Дерандомизация: основные результаты и подходы
Лекция 1: Вычисления с односторонней ошибкой и дисперсеры.

Дмитрий Ицыксон

ПОМИ РАН

1-ое декабря 2006

План

- Вероятностные алгоритмы с односторонней ошибкой (кратко)
- Модель черного ящика
- Дисперсеры и экономия случайных битов
- Конструкция дисперсеров на основе экспандеров
- Конструкция дисперсеров на основе универсального хеширования

RP: вероятностные алгоритмы с односторонней ошибкой

Определение:

Язык $L \in \mathbf{RP}$, если существует полиномиальный вероятностный алгоритм A, такой что

- $A(x) = 0$, при $x \notin L$
- $P\{A(x) = 1\} \geq \frac{1}{2}$, при $x \in L$

Пример

Дана матрица над $\mathbb{Z}[x]$ (элементы — многочлены с целыми коэффициентами). Определить вырождена она или нет?

Модель черного ящика

Задача

Алиса помечает $(1-\varepsilon)N$ из N карт на лицевой стороне. Карты раскладываются рубашкой вверх. Боб должен найти за d попыток хотя бы одну помеченную карту, открыв не более, чем d карт.

Соображения

- Если $d < \varepsilon N$, то детерминированной стратегии у Боба нет.
- Если Боб будет выбирать случайным образом, то вероятность не найти помеченную карту за d попыток равняется ε^d. При этом Боб будет использовать примерно $d \log N$ случайных бит.

Модель черного ящика

Уточнение задачи

- Ограничим число случайных битов, доступных Бобу числом R
- Стратегия Боба: $\phi : \{0,1\}^R \to 2^{\{1,2,\ldots,N\}}, \forall r |\phi(r)| \leq d$
- ϕ называем $(N, \varepsilon, R, d, p)$-стратегией, если для любого хода Алисы вероятность неуспеха Боба меньше p.

Соответствие стратегия ϕ — двудольный граф G_ϕ

$$G_\phi(V = \{0,1\}^R \cup \{1,2,\ldots,N\}, E)$$
$$(r,x) \in E \iff x \in \phi(r).$$

Дисперсер

Определение

Двудольный граф $G(V = (A \cup B), E \subseteq A \times B)$ называется $(1 - \varepsilon)$-попадающим дисперсером с порогом T, если $\forall S \subseteq A, |S| \geq T$, множество соседей множества S по размеру больше, чем $\varepsilon|B|$.

Теорема

Для того, чтобы стратегия ϕ являлась $(N, \varepsilon, R, d, p)$-стратегией необходимо и достаточно, чтобы граф G_ϕ был $(1 - \varepsilon)$-попадающим дисперсером с порогом $p \cdot 2^R$.

Дисперсер

Определение

Двудольный граф $G(V = (A \cup B), E \subseteq A \times B)$ называется $(1 - \varepsilon)$-попадающим дисперсером с порогом T, если $\forall S \subseteq A, |S| \geq T$, множество соседей множества S по размеру больше, чем $\varepsilon|B|$.

Теорема

Для того, чтобы стратегия ϕ являлась $(N, \varepsilon, R, d, p)$-стратегией необходимо и достаточно, чтобы граф G_ϕ был $(1 - \varepsilon)$-попадающим дисперсером с порогом $p \cdot 2^R$.

Доказательство

Необходимость

- Пусть $S \subseteq \{0,1\}^R$, $|S| \geq p \cdot 2^R$, а множество соседей не больше, чем εN.
- Алиса помечает все карты, к которым нет ребра из S (их хотя бы $(1-\varepsilon)N$).
- Тогда вероятность неуспеха Боба не меньше, чем p. Противоречие!!!

Достаточность

Пусть S множество случайных битов, на которых достигается неуспех Боба, если бы $|S| \geq p \cdot 2^R$, то множество соседей S было бы более εN и Боб выиграл бы.

Понижение вероятности ошибки с помощью дисперсера

- Для каждого k можно построить $\frac{1}{2}$-попадающий дисперсер с множеством вершин $A = \{0,1\}^r \cup B = \{0,1\}^r$ и порогом $\frac{2^r}{r^k}$ степени $r^{O(1)}$.
- Пусть для языка L есть алгоритм, проверяющий принадлежность, с вероятностью ошибки $\frac{1}{2}$, использующий r случайных битов.
- Возьмем случайную строчку x из A и для всех ее соседей y в дисперсере запустим алгоритм, где вместо случайных битов будут выступать биты y.
- Вероятность ошибки $\frac{1}{r^k}$
- Время работы увеличилось в полином раз
- Число случайных битов не изменилось

Требование к дисперсеру

Возможность за полиномиальное время построить множество соседей вершины.

Хранить весь дисперсер возможности нет!!!

Конструкция дисперсера, основанная на экспандерах

Определение

Неориентированный d-регулярный граф называется c-экспандером, $(c > 0)$. Если $\forall V' \subseteq V, |V'| \leq \frac{1}{2}|V|$ выполняется: $|V' \cup \Gamma(V')| \geq (1 + c)|V'|$, где $\Gamma(V')$ — это множество вершин, смежных вершинам из V'.

