(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 106368813 B
(45)授权公告日 2018.09.25

(21)申请号 201610772979.3

(22)申请日 2016.08.30

(65)同一申请的已公布的文献号
申请公布号 CN 106368813 A

(43)申请公布日 2017.02.01

(73)专利权人 北京协同创新智能电网技术有限公司
地址 100000 北京市海淀区丰秀中路3号院13号楼3层304室

(72)发明人 王建东 朱迪 黄越 杨子江

(74)专利代理机构 济南圣达知识产权代理有限公司 37221
代理人 张勇

(51)Int.Cl.
F02B 77/08(2006.01)

(56)对比文件
JP 5186322 B2,2013.04.17,
CN 101713395 A,2010.05.26,
CN 105761407 A,2016.07.13,
CN 105241669 A,2016.01.13,
CN 104019000 A,2014.09.03,
US 2015235139 A1,2015.08.20,

审查员 张玉春

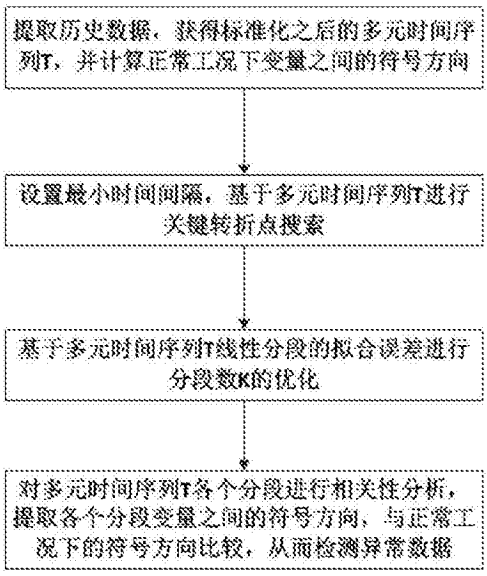
权利要求书2页 说明书6页 附图4页

(54)发明名称

一种基于多元时间序列的异常报警数据检测方法

(57)摘要

本发明公开了一种基于多元时间序列的异常报警数据检测方法，从历史数据中提取多个相关变量的数据，建立多元时间序列，将其标准化，计算正常状态下每个变量之间的符号方向；确定基于关键转折点的时间序列分段描述，设置最小时间间隔，进行关键转折点搜索；表示多元时间序列的分段线性，根据数据点到各个分段的正交距离确定拟合误差，设置损失函数阈值，优化分段数量，得到优化后的分段结果；基于优化之后分段结果，对多元时间序列的各个分段进行相关性分析，提取各个分段变量之间的符号方向，检测与正常状态下的符号方向不一致的异常数据。本发明为实现多变量报警系统的动态报警阈值设计提供有利的条件，从而减少干扰报警。



CN 106368813 B

1.一种基于多元时间序列的异常报警数据检测方法,包括以下步骤:

(1)从历史数据中提取当前工作点之前的一定时间内的多个相关变量的数据,建立多元时间序列,将其标准化,计算正常状态下每个变量之间的符号方向;

其特征是:

(2)确定基于关键转折点的时间序列分段描述,设置最小时间间隔,进行关键转折点搜索;

(3)基于线性分段的多元时间序列,根据数据点到各个分段的正交距离确定拟合误差,设置损失函数阈值,优化分段数量,得到优化后的分段结果;

(4)基于优化之后分段结果,对多元时间序列的各个分段进行相关性分析,提取各个分段变量之间的符号方向,检测与正常状态下的符号方向不一致的异常数据。

2.如权利要求1所述的一种基于多元时间序列的异常报警数据检测方法,其特征是:所述步骤(1)中,具体包括:

(1-1)提取当前工作点之前时间长度为n的多个相关变量的原始数据,建立多元时间序列;

(1-2)求取原始数据的样本均值和标准差,将多元时间序列标准化;

(1-3)根据每两个变量的相关系数确定符号方向,构建符号方向矩阵。

3.如权利要求2所述的一种基于多元时间序列的异常报警数据检测方法,其特征是:所述步骤(1-3)中,利用$\rho_{\Delta T}[X_i,X_j]$代表两时刻之间的子时间序列内变量X_i和X_j的相关系数,任意两个变量X_i和X_j在同一段子时间序列内的符号方向$sign_{\Delta T}(X_i,X_j)$取值1,-1,0分别表示变量之间关系为正相关、负相关、无显著相关。

4.如权利要求1所述的一种基于多元时间序列的异常报警数据检测方法,其特征是:所述步骤(2)中,具体包括:

(2-1)给出基于关键转折点的时间序列分段的数学描述,即多元时间序列被K+1个关键点分为K个不重叠的时间片段;

(2-2)给出由m个变量和时间t构成的m+1维线性空间中点到直线的正交距离的数学描述;

(2-3)设置最小时间间隔,以其作为搜索关键转折点过程的停止条件。

5.如权利要求1所述的一种基于多元时间序列的异常报警数据检测方法,其特征是:所述步骤(3)中,利用在最小时间间隔δ的约束下得到的关键转折点集合对原始时间序列进行分段线性表示,选择恰当的分段数量以避免过拟合,实现在一定拟合误差的约束下使用较少的关键转折点作为最终的分段点。

6.如权利要求1所述的一种基于多元时间序列的异常报警数据检测方法,其特征是:所述步骤(3)中,具体包括:

(3-1)利用线性插值的方法,将多维空间中的数据点在所属分段首尾数据点连线上的投影作为拟合点,并将多元时间序列进行分段线性表示;

(3-2)利用正交距离表示拟合误差;

(3-3)将拟合误差作为损失函数,将不同分段数K对应的损失函数值E(K)绘制到平面直角坐标系中,选取损失函数值小于阈值对应的点所对应的分段数为优化结果;

所述步骤(3-3)中,观察损失函数值随分段数K的收敛情况,在损失函数值减小随K值增

加趋于平稳的区域设置合理的损失函数阈值。

7.如权利要求6所述的一种基于多元时间序列的异常报警数据检测方法，其特征是：所述步骤(3-3)中，设第一个满足损失函数值小于损失函数阈值对应的点坐标为(c,E(c))，则选择该点对应的分段数K＝c作为优化结果和对应的关键转折点集合$Q_c=[q_1,\cdots,q_{c+1}]$作为用于进行相关性趋势提取的分段点。

