



(12) 发明专利

(10) 授权公告号 CN 111667144 B

(45) 授权公告日 2023.04.28

(21) 申请号 202010362136.2

(22) 申请日 2020.04.30

(65) 同一申请的已公布的文献号
申请公布号 CN 111667144 A

(43) 申请公布日 2020.09.15

(73) 专利权人 北京中电普华信息技术有限公司
地址 100192 北京市海淀区清河小营东路
15号科研楼710室
专利权人 国网信息通信产业集团有限公司
国家电网有限公司

(72) 发明人 万泉 陈雁 闫富荣 张文 袁葆
欧阳红 戴永新 张浩 周春

(74) 专利代理机构 北京集佳知识产权代理有限
公司 11227
专利代理师 李慧引

(51) Int.Cl.

G06Q 10/0639 (2023.01)

G06Q 50/06 (2012.01)

G06F 18/23213 (2023.01)

(56) 对比文件

CN 105573997 A, 2016.05.11

CN 110824270 A, 2020.02.21

刘卫新;尹文庆;潘霞;杨金成;.聚类k-means算法在新疆反窃电工作中的应用.南昌大学学报(理科版).2017,(第05期),全文.

杨小铭;花永冬;黄淳驿;冯佳豪;何丽娜;王承民;.低压台区用户的窃电趋势嫌疑度分析方法研究.电器与能效管理技术.2016,(第10期),全文.

审查员 刘宇实

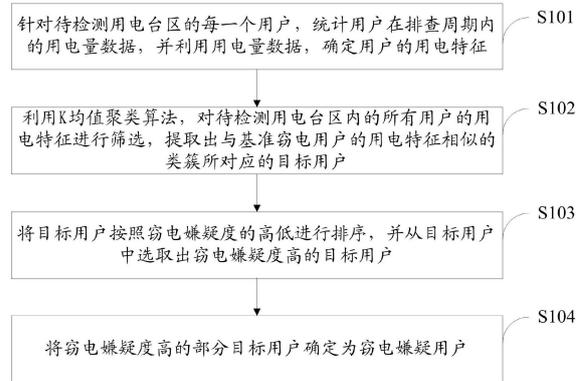
权利要求书2页 说明书10页 附图6页

(54) 发明名称

用户的识别方法及装置

(57) 摘要

本申请提供了一种用户的识别方法及装置,该方法包括:针对待排查用电台区的每一个用户,统计所述用户在排查周期内的用电量数据,并利用所述用电量数据,确定所述用户的用电特征;利用K均值聚类算法,对所述待排查用电台区内的所有用户的用电特征进行筛选,提取出与基准窃电用户的用电特征相似的目标用户;将所述目标用户按照窃电嫌疑度的高低进行排序,并从所述目标用户中选取窃电嫌疑度高的目标用户;其中,所述窃电嫌疑度高的目标用户的数量为预设数量;将所述窃电嫌疑度高的部分目标用户确定为窃电嫌疑用户。



1. 一种用户的识别方法,其特征在于,包括:

针对待排查用电台区的每一个用户,统计所述用户在排查周期内的用电量数据,并利用所述用电量数据,确定所述用户的用电特征,包括:统计在排查周期内,所述用户每天的用电量数据;基于所述用户每天的用电量数据,分别计算得到所述用户的用电量的均值、标准差及四分位数;将所述均值、标准差及四分位数作为所述用户的用电特征;

利用K均值聚类算法,对所述待排查用电台区内的所有用户的用电特征进行筛选,提取出与基准窃电用户的用电特征相似的类簇所对应的目标用户;

将所述目标用户按照窃电嫌疑度的高低进行排序,并从所述目标用户中选取窃电嫌疑度高的目标用户;其中,所述窃电嫌疑度高的目标用户的数量为预设数量;

一个用户的所述窃电嫌疑度的计算方式,包括:计算用户对应的标准差与均值的商,得到所述用户的窃电嫌疑度;

将所述窃电嫌疑度高的部分目标用户确定为窃电嫌疑用户。

2. 根据权利要求1所述的识别方法,其特征在于,所述利用K均值聚类算法,对所述待排查用电台区内的所有用户的用电特征进行筛选,提取与基准窃电用户的用电特征相似的类簇所对应的目标用户,包括:

利用轮廓系数确定多个聚类中心;其中,所述多个聚类中心的每一个聚类中心,聚合了所述待排查用电台区内的至少一个用户的用电特征;

将所述多个聚类中心中,符合预设条件的目标聚类中心确定为嫌疑类簇;其中,所述预设条件为所述目标聚类中心聚合的用户的用电特征满足:平均四分位数最小、且平均标准差最大;

从所述嫌疑类簇中,选取所述标准差大于所述均值的用户,确定为与基准窃电用户的用电特征相似的目标用户。

3. 根据权利要求1所述的识别方法,其特征在于,所述基准窃电用户的用电特征的确定方式,包括:

建立多个样本用户的用电量数据的频次直方图,其中,所述多个样本用户中包括有窃电用户和正常用户;

对每一个样本用户,分析所述样本用户的频次直方图,确定所述每一个样本用户的用电特征;

对所述窃电用户的用电特征与所述正常用户的用电特征进行比较,得到比较结果,并依据所述比较结果确定基准窃电用户的用电特征。

4. 一种用户的识别装置,其特征在于,包括:

统计单元,用于针对待排查用电台区的每一个用户,统计所述用户在排查周期内的用电量数据,并利用所述用电量数据,确定所述用户的用电特征;

所述统计单元,包括:

统计子单元,用于统计在排查周期内,所述用户每天的用电量数据;

第一计算子单元,用于基于所述用户每天的用电量数据,分别计算得到所述用户的用电量的均值、标准差及四分位数;

第一确定子单元,用于将所述均值、标准差及四分位数作为所述用户的用电特征;

筛选单元,用于利用K均值聚类算法,对所述待排查用电台区内的所有用户的用电特征

进行筛选,提取出与基准窃电用户的用电特征相似类簇所对应的目标用户;

排序单元,用于将所述目标用户按照窃电嫌疑度的高低进行排序,并从所述目标用户中选取窃电嫌疑度高的目标用户;其中,所述窃电嫌疑度高的目标用户的数量为预设数量;

第二计算子单元,用于计算用户对应的标准差与均值的商,得到所述用户的窃电嫌疑度;

确定单元,用于将所述窃电嫌疑度高的部分目标用户确定为窃电嫌疑用户。

5. 根据权利要求4所述的识别装置,其特征在于,所述筛选单元,包括:

第二确定子单元,用于利用轮廓系数确定多个聚类中心;其中,所述多个聚类中心的每一个聚类中心,聚合了所述待排查用电台区内的至少一个用户的用电特征;

第三确定子单元,用于将所述多个聚类中心中,符合预设条件的目标聚类中心确定为嫌疑类簇;其中,所述预设条件为所述目标聚类中心聚合的用户的用电特征满足:平均四分位数最小、且平均标准差最大;

第四确定子单元,用于从所述嫌疑类簇中,选取所述标准差大于所述均值的用户,确定为与基准窃电用户的用电特征相似的目标用户。

6. 根据权利要求5所述的识别装置,其特征在于,还包括:

建立单元,用于建立多个样本用户的用电量数据的频次直方图,其中,所述多个样本用户中包括有窃电用户和正常用户;