Экспандеры Gabber-Gallil

Двудольный граф $V = (\mathbb{Z}_m)^2 \cup (\mathbb{Z}_m)^2$. Вершина (x, y) из одной доли соединена с вершинами $(x, y), (x, x + y), (x, x + y - 1), (x + y, y), (x + y + 1, y)$ из другой доли.

Этот граф 5-регулярный c-экспандер для маленького $c > 0$.

Дисперсер по экспандеру

Конструкция Карпа

Пусть $G(V', E')$ — экспандер, строим дисперсер. $A = B = U'$, для параметра l определим множество ребер $E = \{(u, v)|$ есть путь длины не более l от u до v в графе $G\}$.

- Нам хочется, чтобы $\forall S \subseteq A, |S| \geq T$ выполнялось бы $|\Gamma(S)| > \epsilon|V|$.
- Если множество соседей S больше $\frac{1}{2}|A|$, то все хорошо: $|\Gamma(S)| \geq \frac{1}{2}|B|$.
- Иначе $\Gamma(S) \geq (1 + c)^l|S| \geq \frac{1}{2}|B|$, при $T = \frac{1}{2(1+c)^l}|B|$
- Мы получили $(\frac{1}{2} + \varepsilon)$-попадающий дисперсер для любого $\varepsilon > 0$, степень каждой вершины $\leq d^l$.

Дисперсер по экспандеру

Применяем конструкцию Карпа к графам Gabber-Galil

Theorem

Для любой константы k существует явный $\frac{1}{2}$-попадающий дисперсер $E_r \subset V_r \times V_r$, где $|V_r| = 2^r$, степени $r^{O(1)}$ с порогом $2^r/r^k$

Универсальное хеширование

Задача хеширования

Хочется иметь детерминированную функцию, которая “равномерно” разбрасывала элементы большого множества X по k ящикам, $k < |X|$.

Идея универсального хеширования

Если мы не можем считать, что элементы поступают в случайном порядке, то мы можем сами выбирать хешфункцию случайным образом.

Определение

Множество отображений $\mathfrak{H}$ из $\{0,1\}^n$ в $\{0,1\}^m$ называется универсальным семейством хеш-функций, если для любых $x, y \in \{0,1\}^n, x \neq y$ выполняется $P\{h(x) = h(y)\} \leq 2^{-m}$, где h берется равномерно из $\mathfrak{H}$

Универсальное хеширование

Определение

Множество отображений $\mathfrak{H}$ из $\{0,1\}^n$ в $\{0,1\}^m$ называется сильным универсальным семейством хеш-функций, если для любых $x, y \in \{0,1\}^n, x \neq y$ и $a, b \in \{0,1\}^m$ выполняется $P\{h(x) = a \wedge h(y) = b\} = 2^{-2m}$, где h берется равномерно из $\mathfrak{H}$

Пример сильного универсального семейства хеш-функций

$\mathbf{F}$ — поле, $|\mathbf{F}| = 2^n$. $\mathfrak{H} = \{h_{a,b} : \mathbf{F} \to \mathbf{F} | a, b \in \mathbf{F}, h_{a,b}(x) = ax + b\}$. Есть биекция между $\mathbf{F}$ и $\{0,1\}^n$. Можно выкидывать ненужные биты и получить $m<n$

Дисперсер из универсального семейства

Конструкция

Пусть есть сильное универсальное семейство хеш-функций $\mathfrak{H}$, отображающих множество D на B. Рассмотрим двудольный граф с долями $(A = \mathfrak{H}, B)$. $E = \{(h, h(x)) | h \in A, x \in D\}$.

Теорема

Если $D > \lceil \frac{1}{\varepsilon} \rceil$, то этот граф является $\frac{1}{2}$-попадающим дисперсером степени D с порогом $\varepsilon|A|$.

Доказательство

- Возьмем $S \subset A, |S| \geq \varepsilon|A|$, надо показать, что $|\Gamma(S)| \geq |B|/2$. Или, что для любого $B' \subseteq B, |B'| \geq |B|/2$ для слуайной $h \in A$ вероятность существования ребра из h хотя бы $1 - \varepsilon$.
- Фиксируем B'. Случайная переменная Y_i, $Y_i = 1$, если $h(i) \in B'$, иначе $Y_i = 0$
- Существование ребра между h и B' эквивалентно $\sum_{i \in D} Y_i > 0$.
- $E[Y_i] = |B'|/|B| \geq \frac{1}{2}$
- $D[Y_i] = E[Y_i](1 - E[Y_i]) \leq 1/4$

Доказательство

- $E[\sum_i Y_i] = \sum_i E[Y_i] \geq |D|/2$
- Так как Y_i попарно независимы, то $D[\sum_i Y_i] = \sum_i D[Y_i] \leq |D|/4$
- Неравенство Чебышева: $P\{|X - \mu| \geq t\} \leq \sigma^2/t^2$
- $P\{$ нет ребра между h и $B'\} = P\{\sum_{i \in D} Y_i = 0\} = P\{\sum_{i \in D} Y_i \leq E[\sum_{i \in D} Y_i] - |D|/2\} \leq (|D|/4)/(|D|/2)^2 = 1/|D| < \varepsilon$