8.如权利要求1所述的一种基于多元时间序列的异常报警数据检测方法，其特征是：所述步骤(4)中，具体包括：

(4-1)计算每个分段中任意两个变量之间的相关系数；

(4-2)对变量相关性进行单边假设检验，设置显著性水平，根据单边假设检验结果和显著性水平确认变量间的相关性，确定变量符号；

(4-3)根据变量符号，构建变量符号矩阵，将变量符号矩阵和符号方向矩阵中对应位置的元素进行比较，如两者不同，则其对应的分段为异常数据。

9.如权利要求8所述的一种基于多元时间序列的异常报警数据检测方法，其特征是：所述步骤(4-1)中，具体方法为：

假设将时间序列T的第s个分段内变量Xi和Xj的观测值分别从小到大排列，并依次用正整数k＝1，…，z_s标记，记为R_k和S_k，则时间序列T中第s个分段内任意两个变量之间的Spearman样本相关系数为：

$$\rho_s[X_i,X_j]=\frac{\sum_{k=1}^{z_s}(R_k-\bar{R})(S_k-\bar{S})}{\sqrt{\sum_{k=1}^{z_s}(R_k-\bar{R})^2(S_k-\bar{S})^2}}=1-\frac{6\sum_{k=1}^{z_s}d_k^2}{z_s(z_s^2-1)},$$

其中$\bar{R}=\frac{1}{z_s}\sum_{k=1}^{z_s}R_k$，$\bar{S}=\frac{1}{z_s}\sum_{k=1}^{z_s}S_k$，$d_k=R_k-S_k$。

10.如权利要求8所述的一种基于多元时间序列的异常报警数据检测方法，其特征是：所述步骤(4-2)中，单边假设检验：$H0:\rho_s[X_i,X_j]=0$ vs $H1:\rho_s[X_i,X_j]>0$，$H0:\rho_s[X_i,X_j]=0$ vs $H2:\rho_s[X_i,X_j]<0$；

当样本个数n>10时，随机变量Us被定义为：$U_s=\rho_s[X_i,X_j]\sqrt{\frac{z_s-2}{1-(\rho_s[X_i,X_j])^2}}$，其中，$z_s$为第s个分段内的样本个数；给定显著性水平α，如果$U_s>t_\alpha(z_s-2)$，则与H1相对的H0被拒绝，如果$U_s<-t_\alpha(z_s-2)$，则与H2相对的H0被拒绝，其中$t_\alpha(z_s-2)$表示统计量Us的分位数，此时，第s个分段内Xi和Xj的相关性被认为是显著的，符号方向$sign_s(X_i,X_j)$分别取值为1或-1，如果$|U_s|<t_\alpha(z_s-2)$，无论对于H1或者H2，H0都不能被拒绝，此时变量间无显著相关性，符号方向$sign_s(X_i,X_j)$取值为0。

所述步骤(4-2)中，当样本个数n<10时，查询用于小样本容量假设检验的Spearman秩相关系数的临界值，将对应于给定z_s和α的相关系数临界值表示为$\rho_\alpha(z_s)$，如果$|\rho_s[X_i,X_j]|>\rho_\alpha(z_s)$，H0被拒绝，$sign_s(X_i,X_j)$分别取值为1或-1，反之H0不能被拒绝，符号方向$sign_s(X_i,X_j)$取值为0。

一种基于多元时间序列的异常报警数据检测方法

技术领域

[0001] 本发明涉及一种基于多元时间序列的异常报警数据检测方法。

背景技术

[0002] 报警系统对保障燃煤发电机组的安全生产与高效运行发挥着至关重要的作用，由于实际工业过程中关联变量之间的相互影响，传统的单变量报警阈值设计方法可能产生大量干扰报警(漏报警和误报警)并导致“报警过多”的发生，使得现场操作人员的注意力受到影响，增大了在异常生产状况发生时做出正确处置的难度。为了实现多变量报警系统的动态报警阈值设计，寻找一种从历史数据中自动筛选出处于正常状况和异常状况数据段的检测方法是十分必要的。

发明内容

[0003] 本发明为了解决上述问题，提出了一种基于多元时间序列的异常报警数据检测方法，本方法通过建立多元时间序列，从模式异常的角度提出了一种结合时间序列分段线性表示方法和定性趋势分析方法的异常数据检测算法，自动对历史数据进行异常数据检测，为实现多变量报警系统的动态报警阈值设计提供有利的条件，从而减少干扰报警，提高现场操作人员处理报警的效率，保障了生产安全性。

[0004] 为了实现上述目的，本发明采用如下技术方案：

[0005] 一种基于多元时间序列的异常报警数据检测方法，包括以下步骤：

[0006] (1)从历史数据中提取当前工作点之前的一定时间内的多个相关变量的数据，建立多元时间序列，将其标准化，计算正常状态下每个变量之间的符号方向；

[0007] (2)确定基于关键转折点的时间序列分段描述，设置最小时间间隔，进行关键转折点搜索；

[0008] (3)基于线性分段的多元时间序列，根据数据点到各个分段的正交距离确定拟合误差，设置损失函数阈值，优化分段数量，得到优化后的分段结果；

[0009] (4)基于优化之后分段结果，对多元时间序列的各个分段进行相关性分析，提取各个分段变量之间的符号方向，检测与正常状态下的符号方向不一致的异常数据。

[0010] 所述步骤(1)中，具体包括：

[0011] (1-1)提取当前工作点之前时间长度为n的多个相关变量的原始数据，建立多元时间序列；

[0012] (1-2)求取原始数据的样本均值和标准差，将多元时间序列标准化；

[0013] (1-3)根据每两个变量的相关系数确定符号方向，构建符号方向矩阵。

[0014] 所述步骤(1-3)中，利用$\rho_{\Delta T}[X_i,X_j]$代表两时刻之间的子时间序列内变量X_i和X_j的相关系数，任意两个变量X_i和X_j在同一段子时间序列内的符号方向$sign_{\Delta T}(X_i,X_j)$取值1，-1，0分别表示变量之间关系为正相关、负相关、无显著相关。

