(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 106682079 B
(45)授权公告日 2020.06.05

(21)申请号 201611039954.9

(22)申请日 2016.11.21

(65)同一申请的已公布的文献号
申请公布号 CN 106682079 A

(43)申请公布日 2017.05.17

(73)专利权人 云南电网有限责任公司电力科学研究院
地址 650217 云南省昆明市经济技术开发区云大西路105号

(72)发明人 王昕 李川 曹敏 李英娜 黄炜 赵艳峰 李翔 赵旭 蒋婷婷

(74)专利代理机构 北京弘权知识产权代理事务所(普通合伙) 11363
代理人 逯长明 许伟群

(51)Int.Cl.
G06F 16/2458(2019.01)
G06F 16/25(2019.01)
G06F 16/28(2019.01)
G06Q 50/06(2012.01)

(56)对比文件
CN 104850629 A,2015.08.19,
CN 105844294 A,2016.08.10,
CN 105512206 A,2016.04.20,
CN 105868266 A,2016.08.17,
CN 104318316 A,2015.01.28,
CN 106127602 A,2016.11.16,
CN 104573713 A,2015.04.29,
CN 104462802 A,2015.03.25,
CN 103810261 A,2014.05.21,
CN 104881735 A,2015.09.02,
US 9336484 B1,2016.05.10,

审查员 沈晓娟

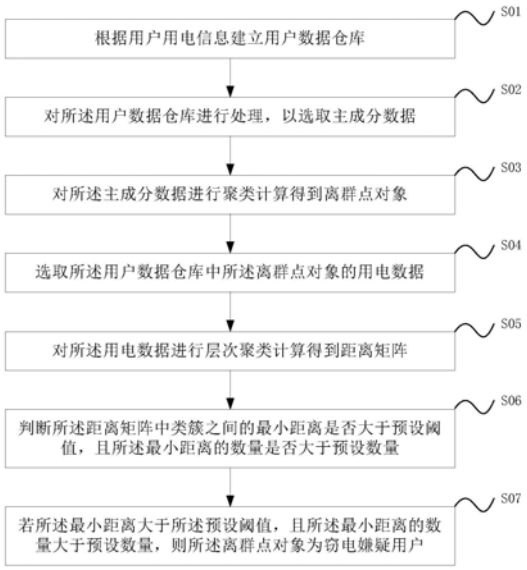
权利要求书3页 说明书7页 附图1页

(54)发明名称

一种基于聚类分析的用户用电行为检测方法

(57)摘要

本发明提供了一种基于聚类分析的用户用电行为检测方法,该检测方法通过建立用户数据仓库以收集用户用电信息,对所建立的用户数据仓库进行处理形成主成分数据,以处理大样本数据,提高挖掘效率;对主成分数据进行聚类计算得到不同的用户用电模式,其中,用户用电模式中的离群点对象可能为窃电用户,进而对离群点对象采用层次聚类方法计算得出窃电嫌疑用户。本发明提供的基于聚类分析的用户用电行为检测方法中,层次聚类方法只需对对象之间的距离排序,并根据距离序列进行聚类,不需要检查和估算对象;同时,离群点的检测与聚类计算能够高度互补,具有很好的伸缩性。



CN 106682079 B

1.一种基于聚类分析的用户用电行为检测方法,其特征在于,所述方法包括:

根据用户用电信息建立用户数据仓库;

对所述用户数据仓库进行处理,以选取主成分数据;

对所述主成分数据进行聚类计算得到离群点对象;

选取所述用户数据仓库中所述离群点对象的用电数据;

对所述用电数据进行层次聚类计算得到距离矩阵;

判断所述距离矩阵中类簇之间的最小距离是否大于预设阈值,且所述最小距离的数量是否大于预设数量;

若所述最小距离大于所述预设阈值,且所述最小距离的数量大于预设数量,则所述离群点对象为窃电嫌疑用户。

2.根据权利要求1所述的基于聚类分析的用户用电行为检测方法,其特征在于,根据用户用电信息建立用户数据仓库包括:

