Advanced Algorithms Ex.4

Dvir Fried

Deadline: July 19 at 2PM

Note: Questions that start with ☺ are not for submission and are given as warm-up and for the sake of your confidence with the materials, but be aware that you are supposed to know how to solve them.

1 Stringology

Either-or Matching

Propose an efficient algorithm for the following problem.

Input: A text $T[0 \dots n-1]$ over alphabet $\mathbb{N}$. A pattern $P[0 \dots m-1]$ over alphabet $\mathbb{N} \times \mathbb{N}$ i.e. every index of P is a pair of integers. Denote $P[j] = (p_j^1, p_j^2)$.

Output: Every index $i \in [0 \dots n-m]$ such that for every $j \in [0 \dots m-1]$ we have $p_j^1 = T[i+j]$ or $p_j^2 = T[i+j]$.

An algorithm whose runtime is $O(n\sqrt{m \log m})$ would get 75% of the points, while an $O(n \log m)$ runtime algorithm would get full points. No need for a formal proof; just explain why your algorithm would work.

k-Edit Distance with Wildcards

For the symbol $\lozenge$ (which can be matched with any other letter in Σ), we denote by $ED_\lozenge(S,T)$ the edit distance between the strings S and T where $\lozenge$ can appear in either T or P.

Propose an algorithm with $O(nk)$ time complexity, for the following problem.

Input: String S and string T, of size n and $m \leq n$ respectively, over alphabet $\Sigma \cup \{\lozenge\}$.

Output: If the $ED_\lozenge(S,T) \leq k$ return $ED_\lozenge(S,T)$. Otherwise, return that $ED_\lozenge(S,T) > k$.

No need for a formal proof; just explain why your algorithm would work.

2 Communication Complexity

Median

☺ In class you defined the two-way communication complexity (CC) median problem, and have seen a protocol with communication complexity of $O(\log^2 n)$ bits.

Show a protocol with communication complexity of $O(\log n)$ bits.

CC Lower bound

Recall that in class we have seen:

1. Reduction from one-way CC *Indexing* ($Indexing_{OW}$) to *Median* in the streaming model.
2. Reduction from *Disjointness* to *Distinct count* (DC) in the streaming model.

- ☺ Show a reduction from $Indexing_{OW}$ to DC.
- ☺ Show a reduction from *Disjointness* to Median.
- Show that the problem of L_∞-approximation (which return a value $\hat{L}$ such that $(1-\varepsilon)L_\infty \leq \hat{L} \leq (1+\varepsilon)L_\infty$) has lower bound of $\Omega(n)$.

3 Streaming Model

Sliding window

The ε-*approximation Distinct Count in a Window* (DCW_ε) streaming problem is defined as follows.

Input: window of size w and a stream $x_1, \ldots, x_n$ of elements from the universe $U = \{0, \ldots, u-1\}$.

Output: When receiving the jth element e_j of the stream, output an ε-approximation of the distinct count of elements in the window $e_{j-(w-1)}, e_{j-(w-2)}, \ldots, e_j$

- Show a space-efficient algorithm for DCW_ε problem with failure probability of $\delta > 0$, and analyze its expected space.
- Prove that any algorithm for the DCW_ε problem requires $\Omega(\frac{1}{\varepsilon})$ space.