

文章编号: 1671- 251X(2011)01- 0068- 05

数据挖掘技术在电力营销系统的应用现状

郝晓弘¹, 朱洁¹, 王维洲², 胡振邦¹

(1. 兰州理工大学电气工程与信息工程学院, 2. 甘肃电力科学研究院, 甘肃 兰州 730050)

摘要: 针对电力营销系统存在数据采集不全、个别数据不准确等问题, 结合电力营销系统的数据特点, 分析探讨了关联规则、分类、时间序列和序列挖掘、聚类、空间挖掘等数据挖掘技术在电力营销系统中的应用现状及前景展望。

关键词: 电力营销系统; 数据挖掘; 关联规则; 分类; 序列挖掘; 聚类; 空间挖掘; 应用现状

中图分类号: TD601 **文献标识码:**A

Application Status of Data Mining Technology in Power Marketing System

HAO Xiao-hong¹, ZHU Jie¹, WANG Wei-zhou², HU Zhen-bang¹

(1. College of Electrical and Information Engineering of Lanzhou University of Technology,
Lanzhou 730050, China. 2. Gansu Power Science Research Institute, Lanzhou 730050, China)

Abstract: For problems of incomplete data collection and certain inaccurate data, the paper analyzed and discussed application status and prospect of data mining technologies such as association rule, classification, time series and series mining, clustering, space mining in power marketing system based on data characteristics of power marketing system.

Key words: power marketing system, data mining, association rule, classification, series mining, clustering, space mining, application status

0 引言

如何从海量的电力营销数据中提取出有价值的信息, 进而快速地为决策者提供尽可能准确、翔实的等量化指标和决策数据, 提高经营管理水平, 指导电网的经济运行工作, 是目前信息技术在电力营销系统应用中面临的主要问题。数据挖掘 (Data Mining, DM) 技术的出现, 无疑将在电力营销领域引起一场革命性的变革。本文从电力营销系统的重要性入手, 介绍数据挖掘的技术特点及其在电力营销系统的应用现状, 从而为电网的经济运行提供决策依据。

1 电力营销系统

电力营销系统以业扩报装、电能计量、用电管理、营业计费和线损管理等关键业务为核心, 在各业务模块之上提供服务模块和分析模块。服务模块包括电话服务、因特网服务和客户中心服务等, 侧重于为电力用户提供各种快捷优质的服务; 而分析模块则包括综合业务查询、基于历史数据的统计、效益分析和决策支持, 侧重于为企业领导提供及时准确的决策依据^[1]。因此, 一切为电力系统正常运行提供决策的原始数据都可归结为营销数据范畴, 比如生产系统规划设计、负荷预测及用户特征提取、经济调度、电力系统故障诊断、动态安全评估、异常数据的挖掘及相应处理等。

1.1 营销系统数据来源

营销系统的海量数据由管理信息系统、地理信息系统、SCADA 系统以及电网运行的实时信息系统(负荷管理系统、电能量计费系统、配变检测系统、计量检定)的运行数据所组成。随着电力企业信息

收稿日期: 2010- 09- 10

基金项目: 国家电网公司 2010 年研究开发项目(国家电网发展 [2010]151 号)

作者简介: 郝晓弘(1960-), 男, 甘肃泾川人, 教授, 博士研究生导师, 研究方向为学习控制、电动机控制技术、现场总线技术等。
E-mail: apple_zhujie@163.com

化建设的快速发展, 各系统已产生并积累了较为庞大的历史数据。

1.2 营销系统的数据特点

(1) 数据多。在电力营销系统中, 数据主要分为由各种装置实时采集的现场数据和由调度中心多种系统在运行过程中产生的大量数据, 数据来源多。另外, 电力系统属于大规模奇异非线性动态大系统, 在对其进行特征描述时往往涉及到上千个状态变量。传统的处理方法是对系统进行降维或简约化处理, 这在一定程度上影响了最终结果的精度^[2]。

(2) 数据种类混杂。营销系统是一个标准的混杂系统, 其上层(如调度中心)给出的(调度)决策主要是逻辑性的操作指令, 而下层控制(如发电机的励磁与调速控制)主要是连续性的, 为了达到系统多目标优化控制的目的, 应将不同性质的上层和下层控制有机地对合起来。

