

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/290447096>

# New unified approach to short-term load forecasting considering correlated factors (in Chinese 考虑相关因素的统一短期负荷预测新方法)

Article in *Dianli Xitong Zidonghua/Automation of Electric Power Systems* · September 1999

CITATIONS  
25

READS  
93

4 authors, including:



Chongqing Kang  
Tsinghua University  
527 PUBLICATIONS 4,469 CITATIONS

SEE PROFILE



Qing Xia  
Tsinghua University  
394 PUBLICATIONS 3,492 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Power system short-term operation optimization and decision-making methodology considering intermittent energy spatio-temporal correlation [View project](#)



Research on the technology development and environment management effects of photovoltaics and energy storage industry in China [View project](#)

# 一种规范化的处理相关因素的短期负荷预测新策略

康重庆 程 旭 夏 清 沈 瑜

(清华大学电机系 100084 北京)

**摘 要** 提出了一种规范化的可以直接考虑各种相关因素的短期负荷预测新策略。这种策略不仅可以考虑已经被详细研究过的气象因素,而且可以同时考虑一般性的分类指标,如工作日/休息日,正常日/节假日等。设计了一个规范化的相关因素映射数据库,用以反映不同影响因素在映射函数上的差异。根据模式识别的基本原理,应用聚类分析方法描述由于相关因素的不同而导致的待预测日与历史日之间的差异程度。作为这种预测策略的一个应用,实现了一种基于相关因素匹配的短期预测方法。实际算例表明,应用这种新的规范化的预测策略可以得到更加精确的预测结果。

**关键词** 短期负荷预测 模式识别 聚类分析

**分类号** TM 715

## 0 引言

负荷预测是从已知的用电需求和对此有影响的经济、气象等因素情况出发,探索用电负荷与主要影响因素之间的内在联系和发展变化规律,对未来用电需求作出预先的预测。负荷预测工作实质上是对电力市场需求的预测,是实现电网安全、经济运行的前提。为了准确地预测市场对电力这一商品的需求,必须认真分析各种信息与相关数据,采用最新的理论进一步提高预测精度。提高预测精度的途径之一是在预测过程中设法计及各种相关因素(如气象因素)对预测结果的影响<sup>[1,2]</sup>。

## 1 基于聚类分析的预测策略

### 1.1 现有方法分析

短期负荷预测是负荷预测的重要组成部分。国内外学者提出了多种短期负荷预测的方法,如多元回归<sup>[3]</sup>、谱分析<sup>[2,3]</sup>、ARMA 模型<sup>[2,3]</sup>、人工神经网络方法(ANN)<sup>[4]</sup>等。总结起来,主要可以分为以下几类:仅利用负荷自身发展规律的方法,如ARMA模型等;负荷发展规律与气象因素结合的方法,如ANN方法;其他方法。

第1类方法仅仅对历史负荷数据进行统计、分析、运算,而对于其他的相关信息特别是对短期负荷影响较大的气象信息没有进行考虑,使正常日的预测精度无法进一步提高,而特殊天气日更会造成误差过大。这是因为仅靠历史数据并不能很好地反映它未来的发展趋势,气象因素对短期负荷的影响很

大,但未能在这类算法中体现。

第2类方法考虑了气象因素,一般采用经验方法,利用粗略的气象情况进行补偿,或者作为相关元进入神经网络模型计算,但由于采用的信息太少和相关方式较弱,结果有时并不理想。此外,该类方法一般不涉及气象以外的因素,并且计及气象因素的方式也不灵活。

事实上,负荷预测考虑的相关因素绝对不仅仅是气象因素,而应该包括:日分类(正常日、国庆、春节等);星期类型(周一~周日);日期差(两日之间相距的天数);日天气类型(晴、阴等);日最高温度、日平均温度、日最低温度;日降雨量;湿度;风速等。随着科学技术的发展,有可能新增加其他相关因素。为了叙述方便,下文中称上述所有相关因素在某日的取值为该日的特征量。

