

电动汽车领域的大数据处理与分析

为推动电动汽车这一新兴行业的发展,需要对其行业特点进行深入研究,通过运用大数据技术对电动汽车产业的运行数据、运营数据进行数据分析处理,揭示电动汽车产业的特性。

潘鸣宇 张 禄 孙 舟 王伟贤 郑静楠/国网北京市电力公司

0 引言

随着电动汽车产业的不断发展以及相关信息系统的不断建设,电动汽车运行数据和运营数据不断增加,如何有效地集中存储分析这些数据,形成数据集中管理、资源统一共享的应用模式,并从中挖掘数据的潜在价值,是目前市场所需。

本文以电动汽车运行数据和运营数据分析处理方式的研究与实现为目的,改善现有数据分析处理方式,便于对充电集中时间、充电电量增量情况和充电电度变化等情况信息的统计、后续充电设施建设指导以及数据挖掘分析等提供一种基于电动汽车运行数据和运营数据的分析处理方法,该方法可利用基于列数据库的Hbase存储技术实现对数据的存储,再基于Hadoop提供的一个数据仓库工具Hive搭建数据仓库,对各个数据系统的存储数据进行整合,然后利用MapReduce、流计算等处理技术进行数据处理,在前端可视化进行多维分析、报表以及数据挖掘等方式为用户展示所需信息,大大提高了数据处理效率,便于数据分析。从而避免了传统表处理的烦琐操作,提高了工作效率并降低了出错率。

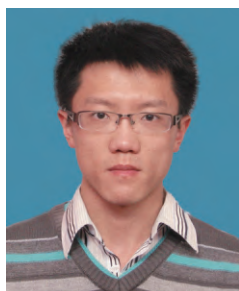
随着电动汽车相关信息系统的不断建设,信息化已经全面融入电动汽车产业的各个方面,积累了大量结构化数据、非结构化数据以及海量历史准实时数据等。本文将通过对电动汽车运营监控系统、充换电站级监控系统和充电桩采集系统等多个电动汽车相关信息系统的研究与分析,筛选出具备分析价值的

1 电动汽车领域的大数据

1.1 电动汽车领域大数据的产生

电动汽车领域的的数据大致可分为两类:一类是电动汽车及相关设备的运行和设备检测数据;另一类是电动汽车产业运营数据。根据数据的内在结构,这些数据可以进一步细分为结构化数据和非结构化数据。结构化数据主要包括存储在关系数据库中的数据,目前电动汽车相关信息系统中的大部分数据是这种形式,随着电动汽车产业的不断发展,这部分数据增长很快。相对于结构化数据而言,不方便使用数据库二维逻辑表来表现的数据即称为非结构化数据,主要包括视频监控、图形图像处理等产生的数据。在电动汽车相关信息系统中,非结构化数据也占到了很大的比重,主要体现在电动汽车充换电站监控系统中。

结构化数据根据处理时限要求又可以划分为实时数据和准实时数据,比如电动汽车充电桩远程控制、充电预约和无卡充电等数据是实时数据,需要快



潘鸣宇/高级工程师

关键词/Keywords

电动汽车·
充换电设施·
大数据·

速、准确地处理；而大量的状态监测数据对实时性要求相对较低，可以作为准实时数据处理。

1.2 电动汽车领域大数据的特点

电动汽车领域的大数据具备如下特征：

1) 数据规模巨大。仅以北京市为例，每天有上千辆电动汽车需要充电，近3 000台充电桩进行实时状态采集，日增长数据量达到太字节级。

2) 数据类型繁多。电动汽车领域数据分布广、种类繁多，包括实时数据、历史数据、文本数据、多媒体数据和时间序列数据等各类结构化、半结构化数据以及非结构化数据，各类数据查询与处理的频度和性能要求也不尽相同。

3) 价值密度低。以视频为例，连续不间断监控过程中，可能有用的数据仅仅有1~2 s。在电动汽车充电站监控中存在同样问题，所采集的绝大部分数据都是正常数据，只有极少量的异常数据，而异常数据是状态检修的最重要依据。

