

动态模糊聚类及其在变压器故障诊断中的应用

陈 舵^{1,2}, 崔杜武¹, 李 雪^{1,3}

CHEN Duo^{1,2}, CUI Du-wu¹, LI Xue^{1,3}

1.西安理工大学 计算机科学与工程学院, 西安 710048

2.唐山学院 计算中心, 河北 唐山 063000

3.陕西师范大学 国际商学院, 西安 710062

1.School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China

2.Computing Centre of Tangshan College, Tangshan, Hebei 063000, China

3.International Business School of Shaanxi Normal University, Xi'an 710062, China

E-mail: vkxtfj@163.net

CHEN Duo, CUI Du-wu, LI Xue. Dynamic fuzzy clustering algorithm and its application to fault diagnosis of power transformer. Computer Engineering and Applications, 2008, 44(20): 26-29.

Abstract: Theoretical analysis and practice indicate that the insulative faults of power transformers are closely related with both the components and the component ratios of various dissolved gas-in-oil. In this paper, a variable length coding genetic fuzzy cluster algorithm is designed to effectively fuse the above two kinds of information and accomplish the dynamic cluster analysis on the faults of power transformer. We adopt a cluster prototype-based dynamic real-coding scheme, in which the variable length chromosomes express cluster prototypes and different length of chromosomes corresponding to different numbers of cluster prototypes. Also, a kind of novel crossover operator and mutation operator are designed to adapt the variable length chromosomes. Theoretical analysis and experimental results indicate that the proposed method is valid.

Key words: dissolved gas analysis; clustering analysis; fault diagnosis; variable-length coding genetic algorithm

摘 要: 理论分析和实践表明, 电力变压器绝缘故障与油中特征气体组分含量及特征气体组分比值密切相关。提出一种基于遗传算法的动态模糊聚类算法, 有效融合特征气体组分含量及组分比值两类故障信息, 完成对变压器故障的动态聚类分析。该算法采用实数编码方案, 染色体长度可变, 不同的长度对应于不同的故障类别数; 并采用一种新的适合于变长染色体的交叉和变异算子。与特征气体法、三比值法进行对比实验, 表明该算法具有较高的判定正确率。

关键词: 溶解气体分析; 聚类分析; 故障诊断; 变长度遗传算法

DOI: 10.3778/j.issn.1002-8331.2008.20.009 **文章编号:** 1002-8331(2008)20-0026-04 **文献标识码:** A **中图分类号:** TM835

1 引言

电力变压器是一种造价较高影响较大的电力设备, 其故障诊断研究具有重要的理论和实用价值。溶解气体分析(Dissolved Gas analysis, DGA)能够有效地发现充油电力变压器内部潜伏性故障以及发展程度, 因而被广泛应用。变压器故障诊断在本质上属模式识别范畴, 聚类分析是多元统计分析的方法之一, 也是统计模式识别中非监督模式分类的一个重要分支^[1]。DGA技术与聚类分析相结合, 是提高变压器故障诊断准确性的一种有效途径。文献[2]根据变压器故障诊断中状态信息的模糊和灰色特征, 借助模糊C均值(FCM)算法, 建立了诊断变压器故障的基于模糊聚类和灰色聚类的集成模型及方法。文献[3]引入反正切型函数分布和加权欧式距离, 结合迭代自组织

分析(ISODATA)聚类算法建立了变压器故障树型分类模型。FCM算法和ISODATA算法属于聚类分析中的划分算法, 有较高的运算效率, 但存在聚类结果依赖初始值和易陷入局部极值的缺点。另外, 这些算法需要事先给定聚类类别数, 为静态聚类, 在实际应用中有一定的局限性。针对上述问题, 一些学者将划分聚类算法与染色体长度可变的遗传算法相结合, 解决动态聚类问题, 并应用于电力变压器的故障诊断^[4], 效果良好。

本文提出一种基于变长编码遗传算法的动态聚类算法, 该算法综合利用特征气体组分含量和组分比值两种故障信息, 自动计算故障模型及故障类别数, 完成变压器故障聚类与诊断。采用实数编码方案, 用染色体直接代表故障聚类原型, 且长度可变, 不同的长度对应于不同的聚类个数, 即故障类别数。采用

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60743009); 陕西省自然科学基金(the Natural Science Foundation of Shaanxi Province of China under Grant No.2006F43)。

