

基于知识发现的电力负荷预测模型

窦全胜^{1,2,4} 史忠植² 于尔铿⁴ 杨斌³ 刘仲尧⁴

(山东工商学院计算机科学与技术学院 烟台 264005)¹ (中国科学院计算技术研究所 北京 100080)²
(黑龙江省电力有限公司 哈尔滨 150010)³ (烟台东方电子信息产业集团有限公司 烟台 264001)⁴

摘要 负荷预测是电力系统的一个传统研究问题。针对黑龙江省的气象和经济特点,提出了基于知识发现的负荷预测模型。首先通过传统的近大远小方法生成基本预测曲线,并根据从历史气象资料及负荷数据中提取的规则加以修正,生成最终预测曲线。该模型在黑龙江省电网公司得以应用,收到了较为理想的效果。

关键词 负荷预测,知识发现

中图分类号 TP301.6 **文献标识码** A

Power Load Forecasting Model Based on Knowledge Discovery

DOU Quan-sheng^{1,2,4} SHI Zhong-zhi² YU Er-keng⁴ YANG Bin³ LIU Zhong-yao⁴

(School of Computer Science and Technology, Shandong Institute of Business and Technology, Yantai 264005, China)¹

(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China)²

(Heilongjiang Electric Power Company Limited, Harbin 150010, China)³

(Yantai Dongfang Electronics Information Industry Group Co., Ltd, Yantai 264001, China)⁴

Abstract Load forecasting is a traditional research field of power system, this work made an analysis for meteorological and economic characteristics of Heilongjiang Province, and put forward the load forecasting model based on knowledge discovery. First of all, generate the initial prediction curve by traditional method, and extract related rules from the historical meteorological data and load data, update the initial prediction curve using these rules, to generate the final prediction curve. The model has been used in software system of load forecasting of Heilongjiang power grid Co., Ltd, obtained desired results.

Keywords Load forecasting, Knowledge discovery

1 引言

负荷预测是电力系统的一个传统研究问题,是指从已知的电力、经济、社会、气象等情况出发,通过对历史数据的分析和研究,探索事物之间的内在联系和发展变化规律,对负荷发展做出预先估计和推测。关于电力负荷预测的方法众多,传统方法包括趋势外推法、回归法、时间序列法等。近年来,基于软计算和智能模型的预测方法不断涌现出来,主要包括神经网络法^[1-4]、专家系统法^[5,6]、模糊推理法^[7,8]、数据挖掘法^[9,10]等,这些方法代表了电力系统负荷预测研究的最新进展。

电力系统负荷预测是一个实践性很强的任务,不同国家、不同地区和不同时期的电力负荷特征有很大区别。电力系统负荷的这种多样性决定了没有一个通用的预测模型可以适用于所有的电力系统。

黑龙江省是我国重要的重工业及商品粮基地,对黑龙江省电力负荷进行准确的预测,不论是从经济角度还是安全角度都是十分重要的。本文对黑龙江省气象、经济特点进行了

分析,提出了基于知识发现技术(knowledge discovery)的电力负荷预测模型,并将此模型应用于黑龙江省电力负荷预报软件中。实践证明,该模型在所有预测模型中效果最好,适合黑龙江省电力负荷变化的具体特点。下面对该模型进行详细阐述。

2 负荷预测模型概述

在负荷预测的众多方法中,历史数据是预测的重要依据。通常情况下,相近日期的负荷曲线有较强的相似性。黑龙江省负荷曲线总体上也具有这样的特征,但是在一些特殊的季节或特殊天气里,相近日期的负荷曲线在局部上却有较大差异,这种差异主要由以下原因造成:

①黑龙江省属寒温带大陆性季风气候,四季分明,冬季漫长,昼夜温差大。在供暖期前后,电取暖会对负荷造成不规律的影响。

②黑龙江省风能资源丰富,各地年平均风速为 2m/s~4m/s。风力 $\geq 3\text{m/s}$ 的时间在松嫩平原、松花江干流谷地和三江平原为 4000h~5000h,主要出现在 3月~6月和 10月~

收稿日期:2009-05-12 返修日期:2009-08-11 本文受国家自然科学基金(60775035),国家重点基础研究发展计划资助项目(2007CB311004),国家高技术研究发展计划项目(2007AA01Z132)资助。

窦全胜(1971-),男,副教授,主要研究方向为智能科学、数据挖掘等, E-mail: li_dou@163.com.

