

基于模糊聚类和粗糙集的仿真可信性模糊综合评估

吴 静, 吴晓燕, 高忠长

(空军工程大学导弹学院, 陕西 三原 713800)

摘要: 为客观有效地解决仿真可信性评估问题, 提出基于模糊聚类和粗糙集的仿真可信性模糊综合评估方法。阐述了基于模糊传递闭包法进行模糊聚类分析的基本步骤, 分析了粗糙集中属性重要性的相关原理; 提出了基于模糊聚类和粗糙集的可信性模糊综合评估模型, 利用模糊聚类和粗糙集中的属性重要性原理客观地进行各因素权重分配, 结合模糊综合评判进行仿真可信性综合评估; 以某飞行视景仿真系统为例, 进行可信性综合评估。结果表明, 该方法具有一定的合理性和可行性。

关键词: 模糊聚类; 粗糙集; 权重分配; 可信性评估

中图分类号: TP 391.9

文献标志码: A

Fuzzy comprehensive evaluation of simulation credibility based on fuzzy clustering analysis and rough sets theory

WU Jing, WU Xiao-yan, GAO Zhong-chang

(Missile Inst. of Air Force Engineering Univ., Sanyuan 713800, China)

Abstract: To execute credibility evaluation of modeling and simulation applications objectively and cost-effectively, a method of fuzzy comprehensive evaluation of simulation credibility based on fuzzy clustering analysis and rough sets theory is proposed. The fuzzy clustering analysis (FCA) based on fuzzy transitive closure method is analyzed and the attribute significance in rough sets theory is introduced briefly. Furthermore, a fuzzy comprehensive evaluation model of simulation credibility based on FCA and rough sets theory is established. With the confirmation of multi-attribute weight allocation by FCA and rough sets theory, simulation credibility evaluation is performed based on fuzzy comprehensive evaluation. Finally, the method is applied to credibility evaluation of a flight visual simulation system. The evaluation result shows the reasonability and validity of the proposed approach.

Keywords: fuzzy clustering; rough set; weight allocation; credibility evaluation

0 引言

随着先进仿真技术的迅猛发展、仿真应用规模的不断扩大以及建模与仿真复杂程度的不断提高, 仿真界对仿真可信性评估提出了迫切的需求。如果对仿真系统不作可信性评估, 用户就不能确定对仿真结果应拥有多大的信心, 不能确定使用仿真结果会产生多大的风险。然而, 仿真可信性评估是一个多人、多层次、多目标综合评估问题的求解过程, 而且是在人员维、层次维、目标维、对象维、时间维的高维空间中进行的, 涉及大量不确定性、模糊性和复杂性因素, 给可信性评估带来了极大难度。缺少一个有效的可信性评估方法已成为制约仿真系统可信性评估发展的重要因素^[1]。

目前, 针对不同的仿真系统, 许多专家学者对仿真可信性的量化研究进行了有益地探索, 提出了新的仿真可信性评估方法。文献[2]提出基于相似理论的系统仿真可信性分析的基本思想和方法, 文献[3]提出了改进的基于相似度理论的仿真可信度评估方法, 由于仿真系统往往是复杂系统, 不容易找出其全部要素来构建相似元, 因此相似元的获取、表示和度量成为该方法的瓶颈。文献[4]分析了层次分析法在某仿真系统可信性评估中的应用, 文献[5]运用模糊层次分析法计算影响系统仿真可信度评估的各因素的权重, 文献[6]讨论了仿真可信度模糊评判法, 文献[7]给出了建立仿真系统可信度评估的评估体系原则和模糊定量评估方法。但是, 层次分析法和模糊综合评判法有一个共同的问题: 无论是构造判断矩阵, 还是指标权重的确定, 都不是