将用户用电信息添加到事实表,以建立逻辑模型,所述用户用电信息包括电流、电压、用户的类型、地理位置、年均用电量、月均用电量和/或日用电量;

将所述逻辑模型转化为数据模型;

使用ETL对所述数据模型中的用电数据进行清洗,形成用户数据仓库。

3.根据权利要求1所述的基于聚类分析的用户用电行为检测方法,其特征在于,对所述用户数据仓库进行处理包括对所述用户数据仓库中的数据进行切块或切片操作。

4.根据权利要求1所述的基于聚类分析的用户用电行为检测方法,其特征在于,对所述主成分数据进行聚类计算得到离群点对象包括通过k-means算法或FCM算法对所述主成分数据进行聚类计算得到离群点对象。

5.根据权利要求4所述的基于聚类分析的用户用电行为检测方法,其特征在于,通过k-means算法对所述主成分数据进行聚类计算得到离群点对象包括:

所述主成分数据构成数据集;

在所述数据集中选取k个对象为初始聚类中心,设定差异程度s;

计算所述数据集中其余数据与各初始聚类中心的欧式距离;

根据所述欧式距离及所述初始聚类中心将所述其余数据划分为k个类簇;

计算所述k个类簇的簇内变差E;

根据所述簇内变差E重新确定聚类中心;

根据重新确定的所述聚类中心重新划分类簇,直至连续两次划分的类簇相同,得到最终类簇及最终聚类中心;

计算所述最终类簇中的每个数据与所述最终聚类中心的距离dist(o,c0)及每个数据与所述最终聚类中心的平均距离l_{c_0};

根据所述距离dist(o,c0)和所述平均距离l_{c_0}计算$\frac{dist(o,c_0)}{l_{c_0}}$;

判断所述$\frac{dist(o,c_0)}{l_{c_0}}$与所述差异程度s的大小;

若所述$\frac{dist(o,c_0)}{l_{c_0}}$大于所述差异程度s，则为离群点对象；

若所述$\frac{dist(o,c_0)}{l_{c_0}}$小于或等于所述差异程度s，则为正常对象。

6.根据权利要求4所述的基于聚类分析的用户用电行为检测方法，其特征在于，通过FCM算法对所述主成分数据进行聚类计算得到离群点对象包括：

将所述主成分数据划分为c个类簇；

分别确立c个所述类簇的初始聚类中心；

设定模糊因子m，差异程度s；

计算所述主成分数据中的每个数据隶属于c个类簇的隶属度；

根据所述隶属度确定隶属度矩阵U；

根据$v_i=\frac{\sum_{j=1}^{n}u_{ij}^{m}x_j}{\sum_{j=1}^{n}u_{ij}^{m}}$计算聚类中心矩阵V；其中，在c个类簇中分别确立初始聚类中心，初始聚类中心为$V=\{v_1,v_2,\cdots,v_c\}$，v_i是c个类簇中第i个类簇；u_{ij}是计算数据集$x=\{x_1,x_2,\cdots,x_n\}$中的每个数据分别隶属于c个类簇的隶属度；m是模糊因子；x_j是数据集$x=\{x_1,x_2,\cdots,x_n\}$中的第j个数据；n是数据集$x=\{x_1,x_2,\cdots,x_n\}$中的数据个数，j是大于或等于1且小于或等于n的整数；

根据所述隶属度矩阵U和所述聚类中心矩阵V计算目标函数J(U,V)；

若连续两次计算的所述目标函数J(U,V)的差值大于或等于终止误差，则重新计算隶属度矩阵U，直至连续两次计算的所述目标函数J(U,V)的差值小于终止误差，得到最终隶属度矩阵U；

