面向电力变压器状态评价的油中溶解气体监测数据补全方法

张若愚[1],齐 波[2],张 鹏[1],李成榕[2]
(1. 华北电力大学 高电压与电磁兼容北京市重点实验室,北京 102206;
2. 华北电力大学 新能源电力系统国家重点实验室,北京 102206)

摘要:基于马尔可夫理论充分考虑相邻时间点系统在所有状态间的转移特性,提出了一种基于马尔可夫模型的变压器油中溶解气体数据补全方法,将油中溶解气体数据时间序列转化为在不同状态间转移的马尔可夫链,利用正、反向的状态转移矩阵计算得到油中溶解气体数据的补全值。从数据挖掘的角度建立了油中溶解气体数据质量的综合评估体系,从多个角度对数据补全的效果进行评估,并基于D-S证据融合理论融合各个角度的评估结果,得到综合评估结果。利用所提方法对某变压器100组油中溶解气体数据中25组随机缺失值进行补全,结果表明补全后的数据与实际值相似度可以达到99.999%。进一步地,验证其中15组极值点、跃变点处缺失数据补全效果,经过综合评估,补全后的数据与实际值相似度可以达到98.956%。经过验证表明所提方法能够在不改变数据特征的前提下对变压器油中溶解气体的缺失值进行准确的补全,有利于提高变压器状态评估方法的准确性。

关键词:电力变压器;状态评价;油中溶解气体;数据补全;马尔可夫过程;综合评估指标
中图分类号:TM 41 **文献标志码:**A DOI:10.16081/j.epae.201910004

0 引言

电力变压器是保证输配电网稳定运行的重要设备,变压器油中溶解气体监测数据是对变压器进行状态评估的重要依据[1-2]。一段时间的监测数据,包含其整体变化趋势、变化中的极值点及跃变点以及数据统计特征,可以从多方面反映变压器内部可能存在的异常状况[3-4],但受到电磁干扰及恶劣的操作环境的影响,会出现变压器中安装的气体传感器故障、通信装置产生数据传输错误、通信中断等情况,导致在线油中溶解气体分析采集的数据含有大量的缺失值及因监测系统故障产生的异常数据。对于异常数据目前常采取删除操作,使得该采样时间点数据缺失[5-7]。数据缺失破坏了在线监测数据时间序列的连续性,一定程度上改变了其数据特征及变化趋势,甚至出现关键的极值点及跃变点缺失的情况。传统的三比值法或者神经网络、支持向量机、灰色系统理论等各种智能诊断分析技术均会因数据缺失而无法对变压器状态进行评价或者因评估依据不符合实际情况,难以对变压器状态进行准确评价,无法充分发挥在线监测系统的作用。

传统的数据补全方法有均值插补法、最近邻插补法、多项式插补法等。均值插补法是利用完整数据的平均值进行插补,最近邻插补法通过距离函数寻找与缺失值距离较近的数值进行插补,多项式插补法是利用多项式对时间序列进行拟合,得到缺失数据的预测值[8]。上述方法未能充分考虑油中溶解气体数据的波动性,对于时间序列的拟合过于粗糙,导致对缺失值的补全精确度不足。近年来,国内外学者在数据补全领域进行了许多研究:华北电力大学的刘沅昆等提出了基于Pearson相关系数与回归模型相结合的缺失数据恢复方法,寻找与缺失值临近数据相关系数最高的历史数据,建立2个时间序列间的一元回归模型,对缺失值进行估计[9];中国电力科学研究院的刁赢龙等提出了利用协同过滤推荐算法,考虑相邻监测点数据变化规律的相似性,根据异常值临近监测点的数据规律对缺失值进行补全[10];湖南大学的毛李帆等提出了利用最小二乘支持向量机,将完整数据作为训练集,将社会经济发展指标作为模型输入,并将负荷值作为输出进行训练,然后根据缺失负荷年份的社会经济指标,对缺失值进行补全[11]。上述方法均需要引入与缺失数据具有较强关联性的其他变量作为辅助参量,间接反映含有缺失值的时间序列数据的变化规律,对需要补全的时间序列数据本身的规律未能够进行充分的挖掘。

马尔可夫过程是描述离散状态时间序列特性的重要方法,马尔可夫过程用转移概率描述系统在不同状态间的变化,给出下一时间点系统处于不同状态的可能,而不是给出某一确定的状态,能够更好地反映系统细微的波动性。

油中溶解气体含量数据随时间在小范围内波动,当前时刻的油中溶解气体含量是一段时间变压

收稿日期:2018-11-06;**修回日期:**2019-08-21

器内气体产生、溶解等过程累积作用的结果，是过去一段时间气体变化情况的综合反映，可认为其未来状态的条件概率分布仅依赖于当前状态，符合马尔可夫特性。通过不同状态间的转移来描述油中溶解气体含量监测数据在数值上的波动，可以将油中溶解气体含量数据时间序列视为马尔可夫过程，将油中溶解气体含量数值的变化转化为其处于不同状态区间的概率的变化，能够更好地分析油中溶解气体含量数据的波动。

