

基于人工智能技术的不完备信息系统 智能诊断方法研究

周 頔

(四川文理学院 智能制造产业技术研究院, 四川 达州 635000)

摘要: 为了提高不完备信息系统故障诊断的正确性与效率, 提出一种基于粗糙集理论、蚁群优化算法和 RBF 神经网络相结合的故障智能诊断方法; 该方法首先利用“条件组合补齐算法”对不完备的数据进行完备化处理, 再利用粗糙集对条件属性进行知识约简, 得到具有最大完备度的最小规则集, 接着用蚁群算法优化 RBF 神经网络的权值, 并将最小规则集用于训练 RBF 神经网络模型, 获得故障智能诊断模型; 通过实际工程数据验证故障智能诊断模型的有效性, 结果表明提出的方法能有效实现系统故障的诊断。

关键词: 智能诊断; 不完备信息系统; 粗糙集理论; 蚁群算法; 神经网络

Study on an Intelligent Diagnosis Method of Fault in Incomplete Information System Based on Artificial Intelligence Technology

Zhou Di

(Industrial Technological Institute of Intelligent Manufacturing, Sichuan University of Arts and Science, Dazhou 635000, China)

Abstract: In order to improve fault diagnosis correctness and efficiency of incomplete information system, an intelligent diagnosis method of fault based on rough set (RS), ant colony optimization (ACO) algorithm and radial basis function (RBF) neural network is proposed. In this intelligent diagnosis method, the combination and condition supplement algorithm is used to deal with the incomplete data with the maximum completeness. The RS as a new mathematical tool is used to remove redundant information in order to obtain the minimum rule set. Then the ACO algorithm is directly used to optimize the weights of RBF neural network in order to establish an optimized RBF neural network model, then the minimum rule set is inputted the optimized RBF neural network model in order to obtain an intelligent diagnosis model. The actual data are used to verify the effectiveness of intelligent diagnosis model. The experiment results show that the proposed intelligent diagnosis method can effectively diagnose the faults of system.

Keywords: intelligent diagnosis; incomplete information system; rough set theory; ant colony optimization; RBF neural network

0 引言

随着计算机技术、通信技术和网络技术的不断发展, 现代计算机系统朝着复杂化、自动化、智能化等方向发展, 其系统结构逐渐复杂, 性能要求逐渐提高^[1]。然而当系统中的设备出现故障时, 一些智能信息处理技术被用于准确定位系统的故障, 并正确处理出现的故障。智能故障诊断是一个极其复杂的信息处理系统, 主要借助信息采集和传输技术等, 得到被测对象的大量数据信息。对于复杂系统而言, 很多问题都是在信息不完备的条件下, 利用各种方法与技术来做出决策的。所以故障智能诊断系统也应该能够在故障信息不完备的条件下, 具有故障检测、诊断、处理的能力^[2]。但在实际的故障诊断中, 由于系统的复杂性, 存在着故障征兆属性空缺、属性值丢失、属性值无从知道、属性值描述不确定、数据稀疏等不完备信息。所以许多学者对知识获取中数据所

表现出的不完备进行了深入研究。文献[3]提出一种近似测度和删除空缺属性值方法, 实现了不完备数据集的完备化处理; 文献[4]提出了容差关系概念, 给出不完备信息系统粗糙近似, 获取决策规则; 在此基础上, 文献[5]提出了量化容差关系概念来刻画样本对象的相似程度; 文献[6]提出了限制容差关系的概念, 更加符合实际情况; 文献[7]提出了基于联系度的拓展粗糙集模型; 文献[8]进一步提出了具有加权联系度的拓展粗糙集模型; 文献[9]研究了不完备信息系统的基于集对分析的粗糙集模型; 文献[10]提出了基于全相容粒度的粗糙集模型; 文献[11]提出了不完备信息系统相似类概念, 用于挖掘关联规则; 文献[12]利用确定属性值来替换具有相同的决策属性的值。然而这些处理方法也可能会使原始数据和经过处理获得的知识存在不同程度的失真等。针对这些情况, 本文首先对原始数据进行完备化处理, 再融合粗糙集、蚁群算法和径向基神经网络等人工智能技术的优势, 提出一种不完备信息系统的故障智能诊断模型与方法。

