Нижняя оценка сложности включающего поиска в классе древовидных схем *

Э.Э.Гасанов
Московский государственный университет

Аннотация

В статье в классе древовидных информационных сетей с базовым множеством переменных приведены примеры таких задач включающего поиска, что нижняя оценка их сложности такова, что по сравнению с ней среднее время перечисления ответа есть о-малое.

1 Введение

Известным фактом в теории информационного поиска является так называемая тривиальная нижняя оценка [1], которая говорит, что временная сложность задач поиска не меньше, чем время, необходимое на перечисление ответа. Для задачи включающего поиска, которая состоит в поиске в некотором конечном подмножестве булева куба, тех точек, которые находятся в отношении "не меньше по-компонентно"с точкой-запросом, в работе [2] получена нижняя оценка сложности, в два раза превышающая тривиальную нижнюю оценку. В данной работе найдены такие задачи включающего поиска, для которых в классе древовидных информационных сетей в случае, когда базовое множество есть множество переменных, получена нижняя оценка, по сравнению с которой тривиальная нижняя оценка есть о-малое. Этот результат так же справедлив в классе, так называемых, бесповторных информационных сетей, для которых в [3] доказана древовидность оптимальных сетей.

Автор выражает благодарность В.Б.Кудрявцеву и А.С.Подколзину за поддержку в работе.

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант 95-01-00597)

2 Основные понятия и формулировка результата

Мы будем использовать терминологию и обозначения из работы [4]. Напомним понятие информационной сети с переключателями (ИСП) и другие необходимые понятия из [4].

Пусть X — множество запросов, причем на X определено вероятностное пространство $\langle X, \sigma, \mathbf{P}\rangle$, где σ — алгебра подмножеств множества X, $\mathbf{P}$ — вероятностная мера на σ; Y — множество записей (объектов поиска); ρ — бинарное отношение на $X \times Y$, называемое отношением поиска; тройку $S = \langle X, Y, \rho\rangle$ (или пятерку $S = \langle X, Y, \rho, \sigma, \mathbf{P}\rangle$, если мы хотим подчеркнуть какое именно вероятностное пространство над X используется) будем называть *типом* или иногда более развернуто *типом задач информационного поиска*; тройку $I = \langle X, V, \rho\rangle$, где V — некоторое конечное подмножество множества Y, в дальнейшем называемое *библиотекой*, будем называть задачей информационного поиска (ЗИП) типа $S = \langle X, Y, \rho\rangle$ (или ЗИП, принадлежащей типу S, и обозначать $I \in S$), и будем считать, что задача $I = \langle X, V, \rho\rangle$ состоит в перечислении для произвольно взятого запроса $x \in X$ всех тех и только тех записей из V, которые находятся в отношении ρ с запросом x, т.е. удовлетворяют запросу x; $O(y, \rho) = \{x \in X : x\rho y\}$ — тень записи $y \in Y$; $N_f = \{x \in X : f(x) = 1\}$, где f — одноместный предикат, определенный на X, т.е. $f : X \to \{0, 1\}$; $\chi_{y,\rho} : X \to \{0, 1\}$ такая, что $N_{\chi_{y,\rho}} = O(y, \rho)$ — характеристическая функция записи y; F — множество символов одноместных предикатов, определенных на множестве X, называемое базовым множеством предикатов; G — множество символов одноместных переключателей, определенных на множестве X. Под переключателями будем понимать функции, областью значений которых являются конечные подмножества натурального ряда.

Пару $\mathcal{F} = \langle F, G\rangle$ назовем базовым множеством.

Если n — натуральное число, а $g(x)$ — некий переключатель, то через $\xi_g^n(x)$ обозначим предикат, определенный на X , такой, что

$$N_{\xi_g^n} = \{x \in X : g(x) = n\}.$$

Обозначим

$$\widehat{G} = \{\xi_g^n : g \in G, n \in \mathbf{N}\}.$$

Определение понятия ИСП разбивается на два этапа. На первом этапе раскрывается структурная (схемная) часть этого понятия, на втором – функциональная.

Определение ИСП с точки зрения ее структуры.

Пусть нам дана ориентированная многополюсная сеть.

