

# 基于粗糙集理论的分布式检测系统数据融合

张素妍 聂伟

(北京化工大学信息科学与技术学院, 北京 100029)

**摘要:** 探讨分布式检测系统数据融合的粗处理方法, 给出基于粗糙集理论的分布式检测系统的数据融合结构和相容算法, 通过不可分辨关系确定属性约简、导出决策规则, 并将其应用于油田输油管测漏系统中, 结果证明了该方法的合理性和有效性。

**关键词:** 分布式检测系统; 数据融合; 粗糙集

**中图分类号:** TP182

## 引言

数据融合作为一种数据综合处理技术, 在军事、医学诊断、故障诊断、金融分析等领域得到了广泛应用, 是目前研究的热点课题。传统的数据处理方法包括 Bayes 决策理论、D-S 证据推理、产生式规则、模糊推理、神经网络以及卡尔曼滤波等。这些方法都各有特点。采用数据融合技术的系统都具有多个信息源、各信息源数据不完备等特点, 融合中心的任务就是对这些不完备、不确定以及随机的数据进行综合分析, 以达到正确分类的目的。

粗糙集理论<sup>[1]</sup>是波兰科学家 Pawlak 于 1982 年提出的, 处理不完整数据和不精确知识的强有力的数学工具, 目前已经成为人工智能领域的一个新的学术热点, 在知识获取<sup>[2]</sup>、知识分析<sup>[3-4]</sup>和决策分析<sup>[5]</sup>等方面得到了广泛的应用, 在数据融合技术<sup>[6]</sup>中也有一定程度的应用, 受到了国内外专家和科研人员的广泛关注。其优点是不需要预先给定检测对象的某些属性或特征的数学描述, 而是直接从给定问题的知识分类出发, 通过不可分辨关系和不可分辨类确定对象的知识约简、导出问题的决策规则。本文将粗糙集理论应用于分布式检测系统, 探讨基于粗糙集理论的分布式检测系统的数据处理方法。文中给出了系统结构和融合方法<sup>[7]</sup>, 该方法以局部决策作为条件属性, 融合中心采用粗糙集理论方法对局部决策进行处理, 包括知识表达、化简, 最后给

出决策规则, 并将其应用于油田输油管测漏系统, 结果证明了该方法的合理性和有效性。

## 1 分布式检测系统结构及数据融合方法

分布式检测系统结构见图 1。系统包括检测对象、局部处理器  $D_1, D_2, \dots, D_n$  和数据融合中心。局部处理器首先通过各自的传感器  $S_1, S_2, \dots, S_n$  采集检测对象不同方面的特征, 然后根据这些特征作出局部判决 ( $I_1, I_2, \dots, I_n$  为局部判决量)。数据融合中心采用一定的方法对局部判决结果作处理, 得出最终判决 ( $O$  为最终判决量)。

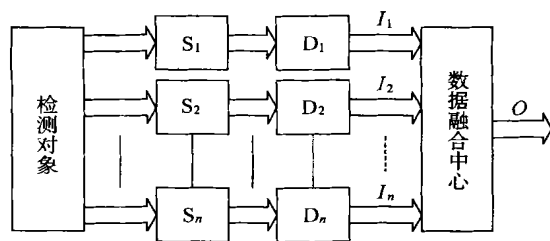


图1 分布式检测系统结构

Fig. 1 Structure of distributing measure system

基于粗糙集理论的分布式检测系统中, 关于融合中心的知识表达采用决策表的形式。决策表由论域中的对象、条件属性和决策属性组成, 论域  $U$  是由各局部检测器的检测次数所组成的集合; 条件属性集  $C = \{a_1, a_2, \dots, a_n\}$  由检测系统中的局部判决量所组成的集合表示, 条件属性值就是局部决策结果  $u_{mij}$  ( $m = 1, 2, \dots, k; i = 1, 2, \dots, n; j = 1, 2, \dots, p$ ); 决策属性集  $D = \{d\}$  由融合中心输出的各个最终决策量所组成的集合表示, 属性值为  $u_{mrb}$ , 见表

