



电信运营商能耗优化的大数据分析模型研究

方艾,金铎,徐雄,梁冰

(中国电信股份有限公司广东研究院 广州 510630)

摘要:针对电信运营商现网的实际情况,提出一种低成本可实施、基于大数据的能耗分析模型,帮助运营商在现有条件下,快速实现能耗分析管控,逐步建立精细化运营管理,为节能减排提供科学依据和决策,未来可发展到智能化的能耗管控。

关键词:能耗分析;大数据;机器学习;运营商

doi: 10.3969/j.issn.1000-0801.2014.10.007

A Big Data Based Analytical Model for Telecom Operator's Energy Usage Optimization

Fang Ai, Jin Duo, Xu Xiong, Liang Bing

(Guangdong Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China)

Abstract: An analytical model was discussed on energy usage analysis and optimization, dedicated to fast implementing at low cost for the telecom operator based on its current facilities, which may also improve power usage monitoring and data supported managing.

Key words: energy usage analysis, big data, machine learning, telecom operator

1 引言

能源,一直是全球热议的话题。节能减排备受各国政府和企业的重视与积极推进。电信运营商在节能减排上亦负有不可推卸的重任。据数据显示^[1],国内三大运营商在2012年的电费支出高达400亿元,占据运营成本的较高比重,南方一省分公司2013年的能耗相当于31.7万吨标准煤。随着移动业务和数据业务的迅猛发展,配套的基站数量、数据中心投入在继续增加,所带来的电力消耗也将持续增长。因此,节能减排对于运营商,不论是责任上还是利益上,都具有重大的意义。

然而,现阶段运营商对于节能减排尚处于粗放型的管理阶段,往往只能通过行政命令和财务管控等手段,效果

不佳。主要的问题如下:

- 能耗测量采用传统电表按月统计的方式,无法精确反映用电变化趋势,难以察觉用电异常,无法形成及时有效的管控;
- 如果对所有机房、局站的用电量都采用智能电表进行监控,投资巨大,维护成本高;
- 能耗评估标准不统一,缺乏有说服力的数据支撑节能减排措施的落地实施。

业界对于能耗监控和能耗优化已有一些解决方案,但并不完全适用于运营商,本文通过对运营商实际情况的分析,结合已具备的条件、实施成本等,提出一种大数据分析的解决思路和分析模型,可帮助运营商解决能耗管理的燃眉之急,并逐步实现精细化运营管理。

2 能耗分析

没有测量就没有管理,开展节能减排、进行能耗管理首先要清楚企业能耗使用的状况,包括各用电机房、基站、各服务网点、办公场地等用电实体在统计周期内的用电量、用电构成、变化趋势、异常等。运营商目前急切的节能需求是:一方面找出存在浪费的环节;另一方面找出可以优化的环节。通过测量到的数据:可进行横向对比,找出同类型机房/局站的能耗差异,帮助发现问题;可进行纵向对比,找出异常点;结合业务负载、办公场地大小、人员配置、地理环境、气候等内外部因素,建立能耗预测模型,进而形成合理的能耗评价标准。

2.1 能耗分析模型

业界普遍使用PUE(电力使用效率)评价数据中心的能耗^[2],本文所述的建模分析方法参考PUE标准,并认为:

$$PUE = \text{机房总用电量} / \text{主设备用电量}$$

其中,机房总用电量是维持机房正常运行的所有能耗(包括主设备、空调、照明、电源等能耗),主设备用电量则包括交换、传输、数据网的各类服务器、存储和网络设备的能耗,PUE的时空粒度取决于上述各项采集的粒度。

业界PUE一般在1.5左右,较高水平的达到1.1^[3]。保持较低的PUE是能耗运营成效的体现之一,理论上为降低PUE,一方面应设法降低主设备能耗;另一方面,应尽量减少主设备以外用电的浪费。对于大部分的通信机房、移动基站和接入局所,主设备能耗可调优的空间相对较小,现阶段可重点关注空调设备的能耗优化。

