

# 基于谱聚类集成的变压器在线故障诊断

刘荣胜<sup>1</sup> 彭敏放<sup>1</sup> 肖祥慧<sup>2</sup>

(1. 湖南大学电气与信息工程学院 湖南长沙 410082; 2. 长沙理工大学电气与信息工程学院 湖南长沙 410114)

**摘 要:** 为了提高基于油中溶解气体分析技术(DGA)的变压器故障诊断准确率,本文提出了一种基于谱聚类集成的变压器在线故障诊断(TOFD-SCE)方法.以加权二次抽样算法抽取样本、构建基础谱聚类的样本集,以基础谱聚类学习问题的局部知识;平衡多样性与正确性选择集成成员;集成多个成员谱聚类的结果来提高变压器故障诊断的准确率.传统变压器故障诊断方法基于历史数据建立模型,缺乏在线学习的能力;TOFD-SCE利用历史与在线新增两种DGA数据来训练、修正模型,提高了其故障诊断的准确率.对SSP300000/500型变压器的故障诊断实验结果表明:TOFD-SCE的准确率优于IEC三比值法、BP神经网络法及支持向量机法,验证了其有效性.

**关键词:** 故障诊断; 变压器; 谱聚类集成; 油中溶解气体分析; 信号处理

**中图分类号:** TM411+.2 **文献标识码:** A **文章编号:** 0372-2112 (2017) 10-2491-07

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.10.025

## The Transformer On-line Fault Diagnosis Based on Spectral Clustering Ensemble

LIU Rong-sheng<sup>1</sup>, PENG Min-fang<sup>1</sup>, XIAO Xiang-hui<sup>2</sup>

(1. College of Electrical and Information Engineering, Hunan University, Changsha, Hunan 410082, China;

2. College of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha, Hunan 410114, China)

**Abstract:** To improve the accuracy of the transformer fault diagnosis based on dissolved gas analysis in oil (DGA), a transformer on-line fault diagnosis based on spectral clustering ensemble (TOFD-SCE) was proposed in this paper. The weighted double sampling algorithm create the samples set of the basic spectral clustering, which learned the local knowledge of the problems. The accuracy was improved by integrating the results of ensemble members, which were picked up from the basic spectral clustering in terms of the accuracy and variety. The conventional models are only trained by the historical data, and can't learn on-line. TOFD-SCE is trained and modified by both historical and new online data, and the accuracy is improved. The TOFD-SCE was validated by diagnosing the fault of SSP300000/500 transformers. Comparing with IEC three ratio, BP-neural networks and support vector machine, TOFD-SCE is more outstanding.

**Key words:** fault diagnosis; power transformer; spectral clustering ensemble; dissolved gas analysis in oil; signal processing

## 1 引言

油中溶解气体分析技术(Dissolved Gas Analysis, DGA)已被广泛应用于油浸式变压器潜伏性故障诊断<sup>[1]</sup>.常规的变压器故障诊断方法包括关键气体法、IEC60599比值法<sup>[2]</sup>、Doernenburg比值法、Rogers比值法和Duval三角法<sup>[3]</sup>等.常规诊断方法简单、实用;但故障征兆与故障类型编码表依赖于专家的经验<sup>[1]</sup>、编码不

全、边界过于绝对、准确率偏低、不能适应多故障并发的诊断<sup>[4]</sup>.模式识别、机器学习及人工智能等被应用于基于DGA的变压器故障诊断,以此来提高故障诊断的准确率,如专家系统<sup>[5]</sup>、层次分析法、BP神经网络(BP-Neural Network, BPNN)<sup>[6]</sup>、模糊逻辑(Fuzzy Logic, FL)<sup>[7]</sup>、支持向量机(Support Vector Machine, SVM)<sup>[8,9]</sup>等.专家系统通过融合领域专家的经验来提高故障诊断的准确率,但诊断规则源于专家的间接经验,不能直

收稿日期:2015-11-28;修回日期:2017-03-06;责任编辑:孙瑶

基金项目:国家自然科学基金(No.61472128, No.61173108);湖南省自然科学基金重点项目(No.14JJ2150);国家电网公司总部科技项目(No.5216A514001K)

接从 DGA 数据中学习知识. 层次分析法降低了主观因素对变压器故障诊断的负面影响, 但不能从新增 DGA 数据中在线学习、调整权重、修正模型. Vladimiro Miranda 等将自相关神经网络应用于基于 DGA 的变压器故障诊断<sup>[6]</sup>, 效果良好, 但神经网络的学习过程需要大量的历史或者仿真样本, 网络结构的选择依赖于个人经验, 亦缺乏在线学习能力. 文献[7]应用 FL 解决变压器故障诊断的边界过于绝对、结论不一致等问题, 但其知识来源于其它诊断方法, 不能直接从 DGA 数据中学习, 隶属度函数与模糊规则不能在线学习, 自适应的调整. SVM 处理小样本学习时具有明显优势, 但易受随机因素影响, 基于 SVM 的变压器故障诊断的适应性和可靠性有待进一步提高, 尺度函数与罚参数的选择需要进一步研究. 组合诊断方法融合多种单体方法的优点来提高其故障诊断的准确率, 已成为变压器故障诊断的研究热点<sup>[10]</sup>. 但多故障并发时通用性较差, 缺乏根据新增 DGA 数据在线学习的能力.

