum_{i=1}^c u_{ik} - 1 \right) \\ &= \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m \|x_k - v_i\|^2 + \sum_{k=1}^n \lambda_k \left(\sum_{i=1}^c u_{ik} - 1 \right) \end{aligned} \quad (2.7)$$

となり、最適性の必要条件として

$$\frac{\partial L}{\partial u_{ik}} = m(u_{ik})^{m-1} \|x_k - v_i\|^2 + \lambda_k = 0 \quad (2.8)$$

が得られる。 x_k について、 $x_k = v_i$ となる v_i が存在しない場合 $\|x_k - v_i\|^2 > 0$ ($i = 1, \dots, c$) であることがわかる。 λ_k を消去するため、上式の添字 i を j に変更して移項すれば、

$$u_{jk} = \left[\frac{-\lambda_k}{m \|x_k - v_j\|^2} \right]^{\frac{1}{m-1}}$$

この両辺を $j = 1, \dots, c$ について加えると $\sum_{j=1}^c u_{jk} = 1$ より、

$$\sum_{j=1}^c \left[\frac{-\lambda_k}{m \|x_k - v_j\|^2} \right]^{\frac{1}{m-1}} = 1$$

この式と (2.8) から λ_k を消去することができて、

$$u_{ik} = \left[\sum_{j=1}^c \left(\frac{\|x_k - v_i\|^2}{\|x_k - v_j\|^2} \right)^{\frac{1}{m-1}} \right]^{-1} \quad (2.9)$$

を得る。

この解は実際 $u_{ik} > 0$ を満たしている。緩い条件のもとでの最適解が、制約 M を満たしているのもとの制約 M のもとでの最適解でもある。よって、(2.9) は制約 M のもとでの最適解であることがわかった。

なお、最適性の必要条件だけからは、一般に最適解ということはできないが、この場合、目的関数 J_m が U に関して凸であるため、実際に最適解であることがわか

る。また、 $x_k = v_i$ を満たす v_i が存在する場合、 $u_{ik} = 1, u_{ik} = 0 (j \neq i)$ が最適であるのは、目的関数 J_m に注目すると明らかである。

つづいて、FCM3 におけるクラスター中心 v_i の最適解について考える。目的関数 J_m は

$$\|x_k - v_i\|^2 = \sum_{j=1}^p (x_k^j - v_i^j)^2$$

を使ってかき直すと

$$J_m(U, V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m \sum_{j=1}^p (x_k^j - v_i^j)^2$$

とかける。 v_i については制約条件はなく、また目的関数 J_m は v_i に関して凸である。このことから J_m を i 番目のクラスター中心の j 成分 v_i^j について偏微分する。

$$\begin{aligned} -\frac{1}{2} \frac{\partial J_m}{\partial v_i^j} &= \sum_{k=1}^n (u_{ik})^m (x_k^j - v_i^j) = 0. \\ v_i^j &= \frac{\sum_{k=1}^n (u_{ik})^m x_k^j}{\sum_{k=1}^n (u_{ik})^m}. \end{aligned}$$

すべての成分 j について上式が成立するので (2.5) が得られる。

2.3.2 エントロピー正則化を用いたファジィc-平均法

工学の諸問題には、真の解が存在しない問題や、観測データが少し変化すると解が大きく変化するような不連続性を持つ問題がしばしば含まれる。このような本質的に解きにくい問題は、適切でない問題あるいは不良設定問題 (ill-posed problem) と呼ばれる。不良設定問題の解き方としてよく知られている考え方に、正則化 (regularization) がある。正則化とは、もとの問題が特異なときにそれを何らかの方法で正則にして近似解を求める考え方であり、ある特別な解法やアルゴリズムを指すものではない。

ここで、ファジィc-平均法はc-平均法の正則化ととらえることはできないか、と考えてみる。正則化は特異な問題を正則にする方法であり、そのために、もとの問題にパラメータなどを導入して少し変形することによって、正則化された問題の解を求める。

通常, c-平均法は特異であるとはいわないが, ここでは c-平均法の解が特異であり, ファジィ化された解が正則であるかのようにとらえてみる. c-平均法においてただ単にメンバシップが0もしくは1の値をとる制約条件 M_c を, 0 から1の連続的な値をとることのできる制約条件 M に緩めただけでは, ファジィ化された解は求められなかった. そこで, 標準的なファジィc-平均法では, 目的関数 J_1 を J_m に変更することによりファジィ化された解が求められた. このことは, 目的関数にべき乗パラメータ m を導入することにより, ファジィ化がなされた, すなわち正則化が行われた, とみることができる.

