

纪念广西计算中心成立二十周年

论文集

1978 — 1998

刘连芳 主编

一九九八年十二月于南宁

序

广西壮族自治区政府主席助理、广西科技厅厅长 张正铀

广西计算中心在改革大潮中经历了创业与发展的过程。在过去的20个年头中，在八桂大地上将计算机科研与开发与广西的经济建设紧密结合，不断地取得新成果，受到国家科委和自治区领导及兄弟省同行的注目。

广西计算中心成立20年来，留下了前进的足迹与工作的业绩，无论在承担国家863计划项目、广西各类基金项目研究，还是承担各类攻关 新产品开发，无论是计算机硬件、软件开发，还是各类服务工作，都取得显著的成就。其中超媒体技术研究保持国内领先水平，多项研究成果填补区内空白。广西计算中心是广西计算机技术研究与开发的中坚力量，也是一支发展广西信息产业，尤其是发展广西软件产业的重要力量，在推动广西的科技进步和促进全区的计算机事业发展作出了积极贡献。

20年来，广西计算中心承担了多项国家、自治区科研攻关项目，到1998年已完成科研项目190多项，其中完成国家“863”计划4项。成果获得各种奖励136项次，其中获广西科技进步一、二等奖12项，获广西计算机开发应用成果奖20项。中心科技人员在国内外学术刊物和会议上共发表论文400余篇。

为纪念广西计算中心成立二十周年，从已发表的论文中精选出学术水平高并具有代表性的文章28篇编成论文集，便于交流和参考。

当前，科学技术突飞猛进，知识经济初显端倪，竞争日趋激烈。这对从事计算机科学技术研究与开发的工作者尤具有挑战性，望广西计算中心高举邓小平理论伟大旗帜，实践“科学技术是第一生产力”的思想，全面贯彻“科教兴国”战略，遵照江总书记的“加强应用技术的开发和推广，促进科技成果更快更好地向现实生产力转化，不断解决经济建设中的重大和关键技术问题，推动经济体制和经济增长方式的根本转变。”的指示，开发信息产品，创立广西信息名牌产品，在实现广西科技与经济结合的新突破中，深化改革，勇于开拓，敢于创新，在攀登科学高峰和实现高科技产业化的征途上，作出新成绩，昂首阔步迈向21世纪。

1998年12月18日

目 录

红外图谱数据库及其检索.....	刘连芳等 (1)
非线性偏回归方程的建立与使用.....	刘连芳等 (8)
排序算法的优化.....	张正铀等 (13)
高科技与软件产业的发展.....	张正铀 (18)
UNIX I/O子系统的移植与实现.....	石文昌 (21)
面向开发与维护过程的软件工程方法.....	刘连芳等 (30)
A Graphics—Based Hypermedia Approach	Lianfang Liu et al. (36)
一种汉字编码方案的设计与实现.....	廖宏等 (42)
UNIX/XENIX系统多用户通信管理软件的设计.....	廖文辉等 (50)
古壮文操作系统和编辑排版系统.....	刘连芳等 (54)
一种超媒体方法和它的导航、检索功能.....	刘连芳等 (59)
软件国际化机制设计技术.....	石文昌 (65)
公费医疗微机网络管理监督系统的设计与实现.....	杨磊等 (72)
计算机网络平台建设.....	石文昌 (78)
以图形为基础的超媒体模型.....	刘连芳 (85)
一种高低压配电设备计算机辅助设计方法.....	熊瑞萍等 (93)
GBH的链—如何超越媒体空间.....	刘连芳等 (98)
集散控制系统在面粉生产中的应用.....	陈大连等 (104)
广西政府信息服务系统的开发 (英文).....	石文昌等 (109)
图形视窗式电子邮件系统R-MAIL的研制.....	石文昌等 (116)
巧用PC机打印机口采集程控交换机计费信息.....	杨磊等 (122)
NEAX2400SDS程控交换机自动计费系统的设计与实现...	杨磊等 (126)
从Dexter模型看GBH.....	刘连芳等 (131)
空间地理信息可视化管理系统的实现与方法.....	熊瑞萍等 (136)
超文本/超媒体技术及其应用.....	刘连芳 (141)
WWW核心技术HTTP协议的研究与实现.....	刘连芳等 (149)
一种新颖的MIS系统集成解决方案.....	杨磊等 (162)
经典Ramsey数的下界.....	罗海鹏等 (167)
附: 其它科技论著索引.....	(174)

