第 30 卷 第 23 期
2006 年 12 月

电 网 技 术
Power System Technology

Vol. 30 No. 23
Dec. 2006

文章编号：1000-3673（2006）23-0083-05　　中图分类号：TM715　　文献标识码：A　　学科代码：790·625

数据挖掘技术在电价预测中的应用

林其友[1]，陈星莺[1]，王之伟[2]

（1. 河海大学 电气工程学院，江苏省 南京市 210098；2. 江苏省电力公司，江苏省 南京市 210024）

Application of Data Mining in Electricity Price Forecasting

LIN Qi-you[1]，CHEN Xing-ying[1]，WANG Zhi-wei[2]

（1．College of Electrical Engineering，Hohai University，Nanjing 210098，Jiangsu Province，China；
2．Jiangsu Electric Power Company，Nanjing 210024，Jiangsu Province，China）

ABSTRACT：The authors relate the features of data mining in brief; analyze the influencing factors of electricity price in detail; and propose a method based on data mining to forecast electricity price. In the proposed method the electricity price is characterized by five characteristic elements, i.e., the relation of market supply and demand, bidding based transaction of generated power, load demand of customers, price of fuel, price index and level of consumption; based on the forecasting tool for these characteristic elements and considering the influence extents of different factors influencing electricity price, the similarity search technique in data mining is adopted; then bringing in weight coefficient the weighted average for searched matching price suite is performed; at last the forecasted electricity price is obtained. The concrete application of the proposed method is demonstrated by case study.

KEY WORDS：power market；electricity price forecasting；data mining；similarity search

摘要：简要叙述了数据挖掘技术的特点，分析了影响电价的因素，提出了一种基于数据挖掘技术的电价预测方法。该方法将电价用市场供求关系、上网竞价发电功率、用户负荷需求、燃料价格、物价指数和消费水平等元素来表征，并考虑了不同电价影响因子的影响程度。利用数据挖掘中的相似性搜索技术，引进权重系数对所搜索到的匹配电价序列进行加权平均，进而得到所预测的电价值。最后举例说明了该方法的具体应用过程。

关键词：电力市场；电价预测；数据挖掘；相似性搜索

0　引言

自20世纪90年代以来，全球掀起了电力工业市场化改革的浪潮，目的就是要打破传统的电力垄断垂直经营体制，在电力企业中引入竞争机制，充分优化资源配置，从而达到提高效率、降低成本、促进电力行业持续健康地发展，最终促使社会效益最大化的目的。在整个电力体制改革过程中，电价作为重要的经济杠杆，在建立和培育电力市场、优化配置电力资源及调整各种利益关系等方面具有不可替代的作用，因此它是电力体制改革的核心内容。虽然目前各国电力市场的模式及电价的形成机制不尽相同，但电价总体上是时变的，并与电力系统的运行情况、原材料价格、气象变化、物价指数以及用户用电情况等因素密切相关。电价随着需求而变化，电价的变化也反作用于需求。在电力市场交易和竞标过程中，如果事先知道电价信息并提前安排生产计划和竞标策略，就可以获得更大的经济利益。准确的电价预测可以为市场参与者提供投资导向并进一步规避风险。因此，电价预测已成为电力市场中亟待研究和解决的课题之一。

由于电力市场出现的时间还很短，关于电价预测的研究还处于起步阶段。文献[1-2]利用人工神经网络进行电价预测。文献[3]分别利用人工智能仿真技术和人工神经网络技术预测电力供给方和需求方电量，并将两者有效地结合起来进行电价预测。文献[4]利用分类技术和自动回归方法并结合贝叶斯概率分布原理来预测电价。文献[5]利用自动平均滑差回归法来预测日前市场电价。近年来，数据挖掘技术在电力系统中的应用越来越广泛。文献[6-7]阐述了数据挖掘技术在电力行业中的应用现状和前景。文献[8-9]分别阐述了数据挖掘技术在机组事故追忆及发电设备状态检修中的应用。文献[10-12]研究了数据挖掘技术在负荷预测及电力调度中的应用。文献[13-14]基于粗糙集理论分别研究了离散化

