

粗糙集在数据融合系统灰色评估中的应用

李 阳, 王朝英, 郭威武

(空军工程大学 电讯工程学院, 陕西 西安 710077)

摘 要: 为了对数据融合系统性能进行综合评估, 建立了相对完备的分布式融合系统性能指标体系. 提出了一种基于粗糙集的数据融合系统灰色评估模型, 通过引入粗糙集理论, 对指标进行约简并确定权值, 减少了主观因素的影响, 提高了评估的公平性. 通过算例证明了模型的可行性和可靠性, 为数据融合系统的开发、论证和实际使用提供了科学的理论决策依据.

关键词: 数据融合; 性能评估; 粗糙集; 灰色关联

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2010)07-0106-04

Application of Rough Set in Grey Relation Evaluation of Data Fusion System

LI Yang, WHANG Zhao-ying, GUO Wei-wu

(Institute of Telecommunication, Air Force Engineering University, Xi'an 710077, China)

Abstract: In order to evaluate the overall performance of data fusion system, an approximately complete set of criterions is established for the distributed kind. An evaluation model of grey relation analysis based on rough set theory is proposed, which by means of applying rough set theory reduce redundant attributes and determine criterion weight values, thus resulting in less subjectivity and more objective evaluation results. The feasibility and credibility of the model is proved by an example. It provides a scientific decision foundation theoretically for development, demonstration and practical application of the data fusion system.

Key words: data fusion; performance evaluation; rough set; grey correlation

1 引言

自 1970 年代以来, 多传感器数据融合技术作为新兴学科诞生并取得迅速发展, 越来越多地应用于包括 C³I 系统和各种武器平台在内的军事领域和许多民事领域^[1]. 1990 年代中后期以来, 我国也进行了大量研究, 但主要集中于目标跟踪、融合算法方面^[2-3]. 在对融合算法和系统性能进行评估方面国内外研究都比较欠缺, 主要原因是融合系统所汇集和处理的信息种类繁多, 信息源之间关系复杂, 应用背景和建设目标差异较大, 使得客观、公正地评估融合系统性能变得比较困难, 导致系统的设计、开发带

有一定的盲目性. 鉴于此, 本文首先针对分布式融合系统建立一套性能评估指标体系; 然后根据数据融合系统性能指标的灰色性, 结合粗糙集和灰色关联理论建立数据融合系统综合性能评估模型. 最后通过算例证明模型的可行性和可靠性.

2 性能指标的建立

数据融合系统通过对来自不同传感器的数据进行自动分析、优化综合等处理, 获得对战场态势的准确了解^[1]. 根据分布式数据融合系统的特点、应用目标和功能, 本文采用如下指标体系^[4], 该指标体系分为航迹起始、航航相关、目标识别、跟踪精度等

4类一级指标和15类二级指标,如图1所示.指标体系的建立为后续的评估过程打下基础.

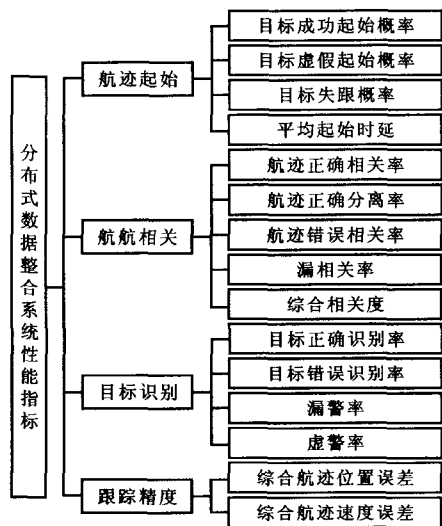


图1 分布式数据融合系统性能指标体系

3 粗糙集理论基础

粗糙集(Rough set)理论^[5]是由波兰数学家 Z. Pawlak 于 1982 年提出来的,它是一种刻画不完全性和不确定性的数学工具.它将知识理解为对数据的划分,由于其仅依赖于原始数据而不需要任何先验信息便可揭示数据内部潜在的规律,因而广泛应用于模式识别、故障诊断等领域.

3.1 知识表达系统

定义1 设四元组 $S = (U, A, V, f)$ 是一个知识表达系统,其中 $U = \{x_1, x_2, \dots, x_m\}$ 为有限非空集合,也称为论域或对象空间; $A = \{a_1, a_2, \dots, a_n\}$ 为属性的有限非空集合. $V = \bigcup V_a$, 其中 $a \in A$, V_a 为属性 a 的值域; $f: U \times A \rightarrow V$ 为信息函数,对于 $\forall a \in A, \forall x \in U, f(x, a) \in V$,它指定了 U 中每一对象的属性值.

