

割合を考慮したファジィ分割表と類似係数 Fuzzy Contingency Table including Total Ratio and Similarity Indices

上江洲 弘明 (金沢工業大・数理工教育研究センター)

Hiroaki UESU, Kanazawa Institute of Technology

E-mail: uesu@neptune.kanazawa-it.ac.jp

The contingency table analysis provides the interrelation between two variables and can help find interactions between them. Previously, we extended the contingency table by applying fuzzy theory and defined a fuzzy 2×2 contingency table. However, this definition cannot be extended to fuzzy $m \times n$ contingency tables. In this paper, we propose the “new” definition of fuzzy $m \times n$ contingency table and apply to similarity index calculation.

1. はじめに

分割表分析は、2 つの変数間の相互作用を見つけるのに有用である。これまで筆者らは、ファジィ集合論を適用して分割表を拡張し、 t -ノルムを用いたファジィ 2×2 分割表を定義した。しかし、この定義では一般のファジィ $m \times n$ 分割表に拡張できない。本論文では、新たなファジィ $m \times n$ 分割表の定義を提案し、この分割表を用いた際のファジィ類似係数を紹介する。

2. ファジィ $m \times n$ 分割表

ファジィ集合の集合演算においては、一般的に相補律(矛盾律と排中律)は成立しない。そこで、ファジィ分割表では、各ファジィ集合の濃度を考え、行周辺合計 (row marginal total) および列周辺合計 (column marginal total) の総計が 1 となるよう定義する。

定義 1 ファジィ集合の濃度

全体集合 $U = \{x_i \mid i = 1, \dots, N\}$ 上のファジィ集合 A の濃度 $|A|$ は以下のように定められる。

$$|A| = \sum_{k=1}^N \mu_A(x_k)$$

ここで、 $\mu_A(x)$ はファジィ集合 A のメンバーシップ関数である。

ファジィ $m \times n$ 分割表を再定義する。筆者らが提案してきた従来のファジィ $m \times n$ 分割表では積演算として代数積を用いていたが、この新たな定義では t -ノルムを用いている。これにより、より柔軟な適用が可能となる。

定義 2 ファジィ $m \times n$ 分割表

全体集合 $U = \{x_i \mid i = 1, \dots, N\}$ 上のファジィ集合 $A_1, \dots, A_n, B_1, \dots, B_m$ について、ファジィ $m \times n$ 分割表は以下のように定められる (Table 1)。

Table 1 Fuzzy $m \times n$ contingency table.

	A_1	...	A_n
B_1	f_{11}	...	f_{1n}
$\vdots$	$\vdots$		$\vdots$
B_m	f_{m1}	...	f_{mn}

ここで、

$$\sum_{i=1}^n \mu_{A_i}(x_k) = 1, \quad \sum_{i=1}^m \mu_{B_i}(x_k) = 1$$

であり、

$$f_{ij} = \sum_{k=1}^N \frac{T(\mu_{A_j}(x_k), \mu_{B_i}(x_k))}{\sum_{s=1}^n \sum_{t=1}^m T(\mu_{A_s}(x_k), \mu_{B_t}(x_k))}$$

である。なお、演算 $T(p, q)$ は t -ノルムである。

定理 1 行列周辺合計

全体集合 $U = \{x_i \mid i = 1, \dots, N\}$ 上のファジィ集合 $A_1, \dots, A_n, B_1, \dots, B_m$ についてのファジィ $m \times n$ 分割表において、以下の等式が成立する。

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = N$$

3. 類似係数とファジィ 2×2 分割表

ここでは類似係数を考えるため、前述のファジィ $m \times n$ 分割表の特別な場合であるファジィ 2×2 分割表を用いる。

定義 3 ファジィ 2×2 分割表

以下のような 2 つのベクトル $\mathbf{x}, \mathbf{y}$ を考える。

$$\mathbf{x} = (x_i), \mathbf{y} = (y_i), 0 \leq x_i, y_i \leq 1, 1 \leq i \leq N$$

このとき、ファジィ 2×2 分割表は以下のように定められる (Table 2)。

Table 2 Fuzzy 2×2 contingency table.

$\mathbf{x} \backslash \mathbf{y}$	1	0	Sum
1	a_{xy}	b_{xy}	$a_{xy} + b_{xy}$
0	c_{xy}	d_{xy}	$c_{xy} + d_{xy}$
Sum	$a_{xy} + c_{xy}$	$b_{xy} + d_{xy}$	N

$$a_{xy} = \sum_{i=1}^N \frac{T(x_i, y_i)}{\kappa_i}, \quad b_{xy} = \sum_{i=1}^N \frac{T(x_i, 1 - y_i)}{\kappa_i},$$

$$c_{xy} = \sum_{i=1}^N \frac{T(1 - x_i, y_i)}{\kappa_i}, \quad d_{xy} = \sum_{i=1}^N \frac{T(1 - x_i, 1 - y_i)}{\kappa_i}$$

ここで、 κ_i は

$$\kappa_i = \sum_{\substack{p_i \in \{x_i, 1-x_i\} \\ q_i \in \{y_i, 1-y_i\}}} T(p_i, q_i)$$

である。

このファジィ2×2分割表から類似係数を定義する. 代表的な類似係数は以下の通りである.

