

基于 GSA 算法的宽带载波抄表用户类型辨识

易永仙^{1,2}, 陈 霄^{1,2}, 崔高颖^{1,2}, 周 玉^{1,2}

(1. 国网江苏省电力公司电力科学研究院, 江苏省南京市, 211103;

2. 国家电网公司电能计量重点实验室, 江苏省南京市, 210019)

摘 要: 采用 GSA 算法确定 K-Means 的聚类数量, 能够提高聚类分析的效率。应用宽带载波技术, 可以方便地实现小区用电信息采集, 利用其高速信道可以得到多种密集度较高的用电数据。对于宽带载波抄表系统采集的用户用电曲线, 采用基于 GSA 算法的聚类方法对数据进行分析, 实现用户类型的辨识。通过该方法可以帮助供电公司充分利用采集的用电数据, 提高大数据分析能力。

关键词: GSA 算法; 聚类分析; 宽带载波技术; 大数据分析

0 引言

随着用电信息采集系统的建设, 海量的电力负荷数据被收集起来。为促进供电服务水平不断提高, 推动有序用电, 电力需求侧管理技术得到广泛应用。电力系统负荷分类是用来挖掘用户用电信息、负荷预测与系统规划、用户个性化服务和互动的重要基础性工作。用户的负荷类型信息可以帮助供电公司分析电力用户的电力需求特点, 从用电行为和用电习惯等特征进行细分, 并制定相应的营销策略, 以提供相应的个性化的服务。通过对电力用户进行类型辨识, 掌握用电负荷的分布和变化规律, 有助于实现有序用电, 提高用电管理服务水平, 具有重要的理论意义和现实意义^[1-2]。

当前有多种方法可以用于电力用户负荷分类, 其方法理论依据多为聚类分析技术, 不同点在于选取的聚类算法。如基于自适应模糊C均值 (Fuzzy C Means, FCM) 对其进行聚类分类^[3-4], 并通过实例进行了分析验证^[5]; 文献[6]基于蚁群优化算法并与 Kohonen 神经网络对比; 文献[7]应用模糊免疫网络和最小生成树方法对样本进行归类。K-Means 算法是一种计算数据聚集的常用算法, 通过不断地取距离种子点最近均值, 输出满足方差最小标准的K个聚类, 计算复杂度低, 能有效地处理大数据集, 在高维的负荷数据聚类中也体现了其优越性, 因此成为负荷分类中经常使用的一种基本算法^[8]。但是 K-means 算法的聚类数量K值需要事先给定, 随机选取的初始聚类中心会严重影响最终结果。ISODATA 算法通过类的自动合并和分裂, 得到较为合理的类型数目K^[9], 间隙统计算法 (Gap Statistic Algorithm,

GSA) 通过计算聚类相似度来确定最佳聚类个数, 可以大大提高K-Means聚类的效率^[10]。

在用电信息采集系统已经初步建成的现实条件下, 供电公司已经得到大量的用户用电数据, 但因现场数据采集手段有限, 目前针对普通用户, 主要采集日冻结电量作为结算依据, 如果用于用户类型辨识则时间间隔过长^[11]。因此, 采用电力线宽带载波 (Broadband Power Line Communication, BPLC) 技术, 采集更为密集的小时甚至分钟数据, 包括电量、电流、电压和功率等, 通过GSA算法实现用户类型的聚类分析, 可以为电力负荷分类、预测和规划提供海量数据依据, 宽带载波技术所带来的高速通信网络也可以为电力用户信息互动等信息和增值业务提供网络基础, 有助于提高用电管理和用电服务水平。

1 聚类算法

1.1 模糊评价基本模型

用于电力系统数据辨识聚类分析方法, 主要是 K-Means 算法, 聚类个数 K 值需要人工给定, 随机选取的初始聚类中心会严重影响最终结果, 而间隙统计算法 GSA 可以用来估计聚类的最佳聚类个数, 将两种方法结合用于电力系统数据辨识, 可取得良好的辨识效果。

K-Means 算法需要事先输入聚类数量 K, 将 N 个数据对象划分为 K 个聚类, 利用各聚类中对象的均值获得一个“中心对象”计算聚类相似度, 使同一聚类中的对象相似度较高, 而不同聚类中的对象相似度较小。