收稿日期: 2003-05-08

第一作者: 女, 1975年生, 硕士生

E-mail: zhangsuyan-z@sohu.com

1。

表 1 融合中心决策表

Table 1 Knowledge table of fusion center

| $U$ | $a_1$     | $a_2$     | $a_3$     | ... | $a_n$     | $d$       |
|-----|-----------|-----------|-----------|-----|-----------|-----------|
| 1   | $u_{11j}$ | $u_{12j}$ | $u_{13j}$ | ... | $u_{1nj}$ | $u_{1rb}$ |
| 2   | $u_{21j}$ | $u_{22j}$ | $u_{23j}$ | ... | $u_{2nj}$ | $u_{2rb}$ |
| ... | ...       | ...       | ...       | ... | ...       | ...       |
| $k$ | $u_{k1j}$ | $u_{k2j}$ | $u_{k3j}$ | ... | $u_{knj}$ | $u_{krb}$ |

为说明问题,这里先介绍三个概念。设  $Q$  为  $U$  上的一个等价关系(即分类)族,若存在另一等价关系族  $P \subset Q$ ,则  $P(P$  中全部等价关系的交集)就称为  $P$  上的不可分辨关系,记为  $\text{ind}(P)$ ;若  $Y$  为  $U$  的一个子集,  $x \in U$ ,  $[x]_Q$  表示所有与  $x$  不可分辨的对象所组成的集合,这样关于  $Y$  的上、下近似集可表示为:

$$Q^+(Y) = \{x \in U \mid [x]_Q \subseteq Y\} \quad (1)$$

$$Q^-(Y) = \{x \in U \mid [x]_Q \cap Y \neq \emptyset\} \quad (2)$$

在表 1 所表达的系统中,根据不可分辨关系,将  $U$  表示为关于决策属性集  $D$  的等价关系  $F = \{X_1, X_2, \dots, X_m\}$ ,  $X_t \subseteq U$ ,  $X_t \cap X_s = \emptyset$  且  $X_1 \cup X_2 \cup \dots \cup X_m = U$ , 其中  $X_t, X_s (t, s = 1, 2, \dots, m; t \neq s)$  称为等价类。等价关系  $F$  关于  $B (B \subseteq C)$  的上、下近似集,可根据下面的两个式子求得:

$$B^+(F) = \{B^+(X_1), B^+(X_2), \dots, B^+(X_m)\} \quad (3)$$

$$B^-(F) = \{B^-(X_1), B^-(X_2), \dots, B^-(X_m)\} \quad (4)$$

进而,可由下式

$${}_B F = \text{card}[B^-(X_t)] / \text{Ucard}[B^-(X_t)] \quad (5)$$

求得  $F$  的分类精度,其中  $\text{card}()$  表示基数。当  ${}_B F = 1$  时,等价关系能准确的表达这个系统关于检测对象的知识(即无边界)。

知识表的简化分两步,首先是条件属性集  $C$  相对于决策属性集  $D$  简化(简称  $C$  的  $D$  简化)。  $C$  的  $D$  简化可用区分矩阵和区分函数来实现。区分矩阵的关系式如下:

$$Q(B) = \{q(v, w) \mid k \times k\} \quad (6)$$

其中,  $q(v, w) = \{a \in B; f(a_v) = f(a_w) \text{ 且 } f(d_v) \neq f(d_w)\}; d \in D; 1 \leq v < w \leq k; k = \text{card}(U), f()$  为信息函数。利用区分矩阵的关系式构造出区分矩阵,该区分矩阵为一下三角形形式的矩阵。根据构

造出的区分矩阵写出区分函数的表达式如下:

$$f_B(D) = \bigwedge_{(v, w) \in k \times k} q(v, w) \quad (7)$$

然后根据区分函数求出属性简化和核,去掉决策表中冗余的条件属性。一般来说,  $C$  的  $D$  简化有多个,  $C$  的  $D$  核是多个简化的交集(当  $C$  的  $D$  简化仅有一个时,该简化就是核)。

