離散値制約によるファジィクラスタリングの研究開発

仲　大樹

クラスタリング（クラスター分析）とは，類似・非類似度に基づいて，個体を少数のクラスターに分類する手法である．クラスタリングは，いくつかの観点で分類されるが，本研究では非階層的でファジィな手法に着目する．その代表例にファジィ c−平均法(Bezdek, 1981)がある．

ファジィ c−平均法では，ファジィネスを導入するために調整パラメータを必要とするが，これとファジィネスとの対応関係は事前にわからない．本研究では，この問題を解決すべく，調整パラメータに代わって，離散値制約による新たなファジィクラスタリングを提案した．

提案手法の目的関数は，

$$f(\mathbf{A},\mathbf{C}|\mathbf{X}) = \|\mathbf{X} - \mathbf{AC}\|^2$$

で表される．ここで，$\mathbf{X}$ は $I\times J$ のデータ行列，$\mathbf{A}$ は $I\times K$ の所属度行列，$\mathbf{C}$ は $K\times J$ の特徴行列である．$\mathbf{A}$ は，要素が 0 以上（$a_{ik}\geqq 0$）で，各行和が 1（$\mathbf{A1}_K = \mathbf{1}_I$）の確率行列である．しかし，この制約条件のままだと確率行列の不定性が生じ，解が一意に定まらない．従って，$a_{ik}\geqq 0$ に代わって，次の制約条件を課す．

$$a_{ik}\in\{0,0.1,\cdots,0.9,1.0\}$$

これにより，$\mathbf{A}$ の要素 a_{ik} が$\{0,0.1,\cdots,0.9,1.0\}$の計 11 個のうち，いずれかの値をとるように制限される．この制約を「離散値制約」と呼ぶ．本研究では，離散値を 0.1 間隔の計 11 個に設定するが，これは一例である．分析者がどのような離散値を設定するかによってファジィネスの程度を決められる．このように，提案手法ではファジィネスの程度を直感的に理解することができる．また，a_{ik} のとり得る値が有限個であることから，ファジィ c−平均法と比較して，解釈のしやすい所属度行列を得られる．　パラメータの推定には，交互最小二乗法を用いる．

提案手法の有用性を検証するため，シミュレーションと実データ解析を行った．シミュレーションでは，人工データを作成し，これに提案手法を適用した．真値と分析結果とを比較した結果，提案手法は，真値の再現性を十分に有していることを確認できた．実データ解析では，提案手法とファジィ c−平均法との比較実験を行った．UCI Machine Learning Repository (Frank & Asuncion, 2010) に所蔵されている，ワインデータを用いて正分類率を比較した．その結果，離散値制約による上記の利点を持ちながら，提案手法はファジィ c−平均法と同程度のクラスタリングが可能であることが例証された．その一方で今後の課題として，高速なアルゴリズムの開発や，離散値制約における離散値の選択基準の開発などが挙げられる．（行動統計科学）