

# 浅析全社会用电信息大数据研究与应用

黄 敏<sup>1</sup>, 梅 沁<sup>2</sup>, 徐 麓<sup>2</sup>

(1.江苏电力公司大数据管理中心, 江苏 南京 210024; 2.无锡供电公司信通分公司, 江苏 无锡 210060)

**摘 要:**如何将大数据有效的应用于电力行业, 本文针对这一课题首先分析了大数据的主要特征和大数据实现的关键技术; 其后, 重点介绍了用电信息大数据系统开发的背景和大数据思维方式, 以及该系统已经实现的部分功能。

**关键词:** 大数据; 用电; 云计算

## 0 引言

近年来, 大数据被人们谈论的越来越多, 各类有关大数据的讯息扑面而来, 大数据在互联网、电信运营、金融投资等相关领域的率先应用已经累积了相当的示范效应, 不论是传统的 IBM、Oracle、SAP、intel 等 IT 厂家, 还是以 google、amazon 为代表的新兴互联网产品供应商, 都推出了各种各样甚至让人眼花缭乱的解决方案和软硬件产品。

那么, 究竟该如何有效的利用大数据服务电力行业的生产管理, 切实借助大数据提高电力企业的运营水平, 开发部署成熟可靠的大数据运行平台, 而不是陷入大数据布下的陷阱中, 就摆在了各大电力公司的面前。早在 2014 年初, 国网江苏公司就成立了大数据管理中心, 专门针对大数据展开相关研究工作。经过一段时间的工作, 管理中心确立了公司运营、电网运行、社会服务三个方面的重点课题, 旨在通过重点课题来带动此项工作的深化应用。其中, 江苏公司报送的两项课题均入围国网公司年度试点课题, 本文将“基于用电信息的大数据应用”课题相关情况进行介绍, 深入探讨大数据在电力行业的典型应用。

## 1 大数据与云计算

大数据 (big data), 或称巨量资料, 指的是所涉及的资料量规模巨大到无法通过目前主流软件工具, 在合理时间内达到撷取、管理、处理、并整理成为帮助企业经营决策更积极目的资讯。大数据通常有四个特点, 即 volume (大量)、velocity (高速)、variety (多样)、value (价值)。

云计算是通过网络提供可伸缩的廉价的分布式

计算能力, 云计算代表了以虚拟化技术为核心, 以低成本为目标的动态可扩展网络应用基础设施, 云计算是近年来最有代表性的网络计算技术与模式。

大数据与云计算是一个问题的两面, 一个是问题, 一个是解决问题的方法。通过云计算对大数据进行分析、预测, 会使决策更为精准, 释放出更多数据的隐藏价值。大数据技术的需求是伴随着云计算平台的出现而出现的, 实际上目前云计算技术是大数据存储与处理技术的重要组成部分。由于大数据的数据量和分布式的特点, 使得传统的数据管理技术难以胜任这种海量数据。

云计算的核心是海量数据存储和数据并行处理技术。其核心思想包括分布式文件系统(Distributed file system FS)和MapReduce 技术, 主要思路由 Google 公司提出。DFS 有着高容错性的特点, 并且是为部署在价格低廉的硬件上而设计的, 而且它为应用程序提供高吞吐量的数据访问, 适合那些有着超大数据集的程序。Hadoop 提供了DFS 的一种开源实现(HDFS), 该分布式文件系统放宽了POSIX 的要求, 可以实现流的形式访问文件系统中的数据(streaming access), 并具有高可靠性、高可扩展性以及负载均衡等能力。MapReduce是2004 年由谷歌公司提出的一个用来进行并行处理和生成大数据集的并行编程模型。Hadoop 包括了MapReduce 的开源实现, 为使MapReduce并行编程模型更易使用, 出现了多种大数据处理高级查询语言, 如Facebook 的Hive、雅虎的Pig、谷歌的Sawzall等。

用电信息大数据系统开发就是采用上述技术路线, 在现实环境中搭建了一个20多台普通PC服务器的云计算平台, 经过实际的数据验证, 计算性能较传统的IOE有较大提高, 相信在将来100台更大规模

的环境中，计算性能将有更大程度的提升。

## 2 用电信息大数据系统

### 2.1 背景

该系统在借助大数据思维方法，应用大数据技术，以江苏电力公司 3000 多万只电表每日实时数据为基础，以十多年来的营销信息系统积累了大量用电历史数据为依托，结合 PMI 等社会经济先行景气指数，探索大数据时代的电量、负荷预测的新方法，同时研究构建一套以用电走势、开工率、业扩为依据的各行各业用电景气指数，用以监测预警经济发展趋势，为政府部门判断宏观经济形势、制定经济政策或投资决策服务，进一步丰富电力公司服务经济社会发展的内涵、提升电力公司的社会形象。