本文提出了基于马尔可夫模型的变压器油中溶解气体数据补全方法，对油中溶解气体含量数据进行状态空间划分，对于缺失数据，根据油中溶解气体数据完整部分建立在时间序列上正向和反向的马尔科夫转移矩阵，经过1次或多次状态转移，计算可能的补全值并进行加权求和得到最终的补全值；根据多个指标对原始数据和补全后的数据进行多方面评估，最后运用D-S证据融合理论对各个角度的评估结果进行整合，得到综合评估结果。

1 油中溶解气体监测数据特征

变压器内部发生热故障、放电性故障或者油、纸老化时，会产生多种特征气体，在线监测系统会依照一定的频率采集油中溶解气体数据，作为变压器状态评价的依据。但变压器中的局部放电等故障产生的气体含量具有一定的随机性，而且油中溶解气体还受到变压器的运行环境、油箱内的温湿度、变压器负载等诸多因素的影响，导致变压器油中溶解气体数据并不是一个平滑增长的时间序列，而是随时间在一定范围内反复波动的时间序列。某变压器400 d内的6种特征气体含量的时间序列数据见附录A中的图A1。

传统的补全方法采用整体数据的均值、众数等统计量或某段相似数据作为补全值，忽略了油中溶解气体含量数据随时间的变化，而拟合方法大多采用平滑的曲线对数据进行拟合，同样难以精确描述油中溶解气体含量数据的波动性，在具体数值和时间序列数据特征方面均存在较大的误差。

2 基于马尔可夫过程的油中溶解气体监测数据补全方法

2.1 马尔可夫过程

马尔可夫过程是记忆无后效的过程，即在已知当前状态下，过程的未来状态与它的过去状态无关。由不同状态间的转移概率所组成的矩阵称为状态转移矩阵，马尔可夫状态转移矩阵$\boldsymbol{P}$是一个$n\times n$阶矩阵，如式(1)所示。

$$\boldsymbol{P}=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \tag{1}$$

$$0\leqslant p_{ij}\leqslant 1,\ \sum_{j=1}^{n}p_{ij}=1 \quad 1\leqslant i,j\leqslant n$$

1步转移概率矩阵可用n阶方阵$\boldsymbol{A}=(a_{ij})$描述，其中a_{ij}是从状态i出发，经过1步转移到达状态j的转移概率，称方阵$\boldsymbol{A}$为1步转移概率矩阵。记$\boldsymbol{A}$的m次幂$\boldsymbol{A}^m=(a_{ij}^{(m)})$，有$a_{ij}^{(m)}=p_{ij}^{(m)}$，记$\boldsymbol{A}$的$m$次幂$\boldsymbol{A}^m$的第$i$行、第$j$列元素为从状态$i$出发，经过$m$步转移后到达状态$j$的$m$步转移概率，由此可以用1步转移概率的幂乘积描述多步转移概率。

初始状态确定后，可以利用转移矩阵逐步预测之后的时间序列数据所处的状态。

2.2 数据补全方法

本文提出了基于马尔可夫过程的油中溶解气体监测数据补全方法，首先将油中溶解气体数据划分到不同的状态区间，构建油中溶解气体的马尔可夫链，然后依据完整数据的马尔可夫链，生成正反向状态转移矩阵，将距离缺失值最近的完整数据所处的状态作为初始状态，经过多步状态转移得到可能的补全值，对正反向生成的补全值分别赋予权重，得到最终的补全值。补全方法流程如附录A中的图A2所示。

2.2.1 数据空间划分

马尔可夫过程是依照一定概率分布在离散状态之间转移的过程，对于油中溶解气体含量时间序列数据，需将每个采样时间点的数据转化为对应的状态，构成一个马尔可夫链。取训练集中油中溶解气体数据的最大值$a_{\max}$和最小值$a_{\min}$，将区间$[a_{\min},a_{\max}]$以一定间隔划分为n个状态区间，状态区间间隔的确定，取决于训练集中样本的数量和油中溶解气体含量数据补全所要求的精度，训练集样本越多，精度要求越高，则状态区间的划分越密集。每个区间对应一个油中溶解气体可能处于的状态，这n个状态的集合称为马尔可夫链的状态空间，根据每个采样时间点的油中溶解气体落入的状态区间确定其所属的状态，赋予其状态标签。对于某变压器的油中溶解H_2含量，其长度为100的时间序列数据见图1。

本文以0.1 μL/L为间隔划分为状态空间，为每个采样时间点赋予状态标签，结果如表1所示。

2.2.2 状态转移矩阵生成

在马尔可夫过程中，从状态i转移到状态j的概率为：

$$P_{ij}=s(j|i)/s(i) \tag{2}$$

图1　油中溶解气体状态空间划分

Fig.1　Spatial division of dissolved gas state in oil

表1　数据空间划分

Table 1　Data space partition

区间范围	状态	区间范围	状态
[6.3,6.4]	1	(6.7,6.8]	5
(6.4,6.5]	2	(6.8,6.9]	6
(6.5,6.6]	3	(6.9,7.0]	7
(6.6,6.7]	4	(7.0,7.1]	8

