Лекция 8.

Логарифмическая нижняя оценка
сложности информационных графов
для задач поиска с коротким ответом
и равновероятными тенями записей.
Вид оптимального информационного графа
для таких задач.

1 Нижняя оценка сложности задач поиска с коротким ответом в случае равновероятных теней записей

Результат теоремы о существовании оптимального древовидного информационного графа для задач поиска с коротким ответом, доказанной в предыдущей лекции, помимо того, что представляет самостоятельный интерес, позволяет получить нижнюю оценку сложности информационных графов для задач поиска с равномощными слабо пересекающимися тенями записей, причем нижнюю оценку значительно лучшую, чем мощностная нижняя оценка.

Дадим строгое определение исследуемого в данном разделе класса задач поиска.

Скажем, что ЗИП $I = \langle X, V, \rho \rangle$ *обладает F-свойством*, если она обладает A-свойством и для любых $y, y' \in V$ справедливо $\mathbf{P}(O(y,\rho)) = \mathbf{P}(O(y',\rho))$.

Обозначим

$$R(k) = 3k[\log_3 k] + 4(k - 3^{[\log_3 k]}) + \max(0, k - 2 \cdot 3^{[\log_3 k]}),$$

где $k \geq 1$.

Справедлива следующая теорема.

Теорема 1. *Если* $I = \langle X, V, \rho \rangle$ *— ЗИП, обладающая* F*-свойством,* $\mathcal{F} = \langle F, \emptyset \rangle$ *— измеримое базовое множество, допустимое для* I*, то*

$$T(I, \mathcal{F}) \geq \mathbf{P}(O(y, \rho)) \cdot R(|V|),$$

где $y \in V$*.*

Прежде чем доказывать теорему, докажем ряд вспомогательных утверждений.

Лемма 1. *Пусть* $0 \leq t \leq t_1 \leq t_2$*,* $t_1 - t \geq 1$ *тогда*

$$R(t_1) + R(t_2) \leq R(t_1 - t) + R(t_2 + t).$$

Доказательство. Нетрудно убедиться, что $R(k)$ есть непрерывная выпуклая функция, представляющая собой ломаную линию, причем точки 3^l и $2 \cdot 3^l$ ($l = 0, 1, 2, \ldots$) являются точками излома. А из свойств выпуклых функций легко выводится утверждение леммы.

Обозначим через $\mathbf{N}$ — множество натуральных чисел.

Лемма 2. *Пусть* $k \in \mathbf{N}$*,* $k \geq 2$*,* $l = [\log_3 k]$*, тогда*

$$\begin{aligned} r_1(k) &\stackrel{\text{def}}{=} \min\{R(t_1) + R(t_2) : t_1 + t_2 = k; t_1, t_2 \in \mathbf{N}\} = \\ &= \begin{cases} R(k) + 4 \cdot 3^{l-1} - 3k, & \text{если } 3^l \leq k \leq 4 \cdot 3^{l-1} \\ R(k) - 2k, & \text{если } 4 \cdot 3^{l-1} \leq k \leq 2 \cdot 3^l \\ R(k) - k - 2 \cdot 3^l, & \text{если } 2 \cdot 3^l \leq k \leq 3 \cdot 3^l. \end{cases} \end{aligned}$$

Доказательство. Из леммы 1 следует, что

$$r_1(k) = \begin{cases} 2 \cdot R(k/2), & \text{если } k \text{ — четно} \\ R((k-1)/2) + R((k+1)/2), & \text{если } k \text{ — нечетно.} \end{cases}$$

Пусть $l = [log_3 k]$, $c = k - 3^{[\log_3 k]}$, то есть $k = 3^l + c$, где $0 \leq c \leq 2 \cdot 3^l$, тогда по определению функции $R(k)$

$$R(k) = \begin{cases} 3 \cdot l \cdot k + 4 \cdot c, & \text{если } 0 \leq c \leq 3^l \\ 3 \cdot l \cdot k + 5 \cdot c - 3^l, & \text{если } 3^l \leq c \leq 2 \cdot 3^l. \end{cases}$$

Докажем утверждение леммы при четном k.

1) Пусть $0 \leq c \leq 3^{l-1}$, тогда

$$\begin{aligned} r_1(k) &= 2\cdot R(k/2) = 2\cdot R(3^{l-1} + (3^{l-1}+c)/2) = \\ &= 2\cdot(3\cdot(l-1)\cdot k/2 + 2\cdot(3^{l-1}+c)) = \\ &= 3\cdot l\cdot k + 4\cdot c - 3\cdot k + 4\cdot 3^{l-1} = R(k) + 4\cdot 3^{l-1} - 3\cdot k, \end{aligned}$$

так как в этом случае $(3^{l-1}+c)/2 \le 3^{l-1}$.