PUE值除了用于评估用电效率,也可用于评估耗电量水平。以上PUE计算式改写为:

$$\text{机房总用电量} = \text{主设备用电量} \times PUE$$

$$\text{主设备以外用电量} = \text{主设备用电量} \times (PUE - 1)$$

如果忽略照明用电及输电损耗等,空调用电量可近似等于:

$$\text{主设备用电量} \times (PUE - 1)$$

对于使用中央空调的机房,其空调用电量无法直接测算,但主设备用电量可以获取,通过与之参数相近的其他机房的PUE(如同类型、同等规模且PUE可测算的),可估算出空调用电量或总电量的参考值,再结合本机楼整体PUE及本机楼其他机房用电量进行分析评估。

2.2 能耗数据

运营商目前可利用的数据来源有多个,其各自的数据特点和质量如下所述。

(1) 财务数据

财务报表数据对应供电局的电表读数,反映局站的总用电量,无法区分办公场地、机房、设备等分项的用电量,因此无法得知用电构成,无法定位浪费的用电,且时间粒度较粗,一般每月一统计,但可作为总电量参考。

(2) 智能电表数据

通过在供电线路加装智能电表,可监控机房总路、分路或设备等指定监控点的用电情况,并通过短信或无线网络实时上传至监控中心。运营商已对重点的机房和接入网点进行了安装部署,但由于投资成本较高,无法大范围地铺设,此外智能电表运行一段时间后可能产生误差,需要定期校准,维护成本高。

(3) 动力环境监控数据

动力环境监控系统(以下简称动环监控系统)可集中对分布的机房主设备电流、电压、功率,环境温湿度等进行实时监测和计算,提供告警等功能。通过其具备的北向接口可获取监控数据,目前运营商基本部署了动环监控系统,只需进行接口开发,即可获得能耗相关数据。更重要的是,动环系统的监控已经细化到了机房或设备级别,无需大规模投入智能电表,即可低成本地实现细粒度(如机房、基站)的能耗监控,此方案的投资较少,可作为现阶段能耗数据的主要来源。

(4) 其他系统数据

某些设备(如PC服务器)的电源模块具备能耗监控及数据接口,可获取其能耗数据作为补充,建议在具备条件的重点机房采集,新建机房时亦考虑该数据的采集。

(5) 互联网数据

外部环境因素,如当地的气候,包括风力风向、日照强度、温湿度等,对于机房能耗也有较大的影响,应该纳入能耗分析模型。这部分数据可通过开放的互联网采集。

以上各数据来源覆盖的粒度不一,可互为补充、综合利用。对在各个数据源都未能获得数据的,在必要时可补充监控采集点,以达到一定的精细粒度要求。例如使用中央空调的多个机房,无法准确区分各机房的空调用电量;对于办公场所或服务网点,可能未有监控;如果要区分这些用电,应考虑加装电表。

2.3 大数据方案

从上述数据源分析可见,运营商已经能够获得大量的能耗数据,如何利用好这些数据,发现问题并找到原因,是实现能耗运营目标的重点。对此,大数据是有效可行的方



案,体现在以下几点。

- 采集足够多的数据是进行有效分析的前提,传统采集频率产生的数据,其量与质都不足以帮助发现问题,提高采集频率可获得更多的细节以及提高数据的精确度。
- 运营商机房、基站种类多、数量多,分布广、相关变量多,使得能耗数据复杂,数据量大。以某省 11 万个接入网点测算,假设每个网点采集 5 项指标,以 5 min 作为采集频率,每天将产生 $110\,000 \times 5 \times 60 / 5 \times 24 \approx 1.5$ 亿条记录,传统方式难以高效地处理此类多源异构的海量数据,且投资、维护成本高,因此必须采用大数据的方式处理。
- 简单的统计对比方式无法发现实质问题,依赖人工分析则投入大,效率低且不可持续。为此,可借助机器学习算法,同时引入更多相关的变量和数据,通过对大量数据的关联分析,寻找数据之间的相关性,发现影响能耗的各种因素,进而找到切实有效的节能方法。

能耗分析架构可基于 Hadoop 实现,Hadoop 是成熟的开源分布式计算框架,专门针对大数据的存储与计算设计,具有高可靠、高效、高扩展、高容错、低成本的优点。如图 1 所示的框架:来自不同数据源(财务、智能电表、动环系统、互联网等)的数据,实时或分时地传输至数据处理集群,经过 Hadoop streaming 的提取与清洗,汇聚整合成便于

