

基于多元大数据平台的用电行为分析构架研究

郝然,艾芊,肖斐

(上海交通大学 电子与电气工程学院,上海 200240)

摘要: 随着智能电网的发展,越来越多的测量装置向底层延伸。高级测量体系和配电网的发展不可避免地使用户用电数据量呈几何倍数增长,另一方面,电网也在积极寻求方法让需求侧可以充分地参与电网调控,增强电网可控性和经济性。在上述背景下,运用配用电数据分析用户用电行为建立相关驱动方法,可充分利用现有资源,为政府政策制定、电力公司业务拓展和用电行为引导提供新的解决思路。在配用电数据采集、聚合、处理和应用等方面提出了以大数据平台为基础的整体构架,设计了基于流处理和批处理的数据驱动方法,提出了适用于多维大数据用电行为分析的随机矩阵相关性算法,最后讨论了用电行为分析面向不同对象的应用场景。

关键词: 智能电网; 用电行为; 数据驱动; 高维大数据; 随机矩阵; 互动机制

中图分类号: TM 761

文献标识码: A

DOI: 10.16081/j.issn.1006-6047.2017.08.003

0 引言

随着智能电网理论与实践的推进,电网建设正朝着数字化、互动化和分布化的方向迈进。目前,新能源技术和信息技术等智能电网关键技术在发电侧和输电侧的渗透较多,而用电侧的智能化与理想目标还有较大差距,特别是对用户侧用电行为的分析与理解不能适应用户用电量的增长和用电方式多元化的趋势。因此,对用户用电行为进行深入分析就显得十分迫切。

在传统的电力系统运作模式下,生产与消费环节具有同时性,电网内不同区域或同一电网内不同时间段的电力消费的严重不平衡问题会使电网的运行状态不稳定,并且会导致能量的损失。现阶段电力系统的发展给生产与消费的平衡带来了新的挑战:

a. 用户侧负荷的增长加剧了电网的峰谷差,削峰填谷的任务强度和难度进一步增加;

b. 用户多样的生活方式对应不同的用电方式,这使得需求侧负荷变化的随机性增大,传统的负荷预测方法的精度越来越难以保证;

c. 大规模新能源发电的接入同样使得发电侧具有了波动性和随机性,系统出力同样难以确定;

d. 用户对供电质量和经济性的要求越来越高。

单纯依靠传统方法,即新增装机容量、输变电扩容和新设备用线路在技术层面和经济层面上都面临着前所未有的挑战。因此,未来电力系统需要需求侧的积极参与,实现需求侧与供应侧资源的双向互动

和协同优化^[1],共同保证电力系统供需平衡和安全经济运行。

用户用电行为分析是在负荷建模的基础上,通过用户用电关联分析的方法识别用户用电模式,挖掘需求侧响应的内在机理,其目的是得到科学的用电行为理解并将其作为价格和激励互动机制制定的依据。

近年来,已有一些学者对用户用电行为展开了相关研究。文献[2]提出改进的 k -means 聚类方法用于用户用电行为分析;文献[3]在云计算平台的基础上,通过智能用户采集系统,采用分布式程序构架实现了用户用电行为的聚类分析;文献[4]考虑用户针对电价变动的用电行为统计规律,建立电网公司与用户博弈的分时电价定价模型;文献[5]分析了影响居民用电的因素,采用模糊综合评价的方法分析了不同观念、收入以及信息偏好用户的需求侧响应能力;文献[6]建立了峰谷电价下用户响应模型并对模型参数进行了估计,提出了随峰谷电价迁移的模型实时更新流程。在应用方面,文献[7]采用大数据的思想,提出了基于随机森林算法的并行负荷预测;文献[8]提出了包括电价制定决策支持、需求方案制定、能效提升方案制定等用户用电行为理解的应用场景。

本文研究了现代智能化配用电信息的提取方式,讨论了传统峰谷定价用户用电行为建模的局限性,根据现有技术条件给出了多元大数据平台融合的改进思路;分析了模式识别、关联分析以及用户用电行为理解的实现方法,率先将批处理和流处理的大数据处理方法应用于用户用电行为分析,提出适用于用电大数据相关性分析的随机矩阵方法,突出随机矩阵在多维大数据分析方面的优势。最后从不同角度给出了用户用电行为分析的应用场景。

收稿日期:2016-09-23;修回日期:2017-06-13

基金项目:国家自然科学基金资助项目(51577115);国家重点研发计划(973计划)资助项目(2016YFB0901304)

Project supported by the National Natural Science Foundation of China(51577115) and the National Key Research and Development Program of China(973 Program)(2016YFB0901304)