因此,需要一种规范化的策略,可以直接考虑各种相关因素(不仅是气象因素)。该策略既可以指导预测人员构造新的短期负荷预测方法,也可以对各种现有的预测方法进行改造,使之可以计及各种因素的影响。

模式识别中的聚类分析是处理这些特征量的一种很好的思路<sup>[1,2,5]</sup>。本文应用聚类分析的基本原理,在不同日之间特征量的差异度、相似度概念的基础上,设计了一种规范化的映射数据库,并举例说明了具体预测方法的实现。

### 1.2 不同日之间特征量的差异度、相似度

这里首先引入各日的量化指标向量:记 $x_i = [x_{i1}, x_{i2}, \dots, x_{im}]$ 为第 $i$ 日的所有特征量的取值(假设共有 $m$ 个特征量),记 $L_i = [l_{i1}, l_{i2}, \dots, l_{iT}]$ 为第 $i$ 日的负荷曲线(假设共有 $T$ 个数据点,例如 $T = 24$ ,

$T = 48$  等), 则可将第  $i$  日的所有量化指标记为:  
 $D_i = (X_i, L_i)$ 。

于是可以引入不同日之间“差异度”的概念, 它描述任意两日由于特征量的差别而表现出的差异程度。抽象地, 设有  $i, j$  两天, 其因素量化指标分别为  $X_i, X_j$ , 则可用下述指标描述两天之间的差异度:

$$d_{ij} = \sqrt[q]{\sum_{k=1}^m |x_{ik} - x_{jk}|^q} \quad (1)$$

当  $q = 2$  时, 该差异度正是  $m$  维空间中两点之间的距离。

类似地, 可以引入“相似度”的概念, 它描述两天之间特征量的接近程度。可用下述指标计算:

$$r_{ij} = \frac{\sum_{k=1}^m x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^m x_{ik}^2} \sqrt{\sum_{k=1}^m x_{jk}^2}} \quad (2)$$

该相似度是  $m$  维空间中两向量之间的夹角余弦。

相似度与差异度可以用一定的尺度进行相互转换, 例如:

$$r_{ij} = 1 - \beta d_{ij}^\alpha \quad (3)$$

其中  $\beta, \alpha$  为适当的参数。

## 2 差异度与相似度计算的的实际处理

差异度  $d_{ij}$  越小(相似度  $r_{ij}$  越大), 表示  $i, j$  两天之间的日类型、天气类型、气象状况等各种因素“综合”意义上更接近, 考虑以下情况:

a. 一般情况下, 应将所有的  $x_{ik}, x_{jk}$  映射到  $[0, 1]$  区间上比较, 这样才有可比性。但是, 为了体现其中某些因素的“主导”差异, 可以使其在  $[0, a]$  上映射, 这里  $a > 1$ , 从而使该分量的差异在  $d_{ij}$  中显得比较强烈。

b. 对于常规的直接量化气象指标, 如日平均温度、日最高温度、日最低温度、降雨量、平均风速和湿度等, 应采取线性映射, 或分段线性映射(按几个阈值分段); 对于近大远小的日期差  $i - j$ , 采用线性映射, 对于星期类型(周一, 周二, ..., 周日)可采用分组映射, 如周一~周五映射为  $0.1 \sim 0.5$ , 周六、周日为  $3.0, 3.2$ , 从而加大休息日和工作日之间的差别; 对于日天气类型(晴、多云、阴等)应按分类映射, 但有一个排序, 表明负荷的递增和递减特性, 使相近的天气类型有较相同的负荷特性。

通过以上分析, 任意  $i, j$  两天, 无论星期类型(工作日、休息日)、日天气类型(阴、晴等)、日期差及各种气象指标的差异如何, 总可以通过  $d_{ij}$  (或者  $r_{ij}$ ) 衡量两天之间的差异度(或者相似度), 从而为相似日的选取打好基础。这是基于聚类分析的一种手段。

