

基于主成分分析和概率神经网络的 变压器局部放电模式识别

苑津莎, 尚海昆

(华北电力大学 电气与电子工程学院, 河北 保定 071003)

摘要: 提出利用主成分分析(PCA)的方法对变压器局部放电原始特征参数进行降维,并提取出新的主成分因子。结果表明,提取出的主成分因子可以很好地表征原始特征向量。通过概率神经网络(PNN)分类器分别对降维前和降维后的特征向量进行训练和识别。研究发现,提取出的新因子有效缓解了分类器负担,且 PNN 分类器的识别效果良好,优于传统 BP 神经网络分类器。

关键词: 主成分分析; 概率神经网络; 变压器; 局部放电; 模式识别

中图分类号: TM 41

文献标识码: A

DOI: 10.3969/j.issn.1006-6047.2013.06.005

0 引言

局部放电是造成电力变压器绝缘故障的主要原因,而在变压器绝缘体系中的放电类型有很多种,不同的放电类型造成的绝缘破坏程度有较大差异,因此对放电类型进行识别是非常重要的。由于变压器超高频检测技术具有检测频率高、信号数据量丰富、抗干扰性好等优点,因此在局部放电检测中具有十分重要的地位^[1-2]。

在局部放电模式识别过程中,模式特征的选择至关重要,直接影响识别结果。目前局部放电常用特征提取方法主要有统计特征参数法^[3-4]、分形特征参数法^[5-6]、图像矩特征参数法^[7]、波形特征参数法^[8]及小波分解法^[9]等,其中统计特征参数法因其物理意义明确及较好的谱图区分能力而得到更多的关注。本文采用统计特征方法提取描述超高频局部放电谱图特征的 37 个统计算子进行分析。

主成分分析 PCA(Principal Component Analysis)方法作为统计学中的有效分析工具,将数据从高维空间变换至低维空间,已被应用到各种领域,并取得了较好的识别和降维效果^[10-14]。本文利用主成分分析方法对局部放电 37 个特征参数进行分析,提取出 12 个更具代表性的新特征因子来表征局部放电图谱。

目前在模式识别领域 BP 神经网络(BPNN)以其固有的模式分类和噪声抑制能力得到了广泛的应用,但存在训练样本需求量大、训练周期长及易陷入局部最小等缺点。本文提出的概率神经网络 PNN(Probabilistic Neural Network)^[15-16]在分类功能上与

最优贝叶斯分类器等价,不像传统神经网络需要用 BP 算法进行反向误差传播的计算,而是完全前向的计算过程,结构简单,训练简洁,有效避免陷入局部极小,容错能力较强,在小样本情况下也能得到满意的识别精度。本文将提取出的主成分因子送入 PNN 进行训练,具有较好的识别效果。

1 主成分分析

主成分分析方法是统计学中对数据分析的有力工具,将高维数据集变换到低维空间,保留最多的原始数据信息。该方法基本思想是通过将原有 m 个数据元素的线性加权组合,构成 r 个互相正交的新元素($r \ll m; r, m \in \mathbf{N}$),并使其能够反映被观测对象的原始特征。这样,由 m 个元素张成的特征空间就被压缩至 r 维,而且在此低维空间中被观测对象会更清楚地表达。此处, r 个互相正交的新元素就称为“主成分”。主成分分析的计算步骤如下。

a. 根据研究问题选择初始分析变量。

选取具有较强相关性的初始变量进行分析,通常采用因子相关性分析的方法进行变量选择,如 KMO(Kaiser-Meyer-Olkin)检验和 Bartlett 球形检验方法。

b. 计算协方差矩阵 C 。

假设 $X = (x_1, x_2, \dots, x_m)^T$ 是 $m \times n$ 的原始特征空间矩阵,其中 m 是数据的样本数, n 是特征参数的个数,协方差矩阵的计算公式如下:

$$C = \frac{1}{m-1} \sum_{j=1}^m (x_j - \bar{x})(x_j - \bar{x})^T \quad (1)$$

其中, $\bar{x}$ 是 X 的平均值, $\bar{x} = \frac{1}{m} \sum_{j=1}^m x_j$, 得到的协方差矩

阵是一个 $m \times m$ 的方阵。

c. 计算协方差矩阵的特征值与特征向量。

首先计算协方差矩阵 C 的特征值 $\lambda_j (j=1, 2, \dots,$

收稿日期:2012-06-21;修回日期:2013-04-03

基金项目:中央高校基本科研业务费专项资金资助项目(13X-S26)

