



(12) 发明专利申请

(10) 申请公布号 CN 111931420 A

(43) 申请公布日 2020.11.13

(21) 申请号 202010786691.8

G06F 119/02 (2020.01)

(22) 申请日 2020.08.07

(71) 申请人 合肥工业大学

地址 230009 安徽省合肥市包河区屯溪路  
193号

(72) 发明人 杨静 朱尤杰 沈安波 樊高金

江刘峰 方宝富

(74) 专利代理机构 安徽省合肥新安专利代理有

限责任公司 34101

代理人 陆丽莉 何梅生

(51) Int. Cl.

G06F 30/27 (2020.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

G01M 15/14 (2006.01)

权利要求书3页 说明书8页

(54) 发明名称

基于核再生希尔伯特空间的燃气轮机故障  
预测方法

(57) 摘要

本发明公开了一种基于核再生希尔伯特空间的燃气轮机故障预测方法,其步骤包括:1、获取所有监测节点向量的数据集;2、将所有的向量用核函数映射的希尔伯特空间,并计算映射之后向量之间的典范相关系数;3、计算任意两个节点在给定其他节点后的偏相关系数;4、设定阈值算出强相关相关节点的搜索空间;5、在受限空间内使用爬山搜索,通过评分函数完成定向工作,确定本监控节点与其他监测节点间的因果关系;6、直到评分次数超过设定值前获得相应的监测系统因果结构图,并用于训练故障预测模型;从而得到故障预测模型,以实现故障进行更加准确的预测。

1. 一种基于核再生希尔伯特空间的燃气轮机故障预测方法,是应用于燃气轮机系统中,并获取燃气轮机 $m$ 个监测点的运行状态数据,每个检测点的运行状态数据组成状态向量 $X$ ,记为 $X = \{X_1, X_2, \dots, X_i, \dots, X_m\}$ ,其中 $X_i$ 表示第 $i$ 个监测点的状态数据, $i \in [1, m]$ ,其特征是,所述燃气轮机故障预测方法是按如下步骤进行:

步骤1、定义状态向量 $X$ 的候选邻接矩阵为 $PN_{m \times m}$ ;

步骤2、通过一定数量的实验找到监测点之间的强相关阈值 $K_\alpha$ ;

步骤3、计算第 $i$ 个监测点的状态数据 $X_i$ 与第 $j$ 个监测点的状态数据 $X_j$ 之间的近似典范相关系数 $\gamma(i, j)$ ,从而得到所有 $m$ 个监测点的状态数据之间的近似典范相关系数并构成近似典范相关系数矩阵 $\gamma_{m \times m}$ ;

步骤4、利用式(1)得到残差矩阵 $C_{m \times m}$ :

$$C_{m \times m} = \gamma_{m \times m}^{-1} \quad (1)$$

式(1)中, $\gamma_{m \times m}^{-1}$ 表示近似典范相关系数矩阵 $\gamma_{m \times m}$ 的逆;

步骤5、利用式(2)得到第 $i$ 个监测点的状态数据 $X_i$ 与第 $j$ 个监测点的状态数据 $X_j$ 之间的近似偏相关系数 $\rho(i, j)$ ,从而得到所有 $m$ 个监测点的状态数据之间的近似偏相关系数并构成近似偏相关系数矩阵 $\rho_{m \times m}$ ;

$$\rho(i, j) = -\frac{C(i, j)}{\sqrt{C(i, i)C(j, j)}} \quad (2)$$

式(2)中, $C(i, j)$ 表示残差矩阵 $C_{m \times m}$ 中第 $i$ 行第 $j$ 列的残差;

步骤6、判断 $\rho(i, j) > K_\alpha$ 是否成立,若成立,则表示第 $i$ 个监测点的状态数据 $X_i$ 与第 $j$ 个监测点的状态数据 $X_j$ 之间是强相关,并将候选邻接矩阵 $PN_{m \times m}$ 中的相应位置的元素 $PN(i, j)$ 置为“1”,否则,表示第 $i$ 个监测点的状态数据 $X_i$ 与第 $j$ 个监测点的状态数据 $X_j$ 之间是弱相关或者不相关,并将候选邻接矩阵 $PN_{m \times m}$ 中的相应位置的元素 $PN(i, j)$ 置为“0”;

步骤7、利用爬山搜索算法对候选邻接矩阵 $PN_{m \times m}$ 进行爬山搜索,从而得到监测点关系网络图,其中,所述监测点关系网络图的每个监测点的父节点和子节点都是相应监测点的相关监测点;

步骤8、任意选择一个监测点的状态数据,并作为LSTM神经网络模型的输出,再将与所选择的监测点相关的监测点的状态数据作为LSTM神经网络模型的输入,从而训练LSTM神经网络模型,从而得到故障预测模型;

步骤9、实时监测任意一个监测点的状态数据并获得相应的燃气轮机状态数据集,再利用所述故障预测模型得到实时监测的监测点的预测值,然后将所述预测值与实时监测的监测点的真实值进行比较,当超过所设定的阈值,则表示相应监测点可能发生故障,并给出预警提示。

2. 根据权利要求1所述的燃气轮机故障预测方法,其特征是,所述步骤3是按如下步骤进行:

步骤3.1、将第 $i$ 个监测点的状态数据 $X_i$ 和第 $j$ 个监测点的状态数据 $X_j$ 分别用核方法映射到核再生希尔伯特空间,从而得到再生希尔伯特空间的第 $i$ 个状态数据 $\psi(X_i)$ 和第 $j$ 个状态数据 $\psi(X_j)$ ;

步骤3.2、利用式(3)得到典范相关系数 $\gamma'(i, j)$ :

$$\gamma'(i, j) = \frac{\text{cov}(\psi(X_i)\psi(X_j))}{\sqrt{\text{cov}(\psi(X_i)\psi(X_i))\text{cov}(\psi(X_j)\psi(X_j))}} \quad (3)$$

式(3)中,  $\text{cov}(\cdot)$  表示协方差, 并有:

$$\text{cov}(X_i X_j) = E(\psi(X_i) - E(\psi(X_i)))(\psi(X_j) - E(\psi(X_j))) \quad (4)$$

式(4)中,  $E(\cdot)$  是期望, 并有:

$$E(\psi(X_i)) = \frac{1}{m} \sum_p^m \kappa(X_i, \cdot) \quad (5)$$

$$E(\psi(X_i)\psi(X_j)) = \frac{1}{m} \sum_p^m \kappa(X_i, \cdot)(\kappa(X_j, \cdot))^T \quad (6)$$

式(4)和式(5)中,  $\kappa(\cdot, \cdot)$  是高斯核函数;

步骤3.3、利用式(7)建立傅里变换与高斯核函数的关系式:

$$Z(X_i)Z(X_j)^T \approx \kappa(X_i, X_j) \quad (7)$$

式(7)中,  $Z(\cdot)$  表示傅里叶变化后得到的特征函数;

步骤3.3、利用式(8)得到近似典范相关系数  $\gamma(i, j)$ :

$$\gamma(i, j) = \frac{\|Z(X_i)Z(X_j)\|_F^2}{\|Z(X_i)Z(X_i)\|_F^2\|Z(X_j)Z(X_j)\|_F^2} \quad (8)$$

3. 根据权利要求1所述的燃气轮机故障预测方法, 其特征是, 所述步骤7是按如下步骤进行:

步骤7.1、利用式(9)构建评分函数MDL:

$$MDL = \sum_{j=1}^n (NLL(X_j, \mathbf{Pa}(X_j), \hat{\theta}_j^{mle}) + \frac{|\hat{\theta}_j^{mle}|}{2} \log m) \quad (9)$$

式(9)中, 评分函数MDL由数据的拟合优度  $\sum_{j=1}^n (NLL(X_j, \mathbf{Pa}(X_j), \hat{\theta}_j^{mle}))$  和网络复杂度  $\frac{|\hat{\theta}_j^{mle}|}{2} \log m$  共同决定;  $\mathbf{Pa}(X_j)$  表示第j个监测点的相关监测点集;  $|\hat{\theta}_j^{mle}|$  表示第j个监测点  $X_j$

的自由变量个数,  $\hat{\theta}_j^{mle}$  是第j个变量  $\theta_j$  在表示  $NLL(X_j, \mathbf{Pa}(X_j), \theta_j)$  取最小值时的最大似然估计, 并有:

$$\hat{\theta}_j^{mle} = \arg \min_{\theta} NLL(X_j, \mathbf{Pa}(X_j), \theta) \quad (10)$$

式(10)中,  $NLL(X_j, \mathbf{Pa}(X_j), \theta_j)$  表示第j个监测点  $X_j$  在具有相关监测点集  $\mathbf{Pa}(X_j)$  和第j个变量  $\theta_j$  时的评分;

步骤7.2、利用式(11)计算第j个监测点  $X_j$  在具有相关监测点集  $\mathbf{Pa}(X_j)$  和参数  $\hat{\theta}_j^{mle}$  时的评分  $NLL(X_j, \mathbf{Pa}(X_j), \hat{\theta}_j^{mle})$ :

$$NLL(X_j, \mathbf{Pa}(X_j), \hat{\theta}_j^{mle}) = \sum_{k=1}^m (x_{jk} - (\hat{\theta}_j^{mle})^T \mathbf{pa}(x_{jk}))^2 \quad (11)$$

式(11)中,  $x_{jk}$  表示第j个监测点  $X_j$  在第k个样本的实际取值,  $\mathbf{pa}(x_{jk})$  表示第j个监测点  $X_j$

在其相关监测点集 $Pa(X_j)$ 在第 $k$ 个样本的实际取值; $(\hat{\theta}_j^{mle})^T \mathbf{pa}(x_{jk})$ 是第 $j$ 个监测点 $X_j$ 在其相关监测点集 $Pa(X_j)$ 取值为 $\mathbf{pa}(x_{jk})$ 和权值为 $\hat{\theta}_j^{mle}$ 时,计算得出的预测值;

步骤7.3、定义 $rs$ 表示重新搜索,并初始化 $rs$ 为1;

定义 $GminScore$ 和 $GminGraph$ 分别表示最小MDL评分及其对应的网络,并初始化 $GminScore$ 为无穷大,定义 $Iter$ 表示迭代的次数,并初始化为0;

步骤7.4、将 $Iter+1$ 赋值给 $Iter$ ,若 $rs=1$ ,则表示执行重新搜索;并执行步骤9.6;若 $rs=0$ 时表示不执行重新搜索;

步骤7.5、随机生成一个不存在环结构的邻接矩阵 $g_{m \times n}$ 作为第 $Iter$ 次迭代的初选网络;

