「控制系统」

变压器油中溶解气体组合预测方案及实例验证

文/国网天津市电力公司检修公司　赵　颖

为准确地预测变压器油中溶解气体浓度的变化趋势，并运用预测结果为变压器状态检修提供更好的指导，提出了一种变压器油中溶解气体含量组合预测方案优选方法。

变压器内部存在潜伏性故障时，故障点的热能将会促使绝缘油中更多的烃类键断裂，进而产生大量的低分子烃类气体和氢气，且产气速率相较于正常情况更快。随着故障的发展，油中溶解气体也逐渐累积。通过分析气体含量变化规律并预测其发展趋势，变压器内部存在的缺陷就可以在故障初发期甚至故障发展之前被确定，继而可以提前进行检修安排。

高精度的预测方法是准确把握变压器状态发展规律的基础。已有学者开展了油气含量预测研究工作，提出将多种单一预测方法进行组合，以期实现中和预测误差、减小误差波动范围、提高预测精度的目标。但是，目前组合预测方法对各单一预测方法权重的确定主要依赖于专家经验，这种方式过于主观，甚至会出现权重系数为负的情况。同时，这些组合方法均是针对某一特定应用场景、以固定的几种单一预测方法进行组合，方法的普适性有待提高。

针对现有方法存在的不足以及现场实际应用的需求，本文提出一种变压器油中溶解气体含量组合预测优选方法。首先对单一预测方法的平均预测误差进行扫描，其次采用交叉熵方法对不同的单一预测方法进行赋权，然后对可能存在的组合方式的平均预测误差进行扫描，最后将预测误差最小的组合方式作为最终的组合方案。

基于交叉熵的组合预测优选方法

1. 预测方法评价指标

不同的预测方法揭示油中气体含量变化规律的能力主要通过其预测准确度进行体现，因此，本文采用预测的平均误差作为衡量每种预测方法以及组合预测方式的指标，定义为

$$\bar{\delta}_i = \frac{1}{n}\sum_{j=1}^{n}\frac{\left|\hat{y}_{ij} - \bar{y}_{ij}\right|}{\hat{y}_{ij}} \times 100\% \qquad (1)$$

式中，$\bar{\delta}_i$表示第i种方法对于训练样本数据进行预测的平均误差；$\hat{y}_{ij}$为实测值；$\bar{y}_{ij}$为预测值；n为训练样本的数量。

2. 组合预测优选方法的原理

由于不同预测方法各自的适用条件不同，在不同的应用场景下其预测精度和鲁棒性往往也不相同，单纯地利用单一预测方法进行预测往往带有一定的片面性和不稳定性，但是这些方法在数据处理及应用方面均有各自的特点和一定的可取之处。因此，本文提出了一种基于交叉熵的组合预测优选方法，具体实施流程如图1所示。

首先，收集目标变压器的油色谱分析数据，选定n种算法，分别采用这n种算法对所收集的油中气体含量数据进行预测，求得每种方法的平均误差（共n个值）。其次，采用交叉熵方法分别对可能的k种组合方式中的各单一预测方法进行权重系数计算（$k = \sum_{i=1}^{n} C_n^i - n$），然后采用不同的组

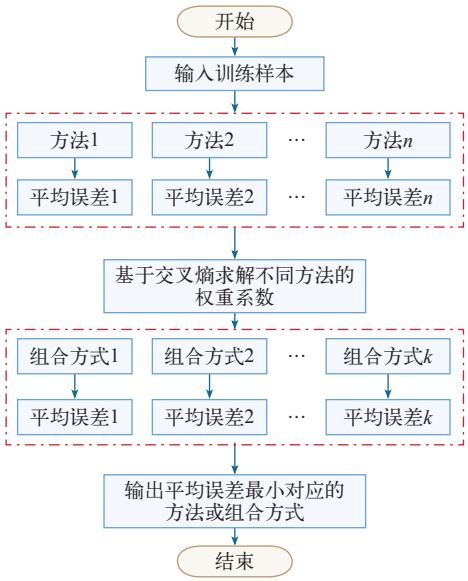

图1　基于交叉熵的组合预测优选方法

合预测方式对所收集的油中气体含量数据进行预测，求得每种方法的平均误差（共k个值）。最后对$n+k$个平均误差进行排序，将平均误差最小对应的方法或组合方式作为最终的输出。

