



(12)发明专利申请

(10)申请公布号 CN 109447309 A

(43)申请公布日 2019.03.08

(21)申请号 201811029046.0

(51)Int.Cl.

(22)申请日 2018.09.05

G06Q 10/04(2012.01)

G06Q 50/06(2012.01)

(71)申请人 中国电力科学研究院有限公司

G06K 9/62(2006.01)

地址 100192 北京市海淀区清河小营东路
15号

申请人 国家电网有限公司

中国科学院沈阳自动化研究所

(72)发明人 张立春 刘彬 唐凤珍 程永锋
姬昆鹏 费香泽 汉京善 斯白露
司佳钧 杨加伦 展雪萍 赵彬
李鹏 李丹煜

(74)专利代理机构 北京安博达知识产权代理有
限公司 11271

代理人 徐国文

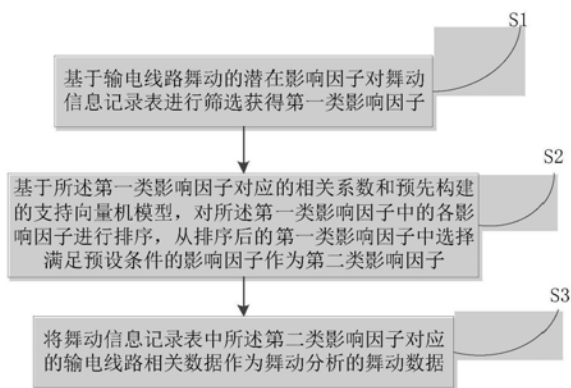
权利要求书3页 说明书10页 附图2页

(54)发明名称

一种舞动数据的挖掘方法及系统

(57)摘要

一种舞动数据的挖掘方法及系统,包括:基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子;基于所述第一类影响因子对应的相关系数和预先构建的支持向量机模型,对所述第一类影响因子中的各影响因子进行排序,从排序后的第一类影响因子中选择满足预设条件的影响因子作为第二类影响因子;将舞动信息记录表中所述第二类影响因子对应的输电线路相关数据作为舞动分析的舞动数据。本发明可以提取出与舞动发生率密切相关的影响因子,提高舞动预测预警的准确率。



1. 一种舞动数据的挖掘方法,其特征在于,包括:

基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子;

基于所述第一类影响因子对应的相关系数和预先构建的支持向量机模型,对所述第一类影响因子中的各影响因子进行排序,从排序后的第一类影响因子中选择满足预设条件的影响因子作为第二类影响因子;

将舞动信息记录表中所述第二类影响因子对应的输电线路相关数据作为舞动分析的舞动数据。

2. 如权利要求1所述的挖掘方法,其特征在于,所述基于所述第一类影响因子对应的相关系数和预先构建的支持向量机模型,对所述第一类影响因子中的各影响因子进行排序,包括:

从所述舞动信息记录表中选择第一影响因子及所述第一影响因子涉及的输电线路相关数据,生成分析数据表;

基于分析数据表计算所述第一类影响因子中各影响因子对应的均值和均方差;

基于各影响因子对应的均值和均方差计算各影响因子的相关系数;

基于各影响因子的相关系数的绝对值对所述第一类影响因子中的各影响因子进行排序得到第一排序结果;

基于分析数据表和支持向量机模型,对所述第一类影响因子中的各影响因子进行排序得到第二排序结果。

3. 如权利要求2所述的挖掘方法,其特征在于,所述从排序后的第一类影响因子中选择满足预设条件的影响因子作为第二类影响因子,包括:

比较第一排序结果中第一类影响因子的各影响因子排序结果,和第二排序结果中第一类影响因子的各影响因子排序结果;

当所述第一排序结果和第二排序结果中顺序为阈值时的多个影响因子相同,则选择相同的影响因子作为第二类影响因子。

4. 如权利要求3所述的挖掘方法,其特征在于,所述基于分析数据表和支持向量机模型,对所述第一类影响因子中的各影响因子进行排序得到第二排序结果,包括:

将第一类影响因子中的全部影响因子设置为原始特征集合,同时设置一个为空集的特征排序集;

将第一类影响因子和所述第一类影响因子对应的输电线路相关数据作为训练样本;

基于所述训练样本和设置的第一公式训练支持向量机分类器得到超平面的权值向量;

基于所述超平面的权值向量和设置的第二公式计算排序准则分数;

将最小的排序准则分数对应的影响因子从所述原始特征集合中删除,并将基于所述最小的排序准则分数得到的结果特征添加到所述特征排序集中;

从当前训练样本中删除最小的排序准则分数对应的影响因子及输电线路相关数据,得到更新后的训练样本;

