

基于 Graph 模型的海量用电数据 并行聚类分析

陶 鹏^{1,2}, 张洋瑞², 李梦宇², 李杰琳²

(1. 国网河北能源技术服务有限公司, 河北 石家庄 050021; 2. 国网河北省电力有限公司电力科学研究院, 河北 石家庄 050021)

摘要:随着智能电网建设的不断深入,在配用电环节收集的监测数据越来越多,逐渐构成智能电网用户侧大数据。传统数据分析模式已经无法满足性能需求,迫切需要新的存储和数据分析模式来应对。提出基于阿里云大数据分析平台 MaxCompute 的海量用电数据聚类分析方法,该方法充分考虑用电数据的特点,设计基于多级分区表的用电数据存储模式,采用三相电压、三相电流、三相功率因数等建立多维数据特征,应用 MaxCompute Graph 框架设计实现高效的海量用电数据的聚类划分算法。实验结果表明,所设计的存储模式可有效提升用电数据的检索效率;通过对不同用电类型的用户进行聚类划分,聚类准确率达到 88%,验证了聚类划分的有效性和高性能。

关键词:用电数据分析;大数据;聚类;Graph;云计算

DOI:10.19781/j.issn.1673-9140.2020.06.019 中图分类号:TM761 文章编号:1673-9140(2020)06-0144-08

Parallel clustering analysis for power consumption data based on graph model

TAO Peng^{1,2}, ZHANG Yangrui², LI Mengyu², LI Jielin²

(1. State Grid Hebei Energy Technology Service Limited Company, Shijiazhuang 050021, China; 2. Electric Power Research Institute, State Grid Hebei Electric Power Co., Ltd., Shijiazhuang 050021, China)

Abstract: As the deepen development of smart grids construction, more and more monitoring data are collected in power distribution network, and gradually forming big data on the user side of smart grids. The traditional data analysis model does no longer meet performance requirements of huge data processing. Thereby new storage and data analysis models are urgently needed to be established. Under this background, this paper proposes a cluster analysis method for massive electricity consumption data based on the Alibaba Cloud big data analysis platform MaxCompute. The characteristics of power consumption data are fully considered, and a multi-level partition table-based power consumption data storage mode are designed by adopting the three-phase voltage, three-phase current, three-phase power factor, etc. to establish multi-dimensional data characteristics. Furthermore, the MaxCompute Graph framework design is employed to achieve an efficient clustering and partitioning algorithm for massive electricity consumption data. Experimental results show that the designed storage mode can effectively improve the retrieval efficiency of electricity consumption data. The clustering accuracy rate reaches 88% for clustering of users with different electricity consumption types. The effectiveness and high efficiency of clustering is verified.

Key words: power consumption data analysis; big data; clustering; Graph; cloud computing

收稿日期:2018-05-16;修回日期:2018-07-14

基金项目:国家自然科学基金(51677072)

通信作者:陶 鹏(1979-),男,硕士,高级工程师,主要从事电力计量和大数据分析的研究;E-mail: hbdyytp@163.com

随着智能电网、通信和传感器技术的发展,电力用户侧数据呈指数级增长,复杂程度也越来越高,逐渐构成了用户侧大数据^[1]。用户侧大数据主要来源于电力公司的用电信息采集系统、电力营销系统等,目前应用范围较窄,主要用于电力公司内部营销业务辅助决策,其数据价值远没有被充分的挖掘。

如何更加深入地了解用电用户属性和特征,构建用户画像,并基于用电用户画像提供个性化、差异化服务是当前用户侧大数据研究的一个热点问题。基于用电信息数据进行聚类分析,可以有效了解、抽取用户的用电特性以及群体用电共性,有助于实现用电用户细分,提升电网公司服务广度和深度,为新型智能用电解决方案奠定基础。针对这一问题,已有一些学者开展了用电用户行为分析和用户分类方面的研究。文献[2]基于典型城市居民智能用电问卷调查数据,采用模糊综合评价的方法研究居民的偏好和智能用电行为,但并未使用能够客观反映用电行为的用电采集数据;文献[3]应用网络化分析方法对短时间尺度用电行为的相关性进行了研究,构造了用电行为相关性聚类分析与等级分析的网络模型;文献[4]针对复杂智能用电环境下智能用电小区的多用户日负荷需求响应问题,研究了用户用电行为聚类方法,并用于互动需求响应;文献[5]研究了基于聚类分析的区域尺度住宅日用电负荷模型和模型检验方法,以单个住户的日均用电和全年最大日负荷为指标进行聚类分析,但并未考虑海量数据造成的数据挖掘性能问题。

