

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2016 年 6 月 30 日 (30.06.2016)



(10) 国际公布号
WO 2016/101690 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2015/092589
- (22) 国际申请日: 2015 年 10 月 23 日 (23.10.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201410806481.5 2014 年 12 月 22 日 (22.12.2014) CN
- (71) 申请人: 国家电网公司 (STATE GRID CORPORATION OF CHINA) [CN/CN]; 中国北京市西城区西长安街 86 号, Beijing 100031 (CN)。 国网山东省电力公司电力科学研究院 (ELECTRIC POWER RESEARCH INSTITUTE OF STATE GRID SHANDONG ELECTRIC POWER COMPANY) [CN/CN]; 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。
- (72) 发明人: 郭志红 (GUO, Zhihong); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。

杜修明 (DU, Xiuming); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。 杨祎 (YANG, Yi); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。 陈玉峰 (CHEN, Yufeng); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。 盛戈峰 (SHENG, Gehao); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。 马强 (MA, Qiang); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。 严英杰 (YAN, Yingjie); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。 辜超 (GU, Chao); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。 朱文兵 (ZHU, Wenbing); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。 朱振华 (ZHU, Zhenhua); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。 郑建 (ZHENG, Jian); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。 王辉 (WANG, Hui); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。 李秀卫 (LI, Xiuwei); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。

[见续页]

(54) Title: TIME SEQUENCE ANALYSIS-BASED STATE MONITORING DATA CLEANING METHOD FOR POWER TRANSMISSION AND TRANSFORMATION DEVICE

(54) 发明名称: 基于时间序列分析的输变电设备的状态监测数据清洗方法

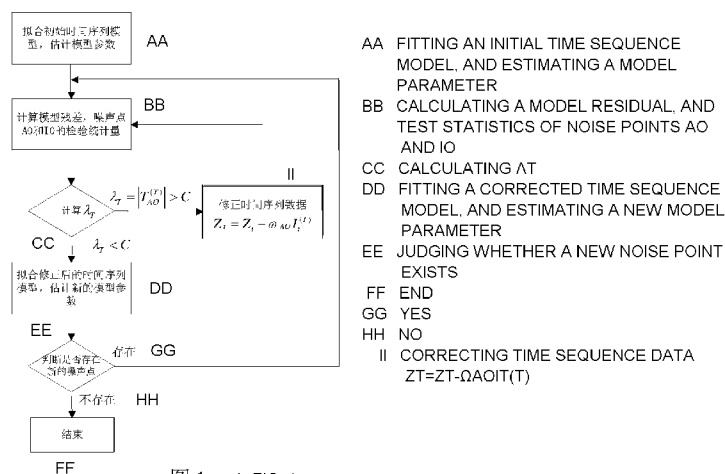


图 1 / FIG. 1

(57) Abstract: A time sequence analysis-based state monitoring data cleaning method for a power transmission and transformation device, comprising the steps of: establishing an initial time sequence model for state data of a power transmission and transformation device, and estimating an initially-fit residual sequence and residual variance by means of the initial time sequence model; calculating a test statistic of each observation point utilizing the initial time sequence model; judging whether a maximum value of the absolute values of test statistics is greater than a set threshold value; and calculating the test statistic of each of the observation points by means of a model residual of a corrected time sequence model, and judging whether a new noise point exists according to whether the maximum value of the absolute values of the test statistics is greater than the set threshold value, until all the noise points are identified. The method has the advantages of a high cleaning efficiency, keeping the data integrity, and avoiding the loss of useful information about data.

(57) 摘要:

[见续页]

Shandong 250002 (CN)。朱孟兆 (ZHU, Mengzhao); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。周加斌 (ZHOU, Jiabin); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。马艳 (MA, Yan); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。段玉兵 (DUAN, Yubing); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。刘辉 (LIU, Hui); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。李明 (LI, Ming); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。赵富强 (ZHAO, Fuqiang); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。许光可 (XU, Guangke); 中国山东省济南市市中区望岳路 2000 号, Shandong 250002 (CN)。

- (74) 代理人: 济南圣达知识产权代理有限公司 (JINAN SHENGDA INTELLECTUAL PROPERTY AGENCY CO., LTD); 中国山东省济南市历下区经十路 17703 号华特广场 B308, Shandong 250061 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第 21 条(3))。

一种基于时间序列分析的输变电设备的状态监测数据清洗方法, 步骤如下: 建立输变电设备状态数据的初始时间序列模型, 通过初始时间序列模型估计初始拟合的残差序列和残差方差; 利用初始时间序列模型, 计算每个观测点的检验统计量; 判断检验统计量的绝对值的最大值是否大于设定阈值; 通过修正后的时间序列模型的模型残差计算每个观测点的检验统计量, 根据检验统计量的绝对值的最大值是否大于设定阈值来判断是否存在新的噪声点, 直到所有的噪声点都被识别出。上述方法具有清洗效率高、保持了数据的完整性、避免数据有用信息丢失的优点。

基于时间序列分析的输变电设备的状态监测数据清洗方法

技术领域

本发明涉及基于时间序列分析的输变电设备的状态监测数据清洗方法。

背景技术

对设备在线监测、带电检测、离线试验等设备全景状态信息进行全方位分析，提升输变电设备评价与异常诊断的准确性是设备状态评估诊断技术的发展趋势。输变电设备全景状态信息呈现来源多、信息异构、数量庞大、属性繁多等特点，其数据往往是不完整的、有噪声的和不一致的。状态量原始的数据质量往往不能满足后续状态评价模型的要求，因此在状态评估或诊断分析之前进行数据清洗是必不可少的。数据清洗通过填充缺失值、平滑噪声数据和识别离群点来提高数据质量，有助于提高数据挖掘过程的准确率和效率。

在输变电设备数据清洗方面,国内外现有的研究较少。文献《基于贝叶斯网络分类器的变压器综合故障诊断方法》在建立故障与信息的映射关系时将海量数据通过粗糙集信息熵的方法进行了约简，从而解决了数据缺失的问题，但是破坏了数据自身信息的完整性。