Выделим в ней один полюс и назовем его *корнем*, а остальные полюса назовем *листьями*.

Выделим в сети некоторые вершины и назовем их *точками переключения* (полюса могут быть точками переключения).

Если β — вершина сети, то через ψ_β обозначим *полустепень исхода* вершины β .

Каждой точке переключения β сопоставим некий символ из G, такой, что максимальное значение переключателя, соответствующего этому символу, не превышает ψ_β. Это соответствие назовем нагрузкой точек переключения.

Для каждой точки переключения β ребрам, из нее исходящим, поставим во взаимоодноозначное соответствие числа из множества $\{\overline{1,\psi_\beta}\}$. Эти ребра назовем переключательными, а это соответствие — нагрузкой переключательных ребер.

Ребра, не являющиеся переключательными, назовем *предикатными.*

Каждому предикатному ребру сети сопоставим некоторый символ из множества F. Это соответствие назовем *нагрузкой предикатных ребер.*

Сопоставим каждому листу сети некоторую запись из множества Y. Это соответствие назовем *нагрузкой листьев.*

Полученную нагруженную сеть назовем *информационной сетью с переключателями* над базовым множеством $\mathcal{F} = \langle F, G \rangle$.

Определение функционирования ИСП.

Пусть нам дана ИСП U .

Последовательность ориентированных ребер сети (α_1, α_2), (α_2, α_3), $\ldots$, (α_{m-1}, α_m) назовем *ориентированной цепью* от вершины α_1 к вершине α_m.

Если c ребро сети, то через $[c]$ обозначим его нагрузку.

Проводимостью ребра (α, β) назовем предикат, равный $[(\alpha, \beta)]$, если ребро предикатное и $\xi_g^{[(\alpha,\beta)]}$, если ребро переключательное, где g — переключатель, соответствующий вершине α.

Проводимостью ориентированной цепи назовем конъюнкцию проводимостей ребер цепи.

Если зафиксировать запрос x, то цепь, проводимость которой на запросе x равна 1, назовем *проводящей цепью* на запросе x.

В ИСП по аналогии с контактными схемами введем для каждой пары вершин α и β *функцию проводимости* $f_{\alpha\beta}$ от вершины α к вершине β следующим образом:

- если $\alpha = \beta$, то $f_{\alpha\beta}(x) \equiv 1$ $(x \in X)$;
- если $\alpha \neq \beta$ и не существует в ИСП ориентированных цепей от α к β, то $f_{\alpha\beta}(x) \equiv 0$;
- если $\alpha \neq \beta$ и множество ориентированных цепей от α к β не пусто, то $f_{\alpha\beta}(x)$ равно дизъюнкции проводимостей всех ориентированных цепей от α к β.

Функцию проводимости от корня ИСП к некоторой вершине β ИСП назовем *функцией фильтра* вершины β и обозначим $\varphi_\beta(x)$.

Через $\mathcal{R}(U), \mathcal{P}(U), \mathcal{L}(U)$ (или просто $\mathcal{R}, \mathcal{P}, \mathcal{L}$) обозначим множества вершин, точек переключения и листьев сети U соответственно.

Пусть $\mathcal{N}$ — некоторая подсеть (т.е. произвольное подмножество вершин и ребер) ИСП U. Через $\langle\mathcal{N}\rangle$ обозначим множество записей, соответствующих листьям этой подсети (в частности, если α — некоторый лист сети U, то под $\langle\alpha\rangle$ будем понимать запись, соответствующую листу α).

Будем говорить, что ИСП U реализует функцию $\mathcal{J} : X \to 2^Y$, называемую функцией ответа сети U и определяемую соотношением :

$$\mathcal{J}(x) = \langle\{\alpha \in \mathcal{L}(U) : \varphi_\alpha(x) = 1\}\rangle.$$

Понятие ИСП полностью определено.

В случае, когда базовое множество переключателей G пусто, т.е. в сетях нет переключателей, то ИСП называются *информационными сетями с дублированием листьев* (ИСД).

В данной статье мы будем в основном исследовать именно ИСД.