在数据分析技术方面,由于用户侧大数据具有海量、结构复杂、处理逻辑复杂、计算频度高、存储周期长^[6]等特征,使得传统的关系型数据库已经难以胜任存储容量、扩展性的需求,传统的集中式数据分析模式无法提供高效的计算性能。因此,一些学者结合云计算和大数据处理技术开展了电力大数据分布式存储和并行处理的研究工作。文献[7]基于 Hadoop MapReduce 设计实现了高效的并行化 EE-MD 算法,并应用于大规模局部放电数据的去噪;文献[8]针对智能小区的居民用电行为展开研究,并结合 MapReduce 对海量用电数据开展了并行化的聚类分析研究;文献[9]应用 Hadoop 实现了多源异构的配用电历史数据的存储和高效处理。上述研究工

作均使用了 MapReduce 算法并行框架实现,但是 MapReduce 并非能够有效适用于各类数据挖掘算法,尤其是对含有复杂迭代计算的问题,MapReduce 算法设计困难,即使使用任务链的方式,多轮迭代也会使得计算过程中 I/O 消耗增长,计算性能将受到严重影响。

MaxCompute Graph 是一套面向迭代的图计算处理框架,使用图进行数据建模,通过迭代对图进行编辑、演化,求解出最终结果。相较于 MapReduce 模型更适合含有大量迭代计算的场景,比如复杂统计分析、数据挖掘等^[10-12]。结合 MaxCompute 表存储和数据集成工具,可以实现海量历史数据的存储和批量并行数据分析,目前,在大型互联网企业的数据仓库和 BI 分析、电子商务网站的交易分析、变压器局部放电相位分析^[13]等领域得到应用。

该文基于海量的用电数据对用电用户行为分析和用户分类开展研究,目标是通过聚类的手段对用户的用电类型进行划分,并在此基础上对相同用电类型的用户进行细分,形成划分群体,从而帮助深入了解用户的用电特性以及群体用电共性,为电网公司制定不同的服务水平和电力需求响应措施奠定基础。针对海量用电信息处理的性能问题,该文提出基于 MaxCompute 表存储模式的用电数据分区存储方法,并应用 Graph 框架设计实现并行化的 k 均值聚类算法,实现海量用电数据的快速聚类划分。

1 用电数据聚类分析整体流程

该文数据分析的目标是基于用户的用电数据和选取的特征,进行两阶段的并行化的聚类分析,首先聚类分析出用户的用电类型,然后,对同一用电类型用户再次进行聚类,细分出精细的用户群体,以支持精细化管理。为了实现分析目标,将阿里云上的 MaxCompute 与 k 均值聚类算法相结合,提出一种阿里云数加平台上的用电数据分析框架,用于完成从用电数据采集、分布式存储、特征选取、并行聚类分析整个过程。通过将聚类结果和用电用户实际类型的对比验证所设计聚类算法的有效性。

用电数据聚类分析的过程主要包括:确定分析目标、数据收集、特征工程、聚类分析、分析结果展示

等环节。数据收集环节主要从用电信息采集系统导出数据,并使用阿里云数据集成工具进行数据同步,将数据导入 MaxCompute 表存储。特征工程环节主要完成缺失值填充、数据转换、归一化、特征抽取等数据处理工作。特征工程之后便获得了可以用于聚类分析的有效特征,并基于 Graph 框架实现高效的聚类分析。分析结果可以使用统计图表或者可视化的形式进行展示。用电数据聚类分析整体流程如图 1 所示。