正則化の要件として, もとの問題に解が存在するときには, 正則化された問題の解は, もとの解の近似解を表現しなければならない. 標準的なファジィc-平均法では, m を1に近づけると, ファジィな解は, c-平均法(もとの問題)の解に収束する.

今, 標準的なファジィc-平均法は, c-平均法の正則化であることを述べた. それでは, c-平均法の正則化の方法として典型的な正則化をみってみる. 典型的な正則化では, 正則化のためのパラメータと関数が加えられる. 正則化パラメータを $\lambda > 0$, 正則化のための関数を K とし, もとの目的関数を J' とすれば, 正則化された目的関数は,

$$J^\lambda = J' + \lambda^{-1}K$$

と表される. ファジィクラスタリングにおいて, このような典型的な正則化を c-平均法についてほどこすために, J' と K を定めよう. 特異な問題(正則化のための関数)は c-平均法の目的関数 ($J' = J_1$) である. 一方, 正則化のための関数として, エントロピー関数を使う. この正則化の方法は, c-平均法をファジィ化(正則化)していることから, ファジィc-平均法の一つと考えられる. そこで, エントロピー関数により c-平均法をファジィ化する方法を, エントロピー正則化によるファジィc-平均法と呼ぶことにする [44, 48].

正則化のための関数をエントロピー

$$K = \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log u_{ik}$$

とすると, 正則化された目的関数は

$$J^\lambda(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} \|x_k - v_i\|^2 + \lambda^{-1} \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log u_{ik}$$

となる.

エントロピー正則化によるファジィc-平均法では、メンバシップ u_{ik} のとることのできる範囲を示した許容集合 \mathcal{M} は、標準的なファジィc-平均法と共通である。ここでもさきほどの標準的なファジィc-平均法と同じく目的関数 J^λ には、2つの変数、メンバシップ u_{ik} とクラスター中心 v_i が含まれるので、エントロピー正則化を用いたファジィc-平均法でも標準的なファジィc-平均法同様の2段階アルゴリズムを使って最適解を求める。

エントロピー正則化を用いたファジィc-平均法のアルゴリズム

FCM1. \bar{U} と \bar{V} の初期値を適当に決める。

FCM2. $\min_{U \in \mathcal{M}} J^\lambda(U, \bar{V})$ の最適解 \hat{U} を求め、 $\bar{U} = \hat{U}$ と置き換える。

FCM3. $\min_{\bar{V}} J^\lambda(\bar{U}, \bar{V})$ の最適解 \hat{V} を求め、 $\bar{V} = \hat{V}$ と置き換える。

FCM4. 最適解 (\bar{U}, \bar{V}) が収束していれば終了、そうでなければステップFCM2に戻る。

2つの変数を最適化するアルゴリズムのステップFCM2とステップFCM3の最適解は、次のようになる。

$$u_{ik} = \frac{e^{-\lambda \|x_k - \bar{v}_i\|^2}}{\sum_{j=1}^c e^{-\lambda \|x_k - \bar{v}_j\|^2}}. \quad (2.10)$$

$$v_i = \frac{\sum_{k=1}^n \bar{u}_{ik} x_k}{\sum_{k=1}^n \bar{u}_{ik}}. \quad (2.11)$$

さて、上の2つの最適解の導出方法をみてみよう。FCM2における最適解の導出では、標準的なファジィc-平均法のところで述べたように制約条件 \mathcal{M} の制約を $\sum_{i=1}^c u_{ik} = 1$ のように緩め、ラグランジュ関数を

$$L = \sum_{i=1}^c \sum_{k=1}^n u_{ik} \|x_k - v_i\|^2 + \lambda^{-1} \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log u_{ik} + \sum_{k=1}^n \nu_k \left(\sum_{i=1}^c u_{ik} - 1 \right)$$

で定義する．ここで， $\nu_k (k = 1, \dots, n)$ はラグランジュ乗数である．

$$\frac{1}{2} \frac{\partial L}{\partial u_{ik}} = \|x_k - v_i\|^2 + \lambda^{-1}(1 + \log u_{ik}) + \nu_k$$