第二步对属性简化表中条件属性值进行简化。方法是:求出条件属性值的相对简化和相对核,这里采用相对区分矩阵和相对区分函数来实现。相对区分矩阵和相对区分函数不同于属性简化中所用到的区分矩阵和区分函数。相对区分矩阵是对称阵的形式,即

$$Q(B) = \{q(v, w) \mid k \times k; q(v, w) = q(w, v)\} \quad (8)$$

其中  $q(v, w) = \{a \in B; f(a_v) = f(a_w) \text{ 且 } f(d_v) \neq f(d_w)\}; d \in D; 1 \leq v, w \leq k; k = \text{card}(U)$ 。相对区分函数有多个(等于论域中的对象数),其求法与属性简化中区分函数的求法相同,不过,它仅仅对应于相对简化矩阵中的某一列,即

$$f_w(D) = \bigwedge_{(v, w) \in k \times 1} q(v, w) \quad (9)$$

其中,  $w$  为相对区分矩阵中列数。

这样就得到了决策表的最小简化。由此可得到知识表达系统的决策规则集,根据规则集就可以对传感器检测到的数据进行决策输出。

## 2 实 例

以油田输油管测漏系统为例说明基于粗糙集理论的多传感器数据融合方法。系统由波峰传感器、压力传感器和流量传感器完成现场数据。波峰传感器和压力传感器的数据波形见图 2,流量传感器通过脉冲计数来完成数据的采集(每  $1 \text{ m}^3$  的容量对应 1000 个脉冲)。经过滤波处理、特征提取等局部处

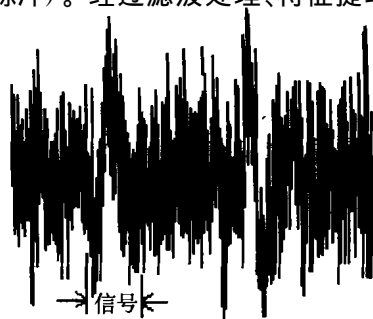


图 2 传感器波形

Fig. 2 Sensor wave

理后得到局部决策。传感器采集到的数据可以分为 8 类,因此,本文取论域元素个数  $k = 8$ 。采用模糊语言来描述基于粗糙集理论的融合中心,得到原始信息系统见表 2。在该检测系统融合中心的知识表达系统中,  $C = \{ a_1, a_2, a_3 \}$  (  $a_1, a_2, a_3$  分别表示波峰、压力、流量),  $D = \{ d \}$  (  $d$  表示最终决策量)。对条件属性和决策属性进行编码 (见表 3),得到知识表达系统的决策表,见表 4。

| 表 2 原始信息系统                          |    |    |    |      |
|-------------------------------------|----|----|----|------|
| Table 2 Original information system |    |    |    |      |
| $U$                                 | 波峰 | 压力 | 流量 | 油管状态 |
| 1                                   | 中  | 小  | 快  | 可能漏  |
| 2                                   | 大  | 小  | 快  | 漏    |
| 3                                   | 大  | 小  | 慢  | 正常   |
| 4                                   | 小  | 大  | 中  | 正常   |
| 5                                   | 小  | 小  | 快  | 可能漏  |
| 6                                   | 中  | 小  | 快  | 可能漏  |
| 7                                   | 中  | 小  | 慢  | 未知   |
| 8                                   | 中  | 中  | 中  | 正常   |

| 表 3 属性编码表              |    |    |     |   |
|------------------------|----|----|-----|---|
| Table 3 Attribute code |    |    |     |   |
| 属性                     | 编码 |    |     |   |
|                        | 0  | 1  | 2   | 3 |
| 波峰 $a_1$               | 小  | 中  | 大   |   |
| 压力 $a_2$               | 小  | 中  | 大   |   |
| 流量 $a_3$               | 慢  | 中  | 快   |   |
| 油管状态 $d$               | 正常 | 未知 | 可能漏 | 漏 |