4) 处理速度快。在几分之一秒内对大量数据进行分析，以支持决策制定。对在线状态数据的处理性能要求远高于离线数据。这种在线的流数据分析与挖掘同传统数据挖掘技术有本质的不同。

可以考虑为各类电动汽车相关数据引入一个新的属性，即数据的真实性。数据的真实性是指与特定类型数据相关的可靠性级别。高质量数据对于数据分析结果的正确性有重要影响，可通过数据筛选提高数据的真实性。

2 大数据关键技术

2.1 数据迁移技术

数据迁移技术可以实现从传统的数据库向分布式的非结构化数据库传递数据，并能够保证数据传输的可靠性、容错性和实时性，且提高数据传输的效率。在数据传输过程中，可以对数据进行简单的处理、过滤，提取有价值、可用数据，减少数据存储所需资源，提高后期数据分析的准确性。

2.2 数据分布存储技术

数据分布存储技术是基于分布式文件系统的非结构化数据存储，它可实现高并发的数据处

理，确保数据存储的高效率、高安全性和高可靠性，并能最大程度地均衡利用所有系统资源。数据分布存储可采用下列方式实现：

1) 分布式文件系统。将数据存储 in 分布式文件系统中，可以保证数据的安全性，并能够在大数据量的情况下采用分布式处理思想，多节点同时处理数据。

2) 非结构化数据库。基于分布式文件系统的非结构化数据库，可以保证高可靠性、高性能、面向列和可伸缩的特性，提供数据分析挖掘所必须的数据安全、高效率数据访问。

2.3 数据挖掘技术

数据挖掘技术实现针对数据的可靠性分析并进行挖掘处理，获取有价值的信息、事物间的关联和隐藏的信息等。数据挖掘技术可以分为以下几个过程：

1) 历史数据分析。采用已经在各案例中被成功使用并得到广泛认可的工具，针对大量数据进行处理，获取其中隐藏的信息，得到可靠的预期结果，并进行后期的数据处理。

2) 数据预测。根据已经获取的可用规律，通过当前所获取的数据，对未来进行可靠性的预测，实现数据挖掘分析的价值。

2.4 数据可视化技术

面对海量的数据，其中包含高准确度和高分辨率数据、时变数据以及多变量数据等，如何在有限的屏幕空间下，以一种直观、容易理解的方式展现给用户，是一项非常有挑战性的工作。数据可视化技术运用计算机图形学和图像处理技术，将数据绘制成高准确度、高分辨率的图像，并提供交互工具，有效利用人的视觉系统，允许实现改变数据处理和算法参数，对数据进行观察及定性和定量分析。通过此操作，可以直观形象地展示抽象的数据，在实现技术价值的同时以艺术的效果展示。

3 电动汽车数据分析处理

本文对电动汽车运行数据和运营数据的分析处理方式进行研究，改善现有数据分析处理方

式, 便于对充电集中时间、充电电量增量情况和充电电度变化情况等信息的统计、后续充电设施建设指导以及数据挖掘分析等提供一种基于电动汽车运行数据和运营数据的分析处理方法, 其技术架构如图1所示。

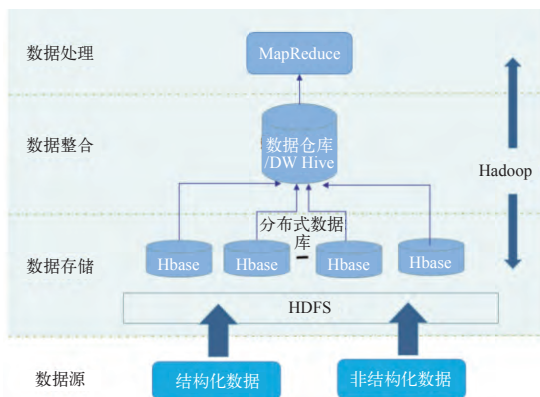


图1 技术架构

本文利用目前流行的Hadoop框架体系，其核心技术包括分布式文件系统（HDFS）、分布式数据库（Hbase）、数据仓库工具（Hive）和处理过程（MapReduce）等。