文献《A Kernel Fuzzy c-Means Clustering-Based Fuzzy Support Vector Machine Algorithm for Classification Problems with Outliers or Noises》和《FSVM-CIL: Fuzzy Support Vector Machines for Class Imbalanced Learning》在处理支持向量机训练集的噪声和异常数据时使用了模型 C 均值聚类方法，通过计算数据到聚类中心的距离来分离出噪声数据。但是这种聚类方法将分离出的噪声数据直接剔除，破坏了状态量数据链的连续性。以上研究在数据清洗过程中都丢弃了造成了数据的丢失，不利于在后续状态评估中对数据本身信息的挖掘。

发明内容

本发明的目的就是为了解决上述问题，提供一种基于时间序列分析的输变电设备的状态监测数据清洗方法，它具有清洗效率高、保持了数据的完整性、避免数据有用信息丢失的优点。

为了实现上述目的，本发明采用如下技术方案：

基于时间序列分析的输变电设备的状态监测数据清洗方法，步骤如下：

步骤（1）：建立输变电设备状态数据的初始时间序列模型，通过初始时间序列模型估计初始拟合的残差序列和残差方差；

步骤（2）：

步骤（2-1）：利用步骤（1）的初始时间序列模型，计算每个观测点的检验统计量；

步骤 (2-2): 判断检验统计量的绝对值的最大值是否大于设定阈值, 如果大于, 则确定存在对初始时间序列模型拟合影响的噪声点, 通过修正时间序列数据来修正噪声点, 通过返回步骤 (2) 继续识别并修正时间序列所有的噪声点, 然后进入步骤 (2-3); 如果小于就进入步骤 (2-3);

步骤 (2-3): 拟合修正后的时间序列模型, 估计修正后的时间序列模型的模型残差;

步骤 (3): 通过步骤 (2-3) 中修正后的时间序列模型的模型残差计算每个观测点的检验统计量, 根据检验统计量的绝对值的最大值是否大于设定阈值来判断是否存在新的噪声点, 直到所有的噪声点都被识别出, 如果存在就返回步骤 (2), 如果不存在就结束。

所述步骤 (1) 的步骤为:

对观测序列 Z_t 建立时间序列模型,

设 Z_t 是单个状态量的无异常值的时间序列, Z_t 服从 $ARIMA(p, d, q)$ 模型, 表示为

$$Z_t = \frac{\hat{\theta}(B)}{\hat{\phi}(B)\nabla^d} e_t \quad (5)$$

$$\hat{\theta}(B) = 1 - \hat{\theta}_1 B - \hat{\theta}_2 B^2 - \dots - \hat{\theta}_q B^q \quad (6)$$

$$\hat{\phi}(B) = 1 - \hat{\phi}_1 B - \hat{\phi}_2 B^2 - \dots - \hat{\phi}_q B^q \quad (7)$$

其中, Z_t 为观测序列, p, q, d 为时间序列 $ARIMA(p, d, q)$ 模型的阶数, $\theta(B)$ 为初始拟合的平稳算子, $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ 是 $\theta(B)$ 中相应的参数, $\varphi(B)$ 为初始拟合的可逆算子, $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_q$ 是 $\varphi(B)$ 中相应的参数, e_t 是不含噪声点的序列初始拟合的残差序列, B 表示延迟算子, $B^d Z_t = Z_{t-d}$, B^d 表示延迟 d 个时刻的延迟算子, Z_t 表示时刻 t 的观测值, Z_{t-d} 表示时刻 $t-d$ 的观测值, ∇^d 是一个表示 t 时刻与 $t-d$ 时刻的观测值间的差值的算子, $\nabla^d = 1 - B^d$ 。

并由所建立的时间序列模型计算初始拟合的残差序列, 即:

$$\hat{e}_t = \pi(B)Z_t = \frac{\varphi(B)\nabla^d}{\theta(B)} Z_t \quad (8)$$

式 (8) 中, Z_t 为观测序列, $\hat{e}_t$ 为含噪声点的序列初始拟合的残差序列, $\theta(B)$ 为初始拟合的平稳算子, $\varphi(B)$ 为初始拟合的可逆算子, B 表示延迟算子, $B^d Z_t = Z_{t-d}$, B^d 表示延迟 d 个时刻的延迟算子, Z_t 表示时刻 t 的观测值, Z_{t-d} 表示时刻 $t-d$ 的观测值, ∇^d 是一个表示 t

时刻与 $t-d$ 时刻的观测值间的差值的算子, $\nabla^d = 1 - B^d$; $\pi(B)$ 为表征残差影响的算子, $\hat{\pi}_1, \hat{\pi}_2, \dots, \hat{\pi}_n$ 为 $\pi(B)$ 相应的参数。

所述 $\pi(B)$ 的定义如 (9) 所示:

$$\hat{\pi}(B) = \frac{\hat{\theta}(B)}{\hat{\phi}(B)\nabla^d} = (1 - \hat{\pi}_1 B - \hat{\pi}_2 B^2 - \dots - \hat{\pi}_n B^n) \quad (9)$$

其中, $\theta(B)$ 为初始拟合的平稳算子, $\phi(B)$ 为初始拟合的可逆算子, n 为残差序列长度, $n \geq 1$; B 表示延迟算子, $B^d Z_t = Z_{t-d}$, $\hat{\pi}_n B^n$ 表示表征滞后 n 个时刻的残差影响的算子;

所述步骤 (1) 的残差方差的初始估计为:

$$\sigma_a^2 = \frac{1}{n} \sum_{j=1}^n \hat{e}_j^2 \quad (10);$$

其中, σ_a^2 表示残差方差, n 表示残差序列的长度, $n \geq 1$, $\hat{e}_j^2$ 的参数含义是残差值的平方, j 的取值范围是 $1 \leq j \leq n$ 。

所述步骤 (2-1) 的步骤为:

利用步骤 (1) 的初始时间序列模型, 计算每个观测点的检验统计量步骤为: 利用步骤 (1) 的初始时间序列模型, 对 $t=1, 2, \dots, n$, 计算每个观测点的检验统计量 T'_{AO} 和检验统计量 T'_{IO} ; 所述检验统计量 T'_{AO} 和检验统计量 T'_{IO} 的计算公式如下:

$$T'_{IO} = \frac{\omega_{IO}}{\sigma_a} \quad (11)$$

$$T'_{AO} = \frac{\omega_{AO}}{\sigma_a} \sqrt{\sum_{j=T}^n \pi_j^2} \quad (12)$$