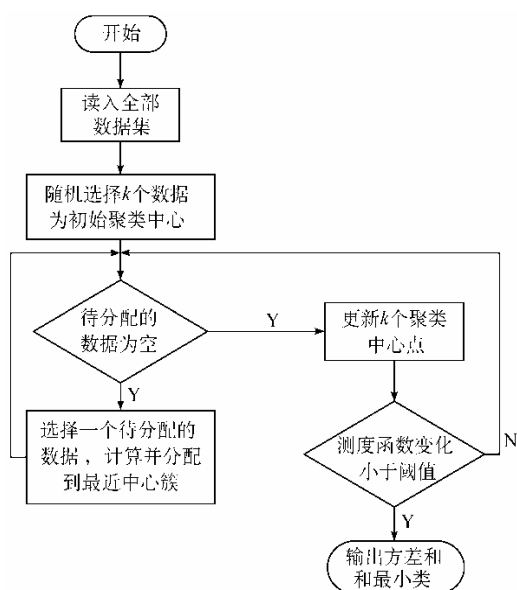


图 1 K-Means 聚类辨识流程

GSA 算法先采用具有典型负荷特性的参考数据进行训练，设数据样本可产生 k 个聚类，定义数据的离散度为

$$W_k = \sum_{i=1}^k \frac{D_i}{2C_i} \quad (1)$$

式中， C_i 为第 i 个聚类中数据点的数量； D_i 为聚类内部各个数据点欧氏距离平方之和为

$$D_i = \sum_{p,q \in F_i} d_{p,q} \quad (2)$$

式中， F_i 为第 i 个聚类所有数据点的集合； p 、 q 为其中的任意两点； $d_{p,q}$ 为任意两个数据点欧氏距离的平方。通过与参考数据集的聚类离散度 $W_r(k)$ 进行对比可得到间隙值为

$$\text{Gap}(k) = E[\ln W_r(k)] - \ln W(k) \quad (3)$$

随着 k 的变化，当 k 满足以下条件时，聚类个数为最佳。

$$\text{Gap}(k) \geq \text{gap}(k+1) - s_{k+1} \quad (4)$$

式中， s_{k+1} 为 $E[\ln W_r(k)]$ 的模拟误差。通过逐步增大 k 值，并利用式 (3) ~ 式 (4) 进行计算比较可最终确定聚类的个数。

2 宽带载波抄表系统

与采用专有线路的各类有线通信方式不同，载波通信借助已有的电力线路传输数据，必然受到电力线路上各类电源、导线、开关、接头和负载本身特性的影响，以及这些线路单元产生的噪声干扰；

电力线路特性的时变性和复杂网络拓扑结构，会导致信号的传输带宽和通信距离受到限制。电力线宽带载波技术采用了 OFDM（正交频分复用）、DMT（多载波调制）等调制技术，在电力线路上实现宽带数据传输，可以使接入电源的终端快速接入通信网络。

图 2 为典型的宽带载波抄表系统体系。在已经建成的居民小区中，将如图 3 所示的宽带载波模块插入电能表中，在变压器出线处安装宽带载波集中器，即可快速建立宽带载波抄表系统。宽带载波模块具有符合行业标准的串行通信接口，可以直接安装在具有载波接口的电能表中，无需对电能表进行更改，同时还可以提供一个额外的以太网接口连接其他的设备，实现共用带宽。

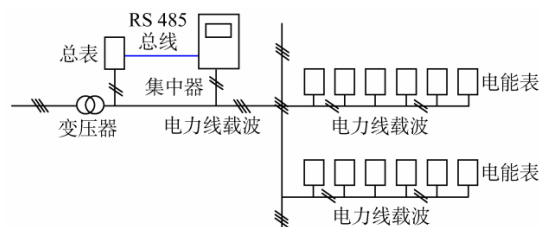


图 2 宽带载波抄表系统

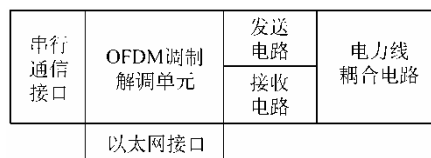


图 3 宽带载波模块功能框图

宽带载波能够提供瞬时高速和大流量的通信，从而具有快速批量抄收台区所有电能表即时和 24 h 整点电量、电压和电流等多个数据项的能力。该系统不仅实现了对每块用户电能表的数据抄收，还具有快速实时数据采集和实时双向互动功能。例如实时的远程断复电等，大大丰富了用电数据采集系统的功能，并为当前基于大数据的用电分析提供了有力的支持。