分析单元,用于对每一个样本用户,分析所述样本用户的频次直方图,确定所述每一个样本用户的用电特征;

比较单元,用于对所述窃电用户的用电特征与所述正常用户的用电特征进行比较,得到比较结果,并依据所述比较结果确定基准窃电用户的用电特征。

用户的识别方法及装置

技术领域

[0001] 本申请涉及数据处理领域,尤其涉及到一种用户的识别方法及装置。

背景技术

[0002] 随着电力需求增长,供电企业在为用户供电的同时,窃电事件越发严重;为保障供电企业的合法经营权益,确定用电用户中的窃电嫌疑用户尤为重要。

[0003] 然传统采用的确定窃电嫌疑用户的方式为:通过大量的工作人员进行拉网式排查,人工排查出窃电嫌疑用户。采用这这种排查方式,由于用电用户数量巨多,拉网式排查窃电嫌疑用户需要大量的工作人员和耗费大量的时间,导致排查成本高和排查大量的时间,导致排除成本高、排查效率低。

发明内容

[0004] 有鉴于此,本申请提供了一种用户的识别方法及装置,以向用户提供一种依据用户用电情况自动筛选出窃电嫌疑用户的服务。

[0005] 为实现上述目的,本申请实施例提供如下技术方案:

[0006] 本申请第一方面提供了一种用户的识别方法,包括:

[0007] 针对待排查用电台区的每一个用户,统计所述用户在排查周期内的用电量数据,并利用所述用电量数据,确定所述用户的用电特征;

[0008] 利用K均值聚类算法,对所述待排查用电台区内的所有用户的用电特征进行筛选,提取出与基准窃电用户的用电特征相似的一类簇所对应的目标用户;

[0009] 将所述目标用户按照窃电嫌疑度的高低进行排序,并从所述目标用户中选取窃电嫌疑度高的目标用户;其中,所述窃电嫌疑度高的目标用户的数量为预设数量;

[0010] 将所述窃电嫌疑度高的部分目标用户确定为窃电嫌疑用户。

[0011] 可选的,所述统计所述用户在排查周期内的用电量数据,确定所述用户的用电特征,包括:

[0012] 统计在排查周期内,所述用户每天的用电量数据;

[0013] 基于所述用户每天的用电量数据,分别计算得到所述用户的用电量的均值、标准差及四分位数;

[0014] 将所述均值、标准差及四分位数作为所述用户的用电特征。

[0015] 可选的,所述利用K均值聚类算法,对所述待排查用电台区内的所有用户的用电特征进行筛选,提取与基准窃电用户的用电特征相似的一类簇所对应的目标用户,包括:

[0016] 利用轮廓系数确定多个聚类中心;其中,所述多个聚类中心的每一个聚类中心,聚合了所述待排查用电台区内的至少一个用户的用电特征;

[0017] 将所述多个聚类中心中,符合预设条件的目标聚类中心确定为嫌疑类簇;其中,所述预设条件为所述目标聚类中心聚合的用户的用电特征满足:平均四分位数最小、且平均标准差最大;

- [0018] 从所述嫌疑类簇中,选取所述标准差大于所述均值的用户,确定为与基准窃电用户的用电特征相似的目标用户。
- [0019] 可选的,一个用户的所述窃电嫌疑度的计算方式,包括:
- [0020] 计算用户对应的标准差与均值的商,得到所述用户的窃电嫌疑度。
- [0021] 可选的,所述基准窃电用户的用电特征的确定方式,包括:
- [0022] 建立多个样本用户的用电量数据的频次直方图,其中,所述多个样本用户中包括有窃电用户和正常用户;
- [0023] 对每一个样本用户,分析所述样本用户的频次直方图,确定所述每一个样本用户的用电特征;
- [0024] 对所述窃电用户的用电特征与所述正常用户的用电特征进行比较,得到比较结果,并依据所述比较结果确定基准窃电用户的用电特征。
- [0025] 本申请第二方面提供了用户的识别装置,包括:
- [0026] 统计单元,用于针对待排查用电台区的每一个用户,统计所述用户在排查周期内的用电量数据,并利用所述用电量数据,确定所述用户的用电特征;
- [0027] 筛选单元,用于利用K均值聚类算法,对所述待排查用电台区内的所有用户的用电特征进行筛选,提取出与基准窃电用户的用电特征相似类簇所对应的目标用户;
- [0028] 排序单元,用于将所述目标用户按照窃电嫌疑度的高低进行排序,并从所述目标用户中选取窃电嫌疑度高的目标用户;其中,所述窃电嫌疑度高的目标用户的数量为预设数量;
- [0029] 确定单元,用于将所述窃电嫌疑度高的部分目标用户确定为窃电嫌疑用户。
- [0030] 可选的,所述统计单元,包括:
- [0031] 统计子单元,用于统计在排查周期内,所述用户每天的用电量数据;
- [0032] 第一计算子单元,用于基于所述用户每天的用电量数据,分别计算得到所述用户的用电量的均值、标准差及四分位数;
- [0033] 第一确定子单元,用于将所述均值、标准差及四分位数作为所述用户的用电特征。
- [0034] 可选的,所述筛选单元,包括:
- [0035] 第二确定子单元,用于利用轮廓系数确定多个聚类中心;其中,所述多个聚类中心的每一个聚类中心,聚合了所述待排查用电台区内的至少一个用户的用电特征;
- [0036] 第三确定子单元,用于将所述多个聚类中心中,符合预设条件的目标聚类中心确定为嫌疑类簇;其中,所述预设条件为所述目标聚类中心聚合的用户的用电特征满足:平均四分位数最小、且平均标准差最大;
- [0037] 第四确定子单元,用于从所述嫌疑类簇中,选取所述标准差大于所述均值的用户,确定为与基准窃电用户的用电特征相似的目标用户。
- [0038] 可选的,上述的任意一项用户的识别装置,还包括:
- [0039] 第二计算子单元,用于计算用户对应的标准差与均值的商,得到所述用户的窃电嫌疑度。
- [0040] 可选的,上述的任意一项用户的识别装置,还包括:
- [0041] 建立单元,用于建立多个样本用户的用电量数据的频次直方图,其中,所述多个样本用户中包括有窃电用户和正常用户;

[0042] 分析单元,用于对每一个样本用户,分析所述样本用户的频次直方图,确定所述每一个样本用户的用电特征;

[0043] 比较单元,用于对所述窃电用户的用电特征与所述正常用户的用电特征进行比较,得到比较结果,并依据所述比较结果确定基准窃电用户的用电特征。

[0044] 与现有技术相比,本申请具有以下优点:

[0045] 本申请提供的用户的识别方法中,针对待排查用电台区的每一个用户,统计所述用户在排查周期内的用电量数据,并利用所述用电量数据,确定所述用户的用电特征;利用K均值聚类算法,分别对所述待排查用电台区内的用户的用电特征进行筛选,提取与基准窃电用户的用电特征相似的类簇所对应的目标用户;将所述目标用户按照窃电嫌疑度的高低进行排序,并从所述目标用户中选取窃电嫌疑度高的目标用户;其中,所述窃电嫌疑度高的目标用户的数量为预设数量;将所述窃电嫌疑度高的部分目标用户确定为窃电嫌疑用户。由此可见,应用本申请提供的用户识别方法,能够依据待排查用电台区内的用户的用电量数据,自动筛选出窃电嫌疑用户,使得工作人员无需进行拉网式排查,节省了大量的人力物力,提高了排查窃电用户的效率。