Project supported by the Fundamental Research Funds for the Central Universities(13XS26)

m)与对应的特征向量 $U_{m \times m}$ (正交方阵);然后将各个特征值按大小排序 $\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_m$,并按照此顺序将特征向量矩阵中的各列进行重排,得到新的变换矩阵 T ;最后将原始特征空间矩阵 X 右乘 T ,便得到 m 个新的特征参数 Y 。矩阵运算表示如下:

$$Y=XT \tag{2}$$

d. 确定主成分的个数 r 。

主成分的选择是根据贡献率来决定的,贡献率又可以通过特征值的大小来计算,特征值越大,它在 Y 中对应的新元素的贡献率就越大。一般采用单一贡献率与累计贡献率 2 个参数进行评价。其中单一贡献率 K_r 是指式(2)中 Y 的某列元素的贡献率,而累计贡献率 K_l 则是指 Y 中前 r 列新元素的总贡献率。两者的计算公式分别为:

$$K_r = \frac{\lambda_r}{\sum_{j=1}^m \lambda_j} \tag{3}$$

$$K_l = \frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^m \lambda_j} \tag{4}$$

通常累计贡献率达到 85% 以上时,认为前 r 个元素可以代替原来的 n 个特征参数,即所谓的主成分。这样既能使损失的信息不太多,又达到减少变量、简化问题的目的。

e. 提取主成分。

确定累计贡献率之后,主成分的个数 r 就确定了,式(1)中 Y 的前 r 列元素 $(y_1, y_2, \cdots, y_r)$ 就是所要求的主成分。

2 局部放电特征参数提取

局部放电相位分布 PRPD(Phase Resolved Partial Discharge)模式是一种广泛应用的局部放电模式。本文对超高频局部放电 PRPD 模式的最大放电电量相位分布 $H_{\text{qmax}}(\Phi)$ 、平均放电电量相位分布 $H_{\text{qm}}(\Phi)$ 、放电次数相位分布 $H_n(\Phi)$ 及局部放电幅值分布 $H_n(q)$ 4 个不同二维图谱分别提取出包括偏斜度 S 、陡峭度 K 、局部峰个数 P_e 、相位中值 M 、放电电量因数 Q 、相位不对称度 Φ 、互相关系数 C_c 共 37 个统计算子,各算子分别从不同角度描述谱图特性,其值可根据文献[4]计算得到。各特征参数如表 1 所示。

3 特征参数的主成分及因子分析

上面提取的局部放电特征量维数过高,各参数之间可能存在信息重叠,直接送入分类器进行模式识别不但会给分类器带来较大的负担,而且也可能产生比较大的识别误差。主成分及因子分析的方法可以使提取出的特征量既全面表征放电谱图特征,又

表 1 参数统计表

Tab.1 Statistic parameters

统计参数	$H_{\text{qmax}}(\Phi)$		$H_{\text{qm}}(\Phi)$		$H_n(\Phi)$		$H_n(q)$
	+	-	+	-	+	-	
S	S_1	S_2	S_3	S_4	S_5	S_6	S_7
K	K_1	K_2	K_3	K_4	K_5	K_6	K_7
P_e	P_{e1}	P_{e2}	P_{e3}	P_{e4}	P_{e5}	P_{e6}	P_{e7}
M	M_1	M_2	M_3	M_4	M_5	M_6	M_7
Q	Q_1		Q_2		Q_3		
Φ	Φ_1		Φ_2		Φ_3		
C_c	C_{c1}		C_{c2}		C_{c3}		

注:“+”、“-”分别表示谱图的正、负半周。

使特征量维数有所下降,从而减小分类器负担。下面以偏斜度 S 为例,描述主成分分析过程。

3.1 因子相关性分析

提取出的特征向量之间可能存在较强的相关性,通过因子相关性分析,可以找出反映变量之间共同特性的少数公共因子变量,判断数据是否适合进行因子分析。若各变量之间相关性较低,则很难找出公共因子。下面对表征谱图偏斜度的特征量 S_1 — S_7 进行相关性分析。常用的检验方法有 KMO 检验和 Bartlett 球形检验。KMO 是用于检验因子分析是否适用的指标,若它在 0.5~1 之间,表示适合;小于 0.5 表示不适合。Bartlett 球形检验是通过 χ^2 检验来完成对变量之间是否相互独立的检验。若统计量取值较大,且对应的显著性水平较小,则因子分析是适用的。运用 SPSS 软件对统计出的 S_1 — S_7 变量进行分析,输出检验结果下:KMO 的值为 0.714,变量之间存在较强的偏相关性;Bartlett 球形检验为 134.34,显著性水平为 0,拒绝零假设,变量之间适合进行因子分析,可以进行主成分分析。

