

# 基于分布式计算的海量用电数据分析技术研究

蒋菱<sup>1</sup>, 王旭东<sup>1</sup>, 于建成<sup>1</sup>, 袁晓冬<sup>2</sup>

(1. 国网天津市电力公司, 天津 300010;

2. 江苏省电力科学研究院, 江苏 南京 210036)

**摘要:**用电行为分析技术对供电企业掌握用户用能方式、调整生产计划以及进行电网规划有着较大的现实意义。传统用电行为分析多利用少量样本数据,由于数据源覆盖面的问题往往容易造成结果偏差。借助大数据技术,可以利用海量用电数据提高用电行为分析的准确性。针对用电行为分析在处理海量数据时效率低下的问题,提出了基于 MapReduce 技术的模糊 C 均值聚类 (FCM) 并行算法,通过将 FCM 算法的迭代过程分解到 Map 和 Reduce 两个步骤中,可以有效地提高聚类过程中数据对象和聚类中心的相似度计算效率。在此基础上,利用所提出的 FCM 并行算法对居民用电数据的四个特征进行聚类分析。实验结果表明,所提算法可以提高海量用电数据聚类分析的效率,证明了计算模型的可行性。

**关键词:** MapReduce; 模糊 C 均值聚类; 用电行为分析; 大数据

中图分类号: TP39

文献标识码: A

文章编号: 1673-629X(2016)12-0176-06

doi: 10.3969/j.issn.1673-629X.2016.12.038

## Research on Power Usage Behavior Analysis Based on Distributed Computing

JIANG Ling<sup>1</sup>, WANG Xu-dong<sup>1</sup>, YU Jian-cheng<sup>1</sup>, YUAN Xiao-dong<sup>2</sup>

(1. State Grid Tianjin Electric Power Company, Tianjin 300010, China;

2. Jiangsu Electric Power Research Institute, Nanjing 210036, China)

**Abstract:** The power usage behavior analysis technology can be used to acquire customer power usage pattern, adjust power generation schedule and plan grid development. Thus, it is meaningful to power grid company. Traditional power usage behavior analysis only uses small volume of data. The limited data will draw to inaccurate result. This problem can be solved by using large scale of data. In allusion to the problem about electricity behavior analysis in the low efficiency of dealing with huge amounts of data, the Fuzzy C-Means clustering (FCM) parallel algorithm based on MapReduce is put forward. By decomposing the iterative process of FCM algorithm into two steps of Map and Reduce, it can effectively improve the efficiency of similarity computing between the data objects and the clustering centers. On this basis, the four characteristics of resident electrical data are clustering analyzed by using the proposed FCM parallel algorithm. The experimental results show that the proposed algorithm can improve the efficiency of mass data clustering analysis and also proves the feasibility of the model.

**Key words:** MapReduce; FCM; analysis of electric behavior; big data

## 0 引言

随着国家电网公司智能电网建设的不断深入推进,先进的信息技术和数字通信技术在电力网络的发电、输电、配电、调度、用电和客户服务等各个环节得到了应用<sup>[1]</sup>。同时,随着经济的快速发展和居民生活水平的日益提高,一方面居民用电量在不断增长,另一方面,居民用户对用电服务的个性化要求也在逐渐提高。

电力企业除了向客户提供电能产品之外,还承担着对用电行为进行专业化指导,提高电能利用效率和利用水平的任务<sup>[2]</sup>。这些需求的满足依赖于用电数据采集和用电数据分析技术。

国网天津市电力公司于 2010 年 1 月启动智能电网综合示范工程“中新天津生态城智能电网创新示范区”建设。工程于 2011 年 9 月建成投运,建设内容包

收稿日期: 2015-09-09

修回日期: 2015-12-23

网络出版时间: 2016-11-21

基金项目: 国家自然科学基金资助项目(51407025); 江苏省科技支撑计划(社会发展)(BE2013737); 天津电力公司科技项目(SGTJDK00DWJS1500033)