式 (11)、(12) 中: $t=1, 2, \dots, n$ 表示噪声点产生的时刻; σ_a 为含噪声点的残差 $\hat{e}_t$ 的标准差; 所述检验统计量 T'_{AO} 和检验统计量 T'_{IO} 的极限分布均为标准正态分布, 服从 t 分布。

所述步骤 (2-2) 的步骤为:

将噪声点的种类分为加性异常值 (Additive Outliers, AO) 和新息异常值 (Innovational Outliers, IO), 缺失值被认为是加性异常值。

定义 $\lambda_t = \max \{ |T'_{AO}|, |T'_{IO}| \}$, λ_t 表示检验统计量 T'_{AO} 和检验统计量 T'_{IO} 的最大值, T 为最大值发

生的时刻;

当 $\lambda_T = |T_{AO}^{(T)}| > C$ 时, 其中 C 是设定阈值, 确定在时刻 T 存在加性异常值 (Additive Outliers, AO), 加性异常值 AO 对模型拟合的影响 ω_{AO} 通过式 (13) 求得;

$$\hat{\omega}_{AO} = \frac{e_T - \sum_{j=1}^{n-T} \pi_j e_{T+j}}{\sum_{j=0}^{n-T} \pi_j^2} \quad (13)$$

其中, $\hat{\omega}_{AO}$ 表示加性异常值的影响, e_T 表示时刻 T 处的残差, e_{T+j} 表示时刻 $T+j$ 处的残差, π_j 表示时刻 j 处残差影响的算子, j 的取值范围是 $1 \leq j \leq n$ 。

$$\hat{\omega}_{IO} = e_T \quad (14)$$

其中, $\hat{\omega}_{IO}$ 表示新息异常值的影响, e_T 表示时刻 T 处的残差;

对于加性异常值, 通过式 (15) 修正初始时间序列数据, 得到新的时间序列 Z_t :

$$Z_t = Z_t - \omega_{AO} I_t^{(T)} \quad (15)$$

其中, Z_t 表示经过一次修正的时间序列, Z_t 表示初始时间序列, ω_{AO} 表示加性异常值的影响, $I_t^{(T)}$ 表示阶跃函数;

并由式 (16) 修正后的残差 $\tilde{e}_t$:

$$\tilde{e}_t = \hat{e}_t - \omega_{AO} \pi(B) I_t^{(T)} \quad (16)$$

其中, $\tilde{e}_t$ 表示经过一次修正的残差, $\hat{e}_t$ 表示初始残差, ω_{AO} 表示加性异常值的影响, $\pi(B)$ 表示残差影响的算子, $I_t^{(T)}$ 表示阶跃函数。

式 (15) 和式 (16) 中 $I_t^{(T)}$ 的定义如式 (17) 所示:

$$I_t^{(T)} = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases} \quad (17)$$

其中, $I_t^{(T)}$ 表示阶跃函数, t 表示时间, T 表示阶跃函数值由 0 转为 1 的时刻;

当 $\lambda_T = |T_{IO}^{(T)}| > C$ 时, 确定在时刻 T 存在新息异常值 (Innovational Outliers, IO), 其对初始时间序列模型拟合的影响 ω_{IO} 通过式 (14) 求得, 利用式 (18) 修正数据, 则 IO 的影响能够消除, 即

$$Z_t = Z_t - \frac{\theta(B)}{\varphi(B)\nabla^d} \omega_{IO} I_t^{(T)} \quad (18)$$

其中, Z_t 表示经过一次修正的时间序列, Z_t 表示初始时间序列, ω_{IO} 表示新息异常值的影响, $I_t^{(T)}$ 表示阶跃函数, $\theta(B)$ 为初始拟合的平稳算子, $\varphi(B)$ 为初始拟合的可逆算子, n 为残差序列长度, $n \geq 1$; B 表示延迟算子, $B^d Z_t = Z_{t-d}$, B^d 表示延迟 d 个时刻的延迟算子, Z_t 表示时刻 t 的观测值, Z_{t-d} 表示时刻 $t-d$ 的观测值, ∇^d 是一个表示 t 时刻与 $t-d$ 时刻的观测值间的差值的算子, $\nabla^d = 1 - B^d$ 。

由式 (19) 修正后的残差:

$$\tilde{e}_t = \hat{e}_t - \omega_{IO} I_t^{(T)} \quad (19)$$

其中, $\tilde{e}_t$ 表示经过一次修正的残差, $\hat{e}_t$ 表示初始残差, ω_{IO} 表示新息异常值的影响, $I_t^{(T)}$ 表示阶跃函数;

使用迭代的方法识别并修正时间序列所有的噪声点。在修正后的残差 $\tilde{e}_t$ 和残差标准差 σ_a^2 的基础上再次计算每个观测点的检验统计量 T_{AO}^t 和检验统计量 T_{IO}^t , 并重复步骤 (2-1), 直到所有的噪声点都被识别出来。

所述步骤 (2-3) 的步骤为:

假设有 k 个异常数据在时刻 $T_1, T_2, \dots, T_k$ 被识别出, 其影响分别为 $\omega_1^{(1)}, \omega_2^{(1)}, \dots, \omega_k^{(1)}$ 同时异常数据被修正而得到了新的时间序列 $z_t^{(1)}$, $z_t^{(1)}$ 代表是第一步迭代得到的序列;

利用新的时间序列拟合得到经过一次修正后的时间序列模型;

此时根据式 (5) 和式 (8) 重新估计该时间序列参数 $\theta^{(1)}(B), \varphi^{(1)}(B), \pi^{(1)}(B)$, 并得到经过一次修正后的时间序列模型残差 $\tilde{e}_t^{(1)}$:

$$\tilde{e}_t^{(1)} = \pi^{(1)}(B) [Z_t^{(1)} - \sum_{j=1}^k \omega_j^{(1)} v_j^{(1)}(B) I_t^{(T_j)}] \quad (20)$$

其中

$$v_j^{(1)}(B) = \begin{cases} \frac{\theta^{(1)}(B)}{\varphi^{(1)}(B)\nabla^d} & IO \\ 1 & AO \end{cases} \quad (21)$$