红外图谱数据库及其检索

刘连芳 张正铀

1 前言

在一般需要对物质作化学分析的单位都有红外分光光度计。它的优点是制作化合物图谱价格低廉(10-15元),样品可以回收。但缺点是要确定未知化合物,若只作它的红外图谱,而后凭人的经验来分析化合物大致结构,再核对数万张标准图谱是很困难的。对结构复杂的物质分析出大致结构后也无从下手查对标准图谱。为了弥补这个不足,往往在作完红外图谱后再借助其它手段,如元素分析,用质谱仪或核磁共振光谱仪再作图谱。元素分析手段落后,分析结果也不够准确,而质谱仪、核磁共振光谱仪价格较高,拥有的单位极少。用质谱仪作一个图谱需50-100元,并且样品不能回收。对极少量的或很宝贵的样品。样品不能回收就是很大的遗憾。

为解决上述红外图谱检索中的困难,应广西医药研究所的要求,我们在C.M.T B档微型机上建立了图谱数据库,并实现了对图谱的检索。C.M.T B档机使用CBASIC语言,它较BASIC语言有很多有益的扩充,特别是增加了文件处理功能,可以建立固定结构及流结构文件,并随需要随机或顺序存取数据。该机用于数据处理较为方便。

2 问题分析

在以下三种情况下需要检索标准图谱:

1. 已知某化合物的红外图谱,要确定其名称、分子式、标准图谱的出处及制作方法。

由于被测化合物中可能含有杂质,因此所作的红外图谱与该物质的标准图谱会有差异。为确保图谱数据库的科学性及检索的准确性,化学分析技术员确定了选择标准图谱中波数及制作的红外图谱中波数的原则,并对图谱的检索提出要求:

- 1) 制作的图谱中一定具备标准图谱中所具有的所有波数;
- 2) 制作的图谱的波数数量可多于标准图谱,超出数量不限;
- 3) 相应波数间误差不超过 $\pm 10\text{cm}^{-1}$ 。

另外,由于每次人工选取波数时,会有人为的误差。为避免这类误差导致检索的错误,要求凡波数相符率达到标准图谱波数数量90%以上者均作为结果

输出，并标明可信度(相符率)，以供进一步分析参考。

2. 查找分子量为某值的全部标准图谱。

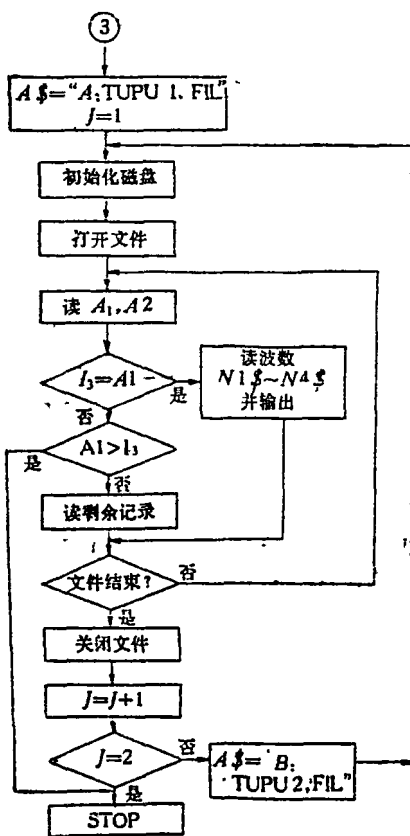
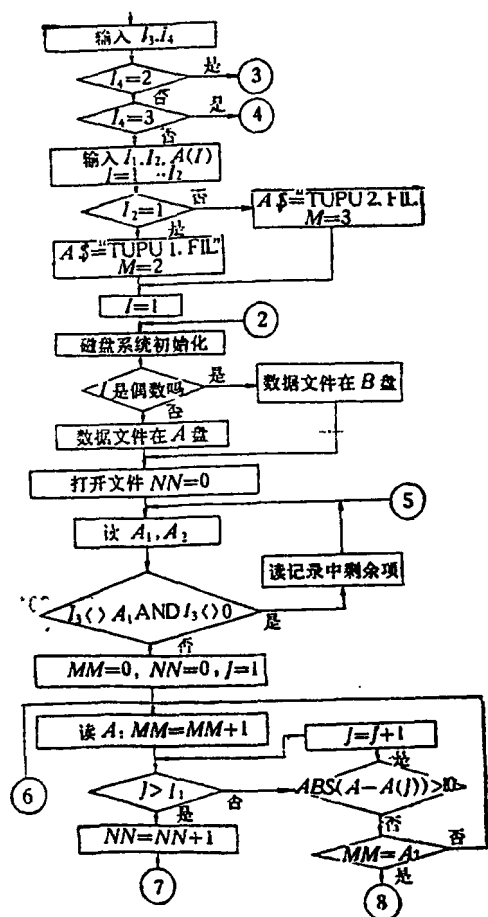
由于不同物质可能会具有相同的分子量，所以输出结果不唯一。

