

論文メモ

文献番号	0005
日付	2021 年 09 月 29 日
名前	武川海斗

文献情報

著者	Sadaaki Miyamoto, Kenta Arai
英文タイトル	Different Sequential Clustering Algorithms and Sequential Regression Models
和文タイトル	逐次型クラスタリングアルゴリズムと逐次型回帰モデルの違い
書誌情報	FUZZ-IEEE, pp1107-1112, 2009
キーワード	なし

1 論文の要約

本稿で注目する項目である、論文のトピック、ベースとなった手法、提案手法のコア要素、実験デザイン・結果と考察、手法の限界・今後の課題についてまとめる。この項目は最後に作成する。

2 論文のトピック

本論文では、クラスタを逐次抽出する手法として、3つのアプローチから4つのアルゴリズムを開発した。

3 ベースとなった手法

3.1 Possibilistic クラスタリング

Possibilistic 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 (2,1) を定義する。この式は、論文 [1] によって提案されている。ただし、 $m = 2, \eta_i = \zeta^{-1}(i = 1, \dots, c)$ とする。

$$J_2(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ki})^m D_{ki} + \sum_{i=1}^c \eta_i \sum_{k=1}^n (1 - u_{ki})^m \quad (1)$$

$$J_e(U, V) = \sum_{k=1}^n \sum_{i=1}^c \{u_{ki} D_{ki} + \lambda^{-1} u_{ki} (\log u_{ki} - 1)\} \quad (2)$$

4 提案手法のコア要素

4.1 Possibilistic クラスタリングと逐次抽出クラスタリング

このアルゴリズムの詳細については、参考文献リスト P0004の方が説明がされている。アルゴリズムについてのみ下記に記す。

SC: General Procedure for Sequential Clustering Algorithms

SC1. 初期の対象物の集合を $X^{(0)} = X$ 、 $k = 0$ とし、対象物の集合 $X^{(k)}$ を持つ関数 $J(v; k) = j_e(v)$ (または $J(v; k) = j_2(v)$) とする。

SC2. $J(v; k)$ の最小化要素を検索します。

$$v^{(k)} = \arg \min_v J(v; k) \quad (3)$$

SC3. 中心 $v^{(k)}$ に属するクラス $G^{(k)}$ を抽出します。

SC4. $X^{(k+1)} = X^{(k)} - G^{(k)}$ とする。 $X^{(k+1)}$ がもう 1 つのクラス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 $k := k + 1$ とし、ステップ SC2 に進む。

END SC.

4.2 Mountain medoid クラスタリング

Mountain クラスタリングとは、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。目的関数に対する更新式が停止するまで繰り返す。

4.3 ノイズクラスタリングに基づいた逐次抽出クラスタリング

ノイズクラスタリングの方法を検討する。この方法では、次の 2 つの目的関数のいずれかを用いる。

$$J_{2n}(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ki})^m D_{ki} + \sum_{k=1}^n (u_{k0})^m \delta \quad (4)$$

$$J_{en}(U, V) = \sum_{k=1}^n \sum_{i=1}^c \{u_{ki} D_{ki} + \lambda^{-1} u_{ki} \log u_{ki}\} + \sum_{k=1}^n u_{k0} \delta \quad (5)$$

ここで、 u_{k0} はノイズクラスへの帰属度であり、 $0; \delta > 0$ は、すべてのオブジェクトがノイズ・クラスターから一定の非類似度 δ を持つことを示すパラメータである。この目的関数を、 $c = 1$ に設定して、アルゴリズム SC で使用する。

4.4 回帰モデルの逐次抽出型クラスタリング

4.3 節と同様の手法で、回帰モデルについて当てはめて考える。そこで、以下の 2 式の目的関数を考える。ここで $(x_k, y_k)_{k=1, \dots, n}$ で、 $x_k \in \mathbf{R}^p$ は p 次元のデータとする。目的関数 (4) を (6) (または (5) を (7)) に置き換えると、同じ逐次アルゴリズム SC を使って複数の回帰モデルを持つことができる。

$$J_{2nr}(U, B) = \sum_{k=1}^n (u_{ki})^m \left(y_k - \sum_{j=1}^p \beta_j x_k^j + \beta_{p+1} \right)^2 + \sum_{k=1}^n (u_{k0})^m \delta \quad (6)$$

$$J_{enr}(U, B) = \sum_{k=1}^n \sum_{i=1}^c u_{ki} \left(y_k - \sum_{j=1}^p \beta_j x_k^j + \beta_{p+1} \right)^2 + \lambda^{-1} \sum_{k=1}^n \sum_{i=1}^c u_{ki} \log u_{ki} + \sum_{k=1}^n u_{k0} \delta \quad (7)$$

5 実験デザイン・結果と考察

ここでは、自身の研究と関連があるものとして、回帰モデルに関する実験についてのみ述べる。この実験では様々な国の GDP とエネルギー消費の関係についての実データを使用している。この図を見ると、実際にクラスタリングできているのが確認できる。

6 手法の限界・今後の課題

逐次抽出型回帰モデルの実験が少ないように感じた。他の実データでも通用するかどうか実験をして確かめていくことが課題である。

7 特に重要な関連研究

論文 [1] は、possibilistic クラスタリングの目的関数について詳細に述べた論文である。また、論文 [2] は、ノイズクラスタに基づくクラスタリングについて記した論文である。逐次抽出法の先駆け的な論文であり、至急読む必要があると考えている。最後に、[3] は mountain クラスタリングについて提唱した論文で、mountain クラスタリングとの比較実験をする場合、読む必要がある論文であると言える。

次に読むべき論文のリスト

- [1] R. Krishnapuram, J.M. Keller, "A possibilistic approach to clustering", IEEE Trans. on Fuzzy Systems , Vol.1, No.2, pp. 98-110, 1993.
- [2] R. N. Dave and R. Krishnapuram, "Robust clustering methods: a unified view," IEEE Trans. Fuzzy Syst, Vol.5, No.2, pp. 270-293, 1997.
- [3] Ronald R. Yager and Dimitar P Filev,"Approximate Clustering Via the Mountain Method", IEEE TRANSACTIONS ON SYSTEMS,MAN, AND CYBERNETICS, 1994