利用式(9)所示的评分函数MDL计算所述第 $Iter$ 次迭代的初选网络的得分 $GraphScore_{Iter}$ ,并更新评分次数 $ENu$ 为 $ENu+n$ ;

步骤7.6、在第 $Iter$ 次迭代的网络中针对第 $i$ 个监测点添加一条通往第 $j$ 个监测点的边,再计算添加边后的初选网络的得分,并存储在添加操作矩阵的相应元素 $AddENu(i, j)$ 中,从而得到所有监测点添加边后的初选网络的得分,并构成添加操作矩阵 $AddENu$ ;

步骤7.7、在第 $Iter$ 次迭代的网络中针对第 $i$ 个监测点,当第 $j$ 个监测点没有子孙节点时,则删除一条第 $i$ 个监测点通往第 $j$ 个监测点的边,再计算删除边后的初选网络的得分,并存储在删除操作矩阵的相应元素 $DelENu(i, j)$ 中,从而得到所有监测点删除边后的初选网络的得分,并构成删除操作矩阵 $DelENu$ ;

步骤7.8、在第 $Iter$ 次迭代的初选网络中针对第 $i$ 个监测点当第 $j$ 个监测点没有子孙节点时,则逆置第 $j$ 个监测点和第 $i$ 个监测点的方向,再计算逆置边后的初选网络的得分,并存储在逆置操作矩阵的相应元素 $RevENu(i, j)$ 中,从而得到所有监测点删除边后的初选网络的得分,并构成逆置操作矩阵 $RevENu$ ;

步骤7.9、分别从所有的添加边后的初选网络的得分、删除边后的初选网络的得分和逆置边后的初选网络的得分中选出最小值,再从三个所选出的最小值中选出一个最小值所对应的操作作为最终操作,并对第 $Iter$ 次迭代的初选网络执行所述最终操作,从而得到最终操作后的第 $Iter$ 次迭代的初选网络,并计算相应的网络分数 $GraphScore_{Iter}$ ;

步骤7.10、将最终操作后的第 $Iter$ 次迭代的初选网络的网络分数 $GraphScore_{Iter}$ 赋值给第 $Iter$ 次迭代的局部最优得分 $LminScore$ ,并令 $rs=0$ ;

步骤7.11、判断 $LminScore < GminScore$ 是否成立,若成立,则将 $LminScore$ 赋值给 $GminScore$ ,并根据最终操作更新邻接矩阵 $g_{m \times n}$ 后,执行步骤7.12;否则,直接执行步骤7.12;

步骤7.12、从最终操作所对应的操作矩阵中删掉相应的最终操作所对应的网络分数,从而更新相应的操作矩阵;

步骤7.13、将 $Iter+1$ 赋值给 $Iter$ ,判断 $Iter > MaxIter$ 是否成立,若成立,若成立,则输出 $GminScore$ 和 $GminGraph$ ,其中 $GminGraph$ 即为最终获得的监测点关系网络图,否则返回步骤7.9。

## 基于核再生希尔伯特空间的燃气轮机故障预测方法

### 技术领域

[0001] 本发明属于数据挖掘领域,具体地说是一种基于核再生希尔伯特空间检测燃气轮机故障之间的偏相关系数的燃气轮机故障预测方法。

### 背景技术

[0002] 随着网络和技术的发展,无论是社会上还是企业中的数据都成几何倍的增长,数据的形式也越来越复杂,关于如何从庞大的数据中提取有用的数据信息的课题的研究也越来越多,显然对于企业而讲这些数据更加有商业价值,比如如何从企业的机组运行检测信息中发现当前机组运行的状态并预测将来机组状态的走势,可以有效的避免企业人力物力的损失,极大的提高企业运转的效率,因此故障诊断是一个非常值得研究的课题。

[0003] 目前国内燃气轮机的监控技术越来越发达,能检测的数据量也越来越多,完全满足从数据中挖掘数据信息的要求,因此燃气轮机机组状态分析、性能监测和故障智能诊断预测研究,具有非常重要的现实意义。我们可以通过数据建模,挖掘当前燃气轮机机组运行的状态信息,进行实时的监控,进一步可以预测燃气轮机机组运行的走势,如果预测发生故障,可以根据数据建模得到的因果网络结构图进行针对性的检修,节约维修时间,提高燃气轮机机组的运行效率,帮助燃气轮机机组安全可靠的运行。然而这些监控的数据往往是非线性的不平稳的异构的,这对于数据建模是一种非常巨大的挑战,所以研究非线性之间的关系就变得非常有必要。

[0004] 而描述复杂网络间关系的杰出模型就是由美国加州大学的Judea Pearl提出的基于概率论和图论的贝叶斯网络模型,并凭杰出的贡献获得2011年度图灵奖。Hoyer等对于贝叶斯网络因果模型进行更进一步扩展,提出了加性噪声模型,该模型可以建模非高斯非线性的数据。而燃气轮机机组的运行数据恰恰也是非高斯非线性的。所以,基于加性噪声模型对燃气轮机机组的运行数据进行分析是非常有意义的研究方向。至于加性噪声模型的结构学习,Hoyer等基于非线性回归和基于HSIC标准提出了识别因果结构的方法,Mooij等提出了基于HSIC回归的算法,Zhang等提出了两阶段的算法,Tillman等人提出了kPC算法,Yamada等提出了最小二乘独立性回归的方法,Mooij等提出了基于最大后验的方法,Zhang等提出一个基于核的条件独立测试,Peters等提出了基于后续独立测试的回归方法,Zhang等提出一个基于回归的条件独立型测试的方法等,Nowzohour等基于惩罚性的似然的方法等等。