其中, $\tilde{e}_t^{(1)}$ 表示经过一次修正后的时间序列模型残差, $\pi^{(1)}(B)$ 表示经过一次修正的残差的影响算子, $z_t^{(1)}$ 表示经过一次修正的时间序列, $\omega_j^{(1)}$ 表示噪声点 AO 或 IO 的影响因子, $\tilde{v}_j^{(1)}(B)$ 表示噪声点 AO 或 IO 的修正因子, $I_j^{(T_j)}$ 表示阶跃函数。

所述步骤 (3) 的步骤为:

找出是否有新的噪声点出现, 直到所有的噪声点都被识别出; 在最后一次迭代后, 针对修正了噪声点的时间序列 z_t 进行联合估计, 得到最终的时间序列模型:

$$Z_t = \sum_{j=1}^k \omega_j v_j(B) I_j^{(T_j)} + \frac{\theta(B)}{\varphi(B) \nabla^d} a_t \quad (22)$$

式 (22) 中, 各参数 $\theta(B), \varphi(B), \omega_j, v_j$ 是在最后一次迭代中得到的;

其中, z_t 表示经过多次修正后最终的时间序列, ω_j 表示噪声点 AO 或 IO 的影响因子, $v_j(B)$ 表示噪声点 AO 或 IO 的修正因子, $I_j^{(T_j)}$ 表示阶跃函数, $\varphi(B)$ 表示最终拟合的可逆算子, $\theta(B)$ 表示最终拟合的平稳算子, a_t 表示最终的残差, B 表示延迟算子, $B^d Z_t = Z_{t-d}$, B^d 表示延迟 d 个时刻的延迟算子, Z_t 表示时刻 t 的观测值, Z_{t-d} 表示时刻 $t-d$ 的观测值, ∇^d 是一个表示 t 时刻与 $t-d$ 时刻的观测值间的差值的算子, $\nabla^d = 1 - B^d$;

至此, 针对输变电设备状态数据的数据清洗完成, 清洗后的序列为 z_t , 清洗后的序列即为修正了噪声点的时间序列。

本发明的有益效果:

1 相比于传统的删除噪声点, 本方法清洗出的数据是不带有噪声点和缺失值的数据, 从而避免了时间序列中 useful 信息的丢失, 更能有效地反映原始时间序列的动态变化, 适应输变电设备状态数据的特点。

2 本方法主要用于清洗输变电设备状态数据。在输变电设备的状态数据序列中噪声点是指由于仪器异常或设备系统的扰动引起的严重偏离期望值的数据, 这些数据不仅会影响模型拟合的精度, 而且会导致后续状态评估出现偏差, 引起误诊。缺失值是指由于传感器的短时失效、通信端口异常、记录失误等因素引起的数据中断, 状态数据中存在的缺失值破坏了系统运行的连续性, 不利于后续的状态评估和趋势检验。通过本方法清洗后的输变电设备状态数据是有效的, 能够为后续的设备状态评估、风险评估、故障预警等提供有力的数据支撑。

3 本方法的时间序列模型充分结合了输变电设备状态数据的特点, 能够确保清洗的效率和清洗后数据的可用性。输变电设备状态量的检测是由各个传感器来完成的, 但是经过底层的

预处理而上传到数据库进行状态评估的原始数据可以认为是按时间序列排列的特征量数据，因此本方法的时间序列模型具有很强的适用性。

4 对输变电设备的状态数据，设备状态信息获取方式的多样性及采集间隔的不确定性使得各状态量时间序列的参数是未知的、异常数据产生的时刻 T 是不确定的，针对噪声点出现时刻与个数未知、预先没有模型参数的情况，本方法能识别出噪声点的产生时刻，并反复修正噪声点，清洗后的序列与正常序列较为相近。

附图说明

图 1 为数据清洗流程图；

图 2 为原始数据和含有异常值的数据；

图 3 为原始数据和初始拟合残差序列；

图 4 为原始数据、最终拟合的数据和最终拟合残差序列。

具体实施方式

下面结合附图与实施例对本发明作进一步说明。

本发明的原理是利用时间序列模型识别各状态量的时间序列，根据序列中噪声点的种类选择不同的修正公式，从而达到修正噪声点数据和填补缺失值的目的，完成对状态量的数据清洗。该方法既能识别出数据中的噪声点和缺失值，又能在分离噪声的过程中噪声点数值进行修正。其总体流程如图 1 所示。

输变电设备状态量的检测是由各个传感器来完成的，但是经过底层的预处理而上传到数据库进行状态评估的原始数据可以认为是按时间序列排列的特征量数据。这些数据的统一格式为“时间.特征量=数值”，因此可认为采集的所有状态量形成了一个单元或多元的连续而完整的时间序列，如矩阵 X 所示：

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix} \quad (1)$$

式中， X_{mn} 代表在时刻 T_n 状态量 m 的数值。

设 x_t 是单个状态量的无异常值的时间序列， x_t 服从 $ARIMA(p, d, q)$ 模型，可表示为

$$X_t = \frac{\theta(B)}{\varphi(B)\nabla^d} a_t \quad (2)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \quad (3)$$

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_q B^q \quad (4)$$

以上式中： $\theta(B)$ 和 $\varphi(B)$ 是没有公共因子的平稳和可逆算子， a_t 是相互独立，具有相同分布 $N(0, \sigma_a^2)$ 的白噪声序列， $\nabla = 1 - B$ 表示延迟算子， $BX_t = X_{t-1}$ ， $B^q X_t = X_{t-q}$ 。

设 Z_t 是单个状态量的无异常值的时间序列， Z_t 服从 $ARIMA(p, d, q)$ 模型，可表示为

$$Z_t = \frac{\hat{\theta}(B)}{\hat{\varphi}(B)\nabla^d} e_t \quad (5)$$

$$\hat{\theta}(B) = 1 - \hat{\theta}_1 B - \hat{\theta}_2 B^2 - \dots - \hat{\theta}_q B^q \quad (6)$$

$$\hat{\varphi}(B) = 1 - \hat{\varphi}_1 B - \hat{\varphi}_2 B^2 - \dots - \hat{\varphi}_q B^q \quad (7)$$

以上式中： $\hat{\theta}(B)$ 和 $\hat{\varphi}(B)$ 是没有公共因子的平稳和可逆算子， e_t 是不含噪声点的序列初始拟合的残差序列， $\nabla = 1 - B$ 与(2)中定义相同。

