



# (12)发明专利申请

(10)申请公布号 CN 108664990 A

(43)申请公布日 2018.10.16

(21)申请号 201810270822.X

(22)申请日 2018.03.29

(71)申请人 清华大学

地址 100084 北京市海淀区清华园

(72)发明人 陈启鑫 郑可迪 王毅 康重庆

夏清

(74)专利代理机构 北京清亦华知识产权代理事

务所(普通合伙) 11201

代理人 张润

(51) Int. Cl.

G06K 9/62(2006.01)

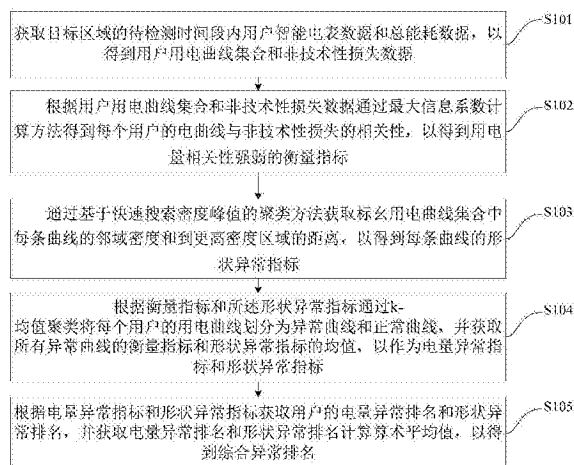
权利要求书3页 说明书8页 附图3页

## (54)发明名称

综合熵方法和密度聚类方法的窃电检测方法  
及装置

## (57)摘要

本发明公开了一种综合熵方法和密度聚类方法的窃电检测方法及装置,其中,方法包括:获取目标区域的待检测时间段内用户智能电表数据和总能耗数据,以得到用户用电曲线集合和非技术性损失数据;并根据用其通过最大信息系数计算方法得到用电量相关性强弱的衡量指标;得到每条曲线的形状异常指标;根据衡量指标和形状异常指标通过k-均值聚类将每个用户的用电曲线划分为异常曲线和正常曲线,得到电量异常指标和形状异常指标;进而获取用户的电量异常排名和形状异常排名,以得到综合异常排名。该方法能够提高各种窃电模式下检测的准确率,实现了大数据下窃电行为的非监督式检测,增大了适用范围。



1. 一种综合熵方法和密度聚类方法的窃电检测方法,其特征在于,包括以下步骤:

获取目标区域的待检测时间段内用户智能电表数据和总能耗数据,以得到用户用电曲线集合和非技术性损失数据;

根据所述用户用电曲线集合和非技术性损失数据通过最大信息系数计算方法得到每个用户的电曲线与非技术性损失的相关性,以得到用电量相关性强弱的衡量指标;

通过基于快速搜索密度峰值的聚类方法获取标么用电曲线集合中每条曲线的邻域密度和到更高密度区域的距离,以得到所述每条曲线的形状异常指标;

根据所述衡量指标和所述形状异常指标通过k-均值聚类将所述每个用户的用电曲线划分为异常曲线和正常曲线,并获取所有异常曲线的所述衡量指标和所述形状异常指标的均值,以作为电量异常指标和形状异常指标;以及

根据所述电量异常指标和所述形状异常指标获取用户的电量异常排名和形状异常排名,并获取所述电量异常排名和所述形状异常排名计算算术平均值,以得到综合异常排名。

2. 根据权利要求1所述的综合熵方法和密度聚类方法的窃电检测方法,其特征在于,所述获取目标区域的待检测时间段内用户智能电表数据和总能耗数据,以得到用户用电曲线集合和非技术性损失数据,进一步包括:

获得所述目标区域的待检测时间段内所有用户的用电曲线;

通过总用电量序列减去各时刻所有用户用电量之和得到所述待检测时间段内目标区域由窃电因素造成的非技术性损失;

将每个时刻的非技术性损失按日整理为对应的区域损失曲线。

3. 根据权利要求2所述的综合熵方法和密度聚类方法的窃电检测方法,其特征在于,其中,

所述根据所述用户用电曲线集合和非技术性损失数据通过最大信息系数计算方法得到每个用户的电曲线与非技术性损失的相关性,以得到用电量相关性强弱的衡量指标,进一步包括:

对待检测的每一日获取所述区域损失曲线和所述每个用户的用电曲线之间的最大信息系数,以得到每一用户曲线与区域损失之间电量相关性的强弱指标;

并且,所述通过基于快速搜索密度峰值的聚类方法获取标么用电曲线集合中每条曲线的邻域密度和到更高密度区域的距离,以得到所述每条曲线的形状异常指标,进一步包括:

对所有的用电曲线分别进行标么化,且保留曲线的形状特点;

获取标么用户曲线之间的欧式距离矩阵,并通过所述基于快速搜索密度峰值的聚类方法生成截断距离,并获取每条标么曲线的密度特征;

根据所述密度特征得到所述每条曲线的形状异常指标。

4. 根据权利要求3所述的综合熵方法和密度聚类方法的窃电检测方法,其特征在于,其中,

所述非技术性损失数据的计算公式为:

$$e_t = E_t - \sum_{i=1}^N x_{it},$$

其中, $e_t$ 为非技术性损失数据, $E_t$ 为观察表所记录的用电量, $x_{it}$ 为智能电表所记录的用户*i*在*t*时段的用电量, $t=1,2,\dots,M \times T$ ;

