

# 論文メモ

文献番号	0006
日付	2021 年 10 月 13 日
名前	武川 海斗

## 文献情報

著者	Rajesh N. Davé and Raghu Krishnapuram
英文タイトル	Robust Clustering Methods: A Unified View
和文タイトル	ロバストなクラスタリング手法についての統一的理解
書誌情報	IEEE TRANSACTIONS ON FUZZY SYSTEMS, Vol. 5, No. 2, pp. 270–293, 1997
キーワード	Clustering validity, fuzzy clustering, robust methods.

## 1 論文の要約

ロバストなモデルとは、ノイズに影響を大きく受けないモデルのことである。実際のモデル運用には外れ値やノイズがつきまとうため、ロバストなモデルが必要となる場面は多い。

本論文では、ロバストなクラスタリングの手法を多数紹介している。しかし、その手法の多くが本質的に同じであることを著者は主張している。

## 2 論文のトピック

本論文では、ロバストなクラスタリングモデルとして、ノイズクラスタリング、可能性クラスタリングを中心に取り上げている。これらのクラスタリングと Ohashi-アルゴリズムや Mountain アルゴリズムなどと比較している。ロバスト統計学についても取り上げている。ロバスト統計学とファジィクラスタリングの関連性について深く述べている。また、ほとんどのクラスタリング手法は目的関数の最小化問題に帰結するため、クラスタリングの妥当性の重要性についても述べている。

## 3 提案手法のコア要素

### 3.1 ノイズクラスタリング

参考文献 [1] にて、Davé が提唱した手法である。  $\delta^2$  はノイズパラメータになっており、  $d^2(\mathbf{x}_j, \beta_i)$  と比較して、ノイズクラスタ項に属するかどうか比較する。この手法は、私自身が行っている研究の基となるため、重要な手法であると言える。

$$J(B, U; X) = \sum_{i=1}^C \sum_{j=1}^N (u_{ij})^m d^2(\mathbf{x}_j, \beta_i) + \sum_{j=1}^N \delta^2 \left( 1 - \sum_{i=1}^C u_{ij} \right)^m \quad (1)$$

式を  $u_{ij}$  について微分を行うと、以下の式が得られる。

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left[ \frac{d^2(\mathbf{x}_j, \beta_i)}{d^2(\mathbf{x}_j, \beta_k)} \right]^{1/(m-1)}} + \left[ \frac{d^2(\mathbf{x}_j, \beta_i)}{\delta^2} \right]^{1/(m-1)} \quad (2)$$

式 (2) は,  $u_{ij}$  について解いた結果である. ここで, 式 (2) の分母の第 2 項について, ノイズパラメータが含まれているのがわかる. つまり, 大きいノイズパラメータを与えた場合,  $u_{ij}$  が大きくなり, 帰属している度合いが小さくなることがわかる.

### 3.2 可能性クラスタリング

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left[ \frac{d^2(\mathbf{x}_j, \beta_i)}{d^2(\mathbf{x}_j, \beta_k)} \right]^{1/(m-1)} + \left[ \frac{d^2(\mathbf{x}_j, \beta_i)}{\delta^2} \right]^{1/(m-1)}} \quad (3)$$

## 4 実験デザイン・結果と考察

### 5 手法の限界・今後の課題

データのノイズとクラスタ数が不明な場合, クラスタリングを困難である. 本論文で紹介した手法は, ノイズをクラスタリングできる手法であるが, クラスタ数が未知の場合は, 信頼性の高いクラスタリングを行うことは難しい. そのため, クラスタ数の代わりとなるパラメータを網羅的に与えてあげる必要がある. 例えば, ノイズクラスタリングでは, クラスタ数の代わりに, ノイズパラメータ  $\sigma^2$  を採用しており, クラスタ数の問題に置き換えたものになっている. しかし, 最適化の手法や妥当性基準が変わるため, 意味のあるものであると言える.

最後に, 本論文では, さまざまなクラスタリング手法を, ロバスト性の観点から考察し, いくつかの手法は等価であると主張した. そのため, 研究者が車輪を再発明しないようにするよう注意する必要がある.

### 6 特に重要な関連研究

参考文献 [1] は, ノイズクラスタリングの手法について提唱した論文である. 本論文でも解説がなされていたが, ノイズパラメータに関する細かい考察がされているため, 重要な論文であると言える.

### 次に読むべき論文のリスト

- [1] R. N. Davé, Characterization and detection of noise in clustering, *Pattern Recognition Letters*, Vol. 12, No. 11, pp. 657–664, 1991