CSE6512: Randomization in Computing

Lecture 5, Sep 13 2011.

Notes by Maryam Shokrniafard

Fingerprinting techniques are used to verify equality of two objects. ($obj_1 \stackrel{?}{=} obj_2$)

Example1:

Input: three n×n matrices A,B and C

Output: "yes" if AB=C

"no" otherwise.

Fact: This problem can be solved in $O(n^{2.376})$ time using the best known deterministic matrix multiplication algorithm.

By using fingerprinting technique, we can solve this problem more efficiently.

A randomized algorithm 1 (Monte Carlo)

Pick a random vector $r=[r_1,r_2,\ldots, r_n]$ where r_i is from $\{0,1\}$ *($r_1,..,r_n$ are picked using a 2-sided fair coin)*

Check if A (Br) = Cr.

End of alg.

Fact: if AB ≠ C then prob.[A(Br)= Cr] ≤ $\frac{1}{2}$

Proof:

Let D=AB-C

Assume that D≠0. This means that there exists at least one non-zero row in D.

Without loss of generality, let this row be $[d_1, d_2, \ldots, d_k, 0, .., 0]$ where k≥1. Obviously we get

Prob.[Dr=0 | D ≠ 0] ≤ prob.[$\sum_{i=1}^{k} d_i \times r_i = 0$ | D≠0]

RHS ≤ prob.[$r_1 = \frac{-\sum_{i=2}^{k} d_i \times r_i}{d_1}$ | D≠0]

We can employ the principle of deferred decisions here. In particular, assume that we have already chosen r_2, .., r_n, and now we are fixing the value of r_1.

Since $r_1 \in \{0,1\}$, we can see that the above probability is $\leq \frac{1}{2}$.

A better algorithm for this problem would be:

Algorithm 2

For i=1 to αlogn do

Pick a random r $\in \{0, 1\}^n$

If A (Br) ≠ Cr then output "AB ≠ C" and quit

End for

Output: "AB=C"

End of alg.

Probability that the above algorithm gives an incorrect answer is ≤ $\left(\frac{1}{2}\right)^{\alpha \log n} = n^{-\alpha}$

$\Longrightarrow$ run time of the algorithm is O(n^2log n).

Example 2.

Input: two degree-n polynomials $p_1(x)$, $p_2(x)$ and one 2n-degree polynomial $p_3(x)$

Output: "yes" if $p_1(x) \times p_2(x) = p_3(x)$ and

"no" otherwise.

Fact: we can solve this problem using DFT in O(n log n) time.

A Randomized algorithm

Let S be a subset of the field F s.t $|S| > 2n$

Let $Q(x)=p_1(x) \times p_2(x)-p_3(x)$

Pick a random $r \in S$

If $Q(r) = 0$ output "yes"

Else output "no"

End of alg.

Analysis:

If $p_1(x)\times p_2(x)=p_3(x)$, the algorithm will never give an correct answer. On the other hand, if $p_1(x)\times p_2(x)\neq p_3(x)$, the algorithm might give an incorrect answer. This algorithm has one-sided error.

Now we show that the probability that the algorithm gives an incorrect answer is very low: In particular, we'll show that $\text{Prob.}[Q(r)=0 \mid Q(x) \neq 0] < n^{-\alpha}$.

Note: Q(x) is a degree-2n polynomial and hence it has at most 2n distinct zeros.

$\Longrightarrow \text{Prob}[Q(r)=0 \mid Q(x) \neq 0] \leq \frac{2n}{|S|}$ **.** We want this to be low.

$\Longrightarrow \frac{2n}{|s|} \leq n^{-\alpha}$. This will happen if

$|s|>2\, n^{\alpha+1}$.

Example 3.

Input: A multivariate polynomial on n variables $Q(x_1, x_2, \ldots, x_n)$

Output: "yes" if $Q(x_1, x_2, \ldots, x_n) \equiv 0$

"no" otherwise

Definitions:

1. Degree of a term is the sum of exponents of the variables in the term.
2. Total degree of $Q(x_1, x2, \ldots, x_n)$ is the maximum degree of its terms.

Theorem: (Schwartz & Zippel)

Let S be a subset of the field F and let $r_1, r_2, \ldots, r_n$ be random elements from S. Then, $\text{prob.}[Q(r_1, r_2, \ldots, r_n)=0 \mid Q(x_1, x_2, \ldots, x_n) \not\equiv 0] \leq \frac{d}{|s|}$ where d is the total degree of $Q(x_1, x_2, \ldots, x_n)$.

Fact: $\text{Prob.}[A] \leq \text{prob.}[A|\overline{B}]+\text{prob.}[B]$ since $\text{prob.}[A]=\text{prob.}[A|B]\ \text{prob.}[B] + \text{prob.}[A|\overline{B}]\ \text{prob.}[\overline{B}] \leq \text{prob.}[A|\overline{B}]+\text{prob.}[B]$

Proof:

We use induction on n.

Base case: The theorem holds for n=1 (proof is given in example 2).

Induction step: Assume that the theorem holds for up to n-1 variables. We show that it holds for n variables as well.

Let $Q(x) = \sum_{i=0}^{k} x_1^i Q_i(x_2, x_3, \ldots, x_n)$ where $k \leq d$.

$\Longrightarrow Q_k(x_2, x_3, \ldots, x_n) \not\equiv 0$ and the total degree of $Q_k(x_2, x_3, \ldots, x_n) \leq d-k$.

By induction hypothesis, we get

$\text{Prob.}[Q_k(r_2, r_3, \ldots, r_n)=0] \leq \frac{d-k}{|s|}$ ----------(1)

Let $Q(x_1, r_2, \ldots, r_n)=q(x_1)= \sum_{i=0}^{k} x_1^i Q_i(r_2, r_3, \ldots, r_n)$

Prob.[$q(r_1)=0 | Q_k(r_2,r_3,\ldots,r_n) \neq 0$] $\leq \frac{k}{|s|}$----------(2)

Let A be the event

$Q(r_1,r_2,\ldots, r_n)=0$

Let B be the event

$Q_k(r_2,\ldots, r_n)=0$

Using the fact prob.[A]<prob.[A|$\overline{B}$]+prob.[B]

We get (from Equations 1 and 2)

Prob.[$Q(r_1,r_2,\ldots, r_n)=0 \mid Q(x_1, x_2,\ldots, x_n) \not\equiv 0\} \leq$

Prob.[$Q(r_1,r_2,\ldots, r_n)=0 \mid Q_k(r_2,\ldots, r_n) \neq 0$]+prob.[$Q_k(r_2,\ldots, r_n)=0$] $\leq \frac{d-k}{|s|}+\frac{k}{|s|} \leq \frac{d}{|s|}$

■