作者简介: 蒋菱(1971-),女,高级工程师,研究方向为配用电与新能源接入。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20161121.1633.010.html>

括分布式发电、微电网、配电自动化等 12 个子项,集中示范智能配电、智能发电、智能用电和信息通信领域的先进技术。在发电侧、电网侧、用户侧的信息通信方面、技术储备方面以及政策支持方面均已取得显著进展<sup>[3-5]</sup>。其中,双向智能电表和用户与电网双向互动技术的应用可以使天津生态城中的居民用户和企业大用户能够获取用电量、费率等用电信息,同时接收电力企业下达的用电指导和负荷控制指令,这使得通过提高终端用电效率和优化用电方式,满足用户用电需求的同时减少电量消耗,从而达到节约能源和保护环境的目的成为可能。

基于智能电表数据,统计并挖掘电力客户的用电模式,是电力企业掌握客户构成,了解用电行为特征的基础,也是提供个性化、精细化用电服务,实现客户智能化、精益化管理的先决条件。近年来,已经有一些专家学者对用电行为分析进行了研究。文献[6]提出了基于  $k$ -means 算法的用电负荷特性分析算法,实现了依据负荷特性对用户负荷的分类。针对传统的单一聚类分析方法对于具有不平衡性以及时序特性的负荷曲线数据存在泛化能力不强、稳健性不高的问题,文献[7]提出利用多种聚类融合的方法获得更优的聚类结果。文献[8]提出了基于模糊聚类的电力负荷特性的分类与综合算法,并通过实验证明基于模糊 C 均值法的聚类能力明显优于基于等价关系的聚类法,而且聚类结果更为合理有效。而文献[9]构建了基于  $k$ -means、 $k$ -medoids、SOM 以及 FCM 等聚类算法的聚类分析模型,实现了对数据集的智能化聚类的分析功能,实验结果表明 FCM 模型对用电行为特征的聚类结果更具归纳性。但是,随着电力通信技术的发展,用电信息采集系统每天产生的用电数据是高频海量的,这就对用户行为特征分析技术提出了要求,即能够高速、高精度处理数量庞大且数据类型众多的用电数据,从中发掘高价值信息。这符合典型的大数据应用特征,同时也意味着使用传统聚类算法无法直接满足上述要求,需要针对大数据的特点进行并行化改进,以适应分

布式计算的需求。文献[10]提出利用 MapReduce 计算模型实现  $k$ -means 聚类算法,但是未实现对于用电数据的分析。文献[11-12]均提出了在 MapReduce 模型下基于  $k$ -means 的用电数据分析算法,但是无法直接应用于聚类效果更好的模糊聚类算法中。

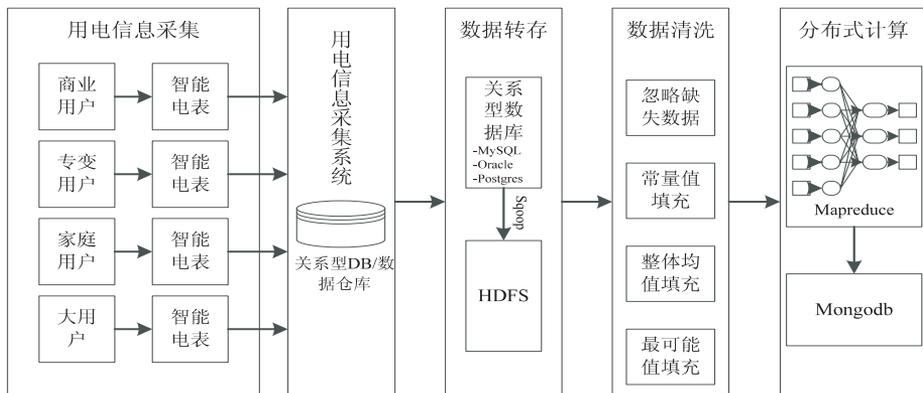
针对智能用电领域对海量数据进行用电行为特征分析的需求,文中提出了一种基于模糊 C 均值聚类(Fuzzy C-Means clustering, FCM)的并行计算算法。该算法在 MapReduce 框架下实现,可以利用 FCM 算法的模糊分析特性对用户用电行为进行更为全面地分析,并利用并行计算提高对海量数据进行分析的效率和可行性。实验结果表明,该算法可以精确用于居民用户用电数据的分析统计,以及对用户的用电模式进行快速、精确的判断。