(3) 数据质量差。在营销系统中, 采集到的数据往往存在含有噪声、缺失、不正确等情况。

(4) 对数据的要求高。当系统处于紧急状态甚至瓦解状态时, 必须制定实时在线快速决策, 使系统重新回到正常状态。

2 数据挖掘技术

数据挖掘又被称为数据库中的知识发现(KDD), 是数据库发展与人工智能技术相结合的产物, 是一门新兴学科。它集统计学、模式识别、人工智能、机器学习、数据库以及高性能并行计算等技术于一体, 把人们对数据的应用从低层次的简单查询提升到从数据中挖掘知识, 为管理者提供决策支持, 具有极为广泛的应用前景, 是当前国际上数据库和信息决策领域的最前沿研究方向之一。

2.1 数据挖掘的基本概念

数据挖掘就是从大量的、不完全的、有噪声的、模糊的、随机的数据中提取隐含在其中的、人们事先不知道的但又是潜在有用的信息和知识的过程^[3]。利用数据挖掘技术可以对电力营销系统产生的海量数据进行分类抽取与优化整合, 合理存储后实时计算出相关指标以提取所关心的营销信息, 其结果将对营销系统的决策起到数据支持作用, 更好地指导企业的管理层和决策层对变化的环境做出快速而科学的市场营销决策。

2.2 数据挖掘的主要技术^[4]

(1) 关联规则。关联规则是数据挖掘中最活跃的研究方法之一, 最早是由 AGRAWAL 等人针对

购物篮分析提出来的, 采用经典的 Apriori 算法, 从大量的数据中挖掘出描述数据项之间相互联系的有价值知识。随着收集和存储在数据库中的数据规模越来越大, 人们对从这些数据中挖掘相应的关联知识越来越有兴趣。关联规则在数据挖掘领域应用广泛, 适于在大型数据中集中发现数据之间有意义的关系, 进而对这些关联规则进行有效评价, 筛选出用户真正感兴趣的、有意义的关联规则。

(2) 分类。分类在数据挖掘中是一项非常重要的任务, 其目的是学会一个分类函数或分类模型(分类器), 该模型能将数据库中的数据项映射到给定类别中的某一个类别。分类可用于预测, 预测的目的是利用历史数据记录自动推导出对给定数据的推广描述, 从而对未来数据进行预测。

(3) 时间序列挖掘和序列挖掘。时间序列挖掘是数据挖掘中的一个重要分支, 有着广泛的应用价值。它是指从大量的时间序列数据中提取人们事先不知道的, 但又潜在有用的、与时间属性相关的信息和知识, 并用于短期、中期或长期预测, 指导人们的社会、经济、军事和生活等行为。序列挖掘又称序列模式挖掘, 是指从序列中发现相对时间或者其它顺序所出现的高频率子序列。作为一般性的方法和技术, 序列挖掘最早由 AGRAWAL 等人提出, 已经成为数据挖掘新的研究分支而被广泛讨论。

(4) 聚类。聚类就是将数据对象分组成多个类或簇, 划分的原则是在同一个簇中的对象之间具有较高的相似度, 而不同簇中的对象差别较大。与分类不同的是, 聚类操作中要划分的类是事先未知的, 类的形成完全是数据驱动的, 属于一种无指导的学习方法。

(5) Web 挖掘。Web 挖掘可简单地定义为针对包括 Web 页面内容、页面之间的结构、用户访问信息、电子商务信息等在内的各种 Web 数据, 应用数据挖掘方法以帮助人们从 WWW 中提取知识, 为访问者、站点经营者以及包括电子商务在内的基于因特网的商务活动提供决策支持。

(6) 空间挖掘。空间挖掘是近年来才发展起来的具有广泛应用前景的数据挖掘技术, 实质上是数据挖掘在空间数据库或空间数据上的应用。简言之, 空间挖掘就是从空间数据库中抽取隐含的知识、空间关系或非显示地存储在空间数据库中的其它模式, 用于理解空间数据、发现数据间(空间或非空间)的关系。

3 数据挖掘技术在电力营销系统中的应用

3.1 关联规则在电力营销系统中的应用

关联规则是目前电力营销数据挖掘研究的主要技术之一, 它帮助决策者分析历史数据和当前数据的特征和规律, 从而预测未来。侯雪波等^[5] 将关联规则引入电力市场的营销分析中, 利用 FP-Growth 算法对经过离散化处理的电力营销数据进行关联规则分析, 描述出各种影响电量销售的外部因素与售电量水平之间的关联特征, 为电力市场营销提供辅助的决策信息。参考文献[6] 提出关联规则可指导电力市场营销策略、项目和投资组合管理, 进而进行需求预测、销售及收入预测、理赔分析等。在采用关联规则与其它方法相结合的方向上, 很多学者相继做了深入研究。张珂等^[7] 提出了基于云模型和关联分析法的电力营销目标市场模糊评价方法, 为确定电力营销目标市场提供了一种简单可行的方法。沈海湄等^[8] 提出了一种采用模糊关联规则挖掘的电力负荷预测新方法, 为电力负荷的非定性、定量分析提供了一种显性、易于理解的方式。牛东晓等^[9] 也在电力负荷预测的研究上做出了努力, 使总体预测精度得到提高。肖俊等^[10] 采用 FP-Growth 算法挖掘负荷数据, 并结合电力行业数据的特殊性, 分析各相关因素对电力负荷的影响, 成功地将关联规则数据挖掘方法应用于城市负荷分析中。