| 表 4 决策表                |       |       |       |     |
|------------------------|-------|-------|-------|-----|
| Table 4 Decision table |       |       |       |     |
| $U$                    | $a_1$ | $a_2$ | $a_3$ | $d$ |
| 1                      | 1     | 0     | 2     | 2   |
| 2                      | 2     | 0     | 2     | 3   |
| 3                      | 2     | 0     | 0     | 0   |
| 4                      | 0     | 2     | 1     | 0   |
| 5                      | 0     | 0     | 2     | 2   |
| 6                      | 1     | 1     | 2     | 2   |
| 7                      | 1     | 0     | 0     | 1   |
| 8                      | 1     | 1     | 1     | 0   |

将表 4 所示的表达系统的论域  $U$  分别表示为关于决策属性  $D$  的等价关系  $F$  和关于条件属性  $C$

的等价关系  $F: F = \{ X_1, X_2, X_3, X_4 \}; R = \{ \{ 1 \}, \{ 2 \}, \{ 3 \}, \{ 4 \}, \{ 5 \}, \{ 6 \}, \{ 7 \}, \{ 8 \} \}$  其中,  $X_1 = \{ 3, 4, 8 \}$ ,  $x_2 = \{ 7 \}$ ,  $X_3 = \{ 1, 5, 6 \}$ ,  $X_4 = \{ 2 \}$ 。由各个等价类关于  $R$  的上、下近似集,可求得等价关系  $F$  的分类精度  $cF = 1$ 。

根据区分矩阵表达式构造该知识表达系统的区分矩阵如下:其中  $e_1$  表示  $a_1$ ,  $e_{1,3}$  表示  $a_1 a_3$ ,  $e_{1,2,3}$  表示  $a_1 a_2 a_3$ ,其余类同。

|   | 1           | 2           | 3           | 4           | 5           | 6         | 7         | 8 |
|---|-------------|-------------|-------------|-------------|-------------|-----------|-----------|---|
| 1 |             |             |             |             |             |           |           |   |
| 2 | $e_1$       |             |             |             |             |           |           |   |
| 3 | $e_{1,3}$   | $e_3$       |             |             |             |           |           |   |
| 4 | $e_{1,2,3}$ | $e_{1,3}$   |             |             |             |           |           |   |
| 5 |             | $e_1$       | $e_{1,3}$   | $e_{2,3}$   |             |           |           |   |
| 6 |             | $e_{1,2}$   | $e_{1,2,3}$ | $e_{1,2,3}$ |             |           |           |   |
| 7 | $e_3$       | $e_{1,3}$   | $e_1$       | $e_{1,2,3}$ | $e_{1,3}$   | $e_{2,3}$ |           |   |
| 8 | $e_{1,3}$   | $e_{1,2,3}$ |             |             | $e_{1,2,3}$ | $e_3$     | $e_{2,3}$ |   |

由区分函数的表达式有:

$$f_C(D) = a_1 \quad (a_1 \quad a_3) \quad a_3 \quad (a_1 \quad a_2 \quad a_3) \\ (a_1 \quad a_3) \quad a_1 \quad (a_1 \quad a_3) \quad (a_2 \quad a_3) \quad (a_1 \\ a_2) \quad (a_1 \quad a_2 \quad a_3) \quad (a_1 \quad a_2 \quad a_3) \quad a_3 \quad (a_1 \\ a_3) \quad a_1 \quad (a_1 \quad a_2 \quad a_3) \quad (a_1 \quad a_3) \quad (a_2 \quad a_3) \\ (a_1 \quad a_3) \quad (a_1 \quad a_2 \quad a_3) \quad (a_1 \quad a_2 \quad a_3) \quad a_3 \\ (a_2 \quad a_3) = a_1 \quad a_3$$

