汽轮发电机定子温度预测模型建模方法实践

赵政雷
（东方电气集团东方电机有限公司　四川　德阳　618000）

摘要：本文以某 1000MW 汽轮发电机真实运行数据为例，介绍了定子线圈出水温度建模方法与过程。使用 KMeans 聚类、BP 神经网络、三次指数平滑等方法实现了定子线圈出水温度的准确预测与动态阈值报警方法的设计。

关键词：线圈出水温度；聚类；神经网络；指数平滑

1　数据来源

对象机组：某 1000MW 汽轮发电机组。

训练样本时间区间：2019 年 3 月 1 日零时至 6 月 1 日零时。

测试样本时间区间：2019 年 6 月 1 日零时至 9 月 1 日零时。

样本时间间隔：约 8s。

样本维度：定子冷却水流量（t/h），A 相电流（A），B 相电流（A），C 相电流（A），定子线圈出水温度（℃），发电机定子冷却水进水温度（℃），发电机励端冷却器冷风温度。

有效样本数 / 训练样本总数：386929/695684。

有效样本数 / 测试样本总数：500059/500059。

2　功能目标

建立针对定子线圈出水温度的计算与异常判断模型，在误差允许范围内实现对象测点温度值的预测性计算；针对该类测点建立基于机组历史运行状态的动态超限报警阈值设计。

3　方案说明

通过机理研究可知，影响定子线圈出水温度运行值的主要因素至少包括电流、冷却水进水温度和冷却水进水流量等。通过对对象机组上述维度数据样本的分析发现，采用单一的神经网络、非线性回归等手段均无法通过电流与冷却水进水流量对较长时间范围（数天）内的温度数据进行有效地表达。

经分析认为，造成这种情况的原因可能是对象机组运行工况的改变，因此，有必要根据机组所处的不同工况及所关联的数据样本进行预测性建模。本文选用 K 均值聚类与 BP 神经网络相结合的方式进行。

同时，由于温度量具有缓变性和延续性等特征，加之通过对定子线圈出水温度测点实测值的分析发现，其数值具有统计学意义上的非稳定性与连续性（滞后性）特征，故本文选用指数平滑法针对单点的趋势进行预测。

4　实现过程

4.1　数据运行情况

为消除数据量纲对后续分析产生的未知影响，将参与试验的部分样本作归一化处理，如 Xstd=(X-Xmin)/(Xmax-Xmin)，该过程不影响数据的整体运行趋势，具体过程不予赘述。

4.2　聚类分析

4.2.1　模型假设

（1）随着机组运行工况的改变，样本特征维度（如电流、流量等）之间的关系模型会发生变化；

（2）可用以表达温度的样本特征维度仅为电流、冷却水流量、进水温度等；

（3）建模所用样本为经过人工筛选的、机组在非故障状态下的运行值；

（4）建模所用样本中的异常样本不足以影响模型效果；

（5）模型误差（归一化后）采用总体方差进行评估，设 $\theta \leqslant 9\times 1e\text{-}4$ 时接受该聚类模型；

4.2.2　K-Means 聚类

经调参后发现，聚类数量在约 200 时损失函数不再显著收敛，为节省计算时间，设定训练的提前退出条件为“50 期内损失函数的极大 / 极小值相等”。经测试，迭代在大约 23 期左右停止，最终损失函数值约 0.000406，说明模型执行情况较好。

4.3　神经网络训练

4.3.1　模型假设

（1）包含聚类模型所有假设；

（2）当机组发生故障时，样本 Si 的数据运行规则发生改变，此时原本经聚类进而标定为类标签 X 的数据样

本可能会被标定为 Y，当类 X 和 Y 是具有较大差异的网络模型的输出且这样的差异显著区别于 Y 的误差概率密度模型时，该故障数据样本 Si 有较高概率被识别出；

（3）模型误差（归一化值）采用样本方差进行评估，训练样本总体的温度预测值与实测值误差在 ±1.0℃内的置信度达到 95% 时接受该模型。

4.3.2　BP 神经网络

经过大量调参后发现，神经元数量为 20 时训练效果较好且执行效率较高；为保证模型效果，设定训练的最大期数为 1000 ；为节省计算时间，设定提前退出条件为“50 期内损失函数极大 / 极小值不大于 0.1 且变化率小于 0.05”。