基于所述更新后训练样本继续查找最小的排序准则分数对应的影响因子,直到所述原始特征集合为空集时,将当前特征排序集对应的影响因子作为第二排序结果。

5. 如权利要求4所述的挖掘方法,其特征在于,所述超平面的权值向量,按下式计算:

$$\omega = \sum_{i=1}^N \alpha_i y_i x_i$$

式中： ω ：超平面的权值向量； α_i ：第*i*个影响因子的拉格朗日乘子； x_i ：第*i*个影响因子的值； y_i ：第*i*个影响因子的类别标签； N ：训练样本个数。

6. 如权利要求4所述的挖掘方法，其特征在于，所述排序准则分数，按下式计算：

$$c_k = \omega_k^2$$

式中： c_k ：第*k*个影响因子的排序准则分数； ω_k^2 ：第*k*个影响因子的超平面的权值向量； k 的取值为原始特征集合中影响因子的个数。

7. 如权利要求4所述的挖掘方法，其特征在于，所述结果特征，按下式计算：

$$p = \operatorname{argmin} c_k$$

式中： p ：由*k*的值组成的结果特征； c_k ：第*k*个影响因子的排序准则分数。

8. 如权利要求3所述的挖掘方法，其特征在于，所述各影响因子的相关系数，按下式计算：

$$R(k) = \frac{\mu_k(+)-\mu_k(-)}{\sigma_k(+)+\sigma_k(-)}$$

式中： $R(k)$ ：第*k*个影响因子的相关系数； $\mu_k(+)$ ：第*k*个影响因子对应(+1)类的均值； $\mu_k(-)$ ：第*k*个影响因子对应(-1)类的均值； $\sigma_k(+)$ ：第*k*个影响因子对应(+1)类的均方差； $\sigma_k(-)$ ：第*k*个影响因子对应(-1)类的均方差。

9. 如权利要求1所述的挖掘方法，其特征在于，所述基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子，包括：

基于输电线路舞动的潜在影响因子获取所述潜在影响因子涉及的输电线路相关数据；

基于潜在影响因子和所述潜在影响因子涉及的输电线路相关数据构建舞动信息记录表；

从所述舞动信息记录表筛选出待分析的影响因子组成第一类影响因子；

其中，所述输电线路舞动的潜在影响因子，包括：风速、风向、线路走向与风向夹角、温度、湿度、降水形式、导线覆冰厚度、地形、海拔高度、电压等级、杆塔回路数、绞线截面积、导线排列方式和导线分裂数。

10. 如权利要求9所述的挖掘方法，其特征在于，所述基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子，还包括：

将第一类影响因子中通过文字描述的影响因子按设置的转换原则转化为数值形式。

11. 一种舞动数据的挖掘系统，其特征在于，包括：

筛选模块，用于基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子；

排序模块，用于基于所述第一类影响因子对应的相关系数和预先构建的支持向量机模型，对所述第一类影响因子中的各影响因子进行排序，从排序后的第一类影响因子中选择满足预设条件的影响因子作为第二类影响因子；

分析模块，用于将舞动信息记录表中所述第二类影响因子对应的输电线路相关数据作为舞动分析的舞动数据。

12. 如权利要求11所述的挖掘系统,其特征在于,所述排序模块,包括:

生成子模块,用于从所述舞动信息记录表中选择第一影响因子及所述第一影响因子涉及的输电线路相关数据,生成分析数据表;

第一计算子模块,用于基于分析数据表计算所述第一类影响因子中各影响因子对应的均值和均方差;

第二计算子模块,用于基于各影响因子对应的均值和均方差计算各影响因子的相关系数;

第一排序结果子模块,用于基于各影响因子的相关系数的绝对值对所述第一类影响因子中的各影响因子进行排序得到第一排序结果;

第二排序结果子模块,用于基于分析数据表和支持向量机模型,对所述第一类影响因子中的各影响因子进行排序得到第二排序结果。

一种舞动数据的挖掘方法及系统

技术领域

[0001] 本发明涉及输电线路防灾减灾领域,具体涉及一种舞动数据的挖掘方法及系统。

背景技术

[0002] 输电线路舞动事故是威胁电网安全稳定运行的重要灾害形式之一,输电线路舞动易导致线路跳闸、导线断股、金具螺栓松动,当发生大范围输电线路舞动事故时,会导致电网大面积停电,严重影响大电网安全稳定运行。近年来,随着人工智能技术的快速发展,其在电网防灾减灾领域的应用日渐深入,人们将人工智能技术引入输电线路舞动预测预警工作中,利用积累的大量舞动相关数据,对舞动事故的发生可能性进行预测预警,从而使得线路运维部门能够提前预判灾害的发生,制定应对措施,降低舞动事故对电网产生的损害。然而,舞动预测的准确率还不尽如人意,原因在于舞动数据的利用和挖掘水平还有待提高,舞动相关数据形式多样,如何从海量复杂的数据中提取出对于舞动研究最具价值的信息,是提高舞动智能预警精度和准确率的基础和关键,因此,需要对舞动数据挖掘技术进行深入研究。

