

基于提升方法的短期负荷预测的研究

彭俊, 谢荣传, 王大刚, 耿波

(安徽大学 计算机学院, 安徽 合肥 230039)

摘要:目前有很多方法可以实现电力负荷预测,但各种方法都有各自的局限性,如何充分利用各种算法,使负荷预测的精确度提高,组合预测方法便应运而生了。然而当前的组合预测方法过于重视算法,而忽略样本。文中给出了一种基于提升方法的电力负荷预测挖掘算法,该方法通过样本权值进行抽样,训练生成模型,并根据模型效率改变样本权值,再次抽样训练生成模型,如此迭代生成多个模型后进行负荷预测。实际应用结果表明该方法比其它方法具有较高的精度。

关键词:提升方法;组合预测方法;样本权值;短期负荷预测;CART算法

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2008)07-0197-03

Study on Short-term Load Forecasting Based on Lifting

PENG Jun, XIE Rong-chuan, WANG Da-gang, GENG Bo

(Department of Computer Science and Engineering, University of Anhui, Hefei 230039, China)

Abstract: There are many algorithms have the ability to forecast short-term load, but each of them has disadvantage. How to use these algorithms to improve the forecasting precision? The combination forecasting method can solve the problem. The current combination forecasting method pay more attention on algorithm than samples. Propose a new method for forecasting short-term load based on the lifting. Use the method sampling by sample weight and then training the model. Change the sample weight by efficiency of model. Doing above repeatedly, generate multiple model to forecast load. The research shows that this method has higher accuracy.

Key words: lifting; combination forecasting method; sample weight; short-term load forecasting; CART

0 引言

电力系统负荷预测是实现电力系统安全和经济运行的前题,是调度部门的重要工作之一,随着现代电网运行调度的发展,特别是电力市场的建设与运营,高精度的负荷预测越来越重要。而短期电力负荷预测通常是指预测未来24小时负荷的需求,是电力系统的一项基本工作,它对于电力系统规划、运行与控制有着重要的意义。短期负荷预测的方法有时间序列分析法、灰色理论法、神经网络法和决策树方法等。其中,时间序列分析法由于难以建立有效的数学模型,致使预测精度普遍不高。灰色预测模型从理论上讲可以适用于任何非线性变化的负荷指标预测,但其微分方程指数解比较适合具有指数增长趋势的负荷指标,对于具

有其他趋势的指标,则拟合灰度大,精度难以提高。利用神经网络方法进行电力负荷预测时,前向神经网络可以通过学习,从复杂的样本数据中拟合出输入输出之间高维、非线性的映射关系,从而进行较高精度的预测。但是,电力负荷受到多种因素的影响,在天气突变、节假日和一些特殊情况下,电力负荷曲线的变化形态同普通工作日电力负荷曲线的变化形态差异较大,用一个前向神经网络无法对上述各种情况下的电力负荷均进行较好精度的预测^[1,2]。

决策树算法是以实例为基础的归纳学习算法,它可以从一组无次序、无规则的训练样本中推理出以决策树表示的分类规则集,然而它偏向选择属性取值较多的属性以及学习简单的逻辑表达式能力差。所以目前并没有一种方法能在任意情况下保证预测精度,目前很多负荷预测采用多种算法建立多个模型,然后分别预测后加权形成预测结果,这就是组合预测方法^[3,4]。然而组合预测方法过于重视算法,而没有充分考虑样本的复杂性。

文中以提升的方法为基础,充分考虑样本的复杂性,提出了一种改进的负荷预测方法。

收稿日期:2007-10-10

基金项目:安徽省自然科学基金项目(070412051);安徽高校省级重点自然科学基金项目(KJ2007A43)

作者简介:彭俊(1981-),男,安徽合肥人,硕士研究生,研究方向为数据仓库与数据挖掘;谢荣传,教授,硕士生导师,研究方向为网络与数据库。

1 提升方法概述

提升方法是一种组合预测方法,一般的组合预测方法以算法为基础,通过不同算法建立多个预测模型,然后根据每个预测模型的精确度对多个预测结果加权形成最终的预测结果。

提升方法以样本为基础,对总体中的样本赋予权值,根据权值抽样生成训练样本,建立预测模型,根据预测模型的精确度改变总体中某些样本的权值,然后再次根据权值抽样,生成训练样本,建立预测模型,从而提高预测精确度。如此迭代多次,建立多个预测模型,利用多个预测模型进行负荷预测得到多个预测结果,然后加权或大数表决形成最终的负荷预测结果。

