論文

ファジィc-Meansによる
ハードクラスタリング†

太田 知宏* 根本 宗記* 市橋 秀友* 三好 哲也*

本論文はラグランジュ乗数法によるファジィ c-Means アルゴリズムのハードクラスタリングへの有効な利用法について述べている。ハードクラスタリングは簡単な組み合わせ最適化問題であるが、数理計画的解法よりもヒューリスティックなアルゴリズムが多く用いられている。よく知られた k -Means アルゴリズムは計算効率が良く短時間にクラスタリング結果を得ることができるが、解は初期値に依存し最適解が得られる頻度はかなり低い。組み合わせ最適化問題の解法としてのホップフィールドニューラルネットワークの研究では、多重一次形式のエネルギー関数についての解空間の超立方体の頂点における解の安定性の議論が行われている。本論文ではハードクラスタリング問題にこれらの知見を応用し、ファジィ c-Means のラグランジュ乗数法や内点法による解法を用いて、閾値を適当に設定することにより最適なクラスタリング結果を高い頻度で得ることができることを示している。ファジィ c-Means の定式化にはエントロピー最大化のファジィクラスタリングを用いて、数値例により提案アルゴリズムの有効性を示している。

キーワード：ホップフィールドニューラルネットワーク、ファジィ c-Means、エントロピー、ハードクラスタリング

1. はじめに

k -Means クラスタリングアルゴリズムは、評価関数の最適化に基づく非階層的ハードクラスタリング法であるが、初期値に依存した解が得られ最適なクラスタリングが得られる頻度はあまり高くない。このことはファジィ c-Means[1][2]も同様に画像処理への応用においても最適なクラスターが得られない問題点が指摘されている[3]。しかし、ファジィ c-Means はパターン認識、判別分析、非線形回帰などさまざまなデータ解析の前処理として多くの応用が行われており、最近では回帰分析とクラスター分析の同時分析法[4]など分類と同時に他のデータ解析手法を適用する研究も行われてきている。これらの応用に共通する課題はクラスタリング結果、すなわちクラスタリング問題の解の質をいかに向上するかであり、最適なクラスターを得るための改善が必要である。

ファジィ c-Means はメンバシップ値のべき乗を評価関数に導入することによってファジィクラスタリングを求める方法である。Li and Mukaidono[5][6]により始められたエントロピー最大化のファジィクラスタリングは、従来のファジィ c-Means と同種のものでありラグランジュ乗数法を用いることもできることが示されている[7][8]。

本論文では、エントロピーの対数関数がホップフィールドニューラルネットワーク[9]における非整数解の出現を制御する項[10]に類似していることから、そこでの解の安定性に関する知見を利用して、ファジィ c-Means によりハードクラスタリングの最適解を得るための評価関数の係数値やメンバシップの閾値の設定法について考察する。

まず組み合わせ最適化問題の勾配法による解法であるホップフィールドニューラルネットワークに基づいて定式化し非整数解の出現を制御する項の効果について述べる。この方法によるクラスタリングでもクラスター中心の初期値に依存した結果が得られ、それが適切に与えられればほぼ最適解に収束する。したがって 0-1 変数については、早く整数解に収束しない方が勾配法によるクラスター中心位置の探索が十分行えて有効であると考えられる。クラスタリングへの応用ではこの項の重み係数は整数解を得にくくするという逆の意味で用いる。

k -Means クラスタリングは線形制約条件付非線形計画問題として定式化されるのが一般的である。そこで、本論文では線形計画問題に対する解法アルゴリズムである内点法[11]の中でも簡単でかつ計算効率が良好であるとされているアフィン変換法[12]を応用し、目的関数の係数値とメンバシップの 0-1 判定閾値を適切に設定すれば k -Means アルゴリズムに比べて高い頻度で最適解を得ることができることを示す。またこのことから、ラグランジュ乗数法によるファジィ c-Means 法も同様に最適解の得られる頻度が高いことを示す。

† Hard Clustering by Fuzzy c-Means
Tomohiro OHTA, Muneki NEMOTO, Hidetomo ICHIIHA-SHI and Tetsuya MIYOSHI

* 大阪府立大学工学部経営工学科
College of Engineering, Osaka Prefecture University

2. ホップフィールドニューラルネットワークによるクラスタリング

簡便化されたホップフィールドニューラルネットワークは次のような常微分方程式で記述される。

$$\frac{du_i}{dt} = \sum_j w_{ij} v_j + h_i \quad (1)$$

$$v_i = s(u_i) = \frac{1}{2} (1 + \tanh \frac{u_i}{x_0}), \quad i=1, \dots, m \quad (2)$$

ただし, u_i, v_i および h_i はそれぞれニューロン i の内部状態, 出力, および閾値, w_{ij} はニューロン j からニューロン i への結合荷重, ニューロン間結合は対称結合 ($w_{ij} = w_{ji}$) で $w_{ii} = 0, x_0$ は定数である。なおニューロンの出力値は $v_i \in [0, 1]$ であり, N 個のニューロンからなる回路網の状態 $\mathbf{v} = (v_1, \dots, v_N) \in [0, 1]^N$ は単位超立方体内部の点で表せる。超立方体の内部の点 $V_0 \in [0, 1]^N$ を初期値として回路網を動作させると, 漸近安定な頂点に収束する。このニューラルネットワークのエネルギー E を

$$E(\mathbf{v}) = -\frac{1}{2} \sum_i \sum_j w_{ij} v_i v_j - \sum_i h_i v_i \quad (3)$$