[0015] 所述步骤(2)中，具体包括：

[0016] (2-1)给出基于关键转折点的时间序列分段的数学描述，即多元时间序列被K+1个关键点分为K个不重叠的时间片段；

[0017] (2-2)给出由m个变量和时间t构成的m+1维线性空间中点到直线的正交距离的数学描述；

[0018] (2-3)设置最小时间间隔，以其作为搜索关键转折点过程的停止条件。

[0019] 所述步骤(3)中，利用在最小时间间隔δ的约束下得到的关键转折点集合对原始时间序列进行分段线性表示，选择恰当的分段数量以避免过拟合，实现在一定拟合误差的约束下使用较少的关键转折点作为最终的分段点。

[0020] 所述步骤(3)中，具体包括：

[0021] (3-1)利用线性插值的方法，将多维空间中的数据点在所属分段首尾数据点连线上的投影作为拟合点，并将多元时间序列进行分段线性表示；

[0022] (3-2)利用正交距离表示拟合误差；

[0023] (3-3)将拟合误差作为损失函数，将不同分段数K对应的损失函数值E(K)绘制到平面直角坐标系中，选取损失函数值小于阈值对应的点所对应的分段数为优化结果。

[0024] 所述步骤(3-3)中，观察损失函数值随分段数K的收敛情况，在损失函数值减小随K值增加趋于平稳的区域设置合理的损失函数阈值。

[0025] 所述步骤(3-3)中，设第一个满足损失函数值小于损失函数阈值对应的点坐标为(c,E(c))，则选择该点对应的分段数K＝c作为优化结果和对应的关键转折点集合Q_c＝$[q_1,\cdots,q_{c+1}]$作为用于进行相关性趋势提取的分段点。

[0026] 所述步骤(4)中，具体包括：

[0027] (4-1)计算每个分段中任意两个变量之间的相关系数；

[0028] (4-2)对变量相关性进行单边假设检验，设置显著性水平，根据单边假设检验结果和显著性水平确认变量间的相关性，确定变量符号；

[0029] (4-3)根据变量符号，构建变量符号矩阵，将变量符号矩阵和符号方向矩阵中对应位置的元素进行比较，如两者不同，则其对应的分段为异常数据。

[0030] 所述步骤(4-1)中，具体方法为：

[0031] 假设将时间序列T的第s个分段内变量Xi和Xj的观测值分别从小到大排列，并依次用正整数k＝1,…,z_s标记，记为R_k和S_k，则时间序列T中第s个分段内任意两个变量之间的Spearman样本相关系数为：

[0032] $$\rho_s[X_i,X_j]=\frac{\sum_{k=1}^{z_s}(R_k-\bar{R})(S_k-\bar{S})}{\sqrt{\sum_{k=1}^{z_s}(R_k-\bar{R})^2(S_k-\bar{S})^2}}=1-\frac{6\sum_{k=1}^{z_s}d_k^2}{z_s(z_s^2-1)},$$

[0033] 其中$\bar{R}=\frac{1}{z_s}\sum_{k=1}^{z_s}R_k$，$\bar{S}=\frac{1}{z_s}\sum_{k=1}^{z_s}S_k$，$d_k$＝$R_k$-$S_k$。

[0034] 所述步骤(4-2)中，单边假设检验：H0：$\rho_s[X_i,X_j]$＝0vs H1：$\rho_s[X_i,X_j]$>0，H0：$\rho_s[X_i,X_j]$＝0vs H2：$\rho_s[X_i,X_j]$<0；

[0035] 当样本个数n>10时，随机变量Us被定义为：$U_s=\rho_s[X_i,X_j]\sqrt{\frac{z_s-2}{1-(\rho_s[X_i,X_j])^2}}$，其中，$z_s$为第s个分段内的样本个数；给定显著性水平α，如果$U_s>t_\alpha(z_s-2)$，则与H1相对的H0被拒绝，如果$U_s<-t_\alpha(z_s-2)$，则与H2相对的H0被拒绝，其中$t_\alpha(z_s-2)$表示统计量Us的分位数，此时，第s个

分段内Xi和Xj的相关性被认为是显著的，符号方向$sign_s(X_i,X_j)$分别取值为1或-1，如果$|U_s|<t_\alpha(z_s-2)$，无论对于H1或者H2，H0都不能被拒绝，此时变量间无显著相关性，符号方向$sign_s(X_i,X_j)$取值为0。

[0036] 所述步骤(4-2)中，当样本个数n<10时，查询用于小样本容量假设检验的Spearman秩相关系数的临界值，将对应于给定z_s和α的相关系数临界值表示为$\rho_\alpha(z_s)$，如果$|\rho_s[X_i,X_j]|>\rho_\alpha(z_s)$，H0被拒绝，$sign_s(X_i,X_j)$分别取值为1或-1，反之H0不能被拒绝，符号方向$sign_s(X_i,X_j)$取值为0。

[0037] 本发明的有益效果为：本发明选取工业变量之间的相关性作为判断工作点状态是否异常的特征，通过建立多元时间序列，从模式异常的角度提出了一种结合时间序列分段线性表示 方法和定性趋势分析方法的异常数据检测算法，自动对历史数据进行异常数据检测，为实现多变量报警系统的动态报警阈值设计提供有利的条件，从而减少干扰报警，提高现场操作人员处理报警的效率，保障了生产安全性。

附图说明

[0038] 图1为本发明所述基于工业历史数据的报警系统异常数据检测方法流程图；

[0039] 图2为本发明具体实施例中变量时间序列和分段结果图；

[0040] 图3为本发明具体实施例中分段数K的决策图；

[0041] 图4(a)为变量在每个分段中相关性分析结果；

[0042] 图4(b)为变量之间符号方向计算结果；

[0043] 图5(a)为用不同线段表示的异常数据检测结果；

[0044] 图5(b)为用不同数值表示的异常数据检测结果。

具体实施方式：

[0045] 下面结合附图与实施例对本发明作进一步说明。

[0046] 图1为本发明所述基于工业历史数据的报警系统异常数据检测方法流程图。

[0047] 如图1所示，一种基于工业历史数据的报警系统异常数据检测方法，包括如下步骤：

[0048] 步骤S1，从历史数据中提取当前工作点之前的时间t内的多个相关变量的数据，建立多元时间序列T'，并将其标准化为时间序列T，同时计算正常状态下各个变量之间的符号方向；

