Дерандомизация: основные результаты и подходы
Лекции 2-3: Вычисления с двусторонней ошибкой и экстракторы. Экстрактор Тревисана

Дмитрий Ицыксон

ПОМИ РАН

15,18-ое декабря 2006

План

- Вероятностные алгоритмы с двусторонней ошибкой (кратко)
- Мажорирующий дисперсер
- Экстрактор
- Экстрактор Trevisana: самовосстанавливающие коды
- Экстрактор Trevisana: конструкция Nisan-Wigderson
- Экстрактор Trevisana: собираем все вместе

BPP: вероятностные алгоритмы с двусторонней ошибкой

Определение:

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

- $P\{A(x) = 0\} \geq \frac{2}{3}$, при $x \notin L$
- $P\{A(x) = 1\} \geq \frac{2}{3}$, при $x \in L$

Мажорирующий дисперсер

Определение

Двудольный граф $G(V = (A \cup B), E \subseteq A \times B)$ называется мажорирующим $(1 - \varepsilon)$-попадающим дисперсером с порогом T, если $\forall S \subseteq A, |S| \geq T$ и для любого $M \subset B, |M| \geq (1 - \varepsilon)|B|$ выполняется $|\Gamma(S) \cap M| \geq \frac{1}{2}|\Gamma(S)|$.

Применение

Понижение вероятности ошибки алгоритмов из **BPP**

Распределения

Определение

Распределения D_1 и D_2 на множестве A называются ε-близкими, если $\sum_{x \in A} |p_1(x) - p_2(x)| \leq \epsilon$.

Определение

Минимальной энтропией распределения D на множестве A называется $H_\infty(D) = min_{x \in D} \log \frac{1}{p(x)}$.

$H_\infty \geq k \iff \forall x \; p(x) \leq 2^{-k}$

Распределения

Определение

Распределения D_1 и D_2 на множестве A называются ε-близкими, если $\sum_{x \in A} |p_1(x) - p_2(x)| \leq \epsilon$.

Определение

Минимальной энтропией распределения D на множестве A называется $H_\infty(D) = min_{x \in D} \log \frac{1}{p(x)}$.

$H_\infty \geq k \iff \forall x \; p(x) \leq 2^{-k}$

Экстрактор

Определение

Отображение $E : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ называется экстрактором с параметрами (k, ε), если для любого распределения D с $H_\infty(D) \geq k$ распределенииeh $E(D, U_d)$ является ε-близким к равномерному. (U_d — равномерное распределение на $\{0,1\}^d$)

Экстрактор — это мажорирующий дисперсер

Предложение

(k, ε)-Экстрактор является мажорирующим $(1/2 + \varepsilon/2)$-попадающим дисперсером с порогом 2^k.

Доказательство

Рассмотрим $V \subset \{0,1\}^n, |V| \geq 2^k$. Пусть D — равномерное распределение на V, тогда в образе D не больше $2^m\varepsilon$ нулей.

Экстрактор по экспандеру

Случайное блуждание по экспандеру — это экстрактор.

Коды с коррекцией ошибки

Определение

Кодом с коррекцией ошибки с минимальным расстоянием δ над множеством S называется отображение $e : S^n \longrightarrow S^m$, где $n < m$ такое, что множество вхождение i, где $e(x)_i \neq e(y)_i$ при $x \neq y$, составляет долю не менее δ.

Код Reed-Solomon

F — конечное поле размера $m > n$. $e_{RS} : F^n \longrightarrow F^m$ Пусть $i_1, i_2, \ldots i_n \in F$. $x \in F^n$, тогда существует ровно один многочлен степени не более $n-1$ такой, что $p(i_j) = x_j, j \in \{1, \ldots, n\}$. Определим $e_{RS}(x) = (p(y))_{y \in F}$. $\delta = \frac{m-n+1}{m}$.

Коды с коррекцией ошибки

Код Адамара

$e_H : \{0,1\}^l \to \{0,1\}^{2^l}$.
$e_H(x) = (< x, y >)_{y \in \{0,1\}^l}$.
$\delta = 1/2$

Конкатенация $e_H \circ e_{RS}$

$|F| = 2^l = m$. $e_H \circ e_{RS} : \{0,1\}^n \to \{0,1\}^{m2^l}$
$y = y_1 \dots y_m = e_{RS}(x)$, $y_i \in F = \{0,1\}^l$
$e_H \circ e_{RS} = e_H(y_1) \circ e_H(y_2) \circ \dots \circ e_H(y_m)$

Коды с коррекцией ошибки

Лемма 1

$n < m = 2^l$, тогда $e = e_H \circ e_{RS}$ — код
$e : \{0,1\}^n \to \{0,1\}^{m^2}$
$\delta = \frac{1}{2} - \frac{n-1}{2m}$

Доказательство

$\delta \geq \frac{m-n+1}{m} \frac{1}{2}$

Следствие

Если n — степень двойки, то существует полиномиально вычислимое кодирование $e : \{0,1\}^n \to \{0,1\}^{n^{10}}$ с $\delta = \frac{1}{2} - \frac{1}{n^4}$.

Лемма 2

При таком кодировании любой шар Хемминга с центром в $\{0,1\}$ и радиусом $n^{10}(\frac{1}{2} - \frac{1}{n^2})$ содержит не более, чем $n^4/2$ кодовых слов.

Коды с потерями

Определение

Кодом с потерями называется пара отображений $c : \{0,1\}^m \to \{0,1\}^l$ и $d : \{0,1\}^l \to \{0,1\}^m$.
D — распределение на $\{0,1\}^m$, тогда искажением называется $\phi = P_{x \in D, i \in U_m}\{d(c(x))_i \neq x_i\}$.