基金项目：教育部博士点基金项目(20060294019)。

方法及小电流接地系统故障选线的有效域问题。

为了给市场参与者提供投资导向，协助其安排生产优化计划并进一步规避风险，本文对电价预测方法进行了研究与探讨。在详细分析了上网竞价发电功率、用电负荷需求以及燃料价格、物价指数与消费水平等电价影响因素之后，本文提出了一种基于数据挖掘中的相似性搜索技术进行电价预测的新方法。该方法基于电价表征特性元素的预测工具，利用相似性来搜索历史数据，对获得相似特性的特征元素追踪相应的电价数据，并根据不同影响程度及相似程度进行加权回归。最后，通过分析英格兰电力市场的数据验证了该方法的有效性和可行性。

1 电价影响因素分析

1.1 上网竞价发电功率

对于某一特定的用电负荷需求而言，当上网竞价发电功率高于该用电负荷需求时，电价会保持相对较低，且上网竞价发电功率越大则竞争越激烈，整个市场出清价就会越低；当上网竞价发电功率低于该用电负荷需求时电价就会抬高，上网竞价发电功率越小则整个市场出清价就会越高。因此，上网竞价发电功率对电价的影响很大，是构成电价波动的主要因素之一。

1.2 用电负荷需求

对于某一特定的上网竞价发电功率而言，用电负荷需求高于它时则电价会保持相对较高，用电负荷需求越大则整个市场出清价就会越高；当用电负荷需求低于上网竞价发电功率时，电价就会降低，用电负荷需求越小则发电侧竞争越充分，整个市场出清价就会越低。因此用电负荷需求对电价的影响很大，是构成电价波动的主要因素之一。

1.3 燃料价格

对于特定的上网竞价发电功率和用电负荷需求，如果燃料价格较高势必造成各发电商发电成本的增加，则上网电价抬高并最终提高市场出清价；如果燃料价格较低就会造成各发电商发电成本降低，则上网电价下降并最终降低市场出清价。因此，燃料价格对电价的影响很大，是构成电价波动的主要因素之一。

1.4 物价指数与消费水平

对于特定的上网竞价发电功率和用电负荷需求而言，如果物价指数与消费水平较高，则整个社会的市场价格都将上升，电价也必然有所提高；相反，当物价指数与消费水平较低时，整个社会的市场价格都将下降，电价也必然有所降低。因此，物价指数与消费水平对电价有一定的影响，是构成电价波动的因素之一。

1.5 其它因素

除上述因素外，电价还容易受天气、市场参与者投机行为、市场力及电力故障等意外因素的影响，假定这些影响均包含在负荷需求及市场供应情况的范围内。对实际电价数据的分析结果表明，未来电价和历史电价有很大关系，电价本身的变化规律及其所受各种因素的影响必然要在历史数据中体现出来。因此，充分利用历史数据并从中挖掘出电价变化规律是很有前途的电价预测方法。虽然神经网络、灰色预测理论和数据挖掘技术都能充分地利用历史数据来进行预测，但在处理“海量”数据时，数据挖掘技术更具优越性，因此本文利用数据挖掘技术中的相似性搜索方法进行电价预测。

2 数据挖掘技术简介

数据挖掘就是从海量数据中提取隐含在其中的、人们事先未知但又是有用的信息和知识，并将其表示成最终能被人理解的模式的高级过程[15]。数据挖掘技术产生于20世纪80年代末期，它是数据库知识发现(knowledge discovery in database，KDD) 的核心技术。数据挖掘技术具有强大的数据处理能力，能从大量的数据中发现有用的规律、规则、联系、模式等知识，其方法包括聚类、分类、相似性、关联度及回归分析等。

电力系统是一个复杂而又庞大的系统，其数据类型多种多样，数据量也极其巨大。电力工业市场化改革以后，数据量更是成倍地增长，尤其是负荷数据和电价数据。在这种情况下，仅利用传统的分析处理方法已不能有效地解决问题。数据挖掘技术正是解决上述问题的有力工具，它在电力用户特征提取和电力系统的负荷预测、运行模型分类、建模、运行状态和设备状态监控及电力调度优化、电价预测等方面都有很好的应用前景。