发明内容

[0003] 为了解决现有技术中所存在的舞动预测准确率低的问题,本发明提供一种舞动数据的挖掘方法及系统。

[0004] 本发明提供的技术方案是:一种舞动数据的挖掘方法,包括:

[0005] 基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子;

[0006] 基于所述第一类影响因子对应的相关系数和预先构建的支持向量机模型,对所述第一类影响因子中的各影响因子进行排序,从排序后的第一类影响因子中选择满足预设条件的影响因子作为第二类影响因子;

[0007] 将舞动信息记录表中所述第二类影响因子对应的输电线路相关数据作为舞动分析的舞动数据。

[0008] 优选的,所述基于所述第一类影响因子对应的相关系数和预先构建的支持向量机模型,对所述第一类影响因子中的各影响因子进行排序,包括:

[0009] 从所述舞动信息记录表中选择第一影响因子及所述第一影响因子涉及的输电线路相关数据,生成分析数据表;

[0010] 基于分析数据表计算所述第一类影响因子中各影响因子对应的均值和均方差;

[0011] 基于各影响因子对应的均值和均方差计算各影响因子的相关系数;

[0012] 基于各影响因子的相关系数的绝对值对所述第一类影响因子中的各影响因子进行排序得到第一排序结果;

[0013] 基于分析数据表和支持向量机模型,对所述第一类影响因子中的各影响因子进行排序得到第二排序结果。

[0014] 优选的,所述从排序后的第一类影响因子中选择满足预设条件的影响因子作为第二类影响因子,包括:

[0015] 比较第一排序结果中第一类影响因子的各影响因子排序结果,和第二排序结果中第一类影响因子的各影响因子排序结果;

[0016] 当所述第一排序结果和第二排序结果中顺序为阈值时的多个影响因子相同,则选择相同的影响因子作为第二类影响因子。

[0017] 优选的,所述基于分析数据表和支持向量机模型,对所述第一类影响因子中的各影响因子进行排序得到第二排序结果,包括:

[0018] 将第一类影响因子中的全部影响因子设置为原始特征集合,同时设置一个为空集的特征排序集;

[0019] 将第一类影响因子和所述第一类影响因子对应的输电线路相关数据作为训练样本;

[0020] 基于所述训练样本和设置的第一公式训练支持向量机分类器得到超平面的权值向量;

[0021] 基于所述超平面的权值向量和设置的第二公式计算排序准则分数;

[0022] 将最小的排序准则分数对应的影响因子从所述原始特征集合中删除,并将基于所述最小的排序准则分数得到的结果特征添加到所述特征排序集中;

[0023] 从当前训练样本中删除最小的排序准则分数对应的影响因子及输电线路相关数据,得到更新后的训练样本;

[0024] 基于所述更新后训练样本继续查找最小的排序准则分数对应的影响因子,直到所述原始特征集合为空集时,将当前特征排序集对应的影响因子作为第二排序结果。

[0025] 优选的,所述超平面的权值向量,按下式计算:

$$[0026] \quad \omega = \sum_{i=1}^N \alpha_i y_i x_i$$

[0027] 式中: ω :超平面的权值向量; α_i :第*i*个影响因子的拉格朗日乘子; x_i :第*i*个影响因子的值; y_i :第*i*个影响因子的类别标签; N :训练样本个数。

[0028] 优选的,所述排序准则分数,按下式计算:

$$[0029] \quad c_k = \omega_k^2$$

[0030] 式中: c_k :第*k*个影响因子的排序准则分数; ω_k^2 :第*k*个影响因子的超平面的权值向量; k 的取值为原始特征集合中影响因子的个数。

[0031] 优选的,所述结果特征,按下式计算:

$$[0032] \quad p = \arg \min c_k$$

[0033] 式中: p :由*k*的值组成的结果特征; c_k :第*k*个影响因子的排序准则分数。

[0034] 优选的,所述各影响因子的相关系数,按下式计算:

$$[0035] \quad R(k) = \frac{\mu_k(+)-\mu_k(-)}{\sigma_k(+)+\sigma_k(-)}$$

[0036] 式中: $R(k)$:第*k*个影响因子的相关系数; $\mu_k(+)$:第*k*个影响因子对应(+1)类的均值; $\mu_k(-)$:第*k*个影响因子对应(-1)类的均值; $\sigma_k(+)$:第*k*个影响因子对应(+1)类的均方差;

$\sigma_k(-)$:第k个影响因子对应(-1)类的均方差。

[0037] 优选的,所述基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子,包括:

