

数据挖掘在智能电网中的应用研究

袁丽娟,袁方

(中国电力工程顾问集团西南电力设计院有限公司,四川成都 610021)

摘要: 电力大数据已成为电网公司进行决策的基础,但单纯数据的积累并不能给电网公司带来益处。因此充分利用这些基于电网实际的数据,对其进行深入分析,发现隐含的信息并加以利用,对指导电力企业做出正确的决策具有重大意义。提出了数据挖掘在智能电网中的应用场景,把数据挖掘技术应用于电力负荷预测,研究了在智能电网中应用数据挖掘技术的方法。

关键词: 数据挖掘; 智能电网; 负荷预测

Abstract: In the age of large power data, large data has become the basis for power companies to make decisions. But the simple accumulation of data can't bring benefits to the grid companies, so making full use of these actual data based on the grid, carrying on the thorough analysis, finding the implied information and using them to guide the power companies to make the right decision-making is of great significance. The application scenario of data mining in smart grid is put forward, and the data mining technology is applied to power load forecasting to study the method of applying the data mining technology to smart grid.

Key words: data mining; smart grid; load forecasting

中图分类号: TM769 文献标志码: A 文章编号: 1003-6954(2017)05-0015-03

DOI: 10.16527/j.cnki.cn51-1315/tm.2017.05.004

0 引言

随着中国智能电网的加快建设和发展,电网系统中积累了大量的电网基础数据,为智能电网建设带来了新的机遇与挑战,同时成为电力生产和运行模式发生变革的重要驱动力。从海量数据中提取出蕴藏的关系和规则等信息,并且根据已有的历史数据,预测未来的发展趋势,为电力企业业务管理向着更精细、更敏捷、更有效发展提供决策支持,成为现今亟待解决的问题。针对这一状况,数据挖掘技术应运而生。

下面将介绍数据挖掘技术及其在智能电网中的应用,侧重研究数据挖掘在细分用户电力负荷预测中的应用。

1 数据挖掘技术

数据挖掘,又称为知识发现(knowledge discovery),数据挖掘技术通过对海量数据进行建模,并运用数理模型对企业的海量数据进行整理与分析^[1]。

通过对这些数据进行微观、中观乃至宏观的统计、分析、综合和推理,发现数据间的关联性、未来趋势以及一般性的概括知识等,这些知识性的信息可用来指导高级决策活动。

1.1 数据挖掘系统的分层

常用的数据挖掘系统,一般分成三层结构^[2],如图1所示。

底层是数据源,包括数据库、数据仓库等。数据挖掘并不完全要建立在数据仓库的基础上,但数据挖掘若能与数据仓库协同工作,则将大大提高数据挖掘的效率。

中间层是数据挖掘工具,利用数据挖掘方法分析数据库中的数据,包括关联分析、聚类分析、神经网络方法、粗糙集方法、遗传算法、模糊数学方法等。

上层是用户层,将获取的信息形象地展示给用户,便于用户理解和观察,可以使用可视化工具。

1.2 数据挖掘过程

为形成数据挖掘产业的统一规范,1999年欧盟等机构联合起草了数据挖掘工业界的标准过程CRISP-DM^[3-4](cross-industry standard process for data mining),即“跨行业数据挖掘标准流程”,它为

数据挖掘项目的生命周期提供了一个综合的描绘,它包括6个阶段:商业理解、数据理解、数据准备、建模、评估和部署,如图2所示。这些阶段之间并不存在绝对的顺序关系,各个阶段一定会有反复的过程。究竟下一步要执行哪个阶段或哪个特定的任务,取决于每个阶段的结果。

