Лекция 13

Сведения задач поиска к задачам распознавания

(Конспект: Ф. Парт)

13.1 Вероятностные алгоритмы и задачи поиска

Дадим определения классов распределенных задач распознавания, решающихся *вероятностными* алгоритмами за полиномиальное в среднем время.

Определение 13.1. Распределенная задача (L, D) содержится в классе $AvgBPP$, если существует вероятностный алгоритм $A(x, n, \delta)$:

1. Время работы A ограниченно $poly(n, 1/\delta)$
2. $\forall x \in supp\, D_n$ верно, что $\Pr_r[A(x, n, \delta) \notin \{L(x), \perp\}] \leq \frac{1}{4}$
3. $\Pr_{x \leftarrow D_n}[\Pr_r[A(x, n, \delta) = \perp] > \frac{1}{4}] < \delta$

Определение 13.2. Распределенная задача (L, D) содержится в классе $HeurBPP$, если существует вероятностный алгоритм $A(x, n, \delta)$:

1. Время работы A ограниченно $poly(n, 1/\delta)$
2. $\Pr_{x \leftarrow D_n}[\Pr_r[A(x, n, \delta) \neq L(x)] > \frac{1}{4}] < \delta$

Замечание 13.1. Константы в этих определениях могут быть заменены на $2^{-\Omega(n)}$

До сих пор мы занимались распределенными задачами распознавания. Определим теперь понятие распределенной задачи поиска. Пусть $\widetilde{L}$ задача поиска решения y для условия x из $L = \{x|\exists y\, Q_L(x,y) = 1\}$, такого что $Q_L(x,y) = 1$. Распределенной задачей поиска $\widetilde{(L,D)}$ будем называть пару из $\widetilde{L}$ и некоторого распределения на входах D.

Неформально, если найдется алгоритм A для (L,D) как в определении $AvgP(HeurP)$ или $AvgBPP(HeurBPP)$, но еще и предоставляющий решение (если $x \notin L$ будем считать, что любое решение правильное), то можно сказать, что задача поиска $\widetilde{(L,D)}$ решается за полиномиальное в среднем время и принадлежит $\widetilde{AvgP}(\widetilde{HeurP})$ или $\widetilde{AvgBPP}(\widetilde{HeurBPP})$ соответственно. Дадим формальные определения.

Определение 13.3. Распределенная задача поиска $\widetilde{(L,D)}$ содержится в классе $\widetilde{AvgP}$, если существует алгоритм $A(x,n,\delta)$:

1. Время работы A ограниченно $poly(n, 1/\delta)$
2. $\forall x \in suppD_n \cap L$ верно, что или $A(x,n,\delta) = \perp$ или $Q_L(x, A(x,n,\delta)) = 1$
3. $\Pr_{x \leftarrow D_n}[A(x,n,\delta) = \perp] < \delta$

Определение 13.4. Распределенная задача поиска $\widetilde{(L,D)}$ содержится в классе $\widetilde{HeurP}$, если существует алгоритм $A(x,n,\delta)$:

1. Время работы A ограниченно $poly(n, 1/\delta)$
2. $\Pr_{x \leftarrow D_n \cap L}[Q_L(x, A(x,n,\delta)) = 0] < \delta$

Определение 13.5. Распределенная задача поиска $\widetilde{(L,D)}$ содержится в классе $\widetilde{AvgBPP}$, если существует вероятностный алгоритм $A(x,n,\delta)$:

1. Время работы A ограниченно $poly(n, 1/\delta)$
2. $\forall x \in supp\ D_n \cap L$ верно, что $\Pr_r[A(x,n,\delta) \neq \perp \wedge Q_L(x, A(x,n,\delta) = 0)] \leq \frac{1}{4}$
3. $\Pr_{x \leftarrow D_n}[\Pr_r[A(x,n,\delta) = \perp] > \frac{1}{4}] < \delta$

Определение 13.6. Распределенная задача поиска $\widetilde{(L,D)}$ содержится в классе $\widetilde{HeurBPP}$, если существует вероятностный алгоритм $A(x,n,\delta)$:

1. Время работы A ограниченно $poly(n, 1/\delta)$
2. $\Pr_{x \leftarrow D_n \cap L}[\Pr_r[Q_L(x, A(x,n,\delta)) = 0] > \frac{1}{4}] < \delta$

13.2 Сведение задач поиска к задачам распознавания

Теорема 13.1. *Если* $(NP,U) \subset AvgBPP(HeurBPP)$, *то* $(\widetilde{NP,U}) \subset \widetilde{AvgBPP}(\widetilde{HeurBPP})$

Доказательство. Будем доказывать утверждение для $AvgBPP$ и $\widetilde{AvgBPP}$ (для $HeurBPP$ и $\widetilde{HeurBPP}$ аналогично).