所述区域损失曲线的表达式为:

$$e_j = [e_{(j-1) \cdot T+1} \text{ L } e_{j \cdot T}],$$

其中,  $e_j$  为区域损失曲线,  $j=1, 2, \dots, M$  代表日下标。

5. 根据权利要求4所述的综合熵方法和密度聚类方法的窃电检测方法, 其特征在于, 其中,

所述标么化用户曲线集合为:

$$\bar{c}_{ij} = \frac{c_{ij}}{\max\{x_{i,(j-1)T+1}, L, x_{i,jT}\}} = \frac{c_{ij}}{\max c_{ij}},$$

其中,  $\bar{c}_{ij}$  为标么化用户曲线集合,  $c_{ij}$  为标么化用户曲线;

所述每条曲线的形状异常指标为:

$$\zeta_p = \frac{\delta_p}{\rho_p + 1},$$

其中,  $\zeta_p$  为每条曲线的形状异常指标,  $\rho_p$  为标么化曲线的邻域密度,  $\delta_p$  到更高密度区域的距离。

6. 一种综合熵方法和密度聚类方法的窃电检测装置, 其特征在于, 包括:

第一获取模块, 用于获取目标区域的待检测时间段内用户智能电表数据和总能耗数据, 以得到用户用电曲线集合和非技术性损失数据;

第二获取模块, 用于根据所述用户用电曲线集合和非技术性损失数据通过最大信息系数计算方法得到每个用户的电曲线与非技术性损失的相关性, 以得到用电量相关性强弱的衡量指标;

第三获取模块, 用于通过基于快速搜索密度峰值的聚类方法获取标么用电曲线集合中每条曲线的邻域密度和到更高密度区域的距离, 以得到所述每条曲线的形状异常指标;

划分模块, 用于根据所述衡量指标和所述形状异常指标通过k-均值聚类将所述每个用户的用电曲线划分为异常曲线和正常曲线, 并获取所有异常曲线的所述衡量指标和所述形状异常指标的均值, 以作为电量异常指标和形状异常指标; 以及

计算模块, 用于根据所述电量异常指标和所述形状异常指标获取用户的电量异常排名和形状异常排名, 并获取所述电量异常排名和所述形状异常排名计算算术平均值, 以得到综合异常排名。

7. 根据权利要求6所述的综合熵方法和密度聚类方法的窃电检测装置, 其特征在于, 所述第一获取模块, 还用于:

获得所述目标区域的待检测时间段内所有用户的用电曲线;

通过总用电量序列减去各时刻所有用户用电量之和得到所述待检测时间段内目标区域由窃电因素造成的非技术性损失;

将每个时刻的非技术性损失按日整理为对应的区域损失曲线。

8. 根据权利要求7所述的综合熵方法和密度聚类方法的窃电检测装置, 其特征在于, 其中,

所述第二获取模块还用于:

对待检测的每一日获取所述区域损失曲线和所述每个用户的用电曲线之间的最大信息系数, 以得到每一用户曲线与区域损失之间电量相关性的强弱指标;

所述第三获取模块还用于：

对所有的用电曲线分别进行标么化，且保留曲线的形状特点；

获取标么用户曲线之间的欧式距离矩阵，并通过所述基于快速搜索密度峰值的聚类方法生成截断距离，并获取每条标么曲线的密度特征；

根据所述密度特征得到所述每条曲线的形状异常指标。

9. 根据权利要求8所述的综合熵方法和密度聚类方法的窃电检测装置，其特征在于：

所述非技术性损失数据的计算公式为：

$$e_t = E_t - \sum_{i=1}^N x_{it},$$

其中， $e_t$ 为非技术性损失数据， $E_t$ 为观察表所记录的用电量， $x_{it}$ 为智能电表所记录的用戶*i*在*t*时段的用电量， $t=1, 2, \dots, M \times T$ ；

所述区域损失曲线的表达式为：

$$e_j = [e_{(j-1) \cdot T+1} \text{ L } e_{j \cdot T}],$$

其中， $e_j$ 为区域损失曲线， $j=1, 2, \dots, M$ 代表日下标。

10. 根据权利要求9所述的综合熵方法和密度聚类方法的窃电检测装置，其特征在于：

所述标么化用户曲线集合为：

$$\bar{c}_{ij} = \frac{c_{ij}}{\max\{x_{i,(j-1)T+1}, \text{L}, x_{i,jT}\}} = \frac{c_{ij}}{\max c_{ij}},$$

其中， $\bar{c}_{ij}$ 为标么化用户曲线集合， $c_{ij}$ 为标么化用户曲线；

所述每条曲线的形状异常指标为：

$$\zeta_p = \frac{\delta_p}{\rho_p + 1},$$

其中， $\zeta_p$ 为每条曲线的形状异常指标， $\rho_p$ 为标么化曲线的邻域密度， $\delta_p$ 到更高密度区域的距离。

## 综合熵方法和密度聚类方法的窃电检测方法及装置

### 技术领域

[0001] 本发明涉及防窃电分析技术领域,特别涉及一种综合熵方法和密度聚类方法的窃电检测方法及装置。

### 背景技术

[0002] 近年来,随着智能电表在电网中的普及,窃电者对电表的攻击方式不再局限于传统的物理切断或破坏,出现了运用数字存储技术和网络通讯技术对智能电表进行篡改的新模式。遭到这类新型攻击时,电网公司所获取的智能电表数据可能经过窃电者精心伪造。另一方面,智能电表可实现每天数十次以上频度的用户用电数据记录,为电网公司提供了海量数据。传统意义上对窃电行为的检测主要依靠电网公司派遣技术人员进行人工筛查或通过摄像头来监控窃电行为。这些方法通常会消耗电网公司较多的人力物力,同时依然无法避免通讯和网络中针对智能电表的高级攻击手段。因此,产生了通过智能电表提供的海量用电数据进行窃电检测的新技术,即通过对用户用电量的分析,提取其中的特征,并找出其中的异常用电行为,对窃电者进行更加精准的定位与更加细致的行为分析。