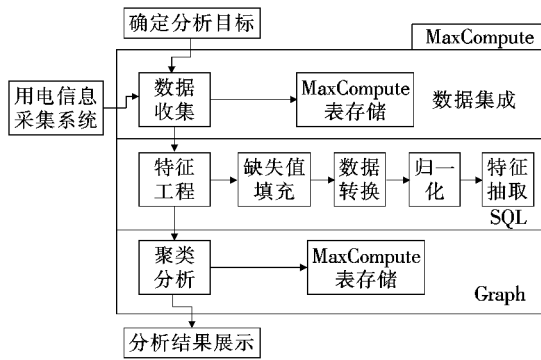


图 1 用电数据聚类分析整体流程

Figure 1 The power consumption analysis flow chart

在图 1 中,原始的用电数据从用电信息采集系统导出后,以文本文件方式存储于本地文件系统。为了将数据上传至 MaxCompute,需要根据设计好的 MaxCompute 表结构进行格式转换,以保证和表结构相匹配。

创建表的操作使用 SQL DDL 语句完成,使用 Tunnel 工具进行数据上传。基于 Graph 框架设计实现并行化的 k -means 算法。程序 jar 包以资源(Resource)的形式上传至 MaxCompute,并使用 jar 命令执行算法。算法的输入和输出均使用 MaxCompute 表。

2 用电数据特征工程和 MaxCompute 表存储设计

2.1 用电数据的特征选择

选取用电信息数据中的三相电压、三相电流、三相功率因素以及有功功率、无功功率、用电量构建多维用电数据特征。以天为时间单位,进行电压电流有效值计算:

$$I_e = \sqrt{\frac{1}{T} \int_0^T i(t)^2 dt} = \sqrt{\frac{II^T}{N}} \quad (1)$$

式中 T 为采集周期,这里使用 1 d; $i(t)$ 、 I 均为采样值; N 为一个采集周期内的采样点数量。

2.2 数据预处理

通过对原始采集数据的观测发现,存在多种类型的数据缺失情况,需要进行针对性的处理。

1) 整体数据缺失。部分采集数据仅包含电表资产号、采集时间等信息,用于聚类的特征值整体缺失。对这些数据执行删除操作。

2) 部分特征值缺失。该文对此类情况使用众数对缺失值进行填充。通过 MaxCompute MapReduce 设计实现并行化的众数计算算法,以资源的方式保存在 MaxCompute 中,以供数据预处理时调用。

在完成缺失值处理之后,对特征量进行归一化处理,以消除特征的量纲对聚类算法性能的影响。该文使用 min-max 归一化方法,即

$$x^* = \frac{x - \min}{\max - \min} \quad (2)$$

2.3 分区存储设计

为了提升数据分析时的检索速度,设计二级分区表结构来存储用户用电数据,如图 2 所示,一级分区列是“用户类型”,二级分区是“数据时间”。采用分区列之后,在以用户类型或时间范围作为检索条件时,可以有效提升数据分析的性能。使用 SQL DDL 实现数据表的创建,如表 1 所示。

在表 1 中,创建了包含 13 个字段的表 PCD,其中第 1 个字段 ID 表示表计编号。CA、CB、CC 分别表示 A、B、C 三相电流; VA、VB、VC 分别表示 A、B、C 三相电压; FA、FB、FC 分别表示 A、B、C 三相

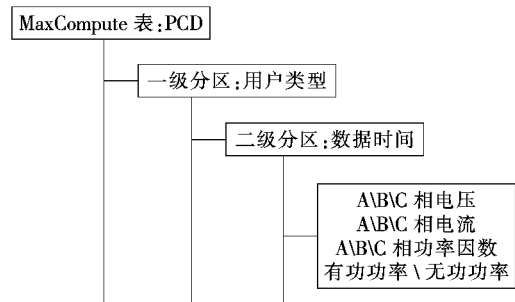


图 2 用电数据二级分区表存储模式

Figure 2 Two-level partition storage table for power consumption data

功率因数;AP、RP 分别表示有功、无功功率;EC 表示用电量;UT 为一级分区列,代表用户类型;Date 为二级分区列,代表时间。

表 1 3 维重构向量表 DDL

Table 1 DDL for 3-dimensional reconstruction vector table

```
Create Table if not exist PCD( // power consumption data table
ID String,
CA double, CB double, CC double,
VA double, VB double, VC double,
FA double, FB double, FC double,
AP double,
RP double
EC double
) partitionedby (
UT String, Date String) life cycle n;
```

