論文メモ

文献番号	0004
日付	2021 年 09 月 27 日
名前	武川海斗

文献情報

著者	Sadaaki Miyamoto, Youhei Kuroda, and Kenta Arai
英文タイトル	Algorithms for Sequential Extraction of Clusters by Possibilistic Method and Comparison with Mountain Clustering
和文タイトル	Possibilistic 法によるクラスターの逐次抽出のアルゴリズムと Mountain Clustering との比較
書誌情報	Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.12 No.5, p448-453,2008
キーワード	possibilistic clustering, sequential extraction of clusters, mountain clustering

1　論文のトピック

本論文では、Possibilistic 法による逐次抽出法を提案を行っている。Possiblistic 法は、目的関数の最適化に基づいており、二つの目的関数についての二章で説明が行われている。

その後、提案手法として、三つの逐次抽出法クラスタリングを提案する。その後、mountain クラスタリングとの性能評価を行い、提案手法の重要性を確認する。

2　ベースとなった手法

2.1　Possiblistic 法クラスタリング

Possiblistic 法クラスタリングは目的関数の最適化に基づいており、以下の 2 つの式 (1,2) を定義する。ここで、λ,ζ はハイパーパラメータでり、u_{ki} はメンバーシップ関数、D_{ki} は個人とクラスタ中心の間の標準的な二乗ユークリッド距離である。

$$J_e(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}\left\{u_{ki}D_{ki} + \lambda^{-1}u_{ki}\left(\log u_{ki} - 1\right)\right\} \tag{1}$$

$$J_2(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}\left\{\left(u_{ki}\right)^2 D_{ki} + \zeta^{-1}\left(1 - u_{ki}\right)^2\right\} \tag{2}$$

2.2　mountain クラスタリング

mountain クラスタリングは、クラスターを順次、すなわち一度に 1 つずつ抽出するものである。特徴として、y は格子点に限定されることが挙げられる。式 (3) は、mountain クラスタリングの更新式である。この更新式が停止するまで繰り返す。

$$M^{(\ell)}(y) = M(\ell-1)(y) - M\left(y^{(\ell-1)}\right)\sum_{k=1}^{n}\exp\left(-\alpha D\left(y^{(\ell-1)}, y\right)\right) \tag{3}$$

3　提案手法のコア要素

3.1　ProcedureA

ProcedureA の手法は以下のようになる。

Procedure A

A1.　候補点 $y_1, ...y_L \in R^p$ を生成する。$X^{(0)} = X$ と $k = 0$ です。

A2.　最小化要素を探す

$$\bar{y} = \arg \min_{v=y_1,\ldots,y_L} J(v;k) \tag{4}$$

A3.　$\bar{y}$ を中心としたクラスタ $G^{(k)}$ を見つけ、$G^{(k)} : X^{(k+1)} = x^{(k)} - G^{(k)}$ を抽出します。$X^{(k+1)}$ がもう一つのクラスタを抽出するのに十分な要素を持っていない場合、停止します; そうでなければ $k := k+1$ で A2 に進みます。

点 $y_1, ..., y_L$ は、mountain クラスタリングと同様に、格子状の点で値を取るか、または X からランダムに選ぶ。この手法は　、

3.2　ProcedureB

この手法は、fussy c−means の手法と最適化方法が似ている。そのため、私が行う研究である、「ガウス過程に基づく c− 回帰逐次抽出法」にも、このアルゴリズムを参考にできると考えている。

Procedure B

B1.　候補点 $y_1, ...y_L \in R^p$ を初期のクラスタとして生成します。$X^{(0)} = X$ と $k = 0$.

B2.　収束するまで uki と vi の計算を繰り返します。収束した点を z1,...zl とします。最小化する要素を求める。

$$\bar{z} = \arg \min_{v=z_1,\ldots,z_\ell} J(v;k) \tag{5}$$

B3.　中心が $\bar{z}$ のクラスタ $G^{(k)}$ を見つける。$G^{(k)} : X^{(k+1)} = X^{(k)} - G^{(k)}$ を抽出する。$X^{(k+1)}$ がもう 1 つのクラスタを抽出するのに十分な要素を持っていない場合は停止し、そうでない場合は $k := k+1$ して B2 に進みます。

3.3　ProcedureC

generalized multi pass possibilistic medoid clustering とも呼ばれる手法である。

Procedure C

C1　候補点 $y_1, ..y_L \in X$ を生成し、$Y = y_1, ...y_L$ から クラスタ中心の初期値 $z_1, ...z_c$ を選択する。$X^{(0)} = X$ と $k = 0$ である。

C2　収束するまで C3 を繰り返す

C3　$z_i(i = 1, ...c)$ に対する K-nearest 要素を $y_{i1}, ...y_{ki} \in Y$ とする。最小化要素を探す.

$$\bar{z}_i = \arg \min_{v=z_i,y_{i1},\ldots,y_{ki}} J(v;k) \tag{6}$$

$z_i = \bar{z}_i$ とおく.

C4

$$\bar{z} = \arg \min_{v=z_i,y_{i1},\ldots,y_{ki}} J(v;k) \tag{7}$$

4　実験デザイン・結果と考察

Mountain クラスタリングの比較実験として、200 個の点を持つ p 次元のデータセットをランダムに生成した数値実験を行った。表 4 は比較実験の結果である。表 4 によると、Mountain クラスタリングは次元数に応じて実行時間も指数関数的に増加する。一方提案手法では実行時間の急激な増加がないことがわかった。

Dimension	Mountain	Procedure A
p(次元)	(ms)	(ms)
2	24.84	30.46
3	772.96	35.62
4	20453.333	42.03
5	485758.59	60.47
6	-	74.68

5　手法の限界・今後の課題

6　特に重要な関連研究

参考文献 [1] は、「一度に一つのクラスタ」を生成する手法について提案された論文である。ノイズクラスタリングについても述べられており、参考文献として読むべき論文であると言える。

次に読むべき論文のリスト

[1] R. N. Dave and R. Krishnapuram, “Robust clustering methods: a unified view,” IEEE Trans. Fuzzy Syst, Vol.5, No.2, pp. 270-293, 1997.