[0003] 目前,相关技术提出了基于极限学习机神经网络的窃电行为分类方法。相关研究人员也提出了类似的有监督式检测方法,并进行了测试,这一类的方法存在需要提供带窃电者标签的数据集对神经网络进行训练的局限性。因此在检测初期,必须依赖于人工检测并对窃电者的用电曲线进行记录与标记。在相关技术中,研究人员将非监督式学习的聚类方法,包括k-均值聚类、高斯混合模型聚类和吸引子传播算法等应用于窃电检测。非监督式学习的方法不需要有标签的数据集,更关注用电曲线的形状异常,但是检测准确率较低。

[0004] 然而,窃电用户对智能电表篡改行为的多样性为窃电检测带来极大的挑战,现在的攻击行为不再直接使电表记录的用电量归零,而是可以由用户来创造数据。用户可以根据实时电价的变化修改自己的用电曲线,最终降低电费。经过用户篡改后的用电曲线可能在形状上偏离原有的曲线,也可能保持了原有形状,但幅值变得更低,以产生电量上的偏差。因此,单纯从曲线形状或电量角度进行的分析无法覆盖所有篡改的方式,难以保证实际应用中的检测准确率。

### 发明内容

[0005] 本发明旨在至少在一定程度上解决相关技术中的技术问题之一。

[0006] 为此,本发明的一个目的在于提出一种综合熵方法和密度聚类方法的窃电方法,该方法可以提高各种窃电模式下检测的准确率。

[0007] 本发明的另一个目的在于提出一种综合熵方法和密度聚类方法的窃电检测装置。

[0008] 为达到上述目的,本发明一方面实施例提出了一种综合熵方法和密度聚类方法的窃电检测方法,包括以下步骤:获取目标区域的待检测时间段内用户智能电表数据和总能耗数据,以得到用户用电曲线集合和非技术性损失数据;根据所述用户用电曲线集合和非技术性损失数据通过最大信息系数计算方法得到每个用户的电曲线与非技术性损失的相

关性,以得到用电量相关性强弱的衡量指标;通过基于快速搜索密度峰值的聚类方法获取标么用电曲线集合中每条曲线的邻域密度和到更高密度区域的距离,以得到所述每条曲线的形状异常指标;根据所述衡量指标和所述形状异常指标通过k-均值聚类将所述每个用户的用电曲线划分为异常曲线和正常曲线,并获取所有异常曲线的所述衡量指标和所述形状异常指标的均值,以作为电量异常指标和形状异常指标;以及根据所述电量异常指标和所述形状异常指标获取用户的电量异常排名和形状异常排名,并获取所述电量异常排名和所述形状异常排名计算算术平均值,以得到综合异常排名。

[0009] 本发明实施例的综合熵方法和密度聚类方法的窃电检测方法,通过获取用户用电曲线集合和非技术性损失数据、电量相关性强弱的衡量指标和每条用电曲线的形状异常指标,并通过k-均值聚类划分用电曲线类别,得到用户的电量异常排名和形状异常排名,并根据其算术平均值得到综合异常排名,进而完成对窃电模式的检测,具有提高各种窃电模式下检测的准确率,实现大数据下窃电行为的非监督式检测,增大适用范围的优点。

[0010] 另外,根据本发明上述实施例的综合熵方法和密度聚类方法的窃电检测方法还可以具有以下附加的技术特征:

[0011] 进一步地,在本发明的一个实施例中,所述获取目标区域的待检测时间段内用户智能电表数据和总能耗数据,以得到用户用电曲线集合和非技术性损失数据,进一步包括:获得所述目标区域的待检测时间段内所有用户的用电曲线;通过总用电量序列减去各时刻所有用户用电量之和得到所述待检测时间段内目标区域由窃电因素造成的非技术性损失;将每个时刻的非技术性损失按日整理为对应的区域损失曲线。

[0012] 进一步地,在本发明的一个实施例中,所述根据所述用户用电曲线集合和非技术性损失数据通过最大信息系数计算方法得到每个用户的电曲线与非技术性损失的相关性,以得到用电量相关性强弱的衡量指标,进一步包括:对待检测的每一日获取所述区域损失曲线和所述每个用户的用电曲线之间的最大信息系数,以得到每一用户曲线与区域损失之间电量相关性的强弱指标;并且,所述通过基于快速搜索密度峰值的聚类方法获取标么用电曲线集合中每条曲线的邻域密度和到更高密度区域的距离,以得到所述每条曲线的形状异常指标,进一步包括:对所有的用电曲线分别进行标么化,且保留曲线的形状特点;获取标么用户曲线之间的欧式距离矩阵,并通过所述基于快速搜索密度峰值的聚类方法生成截断距离,并获取每条标么曲线的密度特征;根据所述密度特征得到所述每条曲线的形状异常指标。

[0013] 进一步地,在本发明的一个实施例中,所述非技术性损失数据的计算公式为:

$$[0014] \quad e_t = E_t - \sum_{i=1}^N x_{it},$$

[0015] 其中, $e_t$ 为非技术性损失数据, $E_t$ 为观察表所记录的用电量, $x_{it}$ 为智能电表所记录的用户*i*在*t*时段的用电量, $t=1,2,\dots,M \times T$ ;