11月。近年来,随着风电的大力推广,其对负荷的影响已经成了不可忽视的因素。

③黑龙江省是我国重要的农业大省、商品粮基地,在春夏季节,农田灌溉对负荷影响较大。

以上这些特点受季节和气象因素影响,且随机性较强。大部分预测方法对这种情况还不能较好地适应。目前,电力部门与气象部门合作,已经能够提供诸如温度、湿度、风力等详细的天气预报数据。随着历史负荷数据和气象等各类外部影响数据的积累,通过新兴的数据挖掘技术,从大量负荷数据和气象数据中发现负荷变化的规律性知识,已经成为可能。图1为本文所采用的基于知识发现技术的负荷预测模型。

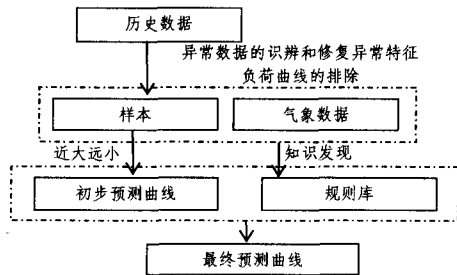


图1 基于知识发现的负荷预测模型

该模型首先根据修复后的历史数据,按照传统的近大远小原则,产生初步的预测曲线,同时从大量的负荷数据和气象数据中提取如下形式的规则,构成规则库:

If 实感温度大于 X and 前一天实感温度大于 Y Then 负荷值提高 $Z\%$

将这样的规则作用于之前得到的初步预测曲线,即可得到最终预测曲线。

3 基本预测曲线的生成

3.1 异常数据的辨识和修复

电力负荷历史数据中含有许多异常数据,它们对负荷特性分析和负荷预报产生不良的影响。异常数据的辨识与修复是负荷预测前必须要做的工作。异常数据的产生是随机的,分布也具有不确定性。在异常数据辨识与修复的众多方法中,小波方法是最常用的方法之一。以往的方法多是采用小波变换对96点数据本身进行分析,这样的方法对于一段相对较长的异常数据点会出现漏判。为此,本文采用双向小波方法对异常数据进行辨识和修复。

本文所采用的小波分解过程如下。设原始信号为 $p(t)$, 其小波分解过程为

$$\begin{cases} C_0 = p(t) \\ C_j[k] = C_{j-1} \bar{h}[2k] \\ D_j[k] = C_{j-1} \bar{g}[2k] \end{cases} \quad j=1, 2, \dots, M$$

其中, $\bar{h}[-k] = h[k]$, $\bar{g}[h] = g[-k]$, $g[k] = (-1)^{k-1} h[k-1]$ 。 $h[k]$ 为低通滤波器, $g[k]$ 为高通滤波器, M 为分解层数, $C_j[k]$, $D_j[k]$, $j=1, 2, \dots, M$ 分别为第 j 层小波变换的低频信号和低频信号。

96点负荷数据可以用一个二维矩阵 $\{L\}_{96 \times 96}$ 表示, 矩阵中的一行是某一天的96点数据。剔除 $\{L\}_{96 \times 96}$ 中的特殊数据, 对 $\{L\}_{96 \times 96}$ 逐行进行小波分解, 并对低频和低频信号进行重构, 可得到特征矩阵 $\{C\}_{96 \times 96}$ 和噪声矩阵 $\{H\}_{96 \times 96}$ 。同样, 对 $\{L\}_{96 \times 96}$ 逐列进行小波分解, 并对低频和低频信号进行重构,