3 基于数据挖掘中相似性分析的电价预测

3.1 电价序列构成

如前所述，电价主要与上网竞价发电功率 P_g、用户负荷需求 P_d、燃料价格 p_f、物价指数和消费水平 a 有关，可将电价特性表示为一组序列空间组合，即

$$p_{(t)}=\{l_{(t)},P_{g(t)},P_{d(t)},p_{f(t)},a_{(t)}\} \quad (1)$$

式中：$p_{(t)}$ 为交易中心在时刻t的市场出清价；$l_{(t)}$ 为时刻t的市场供求关系，且 $l_{(t)}=P_{g(t)}/P_{d(t)}$。$l>1$表明供过于求，$l<1$表明供不应求。$l$ 越大则市场竞争越激烈，价格会较低，l 越小则价格会较高。

由式(1)可知，任一时刻t的电价数据可以由竞价发电功率等5个特性元素来表示，所有历史电价数据可以组成一组电价序列。采用数据挖掘技术可利用上述历史特性元素追踪出相应的历史电价，再根据一定的加权法则即可得到所要预测的电价值。

3.2　相似性分析

假定历史数据均为按照时间顺序排列的集合，本文拟采用时间序列相似性分析(又称相似性查询)技术从大量的历史数据中挖掘出有效信息并进行相应的电价预测。时间序列相似性分析就是在时间序列数据库中发现与给定序列模式相似的序列或在库中查找相似的序列对。此处定义电价的时间序列为等时间间隔的实数值序列(如英国原有POOL模式是每个小时计算一次市场出清价)。当给定阈值$x\geq 0$时，如果序列X和Y之间的距离D满足$D(X,Y)\leq x$，则称序列X和Y 在阈值x内相似，简称序列X和Y相似。在定义序列间距离的公式中，最常用的是欧几里德距离公式，即

$$D(X,Y)=\sqrt{\sum_{i=1}^{n}[X(i)-Y(i)]^2} \tag{2}$$

式中n为序列长度。根据构成电价序列的特性因素可知，电价 p_i 和电价 p_j 在阈值x内相似，则有

$$D(p_i,p_j)=[(l_i-_{l\,j})^2+(P_{gi}-P_{gj})^2+(P_{di}-P_{dj})^2+(p_{fi}-p_{fj})^2+(a_i-a_j)^2]^{\frac{1}{2}}\leq z \tag{3}$$

考虑到各特性元素对电价的影响程度不同，在此取一权重因子进行修正，即

$$D(p_i,p_j)=[b_1(l_i-l_j)^2+b_2(P_{gi}-P_{gj})^2+b_3(P_{di}-P_{dj})^2+b_4(p_{fi}-p_{fj})^2+b_5(a_i-a_j)]^{\frac{1}{2}}\leq z \tag{4}$$

式中 $\sum_{i=1}^{5}b_i=1$。

可见，利用数据挖掘中的时间序列相似性搜索进行电价预测的基本思想是：根据未来时段的负荷、上网竞价发电功率、燃料价格及物价指数和消费水平预测值，在一定范围内的历史数据序列中进行相似性查询，找出与预测序列最相似的历史序列，然后对其所对应的电价序列进行加权平均即可得到预测时段的预测电价。

3.3　利用相似性搜索进行电价预测

3.3.1　数据预处理

时间序列数据一般是不完整、不一致、不规范的，尤其对于上述5个描述电价特性的因素，这种现象更为严重。数据预处理技术可以改进数据的质量，有助于提高后续挖掘过程的精度和性能，因此是数据挖掘过程必不可少的步骤，其工作量通常占整个数据挖掘工作量的70%以上。时间序列数据预处理包括数据的集成、选择、净化及转换等工作。采用下式可对所挖掘的历史数据进行规范化处理：

$$X_i'=\frac{X_i}{\bar{X}} \tag{5}$$

$$\bar{X}=\frac{1}{m}\sum_{i=1}^{m}X_i \tag{6}$$

式中：X_i 表示电价特性中的任一特性元素的实际值；X_i'是规范化后的各数据值；$\bar{X}$ 是所挖掘数据的平均值；m是所挖掘数据的总数，一般m个数据的时间周期正好构成某一电价数据变化的时间间隔(如0.5 h)。