1 用户配用电信息特征提取

1.1 用户配用电数据特征

随着电力系统信息化的推进,智能电表、智能变电站、实时监测检修系统和一大批电力系统生产链甚至是服务于其他行业的信息管理系统^[9]逐步应用于电力系统,它们包含的数据共同构成了用户配用电信息。用户配电数据有以下特征。

a. 数据量巨大、维度多。

庞大的用户群体、多样的数据需求以及为精确分析用户用电行为而越来越高的采样率的要求,使得用户配用电数据体量庞大^[10],如美国天然气电气公司每月采集的 900 万用户的数据信息高达 3 TB^[11]。另外为分析用户负荷构成,采样精度也有巨大的上升潜力,由此会引起数据量的成倍增长。用户配用电信息必然具有大数据的一些基础特征。

b. 复杂的关联性。

用户配用电数据来自分散布置的数据源和不同数据管理采集系统,且具有关联性。首先应该对时空分散且异构的数据进行聚合,为理解用户用电行为后应选取特征信息,运用相关性分析和主成分分析探求用户用电行为与特征量的具体关系。

c. 价值密度低。

用户用电行为分析具有价值,但其数据量大,数据价值密度低,所以需要通过数据挖掘的手段提取用户的用电行为信息,通常采用的方法为基于历史的 MBR 分析法、聚类分析法、回归分析法和决策树分析法等。

d. 存在社会和技术等因素造成的干扰。

由于社会和技术原因(如炒房热带来的房间空置率上升、窃电行为和群居现象,电表本身存在的误差等),用户配用电信息存在误差或错误数据,可通过设置较低的用电阈值进行识别或应用居民普查等其他信息系统判断数据的有效性。

1.2 基于负荷用电监测的用户配用电信息提取

负荷用电监测分为侵入式和非侵入式 2 种^[14]。侵入式监测需要多个传感器进入室内空间,其会占用室内或用电器空间。非侵入式监测称为负荷分解与监测 LDM(Load Disaggregation and Monitoring),最早由 Hart 在 1992 年提出^[15]。LDM 技术无需单独为每个电气设备设置传感器,仅通过采集用户端电压或电流信息,提取每个电气设备的印记特征,并应用机器学习和模式识别方法辨识出每个印记特征的来源,实现 LDM。该技术具有经济性强、简单可靠、单点入户无需通信等优势,对未来精细用电信息提取提供了有效的技术手段。LDM 智能用电分析云构架如图 1 所示。

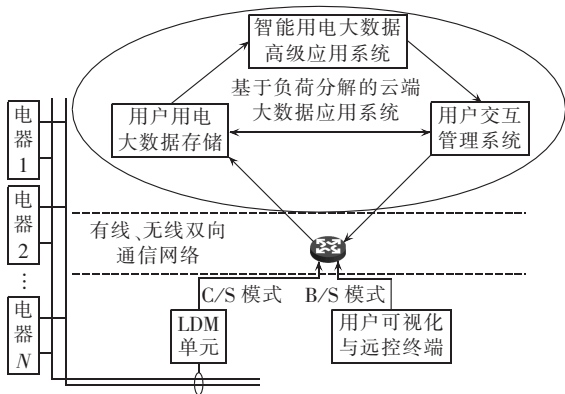


图 1 LDM 智能用电分析云构架图
Fig.1 Cloud architecture of LDM intelligent electricity consumption analysis

LDM 技术对分项用电信息进行在线反馈,将单纯的用电曲线细化为实时的分项用电量,给用户制定节电策略和合理的错峰计划提供了量化的参考依据。另外,基于 LDM 技术的用电信息的可用性和信息的价值密度优于传统采集方法,更适合智能用电云数据的高级应用,如能效分析、用电行为理解等。

1.3 用户配用电信息数据聚合

随着智能传感设备及系统的应用,国家电网在 26 个省的试点项目累计实现了 1.55 亿户信息采集,初步建立了较大规模的高级测量体系(AMI)。截至 2013 年初,全国智能电表覆盖率已达 40.5%,直供直辖的智能电表装设率也已达 55%^[12]。智能传感器的广泛布置为用户配用电信息的采集提供了物质基础,多样的传感器对应多元的数据结构。

通过数据采集、存储和处理技术,首先将时空分散的数据进行分布存储,并形成更新队列,实时更新数据,降低对存储容量的要求;然后,对不同维度异构化的数据进行归一化处理,建立不同的数据库;最后将多元大数据平台的相关数据进行融合。不同数据源的配用电信息数据分类与特点如表 1 所示。

2 传统峰谷电价的用户用电行为建模

传统用户用电行为建模是基于峰谷电价和负荷转移率的研究,受限于数据采集设备的制约,仅可利用电力系统内部电量采集数据,参数辨识也只能依赖于电价调整而引起的用户用电曲线的变化。

2.1 用电需求弹性矩阵建模

为表征用电用户对单时段和多时段响应的敏感程度,该模型构造了电力需求价格弹性矩阵(PEED)^[13],自弹性系数 ρ_{ii} 和交叉弹性系数 ρ_{ij} 分别表示同时段和不同时段电价波动对用电量的影响。

$$\rho_{ij} = \frac{\Delta L_i / L_i}{\Delta p_j / p_j} \tag{1}$$

$$\rho_{ii} = \frac{\Delta L_i / L_i}{\Delta p_i / p_i} \tag{2}$$

表 1 用户用电信息数据分类与特点统计表
Table 1 Statistics of data classification and characteristic of user electricity consumption information

数据源		数据信息	数据特性
电力系统内部数据	用户信息采集系统	用电器类型、电量、电压、电流等实时和历史数据	数据量在 TB、PB 级别,数据量大
	配用电管理系统	配电网拓扑信息、变压器运行信息、用户空间分布信息	数据量大,信息密度小
	生产运行系统	生产计划、检修巡视记录	实时性要求不高
	电力营销系统	用户合同档案、远程抄表数据、电费数据、财务数据、信用评价	数据信息密度大,数据量大
	客户服务系统	服务受理记录、服务处理记录	数据累积,增长快
行业服务数据	地理信息系统	地理信息数据	更新较慢
	天气预报系统	温度、风速、湿度、阳光	实时性要求高
	经济审计系统	地区经济情况、行业发展情况	重要,数据更新较慢
社会数据	城市管理系统	人口分布、城市规划	重要,数据量一般
	日期、习俗	工作日、双休日、节假日	重要,数据量小
	社会普查	用户对互动方案的评价、用户关注点	直接、重要、获取难度大