电动汽车充电运行数据和运营数据分析处理分为数据存储、数据整合、数据分析和数据分析四个环节。

3.1 数据存储

电动汽车运行数据和运营数据包括结构化数据和非结构化数据。结构化数据包括充电桩状态代码、充电电压、充电电流、充电负荷、充电电量、充电时间和充电金额等；非结构化数据包括充电日志、充电视频等文件。Hadoop中的Hbase是一个分布式、面向列的开源数据库，它不同于一般的关系数据库，是一个适合于非结构化数据存储的数据库。因此可将电动汽车的非结构化数据存储于Hbase数据库中，由于该数据库是基于列存储，有利于对数据库进行高效的压缩而减少数据规模，因此有利于电动汽车交易数据存储。

3.2 数据整合

由于电动汽车运行数据和运营数据来源比较分散且数据类型不统一，这就需要对数据进行整合。电动汽车数据整合技术利用基于Hadoop的数

据仓库工具Hive搭建数据仓库，将Hbase数据库的数据全部存储至该数据仓库。该数据仓库存储方式是将结构化的数据文件映射为一张数据库表，并提供类SQL语言，实现完整的SQL查询功能。可以将SQL语句转换为MapReduce任务运行，十分适合数据仓库的统计分析。

3.3 数据处理

由于充电桩时钟跳变、异常充电和窃电等情况,造成电动汽车运行数据和运营数据存在异常,在对电动汽车数据进行挖掘分析前,需先进行数据处理,确保数据的准确性。

使用Hadoop的 MapReduce并行数据处理框架，MapReduce是一种简化的分布式编程模式，将复杂分布式处理过程分解为Map和Reduce过程，提高并发处理能力。让程序自动分布到一个由普通机器组成的超大集群上并发执行。它足够快速的批处理分析仪可以满足业务需求。

MapReduce致力于解决大规模数据处理的问题，因此在设计之初就考虑了数据的局部性原理，利用局部性原理将整个问题分而治之。MapReduce集群由普通PC构成，为无共享式架构。在处理之前，将数据集分布至各个节点，如图2所示。处理时，每个节点就近读取本地存储的数据处理（Map），将处理后的数据进行合并（Combine）、排序（Shuffle and Sort）后再分发（至Reduce节点），避免了大量数据的传输，提高了处理效率。无共享式架构的另一个好处是配合复制（Replication）策略，集群可以具有良好的容错性，一部分节点的Down机对集群的正常工作不会造成影响。它是一个使用简易的软件框架，

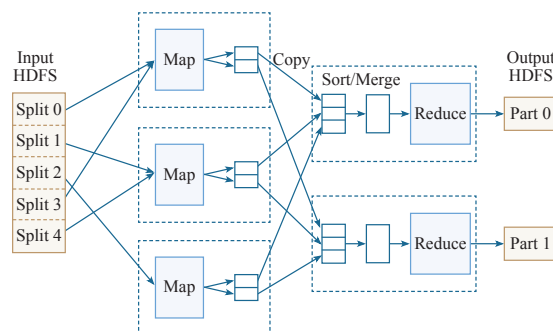


图2 无共享式架构

基于它写出来的应用程序能够运行在由上千个商用机器组成的大型集群上,并以一种可靠容错的方式并行处理上太字节级别的数据集。

3.4 数据分析

目前,北京市电动汽车智能充换电服务网络运营管理系统已积累了大量的电动汽车用户运行数据和运营数据,通过运用本文中提到的大数据分析处理技术,实现对电动汽车充电站使用率、充电站月均负荷和充电站充电时间等数据的分析,持续跟踪观察电动汽车用户充电特征状况,探索电动汽车用户充电行为习惯,研究电动汽车用户充电消费模式,从而更好地开展电动汽车充电业务,提高用户服务的效率和质量。