3 数据分析

对电力用户进行实时用电监测，分析和统计每个用户每日 24 h 整点电量、电压和电流等多项数据，可以为用电营销的精细化管理提供数据支撑，如负荷分类、负荷预测等。同时，还可以通过用户整点电量、电压和电流的比对，随时查出疑似窃电用户，提高用电管理水平。

在某已经安装宽带载波抄表系统的小区，采集

电力用户的日负荷数据进行数据分析,剔除空置房,总计 62 个用户的用电负荷数据用于数据分析,数据采用标么值,原始数据如图 4 所示。图 5 为采用 K-Means 算法得到的在不同 K 值时的 $\log(W_k)$ 曲线。

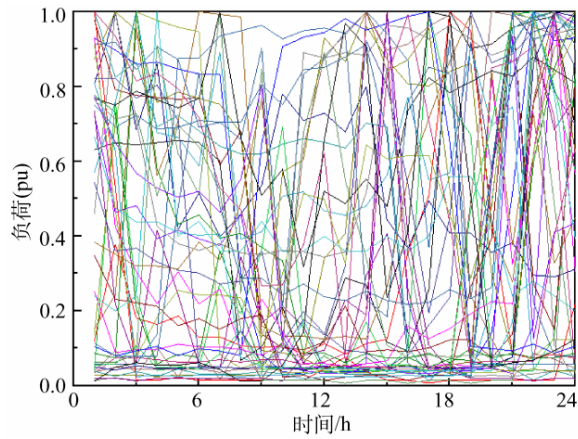


图 4 原始的用户负荷曲线

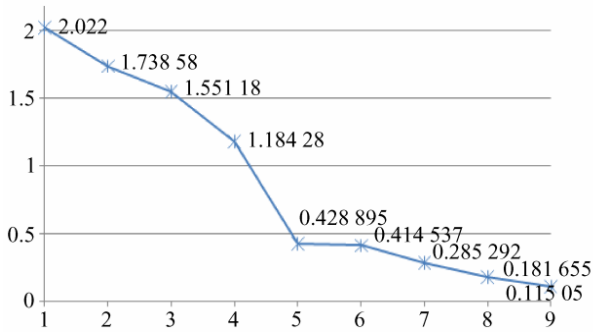


图 5 聚类算法 $\log(W_k)$ -K 曲线

根据 GSA 算法确定最佳聚类个数 K 为 5, 聚类结果如图 6 所示。小区各类用户的负荷曲线分别如图 7~图 11 所示。同类用户的负荷曲线形态相近, 不同类型用户的负荷曲线差异明显, 表明聚类分析得到的用户辨识效果较好。

