

一种信息系统模糊划分方法^{*}

An Information System Fuzzy Parting Method

王小明 冯德民

(陕西师范大学计算机科学系 西安710062)

Abstract The study on the information system parting method is a hotspot in information system development, now. This paper proposes an information system fuzzy parting method (ISFP) based on fuzzy equivalent relation theory. The subsystems parted by ISFP have better system flexibility. The method is simple and practical, and has high application value.

Keywords Information system parting, Fuzzy similar, Fuzzy equivalence, Integration

1 引言

对系统划分一直是计算机信息系统结构分析与设计的重点和难点。本文以模糊等价关系理论为基础^[5],提出了一种信息系统子系统模糊划分方法(Information System Fuzzy Parting,以下简称ISFP)。与80年代初美国IBM公司在精确等价关系理论上提出的信息系统子系统划分方法BSP^[1](Business System Planning, BSP)相比,使用ISFP方法划分的子系统具有更高的系统柔性;而BSP方法是将具有强连通关系(精确等价关系)的过程聚类为一个子系统;划分边界是完全刚性的,子系统局部最优,但不利于系统整体优化和系统伸缩变换(柔性变换)。ISFP方法是将具有一定强度的弱连通关系的过程,按某一阈值聚类,使其具有强连通关系(模糊等价关系),从而将其划归为同一子系统,划分边界是柔性的。ISFP方法克服了BSP方法的不足,更加符合现代信息系统设计对系统柔性要求高的客观需要。

2 ISFP方法的基本思想

ISFP方法的基本思想是把信息系统看作作为一个三元组 $S(R, P, D)$,其中 P 是信息系统所有过程(功能)的集合, D 是所有数据类的集合, R 是 P 对 D 的隶属度(关联度)集合。通过构造 $P \times D$ 空间上的隶属度模糊关系矩阵 R ,由 R 构造 $P \times P$ 空间上的模糊相似矩阵 C ,由 C 演绎出 $P \times P$ 空间上的模糊等价矩阵 Q ,选取阈值 $\lambda(0 \leq \lambda \leq 1)$ 对 Q 进行聚类,所得等价类集即为子系统划分集。为了描述算法,先给出如下定义和定理。

2.1 若干定义、定理

定义1 (模糊关系) 对 $\forall p_i \in P, \forall d_i \in D$,这里,

$1 \leq i \leq m, 1 \leq j \leq n$,存在数 $r_{ij}(0 \leq r_{ij} \leq 1)$ 来表征它们之间的隶属程度,则称 $P \times D$ 空间上确定了一个模糊关系 R , R 可用一个 $m \times n$ 矩阵描述: $R = (r_{ij})_{m \times n}$ 。

定义2 (模糊关系合成) 设 G 是空间 $U \times Y$ 上的模糊关系, S 是空间 $Y \times W$ 上的模糊关系,则在空间 $U \times W$ 上的模糊关系 $G \cdot S$ 定义为:

$$G \cdot S(u, w) = \bigvee G(u, y) \wedge S(y, w)$$

其中“ $\cdot$ ”指按常规矩阵乘法法则运算,但将“ $+$ ”改为“ $\bigvee$ (取大)”,将“ $\times$ ”改为“ $\wedge$ (取小)”运算。

定义3 (模糊相似关系) 若在信息系统过程集合 P 上的模糊关系 F ,满足:(1)自反性,即对 $\forall p_i \in P, F(p_i, p_i) = 1$;(2)对称性,即对 $\forall p_i, p_j \in P$,如果 $p_i F p_j$,则 $p_j F p_i$,这里 F 表示 p_i 与 p_j 之间的耦合度,称 F 为 $P \times P$ 上的模糊相似关系。

定义4 (模糊等价关系) 若 F 为信息系统过程空间 $P \times P$ 上的模糊相似关系,并且 F 满足:(3)传递性,即对 $\forall p_i, p_j, p_k \in P$,如果 $p_i F p_j$,并且 $p_j F p_k$,则 $p_i F p_k$,称 F 为 P 上的模糊等价关系。

