



(12)发明专利申请

(10)申请公布号 CN 108734603 A

(43)申请公布日 2018.11.02

(21)申请号 201810506899.2

(22)申请日 2018.05.24

(71)申请人 河南工业大学

地址 450001 河南省郑州市高新技术产业
开发区莲花街100号河南工业大学科
技处

(72)发明人 陈红梅 唐宇飞 刘楠岿 张会娟

(51)Int.Cl.

G06Q 50/06(2012.01)

G06K 9/62(2006.01)

权利要求书1页 说明书3页 附图4页

(54)发明名称

基于聚类的大数据相位识别方法

(57)摘要

本发明涉及一种基于K-means聚类算法处理供电台区内用户相位信息,来完成供电台区内用户节点的分类。本发明主要步骤包括:采集电力系统运行的智能电表的电压数据,通过PCA处理电压数据,根据数据降维投射散点设定初始聚类中心,利用K-means算法对电压所属相位信息进行聚类分组处理,实现对供电台区内用户相位的分类识别。优点在于基于聚类的大数据相位识别方法直接利用K-means方法将同属相位的电压信息进行聚类,可以准确判断用户相位信息,对配电台区网络拓扑及三相不平衡治理提供支持,合理调整和规划接入每相用户的数量,来维持电力系统运行的稳定;不需要增加额外硬件开销,不需要设置比较阈值,弥补了目前的相位识别技术的弊端。

1. 一种基于聚类的大数据相位识别方法,由K-means聚类算法处理供电台区内用户相位信息,其特征在于:采集电力系统运行的智能电表的电压数据,通过PCA处理电压数据,根据数据降维投射散点设定初始聚类中心,利用K-means算法对电压所属相位信息进行聚类分组处理,实现对供电台区内用户相位的分类识别。

2. 根据权利要求1所述基于聚类的大数据相位识别方法,其技术特征在于:所述聚类算法处理供电台区内用户相位信息具体为:

(1) 对电压数据进行PCA降维处理;

(2) 根据智能电表收集到的用户台区相位数据,同一台区下用户相位为A、B和C三相,设定聚类组别数 $K=3$,根据PCA降维处理获得散点图,选定3个电压数据为相位的初始聚类中心;

(3) 调用K-means程序;

(4) 返回聚类分类结果:用户相位分类结果,聚类中心点坐标,平方误差函数值、迭代次数和各个聚类中点的个数;

(5) 根据Kmeans算法的分类结果将数据投影到2维平面,用不同颜色表示台区用户相位分类,完成相位识别。

3. 根据权利要求2所述基于聚类的大数据相位识别方法,其技术特征在于:调用K-means程序对电压所属相位信息进行聚类分组处理,具体步骤为:

(1) PCA降维处理用户台区电压数据;

(2) 输入选定的PCA散点图初始聚类中心,聚类组别数 $K=3$ 和设置最大迭代次数、平方误差阈值;

(3) 声明Kmeans数组和中间变量;

(4) 利用欧几里得距离公式计算聚类中心至各电压点的距离;

(5) 根据点到最近的聚类中心为一类更新分类结果,重新分配电压数据至相关相位组别;

(6) 根据每个聚类中各点坐标的平均值更新聚类中心坐标;

(7) 更新平方误差函数值;

(8) 往复执行3、4、5、6、7步,直至迭代次数达到最大迭代次数,或平方误差小于阈值,返回分类结果,聚类中心坐标,平方误差函数值,迭代次数,各个聚类中的数据个数。

4. 根据权利要求2或权利3所述基于聚类的大数据相位识别方法,其技术特征在于:通过两次PCA处理电压数据;首次PCA处理电压数据选定3个相位初始聚类中心,具体步骤为通过PCA降维处理电压数据,形成二维平面的用户电压数据散点图,根据散点图形状选取初次初始聚类中心,可有效减少聚类结果的误差;第二次PCA处理用户电压数据降至二维,以便把数据带入K-means聚类算法。