附图说明

[0046] 为了更清楚地说明本申请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,下面描述中的附图仅仅是本申请的实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据提供的附图获得其他的附图。

[0047] 图1为本申请实施例提供的一种用户的识别方法的方法流程图;

[0048] 图2为本申请实施例提供的用户用电量数据的示意图;

[0049] 图3为本申请实施例提供的确定用户的用电特征的方法流程图;

[0050] 图4为本申请另一实施例提供的一种步骤S102的实现方法的流程图;

[0051] 图5为本申请实施例提供的一种基准窃电用户的用电特征的确定方法的流程图;

[0052] 图6-1为本申请实施例提供的一种窃电用户的频次直方图的示意图;

[0053] 图6-2为本申请实施例提供的一种正常用户的频次直方图的示意图;

[0054] 图6-3为本申请实施例提供的另一种正常用户的频次直方图的示意图;

[0055] 图6-4为本申请实施例提供的又一种正常用户的频次直方图的示意图;

[0056] 图7为本申请实施例提供的一种用户的识别装置的结构示意图。

具体实施方式

[0057] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,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他实施例,都属于本申请保护的范围。

[0058] 在本申请中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没

有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0059] 为提供一种排查窃电嫌疑用户的有效方式,本申请实施例提供了一种用户的识别方法,请参照图1所示,包括:

[0060] S101、针对待排查用电台区的每一个用户,统计用户在排查周期内的用电量数据,并利用用电量数据,确定用户的用电特征。

[0061] 本实施例中,用电台区指在电力系统中,(一台)变压器的供电范围或区域,待排查用电台区指代尚未排查窃电用户的用电台区。

[0062] 需要说明的是,与传统的人工拉网式排查方式不同的是,本实施例从后台中心中取得待排查用电台区的所有用户的用电量数据,并统计每一个用户的用电量数据,以确定出在排除周期内,每一个用户的用电特征。

[0063] 可选的,统计用户的用电量数据的方式,可以通过生成统计表的方式,可视化的展示出用户在排查周期内的每一天的用电量数据。例如可参照图2所示;图2为统计了示例用户1524376826,在2019年1月1日至2019年6月7日这个时间周期内,每一天的用电量数据。图中,X轴表示日期,Y轴表示用电量。由图2可看出,该用户的日用电量偏低,大部分天数的用电量趋于0。而在一月份末及二月份初时用电量剧增,呈现出峰值状态。根据历史的排查经验表明,出现这中现象的一种可能性是用户为防止用电检查,仅在排查时间正常用电,从而导致了日用电量普遍为0,且每个月电量少,某几天电量却较大等用电特征。

[0064] 当然,以上示例仅作为一种窃电嫌疑的依据,并不会以此直接判定该用户为窃电用户。在后续中,仍需要进一步进行筛选。

[0065] 可选的,本申请另一实施例中,统计用户的用电量数据,以确定用户的用电特征的方式,可参照图3所示,包括:

[0066] S301、统计在排查周期内,用户每天的用电量数据。

[0067] 本实施例中,排查周期可视情况而定,例如为一个月、一个季度或半年等。通过查询后台数据,可以得到每一个用户在排查周期内,每一天的用电情况。

[0068] S302、基于用户每天的用电量数据,分别计算得到用户的用电量的均值、标准差及四分位数。

[0069] 本申请实施例中,需要明确的是,通过预设的计算程序,可以对每一个用户的用电量数据,完成均值、标准差及四分位数的计算,而非人力为之。

[0070] 其中,均值又称平均数,是表示一组数据集中趋势的量数,是指在一组数据中所有数据之和再除以这组数据的个数。它是反映数据集中趋势的一项指标。例如求得用户的用电量均值为1,则表明用户平均日用电量为1度。

[0071] 标准差又称均方差,是离均差平方的算术平均数的平方根,在概率统计中最常使用作为统计分布程度上的测量,标准差是方差的算术平方根。其计算公式为:所有数减去其平均值的平方和,所得结果除以该组数之个数(或个数减一,即变异数),再把所得值开根号,所得之数就是这组数据的标准差。

[0072] 需要说明的是,标准差能反映一个数据集的离散程度。以两组用电量数据进行示例:集合A{0,0,0,28}和集合B{5,6,8,9}其平均值都是7,但集合B具有较小的标准差,反观集合A,用电量集中在某一天,其余时候则不产生用电,极有可能为不正常的用电情况。

[0073] 四分位数也称四分位点,指把统计的所有数值由小到大排列并分成四等份,处于三个分割点位置的数值。其中,四分位数分别包括:

[0074] 第一四分位数,又称“较小四分位数”,等于该样本中所有数值由小到大排列后,第25%的数字。以总样本n为例,第一四分位数Q1的位置通过公式: $Q1 = (n+1) * 0.25$ 来确定。

[0075] 第二四分位数,又称“中位数”,等于该样本中所有数值由小到大排列后,第50%的数字。以总样本n为例,第二四分位数Q2的位置通过公式: $Q2 = (n+1) * 0.50$ 来确定。

[0076] 第三四分位数,又称“较大四分位数”,等于该样本中所有数值由小到大排列后,第75%的数字。以总样本n为例,第三四分位数Q3的位置通过公式: $Q3 = (n+1) * 0.75$ 来确定。

[0077] 基于此,分别计算得到待排查用电台区内的每个用户的用电量数据所对应的均值、标准差和四分位数(四分位数包括第一四分位数、第二四分位数和第三四分位数)。

[0078] S303、将均值、标准差、及四分位数作为用户的用电特征。

[0079] 在求得了用户用电量数据所对应的均值、标准差及四分位数之后,理论上这些数值在一定程度上反映了该用户的用电情况。但由于本申请处理的是待排查用电台区内的所有用户的用电特征,并不是简单的针对单个用户进行操作,其所操作的对象可能是成千上万个用电用户,因此需要对所有用户的用电特征进行聚合和筛选,以得到窃电嫌疑用户的集合。

[0080] S102、利用K均值聚类算法,对待排查用电台区内的所有用户的用电特征进行筛选,提取出与基准窃电用户的用电特征相似的类簇所对应的目标用户。

[0081] 本实施例中,K均值聚类算法为一种划分聚类算法。在K均值聚类算法中,给定一个数据点集合和需要的聚类中心数k,k由用户指定,k均值算法根据某个距离函数反复把数据分入k个聚类中,从而在终止时得到各个聚类的聚合情况。

[0082] 需要说明的是,在本实施例中,首先会确定出聚类中心K,从而对在待排查用电台区中,所有用户的用电特征进行聚类。其中,均值、标准差、及四分位数不同的用电特征的用户,会自动聚合在不同的聚类中心,形成多个类簇,即多个用户的集合。

[0083] 换言之,通过K均值聚类算法,可将用电特征相似的用户聚合在一起,将用电特征不同的用户聚合在不同的聚类中心中。

[0084] 可选的,在本实例另一实施例中,步骤S102的实施方式可参照图4所示,包括:

