

变压器类设备状态评价的回归建模研究

夏 阳,孙庆山,汤 进,罗 斌

(安徽大学 计算机科学与技术学院,安徽 合肥 230039)

摘 要:在国家电网系统中,电力系统变压器类设备起着至关重要的作用,而对变压器类设备的故障诊断又直接决定着电力变压器类设备的维护工作。文中提出在变压器类设备状态检修中,使用基于线性回归的建模方法,以大量运行中的设备的状态信息为样本,对变压器类设备状态建立回归模型,从而形成对变压器类设备更精确的状态评价,在很大程度上弥补了传统的阶梯性扣分评价体系的不足,以得出更精确的检修方案,并通过实验证明该方法的有效性和正确性。

关键词:变压器类设备;多元线性回归算法;设备故障诊断

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2012)06-0156-03

Research on State Evaluation Modeling of Transformer Equipment

XIA Yang, SUN Qing-shan, TANG Jin, LUO Bin

(College of Computer Science and Technology, Anhui University, Hefei 230039, China)

Abstract: In the national power grid system, transformer equipment of power system plays an important role, and transformer equipment diagnosis directly guides the power transformer equipment maintenance. It proposes to use the nonlinear regression algorithm for equipment condition maintenance, build a transformer nonlinear regression model with a lot of samples on operation equipment state information. The approach can make up for the disadvantage of traditional multistep points in the evaluation system, thus provide a more accurate maintenance scheme. And experiments validate the effectiveness and correctness of the approach.

Key words: transformer equipment; linear regression model; equipment maintenance

0 引言

设备状态检修是指企业以安全、环境、效益等为基础,通过对设备运行状态的评估来决定检修决策并开展相应的设备检修工作,从而达到设备运行安全可靠、检修成本合理的一种设备检修策略。其中,变压器类设备又是最重要的电力设备之一。因对变压器类设备的检修模型的研究,确定设备状态,依据设备状态开展相应的试验、检修工作成为国家电网设备检修工作的重要组成部分^[1]。

近年来国家电网发布试行的变压器类设备状态检修导则^[2]和评价导则^[3],国家电网对变压器类设备的状态检修一直按照该导则对设备进行评价。由于该导则对设备参数采用阶梯型评价的原则,不能精确的体现采集的数据与设备状态的相互关系。例如评价指标的第23条油中溶解气体分析中H₂的选项中H₂含量大于150 $\mu\text{L/L}$ 就要扣除该变压器本体8分,不能准确

的区分H₂含量148 $\mu\text{L/L}$ 和H₂含量152 $\mu\text{L/L}$ 之间的区别,再如对油中溶解气体分析中总烃气体由总烃含量和气体速率多因素决定评价的标准可以进行建模,以更精确地得出多元参数对评分结果的影响。

目前,各种回归模型广泛地应用于油田产量预测^[4]、预测灾害策略研究^[5]、数学数据处理^[6]、生理指标评估^[7]等各种生产应用领域,在生产应用中起到了重要的作用。自然界两变量之间一般存在着非线性关系,由此文中利用多元线性回归算法对变压器类设备状态评价进行回归建模,模拟出变压器类设备扣分评价和各个参数的关系。

1 变压器类设备状态评价的回归模型

1.1 回归算法原理

在回归分析中,当研究的因果关系只涉及因变量和一个自变量时,叫做一元回归分析;当研究的因果关系涉及因变量和两个或者两个以上自变量时,叫做多元回归分析。此外,在回归分析中,又依据描述自变量与因变量之间因果关系的函数表达式是线性的还是非线性的,分为线性回归分析和非线性回归分析。

通常线性回归分析法^[8]是最基本的分析方法,遇

收稿日期:2011-10-21;修回日期:2012-02-02

基金项目:国家自然科学基金(61073116,61003038);安徽省教育自然科学基金重点基金(KJ2010A006)

作者简介:夏 阳(1987-),男,硕士研究生,主要从事图像处理与模式识别研究。

到非线性回归问题可以借助数学手段化非线性回归问题为线性回归问题来处理。

对于变压器类设备,把设备状态评分作为因变量,各种不同设备参数作为自变量,利用大量实际运行中的数据作为分析样本,得出其相互之间的关系模型。

1.2 变压器类设备线性回归算法的数学建模

对于任意的 N 个点 $(x_k, y_k), k=1, 2, \dots, N$, 其中 x_k 互不相同。令 $x = [x_1, x_2, \dots, x_N]^T, y = [y_1, y_2, \dots, y_N]^T$, 如果 x 和 y 之间存在线性关系, 则存在一个一阶线性方程 $y=f(x)$ 拟合这组数据。