收稿日期: 2009-05-21; 修回日期: 2009-09-05。

基金项目: 陕西省自然科学基金(2007F40)资助课题

作者简介: 吴静(1983-), 男, 博士研究生, 主要研究方向为建模仿真及可信性评估。E-mail: ken1983414@126.com

伴随评价过程产生的,而是人为确定的,人的主观判断、选择和偏好对结果影响较大,使得主观成分占很大比重。

模糊聚类法在处理信息时没有任何系统信息的损失,但该理论无法提供基于客观信息的系统各因素的权重分配。应用粗糙集理论来确定因素权重的方法无需提供问题所需处理的数据集合之外的任何先验信息,但它一般只适用于决策表是离散的情形,在处理含有连续条件因素的决策表之前,通常须将其离散化,而这可能会导致信息大量丢失。因此,有必要将二者结合起来进行因素权重客观分配研究,从而有效避免了主观因素对评判结果的影响。本文以模糊集与粗糙集理论为基础,提出基于模糊聚类和粗糙集的仿真可信性模糊综合评估方法,克服了传统评估方法过分依赖专家经验知识的不足,具有一定的实用性和可行性。

1 相关理论

1.1 模糊聚类分析

模糊聚类分析是根据客观事物间的不同特性、亲疏程度和相似性等关系,通过建立模糊相似关系对客观事物进行分类的数学方法^[8-9]。模糊聚类分析的方法主要有模糊传递闭包法、直接聚类法、最大树法、编网法等,下面重点分析模糊传递闭包法。

确定需要处理的样本对象,抽取因素数据。设 $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ 是待分类对象的全体,每个样本用 m 个特征指标向量来表示,即 $x_j = (x_{j1}, x_{j2}, \dots, x_{jm})$,则可用特征值矩阵 $\mathbf{X} = (x_{ij})_{n \times m}$ 对需要处理的样本对象进行聚类。

(1) 数据标准化

为消除特征指标单位的差别和特征指标数量级不同的影响,必须对各指标值进行标准化处理,从而使每一个指标值统一于某种共同的数值特性范围。本文采用平移·标准差变换和 平移·极差变换的数据标准化方法^[9]。

(2) 建立模糊相似矩阵

利用多元分析的方法确定经过数据标准化后的对象 x_i 和 x_j 之间的相似程度: $r_{ij} = R(x_i, x_j) \in [0, 1] (i, j = 1, 2, \dots, n)$, 建立一个对象与对象之间的模糊相似矩阵 $\mathbf{R} = (r_{ij})_{n \times n}$ 。本文采用欧式距离法^[9] 建立模糊相似矩阵(令 $C=1/8$)。

(3) 模糊聚类及确定合理阈值范围

利用模糊等价闭包法求出模糊等价矩阵 $\mathbf{t}(\mathbf{R})$, 对 $\mathbf{t}(\mathbf{R})$ 取适当的 $\lambda \in [0, 1]$, 按 λ 截矩阵 \mathbf{R}_λ 进行动态聚类。

在模糊聚类分析中,对不同的 λ 可得到不同的 λ 分类。但在许多实际问题中,经常遇到如何选择合理的分类阈值 λ , 以保证分类的合理性问题。针对以上问题,本文采用 F -统计量对原始数据进行统计分析,进而确定合理阈值范围。

设 r 为对应于 λ 值的类数, n_i 为第 i 类元素的个数,记

$\bar{x}_{ik} = \frac{1}{n} \sum_{j=1}^{n_i} x_{jk} (k=1, 2, \dots, m)$ 为第 i 类元素的第 k 个特征

的平均值;记 $\bar{x}_k = \frac{1}{n} \sum_{j=1}^n x_{jk} (k=1, 2, \dots, m)$ 为全体样品第 k 个特征的平均值。

引入 F -统计量

$$F = \frac{\sum_{i=1}^r n_i \sum_{k=1}^m (\bar{x}_{ik} - \bar{x}_k)^2 / (r-1)}{\sum_{i=1}^r \sum_{j=1}^{n_i} \sum_{k=1}^m (x_{ik} - \bar{x}_{ik})^2 / (n-r)} \quad (1)$$

它服从自由度为 $r-1, n-r$ 的 F 分布,其分子表征类与类之间的距离,分母表征类内元素间的距离。对给定的置信度 α , 查 F 临界值得 F_α , 如果 $F > F_\alpha$, 根据数理统计方差理论,可知类间差异显著,说明分类比较合理^[8]。