と定義すると, E は時間の変化とともに減少し, 極小となる安定平衡点に収束する。

k -Means クラスタリングはクラスター中心とパターン間のユークリッド距離によりパターンをクラスターに分ける組合せ最適化問題である。この問題を連続緩和型解法で解くために次のような目的関数(エネルギー関数)を考える。

$$\begin{aligned} E(\mathbf{x}, \mathbf{r}) = & \frac{1}{2} \sum_{i=1}^N \left(\sum_{j=1}^M x_{ij} - 1 \right)^2 \\ & + \frac{A}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij} (1 - x_{ij}) \\ & + \frac{B}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij} \| \mathbf{d}_i - \mathbf{r}_j \|^2 \end{aligned} \quad (4)$$

ここで i はパターンの添え字, j はクラスターの添え字, x_{ij} はパターン i のクラスター j へのメンバシップを表す, $\mathbf{d}_i = (d_{i1}, \dots, d_{ik})$ はパターン i , $\mathbf{r}_j = (r_{j1}, \dots, r_{jk})$ はクラスター j の中心ベクトルであり, A, B は定数である。また, 第1項は各パターンが一つのクラスターだけに属することを表す項, 第2項は非整数解の出現を制御する項[10]で, 第3項はクラスター中心からの距離を表す項である。式(1)–(3)の力学系に対するアナログの電気回路による実現は現状では課題が多い。デジタルコンピュータを用いるならばルンゲ・クッタ法で力学系をシミュレートするよりも非線形計画問題として定式化し最急降下法を適用する方が計算コストの

上で適していると考えられる。本章では, 式(4)のエネルギー関数に対する最急降下法(外点法)により次の更新則で解を求めることを考える。

$$x_{ij}^{NEW} = x_{ij}^{OLD} - \tau \frac{\partial E}{\partial x_{ij}} \quad (5)$$

$$\begin{aligned} \frac{\partial E}{\partial x_{ij}} = & \sum_{l=1}^M x_{il} - 1 + A \left(\frac{1}{2} - x_{ij} \right) \\ & + \frac{B}{2} \| \mathbf{d}_i - \mathbf{r}_j \|^2 \end{aligned} \quad (6)$$

$$r_{jk}^{NEW} = r_{jk}^{OLD} - \tau \frac{\partial E}{\partial r_{jk}} \quad (7)$$

$$\frac{\partial E}{\partial r_{jk}} = B \sum_{i=1}^N x_{ij} (r_{jk} - d_{ik}) \quad (8)$$

ただし, τ は小さな正数であり,

$$x_{ij} = \begin{cases} 1 & ; u_{ij} > 1 \\ u_{ij} & ; 0 \leq u_{ij} \leq 1 \\ 0 & ; u_{ij} < 0 \end{cases} \quad (9)$$

と簡便化した。式(4)のエネルギー関数のすべての x_{ij} が 0 または 1 で, 第1項目が 0 となる超立方体の頂点 $X = (X_{11}, \dots, X_{NM}) \in [0, 1]^{N \times M}$ に収束するとき, それを実行可能解と呼ぶ。巡回セールスマン問題や回路ブロック配置問題などでホップフィールドニューラルネットワークの改良[10]として用いられる係数 A は $A=1$ とするとエネルギー関数が $\mathbf{x}$ についての多重1次形式となりいずれかの頂点に収束することが知られている。しかし, 経験的に多重1次形式のエネルギー関数は一般のエネルギー関数に比べてよい解が得にくいとされている。また, $A < 1$ のときは多重1次形式にはならず, 超立方体の内部の極小点に収束する。しかし, 式(4)はクラスタリングのための目的関数であり, x_{ij} の和が 1 であれば第1項目は最小となるので, $A=0$ で第2項目の非整数解の出現を制御する項がなくても, x_{ij} は 0 または 1 に収束し, メンバシップの和が 1 であるという制約条件を満たす実行可能解が得られる。そこで A を負の値として, さらにメンバシップの 0-1 判定閾値 θ_L, θ_U を設定し, 更新中に x_{ij} が小さな正数 θ_L 以下になると 0 に固定し θ_U 以上になると 1 として終了するアルゴリズムを考える。

以下では, 判定閾値 θ_L, θ_U と係数 A, B の定性的な目安を得るために, 解に対応する超立方体の頂点 X と, それに隣接する頂点 X の間におけるエネルギー曲面について考察する。

実行可能解 $(\mathbf{x}, \mathbf{r})$ の要素 x_{ij} は 0 または 1 であるので, $\mathbf{x}$ を超立方体 $[0, 1]^{N \times M}$ の内部から次第に頂点 X に収束させることを考える。解の頂点 X の要素 X_{ab} が 1 である場合と 0 である場合に分ける。文献[13]での超立

方体の頂点における解の安定性についての性質より以下のことが容易に導かれる。

1) x_{ab} が 1 であるとき

実行可能解においてパターン a がクラスター b に属する場合は解の頂点 X において $x_{ab}=1$ である。解の頂点 X と X に隣接する頂点との間におけるエネルギー曲面の極小値の位置を推定する。 $c \neq b$ であるすべての c について $x_{ac}=0$ であるので

$$\frac{\partial E}{\partial x_{ab}} = x_{ab} - 1 + A \left(\frac{1}{2} - x_{ab} \right) + \frac{B}{2} \| \mathbf{d}_a - \mathbf{r}_b \|^2 \quad (10)$$