其中，$s(i)$为训练集中状态i出现的次数；$s(j|i)$为训练集中状态i的下一状态为j的次数。计算每个状态之间的转移概率，最终得到$n\times n$阶状态转移矩阵$\boldsymbol{P}$。

对于油中溶解气体含量所处状态构成的时间序列数据，分别统计相邻时间点油中溶解气体含量在各个状态间转移的次数，根据式(2)计算两两状态之间的转移概率，各个状态之间的转移概率共同构成油中溶解气体含量的状态转移矩阵，将状态时间序列倒序排列，根据相同的原理计算倒序状态转移矩阵，从2个方向描述油中溶解气体含量时间序列数据的状态转移规律。

按照上述方法，对某变压器油中溶解H_2时间序列数据分别生成正向状态转移概率矩阵$\boldsymbol{P}$和反向状态转移矩阵$\boldsymbol{Q}$。

$$\boldsymbol{P}=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/8 & 5/8 & 1/8 & 1/8 & 0 & 0 & 0 & 0 \\ 0 & 1/4 & 1/2 & 1/4 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/4 & 1/3 & 1/12 & 0 & 0 & 0 \\ 0 & 0 & 1/25 & 1/25 & 14/25 & 6/25 & 1/25 & 0 \\ 0 & 0 & 0 & 2/23 & 6/23 & 14/23 & 1/23 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/13 & 11/13 & 1/13 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2/3 & 1/3 \end{bmatrix}$$

$$\boldsymbol{Q}=\begin{bmatrix} 3/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5/8 & 3/8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/12 & 1/2 & 1/3 & 1/12 & 0 & 0 & 0 \\ 0 & 1/12 & 1/4 & 1/4 & 1/4 & 1/6 & 0 & 0 \\ 0 & 0 & 0 & 1/6 & 7/12 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/23 & 6/23 & 14/23 & 2/23 & 0 \\ 0 & 0 & 0 & 0 & 1/26 & 1/26 & 11/13 & 1/13 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2/3 & 1/3 \end{bmatrix} \tag{3}$$

以矩阵$\boldsymbol{P}$中的第2行为例，根据表1的区间划分，在训练集数据中，共有8个采样点的数据位于状态2所属区间(6.4,6.5]内，其中1个采样点的下一采样点数据位于状态1所属的区间[6.3,6.4]，5个采样点的下一采样点数据位于状态2所属的区间(6.4,6.5]内，1个采样点的下一采样点数据位于状态3所属的区间(6.5,6.6]内，1个采样点的下一采样点数据位于状态4所属的区间(6.6,6.7]内。

因此，$s(2)=8$，$s(1|2)=1$，$s(2|2)=5$，$s(3|2)=1$，$s(4|2)=1$，则矩阵$\boldsymbol{P}$的第2行元素为：

$$\begin{cases} P_{21}=\dfrac{s(1|2)}{s(2)}=\dfrac{1}{8},\ P_{22}=\dfrac{s(2|2)}{s(2)}=\dfrac{5}{8} \\ P_{23}=\dfrac{s(3|2)}{s(2)}=\dfrac{1}{8},\ P_{24}=\dfrac{s(4|2)}{s(2)}=\dfrac{1}{8} \\ P_{25}=P_{26}=P_{27}=P_{28}=0 \end{cases} \tag{4}$$

2.2.3　权重分配

对于含有缺失值的油中溶解气体监测数据，设第t个采样时间点数据缺失，选择第$t-1$个采样时间点所处的状态i作为初始状态，生成一个长度为n的向量$\boldsymbol{q}_1$，其中第i个元素为1，其余元素均为0，表示对于已知的数据，其所处状态是确定的，概率为1。将初始状态向量$\boldsymbol{q}_1$与正向状态转移矩阵相乘，得到初始状态的下一个采样时间点，即缺失数据采样时间点处于不同状态的概率。

将处于不同状态的概率作为权重，对各状态对应的状态区间的中间值进行加权求和，得到第1个补全值I_1。将第$t+1$个采样时间点所处的状态作为初始状态$\boldsymbol{q}_2$，若第$t+1$个采样时间点数据仍然缺失，则继续向后寻找，直至找到存在完整数据的采样时间点$t=m$，将其所处状态j作为初始状态$\boldsymbol{q}_2$。利用反向状态转移矩阵，经多步状态转移得到第2个补全值I_2。

经正向、反向状态转移生成的2个可能补全值的生成原理相同，因此2个补全值的初始权重均为0.5。当某一状态在训练集中出现的次数越多时，则表明在相应状态转移矩阵生成过程中对该状态下一步可能的状态考虑得越充分，以此状态作为初始状态进行的状态转移可信度越高。因此，I_1和I_2的权重可以根据训练集中状态i和j出现的次数t_i和t_j进行调整。岭型分布的分布函数可以设置某一固定值作为初始权重，根据输入值在初始权重的基础上进行调整，将正向、反向的初始状态在训练集中出现的次数之差代入岭形分布的分布函数。

$$A(x)=\begin{cases} 0 & x\leqslant a_{\min} \\ \dfrac{1}{2}+\dfrac{1}{2}\sin\dfrac{\pi}{a_{\max}-a_{\min}}\times \\ \left(x-\dfrac{a_{\max}-a_{\min}}{2}\right) & a_{\min}<x\leqslant a_{\max} \\ 1 & x>a_{\max} \end{cases} \tag{5}$$

$$\begin{cases} x=t_i-t_j & t_i\geqslant t_j \\ x=t_j-t_i & t_i<t_j \end{cases}$$