## 1 用电数据分析分布式计算架构

随着智能电表的普及应用,用户用电信息采集频率更加频繁,15 min 甚至 5 min 就需要采集一次数据,且数据呈现双向流动特征,规模和频率呈指数级增长。以天津生态城为例,用电信息采集系统目前已经覆盖 1 500 万用户,数据年增长量约为 12 TB 左右。因此,传统基于单机的分析模式已经无法满足对于海量用电数据的分析需求。

对于海量数据进行分布式批处理计算是提高聚类计算效率的关键,批处理计算框架的理论基础是 Google 的 MapReduce 计算框架。MapReduce 将复杂的并行计算过程高度抽象到两个函数,Map 和 Reduce,并可运行于大规模计算集群上。利用 MapReduce 框架,可以将大规模计算任务分解成许多小的子任务由 Map 步骤处理,由于子任务之间是相互解耦的,因此可以并行处理,Map 输出的结果将通过 Reduce 函数合并生成最终结果。MapReduce 的开源实现的代表就是 Hadoop 平台,目前 Hadoop 广泛被互联网企业用于大规模数据分析。

如图 1 所示,基于 MapReduce 计算框架的智能用



电分析系统分为用电信息采集、数据转存、数据清洗和分布式计算这四个步骤。

(1)用电信息采集:用电信息采集依托于采集终端,包括双向智能电表、转变采集终端、负荷控制终端和分布式能源监控终端等,采集终端实现电能数据的采集、数据管理、数据双向传输以及控制命令执行。采集终端从不同类型的用电用户处以一定频率采集用电数据,包括电压、电流、功率、电能质量和异常事件等,并通过 PLC、无线网络等数据通道保存在用电信息采集系统中,可作为用电行为分析的基础数据。

(2)数据转存:由于用电信息采集系统是基于关系型数据库或数据仓库的,而 MapReduce 计算是基于 HDFS 分布式文件系统的,因此需要通过数据转存将数据从用电信息采集系统转移至 HDFS 中。可以利用基于 Hadoop 平台的数据传输工具完成这一过程,如 Apache 项目 Sqoop,可以用来在 Hadoop 和关系数据库中传递数据。通过 Sqoop,可以方便地将数据从关系数据库导入到 HDFS,或者将数据从 HDFS 导出到关系数据库。

(3)数据清洗:在对用电数据进行聚类分析之前,为了保证结果的可靠性,需要使用数据清洗等数据预处理手段对不完整数据、错误数据和重复数据进行补充、修正和删除,常见数据清洗的方法包括忽略缺失数据、删除负值、用整体均值填充、用最可能值填充、回归方法填充等。

(4)分布式计算:在分布式计算阶段,将通过文中提出的基于 MapReduce 的分布式 FCM 聚类算法完成对用电行为数据的聚类,从中获得的聚类中心可以用于刻画用电用户群体特征,而每一个参与聚类的数据对象对于不同聚类的模糊隶属度关系,可以用来判断用电用户所属的聚类。在完成聚类过程后,可以将结果以键值对的形式保存在非关系型数据库(NoSQL),如 Mongoddb 中,从而方便对于聚类结果的查询或在聚类结果基础上进行进一步的数据挖掘。