[0016] 所述区域损失曲线的表达式为:

$$[0017] \quad e_j = [e_{(j-1) \cdot T+1} \quad e_{j \cdot T}],$$

[0018] 其中, $e_j$ 为区域损失曲线, $j=1,2,\dots,M$ 代表日下标。

[0019] 进一步地,在本发明的一个实施例中,所述标么化用户曲线集合为:

$$[0020] \quad \bar{c}_{ij} = \frac{c_{ij}}{\max\{x_{i,(j-1)T+1}, L, x_{i,jT}\}} = \frac{c_{ij}}{\max c_{ij}},$$

[0021] 其中,  $\bar{c}_{ij}$  为标么化用户曲线集合,  $c_{ij}$  为标么化用户曲线;

[0022] 所述每条曲线的形状异常指标为:

$$[0023] \quad \zeta_p = \frac{\delta_p}{\rho_p + 1},$$

[0024] 其中,  $\zeta_p$  为每条曲线的形状异常指标,  $\rho_p$  为标么化曲线的邻域密度,  $\delta_p$  到更高密度区域的距离。

[0025] 为达到上述目的, 本发明另一方面实施例提出了一种综合熵方法和密度聚类方法的窃电检测装置, 包括: 第一获取模块, 用于获取目标区域的待检测时间段内用户智能电表数据和总能耗数据, 以得到用户用电曲线集合和非技术性损失数据; 第二获取模块, 用于根据所述用户用电曲线集合和非技术性损失数据通过最大信息系数计算方法得到每个用户的电曲线与非技术性损失的相关性, 以得到用电量相关性强弱的衡量指标; 第三获取模块, 用于通过基于快速搜索密度峰值的聚类方法获取标么用电曲线集合中每条曲线的邻域密度和到更高密度区域的距离, 以得到所述每条曲线的形状异常指标; 划分模块, 用于根据所述衡量指标和所述形状异常指标通过k-均值聚类将所述每个用户的用电曲线划分为异常曲线和正常曲线, 并获取所有异常曲线的所述衡量指标和所述形状异常指标的均值, 以作为电量异常指标和形状异常指标; 以及计算模块, 用于根据所述电量异常指标和所述形状异常指标获取用户的电量异常排名和形状异常排名, 并获取所述电量异常排名和所述形状异常排名计算算术平均值, 以得到综合异常排名。

[0026] 本发明实施例的综合熵方法和密度聚类方法的窃电检测装置, 通过获取用户用电曲线集合和非技术性损失数据、电量相关性强弱的衡量指标和每条用电曲线的形状异常指标, 并通过k-均值聚类划分用电曲线类别, 得到用户的电量异常排名和形状异常排名, 并根据其算术平均值得到综合异常排名, 进而完成对窃电模式的检测, 具有提高各种窃电模式下检测的准确率, 实现大数据下窃电行为的非监督式检测, 增大适用范围的优点。

[0027] 另外, 根据本发明上述实施例的综合熵方法和密度聚类方法的窃电检测装置还可以具有以下附加的技术特征:

[0028] 进一步地, 在本发明的一个实施例中, 所述第一获取模块, 还用于: 获得所述目标区域的待检测时间段内所有用户的用电曲线; 通过总用电量序列减去各时刻所有用户用电量之和得到所述待检测时间段内目标区域由窃电因素造成的非技术性损失; 将每个时刻的非技术性损失按日整理为对应的区域损失曲线。

[0029] 进一步地, 在本发明的一个实施例中, 所述第二获取模块还用于: 对待检测的每一日获取所述区域损失曲线和所述每个用户的用电曲线之间的最大信息系数, 以得到每一用户曲线与区域损失之间电量相关性的强弱指标; 所述第三获取模块还用于: 对所有的用电曲线分别进行标么化, 且保留曲线的形状特点; 获取标么用户曲线之间的欧式距离矩阵, 并通过所述基于快速搜索密度峰值的聚类方法生成截断距离, 并获取每条标么曲线的密度特征; 根据所述密度特征得到所述每条曲线的形状异常指标。

[0030] 进一步地, 在本发明的一个实施例中, 所述非技术性损失数据的计算公式为:

$$[0031] \quad e_t = E_t - \sum_{i=1}^M x_{it},$$

[0032] 其中,  $e_t$  为非技术性损失数据,  $E_t$  为观察表所记录的用电量,  $x_{it}$  为智能电表所记录的用户  $i$  在  $t$  时段的用电量,  $t=1, 2, \dots, M \times T$ ;

[0033] 所述区域损失曲线的表达式为:

$$[0034] \quad e_j = [e_{(j-1) \cdot T+1} \quad e_{j \cdot T}],$$

[0035] 其中,  $e_j$  为区域损失曲线,  $j=1, 2, \dots, M$  代表日下标。

[0036] 进一步地, 在本发明的一个实施例中, 所述标么化用户曲线集合为:

$$[0037] \quad \bar{c}_{ij} = \frac{c_{ij}}{\max\{x_{i,(j-1)T+1}, L, x_{i,jT}\}} = \frac{c_{ij}}{\max c_{ij}},$$

[0038] 其中,  $\bar{c}_{ij}$  为标么化用户曲线集合,  $c_{ij}$  为标么化用户曲线;

[0039] 所述每条曲线的形状异常指标为:

$$[0040] \quad \zeta_p = \frac{\delta_p}{\rho_p + 1},$$

[0041] 其中,  $\zeta_p$  为每条曲线的形状异常指标,  $\rho_p$  为标么化曲线的邻域密度,  $\delta_p$  到更高密度区域的距离。

[0042] 本发明附加的方面和优点将在下面的描述中部分给出, 部分将从下面的描述中变得明显, 或通过本发明的实践了解到。

## 附图说明

[0043] 本发明上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解, 其中:

[0044] 图1为根据本发明实施例的综合熵方法和密度聚类方法的窃电检测方法的流程图;