则2个I_1和I_2的权值分别为$0.5+A(x)$和$0.5-A(x)$，对I_1和I_2进行加权求和得到最终的补全值。

$$I=\begin{cases} (0.5+A(x))I_1+(0.5-A(x))I_2 & t_i\geqslant t_j \\ (0.5-A(x))I_1+(0.5+A(x))I_2 & t_i<t_j \end{cases} \tag{6}$$

3 综合评估

3.1 评估指标

数据补全是为了消除数据集中的缺失值对变压器状态评估造成的影响，最大限度地还原在线监测数据的原有数据特征，因此，需要选取合适的指标对数据补全的效果进行综合评估。本文选取了5个评估指标，从数值误差、整体统计相似性、频数分布相似性等方面评估对油中溶解气体监测数据的补全效果。

（1）平均相对误差Δ。

平均相对误差是油中溶解气体缺失部分的补全值与实际值之差的绝对值的平均值，反映了补全值在具体数值上的误差。

$$\Delta=\sum_{i=1}^{n}\frac{\left|x_i-x_{i-\mathrm{int}}\right|}{x_i} \tag{7}$$

其中，x_i为实际值；$x_{i-\mathrm{int}}$为补全值。

（2）均值。

均值是反映数据集集中趋势的特征量，通过均值在一定时间段内可以反映油中溶解气体数据的一般水平。

（3）变异系数C_v。

变异系数指数据集的标准差与均值之比，反映了数据集的离散程度，变异系数越大说明数据的离散程度越高。

$$C_v=\sigma/\mu \tag{8}$$

其中，σ为数据集的标准差；μ为数据集的均值或数学期望。

（4）偏度系数S_k。

偏度系数是数据集均值与中位数或众数之差与标准差之比，是数据分布偏离对称性程度的一个特征量，当分布左右对称时，偏度系数为0，当偏度系数大于0时，该分布为右偏，当偏度系数小于0时，该分布为左偏。

$$S_k=\frac{3(\bar{x}-M_e)}{\sigma} \text{ 或 } S_k=\frac{3(\bar{x}-M_o)}{\sigma} \tag{9}$$

其中，$\bar{x}$为数据集平均值；M_o为数据集众数；M_e为数据集中位数。

（5）峰度系数K。

峰度系数是数据集的四阶中心距与标准差的四次方之比，用来反映数据集频数分布的集中程度。正态分布的峰度系数为0，当峰度系数大于0时，表明数据集的频数分布比正态分布更集中，反之则表明数据集的频数分布比正态分布更分散。

$$K=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^4 f_i}{\sigma^4\sum_{i=1}^{n}f_i}-3 \tag{10}$$

其中，f_i为x_i的概率密度函数。

3.2 评估指标融合

本文提出的5个指标，从多个方面反映了数据特征，将经过补全的油中溶解气体含量时间序列数据与实测数据的各个指标进行对比，可以得到补全后的数据在不同方面与实测数据的相似度。但不同角度得出的相似度结果往往存在一定的差异，因此需要综合各方面的评估结果得到一个整体的相似度作为对补全效果的评估结果。D-S证据融合理论是一种重要的信息融合方法，在处理不确定性信息方面得到了广泛应用，可以将多方面得到的具有一定片面性和不确定性的信息进行融合，得到全面考量后的结果。本文运用D-S证据融合理论对经过补全的油中溶解气体时间序列数据与实测数据在各个指标下的相对误差进行融合，得到对于数据补全效果的综合评估结果。

D-S证据融合理论对研究的问题定义一个识别框架Θ，Θ为包含该问题所有可能结果的集合。对于不同的证据体定义各自的基本信任分配函数，对可能的结果进行基本信任分配，最终根据Dempster合成法则融合不同证据体的基本信任分配，得到综合考虑后的信任分配。设$A\subset\Theta$，m_1为证据体1的基本信任分配函数，m_2为证据体2的基本信任分配函数，A_1、A_2、…、A_k为所有使得$m_1(A)>0$的识别框架子集，B_1、B_2、…、B_l为所有使得$m_2(A)>0$的识别框架子集，若$A_i\cap B_j=A$，则$m_1(A_i)m_2(B_j)$表示分配到A上的一部分信任，确切分配到A上的总信任可以表示为：

$$m(A)=\sum_{A_i\cap B_j=A}m_1(A_i)m_2(B_j) \tag{11}$$

为保证信任总和为1，需要去除$A_i\cap B_j=\varnothing$时的信任，再乘以归一化因子，最终得到由m_1和m_2合成的基本信任分配函数为：

$$m(A)=\frac{\sum_{A_i\cap B_j=A}m_1(A_i)m_2(B_j)}{1-\sum_{A_i\cap B_j=\varnothing}m_1(A_i)m_2(B_j)} \tag{12}$$