から

$$u_{ik} = \exp(-\lambda \|x_k - v_i\|^2 - \lambda \nu_k - 1) \quad (2.12)$$

を得る．この解は $0 \leq u_{ik} \leq 1$ を満たしているから (2.12) は制約 M のもとでの最適解を与えている． $\sum_{i=1}^c u_{ik} = 1$ によって上式から ν_k を消去すれば最適解 (2.10) を得る．目的関数は u_{ik} について凸であることを注意しよう．

FCM3 における最適解の導出では，標準的なファジィc-平均法と同様に，目的関数 J^λ がクラスター中心 V に関して凸であることから， J^λ を i 番目のクラスター中心の j 成分 v_i^j について偏微分し，最適解を求める．

$$\begin{aligned} -\frac{1}{2} \frac{\partial J^\lambda}{\partial v_i^j} &= \sum_{k=1}^n u_{ik} (x_k^j - v_i^j) = 0. \\ v_i^j &= \frac{\sum_{k=1}^n (u_{ik}) x_k^j}{\sum_{k=1}^n (u_{ik})}. \end{aligned}$$

すべての成分 j について上式が成立するので (2.11) が得られる．

2.4 ファジィクラスタリングで用いられる距離空間

前節で紹介した標準的なファジィc-平均法とエントロピー正則化を用いたファジィc-平均法では，通常，ユークリッド距離の二乗が用いられる．しかしながら， L_1 距離を用いたファジィc-平均法についても研究が行われている． L_1 距離を用いたファジィc-平均法は，ユークリッド距離の二乗を用いた場合とは距離空間が違うので，異なる性質を持っている．

2.4.1 ユークリッド距離

ユークリッド距離は，現実の世界の3次元距離空間で測定されるように，実世界での距離を表現できる．ただし，クラスター分析ではユークリッド距離そのもので

はなく，ユークリッド距離の二乗が広く用いられている．また，ユークリッド距離は，測定する距離空間において，各々の成分の尺度がすべて同じであれば自然な距離であるが，異なる尺度の混在する空間では，ユークリッド距離が自然な距離であるとは必ずしもいえない．ユークリッド空間は， L_2 空間とも呼ばれ， p 次元ユークリッド空間での x と y との距離 $d_2(x, y)$ は，

$$d_2(x, y) = \|x - y\|_2 = \sqrt{\sum_{l=1}^p (x^l - y^l)^2} \quad (2.13)$$

で表される．クラスタリングではこのユークリッド距離の二乗が一般的に多く用いられている．ユークリッド距離の二乗を用いたファジィクラスタリングについては前節で述べた．

2.4.2 L_1 距離

クラスタリングにはユークリッド距離の二乗以外に L_1 距離も用いられてきた [56]． x と y の間の L_1 距離 $d_1(x, y)$ は，

$$d_1(x, y) = \|x - y\|_1 = \sum_{l=1}^p |x^l - y^l|. \quad (2.14)$$

と表される．データが，異なる尺度の混在する空間に存在する場合，ユークリッド距離よりも L_1 距離を用いたほうがより自然である． L_1 距離を用いた場合のファジィc-平均法は，ユークリッド距離の二乗を用いた場合のファジィc-平均法と同様に解を導くことはできない． L_1 距離に基づくファジィc-平均法については，Bobrowskiら [8] や Jajuga[34] によって研究されている．この， L_1 距離に基づくファジィc-平均法について説明する．

不確定性を含まない p 個の成分からなるデータを $X = \{x_1, x_2, \dots, x_n\}$ とし，各々の個体は

$$x_k = (x_k^1, x_k^2, \dots, x_k^p)^T \quad (2.15)$$

で表されるとする．これらの個体は c 個のクラスターに分類され，クラスター中心 $V = \{v_1, v_2, \dots, v_c\}$ は

$$v_i = (v_i^1, v_i^2, \dots, v_i^p)^T \quad (2.16)$$

のように表される。また、個体 x_k がクラスター i に所属する度合いは $U = (u_{ik})$ で表され、メンバシップ行列と呼ばれる。メンバシップ行列 $U = (u_{ik})$ はファジィ分割 \mathcal{M} を表している。

$$\mathcal{M} = \{(u_{ik}) \mid u_{ik} \in [0, 1], \sum_{i=1}^c u_{ik} = 1, k = 1, 2, \dots, n\}. \quad (2.17)$$

ファジィc-平均法は、クラスター中心 V とメンバシップ U を変数とする目的関数を最小化することにより、クラスタリングが行われる。標準的なファジィc-平均法の目的関数は

$$J_m(U, V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m d_{ik} \quad (2.18)$$

