国防科技大学学报
JOURNAL OF NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY
一九八六年第四期　总第五十六期　No.4 1986 Sum. 56

数列最优成组剖分的一个动态算法

刘　继　勇

摘　要　数学工作者从铁道调车问题，抽象出一个数学问题，称为数列的成组剖分问题。本文给出一个递推公式，利用递推公式可以逐步减少数列中数值的个数，从而得到求数列最优成组剖分数的一个动态算法。

一、问题的提出及定义

根据铁路运输中列车的编组问题[1]，文献[2]提出了有限数列的最优成组剖分问题，对此问题的一个特殊情况，最优顺序成组剖分，实际工作人员创造了一种简明的方法——“落表法”。作为推广，文献[4]给出了最优拟顺序成组剖分的一个多项式算法。[3]解决了[2]中提出的顺序子序列的计数问题。本文对文献[2]提出的问题给出了数列最优成组剖分的一个动态算法。

设S是一个有限数列

$$S=a_1a_2a_3\cdots a_n$$

其中a_i的取值是自然数，记作$\bar{a}_i(1\leqslant i\leqslant n)$。
a_i称为S的元素，S中不同元素的个数，即n，称为S的长度，记作$L(S)$。a_i的取值$\bar{a}_i$称为S的数值，S中不同数值的个数称为S的宽度，记作$W(S)$。

例　$S=12321$，S有五个元素a_1,a_2,a_3,a_4,a_5，分别取值$1,2,3,2,1$；S有三个数值$1,2,3$；从而$L(S)=5$，$W(S)=3$。

为了书写方便，在不致引起混乱的时候，我们常用数值来表示一个数列。

设p是S的数值，令

$$l(p)=\min\{i|\bar{a}_i=p\}$$
$$r(p)=\max\{i|\bar{a}_i=p\}$$

可见，任取S的一个元素a_i，有$l(\bar{a}_i)\leqslant i\leqslant r(\bar{a}_i)$. 称区间$[l(p),r(p)]=\{x\,|l(p)\leqslant x\leqslant r(p)$，$x$为实数$\}$为数值$p$的码区间。

S_1称为S的子数列，如果S_1是从S中删去若干元素后剩下的数列。S的子数列串$S_1,S_2,\cdots,S_k$称为S的部分，如果

1986年4月17日收到

(*i*) $S_1, S_2, \cdots, S_k$中任何两个子数列没有公共元素；

(*ii*) $L(S_1)+L(S_2)+\cdots+L(S_k)=L(S)$.

其中 k 称为此剖分的剖分数。

对数列定义二元运算一串联，用符号“$\circ$”表示：设 $S_1=a_1a_2\cdots a_n$，$S_2=b_1b_2\cdots b_m$，则

$$S_1\circ S_2=a_1a_2\cdots a_nb_1b_2\cdots b_m$$

S的剖分 $S_1, S_2, \cdots, S_k$ 称为S的成组剖分，如果 $S_1, S_2, \cdots, S_k$满足下面“成组条件”：

“设$S_1\circ S_2\circ\cdots S_k=b_1b_2\cdots b_n$，若$\bar{b}_i=\bar{b}_j$，则对任意的 t $(i\leqslant t\leqslant j)$，有 $\bar{b}_i=\bar{b}_k=\bar{b}_j$。”

应注意，成组剖分中的子数列是有顺序的，例如 $S=12321$，则 $S_1=31$，$S_2=122$ 是S的成组剖分；而 $S_1=122$，$S_2=31$ 就不是成组剖分，因为 $S_1\circ S_2=12231$ 不满足成组条件。

在S的所有成组剖分中，剖分数达到最小的成组剖分称为S的最优成组剖分，其剖分数 k 称为S的最优成组剖分数，记为 $N(S)$。设 S'是 S 的子数列，显然 $N(S')\leqslant N(S)$。

二、算　法

设S是任一数列，a 是S的元素，S的成组剖分 $S_1, S_2\cdots, S_k$,称为S的 a —成组剖分，如果 a 是此剖分中最前面的元素，即S_1中最左端的元素。剖分数达到最小的 a —成组剖分称为S的最优 a —成组剖分。

设$S=a_1a_2\cdots a_n$，且$1, 2, \cdots, m$是S的所有数值。取一特定元素 $a^*, \bar{a}^*\neq p$（对任意 $1\leqslant p\leqslant m$），令

$$S(a^*, i)=a_1a_2\cdots a_{i-1}a^*a_i\cdots a_n \qquad 1\leqslant i\leqslant n+1$$