本文定义识别框架Θ，$A=\{$相似，不相似$\}$，将5个不同的指标视为5个不同的证据体，根据每个指标下经补全的油中溶解气体时间序列数据与实测数据的相对误差进行基本信任分配，计算基本信任分配公式如下：

$$\begin{cases} m_i(\text{相似}) = 1 - \dfrac{\left| G_{\text{int}} - G_{\text{true}} \right|}{G_{\text{true}}} \\ m_i(\text{不相似}) = \dfrac{\left| G_{\text{int}} - G_{\text{true}} \right|}{G_{\text{true}}} \end{cases} \quad i = 1,2,\cdots,5 \qquad (13)$$

将5个不同指标的基本信任分配利用上文所述的Dempster合成法则进行融合，得到对于数据补全效果综合评估的结果。

4　现场验证

选取某地区某变压器2013—2014年200组完整的油中溶解气体监测数据作为训练集，以H_2含量为例，其正向及反向时间序列数据如图2所示。

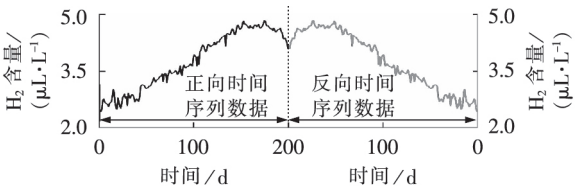


图2　某变压器油中溶解H_2含量正向及反向时间序列数据

Fig.2　Positive- and reverse-sequence time series data of dissolved H_2 content in oil of transformer

综合考量训练集中的样本数及数据精度，对训练集中的油中溶解气体监测数据以0.1 μL／L为间隔进行状态划分，将油中溶解气体监测数据转化为25种状态，根据前文所述公式计算两两状态间的转移概率，组成25×25的正向及反向状态转移矩阵。

另选取该变压器100 d的油中溶解H_2含量在线监测数据作为测试集，随机删除其中25个采样时间点的数据作为缺失值，得到含缺失值的某变压器油中溶解H_2含量时间序列数据如图3所示。

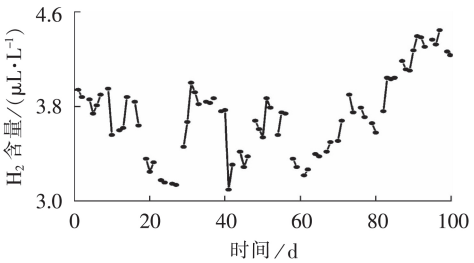


图3　含缺失值的某变压器油中溶解H_2含量时间序列数据

Fig.3　Time series data that containing missing values of dissolved H_2 content in oil of transformer

将距离缺失值最近的实际值所处状态作为初始状态，经过1步或多步状态转移对缺失值从正、反方向进行初步补全，对可能的补全值进行加权求和，对缺失值依次补全，得到完整时间序列数据，如图4所示。

运用多个指标描述补全后的时间序列数据的统计特性，将补全值与实际值进行对比，计算不同指标

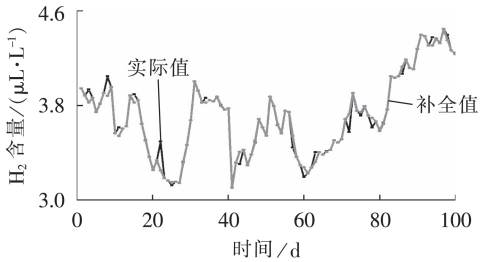


图4　经补全的油中溶解H_2含量时间序列数据

Fig.4　Time series data of dissolved H_2 content in oil after interpolated

下补全值与实际值的误差，最后运用D-S证据融合理论计算补全之后的油中溶解气体含量时间序列与实测时间序列数据的相似度，对补全效果进行综合评估，如表2所示。

表2　补全效果综合评估

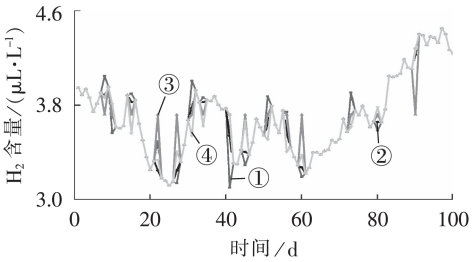
Table 2　Comprehensive evaluation of completion effect

指标	实际值	补全值		
		马尔可夫补全法	保形插值法	均值插值法
平均相对误差/%	—	1.755	1.740	1.674
均值/(μL·L⁻¹)	3.698	3.700	3.703	3.703
均值误差/%	—	0.050	0.133	0.133
变异系数	0.098	0.098	0.098	0.097
变异系数误差/%	—	0.695	0.467	0.589
偏度系数	0.286	0.280	0.250	0.252
偏度系数误差/%	—	2.095	12.534	11.816
峰度系数	−0.582	−0.574	−0.598	−0.594
峰度系数误差/%	—	1.384	2.665	2.065
相似度/%	—	99.999	94.943	95.625