2 负荷预测算法模型

2.1 基本思想

常用的负荷预测算法是寻找相似日,然后根据各种变化因素对相似日负荷进行修正。目前使用频率最高的方法是根据预测日气象因素来寻找相似日,其基理是气象相似的日子负荷相似。然后根据预测日与相似日的气象的差别进行修正。文中正是沿着这一被实践证明有效的思路,并对此思路使用新方法加以延伸。

通过将历史负荷数据与历史气象数据进行关联,生成样本总体,然后利用提升方法为每一个预测时间点(每天共 96 个时间点)建立多个预测模型。通过天气预报寻找历史上的天气相似日,该相似日作为负荷预测基准日,将负荷预测基准日的气象数据、负荷数据以及预测日的气象数据作为输入代入多个预测模型,得到多个预测结果后加权平均或大数表决生成最终预测结果。

2.2 天气相似日

天气相似日是指气象差异很小的那些日子,一般通过计算气象各种因素的欧几里得距离^[5]进行判断是否是天气相似日^[6,7],气象元素有很多,一般选择干球温度、相对湿度、雨量、风速、气压,而气象是按地区分片采集,当负荷预测涉及若干地区时,应该按各地区的负荷比重加权合成综合气象。

下面是计算两日气象相似性的公式:

预测日的天气预报 $X = \{x_1, x_2, \dots, x_M\}$,

某天的真实天气 $Y = \{y_1, y_2, \dots, y_M\}$,

$x_1, x_2, \dots, x_M, y_1, y_2, \dots, y_M$ 是气象的各个因素,如干球温度、雨量、风速、气压等。计算两天之间的相似性公式如下:

$$d = \left[\sum_{i=1}^M (x_i - y_i)^2 \right]^{1/2} \quad i = 1, 2, \dots, M$$

通过天气相似日计算可确定负荷预测基准日,而

距离预测日太长的日子作为预测基准日意义不大,所以一般设定一个时间段,在距离预测日时间段内寻找相似日,作为预测基准日。

2.3 建立预测模型

预测某日负荷就是预测该天的 96 个时间点(每 15 分钟记为一个时间点)的负荷情况。将历史负荷数据与历史气象数据按日期与时间匹配,组成一个数据表:历史电力气象表。表中每天的记录有 96 条,也就是每隔 15 分钟一条电力气象数据。为了预测结果清晰明了,可以分开预测每个时间点的负荷情况。将历史电力气象表按时间点划分成为 96 个历史电力气象子表,每个时间点预测模型所需的训练数据就保存在对应的历史电力气象子表。现在用提升的方法为某个时间点的负荷预测建立 K 个预测模型:

建模初始工作:首先将该时间点对应的历史电力气象子表 $T1$ 的每个样本赋一个初始权值 $1/N$ (N 为子表记录个数)。

建模具体过程:

1)按权值大小从历史电力气象子表 $T1$ 中抽取样本(权值大的样本优先抽取)组成训练样本集,一般抽取总样本个数的 80% 作为规则学习算法(如决策树 CART 算法, Apriori 算法等^[8])的训练样本集。

2)由规则学习算法通过训练样本集学习,提取规则,得到一个规则集,也就是一个预测模型。

3)用上一步骤生成预测模型对该历史电力气象子表 $T1$ 中的样本进行预测分类,这样便出现两种情况:对某个样本正确分类,或对某个样本错误分类。

4)根据以上出现的两种情况,那么正确分类的样本,它们的权值就必须在当前权值的基础上变小,错误分类的样本,它们的权值就相应地增大。权值增大或减小的量与生成模型个数相关。

例如用提升的方法为每个时间点建立 K 个预测模型,那么正确分类的样本的权值就变为: $(1 - 1/k) * q$, q 为该样本当前的权值。

错误分类的样本的权值就变为: $(1 + 1/k) * q$, q 为该样本当前的权值。

5)这样历史电力气象子表 $T1$ 中的样本权值就改变了,然后根据子表 $T1$ 中的样本权值进行再次抽样,组成训练样本集,通过规则学习算法在训练样本集上的学习并提取规则,再次得到一个预测模型。

此次的训练样本集包含了很多被前一个预测模型错误分类的样本(被错误分类的样本,它的权值变大)。因此规则学习算法从这样的一个训练样本集中提取规则,建立预测模型,那么上次被错误分类的样本由新生成的模型进行分类,被误分的可能性大大减小了。

6)以上过程迭代,直到产生的模型数目达到要求的数目。

这样用提升的方法为某个时间点建立 K 个预测模型的工作就结束了,用同样的过程建立其余时间点的 K 个预测模型。

2.4 模型预测流程

模型预测具体流程如下:

步骤1: 计算相似日(预测基准日):通过预测日的天气预报数据,用2.2介绍的欧几里得距离方法找出历史上与预测日气象相似的日子,作为预测基准日。

步骤2: 预测某个时间点的负荷:将预测基准日此时间点的气象数据、负荷数据,以及预测日此时间点的气象数据代入到对应此时间点的 K 个预测模型,共得到 K 个预测结果。通过对此 K 个预测结果进行加权平均或众数表决处理,得出最终的预测结果。

步骤3: 其余时间点的负荷预测与步骤2相同。

3 预测应用实例

此算例使用的是安徽黄山地区气象与负荷数据,利用历史数据预测2006年6月~2007年5月的负荷,用提升的方法与CART决策树算法分别预测,结果见表1。

表1 2006年6月~2007年5月预测结果

时间	基于提升方法的负荷预测 月度累计准确率(%)	基于 CART 决策树算法的负 荷预测月度累计准确率(%)
2006-6	96.83	96.94
2006-7	97.77	97.53
2006-8	98.28	98.48
2006-9	97.80	97.35
2006-10	97.43	96.47
2006-11	96.51	94.75
2006-12	97.40	96.60
2007-1	97.47	97.73
2007-2	97.17	96.13
2007-3	92.49	97.12
2007-4	96.80	96.19
2007-5	97.02	95.87
	97.41	96.76

从上表可看出使用提升方法进行负荷预测年度累计准确率提高了0.65个百分点,总之,该方法以历史数据中每15分钟气象因素变化与负荷变化相关联,通过提升方法建模获取每15分钟气象因素对负荷的影响规律,并据此作出更为准确的预测,实际运行表明,使用使用该方法,预测效果明显改善。

4 结 语

电力负荷的变化受众多的因素影响,如电力用户

特性(工业、农业、城市居民等)、节假日(春节、“五一”、“十一”、元旦等)、气象(气温、气压、雨量、风速、湿度)等等,这就导致了电力负荷预测是一项非常复杂和困难的工作,可以在以下几个方面改进,本算法的精度还可进一步提高,更加实用。

首先,提高气象要素预测的准确性。实践证明,气象因素是引起短期负荷变化的最主要原因之一,而气象预测精度直接影响负荷预测精度。

其次,加入节假日模型^[9]。诸如春节、“五一”、“十一”、元旦等长假,对负荷也有很大影响,但长假是近几年才有的,所以累积的资料不多,不便于进行数据挖掘。随着数据的积累,再考虑加入节假日模型,会使负荷预测精度提高。

最后,要研究系统拉电^[10]。由于系统拉电,造成电力负荷的实际数据不能反映负荷本身的变化规律,也不能反映负荷受其他因素影响的真实关系,由于数据挖掘是建立在历史数据基础上的,所以,如果发生大量的拉电现象,可能会影响数据挖掘效果。初步考虑对受拉电影响的数据进行人工标记,特殊处理,具体实施方案有待认真研究。

值得指出的是文中所提出的提升方法适用于各种规则学习算法。

参考文献:

- [1] 牛东晓,曹树华,赵磊,等.电力负荷预测技术及其应用[M].北京:中国电力出版社,1998.
- [2] 刘晨晖.电力系统负荷预报理论与方法[M].哈尔滨:哈尔滨工业大学出版社,1987.
- [3] 陈耀武,汪乐宇,龙洪玉.基于组合式神经网络的短期电力负荷预测模型[J].中国电机工程学报,2001,21(4):79-82.
- [4] 罗滇生,姚建刚,何洪英,等.基于自适应滚动优化的电力负荷多模型组合预测系统的研究与开发[J].中国电机工程学报,2003,23(5):58-61.
- [5] Han Jiawei, Kamber M. 数据挖掘概念与技术[M]. 范明, 孟小峰, 等译. 北京:机械工业出版社, 2001.
- [6] 卢毅,姜勇.基于相似日的神经网络短期电力负荷预测[J].电力系统及其自动化学报,2001,13(6):35-37.
- [7] 莫维仁,张伯明.短期负荷预测中选择相似日的探讨[J].清华大学学报:自然科学版,2004(1):29-36.
- [8] 闪四清,陈茵,程雁,等.数据挖掘——概念、模型、方法和算法[M].北京:清华大学出版社,2003.
- [9] 冯丽,邱家驹.粗糙集理论及其在短期电力负荷预测中的应用[J].电网技术,2004,16(6):60-63.
- [10] 李天云,孙艳鹤,范广良.基于弹性系数的电力系统短期负荷预测方法[J].吉林电力,2003(3):26-28.