根据所述最终隶属度矩阵U对所述主成分数据划分最终类簇及最终聚类中心；

计算所述最终类簇中的每个数据与所述最终聚类中心的距离$dist(o,c_0)$及每个数据与所述最终聚类中心的平均距离l_{c_0}；

根据所述距离$dist(o,c_0)$和所述平均距离l_{c_0}计算$\frac{dist(o,c_0)}{l_{c_0}}$；

判断所述$\frac{dist(o,c_0)}{l_{c_0}}$与所述差异程度s的大小；

若所述$\frac{dist(o,c_0)}{l_{c_0}}$大于所述差异程度s，则为离群点对象；

若所述$\frac{dist(o,c_0)}{l_{c_0}}$小于或等于所述差异程度s，则为正常对象。

7.根据权利要求1所述的基于聚类分析的用户用电行为检测方法，其特征在于，对所述用电数据进行层次聚类计算得到距离矩阵包括：

将每个所述用电数据设定为一个初始类簇；

计算每两个初始类簇之间的欧式距离；

所述欧式距离形成距离矩阵。

一种基于聚类分析的用户用电行为检测方法

技术领域

[0001] 本发明涉及用电行为检测技术领域，尤其涉及一种基于聚类分析的用户用电行为检测方法。

背景技术

[0002] 窃电是指以不缴或少缴电费为目的而采用非法手段不计量或少计量用电量的行为。窃电行为不但严重损害供电企业的利益，而且还扰乱正常的供用电秩序，严重影响了电力事业的发展，并给用电安全带来了严重威胁。窃电行为一旦形成一定规模会大大增加供电线路的负担，轻则影响整个小区的电压稳定，重则会导致整片区域停电，影响用户使用电量。

[0003] 从近年来的各种窃电案件来看，不但窃电的手段越来越狡诈，窃电的数额也越来越大，同时，还有不少窃电者想方设法逃避应有的法律惩罚，因此，防治窃电问题已成为一个不容忽视的、迫切需要解决的重要课题，引起了各级供电企业和有关部门的重视。窃电行为一般在查抄电表或缴费时才被发现，然而此时已发生窃电行为，供电系统已经遭到破坏，此时只能对窃电者进行惩罚并改正其窃电行为，因此，窃电行为并不能提前知晓，也就不能提前制止窃电行为。

[0004] 目前，我国防窃电的方法主要包括采用专用计量箱或专用电表箱、装设防窃电设备等，然而上述方法需要大量更换计量箱或电表箱、装设新设备，大大浪费人力、物力。若只针对个别用户更换防窃电设备，则只能对安装有更换防窃电设备的用户进行控制，并不能对整片区域的所有用户的用电行为进行检测，因而仍然可能存在有漏检的窃电用户。