3.2 分类在电力营销系统中的应用

在中长期预测中, 除传统的序列预测方法外, 模糊理论^[11]、专家系统^[12] 等方法均被应用。姚李孝等^[13] 提出的基于竞争分类的神经网络方法取得了较满意的预测精度。冯丽等^[14] 提出的模式分类法可提高电力负荷预测的精确度。参考文献[15] 提出了一种将决策树技术同外推算法相结合的短期负荷预测算法, 并运用于福建省日调度计划的编制中, 预测精度较高。管胜利^[16] 利用分类树形成子数据库的方法, 在进行 SCADA 系统不良数据状态估计时, 降低了 SCADA 数据库的规模, 有效提高了计算速度。除上述应用方向外, 神经网络和模糊逻辑控制相结合的技术还在客户关系管理(CRM) 方面得到应用, 参考文献[17] 对此进行了探讨。参考文献[18] 利用回归算法、神经网络以及归纳算法等多种算法得出对于各机组的开停机计划表。

3.3 时间序列挖掘和序列挖掘在电力营销系统中的应用

时间序列挖掘被认为是最经典、最系统、被广泛

采用的一类短期负荷预测方法。而在短期负荷预测中, 应用最广泛、研究最多的是神经网络。在实际应用中, 常常将时间序列挖掘和神经网络结合来分析电力营销数据。参考文献[19] 对此做出了很好的总结。小波神经元网络具有比 BP 神经网络更快的收敛速度, 其改进隶属度聚类方法的应用可改善负荷大波动日的预测精度^[20]。范习辉等^[21] 提出了一种基于时间窗的序列挖掘算法, 并将其应用于警报信息的智能处理, 从而更加有效地指导电力系统的故障定位与诊断。参考文献[22] 提出了一种基于错误模型分析和快速诊断推理的新型数据挖掘算法, 该算法可大大提高系统状态分析和挖掘的能力, 从而有效判定电力系统运行状态, 提高对错误模型分析的精确度。

3.4 聚类在电力营销系统中的应用

聚类集中应用在电力用户分类与信用评价、不良数据的修正、负荷预测及分类、变压器故障诊断等方面。冯璐等^[23] 根据供电企业客户各方面不同的属性, 采用聚类分析法将各客户聚类为不同的组别, 决策分析者可根据聚类的结果分析各组别之间的差异性, 通过对类群特征的研究, 采取差异的营销策略, 从而提高企业的经济效益。参考文献[24] 针对电力客户信用分类的特征, 提出基于模糊聚类分析的电力客户信用评价算法, 得到了不同客户群的聚类中心以及客户的隶属度矩阵, 为客户群的特征分析提供了量化依据。参考文献[25] 在传统聚类算法 CURE 的基础上提出应用信息熵原则选择聚类过程中的基本参数, 并利用 Kohonen 网络提取相关负荷的特征曲线, 并将其用于不良数据的校正。王志勇等^[26] 针对获得的客户用电数据, 在经过一定的预处理后选取最合适的聚类方法以及聚类数目得到典型负荷代表曲线, 进而了解客户的用电模式、制订相应的购电合同, 对于增加电力部门的经济效益有着重要的作用。参考文献[27] 提出采用自适应矢量量化(AVQ) 神经网络模糊聚类方法对电力负荷特性进行分类与综合, 因 AVQ 神经网络模糊聚类算法本身机时小、处理速度更快, 因而结果更为合理。参考文献[28] 主要论述了利用可适应性的决策树来对不同类型的负荷进行聚类。符杨等^[29] 针对模糊 C 均值聚类算法用于溶解气体成分分析时存在的问题, 将加权模糊核聚类方法(WFKC) 引入到电力变压器故障诊断中, 建立一个新的变压器故障诊断模型, 该模型能有效改善复杂数据集的聚类性能, 提高了故障诊断的正确率。郑蕊蕊等^[30]、陈舵等^[31] 也在