模糊数学中已经建立起完善的由模糊相似到模糊等价的理论和方法,其核心是传递闭包理论^[5],限于篇幅,这里不加证明地给出由模糊相似矩阵计算模糊等价矩阵的定理^[5]。

定理1 若 $F = (f_{ij})_{m \times m}$ 为 $P \times P$ 上的模糊相似关系,则其传递闭包(满足传递性的最小包含,即模糊等价矩阵)为:

$$Q = F \cdot F \cdots F = F^{m-1} \quad (2-1)$$

证明:(略)

2.2 ISFP算法描述

步1 列出信息系统 S 的所有过程(功能),得过程集合 $P = \{p_1, p_2, \dots, p_m\}$;

步2 列出信息系统 S 的所有数据类,得数据类集

^{*} 本课题受正林集团技术创新基金资助,王小明 副教授,从事企业信息系统、软件工程、数据库的研究和教学工作,冯德民教授,主要从事企业信息系统、软件算法、软件工程的研究与教学工作。

合 $D = \{d_1, d_2, \dots, d_n\}$;

步3 以数据类集为行,以过程集为列,构造过程与数据类的隶属度矩阵 $R = (r_{ij})_{m \times n}$, r_{ij} 表示 p_i 对 d_j 的隶属程度(即 p_i 产生或使用 d_j 的程度), $0 \leq r_{ij} \leq 1$, 由系统分析确定;

步4 根据定义3,由隶属度集合 R 构造过程空间 $P \times P$ 的耦合对称矩阵(过程间的模糊相似矩阵) $C = (c_{ij})_{m \times m}$

$$C_{ij} = \begin{cases} 1 & \text{当 } i=j \\ 1 - \theta \times \sum_{k=1}^n |r_{ik} - r_{jk}| & \text{当 } i \neq j \end{cases} \quad (2-2)$$

其中, $1 \leq i, j \leq m$, θ 为待定系数,按 $0 \leq c_{ij} < 1$ 来确定,并且 $\theta > 0$ 。

步5 根据定理1和定义2求模糊相似矩阵 C 的模糊等价矩阵 Q , 即 $Q = C^{n-1} = (q_{ij})_{m \times m}$;

步6 选取阈值 λ ($0 \leq \lambda \leq 1$)。一般情况下,按下列公式选取 λ 值,聚类所得等价关系比较理想。

$$\lambda = \frac{\sum_{i=1}^m \sum_{j=1}^m q_{ij}}{m \times m} \quad (2-3)$$

对 Q 聚类,求得模糊等价矩阵 $Q^\lambda = (q_{ij}^\lambda)_{m \times m}$ 。

$$q_{ij}^\lambda = \begin{cases} 1 & \text{当 } q_{ij} \geq \lambda \\ 0 & \text{当 } q_{ij} < \lambda \end{cases} \quad (2-4)$$

根据定义4, Q^λ 中满足等价关系的过程划归为同一子系统。

由于 λ 取值的不同,可以得到多个聚类划分方案,再结合深入的系统分析,选取最理想的划分方案。

3 应用举例

例 在某外资企业信息系统(TTMIS)结构分析设计中,系统分析人员与企业业务人员通过系统调研(为了演算过程简洁,以下只给出部分过程与数据类的抽象描述)得:

步1 过程集合 $P = \{p_1, p_2, p_3, p_4, p_5\}$;

步2 数据类集合 $D = \{d_1, d_2, d_3, d_4\}$;

步3 分析 p_i 与 d_j 之间的隶属关系,得隶属度矩阵 $R_{5 \times 4}$:

	d_1	d_2	d_3	d_4
p_1	0.4	0	0.7	0.3
p_2	0	0.3	0	0.9
p_3	0	0	0.8	0
p_4	0.5	0.9	0	0
p_5	0.3	0.1	0	0.6