## 3 映射数据库的设计

由于各个特征量的量纲各不相同, 因此需要把不同量纲的值通过无量纲化处理, 映射到特定的区间, 使各个量之间可以有数值上的可比性, 从而方便相似度和差异度的定量计算。

### 3.1 映射库的构成

需要考虑的特征量包括 2 类。

a. 原始定量指标: 温度(最高温度、最低温度、平均温度等), 降雨量, 风速, 相对湿度等;

b. 可转化为定量指标: 日天气类型(阴、晴、多云、雨、雪、风等), 星期类型(周一, 周二, ..., 周日), 日期差(历史日与预测日相差天数, 1 天, 2 天, 等), 日分类(正常日, 元旦, 国庆, 春节等)等。

当需要考虑新的特征量时, 预测人员可自行加入。建立指标映射数据库如表 1。

表 1 指标映射数据库  
Table 1 Content of index-mapping database

特征量名称	特征量描述	映射前取值	映射后取值	特征量名称	特征量描述	映射前取值	映射后取值	特征量名称	特征量描述	映射前取值	映射后取值
日分类	正常日	1	0	星期类型	周五	5	0.5	日最高温度	温度 37	37	3.5
日分类	元旦前一天	2	1.4	星期类型	周六	6	3.0	日最高温度	温度 36	36	2.7
日分类	元旦	3	1.9	星期类型	周日	7	3.2	日最高温度	温度 35	35	2.0
日分类	元旦后一天	4	1.5	日期差	昨天	1	0.1	日最高温度	温度 34	34	1.8
日分类	国庆	5	3.5	日期差	...	...	...	日最高温度	温度 33	33	1.6
日分类	春节	6	4.5	日天气类型	晴	1	0.10	日最高温度	温度 32	32	1.4
日分类	...	...	...	日天气类型	阴	2	1.45	日最高温度	...	...	...
星期类型	周一	1	0.1	日天气类型	...	...	...	其他因素	...	...	...
星期类型	...	...	...	日最高温度	温度 38	38	4.5				

### 3.2 映射库的说明

#### 3.2.1 原始定量指标的映射

以最高温度为例。假设本地区高温阈值为 30 和 35。则 0 ~ 30 可以采用线性映射, 映射值有变化, 但变化不大; 而 30 ~ 35 采用另外一组线性映射, 映射值相互之间的差别比较明显; 35 以上采用非线性映射, 气温每增加 1, 其映射值变化很大。低温区域类似。

若日最高温度为主导气象因素, 则其映射区间可以超出  $[0, 1]$  的区间限制, 而对于非主导气象因素, 则其映射区间应限制在  $[0, 1]$ 。

#### 3.2.2 可转化为定量指标的映射

如表 1 中的示例。仅详细讨论以下 2 个指标。

##### 3.2.2.1 星期类型

由于星期类型在短期负荷预测中是占主导作用的影响因素, 故其映射区间可以映射到  $[0, 1, 3.2]$  的映射区间中, 以加大星期因素的作用。周一~周五的映射值很接近, 而周六和周日比较接近, 表明周一~周五是负荷类型相似的正常工作日, 而周六和周日是休息日, 这两组之间有较大的差别。区分工作日和休息日, 在聚类分析时比较有利。

##### 3.2.2.2 日分类属性

日分类属性在短期预测中的作用很大, 特别对于重大节日以及重大节日左右几天的影响比较大, 如果不加考虑, 必然会对预测结果产生较大的影响, 从而产生较大的预测误差。为了将正常日与节假日区别对待, 更有利于节假日的负荷曲线的预测, 下面对日分类属性进行讨论。

表 1 列出了具体的做法, 其特点是: 正常日编号可以编为 1, 均映射为 0。节假日的编号从 2 开始连续编号, 顺序无所谓, 但要进行分组映射。负荷特性比较接近的节假日可以分为一组, 映射的取值也较接近。对于不同的地区, 可能节假日的特点也各不相同, 因此, 不同地区可以设计自己的节假日映射数据库及其映射方式。