3.3.2　相似性搜索

可将已处理好的所有数据按照时间序列进行排序，进而构成由5个特征元素代表的相应电价序列。相似性搜索就是对历史数据进行搜索并提取出与目标数据相似的数据序列。若给定阈值z_i，则对于任一时间序列数据应满足

$$D(X_i,X^*)\leq z_i \tag{7}$$

式中：X_i为电价特性中的任一特性元素的历史实际值所组成的序列；X^* 为电价特性中的任一特元素的预测值所组成的序列。序列 X_i 与 X^* 相似并可作为相似序列参与电价预测。由于市场供求关系对电价的影响很大，因此一般可先对该序列进行相似性搜索，再结合具体的上网竞价发电功率、用户负荷需求、燃料价格、物价指数和消费水平数据进行二次搜索，这样，在搜索海量数据时可以提高搜索速度和处理效率。假定最终提取出的由满足要求的历史数据序列构成的电价序列为

$$p=\{p_1,p_2,\cdots p_n\} \tag{8}$$

式中：n表示满足相似性搜索要求的历史电价数据个数；$p_1,p_2,\cdots,p_n$ 分别表示相应的历史电价数据。

3.3.3　采用加权平均法预测电价

利用相似性搜索所得到的电价序列值对各电价值进行加权平均。两个序列之间的相似程度可以由它们之间的距离来表示，距离越小则相似性越高，

其对电价的影响就越大，故此处以历史数据序列与所预测序列之间距离的倒数作为权值，利用下式计算所预测的电价：

$$p^* = \left[\sum_{i=1}^{n} \frac{p_i}{\sum_{i=1}^{n} b_i D(X_i, X^*)}\right] \Bigg/ \left[\sum_{i=1}^{n} \frac{1}{\sum_{i=1}^{n} b_i D(X_i, X^*)}\right] \quad (9)$$

式中：p^* 表示要预测的电价序列；$\sum_{i=1}^{n} b_i D(X_i, X^*)$ 表示经过权重因子修正后各电价序列的特征元素与相应预测值之间的综合距离。由式(9)还可看出，与预测目标越相似的数据序列在电价预测中所占的权重越大，否则权重就越小。

4 算例分析

目前，将上网竞价发电功率、燃料价格、物价指数和消费水平等各项相结合进行电价预测还未见报导。为了更好地说明本文提出的电价预测方法，算例中电价数据和负荷数据取自2003年2月的英格兰电力市场数据，上网竞价发电功率、燃料价格、物价指数和消费水平数据均为假设数据。另外，在相似性搜索中，x值取得太小则可匹配序列就很少，预测值容易受个别序列随机性的影响；x值取得过大 则可匹配序列将增多，从而淡化了相应特征元素序列下的价格特征。x如何选取还有待于进一步研究，此处取x=0.1。采用本文提出的相似性搜索技术，利用英格兰2003年2月1日—21日的数据对其22日—28日的电价数据进行预测，其中22日的电价预测曲线如图1所示，22日—25日电价预测误差如表1所示。

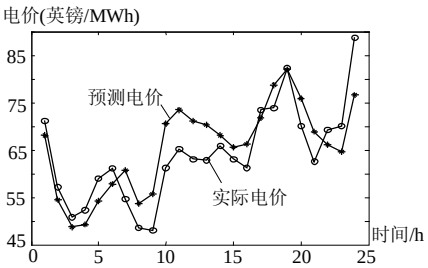

图1 英格兰2003年2月22日的电价预测曲线

Fig. 1 Price forecasting curves of England on Feb.22th, 2003

表1 英格兰2003年2月22—25日电价预测误差

Tab. 1 Error of electricity price forecasting of England from Feb. 22th, 2003 to Feb. 25th, 2003

指 标	22日	23日	24日	25日
最小绝对百分值误差/%	0.49	1.83	0.62	0.36
平均绝对百分值误差/%	3.06	4.89	3.71	4.54
最大绝对值百分值误差/%	8.68	9.12	8.29	9.78

通过图1及表1中的数据可以看出，采用时间序列相似性方法来预测电价的总体效果较好，预测的平均精度也较高。由于缺乏上网竞价发电功率、燃煤等原材料价格、物价指数与消费水平等对短期电价预测影响较大的数据，所以总体预测精度不是很高。如果能够得到上述数据并合理地设置各序列权重系数，则所预测的电价数据精度将会更高。另外，电价还受各种不确定性因素的影响(如发电商的投机行为、市场力及政府政策导向等)，从而增加了预测的难度。与一般的负荷预测相比，电价预测的误差明显偏大，这主要是由于电价受很多人为因素的影响，而这些因素在电价预测中是难以计及的。

