ZHINENG WAJUE DIANLI FUHE YUCE YANJIU JI YINGYONG

智能挖掘电力负荷预测研究及应用

王建军 著

中国水利水电出版社
www.waterpub.com.cn

智能挖掘电力负荷预测研究及应用

王建军　著

中国水利水电出版社

内 容 提 要

本书介绍了电力负荷预测的概念和基本原理并主要介绍了电力负荷预测的方法。首先介绍了典型的电力负荷预测方法，有经典预测方法、回归分析、时间序列分析、灰色预测、组合预测方法、神经网络和支持向量机智能预测方法；接着介绍了如何结合知识挖掘技术和智能预测方法进一步得到精确预测结果，包括寻找相似日、优化智能预测方法的参数和后干预纠偏技术。在此基础上，介绍了预警技术和软件实现。

本书可以作为高等院校电力管理相关专业本科、研究生的学习参考教材，也可以供从事电力规划、调度、营销等管理人员参考使用。

图书在版编目（CIP）数据

智能挖掘电力负荷预测研究及应用 / 王建军著. -- 北京 : 中国水利水电出版社, 2013.9
ISBN 978-7-5170-1295-5

Ⅰ. ①智… Ⅱ. ①王… Ⅲ. ①电负荷－预测－研究 Ⅳ. ①TM715

中国版本图书馆CIP数据核字(2013)第236116号

书　　名	**智能挖掘电力负荷预测研究及应用**
作　　者	王建军　著
出版发行	中国水利水电出版社 （北京市海淀区玉渊潭南路1号D座　100038） 网址：www.waterpub.com.cn E-mail：sales@waterpub.com.cn 电话：（010）68367658（发行部）
经　　售	北京科水图书销售中心（零售） 电话：（010）88383994、63202643、68545874 全国各地新华书店和相关出版物销售网点
排　　版	中国水利水电出版社微机排版中心
印　　刷	三河市鑫金马印装有限公司
规　　格	170mm×240mm　16开本　10.25印张　190千字
版　　次	2013年9月第1版　2013年9月第1次印刷
印　　数	0001—1000册
定　　价	**30.00**元

前　言

电力工业是国家的重大基础行业，对于我国经济建设、国家安全、社会稳定、居民生活质量具有至关重要的作用。电力负荷预测是电力系统调度、用电、计划、规划等管理部门的重要基础工作之一，精确的电力负荷预测对于制定发电计划、制定经济合理的电力调配计划、制定上网竞价计划、控制电网经济运营、降低旋转储备容量、进行电力市场需求分析、避免重大事故、有效化解风险、保障生产和生活用电方面具有十分重要的意义。

负荷预测的核心问题是建立合适的数学预测模型，而电力预测模型自20世纪20年代开始就有学者对其开始进行研究，但由于当时的电力系统规模小，变化较为平稳，因此电力负荷预测没有受到重视。而随着电力系统的市场化进程以及对能源的空前重视，使得负荷预测受到了更加广泛的重视。产生了诸如单耗法、弹性系数法、线性回归模型、时间序列模型、灰色预测法、专家系统法、模糊数学法、优选组合法、小波分析法以及近二三十年来研究较热的以神经网络和支持向量机为代表的智能预测方法等。这些方法均在一定的适用条件下发挥了较好的效果，证明了模型的适用性，但是也同时带来了一定的困惑，使得相关人员越来越难以理解和难以选择合适的模型。因此，本书的第一个目的是介绍较为常用的各类负荷预测方法，研究其模型结构、功能特点、适用范围，了解各类方法在电力系统的实际应用情况，以利于搞好负荷预测的工作，提高负荷预测人员的技术水平。

此外，本书也对深层次的负荷预测问题进行了一定的思考，由于电力负荷的影响除了包括负荷自身特性、经济人口等定量型因素的影响外，同时也包含着不规则事件、日期类型、季节类型、描述性天气等非数字型的定性因素的影响，因此，电力负荷预测的工作