发明内容

[0005] 本发明提供一种基于聚类分析的用户用电行为检测方法，以通过大量的用户用电数据分析该片区域内用户的用电行为，进而确定窃电嫌疑用户。

[0006] 本发明提供了一种基于聚类分析的用户用电行为检测方法，所述方法包括：

[0007] 根据用户用电信息建立用户数据仓库；

[0008] 对所述用户数据仓库进行处理，以选取主成分数据；

[0009] 对所述主成分数据进行聚类计算得到离群点对象；

[0010] 选取所述用户数据仓库中所述离群点对象的用电数据；

[0011] 对所述用电数据进行层次聚类计算得到距离矩阵；

[0012] 判断所述距离矩阵中类簇之间的最小距离是否大于预设阈值，且所述最小距离的数量是否大于预设数量；

[0013] 若所述最小距离大于所述预设阈值，且所述最小距离的数量大于预设数量，则所述离群点对象为窃电嫌疑用户。

[0014] 优选地，根据用户用电信息建立用户数据仓库包括：

[0015] 将用户用电信息添加到事实表，以建立逻辑模型，所述用户用电信息包括电流、电

压、用户的类型、地理位置、年均用电量、月均用电量和/或日用电量；

[0016] 将所述逻辑模型转化为数据模型；

[0017] 使用ETL对所述数据模型中的用电数据进行清洗，形成用户数据仓库。

[0018] 优选地，对所述用户数据仓库进行处理包括对所述用户数据仓库中的数据进行切块或切片操作。

[0019] 优选地，对所述主成分数据进行聚类计算得到离群点对象包括通过k-means算法或FCM算法对所述主成分数据进行聚类计算得到离群点对象。

[0020] 优选地，通过k-means算法对所述主成分数据进行聚类计算得到离群点对象包括：

[0021] 所述主成分数据构成数据集；

[0022] 在所述数据集中选取k个对象为初始聚类中心，设定差异程度s；

[0023] 计算所述数据集中其余数据与各初始聚类中心的欧式距离；

[0024] 根据所述欧式距离及所述初始聚类中心将所述其余数据划分为k个类簇；

[0025] 计算所述k个类簇的簇内变差E；

[0026] 根据所述簇内变差E重新确定聚类中心；

[0027] 根据重新确定的所述聚类中心重新划分类簇，直至连续两次划分的类簇相同，得到最终类簇及最终聚类中心；

[0028] 计算所述最终类簇中的每个数据与所述最终聚类中心的距离dist(o,c_0)及每个数据与所述最终聚类中心的平均距离l_{c_0}；

[0029] 根据所述距离dist(o,c_0)和所述平均距离l_{c_0}计算$\frac{dist(o,c_0)}{l_{c_0}}$；

[0030] 判断所述$\frac{dist(o,c_0)}{l_{c_0}}$与所述差异程度s的大小；

[0031] 若所述$\frac{dist(o,c_0)}{l_{c_0}}$大于所述差异程度s，则为离群点对象；

[0032] 若所述$\frac{dist(o,c_0)}{l_{c_0}}$小于或等于所述差异程度s，则为正常对象。

[0033] 优选地，通过FCM算法对所述主成分数据进行聚类计算得到离群点对象包括：

[0034] 将所述主成分数据划分为c个类簇；

[0035] 分别确立c个所述类簇的初始聚类中心；

[0036] 设定模糊因子m，差异程度s；

[0037] 计算所述主成分数据中的每个数据隶属于c个类簇的隶属度；

[0038] 根据所述隶属度确定隶属度矩阵U；

[0039] 根据$v_i=\frac{\sum_{j=1}^{n}u_{ij}^{m}x_j}{\sum_{j=1}^{n}u_{ij}^{m}}$计算聚类中心矩阵V；

[0040] 根据所述隶属度矩阵U和所述聚类中心矩阵V计算目标函数J(U,V)；

[0041] 若连续两次计算的所述目标函数J(U,V)的差值大于或等于终止误差，则重新计算隶属度矩阵U，直至连续两次计算的所述目标函数J(U,V)的差值小于终止误差，得到最终隶属度矩阵U；

