Лекция 14

Сведение полиномиально моделируемого распределения к равномерному

(Конспект: С. Капулкин)

На сегодняшний день для многих **NP**-трудных задач найдены быстрые в среднем алгоритмы. Таким образом можно предположить, что формально трудные задачи могут оказаться легко разрешимы для всех встречающихся на практике случаев. Встает проблема поиска действительно сложных задач, которые были бы трудны в среднем. При этом объектом для изучения становятся классы задач с заданным по входам рассперделением, с учетом которого сложность в среднем и рассчитывается. Естественным будет рассмотреть задачи (в том числе, но не только, из класса **NP**), входы для которых задаются некоторыми конкретными распределениями.

14.1 Сведения в классах задач, трудных в среднем

Для рассматриваемых классов задач можно определить алгоритмы, сводящие друг к другу как языки с фиксированным распределением, так и классы, заданные одним языком, но над различными распределениями.

Определение 14.1. *PSamp* называется полиномиально моделируемое распределение.

В данной лекции доказывается сводимость задач поиска $(\mathbf{NP}, \widetilde{\mathbf{PSamp}})$ к $(\widetilde{\mathbf{NP}}, \mathbf{U})$. Можно продолжить сведение и дальше, получив следующую цепочку:

$$(\mathbf{NP}, \widetilde{\mathbf{PSamp}}) \longrightarrow (\widetilde{\mathbf{NP}}, \mathbf{U}) \longrightarrow (\mathbf{NP}, \mathbf{U}) \longrightarrow (\mathbf{BH}, \mathbf{U}^{\mathbf{BH}}) \longrightarrow (\mathbf{BH}', \mathbf{U})$$

В первых двух переходах в цепочке используется вероятностное сведение, определение которого дано ниже.

14.2 Вероятностное сведение

Определение 14.2. Задача поиска $\widetilde{(L, D)}$ вероятностно сводится к задаче поиска $\widetilde{(L', D')}$, если существует пара алгоритмов $f(x, n) : L \to L \cup \bar{L}$ — вероятностный и $g(x) : L' \to L$ — детететерменированный такие, что выполняются следующие условия:

Обозначение. $\forall x V_x \subseteq suppf(x, n)$

1. $\forall x, x' \in L, x \neq x' : V_x \cap V_{x'} = \emptyset$
2. найдется полином $q_1(n) : \forall x \in suppD_n \ \Pr[f(x, n) \in V_x] \geq 1/q_1(n)$
3. найдется пара полиномов $p(n)$ и $q_2(n) : D_n(x) \leq q_2(n) D'_{p(n)}(V_x)$
4. для всех длин n $\forall x \in L, \forall y \in V_x, y \in L'$ любая подсказка w для y переводится алгоритмом g в подсказку для $x : g(w)$ — подсказка для x.

Вероятностное сведение действует следующим образом. Функция f переводит слова из языка L в слова из языка L'. При этом f может ошибаться, т.е. переводить слова из L в $\bar{L'}$, но делает это не слишком часто. Фунция g переводит подсказки для языка L' в подсказки для языка L.

Лемма 14.1. *Пусть* $\widetilde{(L, D)}$ *сводится к* $\widetilde{(L', D')}$ *и* $\widetilde{(L', D')} \in$ **HeurBPP**, *тогда* $\widetilde{(L, D)} \in$ **HeurBPP**.

Доказательство. Пусть $A'(x, \delta)$ — алгоритм для $\widetilde{(L', D')}$. Построим алгоритм $A(x, \delta)$ для $\widetilde{(L, D)}$:
Независимо $N = 100 q_1(n)$ раз выберем $y_1 = f(x, n), y_2 = f(x, n) \ldots y_N = f(x, n)$
Для каждого y_i запустим $A'(y_i, \dfrac{\delta}{2q_2(n)}) = w_i$. Если $\exists i : g(w_i)$ — подсказка для x, то возвращаем $g(w_i)$, иначе 0.

Осталось доказать, что A удовлетворяет определению **HeurBPP**.

$$\measuredangle F = \{y : A'(y, p(n), \frac{\delta}{2q_2(n)} \text{ выдает не подсказки для } y \text{ с вер.} \geq 1/4\}$$

$$D'_{p(n)}(F) \leq \frac{\delta}{2q_2(n)}$$

$$\measuredangle B = \{x \in suppD_n : D'_{p(n)}(V_x \cap F) \geq D'_{p(n)}(V_x)/2\}$$

$$D_n(B) \leq q_2(n) \sum_{x \in B} D_{p(n)}(V_x) \leq 2q_2 \sum_{x \in B} D'_{p(n)}(V_x \cap F) \leq 2q_2(n) D'_{p(n)}(F) \leq \delta$$

Для $x \notin B$ в виду многократной (N раз) выборки с константной вероятностью подсказка будет выдаваться достаточно часто (с вер. $\geq 3/4$). $\square$

14.3 Сведение PSamp к равномерному

Обозначение. $S_n = \{0,1\}^n$ — поле из 2^n элементов, сложение покоординатно. $S_k = \{0,1\}^k$ — подмножество S_n, его первые k битов.