[0049] 步骤S2，设置最小时间间隔δ，并基于多元时间序列T进行关键转折点搜索；

[0050] 步骤S3，基于多元时间序列T线性分段的拟合误差进行分段数K的优化；

[0051] 步骤S4，基于优化之后分段结果，对多元时间序列T的各个分段进行相关性分析，提取各个分段变量之间的符号方向，根据其与正常状态下的符号方向是否一致来检测异常数据。

[0052] 在本发明的具体实施例中，步骤S1的具体实现为：

[0053] 步骤S11，提取当前工作点之前时间长度为n的多个相关变量的原始数据，用X(t)来表示变量i在t时刻的数值，建立多元时间序列$T'=\{X'_i(t)\}$，其中$i=1,\cdots,m$，$t=1,\cdots,n$，m表示变量个数，n表示时间长度。

[0054] 步骤S12，将多元时间序列T' 标准化为T，$T=\{X_i(t)\}=\left\{\frac{X_i'(t)-\overline{X_i'}}{\delta_{X_i'}}\right\}$，其中$\overline{X_i'}$代表原始数据X′i(t)的样本均值，$\delta_{X_i'}$代表原始数据X′i(t)的样本标准差。

[0055] 步骤S13，计算任意两个变量Xi和Xj在同一段子时间序列内的符号方向：$sign_{\Delta T}(X_i,X_j)=\begin{cases}1, \rho_{\Delta T}[X_i,X_j]>0\\ 0,\rho_{\Delta T}[X_i,X_j]=0\\ -1,\rho_{\Delta T}[X_i,X_j]<0\end{cases}$，其中ΔT代表时刻t1到时刻t2的子时间序列，1≤t1≤t2≤n，$\rho_{\Delta T}[X_i,X_j]$代表ΔT内变量Xi和Xj的相关系数，$sign_{\Delta T}(X_i,X_j)$取值1，-1，0分别表示变量之间关系为正相关、负相关、无显著相关。在正常情况下，变量之间的符号方向保持不变，符号方向矩阵可被定义为：$R=\begin{bmatrix}1 & sign_T(X_1,X_2) & \cdots & sign_T(X_1,X_m)\\ sign_T(X_2,X_1) & 1 & \cdots & sign_T(X_2,X_m)\\ \cdots & \cdots & 1 & \cdots\\ sign_T(X_m,X_1) & sign_T(X_m,X_2) & \cdots & 1\end{bmatrix}$。

[0056] 在本发明的具体实施例中，步骤S2的具体实现为：

[0057] 步骤S21，给出基于关键转折点的时间序列分段的数学描述。给定整数K，时间序列T可以被K+1个关键点分为K个不重叠的的时间片段，用S＝{p1，…，pK+1}表示，其中pi，i＝1，…，k+1，代表第i个关键转折点的时间标签，并且有p1＝1，pK+1＝n。将S中第j个分段表示为sj＝{{Xi(t)}，pj<t≤pj+1}，其中j＝1，…，K。定义zj＝pj+1-pj为第j个分段中包含的数据点的个数。

[0058] 步骤S22，给出由m个变量和时间t构成的m+1维线性空间中点到直线的正交距离的数学描述。空间中直线AB的参数方程可表示为：$\frac{X_1-X_{1A}}{X_{1B}-X_{1A}}=,\ldots,=\frac{X_i-X_{iA}}{X_{iB}-X_{iA}}=\frac{t-t_A}{t_B-t_A}=\beta$，其中i＝1，…，m。则直线AB上任意一点P0的坐标可表示为[(XiB-XiA)β+XiA，(tB-tA)β+tA]。因此，点P到直线AB的距离可被定义为：$\|\overrightarrow{P_0P}\|=\sqrt{[(X_{iB}-X_{iA})\beta+X_{iA}-X_{iP}]^2+[(t_B-t_A)\beta+t_A-t_P]^2}=\sqrt{\left(\|\overrightarrow{AB}\|\right)^2\beta^2+2(\overrightarrow{AB}\cdot\overrightarrow{AP})\beta+\left(\|\overrightarrow{AP}\|\right)^2}$。其中，$\overrightarrow{AB}=[(X_{iB}-X_{iA}),(t_B-t_A)]$，$\overrightarrow{AP}=[(X_{iP}-X_{iA}),(t_P-t_A)]$。当$\|\overrightarrow{P_0P}\|$取极小值时对应的参数$\beta^*=-(\overrightarrow{AB}\cdot\overrightarrow{AP})/\left(\|\overrightarrow{AB}\|\right)^2$，则点P到直线AB的最小距离即正交距离为$D=\frac{\sqrt{\left(\|\overrightarrow{AP}\|\|\overrightarrow{AB}\|\right)^2-(\overrightarrow{AB}\cdot\overrightarrow{AP})^2}}{\|\overrightarrow{AB}\|}$。

[0059] 步骤S23，设置最小时间间隔δ，0<δ<n，用于减少噪音对分段结果的影响，并作为为搜索关键转折点过程的停止条件。在处理实际工业过程数据时，噪声的干扰会导致关键点之间的时间间隔过短，因此的当满足条件min(zj)<δ，j＝1，…，K时停止搜索关键转折点。

[0060] 在本发明的具体实施例中，步骤S3的具体实现为：

[0061] 利用在最小时间间隔δ的约束下得到的关键转折点集合Q对原始时间序列进行分段线性表示，通常会导致过拟合，为了避免过拟合，有必要选择合适的分段数K，以实现在一定拟合误差的约束下使用较少的关键转折点作为最终的分段点。