1.2 粗糙集相关理论

粗糙集理论是一种新的处理模糊和不确定性知识的工具,其主要思想就是在保持分辨能力不变的前提下,通过知识约简,导出问题的决策或分类规则^[10]。

定义 1 信息系统 $S = \langle U, A, V, f \rangle$, 其中, U 是对象的非空有限集合,称为论域; A 是因素的非空有限集合, $A = C \cup D$, C 和 D 分别称为条件因素集和决策因素集; $V = \bigcup_{a \in A} V_a$, V_a 是因素 a 的值域; $f: U \times A \rightarrow V$ 是一个信息函数,它为每个对象的每个因素赋予一个信息值,即 $\forall a \in A, x \in U, f(x, a) \in V_a$ 。

定义 2 给定知识库 $K = (U, R)$, 对于每个子集 $X \subseteq U$ 和一个等价关系 $R \in \text{ind}(K)$, 定义两个子集,即

$$\underline{R}X = \bigcup \{Y \in U/R \mid Y \subseteq X\} \quad (2)$$

$$\overline{R}X = \bigcup \{Y \in U/R \mid Y \cap X \neq \emptyset\} \quad (3)$$

分别称它们为 X 的 R 下近似集和 R 上近似集。

定义 3 给定知识库 $K = (U, R)$, 且 $B, D \subseteq R$ 。设

$$k = \gamma_B(D) = |\text{POS}_B(D)| / |U| \quad (4)$$

式中, $\text{POS}_B(D) = \underline{B}D$ 。则称知识 D 是 $k (0 \leq k \leq 1)$ 度依赖于知识 B 的。

定义 4 给定一个信息系统 $S = \langle U, C, V, f \rangle$, 且 $B \subseteq C$, $\forall a \in B$, 定义

$$\text{sig}(a, B, C) = \gamma_B(C) - \gamma_{B-\{a\}}(C) \quad (5)$$

即为因素 a 对因素集 B 的重要性。

2 基于模糊聚类和粗糙集的可信性模糊综合评估模型

由模糊聚类理论可知,该理论在处理信息时没有任何系统信息的损失,但该理论无法提供基于客观信息的系统各因素的权重分配,而进行模糊评判时通常需要提供先验的权重分配。与此相对应,粗糙集理论在进行数据处理时无需提供数据之外的先验信息,并能够提供各因素的重要性大小,但该理论在进行模糊数据的处理时采用区间划分方法要忽略较多的系统信息^[11-12]。因此,将这二者结合起来进行模糊综合评估研究是十分必要的。

(1) 根据仿真系统的多次实际运行情况,建立评估样本 $\mathbf{X} = (x_{ij})_{n \times m}$ 。

(2) 基于模糊聚类和粗糙集理论确定因素权重,具体步骤如下。

步骤 1 对 X 按模糊传递闭包法的一般步骤进行聚类,根据 F-统计量方法确定在满足条件 $F > F_\alpha$ 时合理的阈值范围 $\lambda_C \in [a_C, b_C]$ 及相应的分类。

步骤 2 分别删除因素 $c_i (i = 1, 2, \dots, m)$, 得到删除某一因素 c_i 后的特征值矩阵。对该矩阵按模糊聚类分析的方法进行分类,利用 F-统计量确定满足条件 $F > F_\alpha$ 时,合理的阈值范围 $\lambda_{C-(c_i)} \in [a_{C-(c_i)}, b_{C-(c_i)}]$ 及相应的分类。

步骤 3 为保证包含全部因素和删除某一因素时各阈值范围的合理性,确定整体阈值范围 $\lambda_k \in [\max(a_C, a_{C-(c_i)}), \min(b_C, b_{C-(c_i)})], 1 \leq k \leq p$ 。

步骤 4 确定各因素的重要性。以没有删除任何因素的分类为基准,把没有删除任何一个因素视为一种知识分类,将删除各个因素后视为另一种知识分类,讨论删除各因素后的分类相对于总的因素分类的正域。其本质上是删除一个因素后的分类可以准确划分到没有删除任何一个因素的分类中的对象的集合。故根据定义 4 和式(5)(令 $B = C$),在某一阈值水平 λ_k ,某因素 c_i 的重要性可表示为

$$sig^{\lambda_k}(\{c_i\}, C, C) = \gamma_C^{\lambda_k}(C) - \gamma_{C-(c_i)}^{\lambda_k}(C) = 1 - \gamma_{C-(c_i)}^{\lambda_k}(C) \quad (6)$$