作者简介: 陈舵(1962-), 博士生, 主要研究方向: 数据挖掘, 电力电子技术等; 崔杜武(1945-), 教授, 博士生导师, 主要研究方向: 人工智能、数据挖掘、决策支持系统、电力电子技术等; 李雪(1974-), 博士生, 主要研究方向: 智能计算等。

收稿日期: 2008-03-13 **修回日期:** 2008-04-14

新的交叉和变异算子以适合变长遗传编码。提出 FCM 局部优化算子, 加强遗传算法的局部搜索能力, 提高算法效率和求解质量。

2 模糊聚类问题描述

设 $X = \{x_1, x_2, \dots, x_n\}$ 是 n -数据集, $A = \{A_1, A_2, \dots, A_s\}$ 为 X 的属性集。样本 $x_i = [x_{i,1}, x_{i,2}, \dots, x_{i,s}]$, 其中 $x_{i,j} \in \text{DOM}(A_j)$, $\text{DOM}(A_j)$ 是属性 A_j 的值域, $1 \leq i \leq n, 1 \leq j \leq s$ 。对 X 进行模糊聚类分析, 可描述为多约束优化问题:

$$\max f(U, V; \Gamma) \quad (1)$$

$$\begin{cases} \sum_{i=1}^c u_{i,j} = 1 & 1 \leq j \leq n \\ \text{s.t. } 0 \leq u_{i,j} \leq 1 & 1 \leq j \leq n, 1 \leq i \leq c \end{cases} \quad (2)$$

$$\text{s.t. } 0 < \sum_{j=1}^n u_{i,j} < n \quad 1 \leq i \leq c \quad (3)$$

$$\text{s.t. } 0 < \sum_{j=1}^n u_{i,j} < n \quad 1 \leq i \leq c \quad (4)$$

式中, $f(\cdot)$ 为聚类目标函数, Γ 为参数集, $U = [u_{i,j}]_{c \times n}$ 是模糊划分矩阵, 矩阵元素 $u_{i,j}$ 表示样本 x_j 对于第 i 类的隶属度。 $V = \{V_1, V_2, \dots, V_c\}$ 为聚类原型 (Cluster Prototype) 集, 其中 $V_l = (v_{l,1}, v_{l,2}, \dots, v_{l,s})$, $v_{l,j} \in \text{DOM}(A_j)$, $1 \leq l \leq c, 1 \leq j \leq s$ 。根据是否需要预先给定聚类数 c , 分别将上述问题称为静态聚类分析和动态聚类分析。

3 基于变长编码遗传算法的动态模糊聚类

本章讨论变长编码及相关的进化和优化策略, 提出基于变长编码遗传算法的动态模糊聚类算法。

3.1 染色体编码

本文采用基于聚类原型的实数编码方法, 一条染色体由 c 个聚类原型组成:

$$H = V_1 V_2 \dots V_c = v_{1,1} v_{1,2} \dots v_{1,s} \dots v_{c,1} v_{c,2} \dots v_{c,s} \quad (5)$$

$$\text{s.t. } v_{i,j} \in \text{DOM}(A_j), j=1, 2, \dots, s, l=1, 2, \dots, c \quad (6)$$

采用上述直接编码方法, 优点是编码-解码过程简单高效, 且满足正则性。

令类别数 c 在区间 $[c_{\min}, c_{\max}]$ 变化, 构成变长染色体, 求解动态聚类问题。参数 c_{\min} 和 c_{\max} 可根据领域知识来选择, 但这并非总是可行的, 所以很多研究者使用经验值 $c_{\min} = 2$ 和 $c_{\max} = \sqrt{n}^{\text{[9]}}$ 。

3.2 遗传算子

典型的遗传算子包括选择、交叉和变异。

(1) 选择算子

选择算子以对染色体适应度的评价为基础, 对遗传算法的搜索提供导向: 依据适应度的大小决定染色体被遗传到下一代的概率, 主要有适应值比例选择、排序选择和联赛选择等。本文采用比例选择算子, 染色体被选中的概率与其适应值成正比。

(2) 交叉算子

交叉算子作用于被选定的父代染色体, 按某种方式交换其部分基因而形成具有新基因结构的子代染色体。交叉算子是否能够有效地产生有意义的新染色体, 在很大程度上决定了遗传算法性能的好坏。常用的交叉操作包括单点交叉、多点交叉、均匀交叉、算术交叉等。对于变长编码, 传统的交叉操作已经不适用, 因此提出如下改进的交叉算子。