であり、 $x_{ab}=0$ である隣接頂点において

$$\frac{\partial E}{\partial x_{ab}} \Big|_{x_{ab}=0} = -1 + \frac{A}{2} + \frac{B}{2} \| \mathbf{d}_a - \mathbf{r}_b \|^2 \quad (11)$$

である。

K 次元ベクトルで表されるパターン $\mathbf{d}_i$ が超立方体 $[0,1]^K$ の内部に基準化されているとすると、

$$\| \mathbf{d}_a - \mathbf{r}_b \|^2 \leq 1 \quad (12)$$

であるので、 A を負の適当な大きさに選ぶことにより

$$\frac{\partial E}{\partial x_{ab}} \Big|_{x_{ab}=0} < 0 \quad (13)$$

とできる。また

$$\frac{\partial E}{\partial x_{ab}} \Big|_{x_{ab}=1} = -\frac{A}{2} + \frac{B}{2} \| \mathbf{d}_a - \mathbf{r}_b \|^2 \quad (14)$$

であるので、 A を負、 B を正とすると

$$\frac{\partial E}{\partial x_{ab}} \Big|_{x_{ab}=1} > 0 \quad (15)$$

である。このとき x_{ij} についての 2 次関数 E の超立方体の辺(解の頂点とその隣接する頂点との間における辺)上の極小値は区間 $[0,1]$ の内部にある。 $x_{ab}=1$ であるので A を小さくすると極小点が左に移動し、解の頂点から遠ざかる。極小値の位置が 0.5 に近くなるように係数 A を選べば、最急降下法により x_{ij} が 0 または 1 に収束しにくくなりクラスター中心を十分探索できるので、 $\mathbf{r}$ の最適解が得やすいと期待できる。極小値の位置は

$$x_{ab} = \frac{1 - \frac{A}{2} - \frac{B}{2} \| \mathbf{d}_a - \mathbf{r}_b \|^2}{1 - A} \quad (16)$$

である。クラスタリングの目的はクラスター中心 $\mathbf{r}_j$ からの距離を最小化することであるので正数 B は大きくすべきであるが、式(4)の第 1 項が 0 にならず実行可能解を得にくくなる。式(16)より B を大きくすると極小点は解の頂点 X ($x_{ab}=1$) から $x_{ab}=0$ 側に隔たった位置にくる。

2) x_{ab} が 0 のとき

実行可能解においてパターン a がクラスター b に属さないとき解の頂点 X において $x_{ab}=0$ であるので、 $x_{al}=1$ となる l がひとつ存在し、その他は $x_{al}=0$ であるので

$$\frac{\partial E}{\partial x_{ab}} = x_{ab} + A \left(\frac{1}{2} - x_{ab} \right) + \frac{B}{2} \| \mathbf{d}_a - \mathbf{r}_b \|^2 \quad (17)$$

$$\frac{\partial E}{\partial x_{ab}} \Big|_{x_{ab}=0} = \frac{A}{2} + \frac{B}{2} \| \mathbf{d}_a - \mathbf{r}_b \|^2 \approx 0 \quad (18)$$

かつ正である。また

$$\frac{\partial E}{\partial x_{ab}} \Big|_{x_{ab}=1} = 1 - \frac{A}{2} + \frac{B}{2} \| \mathbf{d}_a - \mathbf{r}_b \|^2 > 0 \quad (19)$$

となる。ただし A は負で $|A| \ll 1$ としている。式(18)、(19)より、解の頂点 X と隣接する頂点との間の超立方体 $[0,1]^K \setminus M$ の一辺及びその延長線上におけるエネルギー曲面の極小値

$$x_{ab} = \frac{-\frac{A}{2} - \frac{B}{2} \| \mathbf{d}_a - \mathbf{r}_b \|^2}{1 - A} \quad (20)$$

は $A=0$ のとき $x_{ab}=0$ より左側にある。 $\partial x_{ab} / \partial A < 0$ であるので A を 0 から小さくしていくと極小値の位置は、右側に移動する。

以上の考察は係数 A の値によって極小値の位置を操作しようとするもので、 A を 0 より少し小さくすることにより x_{ij} が 0 または 1 に急速に収束するのを防止できる。

3. アフィン変換内点法によるクラスタリング

式(4)のホップフィールドエネルギーに対する最急降下法は外点法であり、式(4)の係数 A と B および閾値 θ_L, θ_U の 4 つを適切に選ばないと実行可能解が求まらなかったり、最適解の得られる頻度が低い。そこで次の改善を考える。

一般に k -Means 法によるクラスタリングは制約条件付非線形計画問題として定式化されている[1]。 k -Means クラスタリングの問題を次のように表す。

$$\begin{aligned} \min \quad E(\mathbf{x}, \mathbf{r}) = & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij} \| \mathbf{d}_i - \mathbf{r}_j \|^2 \\ & + \frac{A}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij} (1 - x_{ij}) \end{aligned} \quad (21)$$

$$\text{s.t.} \quad \sum_{j=1}^M x_{ij} = 1, \quad i=1, \dots, N \quad (22)$$

$$x_{ij} \geq 0 \quad (23)$$

式(22)、(23)は $M-1$ 次元の N 個の単体を表している。本章ではこれらの単体内部で式(21)を最小化する

解を探索する. 式(21)はホップフィールドニューラルネットワークでのエネルギー関数の第1項目と係数 B がなく, 線形制約条件のついた非線形最小化問題である. 式(21)第2項目の A が0の場合は, $\mathbf{r}$ を固定すると線形計画問題であることから式(21)-(23)の問題の解 $\mathbf{x}$ は退化[14]がなければ端点である. すなわち x_{ij} は0または1に収束することが知られている[1].