以对聚类 10# 的网络模型的训练为例：其最低损失函数为 0.004562，迭代在大约 180 期左右停止。说明该聚类的 BP 神经网络对其中的样本数据进行了有效的表达。

4.4　趋势预测

4.4.1　模型假设

（1）受限于物理条件的缓变型变量（如温度）不会发生突破物理模型的突变；

（2）所谓“突变”，可以量化为基于平滑移动窗口的数据均值所构成的平滑曲线于某点处的短期延伸。

4.4.2　平滑预测

经 ADF 平稳性检验发现，训练样本中的温度数据为非时序型数据，即平稳的时间序列。

由于数据整体运行趋势不会因为滑动平均的处理而丢失，故对于每一样本点，其下一期数据的运行趋势应当可以通过其前期数据的运行情况进行表达。

通过图 1 可见，在采用三次指数平滑法并经大量试验后发现，当滑窗尺寸为 5、趋势预测模型学习率为 0.9 时，预测值与原始值 MSE 较低，即说明此时模型对数据的实际运行情况进行了优秀地表达。

5　模型测试

5.1　测试集检验

测试集数据为对象机组的实测数据，在此条件下模型的预测曲线应当与实测值曲线具有较小偏差，其误差应当至少满足工程需要。通过分别计算实测值与网络计算值、趋势预测值在测试集上的误差分布及累计概率密度可知：在测试集中，当 error ∈ [-1.3，1.3]℃时，网络模型的置信度可达 99% 以上；当 error ∈ [-0.2，0.2]℃时，趋势模型的置信度可达 99% 以上，详见图 2。

5.2　具有人为异常数据的测试集检验

由于上述模型是通过研究历史正常工况数据的运行情况以对新样本为正常数据的概率进行分析，故其判据应当为：新样本可以被归类于模型已学习到的数据样本的聚类，且新样本满足该聚类的统计学特征，即新样本的温度不应显著突破所属聚类的历史极值，且采用该聚类项下的网络模型所得的计算值与新样本的温度的误差显著大于模型固有误差概率密度函数所构成的经验累积概率分布曲线所构成的 95% 置信度所代表的误差值的区间时，可将该样本定义为异常样本；同理，趋势预测计算结果与新样本的温度的误差显著大于模型固有误差