[0062] 步骤S31，定义多元时间序列的分段线性表示。当多元时间序列T被K+1个关键转折点p1，…，pK+1分为K段，则此时多元时间序列T的分段线性表示为：TPLR＝<f1[(Xi(p1)，p1)，(Xi

$(p_2),p_2)]$,…,$f_K[(X_i(p_K),p_K),(X_i(p_{K+1}),p_{K+1})]>$。其中$f_1[(X_i(p_1),p_1),(X_i(p_2),p_2)]$表示在分段$[p_j,p_{j+1}]$内的线性拟合函数。本发明利用线性插值的方法,将m+1维空间中的数据点$(X_i(t),t)$,其中i=1,…,m,t=1,…,n在所属分段首尾数据点连线上的投影作为拟合点,从而得到拟合点$\left(\hat{X}_i(t),\hat{t}(t)\right)$。

[0063] 步骤S32,定义时间序列分段线性表示的拟合误差。对时间序列T进行分段线性表示,采用线性插值得到原始数据的拟合点,则拟合误差为:$E=\sqrt{\frac{\sum_{t=1}^{n}\left(\left\|(X_i(t),t)-(\hat{X}_i(t),\hat{t}(t))\right\|\right)^2}{n}}$,其中i=1,…,m。根据线性空间中点到直线正交距离的定义,拟合误差还可以用正交距离表示为:$E=\sqrt{\frac{\sum_{j=1}^{n}D(j)^2}{n}}$,其中D(j)表示数据点$(X_i(j),j)$到所属分段的正交距离。

[0064] 步骤S33,设置损失函数阈值η,对分段数K进行优化并得到优化的分段结果。将拟合误差E作为损失函数,假定关键转折点集合Q=$[q_1,q_2,\cdots,q_1,q_{1+1}]$,计算当2≤K≤l时,不同分段数K对应的损失函数值E(K),并将所有点(K,E(K))绘制到平面直角坐标系中。

[0065] 观察损失函数E随分段数K的收敛情况,在E值减小随K值增加趋于平稳的区域设置合理的损失函数阈值η。假设第一个满足E(K)<η对应的点坐标为(c,E(c)),则选择该点对应的分段数K=c作为优化结果和对应Q中的关键转折点集合$Q_c=[q_1,\cdots,q_{c+1}]$作为用于进行相关性趋势提取的分段点。

[0066] 在本发明的具体实施例中,步骤S4的具体实现为:

[0067] 步骤S41,获得任意两个变量之间的相关系数。假设将时间序列T的第s个分段内变量Xi和Xj的观测值分别从小到大排列,并依次用正整数k=1,…,z_s标记,记为R_k和S_k。则时间序列T中第s个分段内任意两个变量之间的Spearman样本相关系数为:$\rho_s[X_i,X_j]=\frac{\sum_{k=1}^{z_s}(R_k-\bar{R})(S_k-\bar{S})}{\sqrt{\sum_{k=1}^{z_s}(R_k-\bar{R})^2(S_k-\bar{S})^2}}=1-\frac{6\sum_{k=1}^{z_s}d_k^2}{z_s(z_s^2-1)}$,其中$\bar{R}=\frac{1}{z_s}\sum_{k=1}^{z_s}R_k$,$\bar{S}=\frac{1}{z_s}\sum_{k=1}^{z_s}S_k$,$d_k=R_k-S_k$。

[0068] 步骤S42,对变量相关性进行单边假设检验。单边假设检验:H0:$\rho_s[X_i,X_j]=0$vs H1:$\rho_s[X_i,X_j]>0$,H0:$\rho_s[X_i,X_j]=0$vs H2:$\rho_s[X_i,X_j]<0$。当样本个数n>10时,随机变量Us可被定义为:$U_s=\rho_s[X_i,X_j]\sqrt{\frac{z_s-2}{1-(\rho_s[X_i,X_j])^2}}$,其中,$z_s$为第s个分段内的样本个数。给定显著性水平α,如果$U_s>t_\alpha(z_s-2)$,则与H1相对的H0被拒绝,如果$U_s<-t_\alpha(z_s-2)$,则与H2相对的H0被拒绝,其中$t_\alpha(z_s-2)$表示统计量Us的分位数。此时,第s个分段内Xi和Xj的相关性被认为是显著的,符号方向$sign_s(X_i,X_j)$分别取值为1或-1。如果$|U_s|<t_\alpha(z_s-2)$,无论对于H1或者H2,H0都不能被拒绝,此时变量间无显著相关性,符号方向$sign_s(X_i,X_j)$取值为0。

[0069] 当样本个数n<10时,查询用于小样本容量假设检验的Spearman秩相关系数的临界值,将对应于给定z_s和α的相关系数临界值表示为$\rho_\alpha(z_s)$,如果$|\rho_s[X_i,X_j]|>\rho_\alpha(z_s)$,H0被拒绝,$sign_s(X_i,X_j)$分别取值为1或-1,反之H0不能被拒绝,符号方向$sign_s(X_i,X_j)$取值为0。

[0070] 步骤S43,根据异常数据检测规则判定异常数据。如果对于所有的i,j∈[1,2,…,m],均满足$sign_s(X_i,X_j)=sign_T(X_i,X_j)$,则第s分段被划分为正常数据。如果存在任一i,j∈[1,2,…,m],使得$sign_s(X_i,X_j)\neq sign_T(X_i,X_j)$,则第s分段被划分为异常数据。

[0071] 以下是本发明所述方法在具体示例中的应用,具体应用场景为电厂。

[0072] 选定电厂中的给水泵的进口流量和给水泵汽轮机转速作为相关变量，在电厂中的一次停机事故中，选取停机前采样周期为1秒，样本容量为n＝2239的流量和转速二元时间序列作为历史数据，并将其标准化，标准化之后的二元时间序列记为T＝[Q(t),V(t)]，其中t＝1,…,n。

[0073] 根据其工作原理，正常工况下给水泵进口流量(简称流量，用Q表示)与给水泵汽轮机转速(简称转速，用V表示)保持高度正相关性，因此符号方向$sign_T(Q,V)=1$，符号方向矩阵$R=\begin{vmatrix}1 & 1\\ 1 & 1\end{vmatrix}$。