电力变压器的故障诊断方面进行了研究,取得了可喜的研究成果。

3.5 空间挖掘在电力营销系统中的应用

对电力营销系统而言,大多数情况都需要决策人员快速地分析、诊断,并及时做出正确反应,特别是在电力市场条件下,重要决策的正确性对于电力企业的发展意义是不可估量的。将电网运行数据、负荷位置分布数据和实时变化数据等多目标层次的信息合为一体,利用特殊的空间技术对其进行综合处理,才能实现如设备跟踪、故障定位、模拟停电、损失评价及最优调度等高级功能^[32]。参考文献[33]利用空间分布规则、空间聚类规则、空间特征规则、空间区分规则得到同类及不同类负荷的分布。参考文献[34]指出空间挖掘可以应用于空间负荷预测、设备跟踪维护和故障定位。参考文献[35]提出可以利用空间挖掘到的普遍几何知识(如地理编码)对各类用户进行行业扩报装、查表收费、负荷管理等业务营运工作,还可以根据变压器、线路的实际负荷以及用户的地理位置、负荷可控情况制定各种负荷控制方案,以实现对负荷调峰、错峰和填谷等负荷侧管理功能。

4 结语

数据挖掘在电力营销系统的整体应用上仍处在起步阶段,单一的挖掘算法很难满足实际决策的需要,应在多方面因素的影响下不断改进挖掘算法。目前尚未出现针对特定决策支持系统的专门挖掘软件,只是架构出系统结构,核心的算法体制并未完善,不能满足决策需要,如何保证数据挖掘方法更精确、高效、稳定地应用于营销系统中仍是亟待解决的问题。虽然存在不足,但数据挖掘对潜在问题和规律具有很高的预见性,并且具有高效计算、监测和管理的能力,因此,它适用于电力营销系统大规模非线性问题的求解,必会显示出其强大的生命力,成为电力营销领域发展的重要工具。

参考文献:

- [1] 徐春玲,姜雨,郑孝安.基于应用服务器的三层体系结构的电力营销系统[J].电网技术,2002,26(7):71-73.
- [2] 卢强.数字电力系统[J].电力系统自动化,2004(9):1-4.
- [3] HAN Jiawei, KAMBER M. 数据挖掘:概念与技术[M].范明,孟小峰,译.北京:机械工业出版社,2001:185-195.
- [4] 毛国君,段立娟,王实,等.数据挖掘原理与算法[M].北京:清华大学出版社,2005.
- [5] 侯雪波,田斌,葛少云,等.关联规则技术在电力市场营销分析中的应用[J].电力系统及其自动化学报,2005,17(2):67-72.
- [6] 张斌,孙哲.关联规则技术在电力系统中的应用[J].电网技术,2007,31(1):64-66.
- [7] 张珂,王玉凡,苑津沙,等.基于云模型和关联分析法的电力营销目标市场模糊评价[J].华北电力大学学报,2009,36(4):30-34.
- [8] 沈海湄,王加阳,蒋外文,等.模糊关联规则挖掘在电力负荷预测中的应用[J].计算机工程,2003,29(15):138-140,162.
- [9] 牛东晓,吕佳良.关联分析在电力负荷灰色神经网络预测中的应用[J].华东电力,2007,35(8):60-62.
- [10] 肖俊,张晶,朱涛,等.基于关联分析的城市用电负荷研究[J].电力系统自动化,2007,31(17):103-107.
- [11] 孙洪波,徐国禹,秦翼鸿.模糊理论在电力负荷预测中的应用[J].重庆大学学报:自然科学版,1994,17(1):18-22.
- [12] KANDIL M S, EL-DEBEIKY S M, HASANIEN N E. Long-term Load Forecasting for Fast Developing Utility Using a Knowledge Based Expert System[J]. IEEE Transactions on Power Systems, 2002, 17(2):491-496.
- [13] 姚李孝,姚金雄,李宝庆,等.基于竞争分类的神经网络短期电力负荷预测[J].电网技术,2004,28(10):45-48.
- [14] 冯丽,邱家驹.基于电力负荷模式分类的短期电力负荷预测[J].电网技术,2005,29(4):23-27.
- [15] 姚宇臻,林丽娟.基于决策树的外推算法在福建电网短期负荷预测中的应用[J].福建电力与电工,2007,27(2):24-27.
- [16] 管胜利.应用数据挖掘技术估计SCADA系统不良数据状态[J].电力信息化,2007,5(4):70-72.
- [17] SFORNA M. Data Mining in a Power Company Customer Database [J]. Electric Power System Research, 2000, 55(3): 201-209.
- [18] STEELE J A, MCDONALD J R, D'ARCY C. Knowledge Discovery in Databases: Applications in the Electrical Power Engineering Domain [C]//IEE Colloquium on IT Strategies for Information Overload, 1997, London: 1-4.
- [19] HIPPERT H S, PEDREIRA C E, SOUZA R C. Neural Networks for Short-term Load Forecasting: A Review and Evaluation [J]. IEEE Transactions on Power Systems, 2001, 16(1): 44-55.
- [20] 张步涵,赵剑剑,刘小华,等.一种基于小波神经元网络的短期负荷预测方法[J].电网技术,2004,28(7):15-18.