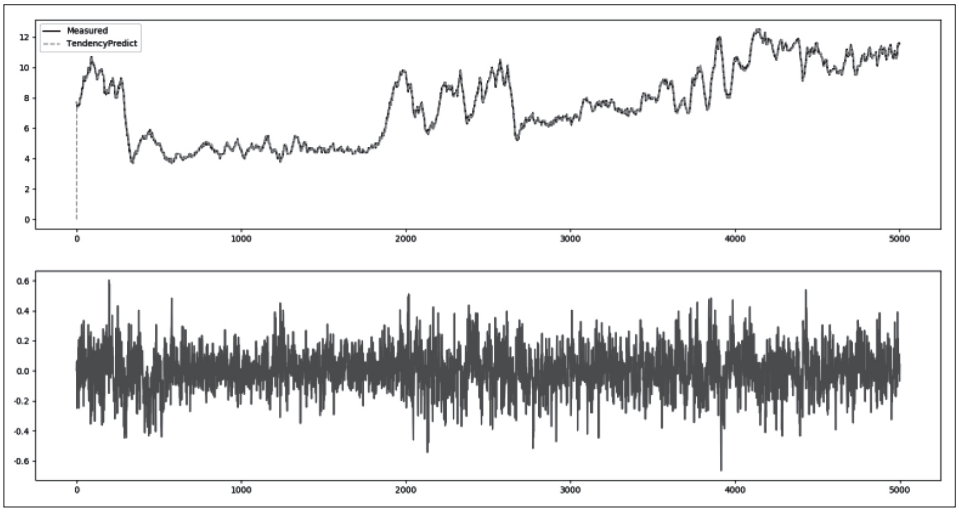

图 1　趋势预测的实验结果

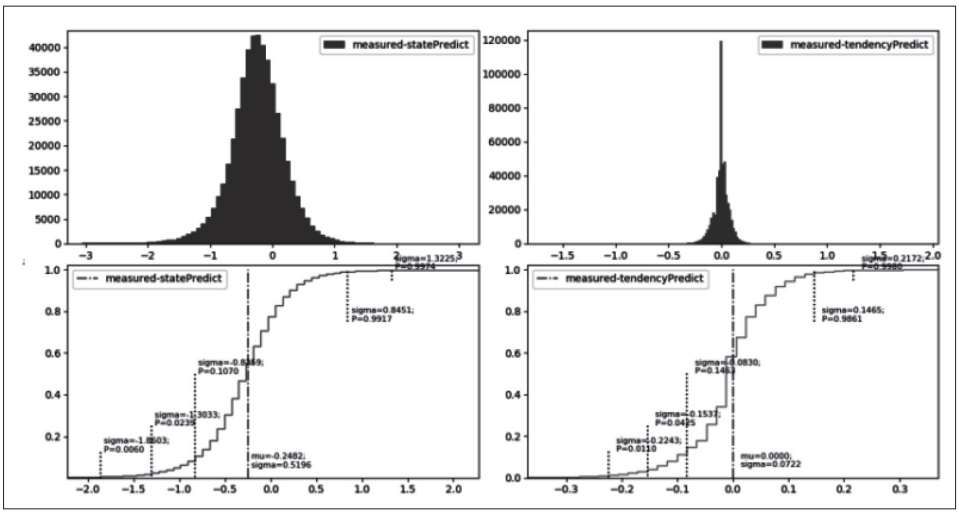

图 2　网络与趋势模型于测试集的表现情况

概率密度函数所构成的经验累积概率分布曲线所构成的95% 置信度所代表的误差值的区间时，可将该样本定义为异常样本。

为了进一步研究该模型在不同工况（稳定工况与非稳定工况）下对不同程度异常数据判别的准确度或模型敏感度，在以下工况及人为创造异常数据条件下进行了试验：

（1）稳定工况——样本数据运行较平稳处

对数据进行如下人为调整：x=300 处，温度高于实测值 3℃；x=400 处，温度低于实测值 3℃；x=500 处，温度高于实测值 2℃；x=600 处，温度低于实测值 2℃；x=700 处，温度高于实测值 1℃；x=800 处，温度低于实测值 1℃；x=900 处，温度高于实测值 0.5℃；x=1000 处，温度低于实测值 0.5℃。

（2）非稳定工况——样本数据变化较剧烈处

对数据进行如下人为调整：x=2000 处，温度高于实测值 3℃；x=2010 处，温度低于实测值 3℃；x=2020 处，温度高于实测值 2℃；x=2030 处，温度低于实测值 2℃；x=2040 处，温度高于实测值 1℃；x=2050 处，温度低于实测值 1℃；x=2060 处，温度高于实测值 0.5℃；x=2070 处，温度低于实测值 0.5℃。