其中, ΔL_i 、 Δp_i 和 ΔL_j 、 Δp_j 分别为激励前后 i 时段和 j 时段的负荷量变化值、电价变化值; L_i 、 p_i 分别为 i 时段的负荷量和电价的参考值。

根据经济学原理,对不同偏好的用户构造弹性矩阵 G :

$$G = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nm} \end{bmatrix}$$

(3)

定义用户满意度是建立在用电量变化 ε 和总电费变化 δ 基础上的^[14],并设置 2 个目标的权值分别为 γ_1 和 γ_2 ,则满意度为 $R = \gamma_1 \varepsilon + \gamma_2 \delta$ 。

2.2 用户心理学建模

心理学研究表明,对消费者的激励只有在达到一定阈值后才会有消费反馈,超过这一阈值后消费反馈会随着激励程度的增大而增大,直到达到饱和值^[15]。通常将这一用户心理现象用分段线性化的方法表示。

这里因变量为负荷转移率,即转移负荷量与原负荷的比值,自变量为电价差额,可表示为:

$$\lambda_{ijm} = \begin{cases} 0 & 0 < \Delta p_{ij} \leq a_{ijm} \\ K_{ijm}(\Delta p_{ij} - a_{ijm}) & a_{ijm} < \Delta p_{ij} \leq \lambda_{ijm}^{\max} / K_{ijm} + a_{ijm} \\ \lambda_{ijm}^{\max} & \Delta p_{ij} > \lambda_{ijm}^{\max} / K_{ijm} + a_{ijm} \end{cases}$$

(4)

其中, λ_{ijm} 为用户 m 在 i 时段转移到 j 时段的负荷转移率; Δp_{ij} 为 i 时段和 j 时段的电费价格差; a_{ijm} 、 K_{ijm} 分别为用户 m 在从 i 时段到 j 时段的线性区斜率响应阈值。用电负荷转移率与电价差额的函数关系如图 2 所示。

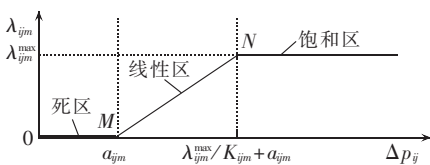


图 2 用户负荷转移率/电价差分段线性化模型图

Fig.2 Piecewise linearized model of user load transfer rate/price difference

用户 m 在 t 时段的负荷可以表示为:

$$L_{tm} = L_{tm0} + \sum_{i \in \Theta_t} \lambda_{itm} \bar{L}_{im} - \sum_{i \in \Phi_t} \lambda_{itm} \bar{L}_{im}$$

(5)

其中, L_{tm0} 为未进行激励时的负荷; $\bar{L}_{im}$ 为 i 时段的平均负荷; Θ_t 为电价高于 t 时段的所有时段集合; Φ_t 为电价低于 t 时段的所有时段集合。

2.3 统计学建模

通过设计科学合理的问卷,从问卷中提取用户信息,直接得到用户对不同营销政策的喜好程度,应用统计学中的回归分析得到不同用户对不同营销政策的响应曲线。

利用统计学算法简单,但对数据要求高,需要通过网络和上门问卷等形式汇总用户的实际意见,数据采集成本高,且存在人为因素造成的数据误差。

3 现有建模方法的不足与改进思路

传统用户用电响应模型(图 3)虽然可以从不同角度描述用户用电行为,但是由于数据分布式采集、存储和处理等技术的局限依然存在很多缺点,不能与当前配用电大数据和用电分析应用相适应。传统用户用电响应模型的缺点主要有:

- a. 传统用户用电响应模型仅考虑了电价随时间分布的单一因素,没有考虑天气、节假日、经济发展水平等其他重要因素;
- b. 传统用户用电响应模型的输出结果一般为用电负荷转移矩阵和用户满意度,只能应用于电价制定、峰谷时段调整和互动机制制定等几个方面;
- c. 电负荷转移矩阵参数辨识困难,需通过多次电价调整且通过一段时间的实践才能得出系统参数,成本较高;
- d. 传统用户用电响应模型一般只能进行离线分

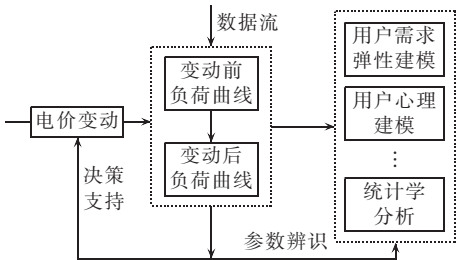


图 3 传统单一数据平台下的用户用电响应建模结构图

Fig.3 Structure of modeling based on conventional single data platform for user electricity consumption response