Лемма

Пусть $e : \{0,1\}^m \to \{0,1\}^{m^{10}}$ — код с коррекцией ошибки. D — распределение на $\{0,1\}^m$, $H_\infty(D) \geq t$, $e(D)$ — распределение на $\{0,1\}^{m^{10}}$, тогда для любого кода с потерями $c : \{0,1\}^{m^{10}} \to \{0,1\}^l$ и искажением $\phi < \frac{1}{2} - \frac{2}{m^2}$ выполняется $l \geq t - 6 \log m$.

Коды с потерями

Определение

Кодом с потерями называется пара отображений $c : \{0,1\}^m \to \{0,1\}^l$ и $d : \{0,1\}^l \to \{0,1\}^m$.
D — распределение на $\{0,1\}^m$, тогда искажением называется $\phi = P_{x \in D, i \in U_m}\{d(c(x))_i \neq x_i\}$.

Лемма

Пусть $e : \{0,1\}^m \to \{0,1\}^{m^{10}}$ — код с коррекцией ошибки. D — распределение на $\{0,1\}^m$, $H_\infty(D) \geq t$, $e(D)$ — распределение на $\{0,1\}^{m^{10}}$, тогда для любого кода с потерями $c : \{0,1\}^{m^{10}} \to \{0,1\}^l$ и искажением $\phi < \frac{1}{2} - \frac{2}{m^2}$ выполняется $l \geq t - 6 \log m$.

Доказательство леммы

y — случайный элемент $e(D)$; G — событие, означающая $\rho_{Ham}(y, d(c(y))) < (\frac{1}{2} - \frac{1}{m^2})m^{10}$.
Имеем $P\{y \in G\} \geq \frac{1}{m^2}$, иначе искажение $\phi > (1 - \frac{1}{m^2})(\frac{1}{2} - \frac{1}{m^2}) > \frac{1}{2} - \frac{2}{m^2}$.
Пусть w — самое популярное значение среди $c(y), y \in G$.
Обозначим событие G': $y \in G, c(y) = w$. $P\{y \in G'\} \geq \frac{1}{2^l m}$.
В шаре радиуса $(\frac{1}{2} - \frac{1}{m^2})m^{10}$ не более, чем $m^4/2$ элементов с ненулевой вероятностью в $e(D)$. Поэтому $P\{G'\} \leq \frac{m^4}{2}2^{-t}$.
Итого $\frac{1}{2^l m} \leq \frac{m^4}{2}2^{-t}$, т.е., $l \geq t - 6 \log m$.

Дизайн Nisan-Wigderson

Дизайн

$S_1, S_2, \ldots, S_m$ — s-элементные подмножества $\{1, 2, \ldots n\}$. Для любых $i \neq j$, $|S_i \cap S_j| \leq l$.

Лемма

Для любой коснтанты $c \geq 1$ существует дизайн с параметрами: $m, n = O(\log m), s = c \log m), l = \log m$. И его можно построить за полиномиальное от m время.

Доказательство

$n = 100c^2 \log m$ Строим индуктивно $S_1, \ldots, S_i$. Для построения S_{i+1} выбираем случайным образом $2c \log m$ элементов: с большой вероятностью, среди них хотя бы $c \log m$ различных и пересечения с остальными меньше $\log m$.

Предсказатель

Определение

Функция f называется ε-предсказателем для распределения D, если для случайного элемента x, полученного по распределению D, $P\{f(x_1, x_2, \ldots x_{i-1}) = x_i\} \geq \frac{1}{2} + \varepsilon$.

Лемма

Если распределение D на $\{0,1\}^m$ не является ε-близким к равномерному, то для D существует $\frac{\varepsilon}{2m}$-предсказатель.

Доказательство

Гибридный метод.

Trevisan экстрактор: собираем все вместе

Экстрактор

Описываем экстрактор $E : U \times A \to V$, где
$U = \{0,1\}^{r^k}, V = \{0,1\}^r, A = \{0,1\}^{c \log r}$. Порог минимальной энтропии r^3, $varepsilon \leq \frac{1}{10}$.
$x \in U$, считает $e(x) \in \{0,1\}^{10r^k}$. Будем представлять $e(x)$, как $g : \{0,1\}^{10k \log r} \to \{0,1\}$.
$S_1, S_2, \ldots S_r$ — дизайн с параметрами
$m = r, n = c \log r, s = 10k \log r, l = \log r$.
$y \in A$ — случайная строка. y_{S_i} — сужение на множество S_i.
$E(x,y) = g(y_{S_1})g(y_{S_2}) \ldots g(y_{S_r})$.

Trevisan экстрактор: собираем все вместе

Доказательство

Пусть распределение D на U имеет min-энтропию не менее r^3.
Если распределение $E(D, U_A)$ не являестя $1/10$ близким к равномерному, то есть $\frac{1}{20(r+1)}$-предсказатель.
Умеем по $g(y_{S_1}), g(y_{S_2}), \dots g(y_{S_{i-1}})$ предсказывать $g(y_{S_i})$.
Зафиксируем все биты y, кроме y_{S_i}. Зная $e(x)$, можно представить $g(y_{S_j})$, как функцию от $y_{S_i \cap S_j}$. На запись уйдет не больше, чем r битов. Тогда все $g(y_{S_j})$ можно записать, используя $O(r^2)$ битов.
Кодирование с потерями: представляем $e(x)$, как эти $O(r^2)$ битов. У нас есть предсказатель, значит искажения будут не более, чем $\frac{1}{2} - \frac{1}{20(r+1)}$, но тогда длина кода должна быть как минимум r^3, а у нас $O(r^2)$.