すべての x_{ij} のうち第 i 番目のパターンに対応する部分を $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iM})^T$ とする. ただし T は転置を表す. また M 次元ベクトル $\mathbf{I}$ を $\mathbf{I} = (1, 1, \dots, 1)^T$, 第 j クラスタの中心ベクトルを $\mathbf{r}_j = (r_{j1}, r_{j2}, \dots, r_{jk})^T$ とする. クラスタ中心 $\mathbf{r}$ の更新は $\Delta \mathbf{r} = -\partial E / \partial \mathbf{r}$ とする. $\Delta \mathbf{r}$ の要素は $-\partial E / \partial r_{jk}$ であり

$$\frac{\partial E}{\partial r_{jk}} = \sum_{i=1}^N x_{ij} (d_{ik} - r_{jk}) \quad (24)$$

である.

式(22)の制約条件式で表される単体の内部に閉じこめて探索するのに勾配射影法[15]を直接用いるといくつかの x_{ij} の値は負になり単体の外部に出るために許容領域へのへりかがり[11]を必要とする. 本章では線形計画問題に対する解法アルゴリズムであるアフィン変換を用いる内点法を k -Means クラスタリング問題の式(21)-(23)に適用する. 問題(21)-(23)は非線形計画問題であるが, 制約条件は線形であり, 実行可能領域は N 個の M -1 次元単体である. したがって単体の内部を探索する Karmarkar 法での射影変換を用いることも考えられるが Karmarkar 標準形への変換はやや複雑で人為的である[14]. そこで内点法の中でも簡単でかつ計算効率が良好であるとされているアフィン変換法[12], [14]を用いて単体の内部にとじ込めて解を探索する. 初期実行可能内点は

$$x_{ij} = \frac{1}{M}, \quad i=1, \dots, N, \quad j=1, \dots, M \quad (25)$$

とする. 線形計画問題におけるアフィン変換法では常に $x_{ij} > 0$ となるように単体の内部を比較的大きなステップ幅で探索できることが特長である. ある実行可能内点 $\mathbf{x}^* > \mathbf{0}$ に対して, 任意の点 $\mathbf{x}$ をすべての要素が正である点 $\mathbf{y} > \mathbf{0}$ に移す写像を

$$y_{ij} = \frac{x_{ij}}{x_{ij}^*}, \quad i=1, \dots, N, \quad j=1, \dots, M \quad (26)$$

とする. x_{ij}^* , $j=1, \dots, M$ を対角要素とする $M \times M$ 対角行列を D_i^* とすると

$$\mathbf{y}_i = (D_i^*)^{-1} \mathbf{x}_i \quad (27)$$

とベクトル表現できる.

$$\mathbf{x}_i = D_i^* \mathbf{y}_i \quad (28)$$

であるので式(21)の目的関数は

$$E(\mathbf{y}, \mathbf{r}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij}^* y_{ij} \|\mathbf{d}_i - \mathbf{r}_j\|^2 + \frac{A}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij}^* y_{ij} (1 - x_{ij}^* y_{ij}) \quad (29)$$

式(22), (23)の制約条件は

$$\sum_{j=1}^M x_{ij}^* y_{ij} = 1, \quad i=1, \dots, N \quad (30)$$

$$y_{ij} \geq 0 \quad (31)$$

と表される. 式(26)は実行可能内点 $\mathbf{x}^*$ を点 $(1, 1, \dots, 1)^T$ に移す写像であるので, $\mathbf{x} = \mathbf{x}^*$ のとき $\mathbf{y} = (1, 1, \dots, 1)^T$ は式(30)を満たす式(29)-(31)の実行可能内点である. 式(29)-(31)に対して勾配射影法を適用する. 最急降下方向 $-\nabla_{\mathbf{y}} E$ を式(30)の制約条件を破らないよう式(30)で定まる部分空間に平行な部分空間上に直交射影し

$$\Delta \mathbf{y}_i = -\nabla_{\mathbf{y}_i} E + \left(\frac{\nabla_{\mathbf{y}_i} E^T \mathbf{x}_i^*}{\mathbf{x}_i^{*T} \mathbf{x}_i^*} \right) \mathbf{x}_i^*, \quad i=1, \dots, N \quad (32)$$

とする. 式(32)は射影行列を用いることによって求められるが, グラム・シュミットの直交化法により容易に求めることができる(付録参照). $\nabla_{\mathbf{y}_i} E$ の要素 $\partial E / \partial y_{ij}$ は $\mathbf{y}_i = \mathbf{I}$ のとき

$$\frac{\partial E}{\partial y_{ij}} = \frac{1}{2} x_{ij}^* \|\mathbf{d}_i - \mathbf{r}_j\|^2 + \frac{A}{2} x_{ij}^* (1 - 2 x_{ij}^*) \quad (33)$$

である. ただし $\mathbf{I}$ は M 次元ベクトル $(1, 1, \dots, 1)^T$ である. そして

$$\mathbf{y}_i = \mathbf{I} + \alpha \frac{\Delta \mathbf{y}_i}{\|\Delta \mathbf{y}_i\|} \quad (34)$$

と更新する. このとき $0 < \alpha < 1$ であれば $\mathbf{y} > \mathbf{0}$ となり, かつ実行可能内点である. 式(34)の $\mathbf{y}_i$ に式(26)の逆変換を施すと

$$\mathbf{x}_i = \mathbf{x}_i^* + \alpha D_i^* \frac{\Delta \mathbf{y}_i}{\|\Delta \mathbf{y}_i\|} \quad (35)$$

として式(21)-(23)の新たな実行可能内点 $\mathbf{x}_i$, $i=1, \dots, N$ を求めることができる. アフィン変換法によるアルゴリズムを以下に示す.