3. 查找已知化合物的标准图谱。

3 方法的选择

本课题属于数据处理，可以使用CBASIC提供的文件。

大量的图谱检索工作是已知化合物的红外图谱，需要确定其名称、分子量等等，并将可信度在90%以上的图谱全部输出，也就是说必须将标准图谱全部查对一遍。那么，顺序存取和随机存取都是可行的。但随机存取只适用于固定结构文件，固定结构的图谱数据库比流结构的多占用48%左右的存储。从节省存储的角度看，流结构优于固定结构。此外，由于顺序存取时计算机不需要计算记录的相对地址，加上流结构文件的记录长度一般都比固定文件的记录长度小许多，因而读取一个流文件的记录所用的时间少于读取一个固定文件的记录的时间。试验也证明了这一点(见附表)。所以在已知红外图谱的情况下使用流文件的顺序存取优于固定文件的随机存取。



在已知分子量或化合物名称的情况下，若在固定文件上随机查找，可使用对分法提高检索速度，使得平均检索时间正比于 $\log_2 N$ (N 为记录总数)。

但考虑到实际工作中，一般已知红外图谱，需检索标准图谱的工作是占主要地位的，所以实际应用还是采用流结构文件的顺序存取。

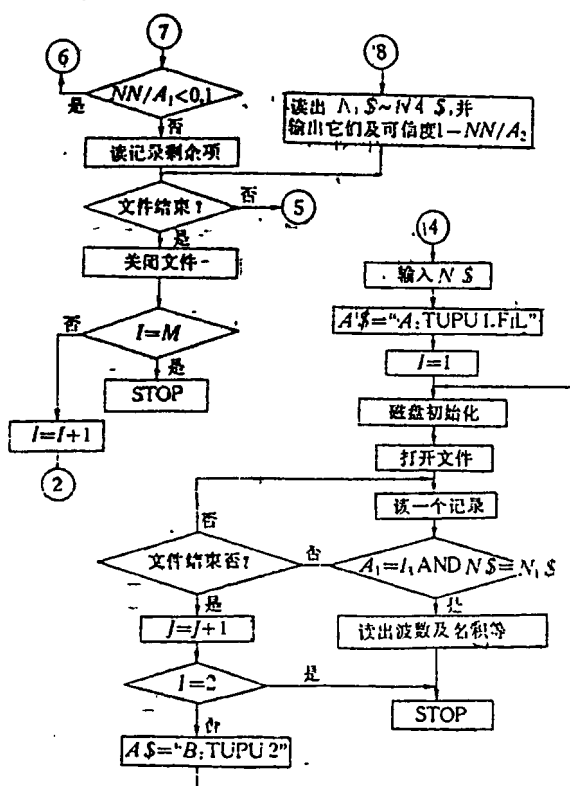
另外，根据图谱特征，将其分为两大类，在 $1500\text{cm}^{-1} - 1600\text{cm}^{-1}$ 内有值为第一类，在此间无值为第二类。按类建数据库，且按类检索，则可节省工作量近50%。每个数据库的记录按分子量大小顺序存放(从小到大)，在已知分子量情况下，标准图谱的分子量一旦大于已知分子量则可停止检索，也可减少不必要的工作。

4 程序及框图

数据库中存入的信息为：分子量 A_1 、波数个数 A_2 ，波数(A_2 个)，化合物名称 N_1 \$，分子式 N_2 \$，图谱来源 N_3 \$，制作方法 N_4 \$。

例如，一化合物，其标准图谱的16个波数，其中有处于 $1500\text{cm}^{-1} - 1600\text{cm}^{-1}$ 范围内的，所以属于第一类。其记录为265, 16, 691, 762, 780, 800, 885, 970, 1157, 1192, 1245, 1275, 1325, 1475, 1490, 1570, 1611, 1680, “3-(SALICYLIDENEAMINO)-4-(3H)-QUINAZOLINONI”，“C15H11N3O2”，2042, “K”。