用$f(a^*, i, \{1, 2, \cdots, m\})$表示$S(a^*, i)$的最优$a^*$—成组剖分数。

定理 1　设S是任意一个数列，则S的最优成组剖分数等于$S(a^*, 1)$的最优a^*—成组剖分数。即

$$N(S)=f(a^*, 1, \{1, \cdots, m\})$$

证明：由定义

$$S(a^*, 1)=a^*a_1a_2\cdots a_n=a^*\circ S$$

设 $S_1, S_2, \cdots, S_k$ 是S的最优成组剖分，$K=N(S)$，则

$a^*\circ S_1, S_2, \cdots, S_k$ 是 $S(a^*, 1)$的 a^*—成组剖分，所以 $f(a^*, 1, \{1, 2, \cdots, m\})\leqslant K=N(S)$。

反过来，设$S_1, S_2, \cdots, S_k$是$S(a^*, 1)$的最优a^*—成组剖分，即$K=f(a^*, 1, \{1, 2, \cdots m\})$。设$S_1=a^*\circ S_0$，则$S_0, S_2, \cdots, S_k$是$S$的成组剖分。

所以　　$N(S)\leqslant K=f(a^*, 1, \{1, 2, \cdots, m\})$

故　　$N(S)=f(a^*, 1, \{1, 2, \cdots, m\})$　　　　证毕

由此可见，欲计算 $N(S)$，只需计算 $f(a^*,1,\{1,2,\cdots,m\})$，下面我们给出计算 $f(a^*,1,\{1,2,\cdots,m\})$的方法。

先考虑特殊情况，当S只有一个数值p时，即 $S=a_1a_2\cdots a_n$，$\bar{a}_1=\bar{a}_2=\cdots=\bar{a}_n=p$。显然$l(p)=1$，$r(p)=n$。这时$S(a^*,i)$有如下最优 a^*—成组剖分：

情况 1 $S(a^*,i)=a^*a_1a_2\cdots a_n$，$i=l(p)$，则$S_1=a^*a_1a_2\cdots a_n$是$S(a^*,i)$的最优$a^*$—成组剖分，$a_{r(p)}$是该剖分的最后元素。

情况 2 $S(a^*,i)=a_1a_2\cdots a_na^*$，$i=r(p)+1$，则$S_1=a^*$，$S_2=a_1a_2\cdots a_n$是$S(a^*,i)$的最优$a^*$—成组剖分，$a_{r(p)}$是该剖分的最后元素。

情况 3 $S(a^*,i)=a_1a_2\cdots a_{i-1}a^*a_i\cdots a_n$，$l(p)<i\leqslant r(p)$，则$S_1=a^*a_i\cdots a_n$，$S_2=a_1a_2\cdots a_{i-1}$是$S(a^*,i)$的最优$a^*$—成组剖分，$a_{i-1}$是该剖分的最后元素。

可见，$S(a^*,i)$的最优a^*—成组剖分数

$$f(a^*,i,\{p\})=\begin{cases}1 & i=l(p)\\ 2 & i=r(p)+1\\ 2 & l(p)<i\leqslant r(p)\end{cases}$$

我们把上面给出的诸剖分的最后元素通记为$a^{**}(p)$。

$$a^{**}(p)=\begin{cases}a_{r(p)} & i=l(p)\\ a_{r(p)} & i=r(p)+1\\ a_t & l(p)<i\leqslant r(p)\end{cases}$$

这里$t=\max\limits_{1\leqslant j<i}\{j\mid\bar{a}_j=p\}$。

对于一般情况，$S=a_1a_2\cdots a_n$，设S的所有数值是$1,2,\cdots,m$，

$$S(a^*,i)=a_1a_2\cdots a_{i-1}a^*a_i\cdots a_n$$

我们给出求$S(a^*,i)$的最优a^*—成组剖分数的递归公式。

首先，设$S''(a^*,\{p\}$是$S(a^*,i)$中由a^*与数值为p的元素所组成的子数列，设$f(a^*,\{p\})$是$S''(a^*,i)$的最优a^*—成组剖分数。然后，我们从$S(a^*,i)$中删去$S''(a^*,\{p\})$，而保留$a^{**}(p)$，得到子数列$S'(a^{**}(p),\{1,2,\cdots,p-1,p+1,\cdots m\})$，该子数列是由$S$中数值为$1,\cdots,p-1,p+1,\cdots,m$的元素与$a^{**}(p)$所组成的。设$f(a^{**}(p),\{1,\cdots,p-1,p+1,\cdots,m\})$为$S'(a^{**}(p),\{1,\cdots,p-1,p+1,\cdots,m\})$的最优$a^{**}(p)$—成组剖分数。我们有如下定理。