步4 根据公式(2-2)构造过程耦合对称矩阵(模糊相似矩阵) $C_{5 \times 5}$, 这里取 $\theta = 0.3$

$$C_{5 \times 5} = \begin{bmatrix} 1 & 0.4 & 0.76 & 0.4 & 0.64 \\ & 1 & 0.4 & 0.4 & 0.76 \\ & & 1 & 0.34 & 0.46 \\ & & & 1 & 0.52 \\ & & & & 1 \end{bmatrix}$$

步5 根据定义2和公式(2-1)求矩阵 $C_{5 \times 5}$ 的模糊

等价矩阵 $Q_{5 \times 5} = C^{5-1} = C^4$

$$C^2 = C \cdot C = \begin{bmatrix} 1 & 0.4 & 0.76 & 0.4 & 0.64 \\ 0.4 & 1 & 0.4 & 0.4 & 0.76 \\ 0.76 & 0.4 & 1 & 0.34 & 0.46 \\ 0.4 & 0.4 & 0.34 & 1 & 0.52 \\ 0.64 & 0.76 & 0.46 & 0.52 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.4 & 0.76 & 0.4 & 0.64 \\ 0.4 & 1 & 0.4 & 0.4 & 0.76 \\ 0.76 & 0.4 & 1 & 0.34 & 0.46 \\ 0.4 & 0.4 & 0.34 & 1 & 0.52 \\ 0.64 & 0.76 & 0.46 & 0.52 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.64 & 0.76 & 0.52 & 0.64 \\ 0.64 & 1 & 0.46 & 0.52 & 0.76 \\ 0.76 & 0.46 & 1 & 0.46 & 0.64 \\ 0.52 & 0.52 & 0.46 & 1 & 0.52 \\ 0.64 & 0.76 & 0.64 & 0.52 & 1 \end{bmatrix}$$

$Q_{5 \times 5} = C^4 = C^2 \cdot C^2 =$

$$\begin{bmatrix} 1 & 0.64 & 0.76 & 0.52 & 0.64 \\ 0.64 & 1 & 0.64 & 0.52 & 0.76 \\ 0.76 & 0.64 & 1 & 0.52 & 0.64 \\ 0.52 & 0.52 & 0.52 & 1 & 0.52 \\ 0.64 & 0.76 & 0.64 & 0.52 & 1 \end{bmatrix}$$

步6 根据公式(2-3)求得 $\lambda = 0.69$, 再根据公式(2-4)对 Q 聚类得:

$$Q^{0.69} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

因此信息系统的子系统划分为 $\{p_1, p_3\}, \{p_2, p_5\}, \{p_4\}$ 。

在实际应用中,可以结合具体问题对聚类阈值 λ 取不同的值,得出几个预选划分方案,通过分析,选取最理想的划分方案。

结束语 ISFP 方法简单、实用,而且易于在计算机上实现,比人工划分子系统更具有精确性和柔性。当信息系统越是庞大复杂时,越能发挥其优越性。在使用 ISFP 方法时,由于过程集合与数据类集合的隶属度矩阵构造不同,或聚类阈值 λ 取值不同,可能得到不同的划分集,系统分析时可以在多个划分方案之间进行分析择优。该方法已成功地应用于多个大中型企业信息分析设计中,取得了良好的效果。

参考文献

- 1 IBM. Business System Planning-Information System Planning Guide, Publication No. GE20-0527, 1984
- 2 Shaw M, Garlan D. Software architecture: perspectives on an emerging discipline. Prentice Hall, Inc. 1996
- 3 Wiederhold G. Mediators in the architecture of future information system. IEEE Computer, 1992, 25(3)
- 4 Mensan K E. Identifying Subsystems Information System Analysis. Information System, 1984, 9(2)
- 5 陈贻源. 模糊数学. 武汉: 华中工学院出版社, 1986
- 6 卢开澄, 等著. 图论及其应用(第二版). 北京: 清华大学出版社, 1995
- 7 方世昌. 离散数学. 西安: 西北电讯工程学院出版社, 1985
- 8 沈继红, 等编. 数学建模. 哈尔滨: 哈尔滨工程大学出版社, 1996