析,不能充分挖掘实时价值。

针对传统用户用电响应模型的不足,结合大数据平台现有条件和用电分析需求,本文提出几点改进设想,改进的用户用电响应模式的结构框图如图 4 所示。

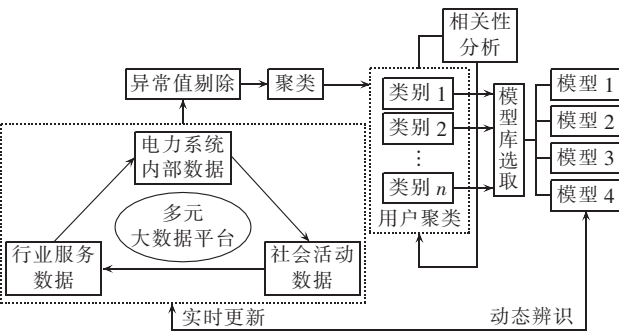


图 4 基于多元大数据平台的用户用电响应建模结构图
Fig.4 Structure of modeling based on multivariate bid data platform for user electricity consumption response

- a. 通过不同领域的信息采集对用电当事人进行聚类分析,使用户用电响应模型更有针对性。
- b. 针对多个因素建立负荷转移率相关性模型,并剔除显著性较低的因素。
- c. 通过引入实时电价进行参数辨识和修正,并根据一年中不同时段的数据建立参数对照表。通过实时数据分析建立用电动态模型,可根据 logistics 函数构造响应模型或建立有不同延时和时间常数的响应模型。
- d. 针对多元大数据平台中的异常或错误数据进行剔除并分析,及时了解房间空置率、电表故障、偷电监测等问题并采取合理对策。

4 基于用电大数据的用户用电行为分析

本文针对离线与实时的用电行为分析分别设计了批处理和流处理^[17]的数据分析方法。

4.1 用户用电行为离线分析

Google 公司在 2004 年提出的 MapReduce 是典型的批处理构架。MapReduce 搭建了用于并行计算的软件处理模型,基于此结构上的程序可以对大规模的分布式智能元件分割而治,批量处理数据,然而其欠缺实时处理数据的能力,仅适合价值不随时间而减小的数据的分析。

4.1.1 用户用电行为模式认知

用户用电行为模式认知需要在噪声、高维度、具有复杂关联关系且随时间平移的信息中识别用电模式。传统建模仅研究电价与用电量这种单一维度的关系,随着目前人们的用电方式和用电心理越来越多样化,在复杂的机理环境下传统模型驱动的弊端日益凸显,而数据驱动方法可以考虑跨领域的多维因素。

聚类是在选取特征和权值的基础上对样本进行相似性搜索,对样本按照特征进行分类,可以依据用电信息进行精细分类。用户用电行为的模式认知一般通过聚类算法实现。常用的聚类算法有模糊 c 均值聚类、k-means 聚类、支持向量机以及竞争学习等方法。

由于智能数据采集设备分布广泛且数据较多,本文选取广州某工业园区用户信息和负荷曲线等信息,提出了一种采用简化并行计算的 MapReduce 分布式程序模型设计大数据云构架。图 5 是基于 MapReduce 架构的用户用电行为模式大数据聚类方案。

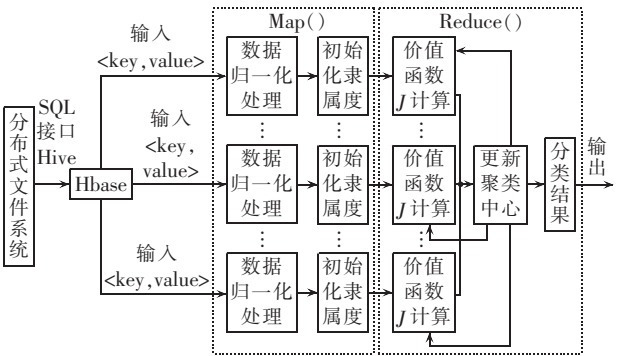


图 5 基于 MapReduce 架构的用户用电模式分析聚类方案流程图

Fig.5 Flowchart of clustering scheme of user electricity consumption behavior analysis based on MapReduce architecture

Hadoop 是 Apache 开发的开源分布式计算平台。分布式文件系统(HDFS)采用 M/S 结构,由多个管理节点和数据节点组成,每个节点均为智能测量终端或计算设备。系统数据库由 Hbase 数据库和 HDFS 组成,Hive 将结构化数据映射为数据表,可将结构化查询语言(SQL)数据库转化为 MapReduce 任务。每个阶段都以<key,value>的数据结构传输,其中,key 为样本编号,value 为用电数据的各个维度的坐标值。

设置 4 个聚类中心和 4 个 MapReduce 并行节点,分析广州某工业示范园区多个工厂 2015 年用电曲线。比较传统的串行聚类算法与基于 MapReduce 架构的聚类算法在 Java 环境下的计算效率,结果见表 2。由表 2 可知,基于 MapReduce 架构的聚类算法

表 2 MapReduce 与传统串行计算效率对比

实验次数		记录数		计算时间/s	
				本文方法	传统方法
1	7300			0.890	1.87
2	14600			1.065	2.45
3	29200			2.791	4.62
4	58400			3.259	8.84
5	175200			3.729	11.5

计算效率较高,且数据量越大算法优势越明显。

4.1.2 基于用户用电行为理解的相关性分析

用户用电行为与用户类别、天气、日期、温度、经济情况等多个维度有着强弱不等的耦合。分析经济社会自然等各种原因对用电时间、电量和频度的复杂相关性和因果性,针对特定的用户用电行为提炼关联性较强的影响要素,对于精细化用户用电响应模型、理解用户用电行为有很大意义。

