

基于支持向量机的变电设备缺陷发生率的预测及应用

陈义刚 徐厚东

(国网四川省电力公司,四川成都 610041)

摘要: 缺陷预测是设备管理中的重要内容。变电设备由于运行环境复杂,设备缺陷发生的随机性较大。基于支持向量机理论,采用数据挖掘技术,通过对设备运行环境和缺陷发生率的统计分析,利用支持向量机建立设备缺陷平均发生率与设备运行环境的回归函数,回归结果与实际情况较为吻合。对给定运行环境下设备缺陷平均发生率进行预测,预测误差小于10%,对设备的运行维护管理具有较高的参考价值。

关键词: 支持向量机; 设备缺陷管理; 发生率; 预测

Abstract: Defect prediction is the important content of the equipment management. Due to the complex operating condition of transmission and distribution equipment, the randomness of the occurrence of defects is high. Based on the theory of support vector machine, the operating condition and defect occurrence rate of the equipment are analysis using data mining technology, and the regression functions for the average defect occurrence rate and the operating condition are established using support vector machine, whose regression results coincide with the actual situation. The prediction for the average occurrence rate of equipment defects under the given operating condition is carried out and the prediction error is less than 10%, which is of valuable reference for the management of the operation and maintenance of the equipment.

Key words: support vector machine; management of equipment defect; occurrence rate; prediction

中图分类号: TM866 文献标志码: A 文章编号: 1003-6954(2013)06-0075-03

0 引言

变电站是电力系统的重要枢纽,变电设备的安全稳定运行是确保电力系统可靠供电的重要基础。设备的缺陷管理是运行管理的重要组成部分,是提高设备完好率的重要手段,是编制设备修、试、校计划的主要依据。设备缺陷管理也是安全生产管理的重要内容之一。随着变电设备的大量增加,每年发生的设备缺陷也急剧增加。大量增加的设备缺陷需要消耗大量的运维力量。科学预测设备缺陷的发生率,及时制定有针对性的措施,对于保障设备的安全稳定运行具有重要的意义。

由于设备的运行环境较为复杂,统计样本量较少,对设备缺陷发生率的预测一般较为困难。近年来,随着支持向量机技术的发展,支持向量机在少量样本的回归分析和预测中得到了快速发展和应用,为科学预测设备缺陷发生率提供了一种新的方法和手段。采用支持向量机,利用设备的历史运行环境

数据和缺陷发生率进行回归分析,建立设备缺陷发生率与设备运行环境的回归函数,对设备缺陷发生率进行预测和分析,为设备的运行维护管理提供参考。

1 支持向量机原理^[1 2]

支持向量机是数学挖掘中的新方法,它是建立在统计学基础之上的通用学习方法,专门针对小样本情况下研究机器学习规律的理论,目前已较多地应用于工程预测和综合评价等诸多领域。

回归问题可以表述为,通过给定的观测样本集 $T = \{(x_1, y_1), \dots, (x_i, y_i)\}$ 寻找一个回归函数 $f \in F$, 使得关系 $y = f(x)$ 损失风险最小。

其中 $x_i \in R^n$ 是输入指标向量; $y_i \in R$ 是输出。

回归问题的目的是,给定一个新的输入 x 根据已知观测集所蕴含的关系,推断它所对应的输出 y 值。

支持向量机方法最初是以解决分类问题为出发点的,通常选择 ε 为不敏感损失函数。即

$$c[x, y, f(x)] = |y - f(x)| - \varepsilon$$

$$= \max\{0, |y - f(x)| - \varepsilon\} \quad (1)$$

当 x 点的观测值 y 与预测值 $f(x)$ 之差不超过事先给定的 ε 时, 认为在该点的预测值 $f(x)$ 是无损失的。

支持向量机进行回归分析时, 限定回归函数在线性函数集合中进行选取。当观测样本不能用线性函数进行回归时, 支持向量机先将观测样本 (x_i, y_i) 映射到一个高维的 Hilbert 空间中 $(\Phi(x_i), y_i)$, 然后再对映射后的样本集进行线性回归。即, 令

$$z = \Phi(x) \quad (2)$$

支持向量机的线性回归函数一般为

$$f(x) = w \cdot z + b, w \in R^n, b \in R \quad (3)$$

基于结构风险最小化原则, 回归问题转化为下列最优化问题。

$$\min_{w \in R^n, \zeta, \zeta^* \in R} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k (\zeta_i + \zeta_i^*)$$

$$\text{s. t. } [(w \cdot z_i) + b] - y_i \leq \varepsilon + \zeta_i$$

$$y_i - [(w \cdot z_i) + b] \leq \varepsilon + \zeta_i^*$$

$$\zeta_i, \zeta_i^* \geq 0 \quad (4)$$

其中, $\|w\|^2$ 表示置信范围, 体现了函数集的表达能; $\sum_{i=1}^k (\zeta_i + \zeta_i^*)$ 体现了经验风险; C 为惩罚因子, 最小化这两项之和体现了结构风险最小化思想; ζ_i, ζ_i^* 为松弛变量的上下限, 体现了不敏感损失函数的应用。