综合 p 个阈值水平,则各因素综合重要性为

$$sig(\{c_i\}, C, C) = \frac{1}{p} \sum_{k=1}^p sig^{\lambda_k}(\{c_i\}, C, C) \quad (7)$$

步骤 5 根据综合重要性的大小确定各因素的权重分配,即

$$w_i = sig(\{c_i\}, C, C) / \sum_{j=1}^m sig(\{c_j\}, C, C) \quad (8)$$

(3) 模糊综合评判。根据式(9)对评估样本进行无量纲化处理,得到 X', 从而使每一个指标值统一于同一数值特性范围[0, 1]。

$$\begin{cases} x'_{ij} = \frac{x_{ij} - \min_{i=1}^n(x_{ij})}{\max_{i=1}^n(x_{ij}) - \min_{i=1}^n(x_{ij})}, & x_{ij} \text{ 为效益型指标} \\ x'_{ij} = \frac{\max_{i=1}^n(x_{ij}) - x_{ij}}{\max_{i=1}^n(x_{ij}) - \min_{i=1}^n(x_{ij})}, & x_{ij} \text{ 为成本型指标} \end{cases} \quad (9)$$

模糊综合评判结果为

表 2 分类及 F 值

λ 值	0.95	0.92	0.91	0.87	0.83	0.82	0.80
r 类	9	7	6	5	4	3	2
F	114.349 1	55.960 5	87.362 4	76.798 6	42.537 7	47.840 8	25.969 0
$F_{0.005}(r-1, n-r)$	23 925	44.84	22.46	15.56	12.92	12.4	14.69
$F - F_{0.005}(r-1, n-r)$	—	11.120 5	64.902 4	61.238 6	29.617 7	35.440 8	11.279 0

由于所有对象各自成类或全部对象并入一类没有实际意义,因此不考虑 $\lambda = 1$ 和 $\lambda = 0.78$ 的情况。从表 2 可以看出,符合条件 $F > F_{0.005}$ 合理的阈值范围为 $\lambda_C \in [0.8, 0.92]$ 。

$$B = w \circ (X')^T = \{b_1, b_2, \dots, b_n\} \quad (10)$$

式中,“ \circ ”代表模糊合成算子。

3 实例分析

以某飞行视景仿真系统^[13]为例,综合分析影响仿真系统可信性的各方面因素。选取有效性(c_1)、逼真度(c_2)、可靠性(c_3)、网络延时(c_4)、输入输出延时(c_5)、算法精度(c_6) 6 个指标,根据仿真系统的多次运行情况,建立评估样本如表 1 所示。其中有效性和逼真度为定性指标,其度量可通过专家评判的方法获得;其余指标均为定量指标,其度量根据仿真系统的多次运行而获得。采用平均故障间隔时间(mean time between failures, MTBF)来衡量系统的可靠性,单位为小时/次,网络延时和输入输出延时的单位分别为 ms 和 μs ,算法精度的度量为仿真数据的小数点后的计算位数。

对该评估样本进行基于模糊传递闭包法的模糊聚类分析,动态聚类图如图 1 所示。

表 1 评估样本

	c_1	c_2	c_3	c_4	c_5	c_6
1	0.90	0.85	5 000	15	300	10
2	0.70	0.80	4 000	25	350	8
3	0.55	0.60	2 800	30	420	6
4	0.40	0.50	2 000	35	500	4
5	0.20	0.40	1 200	50	700	2
6	0.85	0.90	4 800	20	320	10
7	0.60	0.60	3 000	32	450	6
8	0.40	0.55	2 500	40	600	5
9	0.25	0.35	1 000	45	650	2
10	0.35	0.35	1 500	43	630	2

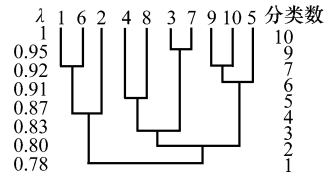


图 1 动态聚类图

基于 F-统计量确定合理阈值范围。利用评估样本的原始数据,根据式(1)分别计算各个 λ 所对应的 F-统计量的值,计算结果如表 2 所示。

类似地,根据步骤 2,用 F-统计量确定删除条件因素 c_i 后合理的阈值范围:删除因素 c_1 后,符合条件 $F > F_{0.005}$ 合理的阈值范围 $\lambda_{C-(c_1)} \in [0.8, 0.93]$;删除因素 c_2 后,合理的