为节省存储，图谱来源、制作方法均为缩写，在打印输出时给出完整内容。如制作方法“C”表示“CELL”，“K”，表示“KBR WAFER”等等。



程序

	INPUT "MOLECULAR WEIGHT, SEARCHING SORT"; I3, I4		IF ABS(A-A (J%))>10 THEN 500
	IF I4=2 THEN 2		IF MM=A2 THEN GOSUB 600: /
	IF I4=3 THEN 10		PRINT "AVAILABILITY: ";
	INPUT "NUMBER. SORT"; I1, I2		1-NN/A2: GOTO 300
	IF I2=1 THEN A \$="TUPU1.FIL": \		READ #1; A
	M=2 ELSE A \$="TUPU2.FIL": \		MM=MM+1
	M=3	500	J1=J%
	DIM A (I1)		NEXT J%
	FOR I=1 TO I1		NN=NN+1
301	PRINT "A("; I; ")="		IF NN/A2 < - 0.1 THEN 700
	INPUT A (I)	100	MM=MM+1
	IF A (I)=1 THEN I=I-1: GOTO 301		FOR II=MM TO A2
	NEXT		READ #1; A
	FOR I=1 TO I1		NEXT
	PRINT A (I);		READ #1; N1 \$, N2 \$, N3 \$, N4 \$
	NEXT	400	WEND
	PRINT		COLSE 1
303	INPUT "AA \$="; AA \$		NEXT I
	IF AA \$="Y" THEN 302		PRINT "END!"
	INPUT N	600	STOP
	FOR J=1 TO N		READ #1; N1 \$, N2 \$, N3 \$, N4 \$
	INPUT I, A (I)		PRINT
	NEXT		PRINT "IT IS"; N1 \$; "-"; N2 \$;
	GOTO 303		": M. \W. "; A1
302	FOR I=1 TO M	900	IF N4 \$="C" THEN N4 \$="CELL": \
	GOSUB 200		GOTO 800
	IF END #1 THEN 400		IF N4 \$="K" THEN N4 \$="KBR WA\
300	WHILE -1	800	FER" ELSE N4 \$="SPLIT MULL"
	READ #1; A1, A2		PRINT "FROM: S. G"; N3 \$; "K"; "\
	IF I3<>A1 AND I3<>0 THEN MM=1: \		MADE TETHOD: "; N4 \$
	GOTO 100	200	RETURN
	MM=0		INITIALIZE
	NN=0		PRINT "OK!"
	J1=1		IN INT(I/2)*2<>I THEN B \$="A: "\
700	READ #1; A		+A \$ ELSE B \$="B: "+A \$
	MM=MM+1		FILB \$ E
	FOR J=%J1 TO I1		RETURN
		2	A \$="TUPU1.FIL"

```

M=2
FOR J=1 TO 2
FOR I=1 TO M
GOSUB 200
IF END #1 THEN 4
3 WHILE -1
  READ #1; A1, A2
  IF I3=A1 THEN PRINT: GOSUB 5: \
  GOTO 3
  IF A1>I3 THEN 4
  FOR II=1 TO A2
  READ #1; A
  NEXT
  READ #1; N1 $; N2 $, N3 $, N4 $
  WEND
4 CLOSE 1
  NTXT I
  IF J=1 THEN A $ "TUPU2. FIL"
  M=3
  NEXT J
  PRINT "END!"
  STOP
5 FOR II=1 TO A2
  READ #1; A
  PRINT A;
  NEXT II
  GOSUB 600
  RETURN
10 INPUT "NAME"; N $
  INPUT "YES/NO?"; AA $
  IF AA $ ="N" THEN 10
  A $ ="TUPU1. FIL"
  M=2
  FOR II=1 TO 2
  FOR I=1 TO M
  GOSUB 200
  IF END #1 THEN 40
20 WHILE -1
  READ #1; A1, A2
  DIM B (A2)

