

基于改进蚁群算法的变压器诊断数据的约简

田冰冰¹, 刘念¹, 刘琨¹, 姜刚²

(1. 四川大学电气信息学院, 四川 成都 610065; 2. 孝感电业局, 湖北 孝感 432100)

摘要: 针对变压器故障诊断提出的各种算法, 多数不能从激增的诊断数据中提取有效的信息, 无法实现快速准确诊断。为提高变压器故障诊断的速度, 基于蚁群算法基本原理和模糊信息熵理论, 对蚁群算法从局部搜索策略、蚂蚁内部状态、信息素更新策略及状态转移规则四方面进行了改进, 提出诊断数据约简新方法。经过变压器故障诊断实例验证, 使用改进方法将数据约简后, 具有较高的诊断正确率, 而与传统的算法相比, 诊断速度快。

关键词: 蚁群算法; 数据约简; 模糊粗糙集; 信息熵; 变压器

The transformer diagnosis data reduction based on improved ant colony algorithm

TIAN Bing-bing¹, LIU Nian¹, LIU Kun¹, JIANG Gang²

(1. School of Electronic Information, Sichuan University, Chengdu 610065, China; 2. Xiaogan Electric Power Bureau, Xiaogan 432100, China)

Abstract: Most algorithms for transformers fault diagnosis can't extract effective information from increasing data, thus it can't diagnose fault fast and accurately. In order to improve the speed of fault diagnosis on the basis of the principle of ant colony algorithm and the information entropy theory of fuzzy rough set, the local search strategy, ant's internal state, pheromone updating and state transition rules of the ant colony model are modified. An improved ant colony algorithm for the reduction of the diagnosis data is proposed. The transformer fault diagnosis experiment indicates that the proposed algorithm has higher diagnosis accuracy rate in data reduction and has fast diagnosis speed compared with traditional algorithm.

Key words: ant colony algorithm; data reduction; fuzzy rough set; information entropy; transformer

中图分类号: TM77 文献标识码: A 文章编号: 1674-3415(2011)01-0096-04

0 引言

电力变压器作为电力系统中最重要和最贵重的设备, 其运行的安全性与电网供电可靠性有直接关系。所以, 电力变压器的故障诊断技术一直都是研究的焦点。目前, 变压器诊断应用较多的是油中溶解气体分析法(DGA), 国内外专家先是将罗杰斯比值法, IEC推荐的三比值法^[1], Dornierbur法等算法应用于其中, 取得一定成果。但无论哪种算法, 共同特征是都需要编码, 这就导致存在“编码盲区”问题。为了弥补这个问题, 一些学者在这些方法的基础上, 引入专家系统^[2]、人工神经网络^[3-4]、模糊数学^[5]、进化遗传算法^[6]、支持向量机^[7]及各种混合诊断系统^[8-9]等方法, 取得了一定进展。

但是电力行业始终处于不断发展之中, 电力设备的维修方式也随之有了大发展, 用于诊断的数据是包括离线试验数据和大量连续监测的在线数据在

内的数据信息^[10], 这些激增的数据背后隐藏着许多重要信息。如不能发现数据中存在的关系和规则, 无法根据现有的数据得知电力设备的故障状态并做出诊断决策。基于蚁群算法的变压器故障诊断系统正是在这一背景下提出的。针对蚁群算法也存在搜索时间长、易陷于局部最优解、收敛速度慢等不足, 国内外学者提出了许多改进的蚁群算法。本文利用蚁群算法的优点并对缺点进行改进, 对诊断数据进行约简, 然后用于变压器的故障诊断, 结果表明这种方法是有用的。

1 蚁群算法简介

蚁群算法(ACO)是一种新型的模拟自然界蚁群行为的进化算法, 由意大利学者 M.Dorigo 等人^[11]于 20 世纪 90 年代初最先提出。ACO 算法是一种模拟蚂蚁群体在寻找食物时能找到距食物最短路径的仿生随机搜索算法。蚁群中每个蚂蚁个体间不直接