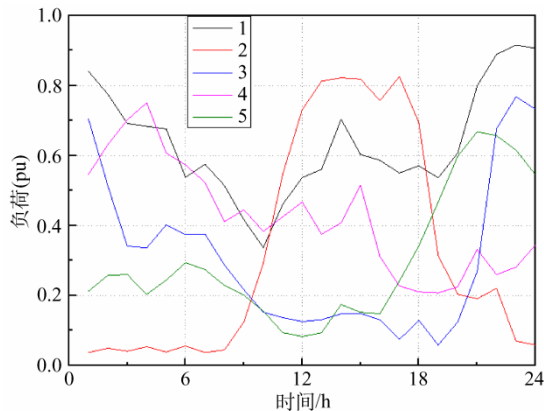


图 6 聚类结果

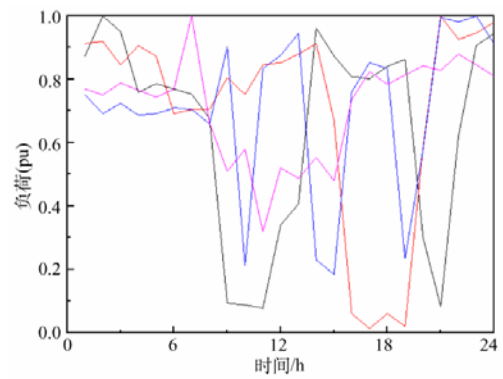


图 7 第一类用户负荷曲线

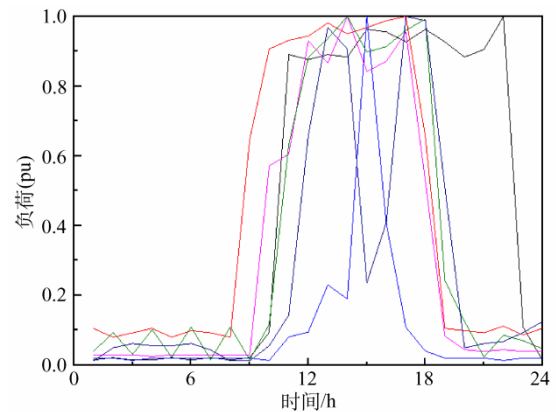


图 8 第二类用户负荷曲线

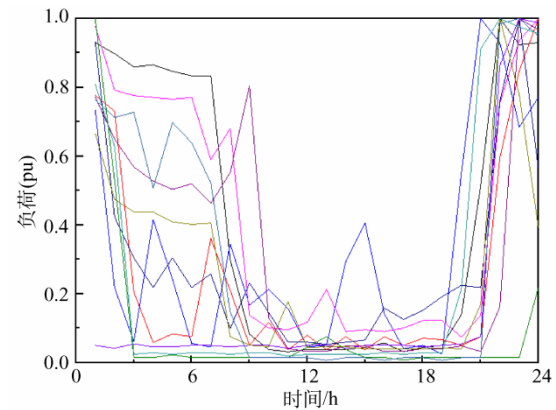


图 9 第三类用户负荷曲线

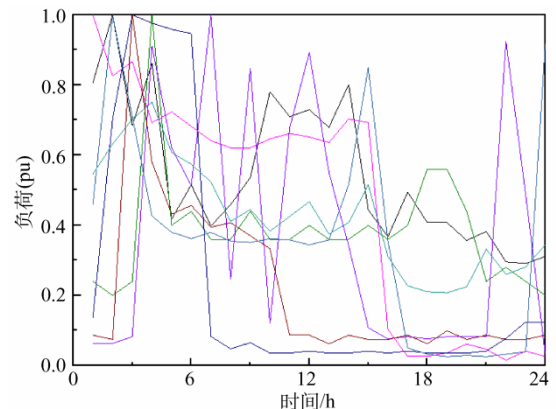


图 10 第四类用户负荷曲线

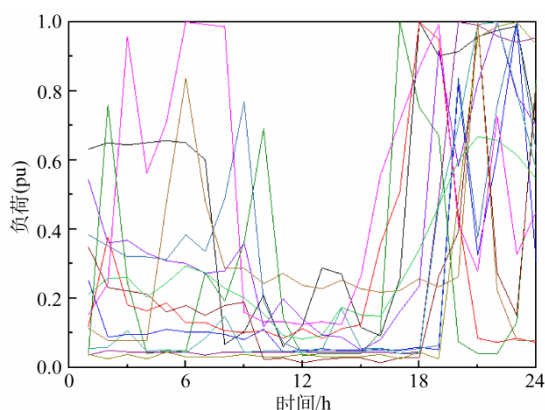


图 11 第五类用户负荷曲线

从负荷分类结果可见,第一类用户是居家用户,夜间用电较多,可能与空调等大功率电器用电有关,白天出入活动,用电具有不确定性;第二类用户是明显的白天用电用户,例如公司、办公室等;第三类用户是错峰用电用户,是比较关注用电成本的一类;第四类用户用电相对平稳,没有突出的峰谷,可能是家中经常有人的用户;第五类用户早晚负荷较大,白天负荷小,应是典型的白天外出工作居民的用电。

4 结束语

将基于 GSA 算法的宽带载波抄表用户类型辨识技术,用于用电数据采集系统的数据分析,可以取得如下效果。

1) 可以实时监测用户即时和 24 h 整点电量、电压和电流对电力用户进行实时用电监测,分析和统计每个用户每日 24 h 整点电量、电压和电流等多项数据,可以为用电营销的精细化管理提供数据支撑,如负荷分类、负荷预测等。同时,还可以通过用户整点电量、电压和电流的比对,随时查出疑似窃电用户,提高用电管理水平。宽带载波能够提供瞬时高速和大流量的通信,从而具有快速批量抄收台区所有电能表即时和 24 h 整点电量、电压和电流等多项数据的能力。