针对现有变压器故障诊断方法缺乏在线学习能力的问题, 本文提出了一种基于谱聚类集成的变压器在线故障诊断方法 (Transformers On-line Fault Diagnosis Based on Spectral Clustering Ensemble, TOFD-SCE). TOFD-SCE 利用历史与在线新增两种 DGA 数据资源来提高其整体性能. 通过对 SSP300000/500 型变压器进行基于 DGA 的变压器故障诊断实验, 验证了 TOFD-SCE 的有效性. 与 IEC 三比值法、BPNN 及 SVM 相比, TOFD-SCE 的准确率有较大提高.

## 2 谱聚类集成分类器模型

受额定功率、电压等级、役龄、型号、生产厂家及负载等因素影响, 变压器油中溶解气体与其故障类型、严重程度之间的关系具有复杂性和模糊性<sup>[11]</sup>. 多分类器集成是一种解决困难学习问题的有效方法, 已成为机器学习、模式识别及数据挖掘领域的重要研究方向. 谱聚类具有良好的数学基础, 可避免因维数过高而引起的奇异性问题, 可在任意形状的样本空间上聚类<sup>[12]</sup>, 通过谱松弛实现全局优化<sup>[13]</sup>. 鉴于谱聚类与集成学习的优越性, 提出了基于谱聚类集成 (Spectral Clustering Ensemble, SCE) 的变压器故障诊断, SCE 包括 3 个阶段: 基础谱聚类生成、成员谱聚类选择与集成分类器构建.

### 2.1 基础谱聚类生成

谱聚类的样本集  $S = \{s_0, s_1, \dots, s_N\}$ , 其中类别已知子集  $S' = \{s_1, \dots, s_N\}$ , 子集规模  $N = |S'|$ ,  $s_0$  为待识别样本. 类别集合  $T = \{t_1, \dots, t_L\}$ , 类别数目  $L = |T|$ . 基础谱聚类分类器根据子集  $S'$  中蕴含的类别信息对样本  $s_0$  进行分类, 其算法如算法 1 所示.

#### 算法 1 谱聚类分类器算法 (Spectral Clustering Classifier Algorithm, SCCA)

输入: 样本集  $S = \{s_0, s_1, \dots, s_N\}$

输出:  $s_0$  的类别

步骤 1: 将样本集  $S$  用谱聚类划分为  $H$  个簇, 谱聚类结果如式 (1);

$$\pi = \{C_1, C_2, \dots, C_H\} \quad (1)$$

步骤 2: 计算样本在簇中的分布矩阵  $M$ ,  $M_{ij}$  表示簇  $C_j$  中类别为  $t_i$  的样本数;

步骤 3: 在聚类结果  $\pi$  中搜索包含  $s_0$  的簇  $C_j$ ;

步骤 4: 按式 (2) 计算簇  $C_j$  中样本数量最多的类别编号  $I$ ;

$$I = \arg \max_{i \in \{1, \dots, L\}} \{M_{ij}\} \quad (2)$$

步骤 5: 输出  $s_0$  的类别  $t_I$ .

SCCA 算法基于谱聚类结果对样本  $s_0$  进行分类, 其结果受聚类簇数、样本集、尺度函数、二次聚类方法及其初值等影响, 为了减少影响谱聚类结果的因素, 提高模型的实用性, SCCA 算法采用基于本征间隙与正交特征向量的自动谱聚类<sup>[14]</sup>. 自动谱聚类根据样本集的特征确定聚类数, 不需要二次聚类, 其聚类结果只受样本集与尺度函数影响. 基础谱聚类集合记为  $D = \{f_1, \dots, f_d\}$ .

### 2.2 成员谱聚类选择

从基础谱聚类集合中选择合适的集成成员是多分类器集成的关键. 准确性与多样性是集成成员选择的两个准则, 成员数为  $d'$  的成员谱聚类集合  $D'$  应满足:  $D' \subseteq D$ ,  $|D'| = d'$  且式 (3) 有最大值<sup>[15]</sup>.

$$\varphi(D') = \lambda \sum_{f_m \in D'} A(f_m) + (1 - \lambda) \sum_{f_m, f_n \in D'} kp(f_m, f_n) \quad (3)$$

其中  $A(f_m)$  为成员分类器  $f_m$  的准确率,  $kp(f_m, f_n)$  为成员分类器  $f_m$  与  $f_n$  的差异性度量指标成对 Kappa<sup>[15]</sup>,  $\lambda \in [0, 1]$  是平衡差异性和准确性的调节因子.