Пусть $L \in NP$. Построим алгоритм $A(x,n,\delta)$ вычисляющий подсказки для слов из L, используя наличие алгоритмов распознавания для (NP,U). Прежде всего заметим, что если бы у каждого слова из L подсказка была бы единственной, то ее можно было бы вычислить ответив на серию вопросов типа "да-нет": "Верно ли, что i-й бит подсказки равен 1?". Это равносильно решению задачи распознавания языка $L' \in NP$:

$$L' = \{(x,i) | x \in L, 1 \le i \le q(n), \exists y\, Q_L(x,y) = 1 \wedge y_i = 1\}$$

Так как $(L',U) \in (NP,U) \subset AvgBPP$, то существует полиномиальный в среднем вероятностный алгоритм $B((x,i),n,\delta')$, решающий (L',U) с константами $\le \frac{1}{4q(n)}$. Алгоритму $A(x,n,\delta)$ достаточно запустить последовательно $B((x,1),n,\frac{\delta}{q(n)})$, $B((x,2),n,\frac{\delta}{q(n)})$, $\ldots$,$B((x,q(n)),n,\frac{\delta}{q(n)})$ и выдать $\perp$ если в результате хотя бы одного вызова B было $\perp$ или конкатенацию результатов запусков B в противном случае. Очевидно, выполнены условия:

1. Алгоритм A работает полиномиальное от n и $\frac{1}{\delta}$ время
2. $\forall x \in supp\, U_n \cap L$ верно, что

$$Pr_r[A(x,n,\delta) \neq \perp \wedge Q_L(x, A(x,n,\delta) = 0)] =$$

$$= \Pr_r\left[\left(\wedge_{i=1}^{q(n)} B\big((x,i),n,\frac{\delta}{q(n)}\big) \neq \perp\right) \wedge \left(\vee_{i=1}^{q(n)} B\big((x,i),n,\frac{\delta}{q(n)}\big) \neq L'\big((x,i)\big)\right)\right] \le$$

$$\le \Pr_r\left[\vee_{i=1}^{q(n)} \left(B\big((x,i),n,\frac{\delta}{q(n)}\big) \neq L'\big((x,i)\big) \wedge B\big((x,i),n,\frac{\delta}{q(n)}\big) \neq \perp\right)\right] \le$$

$$\le \sum_{i=1}^{q(n)} \Pr_r[B\big((x,i),n,\frac{\delta}{q(n)}\big) \neq L'\big((x,i)\big) \wedge B\big((x,i),n,\frac{\delta}{q(n)}\big) \neq \perp] \le q(n)\frac{1}{4q(n)} = \frac{1}{4}$$

3. Вероятность ответа "не знаю":

$$Pr_{x \leftarrow U_n}\left[\Pr_r[A(x,n,\delta) = \perp] > \frac{1}{4}\right] \le \Pr_{x \leftarrow U_n}\left[\exists i \Pr_r[B\big((x,i),n,\frac{\delta}{q(n)}\big) = \perp] > \frac{1}{4q(n)}\right] \le$$

$$\leq \sum_{i=1}^{q(n)} \Pr_{x \leftarrow U_n} \Big[\Pr_r[B\big((x,i), n, \frac{\delta}{q(n)}\big) = \perp] > \frac{1}{4q(n)} \Big] \leq$$

$$\leq \sum_{i=1}^{q(n)} \Pr_{x \leftarrow U_n} \Big[\Pr_r[B\big((x,i), n, \frac{\delta}{q(n)}\big) = \perp] > \frac{1}{4q(n)} \Big] \cdot \Pr_{(x,j) \leftarrow U_{n+log(q(n))}} \Big[j = i \Big] \cdot q(n) =$$

$$= q(n) \cdot \Pr_{(x,i) \leftarrow U_{n+log(q(n))}} \Big[\Pr_r[B\big((x,i), n, \frac{\delta}{q(n)}\big) = \perp] > \frac{1}{4q(n)} \Big] < \delta$$