需要说明的是，由于本文所提出的方法涉及对多种方法的组合，如果所选定的方法过多将会影响组合预测优选方案的输出。因此，本文建议进行组合预测的单一预测方法不应超过5种，同时当某种预测方法的平均预测误差大于其余方法平均预测误差中最大值的2倍，即满足式（2），则将该方法从组合预测中剔除

$$\left|\bar{\delta}_i\right| \geqslant 2\max\left(\left|\bar{\delta}_1\right|,\left|\bar{\delta}_2\right|,\cdots,\left|\bar{\delta}_{i-1}\right|,\left|\bar{\delta}_{i+1}\right|,\cdots,\left|\bar{\delta}_n\right|\right) \quad (2)$$

式中，表示第i种方法预测误差的绝对值。

3. 基于交叉熵的权重系数计算

（1）基本原理

交叉熵是一种全局优化方法，根据概率分布密度函数生成样本解，样本解的整体优劣与概率分布形式及参数有关。该方法并不要求每次生成的解都优于前次迭代的解。有一定的概率生成较差的解，只是在迭代的过程中生成较好解的概率越来越大，生成较差解的概率越来越小。交叉熵方法已在很多领域的优化问题中得到广泛的应用，但是，在电力系统领域，特别是主设备保护状态监测领域，还鲜见相关文献。

对于离散和连续信源在同一概率空间的两 个测度f和g的交叉熵$D(f\|g)$分别定义，离散情况下

$$D(f\|g)=\sum_{i=1}^{n} f_i \log_c\left(\frac{f_i}{g_i}\right) \quad (3)$$

连续情况下

$$D(f\|g)=\int_a^b f(x)\log_c\frac{f(x)}{g(x)}\mathrm{d}x \quad (4)$$

式中，f和g在离散情况下表示概率矢量，连续情况下表示概率密度函数。$D(f\|g)\geqslant 0$，当且仅当$f=g$时等号成立，且$D(f\|g)\neq D(g\|f)$。

（2）求解步骤

步骤1：对历史时刻进行虚拟预测

选择运行年限T内的m个历史时刻，分别用选定的n种算法进行气体含量的虚拟预测，每种方法得到m个气体含量值P_m，形成气体含量预测序列P_{nm}。同时，将这m个历史时刻所测量的实测值组成实测值序列R_m。

步骤2：定义油中溶解气体含量概率密度函数

假设运行年限T内预测结果的误差满足正态分布。$f_n(x)$表示第n种算法所得的油色谱数据预测误差概率密度函数，其满足

$$f_n(x)=\frac{1}{\sqrt{2\pi}\sigma_n}e^{-(x-\mu_n)^2/(2\sigma_n^2)} \quad (5)$$

式中，μ_n为平均值；σ_n为方差。对于第n种算法，将其对历史数据的虚拟预测值P_{nm}与实测值R_m做差得到m个历史时刻的预测误差值$\Delta_{nj}(j=1,2,\cdots,m)$，求其均值 作为式（5）中$\mu_n$。此外，方差$\sigma_n$可以通过式（6）进行计算

$$\sigma_n=\frac{1}{m}\sum_{j=1}^{m}(\Delta_{nj}-\bar{\Delta}_n)^2 \quad (6)$$

再令$f(x)$为融合各影响因素后的油中溶解气体含量预测误差的概率分布密度函数，其满足

$$f(x)=\sum_{i=1}^{n} w_i f_i(x) \quad (7)$$

$$\sum_{i=1}^{n} w_i=1 \quad (8)$$

步骤3：建立支持矢量，确定交叉熵目标函数

支持矢量$\boldsymbol{S}$

$$\boldsymbol{S}=[\boldsymbol{S}_1,\boldsymbol{S}_2,\cdots,\boldsymbol{S}_n] \quad (9)$$

设

$$\begin{cases}\boldsymbol{S}_i=D[f(x)\|f_i(x)]=\int f(x)\ln[f(x)/f_i(x)]\mathrm{d}x \\ \text{s.t.}\quad \int f(x)\mathrm{d}x=1\end{cases} \quad (10)$$