### 2.3 谱聚类集成分类器构建

谱聚类集成以成员谱聚类学习问题的局部知识, 集成多个成员谱聚类的结果来提高整体识别的准确率<sup>[13]</sup>. 成员谱聚类分类器  $f_i$  生成的信息包括谱聚类结果  $\pi_i = \{C_1^i, C_2^i, \dots, C_H^i\}$ 、样本在簇中的分布矩阵  $M$  及样本  $s_0$  的判定类别  $t_i$ . 基于判定类别  $t_i$  进行简单投票方式集成, 计算复杂度小, 速度快; 但忽略了成员谱聚类分类器对最终判决的贡献度, 准确率较低. 为了提高谱聚类集成分类器的整体性能, 本文基于成员谱聚类的聚类结果构建集成分类器, 过程如下:

成员谱聚类集合  $D' = \{f_1, \dots, f_{d'}\}$ , 成员谱聚类的聚类结果  $\Pi = \{\pi_i | i = 1, 2, \dots, d'\}$ , 其中  $\pi_i$  为成员谱聚类  $f_i$  的聚类结果,  $d'$  为集成成员规模. 待识别的数据集  $X$ , 任意数据点  $x \in X$ , 任意的类型  $t \in T$ ,  $x$  的类别为  $t$  的概率如式 (4):

$$P(\text{Typeof}(x) = t) = \frac{\text{Count}(x, t)}{\text{Count}(x)} \quad (4)$$

其中  $\text{Typeof}(x) = t$  表示  $x$  的类别为  $t$ ,  $\text{Count}(x, t)$  表示全体包含  $x$  的簇中, 类型为  $t$  的样本数目之和, 如式 (5);  $\text{Count}(x)$  表示全体包含  $x$  的簇的样本数目之和, 如式 (6)。

$$\text{Count}(x, t) = \sum_{\substack{r \in S, r \neq x, C \in \pi \\ i=1, 2, \dots, d'}} \mu(r, t) \delta(r, C) \delta(x, C) \quad (5)$$

$$\text{Count}(x) = \sum_{\substack{r \in S, r \neq x, C \in \pi \\ i=1, 2, \dots, d'}} \delta(r, C) \& \delta(x, C) \quad (6)$$

其中  $\delta$  为  $x$  属于簇  $C$  的标志函数如式 (7),  $\mu$  为  $x$  的类别为  $t$  的标志函数如式 (8)。

$$\delta(x, C) = \begin{cases} 1, & x \in C \\ 0, & \text{else} \end{cases} \quad (7)$$

$$\mu(x, t) = \begin{cases} 1, & \text{if Typeof}(x) = t \\ 0, & \text{else} \end{cases} \quad (8)$$

如果  $x$  的类别为  $t_i$  的概率远大于其它类别, 即:  $P(\text{Typeof}(x) = t_i) \gg P(\text{Typeof}(x) = t_{\text{else}})$ , 那么  $x$  的类别为  $t_i$ ; 如果  $x$  的类别为  $t_i$  的概率与属于  $t_j$  的概率相当, 并远大于其它类型, 即:  $P(\text{Typeof}(x) = t_i) \approx P(\text{Typeof}(x) = t_j) \gg P(\text{Typeof}(x) = t_{\text{else}})$ , 且  $t_i, t_j$  均为故障类别, 则  $x$  的类别为  $t_i, t_j$  同时发生的多故障类型。

### 3 基于 SCE 的变压器故障诊断

变压器油中溶解气体的种类、含量、比例及增长情况与其故障类型、严重程度密切相关。大型油浸式电力

变压器已基本实现 DGA 信息的在线监测, DGA 数据可在线采集并实时存储。为了利用历史与在线新增两种 DGA 数据提高变压器故障诊断的准确率, 提出了一种基于 SCE 的变压器故障在线诊断方法。

#### 3.1 数据预处理

DGA 数据如表 1 所示, 包括变压器的状态  $T$  及  $\text{CH}_4$ 、 $\text{C}_2\text{H}_6$ 、 $\text{C}_2\text{H}_4$ 、 $\text{C}_2\text{H}_2$ 、 $\text{H}_2$ 、 $\text{CO}$  与  $\text{CO}_2$  等 7 种特征气体的浓度。特征气体的浓度是变压器故障诊断的重要特征, 7 种气体的浓度分别用下标  $1, 2, \dots, 7$  标记; 油中溶解气体含量比值亦是变压器故障诊断的重要指标,  $\text{C}_2\text{H}_2/\text{C}_2\text{H}_4$ 、 $\text{CH}_4/\text{H}_2$ 、 $\text{CH}_4/\text{C}_2\text{H}_6$  等三组气体比值也作为变压器故障诊断的特征, 分别用下标 8、9、10 表示。规模为  $N$  的样本集  $S$  的特征量表示成  $N \times 10$  的矩阵  $V$ ,  $V_{ij}$  代表样本集  $S$  中第  $i$  个样本的第  $j$  个特征的值, 由于不同特征量的数值差异较大, 为了减少不同量纲对变压器故障诊断结果的负面影响, 诊断前先按式 (9) 对各特征量进行归一化处理。

$$x_{\text{new}} = L_0 + \frac{x - x_{\min}}{x_{\max} - x_{\min}}(U_p - L_0) \quad (9)$$

其中  $x_{\min} = \min_{i=1, \dots, N} \{V_{ij}\}$ ,  $x_{\max} = \max_{i=1, \dots, N} \{V_{ij}\}$ ,  $j = 1, \dots, 10$ ,  $x_{\text{new}}$  为归一化后的  $V_{ij}$ ,  $U_p$ 、 $L_0$  分别为  $V_{ij}$  归一化的上界和下界, 本文取 1 和 -1。变压器的状态分为局部放电、低能放电、高能放电、中低温过热、高温过热及正常状态等, 分别记为 PD、D<sub>1</sub>、D<sub>2</sub>、T<sub>12</sub>、T<sub>3</sub> 与 Norm。为便于后续算法描述, 亦记为  $F_1, \dots, F_6$ 。

表 1 变压器 DGA 数据记录

No	CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>2</sub>	H <sub>2</sub>	CO	CO <sub>2</sub>	T
1	128.69	48.37	105.98	0	45.85	451.36	1936.76	T <sub>12</sub>
2	135.24	41.74	105.23	0	36.49	585.03	1822.33	T <sub>12</sub>
...	...	...	...	...	...	...	...	...