设备状态信息获取方式的多样性，采集间隔的不确定性使得各状态量时间序列的参数是未知的、异常数据产生的时刻 T 是不确定的，因此时间序列模型的搭建、模型参数估计、异常数据类型识别是必不可少数据清洗步骤。由于异常数据的存在将使时间序列参数的估计产生偏差，因此针对噪声点出现时刻与个数未知，预先没有模型参数的情况下，对观测的时间序列进行反复的拟合与修正，从而达到数据清洗的目的，其整体流程如图1所示。

其中 C 是预先确定的正常数，通常取3和4之间的值。

南网某输电线路采集的导线温度数据如图2中实线所示，该时间序列不存在噪声点和缺失值。为了检验本文数据清洗方法的实用性，将原来的观测时刻 $t=140$ 的数值剔除（成为缺失点），观测时刻 $t=26$ 和 $t=49$ 分别加入一个AO和IO异常值，从而生成了一个带清洗的时间序列 Z_t ，如图2虚线所示。

首先根据发明内容中所述步骤S1建立导线温度数据的时间序列ARMA模型，如下所示。

$$Z_t = \frac{1}{(1 - 0.914B)(1 - B^{12})} e_t$$

然后根据步骤S2确立数据清洗的总体流程，根据发明内容中所述步骤S3中的内容对时间序列 Z_t 进行数据清洗，其具体如下：

步骤一：首先对时间序列 Z 拟合ARMA模型，得到初始拟合残差序列：

$$e_t = (1 - 0.914B)(1 - B^{12})Z_t$$

初始拟合残差序列如图3中黄线部分所示，初步推断可能存在多个异常值。

步骤二：计算每个观测点的检验统计量 T_{AO} 和 T_{IO} ，逐次迭代直到所有的噪声点都被检验出来，结果如表 1 所示：

表 1 检验出的异常值类型

迭代次数	观测时刻	类型
1	140	AO
2	49	IO
3	26	AO

因此，考虑如下的修正模型：

$$Z_t = \theta_0 + \omega_1 I_t^{(26)} + \omega_2 \frac{1}{(1-\varphi B)\nabla^{12}} I_t^{(49)} + \omega_3 I_t^{(70)} + \frac{1}{(1-\varphi B)\nabla^{12}} a_t$$

根据表中的拟合影响对时间序列的噪声点和缺失值数值进行修正，同时根据前式对修正后的时间序列重新估计其参数，得到第一次修正后的时间序列及残差图如图 3 所示：

步骤三：根据图 3 中的残差可以判断原数据的噪声点和缺失值全部被检验出来。但由于噪声点对于观测时刻的数值拟合残差过大，不符合 AIC 检验，因此需要返回外循环进行迭代计算，进一步修正时间序列，以提高数据清洗质量。

表 2 逐步拟合结果

迭代次数	θ_0	φ	ω_1	ω_2	ω_3
1	1.093	0.868	0.821	-0.513	0.990
2	1.105	0.852	0.133	-0.320	0

步骤四：在通过 2 次外循环的迭代之后，得到最终清洗后的时间序列，如图 4 虚线所示，与原始数据基本符合。

从图 4 可以看出，虽然清洗后的时间序列与原始时间序列在异常值发生时刻附近存在偏差，但是偏差都在 5% 以下，不影响后续的状态评估，属于可接受的范畴。

因此可以得出结论，本发明利用模型拟合残差的迭代检验法来检验出输变电设备状态数据中的噪声点和缺失值，并能在迭代过程中对异常数据进行修正。相比于传统的删除噪声点，本发明清洗出的数据是不带有噪声点和缺失值的数据，从而避免了时间序列中 useful 信息的丢失。

上述虽然结合附图对本发明的具体实施方式进行了描述，但并非对本发明保护范围的限制，所属领域技术人员应该明白，在本发明的技术方案的基础上，本领域技术人员不需要付出创造性劳动即可做出的各种修改或变形仍在本发明的保护范围以内。

权利要求书

1. 基于时间序列分析的输变电设备的状态监测数据清洗方法，其特征是，步骤如下：

步骤（1）：建立输变电设备状态数据的初始时间序列模型，通过初始时间序列模型估计初始拟合的残差序列和残差方差；

步骤（2）：

步骤（2-1）：利用步骤（1）的初始时间序列模型，计算每个观测点的检验统计量；

步骤（2-2）：判断检验统计量的绝对值的最大值是否大于设定阈值，如果大于，则确定存在对初始时间序列模型拟合影响的噪声点，通过修正时间序列数据来修正噪声点，通过返回步骤（2）继续识别并修正时间序列所有的噪声点，然后进入步骤（2-3）；如果小于就进入步骤（2-3）；

步骤（2-3）：拟合修正后的时间序列模型，估计修正后的时间序列模型的模型残差；

步骤（3）：通过步骤（2-3）中修正后的时间序列模型的模型残差计算每个观测点的检验统计量，根据检验统计量的绝对值的最大值是否大于设定阈值来判断是否存在新的噪声点，直到所有的噪声点都被识别出，如果存在就返回步骤（2），如果不存在就结束。

2. 如权利要求1所述的基于时间序列分析的输变电设备的状态监测数据清洗方法，其特征是，所述步骤（1）的步骤为：

对观测序列 z_t 建立时间序列模型，

设 z_t 是单个状态量的无异常值的时间序列， z_t 服从 $ARIMA(p, d, q)$ 模型，表示为

$$Z_t = \frac{\hat{\theta}(B)}{\hat{\phi}(B)\nabla^d} e_t \quad (5)$$

$$\hat{\theta}(B) = 1 - \hat{\theta}_1 B - \hat{\theta}_2 B^2 - \dots - \hat{\theta}_q B^q \quad (6)$$

$$\hat{\phi}(B) = 1 - \hat{\phi}_1 B - \hat{\phi}_2 B^2 - \dots - \hat{\phi}_p B^p \quad (7)$$