5 结论

电力工业市场化改革以后，准确地预测电价对于市场参与者优化其生产经营策略并实现其自身利益最大化均具有重要意义。本文提出的利用数据挖掘中的相似性搜索技术的预测方法考虑了市场供求关系等5种电价影响因子的影响程度，因此预测结果较准确，可为我国电力市场参与者提供一个较好的优化生产辅助决策工具。

由于目前我国电力市场改革刚刚起步，对电价预测的研究还不成熟，本文提出的利用数据挖掘技术进行电价预测也只是初步性探讨。下一步将在设定相似性搜索的阈值z 、表征电价特征元素以及在挖掘时间较长的情况下如何提高挖掘效率等方面进行研究。

参考文献

[1] Szkuta B R，Sanabria L A，Dillon T S．Electricity price short-term forecasting using artificial neural networks[J]．IEEE Trans on Power System，1999，14(3)：851-857.

[2] Ying Yihong， Hsiao Chuanyo．Locational marginal price forecasting in deregulated electric markets using a recurrent neural network [C]．IEEE PES Winter Meeting，Columbus，Ohio，USA，2001.

[3] Bunn D W．Forecasting loads and prices in competitive power markets[J]．IEEE Transactions on Power Systems，2000，88(2)：163-169.

[4] Ni E，Luh P B．Forecasting power market clearing price and its discrete pdf using a bayesian based classification method[C]．IEEE Power Engineering Society Winter Meeting，Columbus，Ohio，USA，2001.

[5] Javier C，Rosario E，Francisco J N，Antonio J C．ARIMA models to predict next-day electricity prices[J]．IEEE Transactions on Power Systems，2003，18(3)：1014-1020.

[6] 路广，张伯明，孙宏斌．数据仓库与数据挖掘技术在电力系统中的应用[J]．电网技术，2001，25(8)：54-57.
Lu Guang，Zhang Boming，Sun Hongbin．Application of data

warehouse and data mining techniques to power systems[J]. Power System Technology，2001，25(8)：54-57.

[7] 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.
Yu Zhihong，Guo Zhizhong. Data mining and power system [J]. Power System Technology，2001，25(8)：58-62.

[8] 于达仁，张志强，鲍文，等. 某机组事故过程中转速信号的复现[J]. 中国电机工程学报，2002，22(8)，113-117.
Yu Daren，Zhang Zhiqiang，Bao Wen，et al. Rate signal recovery of some turbo-generator sets in an accident[J]. Proceedings of the CSEE，2002，22(8)：113-117.

[9] 李凡生，陈庆吉. 决策树分类算法在发电设备状态检修中的应用研究[J]. 电网技术，2003，27(12)：67-70.
Li Fansheng，Chen Qingji. Application of decision tree classification algorithm in condition-based maintenance of power generation equipments[J]. Power System Technology，2003，27(12)：67-70.

[10] 朱六璋，袁林. 运用决策支持对象实现短期电力负荷预测[J]. 电网技术，2004，28(6)：59-62.
Zhu Liuzhang，Yuan Lin. Short-term load forecasting by use of decision support objects[J]. Power System Technology，2004，28(6)：59-62.

[11] 胡政，柳进，胡林献. 电网高峰负荷分析决策平台的设计与实现[J]. 电网技术，2005，29(6)：58-62.
Hu Zheng，Liu Jin，Hu Linxian. Design and implementation of power system peak load analysis and decision-making platform[J]. Power System Technology，2005，29(6)：58-62.

[12] 张文英，束洪春，张叶. 数据仓库技术在昆明电网电量分析中的应用[J]. 电网技术，2005，29(14)：40-44.
Zhang Wenying，Shu Hongchun，Zhang Ye. Application of data warehouse technology in power consumption analysis of Kunming power nerwork[J]. Power System Technology，2005，29(14)：40-44.