设 $H^{(1)}$ 和 $H^{(2)}$ 为待交叉的两个父代染色体, 其中 $H^{(i)} = V_1^{(i)} V_2^{(i)} \dots V_{c_i}^{(i)}$, $i=1, 2$ 分别为 $H^{(1)}$ 和 $H^{(2)}$ 聚类原型基因序列, 不妨设

$$c_1 \leq c_2。$$

步骤 1 参照文献[6]方法, 对 $H^{(1)}$ 和 $H^{(2)}$ 进行最短距离基因匹配: 对 $H^{(1)}$ 中每个代表聚类原型的基因 $V_l^{(1)} = v_{l,1}^{(1)} \dots v_{l,s}^{(1)}$, $l=1, \dots, c_1$, 选择 $H^{(2)}$ 中与 $V_l^{(1)}$ 距离最近的基因 $V_h^{(2)} = v_{h,1}^{(2)} \dots v_{h,s}^{(2)}$, $h=1, \dots, c_2$, 将 $V_h^{(2)}$ 与 $V_l^{(1)}$ 配对, 并且在基因配对的过程中, $V_h^{(2)}$ 中已经配对的基因不再参加后续的配对。按照基因配对的顺序重新排列 $H^{(2)}$ 中的 c_1 个基因, 其余 $(c_2 - c_1)$ 个基因保持原始次序, 这样构成一个新的基因序列, 记为 $\hat{H}^{(2)} = \hat{V}_{c_1+1}^{(2)} \hat{V}_{c_1+2}^{(2)} \dots \hat{V}_{c_2}^{(2)}$ 。

步骤 2 对 $H^{(1)} = V_1^{(1)} V_2^{(1)} \dots V_{c_1}^{(1)}$ 和 $\hat{H}^{(2)}$ 中的 $\hat{V}_1^{(2)} \hat{V}_2^{(2)} \dots \hat{V}_{c_1}^{(2)}$ 这两段等长基因序列, 采用普通算术交叉操作, 生成的子基因序列记为 $V_1^{(3)} V_2^{(3)} \dots V_{c_1}^{(3)}$ 和 $V_1^{(4)} V_2^{(4)} \dots V_{c_1}^{(4)}$ 。

步骤 3 对于 $\hat{H}^{(2)}$ 中的基因序列 $\hat{V}_{c_1+1}^{(2)} \hat{V}_{c_1+2}^{(2)} \dots \hat{V}_{c_2}^{(2)}$, 平均分成两段基因序列, 分别拼接于 $V_1^{(3)} V_2^{(3)} \dots V_{c_1}^{(3)}$ 和 $V_1^{(4)} V_2^{(4)} \dots V_{c_1}^{(4)}$, 构成基因序列 $H^{(3)} = V_1^{(3)} \dots V_{c_1}^{(3)} \hat{V}_{c_1+1}^{(2)} \dots \hat{V}_k^{(2)}$ 和 $H^{(4)} = V_1^{(4)} \dots V_{c_1}^{(4)} \hat{V}_{k+1}^{(2)} \dots \hat{V}_{c_2}^{(2)}$, 其中, $k = c_1 + (c_2 - c_1) / 2$ 。

不难证明: 若父代基因序列 $H^{(1)}$ 和 $H^{(2)}$ 满足约束(6), 则经过步骤 1~步骤 3 生成的子代基因序列 $H^{(3)}$ 和 $H^{(4)}$ 同样满足约束(6)。

交叉算子中, 步骤 1 的聚类原型基因序列重排操作使得不同染色体的等位基因具有相关的聚类原型, 从而避免了较好父代基因交叉后产生极差子代基因而导致的种群退化现象; 步骤 2 是等长基因序列的常规交叉运算, 体现了子代基因对父代基因的继承; 步骤 3 是一种拼接操作, 基因值并不发生改变, 但使得子代基因序列的类别数趋于相同, 是一种对父代“类别数”的继承。