基于聚类的大数据相位识别方法

技术领域

[0001] 本发明涉及一种用户台区节点相位的分类领域,具体涉及一种K-means聚类大数据相位识别方法,属于配电网电参数测量应用领域。

背景技术

[0002] 在电力公司的配电网运营管理中,涉及到很多和用户电能表相位相关的问题,例如:变压器台区的三相功率不平衡调整,当管理人员发现电力系统存在三相功率严重不平衡需要调整用户的相位时,因为不知道用户的具体相位,也就不知道去选择哪些用户去进行调整,无法对新客户接入网络做出评估,使其安装至最合理的相位线路上。又例如在进行线损分析时,发现了某一相线损较大时,存在线路老化隐患和窃电行为,但是因为不知道用户的相位,不能缩小检查范围,工作量非常大。用户相位识别显得非常重要。目前用户相位识别有三种方法:可以通过电力线载波通信技术进行用户节点的安装相位识别;利用相位仪构建二次回路识别;工作人员现场依据接入用户线的颜色进行识别。目前的相位识别技术有很多的不足之处:虽然可以通过电力线载波通信技术可以进行用户节点的安装相位识别,但是对于没有配备电力载波的台区,无法进行相位识别;相位仪的仪器精度,数字化,自动化程度都有待提高,且需要搭建二次回路,投资和工作量大;在用户接入处理过程中缺少电力线颜色的统一要求,无法做到根据接入用户线的颜色准确识别相位。

[0003] 基于大数据分析的用户相位识别方法显得格外重要。目前我国已经基本实现了智能电能表的全覆盖工作,基于大数据分析的用户相位识别方法,不需要增加额外硬件开销,通过编写算法对采集到的用户相位电压信息进行处理、分类,准确得到对配电台区网络拓扑,能大大降低一线运检人员的工作强度和提高了电力公司的运营管理水平。

[0004] 虽然在大数据的相位识别方法领域可以通过线性相关法来完成用户相位的分类,但是此方法会因为阈值的选择影响分类结果;基于聚类的用户相位识别方法是一种有效的大数据相位识别方法,可以作为电力载波等相位识别等方法检验标准。

发明内容

[0005] 一种基于聚类的大数据相位识别方法,由K-means聚类算法处理供电台区内用户相位信息,其特征在于:采集电力系统运行的智能电表的电压数据,通过PCA处理电压数据,根据数据降维投射散点设定初始聚类中心,利用K-means算法对电压所属相位信息进行聚类分组处理,实现对供电台区内用户相位的分类识别。基于聚类的大数据相位识别方法直接利用K-means方法将同属相位的电压信息进行聚类,可以准确判断用户相位信息,对配电台区网络拓扑及三相不平衡治理提供支持,合理调整和规划接入每相用户的数量,来维持电力系统运行的稳定;不需要增加额外硬件开销,不需要设置比较阈值,弥补了目前的相位识别技术的弊端。

[0006] 本发明的有效效果在于:1. 基于聚类的大数据相位识别方法直接利用K-means方法将相位的电压信息进行聚类,可以准确判断用户相位信息,对配电台区网络拓扑及三相

不平衡治理提供支持,来维持电力系统运行的稳定;2.无需载波相位等专用工具识别,减少了工作量,不需要增加额外硬件开销,不需要设置比较阈值,通过智能电能表测量电压功能相位的精确识别,弥补了目前的相位识别技术的弊端;3实现了跨台区识别,可解决电力公司营销系统档案录入错误问题,解决因此带来的线损统计错误问题,合理调整和规划接入用户的数量,对配电台区网络拓扑及三相不平衡治理提供支持,提前对新客户接入网络做出评估。

附图说明

- [0007] 图1为本发明的系统主程序流程图。
[0008] 图2为K-mean K-means子程序流程图。
[0009] 图3 基于聚类的大数据相位识别的仿真实验验证。
[0010] 图4 基于聚类的大数据相位识别的实测实验验证。

具体实施方式

[0011] 下面结合具体实施方式对本发明作进一步详细说明。

[0012] 配电网中变压器台区都安装有总表,变压器台区总表的三相相位是确定的,可以作为参照;

通过智能电能表测量用户相位信息,系统收集用户相位电压数据信息;