[0085] S401、利用轮廓系数确定多个聚类中心。

[0086] 其中,所述多个聚类中心的每一个聚类中心,聚合了所述待排查用电台区内的至少一个用户的用电特征。

[0087] 本实施例利用轮廓系数确定聚类中心数K的值。其中,轮廓系数是聚类效果好与坏的一种评价方式,当假设了K的值后,可相应的计算得到轮廓系数的值i;若i的值代表的内聚度和分离度都相对较优,则相应的K的值也为优;反之,若i的值代表的内聚度和分离度都相对较差,则相应的K的值也为差。基于此,通过设定不同的K值得到相应的不同的轮廓系数i的值,最终根据最优的轮廓系数i确定出聚类中心数K的值,即待排查用电台区内每个用户的用电特征的多个聚类中心。

[0088] 可选的,本申请实施例中,聚类中心K统一设置为6个聚类中心。

[0089] S402、将多个聚类中心中,符合预设条件的目标聚类中心确定为嫌疑类簇。其中,预设条件为目标聚类中心聚合的用户的用电特征满足:平均四分位数最小、且平均标准差

最大。

[0090] 本申请实施例中,在利用多个聚类中心对用户进行聚合后,得到了多个不同的用户集合。以上述的6个聚类中心为例,那么便产生了6个用户集合。基于此,本申请需要从这多个聚类中心中,选出符合预设条件的那一个聚类中心,并将该聚类中心所包含的用户作为目标用户。其中目标用户指代初步具有窃电嫌疑的用户。

[0091] 需要特别说明的是,预设条件作为筛选初步具有窃电嫌疑的用户的方式,它包括的筛选规则有:

[0092] 平均四分位数最小。本实施例中,四分位数最小指代单个用户的第一四分位数+第二四分位数+第三四分位数的和为最小;平均四分位数最小指代目标聚类中心内,其所聚合的所有用户的平均四分位数最小。

[0093] 可选的,由于第三四分位数为所有数值由小到大排列后,第75%的数字,该位置上的数值在历史数据总结出来的经验中,在各个用户的用电情况上都基本相似,因此可以去除第三四分位数,采用第一四分位数+第二四分位数作为计算单个用户的四分位数的大小情况。

[0094] 需要说明的是,四分位数大小在一定程度上反映了用户的用电情况,尤其当某个用户的四分位数多为零时,则表征该用户在所有天数的用电量由小到大排列后,处于25%、50%及75%的位置上的用电量都为零,该用户基本不产生用电行为,因此该用户很有可能是窃电用户,需要通过比较用户的标准差进一步确定。

[0095] 平均标准差最大。本实施例中,平均标准差指代目标聚类中心内,其所聚合的所有用户的平均标准差。前述中提及标准差时,明确了标准差越大,反映了该用户的最高用电量与最低用电量的差值越大,用电情况越不正常。下放到窃电用户的用电情景中,则表现为窃电用户在一个排查周期内都不产生用电行为,而在排查的几天时间里则会产生较大的用电行为。如窃电用户某月的每天的用电量可能为: {0,0,0,0,0,0,0,0,0,0,0,0,0,0,20,20,0,0,0,0,0,0,0,0,0,0,0,0,0};由此可见,窃电用户不仅标准差较大,且四分位数也较小,且多为零。

[0096] 本实施例中,预设条件建立在分析已有的窃电用户的用电行为的基础上,因此在初步筛选窃电嫌疑用户上,具有很强的代表性。

[0097] 在根据预设条件筛选出目标聚类中心,确定为嫌疑类簇后,则可执行步骤S403、从嫌疑类簇中,选取标准差大于均值的用户,确定为与基准窃电用户的用电特征相似的目标用户。

[0098] 需要说明的是,若在嫌疑类簇中,用户的标准差大于均值,表示该用户用电量数据的分布基本不属于正态分布,且用电行为极不稳定,用电数据稀疏,该用户很大可能性属于窃电用户。而目标用户可理解为初步确定的窃电嫌疑用户,且目标用户可以多个。

[0099] 步骤S102中提出了基准窃电用户的用电特征的概念,而在本申请另一实施例中,基准窃电用户的用电特征的确定方式,可参照图5所示,包括:

[0100] S501、建立多个样本用户的用电量数据的频次直方图,其中,多个样本用户中包括有窃电用户和正常用户。

[0101] 本申请实施例中,采用频次直方图的方式来确定出窃电用户在用电情况上与正常用户的区别所在。由此,选择了包括有窃电用户与正常用户的用电量数据作为样本,分别建

立其对应的频次直方图。

[0102] 此处分别列举了一个窃电用户与多个正常用户的频次直方图,请参照图6-1至图6-4所示。图6-1为根据窃电用户的用电量数据所建立的频次直方图,图6-2至图6-4为根据正常用户的用电量数据所建立的频次直方图;其中,X轴为用电量,Y轴为天数,在后续の説明中,将引用图6-1至图6-4作为说明基础。

[0103] S502、对每一个样本用户,分析样本用户的频次直方图,确定所述每一个样本用户的用电特征。

[0104] 承接上述的图6-1至图6-4,从中我们可以看出:

[0105] 图6-1为窃电用户的频次直方图,在用电情况上,窃电用户表现的特征为:处在【0, 0.81】的用电区间的天数高达140天,且统计结果表面,该窃电用户在排查周期内,有89%的天数是不产生用电行为的;同时,处在【8.1, 8.91】的高用电区间的天数却有两天,合理推断为窃电用户为防止检查,在固定时间正常用电,甚至刻意产生高耗电行为以弥补用电量过少的问题。基于此,分析窃电用户的用电特征为:电表为0的天数居多,则其四分位数必然较小且多为0;低值特别稳定,高值波动、差距大,则相应的其用电特征中的标准差也较大。

[0106] 综述,窃电用户的用电特征体现为:四分位数较小且多为0、标准差比较大。

[0107] 下面对图6-2至图6-4表征的正常用户的用电特征进行分析。

[0108] 从图6-2反映的正常用户的用电行为中,可以发现,该用户在各个用电区间内都存在有一定天数,特别是在常见用电区间【1.63, 2.08】内,存在有49个天数。可得出正常用户的用电特征为:用电量分布趋近中部且较为均匀,部分用户的用电量分布还呈现正态分布趋势。

[0109] 从图6-3反映的正常用户的用电行为中,可以发现,该用户相较于图6-2反映的用户的用电情况而言,其用电分布曲线趋于左边,集中分布在【0.79, 1.34】和【1.34, 1.89】这两个用电区间内,分布时长均匀且多,并且,在排查周期内基本无0用电现象,可以确定该用电特征为正常用户中的少量用电的用户。

[0110] 从图6-4反映的正常用户的用电行为中,可以发现,该用户在整个排查周期内的用电均处在【0, 5】用电区间内。结合实际情况进行分析,可以确定该类型的用电特征为空置房居民用户。

[0111] S503、对窃电用户的用电特征与正常用户的用电特征进行比较,得到比较结果,并依据比较结果确定基准窃电用户的用电特征。

[0112] 本申请实施例中,所分析的样本用户包括但不限于上述提到的图6-1及图6-4所对应的用户,而是通过分析了大量的窃电用户与正常用户的频次直方图所展示出来的用电情况,最终总结出来的正常用户与窃电用户在用电特征上的区别,最终,确定出基准窃电用户的用电特征。

[0113] 综述,基准窃电用户的用电特征在后续中用于确定出与之相似的目标用户。在前述内容中,由于初步确定的目标用户可能在数量上比较多,不易于后续的排查工作,因此需要对目标用户进一步进行选取。