(3) 变异算子

变异算子包括两步操作: 基因值变异和类别数变异。

步骤 1 对基因值均匀变异。首先随机选出染色体中变异基因位, 然后考虑到基因值约束于相应属性的值域, 用在相应属性的值域上均匀分布的随机值替代变异基因位的值。

步骤 2 类别数变异采用正态变异运算。设父代染色体的类别数为 c_1 , 则变异后子代类别数为 $c_2 = N(c_1, \sigma)$ 。若 $c_2 \leq c_1$, 则从父代染色体中随机截取 c_2 个基因形成子代染色体; 若 $c_2 > c_1$, 则子代染色体在父代染色体的末尾添加 $(c_2 - c_1)$ 个基因, 添加的基因在满足约束(6)的条件下随机产生。

变异算子中, 步骤 1 和步骤 2 分别对基因值和基因数 (即类别数) 实施某种突变, 使得种群中的染色体在基因值和基因数两方面保持一定的多样性。

3.3 FCM 局部优化算子

FCM 聚类算法是一种交替优化的聚类算法, 在遗传算法中使用该算法作为局部优化算子, 能有效地扩展遗传算法的局部搜索能力, 进而提高算法效率和求解质量。优化过程为: 对于染色体的聚类原型基因 $V_1^{(h)} V_2^{(h)} \dots V_{c_h}^{(h)}$, $h=1, \dots, N$, N 为种群规模, 先求出模糊划分矩阵 U , 再依据 U 对聚类原型基因进行一次更新。

3.4 聚类有效性函数及适应度函数

适应度是个体对于全局最优解可接受程度的一种评价尺

度。适应度函数是个体空间到正实数集的一个映射,通常和具体领域中的目标函数密切相关,对于聚类问题,目标函数可选择为某种适合的聚类有效性函数。文献[7]综合考虑聚类模糊集的模糊度和贴近度,给出一种新的聚类有效性函数 $DS(U;c)$,效果较好。模糊度定义为:

$$D(U;c) = \sum_{i=1}^c D_i(U;c) \quad (7)$$

式中:

$$D_i(U;c) = \sqrt{\frac{1}{n} \sum_{j=1}^n (u_{ij} - (u_{ij})_T)^2} \quad (8)$$

$$(u_{ij})_T = \begin{cases} 1 & u_{ij} \geq \frac{1}{c} \\ 0 & \text{其它} \end{cases} \quad (9)$$

贴近度定义为:

$$N(\tilde{F}_i, \tilde{F}_k) = \min[(\tilde{F}_i \circ \tilde{F}_k), (\tilde{F}_i \hat{\circ} \tilde{F}_k)^c] \quad (10)$$

式中, $\tilde{F}_i, \tilde{F}_k \in U, i, k \in \{1, 2, \dots, c\}$, $(\tilde{F}_i \circ \tilde{F}_k) = \bigvee_{j=1}^n (u_{ij} \wedge u_{kj})$ 为模糊集内积, $(\tilde{F}_i \hat{\circ} \tilde{F}_k) = \bigwedge_{j=1}^n (u_{ij} \vee u_{kj})$ 为模糊集外积, 符号 \wedge 、 \vee 和 $(\cdot)^c$ 分别是取小、取大和求补算子。

聚类有效性函数 DS :

$$DS(U;c) = \frac{\phi(c) \cdot \max_{i=1}^c [D_i(U;c)]}{\min_{h \neq k} [\text{scat}(\tilde{F}_h, \tilde{F}_k)]} \quad (11)$$

式中, $\phi(c) = c/(c-1)$, scat 为聚类离散度:

$$\text{scat}(\tilde{F}_i, \tilde{F}_k) = 1 - N(\tilde{F}_i, \tilde{F}_k) \quad (12)$$

聚类有效性函数 DS 是比例型函数, 分子度量类内紧致性, 数值越小紧致性越好; 分母度量类间分离性, 数值越大, 分离性越好。较小的 DS 值代表较好的聚类结果, 在聚类有效性函数 DS 中, 分子和分母的取值范围都是 $[0, 1]$ 。注意到 DS 的分母不能取零值, 事实上, 对于模糊聚类, $\text{scat}(\tilde{F}_h, \tilde{F}_k)$ 的值一般不为零, 为了避免特殊情况下分母为零函数出现极值, 可令分母为零时函数取一个较大正数 M 。

本文基于聚类有效性函数 DS , 定义染色体 H 的适应度 $J(H)$:

$$J(H) = M - DS(U;c) \quad (13)$$

3.5 变长编码遗传模糊聚类算法

在上述工作基础上, 给出一种基于变长编码遗传算法的动态模糊聚类算法, VLGFC (Variable-Length coding Genetic dynamic Fuzzy Cluster Algorithm)