通常两种分布相交的程度越大，其相互支持程度越高，$\boldsymbol{S}_i$越小，因此有

$$A_i=\boldsymbol{S}_i/\sum_{i=1}^{n}\boldsymbol{S}_i \quad (11)$$

为了用权重系数来反映不同方法之间的支持

程度，令

$$w_i = 1/A_i \Big/ \sum_{i=1}^{n} 1/A_i \tag{12}$$

建立最小交叉熵优化的目标函数

$$F = \min\left(\sum_{i=1}^{n} D[f(x)\|f_i(x)]\right) \tag{13}$$

步骤4：求解最小交叉熵优化问题

目标函数F是关于权重系数w_i的函数，因此其优化问题可以转化为一个对F求极小值的非线性规划问题，利用非线性规划方法即可解出。

交叉熵方法通过求解不同输入量信息差异程度，进而求得对应权重系数，避免了主观因素对结果的影响，具有极强的客观性。同时，所求解出的权重系数并非固定值，当预测的历史时刻变化时（即训练样本发生了变化），权重系数也会动态变化。因此组合预测方法能够通过优化调整权重系数的分配情况来合理地规避单一预测方法误差误差过大的问题。

4. 非等间隔数列处理方法

一般情况下变压器油色谱在线监测是严格按照采样周期（一般为2天）进行的，所以数据均为等间隔的。当通信故障或采样设备故障时，采样数据序列将出现部分缺失，导致数据点非等时间间隔，为了保证本文所提方法在此情况下也能正常进行预测，本文将在进行预测前对非等间隔数据进行预处理。

假设实测值组成的数列为：$x_i=\{x_1,x_2,\cdots,x_n\}$，其对应的测量时间为：$t_i=\{t_1,t_2,\cdots,t_n\}$，时间间隔数列为：$\Delta t_{i+1}=\{\Delta t_2,\Delta t_3,\cdots,\Delta t_n\}$，其中$\Delta t_{i+1}=t_{i+1}-t_i$。则平均时间间隔$\Delta t$为

$$\Delta t = \frac{t_n - t_1}{n-1} \tag{14}$$

实际观测时间间隔与平均时间间隔的偏差系数μ_i为

$$\mu_i = \Delta t - \Delta t_i, i = 2,3,\cdots,N \tag{15}$$

则修正后的等时间间隔实测值数列应为

$$\tilde{x}_{i+1} = x_{i+1} + \frac{\mu_{i+1}}{\Delta t_{i+1}}(x_{i+1} - x_i) \tag{16}$$

其中，$\tilde{x}_1 = x_1$。由此组成间隔为Δt的等间隔数列$\tilde{x}_i=\{\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_n\}$，继而可以进行下一步预测。

5. 预测结果的可信度

由于每次预测结果只是给出具体的数值，结果太过于绝对，不利于实际应用，因此本文给出了预测结果置信区间的描述。

设$y=[y_1, y_2, \cdots, y_{m-1}, y_m]$为前$m$个时刻气体浓度值，$\tilde{y}_{m+i}$为第$m+i$时刻气体浓度预测值

$$\tilde{y}_{m+i} = g(y_m) + \varepsilon_i \tag{17}$$

式中，$g(y_m)$表示组合预测方法的输出；ε_i为预测误差。若ε_i满足正态分布$N(\mu, \sigma)$，即

$$t = \frac{\sqrt{n}(\bar{\varepsilon} - \mu)}{S} \sim t(n-1) \tag{18}$$

式中，$\bar{\varepsilon} = \sum_{i=1}^{n} \varepsilon_i / n$，$S^2 = \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})^2 / (n-1)$为$\sigma_2$的无偏估计。

使用枢轴量法，可得在（$1-\alpha$）水平下μ的置信区间为$[\bar{\varepsilon} - t_{1-\alpha}(n-1)S/\sqrt{n}, \bar{\varepsilon} + t_{1-\alpha}(n-1)S/\sqrt{n}]$。结合（18）可得$m+i$时刻气体浓度预测区间为：$[g(y_m)+\bar{\varepsilon} - t_{1-\alpha}(n-1)S/\sqrt{n}, g(y_m)+\bar{\varepsilon} + t_{1-\alpha}(n-1)S/\sqrt{n}]$。