[0042] 根据所述最终隶属度矩阵U对所述主成分数据划分最终类簇及最终聚类中心；

[0043] 计算所述最终类簇中的每个数据与所述最终聚类中心的距离dist(o,c_0)及每个数据与所述最终聚类中心的平均距离l_{c_0}；

[0044] 根据所述距离dist(o,c_0)和所述平均距离l_{c_0}计算$\frac{dist(o,c_0)}{l_{c_0}}$；

[0045] 判断所述$\frac{dist(o,c_0)}{l_{c_0}}$与所述差异程度s的大小；

[0046] 若所述$\frac{dist(o,c_0)}{l_{c_0}}$大于所述差异程度s，则为离群点对象；

[0047] 若所述$\frac{dist(o,c_0)}{l_{c_0}}$小于或等于所述差异程度s，则为正常对象。

[0048] 优选地，对所述用电数据进行层次聚类计算得到距离矩阵包括：

[0049] 将每个所述用电数据设定为一个初始类簇；

[0050] 计算每两个初始类簇之间的欧式距离；

[0051] 所述欧式距离形成距离矩阵。

[0052] 本发明的实施例提供的技术方案可以包括以下有益效果：

[0053] 本发明提供一种基于聚类分析的用户用电行为检测方法，所述方法包括：根据用户用电信息建立用户数据仓库；对所述用户数据仓库进行处理，以选取主成分数据；对所述主成分数据进行聚类计算得到离群点对象；选取所述用户数据仓库中所述离群点对象的用电数据；对所述用电数据进行层次聚类计算得到距离矩阵；判断所述距离矩阵中类簇之间的最小距离是否大于预设阈值，且所述最小距离的数量是否大于预设数量；若所述最小距离大于所述预设阈值，且所述最小距离的数量大于预设数量，则所述离群点对象为窃电嫌疑用户。本发明提供的基于聚类分析的用户用电行为检测方法通过建立用户数据仓库以收集用户用电信息，对所建立的用户数据仓库进行处理形成主成分数据，以处理大样本数据，提高挖掘效率；对主成分数据进行聚类计算得到不同的用户用电模式，其中，用户用电模式中的离群点对象可能为窃电用户，进而对离群点对象采用层次聚类方法计算得出窃电嫌疑用户。本发明提供的基于聚类分析的用户用电行为检测方法中，层次聚类方法只需对对象之间的距离排序，并根据距离序列进行聚类，不需要检查和估算对象；同时，离群点的检测与聚类计算能够高度互补，具有很好的伸缩性。

[0054] 应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，并不能限制本发明。

附图说明

[0055] 此处的附图被并入说明书中并构成本说明书的一部分，示出了符合本发明的实施

例，并与说明书一起用于解释本发明的原理。

[0056] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，对于本领域普通技术人员而言，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0057] 图1为本发明实施例提供的基于聚类分析的用户用电行为检测方法的流程示意图。

具体实施方式

[0058] 这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明相一致的所有实施方式。相反，它们仅是与如所附权利要求书中所详述的、本发明的一些方面相一致的装置和方法的例子。

[0059] 请参考附图1，附图1示出了本发明实施例提供的基于聚类分析的用户用电行为检测方法的流程示意图，以下检测方法的描述均以附图1为基础。

[0060] 本发明实施例提供了一种基于聚类分析的用户用电行为检测方法，该检测方法包括：

[0061] S01：根据用户用电信息建立用户数据仓库；

[0062] 将电流、电压、用户的类型、地理位置、年均用电量、月均用电量和/或日用电量等的用户用电信息添加到事实表，形成用户数据仓库的逻辑模型，其中，该逻辑模型可以为星型模型。将形成的逻辑模型转化为数据模型，该数据模型在用户数据仓库中以表的形式存放。对转化后的数据模型进行优化，其中，该优化为使用ETL(Extract-Transform-Load)工具抽取数据模型中的电量数据、电压、电流等，ETL工具在抽取的过程中自动过滤数据，所抽取的数据形成用户数据仓库。

[0063] S02：对所述用户数据仓库进行处理，以选取主成分数据；

[0064] 由于用户数据仓库中的用户用电信息种类繁多，而在实际用电情况中，时间、用户、电压、电流、日用电量等数据最能反映用户的用电情况，因此需要对已建立的用户数据仓库中的时间、用户、电压、电流、日用电量等数据进行切块或切片操作，所选取的上述数据形成主成分数据。

[0065] S03：对所述主成分数据进行聚类计算得到离群点对象；

[0066] 在主成分数据中，行为不同于预期对象的点为离群点，该离群点为一个显著不同于其他数据对象的数据对象，且该离群点也不同于噪声数据。在本发明实施例中，通过采用聚类计算主成分数据来判断一个用户是否为一个离群点对象。进一步，在本发明实施例中，所采用的聚类计算方法为k-means算法或FCM(Fuzzy C-means，模糊C均值)算法，其中，k-means算法或FCM算法的具体计算过程为：