#### 3.2 样本选取

样本集  $S$  是影响基础谱聚类对  $s_0$  识别的重要因素, 为了生成多样性的基础谱聚类, 设计了加权二次抽样算法 (Weighted Double Samplings, WDS), 用于构造多样性的样本集。WDS 的具体如算法 2 所示。

##### 算法 2 加权二次抽样算法 (Weighted Double Samplings, WDS)

输入: 类别已知的 DGA 数据集  $Y$ , 类别已知样本子集  $S$  的规模  $N$

输出: 类别已知的样本子集  $S'$

步骤 1: 将  $Y$  按变压器的 6 种状态划分成子集  $Y_1, Y_2, \dots, Y_6$ , 为子集  $Y_i$  中的每个样本指定权重  $W_{ij}$ ,  $j = 1, 2, \dots, |Y_i|$ , 权重与子集  $Y_i$  的几何中心和样本之间的距离反相关,  $Y_i$  中全体样本的权重之和如式 (10)。

$$W_i = \sum_{j=1, 2, \dots, |Y_i|} W_{ij} \quad (10)$$

步骤 2: 抽取  $N$  个状态已知的样本。

for  $i = 1: N$

1、一次抽样: 确定选取类别;

产生一个  $[0, 1)$  之间的随机数  $r_1$ , 如果  $r_1 \in [(i-1)/L, i/L)$ , 则从子集  $Y_i$  中选取, 其中  $L$  为变压器状态数,  $i = 1, \dots, L$ ;

2、二次抽样: 确定选取样本;

产生一个  $[0, 1)$  之间的随机数  $r_2$ ;

用  $r_2 = r_2 \times W_i$  更新  $r_2$  的值;

$j = 1$ ;

while ( $r_2 > 0$ )

$r_2 = r_2 - W_{ij}$ ;

$j++$ ;

end while

选取  $Y_i$  的第  $j-1$  个样本加入样本集  $S'$ ;

end for

步骤 3: 输出类别已知的样本子集  $S'$ 。

WDS 通过给类别中心的样本指派更大的权重使得离类别中心更近的样本具有更大的概率参与基础谱聚

类的构造;通过一次抽样确保样本子集  $S'$  中各种状态的样本数目相近;通过二次抽样实现样本子集  $S'$  的多样性.

### 3.3 参数优化

TOFD-SCE 的性能与调节因子  $\lambda$ 、集成成员数目  $d'$ 、尺度参数  $\sigma$ 、样本集规模  $N$  及样本集  $S$  的实例相关. 尺度参数  $\sigma$  与样本实例用于生成多样性的基础谱聚类. 参照文献 [13]  $\sigma$  取区间  $[0, 1]$  上的随机数;样本实例由 WDS 算法选取. 调节因子  $\lambda$ 、集成成员数目  $d'$  及样本集  $S$  的规模  $N$  是影响 TOFD-SCE 整体性能的 3 个因素. K 折交叉验证 (K-CV) 可以保证所有样本都参与模型的训练与验证. 粒子群优化算法 (PSO) 模型简单易于实现, 收敛速度快<sup>[16]</sup>, 鉴于此提出了基于 K-CV 与 PSO 的参数优化算法. 具体步骤如算法 3 所示.

#### 算法 3 基于 K-CV 与 PSO 的参数优化算法

输入: 类别已知的样本集  $Z$ , 参数  $K$

输出:  $(\lambda, d', N)$  的最优解

步骤 1: 将样本集  $Z$  随机分成样本数量近似的  $K$  个独立的子集  $Z_1, \dots, Z_K$ .

步骤 2:  $Z - Z_k$  作为状态已知的样本集,  $Z_k$  作为验证集, 得到  $K$  个 TOFD-SCE 模型及其在相应验证集上的判正率;

步骤 3: 计算  $K$  个模型的平均判正率  $AC$ , 如式 (11);

$$AC = \frac{1}{K} \sum_{k=1}^K \frac{1}{|Z_k|} \sum_{x_i \in Z_k} \theta(I((Z - Z_k), x_i), t_i) \quad (11)$$

其中  $I((Z - Z_k), x_i)$  表示以  $Z - Z_k$  作为状态已知样本集, TOFD-SCE 对  $x_i$  的分类结果;  $\theta$  为标识函数, 如式 (12)

$$\theta(x, x') = \begin{cases} 1, & \text{if } x = x' \\ 0, & \text{else} \end{cases} \quad (12)$$

步骤 4: 用 PSO 优化 TOFD-SCE 的参数  $(\lambda, d', N)$ . 种群规模设为 30, 最大迭代次数为 200, 适应度函数为  $1 - AC$ , 以连续多次迭代最优适应度函数的差小于给定阈值或达到最大迭代次数为终止条件. 利用均匀分布随机数生成初始种群.  $\lambda$  取  $[0, 1]$  间的随机数,  $d'$  随机取  $[1, 100]$  间的整数,  $N$  随机取  $[1, 500]$  上的整数. 本文以连续 10 次迭代最优适应度函数之间的差小于 0.0001 或迭代次数达到 200 为终止条件.