1 不完备信息系统及其完备化处理方法

虽然复杂信息系统出现故障的种类多种多样, 本文从复杂系统论来看, 这些多种多样的故障在本质上具有不少的共性。因此通过分析归纳总结复杂系统的故障, 具有以下特

收稿日期:2018-02-05; 修回日期:2018-03-22。

基金项目:自然科学基金(2014JY0111);四川省教育厅科技计划项目(18ZA0415)。

作者简介:周 頔(1981-),重庆人,副研究员,硕士,主要从事智能计算、模式识别等方向的研究。

性：(1) 故障的复杂性；(2) 故障的层次性；(3) 故障的相关性；(4) 故障的延时性；(5) 故障的不确定性；(6) 故障的规律性。

对复杂信息系统故障诊断而言，空值存在的影响主要有：(1) 丢失大量有用信息；(2) 增加系统不确定性；(3) 空值导致不可靠的输出。

为了提高故障诊断的正确性与效率，所以需要不完备信息进行完备化处理。本文采用“条件组合补齐算法”将缺失的故障信息进行合理的补充。该方法是通过选取具有相同决策的实例相应属性所有取值试探，选择最好的一种情况，作为补齐的结果。

2 粗糙集理论和 RBF 神经网络

2.1 粗糙集理论

粗糙集理论是 1982 年提出的一种能够有效处理不精确、不确定与不完备等信息的新数学工具。它能发现隐含知识，揭示潜在的规律的特点。经过 30 多年的发展，已在数据挖掘、模式识别、机器学习、智能控制、故障诊断等领域获得广泛的应用粗糙集理论的基本概念描述如下^[13-14]。

定义 1：一信息系统 $S = \langle U, R, V, F \rangle$ ，其中 U 是对象集合， $C \cup D = R$ 是属性集合，子集 C 和 D 分别称为条件属性和决策属性。 $V = \cup_{r \in R} V_r$ 是属性的集合， V_r 表示了属性 $r \in R$ 的属性值范围， $F: U \times R \rightarrow V$ 是一个信息函数，它指定 U 中每一对象 x 的属性值。

定义 2：一信息系统 $S = \langle U, R, V, F \rangle$ ，对于每个属性子集 $B \subseteq R$ ，定义一个不可分辨关系 $\text{Ind}(B)$ 为： $\text{Ind}(B) = \{ (x, y) \in U^2 \}$ 。

定义 3：一信息系统 $S = \langle U, R, V, F \rangle$ ， $B \subseteq R$ ， $X \subseteq U$ ，定义集合 $R_-(X) = \{ x \in U; [x]_{\text{IND}(B)} \subseteq X \}$ 为 X 的 B 下近似集， $R^+(X) = \{ x \in U; [x]_{\text{IND}(B)} \cap X \neq \emptyset \}$ 为 X 的 B 上近似集。

定义 4：一信息系统 $S = \langle U, R, V, F \rangle$ ， $A = CUD$ 且 $B \subset C$ ，定义 D 的 B 正域 $\text{Pos}_B(D) = \cup \{ B(X); X \in \text{IND}(D) \}$ 。

定义 5：一信息系统 $S = \langle U, R, V, F \rangle$ ， $A = CUD$ 且 $B \subset C$ ， B 和 D 间的依赖度定义为： $K(B, D) = \text{card}(\text{Pos}_B(D)) / \text{card}(U)$ 。

2.2 神经网络

人工神经网络 (ANN) 是从信息处理角度对人脑神经网络进行抽象，建立某种简单模型，按不同的连接方式组成不同的网络^[15]。本文采用具有数逼近能力、分类能力和学习速度等优点的 RBF 神经网络，其网络结构如图 1 所示。