通信, 而是通过在其经过的路径上留下一一种叫“信息素”的分泌物来引导其他个体的行为, 某条路径上经过的蚂蚁越多, 留下的信息素也越多, 后来的蚂蚁被吸引到该条路径上的概率也就越大。这样, 通过先行者的先验知识传递给后来者, 将最终导致最优路径选择成功。该算法采用了分布式并行计算机制, 易于与其他优化算法相结合, 而且具有鲁棒性较强、正反馈等许多优点, 具有无中心控制和分布式个体之间间接通信的特征, 所以这种方法一经推出, 便得到了热门研究。随着研究的深入, 蚁群应用范围扩大到更多的组合优化问题。目前, 蚁群算法被证明为是一种有效的启发式优化算法, 可以很好地应用于数据挖掘中。

但正如前述, 蚁群算法也是不完善的, 需要改进。目前, 算法的改进主要是从局部搜索策略、蚂蚁内部状态、信息素更新策略及选择策略四个方面进行, 如自适应蚁群算法、基于信息素扩散的蚁群算法、多态蚁群算法、基于信息熵的改进蚁群算法、基于网格划分策略的连续域蚁群算法、基于交叉变异操作的连续域蚁群算法等。

2 蚁群算法改进

现实世界的的数据一般是含噪声的、冗余的、不完整的和不一致的, 数据预处理是知识发现过程的重要步骤。我们对变压器数据的各种数据进行处理, 能够提高决策的质量。特征选择是数据预处理方法的一种。依据参考文献[12]中采用的方法取得的有效结果, 本文采用基于模糊粗糙集的信息熵理论对蚁群特征算法模型进行改进。

2.1 模糊粗糙集的信息熵理论

1999年 Wierman 提出了测量粗糙集不确定性的信息熵定义。

定义1 设 U 为有限论域, R 是 U 上的等价关系, 关于 R 的等价划分为 $\{X_1, X_2, \dots, X_k\}$, 则近似空间 (U, R) 的信息熵定义为^[13]:

$$E(R) = - \sum_{i=1}^k \frac{|X_i|}{|U|} \ln \frac{|X_i|}{|U|} \quad (1)$$

该定义反映了粗糙分类的信息量, 但却只关注了类别内的信息, 而没有考虑类别外的信息如何度量, 因此该定义并不完整。经过文献[14-15]的进一步的研究和拓展, 得出以下定义和定理。

定义2 设 $U = \{X_1, X_2, \dots, X_n\}$ 为有限论域, R 是 U 上的任意模糊关系, 则近似空间 (U, R) 的信息熵定义为:

$$H(R) = \sum_{i=1}^n \frac{1}{|U|} \left(1 - \frac{|[x_i]_R|}{|U|} \right) \quad (2)$$

其中: $[x_i]_R = \sum_{j=1}^n r_{ij}$ 表示对象 x_i 在模糊关系 R 下的势, r_{ij} 表示两个对象的相似度。

定理1 设 R 是 U 上的自反模糊关系, 则:

$$1) H(R) \text{ 的最大值为 } 1 - \frac{1}{|U|}, \text{ 且 } H(R) = 1 - \frac{1}{|U|} \Leftrightarrow [x_i]_R = \{x_i\}, \forall i \in n;$$

$$2) H(R) \text{ 的最小值为 } 0, \text{ 且 } H(R) = 0 \Leftrightarrow [x_i]_R = U, \forall i \in n.$$

定理2 设 R 是 U 上的等价关系, 则 $H(R) = E(R)$ 。

定义3 设 U 为有限论域, P 和 Q 是 U 上的任意模糊关系, $[x_i]_P$ 和 $[x_i]_Q$ 是 P 和 Q 生成的包含 x_i 的模糊等价类, 则 P 和 Q 的条件熵定义为:

$$H(Q|P) = \sum_{i=1}^n \frac{1}{|U|} \frac{|[x_i]_P - [x_i]_Q|}{|U|} \quad (3)$$

定义4 设 U 为有限论域, P 和 Q 是 U 上的任意模糊关系, $[x_i]_P$ 和 $[x_i]_Q$ 是 P 和 Q 生成的包含 x_i 的模糊等价类, 则 P 和 Q 的互信息定义为:

$$I(Q:P) = \sum_{i=1}^n \frac{1}{|U|} \left(1 - \frac{|[x_i]_P \cup [x_i]_Q|}{|U|} \right) \quad (4)$$