以广州某工业示范园区不同类别用户用电响应能力相关性为例,选取电价调整前后负荷转移电量作为状态量,以经济水平、响应意愿、节假日、温度、用户理想电价偏差 5 个数据源作为特征因素,得到的数据矩阵是一个 3 维的高维矩阵,即特征因素数 $\times$ 同类用户数 $\times$ 采样点数。因此,对于任意一个特征因素,其特征因素是一个矩阵而不是传统意义上的向量。

传统相关性分析方法需准确地提取数据矩阵对应的特征因素、用户数和采样点等信息,将矩阵按照信息分解为多个因素向量后再进行多次分析,执行效率低。本文针对这一问题,提出适用于多元大数据分析的随机矩阵法,该方法根据随机矩阵的统计特性分析相关性,无需将原始数据矩阵分解为多个向量重复计算,在高维大数据处理方面具有优势。

基于随机矩阵分析多元数据相关性的方法简要介绍如下。

对于任意 L 个 K 行 T 列的非 Hermitian 矩阵随机矩阵 $\mathbf{X} \in \mathbf{R}^{K \times T}$,当其行数和列数趋于无穷且保持 $c = K/T, c \in (0, 1]$ 不变时,矩阵标准积特征值 λ 的经验谱分布几乎一定收敛到单环定理(single ring theorem)^[18],其概率密度函数为:

$$f(\lambda) = \begin{cases} \frac{1}{\pi c L} |\lambda|^{2/L-2} (1-c)^{L/2} \leq |\lambda| \leq 1 & (6) \\ 0 & \text{其他} \end{cases}$$

线性特征值统计量 LES(Linear Eigenvalue Statistic)反映一个随机矩阵的特征值的分布,平均谱半径 MSR(Mean Spectral Radius)是 LES 的一种,它描述了随机矩阵的迹。由大数定律(law of large numbers)^[19]可知,矩阵的迹能够反映矩阵元素的统计特性,平均谱半径与单环定理的理论预测值差值越大说明相关性越大,如图 6 所示。

本文依据上述理论,结合小波分解的去噪功能,提出基于随机理论对用户用电行为进行相关性分析。

a. 小波去噪。

用户用电行为有着较大的随机性,笔者希望通过滤波的方法去除用电曲线的随机部分。将正交小波变换等效为一组镜像滤波过程,即令用电数据通过 1 个高通滤波器和 1 个低通滤波器,通过 Mallet 方法去除输入数据的高频分量,降低 Gaussian 随机信

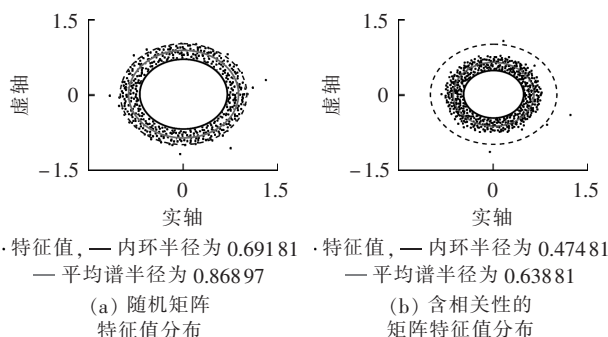


图 6 观测矩阵特征值分布图

Fig.6 Eigenvalue distributions of observation matrix

号能量。

b. 规范化处理。

不同数据源的特征因素的量纲往往是不同的,为使不同数据平台异构的数据具有同等的表现力,需进行规范化处理。

c. 奇异矩阵处理。

从数据源中获取数据矩阵,通过矩阵平铺得到标准非 Hermitian 矩阵 $\tilde{\mathbf{X}}$,再将奇异矩阵 $\tilde{\mathbf{X}}$ 变换为奇异等价阵 $\tilde{\mathbf{X}}_u$ 。

$$\tilde{\mathbf{X}}_u = \sqrt{\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T \mathbf{U}} \quad (7)$$

其中,酉矩阵 $\mathbf{U} = \tilde{\mathbf{X}} \sqrt{(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}}$ 。

d. 构造增广矩阵和参考增广矩阵。

若用户用电行为的状态变量数量为 z_n 个,待研究的影响因素的数量为 g_n 。 z_n 个用户用电状态变量的量测数据可以构成一个状态矩阵 $\mathbf{D}_z$, g_n 个影响因素的量测数据构成因素矩阵 $\mathbf{D}_g$ 。将状态矩阵 $\mathbf{D}_z$ 和因素矩阵 $\mathbf{D}_g$ 合并得到增广矩阵 $\mathbf{A}$,设 $\mathbf{N}$ 为与 $\mathbf{D}_g$ 维度相同的随机噪声矩阵,将状态矩阵 $\mathbf{D}_z$ 与噪声矩阵 $\mathbf{N}$ 增广合并得到参考增广矩阵 $\mathbf{A}_N$ 。

$$\mathbf{A} = \begin{bmatrix} \mathbf{D}_z \\ \mathbf{D}_g \end{bmatrix} \quad (8)$$

$$\mathbf{A}_N = \begin{bmatrix} \mathbf{D}_z \\ \mathbf{N} \end{bmatrix} \quad (9)$$

e. 计算 $\mathbf{A}$ 和 $\mathbf{A}_N$ 的平均谱半径。

平均谱半径的计算公式如式(10)所示。

$$\kappa_{\text{MSR}} = \frac{1}{m} \sum_{i=1}^m |\lambda_i| \quad (10)$$

其中, λ_i 为矩阵特征值。

f. 对平均谱半径之差求积分,以此作为用电量状态与特征因素的相关性度量。

增广矩阵与参考增广矩阵平均谱半径之差 d_{MSR} 定义为:

$$d_{\text{MSR}}(n) = \kappa_{\text{MSR}, \mathbf{A}_N}(n) - \kappa_{\text{MSR}, \mathbf{A}}(n) \quad (11)$$