3.2 协方差矩阵的主成分分析

针对偏斜度 7 个特征参数,根据式(1)构造协方差矩阵求取主成分,通过求解协方差矩阵的特征值和与之对应的特征向量,将原来的相关向量进行线性变换。对得到的特征值进行分析,选择值大于 1 的特征根作为主成分进行分析,从而达到降维的目的。表 2 列出了协方差矩阵的特征值及方差贡献率。

表 2 特征值和贡献率

Tab.2 Eigenvalues and corresponding contribution rates

因子	特征值	贡献率/%	累计贡献率/%
r_1	4.264	60.912	60.912
r_2	2.077	29.674	90.586
r_3	0.333	4.761	95.346
r_4	0.186	2.653	97.999
r_5	0.108	1.544	99.543
r_6	0.024	0.343	99.886
r_7	0.008	0.114	100.000

由表 2 可以看出,前 2 个特征值大于 1,且 2 个值的累计贡献率超过 90%。贡献率随主成分的变化

情况如图 1 所示。

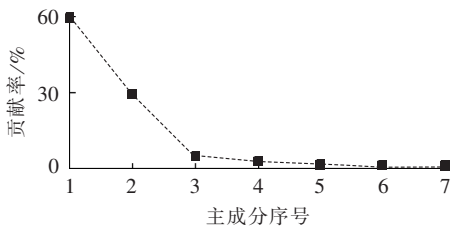


图 1 贡献率随主成分变化图

Fig.1 Variation of components contribution rate along with principal component

由图 1 分析得到,从第 3 个主成分开始,贡献率变化趋势已趋于平稳,且贡献率逐渐减小,其贡献可忽略不计。因此,取前 2 个主成分是比较合适的,可以很好地反映绝大部分变量信息。这样由原来的 7 个指标转化为 2 个新指标,起到了降维作用。

SPSS 分析软件得到主成分系数如表 3 所示。

表 3 主成分系数
Tab.3 Coefficient of principal component

因子	成分系数	
	C_{Y1}	C_{Y2}
r_1	-0.848	0.500
r_2	-0.660	0.680
r_3	0.920	0.356
r_4	0.922	0.150
r_5	0.858	0.437
r_6	0.773	0.382
r_7	-0.281	0.937

由表 3 可得 2 个主成分的线性组合为:

$$Y_1=-0.848r_1-0.660r_2+0.920r_3+0.922r_4+0.858r_5+0.773r_6-0.281r_7 \tag{5}$$

$$Y_2=0.500r_1+0.680r_2+0.356r_3+0.150r_4+0.437r_5+0.382r_6+0.937r_7 \tag{6}$$

用上面两式可计算出主成分的值。

同样的方法可以计算出陡峭度 K 、局部峰个数 P_e 、相位中值 M 、放电量因数 Q 、偏斜度 S 、互相关系数 C_c 的主成分,提取出的成分如表 4 所示。通过主成分分析,提取出 37 个特征变量的 12 个主成分因子,各参数的信息表征率均在 75% 以上。

表 4 各参数主成分提取表
Tab.4 Principal components of parameter

参数	KMO	贡献率 / %	主成分
S	0.714	90.586	P_1, P_2, P_3
K	0.736	89.012	P_4, P_5
P_e	0.639	81.329	P_6, P_7, P_8
M	0.703	75.374	P_9, P_{10}
Q, C_c	0.678	76.150	P_{11}, P_{12}