### 2.2 大数据思维方法

全数据模式，样本=总体：小数据时代的随机采样，最少的数据获得最多的信息，现在一切都改变了，我们需要的是所有的数据，“样本=总体”。

不是精确性，而是混杂性：允许不精确，大数据的简单算法比小数据的复杂算法更有效，纷繁的数据越多越好，混杂性，不是竭力避免，而是标准途径。

不是因果关系，而是相关关系：知道“是什么”就够了，没必要知道“为什么”。在大数据时代，我们不必非得知道现象背后的原因，而是要让数据自己“发声”。

### 2.3 用电信息大数据集成

如图 1 所示，图中将课题的整个过程分解为三个过程，即大数据抽取、大数据清洗、计算和大数据分析，通过搭建一个用电信息大数据平台，对用电的海量数据进行挖掘和分析，从而得到有价值的各类信息。



图 1 用电信息大数据集成过程示意图

## 3 已实现的主要功能介绍

### 3.1 用电量走势分析

计算源为全体用户，而非部分样本；使用营销抄表月电量、历史气象数据、气象电量影响模型推导历史数据（至 2006 年）；完成全省 20 多万负控大用户，06-12 年 4 亿条历史日电量数据的推导；完成历史各地区分行业电量数据推导，生成 1 亿条历史数据。已经导入 13 亿的历史双月抄表数据，分析计算生成 13 个地市 20 亿条历史月电量数据。如图 2 所示。



图 2 用电量走势分析

### 3.2 用电量因素影响模型研究与建立

根据单一用电影响因素以及多因素的重叠影响，利用数据平滑技术搭建用电量因素影响模型。已完成 13 个地市 8000 多类，300 多万条模型数据的生成（目前暂没有考虑风力因素，也没有县级气象数据），预计各模型电量影响关系超过 2 亿条。

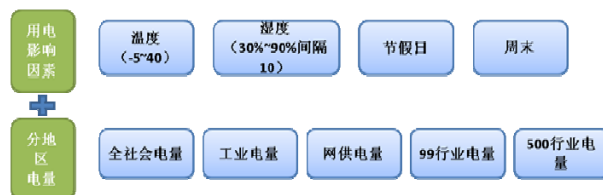


图 3 用电影响分析模型图

### 3.3 负荷特性

通过对全省不同季度典型日负荷曲线数据，分析四季负荷特点；对工业行业下的主要行业、非工业、低压非居民、居民的日用电负荷进行分析展现不同行业用电特性，具体如图 4 所示。



图 4 负荷特性分析模型图

### 3.4 课题的难点分析

该课题的难题主要在于各类分析模型和方法的建立,如负荷特性分析方法、电量预测方法、负荷预测方法、产业结构变化与电量变化的关联分析方法、电量与产值关系分析方法。

### 4 结束语

江苏公司通过开展“基于用电信息的大数据应用”课题的试点研究工作,希望进一步探索大数据在电力行业的应用前景,总结制定完善的设计方法论,确立先进的技术路线,构建公司统一的大数据平台,引领整个公司的大数据应用的深入发展。

#### 参考文献:

- [1] 维克托·迈尔-舍恩伯格,肯尼思·库克耶. 大数据时代[M]. 杭州: 浙江人民出版社, 2013.
- [2] 涂子沛. 大数据[M]. 南宁: 广西师范大学出版社, 2013.
- [3] 赵刚. 大数据: 技术与应用实践指南[M]. 北京: 电子工业出版社, 2013.
- [4] 道格拉斯·W·哈伯德. 数据化决策[M]. 北京: 中国出版集团, 2013.
- [5] 赵国栋,易欢欢,糜万军. 大数据时代的历史机遇[M]. 北京: 清华大学出版社, 2013.

- [6] 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.
- [7] 王春毅. 电力行业的大数据发展解析[J]. 电力信息化, 2013, 11(2): 8-9.
- [8] 衡星辰,周力. 分布式技术在电力大数据高性能处理中的应用[J]. 电力信息化, 2013, 11(9): 40-43.
- [9] 李国杰,程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域[J]. 中国科学院院刊, 2012, 8(9): 647-657.
- [10] 慕晓,高建宏. 基于人工免疫系统聚类算法的用电客户信用分析[J]. 山东电力技术, 2012, 187(4):53-55.
- [11] 杨劲锋,刘涛,陈启冠. 基于海量计量数据的电力客户在线分群研究[J]. 华东电力, 2013, 43(8): 1581-1586.
- [12] 陈娜,徐歆壹,宋红兵,等. 基于 Hadoop 的电信 BSS 大数据平台建设研究[J]. 电信科学, 2013, 36(3): 36-47.

#### 作者简介:

- 黄 敏 (1977—), 男, 湖北仙桃人, 高级工程师, 从事大数据项目管理, E-mail: samsung1634@163.com;
- 梅 沁 (1974—), 男, 江苏无锡人, 高级工程师, 从事信息化项目管理;
- 徐 薏 (1980—), 女, 江苏江都人, 工程师, 从事信息化项目管理。