组合预测的应用方法

通过对华中电网辖区内大量变电站实地调研了解到，运维人员迫切需要一种能够对变电站内设备运行状态变化进行提前预警的工具。因此，为了将预测结果更好地应用于实际生产中，以前文优选方法所确定的组合预测模型为基础，本文提出了基于变预测时间窗的组合预测应用方法，原理如图2所示。

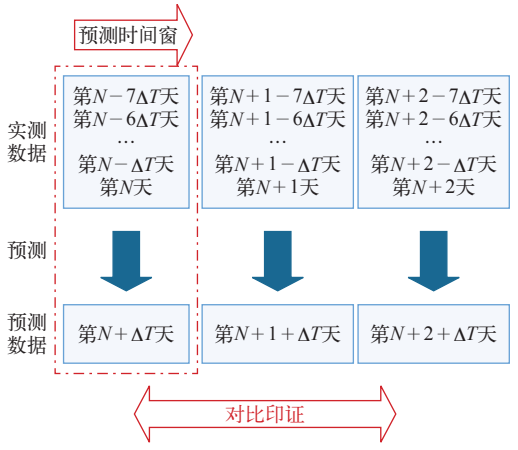

图2　变预测时间窗的组合预测应用方法

具体做法为：假定当前的时标为N，向前追溯7ΔT天的数据（以ΔT为采样间隔）即算法的时间窗为7ΔT，将第$N-7\Delta T$天、第$N-6\Delta T$天、…、第N天的数据填入预测数据窗，输入到组合预测模型中即可得到ΔT天后气体的预测结果（将其记为A），同理可以得到Δ$T+1$天后及Δ$T+2$天后的预测值B和C。如果A的数值超过了注意值，则需观测B和C，若B和C均超过了注意值且呈现增长趋

势，就可以比较可靠地推断该状态量将在ΔT天后发生异常。如果绝大部分的状态量（本文中为油中溶解气体）的预测结果都符合这样的规律，即可以提前ΔT天判断出设备故障的发生。

对于ΔT的取值，本文采用相同的数据样本对ΔT为15天、30天、45天等三种情况下的预测效果分别进行了分析，对比结果如表1所示。

表1　不同ΔT取值下预测结果比较

项目 \ 取值	15天	30天	45天
平均误差（%）	2.37	3.51	6.19
方差	7.26	4.85	3.98

由表中的对比结果可知，预测误差与ΔT的取值成正比关系，即ΔT值越小对应的预测误差也越小。但是，ΔT值越小对应的预测方差却越大，这是因为实际运行过程中电网设备的运行状态总会经历一定幅度的波动，这样的波动也体现在油中气体含量中，当ΔT取值过小时，其所形成的预测序列包含了这一部分的波动数据，将会影响预测效果。因此，综合预测误差和方差的考虑，本文将ΔT取为30天。

借助于变压器油色谱数据的预测结果，运维人员可以提前知晓设备的故障情况，进而有足够的时间从容地安排检修工作，规避应急检修成本，检修的机动性也较高。对预测结果的具体应用模式为：在预测结果揭示设备状态在ΔT天后发生异常，拟提前ΔT_1天（本文中ΔT_1设定为7天，实际应用时根据避免应急检修的需求决定）安排检修计划，在较宽的时间尺度内调集检修资源，安排妥当后进行待命，一旦异常发生即可在第一时间进行处理。如果在ΔT天后并未提示异常，可以根据实际情况继续保持待命状态ΔT_2天（本文暂定为3天，该时间也可根据实际情况进行整定），若仍未发生异常，则取消/解散检修待命。

在预测出设备状态将发生异常时，除了对设备加强巡视以及如上文所述的提前安排检修之外，还可以尝试在系统运行允许的前提下通过改变运行方式，减小目标设备负荷率来规避异常状态的发生，实现流程如图3所示。