4 基于 PNN 的局部放电模式识别

4.1 PNN

PNN 是由径向基函数网络发展而来的一种前馈

神经网络,其理论依据是贝叶斯最小风险准则(贝叶斯决策理论),PNN 作为径向基网络的一种,适合于模式分类。

基于 PNN 的故障诊断方法是概率统计学中被普遍接受的决策方法。假设 2 种已知故障模式 A 、 B ,对于待识别的故障特征样本为 X ,按贝叶斯最小风险准则,若 $H_A L_A F_A(X) > H_B L_B F_B(X)$,则 $X \in A$;若 $H_A L_A F_A(X) < H_B L_B F_B(X)$,则 $X \in B$ 。其中, H_A 和 H_B 代表故障模式 A 和 B 的先验概率, L_A 是将属于 A 的故障特征样本 X 错误划分到模式 B 的代价因子, L_B 为将属于 B 的故障特征样本 X 错误划分到模式 A 的代价因子, F_A 、 F_B 是故障模式 A 、 B 的概率密度函数。一般 F_A 、 F_B 不能精确获得,只能根据现有故障特征样本求其统计值,概率密度函数的估计式如下:

$$F_A(X)=\frac{1}{(2\pi)^{p/2}\delta^p}\frac{1}{m_i}\sum\exp\left[-\frac{(X-X_{Ai})^T(X-X_{Ai})}{2\delta^2}\right] \tag{7}$$

其中, X_{Ai} 为故障模式 A 的第 i 个训练向量; m_i 为故障模式 A 的训练样本总数; δ 为平滑参数,其值确定了以样本为中心的钟状曲线的宽度。

PNN 的层次模型由输入层、模式层、求和层、输出层组成,基本结构如图 2 所示。输入层将输入样本传递给模式层的各个节点,模式层将输入节点传递来的输入向量进行加权求和,经过一个非线性算子运算后传递给求和层;求和层只是简单地将由对应与训练样本中同一类的模式层传来的输入进行求和;输出层选择求和层中输出最大者对应的状态模式为分类结果。

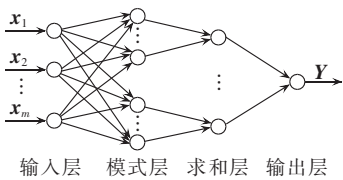


图 2 概率神经网络基本结构

Fig.2 Basic structure of probabilistic neural network

4.2 基于 PNN 的局部放电模式识别

4.2.1 网络结构设计

a. 输入层向量选择。

对于局部放电信号,网络输入即为表征放电信号图谱特征的特征参数,包括上面介绍过的偏斜度、陡峭度、局部峰个数等 37 个统计特征参数,以及经主成分分析方法降维后的 12 个新的特征量用作对比分析,为消除量纲影响,所有输入层向量均采用标准化后的数据。

b. 输出层向量选择。

变压器局部放电类型有许多种,本文采用比较典型的 3 种放电类型作为网络结构的输出,即自由金属颗粒悬浮放电、绝缘表面沿面放电和固体绝缘

内部气隙放电。因此,采用 PNN 进行模式分类时,输入层对应 37 个节点或 12 个节点,输出层对应 3 个输出节点,网络的输出即为诊断出的局部放电类型。文中用作对比分析的 BPNN 输入层及输出层节点与 PNN 相同,隐含层取 16 个(对应 37 个输入向量)和 14 个(对应 12 个输入向量)神经元。

4.2.2 结果分析

本文针对悬浮放电、沿面放电和气隙放电 3 种不同局部放电类型,在各地高压变电站的局部放电在线监测系统中收集得到现场超高频放电信号,对每种类型放电收集得到 10 组历史数据,并对 30 组放电数据进行参数统计,计算出相应 37 个表征二维谱图的统计算子。利用上面的主成分分析方法提取出 12 个新的特征向量,送入 PNN 进行训练。3 种类型放电分别取 5 组数据用来训练,5 组用来测试。此外将未经主成分因子分析的 37 个统计算子送入分类器进行训练和测试,识别结果如图 3 所示。

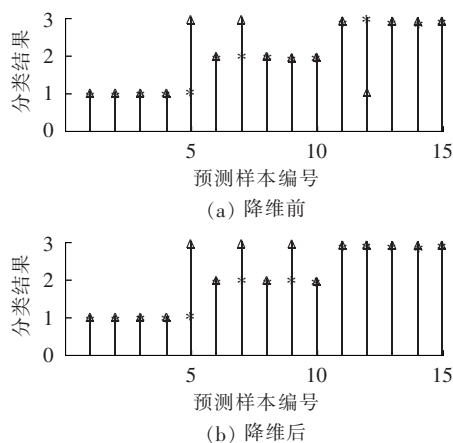


图 3 降维前和降维后的 PNN 识别效果图
Fig.3 Results of PNN recognition before and after dimension reduction