[0074] 给定最小时间间隔δ＝15进行关键转折点搜索，当搜索停止时，K＝116，损失函数阈值与K的分布如图2所示，根据观察，选择η＝0.3作为损失函数阈值，得到优化后的分段数K＝28。

[0075] 根据优化后获得的分段数，绘制变量的分段时间序列图，如图3所示。其中实线代表的是各个关键转折点的位置，虚线代表相关性发生显著变化的时间段。

[0076] 给定α＝0.05，计算每一个分段中Q和V的样本相关系数并进行相关性检验，流量Q与转速V在每个分段中相关性分析结果如图4(a)所示，流量Q与转速V的符号方向计算结果如图4(b)所示，从图中可以看出第24,25和28分段中变量相关性发生了显著变化，即此时$sign_S(Q,V)\neq sign_T(Q,V)$。

[0077] 根据相关性趋势分析结果可知，t＝1911-2126，以及t＝2195-2239之间的数据被检测为异常数据，图5(a)用不同的线段标记了异常数据的检测结果，这些数据所在分段与相关性发生显著变化的第24,25和28分段是一致的，图5(b)用不同数值标记异常数据检测结果。分析可知，第一部分数据出现异常是由于位于给水泵下游的汽包压力异常升高，导致给水泵进出口压差减小，阻力增大，出现转速升高但流量却下降的现象；第二部分数据出现异常是由于前期异常没有得到及时正确处置触发了机组应急停车，这部分数据对于报警系统设计以及故障原因分析的作用十分有限。

[0078] 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。

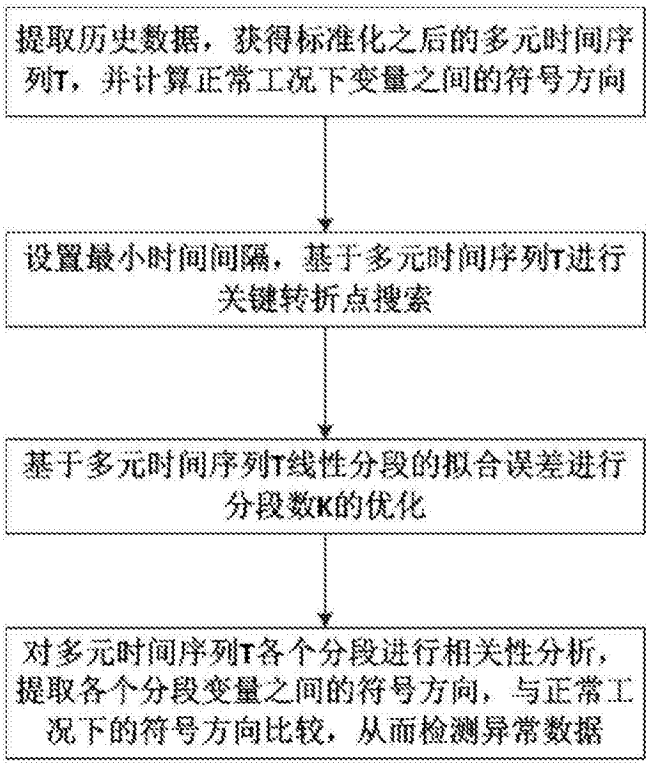

提取历史数据，获得标准化之后的多元时间序列T，并计算正常工况下变量之间的符号方向
设置最小时间间隔，基于多元时间序列T进行关键转折点搜索
基于多元时间序列T线性分段的拟合误差进行分段数K的优化
对多元时间序列T各个分段进行相关性分析，提取各个分段变量之间的符号方向，与正常工况下的符号方向比较，从而检测异常数据