```

```

FOR J=1 TO A2
  READ #1; B (J)
  NEXT J
  READ #1; N1 $, N2 $, N3 $, N4 $
  IF A1=I3 AND N $ =N1 $ THEN 30
  WEND
40 COLSE 1
  NEXT I
  IF II=1 THEN A $ ="TUPU2. FIL"
  M=3
  NEXT
  PRINT "NO FIND!"
  STOP
30 FOR J=1 TO A2
  PRINT B (J);
  NEXT
  PRINT
  GOSUB 900
  STOP
  END

```

注: I₁: 输入的波数数量;

I₂: 输入的图谱类别 (在1500cm⁻¹ - 1600cm⁻¹范围内有值的取值为1, 在1500cm⁻¹ - 1600cm⁻¹范围内无值的取值为2);

I₃: 输入的分子量 (未知则为0);

I₄: 检索类别 (已知图谱为1, 已知分子量为2, 已知化合物名称为3);

N \$: 已知化合物名称;

M: 某类标准图谱文件个数;

A \$: 图谱文件名称.

5 实例

例1: 根据所作的某化合物红外图谱确定其名称、分子式、标准图谱出处等有关信息.

CRUN VER 2. 05

MOLECULAR WEIGHT, SEARCHING SORT

0, 1

NUMBER, SORT 14, 2

A (1)=

? 550

A (2)=	1465)
? 660	FROM: S. G. 65K MADE METHOD: CELL
A (3)=	AVAILABILITY: 0.9
? 630	OK, IT IS HEPTANE - C7H16: M. W. 100
A (4)=	(注: 该标准图谱数据为: 635, 720, 1379, 1465)
? 720	FROM: S. G. 93K MADE METHOD: CELL
A (5)=	AVAILABILITY: 1
? 820	END!
A (6)=	例2: 查找分子量为88的所有化合物的有关资料。
? 957	CRUN VER 2.05
A (7)=	MOLECULAR WEIGHT, SEARCHING SORT\
? 1003	88, 2
A (8)=	OK!
? 1169 \	OK!
A (9)=	770 930 1090 1210 1275 1410 1700
? 1300	OK, IT IS BUTYRIC ACID - C4H8O2: M. W. \
A (10)=	88
? 1343	FROM: S. G. 37K
A (11)=	MADE METHOD: CELL
? 1390	612 870 885 1047 1080 1120 1250 1287 1364\
A (12)=	1450
? 1	OK, IT IS P - DIOXANE - C4H8O2: M. W. 88
A (11)=	FROM: S. G. 21K
? 1380	MADE METHOD: CELL
A (12)=	1050 1240 1370 1740
? 1400	OK, IT IS ETHY ACETATE - C4H8O2: M. W. \
A (13)=	88
? 1450	ROM: S. G. 3K
A (14)=	MADE METHOD: CELL
? 1465	END!
550 660 630 720 820 957 1003 1169 1300 1343\	例3: 查名为HYDROQUINONE的化合物的
1380 1400 1450 1465	标准图谱及其出处。
AA \$ =N	CRUN VER 2.05
? 1	MOLECULAR WEIGHT, SEARCHING SORT\
? 2, 600	110, 3
AA \$ =Y	NAME HYDROQUINONE
OK!	YES/NO? Y
OK, IT IS ISOPROPYL ALCOHOL - C3H8O: \	OK!
M. W. 60 (注: 该标准图谱数据为: 635,	520 610 758 826 1095 1205 1350 1470 1510
813, 950, 1128, 1160, 1305, 1340, 1378, 1405,	

6 小结

两类图谱文件均为流结构。读取一个一类记录需用0.49秒, 读取一个二类记录需0.38秒。一个8寸单面双密度磁盘可放一类图谱4500个, 二类图谱可放6300个。已知图谱, 检索一万个记录平均需要40分钟。按用机费用每小时5元计, 约需3.5元。已知分子量和化合物名称检索一类记录4500个和二类记录6000个, 平均用机时间为18分钟, 费用不足2元。

由上看出用微型机建立图谱数据库并按需要检索比用质谱仪等仪器作图谱费用低, 适应性广, 并能保留样品; 比有的单位利用系统软件建立数据库易于实现; 检索速度及准确性是人工查找所不能比拟的。

该程序允许用户对所输入的波数及化合物名称进行检查, 产生输入错误可以重新输入, 所以可以避免由于输入数据错误造成运行上的浪费。

CBASIC允许在程序执行过程中随需要初始化磁盘。我们利用了这一特点, 解决了检索过程中更换被检索磁盘的问题, 因此, 对图谱数据库的规模没有任何限制, 充分利用了微型机海量外存的特长。若根据各有关专业的需要, 不断扩充红外图谱数据库, 则可使图谱数据库适应于更多的用户。

附表:

	检索一个记录所需时间(秒)		8寸单面双密度盘可容记录数	
	流 结 构	固 定 结 构	流 结 构	固 定 结 构
一类图谱	0.49	0.88	4500	2300
二类图谱	0.38	0.63	6300	3100

编注: 本文发表在《微计算机应用》, 1982, No. 2.