2) Пусть $3^{l-1} \le c \le 3^l$, тогда $3^{l-1} \le (3^{l-1}+c)/2 \le 2\cdot 3^{l-1}$ и

$$\begin{aligned} r_1(k) &= 2\cdot R(k/2) = \\ &= 2\cdot(3\cdot(l-1)\cdot k/2 + 5\cdot(3^{l-1}+c)/2 - 3^{l-1}) = \\ &= 3\cdot l\cdot k + 5\cdot c - 3\cdot k + 3^l = R(k) - 2\cdot k. \end{aligned}$$

3) Пусть $3^l \le c \le 2\cdot 3^l$, тогда $k/2 = 3^l + (c-3^l)/2,\ (c-3^l)/2 \le 3^l$ и

$$\begin{aligned} r_1(k) &= 2\cdot R(k/2) = 2\cdot(3\cdot l\cdot k/2 + 2\cdot(c-3^l)) = \\ &= 3\cdot l\cdot k + 4\cdot c - 4\cdot 3^l = R(k) - k - 2\cdot 3^l. \end{aligned}$$

Для нечетного k лемма доказывается аналогично.

Тем самым лемма доказана.

Следствие 1. *$r_1(k) \ge R(k) - 2\cdot k$, причем равенство достигается только при $4\cdot 3^{[\log_3 k]-1} \le k \le 2\cdot 3^{[\log_3 k]}$.*

Лемма 3. *Пусть $k \in \mathbf{N}$, $k \ge 3$, тогда*

$$\begin{aligned} r_2(k) &\stackrel{\text{def}}{=} \min\{R(t_1) + R(t_2) + R(t_3) : \sum_{i=1}^{3} t_i = k; t_1, t_2, t_3 \in \mathbf{N}\} = \\ &= R(k) - 3\cdot k. \end{aligned}$$

Доказательство. Из леммы 1 следует, что

$$r_2(k) = \begin{cases} 3\cdot R(k/3), & \text{если } k = 3s, \\ 2\cdot R((k-1)/3) + R((k+2)/3), & \text{если } k = 3s+1, \\ R((k-2)/3) + 2\cdot R((k+1)/3), & \text{если } k = 3s+2. \end{cases}$$

Докажем утверждение леммы при $k = 3s$.

Пусть $l = [\log_3 k]$, $c = k - 3^{[\log_3 k]}$, то есть $k = 3^l + c$, $k/3 = 3^{l-1} + c/3$.

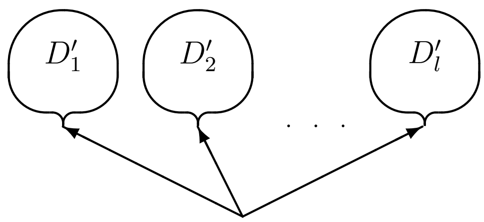

Рис. 1: Ветвь с большой полустепенью исхода корня

1) Пусть $0 \le c \le 3^l$, тогда

$$\begin{aligned} r_2(k) &= 3 \cdot R(k/3) = 3 \cdot (3 \cdot (l-1) \cdot k/3 + 4 \cdot c/3) = \\ &= 3 \cdot l \cdot k + 4 \cdot c - 3 \cdot k = R(k) - 3 \cdot k. \end{aligned}$$

2) Пусть $3^l \le c \le 2 \cdot 3^l$, тогда

$$r_2(k) = 3 \cdot R(k/3) = 3 \cdot (3 \cdot (l-1) \cdot k/3 + 5 \cdot c/3 - 3^{l-1}) = R(k) - 3 \cdot k.$$

Для случаев $k = 3s + 1$ и $k = 3s + 2$ лемма доказывается аналогично. Тем самым лемма доказана.

Рассмотрим функцию $M(D) = \sum_{\beta \in \mathcal{R}(D)} |\beta| \cdot \psi_\beta$, где $D \in \mathcal{D}_0$, $\mathcal{D}_0$ — множество связных ориентированных деревьев с корнем, в которых каждое ребро принадлежит хотя бы одной цепи, ведущей из корня к какой-либо концевой вершине, $|\beta|$ — количество концевых вершин, схемно достижимых из β.

Пусть $N(k) = \min\{M(D) : D \in \mathcal{D}_0 \text{ и } |D| = k\}$, где $|D|$ — количество концевых вершин в дереве D.