由图 3 可见,样本降维后 12 个统计参数的识别效果较之降维前的 37 个统计算子的识别情况相同,错误判断个数均为 3。为进行实验对比,将相同数据输入传统 BPNN 进行训练并测试,对比结果见表 5。

表 5 实验对比结果

Tab.5 Comparison of experimental results

参数	降维前(37 个算子)		降维后(12 个算子)	
	BPNN	PNN	BPNN	PNN
收敛时间/s	1.72	0.85	1.33	0.75
识别率/%	73.3	80.0	66.7	80.0

由表 5 可以看出,数据降维后,PNN 与 BPNN 的收敛时间均有不同程度的降低。PNN 算法在数据降维前后具有相同的识别率,BPNN 算法在数据降维后识别率略有下降。此外,PNN 的总体收敛时间要少于 BPNN,识别效果也要优于 BPNN。

5 结论

本文利用主成分分析方法对局部放电特征数据

进行分析,降维后的新向量可以较好地表征原始特征向量,有效解决了原始特征参数维数过高的问题。通过 PNN 与 BPNN 分类器对降维前后的特征参数进行模式分类,对比得出,降维后的新特征量使得网络结构得到了优化,PNN 与传统的 BPNN 相比具有更好的识别效果。

参考文献:

[1] 王国利,郝艳捧,李彦明. 电力变压器局部放电检测技术的现状和发展[J]. 电工电能新技术,2001,20(2):52-57.
WANG Guoli,HAO Yanpeng,LI Yanming. Present and future development of partial discharge detection in power transformers [J]. Advanced Technology of Electrical Engineering and Energy,2001,20(2):52-57.

[2] JUDD M D,CLEARY G P,BENNOCH C J. Applying UHF partial discharge detection to power transformers[J]. IEEE Power Engineering Review,2002,22(8):57-59.

[3] 胡文堂,高胜友,余绍峰,等. 统计参数在变压器局部放电模式识别中的应用[J]. 高电压技术,2009,35(2):277-281.
HU Wentang,GAO Shengyou,YU Shaofeng,et al. Application of statistic parameters in recognition of partial discharge in transformers[J]. High Voltage Engineering,2009,35(2):277-281.

[4] GULSKI E,KREUGER F H. Computer-aided recognition of discharge sources[J]. IEEE Transactions on Electrical Insulation, 1992,27(1):82-92.

[5] 任先文,薛雷,宋阳,等. 基于分形特征的最小二乘支持向量机局部放电模式识别[J]. 电力系统保护与控制,2011,39(14):143-147.
REN Xianwen,XUE Lei,SONG Yang,et al. The pattern recognition of partial discharge based on fractal characteristics using LS-SVM[J]. Power System Protection and Control,2011,39(14): 143-147.

[6] SATISH L,ZAENGL W S. Can fractal features be used for recognizing 3-D partial discharge patterns?[J]. IEEE Transactions on Dielectrics and Electrical Insulation,1995,2(3):352-359.

[7] 张晓虹,张亮,乐波,等. 基于局部放电的矩特征分析大电机主绝缘的老化[J]. 中国电机工程学报,2002,22(5):94-98.
ZHANG Xiaohong,ZHANG Liang,LE Bo,et al. Analysis on aging condition of stator winding insulation of generator based on the moment characteristics of partial discharge [J]. Proceedings of the CSEE,2002,22(5):94-98.

[8] MAZROUA A A. Discrimination between PD pulse shapes using different neural network paradigms [J]. IEEE Transactions on Dielectrics and Electrical Insulation,1994,1(6):1119-1131.

[9] 淡文刚,陈祥训,郑健超. 采用小波分析与神经网络技术的局部放电统计识别方法[J]. 中国电机工程学报,2002,22(9):1-5.
DAN Wengang,CHEN Xiangxun,ZHENG Jianchao. Classification of partial discharge distribution patterns using wavelet transform and neural network[J]. Proceedings of the CSEE,2002,22(9):1-5.

[10] 刘宝英,杨仁刚. 基于主成分分析的最小二乘支持向量机短期负荷预测模型[J]. 电力自动化设备,2008,28(11):13-17.
LIU Baoying,YANG Rengang. Short-term load forecasting model based on LS-SVM with PCA [J]. Electric Power Automation Equipment,2008,28(11):13-17.