是一件极其复杂的问题。如果不考虑这些定性因素的影响，无论如何改进负荷预测模型，预测精度都很难有根本性的提高，负荷预测理论也难以有较大的突破。因此，本书提出了基于知识挖掘技术的智能协同电力负荷预测研究思想，旨在结合知识挖掘技术和智能电力负荷预测方法进行协同电力负荷预测，通过知识挖掘直接对数据库中的负荷变量属性及对应的各类影响因素变量属性进行分析处理，在预测时通过计算与预测目标各类知识特征的总体关联程度大小，自动提取具有高度相似性综合知识特征的同类历史数据，再结合智能算法和电力负荷预测方法建立具有针对性的自适应结构的智能预测模型对负荷进行预测，在遇到有少部分具有较大的预测误差点时利用知识挖掘形成的纠偏规则进行相应的后干预工作，能够进一步克服以前的预测方法的不足，使预测精度得到突破性的提高，以期为负荷预测的深层次研究起到抛砖引玉的作用。

本书是在笔者的博士论文基础之上扩编而成的，编写过程中既考虑介绍基本的电力负荷预测概念和理念，以供相关专业本科生和从事电力工作并打算了解电力负荷预测的读者使用，也提供了结合知识挖掘后的负荷预测研究供研究生深入研究参考使用。书中的例子均引用自电力系统中的实际案例，具有较好的代表性。为了增强书中例子的实用性，其中大部分例子均配有相应的 matlab 源程序，源程序的获取可在笔者的个人网页 www. wangjianjun. net 上下载。

书中的模型和方法，基础部分参考笔者博士导师华北电力大学长江学者牛东晓教授所著《电力负荷预测技术及其应用（第二版）》一书，其余部分模型借鉴博士期间笔者和华北电力大学预测研究所等人的研究成果，在此对他们表示诚挚的谢意。

笔者由衷地感谢中国水利水电出版社的大力支持，感谢谢维编辑的热忱工作，不断为本书润色、修正，并提出许多宝贵意见。华北电力大学经济与管理学院，以及信息管理教研室的各位老师为本书的撰写创造了良好的条件，笔者在此一并表示感谢，本书中结合知识挖掘技术的智能协同电力负荷预测研究部分得到了国家自然基金项目（No. 71071052）以及中央高校基本科研业务费资助的支持，特此致谢。

北京信息科技大学李莉博士审阅了全书的初稿并进行了校对，

在此表示深切的谢意。

由于作者的研究水平及经验有限，书中仍然可能存在不足之处，希望读者给予批评指正。

王建军

2013 年 5 月

于华北电力大学

目 录

第1章 电力负荷预测概述

1.1 研究电力负荷预测的目的和意义

电力工业是国家在能源领域的重大基础行业，电力是国民经济的命脉，经济要发展，电力是先行，电力对于我国经济建设、国家安全、社会稳定、生活质量具有至关重要的作用。随着国际上电力市场的逐步建立以及国内经济形势的快速发展，电力供求矛盾日益严峻，电力需求的影响因素逐渐增多，传统的电力工业负荷预测理论方法已经不再适用，适应新环境下的负荷预测理论方法研究迫在眉睫。准确及时的电力负荷预测对于制定经济优化的发电计划、制定经济合理的电力调配计划、制定上网竞价计划、在竞价上网中取得优势、最优制定电力现货和期货报价、控制电网经济运营、降低旋转储备容量、进行电力市场需求分析、搞好电力市场营销电力客户关系管理、避免重大事故、有效化解风险、保障生产和生活用电等方面均具有十分重要的意义。

电力负荷预测工作作为电网管理部门的基础工作之一，能够为电网企业以及整个电力的发、输、配、送的电力工业链条直接产生重大的经济效益和社会效益。国外学者早在20世纪80年代中期的研究成果中就已经表明，负荷误差提高1%将会增加10000000英镑的电力经营成本。类似地，对于我国的一个中等规模的省级电网而言，按照常规假设其平均供电负荷为4500MW，如果将系统日负荷预测精度提高1%，就表示在系统发供电可靠率相同的条件下，电网发电出力富裕时可减少50MW的旋转备用容量，电网发电出力不足时可减少非计划限电45MW，由此产生的主要效益为：因系统减少旋转备用容量产生的年经济效益2000万元，减少非计划限电增加售电量产生的年经济效益为446万元，共创电网年经济效益2246万元，假设每千瓦时电产值为5.3元、每千瓦时电的边际利润为0.07元、非计划限电电量损失率为0.8元、全年限电日253天考虑，因系统减少非计划限电产生年社会上的经济效益将高达3.38亿元。此外，按照常规假设，我国火力发电所占比率为75%，以我国2009年上半年的统计数据为依据，供电的煤耗率为341g/(kW·h)，火电设备平均利用小时为2934h，由此计算得的由于负荷精度提高1%而产生的减少50MW备用容量将节约年用煤量60000余t。由此可见，负荷预测精度的提高不但具有十分巨大的经济效益和社会效益，同时也对于我国的节能减排、资源