定理 2 $f(a^*,i,\{1,\cdots,m\})$

$$=\min_{1\leqslant p\leqslant m}\{f(a^*,\{p\})+f(a^{**}(p),\{1,\cdots,p-1,p+1,\cdots,m\})-1\}$$

证明：对任意的$p\,(1\leqslant p\leqslant m)$；设$S_1,S_2,\cdots,S_k$是$S'(a^{**}(p),\{1,\cdots,p-1,p+1,\cdots,m\})$的最优$a^{**}(p)$—成组剖分，$K=f(a^{**}(p),\{1,2,\cdots,p-1,p+1,\cdots,m\})$。设$S_1=a^{**}(p)\circ S_0$。

①若$i\leqslant l(p)$，由特殊情况 1，则$f(a^*,\{p\})=1$，$a^{**}(p)=a_{r(p)}$。

令 $S_1'=a^*a_{l(p)}\cdots a_{r(p)}\circ S_0$，$S_2'=S_2$，$S_k'=S_k$

这时$S_1',S_2',\cdots,S_k'$是$S(a^*,i)$的a^*—成组剖分。

所以 $f(a^*,i,\{1,\cdots,m\})\leqslant K=1+K-1$

$$=f(a^*,\{p\})+f(a^{**}(p),\{1,\cdots,p-1,p+1,\cdots,m\})-1$$

②若 $i>l(p)$，则 $f(a^*,\{p\})=2$，$a^{**}(p)=a_t$，$t=\max\limits_{1<j<i}\{j\,|\bar{a}_j=p\}$。设 $S^{(l)}$ 是由 $S(a^*,i)$ 中所有位于 a^* 左边的数值为 p 的元素生成的子数列，即

$$S^{(l)}=p\cdots a_t=p\cdots a^{**}(p).$$

设 $S^{(r)}$ 是由 $S(a^*,i)$ 中所有位于 a^* 右边的数值为 p 的元素生成的子数列。

令 $S_1'=a^*\circ S^{(r)}$，$S_2'=S^{(l)}\circ S_0$，$S_3'=S_2,\cdots,S_{k+1}'=S_k$

则 $S_1',S_2',\cdots,S_{k+1}'$ 是 $S(a^*,i)$ 的 a^*—成组剖分。

所以
$$f(a^*,i,\{1,\cdots,m\})\leqslant K+1=2+K-1$$
$$=f(a^*,\{p\})+f(a^{**}(p),\{1,\cdots,p-1,p+1,\cdots m\})-1$$

因而对任意的 $p\,(1\leqslant p\leqslant m)$，有

$$f(a^*,i,\{1,\cdots,m\})\leqslant f(a^*,\{p\})+f(a^{**}(p),\{1,\cdots,p-1,p+1,\cdots m\})-1$$

所以 $f(a^*,i,\{1,\cdots,m\})$

$$\leqslant\min_{1<p<m}\{f(a^*,\{p\})+f(a^{**}(p),\{1,\cdots,p-1,p+1,\cdots,m\})-1\}$$

另一方面，设 $S_1,S_2,\cdots,S_k$ 是 $S(a^*,i)$ 的最优 a^*—成组剖分，$K=f(a^*,i,\{1,\cdots,m\})$。设

$$S_1\circ S_2\circ\cdots\circ S_k=a^*p_1\cdots p_1\cdots p_m\cdots p_m$$

①若 $i\leqslant l(p)$，由特殊情况 1，$f(a^*,\{p_1\})=1$，$a^{**}(p_1)=a_{r(p_1)}$。令 S_j' 是从 S_j 中删去 $S''(a^*,\{p\})$，而保留 $a^{**}(p_1)$ 得到的子数列 $(1\leqslant j\leqslant K)$，则 $S_1',S_2',\cdots,S_k'$ 是 $S(a^{**}(p_1),\{1,\cdots,p-1,p+1,\cdots,m\})$ 的 $a^{**}(p_1)$—成组剖分。所以

$$f(a^{**}(p_1),\{1,\cdots,p-1,p+1,\cdots,m\})\leqslant K=f(a^*,i,\{1,\cdots,m\})$$