[0045] 图2为根据本发明一个实施例的综合熵方法和密度聚类方法的窃电检测方法的流程图;

[0046] 图3为根据本发明实施例的综合熵方法和密度聚类方法的窃电检测装置的结构示意图。

## 具体实施方式

[0047] 下面详细描述本发明的实施例, 所述实施例的示例在附图中示出, 其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的, 旨在用于解释本发明, 而不能理解为对本发明的限制。

[0048] 下面参照附图描述根据本发明实施例提出的综合熵方法和密度聚类方法的窃电检测方法, 首先将参照附图描述根据本发明实施例提出的综合熵方法和密度聚类方法的窃电检测方法。

[0049] 图1是本发明实施例的综合熵方法和密度聚类方法的窃电检测方法流程图, 如图1所示, 该综合熵方法和密度聚类方法的窃电检测方法包括以下步骤:

[0050] 在步骤S101中: 获取目标区域的待检测时间段内用户智能电表数据和总能耗数



据,以得到用户用电曲线集合和非技术性损失数据。

[0051] 具体而言,如图2所示,在本发明的实施例中,定义用户智能电表在一天中记录的用电读数序列为“用电曲线”;定义记录区域内总用电量的高安全度智能电表为“观察表”;定义N为区域内用户数,M为待检测的天数。

[0052] 可以理解的是,获得该区域待检测时间段内所有用户的用电曲线,以及观察表所记录的总用电量序列。使用总用电量序列减去各时刻所有用户用电量之和,计算待检测时间段内该区域由窃电等因素造成的电量非技术性损失。

[0053] 具体地,如图2所示,根据观察表和智能电表的读数计算区域的非技术性损失;设T为每天智能电表或观察表采样的次数,设t代表智能电表或观察表的一个记录周期,区域的非技术性损失 $e_t$ 非技术性损失数据的计算公式为:

$$[0054] \quad e_t = E_t - \sum_{i=1}^N x_{it},$$

[0055] 其中, $e_t$ 为非技术性损失数据, $E_t$ 为观察表所记录的用电量, $x_{it}$ 为智能电表所记录的用户i在t时段的用电量, $t=1,2,\dots,M \times T$ 。

[0056] 在步骤S102中:根据用户用电曲线集合和非技术性损失数据通过最大信息系数计算方法得到每个用户的电曲线与非技术性损失的相关性,以得到用电量相关性强弱的衡量指标。

[0057] 可以理解的是,如图2所示,根据户用电曲线集合和非技术性损失数据,并通过对待检测的每一日计算区域损失曲线和各用户用电曲线之间的最大信息系数,得到每一用户曲线与区域损失之间的用电量相关性强弱的衡量指标。

[0058] 在步骤S103中:通过基于快速搜索密度峰值的聚类方法获取标么用电曲线集合中每条曲线的邻域密度和到更高密度区域的距离,以得到每条曲线的形状异常指标。

[0059] 其中,在本发明的一个实施例中,对所有的用电曲线分别进行标么化,仅保留曲线的形状特点;计算标么用户曲线之间的欧式距离矩阵,随后使用基于快速搜索密度峰值的聚类方法生成截断距离,并计算每条标么曲线的密度特征;随后根据密度特征,计算每条曲线的形状异常指标。

[0060] 具体而言,如图2所示,对每个用户i, $i=1,2,\dots,N$ ,令用电曲线 $c_{ij} = [x_{i,(j-1)T+1} \sim x_{i,jT}]$ ,分别计算 $e_j$ 和 $c_{ij}$ 之间的最大信息系数 $MIC(e_j, c_{ij})$ ,具体过程如下所示:

[0061] 对所有的用户曲线进行标么化,即除以当日所有时段用电量的最大值,得到标么化的用户曲线集合,:

$$[0062] \quad \bar{c}_{ij} = \frac{c_{ij}}{\max\{x_{i,(j-1)T+1}, \dots, x_{i,jT}\}} = \frac{c_{ij}}{\max c_{ij}};$$

[0063] 对 $M \times N$ 条标么化的曲线 $\bar{c}$ ,计算其两两之间的欧式距离,并从小到大排列,得到 $MN(MN-1)$ 组距离值 $d_{1,2}, \dots, d_{(MN-1),MN}$ ;

[0064] 使用基于快速搜索密度峰值的聚类方法计算截断距离 $d_c$ 以及各标么化曲线的邻域密度 $\rho_p$ 和到更高密度区域的距离 $\delta_p$ , $p=1,2,\dots,M \times N$ ;

[0065] 根据邻域密度和到更高密度区域的距离,计算曲线p的形状异常程度 $\zeta_p$ :

$$[0066] \quad \zeta_p = \frac{\delta_p}{\rho_p + 1};$$

[0067] 将 $\zeta_p$ 与用户 $i$ 和天数 $j$ 相对应,得到用户 $i$ 第 $j$ 天曲线的形状异常指标 $Z_{i,j}$ 。

[0068] 在步骤S104中:根据衡量指标和形状异常指标通过 $k$ -均值聚类将每个用户的用电曲线划分为异常曲线和正常曲线,并获取所有异常曲线的衡量指标和形状异常指标的均值,以作为电量异常指标和形状异常指标。

[0069] 在步骤S105中:根据电量异常指标和形状异常指标获取用户的电量异常排名和形状异常排名,并获取电量异常排名和形状异常排名计算算术平均值,以得到综合异常排名。