ステップ0 初期実行可能内点を

$$x_{ij} = \frac{1}{M}, \quad i=1, \dots, N, \quad j=1, \dots, M$$

とする.

ステップ1 式(32)と式(35)により $\mathbf{x}$ を更新する. 式(24)より $\Delta \mathbf{r}$ を求め

$$\mathbf{r} := \mathbf{r} + \beta \Delta \mathbf{r}$$

としてクラスタ中心 $\mathbf{r}$ を更新する.

ステップ2 $x_{ij} > \theta_U$ であれば $x_{ij} := 1$ とし, 第 i 単体に対応するその他の x_{ij} は $x_{ij} := 0$ とする.

ステップ3 反復回数が事前に定めた値より大きくなると係数 A を定数倍して 0 に近づける.

ステップ4 すべての x_{ij} が 0 または 1 となればステップ5へ, その他はステップ1へ行く.

ステップ5 M 個のクラスター中心を求め終了.

A を固定して, メンバシップの最大となるクラスターに分類する方法や, A についての勾配を用いて A を次第に 0 に近づける方法では最適解の得られる頻度があまり高くないので, ステップ2の θ_0 とステップ3を設けている.

クラスタリング問題(21)–(23)をラグランジュ乗数法で解く場合は,

$$L = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij} \| \mathbf{d}_i - \mathbf{r}_j \|^2 + \sum_{i=1}^N \lambda_i \left(\sum_{j=1}^M x_{ij} - 1 \right) + \frac{A}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij} (1 - x_{ij}) \quad (36)$$

として, 最適性の必要条件より

$$x_{ij} = \frac{\| \mathbf{d}_i - \mathbf{r}_j \|^2}{2A} - \frac{\sum_{k=1}^M \| \mathbf{d}_i - \mathbf{r}_k \|^2}{2MA} + \frac{1}{M} \quad (37)$$

$$\mathbf{r}_j = \frac{\sum_{i=1}^N x_{ij} \cdot \mathbf{d}_i}{\sum_{i=1}^N x_{ij}} \quad (38)$$

と求まるので, 式(37)と式(38)を交互に計算することにより解を求める.

この方法では, 上述のアフィン変換が使えないために反復計算中に式(37)の x_{ij} が負になることがしばしば起こる. そのために良くない局所解に収束することが多く最適解を得る頻度はあまり高くない.

4. ファジィ c-Means によるハードクラスタリング

Li and Mukaidono[5][6]はエントロピー最大化によるファジィクラスタリング法を提案しているが, この方法は従来のファジィ c-Means と同種の正則化条件を付加したものであり, ラグランジュ乗数法を用いることができることが示されている[16]. クラスタリングを次の最適化問題とする.

$$\min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij} \| \mathbf{d}_i - \mathbf{r}_j \|^2 - \frac{A}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij} \log x_{ij} \quad (39)$$

$$\text{s.t.} \quad \sum_{j=1}^M x_{ij} = 1, \quad i=1, \dots, N \quad (40)$$

$$x_{ij} \in \{0, 1\} \quad (41)$$

最小化問題であるので $A < 0$ のときエントロピー最大化によるファジィクラスタリングとなる. 図1に示すように式(21)第2項目の非整数解の出現を制御する項とエントロピーの $-x_{ij} \log x_{ij}$ は類似した形の関数であるのでエントロピー最大化によるファジィ c-Means と3章の解は類似していると考えられる. またこのことからアフィン変換 $y_{ij} = x_{ij}/x_{ij}^*$ を施し,

$$\min \sum_{i=1}^N \sum_{j=1}^M x_{ij}^* y_{ij} \| \mathbf{d}_i - \mathbf{r}_j \|^2 - A \sum_{i=1}^N \sum_{j=1}^M x_{ij}^* y_{ij} \log(x_{ij}^* y_{ij}) \quad (42)$$

$$\text{s.t.} \quad \sum_{j=1}^M x_{ij}^* y_{ij} = 1, \quad i=1, \dots, N \quad (43)$$

$$y_{ij} \geq 0 \quad (44)$$

とし, 3章と同様の内点法で解くことができる. 勾配法によるので, x_{ij} が 0 に近くなることがあり数値計算が不安定になるためにアフィン変換を用いる必要がある. ラグランジュ乗数法による場合は

$$L = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij} \| \mathbf{d}_i - \mathbf{r}_j \|^2 + \sum_{i=1}^N \lambda_i \left(\sum_{j=1}^M x_{ij} - 1 \right) - \frac{A}{2} \sum_{i=1}^N \sum_{j=1}^M x_{ij} \log x_{ij} \quad (45)$$

すると, 最適性の必要条件より

$$x_{ij} = \frac{e^{\frac{1}{2} \| \mathbf{d}_i - \mathbf{r}_j \|^2 / A}}{\sum_{k=1}^M e^{\frac{1}{2} \| \mathbf{d}_i - \mathbf{r}_k \|^2 / A}} \quad (46)$$