文章编号: 1671- 251X(2011)01- 0072- 03

基于指纹识别的煤矿安全管理系统的设计探讨

许芹^{1,2}

(1. 中国矿业大学信电学院, 江苏 徐州 221116; 2. 安徽科技学院工学院, 安徽 滁州 233100)

摘要:从煤矿安全管理的现状入手, 提出了一种基于指纹识别技术的煤矿安全管理系统, 给出了该系统的总体设计思路、设计方法及软硬件设计方案。该系统采用指纹识别技术来验证煤矿下井员工的身份和约束员工培训工作, 有利于规范职工行为和提高对管理人员的监管力度。

关键词: 煤矿; 安全管理; 指纹识别; 身份验证

中图分类号: TD79 **文献标识码:** B

Discussion of Design of Coal Mine Safety Management System Based on Fingerprint Identification

XU Qin^{1,2}

(1. School of Information and Electrical Engineering of CUMT., Xuzhou 221116, China.
2. College of Engineering of Anhui Science and Technology University, Chuzhou 233100, China)

Abstract: By analyzing existing status of coal mine safety management, the paper proposed a coal mine safety management system based on fingerprint identification technology and gave the general design idea,

收稿日期: 2010- 10- 10

基金项目: 安徽科技学院教研项目(X201096)

作者简介: 许芹(1980-), 女, 山东莱芜人, 讲师, 硕士研究生, 研究方向为信号处理、图像处理、模式识别。E-mail: xuqincher@163.com

- [21] 范习辉, 张焰. 序列模式挖掘在电力系统警报信息处理中的应用 [J]. 电力系统自动化, 2005, 29(13): 49-53.
- [22] HE Yueshun, DING Qiulin. Fault Mode Analyze of Power System Based on Data Mining [C]// Proceedings of the 2009 International Symposium on Web Information Systems and Applications, 2009, Nanchang: 318-320.
- [23] 冯璐, 王成文, 申晓留, 等. 基于数据挖掘的供电企业客户关系管理系统研究与设计 [J]. 电力信息化, 2007, 5(7): 86-89.
- [24] 李频. 基于模糊聚类分析的电力客户信用评价算法 [J]. 计测技术, 2010, 30(1): 18-21, 34.
- [25] 冯丽, 邱家驹. 离群数据挖掘及其在电力负荷预测中的应用 [J]. 电力系统自动化, 2004, 28(11): 41-44.
- [26] 王志勇, 曹一家. 电力客户负荷模式分析 [J]. 电力系统及其自动化学报, 2007, 19(3): 62-65.
- [27] 黎灿兵, 曲芳, 王晓宁, 等. 基于模糊聚类的电力系统负荷特性分析 [J]. 郑州大学学报: 工学版, 2010, 31(1): 107-110.
- [28] PITTB D, KITSCHEN D S. Application of Data Mining Techniques to Load Profiling [C]// Proceedings of the 21st IEEE International Conference of Power Industry Computer Applications, 1999, Santa Clara: 134-136.
- [29] 符杨, 田振宁, 江玉蓉. 加权模糊核聚类法在电力变压器故障诊断中的应用 [J]. 高电压技术, 2010, 36(2): 374-374.
- [30] 郑蕊蕊, 赵继印, 王志男, 等. 基于改进灰色聚类分析的电力变压器故障诊断 [J]. 吉林大学学报: 工学版, 2008, 38(5): 1237-1241.
- [31] 陈舵, 崔杜武, 李雪, 等. 加权模糊聚类及其在电力变压器故障诊断中的应用 [J]. 西安理工大学学报, 2008, 24(2): 182-186.
- [32] 郭颖奇, 陈琳玲. 空间数据技术在 10 kV 配电网调度 GIS 中的应用 [J]. 计算机应用, 2002, 22(8): 91-93.
- [33] 王金凤, 杨丽徙, 陈根永. 空间数据挖掘在电力系统中的应用探讨 [J]. 继电器, 2004, 32(18): 79-82.
- [34] 赵特鹏. 空间数据挖掘在电力 GIS 中的应用 [J]. 电工技术, 2006(2): 28-30.
- [35] 王明俊, 于尔铿, 刘广一, 等. 配电系统自动化及其发展 [M]. 北京: 中国电力出版社, 2000.