3 基于 Graph 的并行化聚类分析

3.1 用电数据的聚类分析

用电信息采集系统收集的数据仅包含用户大类的标签信息,对同一大类的用户用电数据不具备标签,因此,适合采用聚类的方法进行分析。通过将用电行为数据构建多维特征并进行聚类,可以对大类内部的用户进行细分,进一步获取用户的用电特性以及群体用电共性,有助于实现用电用户细分,提升电网公司服务广度和深度,为新型智能用电解决方案奠定基础。

k 均值算法是基于距离的硬聚类算法,具有收敛速度快、算法简单的特点,广泛应用于各种聚类分析场景。它利用最近邻质心决策规则将数据划分成为 k 个簇,并迭代计算每个簇的质心。 k 均值算法可简要描述:①选择 k 个点作为初始质心;②进行迭代计算,将每个点指派到最近的质心,形成 k 个簇,重新计算每个簇的质心,直到质心不再发生变化或者满足终止条件;③输出聚类结果。其中,点之间的相似性度量采用欧氏距离。但传统聚类算法在处理海量的用电信息数据时,存在存储容量和计算性能的问题。文献[14]研究了基于 MapReduce 的并行化 k 均值聚类算法,并在自建的 Hadoop 平台上进行了分析实验。但从 k 均值聚类算法的过程可知,算法的核心是通过多轮的迭代计算完成的,MapReduce 则主要采用“数据并行”的方式,与 k 均值

算法过程并不十分吻合。相较而言,MaxCompute Graph 是一种基于图的迭代计算框架,更适合包含大量迭代计算的应用场景。该文应用 MaxCompute Graph 设计实现了海量用电数据的聚类分析。

3.2 MaxCompute Graph

MaxCompute Graph 是一种面向图的并行迭代计算框架,使用图对分布式内存中的数据进行组织,将数据分布式存储在多个计算节点中,通过消息通信,全局聚合等机制实现节点通信,协调完成迭代计算。

MaxComputeGraph 使用顶点(Vertex)和边(Edge)组织数据。Vertex 是一个四元组结构,可表示为 $\langle ID, Value, Halted, Edges \rangle$ 。其中, ID 是标识符, Value 表示顶点的值; Halted 表示顶点状态,如果值为 false,表示该点不参加本轮的迭代计算; Edges 表示从该点出发的有向边的集合。Edge 可表示为二元组 $\langle DestVertexID, Value \rangle$,其中, DestVertexID 表示该边指向的目标顶点 ID; Value 则表示边的权值。

3.3 基于 Graph 的并行化 k 均值算法

k 均值算法过程是反复的迭代计算过程,直至达到收敛评判准则为止,相对 MapReduce,更适合采用图模型 Graph 实现。Graph 计算作业使用图对数据进行建模,从 MaxCompute 表中读取输入数据,在内存中使用图 Vertex 和 Edge 对数据进行建模描述,借助节点间通信实现状态更新,使用超步(super step)描述迭代计算,并将节点分布在多个计算节点上完成并行的迭代计算。基于 Graph 设计实现了并行化的 k 均值算法,算法描述如下:

1) 算法输入。MaxCompute 数据表 PCD; 初始质心点资源文件; 簇的数目 $k \in N$ 。

2) 图加载。Graph 框架调用自定义 Graph-Loader 加载输入表记录,解析为 Vertex 和 Edge。使用记录的 Recordnum 作为 Vertex 的 ID; 将记录内容转换为 DenseVector 对象封装在 Vertex 的 Value 中。每条记录对应一个 Vertex,多个 Vertex 会被分配给到 MaxCompute 集群中的多个工作节点(Worker)上,以支持后续的并行化聚类。数据表加载为图的过程如图 3 所示。