[0038] 基于输电线路舞动的潜在影响因子获取所述潜在影响因子涉及的输电线路相关数据;

[0039] 基于潜在影响因子和所述潜在影响因子涉及的输电线路相关数据构建舞动信息记录表;

[0040] 从所述舞动信息记录表筛选出待分析的影响因子组成第一类影响因子;

[0041] 其中,所述输电线路舞动的潜在影响因子,包括:风速、风向、线路走向与风向夹角、温度、湿度、降水形式、导线覆冰厚度、地形、海拔高度、电压等级、杆塔回路数、绞线截面积、导线排列方式和导线分裂数。

[0042] 优选的,所述基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子,还包括:

[0043] 将第一类影响因子中通过文字描述的影响因子按设置的转换原则转化为数值形式。

[0044] 基于同一发明构思,本发明还提供了一种舞动数据的挖掘系统,包括:

[0045] 筛选模块,用于基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子;

[0046] 排序模块,用于基于所述第一类影响因子对应的相关系数和预先构建的支持向量机模型,对所述第一类影响因子中的各影响因子进行排序,从排序后的第一类影响因子中选择满足预设条件的影响因子作为第二类影响因子;

[0047] 分析模块,用于将舞动信息记录表中所述第二类影响因子对应的输电线路相关数据作为舞动分析的舞动数据。

[0048] 优选的,所述排序模块,包括:

[0049] 生成子模块,用于从所述舞动信息记录表中选择第一影响因子及所述第一影响因子涉及的输电线路相关数据,生成分析数据表;

[0050] 第一计算子模块,用于基于分析数据表计算所述第一类影响因子中各影响因子对应的均值和均方差;

[0051] 第二计算子模块,用于基于各影响因子对应的均值和均方差计算各影响因子的相关系数;

[0052] 第一排序结果子模块,用于基于各影响因子的相关系数的绝对值对所述第一类影响因子中的各影响因子进行排序得到第一排序结果;

[0053] 第二排序结果子模块,用于基于分析数据表和支持向量机模型,对所述第一类影响因子中的各影响因子进行排序得到第二排序结果。

[0054] 与现有技术相比,本发明的有益效果为:

[0055] 本发明提供的技术方案,基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子;基于所述第一类影响因子对应的相关系数和预先构建的支持向量机模型,对所述第一类影响因子中的各影响因子进行排序,从排序后的第一类影响因子中选择满足预设条件的影响因子作为第二类影响因子;将舞动信息记录表中所述第二类

影响因子对应的输电线路相关数据作为舞动分析的舞动数据,可以提取出与舞动发生率密切相关的影响因子,提高舞动预测预警的准确率。

[0056] 本发明提供的技术方案,可以适用于所有不同区域和不同气象条件下,分析输电线路舞动的影响因子。

[0057] 本发明提供的技术方案,可以有效的挖掘出与输电线路舞动高度相关的影响因子,为输电线路舞动预测建模做出重大贡献,为提前部署输电线路防舞措施争取了宝贵的时间。

附图说明

[0058] 图1为本发明提供的一种舞动数据的挖掘方法流程图;

[0059] 图2为本发明实施例提供的舞动信息记录表示意图;

[0060] 图3为本发明实施例中线路走向坐标转换示意图;

[0061] 图4为本发明实施例中第一排序结果的柱状示意图。

具体实施方式

[0062] 为了更好地理解本发明,下面结合说明书附图和实例对本发明的内容做进一步的说明。

[0063] 实施例1:

[0064] S1、基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子;

[0065] S2、基于所述第一类影响因子对应的相关系数和预先构建的支持向量机模型,对所述第一类影响因子中的各影响因子进行排序,从排序后的第一类影响因子中选择满足预设条件的影响因子作为第二类影响因子;

[0066] S3、将舞动信息记录表中所述第二类影响因子对应的输电线路相关数据作为舞动分析的舞动数据。

[0067] 本实施例中采用的文本数据为“舞动信息记录表”为例,通过本发明提供的挖掘方法获取输电线路进行舞动分析的舞动数据,其中部分数据如图2所示。

[0068] S1、基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子,具体包括:

[0069] 1.1舞动数据描述

[0070] 1.1.1原始数据及其格式化

[0071] 输电线路舞动相关数据庞杂,形式多元,包括图像、文本、视频等多种类型,本发明只针对文本数据进行研究。

[0072] 基于输电线路舞动的潜在影响因子获取潜在影响因子涉及的输电线路相关数据;