通过这种处理, 对于预测日为正常日/节假日寻找最佳历史匹配时有较大的不同, 而且可以自动识别最相近的节假日, 故可以将正常日/节假日统一在一起进行预测。

事实上, 这种方法还实现了一种历史数据的修正。例如, 预测 10 月 8 日的负荷时, 10 月 1 日在“星期类型”上最匹配, 但是在“日分类”上有可能不匹配, 其结果是 10 月 1 日可能对预测日的影响很小。

可以根据本地区特点添加或删除日分类, 视各地区的情况而定。但一般增加一个日分类后无需删除, 只需在实际日特征库中将该日置为 1 即可, 即当

做正常日处理。

不同地区认定的日分类不同, 故该数据表 1 可以包含地区索引。如有些地区认为端午节与其他正常日差别较大, 可将其列入重大节假日, 但其余地区不将其列入。

#### 3.3 映射数据库的修改

可以专门设计一个用户设置界面, 供用户修改映射数据库。用户可以通过这个界面反复预测、修改映射数据库的值, 摸索出适合本地区的映射关系, 直到满意为止。当然, 在以后不断的预测中, 各地区的负荷特性仍有可能改变, 虽然这种改变不是很剧烈, 因此用户仍可以根据负荷特性的变化将认为已经适合本地区的映射关系改变为更适应的映射关系。

若有新的气象因素指标, 可以直接在数据库中加入, 而不用修改计算程序, 大大减轻了工作量, 提高了程序的灵活性。

映射表可以由用户设置, 对于映射前取值不匹配的值, 可以采用最临近两点的线性插值。如温度为 32.5, 则可以在 32 到 33 之间找出对应的 1.40 到 1.60 之间的线性插值。

不同季节可以分别建立相应的合适的映射表, 比如冬季可能风速和温度是主导的气象因素, 而夏季可能主要的影响因素又变成了温度和湿度, 总之主导的气象因素可能会不相同, 而且相同因素的映射值在不同季节也有可能不相同, 这些改变都可以通过修改数据库很容易地进行。

由此可见, 建立此映射数据库的作用是: 使用户通过输入数据可以随意改变和建立复杂的映射模型, 而不是在程序中固化一段“死”的代码。这样, 若需要修改映射方程关系, 只需要修改数据库中的映射关系就可以了。

## 4 应用举例

这里仅举一种使用本文策略的新方法——基于相关因素匹配的短期预测方法。

对历史上的若干天, 取近期  $n$  天的数据作为预测样本集。已知条件为: 历史上各日的特征量  $x_{ik}$  ( $i = 1, 2, \dots, n; k = 1, 2, \dots, m$ ) 与负荷  $l_{it}$  ( $t = 1, 2, \dots, T$ )。这里  $m$  为特征量数目,  $T$  为每日的负荷采样点数,  $x_{ik}$  已经是量化值, 并且做了映射变换。已知待预测日的特征量为  $x_{0k}$ , 待求解的量是  $l_{0t}$ , 则预测步骤如下。

a. 求历史上各日与待预测日的相似度  $r_{i0}$ 。可以直接按式 (2) 求取  $r_{i0}$ , 也可按式 (1) 求得差异度  $d_{i0}$  后再变换为相似度  $r_{i0}$ 。

b. 相似度归一化:

$$r_{i0} = \frac{r_{i0}}{n} \quad (4)$$

$$r_{j0}$$

$$j=1$$

c. 待预测日负荷是历史上各日负荷的加权平均。(周一~周日及其他因素已体现在  $r_{i0}$  中, 这里无需区分)。

$$\hat{l}_{0t} = \sum_{i=1}^n (r_{i0} l_{it}) \quad t = 1, 2, \dots, T \quad (5)$$