其中， z_t 为观测序列， p, q, d 为时间序列 $ARIMA(p, d, q)$ 模型的阶数， $\theta(B)$ 为初始拟合的平稳算子， $\phi(B)$ 为初始拟合的可逆算子， e_t 是不含噪声点的序列初始拟合的残差序列， $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ 是 $\theta(B)$ 中相应的参数， $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p$ 是 $\phi(B)$ 中相应的参数， B 表示延迟算子， $B^d Z_t = Z_{t-d}$ ， B^d 表示延迟 d 个时刻的延迟算子， Z_t 表示时刻 t 的观测值， Z_{t-d} 表示时刻 $t-d$ 的观测值， ∇^d 是一个表示 t 时刻与 $t-d$ 时刻的观测值间的差值的算子， $\nabla^d = 1 - B^d$ ；

并由所建立的时间序列模型计算初始拟合的残差序列，即：

$$\hat{e}_t = \pi(B)Z_t = \frac{\varphi(B)\nabla^d}{\theta(B)}Z_t \quad (8)$$

式(8)中, Z_t 为观测序列, $\hat{e}_t$ 为含噪声点的序列初始拟合的残差序列, $\theta(B)$ 为初始拟合的平稳算子, $\varphi(B)$ 为初始拟合的可逆算子, B 表示延迟算子, $B^d Z_t = Z_{t-d}$, B^d 表示延迟 d 个时刻的延迟算子, Z_t 表示时刻 t 的观测值, Z_{t-d} 表示时刻 $t-d$ 的观测值, ∇^d 是一个表示 t 时刻与 $t-d$ 时刻的观测值间的差值的算子, $\nabla^d = 1 - B^d$; $\pi(B)$ 为表征残差影响的算子, $\hat{\pi}_1, \hat{\pi}_2, \dots, \hat{\pi}_n$ 为 $\pi(B)$ 相应的参数。

3. 如权利要求2所述的基于时间序列分析的输变电设备的状态监测数据清洗方法, 其特征是, 所述 $\pi(B)$ 的定义如(9)所示:

$$\hat{\pi}(B) = \frac{\hat{\theta}(B)}{\hat{\varphi}(B)\nabla^d} = (1 - \hat{\pi}_1 B - \hat{\pi}_2 B^2 - \dots - \hat{\pi}_n B^n) \quad (9)$$

其中, $\theta(B)$ 为初始拟合的平稳算子, $\varphi(B)$ 为初始拟合的可逆算子, n 为残差序列长度, $n \geq 1$; B 表示延迟算子, $B^d Z_t = Z_{t-d}$, $\hat{\pi}_n B^n$ 表示表征滞后 n 个时刻的残差影响的算子。

4. 如权利要求1所述的基于时间序列分析的输变电设备的状态监测数据清洗方法, 其特征是, 所述步骤(1)的残差方差的初始估计为:

$$\sigma_a^2 = \frac{1}{n} \sum_{j=1}^n \hat{e}_j^2 \quad (10);$$

其中, σ_a^2 表示残差方差, n 表示残差序列的长度, $n \geq 1$, $\hat{e}_j^2$ 的参数含义是残差值的平方, j 的取值范围是 $1 \leq j \leq n$ 。

5. 如权利要求1所述的基于时间序列分析的输变电设备的状态监测数据清洗方法, 其特征是, 所述步骤(2-1)的步骤为:

利用步骤(1)的初始时间序列模型, 计算每个观测点的检验统计量步骤为: 利用步骤(1)的初始时间序列模型, 对 $t=1, 2, \dots, n$, 计算每个观测点的检验统计量 T_{Ao}^t 和检验统计量 T_{Io}^t ; 所述检验统计量 T_{Ao}^t 和检验统计量 T_{Io}^t 的计算公式如下:

$$T_{Io}^t = \frac{\omega_{Io}}{\sigma_a} \quad (11)$$

$$T_{AO}^t = \frac{\omega_{AO}}{\sigma_a} \sqrt{\sum_{j=T}^n \pi_j^2} \quad (12)$$

式 (11)、(12) 中: $t=1,2,\dots,n$ 表示噪声点产生的时刻; σ_a 为含噪声点的残差 $\hat{e}_t$ 的标准差; 所述检验统计量 T_{AO}^t 和检验统计量 T_{IO}^t 的极限分布均为标准正态分布, 服从 t 分布。

6. 如权利要求 1 所述的基于时间序列分析的输变电设备的状态监测数据清洗方法, 其特征是, 所述步骤 (2-2) 的步骤为:

将噪声点的种类分为加性异常值 AO 和新息异常值 IO, 缺失值被认为是加性异常值;

定义 $\lambda_T = \max_t \{ |T_{AO}^t|, |T_{IO}^t| \}$, λ_T 表示检验统计量 T_{AO}^t 和检验统计量 T_{IO}^t 的最大值, T 为最大值发生的时刻;

当 $\lambda_T = |T_{AO}^{(T)}| > C$ 时, 其中 C 是设定阈值, 确定在时刻 T 存在加性异常值 AO, 加性异常值 AO 对模型拟合的影响 ω_{AO} 通过式 (13) 求得;

$$\hat{\omega}_{AO} = \frac{e_T - \sum_{j=1}^{n-T} \pi_j e_{T+j}}{\sum_{j=0}^{n-T} \pi_j^2} \quad (13)$$

其中, $\hat{\omega}_{AO}$ 表示加性异常值的影响, e_T 表示时刻 T 处的残差, e_{T+j} 表示时刻 T+j 处的残差, π_j 表示时刻 j 处残差影响的算子, j 的取值范围是 $1 \leq j \leq n$;

$$\hat{\omega}_{IO} = e_T \quad (14)$$

其中, $\hat{\omega}_{IO}$ 表示新息异常值的影响, e_T 表示时刻 T 处的残差;

对于加性异常值, 通过式 (15) 修正初始时间序列数据, 得到新的时间序列 Z_t :

$$Z_t = Z_i - \omega_{AO} I_t^{(T)} \quad (15)$$

其中, Z_t 表示经过一次修正的时间序列, Z_i 表示初始时间序列, ω_{AO} 表示加性异常值的影响, $I_t^{(T)}$ 表示阶跃函数;

并由式 (16) 修正后的残差 $\tilde{e}_t$:

$$\tilde{e}_t = \hat{e}_t - \omega_{AO} \pi(B) I_t^{(T)} \quad (16)$$

其中, $\tilde{e}_t$ 表示经过一次修正的残差, $\hat{e}_t$ 表示初始残差, ω_{AO} 表示加性异常值的影响, $\pi(B)$ 表示残差影响的算子, $I_t^{(T)}$ 表示阶跃函数;