平均谱半径之差越大,则因素矩阵相对于噪声矩阵的规律性越强,即状态量与因素量相关性程度越大。

上述随机矩阵相关性分析过程可以归纳为图 7 所示的流程示意图。

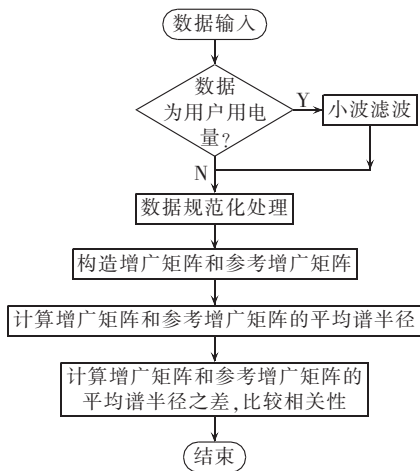


图 7 基于随机矩阵的相关性分析流程图
Fig.7 Flowchart of correlation analysis based on stochastic matrix

选取经济水平、响应意愿、节假日、温度和理想电价这 5 个特征因素,分析特征因素对 4 类用户响应能力的相关性,每类用户选取 20 个典型用户,每个典型用户采样 8760 个数据点,分析广州工业示范园区用户用电行为相关性,结果见表 3。

表 3 基于随机矩阵的广州某工业园区用电相关性分析结果

Table 3 Results of electricity consumption correlation analysis based on stochastic matrix for Guangzhou Industrial Park

特征因素	d_{MSR}			
	用户类型 1	用户类型 2	用户类型 3	用户类型 4
经济水平	0.1986	0.2457	0.2776	0.3075
响应意愿	0.2046	0.1459	0.2336	0.2641
节假日	0.0612	0.0949	0.1864	0.2053
温度	0.0276	0.0413	0.0212	0.0394
理想电价	0.0705	0.1010	0.1632	0.1348

由表 3 可知,用户经济水平和响应意愿对用户响应能力影响较大,温度和用户理想电价对于大部分工业用户响应能力的影响较小,不同类型的用户对于节假日因素的敏感程度差别较大,其中类型 3 和类型 4 的用户对于节假日的敏感程度明显大于类型 1 和类型 2 的用户。

4.2 用户用电行为实时分析

现有的分析手段一般是采集数据后进行集中数据分析,然而,很多用电数据的价值随着时间的流逝

而降低^[20],所以事件出现后必须尽快地对它们进行处理,单纯的离线分析不能满足及时性数据信息价值的准确提取。

本文提出的离线分析与实时分析融合的分析方法充分挖掘了实时数据的价值,运用于短期负荷预测和实时电价制定等方面的应用场景。基于此,本文设计了流处理的方法解决实时用电数据分析问题的计算框架,对应的流程图见图 8。

传统的分布式计算一般是首先拿到一个大的积累后的数据,再对数据进行拆分和聚合,而流处理的机制类似流管道,其通过事件机制,接收到消息后立即进行处理。

Puma、Storm 和 S4 均为目前比较典型的数据流计算系统,在互联网中得到了较为广泛的应用。拓扑是 Storm 中最高层次的抽象概念,它可以被提交到 Storm 集群执行,一个拓扑就是一个流转换图,图中每个节点是 1 个 Spout 或者 Bolt,当 Spout 或者 Bolt 发送元组到流时,它就发送元组到每个订阅了该流的 Bolt。Bolt 是 Storm 中的消息处理器,用于为拓扑进行消息的处理,Bolt 可以执行过滤、聚合、查询数据库等操作。

4.3 多元大数据用户用电行为分析新应用

通过对各个数据系统进行数据聚合、选取行为特征并对用户行为进行模式认知和关联性分析后,对于用户用电行为的理解和用电分析的结论可为政府、各类用户和电力公司提供制定决策的依据。

a. 面向政府。用户用电行为分析的结论可以为电费补贴政策、经济发展预测和房间置空率估计等提供依据。

b. 面向电力公司。建立基于大数据机器学习的负荷预测模型,为互动机制的制定和电网调度原型提供决策支持。

c. 面向用电用户。根据电网需要的辅助服务建立电网与用户合作博弈模型,引导用户选择合适的用电套餐,为用户用电进行合理规划。

5 用户用电行为分析技术框架

本文在用户侧数据采集、汇总、处理的基础上,归纳总结了用户用电行为分析的目标、驱动方法、技术路线和应用领域,归纳总结出用户用电行为分析的总体技术框架,如图 9 所示。