[0073] 基于潜在影响因子和潜在影响因子涉及的输电线路相关数据构建舞动信息记录表。

[0074] 图2中标记为舞动的记录记为正样本,标记为非舞动的记录记为负样本,显然正负样本分布不均衡,比例约为1:7。

[0075] 由于本发明旨在分析输电线路舞动的主要影响因素,因此,只选取与舞动成因相

关的数据进行分析。输电线路舞动的潜在影响因子可以归为三大类,一是气象因素,二是地理因素,三是输电线路自身的结构特征。

[0076] 1) 气象因素:风速、风向、线路走向与风向夹角、温度、湿度、降水形式、导线覆冰厚度。

[0077] 2) 地理因素:地形、海拔高度。

[0078] 3) 输电线路自身的结构特征:电压等级、杆塔回路数、绞线截面积、导线排列方式、导线分裂数。

[0079] 根据舞动发生机理以及舞动规律的理论分析结果,筛选以下12个影响因子组成第一类影响因子进行分析:电压等级、海拔高度、线路走向、风速、风向、风向与线路夹角、温度、湿度、降水形式、导线覆冰厚度、外绞线面积与导线分裂数。

[0080] 从舞动信息记录表中选择第一影响因子及第一影响因子涉及的输电线路相关数据,生成分析数据表。

[0081] 将第一类影响因子中通过文字描述的影响因子按设置的转换原则转化为数值形式。

[0082] 由图2可知,线路走向、风向、降水形式3个影响因子的值为文字描述,其余9个影响因子均为数值型信息,为了能够被数据挖掘工具所识别,需要将文字描述转换为数值型信息。其中线路走向和风向按照图3所示的坐标进行转换。具体的格式转换原则如表1.1至表1.3所示。

[0083] 表1.1线路走向格式转换原则

线路走向	对应数值 (°)
南北/北南	90
东西/西东	0
东北-西南/西南-东北	45
西北-东南/东南-西北	135
东西偏北 N°度	N
东西偏南 N 度	180-N
南北偏东 N 度	90-N
南北偏西 N 度	90+N
注*: N 为整数, 且 $0 < N < 180$	

[0085] 表1.2风向格式转换

[0086]

风向	对应数值
北(偏北)	90
东北	45
西北	135

[0087] 表1.3降水形式格式转换

[0088]

降水形式	对应数值
雾凇	1
雨凇	2
湿雪	3

[0089] S2、基于所述第一类影响因子对应的相关系数和预先构建的支持向量机模型,对所述第一类影响因子中的各影响因子进行排序,从排序后的第一类影响因子中选择满足预设条件的影响因子作为第二类影响因子,具体包括:

[0090] 1.2舞动影响因子与是否舞动的相关程度

[0091] 为了挖掘出与舞动事故发生高度相关的最主要影响因子,需要分析各个影响因子对于区分舞动样本与非舞动样本的重要性。本实施例分别采用相关性分析方法和递归特征消除法来分析各个影响因子对于舞动发生与否的重要性。

[0092] 1.2.1相关性分析法

[0093] 相关性分析是利用统计学原理对数据进行分析,利用相关系数来度量变量之间的相关程度,在本实施例中即计算各个影响因子对应的相关系数,按照相关系数的绝对值进行排序,从而判断各个影响因子对舞动事故发生率的相关性大小。

[0094] 是否舞动属于两类问题,即二值(binary)离散目标变量,对应于两个值+1或者-1,其相关系数的计算公式如下所示:

$$[0095] \quad R(k) = \frac{\mu_k(+)-\mu_k(-)}{\sigma_k(+)+\sigma_k(-)}$$

[0096] 其中R(k):第k个影响因子的相关系数,R>0表示与(+1)类相关,R<0表示与(-1)类相关,R的绝对值越大,相关性越强; $\mu_k(+)$:第k个影响因子对应(+1)类的均值; $\mu_k(-)$:第k个影响因子对应(-1)类的均值; $\sigma_k(+)$:第k个影响因子对应(+1)类的均方差; $\sigma_k(-)$:第k个影响因子对应(-1)类的均方差。

[0097] 计算结果如图4所示,从排序结果来看,就当前数据记录而言,电压等级、外绞线面积、导线分裂数、线路走向、风速等对于区分舞动发生与否重要性较高,而风向、风向与线路夹角等对于识别舞动的重要性次之。

[0098] 第一排序结果为:电压等级、外绞线面积、导线分裂数、线路走向、风速、降水形式、海拔高度、导线覆冰厚度、温度、风向与线路夹角以及风向。

[0099] 1.2.2递归特征消除法

[0100] 递归特征消除(Recursive Feature Elimination,RFE)的主要思想是使用一个机器学习模型(如支持向量机或者回归模型)反复训练多次,每次训练结束后,消除最差的特征(可以根据系数来选),然后在剩余的特征上重复这个过程,直到所有特征都遍历了。这个过程中特征被消除的次序就是特征的排序。最先消除的特征最不重要,最后消除的特征最为重要,该方法是一种寻找最优特征子集的贪心算法。

