Data Structures for Approximate Orthogonal Range Counting

Yakov Nekrich

Dept. of Computer Science
University of Bonn
yasha@cs.uni-bonn.de

Abstract. We present new data structures for approximately counting the number of points in an orthogonal range. There is a deterministic linear space data structure that supports updates in $O(1)$ time and approximates the number of elements in a 1-D range up to an additive term $k^{1/c}$ in $O(\log \log U \cdot \log \log n)$ time, where $k$ is the number of elements in the answer, $U$ is the size of the universe and $c$ is an arbitrary fixed constant. We can estimate the number of points in a two-dimensional orthogonal range up to an additive term $k^\rho$ in $O(\log \log U + (1/\rho) \log \log n)$ time for any $\rho > 0$. We can estimate the number of points in a three-dimensional orthogonal range up to an additive term $k^\rho$ in $O(\log \log U + (\log \log n)^3 + (3^v) \log \log n)$ time for $v = \log \frac{1}{\rho} / \log \frac{3}{2} + 2$.

1 Introduction

Range reporting and range counting are two variants of the range searching problem. In the range counting problem, the data structure returns the number of points in an arbitrary query range. In the range reporting problem the data structure reports all points in the query range. Both variants were studied extensively and in many cases we know the matching upper and lower bounds for those problems for dimension $d \leq 4$. Answering an orthogonal range counting query takes more time than answering the orthogonal range reporting query in the same dimension. This gap cannot be closed because of the lower bounds for the range counting queries: while range reporting queries can be answered in constant time in one dimension and in almost-constant time in two and three dimensions (if the universe size is not too big), range counting queries take super-constant time in one dimension and poly-logarithmic time in two and three dimensions.

Approximate range counting queries help us bridge the gap between range reporting and counting: instead of exactly counting the number of points (elements) in the query range, the data structure provides a good estimation. There are data structures that approximate the number of points in a one-dimensional interval [4, 19] or in a halfspace [7, 15, 2, 8] up to a constant

---

1 For simplicity, we consider only emptiness queries. In other words, we ignore the time needed to output the points in the answer: if range reporting data structure supports queries in $O(f(n) + k)$ time, we simply say that the query time is $O(f(n))$. 

Y. Dong, D.-Z. Du, and O. Ibarra (Eds.): ISAAC 2009, LNCS 5878, pp. 183–192, 2009. © Springer-Verlag Berlin Heidelberg 2009
factor: given a query $Q$, the data structure returns the number $k'$ such that 
$(1 - \varepsilon)k \leq k' \leq (1 + \varepsilon)k$, where $k$ is the exact number of points in the answer and $\varepsilon$ is an arbitrarily small positive constant. In this paper we consider the following new variant of approximate range counting: If $k$ is the number of points in the answer, the answer to a query $Q$ is an integer $k'$ such that $k - \varepsilon k^\alpha \leq k' \leq k + \varepsilon k^\alpha$ for some constant $\alpha < 1$. Thus we obtain better estimation for the number of points in the answer for large (superconstant) values of $k$. On the other hand, if the range $Q$ is empty, then $k' = 0$. We present data structures that approximate the number of points in a $d$-dimensional orthogonal range for $d = 2, 3$. We also describe a dynamic one-dimensional data structure.

Dynamic 1-D Data Structure. A static data structure that answers 1-D reporting queries in $O(1)$ time is described in [4]. In [4] the authors also describe a static data structure that approximates the number of points in a 1-D range up to an arbitrary constant factor in constant time. Pătraşcu and Demaine [24] show that any dynamic data structure with polylogarithmic update time needs $\Omega((\log n / \log \log U)$ time to answer an exact range counting query; henceforth $U$ denotes the size of the universe. The dynamic randomized data structure of Mortensen [19] supports approximate range counting queries in $O(1)$ time and updates in $O(\log^2 U)$ time; see [19] for other trade-offs between query and update times. In this paper we present a new result on approximate range counting in 1-D:

- There is a deterministic data structure that can answer one-dimensional approximate range counting queries using the best known data structure for predecessor queries, i.e. dynamic data structure supports range reporting queries in $O(dpred(n, U))$ time, where $dpred(n, U)$ is the time to answer a predecessor query in the dynamic setting; currently $dpred(n, U) = O(\min(\log \log U \cdot \log \log n, \sqrt{\log n / \log \log n}))$. We show that we can approximate the number of points in the query range up to an additive factor $k^{1/c}$, where $k$ is the number of points in the answer and $c$ is an arbitrary constant, in $O(dpred(n, U))$ time. We thus significantly improve the precision of the estimation; the query time is still much less than the lower bound for the exact counting queries in the dynamic scenario.

Using the standard techniques, we can extend the results for one-dimensional approximate range counting to an arbitrary constant dimension $d$. There is a data structure that approximates the number of points in a $d$-dimensional range up to an additive term $k^c$ for any $c > 0$ in $O(\log \log n (\log n / \log \log n)^{d-1})$ time and supports updates in $O(\log^{d-1+c} n)$ time. For comparison, the fastest known dynamic data structure [18] supports emptiness queries in $O((\log n / \log \log n)^{d-1})$ time. Dynamic data structures are described in section 2.

Approximate Range Counting in 2-D and 3-D. We match or almost match the best upper bounds for 2-D and 3-D emptiness queries. Best data structures for exact range counting in 2-D and 3-D support queries in $O(\log n / \log \log n)$ and $O((\log n / \log \log n)^2)$ time respectively [14].