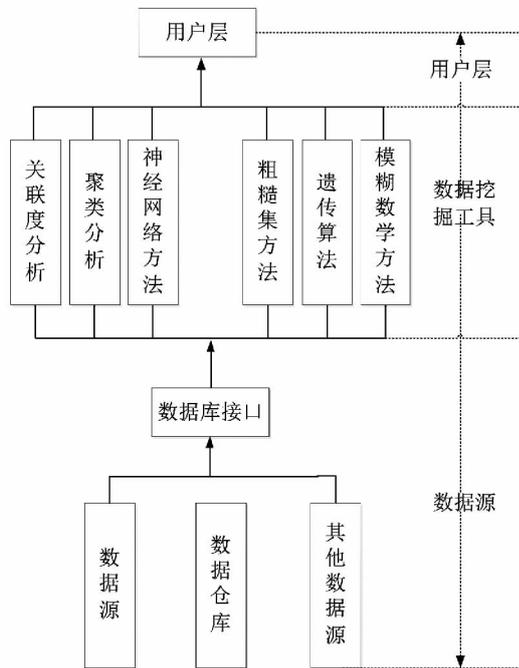


图1 数据挖掘分层结构

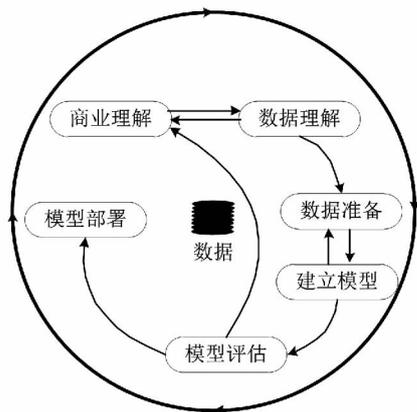


图2 CRISP-DM数据挖掘过程模型

1) 商业理解: 这是数据挖掘的初始阶段,确定数据挖掘的目的,明确需要解决的问题,并将这些目的与数据挖掘的定义以及结果结合起来。

2) 数据理解: 在业务分析目标的指导下,采集与业务对象有关的原始数据,并对可用的数据进行评估,并从中选择出用于数据挖掘应用的专门分析型数据库,以便提高数据挖掘的效率和结果的正确性。建立数据库可分为以下几个部分: 数据收集、数

据描述、数据选择、数据质量评估、数据清理、数据合并与整合、构建元数据、加载数据挖掘数据库、维护数据挖掘库。

3) 数据准备: 对可用的原始数据进行一系列的组织和清洗,以便达到建模的需求。此步骤可以划分为4个部分: 选择变量、选择记录、创建新变量、转换变量。

4) 建模: 为实现数据分析目的,对预处理后的数据执行数据挖掘算法的过程。在实际应用中,通常在同一数据集上,应用多种数据挖掘算法,并通过多个模型进行比较和选择,最终确定在当前数据集上使用效果最好的模型。

5) 评估: 对数据挖掘过程中的每个处理环节、步骤进行验证;对数据挖掘结果的质量、合理性、有效性等进行评价,重点考虑得出的结果是否符合第一步的商业目的。

6) 部署: 将发现的结果以及过程组织成为可读文本形式。

在模型部署和应用之后,还要不断监控模型的效果,随着使用时间的增加,可能需不断地对模型重新测试,甚至需重新建立模型。

2 数据挖掘在智能电网中的应用

2.1 智能电网数据特征分析

电网业务数据大致分为三类^[5-6]: 1) 电力企业生产数据,如电压质量指标、发电量、电网检测等方面的数据; 2) 电力企业运营数据,如交易电价、售电量等方面的数据; 3) 电力企业管理数据,如一体化平台、协同办公等方面的数据。智能电网相比传统电网具有更高的智能化水平,需要采集的电网数据从TB级逐渐上升至PB级^[7]。

2.2 数据挖掘在细分用户电力负荷预测中的应用

随着中国售电市场^[8-9]的开放,今后用户可自由选择售电公司。电力企业“以人为本”的行业意识就需要提高,为了更好地了解电力用户用电的需求,电力负荷预测成为一项重要的工作。下面重点研究数据挖掘在细分用户电力负荷预测中的应用。

根据电力系统相关标准,按照数据挖掘工业界的标准过程CRISP-DM,电力负荷预测数据挖掘系统的总体结构如图3所示。由图3可以看出该系统主要包括原始数据层、数据获取层、数据存储层和前