[0101] 本实施例中采用基于支持向量机(Support Vector Machine,SVM)模型的递归特征消除法(SVM-RFE),支持向量机(SVM)原理简述,详细可参见文献《数据挖掘导论》(人民邮电出版社)5.2.2节内容,包括:

[0102] 已知训练样本集 $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, 其中, $\mathbf{x}_i \in \mathbb{R}^D, y_i \in \{+1, -1\}$ 为 \mathbf{x}_i 的类别标签, N 为训练样本个数, D 为样本的特征维数。

[0103] SVM 寻求最优分类面 $\omega \cdot \mathbf{x} + b = 0$, 其中, ω 是超平面的权值向量, b 为阈值, 使得两类样本之间的分类间隔最大化, 为获得权值向量和阈值, 需求解下面的优化问题:

$$[0104] \quad \min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \zeta_i \quad (\text{式 1})$$

$$[0105] \quad \text{s. t.} \quad \begin{aligned} y_i(\omega \cdot \mathbf{x}_i + b) &\geq 1 - \zeta_i; i = 1, 2, \dots, N \\ \zeta_i &\geq 0; i = 1, 2, \dots, N \end{aligned} \quad (\text{式 2})$$

[0106] 其中, $C > 0$ 为惩罚参数, ζ_i 是松弛变量。

[0107] 参数 C 起到控制对错误划分样本进行惩罚程度的作用, 实现在错分样本的比例与算法复杂度之间的折中, 通过引入拉格朗日乘子, 可以将 SVM 的优化问题转变为以下的对偶规划问题:

$$[0108] \quad \min \left(\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) - \sum_{i=1}^N \alpha_i \right) \quad (\text{式 3})$$

$$[0109] \quad \text{s.t.} \quad \sum_{i=1}^N y_i \alpha_i = 0; 0 \leq \alpha_i \leq C, i = 1, 2, \dots, N$$

[0110] 其中, α_i 是拉格朗日乘子。

[0111] 权值向量和对偶优化 (式 11) 的解之间的关系为:

$$[0112] \quad \omega = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \quad (\text{式 4})$$

[0113] SVM 的判别函数为:

$$[0114] \quad f(x) = \text{sgn} \left(\sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b \right) \quad (\text{式 5})$$

[0115] 其中, $\text{sgn}(\cdot)$ 是符号函数。

[0116] SVM-RFE 特征选择算法是把需要的所有特征初始化为一个特征集合, 然后每次迭代剔除一个排序准则分数最小的特征, 直到获取最终的特征集, 因此, SVM-RFE 是一个基于 SVM 的最大间隔原理的序列后向选择 (Sequential Backward Selection, SBS) 算法。

[0117] 在 SVM-RFE 中, 第 i 个特征排序准则分数定义为:

$$[0118] \quad c_i = \omega_i^2 \quad (\text{式 6})$$

[0119] 在每次迭代中去掉排序准则分数最小的特征, 然后在用剩下的特征训练 SVM, 进行下一次迭代。

[0120] 本实施例中采用 SVM-RFE 特征选择算法获得第二排序结果, 包括:

[0121] 输入: 训练样本 $\{\mathbf{x}_i, y_i\}_{i=1}^N, y \in \{-1, +1\}$;

[0122] 输出: 特征排序集 R ;

[0123] 1) 初始化, 原始特征集合 $S = \{1, 2, \dots, D\}$, 特征排序集 $R = \emptyset$;

[0124] 2) 循环以下过程直至 $S = \emptyset$;

[0125] ① 获取带候选特征集合的训练样本, 本实施例中的训练样本为分析数据表;

- [0126] ②用式(4)训练SVM分类器,得到 ω ;
- [0127] ③用式(6)计算排序准则分数 $c_k = \omega_k^2, k=1, 2 \dots D$;
- [0128] ④找出排序准则分数最小的特征 $p = \arg \min_k c_k$, 即当 c_k 取最小值时, k 的取值;
- [0129] ⑤更新特征集 $R = \{p\} \cup R$;
- [0130] 在 S 中去除此特征 $S = S/p$ 。
- [0131] 最终得到特征排序集合 R , R 中 k 对应原始特征集合的影响因子。
- [0132] 采用REF法来对影响因子进行排序, 其第二排序结果如表4所示:
- [0133] 表4 RFE方法影响因子排序结果
- [0134]

排序	影响因子
1	风速
2	电压等级
3	线路走向
4	外绞线面积
5	导线分裂数
6	湿度
7	海拔高度
8	风向与线路夹角
9	导线覆冰厚度
10	降水形式
11	风向
12	温度