根据对偶原理, 引入式 (3) 的对偶问题如下。

$$\min_{\alpha, \alpha^* \in R^k} \frac{1}{2} \sum_{i=1}^k (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (z_i \cdot z_j) +$$

$$\varepsilon \sum_{i=1}^k (\alpha_i^* + \alpha_i) - \sum_{i=1}^k y_i (\alpha_i^* - \alpha_i)$$

$$\text{s. t. } \sum_{i=1}^k (\alpha_i^* - \alpha_i) = 0$$

$$0 \leq \alpha_i^*, \alpha_i \leq C \quad (5)$$

利用序列最小最优化算法可得到最优解 $\overline{\alpha_i^*}$ 和 $\overline{\alpha_i}$, 并根据 KKT 条件计算得到 \overline{b} , 从而回归函数的表达式如下。

$$f(x) = \sum_{i=1}^k (\overline{\alpha_i^*} - \overline{\alpha_i}) (z_i \cdot z) + \overline{b} \quad (6)$$

即

$$f(x) = \sum_{i=1}^k (\overline{\alpha_i^*} - \overline{\alpha_i}) (\Phi(x_i) \cdot \Phi(x)) + \overline{b}$$

$$+ \sum_{i=1}^k (\overline{\alpha_i^*} - \overline{\alpha_i}) K(x_i, x) + \overline{b} \quad (7)$$

式中 $K(x_i, x)$ 为支持向量机的核函数。常用的核函数有多项式核函数、高斯径向基函数核函数、B 样条

核函数、傅里叶核函数、Sigmoid 核函数等。

2 变电设备缺陷发生率的支持向量机模型

变电设备缺陷的产生除设备本身质量原因外, 与设备的运行环境具有很大的关系。设备本身的质量原因而引起的设备缺陷的概率接近于一个常数, 因此, 变电设备的缺陷发生率受设备的运行环境影响较大, 是设备运营环境的函数。

根据变电设备的运行特点, 影响设备的运行环境通常主要包括设备的运行负荷、运行年限、温度、湿度、污秽情况、海拔等。将不同地区变电设备的运行环境作为输入向量, 将各地区变电设备的缺陷发生率作为输出量, 根据观测到的历史数据, 通过支持向量机理论进行机器学习, 建立回归模型, 并利用该模型, 对给定的运营环境下的设备缺陷发生率进行预测, 从而进一步指导变电设备的运行和维护。

近年来, 随着中国电力设备制造企业的不断发展, 技术不断进步, 电力设备运行中发生缺陷的情况越来越少。对于单台变电设备而言, 发生缺陷具有较大的偶然性。根据统计学原理, 为了得到更为准确的概率, 需要更多的运行样本。但是, 根据电力系统设备的运行特点, 相同运行环境下, 型号、容量等参数完全相同的设备几乎没有。因此, 需要对设备和设备的运行环境做一些近似处理。

设备缺陷发生率, 通常按照设备类型, 如变压器、断路器等, 来统计设备的平均缺陷发生率。在电力系统的生产运行中, 也常常采用平均缺陷发生率。

设备的平均缺陷发生率定义如下。

$$P_i = N_i / T_i \quad (8)$$

式中, P_i 表示某类设备的缺陷发生率, 项/台; N_i 表示该类设备当期发现的缺陷总数, 项; T_i 表示当期该类设备的总数量, 台。

根据设备平均缺陷发生率的概念, 设备的运行环境可采用统计平均值来近似。根据统计学的原理, 通常采用统计区域内设备的平均负载率、平均运行年限、平均运行温度、平均运行湿度、平均污秽等级、平均海拔等, 作为本地区设备的运行环境。

$$E_i = \sum (e_i \cdot n_i) / N \quad (9)$$

式中, n_i 表示运行于环境值为 e_i 的设备数量; N 为该类设备总数量。

将历史统计得到的设备运行环境参数平均值和设备缺陷平均发生率进行归一化处理,构造样本集 (E, P) 利用支持向量机进行机器训练后,得到设备缺陷平均发生率的预测模型为:

$$P(e) = \sum_{i=1}^k (\bar{\alpha}_i^* - \bar{\alpha}_1) K(e_i, e) + \bar{b} \quad (10)$$