定理1 对任一属性子集 $P \subseteq A$ 且 $P \neq \emptyset$, $\exists$ 对象 $x_1, x_2 \in U$, 对 $\forall b \in P$, 如果满足 $f(x_1, b) = f(x_2, b)$, 则称对象 x_1, x_2 在属性集 P 上不可分辨,记作 $\text{Ind}(P)$.

不可分辨关系是一种等价关系,具有自反性、对称性和传递性.符号 U/A 表示 U 中的对象根据属性集 A 构成的所有等价关系集.

定义2 对属性 $a_i \in A$, 如果 $\text{Ind}(A) = \text{Ind}(A - \{a_i\})$, 则称属性 a_i 是冗余的, 否则, 称 a_i 是独立的或者必要的. 当对任意 $a \in A$, 如果 a 是独立的, 则称 A 为独立的. 当 A 为独立的, 且 $B \subseteq A$, 则 B 也

是独立的.

3.2 属性约简及重要性

定义3 若 $\forall P \subseteq A$ 是独立的, 并且 $\text{Ind}(P) = \text{Ind}(A)$, 则称 P 是 A 的一个约简.

A 的约简不是唯一的, A 的约简记为 $\text{red}(A)$.

定义4 对 $\forall P \subseteq A, \forall p \in P$, 属性 p 的重要性为

$$\text{sig}_P^1(p) = r(P) - r(P - \{p\})$$

式中, $r(P) = |\text{pos}_P(A)| / |U|$ 称为 P 的近似质量. $\text{sig}_P^1(p)$ 表示从属性集 P 中去掉某些属性 p 后, 分类 U/P 的正域受到影响的程度, 是从决策分类的角度对属性 p 重要性的一种度量.

定义5 设 $S = (U, A, V, f)$ 为一个信息系统, $\forall P \subseteq A, a \in P, U/P = \{p_1, p_2, \dots, p_t\}$, 属性 a 在属性集 P 中的重要性定义为

$$\text{Sig}_{P \setminus \{a\}}^2 = I(P) - I(P \setminus \{a\})$$

式中, $I(P) = 1 - \frac{1}{|U|^2} \sum_{i=1}^t |P_i|^2$, 表示属性集 P 的信息量. $\text{sig}_{P \setminus \{a\}}^2(a)$ 表示从属性集 P 中去掉属性 a 后引起的信息量变化的程度, 是属性在属性集中重要性的一种度量.

定义6 当属性 $p \in P$ 时, 综合考虑属性 p 对决策分类的重要性和其在属性集 P 中的重要性, 定义加权平均属性重要性为^[6]

$$\text{sig}(p) = w \text{sig}_P^1(p) + (1 - w) \text{sig}_{P \setminus \{p\}}^2(p)$$

式中, $0 \leq w \leq 1$.

3.3 属性权重计算

权重反映属性之间重要性的差异. 根据定义 4 ~ 6, 约简后属性 p 的权重可由其加权平均属性重要性来度量, 它综合考虑了属性对决策分类的重要性的信息量, 物理意义明确且不受主观因素干扰, 增强了评估的客观性.

定义7 若 $P \subseteq A, P = \{p_1, p_2, \dots, p_m\}$ 且 $\text{Ind}(P) = \text{Ind}(A)$, 则属性 p_i 的权重为

$$w_i = \text{sig}(p_i) / \sum_{j=1}^m \text{sig}(p_j)$$

式中, $0 \leq w_i \leq 1$ 且 $\sum_{i=1}^m w_i = 1$.

4 评估模型的建立

设有 m 个数据融合系统组成评估对象 $O_1, O_2, \dots, O_m$, 及 n 个一级指标 $V_1, V_2, \dots, V_n$, 各一级指标下辖若干二级指标, 共同构成多层次多属性决策评估系统. 现利用粗糙集和灰色关联理论建立评估

模型:

4.1 属性约简、权重确定

对各一级指标分别应用粗集理论进行属性约简,并确定权值.设对一级指标 V_i ,根据定义 1~7,利用粗集理论进行属性约简后,得

$$V_i = (v_{i1}, v_{i2}, \dots, v_{it_i}).$$

对应的权重向量为

$$W_i = (w_{i1}, w_{i2}, \dots, w_{it_i}).$$

式中, t_i 为一级指标 V_i 经属性约简后所剩属性的个数.

n 个一级指标 $V_1, V_2, \dots, V_n$ 的属性权重向量为

$$W = [w_1, w_2, \dots, w_n].$$

W 反映了各一级指标的相对重要性,一般由专家调查法得出.