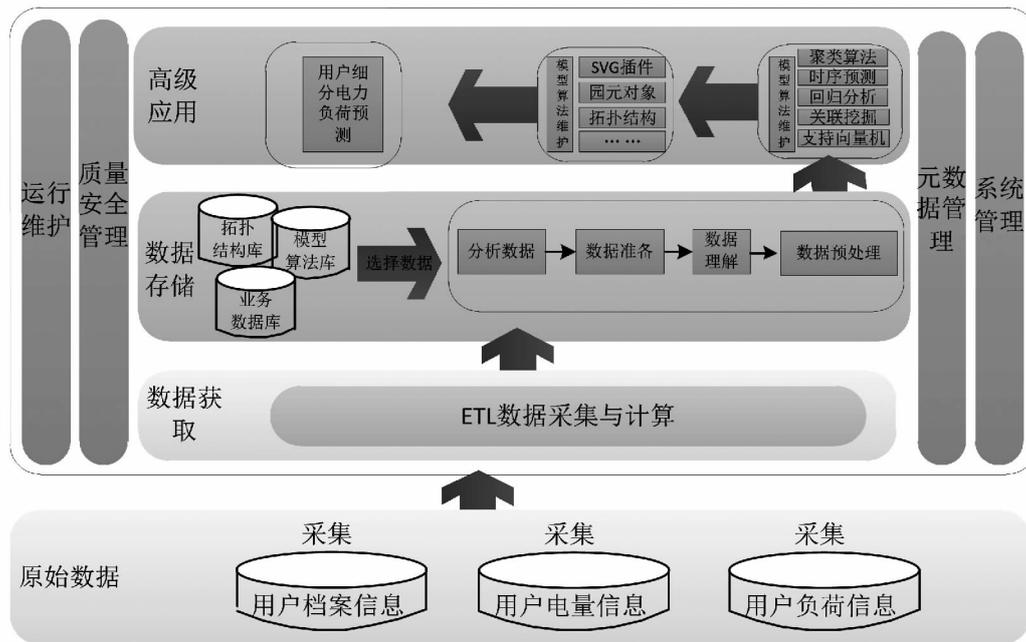


图3 电力负荷海量数据挖掘系统

端应用层。

这里提出基于细分用户负荷预测方法。该方法分为用户细分与负荷预测两个主要步骤:

1) 用户细分

用户细分中涉及的主要指标包括行业分类、用电规模分类、用电类别分类、客户信用分类、用户地理位置分类等,因此大量的用户细分工作还需要通过聚类分析来自动地进行聚合。

数据预处理:基于需求侧用户个性化的客户价值、客户的用电可靠性要求和客户行为的特征,先用聚类算法对行业用户进行了分组,将具有类似用电特征的用户聚集在同一组内,以便充分了解群组的用电特征^[10]。

常用的聚类算法包括:①K - Means 算法,通过各个属性间的距离度量值对输入的数据对象进行划分;②基于统计的 COBWEB 聚类,是一种增量式概念聚类方法,输入对象是采用符号量(属性 - 值)对来加以描述的,采用分类树的形式来创建一个层次聚类;③EM 算法,使用统计方法来确定各个分类中存在该用户的概率,由此可以对用户依照概率大小进行区分。

2) 负荷预测模型

针对分组后的用户组分别利用不同的时序预测算法进行负荷建模与预测,再将不同预测方法得到的预测结果进行加权平均。使用的预测方法主要有:RBF 神经网络、BP 神经网络、回声状态网络、支

持向量机等。

首先将各个用户群组的时间序列历史数据进行汇总分析,对每个群组内的计量点进行实时负荷叠加,分别得到叠加后的各点实时负荷序列数据。建立每个群组的实时负荷序列数据后,计算出该群组内每个日最大负荷值。不同日最大负荷构成该群组的日最大负荷序列,这样就得到每个群组的最大日负荷历史序列数据。在构建出每个群组的实时负荷历史序列数据后,计算出该群组内每个月最大负荷值,不同月最大负荷构成该群组的月最大负荷序列,得到每个群组的最大月负荷历史序列数据。这样,就分别得出了各用户群组的实时负荷、日最大负荷及月最大负荷历史数据,由此得到完整的负荷预测模型。

该方法在可接受的时间成本下,能尽可能多地了解不同行业、不同类别用户的用电需求,对发电调度、有序用电、精益化市场营销具有重要的指导价值。

3) 负荷预测模型分析

在进行负荷预测时,遵循“先细分、再预测”的原则,即先对用户属性进行选择,并利用聚类算法将用户负载细分为不同特性分类;再分别利用不同时序预测算法进行负荷预测,最后将各种预测算法下的负载预测结果进行加权求和,形成最终预测结果。另外还将预测结果与实际数据进行比较,对预测结果进行评价,并反馈至预测模型,通过调整建模参

(下转第46页)