由表2可见，马尔可夫补全法对于随机缺失的25组数据的时间序列补全的相似度可达到99.999%，比传统方法更加准确。

在变压器状态评价中，不仅要考虑油中溶解气体含量时间序列的整体变化趋势，时间序列中的极值点、跃变点处油中溶解气体含量最有可能超出注意值或发生跃变，可反映数据的波动情况，因此也是状态评价的重要依据。为验证本文所提补全方法在时间序列中极值点、跃变点处关键数据的补全效果，在100组数据中，删除15个油中溶解气体含量时间序列极值点、跃变点处的数据，作为缺失值，分别用马尔可夫补全方法、均值插补法，保形插值法进行补全，得到完整时间序列数据，并利用本文提出的综合评估体系，对不同方法的补全效果进行对比，补全效果见图5。对补全效果进行综合评估，结果见表3。

马尔可夫补全方法对于时间序列补全的相似度可达到98.956%，与对随机缺失的补全相比，本文提出的方法对于极值点与跃变点处的关键数据补全效果略有下降，但相比于传统方法有较大提高。



① 实际值，② 马尔可夫补全法
③ 均值插值法，④ 保形插值法

图5 经补全的关键节点数据缺失油中溶解H_2含量时间序列数据

Fig.5 Time series data of dissolved H_2 content in oil with missing key values after interpolated

表3 关键节点补全效果综合评估

Table 3 Comprehensive evaluation of key point completion effect

指标	实际值	补全值		
		马尔可夫补全法	保形插值法	均值插值法
平均相对误差/%	—	4.712	7.992	8.528
均值/(μL·L⁻¹)	3.698	3.702	3.714	3.701
均值误差/%	—	0.114	0.128	0.215
变异系数	0.098	0.088	0.087	0.083
变异系数误差/%	—	3.242	4.196	8.657
偏度系数	0.286	0.414	0.487	0.421
偏度系数误差/%	—	44.584	70.067	46.805
峰度系数	-0.582	-0.434	-0.375	-0.099
峰度系数误差/%	—	25.360	35.510	82.865
相似度/%	—	98.956	86.219	48.647

5 结论

(1)本文提出了基于马尔可夫过程的数据补全方法，依据油中溶解气体变化过程具有的马尔可夫特性，将油中溶解气体数据时间序列转化为马尔可夫链，基于部分完整数据获取油中溶解气体数据在时间上的规律，生成状态转移概率矩阵，通过多步状态转移对油中溶解气体数据缺失部分进行补全。

(2)本文从数理统计角度建立了对油中溶解气体数据质量的综合评估体系，利用平均相对误差、均值、变异系数、偏度系数、峰度系数5个指标描述数据特征，并运用D-S证据理论进行融合，从而对数据质量进行评估。通过原始数据与补全数据数据质量的对比，验证了基于马尔可夫过程的油中溶解气体含量数据补全方法的有效性与准确性。

(3)利用现场数据对本文提出的方法进行了验证，对于油中溶解气体监测数据的缺失值补全的相似度可达到99.999%，对于极值点、跃变点处的关键数据进行补全的形似度可以达到98.956%。

附录见本刊网络版(http://www.epae.cn)。

参考文献：

[1] MACKENZIE E A, CROSSEY J, DE PABLO A, et al. On-line monitoring and diagnostics for power transformers[C]//2010 IEEE International Symposium on Electrical Insulation (ISEI). San Diego, USA: IEEE, 2010: 1-5.

[2] 程鹏，佟来生，吴广宁，等. 大型变压器油中溶解气体在线监测技术进展[J]. 电力自动化设备，2004，24(1)：90-93.
CHENG Peng, TONG Laisheng, WU Guangning, et al. Technical achievements of on-line monitoring of dissolved gas in transformer oil[J]. Electric Power Automation Equipment, 2004, 24(1): 90-93.

[3] 郝思鹏，张济韬，张仰飞，等. 融合在线监测数据的变压器状态评估[J]. 电力自动化设备，2017，37(11)：176-181.
HAO Sipeng, ZHANG Jitao, ZHANG Yangfei, et al. State evaluation of transformer based on information fusion of on-line monitoring data[J]. Electric Power Automation Equipment, 2017, 37(11): 176-181.

[4] 邹建明. 在线监测技术在电网中的应用[J]. 高电压技术，2007，33(8)：203-206.
ZOU Jianming. Application of on-line monitoring technology on power grid[J]. High Voltage Engineering, 2007, 33(8): 203-206.

[5] 刘佳佳，基于DGA技术的变压器故障在线监测方法研究[D]. 长春：吉林大学，2017.
LIU Jiajia. Research on on-line monitoring method of transformer faults based on DGA technology[D]. Changchun: Jilin University, 2017.

[6] 齐波，张鹏，荣智海，等. 基于数据驱动和多判据融合的油色谱监测传感器有效性评估方法[J]. 电网技术，2017，41(11)：3662-3669.
QI Bo, ZHANG Peng, RONG Zhihai, et al. Validity assessment method of DGA sensors based on data driven and multiple criterion integration[J]. Power System Technology, 2017, 41(11): 3662-3669.

