Partitioning Number Sequences into Optimal Subsequences

J. Zobel

Department of Computer Science
RMIT University
GPO Box 2476V
Melbourne 3001 Australia
email: jz@cs.rmit.edu.au

P. Dart

Department of Computer Science and Software Engineering
The University of Melbourne
Parkville 3052 Australia
email: philip@cs.mu.oz.au

We consider how to partition finite sequences of positive numbers into subsequences such that each resulting subsequence has a sum of at least a given minimum. Given several different optimality criteria, based on maximising the number of subsequences and minimising their variance in size, we develop and analyse a series of algorithms that yield optimal solutions.

Keywords: algorithms, analysis of algorithms, computational complexity.

1. INTRODUCTION

Consider a finite sequence of positive numbers that is to be partitioned into subsequences as follows. Each subsequence must contain numbers whose sum is at least L, where L is fixed; the variation between the sums is to be kept small; and the sums should be close to L. That is, the criteria for the partitioning are that the number of sums should be maximised and the variance should be minimised.

The problem of partitioning arose from our work in information retrieval, where it can be desirable to divide documents consisting of small units of text (such as sentences or paragraphs) into larger units of consistent length, for retrieval or presentation (Kaszkiel *et al*, 1999; Zobel *et al*, 1995). First, enforcing a minimum length helps ensure that there is sufficient content to allow accurate matching of queries and documents. Second, when units of text significantly vary in length, estimation of their relevance to a query is inaccurate, so it is desirable to reduce the variance. Partitioning can also be used for division of irregularly-sized data such as web pages into packets for network transmission, where there is a trade-off between response time and total cost of transmission.

Similar problems have been described by other authors. Larmore and Hirschberg (1985) have considered how to break scrolls into pages whose lengths must fall within a certain range. In

Manuscript received: June 2000
Editor: John Roddick