[0070] 具体地,步骤S104和步骤S105结合,如图2所示,通过使用 $k$ -均值聚类,将每个用户各日的电量相关性指标分成异常和正常两类,计算异常指标的均值作为该用户的电量异常指标,并根据该指标给出每个用户电量异常度排名;使用 $k$ -均值聚类将每个用户的用电曲线划分为异常曲线和正常曲线,异常指标的均值作为该用户的形状异常指标,并根据该指标给出每个用户形状异常度排名;计算用户电量异常度排名和形状异常度排名的算术平均值,作为用户的综合异常排名,其具体过程如下所示:

[0071] 对用户 $i$ 的 $M$ 个MIC( $e_j, c_{ij}$ )值,使用 $k$ -均值聚类并取 $k=2$ ,将 $M$ 个值划分为两类,令取值较大类的均值为 $Q_i$ ,根据 $Q_i$ 的大小给出 $N$ 个用户的可疑度排名 $Rank_1$ , $Q_i$ 越大对应用户 $i$ 的 $Rank_1$ 越大;

[0072] 对用户 $i$ 的 $M$ 个 $Z_{i,j}$ 值,使用 $k$ -均值聚类并取 $k=2$ ,将 $M$ 个值划分为两类,令取值较大类的均值为 $S_i$ ,根据 $S_i$ 的大小给出 $N$ 个用户的可疑度排名 $Rank_2$ , $S_i$ 越大对应用户 $i$ 的 $Rank_2$ 越大;

[0073] 计算 $Rank_1$ 和 $Rank_2$ 的算术平均值 $Rank$ ,得到用户的综合异常排名:

$$[0074] \quad Rank = \frac{1}{2}(Rank_1 + Rank_2);$$

[0075] 进一步地,通过用户的综合异常排名进而完成对窃电模式的检测,可以提升检测的准确性和适用性。

[0076] 本发明实施例的综合熵方法和密度聚类方法的窃电检测方法,通过获取用户用电曲线集合和非技术性损失数据、电量相关性强弱的衡量指标和每条用电曲线的形状异常指标,并通过 $k$ -均值聚类划分用电曲线类别,得到用户的电量异常排名和形状异常排名,并根据其算术平均值得到综合异常排名,进而完成对窃电模式的检测,具有提高各种窃电模式下检测的准确率,实现大数据下窃电行为的非监督式检测,增大适用范围的优点。

[0077] 其次参照附图描述根据本发明实施例提出的综合熵方法和密度聚类方法的窃电检测装置。

[0078] 图3是本发明一个实施例的综合熵方法和密度聚类方法的窃电检测装置的结构示意图。

[0079] 如图3所示,该综合熵方法和密度聚类方法的窃电检测装置10包括:第一获取模块100、第二获取模块200、第三获取模块300、划分模块400和计算模块500。

[0080] 其中,第一获取模块100用于获取目标区域的待检测时间段内用户智能电表数据和总能耗数据,以得到用户用电曲线集合和非技术性损失数据。第二获取模块200用于根据用户用电曲线集合和非技术性损失数据通过最大信息系数计算方法得到每个用户的电曲线与非技术性损失的相关性,以得到用电量相关性强弱的衡量指标。第三获取模块300用于通过基于快速搜索密度峰值的聚类方法获取标么用电曲线集合中每条曲线的邻域密度和

到更高密度区域的距离,以得到每条曲线的形状异常指标。划分模块400用于根据衡量指标和形状异常指标通过k-均值聚类将每个用户的用电曲线划分为异常曲线和正常曲线,并获取所有异常曲线的衡量指标和形状异常指标的均值,以作为电量异常指标和形状异常指标。计算模块500用于根据电量异常指标和形状异常指标获取用户的电量异常排名和形状异常排名,并获取电量异常排名和形状异常排名计算算术平均值,以得到综合异常排名。本发明实施例的窃电检测装置10能够提高各种窃电模式下检测的准确率,实现了大数据下窃电行为的非监督式检测,增大了适用范围。

[0081] 进一步地,在本发明的一个实施例中,第一获取模块100还用于获得目标区域的待检测时间段内所有用户的用电曲线;通过总用电量序列减去各时刻所有用户用电量之和得到待检测时间段内目标区域由窃电因素造成的非技术性损失;将每个时刻的非技术性损失按日整理为对应的区域损失曲线。

[0082] 进一步地,在本发明的一个实施例中,第二获取模块200还用于对待检测的每一日获取区域损失曲线和每个用户的用电曲线之间的最大信息系数,以得到每一用户曲线与区域损失之间电量相关性的强弱指标。第三获取模块还用于对所有的用电曲线分别进行标么化,且保留曲线的形状特点,获取标么用户曲线之间的欧式距离矩阵,并通过基于快速搜索密度峰值的聚类方法生成截断距离,并获取每条标么曲线的密度特征;根据密度特征得到每条曲线的形状异常指标。

[0083] 进一步地,在本发明的一个实施例中,非技术性损失数据的计算公式为:

$$[0084] \quad e_t = E_t - \sum_{i=1}^N x_{it},$$

[0085] 其中, $e_t$ 为非技术性损失数据, $E_t$ 为观察表所记录的用电量, $x_{it}$ 为智能电表所记录的用户*i*在*t*时段的用电量, $t=1,2,\dots,M \times T$ ;

[0086] 区域损失曲线的表达式为:

$$[0087] \quad e_j = [e_{(j-1) \cdot T+1} \quad L \quad e_{j \cdot T}],$$

[0088] 其中, $e_j$ 为区域损失曲线, $j=1,2,\dots,M$ 代表日下标。

[0089] 进一步地,在本发明的一个实施例中,标么化用户曲线集合为:

$$[0090] \quad \bar{c}_{ij} = \frac{c_{ij}}{\max\{x_{i,(j-1)T+1}, L, x_{i,jT}\}} = \frac{c_{ij}}{\max c_{ij}},$$