定理3 给定模糊信息系统 $\langle U, A, V, f \rangle$, $A = C \cup D$, B 是 C 的子集。如果 $H(D|B) = H(D|B-a)$, $\forall a \in B$, 则 a 在 B 中相对 D 是多余的; 如果 $H(D|B-a) > H(D|B)$, 则 B 是独立的, 称 B 是 C 的一个 D 相对约简, 如果 B 满足以下两条:

- 1) $H(D|B) = H(D|C)$
- 2) $\forall a \in B, H(D|B-a) > H(D|B)$

基于该定理, 随着约简的逐步深入, 子集的条件熵逐步减小, 则最优子集的条件熵应达到最小值, 故后面设计算法运行到条件熵不再减小时搜索停止。

2.2 基于蚁群模型的改进算法

以前述信息熵理论为基础对蚁群系统模型进行改进, 提出一种新的特征选择算法, 主要分为初始化设置阶段、特征子集搜索阶段和子集性能评价阶段。具体步骤如图1所示。

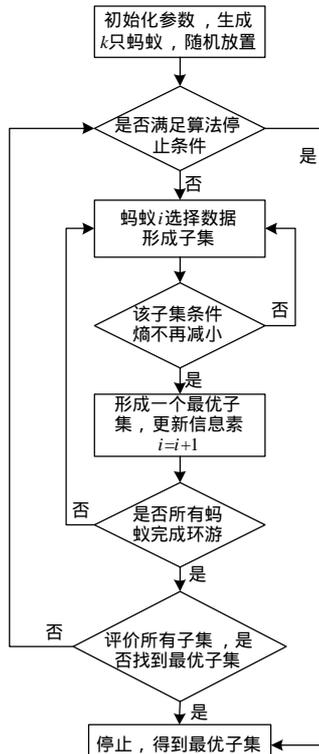


图 1 算法流程图

Fig.1 Flow chart of algorithm

2.2.2.1 启发函数的改进

蚂蚁在搜索过程中总是希望最先遍历最能满足评价函数的节点,尤其在初始时,所有节点的信息素都是一致的,更需要启发函数对蚂蚁进行引导。由于特征和类之间的关系能较好地反映特征的性能,本文将每个特征节点的启发函数定义为如下表示的该特征与决策类的互信息。

$$h_s = I(s:d) \quad (5)$$

其中: s 代表当前条件特征; d 表示分类特征。

2.2.2.2 状态转移规则

为了实现蚂蚁从当前节点到下一个节点之间的运动,需要计算所有候选节点的转移概率,本文基于信息素和启发函数定义蚂蚁的状态转移规则为:

$$s = \begin{cases} \operatorname{argmax}\{t_u(t)[h_u]^a\}, \\ q \leq q_0 (\text{Exploitation}) \\ p_s(t), \\ \text{其他 } s \in j_k(r) (\text{Exploration}) \end{cases} \quad (6)$$

$$P_s(t) = \frac{b \cdot t_s(t) + (1-b) \cdot h_s}{\sum_{u \in j_k(r)} [b \cdot t_u(t) + (1-b) \cdot h_u]} \quad (7)$$

其中: 参数 q_0 为 Exploitation 概率因子, 确定规则两部分的相对重要性; a 、 β 分别表示不同转移条件

下信息素和启发信息的相对重要性; $j_k(r)$ 指当前搜索所有可选择的特征集合。规则的 Exploitation 部分, 蚂蚁选择能使信息素和启发信息的乘积最大的特征; 而 Exploration 部分, 蚂蚁则根据每个特征的转移概率来确定, 选择具有最大转移概率的特征, 以此保证蚂蚁在利用以往信息的同时也具有开拓新搜索空间的能力, 防止过快收敛到局部最优。