步骤 1 初始化。输入数据集 X ; 设定种群规模 N , 交叉概率 P_c , 变异概率 P_m , 最大代数 $Maxgen$, 模糊加权指数 m , 置进化代数 $t:=0$; 初始化种群 $G(0)$: 随机生成类别数, 并根据类别数在 X 中随机选择相应数目的样本生成染色体。令 $\tilde{H}(0)$ 为 $G(0)$ 中的最佳染色体。

步骤 2 对当代种群 $G(t)$ 实施选择、交叉、变异和 FCM 局部优化操作, 生成下代种群 $G(t+1)$ 。

步骤 3 设 $\tilde{H}(t+1)$ 是 $G(t+1)$ 中最佳染色体, 则令 $\hat{H}(t+1) = \max\{\hat{H}(t), \tilde{H}(t+1)\}$, 即 $\hat{H}(t+1)$ 是到目前为止的最佳染色体; 若 $\hat{H}(t+1)$ 优于 $\tilde{H}(t+1)$, 则用 $\hat{H}(t+1)$ 替换 $\tilde{H}(t+1)$ 。

步骤 4 若 $t=Maxgen$, 则输出最佳染色体 $\hat{H}(t+1)$, 算法结束; 否则, 令 $t:=t+1$, 转到步骤 2。

在算法 HVGFC 中, 为了不丢失最优解并保证算法收敛, 采用精英记录策略, 即算法总是记录到目前为止发现的最佳染色体, 所以, 适应值序列 $\{J(\hat{H}(t))\}$ 必然具有单调性, 算法收敛。

4 特征气体组分含量和组分比值的融合

4.1 特征气体的选取及规格化

本文选用油中溶解气体 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 和 C_2H_2 为故障特征气体。由于电压等级、变压器容量等存在差异, 变压器油中各溶解气体组分含量差别很大, 在分析之前需要规格化处理, 即:

$$\bar{u}_{i,j} = \frac{u_{i,j}}{\sum_{k=1}^5 u_{i,k}}, 1 \leq i \leq n, 1 \leq j, k \leq 5 \quad (14)$$

式中, $u_{i,j}$ 是原始样本数据; $\bar{u}_{i,j}$ 表示处理后的数据; i 是样本序号; j 为属性序号, 用 1~5 依次表示上述 5 种特征气体。

4.2 特征气体比值的选取与处理

DL/T722-2000《变压器油中溶解气体分析和判断导则》推荐改良三比值法作为设备故障诊断的主要方法, 据此, 本文使用三对比值, 即 C_2H_2/C_2H_4 、 CH_4/H_2 和 C_2H_4/C_2H_6 作为特征气体组分比值。特征气体的组分比值与组分含量是性质不同的两类数据, 所以, 需对比值进行必要的数据处理。

文献[8]通过对特征气体的组分比值故障判别能力的深入分析, 提出下述观点: (1) 比值 C_2H_2/C_2H_4 的故障判断意义明确, 该比值越小, 发生热性故障的可能性越大; 反之, 比值越大, 发生电性或电性兼热性故障的可能性越大。 (2) 当比值 C_2H_2/C_2H_4 的值较小 (小于 0.1, 编码值为 0) 时, C_2H_4/C_2H_6 比 CH_4/H_2 有更强的故障判断能力, 该值越大, 所对应的故障温度越高; (3) 当比值 C_2H_2/C_2H_4 的值较大 (大于 0.1, 编码值为 1, 2) 时, CH_4/H_2 的故障判断能力远大于 C_2H_4/C_2H_6 , 后者几乎失去了故障判断能力。据此, 针对特征气体组分比值进行规格化及提升与压缩处理。

(1) 规格化

$$\bar{u}_j = \frac{1}{c_1} \min(u_j, k_1) \quad j=6, 7, 8 \quad (15)$$

式中, u_j 、 \bar{u}_j ($j=6, 7, 8$) 分别表示三对比值 C_2H_2/C_2H_4 、 CH_4/H_2 和 C_2H_4/C_2H_6 原始值和规格化后的值。取 $k_1=3.3$ 。