可得到特征矩阵 $\{\hat{C}\}_{96 \times 96}$ 和噪声矩阵 $\{\hat{H}\}_{96 \times 96}$ 。对于 $\{L\}_{96 \times 96}$ 中的异常数据点无论从行还是从列上看都会与正常数据点存在差异。令

$$\Lambda = H + \hat{H}^T = \{\lambda\}_{96 \times 96}$$

$$\tilde{L} = \frac{1}{2} (C + \hat{C}^T) = \{\tilde{l}\}_{96 \times 96}$$

设置阈值 α , 当 $\lambda_{ij} > \alpha$ 时, $\{L\}_{96 \times 96}$ 中数据 l_{ij} 是奇异点, 用 \tilde{l}_{ij} 替换它。

3.2 初步预测曲线的生成

对异常数据进行修复之后, 还要进行异常曲线的排除。这里所说的异常曲线, 并不是异常数据造成的, 而是一些影响因素造成与众不同的负荷曲线。这样的负荷曲线会对负荷预报造成负面影响, 因此本文采用以下方法将其剔除。

将样本集中的每个96点数据看成是一个向量, 按欧式距离对样本集进行聚类, 通过聚类可以得到形如图2的聚类谱系。

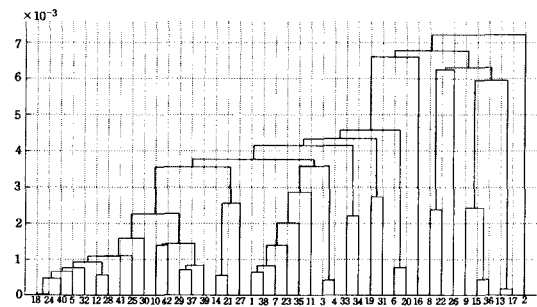


图2 样本聚类谱系

按照一定测度对聚类谱系进行切割, 除去特征异常的负荷曲线。

对处理后的样本按近大远小的原则加权处理, 生成初步预测曲线。

4 基于遗传算法的规则萃取

4.1 样本数据处理

将处理后的负荷数据融入相应的气象数据, 形成样本集。在样本集中, 气象等影响因素的变化决定了负荷曲线的变化。因此, 每一个样本都确定了一条规则, 如下式所示

$$X_1 \wedge X_2 \wedge \dots \wedge X_n \rightarrow Y$$

其中, X_1, \dots, X_n 为气象等影响因素, 称为条件属性; Y 是与之对应的负荷值, 称为决策属性。在负荷预测的过程中, 气象因素的累积效应是需要考虑的, 这里考虑了前3天气象变化对负荷的影响, 将前3天和当前的温度、风力、降水等影响因素作为条件属性。由于要考查的是相关气象因素的变化对负荷的影响, 因此在这里决策属性并不是负荷本身, 而是当天负荷值与前一天负荷值变化的百分比。

尽管人们通常用气温来反映环境的冷热, 但人体对外界冷热的舒适感并不能通过温度一项因素来评价。本文条件属性中的温度使用的是实感温度, 由以下公式计算

$$T_e = 37 - Q_1 / Q_2 - T_r$$

$$Q_1 = \Delta T (1.76 + 1.4V^{0.75})$$

$$Q_2 = (0.68 - 0.41R^h) (1.76 + 1.4V^{0.75}) + 1$$

$$T_r = 0.29T' (1 - R^h)$$

$$\Delta T = 37 - T_e$$

其中, T_e 为实感温度, T_a 为测量温度, V 为风速, R^h 为相对湿度, $Q_1, Q_2, T_c, \Delta T$ 为中间计算量。从定义上可以看出: 风速可以使实感温度降低, 但随着气温的升高, 对实感温度的影响越来越小。当温度超过 37°C 时, 风速增大会使实感温度升高; 当风速一定、气温较高时, 湿度的增加会使实感温度升高; 而气温较低时, 湿度增大会使实感温度降低。目前, 这些气象因素都可以从气象部门获得。