[0067] k-means算法对主成分数据进行计算得到离群点对象的具体内容包括：

[0068] 选取主成分数据为数据对象o，所有数据对象o构成数据集D；

[0069] 在数据集D中随机选取k个对象为初始聚类中心，设定差异程度s；

[0070] 计算数据集D中其余数据与各初始聚类中心的欧式距离，欧氏距离的计算公式为：

$$d(x_i,x_j)=\sqrt{(v_{i1}-v_{j2})^2+\cdots+(v_{in}-v_{jn})^2}=\sqrt{\sum_{q=1}^{n}(v_{iq}-v_{jq})^2}$$ 其中，x_i、x_j为两个n维对象，$A=(a_1,a_2,\ldots\ldots,a_n)$为每个对象的n维属性集，$x_i(v_{i1},v_{i2},\ldots\ldots,v_{in})$，$x_j(v_{j1},v_{j2},\ldots\ldots,v_{jm})$，$v_i$为属性$a_i$的值；

[0071] 根据欧式距离及初始聚类中心将其余数据按照相似程度划分为k个类簇；

[0072] 计算k个类簇的簇内变差E，其中，簇内变差E的计算公式为：

[0073] $$E=\sum_{i=1}^{k}\sum_{X\in C_i}dist(X,C_i)^2$$ ；

[0074] 根据簇内变差E重新确定聚类中心；

[0075] 根据重新确定的聚类中心重新划分类簇，迭代计算，直至连续两次划分的类簇相同，得到最终类簇及最终聚类中心；

[0076] 计算最终类簇中的每个数据与最终聚类中心的距离$dist(o,c_0)$及每个数据与最终聚类中心的平均距离l_{c_0}；

[0077] 根据距离$dist(o,c_0)$和平均距离l_{c_0}计算$\frac{dist(o,c_0)}{l_{c_0}}$；

[0078] 判断$\frac{dist(o,c_0)}{l_{c_0}}$与差异程度s的大小；

[0079] 若$\frac{dist(o,c_0)}{l_{c_0}}$大于差异程度s，则为离群点对象；

[0080] 若$\frac{dist(o,c_0)}{l_{c_0}}$小于或等于差异程度s，则为正常对象。

[0081] FCM算法对主成分数据进行计算得到离群点对象的具体内容包括：

[0082] 将主成分数据设为数据集$x=\{x_1,x_2,\cdots,x_n\}$，并将该数据集划分为c个类簇；

[0083] 在c个类簇中分别确立初始聚类中心，该初始聚类中心为$V=\{v_1,v_2,\cdots,v_c\}$；

[0084] 设定模糊因子m，差异程度s；

[0085] 计算数据集$x=\{x_1,x_2,\cdots,x_n\}$中的每个数据分别隶属于c个类簇的隶属度u_{ij}，u_{ij}的计算公式为 $$u_{ij}=\frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}}$$，其中，d_{ij}为第i个样本与第j个聚类中心的距离，d_{ik}为第i个样本与第k个类簇中心的距离，约束条件为：$u_{ij}\in[0,1]$，$\sum_{k=1}^{c}u_{uk}=1$；

[0086] 根据隶属度确定隶属度矩阵U，该隶属度矩阵U为$U=\begin{bmatrix}u_{11}&\cdots&u_{1c}\\ \vdots&\ddots&\vdots\\ u_{n1}&\cdots&u_{nc}\end{bmatrix}$，$u_{ij}$表示第i个

样本属于第j类的隶属度；

[0087] 根据$v_i=\frac{\sum_{j=1}^{n}u_{ij}^{m}x_j}{\sum_{j=1}^{n}u_{ij}^{m}}$计算聚类中心矩阵V；

[0088] 根据隶属度矩阵U和聚类中心矩阵V计算目标函数J(U,V)，该目标函数J(U,V)的计算公式为：$J(U,V)=\sum_{i=1}^{n}\sum_{j=1}^{c}u_{ij}^{m}d_{ij}^{2}$；

[0089] 若连续两次计算的目标函数J(U,V)的差值大于或等于终止误差，则重新迭代计算隶属度矩阵U，直至连续两次计算的目标函数J(U,V)的差值小于终止误差，得到最终隶属度矩阵U；