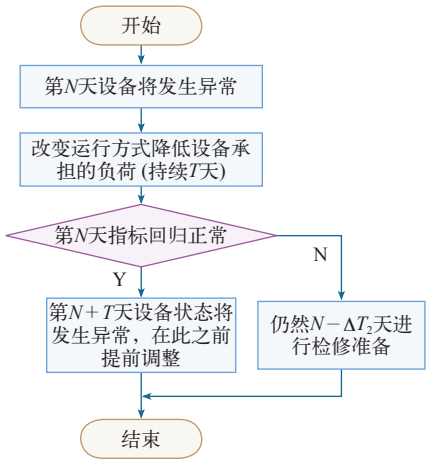

图3　组合预测应用方法的实现流程

组合预测方法的实例

反映变压器状态的油中溶解气体主要有H_2、CH_4、C_2H_2、C_2H_4、C_2H_6、CO以及CO_2。本文采用某两台220 kV主变压器2007—2012年油色谱数据，每台主变有时间间隔为30天的60组数据，前47组为训练样本，后13组为测试样本，且1#主变压器测试样本存在低温过热，2#主变压器测试样本存在高温过热。以H_2为例对所提出的组合预测优选方法进行说明。

首先，选定三种典型的智能预测方法，即径向基函数神经网络（RBFNN）、最小二乘支持矢量机（LSSVM）以及灰色理论（GM），采用充分的训练样本，对以上方法进行扫描并形成组合预测的优选。

由于三种算法存在4种组合方式，即3种两两组合的形式以及三种方法均参与组合的形式。因此由前面所述步骤确定这几种组合形式中算法对应的权重系数，继而得到4种基于交叉熵的组合预测模型。

最后采用两台主变压器的测试样本对不同预测方法的结果进行对比，如表2及表3所示。对比方法包括基于单一算法（RBFNN、LSSVM、GM）的预测方法、3种基于两种算法的组合预测（RBFNN与LSSVM组合，RBFNN与GM组合，GM与LSSVM组合）以及基于三种算法的组合预测，各单一预测方法的预测结果分别以x_1、x_2、x_3表示，各组合预测方法的预测结果分别以y_1、y_2、y_3及y_4表示。

从表2及表3的对比结果可以看出，根据平均误差最小的原则，对于存在低温过热的1#主变压器而言，采用基于RBFNN及LSSVM组合预测的方法平均误差仅为1.85%，能够取得与实测值最为贴近的效果。对于存在高温过热的2#主变压器而言，采用基于RBFNN及GM组合预测的方法为

表2　1#变压器油气含量值（单位：μL/L）

序号	实测值	单一预测方法			组合预测方法			
		x_1	x_2	x_3	y_1	y_2	y_3	y_4
1	10.5	10.6	10.3	11.0	10.4	10.8	10.6	10.6
2	13.7	14.1	14.4	13.1	14.3	13.6	13.8	13.9
3	12.9	12.8	13.3	14.4	13.1	13.6	13.8	13.5
4	17.3	16.2	16.4	15.2	16.3	15.7	15.9	16.0
5	22.4	24.2	21.8	20.6	22.9	22.5	21.3	22.2
6	32.7	31.0	34.2	29.5	32.8	30.3	32.2	31.8
7	48.9	47.1	47.5	44.8	47.3	46.0	46.3	46.6
8	59.1	60.7	56.4	72.2	58.3	66.3	63.2	62.4
9	68.8	69.6	67.3	80.2	68.3	74.8	72.9	71.8
10	77.2	71.1	80.4	81.9	76.3	76.3	81.0	77.9
11	69.3	67.3	65.9	88.6	66.5	77.7	75.7	73.1
12	61.7	64.6	58.9	73.7	61.4	69.0	65.3	65.1
13	63.2	60.0	65.9	74.7	63.3	67.1	69.7	66.7
$\bar{\delta}$(%)		4.21	3.72	5.42	1.85	4.59	4.19	2.85