[7] 季俊杰. 基于数据驱动的气体传感器故障诊断方法研究[D]. 重庆：重庆大学，2017.
JI Junjie. Research on fault diagnosis method of gas sensor based on data-driven[D]. Chongqing: Chongqing University, 2017.

[8] 庞达. 基于厄米特插值和时间序列模型在沉降监测中的应用[J]. 测绘与空间地理信息，2018，41(3)：171-177.
PANG Da. Based on the Hermitian interpolation and time series model in the subsidence monitoring research[J]. Geomatics & Spatial Information Technology, 2018, 41(3): 171-177.

[9] 刘沅昆，栾文鹏，徐岩，等. 针对配电变压器的数据清洗方法[J]. 电网技术，2017，41(3)：1008-1014.
LIU Yuankun, LUAN Wenpeng, XU Yan, et al. Data cleaning method for distribution transformer[J]. Power System Technology, 2017, 41(3): 1008-1014.

[10] 刁赢龙，盛万兴，刘科研，等. 大规模配电网负荷数据在线清洗与修复方法研究[J]. 电网技术，2015，39(11)：3135-3140.
DIAO Yinglong, SHENG Wanxing, LIU Keyan, et al. Research on online cleaning and repair methods of large-scale distribution network load data[J]. Power System Technology, 2015, 39(11): 3135-3140.

[11] 毛李帆，姚建刚，金永顺，等. 中长期负荷预测的异常数据辨识与缺失数据处理[J]. 电网技术，2010，34(7)：148-153.
MAO Lifan, YAO Jiangang, JIN Yongshun, et al. Abnormal data identification and missing data filling in medium- and long-term load forecasting[J]. Power System Technology, 2010, 34(7): 148-153.

[12] LAWLER G F. Introduction to stochastic processes[M]. 张景肖，译. 北京：机械工业出版社，2010.

[13] 张施令，姚强. 基于WNN-GNN-SVM 组合算法的变压器油色

谱时间序列预测模型[J]. 电力自动化设备，2018，38(9)：155-161.
ZHANG Shiling，YAO Qiang. Predicting model of transformer DGA time series based on WNN-GNN-SVM combined algorithm[J]. Electric Power Automation Equipment，2018，38(9)：155-161.
[14] 丁明，鲍玉莹，毕锐. 应用改进马尔科夫链的光伏出力时间序列模拟[J]. 电网技术，2016，40(2)：459-464.
DING Ming，BAO Yuying，BI Rui. Simulation of PV output time series used improved Markov chain[J]. Power System Technology，2016，40(2)：459-464.
[15] 林峻，严英杰，盛戈皞，等. 考虑时间序列关联的变压器在线监测数据清洗[J]. 电网技术，2017，41(11)：3733-3740.
LIN Jun，YAN Yingjie，SHENG Gehao，et al. Online monitoring data cleaning of transformer considering time series correlation[J]. Power System Technology，2017，41(11)：3733-3740.
[16] MOSINSKI F，PIOTROWSKI T. New statistical methods for evaluation of DGA data[J]. IEEE Transactions on Dielectrics and Electrical Insulation，2003，10(2)：260-265.
[17] POGGI P，NOTTON G，MUSELLI M. Stochastic study of hourly total solar radiation in Corsica using a Markov model[J]. International Journal of Climatology，2000，20(14)：1843-1860.
[18] CHEN J，LI W，LAU A，et al. Automated load curve data cleansing in power systems[J]. IEEE Transactions on Smart Grid，2010，1(2)：213-221.
[19] 中华人民共和国国家质量监督检验检疫总局. 变压器油中溶解气体分析和判断导则：GB／T7252—2001[S]. 北京：中国标准出版社，2001.

作者简介：

张若愚

张若愚（1995—），男，山东临沂人，硕士研究生，主要研究方向为电气设备在线检测与故障诊断（**E-mail**：z1045087967@126.com）；

齐　波（1980—），男，山东临沂人，教授，博士，主要研究方向为电气设备在线检测与故障诊断（**E-mail**：lqicb@ncepu.edu.com）；

张　鹏（1990—），男，江苏南通人，博士研究生，主要研究方向为电气设备在线检测与故障诊断（**E-mail**：zpkkzp@163.com）；

李成榕（1957—）男，福建福州人，教授，博士，主要研究方向为电气设备在线检测与故障诊断、电气绝缘技术、高压测量技术、新型电介质材料的研制与应用等（**E-mail**：lcr@ncepu.edu.cn）。

Method for interpolating monitoring data of dissolved gas in oil for power transformer state assessment

ZHANG Ruoyu[1]，QI Bo[2]，ZHANG Peng[1]，LI Chengrong[2]

（1. Beijing Key Laboratory of High Voltage & EMC，North China Electric Power University，Beijing 102206，China；
2. State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources，
North China Electric Power University，Beijing 102206，China）