系统数据导入算法,调用PCA首次处理电压数据,数据降维,将数据投射到二维平面(散点图表示);

根据散点图形状选取初始聚类中心,记录初始聚类中心坐标值;

同一台区下用户相位为A、B和C三相,设置聚类组别数 $K=3$,声明K-means聚类算法中间变量;

输入初始聚类中心坐标值,聚类中心数、迭代次数;

进行K-means程序处理;

返回聚类分类结果:用户相位分类结果,聚类中心点坐标,平方误差函数值、迭代次数和各个聚类中点的个数;

根据Kmeans算法的分类结果将数据投影到2维平面,用不同颜色表示台区用户相位分类,完成相位识别;

调用K-means程序对电压所属相位信息进行聚类分组处理,具体步骤为:

(1)PCA降维处理用户台区电压数据;

(2)输入选定的PCA散点图初始聚类中心,聚类组别数 $K=3$ 和设置最大迭代次数、平方误差阈值;

(3)声明Kmeans数组和中间变量;

(4)利用欧几里得距离公式计算聚类中心至各电压点的距离;

(5)根据点到最近的聚类中心为一类更新分类结果,重新分配电压数据至相关相位组别;

(6)根据每个聚类中各点坐标的平均值更新聚类中心坐标;

(7)更新平方误差函数值;

(8) 往复执行3、4、5、6、7步,直至迭代次数达到最大迭代次数,或平方误差小于阈值,返回分类结果,聚类中心坐标,平方误差函数值,迭代次数,各个聚类中的数据个数。

[0013] 本发明的系统主程序流程图如附图1所示。

[0014] K-means程序编程框图如附图2所示。

[0015] 本发明提出的方案通过如下仿真实验和实测数据加以验证:

仿真实验数据里包含了300用户(包含了A、B、C三相采集了300个用户数据),每个用户采集1000个数据,将数据带入程序得到以下结果:

其中,附图3所示中三种颜色代表不同的簇,红点代表最终聚类中心。输出结果自动保存在数据文件夹下,显示每个簇内的聚类组别数和最终聚类中心,以及给每个数据加上所属类别的标签,将每个簇内的数据保存为CSV形式,得到关于用户相位聚类结论与仿真设计一致。

[0016] 实测数据包含了25用户(包含了A、B、C三相采集了25个用户数据),每个用户采集500个数据,将数据带入程序得结果如附图4所示,用户相位聚类结论与实测实验结论一致。

