

基于大数据分析的综合能源服务平台数据应用研究

文/国网江苏省电力有限公司苏州供电公司 王小蕾

海量用户数据对传统的数据处理方式提出了新的挑战。阐述了综合能源服务平台的数据采集、数据处理及数据分析方法。针对海量用户数据特点,提出应用统计学方法、机器学习等数据挖掘方法开展大数据分析,可为综合能源服务平台背景下海量数据应用提供参考。

海量用户数据对传统的数据处理方式提出了新的挑战。数据采集频次的增加对采集设备的灵敏度、系统的带宽和存储量都提出了更高要求。依托泛在电力物联网技术、智能量测技术的发展,目前智能电能表、低压故障指示器等设备已经可以实现分钟级采集、停电上送等功能。但数据处理仍较为粗放,漏报率和误报率较高,对数据应用的准确度和效率产生了严重影响。本文系统地阐述了综合能源服务平台的数据采集、数据处理及数据分析方法,为海量数据应用提供参考。

综合能源服务平台数据采集

综合能源服务平台数据来源包括用户申报数据和系统采集数据。申报数据由用户录入、平台审核,包括能源供应商注册信息、用户资产信息、能效及排放等数据。采集数据由安装在需要监控的供应商和消费者入网点处的采集装置获取,涉及能源生产、消费和交易等全生命周期数据信息。

平台数据处理

1. 数据处理模式

大数据的处理模式可分为批处理和流处理。批处理是先分块存储数据,然后再分批进行处理,实

现简单,对硬件要求不高。流处理则是直接对数据进行处理,可以实现数据的实时处理,处理结果立即可用,并且随着新数据的抵达会持续更新。

(1) 批处理

批处理的优点是支持对历史时序数据的处理,实现方式便捷,可以在一定程度上降低数据传输过程中的通信成本。但是批处理具有查询数据量大、非实时的缺点。MapReduce是目前最具代表性、应用最广泛的批处理模型,该模型首先对初始数据进行分块存储,然后分配给不同任务区块进行批次处理。由于这种方法严重依赖存储设备,每个任务需要多次执行读写操作,速度相对较慢。另一方面,由于MapReduce可以处理海量的数据集,适合在时效性要求不高的场景下处理大规模数据集。

(2) 流处理

流处理模式具有时效性高、无须查询原始数据和减少磁盘读取次数等优点,在具有实时处理需求的任务场景表现较好。比如,需要实时结算的金融市场、网站流量实时统计及其他基于时序变化的高级应用。由于流处理模式下实时响应速度快,其计算基本在内存中完成,因此对硬件要求较高,内存容量等硬件设施要求较高成为限制该模式发展的原因。

2. 数据清洗

传统的能源系统彼此独立，数据分散在单一的数据管理系统中。随着多源能源系统的融合，综合能源系统数据呈现几何式增长，从单一的结构化数据类型向结构化、半结构化和非结构化融合数据类型转变。用户不同类型的数据在集成过程中会产生大量的冗余数据和干扰信息，为了冗余数据、甄别失真数据和排除干扰信息，数据清洗过程至关重要。一方面，清洗颗粒过细，可能会导致关键重要信息缺失；另一方面，清洗颗粒过粗，则无法实现有效的数据清洗。因此，针对具体应用场景应选择合适的算法工具。

大数据分析的方法

开展大数据分析是为了探索海量集成数据间的关联关系，充分挖掘数据价值，对未来发展进行合理评估。数据分析的任务一般可分为两类：解释型和预测型。解释型是对数据规律进行归纳总结，并做规范化呈现。预测型则是在研究现有数据规律的基础上，对未来的变化趋势进行合理预测，以指导现阶段的业务决策。传统的数据分析技术主要针对结构化数据进行处理，目前已有较为成熟且有效的分析体系。然而综合能源服务平台的海量数据具有规模大、类型多样化和分布广泛等特点，需要结合聚类分析、统计学方法、模糊处理及机器学习等数据挖掘方法对海量数据进行精细化分析处理。

规模效应下，引入聚类算法对海量数据进行预处理，可以缓解服务器运算压力，提高系统对大规模数据的处理能力。聚类方法作为一种数据分析工具，其目的是将数据点划分为不同的类别。根据算法设计思想的差异，可将聚类方法分为基于层次聚类、基于密度聚类和高维聚类等。通过聚类方法，可以实现对用户用能数据的多维度分析和用能特性画像，便于开展差异化的营销策略，进一步巩固存量用户，拓展增量用户。应用统计学方法，对多维数据进行相关性、因果性分析，深入挖掘海量数

据中的关联信息，建立数据逻辑链，为辅助决策提供数据支撑。在实时能源市场中，能源供应商的出清价、投标价均为不确定性因素。模糊理论是目前较为成熟的处理不确定性问题的方法，其算法的核心是确定隶属度函数，该函数表征了模糊集合的特征，用于定义模糊的边界。模糊决策可以在不确定情形下得到多种近似的结果及其可能性，适用于能源价格预测和辅助决策。机器学习是人工神经网络、决策树等智能方法的统称。其理论基础是通过大量数据的训练来模拟人脑的神经网络决策链，基于训练出来的模型来推导预测结果，目前已获得广泛应用。

典型应用

（1）建立用户群体用能特征模型

基于数据聚类技术，对用户群体进行细分，并对不同产业、不同规模和不同区域用户的用能数据进行横纵向多维度类比，从而细致刻画不同用户群体的用能特性。在此基础上，应用统计学方法、机器学习等技术研究用户多维数据的关联性，挖掘用户用能行为特点，建立多时空尺度模型、多类型能源需求模型，为用户提供定制化服务。

（2）数据共享与增值服务

综合能源系统运行过程中积累了大量交易数据、用户用能数据。在基于大数据的综合能源服务系统中，数据不仅是分析处理的对象，更是一种资源。对国家而言，能源数据是社会经济发展的风向标，平台可建立基础数据共享机制和增值服务机构，提供业务咨询、节能评估以及需求优化等多种类型增值服务。

结束语

随着综合能源服务的发展，平台将集成海量用户用能数据。综合能源服务背景下的平台数据采集、处理和分析方法，可为平台数据价值挖掘、业务应用拓展提供参考。EA

欢迎订阅 欢迎投稿 欢迎刊登广告