と定義される。ここで、 $m > 1$ は定数パラメータである。エントロピー正則化を用いたファジィc-平均法の目的関数は

$$J^\lambda(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} d_{ik} + \lambda^{-1} \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log u_{ik} \quad (2.19)$$

と定義される。パラメータ $\lambda > 0$ は定数である。目的関数 (2.18), (2.19) で用いられる d_{ik} は個体 x_k とクラスター中心 v_i との間の L_1 距離を表している。

$$d_{ik} = \|x_k - v_i\|_1 = \sum_{l=1}^p |x_k^l - v_i^l|. \quad (2.20)$$

目的関数 (2.18), (2.19) を最小化することにより、クラスタリングが実現されるが、各々の目的関数は2つの変数、メンバシップ U とクラスター中心 V を含んでいる。そこで次のようなファジィc-平均法のアルゴリズムを用いて、2つの変数に関して交互に最適化をおこなうことにより、目的関数の最小化を行う。

$$J(U, V) = J_m(U, V) \quad \text{もしくは} \quad J(U, V) = J^\lambda(U, V)$$

と目的関数を置き換え、アルゴリズム FCM を適用する。

ファジィc-平均法のアルゴリズム

FCM1. \bar{U} の初期値を定める。

FCM2. $V = \bar{V}$ を固定して

$$\min_{U \in \mathcal{M}} J(U, \bar{V})$$

を解き、最適解を \bar{U} とする。

FCM3. $U = \bar{U}$ を固定して

$$\min_V J(\bar{U}, V)$$

を解き、最適解を \bar{V} とする.

FCM4. 解 (\bar{U}, \bar{V}) が収束すれば終了. そうでなければ FCM2 へ.

End of FCM

標準的なファジィc-平均法の FCM2 における解は

$$\bar{u}_{ik} = \left[\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{1}{m-1}} \right]^{-1} \quad (2.21)$$

であり、エントロピー正則化を用いたファジィc-平均法の FCM2 における解は

$$\bar{u}_{ik} = \frac{e^{-\lambda d_{ik}}}{\sum_{j=1}^c e^{-\lambda d_{jk}}} \quad (2.22)$$

である. しかしながら, L_1 距離に基づく標準的なファジィc-平均法とエントロピー正則化を用いたファジィc-平均法における FCM3 の解 (クラスター中心の最適解 \bar{V}) はユークリッド距離に基づくファジィc-平均法のように簡単には求めることはできない. 不確定性を含まないデータに対し, L_1 距離を用いた場合の FCM3 の解に関して Bobrowski ら [8] や, Jajuga [34] によって研究されており, より効率的なアルゴリズムが宮本ら [43] によって提案されている. 宮本らにより提案されたクラスター中心を求めるアルゴリズムは, 以下のようなものである.

L_1 距離に基づくファジィc-平均法の目的関数のクラスター中心に関する偏導関数を考えると, 各々の成分は単調増加で階段の形状をした区分関数となっている. このアルゴリズムでは, 階段状の区分関数である偏導関数をクラスター中心の第 l 成分 v_k^l に関して区分点を順に探索していき, 偏導関数の値が負から非負に転じる場所を探索する. つまり, 目的関数がクラスター中心の1つの成分に関して凸であるので, 最適解を求めるために偏導関数が負から非負に転じる場所を求めている.

個体 x_k の第 l 成分 x_k^l について, アルゴリズムの前段階で, まずソーティングが行われる.

$$x_1^l \leq \dots \leq x_k^l \leq \dots \leq x_n^l$$

のように並べかえる.

Algorithm (Searching v_i^l)

```

begin
   $S := -\sum_{k=1}^n (u_{ik})^m$ ;
   $k := 0$ ;
  while ( $S < 0$ ) do begin
     $k := k + 1$ ;
     $S := S + 2(u_{ik})^m$ ;
  end;
  output  $v_i^l = x_k^l$  as the  $l$ -th coordinate of the cluster center  $v_i$ ;
end.
```

このアルゴリズムを用いることにより, 目的関数のクラスター中心に関する厳密な最適解が得られる.