4 大数据在电动汽车领域的应用

4.1 电动汽车及相关设备的运行和检测数据分析

(1) 充电桩使用率分析

衡量经济效益的标准是单位投入的产出,通常情况下,充电设施的使用率越高,其经济效益越显著。

令 $N_{\text{桩}}$ 表示充电桩的使用率,则其计算公式为

$$N_{\text{桩}} = (T_{\text{桩}} / T_{\text{All}}) \times 100\% \quad (1)$$

式中, $T_{\text{桩}}$ 为在某个时间段内,该充电桩充电时长; T_{All} 为在某个时间段内的时长。

举例说明:统计某一天内某站的充电桩使用率,假设该站有充电桩1和充电桩2,通过式(1)可分别计算使用率。 T_{All} 为1天 $\times 24$ h(以小时计); $T_{\text{桩}}$ 为该天内,该充电桩充电时长,即该充电桩该天每条交易记录每条结束时间减开始时间的和(以小时计),即

$$T_{\text{桩}} = \sum_{i=0}^{A-1} (T_{\text{结束}A} - T_{\text{开始}A}) \times 24 \text{ h} \quad (2)$$

式中, $T_{\text{结束}A}$ 为充电桩第 A 次充电时的充电结束时间; $T_{\text{开始}A}$ 为充电桩第 A 次充电时的充电开始时间。

若计算该充电站的总使用率,表达式为

$$N_{\text{站}} = \left(\sum_{i=0}^{B-1} T_{\text{桩}B} / T_{\text{All}} \right) \times 100\% \quad (3)$$

式中, $\sum_{i=0}^{B-1} T_{\text{桩}B}$ 为该站内所有桩充电时长的和; T_{All}

为在某个时间段内的时长; B 为该站的充电桩数。

充电站使用率分析如图3所示。通过对北京市9个区充电站的整体充电桩使用情况进行统计分析,发现如下特征:

- 1) 各区域充电站的使用率普遍低于40%。
- 2) 4~6月份和11月~次年2月份的使用率较其他月份明显下降。
- 3) 除房山外,其他地区的充电站使用率变化较为平稳。

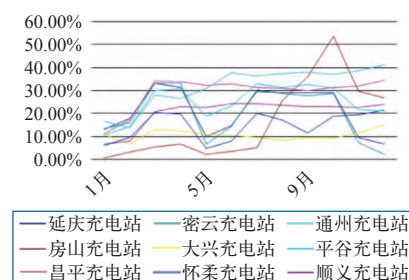


图3 充电站使用率分析结果

(2) 充电站负荷情况分析

为了进一步了解各区充电站内充电桩的使用情况,以月为单位,计算月均负荷。

令 $P_{\text{桩}}$ 表示充电桩单次充电负荷,则其计算公式为

$$P_{\text{桩}} = Q_{\text{桩}} / (T_{\text{结束}} - T_{\text{开始}}) \quad (4)$$

式中, $Q_{\text{桩}}$ 为充电桩充电电量; $T_{\text{结束}}$ 为充电桩充电时的充电结束时间; $T_{\text{开始}}$ 为充电桩充电时的充电开始时间。

若计算充电站月均负荷,表达式为

$$P_{\text{站}} = \sum_{i=0}^{B-1} P_{\text{桩}B} / T_{\text{All}} \quad (5)$$

式中, $\sum_{i=0}^{B-1} P_{\text{桩}B}$ 为该站内所有充电桩该月负荷的和; T_{All} 为该月的总天数。

充电站负荷情况分析结果如图4所示。通过对北京市9个区充电站的月均负荷情况进行统计分析,发现如下特征:

- 1) 充电站负荷曲线较为平稳,整体呈现缓慢上升趋势。
- 2) 充电站负荷曲线变化与充电站使用率基本吻合。

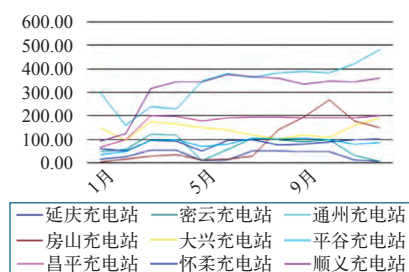


图4 充电站负荷情况分析结果

4.2 电动汽车产业运营数据分析

(1) 充电电量分析

充电电量分析如图5所示。通过对北京市9个区充电站充电交易记录中的充电电量数据进行分析,发现存在如下特征:

- 1) 电动汽车单次充电电量集中在10~20 kW·h范围内。
- 2) 由于充电桩测试及用户异常充电操作等原因,存在单次充电电量为0~1 kW·h的情况。
- 3) 有少数车辆电容量较一般电动出租车大,存在少数单次充电电量达到30 kW·h以上的情况。

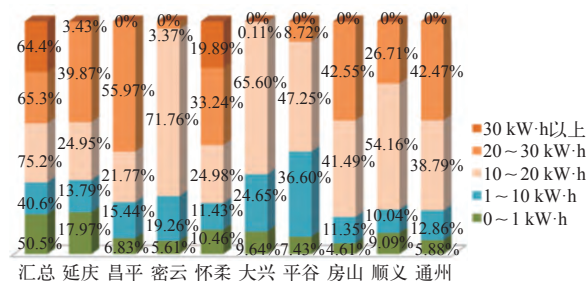


图5 充电电量分析

(2) 充电桩占用时长分析

通过对北京市9个区充电站充电交易记录中的充电占用时长进行统计分析,发现存在如下特征:

- 1) 快充充电桩充电占用时长主要为1 h之内。
- 2) 慢充充电桩充电占用时长主要为3~12 h。
- 3) 由于充电桩测试及用户异常充电操作等原因,存在充电桩占用时长为0~3 h的情况。
- 4) 部分用户存在充电完成后不及时结算拔电的情况,存在充电桩占用时长超12 h的充电记录。

(3) 充电时间特征

通过对电动汽车用户充电时间进行统计分析,了解电动汽车用户的充电行为规律,结合各

区域充电站的布局分布提出改进建议,充电时间特征如图6所示。

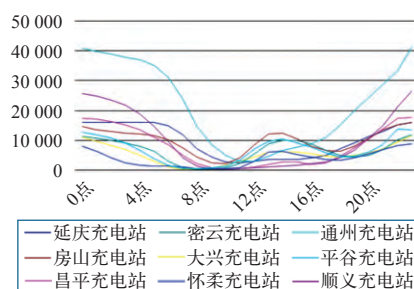


图6 充电时间特征

将1天24 h按整点分为24个节点,通过电动汽车用户每次充电时的充电开始时间与结束时间进行统计。

令 C_n 表示统计时间段内 n 点时该桩的总充电次数,则其计算公式为

$$C_n = \sum_{i=0}^{A-1} T_A \quad (6)$$

式中, T_A 为 n 点时第 A 条充电记录是否处于充电状态,若正在充电,则计为1,否则记为0。

若计算充电站的充电时间,表达式为

$$T_{\text{站}} = \sum_{i=0}^{B-1} C_{nB} \quad (7)$$

式中, C_{nB} 为 n 点时第 B 个充电桩的充电次数。

运用式(7)计算充电站1天24个整点的充电次数,并通过折线图的形式进行展现。

通过对北京市9个区充电站每日整点处在充电状态的交易数据进行统计分析,发现存在以下特征:

- 1) 电动汽车用户主要集中在晚上对电动汽车进行充电。
- 2) 部分电动汽车用户选择下午对电动汽车充一次电。

5 结束语

随着电动汽车产业的不断发展、充电设施的不断建设以及电动汽车业务的不断推广,电动汽车的相关数据量将呈现迅速上升的态势。一般的数据分析已无法满足在海量电动汽车数据上的应用
(下转第33页)