(2) 压缩与提升

$$\hat{u}_7 = k \times \bar{u}_7 \times f(u_6) \quad (16)$$

$$\hat{r}_8 = k \times \bar{u}_8 \times [2 - f(u_6)] \quad (17)$$

式中, \hat{u}_7 和 \hat{u}_8 分别表示对组分比值 u_7 和 u_8 进行规格化和提升压缩处理后的值, 且:

$$f(x) = \frac{2}{\pi} \arctg[k_3(x - k_2)] + 1 \quad (18)$$

利用 $f(x)$ 函数可以实现根据 u_6 的值对 u_7 和 u_8 的提升和压缩。当 $u_6 > k_2$ 时, $f(x) > 1$, \hat{u}_7 的值增加, \hat{u}_8 的值减小, 即提升 \hat{u}_7 压缩 \hat{u}_8 , 当 $r_1 >> k_2$ 时, $f(x) \rightarrow 2$, \hat{u}_7 获得接近 2 倍的提升, 同时 $\hat{u}_8 \rightarrow$

表 1 四种诊断方法实验结果对比

故障类型	统计台数	特征气体聚类		改良三比值法		综合聚类		综合动态聚类	
		判别正确台数	正确率/%	判别正确台数	正确率/%	判别正确台数	正确率/%	判别正确台数	正确率/%
局部放电	15	14	93.33	11	73.33	14	93.33	14	93.33
低能放电	35	19	54.29	26	74.29	26	74.29	26	74.29
电弧放电	51	11	21.57	47	92.16	46	90.20	47	92.16
低温过热	22	15	68.18	16	72.73	17	77.27	17	77.27
中温过热	26	23	88.46	24	92.31	25	96.15	25	96.15
高温过热	64	48	75.00	63	98.44	63	98.44	63	98.44
合计/正确率	213	130	61.03	187	87.79	191	89.67	192	90.14

0; 同理, 当 $u_7 < k_2$ 时, 提升 \hat{u}_8 压缩 \hat{u}_7 。参数 k_2 确定了提升与压缩的分界点, 参照三比值法编码规则, 取 $k_2 = 0.1$ 。参数 k_3 用来调节函数的梯度, 其数值越大曲线越陡峭, 估算方法为: 当自变量增大(减小)10%时, 函数值增大(减小)80%, 据此估算 $k_3 = 296$ 。

系数 k 用来限定 \hat{u}_7 和 \hat{u}_8 的取值范围, 取 $k = 0.5$, 则 $0 \leq \hat{u}_7, \hat{u}_8 \leq 1$ 。

4.3 特征气体组分含量和组分比值的动态聚类分析

对故障样本经过上述处理, 得到聚类数据集 $X = \{x_1, x_2, \dots, x_n\}$, 样本 $x_i = (x_{i,1}, \dots, x_{i,p}, x_{i,p+1}, \dots, x_{i,s})$, $p = 5, s = 8, 1 \leq i \leq n$, 其中, $x_{i,j} = \bar{u}_{i,j}$, $j \leq p$, 是规格化的组分数据, $x_{i,6} = \bar{u}_{i,6}$ 是规格化的 C_2H_2/C_2H_4 比值数据, $x_{i,7} = \hat{u}_{i,7}$, $j = 7, 8$, 分别是经规格化和提升与压缩处理的 CH_4/H_2 及 C_2H_4/C_2H_6 比值数据。

聚类数据集 X 中同时包含特征气体组分含量和组分比值数据, 体现了两类信息的融合。应用本文算法 VLGFA 对数据集 X 进行动态模糊聚类分析, 实现故障聚类, 求得故障数和故障模型。

5 实例分析

充油电力变压器中典型故障类型有低温过热 ($t < 300^\circ C$)、中温过热 ($300^\circ C < t < 700^\circ C$)、高温过热 ($t > 700^\circ C$)、局部放电、低能放电、高能放电 6 种。本文从公开刊物上发表的故障变压器 DGA 数据和唐山供电公司故障变压器 DGA 数据中随机选取 213 台变压器样本数据, 按如下方法进行对比分析:

- (1) 特征气体法: 以上述 5 种特征气体的组分含量为特征量, 按公式(14)进行规格化处理, 应用 FCM 算法进行故障聚类。
- (2) 三比值法: 采用 DL/T722-2000 导则推荐的改良三比值法进行故障判断。