[0091] 其中, $\bar{c}_{ij}$ 为标么化用户曲线集合, $c_{ij}$ 为标么化用户曲线;

[0092] 每条曲线的形状异常指标为:

$$[0093] \quad \zeta_p = \frac{\delta_p}{\rho_p + 1},$$

[0094] 其中, $\zeta_p$ 为每条曲线的形状异常指标, $\rho_p$ 为标么化曲线的邻域密度, $\delta_p$ 到更高密度区域的距离。

[0095] 需要说明的是,前述对综合熵方法和密度聚类方法的窃电检测方法实施例的解释说明也适用于该实施例的装置,此处不再赘述。

[0096] 本发明实施例的综合熵方法和密度聚类方法的窃电检测装置,通过获取用户用电曲线集合和非技术性损失数据、电量相关性强弱的衡量指标和每条用电曲线的形状异常指标,并通过k-均值聚类划分用电曲线类别,得到用户的电量异常排名和形状异常排名,并根

据其算术平均值得到综合异常排名,进而完成对窃电模式的检测,具有提高各种窃电模式下检测的准确率,实现大数据下窃电行为的非监督式检测,增大适用范围的优点。

[0097] 在本发明的描述中,需要理解的是,术语“中心”、“纵向”、“横向”、“长度”、“宽度”、“厚度”、“上”、“下”、“前”、“后”、“左”、“右”、“竖直”、“水平”、“顶”、“底”“内”、“外”、“顺时针”、“逆时针”、“轴向”、“径向”、“周向”等指示的方位或位置关系为基于附图所示的方位或位置关系,仅是为了便于描述本发明和简化描述,而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作,因此不能理解为对本发明的限制。

[0098] 此外,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本发明的描述中,“多个”的含义是至少两个,例如两个,三个等,除非另有明确具体的限定。

[0099] 在本发明中,除非另有明确的规定和限定,术语“安装”、“相连”、“连接”、“固定”等术语应做广义理解,例如,可以是固定连接,也可以是可拆卸连接,或成一体;可以是机械连接,也可以是电连接;可以是直接相连,也可以通过中间媒介间接相连,可以是两个元件内部的连通或两个元件的相互作用关系,除非另有明确的限定。对于本领域的普通技术人员而言,可以根据具体情况理解上述术语在本发明中的具体含义。

[0100] 在本发明中,除非另有明确的规定和限定,第一特征在第二特征“上”或“下”可以是第一和第二特征直接接触,或第一和第二特征通过中间媒介间接接触。而且,第一特征在第二特征“之上”、“上方”和“上面”可是第一特征在第二特征正上方或斜上方,或仅仅表示第一特征水平高度高于第二特征。第一特征在第二特征“之下”、“下方”和“下面”可以是第一特征在第二特征正下方或斜下方,或仅仅表示第一特征水平高度小于第二特征。

[0101] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不必针对的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0102] 尽管上面已经示出和描述了本发明的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本发明的限制,本领域的普通技术人员在本发明的范围内可以对上述实施例进行变化、修改、替换和变型。