从图 1 可知，RBF 神经网络从输入到输出的映射为：

$$f(x) = \sum_{i=1}^m \omega_i \varphi(x, c_i) \quad (1)$$

其中： $x = (x_1, x_2, \dots, x_n)^T \in R^n$ 为输入矢量， $W = (\omega_1, \omega_2, \dots, \omega_m)^T \in R^m$ 是权值， φ 为径向基函数， f 为输出， c_i 是 RBF 的第 i 个聚类中心。RBF 神经元的传递函数有多种形式，但最常用的是高斯函数，即：

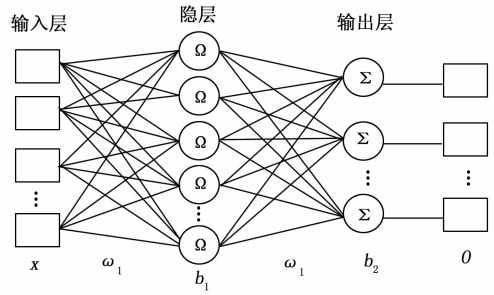


图 1 RBF 神经网络结构

$$\varphi(x, c_i) = \exp\left(-\frac{\|x - c_i\|^2}{\sigma_i^2}\right) \quad (2)$$

影响 RBF 神经网络性能的参数主要是隐含层节点数及相应中心节点位置和宽度。因此需要进一步研究 RBF 神经网络参数优化问题。

3 基于蚁群算法优化神经网络参数

蚁群优化 (ant colony optimization, ACO) 算法是一种典型的群智能算法^[16]。ACO 算法通过信息素的调整较好地控制了种群的多样性，具有较强的全局优化能力。因此充分利用蚁群优化算法的优点，对神经网络参数和学习规则进行优化处理，为提高 RBF 神经网络性能提供了一种新的途径。由于蚁群算法具有全局搜索能力，所以用蚁群算法优化网络隐含层，能提高 RBF 神经网络的性能。

基于蚁群算法优化神经网络参数和学习规则的步骤如下：

- 1) 蚁群优化算法的参数初始化。
- 2) 构造适应度函数：取为总失真的倒数：

$$f = \left[\sum_{k=1}^K \sum_{x_i \in \Omega} \|x_i - c_j\|^2 \right]^{-1} \quad (3)$$

- 3) 蚁群优化算法寻优过程。

4) 径向基函数中心和宽度：将蚁群优化算法所得的每一类都作为隐含层节点，分别求取径向基函数的中心 (c_j) 和宽度 (σ_j^2)：

$$c_j = \frac{1}{n_j} \sum_{x_i \in \Omega} X_i \quad (4)$$

$$\sigma_j^2 = \frac{1}{n_j} \sum_{x_i \in \Omega} \|x_i - c_j\|^2 \quad (5)$$

5) 隐含层到输出层权值的调整：用梯度下降法训练网络隐含层与输出层之间的权值。

4 基于粗糙集、蚁群优化算法和神经网络的故障智能诊断方法

本文将粗糙集理论、蚁群优化算法和神经网络引入到故障诊断中，充分利用粗糙集理论知识约简能力，蚁群优化算法的全局优化能力和 RBF 神经网络的自学习与知识获取的能力，提出一种基于粗糙集理论、蚁群优化算法和 RBF 神经网络相结合的智能故障诊断方法，其故障智能诊断模型，如图 3 所示。

基于粗糙集、蚁群优化算法和神经网络的故障智能诊断

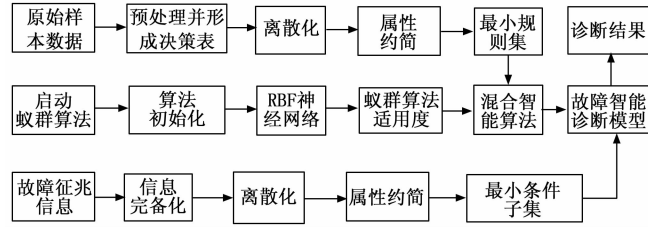


图 3 基于粗糙集、蚁群算法和神经网络的故障智能诊断模型

模型的具体步骤如下。

Step 1: 输入原始样本数据，即从关系数据库中随机读取历史故障样本信息。

Step 2: 通过数据预处理，建立系统关系数据模型，形成原始故障信息的二维决策表。

Step 3: 连续属性离散化：是在特定连续属性的值域范围内设定若干个离散化的划分点，获得若干个离散化区间，使用不同的符号来表示子区间的属性值。从本质看，就是对初始断点的选择与合并过程。