式 (15) 和式 (16) 中 $I_t^{(T)}$ 的定义如式 (17) 所示:

$$I_t^{(T)} = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases} \quad (17)$$

其中, $I_t^{(T)}$ 表示阶跃函数, t 表示时间, T 表示阶跃函数值由 0 转为 1 的时刻;

当 $\lambda_T = |T_{IO}^{(T)}| > C$ 时, 确定在时刻 T 存在新息异常值 IO , 其对初始时间序列模型拟合的影响 ω_{IO} 通过式 (14) 求得, 利用式 (18) 修正数据, 则 IO 的影响能够消除, 即

$$Z_t = Z_t - \frac{\theta(B)}{\varphi(B)\nabla^d} \omega_{IO} I_t^{(T)} \quad (18)$$

其中, Z_t 表示经过一次修正的时间序列, Z_t 表示初始时间序列, ω_{IO} 表示新息异常值的影响, $I_t^{(T)}$ 表示阶跃函数, $\theta(B)$ 为初始拟合的平稳算子, $\varphi(B)$ 为初始拟合的可逆算子, n 为残差序列长度, $n \geq 1$; B 表示延迟算子, $B^d Z_t = Z_{t-d}$, B^d 表示延迟 d 个时刻的延迟算子, Z_t 表示时刻 t 的观测值, Z_{t-d} 表示时刻 $t-d$ 的观测值, ∇^d 是一个表示 t 时刻与 $t-d$ 时刻的观测值间的差值的算子, $\nabla^d = 1 - B^d$;

由式 (19) 修正后的残差:

$$\tilde{e}_t = \hat{e}_t - \omega_{IO} I_t^{(T)} \quad (19)$$

其中, $\tilde{e}_t$ 表示经过一次修正的残差, $\hat{e}_t$ 表示初始残差, ω_{IO} 表示新息异常值的影响, $I_t^{(T)}$ 表示阶跃函数;

使用迭代的方法识别并修正时间序列所有的噪声点; 在修正后的残差 $\tilde{e}_t$ 和残差标准差 σ_a^2 的基础上再次计算每个观测点的检验统计量 T_{AO}^t 和检验统计量 T_{IO}^t , 并重复步骤 (2-1), 直到所有的噪声点都被识别出来。

7. 如权利要求 2 所述的基于时间序列分析的输变电设备的状态监测数据清洗方法, 其特征是, 所述步骤 (2-3) 的步骤为:

假设有 k 个异常数据在时刻 $T_1, T_2, \dots, T_k$ 被识别出, 其影响分别为 $\omega_1^{(1)}, \omega_2^{(1)}, \dots, \omega_k^{(1)}$ 同时异常数据被修正而得到了新的时间序列 $Z_t^{(1)}$, $Z_t^{(1)}$ 代表是第一步迭代得到的序列;

利用新的时间序列拟合得到经过一次修正后的时间序列模型;

此时根据式 (5) 和式 (8) 重新估计该时间序列参数 $\theta^{(1)}(B), \varphi^{(1)}(B), \pi^{(1)}(B)$ ，并得到经过一次修正后的时间序列模型残差 $\tilde{e}_t^{(1)}$ ：

$$\tilde{e}_t^{(1)} = \pi^{(1)}(B) \left[Z_t^{(1)} - \sum_{j=1}^k \omega_j^{(1)} \tilde{v}_j^{(1)}(B) I_t^{(T_j)} \right] \quad (20)$$

其中

$$\tilde{v}_j^{(1)}(B) = \begin{cases} \frac{\theta^{(1)}(B)}{\varphi^{(1)}(B) \nabla^d} & IO \\ 1 & AO \end{cases} \quad (21)$$

其中， $\tilde{e}_t^{(1)}$ 表示经过一次修正后的时间序列模型残差， $\pi^{(1)}(B)$ 表示经过一次修正的残差的影响算子， $Z_t^{(1)}$ 表示经过一次修正的时间序列， $\omega_j^{(1)}$ 表示噪声点 AO 或 IO 的影响因子， $\tilde{v}_j^{(1)}(B)$ 表示噪声点 AO 或 IO 的修正因子， $I_t^{(T_j)}$ 表示阶跃函数。

8. 如权利要求 1 所述的基于时间序列分析的输变电设备的状态监测数据清洗方法，其特征是，所述步骤 (3) 的步骤为：

找出是否有新的噪声点出现，直到所有的噪声点都被识别出；在最后一次迭代后，针对修正了噪声点的时间序列 Z_t 进行联合估计，得到最终的时间序列模型：

$$Z_t = \sum_{j=1}^k \omega_j \tilde{v}_j(B) I_t^{(T_j)} + \frac{\theta(B)}{\varphi(B) \nabla^d} a_t \quad (22)$$

式 (22) 中，各参数 $\theta(B), \varphi(B), \omega_j, \tilde{v}_j$ 是在最后一次迭代中得到的；

其中， Z_t 表示经过多次修正后最终的时间序列， ω_j 表示噪声点 AO 或 IO 的影响因子， $\tilde{v}_j(B)$ 表示噪声点 AO 或 IO 的修正因子， $I_t^{(T_j)}$ 表示阶跃函数， $\varphi(B)$ 表示最终拟合的可逆算子， $\theta(B)$ 表示最终拟合的平稳算子， a_t 表示最终的残差， B 表示延迟算子， $B^d Z_t = Z_{t-d}$ ， B^d 表示延迟 d 个时刻的延迟算子， Z_t 表示时刻 t 的观测值， Z_{t-d} 表示时刻 $t-d$ 的观测值， ∇^d 是一个表示 t 时刻与 $t-d$ 时刻的观测值间的差值的算子， $\nabla^d = 1 - B^d$ ；