2.2.3 信息素更新规则

留在路径上的信息素对蚂蚁选择特征十分重要,且随着蚂蚁的经过,信息素的量也在不断变化。从初始信息素开始,一方面信息素随着时间的推移不断挥发,另一方面,在蚂蚁经过的路径上,信息素按某种更新规则不断增加,增加的信息素多少根据该路径的质量而定,也即根据选择的特征子集性能而定。本文以式(8)(9)作为信息素更新公式,由更新路径信息素改为更新节点信息素,且在每次蚂蚁搜索结束后,以本次搜索获得的特征子集的条件熵的倒数作为本次信息素的更新量:

$$t_s(t+1) = (1-r(t)) \cdot t_s(t) + \Delta t_s(t) \quad (8)$$

$$\Delta t_s(t) = \frac{1}{H(d/S)} \quad (9)$$

其中: 信息素挥发系数随着当前路径信息素的多少而变化,定义为 $r(t) = 2 \arctan(t_s(t)) / \rho$, 当信息素多时,挥发系数较大;反之,则相对较小。另外,为防止搜索过快停滞或收敛速度较慢,将所有节点信息素的总量限定在一定范围内。

$$t_{\max} = \frac{1}{1-r(t)} \cdot \frac{1}{H(d/F)} \quad (10)$$

$$t_{\min} = t_{\max} / (2N) \quad (11)$$

其中: $H(d/F)$ 为所有特征相对于决策类的条件熵; N 为特征总数。

3 实例验证及结果分析

变压器故障一般以单一类型故障为主,多故障也时有发生。常见的单一故障类型有:低温过热 ($t < 300$)、中温过热 ($300 < t < 500$)、高温过热 ($t > 500$)、局部放电、低能放电、高能放电等六种潜伏性故障类型。多故障发生的情况主要有高能放电兼过热、低能放电兼过热等。

为了测试诊断模型的确诊情况,文章以额定电压为 220 kV 的变压器为研究对象,收集 2007 年 1 月到 2009 年 1 月湖北省某变电站 500 条未经剔除、筛选的变压器油中气体浓度数据,数据集中包含混合或连续数据,去掉测试结果明显有误的数据,并对冲突数据进行处理之后,得到 450 条可用数据,

其中 312 条变压器正常运行的数据, 138 条有明确故障的数据。按照上述的数据结构建立数据库, 作为研究对象, 可以同时验证算法处理不同数据的能力。

如图 2 所示, 当概率因子取 0.7 或 0.8 时, 算法取得子集分类误差最小, 但取 0.8 时相对更少, 故概率因子 q_0 取 0.8, 设置 $\alpha=2$, $\beta=0.6$, $\rho=0.5$ 时, 蚂蚁数与特征数相同, 由此取得约简后的变压器诊断数据。

当用这些数据进行变压器诊断时, 得到的诊断结果如表 1 所示。

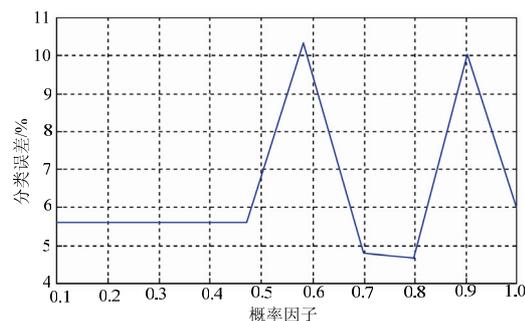


图 2 概率因子与分类误差的关系

Fig.2 Relationship between probability factor and classification errors

表 1 优化算法诊断结果

Tab. 1 Diagnostic results using optimization algorithm

H ₂ /(μ L/L)	CH ₄ /(μ L/L)	C ₂ H ₆ /(μ L/L)	C ₂ H ₄ /(μ L/L)	C ₂ H ₂ /(μ L/L)	诊断结果	实测结果
238	74	149	89	208	电弧放电伴有 过热	外部线圈烧毁
100	94	304	22	118	油中电弧放电	金属异物坠落引起裸金属 过热
48	230	810	160	7	裸金属过热	套管未接地
131	202	373	192	105	电弧放电	分接开关烧毁
980	73	12	58	0	裸金属过热	套管未接地
260	13	84	29	92	油中电弧放电	电弧放电
60	40	110	9.9	70	油中电弧放电	分接开关电弧放电
72	116	70	36	19	裸金属过热	引线绝缘损坏
50	90	260	18	5.9	裸金属过热	铁芯多点接地