将条件属性离散化。设 R 为一条件属性, W 为属性 R 所有取值的集合。按实际需要将 W 划分成 N 个互不相交的子集, 即 $\bigcup_{n=1}^N U_n = W, U_i \cap U_j = \Phi, i \neq j$ 。以此为标准将属性 R 拆分成 N 个属性, 如图 3 所示。

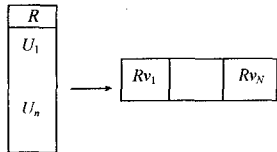


图 3 条件属性的离散

按实际需要, 将决策属性分成若干类, 条件属性与这些类别相对应, 进而形成新样本集。

4.2 基于遗传算法的知识萃取

1975 年 Holland 首次提出了遗传算法 (Genetic Algorithms GA) 的概念, 之后遗传算法在各个领域广为应用, 已经形成了一个比较活跃的研究领域。关于遗传算法的研究较多, 这里不再对其过程进行阐述。以下重点阐述遗传算法应用于负荷数据知识萃取所涉及的个体编码策略、评价策略和遗传操作。

1) 编码机制

对于任意条件属性 X_i , 假设离散后 X_i 被划成 M 个互不相交的子集, 用 M 位二进制编码即可对 X_i 进行表示, 则称这种刻画一个条件属性的二进制串为基因片段。显然, 在一个基因片段中只能有一位是 1。将决策属性也用基因片段表示, 放在染色体的末尾。表示决策属性的基因片段的长度与决策属性的分类数有关, 染色体具体形式如下

$$X_1 \wedge X_2 \wedge \dots \wedge X_i \wedge \dots \wedge X_n \rightarrow Y$$

$$001 \dots 001 \dots 0 \dots 010$$

$\underbrace{\hspace{10em}}_M$

2) 交叉和变异操作

遗传算法的遗传操作包括交叉和变异, 由于染色体的编码方式, 遗传操作不能破坏染色体基因片段, 因此个体的交叉操作是以基因片段为基本单位的。将染色体用基因片段表示, 其交叉过程为

$$\begin{matrix} S_1 S_2 \dots & S_i \dots S_n & \rightarrow & S_1 S_2 \dots S_i' \dots S_n' \\ S_1' S_2' \dots & S_i' \dots S_n' & \rightarrow & S_1' S_2' \dots S_i \dots S_n \end{matrix}$$

变异操作也是以基因片段为单位的, 表示决策属性的基因片段不发生变异, 而其他基因片段在变异时必须保证基因片段中至多只有一位为 1。

3) 个体的评价

为叙述上的方便, 以下用 X 表示 $X_1 \wedge X_2 \wedge \dots \wedge X_n$, 设每一个样本都确定的相应规则为

$$X \rightarrow Y$$

令

$$TP = |X \& Y|$$

$$FP = |\neg X \& Y|$$

$$FN = |X \& \neg Y|$$

其中, TP 为样本集中满足规则 $X \rightarrow Y$ 的样本个数, FP 为样本集中满足规则 $\neg X \rightarrow Y$ 的样本个数, FN 为样本集中满足规则 $X \rightarrow \neg Y$ 的样本个数。定义评价函数为

$$Fitness = \frac{TP}{(TP+FP)} \times \frac{TP}{(TP+FN)}$$

从上式可知, 评价函数由两部分作用而成, 其中 $\frac{TP}{(TP+FP)}$ 通过决策属性来对 $X \rightarrow Y$ 进行估价, 反映了条件属性是 X 且决策属性是 Y 的样本与所有决策属性为 Y 的样本的比例; 第二项 $\frac{TP}{(TP+FN)}$ 则从条件属性出发, 反映了所有条件属性是 X 且决策属性是 Y 的样本与所有条件属性为 X 的样本的比例。以下举例说明:

条件属性	决策属性
X	Y
X	Z
W	Y
X	Y
W	Z
U	Y

对于规则 $X \rightarrow Y$, 相应的 $TP=2, FP=2, FN=1$, 从而表示该规则的染色体评价函数值为 0.167。

4) 选择策略

本文采用赌盘选择策略来对下一代个体进行选择。令 f_i 表示种群中第 i 个染色体的适应度值, $\sum f_i$ 表示群体的适应度值之总和。由此, 染色体 i 存活到下一代的概率为 $P_i = \frac{f_i}{\sum f_i}$ 。染色体适应度值越大, 那么它被选择到的机会就越大。

采用上面所述遗传算法, 对黑龙江省历史负荷数据和气象数据进行处理。初始时刻种群规模为 20, 交叉和变异的概率分别是 0.7 和 0.05。多次运行, 每次运行都选择最后一代种群中估价函数值大于某一阈值的个体, 将个体解码后形成相应的规则。

在这里需要说明的是, 对于个体的评价需要对样本集进行遍历, 因此每一次规则的萃取都需要较长的时间。事实上, 并不是每一次负荷预报都需要进行规则的萃取, 只有在数据积累到一定程度时, 才重新对规则库进行修订, 这段时间可以是几个月或一年。

在具体预报的时候, 首先根据传统方法生成初步预测曲线 l_{init} , 然后考察相关影响因素是否与规则集中的某一规则的条件相符, 如果存在这样的规则, 则使用该规则对 l_{init} 加以修正, 生成最终的预测曲线。若不存在这样的规则, l_{init} 即为最终预报结果。

以 2005 年—2007 年黑龙江省的气象及负荷数据为样本, 采用上述方法进行规则萃取, 同时对 2008 年负荷进行预报。预报结果如表 1 所列。

表 1 预测结果统计

统计类型	比率
误差小于等于 1% 的点的比例	74% (26001 点)
大于 1% 小于等于 3% 点的比例	17% (5973 点)
误差大于 3% 点的比例	9% (3162 点)
平均误差	1.93%

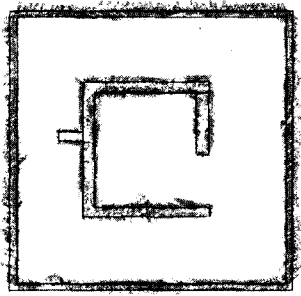


图6 DSmT 实验结果与实际地图比较(无串扰)

3)由图5的椭圆标记内的区域可以看出,多次反射和串扰现象在复杂环境下且有多个声纳同时工作时非常严重,极大地影响了环境信息的准确度。图5中椭圆标记区域内的“障碍物”都是多次反射和串扰所造成的。

4)图6则显示了本文提出的传感器管理方法的有效性,它消除了多次反射和串扰所带来的干扰,使得声纳信息更为精确。

5)由于在更新地图时采用了限制传播算法,因此在一个通信周期内只需要更新地图中很小的一部分,运算效率就会大为提高,加上计算简便的信度赋值函数,系统在构建地图时无延迟甚至崩溃现象。

结束语 本文采用最新的 DSmT 理论利用不精确传感器(声纳)的 DSmT 模型,构造了一组计算简便的基本信度赋值函数,有效地融合了移动机器人在探测动态未知环境时使用声纳获取到的不精确、不确定和高冲突环境信息。最后进行了经典 DSm 模型(有传感器管理)和 DST(无传感器管理)两种算法构建环境地图的实验,分别构建出了 DSmT 和 DST 框架下的二维基本信度赋值地图。两个实验的结果表明,采用相同的模型和融合步骤,在同一未知环境中,带传感器管理的 DSmT 融合算法的表现远优于无传感器管理的 DST 算法,是一种优秀的处理不精确、不确定和高冲突信息的融合方法,为移动机器人探测动态未知环境提供了有力的理论依据。