## 2 基于 FCM 算法的用电行为特征分析

### 2.1 FCM 聚类算法

聚类是一种最常见的对大规模数据集进行检验和分类的无监督学习算法(Unsupervised Learning Algorithm)。在无监督学习中,不需要预先对群体进行分类或设置辅助聚类过程的样本,而是根据数据元素自身特性的自动化分组,同一聚类中的数据对象将比来自于其他聚类中的数据对象实例更加“接近”。目前,已有很多聚类算法被应用于不同的领域,其中,模糊聚类(Fuzzy Clustering)算法考虑到了真实数据的不确定性,并且与硬划分(Hard Clustering)相比,模糊聚类算

法允许一个数据对象属于多个不同的聚类,数据对象与每个聚类中心的接近程度可以使用隶属度来衡量,因此其应用方式更为灵活。其中,文中提出使用 FCM 聚类算法进行用电行为分析,基于目标函数的 FCM 聚类算法适用于处理大量数据,而且算法过程简单,因此易于在计算机上实现,适合对基于时间序列的复杂数据集进行划分,这一特性与用电数据的特性吻合。FCM 算法的核心思想是通过求解  $J_m(U, P)$  的极小值解  $\min\{J_m(U, P)\}$ ,从而获得最佳的划分矩阵和聚类中心矩阵。对于模式空间中包含  $n$  个成员的待分类对象集合  $X = \{x_1, x_2, \dots, x_n\}$  而言,划分举证  $U$  可以表示为:

$$U = [\mu_{ik}] = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1n} \\ \vdots & \mu_{ik} & \vdots \\ \mu_{c1} & \cdots & \mu_{cn} \end{bmatrix}, 1 \leq i \leq c, 1 \leq k \leq n \tag{1}$$

其中,  $\mu_{ik} = \mu_{X_i}(x_k)$  表示样本  $x_k$  与子集  $X_i (1 \leq i \leq c)$  之间的隶属关系,对于 FCM 而言,  $\mu_{ik}$  的取值范围为  $[0, 1]$ ,即每个样本与子集  $X_i$  之间的隶属关系可以由一个  $0 \sim 1$  之间的实数模糊表示。而  $P = \{p_i, 1 \leq i \leq c\}$  表示第  $i$  类子集  $X_i$  的聚类中心矩阵。

优化目标可以表示为:

$$J_m(U, P) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m (d_{ik})^2, 1 \leq m \leq \infty \tag{2}$$

其中,  $m$  为平滑因子,  $m$  控制模式在类子集之间的分享程度,  $m$  越大,得到的聚类结果越模糊,一般情况下,为了控制聚类结果不要太模糊,将  $m$  设为 2;  $d_{ik}$  表示样本  $k$  到第  $i$  个聚类中心  $p_i$  之间的距离,可以用不同类型的范式距离表示,文中使用欧氏距离表征:

$$d_{ik} = \left[ \sum_{j=1}^s |x_{ij} - x_{kj}|^2 \right]^{1/2} \tag{3}$$

FCM 算法通过迭代不断更新隶属度  $\mu_{ik}$  和聚类中心  $p_i$ ,当迭代收敛时,获得的隶属度和聚类中心可以用于对数据集进行分类并确定数据对象与分类之间的隶属关系,迭代过程通过在停止域和迭代次数  $b$  的控制下,对下式进行求解进行:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left[ \frac{d_{ik}}{d_{jk}} \right]^{\frac{2}{m-1}}} \tag{4}$$

$$p_i = \frac{\sum_{k=1}^n (\mu_{ik})^m x_k}{\sum_{k=1}^n (\mu_{ik})^m} \tag{5}$$

### 2.2 基于 FCM 的用电行为分析

居民用户、大用户安装的智能电表借助 PLC 和无线通信等技术,以一定频率向用电信息采集系统传输

用户用电数据,从中选取四类特征作为聚类分析的数据对象:

- (1) 负荷量  $x_{i1}$ : 采集时刻的用电负荷;
- (2) 负荷率  $x_{i2}$ : 平均负荷/最大负荷;
- (3) 峰电系数  $x_{i3}$ : 峰时用电量/日用电总量;
- (4) 谷电系数  $x_{i4}$ : 谷时用电量/日用电总量。