[0090] 根据最终隶属度矩阵U，若样本x_i属于第j个类簇的隶属度最大，则样本属于第j个类，从而完成对主成分数据所属最终类簇及最终聚类中心的划分‘

[0091] 计算最终类簇中的每个数据与最终聚类中心的距离dist(o,c_0)及每个数据与最终聚类中心的平均距离l_{c_0}；

[0092] 根据距离dist(o,c_0)和平均距离l_{c_0}计算$\frac{dist(o,c_0)}{l_{c_0}}$；

[0093] 判断$\frac{dist(o,c_0)}{l_{c_0}}$与差异程度s的大小；

[0094] 若$\frac{dist(o,c_0)}{l_{c_0}}$大于差异程度s，则为离群点对象；

[0095] 若$\frac{dist(o,c_0)}{l_{c_0}}$小于或等于差异程度s，则为正常对象。

[0096] S04：选取所述用户数据仓库中所述离群点对象的用电数据；

[0097] 上述所选出的离群点对象即为用电异常用户。在实际用电情况中，窃电用户的窃电方式有多种，如欠电压法、欠电流法、三相不平衡率及线损率等，但以欠电压法和欠电流法为主，因此本发明实施例通过电压或电流两方面来分析离群点对象，下述用电对象的选取以电压数据为例，欠电流法的分析方法同欠电压法。在选取数据时，选取用户数据仓库中离群点对象的电压数据。

[0098] S05：对所述用电数据进行层次聚类计算得到距离矩阵；

[0099] 将每个电压数据设定为一个初始类簇；

[0100] 采用最小距离的方法计算每两个初始类簇之间的欧式距离，两个初始类簇之间的欧式距离表示某两天同一时段的电压数据的变化；

[0101] 将所计算得到的所有欧式距离保存成距离矩阵。

[0102] S06：判断所述距离矩阵中类簇之间的最小距离是否大于预设阈值，且所述最小距离的数量是否大于预设数量；

[0103] 将距离矩阵距离最小的两个类簇合并形成新的类簇，同时更新距离矩阵，迭代计

算类簇之间的距离及更新距离矩阵，在迭代计算及更新距离矩阵的过程中，类簇之间的最小距离逐渐增大，此时，判断距离矩阵中类簇之间的最小距离是否大于预设阈值，且最小距离的数量是否大于预设数量，其中，该预设阈值为电压值变化最大值，该预设数量标压的-10％—+7％。

[0104] S07：若所述最小距离大于所述预设阈值，且所述最小距离的数量大于预设数量，则所述离群点对象为窃电嫌疑用户。

[0105] 依据上述判定，若最小距离大于预设阈值，且最小距离的数量大于预设数量，则表示某几天的一个时段中的电压数据跟近一个月同一时段的电压数据有差异的天数较多，即可判定用户在某一天或几天可能发生了欠电压的情况，如果这种情况长期发生，则可把该用户列为欠电压法窃电的嫌疑用户。

[0106] 本发明实施例提供的基于聚类分析的用户用电行为检测方法通过建立用户数据仓库以收集用户用电信息，对所建立的用户数据仓库进行处理形成主成分数据，以处理大样本数据，提高挖掘效率；对主成分数据进行聚类计算得到不同的用户用电模式，其中，用户用电模式中的离群点对象可能为窃电用户，进而对离群点对象采用层次聚类方法计算得出窃电嫌疑用户。本发明提供的基于聚类分析的用户用电行为检测方法中，层次聚类方法只需对对象之间的距离排序，并根据距离序列进行聚类，不需要检查和估算对象；同时，离群点的检测与聚类计算能够高度互补，具有很好的伸缩性。

[0107] 本领域技术人员在考虑说明书及实践这里发明的公开后，将容易想到本发明的其它实施方案。本申请旨在涵盖本发明的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本发明的一般性原理并包括本发明未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的，本发明的真正范围和精神由下面的权利要求指出。

[0108] 应当理解的是，本发明并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。本发明的范围仅由所附的权利要求来限制。