### 3.4 变压器故障诊断步骤

TOFD-SCE 的流程如图 1, 包括 6 个步骤:

- 步骤 1: DGA 数据预处理;
- 步骤 2: 基于 K-CV 与 PSO 的 TOFD-SCE 参数优化;
- 步骤 3: 具有多样性的基础谱聚类分类器生产;
- 步骤 4: 从基础谱聚类中选择谱聚类集成成员;
- 步骤 5: 谱聚类集成分类器模型构造;
- 步骤 6: 应用谱聚类集成分类器诊断变压器故障.

## 4 实验及其结果分析

### 4.1 实验介绍

实验的 DGA 数据包括两部分: 一部分是由某省电

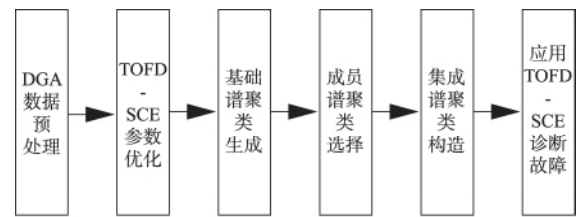


图1 TOFD-SCE模型流程图

力科学院提供的实测数据, 另一部分来源于公开发表的相关文献. 经整理共获得 2190 组有效数据, 其中正常、局部放电、低能放电、高能放电、中低温过热、高温过热的 DGA 记录别为 981、334、321、187、205 与 162 组. 各种类型的 DGA 数据均随机抽取 25 组, 作为测试集  $X$ , 其余 DGA 数据作为类别已知的样本集  $Z$ , 将  $Z$  划分成 10 个子集  $Z_1, \dots, Z_{10}$ , 子集之间同类别的样本数目相近. 基于 10 折交叉验证与 PSO 优化 TOFD-SCE 的参数  $(\lambda, d', N)$ , 共试验 30 次. 限于篇幅, 图 2 列出了其中 5 次试验的适应度函数变化曲线.

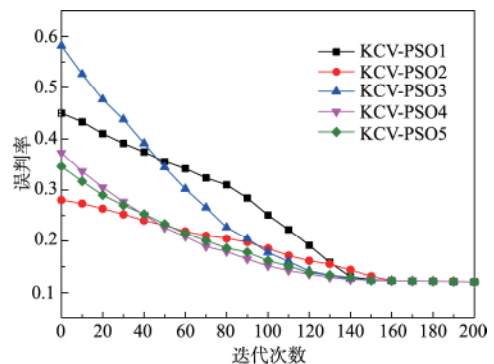


图2 TOFD-SCE参数优化试验

由于粒子采用随机初始化, 初始阶段各次实验的适应度函数差异较大, 随着迭代次数的增加, 适应度函数逐步减小, 并逐步收敛于最优解. 虽然各次试验的收敛速度有差异, 但迭代 150 次后, 都能收敛于 0.0871 附近. 对 30 次试验求均值, 调节因子的均值  $\bar{\lambda} = 0.6473$ , 集成成员数目的均值  $\bar{d}' = 7.1000$ , 基础谱聚类类别已知样本集的规模  $N$  的均值  $\bar{N} = 248.3185$ . 集成成员数目与样本集的规模只能为整数, 对  $\bar{d}'$  与  $\bar{N}$  取整, TOFD-SCE 参数  $(\lambda, d', N)$  的近似最优解为  $(0.6473, 7, 248)$ .

为了验证基于 K-CV 与 PSO 的参数优化算法的有效性, 设定集成成员数目  $d' = 7$ , 样本集的规模  $N = 248$ , 调节因子  $\lambda$  以 0.05 为步长, 从 0 增加到 1 进行试验. 集合  $Z$  为类别已知样本集, 对测试集  $X$  进行测试, 进行 20 次实验, 取平均判正率, 实验结果如图 3.

图 3 显示:  $\lambda = 0$  时, 只考虑了集成成员的差异性, 未考虑正确性, 低质量的集成成员影响了谱聚类集成的整体性能, TOFD-SCE 的判正率较低; 随着  $\lambda$  增大, 综合考虑了谱聚类集成成员的差异性与准确性, TOFD-SCE 的

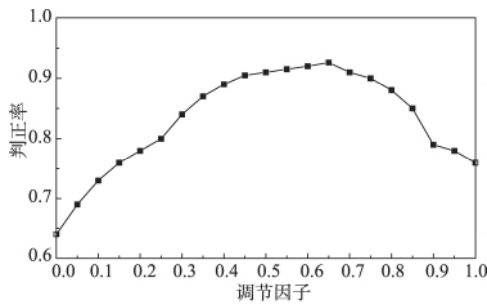


图3 调节因子优化验证实验

判正率逐步提高; 当  $\lambda$  增大到 0.65 时, 判正率达到最大值; 而当  $\lambda > 0.65$  后, 集成成员的差异性成为影响 TOFD-SCE 性能的主要因素, 其准确率开始下降。