[0005] 目前这些方法的主要局限包括:

[0006] (1) 燃气轮机的故障预测采用直接知识表示的方法,速度快,但知识库中的故障较少,面对新的故障时候就不能有效的进行故障预测,很大可能会导致诊断失误;

[0007] (2) 单独的神经网络诊断方法,不能对燃气轮机系统内部的各部件关系进行准确的揭示,只能通过数据进行表面分析,这也给故障预测带来很大的不确定性。

## 发明内容

[0008] 本发明为克服现有技术存在的不足之处,提出了一种基于核再生希尔伯特空间的燃气轮机故障预测方法,以期适应于线性和非线性的故障预测,并能提高对故障的预测准确度。

[0009] 本发明为达到上述发明目的,采用如下技术方案:

[0010] 本发明一种基于核再生希尔伯特空间的燃气轮机故障预测方法,是应用于燃气轮机系统中,并获取燃气轮机 $m$ 个监测点的运行状态数据,每个检测点的运行状态数据组成状态向量 $X$ ,记为 $X = \{X_1, X_2, \dots, X_i, \dots, X_m\}$ ,其中 $X_i$ 表示第 $i$ 个监测点的状态数据, $i \in [1, m]$ ,其特点是,所述燃气轮机故障预测方法是按如下步骤进行:

[0011] 步骤1、定义状态向量 $X$ 的候选邻接矩阵为 $PN_{m \times m}$ ;

[0012] 步骤2、通过一定数量的实验找到监测点之间的强相关阈值 $K_\alpha$ ;

[0013] 步骤3、计算第 $i$ 个监测点的状态数据 $X_i$ 与第 $j$ 个监测点的状态数据 $X_j$ 之间的近似典范相关系数 $\gamma(i, j)$ ,从而得到所有 $m$ 个监测点的状态数据之间的近似典范相关系数并构成近似典范相关系数矩阵 $\gamma_{m \times m}$ ;

[0014] 步骤4、利用式(1)得到残差矩阵 $C_{m \times m}$ ;

$$[0015] \quad C_{m \times m} = \gamma_{m \times m}^{-1} \quad (1)$$

[0016] 式(1)中, $\gamma_{m \times m}^{-1}$ 表示近似典范相关系数矩阵 $\gamma_{m \times m}$ 的逆;

[0017] 步骤5、利用式(2)得到第 $i$ 个监测点的状态数据 $X_i$ 与第 $j$ 个监测点的状态数据 $X_j$ 之间的近似偏相关系数 $\rho(i, j)$ ,从而得到所有 $m$ 个监测点的状态数据之间的近似偏相关系数并构成近似偏相关系数矩阵 $\rho_{m \times m}$ ;

$$[0018] \quad \rho(i, j) = -\frac{C(i, j)}{\sqrt{C(i, i)C(j, j)}} \quad (2)$$

[0019] 式(2)中, $C(i, j)$ 表示残差矩阵 $C_{m \times m}$ 中第 $i$ 行第 $j$ 列的残差;

[0020] 步骤6、判断 $\rho(i, j) > K_\alpha$ 是否成立,若成立,则表示第 $i$ 个监测点的状态数据 $X_i$ 与第 $j$ 个监测点的状态数据 $X_j$ 之间是强相关,并将候选邻接矩阵 $PN_{m \times m}$ 中的相应位置的元素 $PN(i, j)$ 置为“1”,否则,表示第 $i$ 个监测点的状态数据 $X_i$ 与第 $j$ 个监测点的状态数据 $X_j$ 之间是弱相关或者不相关,并将候选邻接矩阵 $PN_{m \times m}$ 中的相应位置的元素 $PN(i, j)$ 置为“0”;

[0021] 步骤7、利用爬山搜索算法对候选邻接矩阵 $PN_{m \times m}$ 进行爬山搜索,从而得到监测点关系网络图,其中,所述监测点关系网络图的每个监测点的父节点和子节点都是相应监测点的相关监测点;

[0022] 步骤8、任意选择一个监测点的状态数据,并作为LSTM神经网络模型的输出,再将与所选择的监测点相关的监测点的状态数据作为LSTM神经网络模型的输入,从而训练LSTM神经网络模型,从而得到故障预测模型;

[0023] 步骤9、实时监测任意一个监测点的状态数据并获得相应的燃气轮机状态数据集,再利用所述故障预测模型得到实时监测的监测点的预测值,然后将所述预测值与实时监测的监测点的真实值进行比较,当超过所设定的阈值,则表示相应监测点可能发生故障,并给出预警提示。

[0024] 本发明所述的燃气轮机故障预测方法的特点是,所述步骤3是按如下步骤进行:

[0025] 步骤3.1、将第*i*个监测点的状态数据 $X_i$ 和第*j*个监测点的状态数据 $X_j$ 分别用核方法映射到核再生希尔伯特空间,从而得到再生希尔伯特空间的第*j*个状态数据 $\psi(X_i)$ 和第*j*个状态数据 $\psi(X_j)$ ;