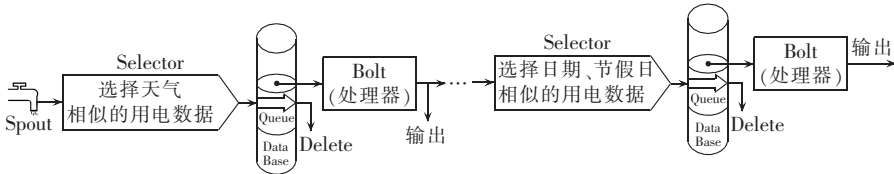


图 8 实时用电信息流处理流程图
Fig.8 Flowchart of real-time stream processing of electricity consumption information

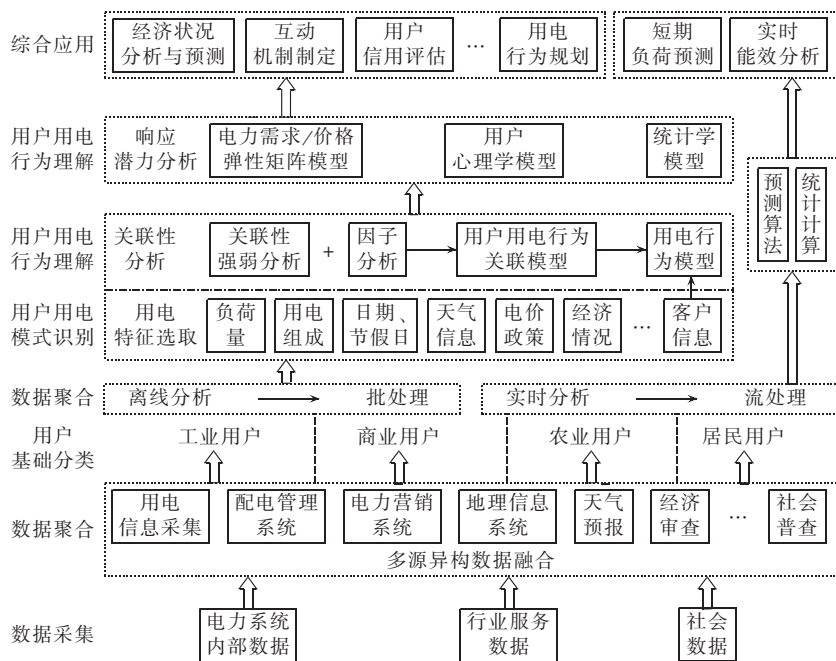


图 9 用电行为分析及应用技术框架图

Fig.9 Frame diagram of electricity consumption behavior analysis and application technology

6 结论

通过研究用户行为分析的框架和基本方法,本文得出的结论如下。

a. 将配用电大数据应用于用户用电行为分析。将新型的测量系统和数据驱动相融合,实现数据共享,使数据真正地为提高配网智能化和互动化服务,为现有的技术和业务瓶颈提供了一种新思路。

b. 智能采集设备已迅速普及,然而数据存储、数据集成和数据挖掘等技术亟需突破,建立统一的大数据的处理平台才能真正发挥智能采集设备的效用。

c. 流处理、批处理的数据驱动方法可有效地处理多元异构大数据。随机矩阵分析方法在多元大数据平台下,如果数据量和维度都较大则优势明显。

d. 用户用电行为分析综合应用范围广,对于政府政策、电力公司机制的制定和用户用电行为的引导具有数据支持的作用。

参考文献:

- [1] 曾鸣,武庚,李冉,等. 能源互联网中综合需求侧响应的关键问题及展望[J]. 电网技术,2016,40(11):3394-3398.
ZENG Ming,WU Geng,LI Ran,et al. Key problems and prospects of integrated demand response in energy internet[J]. Power System Technology,2016,40(11):3394-3398.
- [2] XUE L,LUAN W. Improved K-means algorithm in user behavior analysis[C]//2015 Ninth International Conference on Frontier of Computer Science and Technology. Dalian,China:IEEE,2015:339-342.
- [3] 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.

- [4] 董军,张晓虎,李春雪,等. 自动需求响应背景下考虑用户满意度的分时电价最优制定策略[J]. 电力自动化设备,2016,36(7):67-73.
DONG Jun,ZHANG Xiaohu,LI Chunxue,et al. Optimal TOU pricing strategy considering user satisfaction in automated demand response background[J]. Electric Power Automation Equipment,2016,36(7):67-73.
- [5] 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.
- [6] 阮文骏,王蓓蓓,李扬,等. 峰谷分时电价下的用户响应行为研究[J]. 电网技术,2012,36(7):86-93.
RUAN Wenjun,WANG Beibei,LI Yang,et al. Customer response behavior in time-of-use price[J]. Power System Technology,2012,36(7):86-93.
- [7] 王德文,孙志伟. 电力用户侧大数据分析与并行负荷预测[J]. 中国电机工程学报,2015,35(3):527-537.
WANG Dewen,SUN Zhiwei. Big data analysis and parallel load forecasting of electric power user side[J]. Proceedings of the CSEE,2015,35(3):527-537.
- [8] 王继业,季知祥,史梦洁,等. 智能配用电大数据需求分析与应用研究[J]. 中国电机工程学报,2015,35(8):1829-1836.
WANG Jiye,JI Zhixiang,SHI Mengjie,et al. Scenario analysis and application research on big data in smart power distribution and consumption systems[J]. Proceedings of the CSEE,2015,35(8):1829-1836.
- [9] 葛磊蛟,王守相,瞿海妮. 智能配用电大数据存储架构设计[J]. 电力自动化设备,2016,36(6):194-202.
GE Leijiao,WANG Shouxiang,ZHAI Haini. Design of storage