Дерево $D \in \mathcal{D}_0$ назовем *простым*, если $M(D) = N(|D|)$.

Лемма 4. *Для любого натурального k существует простое дерево с k концевыми вершинами, полустепень исхода любой внутренней вершины которого равна либо 2, либо 3.*

Доказательство. Предположим, что дерево D содержит ветвь D_1 вида изображенного на рисунке 1, где $l \ge 4$.

Рассмотрим дерево D_2 вида, указанного на рисунке 2, где $r = [l/2]$.

Покажем, что $M(D_2) \le M(D_1)$.

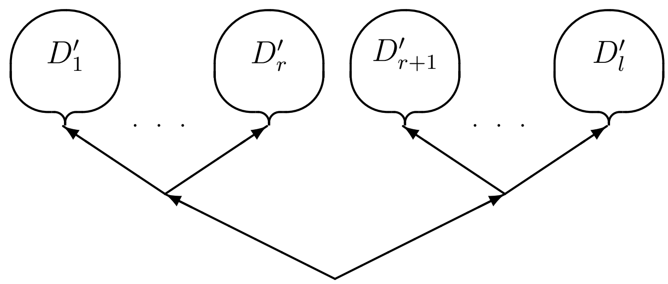

Рис. 2: Уменьшение полустепени исхода ветвей

Пусть $t_i = |D'_i|$ $(i = \overline{1, l})$, $t_i \geq 1$.

Возможны два случая:

1) l — четно, то есть $l = 2 \cdot r$, где $r \geq 2$, тогда

$$\begin{aligned} M(D_2) &= 2 \cdot \sum_{i=1}^{l} t_i + r \cdot \sum_{i=1}^{l} t_i + \sum_{i=1}^{l} M(D'_i) \leq M(D_1) = \\ &= l \cdot \sum_{i=1}^{l} t_i + \sum_{i=1}^{l} M(D'_i), \end{aligned}$$

причем равенство достигается лишь при $r = 2$.

2) l — нечетно, то есть $l = 2 \cdot r + 1$, где $r \geq 2$, тогда

$$M(D_2) = 2 \cdot \sum_{i=1}^{l} t_i + r \cdot \sum_{i=1}^{r} t_i + (r+1) \cdot \sum_{i=r+1}^{l} t_i + \sum_{i=1}^{l} M(D'_i) < M(D_1).$$

Таким образом, мы показали, что если в дереве есть ветвь вида D_1, то ее всегда можно заменить на ветвь вида D_2, и дерево от этого не усложнится.

Осталось заметить, что ветвь вида D_3, изображенного на рисунке 3, всегда можно заменить на ветвь D_4, не усложняя при этом дерева.

Тем самым лемма доказана.

Лемма 5. *Для любого натурального k выполняется $N(k) = R(k)$.*

Доказательство. Доказывать будем индукцией по числу k.

Базис индукции. $k = 1$; $N(1) = 0 = R(1)$.
$k = 2$; $N(2) = 4 = R(2)$. $k = 3$; $N(3) = 9 = R(3)$.

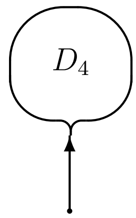

Рис. 3: Ветвь с единичной полустепенью исхода корня

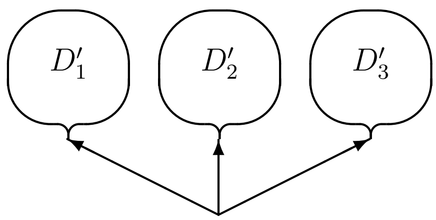

Рис. 4: Дерево D_1

Шаг индукции. Пусть для любого $t < k$ $N(t) = R(t)$.

Согласно лемме 4, существует простое дерево с k концевыми вершинами, которое имеет либо вид D_1, изображенный на рисунке 4, либо вид D_2, изображенный на рисунке 5.

Согласно предположению индукции, лемме 3 и следствию 1,

$$\begin{aligned} M(D_1) &= 3 \cdot k + M(D_1') + M(D_2') + M(D_3') \geq \\ &\geq 3 \cdot k + R(|D_1'|) + R(|D_2'|) + R(|D_3'|) \geq \\ &\geq 3 \cdot k + \min\{R(t_1) + R(t_2) + R(t_3) : t_1 + t_2 + t_3 = k, \\ &\quad t_1, t_2, t_3 \in \mathbf{N}\} = 3 \cdot k + R(k) - 3 \cdot k = R(k), \end{aligned}$$