令 Y 为该变压器类设备在该参数项目上的应扣分值, X_1, X_2, \dots, X_p (变化率, 气体容量, 温度, 湿度等) 分别为各参数数值, 设变量 Y 与变量 X_1, X_2, \dots, X_p 之间存在线性关系。则 $y=\beta_0+\beta_1x_1+\beta_2x_2+\dots+\beta_px_p+\varepsilon$, 其中 $\varepsilon \sim N(0, \sigma^2), \beta_0, \beta_1, \dots, \beta_p$ 和 σ^2 是未知参数。

若令

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \beta=\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, X=\begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

则上述多元线性回归模型用矩阵的形式表示为:

$Y=X\beta+\varepsilon,$

其中参数 β 的估计值 $\hat{\beta}$, 就是求最小二乘函数 $Q(\beta)=(Y-X\beta)^T(Y-X\beta)$ 达到最小值的 β 值。

可以证明最小二乘估计的

$$\hat{\beta}=(X^T X)^{-1} X^T Y,$$

从而最终可得回归方程为

$$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1 X_1+\hat{\beta}_2 X_2+\cdots+\hat{\beta}_p X_p$$

1.3 一元回归方程的非线性变换

相对于一元线性方程, 非线性方程会更加符合在实际生产中的应用, 而对非线性方程的求解一般转换为线性方程进行求解。处理非线性回归的基本方法是, 通过变量变换, 将非线性回归化为线性回归, 然后用线性回归方法处理^[9]。设变量 y 和变量 x 满足一元非线性方程 $y=c e^{a x}$, 则 $\ln y=\ln c+a x$, 令 $Y=\ln y, b=\ln c$ 。则上述方程可转化为一元线性方程 $Y=a x+b$ 。

2 用线性回归算法对变压器类设备故障诊断的建模实现

2.1 一元非线性回归算法的建模实验

阶梯式的评价模型对数值进行建模不能充分反应扣分和采集参数之间的关系, 以变压器本体的短路冲击电流在容许短路电流的百分比为例, 国家电网下发的标准中根据短路冲击电流在允许短路电流的百分比给该变压器的本体扣除相应的分数, 见表 1:

表 1 短路冲击电流在允许短路电流的百分比项目评分表

短路冲击电流在允许短路电流的百分比	0 ~ 50%	50% ~ 70%	70% ~ 90%	90% 以上
扣分	0	6	12	30

用线性回归算法^[10]对该项评分进行一元线性回归建模, 表示为 $y=x_1 \beta_1+\beta_0$, 其中 x_1 为短路冲击电流在允许短路电流的百分比, y 为该变压器的本体应扣除的相应分数, 通过软件依据国家电网标准模拟出 400 条样本数据进行回归训练, 用最小二乘法估计出参数, 得出线性方程 $Y1=19.33 * X-5.28$ 。在实验中, 用 $y+1=C e^{a x}$ 模型进行非线性变换得出非线性模型 $Y2=\exp (3.857 * X-0.846)-1$; 并将当 $Y1$ 或者 $Y2$ 小于零时置零^[11]。图 1 为建模函数的图像表示, 其中散列点为原始数据样本, 直线为线性回归模型 $Y1$, 曲线为非线性模型 $Y2$ 。