得到决策表只有一个简化 (  $a_1, a_3$  ),因此,该简化也是条件属性的核。

属性化简后得到的简化知识表见表 5。

| 表 5 简化知识表                   |       |       |     |
|-----------------------------|-------|-------|-----|
| Table 5 Knowledge reduction |       |       |     |
| $U$                         | $a_1$ | $a_3$ | $d$ |
| 1                           | 1     | 2     | 2   |
| 2                           | 2     | 2     | 3   |
| 3                           | 2     | 0     | 0   |
| 4                           | 0     | 1     | 0   |
| 5                           | 0     | 2     | 2   |
| 6                           | 1     | 2     | 2   |
| 7                           | 1     | 0     | 1   |
| 8                           | 1     | 1     | 0   |

根据简化知识表可以构造出相对区分矩阵如下 (表示法与区分矩阵同):

|   | 1         | 2         | 3         | 4         | 5         | 6         | 7         | 8         |
|---|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | $e_1$     | $e_{1,3}$ | $e_{1,3}$ |           |           | $e_3$     | $e_3$     |           |
| 2 | $e_1$     |           | $e_3$     | $e_{1,3}$ | $e_1$     | $e_1$     | $e_{1,3}$ | $e_{1,3}$ |
| 3 | $e_{1,3}$ | $e_3$     |           |           | $e_{1,3}$ | $e_{1,3}$ | $e_1$     |           |
| 4 | $e_{1,3}$ | $e_{1,3}$ |           |           | $e_3$     | $e_{1,3}$ | $e_{1,3}$ |           |
| 5 |           | $e_1$     | $e_{1,3}$ | $e_3$     |           | $e_{1,3}$ | $e_{1,3}$ |           |
| 6 |           | $e_1$     | $e_{1,3}$ | $e_{1,3}$ |           | $e_3$     | $e_3$     |           |
| 7 | $e_3$     | $e_{1,3}$ | $e_1$     | $e_{1,3}$ | $e_{1,3}$ | $e_3$     |           | $e_3$     |
| 8 | $e_3$     | $e_{1,3}$ |           |           | $e_{1,3}$ | $e_3$     | $e_3$     |           |

同理可写出相对区分函数如下：

$$f_1(D) = a_1 \quad (a_1 \quad a_3) \quad (a_1 \quad a_3) \quad a_3 \quad a_3 =$$

$$a_1 \quad a_3$$

$$f_2(D) = a_1 \quad a_3 \quad (a_1 \quad a_3) \quad a_1 \quad a_1 \quad (a_1$$

$$a_3) \quad (a_1 \quad a_3) = a_1 \quad a_3$$

$$f_3(D) = (a_1 \quad a_3) \quad a_3 \quad (a_1 \quad a_3) \quad (a_1$$

$$a_3) \quad a_1 = a_1 \quad a_3$$

$$f_4(D) = (a_1 \quad a_3) \quad (a_1 \quad a_3) \quad a_3 \quad (a_1$$

$$a_3) \quad (a_1 \quad a_3) = a_3$$

$$f_5(D) = a_1 \quad (a_1 \quad a_3) \quad a_3 \quad (a_1 \quad a_3) \quad (a_1$$

$$a_3) = a_1 \quad a_3$$

$$f_6(D) = a_1 \quad (a_1 \quad a_3) \quad (a_1 \quad a_3) \quad a_3 \quad a_3 =$$

$$a_1 \quad a_3$$

$$f_7(D) = a_3 \quad (a_1 \quad a_3) \quad a_1 \quad (a_1 \quad a_3) \quad (a_1$$

$$a_3) \quad a_3 \quad a_3 = a_1 \quad a_3$$

$$f_8(D) = a_3 \quad (a_1 \quad a_3) \quad (a_1 \quad a_3) \quad a_3 \quad a_3 =$$

$$a_3$$

这样,得到知识表的最小简化表见表 6。

表 6 最小简化表

Table 6 Minimum reduction

| $U$ | $a_1$ | $a_3$ | $d$ |
|-----|-------|-------|-----|
| 1   | 1     | 2     | 2   |
| 2   | 2     | 2     | 3   |
| 3   | 2     | 0     | 0   |
| 4   | -     | 1     | 0   |
| 5   | 0     | 2     | 2   |
| 6   | 1     | 2     | 2   |
| 7   | 1     | 0     | 1   |
| 8   | -     | 1     | 0   |