3 变电设备缺陷发生率的预测和应用

对国网四川省电力公司2012年1月至2013年9月变压器的平均缺陷发生率以及变压器的平均运行负载率、平均运行温度、平均运行湿度、平均海拔高度等运行环境进行统计,并作为样本,基于支持向量机理论,选择高斯径向基函数作为核函数,通过适当调整支持向量机的训练参数(本例中选取 $e - SVR$ 核函数 γ 函数设置取7,损失函数设置选5),建立支持向量机变电设备缺陷发生率回归模型。回归分析得到的变压器缺陷平均发生率与实际缺陷平均发生率结果如图1。

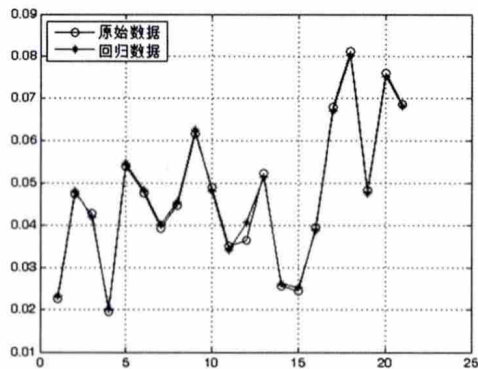


图1 支持向量机对变压器缺陷平均发生率的回归分析

通过图1可以看出,利用支持向量机对变电设备缺陷进行回归分析,大部分回归点可以得到较为理想的结果,回归分析平均方差为0.02%。设备缺陷平均发生率与实际情况的变化趋势基本一致。

根据四川电网的负荷分析和气象分析,得到四川电网的2013年10月设备运行环境参数,代入上述回归模型,得到国网四川省电力公司10月份变压器缺陷平均发生率的预测值,输出如下。

$$\text{Mean squared error} = 0.287\ 408 \text{ (regression)}$$

$$\text{预测数据} = 0.047\ 7$$

根据国网四川省电力公司2013年10月份变压器实际运行缺陷统计,计算得到10月份变压器的实际缺陷平均发生率为0.0519。回归模型预测值与

实际值相差仅0.42个百分点,回归预测误差为

$$\left| \frac{\Delta p}{p} \right| = \left| \frac{P_{\text{预测}} - P_{\text{实际}}}{P_{\text{实际}}} \right| \times 100\% = 8.1\% \quad (11)$$

因此,该回归模型可以在变电设备运维管理中提供较为准确的参考。

由于设备缺陷发生率是一种概率,实际设备缺陷的发生还受很多随机因素的影响,因此,预测值与实际值之间还存在着一定的偏差,尤其是在设备量较少的地区。

根据全省变电设备缺陷发生率的统计数据得到回归模型,也可以对地市公司设备的运维进行预测和分析。对某一地区设备未来的运维环境进行预测后,代入支持向量机的预测模型式(10),便可以得到该地区设备故障平均发生率,从而可以进一步指导该地区设备的运维。

4 结 论

变电设备由于运行环境复杂、技术更新快,具有参考价值的设备缺陷平均发生率统计样本较少。支持向量机由于在解决小样本条件下机器学习的问题具有较好的应用。因此,利用支持向量机理论,对变电设备运行环境和缺陷平均发生率进行回归分析,建立变电设备缺陷平均发生率预测模型,得到的预测值与实际值较为接近,预测效果较为理想。在给定运行环境下,能够较为准确地预测出设备缺陷的平均发生率,对设备的运行、维护和抢修等管理具有较大的参考价值。

参考文献

- [1] 田英杰. 支持向量回归机及其应用研究[D]. 北京: 中国农业大学, 2005.
- [2] 刘庆彪,张步涵,王凯,等. 电价预测的自适应支持向量机方法研究[J]. 电力系统保护与控制, 2008, 36(22): 34-39.
- [3] 张华,曾杰. 基于支持向量机的风速预测模型研究[J]. 太阳能学报, 2010, 31(7): 928-932.
- [4] 陈沅涛,徐蔚鸿,吴佳英. 一种增量向量支持向量机学习算法[J]. 南京理工大学学报:自然科学版, 2012, 36(5): 873-878.
- [5] 邓乃杨,田英杰. 数据挖掘中的新方法——支持向量机[M]. 北京: 科学出版社, 2004.

作者简介:

陈义刚(1981),男,工程师,主要从事电力系统管理。

(收稿日期:2013-07-04)