利用，以及可持续发展具有非常重大的意义。

1.2 电力负荷预测的相关概念

1.2.1 电力负荷预测概念及基本原理

电力负荷预测是指根据电力系统的运行特性、增容决策、自然条件与社会影响等诸多因素，在满足一定精度要求的条件下，确定未来某特定时刻或某些特定时刻的负荷值。其中的电力负荷指的是广义电力负荷，即除了电力用户常说的某一时刻的用电功率的狭义电力负荷外，还包括在电力工业整个环节中的电能生产的发电量、电能供给的供电量、电力企业销售给用户的售电量以及用户消耗的用电量等。由于电力负荷预测是根据电力负荷的历史数据和目前已知的或者是可预测的相关变量状态对未来的数值进行推测，因此需要利用预测工作的基本原理，用于指导负荷预测工作，其基本原理如下。

1. 可知性原理

预测对象的发展规律未来的状况是可以为人们预知的，其所在的客观世界是可以被认识的，人们不但可以认识它的过去和现在，而且可以通过总结它的过去和现在推测其未来的发展趋势。这是人们进行预测活动的基本依据。

2. 可能性原理

因为事物的发展变化是在内因和外因共同作用下进行的。内因的变化及外因作用力大小不同，会使事物发展变化有多种可能性。因此，对某一具体指标的预测往往是按照其发展变化的多种可能性进行多方案预测的。

3. 连续性原理

预测对象的发展是一个连续统一的过程，其未来发展是这个过程的延续。它强调了预测对象总是从过去发展到现在，再从现在发展到未来。可以认为事物发展变化过程中会将某些原有的特征保留下来，延续下去。电力系统的发展变化同样存在着惯性，如某些负荷指标会以原有的趋势和变化率发展下去。这种惯性正是负荷预测的主要依据。因此，了解事物的过去和现在，并掌握其变化规律，就可以对其未来的发展情况利用连续性原理进行预测。

4. 相似性原理

尽管客观世界中各种事物的发展各不相同，但一些事物发展之间还是存在着相似之处，可以利用这种相似性进行预测。在很多情况下，作为预测对象的现在发展过程和发展状况可能与另一事物过去一定阶段的发展过程和发展状况相类似，这时就根据后一事物的已知发展过程和状况，来预测所预测对象的未来发展过程和状况，这就是相似性原理，如传统预测技术中使用的类推法或历

史类比法就是基于这个原理的预测方法。例如，当我们预测一个新的经济开发区的用电量时，由于其建成时期较短，没有很多历史数据可利用时，就难以利用趋势外推、回归分析等方法建模预测。这种情况下，可以参考一个早已建成的、规模和条件具有可比性的其他经济开发区，用其发展时期相对应的用电量作为预测新经济开发区用电量的基础，从而可以作出相应的预测结果。

5. 反馈性原理

反馈是利用输出返回到输入端再调节输出结果的过程。预测的反馈性原理实际上是为了不断提高预测的准确性而进行的反馈调节。人们在预测活动实践中发现，当预测的结果和经过一段实践所得到的实际值存在着差距时，可利用这个差距，对远期预测值进行反馈调节，以提高预测的准确性。在进行反馈调节时，可以首先认真分析预测值和实际值之间的差距及产生差距的原因，然后根据已经查明的原因，适当改变输入数据，进行反馈来调节远期预测结果。反馈性预测实质上就是将预测的理论值与实际相结合，在实践中检验，然后进行修改、调整，使预测质量进一步提高。

