

合理的選択理論による決定表のグループ形成解析

学籍番号：90106027 乾口研究室 榎本 隆太

1. はじめに

個人見解を表す複数の決定表を分類することにより、意見分布を概括的に把握しようとする試みがなされている[1]。決定表間の類似度を定義し、階層的クラスタリング[2]を適用している。この方法では、形成されるクラスター数が決めにくいという欠点がある。本研究では、分類対象が決定表という特殊性を利用して、クラスター統合による各個人の損益を考察し、クラスター数が自動的に決定されるクラスタリング法を提案する。実データを用いて提案法の特徴や有用性について考察する。

2. 合理的選択理論による決定表のグループ形成

決定表のクラスターが形成されたとき、クラスター内の過半数の決定表に支持される意見は、クラスター意見として採用される。ある決定表の意見がクラスター意見として採用されると、その意見が全体に反映されやすくなるので、その決定表にとって有利になる。したがって、同様な意見をもつ決定表はクラスターを形成しようとする。一方、クラスターが大きくなり、意見がある程度ばらつくようになると、各決定表の意見がクラスター意見として採用されにくくなり、大きなクラスターは必ずしも各決定表にとって有利とは限らない。本研究では、このような観点から、適当なクラスター数で停止する決定表の階層的クラスタリング法を提案する。

いま、二つの決定表のクラスター K_a, K_b を考える。この二つのクラスターが結合して、新たなクラスター $K = K_a \cup K_b = \{T_1, \dots, T_p\}$ を形成する際の利得について考察する。決定表に現れる対象評価を (w, v_d) と表す。ここで、 w は条件属性パターンで、 v_d は決定属性値を示している。評価 (w, v_d) をもつ決定表 T_i 内の対象は唯一とは限らず、複数存在しうる。その出現回数を $\sigma^{T_i}(w, v_d)$ と記す。また、決定表 T_i 内の対象評価は整合しているとは限らず、 T_i 内で矛盾しない対象評価を意見と呼び、意見の集合を \underline{T}_i と表す。クラスター K 内で意見 (w, v_d) をもつ決定表数は、

$$x_K(w, v_d) = |\{T_i \in K \mid (w, v_d) \in \underline{T}_i\}| \quad (1)$$

と定められる。また、クラスター K 内で過半数を得ている意見の集合を $Agg(K)$ と記すと、 $Agg(K) = \{(w, v_d) \mid 2x_K(w, v_d) > |K|\}$ となる。ただし、 $|K|$ は K の基数である。以上で定義された記号を用いると、決定表 $T_i \in K$ にとってのクラスター K の価値は次式で定められる。

$$Val_K(T_i) = \sum_{(w, v_d) \in Agg(K)} \sigma^{T_i}(w, v_d) S^{x_K(w, v_d) - 1} \quad (2)$$

ただし、 S は予め定められた定数である。

決定表 $T_i \in K_a$ にとってのクラスター K_a の価値 $Val_{K_a}(T_i)$ も、決定表 $T_j \in K_b$ にとってのクラスター K_b の価値 $Val_{K_b}(T_j)$ も同様に定義できるので、より大きなクラスター K の形成による $T_i \in K_a, T_j \in K_b$ の利得は、それぞれ、次のように定められる。

$$Gain_K(T_i|K_a) = Val_K(T_i) - Val_{K_a}(T_i) \quad (3)$$

$$Gain_K(T_j|K_b) = Val_K(T_j) - Val_{K_b}(T_j) \quad (4)$$

$Gain_K(T_y|K_z)$ が正であれば、 $T_y \in K_z$ にとって、クラスター K の形成は有利であり、負であれば不利となる。全会一致

の原則に従えば、クラスター K_a, K_b 内のすべての決定表が有利にならないとクラスター K を形成しないことになる。

通常の階層的クラスタリング法のクラスター形成に、上述の考え方を適用した手法を提案法 1 と呼ぶ。

提案法 1 では、クラスター内の決定表の意見とクラスター意見との対立度については考察していない。そこで、クラスター意見との対立度を考慮し、価値が対立度による損失より大きい限り、クラスターの結合を受諾するモデルを考える。

クラスター意見との対立度は次式で定義できる。

$$Conf_K(T_i) = \sum_{w:(w, v_d^i) \in Agg(K), w \in C(\{v_d^i\}, v_d^i \neq v_d^j)} F \sigma^{T_i}(w, v_d^i) \quad (5)$$

ただし、 F は定数である。これにより、大きなクラスター K の形成による $T_i \in K$ の総合評価値はそれぞれ、次のようになる。

$$Total_K(T_i) = Val_K(T_i) - Conf_K(T_i) \quad (6)$$

$Total_K(T_i)$ が非負である限り、 T_i はクラスター K の形成を受諾することになる。

階層的クラスタリング法において、 $\min_{T_i \in K} Total_K(T_i)$ の大きい順に、すべての可能な結合に対して $\min_{T_i \in K} Total_K(T_i)$ が負となるまで、クラスターの結合を行う方法を提案法 2 と呼ぶ。

3. 実データへの適用

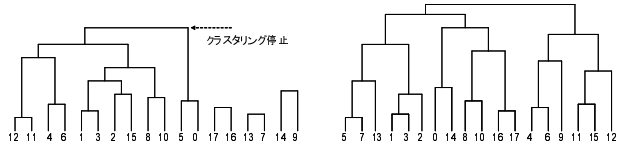


図 1: 提案法 2 の結果

図 2: 従来法の結果

本研究では、学生の就職志望に関するアンケートデータと図柄の好みに関するアンケートデータに対して、提案法と従来法[1]を適用し、提案法の特徴を検討した。まず、定数 S, F を調整することにより、図 1 に示すように、提案法 2 では、適当なクラスター数で停止することが確認できた。

形成されたグループの特徴を調べるため、従来法で得られた樹形図を同じクラスター数で切断して得られるグループと比較した。クラスター意見と決定表の意見との一致度(評価 1)、クラスター意見と決定表の意見との対立度(評価 2)、クラスターサイズのバラツキ度合(評価 3)の 3 点について考察した。提案法 1 と従来法との比較では、いずれの評価においても大きな差は認められなかった。提案法 2 と従来法の比較では、図 1、図 2 に示すように、提案法 2 がクラスターサイズのバラツキ度合が大きく、少数の大きなクラスターができる傾向にある。他の評価においては大きな差は認められなかった。

参考文献

1. M. Inuiguchi, T. Furudono: Clustering Analysis of Individual Opinions Given by Decision Tables, *Proceedings of 10th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty*, pp.41-53 (2007)
2. 宮本 定明: “クラスター分析入門, ファジークラスタリングの理論と応用,” 森北出版 (1999)