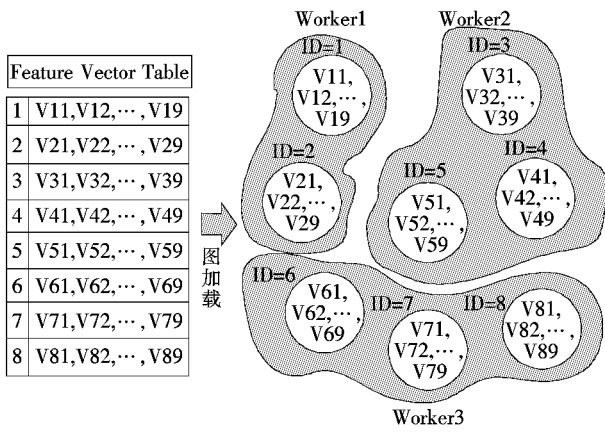


图 3 图加载
Figure 3 Graph load

3)初始化。创建并初始化 AggregatorValue, 用于存储迭代计算过程数据,包括 k 个质心点,最近邻点。AggregatorValue 用于存储迭代计算的局部聚合结果。

4)迭代计算。worker 执行 createInitialValue, 调用 context.getLastAggregatedValue 方法获取上一轮迭代的质心计算结果来初始化本轮迭代计算质心,最邻近点集合初始化为空集。每个节点在各自的 worker 上并行执行 Compute 方法,完成本轮迭代计算。调用 context.aggregate 方法来执行局部迭代逻辑:计算本 Worker 上节点与质心的距离,判别所属类别,并更新 AggregatorValue。各 Worker 将局部迭代结果发送给 AggregatorOwner 汇总,计算本轮质心,判断是否满足迭代终止条件。如果终止,将质心写入 MaxCompute 表;否则进入下一轮迭代。

上述算法执行过程是在集群环境下以分布式并行方式完成的,除了节点内部的并行计算,算法使用了用于收集全局信息的 Aggregator 机制,收集每个 Worker 的局部计算结果,并汇总,从而可以判断是否满足收敛准则。处理的流程和涉及的组件如图 4 所示。

在图 4 中,计算逻辑包括两部分:一部分是分布式执行,在所有 Worker 上执行局部汇总,另一部分在 AggregatorOwner 所在的 Worker 上执行,完成总体汇总,计算质心,并判断迭代终止条件。

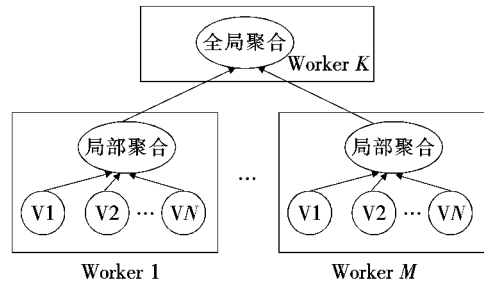


图 4 基于 Graph 的并行化 k 均值聚类算法组件和流程
Figure 4 Graph-based parallel k -means clustering algorithm components and processes

3.4 两阶段用电用户聚类划分

应用所设计的 Graph 框架下的并行化 k 均值聚类算法,对用户用电数据进行聚类分析,分析过程分为 2 个阶段:①基于用电数据特征表中电压、电流、功率因数等特征进行聚类,识别用户的用电类型,完成初步划分;②对于同类用电类型,根据用户的用电量信息进行再次聚类,实现用户细分。使用任务管道连接这 2 个阶段,按顺序完成分析。用电类型和用户类型信息如表 2 所示。

表 2 用电和用户类型

Table 2 Power consumption and user type	
用电类型	用户类型
农业排灌	大型专变用户
农业生产用电	中小型专变用户
商业用电	三项一般工商用户
中小学教学用电	居民用户
普通工业用电	小水电上网用户
居民生活用电	居民用户
一般工商用电	三项一般工商用户
大工业用电	大型专变用户