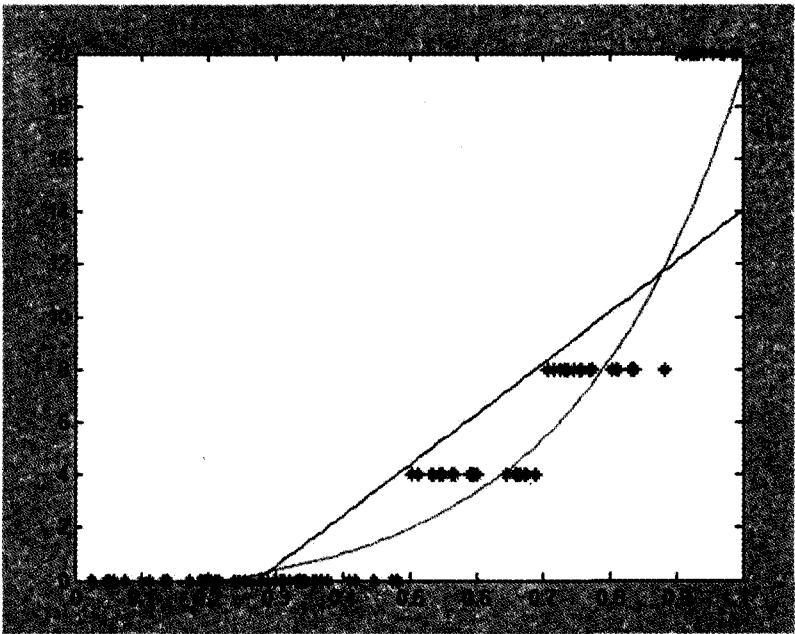


图 1 短路冲击电流在允许短路电流的百分比的建模图示

可以看出曲线 $Y2$ 能更好地反映参数和扣分之间的关系。样本数据比对如表 2 所示

当 $Y1, Y2$ 扣分为负数时, 将 $Y1, Y2$ 置零, 避免了扣分为负数的情况, 由表中数据可见用非线性回归模型

对该项参数进行评价最为合理准确。

表 2 短路冲击电流在允许短路电流的百分比项目建模样本表

序号	样本参数(短路冲击电流在允许短路电流的百分比)	阶梯扣分	线性回归模型	非线性回归模型
1	0.01979	0	-4.9054593	0
2	0.03147	0	-4.6796849	0
3	0.11801	0	-3.0068667	0
4	0.14771	0	-2.4327657	0
5	0.25574	6	-0.3445458	0.15072155
6	0.3053	6	0.613449	0.393114795
7	0.42083	6	2.8466439	1.175245018
8	0.45746	6	3.5547018	1.505336909
9	0.53456	12	5.0450448	2.372984363
10	0.56311	12	5.5969163	2.765630347
11	0.58848	12	6.0873184	3.152735858
12	0.68744	12	8.0002152	5.082745034
13	0.75064	12	9.2218712	6.761832708
14	0.82241	12	10.6091853	9.237273967
15	0.97805	30	13.6177065	17.65919119

2.2 多元线性回归算法的建模实现

有些时候,会有多参数变量的评价标准,见表 3:

表 3 油中溶解气体总烃项目评分表

油中溶解气体总烃	总烃含量大于 150μl/l	产气速率大于 10%/月	总烃含量大于 150μl/l,且产气速率大于 10%/月
	12	24	30

用多元线性回归算法对该项评分进行建模,表示 $y = x_1\beta_1 + x_2\beta_2 + \beta_0$,其中 x_1 为总烃密度含量, x_2 为产气速率, y 为该变压器的本体油中气体分析应扣除的相应分数,通过软件依据国家电网标准模拟出 400 条样本数据进行回归训练,用最小二乘法估计出参数,得出线性方程 $Y = 0.137X_1 + 160.131X_2 - 19.921$,并将当 X_1 或者 X_2 小于零时置零。绘图如图 2 所示:

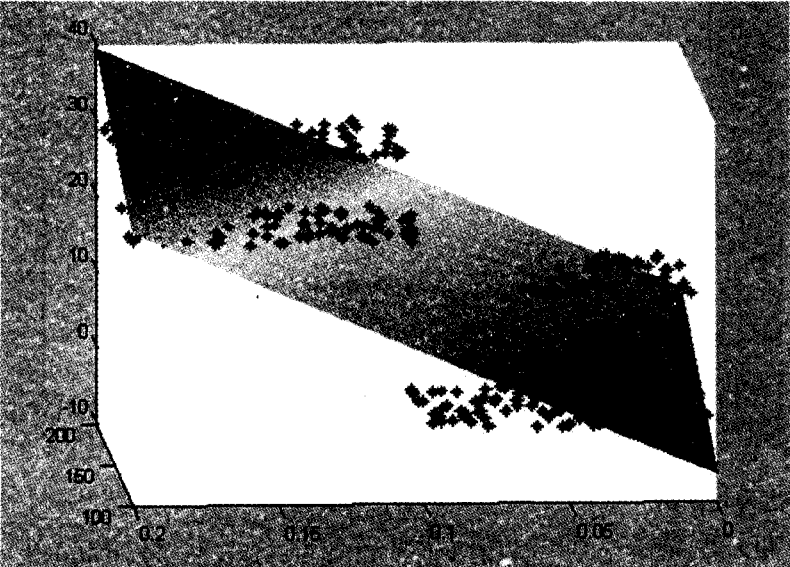


图 2 油中溶解气体总烃的建模图示

如上图平面即是模拟出的用多元线性回归算法模拟出的模型。横轴为油中溶解气体总烃产气速率,竖

轴为总烃密度含量,纵轴为应扣分数。散列点为样本数据。表 4 为部分样本实验数据:

表 4 油中溶解气体总烃项目建模样本表

序号	油中溶解气体总烃		阶梯扣分	线性回归模型
	总烃密度含量(μl/l)	产气速率		
1	3	0.09456	0	0
2	123	0.04882	0	4.74759542
3	158	0.10528	30	18.58359168
4	126	0.22311	24	33.06782741
5	232	0.19079	30	42.41439349
6	40	0.03886	0	0
7	48	0.06763	0	0
8	78	0.03679	0	0
9	24	0.03999	0	0
10	111	0.03159	0	0.34453829
11	156	0.10553	30	18.34962443
12	112	0.09387	0	10.45449697
13	227	0.07653	24	23.43282543
14	76	0.01123	0	0
15	54	0.10303	24	3.97529693

可见用多元线性回归算法对多参数和扣分进行建模更加合理,更能体现不同参数下该变压器本体的扣分状况。

此外文中还对其他不同的评分标准进行建模如表 5:

表 5 不同评分标准下建模

评价项目	线性回归模型	非线性回归模型	备注
绕组电容变化率	$y_1 = 6.119x - 10.751$	$y_2 = \exp(57.59 * X - 87.99) - 1$	x 电容为变化率
铁心接地电流	$y_1 = 29.516x + 3.065$	$y_2 = \exp(3.13 * X - 4.89) - 1$	x 为电流值(安培)
泄露电流变化率	$y_1 = 0.298x - 6.916$	$y_2 = \exp(59.56 * X - 90.00) - 1$	x 电流为变化率
乙炔含量	$y_1 = 6.100x - 22.473$	$y_2 = \exp(16.30 * X - 6.19) - 1$	x 为乙炔含量(μl/l)
氢气含量	$y_1 = 0.118x - 13.781$	$y_2 = \exp(19.24 * X - 17.84) - 1$	x 为氢气含量(μl/l)

实践证明,非线性回归模型比现有的阶梯性模型更具有优势。

3 结束语

文中通过把线性回归算法应用于变压器类设备的评分标准中,以大量现实生产中的样本数据为依据,对原本阶梯式的评分进行改进,建模示例可以表明非线性回归模型和多元线性回归模型都可以更好地反映参数和扣分的关系,用以对实时参数做出可靠具体的判断^[12]。其实不仅仅是变压器类设备,所有变电站的一次设备均可通过相应的方法建模,得出比原有模型更精确的扣分,给生产提供更加实际的参考数据。

结合语义距离和共同祖先节点的知识,即公式(1)和(2),进一步得出相似度函数(Similar Function):

$$F(X,Y) = e^{-\alpha \text{Distance}(X,Y)} + \beta \text{Desity}(\text{NearConAncestor}(X,Y)) + \gamma \frac{\text{Depth}(\text{NearConAncestor}(X,Y))}{\text{Height}(\text{NearCoAncestor}(X,Y))} \quad (3)$$

其中 $\alpha + \beta + \gamma = 1$, $F(X,Y)$ 的值域为 $(0,1]$ 。

对于相似度计算公式(3)而言,假设,取 $\alpha = 0.2$, $\beta = 0.2$, $\gamma = 0.2$,对图 3 的节点 E1 进行相似度计算,结果如表 3 所示:

表 3 各节点与 E1 的相似度计算

序号	查询节点	节点名称	相似度
1	E1	B1	0.3654
2	E1	C2	0.7467
3	E1	C4	0.5643
4	E1	D3	0.8675
5	E1	E2	0.9102

4 结束语

由此看出,可以根据构建出来的概念树,来分析概念与概念之间的上下位关系,再通过相似度的计算方法,用具体的数值表示出概念间的关系。从而可以针对某一个特定的领域来研究短文本,更便于找出短文本的相似性和唯一性。