- [0135] 排序结果显示, 风速、电压等级、外绞线面积、线路走向和导线分裂数对于是否发生舞动的影响较为重要, 与相关系数法得到的结果一致。
- [0136] 从第一排序结果和第二排序结果中选择风速、电压等级、外绞线面积、线路走向和导线分裂数组成第二类影响因子;
- [0137] 将第二类影响因子对应的输电线路相关数据作为舞动分析的舞动数据。
- [0138] 实施例2:
- [0139] 基于同一种发明构思, 本发明还提供了一种舞动数据的挖掘系统, 包括:
- [0140] 筛选模块, 用于基于输电线路舞动的潜在影响因子对舞动信息记录表进行筛选获得第一类影响因子;
- [0141] 排序模块, 用于基于所述第一类影响因子对应的相关系数和预先构建的支持向量机模型, 对所述第一类影响因子中的各影响因子进行排序, 从排序后的第一类影响因子中选择满足预设条件的影响因子作为第二类影响因子;
- [0142] 分析模块, 用于将舞动信息记录表中所述第二类影响因子对应的输电线路相关数据作为舞动分析的舞动数据。
- [0143] 实施例中, 所述排序模块, 包括:
- [0144] 生子模块, 用于从所述舞动信息记录表中选择第一影响因子及所述第一影响因

子涉及的输电线路相关数据,生成分析数据表;

[0145] 第一计算子模块,用于基于分析数据表计算所述第一类影响因子中各影响因子对应的均值和均方差;

[0146] 第二计算子模块,用于基于各影响因子对应的均值和均方差计算各影响因子的相关系数;

[0147] 第一排序结果子模块,用于基于各影响因子的相关系数的绝对值对所述第一类影响因子中的各影响因子进行排序得到第一排序结果;

[0148] 第二排序结果子模块,用于基于分析数据表和支持向量机模型,对所述第一类影响因子中的各影响因子进行排序得到第二排序结果。

[0149] 实施例中,所述排序模块,还包括:

[0150] 比较子模块,用于比较第一排序结果中第一类影响因子的各影响因子排序结果,和第二排序结果中第一类影响因子的各影响因子排序结果;

[0151] 重组子模块,用于当所述第一排序结果和第二排序结果中顺序为阈值时的多个影响因子相同,则选择相同的影响因子作为第二类影响因子。

[0152] 实施例中,所述第二排序结果子模块,包括:

[0153] 初始单元,用于将第一类影响因子中的全部影响因子设置为原始特征集合,同时设置一个为空集的特征排序集;

[0154] 训练样本单元,用于将第一类影响因子和所述第一类影响因子对应的输电线路相关数据作为训练样本;

[0155] 计算权值向量单元,用于基于所述训练样本和设置的第一公式训练支持向量机分类器得到超平面的权值向量;

[0156] 计算排序准则分数单元,用于基于所述超平面的权值向量和设置的第二公式计算排序准则分数;

[0157] 处理单元,用于将最小的排序准则分数对应的影响因子从所述原始特征集合中删除,并将基于所述最小的排序准则分数得到的结果特征添加到所述特征排序集中;

[0158] 更新单元,用于从当前训练样本中删除最小的排序准则分数对应的影响因子及输电线路相关数据,得到更新后的训练样本;

[0159] 循环单元,用于基于所述更新后训练样本继续查找最小的排序准则分数对应的影响因子,直到所述原始特征集合为空集时,将当前特征排序集对应的影响因子作为第二排序结果。

[0160] 实施例中,所述筛选模块,包括:

[0161] 获取单元,用于基于输电线路舞动的潜在影响因子获取所述潜在影响因子涉及的输电线路相关数据;

[0162] 构建单元,用于基于潜在影响因子和所述潜在影响因子涉及的输电线路相关数据构建舞动信息记录表;

[0163] 重组单元,用于从所述舞动信息记录表筛选出待分析的影响因子组成第一类影响因子;

[0164] 其中,所述输电线路舞动的潜在影响因子,包括:风速、风向、线路走向与风向夹角、温度、湿度、降水形式、导线覆冰厚度、地形、海拔高度、电压等级、杆塔回路数、绞线截面

积、导线排列方式和导线分裂数。

[0165] 显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0166] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0167] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0168] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0169] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0170] 以上仅为本发明的实施例而已,并不用于限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均包含在申请待批的本发明的权利要求范围之内。