4 实验结果

4.1 实验环境

该文聚类分析实验在阿里云数加平台上完成。数加平台是一个用于大数据分析的多功能平台,其基础的计算引擎是 MaxCompute。在实验过程中,采用 MaxCompute SQL DDL 完成数据表的创建,数加采集和同步使用数据集成工具,基于 MaxCompute Graph 框架完成并行聚类分析,使用 Max-

Compute SQL DML 完成特征工程。使用组件的层次关系如图 5 所示,MaxCompute 表是数据存储容器,配置文件、程序代码等作为 MaxCompute 资源。Graph 是基于图模型的并行迭代计算框架,SQL DDL 组件用于实现数据定义,SQL DML 组件用于实现数据 SQL 操作,数据集成工具用于实现数据同步、导入导出。

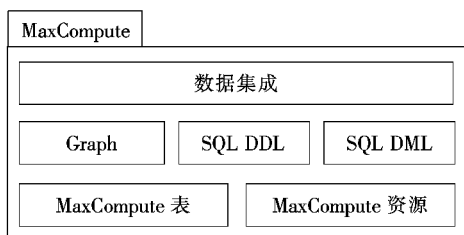


图 5 MaxCompute 功能组件

Figure 5 MaxCompute components

4.2 分析结果

该文实验数据来源于某省电网公司的用电信息采集系统,涵盖 8 种用电类型和 5 种用户类型,从 2017 年 4 月至 7 月,共 16 万条数据,用于聚类分析。对采集的数据使用所设计的算法进行聚类分析,验证聚类分析的正确率和聚类性能。

实验 1 聚类效果验证。

使用用电信息采集系统的多种用电类型数据,应用所设计的基于 Graph 的 k 均值聚类算法,对用

电数据进行聚类分析,验证聚类分析的准确率。聚类分析模型和参数如表 3 所示,initCenterMethod 代表初始质心的来源,这里是通过制定外部表或者资源文件的方式提供初始质心;Loop 代表最大迭代次数,设定为 100;聚类数量 k 取 4;采用欧氏距离进行距离度量;accuracy 取 0.1。

第 1 阶段用电类型的聚类分析结果以及各用电类型的聚类分析正确率曲线如图 6 所示,分别统计了各用电类型的实际样本数和聚类样本数。

从图 6 的聚类分析结果可以看出,该文所设计的并行聚类算法可以正确划分出大多数的用电数据。用电类型的总体聚类正确率达到 88.3%。引起聚类错误的一个因素是采集数据中存在一定的数据值缺失,该文使用取众数的方式对缺失值进行填充,引入了一定的聚类误差。

在完成用电类型聚类划分后,第 2 阶段针对每个用电类型大类进行再次聚类分析,统计每个用电类型下用户类型被正确聚类的样本数量,计算每个用电类型下的用户类型聚类正确率,如图 7 所示。

实验 2 聚类性能验证。

将基于 Graph 框架的 k 均值算法与传统的 k 均值算法的执行性能进行对比分析,以验证所设计算法是否可以有效用于大规模用电数据分析。通过不断增加输入数据集的规模,验证并行算法的聚类性能。实验结果如图 8 所示。

表 3 聚类分析模型

Table 3 Clustering analysis model

模型名称	算法名称	特征属性	参数
k 均值聚类	基于 Graph 的 k 均值聚类	ca,cb,cc,va,vb,vc,fa,fb,fc,ap,rp	enableSparse: false;initCenterMethod: external;loop: 100;centerCount: 4;distanceType: euclidean;accuracy: 0.1