[0017] 本发明说明书中未作详细描述的内容属于本专业领域技术人员公知的现有技术。

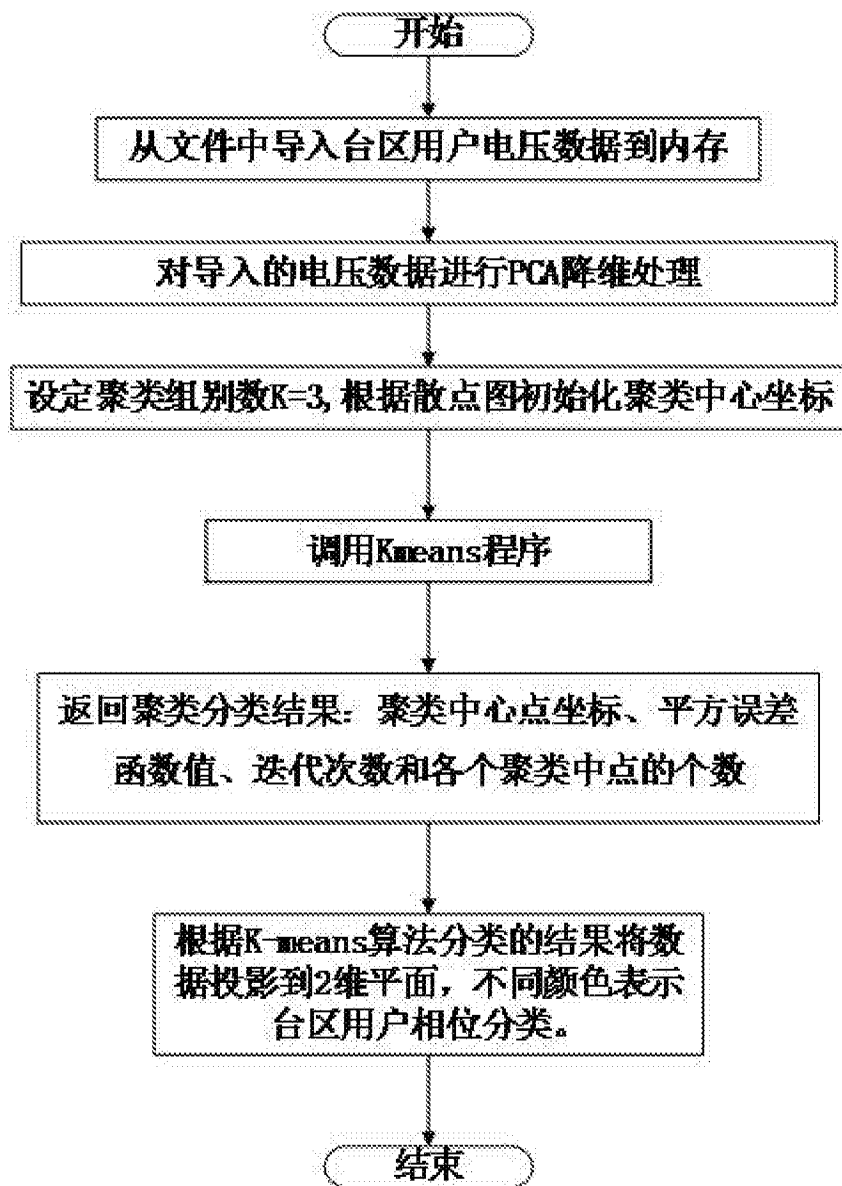