Step 4: 属性约简：采用粗糙集理论对属性进行知识约简，以获得能够覆盖原数据特征的最小条件属性规则集。其属性的重要程度可以用公式 (6) 和 (7) 来计算^[17]：

$$r_c(D) = \text{card}(\text{pos}_c(D)) / \text{card}(U) \quad (6)$$

$$r_{c-a}(D) = \text{card}(\text{pos}_{c-a}(D)) / \text{card}(U) \quad (7)$$

Step 5: 启动蚁群算法，用于优化 RBF 神经网络；

Step 6: 把粗糙集理论处理得到的具有最大完备度的最小规则集输入优化得 RBF 神经网络，训练得到 RARNN 模型；

Step 7: 根据 RARNN 模型的诊断误差，判断是否满足要求。如果满足，转至 Step 8，否则转至 Step 4；

Step 8: 将待诊故障信息输入 RARNN 模型，获得故障诊断的结果。

5 故障诊断模型的应用

5.1 数据来源

以电机子系统为例进行振动故障诊断^[18]，由于系统的复杂性，导致采集的数据有些存在一些缺失。 U 是论域， $a \sim e$ 是条件属性，分别表示信号频域特征频谱 $< 0.5 f, 0.5 f \sim 0.6 f, 1 f, 2 f, \geq 3 f$ 的幅值分量能量。 h 是决策属性，表示发电机不同的故障， q 分别对应油膜振荡故障、不平衡故障和不对中故障。本文选取实验部分数据进行研究。对采集的信号进行归一化处理，构成不完备故障诊断决策表，如表 1 所示。

5.2 数据完备化处理

首先采用“条件组合补齐算法”将缺失的数据进行合理有效的完备化处理，获得完备化的故障诊断决策表，如表 2 所示。

5.3 数据离散化

离散化数据的连续属性，是为了使用不同的符号来表示子区间的属性值。这里采用通过模拟自然进化过程搜索

表 1 不完备故障诊断决策表

U	a	b	c	d	e	q
1	0.051	0.793	0.227	—	0.012	1
2	0.242	0.971	0.316	0.057	—	1
3	0.164	—	0.288	0.023	0.015	1
4	0.031	0.063	—	—	0.059	2
5	0.046	0.024	0.983	0.313	—	2
6	0.012	0.052	0.877	0.185	0.068	2
7	0.035	0.039	0.383	0.534	0.233	3
8	—	0.021	—	0.459	0.112	3
9	0.013	—	0.425	0.495	0.179	3

表 2 完备化的故障诊断决策表

U	a	b	c	d	e	q
1	0.051	0.793	0.227	0.036	0.012	1
2	0.242	0.971	0.316	0.057	0.014	1
3	0.164	0.583	0.288	0.023	0.015	1
4	0.031	0.063	0.727	0.429	0.059	2
5	0.046	0.024	0.983	0.313	0.164	2
6	0.012	0.052	0.877	0.185	0.068	2
7	0.035	0.039	0.383	0.534	0.233	3
8	0.158	0.021	0.376	0.459	0.112	3
9	0.013	0.643	0.425	0.495	0.179	3

最优解的遗传算法，在保持数据不可分辨关系的前提下，以最小断点集为目标，实现数据连续属性的离散化处理，获得离散化的故障诊断决策表，如表 3 所示。

表 3 离散化的故障诊断决策表

U	a	b	c	d	e	q
1	1	2	1	1	1	1
2	3	2	2	1	1	1
3	2	2	1	1	1	1
4	1	1	3	3	1	2
5	1	1	3	2	2	2
6	1	1	3	1	3	2
7	1	1	2	3	3	3
8	2	1	2	3	2	3
9	1	2	3	3	2	3