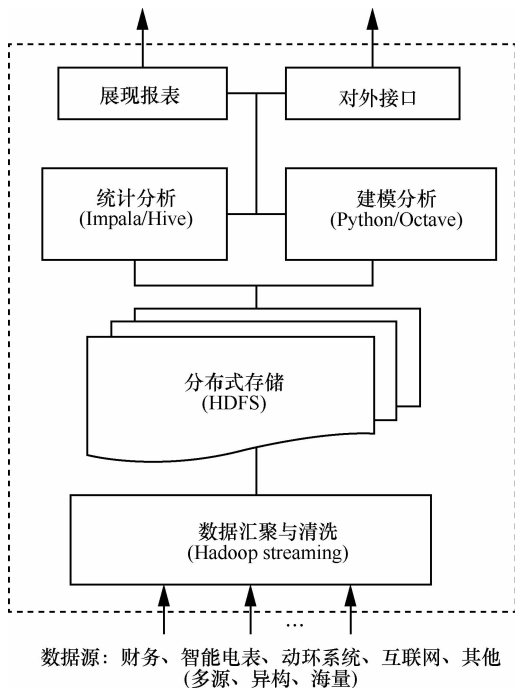


图 1 能耗分析架构

分析建模的基础数据,存储于集群的分布式文件系统上。基础数据用于:使用数据库工具(如 Impala/Hive)进一步统计分析得到报表数据;使用建模分析工具,如 Python scikit/Octave 进行建模分析、预测等应用。最后,报表及分析模型均可通过展现模块在 Web 展示或导出,也可通过接口供内、外部系统调用。上述软件均为成熟的开源工具,硬件采用普通配置的 PC 服务器,投资成本低。

2.4 能耗建模

影响机房能耗的因素有很多,包括机房外部环境(如温湿度、风力风向等)、机房配置(如冷气温度、开启的风柜/风扇数量、设备数量、功耗等参数),机器学习可帮助发现这些因素与能耗之间可能存在的关系。针对不同类型的机房,如核心机房(交换、传输、数据)、接入机房、基站等,可建立对应的模型。基于神经网络的分析模型特别适合处理此类需要同时考虑多种因素和条件、不精确和模糊的问题,此外还具备大规模并行、分布式存储和处理、自组织、自适应和自学能力。

图 2 是一个简化的神经网络模型^[4],包括输入层 X 、隐含层 a 和输出层 $h_\theta(x)$, θ^1 、 θ^2 是模型参数。

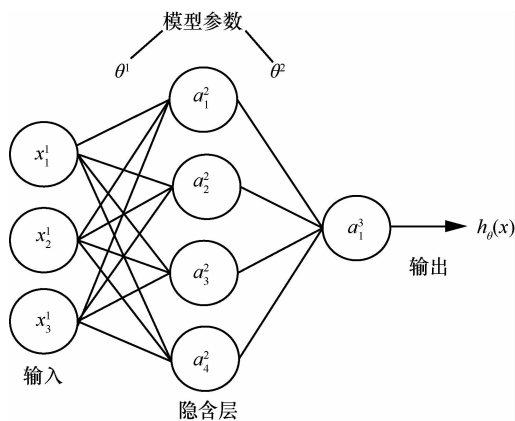


图 2 神经网络模型示意

神经网络模型可通过前向传递函数计算,由上一层激发下一层:

$$a^2 = g((\theta^1)^T X) \quad (1)$$

$$h_\theta(x) = a^3 = g((\theta^2)^T a^2) \quad (2)$$

其中, X 为模型输入矩阵, θ^1 为输入层到隐含层的参数矩阵, θ^2 为隐含层到输出层的参数矩阵, $g(\cdot)$ 为 Sigmoid (S型) 函数, 它的输出取值位于 (0,1) 区间。

在能耗分析模型里,输入变量 X 包括但不限于主设备数量、主设备功耗、空调设置温度、外部温度、外部湿度、外部风力、机房面积、开启的风柜数量、时间等多种因素。输

出 $h_{\theta}(x)$ 一般针对 PUE 值。

为得到模型参数 θ^1, θ^2 , 可通过求成本函数 $J(\theta)$ 的全局最小值进行^[5], $J(\theta)$ 表示为:

$$J(\theta) = \frac{1}{2m} \left[\sum_{i=1}^m (h_{\theta}(x^i) - y^i)^2 + \lambda \sum_{i=1}^{l-1} \sum_{j=1}^n \theta_{ij}^2 \right] \quad (3)$$