由最小简化表可以得到该知识表达系统的决策规则集如下：

$$(a_1 = 2 \quad a_3 = 0) \quad a_3 = 1 \Rightarrow d = 0$$

$$a_1 = 1 \quad a_3 = 0 \Rightarrow d = 1$$

$$(a_1 = 0 \quad a_3 = 2) \quad (a_1 = 1 \quad a_3 = 2) = (a_1 = 0$$

$$a_1 = 1) \quad a_3 = 2 \Rightarrow d = 2$$

$$a_1 = 2 \quad a_3 = 2 \Rightarrow d = 3$$

在油田输油管测漏系统中,基于粗糙集的多传感器数据融合方法得到了很好的应用,取得了较好的效果,规则正确率可达 98.3 %。

### 3 结束语

粗糙集理论给数据处理领域提供了一种新的处理方法,尤其是对不完备数据。本文主要探讨了相容状况下分布式监测系统数据融合的粗处理方法,从结果来看,采用该方法不需要概率论方法中的先验概率,也不需要模糊推理中的隶属函数,同时还可以去除冗余的局部检测器和局部决策,使系统的配置最优化,提高系统的运行速度,这些都是其它方法不可比拟的。为了使系统更有效、更准确,论域要取的足够大,这样会出现不相容的情况,限于篇幅另文处理。

### 参 考 文 献

- [1] 曾黄麟. 粗糙及理论及其应用[M]. 重庆: 重庆大学出版社, 1996
- [2] 刘红艳, 王万军. 基于粗糙集的数据挖掘[J]. 甘肃教育学院学报, 2002, 16(3): 17 - 20
- [3] Zdzislaw Pawlak. Rough sets and intelligent data analysis[J]. Information Sciences, 2002, 147: 1 - 12
- [4] Francis E H Tay, Shen L X. Economic and financial prediction using rough sets model[J]. European Journal of Operational Research, 2002, 141: 641 - 659
- [5] Jackson A G, Pawlak Z, LeClair S R. Rough sets applied to the discovery of materials knowledge[J]. Journal of Alloys and Compounds, 1998, 279: 14 - 21
- [6] 张兆礼, 孙圣和. 粗神经网络及其在数据融合中的应用[J]. 控制与决策, 2002, 16(1): 76 - 82
- [7] Walczak B, Massart D L. Rough sets theory[J]. Chemometrics and Intelligent Laboratory Systems, 1999, 47: 1 - 16

(下转第 94 页)

- (5): 7 - 8
- [5] 张钟华, 贺 青. 用单片机实现的高精度音频频率计 [J]. 电测与仪表, 1997(11): 3 - 5
- [6] 朱大奇. 一种高精度测频电路设计 [J]. 电子与自动化, 1997(4): 39 - 40

## Wide range frequency measurement technique and its realization based on automatic frequency divider

Zhu Xue-zhi Wang Xue-wei

(College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China)

**Abstract:** A technique for an automatic wide range frequency measurement was introduced. The technique is that the frequency factor of the divider is adjusted automatically according to the signal frequency, to magnify properly the signal period, and realize the frequency measurement on the range of 40 Hz ~ 10 kHz. The principle of measurement circuit and frequency factor's optimum choice were discussed and the realization of software and hardware was proposed. The technique is very precise and practical with low cost.

**Key words:** frequency measurement; automatic frequency divider method; microcontroller

(责任编辑 刘同帅)

(上接第 90 页)

## Data fusion based on rough set theory in distributing measure system

Zhang Su-yan Nie Wei

(College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China)

**Abstract:** The consistent datum was disposed by the rough set theory in the data fusion center of a distributing measure system. The data fusion structure and crisp fusion algorithm of the distributing measure system based on the roughs set theory were proposed. Reduction of attributes was ascertained by an indiscernibility relation, and decision rules were deduced. They were applied to a measure leak system of oil pipe in oil field. The results illustrate that the method is rational and valid.

**Key words:** distributing measure system; data fusion; rough set

(责任编辑 刘同帅)