考虑;距换流站40%~70%范围内,悬垂串和跳线串按5片、耐张串6片考虑;其余段悬垂串和跳线串按3片、耐张串4片考虑。

2.3 绝缘间隙

1) 塔头间隙:按绝缘配合要求,塔头的间隙不应小于招弧角间隙值,即在相应风偏情况下塔头间隙暂按招弧角间隙的相同值考虑。

2) 直流耐压间隙:由于接地极线路的最高工作电压(即直流运行电压)未变,故直流耐压间隙仍按原设计值0.12 m考虑。

3) 雷电过电压间隙:由于接地极线路的雷电过电压耐受水平低,雷电间隙仍按原设计值0.5 m考虑。

3 结 语

通过对楚雄站内、楚雄—穗东±800 kV线路各种故障工况下,在楚雄接地极线路上引起的操作过电压进行仿真计算,提出了楚雄接地极线路提高绝缘应满足的要求,同时考虑操作过电压的分布特性,

(上接第17页)

数,提高预测模型的精度。

3 结 语

智能电网中的大数据产生于电力系统的各个环节,通过有效地利用数据挖掘技术,从电力企业积累的数据中获得有助于管理决策的知识,帮助企业解决面临的现实问题,提高企业基于数据的业务管控能力和运营效率。

数据挖掘在智能电网的很多方面都有应用,并且效果明显。以数据挖掘在细分用户电力负荷预测中的应用为切入点,分析在售电侧改革的背景下,利用数据挖掘技术提高电网企业的经济效益、实现管理创新。

从中国目前的应用来看,数据挖掘技术在智能电网中还没有大规模和普遍使用,并且数据挖掘在实际的实现过程中仍然存在着一些亟待解决的问题。因此,仍需要进一步开展数据挖掘技术理论与应用方面的研究。

参考文献

[1] 温满华. 刍议大数据时代数据挖掘技术在电力企业中的应用[J]. 现代国企研究, 2015(24): 26.

对原线路的招弧角间隙、绝缘子片数、塔头间隙进行了加强改造,进一步加强了楚雄接地极线路的绝缘配置水平,保证了楚雄—穗东±800 kV线路的安全运行。

参考文献

[1] 赵婉君. 高压直流输电工程技术[M]. 北京: 中国电力出版社, 2004.
[2] 周沛洪, 修木洪, 谷定燮, 等. ±800 kV 直流系统过电压保护和绝缘配合研究[J]. 高电压技术, 2006, 32(12): 125-383.
[3] 王彩芝, 姜映辉, 王俊江, 等. 基于PSCAD/EMTDC的HVDC接地极线路故障仿真[J]. 中国电力, 2014, 47(2): 69-72.

作者简介:

胡全(1980), 硕士研究生、高级工程师, 从事输电线路设计工作;

鄢艺(1987), 硕士研究生、工程师, 从事输电线路设计工作。

(收稿日期: 2017-09-15)

[2] 于之虹, 郭志忠. 数据挖掘与电力系统[J]. 电网技术, 2001, 25(8): 58-62.
[3] 陈星莺, 张晓花, 瞿峰, 等. 数据挖掘在电力系统中的应用综述[J]. 电力科学与技术学报, 2007, 22(3): 51-56.
[4] 罗辑, 杨劲锋, 肖勇, 等. 用电数据挖掘技术与应用[M]. 北京: 中国电力出版社, 2015.
[5] 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.
[6] 邓焱琪. 智能电网大数据处理技术现状与挑战[J]. 中外企业家, 2015(6): 126.
[7] 中国电机工程学会信息化专业委员会. 中国电力大数据发展白皮书[M]. 北京: 中国电力出版社, 2013.
[8] 白杨, 谢乐, 夏清, 等. 中国推进售电侧市场化的制度设计与建议[J]. 电力系统自动化, 2015, 39(14): 104-109.
[9] 张晓萱, 薛松, 杨素, 等. 售电侧市场放开国际经验及其启示[J]. 电力系统自动化, 2016, 40(9): 1-8.
[10] 张晓春, 倪红芳, 李娜. 基于数据挖掘的供电企业客户细分方法及模型研究[J]. 科技与管理, 2013, 15(6): 104-109.

作者简介:

袁丽娟(1981), 大学本科, 从事电力通信工作;

袁方(1990), 硕士研究生, 从事信号与信息处理。

(收稿日期: 2017-06-27)