4.2 建立评估矩阵

根据上步所约简的指标,确定一级指标下的评估矩阵:

$$A_i = [x'_{ij}]_{t_i \times m} = \begin{bmatrix} x_{i11} & x_{i12} & \dots & x_{i1m} \\ x_{i21} & x_{i22} & \dots & x_{i2m} \\ \vdots & \vdots & & \vdots \\ x_{it_i1} & x_{it_i2} & \dots & x_{it_i m} \end{bmatrix}$$

式中, $x'_{ij} = x_i(O_j)(i' = 1, 2, \dots, t_i; j = 1, 2, \dots, m)$ 为评估对象 O_j 关于指标 $v_{ii'}$ 的观测值.

总评估矩阵为

$$A = [A_1; A_2; \dots; A_n].$$

4.3 确定参考向量

根据指标的性质(极大型,极小型),选取各项指标的最优值构成参考向量 y_0 :

$$y_0 = [y_1, y_2, \dots, y_l],$$

式中, $l = t_1 + t_2 + \dots + t_n$, 为约简后所有二级指标个数之和.

4.4 评估指标的规范化

为了消除不同性质指标间量纲和数量级的影响,需对指标进行规范化处理,将其规范化为 $[0, 1]$ 区间的极大型指标^[7].

对于极大型指标:

$$x'_{ij} = \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}} \quad (1)$$

对于极小型指标:

$$x'_{ij} = \frac{\max_i x_{ij} - x_{ij}}{\max_i x_{ij} - \min_i x_{ij}} \quad (2)$$

为叙述方面,以下 $x'_{ij} = x'_{ij}$.

4.5 计算关联度

评估对象 O_j 于指标 $v_{ii'}$ 下的关联系数 $d_{ij}(i' = 1, 2, \dots, t_i; j = 1, 2, \dots, m)$ 由下式确定:

$$d_{ij} = \frac{\Delta_1 + \lambda \Delta_2}{\Delta_3 + \lambda \Delta_2} \quad (3)$$

$\Delta_1 = \min_i \min_j |x_{ij} - y_i'|, \Delta_2 = \max_i \max_j |x_{ij} - y_i'|, \Delta_3 = |x_{ij} - y_i'|, \lambda$ 为分辨系数,一般取 $\lambda = 0.5$.

于是,在一级指标 V_i 下各评估对象相对于最优系统 y_0 的灰色关联系数矩阵为

$$D_i = [d_{ij}]_{t_i \times m} = \begin{bmatrix} d_{i11} & d_{i12} & \dots & d_{i1m} \\ d_{i21} & d_{i22} & \dots & d_{i2m} \\ \vdots & \vdots & & \vdots \\ d_{it_i1} & d_{it_i2} & \dots & d_{it_i m} \end{bmatrix}$$

总灰色关联系数矩阵为

$$D = [D_1; D_2; \dots; D_n]$$

4.6 多层次关联度合成

对一级指标 V_i ,各评估对象相对参考向量 y_0 的关联度向量为

$$R_i = W_i * D_i = (r_{i1}, r_{i2}, \dots, r_{im}) \quad (4)$$

式中, r_{ij} 为评估对象 O_j 在一级指标 V_i 下与对应参考向量的关联度.

所有评估对象相对参考向量 y_0 的关联度矩阵为

$$R = [R_1; R_2; \dots; R_n] = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ r_{21} & r_{22} & \dots & r_{2m} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nm} \end{bmatrix}$$

进一步可求得各评估对象总的关联度向量为

$$R' = W * R = [r_1, r_2, \dots, r_m] \quad (5)$$

式中, r_i 为评估对象 O_i 与参考向量的最终相关度.对 r_i 进行大小排序,就各数据融合系统综合效能的优劣顺序.

5 实例应用

为了说明该方法使用步骤,现假设有某 7 种类型数据融合系统,采用本文提出的综合效能评估模型对其进行效能排序.各指标名称依次以字母代替,指标数据如表 1 所示.