图1

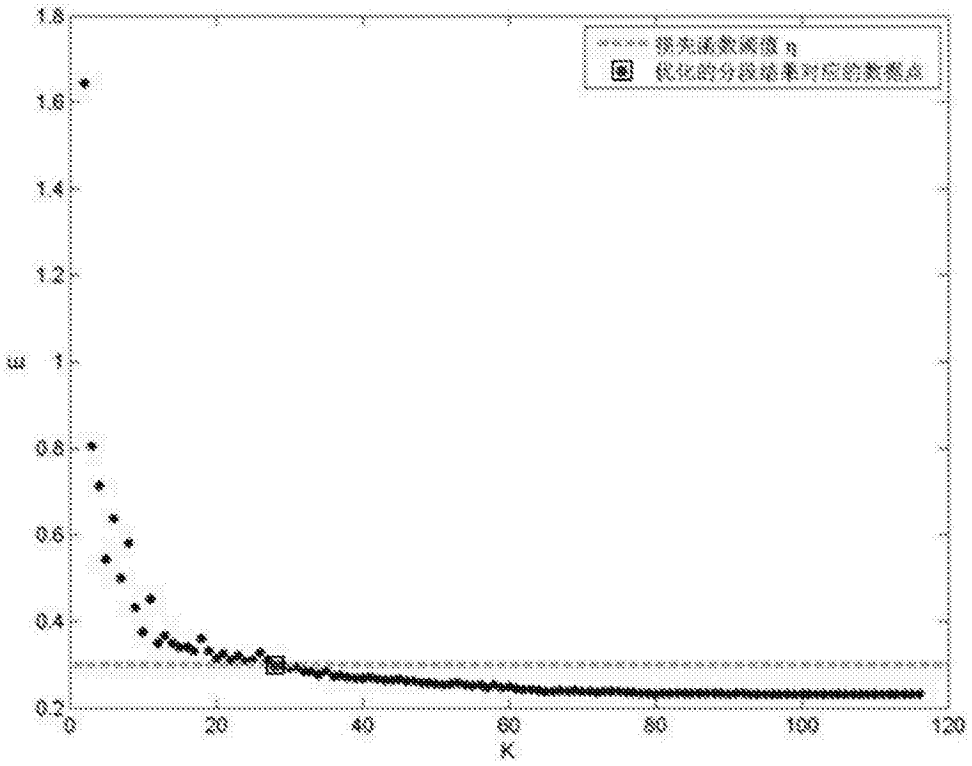

1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
20
40
60
80
100
120
K

图2

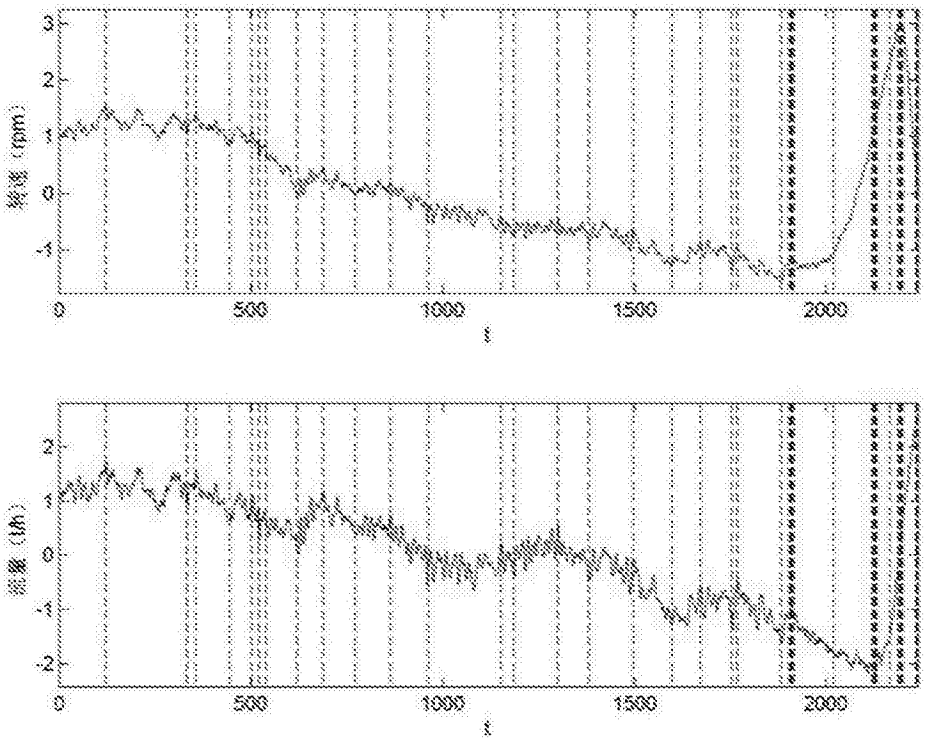

3
2
1
0
-1
0
500
1000
1500
2000
t
2
1
0
-1
-2
0
500
1000
1500
2000
t

图3

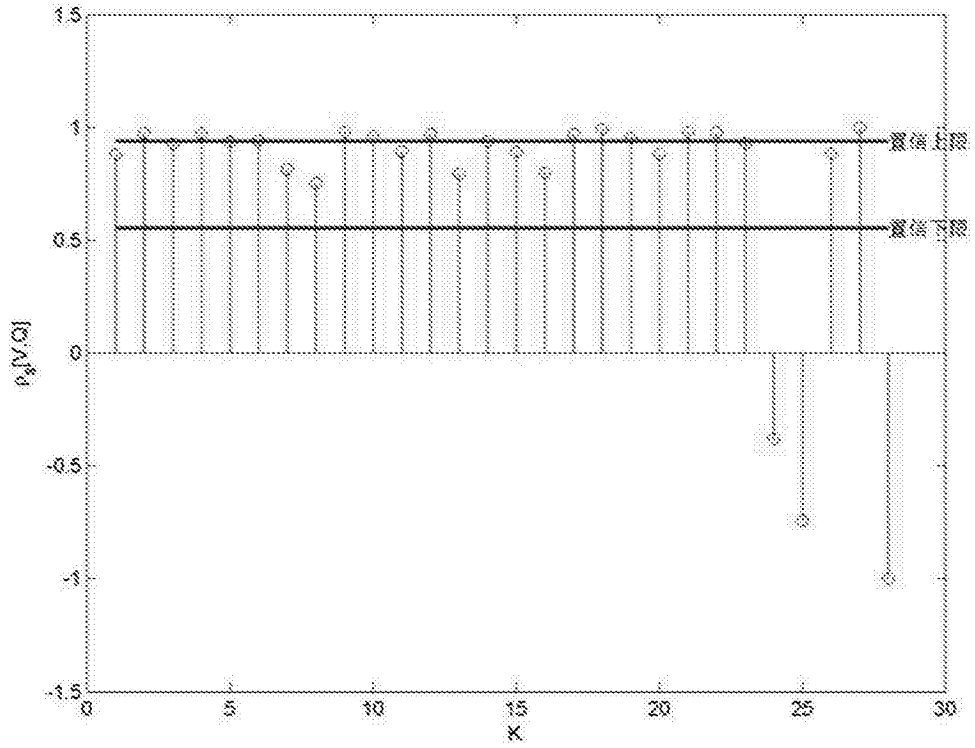

1.5
1
0.5
0
-0.5
-1
-1.5
0
5
10
15
20
25
30
K

图4(a)

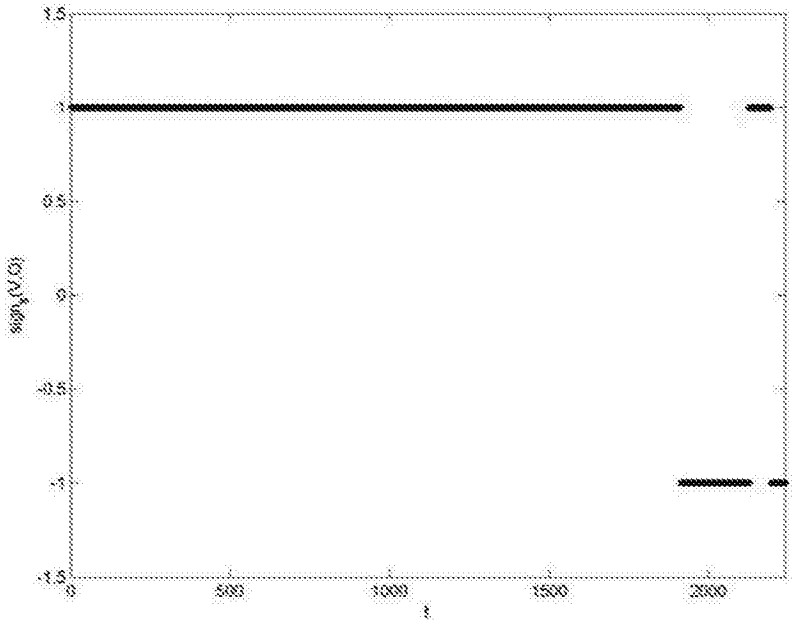

1.5
1
0.5
0
-0.5
-1
-1.5
0
500
1000
1500
2000
t

图4 (b)