[0026] 步骤3.2、利用式(3)得到典范相关系数 $\gamma'(i, j)$ :

$$[0027] \quad \gamma'(i, j) = \frac{\text{cov}(\psi(X_i)\psi(X_j))}{\sqrt{\text{cov}(\psi(X_i)\psi(X_i))\text{cov}(\psi(X_j)\psi(X_j))}} \quad (3)$$

[0028] 式(3)中, $\text{cov}(\cdot)$ 表示协方差,并有:

$$[0029] \quad \text{cov}(X_i X_j) = E(\psi(X_i) - E(\psi(X_i)))(\psi(X_j) - E(\psi(X_j))) \quad (4)$$

[0030] 式(4)中, $E(\cdot)$ 是期望,并有:

$$[0031] \quad E(\psi(X_i)) = \frac{1}{m} \sum_p \kappa(X_i, \cdot) \quad (5)$$

$$[0032] \quad E(\psi(X_i)\psi(X_j)) = \frac{1}{m} \sum_p \kappa(X_i, \cdot)(\kappa(X_j, \cdot))^T \quad (6)$$

[0033] 式(4)和式(5)中, $\kappa(\cdot, \cdot)$ 是高斯核函数;

[0034] 步骤3.3、利用式(7)建立傅里变换与高斯核函数的关系式:

$$[0035] \quad Z(X_i) Z(X_j)^T \approx \kappa(X_i, X_j) \quad (7)$$

[0036] 式(7)中, $Z(\cdot)$ 表示傅里叶变化后得到的特征函数;

[0037] 步骤3.3、利用式(8)得到近似典范相关系数 $\gamma(i, j)$ :

$$[0038] \quad \gamma(i, j) = \frac{\|Z(X_i)Z(X_j)\|_F^2}{\|Z(X_i)Z(X_i)\|_F^2 \|Z(X_j)Z(X_j)\|_F^2} \quad (8)$$

[0039] 所述步骤7是按如下步骤进行:

[0040] 步骤7.1、利用式(9)构建评分函数MDL:

$$[0041] \quad MDL = \sum_{j=1}^n (NLL(X_j, \mathbf{Pa}(X_j), \hat{\theta}_j^{mle}) + \frac{|\hat{\theta}_j^{mle}|}{2} \log m) \quad (9)$$

[0042] 式(9)中,评分函数MDL由数据的拟合优度 $\sum_{j=1}^n (NLL(X_j, \mathbf{Pa}(X_j), \hat{\theta}_j^{mle}))$ 和网络复杂度

$\frac{|\hat{\theta}_j^{mle}|}{2} \log m$ 共同决定; $\mathbf{Pa}(X_j)$ 表示第*j*个监测点的相关监测点集; $|\hat{\theta}_j^{mle}|$ 表示第*j*个监测点 $X_j$

的自由变量个数, $\hat{\theta}_j^{mle}$ 是第*j*个变量 $\theta_j$ 在表示 $NLL(X_j, \mathbf{Pa}(X_j), \theta_j)$ 取最小值时的最大似然估计,并有:

$$[0043] \quad \hat{\theta}_j^{mle} = \arg \min_{\theta} NLL(X_j, \mathbf{Pa}(X_j), \theta) \quad (10)$$

[0044] 式(10)中, $NLL(X_j, \mathbf{Pa}(X_j), \theta_j)$ 表示第*j*个监测点 $X_j$ 在具有相关监测点集 $\mathbf{Pa}(X_j)$ 和第*j*个变量 $\theta_j$ 时的评分;

[0045] 步骤7.2、利用式(11)计算第*j*个监测点 $X_j$ 在具有相关监测点集 $\mathbf{Pa}(X_j)$ 和参数 $\hat{\theta}_j^{mle}$ 时的评分 $NLL(X_j, \mathbf{Pa}(X_j), \hat{\theta}_j^{mle})$ :

$$[0046] \quad NLL(X_j, \mathbf{Pa}(X_j), \hat{\boldsymbol{\theta}}_j^{mle}) = \sum_{k=1}^m (x_{jk} - (\hat{\boldsymbol{\theta}}_j^{mle})^T \mathbf{pa}(x_{jk}))^2 \quad (11)$$

[0047] 式(11)中,  $x_{jk}$ 表示第j个监测点 $X_j$ 在第k个样本的实际取值,  $\mathbf{pa}(x_{jk})$ 表示第j个监测点 $X_j$ 在其相关监测点集 $\mathbf{Pa}(X_j)$ 在第k个样本的实际取值;  $(\hat{\boldsymbol{\theta}}_j^{mle})^T \mathbf{pa}(x_{jk})$ 是第j个监测点 $X_j$ 在其相关监测点集 $\mathbf{Pa}(X_j)$ 取值为 $\mathbf{pa}(x_{jk})$ 和权值为 $\hat{\boldsymbol{\theta}}_j^{mle}$ 时, 计算得出的预测值;

[0048] 步骤7.3、定义rs表示重新搜索, 并初始化rs为1;

[0049] 定义G min Score和G min Graph分别表示最小MDL评分及其对应的网络, 并初始化G min Score为无穷大, 定义Iter表示迭代的次数, 并初始化为0;

[0050] 步骤7.4、将Iter+1赋值给Iter, 若rs=1, 则表示执行重新搜索; 并执行步骤9.6; 若rs=0时表示不执行重新搜索;

[0051] 步骤7.5、随机生成一个不存在环结构的邻接矩阵 $g_{m \times m}$ 作为第Iter次迭代的初选网络;

