

クラスタサイズ調整変数を導入した クラスタリング手法の特性比較及び精度評価

Characteristic Comparison and Accuracy Evaluation of Clustering Method
with Cluster Size Adjustment Variable

AF16009 池辺 颯一
Soichi Ikebe

指導教員 神澤 雄智
Yuchi Kanzawa

1 はじめに

近年、情報通信社会の発展に伴いデータ量が増大し、日々多様なデータがコンピュータに蓄積されている。この大量のデータから有益な情報を抽出する手法として、データを類似度に基づきグループ化するクラスタリングに注目が集まっている。既存の手法における課題として、各クラスタのサイズに差がある場合、クラスタリングから有意な結果が得られないというものがある。そこで、各クラスタのサイズを考慮してクラスタリングを行う手法が複数提案されており、本研究はそれらの手法について各手法の特性を把握するとともに、最も有用な手法を発見することを目的とする。

2 提案内容

各クラスタのサイズを考慮するために、既存の手法にクラスタサイズ調整変数を導入した eFCMA, qFCMA, sFCMA の 3 手法について実験を行う。

まず、これらの手法についてそれぞれの特性を把握するため、人工データを用いて実験を行う。複数のパラメータで実験を行い、それぞれで算出された分類関数から比較及び評価を行う。

次に、これらの手法から最も有用なものを発見するために、実データを用いて ARI(Adjusted Rand Index) を算出し、その値が最も高いものを有用な手法と評価する。

3 人工データの実験結果

人工データとして、クラス数 2、各クラスのデータ数 50、合計データ数 100 のデータを平均値 $(-1, -1)$ 、標準偏差 $(0.5, 0.5)$ 及び平均値 $(1, 1)$ 、標準偏差 $(0.5, 0.5)$ のガウスサンプリングで生成したデータを用いた。

eFCMA の実験結果を図 1a, 1b に示す。垂直軸は分類関数値を、底面はデータ空間を表す。網掛けで示されるのが分類関数であり、各点がデータを表している。パラメータ λ を 1 から 10000 に変化させたところ、 λ の値を大きくすることにより分類関数がクリスプになることが分かった。

次に、sFCMA の実験結果を図 2a, 2b に示す。パラメータ m を 2 から 1.01 に変化させたところ、 m の値を大きくすることにより分類関数がファジィになることが分かった。

qFCMA の実験結果を図 3a, 3b, 3c に示す。こちらは、パラメータ (m, λ) の組み合わせとして、 $(2, 10), (1.01, 10), (1.01, 10000)$ の 3 通りでクラスタリングを行った。分類関数より、 m の値を大きくすることによりファジィになり、 λ の値を大きくするほどクリスプに

なることが分かった。また、qFCMA において $\lambda \rightarrow +0$ とすると sFCMA と同じ特性が得られることがわかり、qFCMA は $m \rightarrow +0$ とすると eFCMA と同様の特性を示すことがわかった。これらの実験結果より qFCMA は sFCMA と eFCMA の特性を併せ持つと言える。

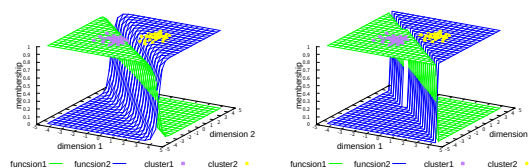


図 1: eFCMA の人工データの実験結果

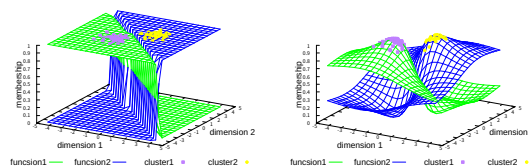
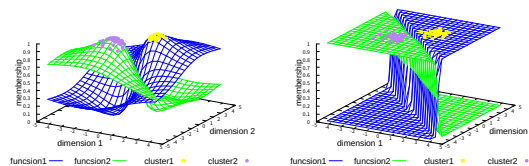


図 2: sFCMA の人工データの実験結果



(a) $m = 2, \lambda = 10$ (b) $m = 1.01, \lambda = 10$

(c) $m = 1.1, \lambda = 10000$
図 3: qFCMA の人工データの実験結果

4 実データの実験結果

eFCMA, sFCMA, qFCMA の実データ実験の結果について、それぞれ図 ??, ??, ?? に示す。eFCMA では λ の値を 0.1 から 100 まで 1 刻み、sFCMA では m の値を 1.01 から 2.00 まで 1.01 刻み、qFCMA では m の値を 1.01 から 2.00 まで 0.01 刻み、 λ の値を 0.1 から 100 まで 1 刻みで変化させた。それぞれの手法の最高 ARI

を表 1 に示す .

表 1: 各手法の ARI の最高値とパラメータ

手法名	ARI の最高値	パラメータ値
eFCMA	0.352349	$\lambda = 67$
qFCMA	0.339808	$\lambda = 99, m = 1.12$
sFCMA	0.715312	$m = 1.01$

5 まとめと今後の課題

参考文献

- [1] 河崎めぐみ, 千葉翔太: “加速度センサーを用いたユーザの行動状態推定方式,” 信学論 B, Vol. J85-B, No. 5, pp. 755–767, (2011).
- [2] 山田一郎, 中村仁: “モバイル環境におけるマルチメディア通信品質の研究,” 信学技報, Vol. 109, No. 204, pp. 27–32, (2010).
- [3] Seki, Y. and Kirii, Y.: “Fast Handover Scheme using User’s preference,” Wireless Communications and Mobile Computing, Vol. 7, Issue 5, pp. 553–568, (2009).