即 $f(a^*,\{p_1\})+f(a^{**}(p_1),\{1,\cdots,p-1,p+1,\cdots,m\})-1\leqslant f(a^*,i,\{1,\cdots,m\})$

②若 $i>l(p_1)$，则 $f(a^*,\{p_1\}=2$，$a^*(p_1)=a_t$，其中 $t=\max\limits_{1<j<i}\{j\,|\bar{a}_j=p\}$。因为在 $S(a^*,i)$ 中 $a^{**}(p_1)$ 位于 a^* 的左边，而 a^* 是 S_1 中最左端的元素，所以 $a^{**}(p_1)$ 不在 S_1 中，必在 S_2 中；设

$$S_1=a^*p_1\cdots p_1,\ S_2=p_1\cdots p_1\circ S_0=p_1\cdots a^{**}(p_1)\circ S_0$$

令 $S_1'=a^{**}(p_1)\circ S_0$，$S_2'=S_3,\cdots,S_{k-1}'=S_k$

则 $S_1',S_2',\cdots,S_{k-1}'$ 是 $S'(a^{**}(p_1),\{1,\cdots,p_1-1,p_1+1,\cdots,m\})$ 的 $a^{**}(p_1)$—成组剖分，其剖分数为 $K-1$。所以

$$f(a^{**}(p_1),\{1,\cdots,p_1-1,p_1+1,\cdots,m\})\leqslant K-1=f(a^*,i,\{1,\cdots,m\})-1$$

即
$$f(a^*,i,\{1,2,\cdots,m\})\geqslant f(a^{**}(p_1),\{1,\cdots,p-1,p+1,\cdots,m\})+1$$
$$=f(a^*,\{p_1\})+f(a^{**}(p_1),\{1,\cdots,p_1-1,p_1+1,\cdots,m\})-1$$

于是有

$$f(a^*,i,\{1,2,\cdots,m\})\geqslant\min_{1<p<m}\{f(a^*,\{p\})+f(a^{**}(p),\{1,\cdots,p-1,p+1,\cdots,m\})-1\}$$

从而定理得证。

三、结 束 语

上节定理 2 给出了计算 $f(a^*, i, \{1, \cdots, m\})$ 的一个递推公式，通过递推公式可以逐次减少数值的个数。一直到只有一个数值，这时由特殊情况，即边界值，就可以求出结果。但我们容易看出，用这个算法来计算 $N(S)$，其复杂程度是 $O(m!)$。我们在 PC—1500 袖珍计算机上用 BASIC 语言计算了一些例子，见表 1。

表 1

S	$N(S)$	计算时间
123214543	8	2′29″
142135363	2	15′29″
134242124	8	31″
6456246123264	8	16′
646425135	2	15′18″
212514123154234	4	2′38″
53453423415243	8	2′34″
51434521613	8	15′49″
13143241424	8	29″
461536243132531	4	15′42″
214643516346	8	16′13″
471236725464	8	1:54′51″

从上表可以看出，当 $m=4$ 时，计算时间 $t_4 \approx 30''$，而 $t_5 \approx 2'30''$，$t_6 \approx 15'30''$，$t_7 \approx 2$ 小时，我们估计 $t_{10} \approx 2$ 个月，$t_{11} \approx 2$ 年。文献[5]给出了数列最优完全成组剖分的概念，从而得到了数列最优成组剖分数的界的估计。

本文是在许国志、陈庆华二位老师指导下完成的。

参 考 文 献

〔1〕李国风，表格移动调车钩计划方法，铁道科学技术，1963年。

〔2〕中国科学院数学所运筹室二组，铁道调车问题中数学方法的初步研究，应用数学学报 1 (1978) No2, p91—p105。

〔3〕蔡茂诚，顺序子序列的计数问题，应用数学学报，4(1981) No2, p140—p150.

〔4〕朱永津、朱若鹏，任意序列重新组合为某种有序序列的研究，中国科学A辑，1983.2 第二期，p119—p127.

〔5〕许国志、陈庆华、刘继勇，科学通报，第15期—P1128—P1131 1986.

A Dynamic Algorithm Finding the Optimal Digit—Grouped Partition Number of Sequence

Liu Jiyong

Abstract

From the train marshalling problem, mathematicians autline a mathematics model, called as "the digit—grouped partition of sequence", and give some results. In this paper, we get a recursion formula, by which the number of digit of a sequence is reduced step by step, and obtain a dynamic algorithm solving a number of the optimal digit—grouped partition of sequence.