[0052] 利用式(9)所示的评分函数MDL计算所述第Iter次迭代的初选网络的得分GraphScore<sub>Iter</sub>, 并更新评分次数ENu为ENu+n;

[0053] 步骤7.6、在第Iter次迭代的网络中针对第i个监测点添加一条通往第j个监测点的边, 再计算添加边后的初选网络的得分, 并存储在添加操作矩阵的相应元素AddENu(i, j)中, 从而得到所有监测点添加边后的初选网络的得分, 并构成添加操作矩阵AddENu;

[0054] 步骤7.7、在第Iter次迭代的网络中针对第i个监测点, 当第j个监测点没有子孙节点时, 则删除一条第i个监测点通往第j个监测点的边, 再计算删除边后的初选网络的得分, 并存储在删除操作矩阵的相应元素DelENu(i, j)中, 从而得到所有监测点删除边后的初选网络的得分, 并构成删除操作矩阵DelENu;

[0055] 步骤7.8、在第Iter次迭代的初选网络中针对第i个监测点当第j个监测点没有子孙节点时, 则逆置第j个监测点和第i个监测点的方向, 再计算逆置边后的初选网络的得分, 并存储在逆置操作矩阵的相应元素Re vENu(i, j)中, 从而得到所有监测点删除边后的初选网络的得分, 并构成逆置操作矩阵Re vENu;

[0056] 步骤7.9、分别从所有的添加边后的初选网络的得分、删除边后的初选网络的得分和逆置边后的初选网络的得分中选出最小值, 再从三个所选出的最小值中选出一个最小值所对应的操作作为最终操作, 并对第Iter次迭代的初选网络执行所述最终操作, 从而得到最终操作后的第Iter次迭代的初选网络, 并计算相应的网络分数GraphScore<sub>Iter</sub>;

[0057] 步骤7.10、将最终操作后的第Iter次迭代的初选网络的网络分数GraphScore<sub>Iter</sub>赋值给第Iter次迭代的局部最优得分L min Score, 并令rs=0;

[0058] 步骤7.11、判断L min Score < G min Score是否成立, 若成立, 则将L min Score赋值给G min Score, 并根据最终操作更新邻接矩阵 $g_{m \times m}$ 后, 执行步骤7.12; 否则, 直接执行步骤7.12;

[0059] 步骤7.12、从最终操作所对应的操作矩阵中删掉相应的最终操作所对应的网络分数, 从而更新相应的操作矩阵;

[0060] 步骤7.13、将Iter+1赋值给Iter, 判断Iter > MaxIter是否成立, 若成立, 若成立, 则输出G min Score和G min Graph, 其中G min Graph即为最终获得的监测点关系网络图,

否则返回步骤7.9。

[0061] 与已有技术相比,本发明的有益效果体现在:

[0062] 1、针对燃气轮机机组运行监测数据分布往往是任意的,彼此之间的关系往往具有非线性的特点,本发明基于希尔伯特空间映射改变状态数据之间的函数关系,是一个新研究,大大减少了爬山搜索的学习复杂度,从而满足了燃气轮机状态实时监测的需要。

[0063] 2、偏相关只能应用于线性的检测数据中,燃气轮机机组运行监测数据分布往往是任意的,本发明利用映射将非线性关系转化为无限维的线性关系后,用无限维状态数据的均值估计典范相关性公式中的期望,从而得到近似典范相关系数,进一步得到近似的偏相关系数,可以准确的反应检测点之间的相关程度,克服了偏相关不能识别非线性之间相关程度的缺点,并且保留了偏相关可以识别线性之间的相关程度的优点,从而满足了监测数据的要求。

### 具体实施方式

[0064] 本实施例中,一种基于核再生希尔伯特空间的燃气轮机故障预测方法,是应用于燃气轮机系统中,并获取燃气轮机 $m$ 个监测点的运行状态数据,每个检测点的运行状态数据组成状态向量 $X$ ,记为 $X = \{X_1, X_2, \dots, X_i, \dots, X_m\}$ ,其中 $X_i$ 表示第 $i$ 个监测点的状态数据, $i \in [1, m]$ ,该燃气轮机故障预测方法目的是为了找出监测节点间的关系,找到与任意监测节点相关性较强的监测节点,并在该方法的基础上,使用神经网络的方法对于监测节点的未来趋势进行预测,从而对为燃气轮机的运行状态进行监测和故障预警。具体的说,该燃气轮机故障预测方法是按如下步骤进行的:

[0065] 步骤1、定义状态向量 $X$ 的候选邻接矩阵为 $PN_{m \times m}$ ;

[0066] 步骤2、通过一定数量的实验找到监测点之间的强相关阈值 $K_\alpha$ ;

[0067] 步骤3、计算第 $i$ 个监测点的状态数据 $X_i$ 与第 $j$ 个监测点的状态数据 $X_j$ 之间的近似典范相关系数 $\gamma(i, j)$ ,从而得到所有 $m$ 个监测点的状态数据之间的近似典范相关系数并构成近似典范相关系数矩阵 $\gamma_{m \times m}$ ;

[0068] 步骤3.1、将第 $i$ 个监测点的状态数据 $X_i$ 和第 $j$ 个监测点的状态数据 $X_j$ 分别用核方法映射到核再生希尔伯特空间,从而得到再生希尔伯特空间的第 $i$ 个状态数据 $\psi(X_i)$ 和第 $j$ 个状态数据 $\psi(X_j)$ ;

