

虫情数据文件管理系统

魏 鹰

(广西计算中心)

摘 要

本文介绍一个虫情数据文件管理系统的设计思想和实现。鉴于考虑磁盘容量小、处理的数据量大,系统采用压缩数据处理(一个县每年的数据不压缩需占7.7K字节,压缩后只占3.9K字节)。数据文件中的数据存贮紧凑。系统功能实现主要由主模块、输入模块和取出模块组成。

一、引 言

对农作物虫情的预报是极为重要的工作。区植保总站目前仍采用人工预报方法,速度慢,容易出差错,并且积压了大量急需处理的数据,失去了预报的意义。为适应生产的发展迫切需要计算机来管理数据,快速而准确地进行农作物虫情的预报。

二、系统概况

本系统在TRS-80Ⅲ微型机上实现。系统编制采用TRS-80Ⅲ型MICROCOMPUTER BASICⅡ语言,它具有较强的会话功能和文件处理能力,同时调试修改方便。该机使用五吋单面双密度软磁盘,可供用户使用的容量约有175K,另外系统盘还有124K的空间空闲。

需存贮的有23个县(试点)的原始数据,每个县有13个数据表,各表规格不一,始年也不尽相同(1961、1971或1981年开始),另外还有5个气象表(1961年开始),要逐表分析。数据量很大,一个县每年就有近1400个数据,约需占5K字节;气象表每年有近730个,约需占2.9K字节。数据类型不同(整型、实型和字符串型),有的数据(整型)只用一个二进制位即可,而有的数据(实型、字符串)用二个字节还存不下。数字字符串数据可转换为整数存贮,字母字符串数据却不行(种类太多,处理起来占用内存、时间过多)。这样复杂、量大的数据,如果不采用压缩存贮,那么一个磁盘还存不下两个县的数据,10种虫23个县至少要120个盘,气象表数据也得用半个盘,如此实现起来既浪费软盘,效率又低。

为了能在如此小的磁盘容量上实现这么大的数据存贮，降低成本（节约磁盘用量），提高效率（减少换盘次数），系统采用压缩技术。

三、数据压缩处理

除气象表中“风向”这项数据是用字符串存贮外，其他数据都可化整存贮。整数范围为-32768~32767（以下讨论已转化为整数的数据压缩问题）。

1. 对调查地点 d （二种）和类型 l （三种）可压缩为一个十进位 I ，它们的转换关系为：

$$I = 3d + l$$

其中 $d = \begin{cases} 0 & \text{(观测区)} \\ 1 & \text{(面上)} \end{cases} \quad l = \begin{cases} 0 & \text{(早)} \\ 1 & \text{(中)} \\ 2 & \text{(迟)} \end{cases}$

另外，对调查日期中的月数 M ，有的表数据一月份和十二月份不作调查，可用一个十进位：

$$I = M - 2$$

2. 两个数合并存贮：

① 把整数单元化分为二部分：

$$I \begin{array}{|c|c|} \hline A & B \\ \hline \end{array}$$

-16-a位 a位

根据 A 、 B 两数的范围来确定 a 的值。

$$I = A * 2^a + B$$

例如： A 和 B 都小于256，则取 $a = 8$

$$I = A * 256 + B$$

② 数 A 小于654可与小于100的数 B 合并：

$$I = A * 100 + B$$

③ 对调查日期的日数 A （ ≤ 31 ）可与任一个三位数（即小于1000的数） B 合并：

$$I = A * 1000 + B$$

3. 由于虫情数据在处理中对数据较大的可相应地舍去十位或百位以下数字，所以对上万的数据 x 可表示与存贮为：

数范围	表示	存贮
① $x < 10^4$	x	x
② $10^4 \leq x < 10^5$	$\times \times , \times$	$10 \times \times \times$
③ $10^5 \leq x < 10^6$	$\times \times \times ,$	$20 \times \times \times$

逗号左边第一位为千位数。

4. 若整个表数据都是实数，且有效数字为三位，形如 $\times \times . \times$ ，则采用整数部分独立合并（两数合一），小数部分另外合并（四数合一）。

以上压缩后的数 I 存贮时若大于32767则影射到负值：

$$\text{存贮值} = 32767 - I$$

取出存贮数值时若为负值则按如下还原原值 I ：

I = 32767 - 存储值

四、文件结构

本系统文件采用定长记录结构，每个记录长度为256个字节。每个县建立三个数据文件：LIST1（2~5表）、LIST2表（6）、LIST3（45~54表），存于用户盘内，每张磁盘可存12个县（1961~1980年或1981~1985年或1986~1990年）。气象表建立一个数据文件LIST4（72~76表），存于系统盘内。另外还建立了一个制表静态数据文件DA（只占一个记录）。数据文件结构见如下各表：