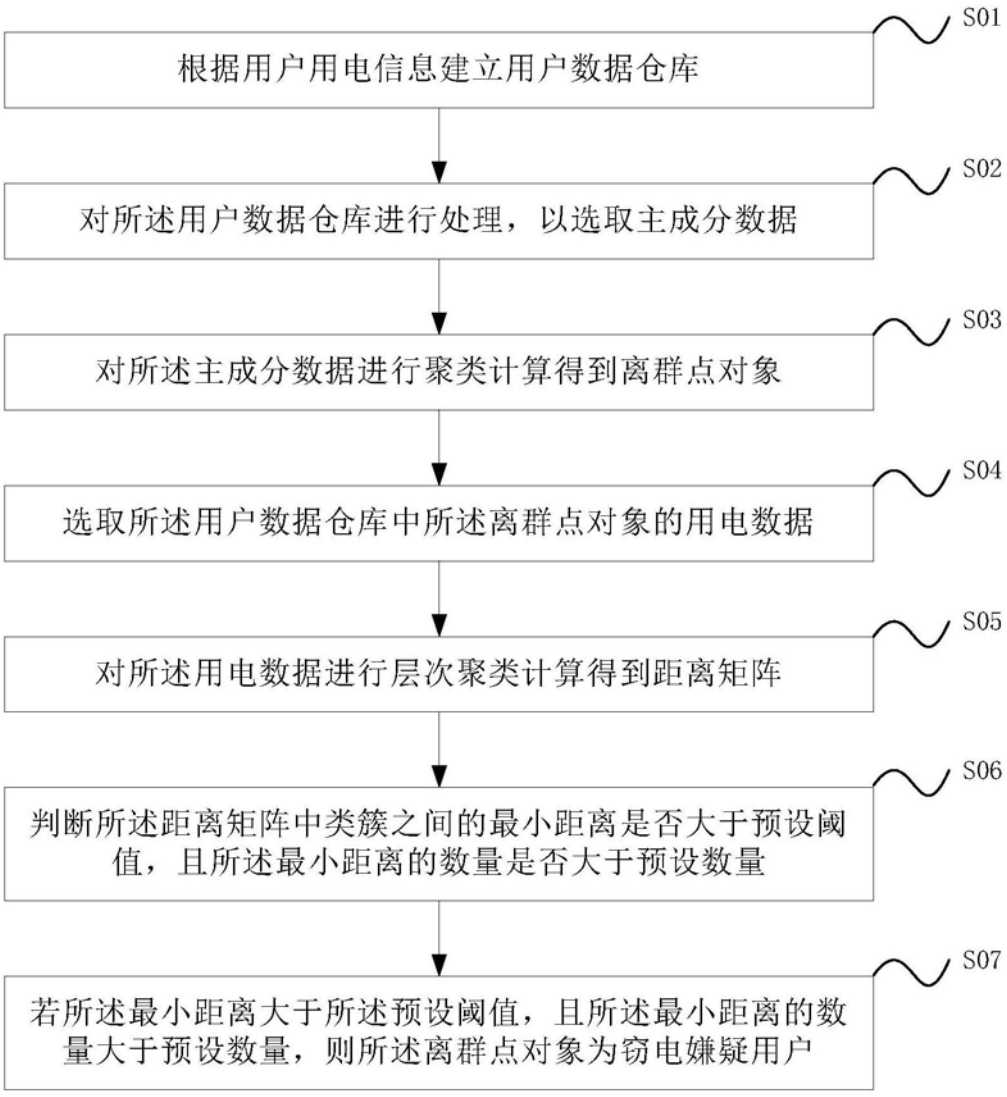

S01
根据用户用电信息建立用户数据仓库
S02
对所述用户数据仓库进行处理，以选取主成分数据
S03
对所述主成分数据进行聚类计算得到离群点对象
S04
选取所述用户数据仓库中所述离群点对象的用电数据
S05
对所述用电数据进行层次聚类计算得到距离矩阵
S06
判断所述距离矩阵中类簇之间的最小距离是否大于预设阈值，且所述最小距离的数量是否大于预设数量
S07
若所述最小距离大于所述预设阈值，且所述最小距离的数量大于预设数量，则所述离群点对象为窃电嫌疑用户

图1