智能电表的采集频率是每 15 min 一个点,因此每日采集 96 个,日用电总量为 96 点数据之和,平均负荷为日用电总量/96,峰时用电量和谷时用电量分别为峰谷时间内的用电总量。因此聚类分析的每一个样本  $x_k$  均为一个四维向量。基于 FCM 算法的用电行为分析流程如图 2 所示。

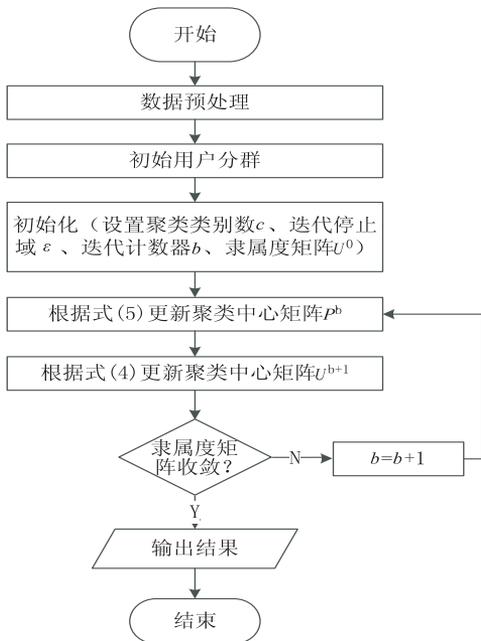


图 2 基于 FCM 的用电行为分析算法

在数据预处理阶段,需要对缺失数据利用差值算法进行补齐处理,对于超出阈值的数据进行修正。接下来,对 FCM 聚类算法进行初始化,包括设置聚类类别  $c$ , 迭代停止域  $\varepsilon$  和迭代步数  $b=0$ , 以及隶属度矩阵  $U^0$ , 可根据用电历史数据进行初始用户分群并计算  $U^0$ 。接下来,根据式(4)和式(5)在迭代过程中不断更新隶属度和聚类中心,直到满足设定的停止域条件  $\|U^{(b)} - U^{(b+1)}\| < \varepsilon$  为止。此时,输出的聚类中心即为用电行为特征,而隶属度矩阵决定了每个样本与用电行为特征的接近程度。

### 2.3 聚类有效性验证

聚类分析的结果与数据样本和参数设定密切相关,由于聚类是一个无监督的学习过程,因此无法获取数据对象相关的标签信息。因此,对于聚类算法对一个数据集产生某种划分结果后,通常难以直观评价一个特定聚类划分的优劣,因此需要引入聚类有效性验证算法对聚类结果进行有效评价。评价的内容包括量

化聚类的簇内紧凑度和簇间分离度。对于模糊聚类算法而言,代表性的聚类有效性验证方法包括 Xie-Beni 指标  $V_{xie}^{[13]}$ 。 $V_{xie}$  基于几何结构,采用“紧凑度”和“分离度”衡量不同划分的聚类质量。对于文中所应用的 FCM 算法而言, $V_{xie}$  通过获取式(6)的最小值完成对聚类有效性的验证。

$$V_{xie}(U) = \frac{\sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m (d_{ik})^2}{n \cdot (\min\{c_i - c_j\})} \quad (6)$$

### 3 FCM 的分布式计算方法

传统的用电行为聚类算法需要将样本数据放入计算机内存再进行计算,受限于计算机的内存大小和运算速度,无法对大量的用电历史数据进行全局计算,只能从中抽取样本,通过对抽样数据集的聚类分析用电行为特征,其结果的准确性由于样本缺失而无法得到保证。而文中所提出的混合计算架构中的批量计算层,可以使用 Hadoop 平台对大规模数据进行分布式计算,由于使用了分布式文件系统 (Hadoop Distributed File System, HDFS) 和 MapReduce 计算模型,分布式计算可以对基于文件的海量历史用电整体数据进行直接计算并获得用电行为特征。