设定  $\lambda = 0.6473$ , 将  $d'$  以 1 为步长, 从 1 增大到 15; 类别已知样本集的规模  $N$  以 5 为步长, 从 5 增大到 500, 对集合  $X$  进行测试, 共进行 30 次试验, 对正确率取平均值, 结果如图 4。

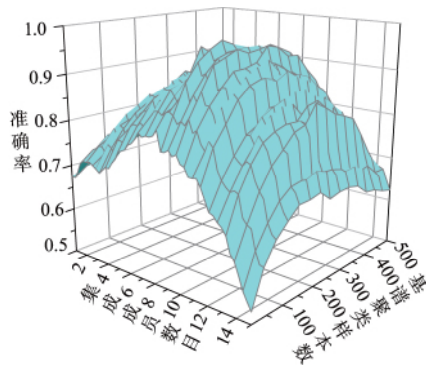


图4 诊断准确率-集成成员数目与样本集规模

实验结果显示: 当集成成员数目较少(尤其当成员分类器数目为 1)时, 谱聚类集成分类器退化为单体谱聚类分类器, TOFD-SCE 的准确率较低, 且稳定性较差; 随着集成成员数目与样本集规模的增大, TOFD-SCE 的准确率有逐步提升的趋势, 偶尔准确率较高, 但稳定性较差; 当  $N = 245$ 、 $d' = 7$  时, TOFD-SCE 的准确率稳定在较高的水平, 与基于 K-CV 与 PSO 的参数优化结果基本一致, 验证了基于 K-CV 与 PSO 的参数优化算法的有效性。

为了比较集成谱聚类与单体谱聚类的故障诊断效果, 对 TOFD-SCE 与基于单体谱聚类的故障诊断结果进行了统计分析, 统计结果如图 5。

图 5 中 TOFD-SCE 的故障诊断准确率高于最好的成员谱聚类分类器, 且稳定性较好。说明谱聚类集成比单体谱聚类更合适于基于 DGA 的变压器故障诊断, 准确率更高、稳定性更好。

设定  $d' = 7$ ,  $N = 248$ ,  $\lambda = 0.6473$ , 以集合  $Z$  为类别已知的样本集, 用 TOFD-SCE 诊断文献 [9] 中 6 条 DGA 数据的实验结果如表 2。

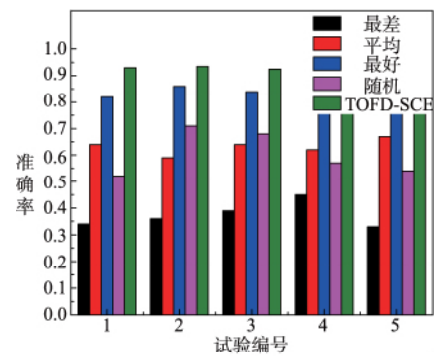


图5 谱聚类集成与单体谱聚类对比分析

表 2 的第 1 组数据显示: 其为正常状态的概率为 0.7829, 低能放电概率为 0.0857, 高能放低概率为 0.0600, 中低温过热概率为 0.0286, 高温过热为 0, 局部放电概率为 0.0429, 其中正常状态的概率远大于其他状态的概率, 因此判断该组数据的综合诊断结果为正常状态。TOFD-SCE 对 6 组数据的诊断结果均与实际情况一致, 验证了其有效性。此外, TOFD-SCE 还给出各种状态可能的概率, 可用于指导检修人员的工作, 提高其工作效率。需要指出的是: 基础谱聚类样本集是通过随机选取样本构成, 具有不稳定性, 各次实验判断变压器所属状态的概率有所波动, 但波动范围较小, 不影响最终诊断结果, TOFD-SCE 的准确率较为稳定。

表 2 TOFD-SCE 变压器故障诊断结果

序号	1	2	3	4	5	6
H <sub>2</sub>	80	115.9	673.6	166.6	70.4	260
CH <sub>4</sub>	10	75	423.5	28.31	69.5	8
C <sub>2</sub> H <sub>6</sub>	4	14.7	77.5	6.72	28.9	2.5
C <sub>2</sub> H <sub>4</sub>	1.5	25.3	988.9	12.4	241.2	2
C <sub>2</sub> H <sub>2</sub>	0	6.8	344.4	0.3	10.4	0
CO	438.17	308.85	456.76	437.63	395.85	365.91
CO <sub>2</sub>	1112.4	1121.89	1860.05	2239.33	1437.34	3634.89
Norm	0.7829	0	0	0.0526	0.0091	0.0394
D <sub>1</sub>	0.0857	0.6859	0.0161	0.0877	0.0030	0.1457
D <sub>2</sub>	0.0600	0.1614	0.8226	0.0772	0.0122	0.1260
T <sub>12</sub>	0.0286	0.0317	0	0.6912	0.0638	0.0472
T <sub>3</sub>	0	0.0720	0	0.0526	0.9058	0.6181
PD	0.0429	0.0490	0.1613	0.0386	0.0061	0.0236
综合	Norm	D <sub>1</sub>	D <sub>2</sub>	T <sub>12</sub>	T <sub>3</sub>	PD