[0114] S103、将目标用户按照窃电嫌疑度的高低进行排序,并从目标用户中选取窃电嫌疑度高的目标用户。其中,窃电嫌疑度高的目标用户的数量为预设数量。

[0115] 本实施例中,可理解为对目标用户进一步进行筛选,从中选出窃电嫌疑度较高的

一部分数量的用户。其中,该数量为预留数量,在本实施例中,则采用该待排查用电台区的用户数的5%的数量作为预留数量。

[0116] 还需要说明的是,窃电嫌疑度用于从目标用户中筛选目标。可选的,窃电嫌疑度的计算方式为:计算用户对应的标准差与均值的商,以获得该用户的窃电嫌疑度。

[0117] S104、将窃电嫌疑度高的部分目标用户确定为窃电嫌疑用户。

[0118] 至此,以上一系列流程便自动完成了对待排查用电台区的所有用户,进行窃电嫌疑用户的识别和筛选,整个过程无需人工干预,只需引入待排查用电台区内的所有用户的用电量数据,便可从中筛选出少部分的窃电嫌疑度高的窃电嫌疑用户。在后续的排查过程中,可将窃电嫌疑用户名单交由工作人员进行逐一排查,进一步确定窃电嫌疑用户是否真正存在绕越窃电的行为。

[0119] 本申请实施例提供的用户的识别方法中,针对待排查用电台区的每一个用户,统计所述用户在排查周期内的用电量数据,并利用所述用电量数据,确定所述用户的用电特征;利用K均值聚类算法,分别对所述待排查用电台区内的用户的用电特征进行筛选,提取与基准窃电用户的用电特征相似的类簇所对应的目标用户;将所述目标用户按照窃电嫌疑度的高低进行排序,并从所述目标用户中选取窃电嫌疑度高的目标用户;其中,所述窃电嫌疑度高的目标用户的数量为预设数量;将所述窃电嫌疑度高的部分目标用户确定为窃电嫌疑用户。由此可见,应用本申请提供的用户识别方法,能够依据待排查用电台区内的用户的用电量数据,自动筛选出窃电嫌疑用户,使得工作人员无需进行拉网式排查,节省了大量的人力物力,提高了排查窃电用户的效率。

[0120] 本申请实施例还提供了一种用户的识别装置,请参照图6所示,包括:

[0121] 统计单元701,用于针对待排查用电台区的每一个用户,统计用户在排查周期内的用电量数据,并利用用电量数据,确定用户的用电特征。

[0122] 筛选单元702,用于利用K均值聚类算法,对待排查用电台区内的所有用户的用电特征进行筛选,提取出与基准窃电用户的用电特征相似的类簇所对应的目标用户。

[0123] 排序单元703,用于将目标用户按照窃电嫌疑度的高低进行排序,并从目标用户中选取窃电嫌疑度高的目标用户;其中,窃电嫌疑度高的目标用户的数量为预设数量。

[0124] 确定单元704,用于将窃电嫌疑度高的部分目标用户确定为窃电嫌疑用户。

[0125] 本申请实施例提供的用户的识别装置中,统计单元701针对待排查用电台区的每一个用户,统计用户在排查周期内的用电量数据,并利用用电量数据,确定用户的用电特征。筛选单元702利用K均值聚类算法,对待排查用电台区内的所有用户的用电特征进行筛选,提取出与基准窃电用户的用电特征相似的类簇所对应的目标用户。排序单元703将目标用户按照窃电嫌疑度的高低进行排序,并从目标用户中选取窃电嫌疑度高的目标用户;其中,窃电嫌疑度高的目标用户的数量为预设数量。确定单元704最终将窃电嫌疑度高的部分目标用户确定为窃电嫌疑用户。应用本实施例提供的用户的识别装置,能在对待排查用电台区进行排查之前,依据待排查用电台区内的用户的用电量数据,自动筛选出窃电嫌疑用户,从而无需对所有的用户进行拉网式排查,节省了大量人力物力,提高了排查效率。

[0126] 本实施例中,统计单元701、筛选单元702、排序单元703及确定单元704的具体执行过程,请参照对应图1的方法实施例的内容,此处便不再赘述。

[0127] 可选的,本申请另一实施例中,统计单元701,包括:

- [0128] 统计子单元,用于统计在排查周期内,用户每天的用电量数据。
- [0129] 第一计算子单元,用于基于用户每天的用电量数据,分别计算得到用户的用电量的均值、标准差及四分位数;
- [0130] 第一确定子单元,用于将均值、标准差及四分位数作为用户的用电特征。
- [0131] 本实施例中,统计子单元、第一计算子单元及第一确定子单元的具体执行过程,请参照对应图3的方法实施例的内容,此处便不再赘述。
- [0132] 可选的,本申请另一实施例中,筛选单元702,包括:
- [0133] 第二确定子单元,用于利用轮廓系数确定多个聚类中心;其中,多个聚类中心的每一个聚类中心,聚合了待排查用电台区内的至少一个用户的用电特征。
- [0134] 第三确定子单元,用于将多个聚类中心中,符合预设条件的目标聚类中心确定为嫌疑类簇;其中,预设条件为目标聚类中心聚合的用户的用电特征满足:平均四分位数最小、且平均标准差最大。
- [0135] 第四确定子单元,用于从嫌疑类簇中,选取标准差大于均值的用户,确定为与基准窃电用户的用电特征相似的目标用户。
- [0136] 本实施例中,第二确定子单元、第三确定子单元及第四确定子单元的具体执行过程,请参照对应图4的方法实施例的内容,此处便不再赘述。
- [0137] 可选的,本申请另一实施例中,用户的识别装置,还包括:
- [0138] 第二计算子单元,用于计算用户对应的标准差与均值的商,得到用户的窃电嫌疑度。
- [0139] 可选的,本申请另一实施例中,用户的识别装置,还包括:
- [0140] 建立单元,用于建立多个样本用户的用电量数据的频次直方图,其中,多个样本用户中包括有窃电用户和正常用户。
- [0141] 分析单元,用于对每一个样本用户,分析样本用户的频次直方图,确定每一个样本用户的用电特征。
- [0142] 比较单元,用于对窃电用户的用电特征与正常用户的用电特征进行比较,得到比较结果,并依据比较结果确定基准窃电用户的用电特征。
- [0143] 本实施例中,建立单元、分析单元及比较单元的具体执行过程,请参照对应图5的方法实施例的内容,此处便不再赘述。
- [0144] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统或系统实施例而言,由于其基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。以上所描述的系统及系统实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。
- [0145] 需要注意的是,本说明书中提及的“第一”、“第二”等概念仅用于对不同的装置、模块或单元进行区分,并非用于限定这些装置、模块或单元所执行的功能的顺序或者相互依存关系。

[0146] 专业人员还可以进一步意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、计算机软件或者二者的结合来实现,为了清楚地说明硬件和软件的可互换性,在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0147] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本申请。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本申请的精神或范围的情况下,在其它实施例中实现。因此,本申请将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