Увы, в таком виде алгоритм A корректен лишь в том случае, если подсказки единственны. Единственности подсказки можно добиться посредством усиления предиката $Q_L(x,y)$ заменив его на конъюнкцию $Q'_L(x,y) = Q_L(x,y) \wedge P(y)$ с правильно подобранным предикатом $P(y)$. В выборе $P(y)$ нам поможет лемма Вэлианта-Вазирани. Она утверждает, что если $S \subset \{0,1\}^{q(n)}$ - множество подсказок, $2^{k-2} \leq |S| \leq 2^{k-1}$ и $H_{q(n),k}$ - семейство попарно независимых хэш функций, то $\Pr_{h \leftarrow U_{H_{n,k}}}[\exists! y \in S : h(y) = 0^k] \geq \frac{1}{8}$. Таким образом, если взять случайную хэш функцию h из $H_{q(n),k}$, то с константной вероятностью предикат $P(y) = (h(y) = 0^k)$ будет выполнен ровно одной подсказкой и $Q'_L(x,y)$ будет одновыполним.

Принимая во внимание вышесказанное, вместо языка L' будем иметь дело с языком $L'' \in NP$:

$$L'' = \{(x,i,k,h) | x \in L, 1 \leq i \leq q(n), 1 \leq k \leq q(n), \exists y \left(Q_L(x,y) \wedge \right.$$

$$\left. \wedge (h(y) = 0^k)\right) = 1 \wedge y_i = 1\}$$

Пусть вероятностный алгоритм C решает (L'', U) за полиномиальное в среднем время с константами $\leq \frac{7}{32q(n)}$. Выберем случайно h и для каждого k как и раньше попытаемся получить подсказку в виде $y = C((x,1,k,h), n, \frac{\delta}{q^2(n)}) C((x,2,k,h), n, \frac{\delta}{q^2(n)}) \ldots C((x,q(n),k,h), n, \frac{\delta}{q^2(n)})$. Нетрудно убедится, что полученный алгоритм удовлетворяет условиям полиномиальности в среднем и решает задачу $\widetilde{(L,U)}$, следовательно $\widetilde{(NP,U)} \subset \widetilde{AvgBPP}$.

□

13.3 Лемма Вэлианта-Вазирани

Определение 13.7. Универсальным семейством попарно независимых хэш функций $H_{n,k}$ называется множество функций $\{0,1\}^n \to \{0,1\}^k$, $k \leq n$ такое, что:

1. $\forall x \in \{0,1\}^n$, $\forall a \in \{0,1\}^k$ верно, что $\Pr_{h \leftarrow H_{n,k}}[h(x) = a] = 2^{-k}$
2. $\forall x' \neq x'' \in \{0,1\}^n$, $\forall a', a'' \in \{0,1\}^k$ верно, что $\Pr_{h \leftarrow H_{n,k}}[h(x') = a' \wedge h(x'') = a''] = 2^{-2k}$

Пример 13.1. Множество линейных функций $h_{a,b} = (ax + b)_{\leq k}$ (операции в поле F_{2^n}) с усеченными образами представляет собой семейство попарно независимых хэш функций.

Лемма 13.1 (Вэлианта-Вазирани). *Пусть* $S \subset \{0,1\}^n$, $2^{k-2} \leq |S| \leq 2^{k-1}$ *и* $H_{n,k}$ *- семейство попарно независимых хэш функций. Тогда* $\Pr_{h \leftarrow H_{n,k}}[\exists! x \in S : h(x) = 0^k] \geq \frac{1}{8}$

Доказательство. Обозначим $p = 2^{-k}$, $\frac{1}{4} \leq p|S| \leq \frac{1}{2}$. Пусть N - количество $x \in S$ таких, что $h(x) = 0^k$. Оценим вероятности $\Pr[N \geq 1]$ и $\Pr[N \geq 2]$:

- По формуле включений-исключений

$$Pr[N \geq 1] \geq \sum_{x \in S} \Pr[h(x) = 0^k] - \sum_{x \neq y \in S} \Pr[h(x) = 0^k \wedge h(y) = 0^k] = p|S| - p^2 C^2_{|S|}$$

- $P[N \geq 2] \leq \sum_{x \neq y \in S} \Pr[h(x) = 0^k \wedge h(y) = 0^k] = p|S| - p^2 C^2_{|S|}$

Искомая вероятность равна $\Pr[N = 1] = \Pr[N \geq 2] - \Pr[N \geq 1] \geq p|S| - 2p^2 C^2_{|S|} \geq p|S| - p^2|S|^2 \geq \frac{1}{8}$ □