6. 系统性原理

预测对象本身不但具有内在的系统，而且由于预测对象和外界事物之间的联系同样也形成了它的外在系统。这些系统综合成一个完整的总系统，在预测时都要加以考虑。即预测对象的未来发展应该是系统整体的动态发展，并且整个系统的动态发展应该与它的各个组成部分和影响因素之间的相互作用和相互影响密切相关。系统性原理同时还强调系统整体最佳，只有系统整体最佳的预测，才是高质量的预测，才能为决策者提供最佳的预测方案。

1.2.2 电力负荷预测的分类

1. 按时间进行分类

在电力负荷预测的研究中，按时间期限进行分类的方法是最常见的分类方法，按照这类分类方法，电力负荷预测可分为中长期、短期和超短期负荷预测。其中中长期负荷预测一般指 1 年以上、10 年以下的以年为单位的预测。短期预测包括一年以内以月、周、日为单位的电力负荷预测，一般预测的是对未来一个月度、未来一周、未来一天的最高负荷、最低负荷或平均负荷等负荷特性指标。超短期负荷预测指的是预测一天以内的间隔 1h、0.5h、15min 乃至更短的时间内的电力负荷预测。

2. 按国民经济行业用电进行分类

国民经济用电分类是说明国民经济各部门用电情况和变化规律，它是反映电气化的发展水平和趋势的指标，用于分析研究经济增长与电力生产增长、社会产品增长与电力消耗量增长的相互关系，是负荷预测和电力分配的依据。我

国的《国民经济行业分类》标准于 1984 年首次发布，分别于 1994 年和 2002 年进行修订，最近一次是由 2011 年第 3 次修订。该国民经济行业分类的标准是由国家统计局起草，国家质量监督检验检疫总局、国家标准化管理委员会批准发布，并将于 2011 年 11 月 1 日实施。一共分为 20 个门类、95 项大类、396 个中类和 913 项小类。详细分布见表 1－1。

表 1－1　《国民经济行业分类》(GB/T 4754—2002) 标准分布图 (2011 年标准)

门　类	大　类	中　类	小　类
A 农、林、牧、渔业	5	18	38
B 采矿业	6	15	33
C 制造业	30	169	482
D 电力、燃气及水的生产和供应业	3	7	10
E 建筑业	4	7	11
F 交通运输、仓储和邮政业	9	24	37
G 信息传输、计算机服务和软件业	3	10	14
H 批发和零售业	2	18	93
I 住宿和餐饮业	2	7	7
J 金融业	4	16	16
K 房地产业	1	4	4
L 租赁和商务服务业	2	11	27
M 科学研究、技术服务和地质勘查业	4	19	23
N 水利、环境和公共设施管理业	3	8	18
O 居民服务和其他服务业	2	12	16
P 教育	1	5	13
Q 卫生、社会保障和社会福利业	3	11	17
R 文化、体育和娱乐业	5	22	29
S 公共管理和社会组织	5	12	24
T 国际组织	1	1	1
合　计	95	396	913

按照上述的国民经济行业分类，可以对相应的用电负荷进行分类，较为常用的负荷预测有工业负荷预测、商业负荷预测、居民用电负荷预测，以及国民经济用电量较大的用户的大用户负荷预测等。

3. 按特性分类

根据负荷预测表示的不同特性，电力负荷预测常常又分为最大负荷（峰荷）、最小负荷（谷荷）、平均负荷、全网负荷、母线负荷等。

1.2.3 电力负荷预测的误差衡量

由于电力负荷预测工作是一种对未来电力负荷数值的不确定估算工作，因此必然会和客观实际数值存在着一定的预测误差，目前衡量负荷预测的误差采用的指标有很多，下面介绍几种较为常用的误差衡量方法。

1. 绝对误差与相对误差

绝对误差用于衡量预测值和实际值之间的差异程度，设 Y 表示实际值，Y' 表示预测值，则称 $Y-Y'$ 为绝对误差，通常用大写字母 E 表示，相对误差值用于衡量预测值和实际值的差异比例，即在绝对误差的基础上除以 Y，得到相应的比值，通常用小写字母 e 表示。其计算公式如下：

$$e=\frac{Y-Y'}{Y}$$

绝对误差与相对误差是最直观的表现电力负荷预测误差的表示方法。在电力系统中作为一种考核指标经常被使用。而相对误差经常也用百分比的形式予以体现，即在相对误差公式的基础上乘以 100%。