从表 1 中可以看出, 使用约简后的数据对变压器进行诊断时, 也得到了较准确的结果, 说明这种诊断方法是有效的。另外, 因为使用的是约简后的数据, 所以运算速度也有提高。

4 结论

本文利用蚁群算法 (ACO) 在解决离散问题时优良的启发式学习和全局寻优能力, 并提出改进方法, 进一步将该改进方法应用于变压器诊断数据的约简, 通过变压器诊断实例验证了这种方法的有效性, 并提高了运算速度。蚁群算法虽然具有很多的优点, 但出现的时间毕竟不长, 对其研究还没有像其他启发式算法那样形成系统的理论, 有很多问题还需进一步研究。

参考文献

[1] IEC Publication 599. Interpretation for the analysis of gases in transformers and other oil-filled electrical equipment in service[S]. 1978.
[2] 杨启平, 薛五德. 专家系统在变压器故障诊断中的应

用[J]. 变压器, 1996, 33 (6): 35-38.
YANG Qi-ping, XUE Wu-de. The application of expert system in power transformer fault diagnosis[J]. Transformer, 1996, 33 (6): 35-38.
[3] Zhang Y, Liu Y. An artificial neural network approach to transformer fault diagnosis[J]. IEEE Trans on Power Delivery, 1996, 11 (4): 1836-1841.
[4] 顾民, 葛良全. 基于自组织神经网络的变压器故障诊断[J]. 继电器, 2007, 35(23): 28-30.
GU Min, GE Liang-quan. Fault diagnosis of power transformers based on the neural network of self-organizing map[J]. Relay, 2007, 35(23): 28-30.
[5] Gen Wei, Sun Cai-xin. Study on fault diagnoses methods of transformer DGA with fuzzy model hierarchy classification[J]. ICMEP, 2000: 167-171.
[6] 谢可夫, 罗安. 遗传算法在变压器故障诊断中的应用[J]. 电力自动化设备, 2005, 25 (4): 55-58.
XIE Ke-fu, LUO An. Application of genetic algorithm in transformer fault diagnosis[J]. Electric Power Automation Equipment, 2005, 25 (4): 55-58.
[7] 吴晓辉, 刘炯, 梁永春, 等. 支持向量机在电力变压器故障诊断中的应用[J]. 西安交通大学学报, 2007, 41 (6): 722-726.

(下转第 122 页 continued on page 122)

参考文献

[1] 丁明, 安玲, 齐先军. 电力市场环境考虑系统可靠性的备用调度[J]. 继电器, 2007, 35(15):14-17.
DING Ming, AN Ling, QI Xian-jun. Reserve dispatch considering system reliability in electricity market environment[J]. Relay, 2007, 35(15):14-17.

[2] 胡扬宇, 李大鹏, 王子琦, 等. CPS考核标准下河南电网AGC控制策略[J]. 继电器, 2006, 34(14):32-34.
HU Yang-yu, LI Da-peng, WANG Zi-qi, et al. AGC control strategy based on CPS standard in Henan Power Grid[J]. Relay, 2006, 34(14):32-34.

[3] 吴国丙, 任震, 祁达才. 电力市场环境下的无功服务及其成本分析[J]. 继电器, 2002, 30(8):14-17.
WU Guo-bing, REN Zhen, QI Da-cai. The analysis of reactive power service and its cost under electricity market environment[J]. Relay, 2002, 30(8):14-17.

[4] 姚建刚, 章建, 银车来. 电力市场运营及其软件开发[M]. 北京: 中国电力出版社, 2001.

[5] 张少华, 方勇, 李渝曾. 电力市场中的激励性机制设计[J]. 电网技术, 2003, 27(1):52-56.
ZHANG Shao-hua, FANG Yong, LI Yu-zeng. Incentive mechanism design in electricity markets[J]. Power

System Technology, 2003, 27(1): 52-56.