非线性偏回归方程的建立与使用

刘连芳 张正铀 乔中南

摘 要 文章通过对水能计算中使用逐步回归方法的情况的介绍, 阐明了建立最佳多元非线性偏回归方程的一些方法, 并介绍了一个用回归方程代替水利学万宁公式进行水能计算的实例。

1 概述

设计中的马骝滩电站所处地形比较复杂, 必须采用长系列(三十年)水文资料对水能进行推算, 以期使电站的设计较为合理。在对三十年的水能资料的处理中多次使用了逐步回归方法。

2 最佳多元非线性偏回归方程的建立

水能计算的关键是根据一年多的水文观测数据用逐步回归方法建立上游、下游各处流量、水位之间的数学表达式——多元非线性偏回归方程, 以便推算三十年的水能数据, 为选择装机方案提供可靠依据。

为节省程序设计工作量, 凡属曲线拟合问题也采用逐步回归方法予以解决。所以课题中各种方法、各个方案总共需建立31个偏回归方程。方法如下:

2.1 函数类型的选择

要建立各个回归方程, 首先要根据水利工作者的经验选择参加回归的自变量。本课题所涉及的回归方程最多两个自变量, 所以自变量的选取并不困难。关键是自变量的函数类型不易确定。自变量与因变量之间的函数关系往往不是某种单一的函数关系, 而是某些函数关系的组合。例如多项式与正弦、多项式与对数、多项式与倒数等等。

a、由于有许多函数在一定条件下可展开为多项式, 所以多项式是最普遍适用的函数关系。在各回归方程的建立中, 我们一般以多项式作为基础, 然后再考虑其它函数的影响。

例1: 12500KW发电机效率 η 与净水头 H 的关系

$$\eta = 1.195127 - 0.2003597H + 0.02243H^2 - 0.07876 \cdot \sin(H) \quad (H \leq 6.3)$$

$$\rho = 0.99911$$

b、因变量有时也会以函数形式出现, 所以在考虑自变量函数类型的同

时,也需要考虑因变量的函数类型。

例2: 坝后水位 Z_3 大于正常高水位时, Z_3 与毛水头 HM 的关系

$$\ln(HM) = -14.5585 - 2.1843 \ln(Z_3 - 21) + 2.0962 \ln(Q_3)$$

$$\rho = 0.9902658$$

究竟自变量和因变量的函数类型是什么,要根据散点图予以确定,并经过多次回归试验,逐步修正,最后找到一个最佳非线性偏回归方程。

例3: 求糙率 N 与平均水位 H_{cp} 的关系

a、从散点图看,两者关系大致为倒数曲线,所建立方程为:

$$\frac{1}{N} = B_0 + \frac{B_1}{H_{cp}} \quad (1)$$

$$\rho = 0.780828$$

b、根据对图形的分析加入对数曲线的影响和 H_{cp} 项

$$\frac{1}{N} = B_0 + \frac{B_1}{H_{cp}} + B_2 \cdot \ln(H_{cp}) + B_3 \cdot H_{cp} \quad (2)$$

$$\rho = 0.96034437$$

c、提高 H_{cp} 的方次,对数项消失。

$$\frac{1}{N} = B_0 + \frac{B_1}{H_{cp}} + B_3 \cdot H_{cp} + B_4 \cdot H_{cp}^2 \quad (3)$$

$$\rho = 0.9603516$$

d、去掉对数项,再提高 H_{cp} 的方次

$$\frac{1}{N} = B_0 + \frac{B_1}{H_{cp}} + B_2 \cdot H_{cp} + B_3 \cdot H_{cp}^2 + B_4 \cdot H_{cp}^3 \quad (4)$$

$$\rho = 0.961828$$

e、提高 $\frac{1}{H_{cp}}$ 方次,进一步提高 H_{cp} 方次

$$\frac{1}{N} = B_0 + \frac{B_1}{H_{cp}} + \frac{B_2}{H_{cp}^2} + B_3 \cdot H_{cp}^2 + B_4 \cdot H_{cp}^3 + B_5 \cdot H_{cp}^4 \quad (5)$$

$$\rho = 0.961833$$

从(3)、(4)、(5)看来,再用提高方次来改进回归结果,效果就不理想了,这几步复相关系数均在0.96以上,说明方程已较好地反映了 N 及 H_{cp} 之间的关系,可以实用。所以试验到此结束。