表3　2#变压器油气含量值（单位：μL/L）

序号	实测值	单一预测方法			组合预测方法			
		x_1	x_2	x_3	y_1	y_2	y_3	y_4
1	13.2	13.3	13.8	12.9	13.6	13.0	13.2	13.2
2	17.3	17.8	16.5	18.0	17.2	17.9	17.5	17.6
3	16.0	15.9	17.9	16.4	16.8	16.2	16.8	16.6
4	21.7	20.4	19.1	20.4	19.8	20.4	20.0	20.1
5	28.1	30.4	25.9	27.0	28.3	28.1	26.6	27.6
6	40.4	38.3	36.5	43.2	37.5	41.6	41.1	40.4
7	60.4	58.2	55.3	59.0	56.8	58.7	57.9	58.0
8	73.9	75.9	90.2	70.6	82.6	72.3	76.5	76.4
9	87.3	88.3	101.7	83.2	94.6	84.9	88.8	88.7
10	96.0	88.4	101.9	100.8	94.7	96.7	101.1	97.9
11	86.9	84.4	111.1	81.7	96.9	82.6	90.6	89.0
12	76.4	80.0	91.3	74.0	85.3	76.0	79.2	79.4
13	79.6	75.6	94.1	82.9	84.3	80.5	86.3	83.6
$\bar{\delta}$(%)		4.03	8.56	4.14	4.52	2.33	3.66	2.58

2.33%，能够取得与实测值最为贴近的效果。这也说明了固定的组合方式并非适用于所有的场合，而本文所提出的组合预测方式优选方法对所有可能存在的组合方式进行充分扫描，能保证在不同的场景下均能选用平均误差最小的方法对油中气体含量变化规律进行预测。

此外，对于智能类算法，有一个指标非常关键，即样本的泛化能力——机器学习算法对新鲜样本的适应能力。

通过上述的算例结果分析可以看出，所有单一算法及组合算法的训练样本均为同一样本集（某两台220 kV主变压器47组2007—2012年的油色谱数据），预测的对象均为此时间区间内的另外24组油色谱指标，对这24组油色谱指标的预测精度高低则直接体现了算法的样本泛化能力的大小。从上面的分析可知，在同一比较基础上，根据本文提出的优选方法所确定的算法的泛化能力最强。进一步地，以1#变压器2013—2015年的36组油色谱监测数据作为新的测试样本对本文所选定的基于RBFNN及LSSVM组合预测方法与其他对比算法进行测试，该算法的平均误差为2.01%，显著优于其他算法，由此也证明本文提出的优选方法所确定的算法具备较强的泛化能力。

以1#主变为例，本文构造了H_2浓度在95%置信度下的预测区间，如图4所示。从图中可以看出，预测时间段内H_2的实测值均落在该区间内。置信水平越高，预测区间包含未来真实数据的概率越大，区间范围更准确，实际应用中应根据实际情况选取最优的置信水平。

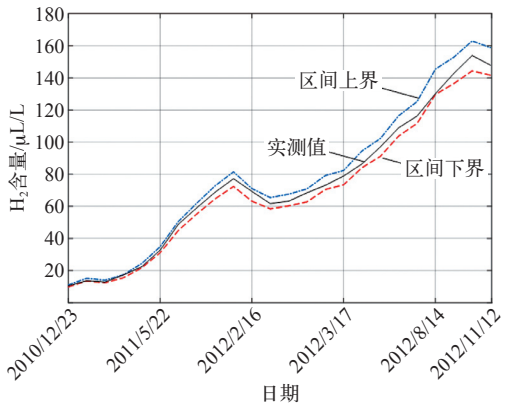

图4　95%置信水平的H_2浓度的预测区间

结束语

为准确地预测变压器油中溶解气体浓度的变化趋势，并运用预测结果为变压器状态检修提供更好的指导，提出了一种变压器油中溶解气体含量组合预测方案优选方法。首先对单一预测方法的平均预测误差进行扫描，其次采用交叉熵方法对不同的单一预测方法进行赋权，然后对所有可能的组合方式的预测平均误差进行扫描，最后将平均误差最小的组合方式作为最终的组合方案。在确定了组合方案后，提出基于可变预测时间窗的预测结果应用方法。基于该方法适当提前安排检修，规避应急检修带来的成本增加和检修不及时带来的潜在损失的增加。**EA**

（收稿日期：2017.05.30）

（编辑：白文亭）