至此，针对输变电设备状态数据的数据清洗完成，清洗后的序列为 Z_t ，清洗后的序列即为修正了噪声点的时间序列。

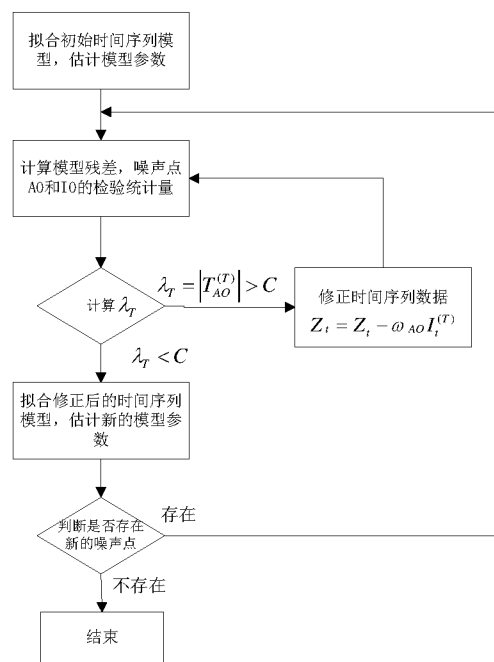


图 1

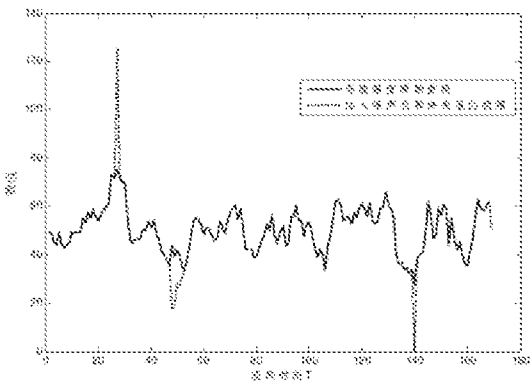


图 2

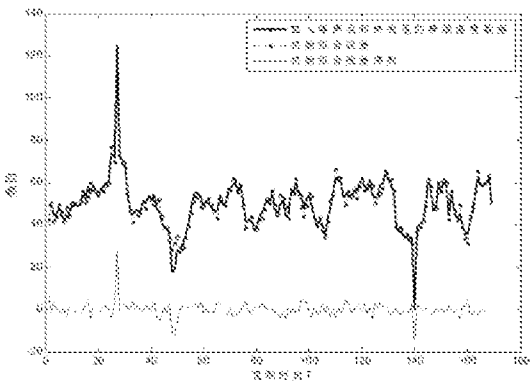


图 3

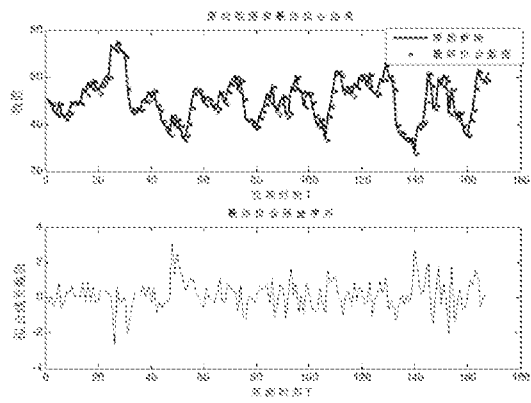


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/092589**A. CLASSIFICATION OF SUBJECT MATTER**

G06F 17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; G06Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI; EPODOC; CNPAT; CNKI; IEEE: abnormal, power transmission and transformation, power grid, sequence, electric+, transmission, transformation, time series, noise, amend, data, cleaning, residual, variance

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 104537034 A (STATE GRID CORPORATION OF CHINA et al.), 22 April 2015 (22.04.2015), claims 1-8	1-8
A	CN 102651093 A (SHANGHAI OCEAN UNIVERSITY), 29 August 2012 (29.08.2012), description, pages 2-4	1-8
A	CN 104156615 A (HARBIN INSTITUTE OF TECHNOLOGY), 19 November 2014 (19.11.2014), the whole document	1-8
A	CN 104036362 A (STATE GRID CORPORATION OF CHINA et al.), 10 September 2014 (10.09.2014), the whole document	1-8
A	US 6892163 B1 (INTELLECTUAL ASSETS LLC), 10 May 2005 (10.05.2005), the whole document	1-8

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 21 December 2015 (21.12.2015)	Date of mailing of the international search report 27 January 2016 (27.01.2016)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer XIE, Xin Telephone No.: (86-10) 62413670

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2015/092589

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104537034 A	22 April 2015	None	
CN 102651093 A	29 August 2012	None	
CN 104156615 A	19 November 2014	None	
CN 104036362 A	10 September 2014	None	
US 6892163 B1	10 May 2005	US 7082379 B1	25 July 2006
		US 7415382 B1	19 August 2008
		US 7158917 B1	02 January 2007

国际检索报告

国际申请号

PCT/CN2015/092589

A. 主题的分类

G06F 17/30(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F; G06Q

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

WPI;EPDOC;CNPAT;CNKI;IEEE: 时间序列, 数据, 清洗, 噪声, 异常, 修正, 输变电, 电网, 残差, 方差, 序列, electric+, transmission, transformation, time series, noise, amend, data, cleaning, residual, variance

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 104537034 A (国家电网公司 等) 2015年 4月 22日 (2015 - 04 - 22) 权利要求1-8	1-8
A	CN 102651093 A (上海海洋大学) 2012年 8月 29日 (2012 - 08 - 29) 说明书第2-4页	1-8
A	CN 104156615 A (哈尔滨工业大学) 2014年 11月 19日 (2014 - 11 - 19) 全文	1-8
A	CN 104036362 A (国家电网公司 等) 2014年 9月 10日 (2014 - 09 - 10) 全文	1-8
A	US 6892163 B1 (INTELLECTUAL ASSETS LLC) 2005年 5月 10日 (2005 - 05 - 10) 全文	1-8

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2015年 12月 21日

国际检索报告邮寄日期

2016年 1月 27日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

解欣

电话号码 (86-10)62413670

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/092589

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	104537034	A	2015年 4月 22日	无			
CN	102651093	A	2012年 8月 29日	无			
CN	104156615	A	2014年 11月 19日	无			
CN	104036362	A	2014年 9月 10日	无			
US	6892163	B1	2005年 5月 10日	US	7082379	B1	2006年 7月 25日
				US	7415382	B1	2008年 8月 19日
				US	7158917	B1	2007年 1月 2日

表 PCT/ISA/210 (同族专利附件) (2009年7月)