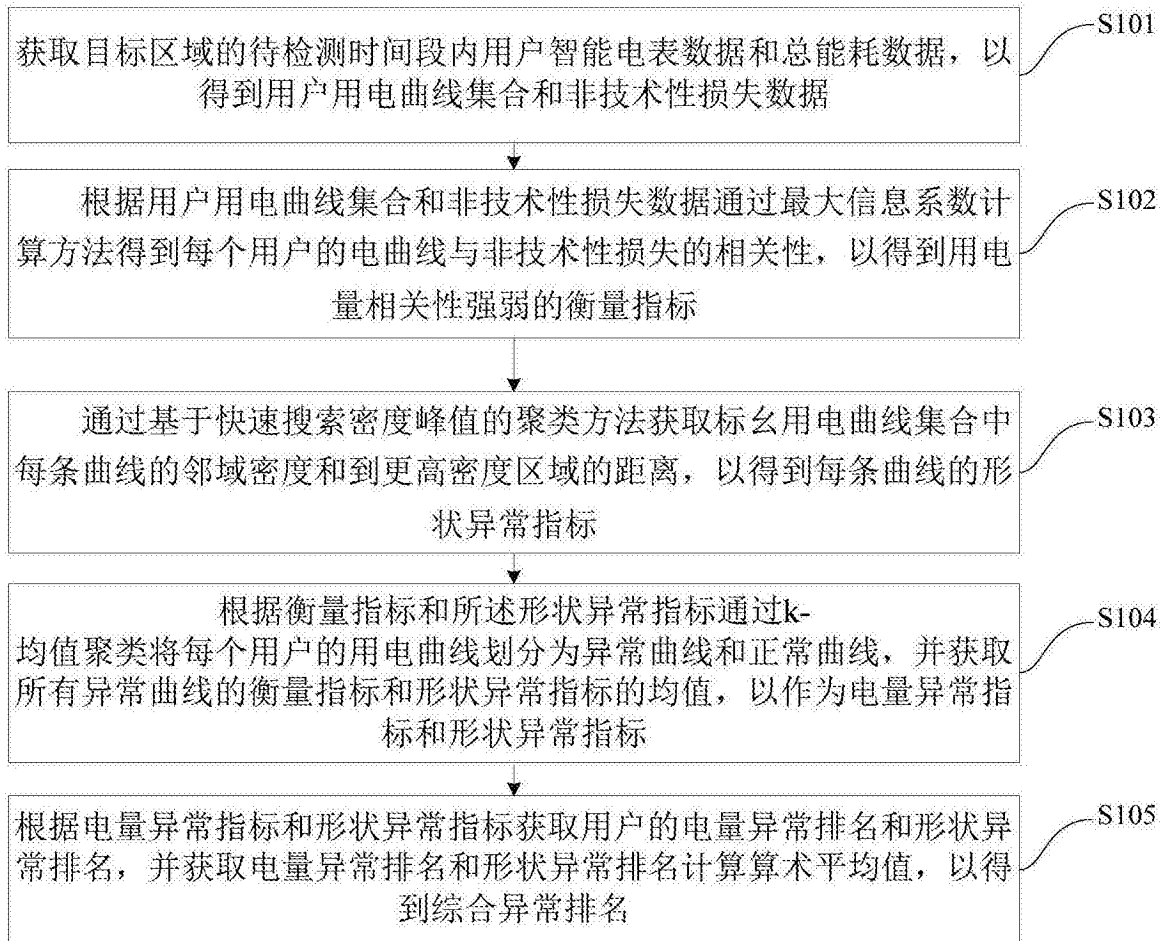


图1

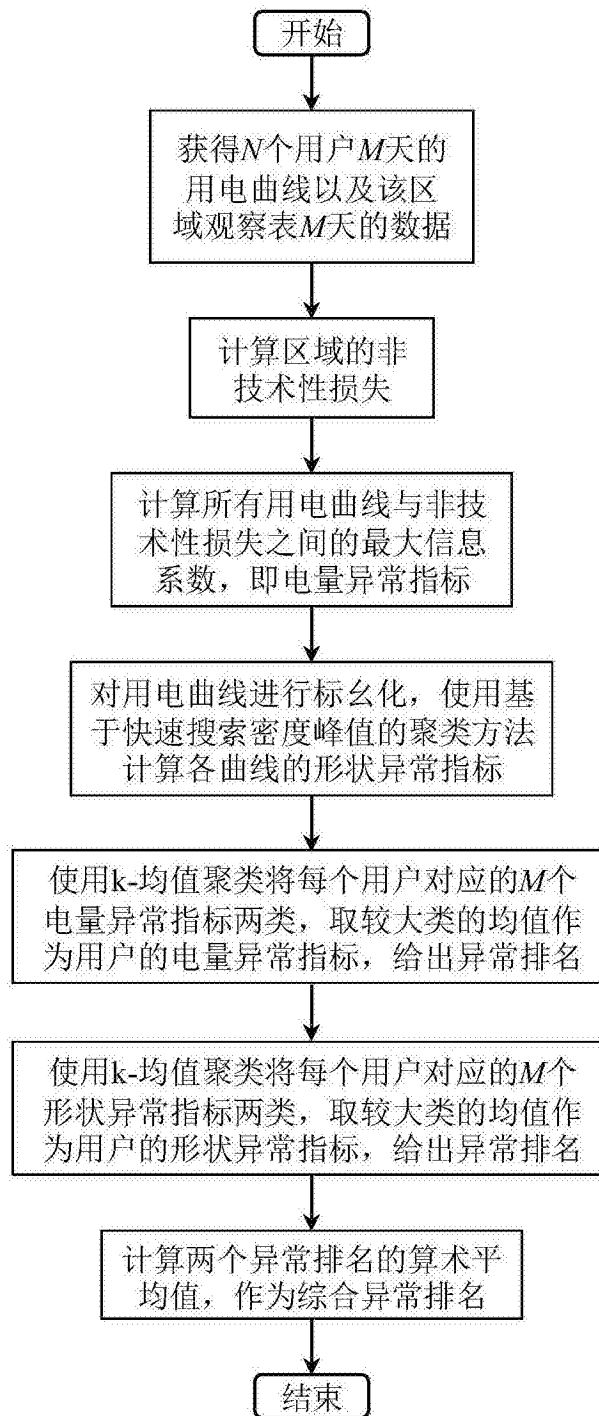


图2

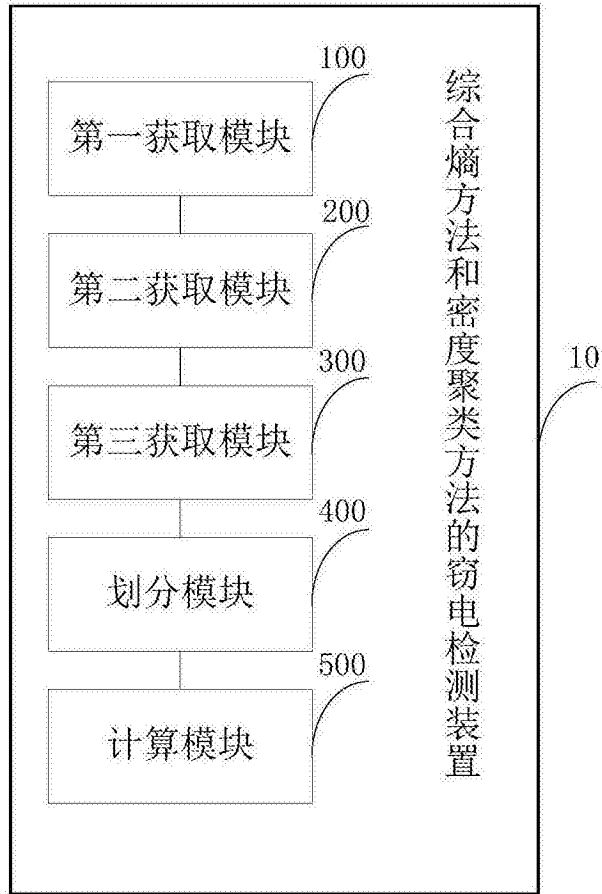


图3