方程(5)的复相关系数略大于方程(4)的,但按方程(5)进行计算要增加五次乘除、一次加减。此时,需根据实际需要选取方程。由于 N 、 H_{cp} 散点图上的

点分布较散，回归结果各点残差相对来讲就稍大些，所以希望选择尽可能好的方程进行运算，就选了方程(5)作为推算N的数学表达式。

2.2 分段回归

为提高精度，对某些问题采用分段回归的方法，可以取得较好的效果。此时，在各段交接处选几个公共点参加前、后两段的回归，然后根据这几点的残差划分使用偏回归方程的范围。这样做，一方面改善了总的回归效果，另一方面使各段之间过渡较为光滑，并尽量减少交点的残差。

例4: 12500KW发电机效率曲线 $\eta = f(H)$ 的建立。

由于H取值在2.88 - 6.3与6.3 - 10范围时， η 的发展趋势明显不同，所以以6.2、6.3、6.35作为公共点分两段进行回归，就得到两个回归方程：

$$\eta = 1.195127 - 0.20036H + 0.02224H^2 - 0.07876\sin(H) \quad (H \leq 6.3)$$

$$\rho = 0.99911$$

$$\eta = -0.96612 - 0.56427H + 0.155134H^2 + 2.607781\ln(H) \quad (H > 6.3)$$

$$\rho = 0.99853$$

2.3 异常数据的取舍

异常数据的取舍主要根据工程技术方面的分析而不是统计学的规则。

a、超出正常值范围的数据不予参加回归。如糙率的正常值小于0.13，则凡大于0.13的数据均舍去。

b、着重对回归结果中残差异常大的点进行分析，确认为观测数据异常的则舍去，再对剩下数据进行回归。

课题中按上述方法建立的31个方程的复相关系数均在0.9618以上。

3 用偏回归方程取代水利学万宁公式

根据角咀水位求坝水位，在水利学上应使用万宁公式

$$Q = \frac{1}{N} \overline{WR}^{2/3} \sqrt{I} \quad (1)$$

其中，Q: 已知坝址流量

N: 糙率， $N = f_1(Z_1, Z_3)$

$\overline{W}$: 平均过水断面面积

$\overline{R}$: 平均水利半径 $\overline{WR}^{2/3} = f_2(Z_1, Z_3)$

I: 比降， $I = f_3(Z_1, Z_3) = \sqrt{\frac{Z_3 - Z_2}{4000}}$

Z_2 : 已求出的角咀水位

Z_3 : 未知坝后水位

$\bar{W}$ 、 $\bar{R}$ 分别根据两张断面图积分求得。 f_1 、 f_2 、 f_3 均系用逐步回归方法建立的多项式形式的数学表达式。

总括以上诸关系，万宁公式最终可表达为：

$$Q = g_1(Z_1, Z_3) \quad (2)$$

(Z_1 、 Q 为已知， Z_3 为所求)

即，求出 $\bar{W}$ 、 $\bar{R}$ 、 f_1 、 f_2 、 f_3 之后，最终需解一个关于 Z_3 的高次方程。

但是从 (2) 式可以看出， Q 、 Z_1 、 Z_3 之间直接存在着某种关系，现又有这三个量的 409 组观测数据，那么用逐步回归方法应该可以找出这种关系。

一旦找出这种关系，求 Z_3 就不必经过上述的 N 、 $\bar{W}$ 、 $\bar{R}$ 等等中间计算。又因为寻找这种关系的目的在于推算 Z_3 ，为使用上的方便，就应以 Z_3 作为因变量来建立偏回归方程。

3.1 试建立偏回归方程

为验证上述思想，我们试建立了偏回归方程

$$Z_3 = 0.29155 + 0.98803Z_1 + 5.54698E-05Q + 2.01617E-09Q^2 \quad (3)$$

$$\rho = 0.99879$$

其中 Q 、 Q^2 项系数仅为 10^{-5} 、 10^{-9} 数量级的数，说明 Q 对 Z_3 有影响，但影响较小。这与交通、水利部门平时使用的散点图一致。