2. 绝对值相对误差（*APE*）和平均绝对值相对误差（*MAPE*）

在进行相对误差的衡量上，由于负荷预测主要关注的是预测精度，从而使得相对误差的正负偏离衡量意义并不是很大。此外，当对于多个预测值进行综合评价时，正负偏离带来的加和会出现一定的抵消情况，因此，一般情况下，对于负荷逐点的评价指标可以采用绝对值相对误差代替，其计算公式如下所示：

$$APE=\left|\frac{Y-Y'}{Y}\right|\times 100\% \tag{1-1}$$

对于多个负荷点或者是负荷预测模型对于多点误差采用的评价指标是平均相对误差（*MAPE*），其计算公式分别如下所示：

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y-Y'}{Y}\right|\times 100\% \tag{1-2}$$

式中：n 为数值个数。

3. 均方误差

$$MSE=\frac{1}{n}\sum_{i=1}^{n}E_i^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2 \tag{1-3}$$

式中：*MSE* 为均方差；其他符号意义同前，用于还原平方失真程度。

均方误差是预测误差平方之和的平均数，它避免了正负误差不能相加的问题。是误差分析的综合指标法之一。

4. 均方根误差

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} E_i^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{1-4}$$

式中：$RMSE$ 为均方根误差；其他符号意义同前。

均方根误差在均方误差进行了开方处理，将误差的量纲和负荷的量纲进行了统一。

5. 标准误差

$$S_Y = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{n-m}} \quad (i=1, 2, \cdots, n) \tag{1-5}$$

式中：S_Y 为预测标准误差；n 为历史负荷数据个数；m 为自由度，也就是变量的个数，即自变量和因变量的个数的总和。

标准误差是利用统计上的误差衡量方法对负荷预测结果的误差进行衡量。

1.3　电力负荷预测的基本步骤

电力负荷预测的基本步骤如下。

1. 确定负荷预测目的，制订预测计划

首先需要依据负荷预测的需求，紧密联系实际情况，确定预测的目的，并根据目的制定负荷预测工作计划。在确定目的时需要考虑的问题有：

(1) 确定预测精度。在制定计划时，需要明确预测的类型属于哪一类，其要求的精度大致有多少，如果是短期负荷预测，一般要求误差在 3%以内，如果是中长期负荷预测，一般要求误差在 10%以内。

(2) 确定搜集历史资料序列的数量。按照预测类型，需要明确搜集的历史资料类型（按年、按季、按月、按周或按日）、需要多少项资料、资料的来源和搜集资料的方法、预测的方法、预测工作完成时间、所需经费来源，等等。关于所需资料项数多少，说法不一。有人主张外推预测的时期数不能超过历史资料的时期数，如设 d＝历史资料时期数，h＝外推预测时期数，则有 $d \geqslant h$。也有人认为，这种要求低估了短期预测所需项数和高估了长期预测所需项数，主张用 $d=4$ 计算。按此式，如向前预测 1 期，则 $A=1$，$d=4$，即需要 4 期历史资料；如向前预测 4 期，则需 8 期历史资料；如向前预测 100 期，就要用 40 期历史资料即可。可见，用这个公式计算，照顾了短期预测的需要，却不利于长期预测。实际上，根据长期的历史资料进行短期预测，要比根据短期的历史资料进行长期预测更可靠些，因为这样根据更充分些。

2. 搜集资料并整理

由于影响负荷预测的因素有很多，因此，在搜集资料时，不但要搜集负荷

序列本身的资料，还要搜集其他相关数据的资料，如搜集国民经济有关部门的资料。在挑选资料时，基本遵循的原则是相关性、可靠性、资料的新旧程度。一般认为，负荷预测的数据需要依据“近大远小”的特征，即离负荷时点越近的数据，其预测参考价值越大。资料搜集质量的高低将直接影响预测精度的高低。