5.4 属性约简

采用具有实现最大完备度的最小规则集的粗糙集理论，对离散化的故障诊断决策表。通过对原始样本决策表属性的计算，得到属性集 $\{a, c, d, e\}$ 对于原始样本决策表来说是重要的，并且可以证明 $\{a, c, d, e\}$ 是该决策表的最小约简。说明 $\{a, c, d, e\}$ 保持该决策表 5 个属性的分类能力，因此， $\{a, c, d, e\}$ 是该决策系统必不可少的 4 个属性。整理属性约简之后的决策表，合并相同规则，删除冗余的规则，获得能够覆盖原数据特征的最小条件属性规则集。

5.5 RBF 神经网络的优化

采用具有全局寻优能力的蚁群优化算法，来优化 RBF 神经网络的基函数的中心、方差和隐含层到输出层的权值 3 个参数，获得 3 个参数组合的最优参数值，作为 RBF 神经

网络的初始参数值；然后采用属性约简获得能够覆盖原数据特征的最小条件属性规则集训练 RBF 神经网络，以获得优化的 RBF 神经网络模型（故障诊断模型 RARNN）。

5.6 故障诊断验证

把采集到的待诊信息 $\{(0.024, *, 0.748, 0.122, 0.023), (0.015, 0.055, 0.483, *, 0.311)\}$ 经过完备化和属性约简处理后，输入 RARNN 模型进行故障诊断，得到故障诊断结果：第一组数据是不平衡故障，第二组数据是不对中故障。

通过电机子系统故障诊断实例的分析可知，基于将粗糙集理论、蚁群优化算法和 RBF 神经网络相结合的故障智能诊断方法（RARNN）能够正确诊断电机子系统的故障，并且该方法能快速收敛，具有较高的故障诊断正确率。

6 结语

本文利用“条件组合补充算法”进行了信息的不完备化处理，同时根据复杂系统故障诊断困难的特点，提出了基于将粗糙集理论、蚁群优化算法和 RBF 神经网络相结合的故障智能诊断方法。该方法通过复杂系统故障诊断实例，给出了该方法的应用步骤。通过仿真实验，结果表明故障智能诊断方法能够加速收敛，具有较高的故障诊断正确率。为其它设备故障诊断提供了一种新的思路。

参考文献：

[1] 李 聪, 梁昌勇, 杨善林. 基于粗糙集的不完备信息系统空值估算方法 [J]. 计算机集成制造系统, 2009, 15 (3): 604-608.
 [2] 蒙祖强, 史忠植. 不完备信息系统中基于相容粒度计算的知识获取方法 [J]. 计算机研究与发展, 2008, 45 (S1): 264-267.
 [3] Peterson N W, et al. The rule induction system lers—aversion for personal computers [J]. Foundations of Computing and Decision Sciences, 1993, 18 (3-4): 181-212.
 [4] Kryszkiewicz M. Properties of incomplete information systems in the framework of rough sets [A]. In: Polkowski L., Skowron A. (Eds.), Rough Sets in Knowledge Discovery. Physica [C]. Heidelberg, 1998 (1): 422-450.

(上接第 4 页)

参考文献：

[1] Duyar A, Eldem V, Merrill W, et al. A Simplified Dynamic Model of the Space Shuttle Main Engine [R]. NASA—TM—104421, 1991.
 [2] Norman A M. Rocketdyne Safety Algorithm: Space Shuttle Main Engine Fault Detection [R]. NASACR—195356, 1994.
 [3] 刘 昆. 分级燃烧循环液氧/液氢发动机系统分布参数模型与通用仿真研究 [D]. 长沙: 国防科学技术大学, 1999.
 [4] 程谋森. 液氢液氧发动机预冷与起动过程模型及 PVM 仿真研究 [D]. 长沙: 国防科学技术大学, 2000.
 [5] Williams B C, Nayak P P. A Model—based Approach to Reactive Self—Configuring Systems [C]. Proceedings of the National Conference on Artificial Intelligence, American Association for Artificial Intelligence, Menlo Park, CA, 1996.
 [6] Kurien J, Nayak P P. Back to the Future for Consistency—based Trajectory Tracking [C]. Proceedings of the National