使用上述预测方法对某电网实际数据进行了模拟预测。预测结果如图 1、图 2 所示。

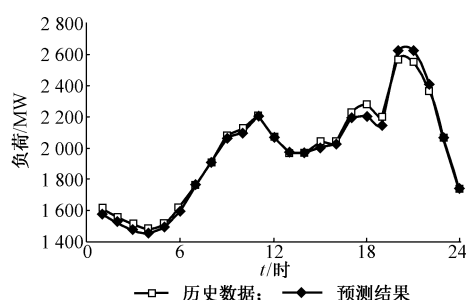


图 1 预测效果与实际历史数据的对比  
Fig. 1 Comparison of forecasting results with historical data

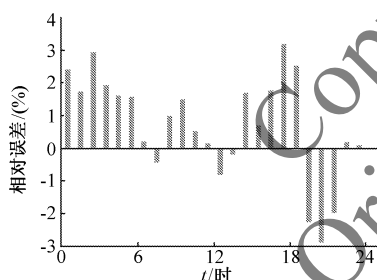


图 2 各时的预测误差  
Fig. 2 Hourly percent error of forecasting result

由图可见, 预测结果令人满意, 相对误差绝对值小于 1% 的有 10 个点, 1%~2% 之间有 8 个点, 大于 2% 的有 6 个点, 最大误差 3.16%, 平均的绝对值相对误差为 1.40%。

## A NEW UNIFIED APPROACH TO SHORT-TERM LOAD FORECASTING CONSIDERING CORRELATED FACTORS

Kang Chongqing, Cheng Xu, Xia Qing, Shen Yu (Tsinghua University, 100084, Beijing, China)

**Abstract** A new unified approach to short-term load forecasting is proposed which is capable of dealing with most of the factors correlated with power system load. This approach can consider not only weather which has been studied thoroughly in some previous literatures, but also some classification indexes such as weekday/weekend, workday/holiday, etc. An index-mapping database is designed to show the different mapping functions for various factors. According to the basic principles of pattern recognition, cluster analysis technique is used to describe the difference between the future day and past days. As its application, a new short-term load forecasting method named factor matching method is put forward. Case study shows that more accurate forecasting results can be got by the proposed method.

**Keywords** short-term load forecasting pattern recognition cluster analysis

## 5 结语

本文提出了一种规范化的可以直接考虑各种相关因素的短期负荷预测策略。考虑的相关因素非常广泛, 不仅仅是气象因素。利用该策略, 预测人员既可以构造新的短期负荷预测方法, 也可以对各种现有的预测方法进行改造, 使之可以计及各种因素的影响。该策略的核心思想是: 设计一个规范化的相关因素映射数据库, 可根据不同地区确定适当的参数, 然后应用聚类分析的原理, 对待预测日与任意的历史日之间求取相似度, 作为负荷预测的依据。这一策略为在短期负荷预测中考虑更复杂的相关因素提供了新思路。文中以一种采用这种策略的短期负荷预测的新方法为例, 并通过实际预测表明其有效性。

## 参考文献

- 1 Dehdashti A S, Tudor J R, Smith M C. Forecasting of Hourly Load by Pattern Recognition—A Deterministic Approach. IEEE Trans on Power Apparatus and Systems, 1982, 101(9): 3290~3294
- 2 孙洪波. 电力网络规划. 重庆: 重庆大学出版社, 1996
- 3 Mohamed A A, Naresh S K. Short-Term Load Demand Modeling and Forecasting: A Review. IEEE Trans on Systems, Man, and Cybernetics, 1982, 12(3): 370~382
- 4 Park D C, El-Sharkawi M A, Marks R J. Electric Load Forecasting Using Artificial Neural Network. IEEE Trans on Power Systems, 1991, 6(2): 442~449
- 5 黄振华, 吴诚一. 模式识别. 杭州: 浙江大学出版社, 1991

康重庆, 男, 1969 年生, 博士, 讲师, 从事电力系统自动化的研究。

程旭, 男, 1977 年生, 硕士研究生, 从事电力市场的研究。

夏清, 男, 1957 年生, 教授, 博士生导师, 长期从事电力系统规划与运行的研究, 近期主要研究方向为电力市场。