2) 通过用户类型辨识,可以针对不同类型的用户分别进行分析,对于目前智能电网互动化的趋势提供基础数据,有助于为用户提供定制服务,提高供电服务水平。

3) 除了对小区内部不同用户进行类型辨识,还可以对配电网覆盖的不同用户以及和沿线用电节点进行类型辨识,通过对不同类型用户进行组合供电,可以最大限度地利用线路资源,降低建设投资,有

助于完善配电网建设,提高供电质量。

4) 可以快速稳定地实现用户的断复电。根据对试点小区用户电能表的实时断复电操作,停电和恢复供电的运行时间都在 1 s 左右,这对于实行用电费控提供了基本保障。基于宽带载波构建的网络,对于今后对电力用户的个性化服务和双向互动都能够提供高速带宽。

参考文献:

- [1] 陆婷婷,高赐威,苏卫华,等.有序用电错峰预案优化编制方法研究[J].电网技术,2014,38(09):2315-2321.
- [2] 鞠平,陈谦,熊传平,等.基于日负荷曲线的负荷分类和综合建模[J].电力系统自动化,2006,30(16):6-9.
- [3] 周开乐,杨善林,王晓佳,等.基于自适应模糊度参数选择改进 FCM 算法的负荷分类[J].系统工程理论与实践,2014,34(5):1283-1289.
- [4] 王文生,王进,王科文. SOM 神经网络和 C-均值法在负荷分类中的应用[J].电力系统及其自动化学报,2011,23(4):36-39.
- [5] 李培强,李欣然,陈辉华,等.基于模糊聚类的电力负荷特性的分类与综合[J].中国电机工程学报,2005,25(24):73-78.
- [6] 孙雅明,王晨力,张智晟,等.基于蚁群优化算法的电力系统负荷序列的聚类分析[J].中国电机工程学报,2005,25(18):40-45.
- [7] 顾丹珍,艾芊,陈陈.一种基于免疫网络理论的负荷分类方法[J].电网技术,2007,31(S1):6-9.
- [8] 白雪峰,蒋国栋.基于改进 K-means 聚类算法的负荷建模及应用.电力自动化设备[J],2010,30(7):80-83.
- [9] 赵永良.用电信息采集系统本地通信方式对比研究[J].电力系统通信,2010,31(26):50-54.
- [10] 洪军,余立莱,黄兆星,等. Fuzzy ISODATA 聚类分析方法的设计[J].计算机与数字工程,2009,37(6):19-20,28.
- [11] 吴军基,杨伟,葛成,等.基于 GSA 的肘形判据用于电力系统不良数据辨识[J].中国电机工程学报,2006,26(22):23-28.

作者简介:

易永仙(1988-),男,浙江苍南人,工程师,主要研究电测量技术和用电信息采集技术;

陈霄(1985-),男,江苏连云港人,高级工程师,主要研究电测量技术和用电信息采集技术;

崔高颖 (1980-), 女, 山东沂水人, 高级工程师, 主要研究
电力信息化和智能用电技术;

周 玉 (1982-), 男, 江苏镇江人, 工程师, 主要研究电测
量技术和用电信息采集技术。

The Type Identification Of BPLC Reading Based on GSA algorithm

Yongxian Yi, Xiao Chen, Gaoying Cui, Yu Zhou

(1. State Grid Jiangsu Electric Power Research Institute, Nanjing 210019, China;

2. Key Laboratory of Electric Energy Metering, State Grid Corporation, Nanjing 210019, China)

Abstract: The GSA algorithm is used to determine the number of K-Means clustering, which can improve the efficiency of clustering analysis. Application of BPLC technology, can collect the residential electricity information easily, the use of its high speed channel can achieve a variety of intensive electricity data. The residential electricity curve, which is read by the technology of BPLC, can be analyzed based on GSA algorithm, the type identification can be achieved through the analytic result. This method can help electric power company make full use of the collected electricity data and improve the ability of large data analysis.

Key words: GSA algorithm; cluster analysis; BPLC technology; large data analysis