Скажем, что ИСП U разрешает ЗИП $I = \langle X, V, \rho\rangle$, если для $\forall x \in X$

$$\mathcal{J}(x) = \{y \in V : x\rho y\}.$$

Сложностью ИСП U на запросе x назовем число

$$T(U, x) = b \cdot \sum_{\beta\in\mathcal{R}\setminus\mathcal{P}} \psi_\beta \cdot \varphi_\beta(x) + a \cdot \sum_{\beta\in\mathcal{P}} \varphi_\beta(x),$$

где константа a характеризует сложность вычисления одного переключателя, а константа b — одного предиката.

Далее всюду будем предполагать, что мы находимся в условиях леммы 1 из [4], т.е. будем считать, что алгебра σ содержит все множества N_f, где $f \in F \cup \widehat{G}$.

При этих условиях $T(U, x)$ — случайная величина по x, и можно ввести понятия сложности сети и сложности ребра сети.

Сложностью ИСП U назовем математическое ожидание величины $T(U, x)$, т.е. число

$$T(U) = \mathbf{M}\, T(U, x).$$

Если (β, α) — ребро ИСП, то *сложностью этого ребра* назовем число

- $b \cdot \mathbf{P}(N_{\varphi_\beta})$ — если (β, α) — предикатное ребро;
- $a \cdot \mathbf{P}(N_{\varphi_\beta})/\psi_\beta$ — если это ребро переключательное.

Легко видеть, что сложность ИСП равна сумме сложностей ребер ИСП, т.е.

$$T(U) = b \cdot \sum_{\beta\in\mathcal{R}\setminus\mathcal{P}} \psi_\beta \mathbf{P}(N_{\varphi_\beta}) + a \cdot \sum_{\beta\in\mathcal{P}} \mathbf{P}(N_{\varphi_\beta}).$$

Далее всюду будем предполагать, что $a = b = 1$.

Пусть нам дана некая ЗИП I. *Сложностью задачи I при базовом множестве $\mathcal{F}$* назовем число

$$T(I, \mathcal{F}) = \inf\{T(U) : U \in \mathcal{U}(I, \mathcal{F})\},$$

где $\mathcal{U}(I, \mathcal{F})$ — множество всех ИСП над базовым множеством $\mathcal{F}$, разрешающих ЗИП I.

Скажем, что базовое множество $\mathcal{F}$ полно для типа S, если для любой ЗИП I типа S $\mathcal{U}(I, \mathcal{F}) \neq \emptyset$.

Пусть M некоторое множество. Через $|M|$ обозначим число элементов во множестве M, называемое мощностью множества M.

Будем писать $\alpha(n) = \bar{o}(1)$, если $\lim\limits_{n\to\infty} \alpha(n) = 0$.

Будем писать $A(n) = \bar{o}(B(n))$, если $A(n) = B(n) \cdot \bar{o}(1)$.

Скажем, что $A(n)$ асимптотически не превосходит $B(n)$ при $n \to \infty$ и обозначим $A \lesssim B$, если существует $\alpha(n) = \bar{o}(1)$ такое, что начиная с некоторого номера n_0, $A(n) \leq (1 + \alpha(n)) \cdot B(n)$.

Если $A \lesssim B$ и $B \lesssim A$, то будем говорить, что A и B асимптотически равны при $n \to \infty$ и обозначать $A \sim B$.

Через $\binom{n}{k}$ будем обозначать число сочетаний из n элементов по k.

Рассмотрим следующий тип задач поиска:

$$S_b = \langle B^n, B^n, \succeq, \sigma, \mathbf{P} \rangle,$$

где B^n — *единичный n-мерный куб*, то есть

$$B^n = \{(\alpha_1, \ldots, \alpha_n) : \ \alpha_i \in \{0, 1\} \ (i = \overline{1, n})\},$$

$\succeq$ — отношение поиска на $B^n \times B^n$, определяемое следующим соотношением

$$(x_1, \ldots, x_n) \succeq (y_1, \ldots, y_n) \Longleftrightarrow x_i \geq y_i, \ i = \overline{1, n},$$

σ — алгебра подмножеств B^n, представляющая собой множество всех подмножеств B^n, $\mathbf{P}$ — равномерная вероятностная мера на σ, то есть такая мера, что для $\forall x \in B^n$ $\mathbf{P}(x) = 1/2^n$ и $\forall A \subseteq B^n$ $\mathbf{P}(A) = |A|/2^n$.