[11] 廖瑞金,杨丽君,孙才新,等. 基于局部放电主成分因子向量的油纸绝缘老化状态统计分析[J]. 中国电机工程学报,2006,26

- (14):114-119.
- LIAO Ruijin,YANG Lijun,SUN Caixin,et al. Aging condition assessment of oil-paper based on principal component and factor analysis of partial discharge[J]. Proceedings of the CSEE, 2006,26(14):114-119.
- [12] 李岳,吕克洪. 主成分分析在铁谱磨粒识别中的应用研究[J]. 国防科技大学学报,2004,26(1):89-94.
- LI Yue,LÜ Kehong. The application of the Principal Components Analysis (PCA) to debris recognition[J]. Journal of National University of Defense Technology,2004,26(1):89-94.
- [13] 许力,竺鹏东,顾宏杰,等. 基于变尺度 PCA 的电力设备载流故障早期预警[J]. 电力自动化设备,2012,32(5):147-151.
- XU Li,ZHU Pengdong,GU Hongjie,et al. Early warning of electric equipment current-carrying faults based on variable-scale principal component analysis[J]. Electric Power Automation Equipment,2012,32(5):147-151.
- [14] 王辉,王克奇. 基于主分量分析木材纹理的特征提取[J]. 森林工程,2006,22(6):14-16.
- WANG Hui,WANG Keqi. The extraction of wood texture features based on PCA[J]. Forest Engineering,2006,22(6):14-16.
- [15] 石敏,吴正国,徐袭. 基于概率神经网络和双小波的电能量扰动自动识别[J]. 电力自动化设备,2006,26(3):5-8.
- SHI Min,WU Zhengguo,XU Xi. Auto recognition of power quality disturbance based on probabilistic neural networks and double wavelet[J]. Electric Power Automation Equipment,2006,26(3):5-8.
- [16] 王柏生,倪一清,高赞明. 用概率神经网络进行结构损伤位置识别[J]. 振动工程学报,2001,14(1):60-64.
- WANG Baisheng,NI Yiqing,GAO Zanming. Structural damage localization using probabilistic neural network[J]. Journal of Vibration Engineering,2001,14(1):60-64.

作者简介:

苑津莎(1957-),男,河北保定人,教授,博士研究生导师,研究方向为智能信息处理、电磁场理论及应用;
尚海昆(1984-),男,河北保定人,博士研究生,研究方向为电气设备故障诊断(E-mail:shikey007@sina.com)。

Pattern recognition based on principal component analysis and probabilistic neural networks for partial discharge of power transformer

YUAN Jinsha,SHANG Haikun

(College of Electrical and Electronic Engineering,North China Electric Power University,Baoding 071003,China)

Abstract: For reducing the high dimension of original characteristic parameters in partial discharge pattern recognition of power transformer,PCA(Principal Component Analysis) is applied to extract new principal component factors,which represents the original characteristic parameters sufficiently. PNN(Probabilistic Neural Network) classifier is used to train and recognize the characteristic vectors before and after the dimension reduction respectively. It is found that,the extraction of new principal component factors mitigates effectively the load of PNN classifier and its effectiveness of pattern recognition is better than that of traditional BPNN classifier.

Key words: principal component analysis; probabilistic neural network; power transformers; partial discharge; pattern recognition

(上接第 20 页 continued from page 20)

Two-level planning of transmission system with optimal placement of efficiency power plants

ZHENG Jing¹,WEN Fushuan¹,LI Li²,WANG Ke²,GAO Chao²

(1. School of Electrical Engineering,Zhejiang University, Hangzhou 310027, China;

2. Guangdong Power Dispatching Center, Guangzhou 510600, China)

Abstract: The principle of optimal EPP(Efficiency Power Plant) placement is proposed and with the consideration of the optimal EPP placement,a two-level planning model based on two-level planning theory is developed for transmission system,of which,the upper model takes the minimum total cost of transmission system and EPP investment as its objective while the lower model takes the minimum cost of EPP investment as its objective. With the interaction between two levels,the transmission system planning scheme is finally determined by the upper model. A hybrid algorithm combining genetic algorithm and primal-dual interior method is adopted to solve the proposed planning model. Results of case study for 18-bus and 46-bus systems verify the feasibility and efficiency of the proposed model and algorithm.

Key words: transmission system; efficiency power plant; planning; genetic algorithms; primal-dual interior method; models