文中虽然可以把短文本局限在某个特定的领域,但是,短文本的数量众多,且长度都非常短,样本特征非常稀疏,且词性灵活多变。所以,对于短文本相似度的研究,还存在众多难点,例如:怎样对短文本提取有效的特征词,怎样灵活处理短文本的词性问题,怎样更为周全的考虑到影响相似度的因素。这些都要进行下一步的研究。

参考文献:

- [1] 邹文科. 基于本体技术的语义检索及其语义相似度研究[D]. 北京:北京邮电大学,2008.
- [2] 黄 果,周竹荣. 基于领域本体的概念语义相似度计算研究[J]. 计算机工程与设计, 2007,28(10):2460-2463.
- [3] 张凯勇. 基于 WordNet 的词语及短文本语义相似度算法研究[D]. 长春:吉林大学,2008.
- [4] 贾文娟,何 丰. 基于 HowNet 的中文本体学习方法的研究[J]. 计算机技术与发展,2011,21(6):120-125.
- [5] 江 磊. 基于概念树的语义相似度计算的研究[J]. 广东通信技术,2010,30(8):124-131.
- [6] 樊兴华,孙茂松. 一种高性能的两类中文文本分类方法[J]. 计算机学报,2006,29(1):124-131.
- [7] 王洪伟,吴家春. 本体的形式化模型及在语义查询中的应用[M]. 北京:高等教育出版社,2003.
- [8] 冉 婕,孙 瑜. 语义检索中的词语相似度计算[J]. 计算机技术与发展,2011,21(4):40-48.
- [9] 郑 诚,秦多荣. 本体映射中的概念相似度计算[J]. 计算机技术与发展,2008,18(11):120-126.
- [10] 谢信喜,王士同. 适用于数据的基于相互距离的相似性传播聚类[J]. 计算机应用,2008,28(6):120-124.
- [11] Montejo-Raez A, Dallman D. Experience in Automatic Key-wording of Particle Physics Literature[J]. High Energy Physics Libraries Webzine,2001(5):124-131.
- [12] Zelikovitz S, Marquez F. Transductive Learning for Short-text Classification Problems Using Latent Semantic Indexing[J]. International Journal of Pattern Recognition and Artificial Intelligence,2005,19(2):143-163.
- [13] Sinha R, Mihaleca R. Unsupervised graph-based word sense disambiguation using measure of word semantic similarity [C]//Proceeding of the International Conference on Semantic Computing (ICSC07). Washington, DC, USA: IEEE Computer Society,2007.

(上接第 158 页)

参考文献:

- [1] Guo Liang. Application of Improved Multivariate Linear Regression to Output Prediction of an Oilfield[J]. Journal of Xidian University (Social Science Edition), 2009,19(3):71-75.
- [2] 王瑞兵,范柱国. 滇中红层区滑坡灾害多元线性回归模型构建[D]. 昆明:昆明理工大学,2010.
- [3] 董 冬. 缺失数据下线性模型回归系数岭估计的大样本性质[D]. 桂林:广西师范大学,2010.
- [4] 林升梁. 多元线性回归模型在骨龄评估中的应用[J]. 吉林医学,2011,32(24):5107-5108.
- [5] Sun Baoqin. The Application and Strategy of Condition Based Maintenance of Power Equipment [J]. the Journal of Jilin Chemical Institute,2007,24(1):21-25.
- [6] 安徽省电力公司. 变电站内 35kV 及以下电压等级变压器类设备状态检修导则(试行)[S]. 2009.
- [7] 安徽省电力公司. 变电站内 35kV 及以下电压等级变压器类设备状态评价导则(试行)[S]. 2009.
- [8] Mathews J H, Fink K D. Numerical Methods Using MALTAB [M]. 4th ed. 北京:电子工业出版社,2009.
- [9] 陈永胜,宋立新. 多元线性回归模型以及 SPAA 软件求解[J]. 通化师范学院学报,2001,28(12):8-9.
- [10] 田 兵. 多元线性回归分析及其实际应用[J]. 阴山学刊, 2011,25(1):16-19.
- [11] 张凤莲. 多元线性回归中多重共线性问题的解决办法探讨[D]. 广州:华南理工大学,2010.
- [12] 杨永生. 基于状态监测的机械设备可靠性评估模型[J]. 四川兵工学报,2010(7):49-52.