となる. またクラスター中心 $\mathbf{r}_j$ は

$$\mathbf{r}_j = \frac{\sum_{i=1}^N x_{ij} \mathbf{d}_i}{\sum_{i=1}^N x_{ij}} \quad (47)$$

図1 $x(1-x)$ と $-x \log x$

と求まるので、式(46)、(47)を交互に計算することにより解を求める。

この方法では対数関数を用いているので x_{ij} が負になることがなく、5章の数値実験に示すように高い頻度で最適解が得られるが、反復計算中に x_{ij} が0に近くなると数値計算が不安定になり実行可能解が得られない場合がある。

5. 数値実験

2次元10個のパターン集合6つを用いて数値実験を行った。ただし、パターンはすべて $[0,1] \times [0,1]$ の内部に基準化されている。表1に数値実験に使用したパターン集合を示す。クラスターの中心はすべて同一の系列の乱数により与えた。Bestは最適解の得られた回数、Secondは2番目に良い解の得られた回数である。判定閾値 θ_v は予備的な実験から0.6として数値実験を行った。係数Aの値は反復100回以後0.2倍に大きくし0に近づけた。図2は3章の非整数解の出現を制御する項を用いた内点法による場合の最適な係数Aの値でのk-Meansアルゴリズムとの比較を示している。図3はパターン集合1の場合の最適解を示している。図4はエントロピーを用いる場合の、アフィン変換法とラグランジュ乗数法との比較を示している。係数Aを適切に選ばばかなりの頻度で最適解が得られている。アフィン変換を用いないと局所解に陥ることが多く、最適解が得られる割合が小さかった。またAの値を次第に0に近づける緩和型解法よりも閾値 θ_v を設ける方が最適解を得る割合が高かった。図2と図4に示すように係数Aと閾値 θ_v をうまく選んだファジィc-Means法はハードクラスタリングの最適解を得る割合がk-Meansアルゴリズムに比べてかなり高くなっている。以上のシミュレーション結果をまとめると

1) ホップフィールドニューラルネットによる外点法では設定すべき2つのパラメータA、Bと2つの閾値 θ_L 、 θ_U が必要でその最適値を経験的に定めるのが困難である[17]。一方、内点法では各々1つずつであり、制約領域の内部に閉じ込めて探索するので、最適解を得る頻度も大幅に改善される。

図2 k-Means法とホップフィールドアプローチによるアフィン変換法の比較
(左: k-Means法、右: アフィン変換法)

図3 パターン集合1の最適解

表1 2次元10個のパターン集合

	1	2	3	4	5	6
1	0.2439 0.1463	0.1358 0.6079	0.2534 0.7022	0.5393 0.8906	0.7394 0.5776	0.1927 0.6295
2	0.8488 0.3609	0.7830 0.5522	0.5161 0.6295	0.4366 0.6922	0.5462 0.3370	0.4008 0.7844
3	0.6683 0.2536	0.7255 0.7072	0.0996 0.4451	0.2271 0.2645	0.0896 0.1198	0.5408 0.5531
4	0.6683 0.2536	0.5849 0.8508	0.5933 0.1323	0.7509 0.3394	0.1934 0.4894	0.7502 0.7507
5	0.1707 0.2293	0.1870 0.7702	0.6905 0.8364	0.0983 0.7566	0.8342 0.5881	0.8187 0.0949
6	0.2293 0.7610	0.3657 0.3376	0.3907 0.4007	0.6683 0.6017	0.4743 0.7402	0.6633 0.4380
7	0.4000 0.4439	0.4331 0.1722	0.4964 0.2939	0.7698 0.0851	0.8478 0.2528	0.4259 0.3608
8	0.8732 0.6536	0.2647 0.3235	0.7658 0.9098	0.8477 0.2441	0.3822 0.8657	0.1557 0.8633
9	0.5171 0.9414	0.6353 0.4251	0.9060 0.5250	0.3300 0.4684	0.2726 0.1886	0.8484 0.2753
10	0.6195 0.2634	0.8426 0.1472	0.2156 0.2088	0.2657 0.5703	0.6629 0.8002	0.3105 0.1607

図4 エントロピー最大化のファジィクラスタリングに対するアフィン変換法とラグランジュ乗数法の比較 (左: アフィン変換法, 右: ラグランジュ乗数法)

2) 非整数解の出現を制御する項, またはエントロピー項の係数を次第に 0 に近づける緩和型解法や, ファジィクラスタリング結果を基に最大メンバシップ値でハードクラスターに分類する方法よりも, メンバシップの閾値 θ_v を設ける方法が最適解を得る頻度が高い。

3) エントロピー最大化によるファジィクラスタリングに勾配斜影法を直接適用すればメンバシップ値 x_{ij} が非常に 0 に近くなるため数値計算が不安定になる。そのためにアフィン変換を用いる必要がある。

4) エントロピー最大化によるファジィクラスタリングで係数 A と閾値 θ_v を適切に設定したラグランジュ乗数法は, 計算効率が良く, 最適解を得る頻度も高い。一方, A が 0 に近いとメンバシップ x_{ij} が 0 に近づきすぎて解が得られない場合も発生する。

6. おわりに

本研究では混合整数計画問題としてのクラスタリング問題を内点法やラグランジュ乗数法により解き, 目的関数の係数とメンバシップの閾値をうまく設定することにより, 高い頻度で最適解が得られることをシミュレーションで確認した。