搜集来的质量需要进行必要的统计审核和加工整理，为预测创造合适的数据条件，搜集来的资料在整理时需要剔除明显不正常的数据、以及弥补缺失值。在利用模型进行预测时，需要依据模型特点对影响因素的数据项进行时间间隔一致化、标准化和量纲一致化处理。进行整理数据序列时依据的数学方法有很多，有时为了精度需要甚至在进行非正常值和缺失值计算时就需要建立预测模型进行修正，即大的预测问题下利用小的预测模型对子问题进行预测。这部分问题可以利用知识挖掘的手段予以解决，详见本书知识挖掘的数据预处理部分。在这里，简单介绍一下电力工业中普遍利用的对数据进行整理的经验方法。

(1) 缺失值的经验补缺。如果缺失值的两端时刻有数据，可利用相邻两边资料取平均值近似代替，如果缺失值处于序列的开头或结尾，可以利用趋势比例计算予以代替。

(2) 异常值的处理。首先可以对搜集来的同一类型的资料数据进行散点图的绘制，从图形中区分资料数据序列中波动比较大的点，一般在电力负荷预测中认为波动超过附近点 20%以上的数据为异常值，需要仔细核实。确认是异常值后，需要对其进行处理，处理的规则是：如果设负荷历史数据为 x_1，…，x_n，令 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$，若 $x_i > \bar{x}(1+20\%)$，取 $x_i = \bar{x}(1+20\%)$；若 $x_i < \bar{x}(1-20\%)$，取 $x_i = \bar{x}(1-20\%)$，从而使历史数据序列趋于平稳。

此外，对于异常值的处理，也可以将异常值删除掉，按照缺失值的处理方法进行处理。

(3) 量纲规范统一处理。对于不同的负荷预测影响因素，有可能统计的口径和量纲不尽相同，因此，在预测时需要对量纲进行统一，比较简单的做法是将每一类的资料统计值均除以该序列的最大值，将所有的影响因素都归为[0，1]区间内，使其量纲为 1。

3. 对资料进行初步分析，确定影响因素

在对资料经过整理之后，还要对所用资料进行初步分析，可以利用相关性分析找出和预测序列相近的影响因素，将其作为自变量。在选取因素的时候，可以利用统计学中的相关性进行分析，其原理是基于向量空间模型（VSM）的分类过程中，待预测特征向量与各类代表向量的夹角是决定该特征向量归属

的重要依据之一，这些夹角的余弦被称作“相似度”。其广泛地用于相关性分析中。其定义为：假设向量的特征空间为 $U=(u_1,u_2,\cdots,u_n)$，任取两个向量 $U_i=(u_1,w_{i1};u_2,w_{i2};\cdots,u_n,w_{in})$，$U_j=(u_1,w_{j1};u_2,w_{j2};\cdots,u_n,w_{jn})$；其中：$w_{ik}$ 为向量 U_i 在属性 $u_k(k=1,\cdots,n)$ 处的权重；简记为 $U_i=(w_{i1},w_{i2},\cdots,w_{in})$；其中：$w_{jk}$ 为向量 U_j 在属性 $u_k(k=1,\cdots,n)$ 处的权重；简记为 $U_j=(w_{j1},w_{j2},\cdots,w_{jn})$，则这两个向量的相似度用它们的夹角余弦值表示，其计算公式如下：

$$Sim(U_i,U_j)=\cos\theta=\frac{\sum_{k=1}^{n}w_{ik}\times w_{jk}}{\sqrt{\left(\sum_{k=1}^{n}w_{ik}^2\right)\left(\sum_{k=1}^{n}w_{jk}^2\right)}} \tag{1-6}$$

其计算的结果值越大，说明向量的相似度越高。

【例】 试计算我国电力消费量和国内生产总值之间的相似关系，其数据见表 1-2。

表 1-2　　电力消费量与国内生产总值统计值（1978—2006 年）

年份	电力消费量（亿 kW·h）	国内生产总值（亿元）	年份	电力消费量（亿 kW·h）	国内生产总值（亿元）
1978	2565	3624.1	1993	8300	35333.9247
1979	2819	4038.2	1994	9260.4	48197.8564
1980	3006.3	4545.62397	1995	10023.4	60793.7292
1981	3093	4891.56106	1996	10764.3	71176.5917
1982	3277	5323.35096	1997	11284.4	78973.035
1983	3515	5962.65157	1998	11598.4	84402.2798
1984	3777.6	7208.05172	1999	12305.23	89677.0548
1985	4117.6	9016.03658	2000	13471.38	99214.5543
1986	4507	10275.1792	2001	14723.46	109655.171
1987	4985.2	12058.6151	2002	16465.45	120332.689
1988	5466.8	15042.823	2003	19031.6	135822.756
1989	5865.3	16992.3191	2004	21971.37	159878.338
1990	6230.4	18667.8224	2005	24940.39	183867.883
1991	6804	21781.4994	2006	28587.9729	210870.995
1992	7589.2	26923.4765			