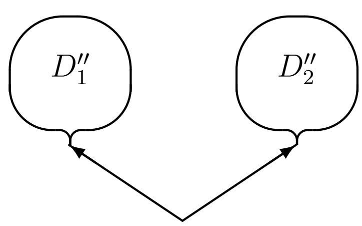

Рис. 5: Дерево D_2

$$\begin{aligned}
M(D_2) &= 2\cdot k + M(D_1'') + M(D_2'') \geq \\
&\geq 2\cdot k + R(|D_1''|) + R(|D_2''|) \geq \\
&\geq 2k + \min\{R(t_1) + R(t_2) : t_1 + t_2 = k, t_1, t_2 \in \mathbf{N}\} \geq \\
&\geq R(k).
\end{aligned}$$

С другой стороны, если D_1', D_2', D_3' — простые и такие, что их мощности (то есть количество концевых вершин) различаются не более чем на 1, то $M(D_1) = R(k)$ и D_1 — простое, а если $4 \cdot 3^{[\log_3 k]-1} \leq k \leq 2 \cdot 3^{[\log_3 k]}$ и D_1'' и D_2'' — простые и такие, что $||D_1''| - |D_2''|| \leq 1$, то $M(D_2) = R(k)$ и D_2 — простое дерево.

Тем самым лемма доказана.

Перейдем к *доказательству теоремы* 1.

Возьмем произвольный ИГ $U \in \mathcal{U}(I, \mathcal{F})$. Согласно теореме о существовании оптимального древовидного информационного графа для задач поиска с коротким ответом существует такое информационное дерево $D \in \mathcal{D}^I$, что $T(U) \geq T(D)$. Оценим сложность ИД D.

$$\begin{aligned}
T(D) &= \sum_{\beta\in\mathcal{R}(D)} \psi_\beta \cdot \mathbf{P}(N_{\varphi_\beta}) \geq \sum_{\beta\in\mathcal{R}(D)} \psi_\beta \cdot \mathbf{P}(N_{\vee_{y\in V_\beta}\chi_{y,\rho}}) = \\
&= \sum_{\beta\in\mathcal{R}(D)} \psi_\beta \cdot \mathbf{P}(\bigcup_{y\in V_\beta} O(y,\rho)) = \\
&= \sum_{\beta\in\mathcal{R}(D)} \psi_\beta \cdot |V_\beta| \cdot \mathbf{P}(O(y',\rho)) = \\
&= \mathbf{P}(O(y',\rho)) \cdot \sum_{\beta\in\mathcal{R}(D)} \psi_\beta \cdot |V_\beta| = \\
&= \mathbf{P}(O(y',\rho)) \cdot M(D) \geq \mathbf{P}(O(y',\rho)) \cdot N(|V|) = \\
&= \mathbf{P}(O(y',\rho)) \cdot R(|V|),
\end{aligned}$$

где y' некоторая запись из V.

В силу произвольности U теорема доказана.

Следствие 2. *В условиях теоремы 1*

$$T(I, \mathcal{F}) \geq c \cdot [\log_3 |V|],$$

где $c \leq 3$.

Справедливость утверждения следует из того, что

$$k \cdot \mathbf{P}(O(y, \rho)) \leq 1.$$

Для сравнения отметим, что мощностная нижняя оценка сложности для задач, обладающих F-свойством, равна константе, которая не превышает 1.

Результат следствия 2 может показаться аналогичным теоретико-информационной нижней оценке, но он имеет принципиальные отличия: во первых, он касается средней сложности, а не сложности в худшем случае, во-вторых, доказывался для произвольных сетей, а не бинарных деревьев.

Следствие 3. *Если $I = \langle X, V, \rho \rangle$ — ЗИП, обладающая F-свойством, $\mathcal{F} = \langle F, \emptyset \rangle$ — измеримое базовое множество, допустимое для I, такое, что $F_0^I \subseteq F$, то*

$$\mathbf{P}(O(y, \rho)) \cdot R(k) \leq T(I, \mathcal{F}) \leq \mathbf{P}(O(y, \rho)) \cdot R(k) + 1,$$

где $k = |V|, \ y \in V$.

Доказательство. Нижняя оценка следует из теоремы 1, а верхнюю дает ИД, изображенный на рисунке 3, где D_4 — простое дерево с k концевыми вершинами, которые одновременно являются листьями графа, нагрузка листьев записями из V выбрана произвольно, а нагрузка ребер осуществляется по методу, описанному в доказательстве теоремы о существовании оптимального древовидного информационного графа для задач поиска с коротким ответом, так, чтобы выполнялось C-свойство.

Тем самым следствие доказано.