为了适应 MapReduce 计算模型,需要对基于 FCM 算法的用电行为特征分析算法进行并行化改造,将 FCM 的迭代过程分解为 Map 和 Reduce 两个阶段。Map 阶段在不同的数据节点上将同一个函数作用于不同的数据集,输出的数据集以  $\langle \text{key}, \text{value} \rangle$  形式保存在数据节点上,在 Map 阶段结束后,计算模型将传输至承担 Reduce 工作的节点,并对 Map 阶段输出的键值对进行合并等处理,并输出  $\langle \text{key}, \text{value} \rangle$  形式的最终结果。由于 Map 和 Reduce 步骤都是可以在多台计算机上分布式运行的,且对分布式计算过程进行了高度抽象,所以 MapReduce 计算模型可以方便、高效地对大规模数据 (1 TB 以上) 进行分析计算。

通过对 MapReduce 计算模型的研究,结合 FCM 算法的流程,可以发现相似度计算,即利用式(3)计算样本到当前聚类中心的距离是最为频繁的计算。对于  $n$  个样本对象在  $k$  个分类中的 FCM 聚类过程,每次迭代需要进行  $n * k$  次距离计算,且每次计算都要对  $s$  个维度的特征进行方差运算。如果能够将距离计算并行化处理,将极大地提高 FCM 的工作效率。根据这一思路,提出的基于 MapReduce 的 FCM 聚类算法流程如图 3 所示。

(1) 将用电数据从关系数据库 (如 Oracle) 拷贝到 HDFS 中,根据聚类的需要确定聚类个数  $c$  和停止域  $\varepsilon$ ;

(2) 根据上一次的聚类结果确定初始聚类中心, 并将这些数据传输至参与分布式计算的数据节点;

(3) 对用电数据进行预处理, 并产生键值对 <user, profile>, 其中 user 为用户的唯一标识, profile 中包含了数据对象的特征  $x_{i1} \sim x_{i4}$ ;

(4) 将所有的键值对 <user, profile> 划分为若干个数据子集, 并传输至 Map 函数, Map 函数根据式(4) 进行隶属度计算, 产生的结果保存在中间键值对 <  $i, \mu_i$  > 中, 其中  $i$  为聚类编号,  $\mu_i$  为数据子集中所有数据对象对第  $i$  个聚类的隶属度;