其中, x, y 为实测数据。将实测数据分为训练集和测试集两部分, 使用训练集求解使得 $J(\theta)$ 最小时的 θ^1, θ^2 , 得到模型的参数, 此时, 代入测试集的输入 X_{test} , 可计算出预测值 $h_{\theta}(x)$, 与测试集的输出 y_{test} 进行对比, 计算其误差, 评估模型的准确性。由于 $h_{\theta}(x)$ 由 Sigmoid 函数获得, 其值位于 (0,1) 区间, 因此训练模型时应将 y 也变换到此区间上, 验证时再将预测值变换回正确的区间。

为加快模型训练时的收敛速度, 可将 X 归一化, 变换到区间 [0,1] 或者 [-1,1] 上。为避免训练过拟合, 可加入适当的正值参数 λ , λ 的取值可通过对训练样本误差和验证样本误差的评估得出。另外可加入反向误差传递函数, 提高模型的准确性。

实际应用中, 输入变量可达到十几个, 隐含层可有多层, 每层可有数十个神经元。由于运营商机房类型较多, 机房的大小、业务承载、内部配置、外部环境各不相同, 可分为几大类, 分别定制合适的模型。模型的具体结构设计需根据不同机房的实测数据进行试验, 结合运算成本制定合适的方案。对采集的数据需进行异常检测, 现网环境中可能存在一些因素, 导致能耗数据异常或者不准确。例如生产机房与办公用电混用、基站非独立用电、机房内违规用电、电表长时间无校准产生大误差等, 导致采集的数据不符合实际。通过在同类型机房中的横向对比与排名以及同一机房在时间上的变化曲线, 可发现数值上的异常, 指引进一步排查确认。确认发生了错误的的数据, 在建模分析时应剔除; 另外, 通过计算数据的均值、方差, 了解数据的分布规律, 剔除偏差过大、远离分布区间的异常数据。随着时间的推移, 新产生的数据不断训练模型, 可使其准确度不断提高。

3 能耗分析应用

3.1 能耗显示

形成一系列报表, 供不同角色人员查看, 展现各级别监控粒度下的电力使用情况及变化趋势, 便于进行横向比较和排名, 可按照多级组织结构, 区分本地网、区县、局站、机楼/机房、基站等不同级别, 生成能耗变化曲线、能耗构

成饼图、能耗均值、方差, 显示异常值、异常时点等, 并能对比不同用电实体的能耗情况, 帮助管理者了解全局与局部情况, 帮助用电单位了解自身及兄弟单位的情况。

3.2 能耗预测

(1) 根据模型的输入, 即机房的配置参数, 如设备数量、设备功率、机房温度设置等变量, 预测 PUE 曲线或能耗曲线。例如, 当机房进行扩容时, 预测 PUE 将如何变化; 当部分设备需要停机检修时, PUE 是否会发生变化, 是否需要减少制冷以节能。

(2) 通过标杆机房的模型, 预测同类机房的 PUE 参考值或用电量参考值, 并与实际情况比对, 评估合理性。

(3) 新建机房时, 通过同类型的机房预测 PUE 或能耗值, 作为设计参考。

3.3 能耗优化

(1) 找出影响 PUE 或能耗的敏感因素, 即通过连续调整模型的某个输入参数 (如空调温度、空调数量、设备数量、设备密度等), 观察 PUE 相应的变化和变化趋势, 对比不同参数对 PUE 的影响, 寻求降低 PUE 或能耗的各种可能。图 3 是对一组模拟数据进行的调优预测, 显示了不同空调温度下 PUE 的变化。由图 3 可见, 适当提高空调温度可降低 PUE。在实际实施过程中, 可通过实测数据, 计算出 PUE 在不同参数设置下的变化曲线, 再结合实际情况综合考虑最优化的配置。采用模型预测指导调优的成本低, 无需在机房实地实验, 可保证生产安全, 大大缩短实验与验证的过程, 节约时间。此外可反复尝试各种调优的可能性。