Задачи поиска, принадлежащие данному типу, есть разновидность задач, именуемых в литературе задачами включающего поиска (см., например, [5]), поэтому тип S_b мы назовем *типом включающего поиска*, а задачи, принадлежащие этому типу, — *задачами включающего поиска*.

Напомним, что в соответствии с терминологией [6] *весом набора* $(\alpha_1, \ldots, \alpha_n) \in B^n$ называют число его координат, равных 1. Формула $x_{i_1}^{\sigma_1} \& x_{i_2}^{\sigma_2} \& \ldots \& x_{i_r}^{\sigma_r}$, где & - знак конъюнкции, $\sigma_k \in \{0, 1\}$, $x_{i_k}^0 = \overline{x}_{i_k}$, $x_{i_k}^1 = x_{i_k}$, $i_k \in \{1, 2, \ldots, n\}$, $k = 1, 2, \ldots, r$ ($r \geq 1$ и $n \geq 1$), называется *конъюнкцией над множеством переменных* $X^n = \{x_1, x_2, \ldots, x_n\}$. Если $x_{i_j} \neq x_{i_k}$ при $j \neq k$, то конъюнкция называется *элементарной*. Элементарная конъюнкция называется *монотонной*, если она не содержит отрицаний переменных. Множество элементарных монотонных конъюнкций от n переменных будем обозначать через $\mathcal{K}^n$.

Рассмотрим произвольную запись $y \in B^n$. Пусть $\{i_1, \dots, i_k\}$ есть множество номеров координат вектора y, которые равны 1. Нетрудно заметить, что

$$\chi_{y,\succeq}(x_1, \dots, x_n) = x_{i_1} \& x_{i_2} \& \dots \& x_{i_k},$$

т.е. характеристическая функция записи y является элементарной монотонной конъюнкцией.

Отсюда согласно [4, теорема 1] сеть, разрешающая некоторую задачу включающего поиска, представляет собой сеть, реализующую некоторую систему элементарных монотонных конъюнкций. Откуда согласно [1, теорема 2] следует, что базовое множество $\langle X^n, \emptyset \rangle$ является полным для типа S_b.

Справедлива следующая теорема.

Теорема 1 *Если базовое множество имеет вид $\mathcal{F} = \langle X^n, \emptyset \rangle$ и $k(n)$, $m(n)$ и $\alpha(n)$ такие числа, зависящие от n, что*

$$k(n) \sim \binom{n}{m-1} \alpha(n),\ m(n) \geq 1,\ m(n) = \bar{o}(\log n / \log \log n),$$

$$\alpha(n) \to \infty,\ (\log n - \log \alpha(n))/(m(n) \log(m(n)+1)) \to \infty$$

при $n \to \infty$, то для любого такого $k(n)$ существуют такие ЗИП $I_k = \langle B^n, V_k, \succeq \rangle$ типа S_b, у которых $|V_k| = k$, что для них при $n \to \infty$ справедливо соотношение

$$\inf\{T(U) : U \in \mathcal{U}_d(I, \mathcal{F})\} = \left(\sum_{y \in V} \mathbf{P}(O(y, \succeq)) \right) / \bar{o}(1),$$

где $\mathcal{U}_d(I, \mathcal{F})$ — множество всех ИСД над базовым множеством $\mathcal{F}$, разрешающих ЗИП I и граф которых имеет вид дерева.

3 Доказательство теоремы

Приведем доказательство теоремы 1.

Пусть $p(n)$ — такое, что $p^m \geq k > (p-1)^m$. Оценим величину p/n. Для этого рассмотрим два случая: когда $m = const$, и $m \to \infty$ при $n \to \infty$.

В первом случае получим

$$\begin{aligned} \frac{p}{n} &\sim \frac{k^{1/m}}{n} \sim n^{-1} \left(\frac{\alpha n(n-1) \cdots (n-m+2)}{(m-1)!} \right)^{1/m} \sim \\ &\sim \left(\frac{n^{m-1}\alpha}{n^m (m-1)!} \right)^{1/m} = \left(\frac{\alpha}{n(m-1)!} \right)^{1/m} = \bar{o}(1), \end{aligned} \tag{1}$$

так как, как легко видеть, $\alpha(n) = \bar{o}(n)$.