表1 数据融合系统评估指标数据

序号	目标起始(C_1)				航迹相关(C_2)					目标识别(C_3)				跟踪精度(C_4)	
	$C_{11}/\%$	$C_{12}/\%$	$C_{13}/\%$	$C_{14}/\%$	$C_{21}/\%$	$C_{22}/\%$	$C_{23}/\%$	$C_{24}/\%$	$C_{25}/\%$	$C_{31}/\%$	$C_{32}/\%$	$C_{33}/\%$	$C_{34}/\%$	$C_{41}/$ (m/s)	$C_{42}/$ (m/s ²)
1	100	1.27	0.67	3.45	98.08	98.00	1.92	2.00	98.62	91.70	2.40	3.20	2.70	44.97	2.12
2	98.00	2.05	0.72	3.05	96.56	97.32	3.44	2.68	96.88	93.40	2.50	1.80	2.30	43.21	3.01
3	89.00	1.79	0.57	2.65	98.03	96.87	1.97	3.13	98.51	96.00	2.50	0.70	0.80	50.64	2.87
4	92.00	1.52	0.60	2.45	97.05	97.65	2.95	2.35	97.87	97.05	1.50	0.45	1.00	55.00	3.21
5	100	1.79	0.57	2.55	96.56	96.87	3.44	3.13	97.21	96.00	2.48	1.24	0.28	42.15	2.65
6	94.00	2.05	0.53	2.30	95.60	95.50	4.40	4.50	96.25	95.00	1.70	2.10	1.20	45.45	2.64
7	90.00	2.20	0.70	3.00	97.03	96.80	2.97	3.20	97.20	94.36	2.32	1.21	2.11	52.00	3.00

现以一级指标目标起始为例,说明属性约简及权重确定方法。

以各属性的中间值为阈值对指标进行离散化。计算加权平均属性重要性时,取 $w = 0.9$, $\text{Ind}(\{C_{11}, C_{12}, C_{13}\}) = \text{Ind}(C_1)$, 因而对目标起始一级指标而言,约简后属性集为 $C'_1 = \{C_{11}, C_{12}, C_{13}\}$, 将属性 C_{14} 去除后,由定义4~6得 $\text{sig}(C_{1i}) = [0.655, 0.655, 0.522]$; 由定义7对其归一化后作为各属性的权值 $W_1 = [0.357, 0.357, 0.286]$. 同理可得其他一级指标的约简及属性权值: $C'_2 = \{C_{23}, C_{24}, C_{25}\}$, $W_2 = [0.273, 0.546, 0.181]$, $C'_3 =$

$\{C_{32}, C_{33}, C_{34}\}$, $W_3 = [0.273, 0.454, 0.273]$. 一级指标跟踪精度下只有两个二级指标,数量较少因而略去属性约简步骤,且赋各指标等权值得: $C'_4 = \{C_{41}, C_{42}\}$, $W_4 = [0.500, 0.500]$. 为简便起见,视各一级指标同等重要,则 $W = [0.5, 0.5, 0.5, 0.5]$.

然后,根据指标性质得参考向量 y_0 :

$$y_0 = [100, 1.27, 0.53, 1.92, 2.00, 98.62, 1.50, 0.45, 0.28, 42.15, 2.12].$$

进而由式(1)、式(2)对指标值进行规划化处理,由式(3)计算总灰色关联系数矩阵为

$$D = \begin{bmatrix} 1 & 0.7333 & 0.3333 & 0.4074 & 1 & 0.4784 & 0.3548 \\ 1 & 0.3735 & 0.4721 & 0.6503 & 0.4721 & 0.3735 & 0.3333 \\ 0.4043 & 0.3333 & 0.7073 & 0.5758 & 0.7037 & 1 & 0.3585 \\ 1 & 0.4493 & 0.9612 & 0.5463 & 0.4493 & 0.3333 & 0.5415 \\ 1 & 0.6477 & 0.5252 & 0.7812 & 0.5252 & 0.3333 & 0.5102 \\ 1 & 0.4051 & 0.9151 & 0.6124 & 0.4566 & 0.3333 & 0.4549 \\ 0.3333 & 0.4595 & 0.4595 & 1 & 0.4645 & 0.8095 & 0.5090 \\ 0.3333 & 0.4194 & 0.7959 & 1 & 0.5524 & 0.3714 & 0.5620 \\ 0.3333 & 0.3746 & 0.6994 & 0.6269 & 1 & 0.5681 & 0.3980 \\ 0.6950 & 0.8584 & 0.4308 & 0.3333 & 1 & 0.6607 & 0.3948 \\ 1 & 0.3787 & 0.4208 & 0.3333 & 0.5070 & 0.5117 & 0.3825 \end{bmatrix}$$