2.5 集合間の距離

2.5.1 最長距離法と最短距離法

2つの集合間の距離として考えられる距離の最大値と最小値として最長距離と最短距離が考えられる. これらの距離は, 他の集合間の距離の上限と下限という意味を持つ. ここで扱う不確定性を含むデータのファジィc-平均法では, クラスター中心と不確定性を含むデータ間の距離が, 最長距離と最短距離それぞれを用いて定義される. 集合 K と集合 L の間の最長距離, 最短距離 $d(K, L)$ は

最長距離法 (Farthest Neighbor)

$$d(K, L) = \max\{d(x, y) : \text{for all } x \in K, y \in L\}$$

最短距離法 (Nearest Neighbor)

$$d(K, L) = \min\{d(x, y) : \text{for all } x \in K, y \in L\}$$

で与えられる.

2.5.2 ハウスドルフ距離

最長距離や、最短距離の間の性質を持つハウスドルフ距離を紹介する。ただし、本論文では、クラスター中心は不確定性を含まないと仮定するので、この仮定をおくと、ハウスドルフ距離が最長距離に一致する。

任意の2つの集合 K と L の間のハウスドルフ距離 $D(K, L)$ を求めることにする。 K の要素 x と L の要素 y を使って次のようにして求められる [57]。

(i) K の要素 x と集合 L との間の距離 $d(x, L)$ を与える。

$$d(x, L) = \bigwedge_{y \in L} d(x, y).$$

(ii) 集合 K と集合 L との間の非対称距離である $\delta(K, L)$ を次のように定義する。

$$\delta(K, L) = \bigvee_{x \in K} d(x, L).$$

(iii) (ii) の非対称距離からハウスドルフ距離 $D(A, B)$ が与えられる。

$$D(K, L) = \delta(K, L) \vee \delta(L, K).$$

ここで、(ii) の $\delta(K, L)$ は非対称距離であることに注意しよう。このようにして、ハウスドルフ距離 $D(K, L)$ が求められる。不確定性を含むデータに対するハウスドルフ距離を用いたクラスタリング手法の研究 [59] も行われている。

ところで、ファジィc-平均法では、クラスター中心は不確定性を持たないと仮定している。これを集合の観点からみると、クラスター中心に対応する集合は、ただ1つの要素からなることになる。この場合、集合間の距離として、最長距離法とハウスドルフ距離を用いた場合は、距離は全く同一となる。つまり、集合間の距離が片方の集合がただ1つの要素からなっている場合、最長距離とハウスドルフ距離が一致することを示す。

2つの集合 K と L を考え、このうち集合 K には、ただ一つの要素 x からなるとする。

$$K = \{x\}.$$

この場合に最長距離法で距離を求めてみると

$$d(K, L) = \max\{d(x, y) : \text{for all } y \in L\} \quad (2.23)$$

となる。また、ハウスドルフ距離では、

(i) K の要素 x と集合 L との間の距離 $d(x, L)$ を与える。

$$d(x, L) = \bigwedge_{y \in L} d(x, y).$$

また、逆の L の要素 y と集合 K との間の距離 $d(y, K)$ を与える。

$$d(y, K) = \bigwedge_{x \in K} d(y, x) = d(y, x).$$

(ii) 集合 K と集合 L との間の非対称距離である $\delta(K, L)$ を次のように求める。

$$\delta(K, L) = \bigvee_{x \in K} d(x, L) = d(x, L).$$

逆の、集合 L と集合 K との間の非対称距離である $\delta(L, K)$ を次のように求める。

$$\delta(L, K) = \bigvee_{y \in L} d(y, K).$$

(iii) (ii) の非対称距離からハウスドルフ距離 $D(A, B)$ が与えられる。

$$\begin{aligned} D(K, L) &= \delta(K, L) \vee \delta(L, K) \\ &= \left\{ \bigwedge_{y \in L} d(x, y) \right\} \vee \left\{ \bigvee_{y \in L} d(y, x) \right\} \\ &= \bigvee_{y \in L} d(y, x) \\ &= \bigvee_{y \in L} d(x, y). \end{aligned} \tag{2.24}$$

ここで、(2.13) より、 $d(y, x) = d(x, y)$ であることに注意しよう。また、 \vee は、 \max の記号なので、今回扱ったファジィc-平均法に限れば、最長距離法を用いた場合 (2.23) と、ハウスドルフ距離を用いた場合 (2.24) は、距離が全く同一となることが示された。

しかし、一般的には集合間の距離の定義が違えば、求める距離も違ったものになるので、今回同一となることが示された2つの距離の定義は、違ったものとして扱わなければならない。