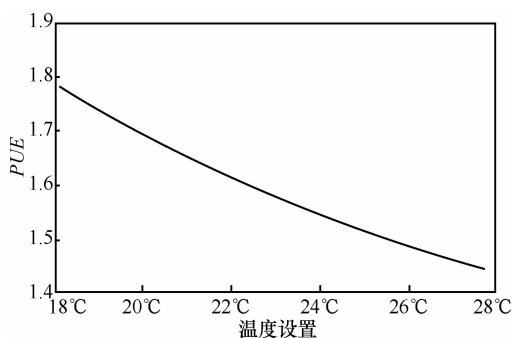


图3 调优示意

(2) 机房发生变更时, 如增减设备导致 PUE 升高时, 可能是制冷发生了浪费, 通过模型分析得出应调整哪些设置, 调整到什么值, 使得 PUE 尽量低。

3.4 智能化管理

未来可形成闭环反馈控制, 实现智能化节能, 特别是对大型的数据中心等能耗大户。通过分析模型的反馈, 自



动化、适时调整机房设置参数,使其能够常年保持在较低的 PUE 水平,实现更实时、智能的节能。

4 结束语

对于标准化数据中心的节能减排,业界已有相当成功的案例和推广经验^[6],但运营商并不能照搬。运营商除了数据中心,更多的是通信机房、移动基站等,且其标准化程度不一。本文根据运营商的实际情况进行变通应用,解决现阶段棘手的问题。PUE 的准确测量与计算是首要任务,需要一定的投入;能耗分析的准确性依赖大量的实测数据,可借助大数据方法和机器学习进行;运营商还需要特别针对机房、基站、服务网点等设备的特点,尝试各种节能方式,采取切实有效的措施,这是一个逐步推进的过程。

参考文献

- 1 能耗过大 电信运营商不堪重负. <http://www.d1net.com/carrier/market/236751.html>
- 2 Power_usage_effectiveness. http://en.wikipedia.org/wiki/Power_usage_effectiveness
- 3 Data centers efficiency. <http://www.google.com/about/datacenters/efficiency/>
- 4 Gao J. Machine-learning-applicationsfor-datacenter-optimization. <http://www.doc88.com/p-9886702774217.html>
- 5 Andrew Ng. Machine learning. <https://class.coursera.org/ml-006>

(上接第 37 页)

运营思路和方法,涵盖了基础平台选型与集成,基础平台运营设计,旨在为企业规模化建设运营大数据应用提供参考。

大数据基础平台是云计算的一种形式,通过共享的基础设施和相应的运营流程,实现了软硬件资源的高效利用、业务应用的快速提供,为客户提供更优质服务,从而提升企业的整体市场竞争能力。

参考文献

- 1 Gartner. Gartner's hype cycle special report for 2014. <https://www.gartner.com/doc/2816917>, 2014
- 2 黄勇军,冯明,丁圣勇等. 电信运营商大数据发展策略探讨. 电信科学,2013,29(3)
- 3 崔艳堂. 腾讯大规模 Hadoop 集群实践. 程序员,2014(2): 99~103
- 4 中国电信集团公司. 分布式平台实施技术指南,2014
- 5 中国电信股份有限公司江西分公司. Hadoop 平台规划方案

- 6 Data centers efficiency how others can do it. <http://www.google.com/about/datacenters/efficiency/external/>

[作者简介]



方艾,男,现就职于中国电信股份有限公司广东研究院,主要研究方向为大数据分析应用、机器学习。



金铎,男,现就职于中国电信股份有限公司广东研究院,主要研究方向为搜索引擎技术、互联网应用、云计算、电子商务。

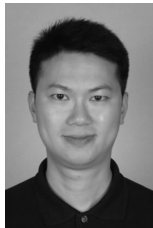
徐雄,男,现就职于中国电信股份有限公司广东研究院,主要研究方向为大数据应用、互联网应用、电子商务。

梁冰,女,现就职于中国电信股份有限公司广东研究院,主要研究方向为云计算、搜索引擎技术、互联网应用。

(收稿日期:2014-09-26)

V2.0, 2014

[作者简介]



陈泳,男,中国电信股份有限公司广东研究院 IT 评测工程师,主要研究方向为 IT 基础设施规划与选型等。



高昂,男,中国电信股份有限公司广东分公司计费业务高级工程师,主要研究方向为分布式平台的规划与建设。

(收稿日期:2014-09-25)