数据运行情况如图 3 所示。

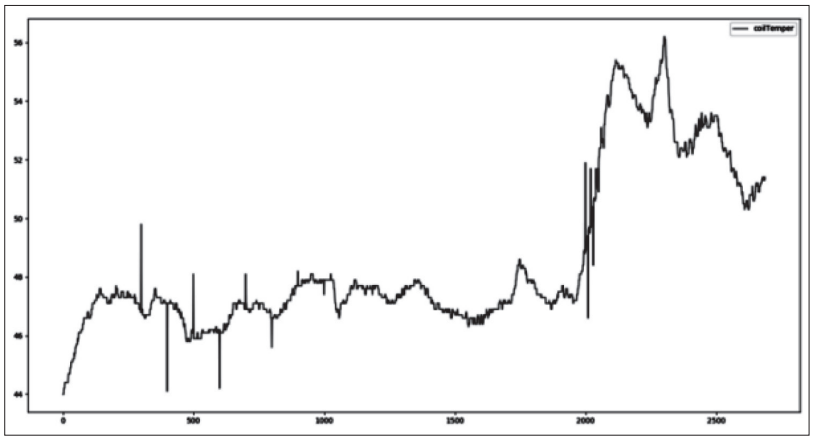

图 3　增加人为异常数据的温度运行图

通过图 3 可以发现，±1℃的人为偏差在模型中得到了凸显，根据 5.1 中提到的置信度所代表的残差分布区间设置报警阈值，阈值设置如表所列时较为合理。

表　阈值设置

	趋势模型	网络模型
μ	0.0	-0.2482
σ	0.0722	0.5196
k	3	1.8
阈值	μ ±k× σ	

试验结果如图 4、图 5 所示。

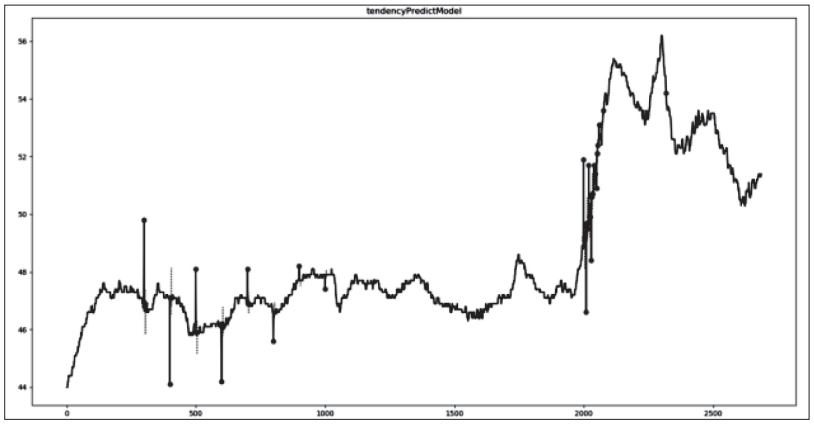

图 4　趋势预测模型检验结果图

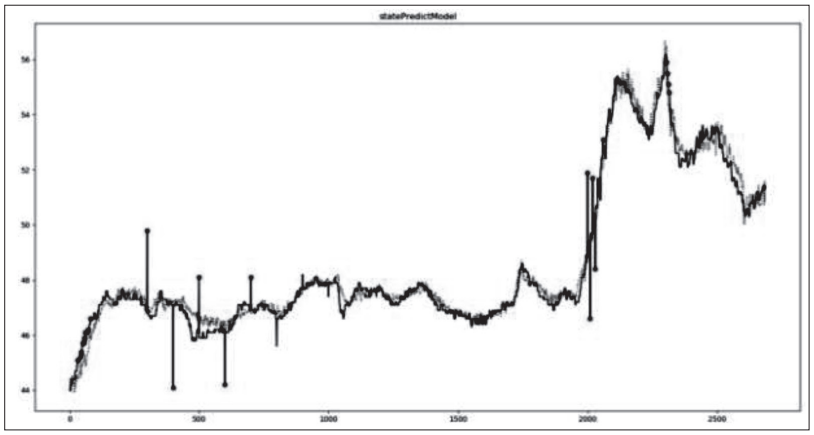

图 5　网络计算模型检验结果图

6　结论

经检验，人为故障样本在不同程度上被检出（使用圆点进行标识），同时也有少量样本被错误地识别或未识别。可以预见的是，以前述两种方法作为综合判据的筛选逻辑将在保证已有异常样本筛出率的条件下进一步降低正常样本的筛出错误率。

综上所述，针对发电机定子线圈出水温度的模型开发结果达到预期要求。

参考文献：

[1] 邓建国，罗隆福．基于 BP 神经网络的大型水轮发电机定子铁心的温度预测 [J]. 湖南大学学报（自然科学版）,2000(01):72-75.

[2] 黎锁平，刘坤会．平滑系数自适应的二次指数平滑模型及其应用 [A]. 系统工程理论与实践 ,2004(02):95-79.