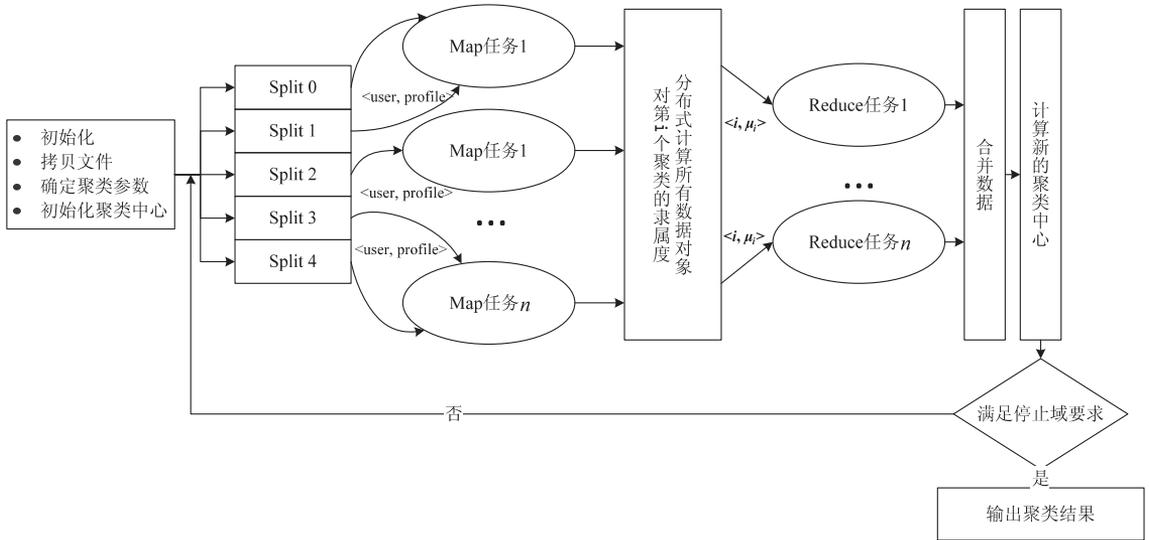


图 3 基于 MapReduce 的 FCM 并行计算过程

(5) 将 Map 函数计算的结果传输至 Reduce 节点, Reduce 将 Map 产生的中间键值根据聚类编号进行合并后, 根据式(5) 进行计算, 获得新的聚类中心;

(6) 重复步骤(2) ~ (5), 直到隶属度矩阵满足停止域条件, 分布式 FCM 算法结束, 输出聚类结果, 包括聚类编号、聚类中心和每个用户对于各个聚类的最终隶属度。

通过上述步骤, 可以实现在 Hadoop 平台上利用 MapReduce 计算模型对用户用电信息的分布式聚类分析, 通过最终的聚类中心矩阵获得对用户群体的分类, 并获得每一个样本数据对于聚类的隶属度, 从而确定其所属分类<sup>[14-18]</sup>。

### 4 算例分析

为了验证利用 FCM 聚类算法在分布式计算架构上实现用电行为聚类分析, 在实验室环境中搭建了由五个节点组成的分布式计算环境。其中一台为 Name-Node, 四台为 DataNode, 安装了 CentOS 6.5 版 32 位 Linux 操作系统, 以及 2.6.0 版本 Hadoop 并行计算环境, 节点间采用千兆以太网通信。收集了天津生态城某小区及周边商户共 457 户居民的家庭用电数据, 数据覆盖范围为 2014 年 7 月 3 日至 2014 年 10 月 28 日, 采样间隔为 15 min, 即每户居民每天采样 96 点数据, 以此为基础对居民用户类型展开研究。

在聚类分析之前, 使用式(7) 对数据进行了归一化处理: 万方数据

$$x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, i = 1, 2, \dots, n \quad (7)$$

其中,  $x_i$  为采集的每点数据;  $x_{\min}$  为整体数据的最小值;  $x_{\max}$  为整体数据的最大值;  $x_i$  为归一化后的值。

利用图 3 的算法流程对所采集的居民用电数据进行并行聚类分析后, 剔除因所包样本量过小而明显不合理的坏数据, 获得了四类典型用户, 如图 4 所示。

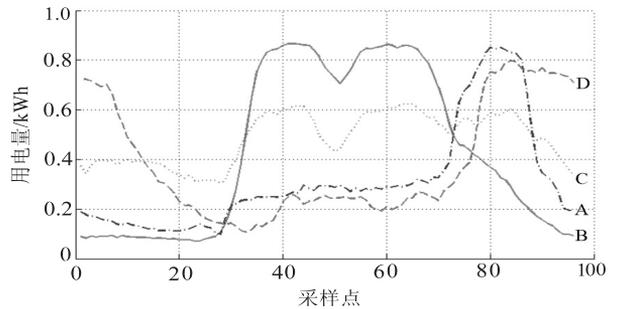


图 4 居民用电行为聚类结果

(1) A 类用户早、晚高峰时期用电负荷量大, 特别是晚间用电达到顶峰, 其余时间用电量较小, 为典型的上班族家庭用户, A 类用户共 274 户。

(2) B 类用户白天整体用电量较高, 而 19:00 后用电量开始回落, 符合在小区中租用办公室白天办公的公司特征, B 类用户共 84 户。

(3) C 类用户全天负荷处于较为平均的水平, 中午和晚间略高, 属于全天都要经营的商户特征, C 类用户共 67 户。

(4) D 类用户谷电系数高而峰电系数低, 且谷时

用电量处于较高水平,应为避免峰时高电价而选择谷时生产的小型生产加工企业,D类用户共32户。

随着智能电网的发展以及电网与用户双向互动业务的增加,未来可针对上述四种不同类型的用户提供不同的电价产品或套餐,或者根据其用电行为特征进行需求侧响应方案的设计,指导最优用电策略的制定。