[0069] 步骤3.2、利用式(1)得到典范相关系数 $\gamma'(i, j)$ :

$$[0070] \quad \gamma'(i, j) = \frac{\text{cov}(\psi(X_i)\psi(X_j))}{\sqrt{\text{cov}(\psi(X_i)\psi(X_i))\text{cov}(\psi(X_j)\psi(X_j))}} \quad (1)$$

[0071] 式(1)中, $\text{cov}(\cdot)$ 表示协方差,并有:

$$[0072] \quad \text{cov}(X_i X_j) = E(\psi(X_i) - E(\psi(X_i)))(\psi(X_j) - E(\psi(X_j))) \quad (2)$$

[0073] 式(2)中, $E(\cdot)$ 是期望,并有:

$$[0074] \quad E(\psi(X_i)) = \frac{1}{m} \sum_p^m \kappa(X_i, \cdot) \quad (3)$$

$$[0075] \quad E(\psi(X_i)\psi(X_j)) = \frac{1}{m} \sum_p^m \kappa(X_i, \cdot)(\kappa(X_j, \cdot))^T \quad (4)$$

[0076] 式(3)和式(4)中, $\kappa(\cdot, \cdot)$ 是高斯核函数;

[0077] 步骤3.3、利用式(5)建立傅里变换与高斯核函数的关系式:

$$[0078] \quad Z(X_i) Z(X_j)^T \approx \kappa(X_i, X_j) \quad (5)$$

[0079] 式(5)中, $Z(\cdot)$ 表示傅里叶变化后得到的特征函数;

[0080] 步骤3.3、利用式(6)得到近似典范相关系数 $\gamma(i, j)$ :

$$[0081] \quad \gamma(i, j) = \frac{\|Z(X_i) Z(X_j)\|_F^2}{\|Z(X_i) Z(X_i)\|_F^2 \|Z(X_j) Z(X_j)\|_F^2} \quad (6)$$

[0082] 步骤4、利用式(7)得到残差矩阵 $C_{m \times m}$ :

$$[0083] \quad C_{m \times m} = \gamma_{m \times m}^{-1} \quad (7)$$

[0084] 式(7)中, $\gamma_{m \times m}^{-1}$ 表示典范相关系数矩阵 $\gamma_{m \times m}$ 的逆;

[0085] 步骤5、利用式(8)得到第*i*个监测点的状态数据 $X_i$ 与第*j*个监测点的状态数据 $X_j$ 之间的偏相关系数 $\rho(i, j)$ ,从而得到所有*m*个监测点的状态数据之间的偏相关系数并构成偏相关系数矩阵 $\rho_{m \times m}$ ;

$$[0086] \quad \rho(i, j) = -\frac{C(i, j)}{\sqrt{C(i, i)C(j, j)}} \quad (8)$$

[0087] 式(8)中, $C(i, j)$ 表示残差矩阵 $C_{m \times m}$ 中第*i*行第*j*列的残差;

[0088] 步骤6、判断 $\rho(i, j) > K_\alpha$ 是否成立,若成立,则表示第*i*个监测点的状态数据 $X_i$ 与第*j*个监测点的状态数据 $X_j$ 之间是强相关,并将候选邻接矩阵 $PN_{m \times m}$ 中的相应位置的元素 $PN(i, j)$ 置为“1”,否则,表示第*i*个监测点的状态数据 $X_i$ 与第*j*个监测点的状态数据 $X_j$ 之间是弱相关或者不相关,并将候选邻接矩阵 $PN_{m \times m}$ 中的相应位置的元素 $PN(i, j)$ 置为“0”;

[0089] 步骤7、利用爬山搜索算法对候选邻接矩阵 $PN_{m \times m}$ 进行爬山搜索,从而得到监测点关系网络图,其中,监测点关系网络图的每个监测点的父节点和子节点都是相应监测点的相关监测点;

[0090] 步骤7.1、利用式(9)构建评分函数MDL:

$$[0091] \quad MDL = \sum_{j=1}^n (NLL(X_j, \mathbf{Pa}(X_j), \hat{\theta}_j^{mle}) + \frac{|\hat{\theta}_j^{mle}|}{2} \log m) \quad (9)$$

[0092] 式(9)中,评分函数MDL由数据的拟合优度 $\sum_{j=1}^n (NLL(X_j, \mathbf{Pa}(X_j), \hat{\theta}_j^{mle}))$ 和网络复杂度 $\frac{|\hat{\theta}_j^{mle}|}{2} \log m$ 共同决定; $\mathbf{Pa}(X_j)$ 表示第*j*个监测点的相关监测点集; $|\hat{\theta}_j^{mle}|$ 表示第*j*个监测点 $X_j$

的自由变量个数, $\hat{\theta}_j^{mle}$ 是第*j*个变量 $\theta_j$ 在表示 $NLL(X_j, \mathbf{Pa}(X_j), \theta_j)$ 取最小值时的最大似然估计,并有:

$$[0093] \quad \hat{\theta}_j^{mle} = \arg \min_{\theta} NLL(X_j, \mathbf{Pa}(X_j), \theta_j) \quad (10)$$