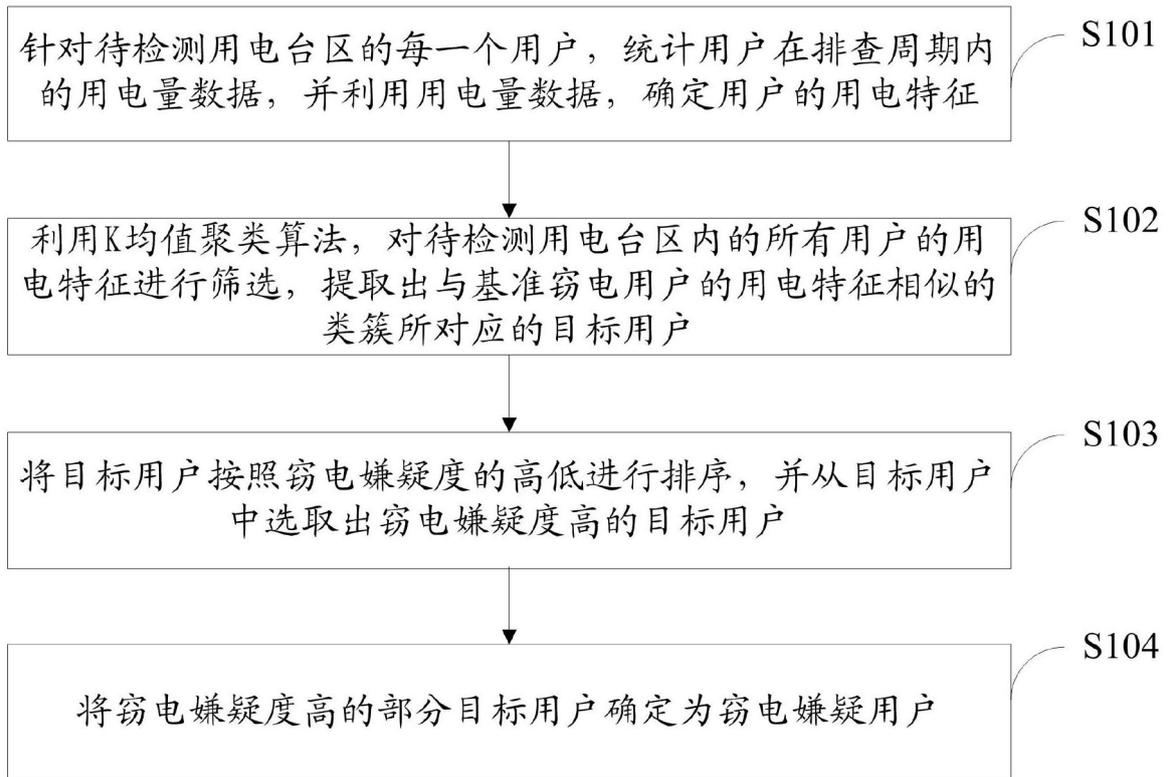


图1

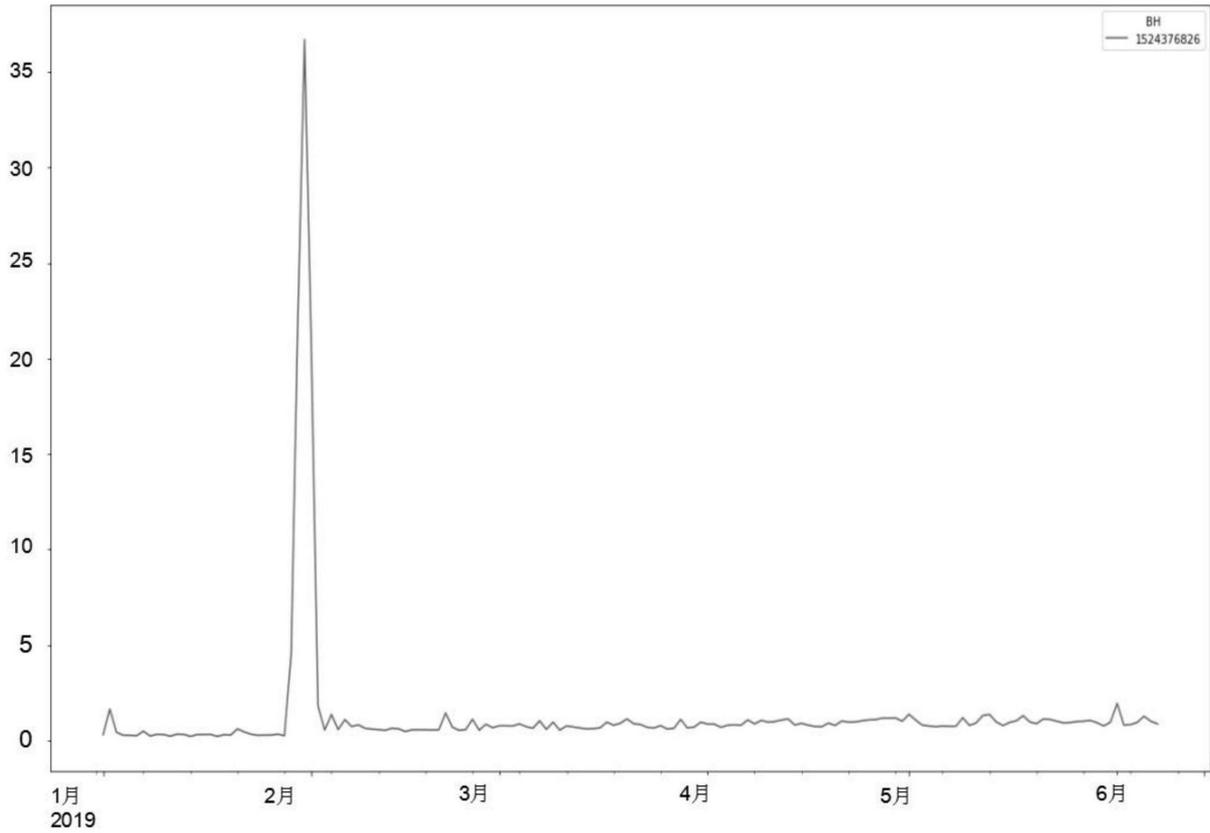


图2

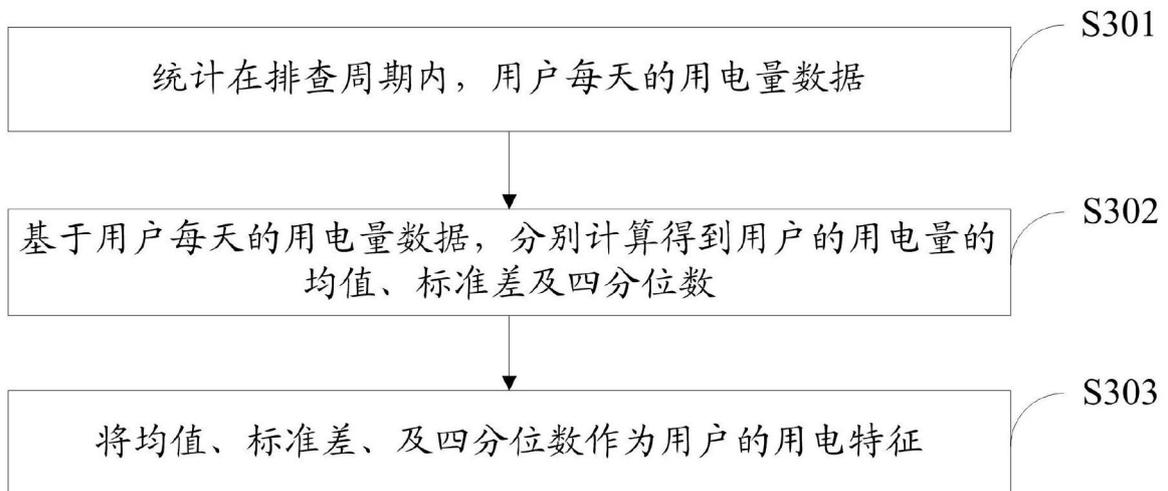


图3