ファジィ c -Means はクラスター毎の主成分ベクトルの張る空間への距離に基づいて分類するファジィ c -Varieties[18]や, 回帰とクラスタリングの同時分析法[4]などに発展してきている。これらは見方をかえると主成分分析や重回帰分析などの多変量解析とクラスタ分析の同時分析法である。このような観点から数量化分析との同時分析法も試みられており[19], 今後さまざまなデータ解析手法との同時分析法へ発展することが期待できる。その際にも最適なクラスタリング結果が得られるように提案手法のさらなる改善を図っ

ていくことが今後の課題である。

参考文献

- [1] 宮本 定明: “階層的クラスタリングの方法 IV,” システム/制御/情報, Vol.34, No.6, pp.347-355(1990).
- [2] J.C.Bezdek: “Pattern recognition with fuzzy objective function algorithms,” Plenum, New York, (1981).
- [3] 宮腰 政明: “ファジィクラスタリングの画像解析への応用,” 日本ファジィ学会誌, Vol.8, No.3, pp.440-447(1996).
- [4] 領家 美奈, 中森 義輝: “適応型ファジィクラスタリングによる分類と回帰の同時分析,” 日本ファジィ学会誌, Vol.8, No.1, pp.136-146(1996).
- [5] M.Mukaidono and R.P.Li: “Gaussian clustering method based on maximum -fuzzy- entropy interpretation,” Fuzzy Sets and Systems (to appear).
- [6] R.P.Li and M.Mukaidono: “A maximum entropy approach to fuzzy clustering,” Proc. of the 4th IEEE Intern. Conf. on Fuzzy Systems (FUZZ-IEEE/IFES'95), pp.2227-2232(1995).
- [7] 浦浜 喜一: “組合せ最適化のニューラル解法の数理論的基礎,” 電子情報通信学会論文誌 A, Vol.J78-A, No.1, pp.47-53(1995).
- [8] 宮本 定明, 向殿 政男: “エントロピー正則化法によるファジィクラスタリング,” 第13回ファジィシステムシンポジウム講演予稿集, pp.121-122(1997).
- [9] J.Hopfield and D.Tank: “Neural computation of decisions in optimization problems,” Biol. Cybern., Vol.52, pp.141-152(1987).
- [10] 喜多 一, 小谷 英之, 西川 緯一: “神経回路モデルによる最適化計算—回路ブロック配置問題を例として,” システム制御情報学会論文誌, Vol.4, No.1, pp.48-56(1991).
- [11] 今野 浩, 山下 浩: “非線形計画法,” 日科技連(1978).
- [12] I.I.Dikin: “Iterative solution of problems of linear and quadratic programming,” Soviet Mathematics Doklady, Vol.8, pp.674-675(1967).
- [13] 松田 聖: “対称結合神経回路網における解の安定性,” 電子情報通信学会論文誌 D-II, Vol.J77-D-II, No.7, pp.1366-1374(1994).
- [14] 茨木 俊秀, 福島 雅夫: “最適化の手法,” 共立出版(1993).
- [15] J.B.Rosen: “The gradient projection method for non-linear programming. Part I, Linear constraints,” SIAM J., Vol.8, pp.181-217(1960).
- [16] S.Miyamoto and Y.Agusta: “An efficient algorithm

m for fuzzy c-means and its termination," J. Cont. Cybern., Vol.24, No.4, pp.421-436(1995).

- [17]市橋 秀友, 李 九炯, 三好 哲也: "内点法による k -Means クラスタリング問題の連続緩和型解法," 第12回ファジィシステムシンポジウム講演論文集, pp.871-872(1996).
- [18]J.C.Bezdek.: "Detection and characterization of cluster substructure II. Fuzzy c-varieties and convex combinations thereof," SIAM J. Appl. Math., Vol. 40, No.2, pp.358-372(1981).
- [19]金海 好彦, 市橋 秀友, 三好 哲也: "ごみ箱クラスターによるコレスポネンス分析," 第2回ファジィOR ミニシンポジウム講演論文集, pp.86-89(1997).

[付録]

$\nabla_{x_i} E$ の単体への直交射影を求めるために $\mathbf{x}_i$ の修正量を次の凸2次計画問題の解として求める.

$$\min \frac{1}{2} \Delta \mathbf{x}_i^T \Delta \mathbf{x}_i \quad (48)$$

$$\text{s.t. } \mathbf{I}^T \Delta \mathbf{x}_i = 0, \quad (49)$$

$$\nabla_{x_i} E^T \Delta \mathbf{x}_i = \varepsilon \quad (50)$$

式(48)は $\mathbf{x}$ の修正量 $\Delta \mathbf{x}$ の長さを最小にすること, 式(49)は

$$\mathbf{I}^T \mathbf{x}_i - 1 = 0, \quad i=1, \dots, N \quad (51)$$

が $\mathbf{x}_i$ が $\Delta \mathbf{x}_i$ だけ修正されても

$$\mathbf{I}^T (\mathbf{x}_i + \Delta \mathbf{x}_i) - 1 = 0 \quad (52)$$

となるための条件を表している. 式(50)は目的関数 E の値が

$$E(\mathbf{x}_i + \Delta \mathbf{x}_i) \approx E(\mathbf{x}_i) + \nabla_{x_i} E^T \Delta \mathbf{x}_i \quad (53)$$