图1

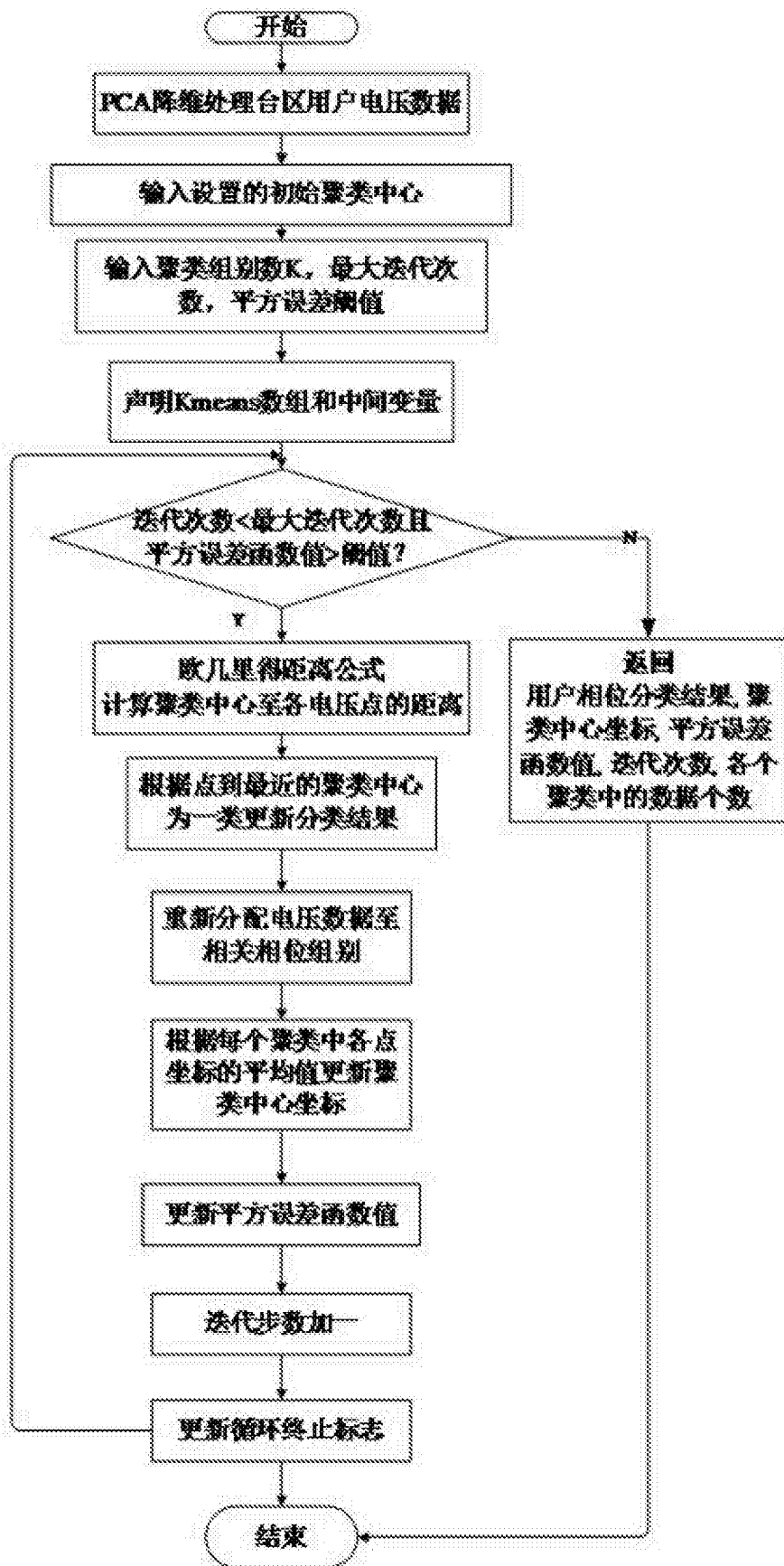


图2