(上接第 214 页)

以上预报结果好于黑龙江省以往使用的负荷预测系统产生的结果。目前该模型已经在黑龙江省电网公司的负荷预报软件中得以应用。

结束语 本文对黑龙江省的气象和经济特点进行分析,提出了基于知识发现的负荷预测模型。首先通过传统近大远小方法生成基本预测曲线,再采用遗传算法从历史气象资料和负荷数据中提取相应的规则。这些规则作用于基本预测曲线,从而生成最终的预测结果。该模型在黑龙江省电网公司得以应用,收到了理想效果。

参考文献

[1] Hippert H S, Pedreira C E, Souza R C. Neural networks for short-term load forecasting: A review and evaluation[J]. IEEE Transaction on Power System, 2001, 16(1): 44-55

[2] Taylor J W, Buizza R. Neural network load forecasting with weather ensemble predictions[J]. IEEE Trans on Power Electron, 2002, 17(3): 626-632

[3] Kandil N, Wamkeue R, Maarouf Saad and Semaan Georges. An

参考文献

[1] 王卫华,陈卫东,席裕庚.基于不确定信息的移动机器人地图构建研究进展[J].机器人,2001,23(6):563-568

[2] Grisetti G, Tipaldi G D, Stachniss C, et al. Fast and accurate SLAM with Rao-Blackwellized particle filters [J]. Robotics and Autonomous Systems, 2007, 55(1): 30-38

[3] Begum M, Mann G K I, Gosine R G. Integrated fuzzy logic and genetic algorithmic approach for simultaneous localization and mapping of mobile robots [J]. Applied Soft Computing, 2008, 8(1): 150-165

[4] Dezert J. Foundations for a new theory of plausible and paradoxical reasoning [J]. Information and Security, 2002, 9: 13-57

[5] Burgard W, Fox D, Hennig D, et al. Estimating the absolute position of a mobile robot using position probability grids[A]// Proceeding of The National Conference on Artificial Intelligence (AAAI-1996) [C]. Oregon, 1996: 896-901

[6] Shafer G. A mathematical theory of evidence [M]. Princeton N. J.: Princeton University Press, 1976

[7] Dezert J, Smarandache F. Advances and Applications of DSmT for Information Fusion [M]. Rehoboth: American Research Press, 2004

[8] Dezert J, Smarandache F. Advances and Applications of DSmT for Information Fusion [M]. Rehoboth: American Research Press, 2006

[9] 李新德,黄心汉,王敏.基于经典 DSmT 的 Sonar 栅格地图创建[J]. 计算机应用研究, 2007, 24(3): 209-212

[10] Elfes A, Moravec H. High resolution maps from wide angle sonar[A]// IEEE Int. Conf. Robotics and Automation [C]. Leuven: IEEE, 1985: 116-121

[11] 王卫华.未知环境中移动机器人创建地图的研究[D].上海:上海交通大学, 2003

efficient approach for term load forecasting using artificial neural networks[J]. International Journal of Electrical Power & Energy Systems, 2006, 28(8): 525-530

[4] Mandal P, Senjyu T, Funabashi T. A neural network based several-hour-ahead electric load forecasting using similar days approach[J]. International Journal of Electrical Power & Energy Systems, 2006, 28(6): 367-373

[5] Han Hak-Guhn, Park Jong-Keun, Kang Sang-Hee, et al. Implementation of short-term load forecasting expert system in a real environment[J]. International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications, 2000, 8(3): 139-144

[6] Rao M S S, Soman S A, Menezes B L, et al. An Expert System Approach to Short-term Load Forecasting For Reliance Energy Limited, Mumbai [C] // IEEE Power India Conference. April 2006

[7] Torres T C. A Soft computing techniques for short term load forecasting[C] // Power Systems Conference and Exposition, 2004