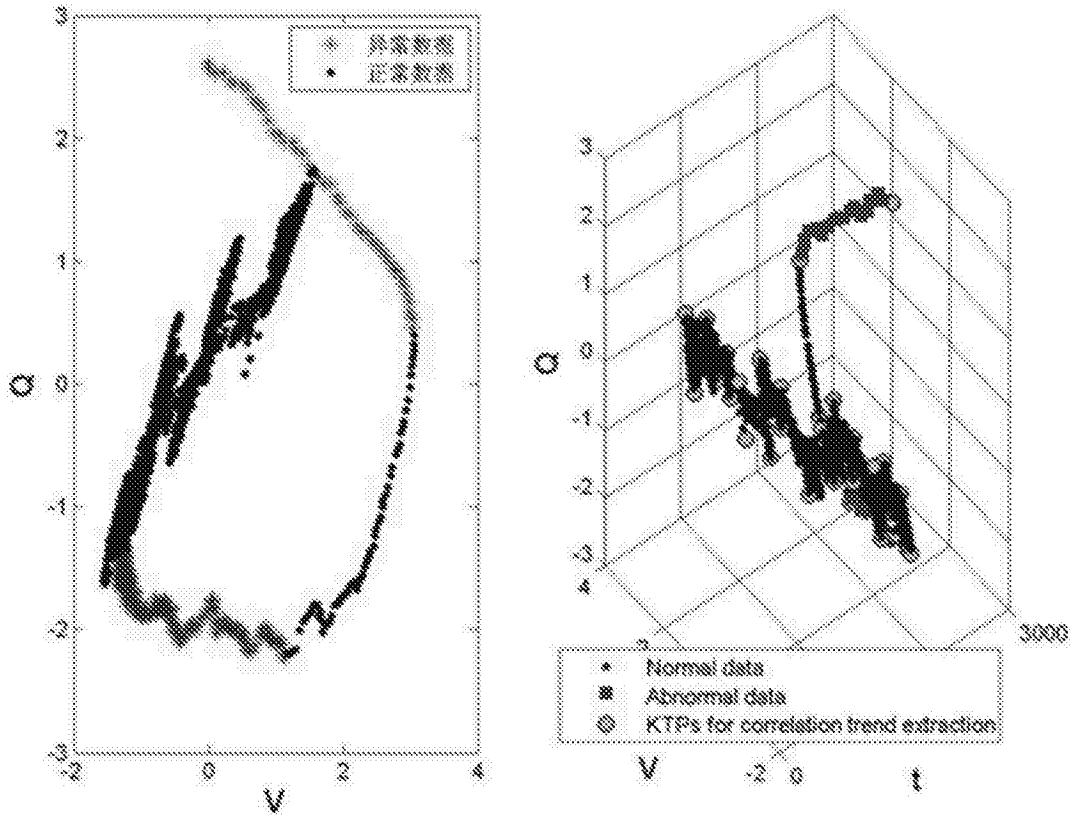

异常数据
正常数据
Q
V
t
3000
Normal data
Abnormal data
KTPs for correlation trend extraction

图5 (a)

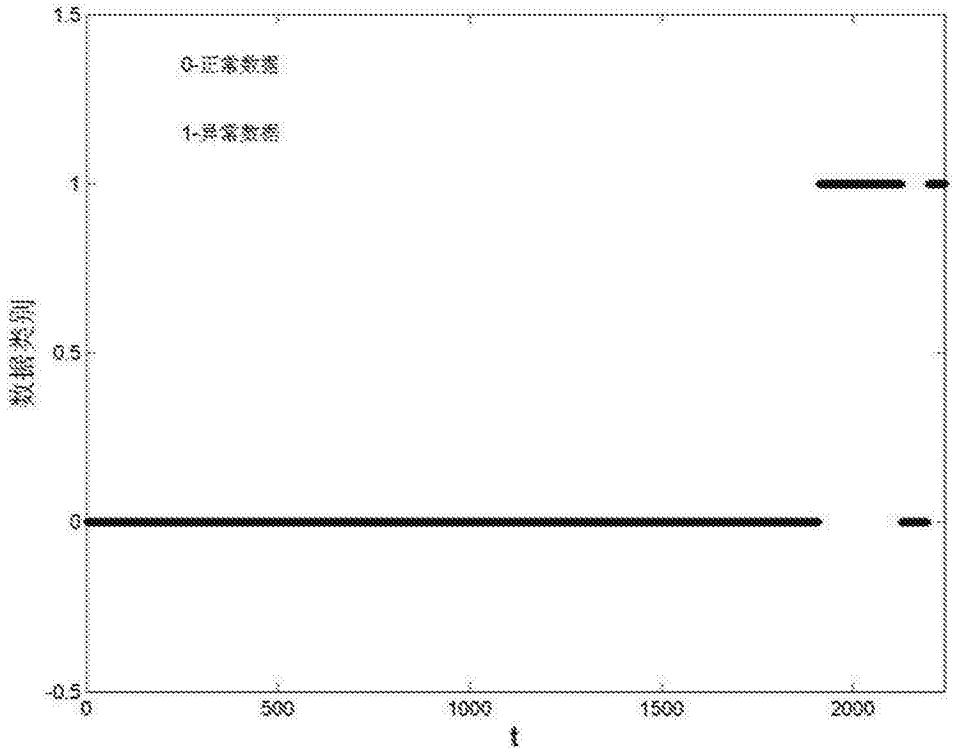


图5 (b)