[13] 于达仁，胡清华，鲍文. 融合粗糙集和模糊聚类的连续数据知识发现[J]. 中国电机工程学报，2004，24(6)：205-210.
Yu Daren，Hu Qinghua，Bao Wen. Combining rough set methodology and fuzzy clustering for knowledge discovery from quantitative data[J]. Proceedings of the CSEE，2004，24(6)：205-210.

[14] 齐郑，艾欣，王炳革，等. 基于粗糙集理论的小电流接地系统故障选线方法的有效域[J]. 电网技术，2005，29(12)：43-46.
Qi Zheng，Ai Xin，Wang Bingge，et al. Effective domain of faulty line detection in small current grounding system based on rough set theory[J]. Power System Technology，2005，29(12)：43-46.

[15] Jiawei Han，Micheline Kamber. Data mining concepts and techniques[M]. 北京：机械工业出版社，2001.

收稿日期：2006-09-30。

作者简介：

林其友(1976—)，男，硕士，助理工程师，从事电力市场电价及输配电自动化方面的研究，E-mail：njhnlqy@sohu.com；

陈星莺(1964—)，女，博士，教授，博士生导师，主要从事电力系统分析与控制、电力系统经济运行(优化)、电力市场与电力经济等方面的研究与分析；

王之伟(1966—)，男，工学硕士，高级工程师，主要从事电力系统规划及电力市场方面的研究。

（编辑　王金芝）

（上接第 71 页 continued from page 71）

[10] 隼军平，盛万兴，王孙安. 新一代变电站自动化网络通信系统研究[J]. 中国电机工程学报，2003，23(3)：16-19.
Sun Junping，Sheng Wanxing，Wang Sun'an. Study on the new substation automation network communication system[J]. Proceedings of the CSEE，23(3)：16-19(in Chinese).

[11] 李兰欣，孙培略，余英. 基于 IEC 61850 的变电站自动化系统通信体系的研究[D]. 北京：中国电力科学研究院，2003.

[12] 丁书文. 数字化变电站自动化系统的网络选型[J]. 继电器，2003，31(7)：37-40.
Ding Shuwen. Choosing internal communication network of digital substation integrated automation system[J]. Relay，2003，31(7)：37-40(in Chinese).

[13] 高翔，刘韶俊. 继电保护状态检修及实施探讨[J]. 继电器，2005，33(20)：23-27.
Gao Xiang，Liu Shaojun. Condition maintenance and implementation of relay protection[J]. Relay，2005，33(20)：23-27(in Chinese).

[14] 曾庆禹. 变电站自动化技术的未来发展二——集成自动化、寿命周期成本[J]. 电力系统自动化，2000，24(20)：1-5.
Zeng Qingyu. The development of substation automation in the near feature part two-integrated automation，life cycle costs[J]. Automation of Electric Power Systems，2000，24(20)：1-5(in Chinese).

[15] NxtPHASE. Arizona public service-case study[DB/OL]. http：//www.nxtphase.com，2006，1.

[16] NxtPhase. NxtPhase optical sensors measurement stability over time and temperature[DB/OL]. http：//www.nxtphase.com，2006，1.

[17] Sanders G A，Blake J N，Rose A H，et al. Commercialization of fiber-optic current and voltage sensors at NxtPhase[C]. Optical Fiber Sensors Conference Technical Digest，2002.

[18] Skeie T，Johannessen S，Brunner C. Ethernet in substation automation[J]. IEEE Control Systems Magazine，2002，22(3)：43-51.

[19] Frances Cleveland，Xan thus Consulting International. IEC TC57：Security standards for the power system's information infrastructure -beyond simple encryption[S]. 2005.

[20] 朱大新. 数字式变电站综合自动化系统的发展[J]. 电工技术杂志，2001，4：20-22.
Zhu Daxin. The development of integrated automation system of digital transformer station[J]. Electrotechnical Journal，2001，4：20-22(in Chinese).

收稿日期：2006-04-23。

作者简介：

高　翔(1962—)，男，硕士，高级工程师，从事电网继电保护与自动化运行与管理工作，Email：gao_x@ec.sp.com.cn；

张沛超(1970—)，男，博士，副教授，主要研究方向为专家系统在电力系统中的应用及电网调度自动化技术。

（编辑　王金芝）