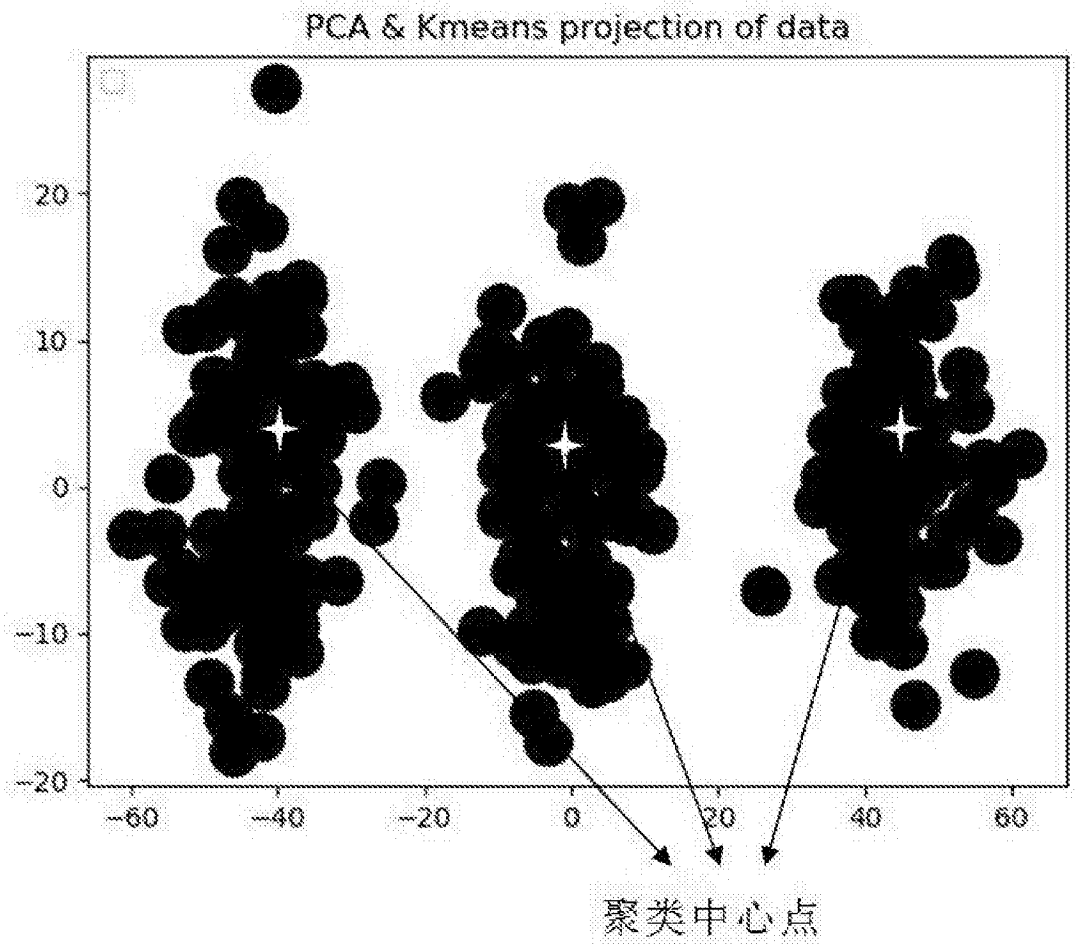


图3

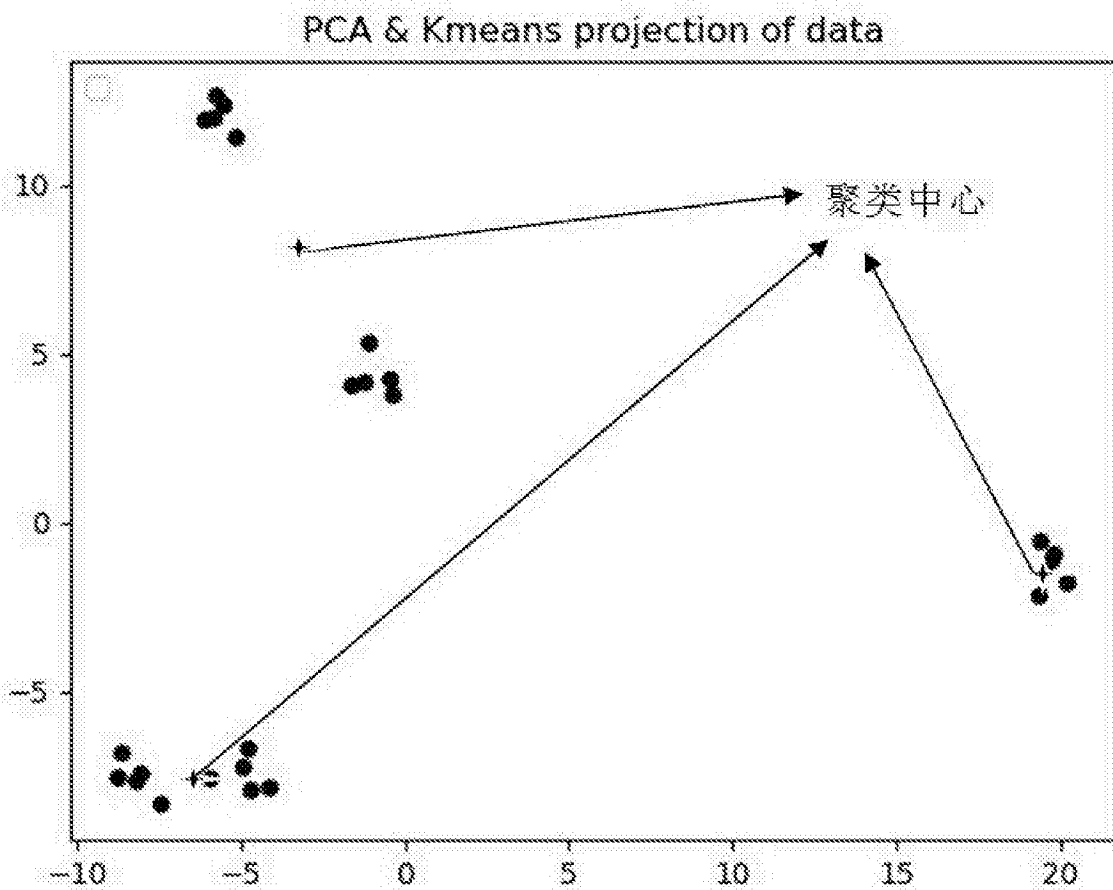


图4