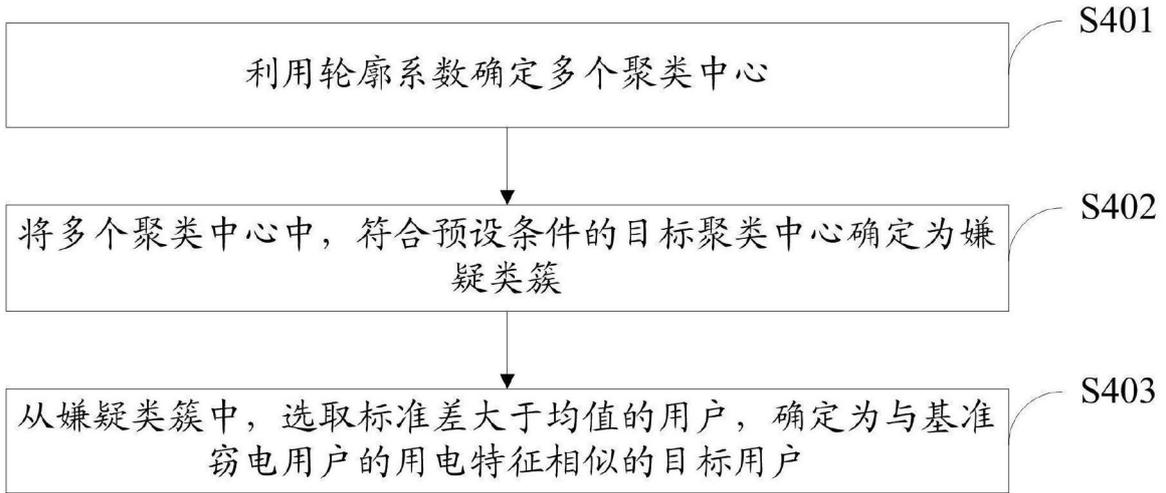


图4

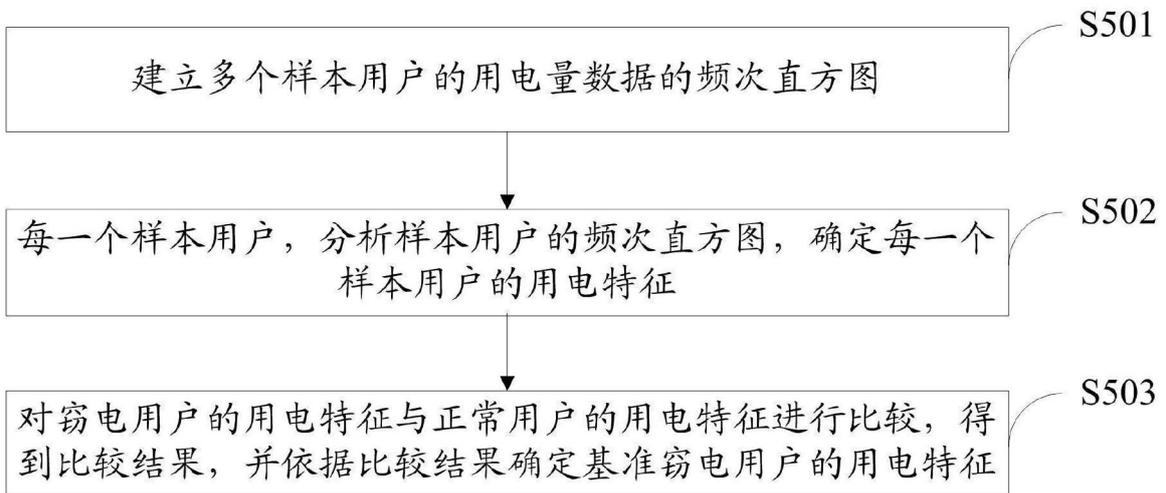


图5

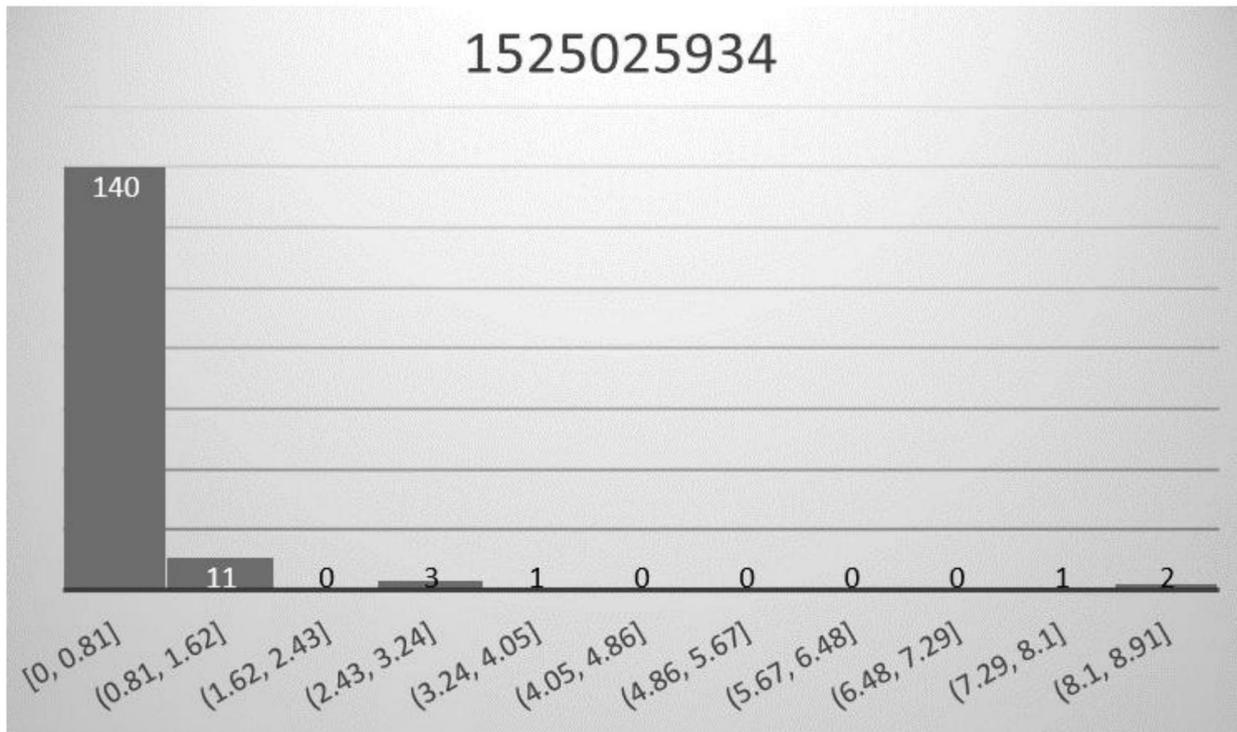


图6-1

