Accepted Manuscript

Interval selection in the streaming model

Sergio Cabello, Pablo Pérez-Lantero

PII: S0304-3975(17)30624-2
DOI: http://dx.doi.org/10.1016/j.tcs.2017.08.015
Reference: TCS 11289

To appear in: *Theoretical Computer Science*

Received date: 8 February 2017
Revised date: 18 August 2017
Accepted date: 21 August 2017

Please cite this article in press as: S. Cabello, P. Pérez-Lantero, Interval selection in the streaming model, *Theoret. Comput. Sci.* (2017), http://dx.doi.org/10.1016/j.tcs.2017.08.015

Interval Selection in the Streaming Model

Sergio Cabello[a,1], Pablo Pérez-Lantero[b,2]

[a] *Department of Mathematics, FMF, University of Ljubljana, Slovenia.*
[b] *Departamento de Matemática y Ciencia de la Computación, Universidad de Santiago (USACH), Chile.*

Abstract

A set of intervals is independent when the intervals are pairwise disjoint. In the interval selection problem we are given a set $\mathbb{I}$ of intervals and we want to find an independent subset of intervals of largest cardinality. Let $\alpha(\mathbb{I})$ denote the cardinality of an optimal solution. We discuss the estimation of $\alpha(\mathbb{I})$ in the streaming model, where we only have one-time, sequential access to the input intervals, the endpoints of the intervals lie in $\{1, \dots, n\}$, and the amount of the memory is constrained.

For intervals of different sizes, we provide an algorithm in the data stream model that given $\varepsilon \in (0, 1/2)$ computes an estimate $\hat{\alpha}$ of $\alpha(\mathbb{I})$ that, with probability at least 2/3, satisfies $\frac{1}{2}(1-\varepsilon)\alpha(\mathbb{I}) \leq \hat{\alpha} \leq \alpha(\mathbb{I})$. For same-length intervals, we provide another algorithm in the data stream model that given $\varepsilon \in (0, 1/2)$ computes an estimate $\hat{\alpha}$ of $\alpha(\mathbb{I})$ that, with probability at least 2/3, satisfies $\frac{2}{3}(1-\varepsilon)\alpha(\mathbb{I}) \leq \hat{\alpha} \leq \alpha(\mathbb{I})$. The space used by our algorithms is bounded by a polynomial in ε^{-1} and $\log n$. We also show that no better estimations can be achieved using $o(n)$ bits of storage.

We also develop new approximate solutions to the interval selection problem, where the intervals have real endpoints and we want to report a feasible solution, that use $O(\alpha(\mathbb{I}))$ space. Our algorithms for the interval selection problem match the optimal results by Emek, Halldórsson and Rosén [Space-Constrained Interval Selection, TALG 2016], but are much simpler.

Keywords: Intervals, Independent set, Data stream, Random estimation, Wise independent hash functions, Approximation algorithms

1. Introduction

Several fundamental problems have been explored in the data streaming model; see [3, 16] for an overview. In this model we have bounds on the amount of available memory, the data arrives sequentially, and we cannot afford to look at input data of the past, unless it was stored in our limited memory. This is effectively equivalent to assuming that we can only make one pass over the input data.

In this paper, we consider the interval selection problem. Let us say that a set of intervals is ***independent*** when all the intervals are pairwise disjoint. In the ***interval selection problem***, the input is a set $\mathbb{I}$ of intervals with real endpoints and we want to find an independent subset of largest cardinality. Let us denote by $\alpha(\mathbb{I})$ this largest cardinality. There are actually two different problems: one problem is finding (or approximating) a largest independent subset,

Email addresses: `sergio.cabello@fmf.uni-lj.si` (Sergio Cabello), `pablo.perez.l@usach.cl` (Pablo Pérez-Lantero)

[1] Supported by the Slovenian Research Agency, program P1-0297, projects J1-4106 and L7-5459; by the ESF EuroGIGA project (project GReGAS) of the European Science Foundation.

[2] Supported by projects CONICYT FONDECYT/Regular 1160543 (Chile), and Millennium Nucleus Information and Coordination in Networks ICM/FIC RC130003 (Chile).