阈值范围 $\lambda_{C-(c_2)} \in [0.83, 0.92]$; 删除因素 c_3 后合理的阈值范围 $\lambda_{C-(c_3)} \in [0.84, 0.93]$; 删除因素 c_4 后合理的阈值范围 $\lambda_{C-(c_4)} \in [0.82, 0.94]$; 删除因素 c_5 后合理的阈值范围 $\lambda_{C-(c_5)} \in [0.83, 0.93]$; 删除因素 c_6 后合理的阈值范围 $\lambda_{C-(c_6)} \in [0.83, 0.92]$ 。

根据步骤3, 确定整体合理阈值范围 $\lambda_k \in [0.84, 0.92]$, $1 \leq k \leq 8$ 。

根据步骤4, 在整体合理阈值范围内, 根据相应的分类情况, 确定各因素的重要性。

当 $\lambda_1 = 0.92$ 时, 包含全部因素的样本分为7类: $\{1, 6\}$ 、 $\{3, 7\}$ 、 $\{9, 10\}$ 、 $\{2\}$ 、 $\{4\}$ 、 $\{5\}$ 、 $\{8\}$; 删除因素 c_1 后的样本分为6类: $\{1, 6\}$ 、 $\{3, 7\}$ 、 $\{5, 9, 10\}$ 、 $\{2\}$ 、 $\{4\}$ 、 $\{8\}$; 删除因素 c_2 后的样本分为6类: $\{1, 6\}$ 、 $\{3, 7\}$ 、 $\{5, 9, 10\}$ 、 $\{2\}$ 、 $\{4\}$ 、 $\{8\}$; 删除因素 c_3 后的样本分为7类: $\{1, 6\}$ 、 $\{3, 7\}$ 、 $\{9, 10\}$ 、 $\{2\}$ 、 $\{4\}$ 、 $\{5\}$ 、 $\{8\}$; 删除因素 c_4 后的样本分为6类: $\{1, 6\}$ 、 $\{3, 7\}$ 、 $\{5, 9, 10\}$ 、 $\{2\}$ 、 $\{4\}$ 、 $\{8\}$; 删除因素 c_5 后的样本分为6类: $\{1, 6\}$ 、 $\{3, 7\}$ 、 $\{5, 9, 10\}$ 、 $\{2\}$ 、 $\{4\}$ 、 $\{8\}$; 删除因素 c_6 后的样本分为7类: $\{1, 6\}$ 、 $\{3, 7\}$ 、 $\{9, 10\}$ 、 $\{2\}$ 、 $\{4\}$ 、 $\{5\}$ 、 $\{8\}$ 。

根据各分类情况及式(6)可知, 各因素的重要性分别为 $(0.3, 0.3, 0.0, 0.3, 0.3, 0)$ 。

类似地, 当 $\lambda_2 = 0.91$ 时, 在阈值水平 λ_2 下, 各因素的重要性分别为 $(0, 0, 0, 0, 0.2, 0)$; 当 $\lambda_3 = 0.89$ 时, 各因素的重要性分别为 $(0, 0, 0, 0.2, 0.2, 0)$; 当 $\lambda_4 = 0.88$ 时, 各因素的重要性分别为 $(0, 0.2, 0.2, 0.2, 0.2, 0.2)$; 当 $\lambda_5 = 0.87$ 时, 各因素的重要性分别为 $(0, 0, 0, 0, 0, 0)$; 当 $\lambda_6 = 0.86$ 时, 各因素的重要性分别为 $(0.3, 0, 0, 0, 0, 0.3)$; 当 $\lambda_7 = 0.85$ 时, 各因素的重要性分别为 $(0.7, 0.3, 0.3, 0, 0.4, 0.3)$; 当 $\lambda_8 = 0.84$ 时, 各因素的重要性分别为 $(0.7, 0.3, 0.7, 0.3, 0.7, 0.7)$ 。