#### 4.2 实验对比分析

为了验证 TOFD-SCE 方法的优越性, 进行了与三比值法、BPNN 法及 SVM 法的变压器故障诊断对比实验。文中三比值法采用 IEC60599 中的三比值方法; BPNN 法参考文献 [9], 输入层神经元数为 5, 隐含层神经元数为 15, 输出层神经元数为 6, 隐含层激励函数选用

tansig 输出层激励函数选用 purelin, 以 trainlm 为训练函数, 以 learnngdm 为权重学习函数, 以均方误差 (Mean Square Error, MSE) 为性能评价指标; SVM 核函数选用 RBF, 采用“一对一”的分类方法, 应用网格搜索与 10 折交叉验证优化正则化系数与核函数参数; TOFD-SCE 的集成成员数目  $d' = 7$ , 基础谱聚类样本集的规模  $N = 248$ , 调节因子  $\lambda = 0.6473$ . 不同方法对文献 [9] 的 6 组 DGA 数据的诊断结果如表 3. 表 3 的第一条记录用 SVM 方法诊断的结论为局部放电故障, 实际没有故障, 属于正常状况, TOFD-SCE 诊断的结论与实际情况一致. 表

表 3 各方法变压器故障诊断结果对比

实际状态	DGA 数据( $\mu\text{L/L}$ )							诊断结果			
	H <sub>2</sub>	CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>2</sub>	CO	CO <sub>2</sub>	TR	BP	SVM	TOFD-SCE
Norm	80	10	4	1.5	0	438.17	1112.4	Norm	Norm	PD	Norm
D <sub>1</sub>	115.9	75	14.7	25.3	6.8	308.85	1121.9	D <sub>2</sub>	D <sub>1</sub>	D <sub>1</sub>	D <sub>1</sub>
D <sub>2</sub>	673.6	423.5	77.5	988.9	344.4	456.76	1860.1	D <sub>2</sub>	D <sub>2</sub>	D <sub>2</sub>	D <sub>2</sub>
T <sub>12</sub>	166.6	28.31	6.72	12.4	0.3	437.63	2239.3	U	T <sub>3</sub>	T <sub>12</sub>	T <sub>12</sub>
T <sub>3</sub>	70.4	69.5	28.9	241.2	10.4	395.85	1437.3	T <sub>3</sub>	D <sub>2</sub>	T <sub>3</sub>	T <sub>3</sub>
PD	260	8	2.5	2	0	365.91	3634.9	PD	PD	PD	PD

表 4 正确率对比

实际状态	测试样本	IEC 三比值		BPNN		SVM		TOFD-SCE	
		数量	正确率	数量	正确率	数量	正确率	数量	正确率
Norm	25	15	0.6	19	0.76	20	0.80	21	0.84
D <sub>1</sub>	25	13	0.52	21	0.84	23	0.92	24	0.96
D <sub>2</sub>	25	20	0.8	22	0.88	22	0.88	22	0.88
T <sub>12</sub>	25	21	0.84	23	0.92	23	0.92	25	1
T <sub>3</sub>	25	14	0.56	21	0.84	21	0.84	23	0.92
PD	25	17	0.68	21	0.84	22	0.88	24	0.96
合计	150	100	0.6667	127	0.8467	131	0.8733	139	0.9267

表 5 列出了多种变压器故障诊断方法的功能和性能, 其中 TOFD-SCE 的准确率较高, 可给出各类故障的概率, 可分析多故障同时发生的情况.

表 5 功能性能比较

方法 \ 属性	准确率	概率	多故障
三比值	0.6667	NO	NO
BPNN	0.8467	NO	NO
SVM	0.8733	NO	NO
TOFD-SCE	0.9267	YES	YES

## 5 结论

(1) 提出了一种谱聚类集成分类器模型, 其以基础

谱聚类学习问题的局部知识, 平衡多样性与准确性选择集成成员, 集成多个集成成员的聚类结果, 以提高谱聚类集成分类器的整体性能.

(2) 研究了影响 TOFD-SCE 性能的因素, 通过随机选取样本实例与尺度参数生成多样性的基础谱聚类, 基于 K-CV 与 PSO 优化其它参数, 以提高 TOFD-SCE 的性能.

(3) TOFD-SCE 利用历史与在线新增两种 DGA 数据构建、修正模型, 以此来提高变压器故障诊断的准确率, 实验结果表明, TOFD-SCE 方法优于三比值法、BPNN 法及 SVM 法.

