加法的クラスタリングモデル

佐 藤 義 治
北海道大学大学院工学研究科
システム情報工学専攻
ysato@main.eng.hokudai.ac.jp

ADCLUS(ADditive CLUStering) モデルは, 重複を許容するクラスターを形成するための方法として, Shepard et al.(1979) によって提案されたものである. このモデルにおいて, クラスターとは何らかの共通の性質を有する対象 (個体) の集まりであると定義される. 各個体の対 (pair) が共通の性質を有する度合いが類似度として観測され, 個体対がいくつかの性質を共有するとき, それらが類似度に寄与する度合は加法的であるものと仮定されている. 一般性を失うことなく, 個体間の類似度 s_{ij} は $0 \leq s_{ij} \leq 1$ であるものとし, ADCLUS モデルはつぎのように与えられる.

$$s_{ij} = \sum_{k=1}^{K} w_k p_{ik} p_{jk} + \varepsilon_{ij}.$$

ここに, K はクラスターの個数であり, 未知であるが与えられたものとする. w_k はクラスター k が有する性質の相対的顕著さを示す重みであり, p_{ik} は個体 i がクラスター k に属すか否かを示すもので, 1(属す) または 0(属さない) の値をとる. ε_{ij} は観測値 s_{ij} とモデルの誤差を表す.

ADCLUS モデルの基本的な問題点は, モデルの推定において, 未知パラメータである p_{ik} が離散値をとること, さらに観測値 s_{ij} の自由度とモデルの自由度との関係から, 十分な当てはまりの良さを得るためには多数のクラスターを必要とすることである.

本研究では, これらの点を考慮して, モデルの条件を緩和すると共に, 加法性は保存するが, 可能な関数形を与えることを目的として, 加法的ファジィクラスタリングモデルについて論ずる. 個体 i がクラスター k に属すか否かを緩めて, ここでは個体 i がクラスター k に属す度合 (確率) を u_{ik} 導入する. すなわち,

$$u_{ik} \geq 0, \quad \sum_{k=1}^{K} u_{ik} = 1$$

とする. さらに個体 i と j がクラスター k に属す度合をそれぞれ u_{ik}, u_{jk} とするとき, これが類似度に寄与する度合は単に積の形 $u_{ik}u_{jk}$ のみではなく, より一般にある関数 $\rho(u_{ik}, u_{jk})$ で与えられるものとし, 以下のようなモデルを考える.

$$s_{ij} = \sum_{k=1}^{K} \rho(u_{ik}, u_{jk}) + \varepsilon_{ij}$$

この関数 $\rho(x, y)$ の満たすべき条件, および s_{ij} が非対称な場合についてのモデル等について考察する.