(3) 特征气体组分含量及组分比值综合聚类法: 以 5 种特征气体组分含量和它们的 3 对比值为聚类特征量, 按公式(15)~(18)分别进行规格化和提升与压缩处理, 指定故障类数 $c = 6$, 用 FCM 算法进行故障聚类, 求得故障模型。

(4) 特征气体组分含量及组分比值综合动态聚类法: 使用与方法(3)相同的聚类特征量及特征量处理方法, 用 VLGFA 算法进行动态故障聚类, 求得故障类数和故障模型。

算法参数: $m = 2.2, N = 50, M = 2, P_c = 0.8, P_m = 0.25, Maxgen = 300$ 。

图 1 显示了在 VLGFA 算法中最佳个体适应度和类别数的进化过程。在进化初期, 适应度迅速增加, 随着进化的进行, 适应度的增加逐步减缓而趋于稳定; 最佳类别数在进化初期出现振荡, 随着进化的进行振荡现象消失, 最终稳定于最佳类别数 $c_{opt} = 6$ 。

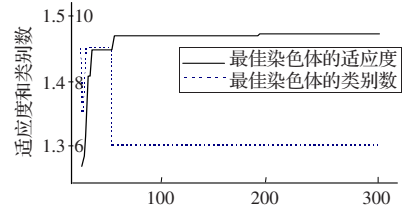


图 1 VLGFA 算法的进化过程图

表 1 为上述四种方法实验结果对比。表中数据表明, 对于局部放电故障, 特征气体聚类法、综合聚类法和综合动态聚类法的正确率明显高于改良三比值法, 这是因为局部放电属于 H_2 主导型故障, 即 H_2 的含量远大于其它特征气体, 所以较容易根据特征气体含量的分布进行故障判断, 同时由于特征气体分布强烈的偏斜性, 使得比值数据的故障判别能力下降; 对于其它故障类型, 仅利用特征气体组分含量的特征气体聚类法的正确率远低于其它三种方法, 说明对于这些故障类型, 比值数据具有更强的判别能力。在低能放电、电弧放电和高温过热故障识别中, 后三种方法结果大体相当; 对于低温过热和中温过热故障, 综合聚类和综合动态聚类法的正确率略高于改良三比值法。从总体上看, 由于综合聚类和综合动态聚类法利用了更多的故障信息, 因而兼有特征气体聚类法和比值法的优点, 故障划分正确率较高, 本文提出的 VLGFA 综合动态聚类法获得最高故障划分正确率, 且能够自动确定故障类别数, 性能优良。

表 2 给出了 6 种故障的聚类原型。注意到聚类数据经过了规格化和提升与压缩处理, 所以在参考上述聚类原型进行故障诊断时, 应首先对故障样本做同样的处理, 计算样本对各故障类型的隶属度, 然后根据隶属度最大原则进行故障诊断。

表 2 VLGFA 算法计算的 6 个故障类型的聚类原型

故障类型	聚类原型							
	\bar{u}_1	\bar{u}_2	\bar{u}_3	\bar{u}_4	\bar{u}_5	\bar{u}_6	\bar{u}_7	\bar{u}_8
局部放电	0.843 8	0.069 7	0.045 9	0.031 9	0.008 7	0.023 0	0.007 0	0.176 2
低能放电	0.435 8	0.096 9	0.048 7	0.075 2	0.343 4	0.960 6	0.081 9	0.001 8
电弧放电	0.456 0	0.139 2	0.041 1	0.189 2	0.174 5	0.305 1	0.102 5	0.010 8
低温过热	0.323 1	0.429 5	0.110 2	0.112 3	0.024 9	0.039 0	0.034 3	0.220 9
中温过热	0.126 9	0.320 3	0.171 5	0.375 6	0.005 8	0.007 4	0.012 5	0.651 9
高温过热	0.127 2	0.242 4	0.085 4	0.533 0	0.011 9	0.007 4	0.009 5	0.976 9

用本文方法和其它方法还对唐山市贾安子 3 号变压器故障实例进行了分析, 以下是情况介绍: 贾安子 3 号变压器, 型号 SFPSZ3-120000/220, 保定变压器厂生产。2000 年 6 月 11 日事故后油色谱数据为 ($\mu L/L$): $H_2 = 56, CH_4 = 25, C_2H_6 = 6, C_2H_4 = 85, C_2H_2 = 10$, 三比值编码 102, 属电弧放电。本文方法判断结果为