3.2 方程 (3) 与万宁公式的对比

a、使用万宁公式

①求 f_2 ：需先求与 409 组 Z_1 、 Z_3 相应的 W_1 、 W_3 、 R_1 、 R_3 ，即共求 818 个面积和 818 个湿周，从而得到 409 个 $\overline{WR}^{2/3}$ ，用这 409 个数据与对应的 Z_1 、 Z_3 一起参加回归，确定 f_2 。

②求 f_1 ：将 409 组 $\overline{WR}^{2/3}$ 、 Z_1 、 Z_3 与相应的 Q 代入 (1)，求出相应的 N ，再用这些 N 与 Z_1 、 Z_3 共同参加回归，确定 f_1 。

③将 f_1 、 f_2 、 f_3 代入 (1)，用对分法解高次方程，求出 Z_3 。求一个 Z_3 平均需作 211 次乘除、81 次加减。

b、使用偏回归方程 (3)

只需进行一次逐步回归即可建立方程 (3)。直接代入已知的 Q 、 Z_1 ，即可求得 Z_3 。推算一个 Z_3 仅需作 7 次乘除、4 次加减，工作量是使用万宁公式的三十分之一。

用两种方法计算 365 个 Z_3 进行对比，结果证明方法 b 比方法 a 精度也有所提高。

		万宁公式	偏回归方程
残 差	~ 0.05	80.821%	86.85%
	0.2 ~ 0.4	1.917%	0
	max	0.799997598	0.725829743
	min	2.57559E-06	3.3736229E-07

· 残差单位: cm

综上所述及对比试验,说明使用偏回归方程(3)代替水利学万宁公式推算坝后水位 Z_3 不仅可行,而且大大节省工作量,计算精度也有所提高。在对三十年的 Z_3 的推算中,我们就使用了偏回归方程(3)

4 结语

逐步回归方程应用领域较多,凡需根据实验、观察、测量等得来的数据建立某些量的数学关系式的问题均可适用。

我们在马骝滩电站的水能计算中,初步摸索出一些用逐步回归方法建立最佳多元非线性偏回归方程的方法,并尝试用逐步回归方法建立的偏回归方程代替水利学的万宁公式,收到了良好的效果。为今后该方法的使用积累了点滴经验。

编注:本文发表在《广西科学院学院》,1982, No.1.

排序算法的优化

张正铀 刘连芳

THE OPTIMIZATION OF A SORTING ALGORITHM

Zhang Zheng-you Liu Lian-fang

Abstract

The paper constructs a finer algorithm that makes the most of the information produced in sorting procedure and appraises its complex nature by analysis of sorting algorithms whose complex nature is $O(N^2)$.

ARNE THESEN 在讨论排序算法时指出,数据排序是极其有用的。如果没有排序,要修改磁盘上的数据几乎是不可能的^[1]。在已出版的Knuth的《计算机程序设计技巧》^[2]第三卷中,就以半卷的篇幅讨论六类二十五种排序算法(Sorting)。

算法的优劣通常用算法的时间复杂性和空间复杂性来衡量,即用程序执行时间和占用的存贮空间量来衡量。由问题长度函数表示的算法执行时间称为算法的时间复杂性,由问题长度函数表示的占用空间称为算法的空间复杂性。排序问题的长度为排序量 N ,所以记时间复杂性为 $T(N)$,空间复杂性为 $S(N)$ 。对时间复杂性的分析有各种方法,其中期望复杂性更可反映算法的动态执行情况,但期望复杂性很难估计。因此,很多算法的分析只能给出最好情况时间复杂性和最坏情况时间复杂性^[3,4]。本文在分析排序算法时,也对算法在最好和最坏情况下的执行时间和占用空间作出分析,并根据实验结果对一般情况作出估计。分析中定义 $T(N)$ 以一次比较操作所用CPU时间为单位。由于排序算法中主要是比较和赋值操作,而一次赋值操作所用CPU时间小于一个单位(见附表1),因此在算法分析中只将赋值次数作为参考值。 $S(N)$ 以存贮一个数据所占用存贮单元数为单位。

对无特殊规律数据集的排序算法,如气泡漂浮^[1,6]、挑选插入^[2,5]等算法的