可知, 输出电流良好地跟踪给定无功参考电流, 直流侧电容电压基本稳定在设定值0.82 kV处, 且波动范围在允许的 $\pm 10\%$ 以内。

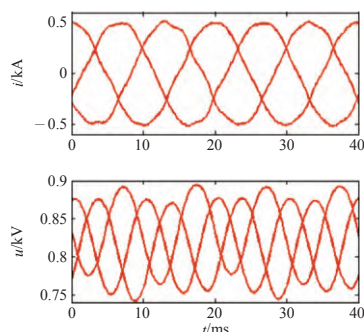


图7 三相输出电流及直流侧电容电压波形

6 结束语

本文提出了在三相静止坐标系下的准比例谐振控制方法。该方法无须进行复杂的坐标旋转变换, 实现了STATCOM无功补偿及滤波功能。仿真结果和工程实例都验证了该方法的有效性和可行性。

参考文献

- [1] 郑重, 杨耕, 耿华. 配置STATCOM的DFIG风电场在不对称电网故障下的控制策略[J]. 中国电机工程学报, 2013(19): 27-38, 22.
- [2] 田桂珍, 王生铁, 刘广忱, 等. 采用STATCOM改善笼型机风电场性能的机理分析与实验研究[J].

(上接第29页)

用, 如何对多源异构的电动汽车数据进行挖掘分析, 寻找其中的潜在价值引人深思。大数据技术在电动汽车数据领域的运用将势在必行。

电动汽车产业进入大数据移动互联网时代, 应该用大数据的思维观念来处理数据, 挖掘数据的潜在价值。电动汽车作为通信、计算机、电力电子、动力控制和驱动技术以及新材料技术等交通运输领域集成应用的产物, 也为云计算、大数据和智能终端等新技术提供了率先应用的环境。大数据的开发应用以及互联网思维的充分运用, 将会推动电动汽车产业更加快速发展, 创造美好的环境以及美好的生活。

电网技术, 2013(7): 1971-1977.

- [3] 许湘莲. 基于级联多电平逆变器的STATCOM及其控制策略研究[D]. 武汉: 华中科技大学, 2006.
- [4] 张数全. 基于分频控制的并联型APF关键技术研究[D]. 武汉: 华中科技大学, 2011.
- [5] 刘钊, 刘邦银. 系统电压不平衡下链式静止同步补偿器控制研究[J]. 中国电机工程学报, 2011, 31(9): 1-7.
- [6] 杭丽君, 李宾, 黄龙, 等. 一种可再生能源并网逆变器的多谐振PR电流控制技术[J]. 中国电机工程学报, 2012, 32(12): 51-58.
- [7] 赵国生, 路向东, 刘楷. 链式SVG直流侧电压分层控制策略的研究[J]. 电测与仪表, 2014(9): 45-49.
- [8] 盛占石, 蒋栗彦, 虞礼贵. 直接电流控制的静止无功发生器的研究[J]. 电测与仪表, 2009(9): 53-56.
- [9] 黄知超, 杨升振, 延红艳, 等. 基于dq坐标系有功无功电流解耦空间电压矢量的STATCOM控制方法[J]. 电测与仪表, 2012(12).
- [10] 陶礼学, 姚钢, 周荔丹, 等. 基于IGBT的 ± 10 kvar STATCOM装置设计[J]. 电气自动化设备, 2006(5): 61-65.

(收稿日期: 2015-11-25) **EA**

参考文献

- [1] 金澈清, 钱卫宁, 周傲英. 流数据分析与管理综述[J]. 软件学报, 2004, 5(8): 1172-1181.
- [2] Stonebraker M, Abadi D J, Madden S, et al. MapReduce and parallel DBMSS: friends or foes[J]. Communications of the ACM, 2010, 53(1): 64-71.
- [3] Wong P C, Shen H W, Chen C, et al. Top ten interaction challenges in extreme-scale visual analytics[J]. Computer Graphics and Applications, 2012, 32(4): 63-67.

(收稿日期: 2015-08-26) **EA**