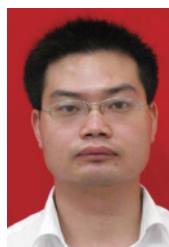
## 参考文献

- [1] Norazhar Anu BAKAR, ABU-SIADA A, ISLAM S. A review of dissolved gas analysis measurement and interpreta-



- tion techniques [J]. IEEE Electrical Insulation Magazine, 2014, 30(3): 39–49.
- [2] LIU Zhi-xiong, SONG Bin, LI En-wen, et al. Study of “code absence” in the IEC three-ratio method of dissolved gas analysis [J]. IEEE Electrical Insulation Magazine, 2015, 31(6): 6–12.
- [3] IEC60599. Mineral Oil-Impregnated Electrical Equipment in Service Guide to the Interpretation of Dissolved and Free Gases Analysis [S].
- [4] Vinicius G M CRUZ, Andre L H COSTA, Marcio L L PAREDES. Development and evaluation of a new DGA diagnostic method based on thermodynamics fundamentals [J]. IEEE Transactions on Dielectrics and Electrical Insulation 2015, 22(2): 888–894.
- [5] 廖瑞金, 姚陈果, 等. 多专家合作诊断变压器绝缘故障的黑板型专家系统 [J]. 电工技术学报, 2002, 17(1): 85–90.
- LIAO Ruijin, YAO Chenguo, et al. Blackboard expert system of multi-expert cooperating diagnosis in transformers' insulation [J]. Transactions of China Electrotechnical Society 2002, 17(1): 85–90. (in Chinese).
- [6] Vladimiro MIRANDA, et al. Diagnosing faults in power transformers with autoassociative neural networks and mean shift [J]. IEEE Transactions on Power Delivery 2012, 27(3): 1350–1356.
- [7] ABU-SIADA A, HMOOD S, ISLAM S. A new fuzzy logic approach for consistent interpretation of dissolved gas-in-oil analysis [J]. IEEE Transactions on Dielectrics and Electrical Insulation 2013, 20(6): 2343–2349.
- [8] WEI Chenghao, TANG Wenhua, WU Qinghua. Dissolved gas analysis method based on novel feature prioritization and support vector machine [J]. IET Electric Power Applications 2014, 8(8): 320–328.
- [9] 朱永利, 尹金良. 组合核相关向量机在电力变压器故障诊断中的应用研究 [J]. 中国电机工程学报, 2013, 33(22): 68–74.
- ZHU Yongli, YIN Jinliang. Study on application of multi-kernel learning relevance vector machines in fault diagnosis of power transformers [J]. Proceedings of the CSEE 2013, 33(22): 68–74. (in Chinese)
- [10] Shakeb A KHAM, Md Danish EQUBAL, Tarikul ISLAM. A comprehensive comparative study of DGA based transformer fault diagnosis using fuzzy logic and ANFIS models [J]. IEEE Transactions on Dielectrics and Electrical Insulation 2015, 22(1): 590–596.
- [11] YANG Ming-Ta, HU Li-Siang. Intelligent fault types diagnostic system for dissolved gas analysis of oil-immersed power transformer [J]. IEEE Transactions on Dielectrics and Electrical Insulation 2013, 60(6): 2317–2324.
- [12] Hiba ZBIB, Sandrine MOUYSSET, Simon STUTE, et al. Unsupervised spectral clustering for segmentation of dynamic PET images [J]. IEEE Transactions on Nuclear Science 2015, 28(5): 830–853.
- [13] 邓晓政, 焦李成, 卢山. 基于非负矩阵分解的谱聚类集成 SAR 图像分割 [J]. 电子学报, 2011, 39(12): 2905–2909.
- DENG Xiao-zheng, JIAO Li-cheng, LU Shan. Spectral clustering ensemble applied to SAR image segmentation using nonnegative matrix factorization [J]. Acta Electronica Sinica 2011, 39(12): 2905–2909. (in Chinese)
- [14] 孔万增, 孙志海, 等. 基于本征间隙与正交特征向量的自动谱聚类 [J]. 电子学报, 2010, 38(8): 1880–1891.
- KONG Wan-zeng, SUN Zhi-hai, et al. Automatic spectral clustering based on eigengap and orthogonal eigenvector [J]. Acta Electronica Sinica 2010, 38(8): 1880–1891. (in Chinese)
- [15] 于重重, 田蕊, 谭励, 等. 非平衡样本分类的集成迁移学习算法 [J]. 电子学报, 2012, 40(7): 1358–1363.
- YU Chong-chong, TIAN Rui, TAN Li, et al. Integrated transfer learning algorithm for unbalanced samples classification [J]. Acta Electronica Sinica 2012, 40(7): 1358–1363. (in Chinese)
- [16] 夏学文, 桂凌, 戴志锋. 基于多尺度选择性学习和探测收缩机制的 PSO 算法 [J]. 电子学报, 2016, 5(1): 1090–1100.
- XIA Xue-wen, GUI Ling, DAI Zhi-feng, et al. A PSO algorithm based on multiscale-selective-learning and detecting-shrinking strategies [J]. Acta Electronica Sinica, 2016, 5(1): 1090–1100. (in Chinese)

## 作者简介



刘荣胜 男, 1981 年生, 湖南娄底人, 现为湖南大学博士研究生, 主要研究方向为电力系统故障分析及其设备状态监测。  
E-mail: 258920025@qq.com



彭敏放 (通讯作者) 女, 1964 年生, 湖南湘乡人, 2006 年博士毕业于湖南大学, 现为湖南大学教授、博士生导师, 主要研究方向为电路测试与诊断、电力系统故障分析及其设备状态监测等。  
E-mail: lrs0623@163.com