Abstract：Fully considering the system shifting characteristics between all states at adjacent time points，a Markov model based method for interpolating monitoring data of dissolved gas in transformer oil for power transformer is proposed. The time series of data of dissolved gas in transformer oil is converted into a Markov chain that transfers between different states，and the interpolating values of the dissolved gas data in transformer oil are calculated by using the forward and reverse state transition matrix. A comprehensive evaluation system for quality of dissolved gas data in oil is established from the perspective of data mining to evaluate the effect of data interpolation from multiple aspects. Based on the D-S evidence fusion theory，the evaluation results from multiple aspects are combined to obtain a comprehensive evaluation result. 25 groups of random missing data in 100 groups of data of dissolved gas in transformer oil are complemented by the proposed method. The results show that the similarity ratio between the completed data and the measured value can reach 99.999%. Furthermore，the data interpolation effect of 15 groups of missing data at extreme points and mutation points are verified. Through comprehensive evaluation，the similarity ratio between the completed data and the measured value at extreme points and mutation points can reach 98.956%. It shows that the proposed method can accurately complement the missing data of dissolved gas in transformer oil without changing the data characteristics，which is beneficial to improve the accuracy of transformer state evaluation method.

Key words：power transformers；state assessment；dissolved gas in oil；data interpolation；Markov process；comprehensive evaluation index

附录

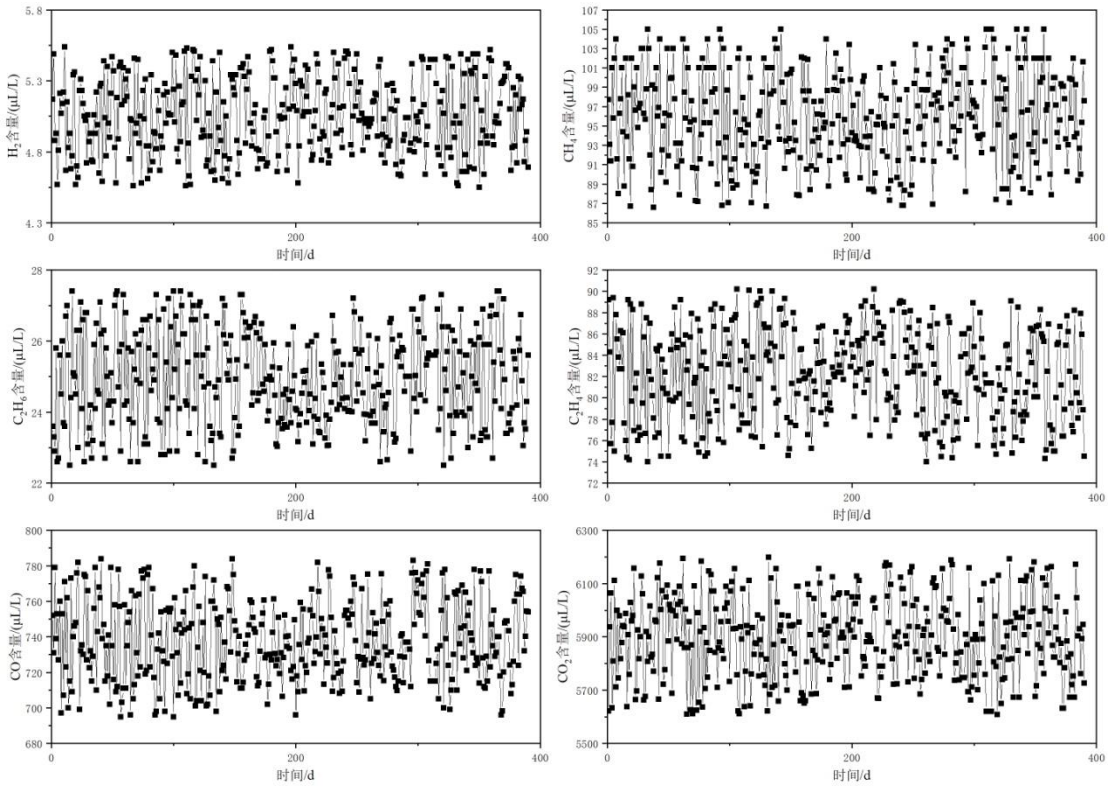


图 A1 某变压器油中溶解气体时间序列数据

Fig.A1 Transformer's time series of dissolved gas in oil

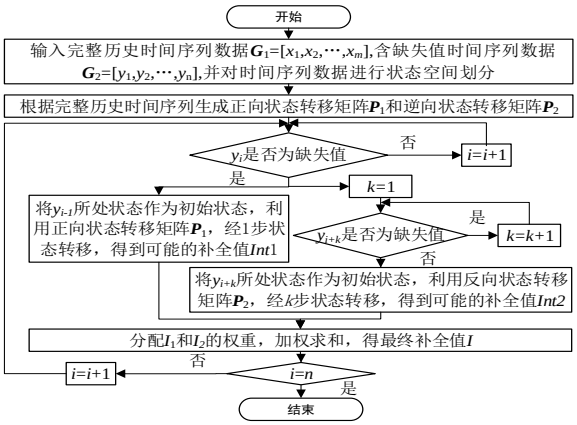


图 A2 基于马尔可夫过程的油中溶解气体监测数据补全流程图

Fig.A2 Flowchart of monitoring data completion of dissolved gases in oil based on Markov process