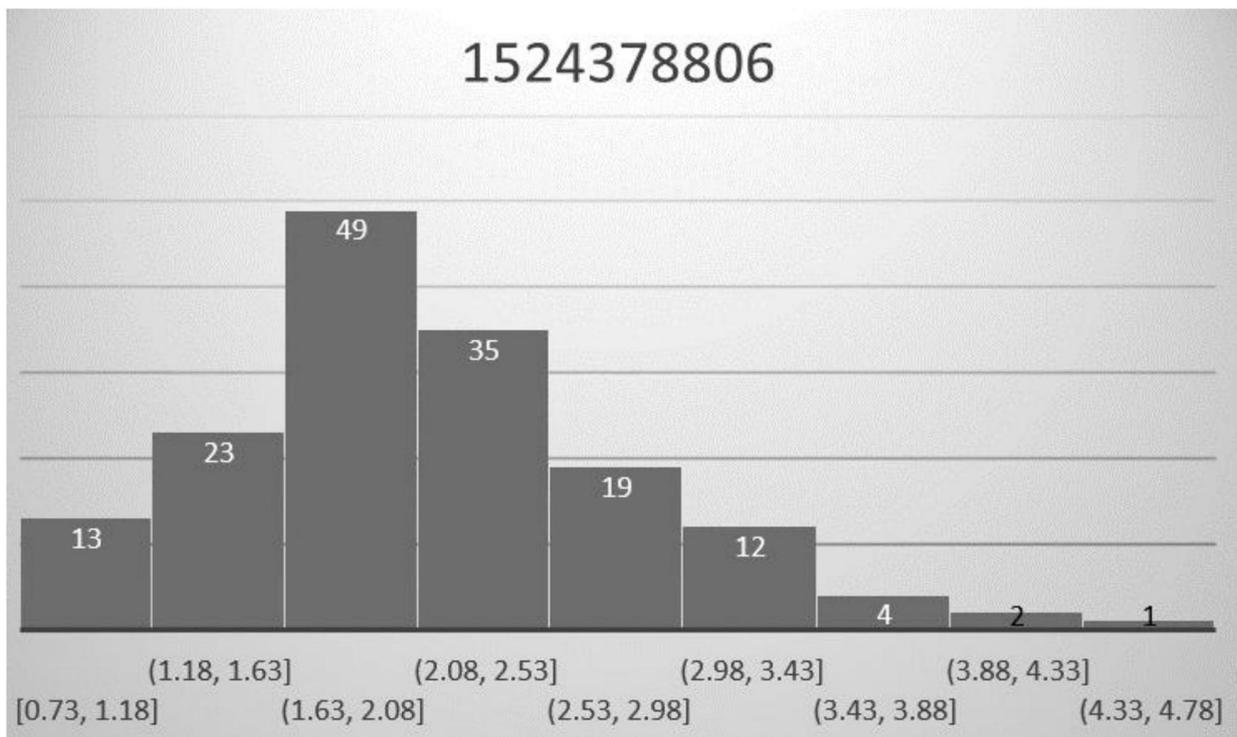


图6-2

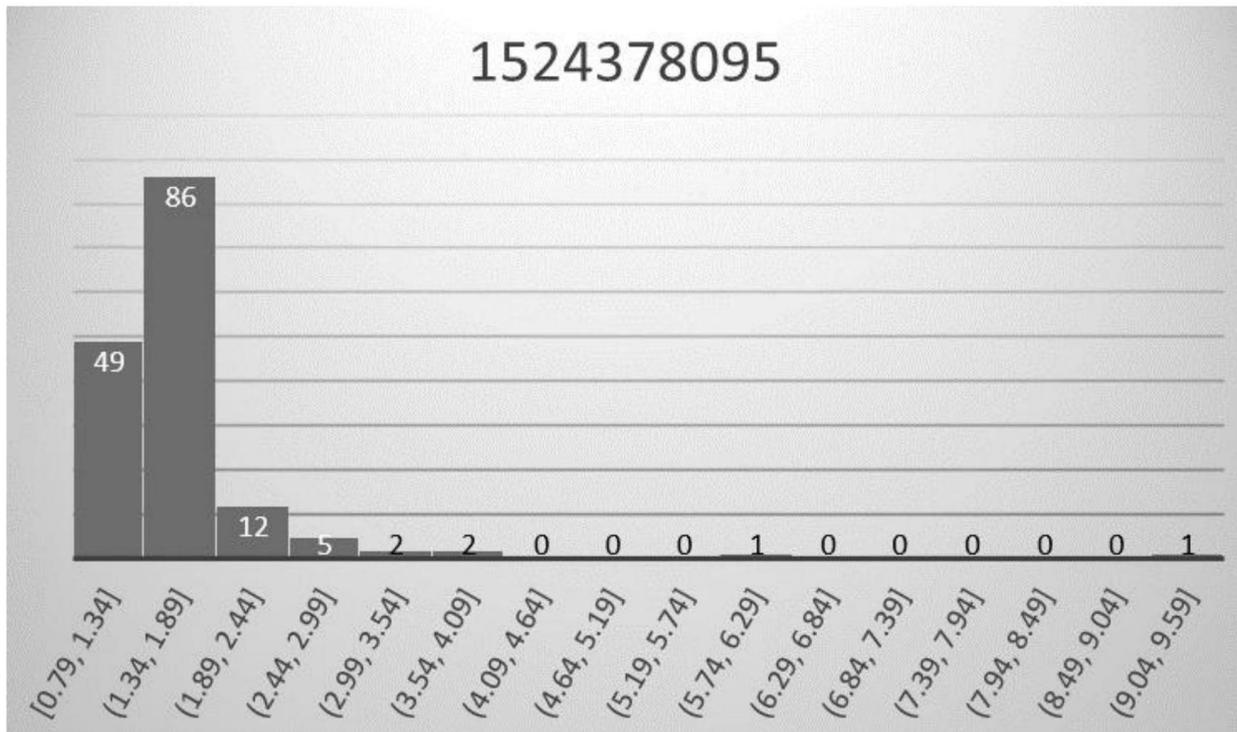


图6-3



图6-4

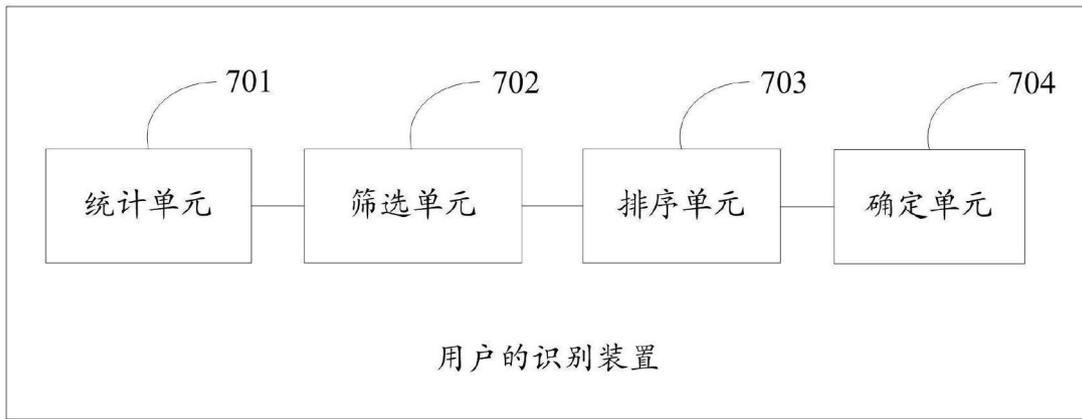


图7