由式(7)可知, 各因素的综合重要性分别为 $(0.2146, 0.1199, 0.1274, 0.1103, 0.2175, 0.1596)$ 。由步骤5和式(8)可知, 各因素权重分配为 $\omega = (0.2273, 0.1250, 0.1364, 0.1136, 0.2273, 0.1705)$ 。

评估样本中, c_4 和 c_5 属于成本型指标, 其余均属效益型指标, 由式(9)对评估样本进行数据无量纲化处理, 得 \mathbf{X}' 。采用主因素突出型模糊算子 $M(\cdot, \vee)$ 进行运算, 则由式(10)可得最终的评价值为

$$B = \omega \cdot (\mathbf{X}')^T = (2.00, 1.75, 1.40, 1.00, 0.10, 1.90, 1.25, 0.57, 0.25, 0.43)$$

因此, 仿真可信性排序结果为 $b_1 > b_6 > b_2 > b_3 > b_7 > b_4 > b_8 > b_{10} > b_9 > b_5$, 由此看出该仿真系统第1次运行的可信性最高, 第5次最低。分析评估样本, 第1次运行的各项指标均明显优于第5次, 而最终仿真可信性排序结果与系

统实际运行情况也基本吻合。因此, 本文所提出的评估方法是合理可行的。

4 结论

为客观而有效地解决仿真可信性评估问题, 减少个人的主观判断、选择和偏好对评判结果的影响, 提出基于模糊聚类和粗糙集的仿真可信性模糊综合评估方法。该方法利用模糊聚类和粗糙集中的属性重要性原理客观地进行各因素权重分配, 并结合模糊综合评判进行仿真可信性综合评估, 客观而充分地挖掘出各因素之间的数据关系, 克服了传统评估方法过分依赖专家经验知识的不足。应用该方法解决了仿真可信性评估问题涉及的大量数据和计算, 提高了评估效率。下一步有必要进行可信性评估自动化及计算机辅助工具的研究。

参考文献:

- [1] 刘飞, 马萍, 杨明, 等. 大型复杂仿真系统可信度量化方法研究[J]. 四川大学学报: 工程科学版, 2006, 38(5): 170-174.
- [2] 徐迪. 基于相似理论的系统仿真可信性分析[J]. 系统工程理论与实践, 2001, 4: 49-53.
- [3] 焦鹏, 唐见兵, 查亚兵. 仿真可信度评估中相似度方法的改进及其应用[J]. 系统仿真学报, 2007, 19(12): 2658-2661.
- [4] 杨惠珍, 康凤举, 李俊. 层次分析法在水下航行器系统仿真可信性评估中的应用研究[J]. 系统仿真学报, 2002, 14(10): 1299-1301.
- [5] 杨惠珍, 康凤举, 李俊. 基于模糊 AHP 的系统仿真可信度评估方法[J]. 计算机仿真, 2003, 20(8): 43-46.
- [6] 张伟, 王行仁. 仿真可信度模糊评判[J]. 系统仿真学报, 2001, 13(4): 473-476.
- [7] 柳世考, 刘兴堂. 一种有效评估大型复杂仿真系统可信度的新方法[J]. 系统仿真学报, 2001, 13(5): 666-670.
- [8] 梁保松, 曹殿立. 模糊数学及其应用[M]. 北京: 科学出版社, 2007.
- [9] 陈水利, 李敬功, 王向公. 模糊集理论及其应用[M]. 北京: 科学出版社, 2005.
- [10] 苗夺谦, 李道国. 粗糙集理论、算法与应用[M]. 北京: 清华大学出版社, 2008.
- [11] 黄定轩, 武振业, 宗蕴璋. 基于属性重要性的多属性客观权重分配方法[J]. 系统工程理论方法应用, 2004, 13(3): 203-208.
- [12] 黄定轩. 基于客观信息熵的多因素权重分配方法[J]. 系统工程理论方法应用, 2003, 12(4): 321-325.
- [13] Wu Jing, Wu Xiaoyan, Zhou Yanyan. Research on design and realization of steering flight visual simulation system [C] // IEEE Asia Simulation Conference 2008/the 7th International Conference on System Simulation and Scientific Computing, Beijing, China, 2008: 1567-1570.