LIST1

记录号	内 容
1	2表8行
2	2表8行
3	2表8行
4	2表6行、4表8行
5	3表12行
6	3表8行4表12行
7	5表5行

2~5表一年的情况

LIST3

记录号	内 容
1	48、49、51、53、54表
2	45、46、47表

45~54表一年的情况

附表一

表	字节/行
2	32
3	20
4	8
5	47
6	36
75	42
76	16

LIST2

记录号	内 容
1	71年7行
2	71年3行、72年4行
3	72年6行

6表二年的情况

LIST4

记录号	内 容
1	74表4个月
2	74表4个月
3	74表4个月
4	72表5个月
5	72表5个月
6	72表2个月、73表5个月、76表一行
7	73表7个月、75表一行

72~76表一年的情况

附表二

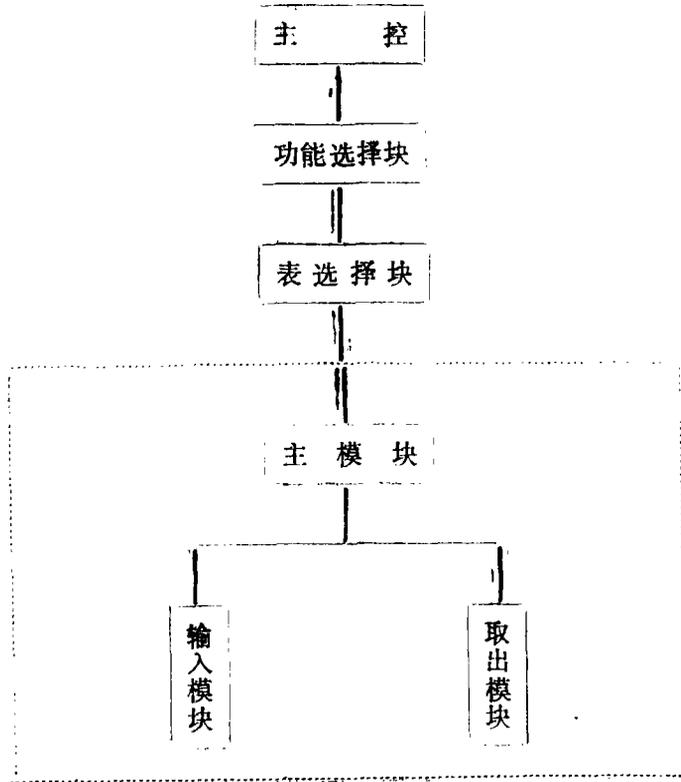
表	字节/月
72	50
73	28
74	60

附表三

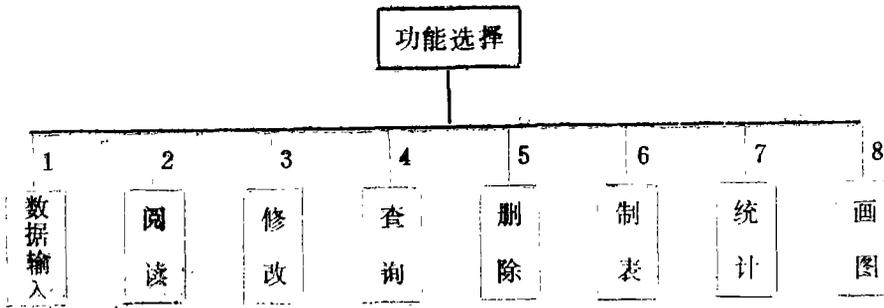
表	字节/表
45	120
46	50
47	79
48	40
49	48
51	60
53	52
54	54

五、系统程序结构

考虑到该机内存只有48K字节，我们把整个系统分为两个程序存于系统盘。制表程序文件“TAB”占用7K字节，制表时先用它来制各表表头，传递数据表打印格式的数据到文件“DA”中，然后转主程序的制表功能制表头下的数据表部分。主程序文件“MAIN”占用31K字节，具有八个功能：



框图一



功能框图

主程序如框图一，其中主模块、输入模块、取出模块是各表程序的公共部分。此外，还有修改子模块、查询子模块、制数据表子模块、统计子模块、画图子模块。

六、结束语

本系统对各种表格数据存贮问题均有一定的适用价值，特别是数据压缩存贮技术以及数据文件的组织部分。当然，由于时间仓促，作者忽略考虑这样一个问题：把所有表程序中主模块合并为一个公共模块（这主要是各表数据复杂，且还需考虑到数据压缩存贮、数据文的组织等问题，故处理起来比较麻烦）。还需要继续努力，改进此系统。

* 本系统的设计是在苏德富老师的悉心指导下进行的，在此表示感谢！

参 考 文 献

1. 计算机应用与软件 1984年一、二期
2. 上海电子计算机应用技术资料汇编 第五辑。
3. 冯懿生：多规格数据表格的处理方法。