Лемма 14.2. *Пусть* $X \subseteq \{0,1\}^n, |X| \geq 2^{n+1-k}$. *Тогда число пар* $a,b \in \{0,1\}^n : \exists r \in \{0,1\}^{k-1} : ar+b \in X$ *больше либо равно* $2^{2n}/2$.

Доказательство. $\measuredangle r \in \{0,1\}^{k-1}$ и $p \in X$ — всего $\geq 2^n$ различных пар (r,p). Для каждой пары (r,p) существует 2^n пар (a,b), удовлетворяющих условию $ar+b=p$ — действительно, для каждого a из $\{0,1\}^n$ однозначно определяется $b = p - ar$. При этом для двух различных пар $(r_1,p_1),(r_2,p_2)$ может существовать не более одной подходящей пары (a,b):

$$\begin{cases} ar_1 + b = p_1 \\ ar_2 + b = p_2 \end{cases} \Longrightarrow a(r_1 - r_2) = p_1 - p_2 \Longrightarrow \text{однозначно определяется } a \text{ и } b$$

Тогда по формулле включений-исключений существует не менее $2^n \times 2^n - 2^n \times (2^n-1)/2 > 2^{2n}/2$ пар (a,b), удовлетворяющих условию. □

Теорема 14.1. *Для любого* $(L, \widetilde{PSamp}) \in (\mathbf{NP}, \widetilde{\mathbf{PSamp}})$ *найдется язык* $L' \in \mathbf{NP}$: $(L, \widetilde{PSamp})$ *сводится к* $(\widetilde{L', U})$

Доказательство. Выполним промежуточное сведение. Пусть D — распределение для языка L из $PSamp$ и S — семплер, пол. выч. функция, задающая распределение D. D_n — распределение на строках длины $\leq p(n)$. Заменим его на распределение $\hat{D}_n$ — распределение на строках длины $p(n)+1$, добавив необходиммый *padd*-инг из нулей, дополнительный бит будем использовать для запятой.

Следующий шаг: $\hat{D}_n \longrightarrow \hat{\hat{D}}_{p(n)}$ — распределение $\hat{D}_n$, использующее для генерации n бит, заменим на распределение, использующее $p(n)$ бит. При этом $\hat{\hat{D}}_q : q$ не равно $p(n)$ ни для какого n, положим равномерным.

Выполнив описанное сведение будем считать, что D_n — распределение на строчках длины $n : \{0,1\}^n \to \{0,1\}^n$, где n - число бит, используемое сэмплером S. $D_n = S(U_n)$.

Построим язык L' из **NP** и докажем сводимость L к L'.

$L = \{x :$ для x существует подсказка $w\}$

$L' = \{h_k(x), h, a, b, k | x \in L, \exists r : x = S(ar+b)\}$, где

- h — хеш-функция
- $h_k : \{0,1\}^n \to \{0,1\}^k$, где $1 \leq k \leq n$ - строчка из $\log n$ битов.
- $a,b \in \{0,1\}^n$

$L' \in \mathbf{NP}$, т.к. сертификатом для $\{h_k(x), h, a, b, k\}$ является x, подсказка w и r.

Сведение:

$f(x,n) = (h_k(x), h, a, b, k)$, где $h \leftarrow H$, $a,b \leftarrow U_n$, $k \leftarrow U_{\log n}$. Проверим выполнимость условий сведения.

1. для выполнения условий леммы 14.2 подберем $k : 2^{n+2-k} \geq S^{-1}(x) \geq 2^{n+1-k}$. Такое k угадаем с вероятностью $\geq 1/n$
2. $\exists r : S(ar+b) = x$ — такое r по лемме найдется не меньше, чем для половины a, b
3. Если $S(ar+b) \neq x$, то $h_k(S(ar+b)) \neq h_k(x)$ с вероятностью $\geq 1 - 1/2^k$

Условия 1)–3) выполняются с вероятностью $\geq 1/poly(n)$. Так же из условий 1)–2) следует существование полинома $q_1(n)$ из определения сводимости.

Заметим, что $V_x \cap V_y = \emptyset$ для $x \neq y$, т.к.

$$\begin{array}{ll} x \rightsquigarrow (h_k(x), h, a, b, k) & x = S(ar+b) \\ y \rightsquigarrow (h_k(y), h, a, b, k) & y = S(ar'+b) \end{array} \Rightarrow^{h_k(x)=h_k(y)} \text{получаем противоречие с 3)}$$

Найдем $q_2(n) : D_n(x) \leq q_2(n) U_{p(n)}(V_x)$. Существование полинома следует из следующих утверждений:

$$\begin{array}{l} U_{p(n)}(V_x) \geq 2^{-k} \dfrac{1}{q(n)} \\ D_n(x) \leq 2^{-k+1} \end{array}$$

□