然后,由式(4)计算评估对象于各一级指标层下的关联度向量:

$$R_1 = W_1 * D_1 = [0.8302, 0.4907, 0.4885, 0.5422, 0.7268, 0.5895, 0.3482]$$

$$R_2 = W_2 * D_2 = [1.0000, 0.5496, 0.7148, 0.6865, 0.4921, 0.3333, 0.5087]$$

$$R_3 = W_3 * D_3 = [0.3333, 0.4181, 0.6777, 0.8981, 0.6506, 0.5447, 0.5028]$$

$$R_4 = W_4 * D_4 = [0.8475, 0.6186, 0.4258, 0.3333, 0.7535, 0.5862, 0.3886]$$

最后,由式(5)求得各评估对象总关联度向量:

$$R' = W * R = [0.3926, 0.2597, 0.2908, 0.3209, 0.3232, 0.2515, 0.2176]$$

根据评估值得出优劣次序为 $1 > 5 > 4 > 3 > 2 > 6 > 7$. 此结果与文献[7](结果为 $1 > 5 > 3 > 4 > 2 > 6 > 7$)相比,约简了评估指标,消除了指标间依赖关系;与文献[8](结果为 $1 > 5 > 4 > 3 > 2 > 6 > 7$)相

比,本文综合考虑了属性对决策分类的重要性的信息量,提高了指标权值确定的客观性;因而本文评估结果更具可信性,与专家分析判断也相一致。

(下转第113页)

用LS-SVM判别方法进行判别,提高了识别速度.在ORL人连数据库上的实验表明,该方法具有较高的识别率和识别速度.有效解决了利用传统支持向量机方法进行识别中存在的识别速度慢的问题.如何进一步提高识别率有待于进一步的探索研究.

参考文献:

- [1] 刘青山,卢汉清,马颂德.综述人脸识别中的子空间方法[J].自动化学报,2003,29(6):900-911.
- [2] 张燕昆,杜平,刘重庆.基于主元分析与支持向量机的人脸识别方法[J].上海交通大学学报,2002,36(6):884-886.
- [3] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000(290):2323-2326.
- [4] 郭锋,刘丽丽,吕凝.基于LLE和SVM的人像识别方法[J].吉林大学学报:信息科学版,2008,26(1):48-54.
- [5] 王自强,钱旭,孔敏.流形学习算法综述[J].计算机工程与应用,2008,44(35):9-12.
- [6] 周贤君,胡金演,郭凤.基于PCA算法的人脸图像预处理研究[J].电视技术,2008,32(7):17-19.
- [7] Vapnik V. The nature of statistical learning theory[M]. New York:Springer, 1995.
- [8] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers [J]. Neural Processing Letter, 1999,9(3):293-300.
- [9] Blanz V, Sch EL Kopf B, Vapnik V N. Extracting support data for a given task[C]// Proceedings of the 1st International Conference on Knowledge Discovery & Data Mining. Menlo Park, CA: AAAI Press, 1995: 252-257.
- [10] Hsu C W, Lin C J. A comparison of methods for Multi-class support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13 (2):415-425.

作者简介:

谈书才 男,(1984-),硕士研究生.研究方向为模式识别与智能系统.
黄景涛 男,(1977-),博士,副教授.研究方向为数据挖掘和智能优化算法及应用.

(上接第109页)

6 结束语

利用粗糙集和灰色关联理论对数据融合系统综合效能进行评估,在不影响关键信息的前提下,最大限度地消除了冗余指标,降低了计算量,并根据指标的重要性和所含信息量确定其权值,减少了评估中主观因素的影响,提高了评估的有效性和客观性.对数据融合系统的研制、论证和改进具有重要意义.

参考文献:

- [1] 何友,王国宏.多传感器信息融合及应用[M].北京:电子工业出版社,2000:1-10.
- [2] 彭云辉,缪栋,刘冬,等. UPF 算法在状态估计中的应用[J].微电子学与计算机,2006,23(11):41-43.
- [3] 郝晶,李卫忠,白剑林.一种多速率的交互式多模型粒子滤波算法[J].微电子学与计算机,2009,26(2):163-166.
- [4] 田杰,陈杰,窦丽华,等.多传感器信息融合性能的测试与评估[J].计算机工程,2003,29(10):84-85.
- [5] 营利荣.面向不确定性决策的杂合粗糙集方法及其应用[M].北京:科学出版社,2008:12-22.
- [6] 李永华,蒋芸,王小菊.一种基于rough集的属性约简的改进算法[J].计算机应用,2008,28(8):2001-2002.
- [7] 程钦文,沈云春,唐纯洁.潜射鱼雷作战能力综合评价模型[J].火力与指挥控制,2004,29(1):110-111.
- [8] 胡方,黄建国,褚福照.基于粗糙集的武器系统灰色关联评估模型[J].兵工学报,2008,29(2):254-255.

作者简介:

李阳 男,(1985-),硕士研究生.研究方向为数据融合.
王朝英 女,(1965-),副教授,硕士生导师.研究方向为数据融合、图像处理等.
郭威武 男,(1983-),硕士研究生.研究方向为军事通信.