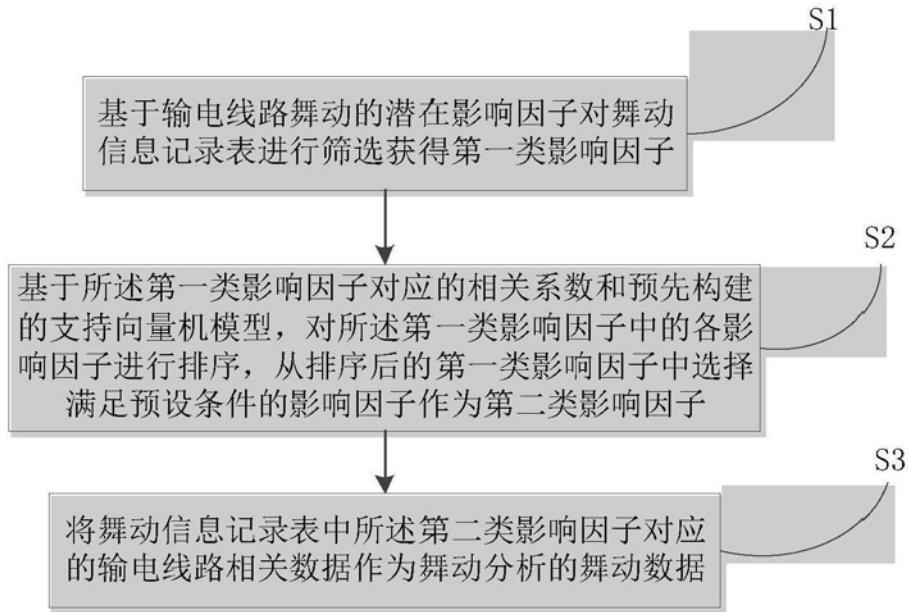


图1

序号	单位	线路名称	电压等级 (kV)	导线舞动区地形 (山区、丘陵区、平原)	海拔高度 (m)	舞动段线路走向	风速 (m/s)	风向	风向与舞动段线路走向夹角 (°)	温度 (°C)	湿度	降水形式 (雨、雾、雪、湿雪)	导线覆冰厚度 (mm)	杆塔回路数	导线型号	导线排列方式 (垂直、水平、三角)	导线分裂数	是否舞动
1	长春超高压局	丰徐2号线	500	丘陵	353	东西	6	西北	26	15	98%	雨淞	5	1	400/35	水平	4	是
2	锦州超高压局	董辽一线	500	山区	200	东西	12m-18m/s	北风	90	2~-7	50%-70%	雨淞	15	单回	LGJQ-300	水平	4	是
3	锦州超高压局	董辽一线	500	山区	250	东西	4m-6m/s	北风	90	0~-8	50%-70%	雨淞	5-8	单回	LGJQ-300	水平	4	否
4	锦州超高压局	董辽二线	500	山区	250	东西	4m-6m/s	北风	90	0~-8	50%-70%	雨淞	5-8	单回	LGJ-300/40	水平	4	否
5	锦州超高压局	沙董一线	500	平原、丘陵	100	东西	4m-6m/s	北风	90	0~-8	50%-70%	雨淞	5-8	单回	LGJ-400/35	水平	4	否

图2

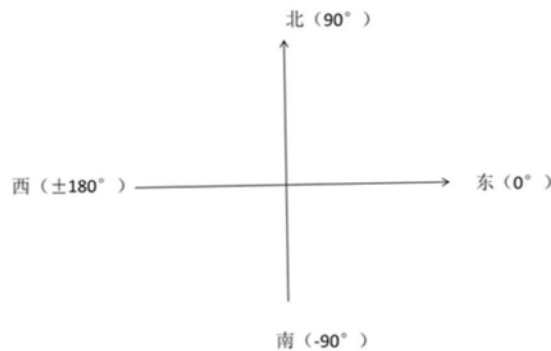


图3

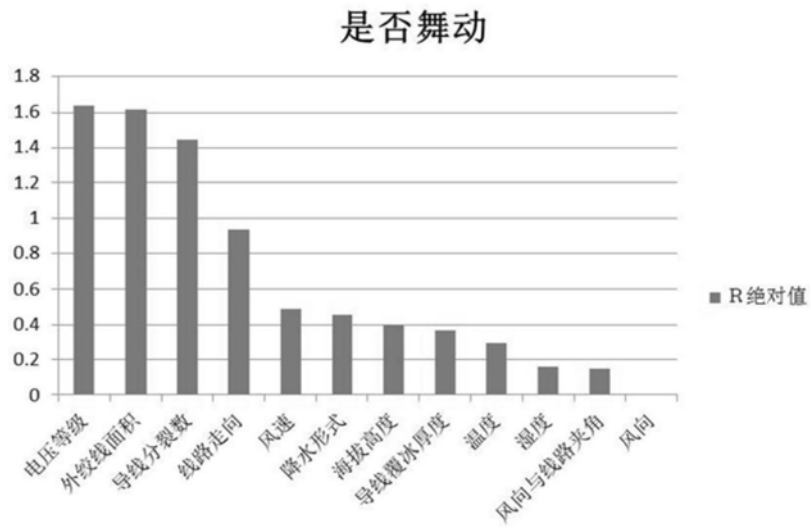


图4