为了测试算法的性能,将数据复制后形成10 GB数据集,在五节点的并行计算平台上进行聚类计算的结果与传统聚类算法相比获得了4倍以上的加速比。这证明了随着数据量的增大,聚类问题由多个处理器协同求解,待分类数据被分为若干个部分分别计算后再进行结果合并,从而使得聚类效率大大提升。

## 5 结束语

针对中新天津生态城中用户用电行为聚类分析的应用场景,提出利用并行计算技术进行聚类分析的计算过程,并具体实现了FCM聚类算法的并行化设计。实验结果表明,该算法能够较为准确地完成天津生态城内的用户分类,挖掘出了海量用电数据中潜在的价值,为用户参与需求侧响应和制定最优用电策略提供了有益的参考。

### 参考文献:

[1] 曹军威,万宇鑫,涂国煜,等.智能电网信息系统体系结构研究[J].计算机学报,2013,36(1):143-167.

[2] 胡学浩.智能电网——未来电网的发展态势[J].电网技术,2009(14):1-5.

[3] 尹倩.中新天津生态城运作模式研究[D].天津:天津理工大学,2009.

[4] 谢开,刘明志,于建成.中新天津生态城智能电网综合示范工程[J].电力科学与技术学报,2011,26(1):43-47.

[5] 李晓途.智能电力设备在中新生态城电网中的应用[D].保定:华北电力大学,2013.

[6] 王春雷,梁小放,章坚民,等.基于用电采集系统的负荷特性曲线聚类分析[J].浙江电力,2014,33(7):6-10.

[7] 林锦波.聚类融合与深度学习在用电负荷模式识别的应用研究[D].广州:华南理工大学,2014.

[8] 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.

[9] 彭显刚,赖家文,陈奕.基于聚类分析的客户用电模式智能识别方法[J].电力系统保护与控制,2014,42(19):68-73.

[10] Anchal P P. Improved MapReduce k-means clustering algorithm with combiner[C]//2014 UKSim-AMSS 16th international conference on computer modelling and simulation. Cambridge:IEEE,2014:12-17.

[11] 张素香,刘建明,赵丙镇,等.基于云计算的居民用电行为分析模型研究[J].电网技术,2013,37(6):1542-1546.

[12] 赵莉,候兴哲,胡君,等.基于改进k-means算法的海量智能用电数据分析[J].电网技术,2014,38(10):2715-2720.

[13] Xie X L, Beni G. A validity measure for fuzzy clustering[J]. IEEE Transactions on Pattern Analysis and machine Intelligence, 1991, 13(8):841-847.

[14] Rusitschka S, Eger K, Gerdes C. Smart grid data cloud: a model for utilizing cloud computing in the smart grid domain [C]//First IEEE international conference on smart grid communications. Gaithersburg, MD:IEEE,2010:483-488.

[15] Silva L, Moura R, Canuto A M P, et al. An interval-based framework for fuzzy clustering applications[J]. IEEE Transactions on Fuzzy Systems, 2015, 23(6):2174-2187.

[16] O'Malley M J, Abel M F, Damiano D L, et al. Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters[J]. IEEE Transactions on Rehabilitation Engineering, 1997, 5(4):300-309.

[17] Anderson D T, Zare A, Price S. Comparing fuzzy, probabilistic, and possibilistic partitions using the earth mover's distance [J]. IEEE Transactions on Fuzzy Systems, 2013, 21(4):766-775.

[18] Suh I H, Kim Jae-Hyun, Rhee F C. Convex-set-based fuzzy clustering[J]. IEEE Transactions on Fuzzy Systems, 1999, 7(3):271-285.