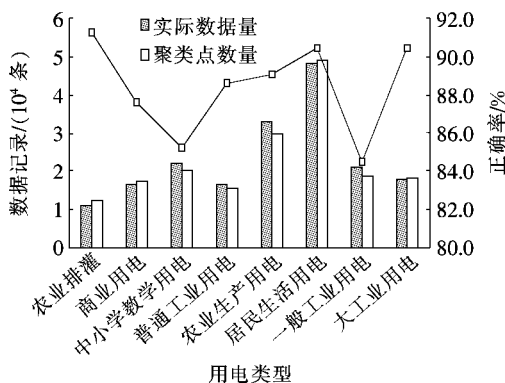


图 6 用电类型聚类分析结果

Figure 6 Clustering analysis of power consumption data

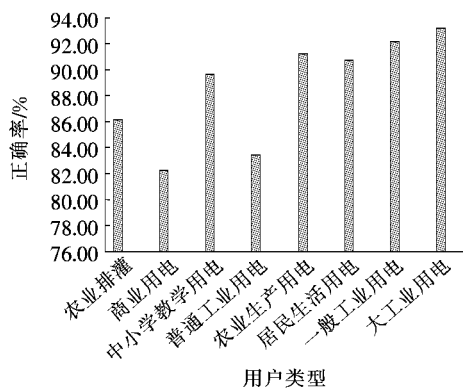


图 7 用户类型聚类分析正确率

Figure 7 Clustering analysis of power consumption user

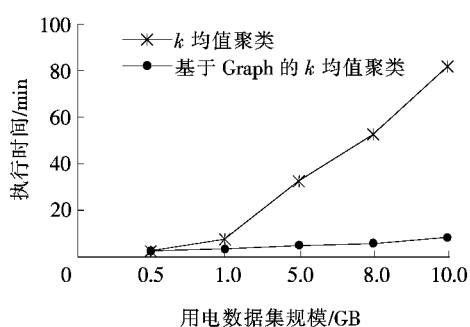


图 8 算法性能对比

Figure 8 Algorithm performance comparison

从图 8 中可以看出, k 均值聚类分析过程包含大量的迭代计算, 过程复杂, 数据规模较大情况下, 执行时间在分钟级。在单机环境下, 算法执行时间随着数据集规模增长而快速增长, 不能有效应对大数据分析; 基于 Graph 的 k 均值算法执行在 Max-Compute 平台上, 算法总体执行时间变化波动很小。从任务的 logview 日志中可以看出, 随着数据规模的增加, 并行分析任务的并行粒度也在不断加大, 平台为分析任务分配的硬件资源(CPU 核心数、内存)接近线性增长, 因此保证了算法执行时间总体非常平稳。

5 结语

智能电网中配用电环节收集的监测数据体量大, 需要大数据技术进行数据分析。该文研究了基于阿里云 MaxCompute 的用电数据分析方法, 在数加平台上设计实现了数据收集、数据预处理、特征工程、聚类分析的整体过程。基于 Graph 并行迭代计算框架, 设计实现了 k 均值聚类分析, 用于大规模用电数据的聚类。基于某省用电信息采集系统的实测用电数据进行验证, 分别考察了各用电类型用户样本数据是否能够正确聚类致相应的簇。第 1、2 阶段总体聚类正确率分别达到 88.3%、90.03%。在计算性能方面, 所设计的聚类分析算法执行时间保持平稳趋势, 当输入数据规模从 0.5 增长至 10 GB 时, 执行时间仅增长 2.45 倍, 运行时间始终控制在 10 min 之内, 从而可以有效应对用电大数据分析的挑战。

出于数据安全性的考虑, 可将该文设计的聚类分析框架和实现部署在电力系统专有云上, 实现高效、安全的用电数据分析。

参考文献:

- [1] 孙毅, 黄婷, 崔晓昱, 等. 互联网视角下的泛在电力物联网[J]. 电力科学与技术学报, 2019, 34(4): 3-12.
SUN Yi, HUANG Ting, CUI Xiaoyu, et al. Ubiquitous power internet of things from the perspective of internet [J]. Journal of Electric Power Science and Technology, 2019, 34(4): 3-12.
- [2] 何永秀, 王冰, 熊威, 等. 基于模糊综合评价的居民智能用电行为分析与互动机制设计[J]. 电网技术, 2012, 36(10): 247-252.
HE Yongxiu, WANG Bing, XIONG Wei, et al. Analysis of residents' smart electricity consumption behavior based on fuzzy synthetic evaluation and the design of interactive mechanism[J]. Power System Technology, 2012, 36(10): 247-252.
- [3] 陈鹏伟, 陶顺, 肖湘宁, 等. 短时间尺度用电行为相关性分析网络模型[J]. 电力系统自动化, 2017, 41(3): 61-69.
CHEN Pengwei, TAO Shun, XIAO Xiangning, et al. Network model for correlation analysis of short-term electricity consumption behavior[J]. Automation of Electric Power System, 2017, 41(3): 61-69.
- [4] 陆俊, 朱炎平, 彭文昊, 等. 计及用电行为聚类的智能小区互动化需求响应方法[J]. 电力系统自动化, 2017, 41(17): 113-120.
LU Jun, ZHU Yanping, PENG Wenhao, et al. Interactive demand response method of smart community considering clustering of electricity consumption behavior [J]. Automation of Electric Power System, 2017, 41(17): 113-120.
- [5] 徐杰彦, 许雯旻, 褚渊, 等. 区域尺度住宅建筑日用电负荷模型构建方法研究[J]. 中国电力, 2020, 53(8): 29-39.
XU Jieyan, XU Wenyang, CHU Yuan, et al. Residential electricity load model construction in district scale [J]. Electric Power, 2020, 53(8): 29-39.
- [6] 王毅, 张宁, 康重庆, 等. 电力用户行为模型: 基本概念与

- 研究框架[J]. 电工技术学报, 2019, 34(10): 2056-2068.
- WANG Yi, ZHANG Ning, KANG Chongqing, et al. Electrical consumer behavior model: Basic concept and research framework[J]. Transactions of China Electrotechnical Society, 2019, 34(10): 2056-2068.
- [7] 宋亚奇, 周国亮, 朱永利, 等. 云平台下并行总体经验模态分解局部放电信号去噪方法[J]. 电工技术学报, 2015, 30(18): 213-222.
- SONG Yaqi, ZHOU Guoliang, ZHU Yongli, et al. Research on parallel ensemble empirical mode decomposition denoising method for partial discharge signals based on cloud platform[J]. Transactions of China Electrotechnical Society, 2015, 30(18): 213-222.
- [8] 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.
- ZHANG Suxiang, LIU Jianming, ZHAO Bingzhen, et al. Cloud computing-based analysis on residential electricity consumption behavior[J]. Power System Technology, 2013, 37(6): 1542-1546.
- [9] 葛磊蛟, 王守相, 王尧, 等. 多源异构的智能配用电数据存储处理技术[J]. 电工技术学报, 2015, 30(Z2): 159-168.
- GE Leijiao, WANG Shouxiang, WANG Yao, et al. Storage and processing technology of the multi-source isomerized data for smart power distribution and utilization[J]. Transactions of China Electrotechnical Society, 2015, 30(S2): 159-168.
- [10] Bryant R E. Graph-based algorithms for boolean function manipulation [J]. IEEE Transactions Comput, 2006, C-35(8): 677-691.
- [11] Mitliagkas I, Borokhovich M, Dimakis A G, et al. Frog-Wild!-Fast pagerank approximations on graph engines [J]. Computer Science, 2015, 8(8): 874-885.
- [12] Asma K, Abdulsahib. Anomaly detection in text data that represented as a graph using DBSCAN algorithm [J]. Journal of Theoretical & Applied Information Technology, 2017, 95(9): 2096-2101.
- [13] 朱永利, 李莉, 宋亚奇, 等. ODPS 平台下的电力设备监测大数据存储与并行处理方法[J]. 电工技术学报, 2017, 32(9): 199-210.
- ZHU Yongli, LI Li, SONG Yaqi, et al. Storage and parallel processing of big data of power equipment condition monitoring on ODPS platform[J]. Transactions of China Electrotechnical Society, 2017, 32(9): 199-210.
- [14] 刘晓悦, 郭强. 海量用电数据并行聚类分析[J]. 辽宁工程技术大学学报: 自然科学版, 2016, 35(1): 76-80.
- LIU Xiaoyue, GUO Qiang. Cloud computing based cluster analysis on data of power utilization[J]. Journal of Liaoning Technical University: Natural Science, 2016, 35(1): 76-80.