- framework for big data of SPDU[J]. Electric Power Automation Equipment, 2016, 36(6): 194-202.
- [10] 张东霞, 苗新, 刘丽平, 等. 智能电网大数据技术发展研究[J]. 中国电机工程学报, 2015, 35(1): 2-12.
- ZHANG Dongxia, MIAO Xin, LIU Liping, et al. Research on development strategy for smart grid big data[J]. Proceedings of the CSEE, 2015, 35(1): 2-12.
- [11] FERGUSON M. Architecting a big data platform for analytics. A whitepaper prepared for IBM[M]. Wilmslow, UK: Intelligent Business Strategies, 2012: 30.
- [12] 张东霞, 姚良忠, 马文媛. 中外智能电网发展战略[J]. 中国电机工程学报, 2013, 33(31): 1-14.
- ZHANG Dongxia, YAO Liangzhong, MA Wenyuan. Development strategies of smart grid in China and abroad[J]. Proceedings of the CSEE, 2013, 33(31): 1-14.
- [13] 丁伟, 袁家海, 胡兆光. 基于用户价格响应和满意度的峰谷分时电价决策模型[J]. 电力系统自动化, 2005, 29(20): 10-14.
- DING Wei, YUAN Jiahai, HU Zhaoguang. Time-of-use price decision model considering users reaction and satisfaction index[J]. Automation of Electric Power Systems, 2005, 29(20): 10-14.
- [14] 秦祯芳, 岳顺民, 余贻鑫, 等. 零售端电力市场中的电量电价弹性矩阵[J]. 电力系统自动化, 2004, 28(5): 16-19.
- QIN Zhenfang, YUE Shunmin, YU Yixin, et al. Price elasticity matrix of demand in current retail power market[J]. Automation of Electric Power Systems, 2004, 28(5): 16-19.
- [15] 李扬, 王治华, 卢毅, 等. 峰谷分时电价的实施及大工业用户的响应[J]. 电力系统自动化, 2001, 25(8): 45-48.
- LI Yang, WANG Zhihua, LU Yi, et al. The implementation of peak and valley time price for electricity and the response of large industries[J]. Automation of Electric Power Systems, 2001, 25(8): 45-48.
- [16] 宋亚奇, 周国亮, 朱永利. 智能电网大数据处理技术现状与挑战[J]. 电网技术, 2013, 37(4): 927-935.
- SONG Yaqi, ZHOU Guoliang, ZHU Yongli. Present status and challenges of big data processing in smart grid[J]. Power System Technology, 2013, 37(4): 927-935.
- [17] 宋亚奇, 刘树仁, 朱永利, 等. 电力设备状态高速采样数据的云存储技术研究[J]. 电力自动化设备, 2013, 33(10): 150-156.
- SONG Yaqi, LIU Shuren, ZHU Yongli, et al. Cloud storage of power equipment state data sampled with high speed[J]. Electric Power Automation Equipment, 2013, 33(10): 150-156.
- [18] QIU R C, HU Z, CHEN Z, et al. Cognitive radio network for the smart grid: experimental system architecture, control algorithms, security, and microgrid testbed[J]. IEEE Transactions on Smart Grid, 2011, 2(4): 724-740.
- [19] WIGNER E P. On the distribution of the roots of certain symmetric matrices[J]. Annals of Mathematics, 1958, 67(2): 325-327.
- [20] AGRAWAL D, DAS S, EL ABBADI A. Big data and cloud computing: current state and future opportunities[C]//Proceedings of the 14th International Conference on Extending Database Technology. Uppsala, Sweden: ACM, 2011: 530-533.

作者简介:



郝然

郝然(1993—),男,江苏徐州人,博士研究生,主要研究方向为电力系统大数据(E-mail: haoransjtu@sjtu.edu.cn);

艾芊(1969—),男,上海人,教授,博士研究生导师,博士,主要研究方向为电力系统建模、电力系统状态辨识(E-mail: aiquan@sjtu.edu.cn);

肖斐(1991—),男,湖南常德人,博士研究生,主要研究方向为可行域分析、电力系统大数据(E-mail: xiaofeisjtu@163.com)。

Architecture based on multivariate big data platform for analyzing electricity consumption behavior

HAO Ran, AI Qian, XIAO Fei

(School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: With the development of smart grid, more and more measuring devices extend to bottom layer. The development of advanced measurement system and distribution network inevitably leads to the geometric increase of user data. On the other hand, the power grid is also actively seeking ways to allow the demand side to fully participate in the grid regulation for enhancing the grid controllability and economy. In the above background, the electricity consumption behavior of users is analyzed based on the data of electricity distribution and consumption to set the related driving methods for the full use of existing resources to provide new solutions to the formulation of government policy, the business development of electricity utilities and the guidance of electricity consumption behavior. An overall architecture based on the large data platform is given for the acquisition, aggregation, processing and application of electricity distribution and consumption data, a data driven method based on the stream processing and batch processing is designed, a stochastic matrix correlation algorithm suitable for the electricity consumption behavior analysis based on the multidimensional big data is proposed, and the application scenarios of electricity consumption behavior analysis are discussed for different objects.

Key words: smart grid; electricity consumption behavior; data-driven; high-dimensional large data; random matrix; interaction mechanism