注　数据来源于《中国统计年鉴》(2007) 和《中国电力年鉴》(2005)。

首先画出电力消费量和国内生产总值之间的散点连线图，如图 1-1 所示。

从图 1-1 中可以看出，GDP 与电力消费之间存在着显著相关关系。按照

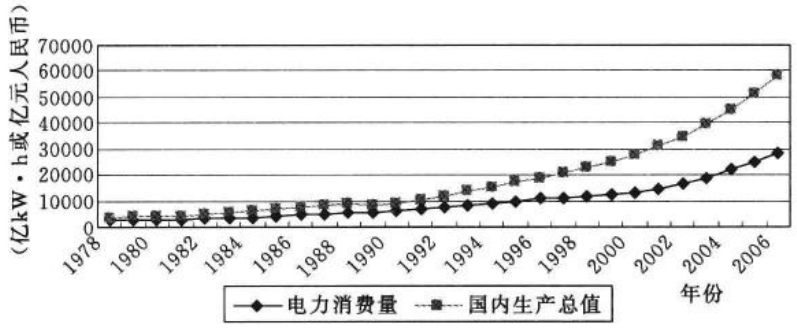

图 1-1 电力消费量与国内生产总值（归算至 1978 年）曲线图

相似度的公式进行计算，可得

$$Sim(U_i,U_j)=\cos\theta=\frac{\sum_{k=1}^{n}w_{ik}\times w_{jk}}{\sqrt{\left(\sum_{k=1}^{n}w_{ik}^2\right)\left(\sum_{k=1}^{n}w_{jk}^2\right)}}=0.9807$$

由此可见，GDP 与电力消费之间的相关关系较强。

也可以利用灰色预测模型中的关联性分析进行分析，判断序列的相近程度，其原理如下：

假定参考数列为 x_0，被比较数列（又称预测数列或因素数列）为 x_i，其中 $i=1,2,\cdots,m$，且

$$x_0=\{x_0(1),x_0(2),\cdots,x_0(n)\}$$
$$x_i=\{x_i(1),x_i(2)\cdots,x_i(n)\}$$
$$(i=1,2,\cdots,m)$$

则称

$$\xi_i(k)=\frac{\min\limits_i\min\limits_k|x_0(k)-x_i(k)|+\rho\max\limits_i\max\limits_k|x_0(k)-x_i(k)|}{|x_0(k)-x_i(k)|+\rho\max\limits_i\max\limits_k|x_0(k)-x_i(k)|}\qquad(1-7)$$

为曲线 x_0 与 x_i 在第 k 点的关联系数。上式中 $|x_0(k)-x_i(k)|=\Delta i(k)$ 称为第 k 点 x_0 与 x_i 的绝对差；$\min\limits_i\min\limits_k|x_0(k)-x_i(k)|$ 称为两级最小差，其中 $\min\limits_k|x_0(k)-x_i(k)|$ 是第一级最小差，这表示在第 x_i 曲线上，找各点与 x_0 的最小差，$\min\limits_i\min\limits_k|x_0(k)-x_i(k)|$ 是第二级最小差，表示在各条曲线中找出的最小差基础上，再按 $i=1$，$i=2$，…，$i=m$ 找所有曲线 x_i 中的最小差；$\max\limits_i\max\limits_k|x_0(k)-x_i(k)$ 是两级最大差，其意义与最小差相似；ρ 称为分辨系数，是 0 与 1 间的数，一般取 $\rho=0.5$。

综合各点的关联系数，可得出整个 x_i 曲线与参考曲线 x_0 的关联度 r_i 为：

$$r_i=\frac{1}{n}\sum_{k=1}^{n}\xi_i(k)\qquad(1-8)$$