[0094] 式(10)中, $NLL(X_j, \mathbf{Pa}(X_j), \theta_j)$ 表示第*j*个监测点 $X_j$ 在具有相关监测点集 $\mathbf{Pa}(X_j)$ 和第*j*个变量 $\theta_j$ 时的评分;

[0095] 步骤7.2、利用式(11)计算第*j*个监测点 $X_j$ 在具有相关监测点集 $\mathbf{Pa}(X_j)$ 和参数 $\hat{\theta}_j^{mle}$

时的评分  $NLL(X_j, \mathbf{Pa}(X_j), \hat{\boldsymbol{\theta}}_j^{mle})$  :

$$[0096] \quad NLL(X_j, \mathbf{Pa}(X_j), \hat{\boldsymbol{\theta}}_j^{mle}) = \sum_{k=1}^m (x_{jk} - (\hat{\boldsymbol{\theta}}_j^{mle})^T \mathbf{pa}(x_{jk}))^2 \quad (11)$$

[0097] 式(11)中,  $x_{jk}$ 表示第j个监测点 $X_j$ 在第k个样本的实际取值,  $\mathbf{pa}(x_{jk})$ 表示第j个监测点 $X_j$ 在其相关监测点集 $\mathbf{Pa}(X_j)$ 在第k个样本的实际取值;  $(\hat{\boldsymbol{\theta}}_j^{mle})^T \mathbf{pa}(x_{jk})$ 是第j个监测点 $X_j$ 在其相关监测点集 $\mathbf{Pa}(X_j)$ 取值为 $\mathbf{pa}(x_{jk})$ 和权值为 $\hat{\boldsymbol{\theta}}_j^{mle}$ 时,计算得出的预测值;

[0098] 步骤7.3、定义rs表示重新搜索,并初始化rs为1;

[0099] 定义G min Score和G min Graph分别表示最小MDL评分及其对应的网络,并初始化G min Score为无穷大,定义Iter表示迭代的次数,并初始化为0;

[0100] 步骤7.4、将Iter+1赋值给Iter,若rs=1,则表示执行重新搜索;并执行步骤9.6;若rs=0时表示不执行重新搜索;

[0101] 步骤7.5、随机生成一个不存在环结构的邻接矩阵 $g_{m \times m}$ 作为第Iter次迭代的初选网络;

[0102] 利用式(9)所示的评分函数MDL计算第Iter次迭代的初选网络的得分GraphScore<sub>Iter</sub>,并更新评分次数ENu为ENu+n;

[0103] 步骤7.6、在第Iter次迭代的网络中针对第i个监测点添加一条通往第j个监测点的边,再计算添加边后的初选网络的得分,并存储在添加操作矩阵的相应元素AddENu(i, j)中,从而得到所有监测点添加边后的初选网络的得分,并构成添加操作矩阵AddENu;

[0104] 步骤7.7、在第Iter次迭代的网络中针对第i个监测点,当第j个监测点没有子孙节点时,则删除一条第i个监测点通往第j个监测点的边,再计算删除边后的初选网络的得分,并存储在删除操作矩阵的相应元素DelENu(i, j)中,从而得到所有监测点删除边后的初选网络的得分,并构成删除操作矩阵DelENu;

[0105] 步骤7.8、在第Iter次迭代的初选网络中针对第i个监测点当第j个监测点没有子孙节点时,则逆置第j个监测点和第i个监测点的方向,再计算逆置边后的初选网络的得分,并存储在逆置操作矩阵的相应元素Re vENu(i, j)中,从而得到所有监测点删除边后的初选网络的得分,并构成逆置操作矩阵Re vENu;

[0106] 步骤7.9、分别从所有的添加边后的初选网络的得分、删除边后的初选网络的得分和逆置边后的初选网络的得分中选出最小值,再从三个所选出的最小值中选出一个最小值所对应的操作作为最终操作,并对第Iter次迭代的初选网络执行最终操作,从而得到最终操作后的第Iter次迭代的初选网络,并计算相应的网络分数GraphScore<sub>Iter</sub>;

[0107] 步骤7.10、将最终操作后的第Iter次迭代的初选网络的网络分数GraphScore<sub>Iter</sub>赋值给第Iter次迭代的局部最优得分L min Score,并令rs=0;

[0108] 步骤7.11、判断L min Score < G min Score是否成立,若成立,则将L min Score赋值给G min Score,并根据最终操作更新邻接矩阵 $g_{m \times m}$ 后,执行步骤7.12;否则,直接执行步骤7.12;

[0109] 步骤7.12、从最终操作所对应的操作矩阵中删掉相应的最终操作所对应的网络分数,从而更新相应的操作矩阵;

[0110] 步骤7.13、将Iter+1赋值给Iter,判断Iter>MaxIter是否成立,若成立,若成立,则输出G min Score和G min Graph,其中G min Graph即为最终获得的监测点关系网络图,否则返回步骤7.9。

[0111] 步骤8、任意选择一个监测点的状态数据,并作为LSTM神经网络模型的输出,再将与所选择的监测点相关的监测点的状态数据作为LSTM神经网络模型的输入,从而训练LSTM神经网络模型,从而得到故障预测模型;

[0112] 步骤9、实时监测任意一个监测点的状态数据并获得相应的燃气轮机状态数据集,再利用故障预测模型得到实时监测的监测点的预测值,然后将预测值与实时监测的监测点的真实值进行比较,当超过所设定的阈值,则表示相应监测点可能发生故障,并给出预警提示。