[5] Stefanowski J., Tsoukias. A. Valued Tolerance and Decision Rules [A]. In W. Ziarko, Y. Yao (eds.), Rough Sets and Current Trends in Computing [C]. Springer Verlag, Berlin, 2001: 212-219.
 [6] 王国胤. Rough 集理论在不完备信息系统中的扩充 [J]. 计算机研究与发展, 2002, 39 (10): 1238-1243.
 [7] 黄 兵, 周献中. 不完备信息系统中基于联系度的粗集模型拓展 [J]. 系统工程理论与实践, 2004, 24 (1): 88-92.
 [8] 赵 翔, 刘同明, 向一丹. 不完备信息系统中基于加权联系度的粗糙集模型 [J]. 计算机应用, 2005, 25 (4): 824-826.
 [9] 邓毅雄, 黄兆华不完备信息系统的基于集对分析的粗糙集模型 [J]. 华东交通大学学报, 2005, 22 (2): 100-103.
 [10] 吴 陈, 杨习贝, 傅 凡. 基于全相容粒度的粗糙集模型 [J]. 系统工程学报, 2006, 21 (3): 292-298.
 [11] Leung Y, Wu Weizhi, Zhang Wenxiu. Knowledge acquisition in incomplete information systems: A rough set approach [J]. European Journal of Operational Research, 2006, 168 (1): 164-180.
 [12] Li Jingrong, Pheng K L, Beng T S. Rmine: A rough set based data mining prototype for the reasoning of incomplete data in condition—based fault diagnosis [J]. Journal of Intelligent Manufacturing, 2006, 17 (1): 163-176.
 [13] 张 铮. 不完备不协调信息条件下的设备智能故障诊断 [D]. 武汉: 华中科技大学, 2007.
 [14] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Science, 1982, 11 (5): 341-356.
 [15] 毛 健, 赵红东, 姚婧婧. 神经网络的发展及应用 [J]. 电子设计工程, 2011, 19 (24): 62-65.
 [16] Dorigo M, Gambardella LM. Ant colony system: a cooperative learning approach to the traveling salesman problem [J]. IEEE Transactions on Evolutionary Computation, 1997, 1 (1): 53-66.
 [17] 庄静芸, 徐中伟, 喻 钢. 一种粗糙集属性约简算法 [J]. 计算机工程, 2009, 35 (15): 67-69.
 [18] 李 化. 汽轮发电机组振动故障智能诊断模型的理论及方法研究 [D]. 重庆: 重庆大学, 1999.
 Conference on Artificial Intelligence, American Association for Artificial Intelligence, Menlo Park, CA, 2000.
 [7] A Kurien J. Diagnosis and Planning With Resource Constraints [D]. Providence, United States: Brown University, Ph. D. dissertation, 2003.
 [8] Narasimhan S and Brownston L. HyDE—A General Framework for Stochastic and Hybrid Model—based Diagnosis [R]. 18th International Workshop on Principles of Diagnosis (DX—07), 2007.
 [9] 黄敏超 液体火箭发动机故障的神经网络诊断研究 [D]. 长沙: 国防科技大学, 1998.
 [10] 刘 冰. 液体火箭发动机故障诊断中的特征提取 [J] 推进技术. 1997, 18 (6): 1-4.
 [11] 李艳军. 新一代大推力液体火箭发动机故障诊断检测与诊断关键技术研究 [D]. 长沙: 国防科技大学, 2014.
 [12] 黄 强. 高压补燃液氧煤油发动机故障检测与诊断技术研究 [D]. 长沙: 国防科技大学, 2012.