(1) Jaccard :

$$s_{xy} = \frac{a_{xy}}{a_{xy} + b_{xy} + c_{xy}}$$

(2) Sorensen-Dice :

$$s_{xy} = \frac{2a_{xy}}{2a_{xy} + b_{xy} + c_{xy}}$$

(3) Russell and Rao :

$$s_{xy} = \frac{a_{xy}}{a_{xy} + b_{xy} + c_{xy} + d_{xy}}$$

(4) Simple Matching :

$$s_{xy} = \frac{a_{xy} + d_{xy}}{a_{xy} + b_{xy} + c_{xy} + d_{xy}}$$

(5) Rogers-Tanimoto :

$$s_{xy} = \frac{a_{xy} + d_{xy}}{a_{xy} + 2b_{xy} + 2c_{xy} + d_{xy}}$$

(6) Simpson :

$$s_{xy} = \frac{a_{xy}}{(a_{xy} + b_{xy}) \wedge (a_{xy} + c_{xy})}$$

(7) Ochiai :

$$s_{xy} = \frac{a_{xy}}{\sqrt{(a_{xy} + b_{xy})(a_{xy} + c_{xy})}}$$

(8) Phi :

$$s_{xy} = \frac{a_{xy}d_{xy} - b_{xy}c_{xy}}{\sqrt{(a_{xy} + b_{xy})(a_{xy} + c_{xy})(b_{xy} + d_{xy})(c_{xy} + d_{xy})}}$$

(9) Yule :

$$s_{xy} = \frac{a_{xy}d_{xy} - b_{xy}c_{xy}}{a_{xy}d_{xy} + b_{xy}c_{xy}}$$

4. 数値例

$x = \begin{pmatrix} 1 \\ 0.4 \\ 0.3 \\ 0.2 \\ 0.1 \end{pmatrix}, y = \begin{pmatrix} 0.7 \\ 0.6 \\ 0.9 \\ 0.5 \\ 0.6 \end{pmatrix}$ に対し, 以下のファジィ2×2分

割表が得られる. t-ノルムは従来通り代数積と新たに min で算出した.

Table 3 Fuzzy 2×2 contingency table(Algebraic Product).

$x \backslash y$	1	0	Sum
1	1.37	0.63	2
0	1.93	1.07	3
Sum	3.3	1.7	5

Table 4 Fuzzy 2×2 contingency table(Logical Product).

$x \backslash y$	1	0	Sum
1	1.3984	0.8317	2.2302
0	1.7738	0.996	2.7698
Sum	3.1722	1.8278	5

これらの分割表から類似係数を計算すると, 以下の値が得られる.

Table 5 Values of some similarity indices.

t-ノルム	Simple Matching	Jaccard	Sorensen-Dice
代数積	0.488	0.3486	0.5170
min	0.4789	0.3493	0.5177

今回の数値例では, 類似係数の値は t-ノルムによって大きな違いは見られなかった.

7. まとめ

今回, t-ノルムを用いて再定義したファジィ分割表を提案した. 数値シミュレーションでは大きな違いがなかったが, もっと多くの例を試して検証していきたい. また, Type-2 ファジィにまで拡張したいと考えている.

参考文献

1) Zadeh L A, Inform and Control, pp.338-353, 1965.

2) Zadeh L.A: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and System, 1, pp.3-28, 1978.

3) De Luca A, Termini S, A definition of nonprobabilistic entropy in the settings of fuzzy set theory, Information and Control, pp.301-312, 1972.

4) H.Uesu : Type-2 Fuzzy Contingency Table Analysis and its Application the Seventh International Conference on Dynamic Systems and Applications, 2015.

5) H.Uesu : Contingency Table Analysis Applying Fuzzy Number and Its Application, the 8th International Joint Conference on Computational Intelligence, pp.93-99, 2016.

6) 上江洲:type-2 ファジィ 2×2 分割表と類似係数について, 第 33 回ファジィシステムシンポジウム, pp.723-726, 2017.

7) H.Uesu : Type-2 Fuzzy Contingency Table and Similarity Indices, ICIC Express Letters Volume 12, Number 8, pp.791-798, 2018.

8) H.Uesu : Bubble Chart Analysis for a Lecture Applying Type-2 Fuzzy Contingency Table, ICIC Express Letters Volume 13, Number 6, pp.483-491, 2019.

9) 上江洲:n 次元ファジィ数ベクトルの順序, 第 65 回理論応用力学講演会・第 22 回土木学会応用力学シンポジウム講演論文集, pp.63-64, 2019.

10) 上江洲:2 次元ファジィ数ベクトルの順序を取り入れた Type-2 ファジィ分割表分析, 知能と情報(日本知能情報ファジィ学会誌)第 32 巻 第 1 号(2020 年 2 月), pp.528-533, 2020.

11) 上江洲:ファジィ分割表の再考と類似係数, 第 30 回ソフトウェア・ワークショップ, pp.50-51, 2020.