[6] 姚诸香, 邹根华, 罗奇. AGC机组辅助服务确定性指标及应用[J]. 华中电力, 2007, 20(5):10-22, 24.
YAO Zhu-xiang, ZHOU Gen-hua, LUO Qi. Auxiliary service deterministic performance index of AGC units and its usage[J]. Central China Electric Power, 2007, 20(5):10-22, 24.

[7] 华中电监会. 华中区域并网发电厂辅助服务管理实施细则(试行) [R].(华中电监市场[2009]11号).2009年1月.

[8] 华中电监会. 华中区域并网发电厂并网运行管理实施细则(试行) [R].(华中电监市场份[2009]11号).2009年1月.

收稿日期: 2010-01-22; 修回日期: 2010-07-05

作者简介:

田雄(1986-), 男, 硕士研究生, 从事电力市场及其相关软件的开发设计; E-mail: bailaoban1986@126.com

姚建刚(1952-), 男, 教授, 博士生导师, 主要从事电力系统自动化、电力市场和高压外绝缘方向的教学与研究;

龚陈雄(1986-), 男, 硕士研究生, 从事电力市场辅助服务及其相关软件的开发设计。

(上接第 99 页 continued from page 99)

WU Xiao-hui, LIU Jiong, LIANG Yong-chun, et al. Application of support vector machine in transformer fault diagnosis[J]. Journal of Xi'an Jiaotong University, 2007, 41 (6): 722-726.

[8] 厉劫翀, 周宁, 吕彬. 基于神经网络、模糊理论的变压器油中溶解气体诊断专家系统[J]. 电网技术, 2006, 30 (S1) .
LI Jie-chong, ZHOU Ning, Lü Bin. Transformer DGA diagnosis expert system based on neural network and fuzzy theory[J]. Power System Technology, 2006, 30 (S1) .

[9] 潘超, 马成廉, 郑玲峰, 等. 一种结合模糊 TOPSIS 法和 BP 神经网络的变压器故障诊断方法[J]. 电力系统保护与控制, 2009, 37 (9): 20-24.
PAN Chao, MA Cheng-lian, ZHENG Ling-feng, et al. A new method based on fuzzy TOPSIS and BP neural network for power transformer fault diagnosis[J]. Power System Protection and Control, 2009, 37(9): 20-24.

[10] 赵笑笑, 云玉新, 陈伟根. 变压器油中溶解气体的在线监测技术评述[J]. 电力系统保护与控制, 2009, 37(23): 187-191.
ZHAO Xiao-xiao, YUN Yu-xin, CHEN Wei-gen. Comment on on-line monitoring techniques for dissolved gas in transformer oil[J]. Power System Protection and Control, 2009, 37(23): 187-191.

[11] Colomi A, Dorigo M, Maniezzo V, et al. Distributed optimization by ant colonies[C]. //Proceedings of the 1st

European Confer Artificial Life. Paris: 1991.

[12] 周书敬, 李彦苍, 崔郁龙. 基于信息熵的改进蚁群算法及其应用[J]. 数量经济技术经济研究, 2004, 10: 104-109.
ZHOU Shu-jing, LI Yan-cang, CUI Gan-long. Improved ant colony algorithm and its application based on information entropy[J]. The Journal of Quantitative & Technical Economics, 2004, 10: 104-109.

[13] WIERMAN M J. Measuring uncertainty in rough set theory[J]. International Journal of General Systems, 1999, 28(1): 283-297.

[14] LIANG JY, CHINK S, DANG CY, et al. A new method for measuring uncertainty and fuzziness in rough set theory[J]. International Journal of General Systems, 2002, 31(4): 331-342.

[15] ZHAO J Y, ZHANG Z L. Fuzzy rough data reduction based on information entropy[C]. //Proceedings of the Sixth International Conference on Machine Learning and Cybernetics. Hong Kong: 2007: 3708-3712.

收稿日期: 2010-01-08; 修回日期: 2010-03-04

作者简介:

田冰冰(1984-), 女, 硕士研究生, 研究方向为电力系统安全监控和故障诊断新技术; E-mail: tbb123.123@163.com

刘念(1956-), 男, 教授, 博士, 研究方向为电力系统安全监控和故障诊断新技术. E-mail: liunianww@163.com