と変化したとき, その変化量 $\nabla_{x_i} E^T \Delta \mathbf{x}_i$ を $\varepsilon (< 0)$ に等しくすることを表していて, $\Delta \mathbf{x}_i$ は $\nabla_{x_i} E$ を式(22)の単体に直交射影した方向ベクトルとなる. 凸2次計画問題(48)-(50)を次のラグランジュ関数 $L(\Delta \mathbf{x}, \lambda)$ の最小化問題とする.

$$\begin{aligned} \min L(\Delta \mathbf{x}, \lambda) = & \frac{1}{2} \Delta \mathbf{x}^T \Delta \mathbf{x} \\ & + \lambda_i \mathbf{I}^T \Delta \mathbf{x}_i + \lambda_{N+1} (\nabla_{x_i} E^T \Delta \mathbf{x}_i - \varepsilon) \end{aligned} \quad (54)$$

式(54)を最小化するための必要十分条件は

$$\frac{\partial L}{\partial \Delta \mathbf{x}_i} = \Delta \mathbf{x}_i + \lambda_i \mathbf{I} + \lambda_{N+1} \nabla_{x_i} E = 0, \quad (55)$$

$$\frac{\partial L}{\partial \lambda_i} = \mathbf{I}^T \Delta \mathbf{x}_i = 0, \quad (56)$$

$$\frac{\partial L}{\partial \lambda_{N+1}} = \nabla_{x_i} E^T \Delta \mathbf{x}_i - \varepsilon = 0 \quad (57)$$

式(55)より $\Delta \mathbf{x}_i$ は $\nabla_{x_i} E$ と $\mathbf{I}$ に1次従属であり, 式(56)より $\mathbf{I}$ に直交する. その符号と大きさは式(57)より決まる. したがって式(55), (56)を満たす方向ベクトル $\Delta \mathbf{x}_i$ はグラム・シュミットの直交化法により

$$\begin{aligned} \Delta \mathbf{x}_i &= \nabla_{x_i} E - \left(\frac{\nabla_{x_i} E^T \mathbf{I}}{\mathbf{I}^T \mathbf{I}} \right) \mathbf{I} \\ &= \nabla_{x_i} E - \left(\frac{\nabla_{x_i} E^T \mathbf{I}}{M} \right) \mathbf{I} \end{aligned}$$

となる.

(1997年10月30日 受付)

[問い合わせ先]

〒599-8531

大阪府堺市学園町1-1

大阪府立大学工学部経営工学科

市橋 秀友

TEL : 0722-54-9352

FAX : 0722-54-9915

E-mail : ichi@ie.osakafu-u.ac.jp

著者紹介

太田 知宏 (おおた ともしろ)

シャープ(株)

1998年 大阪府立大学大学院 工学研究科博士前期課程電気・情報系専攻修了, 同年シャープ(株)入社, 現在に至る. ファジィクラスタリング, ファジィモデリングに興味を持つ.

根本 宗記 (ねもと むねき)

大阪府立大学 大学院工学研究科

1997年 大阪府立大学工学部 経営工学科卒業, 同年同大学院工学研究科博士前期課程電気・情報系専攻入学, 現在に至る. JAVAプログラミング, ファジィクラスタリングに興味を持つ.

市橋 秀友 (いちはし ひでとも)

大阪府立大学 工学部経営工学科

1971年 大阪府立大学工学部経営工学科卒業。同年松下電器産業株式会社入社。1981年大阪府立大学工学部経営工学科助手、1987年同講師、1989年同助教授、1993年同教授、現在に至る。工学博士、ファジィ多目的計画法、ファジィ制御、ニューロ・ファジィなどのファジィ理論の応用研究に従事。国際ファジィシステム学会、日本ファジィ学会、システム制御情報学会、電子情報通信学会、日本経営工学会などの会員。

三好 哲也 (みよし てつや)

大阪府立大学 工学部経営工学科

1989年 大阪府立大学大学院工学研究科博士前期課程経営工学専攻修了。同年大阪府入庁、1993年大阪府立大学工学部経営工学科助手、現在に至る。ファジィ制御、ニューロ・ファジィなどのファジィ応用研究に従事。日本経営工学会、日本ファジィ学会などの会員。

Hard Clustering by Fuzzy c-Means

by

Tomohiro OHTA, Muneki NEMOTO, Hidetomo ICHIHASHI and Tetsuya MIYOSHI

Abstract :

This paper proposes an efficient usage of the fuzzy c -Means clustering algorithm to obtain optimum solutions of the k -Means hard clustering problem with reasonable certainty. The k -Means clustering problem is formulated as a mixed integer programming problem. Based on the studies about the stability of solution in a multi-linear form of the energy function of the Hopfield neural network, it is shown that by estimating a local minimum solution in the hypercube of solution space, the coefficients of energy function and a threshold value for deciding 0-1 integer valued solution can be properly estimated. The fuzzy c -Means problem is solved by the affine scaling interior point method for linear programming problems and the Lagrangian multiplier method for maximizing entropy fuzzy clustering. It is shown by numerical simulations that both of the methods outperform the conventional k -Means algorithm in the quality of solutions found.

Keywords : Hopfield Neural Network, Fuzzy c -Means, Entropy, Hard Clustering

Contact Address : **Hidetomo ICHIHASHI**

College of Engineering, Osaka Prefecture University

1-1 Gakuen-cho, Sakai, Osaka 599-8531, Japan

TEL : 0722-54-9352

FAX : 0722-54-9915

E-mail : ichi@ie.osakafu-u.ac.jp