Во втором случае, используя формулу Стирлинга (см., например, [6, стр. 282]), можем получить аналогичный результат

$$\begin{aligned} \frac{p}{n} \quad &\sim \quad \frac{k^{1/m}}{n} \sim n^{-1}\left(\frac{\alpha}{\sqrt{2\pi(m-1)}}\left(\frac{n\,e}{m-1}\right)^{m-1}\right)^{1/m} = \\ &= \quad \left(\frac{n^{m-1}\alpha}{n^m}\right)^{1/m} \cdot \left(\frac{e}{m-1}\right)^{\frac{m}{m-1}} \cdot (2\pi(m-1))^{-1/2m} = \bar{o}(1). \qquad (2) \end{aligned}$$

Пусть $r(n)$ такое, что

$$pm + p\sum_{i=1}^{r} m^i \le n. \qquad (3)$$

Договоримся обозначать запись $y \in B^n$ через ее характеристическую функцию, то есть представлять ее в виде конъюнкции переменных, номера которых совпадают с номерами разрядов вектора y, в которых стоят 1.

Разобьем все множество переменных X^n на $m + r + 1$ взаимно непересекающиеся части так, что в первых m частях находится по p переменных, в $(m + i)$-й части $(i = \overline{1, r})$ находится pm^i переменных, а в $(m + r + 1)$-й части все оставшиеся переменные. В соответствии с неравенством 3 такое разбиение возможно. Переменные из i-й части X^n $(i = \overline{1, m + r + 1})$ будем обозначать через x^i_q, то есть верхний индекс переменной будет говорить какой части принадлежит переменная.

Рассмотрим библиотеку

$$\begin{aligned} V \quad = \{ \quad & x^1_{q_1} x^2_{q_2} \ldots x^m_{q_m} x^{m+1}_{f_1} \ldots x^{m+r}_{f_r} : q_j \in \{\overline{1, p}\}, j = \overline{1, m}; \\ & f_j(q_1, \ldots, q_m) = \sum_{i=1}^{m} q_i \cdot i^{j-1}, j = \overline{1, r}\}. \end{aligned}$$

Поскольку каждая запись $x^1_{q_1} x^2_{q_2} \ldots x^m_{q_m} x^{m+1}_{f_1} \ldots x^{m+r}_{f_r}$ полностью определяется набором $(q_1, \ldots, q_m)$, то мощность библиотеки V равна $|V| = p^m \ge k$. Отметим также, что для любого $j \in \{\overline{1, r}\}$

$$f_j = \sum_{i=1}^{m} q_i \cdot i^{j-1} \le p \cdot m \cdot m^{j-1} = pm^j.$$

Покажем, что любая пара различных векторов из V имеет не более, чем $m-1$ одинаковую переменную. Возьмем произвольную пару векторов из V:

$$x^1_{q_1} x^2_{q_2} \ldots x^m_{q_m} x^{m+1}_{f_1} \ldots x^{m+r}_{f_r},$$

$$x^1_{q'_1} x^2_{q'_2} \ldots x^m_{q'_m} x^{m+1}_{f'_1} \ldots x^{m+r}_{f'_r}.$$

Пусть $q_{i_l} \neq q'_{i_l}$, $i_l \in \{\overline{1,m}\}$, $l=\overline{1,s}$, и $q_j = q'_j$, если $j \in \{\overline{1,m}\} \setminus \{i_l : l = \overline{1,s}\}$. Так вектора различные, то $s > 0$.

Пусть $f_{j_l} = f'_{j_l}$, $j_l \in \{\overline{1,r}\}$, $l=\overline{1,t}$, и $f_j \neq f'_j$, если $j \in \{\overline{1,r}\} \setminus \{j_l : l = \overline{1,t}\}$. Покажем, что $t < s$.

Предположим, что это не так, то есть $t \geq s$.

Рассмотрим систему уравнений

$$\left\{ \begin{array}{c} f_{j_1} = f'_{j_1}, \\ \ldots\ldots \\ f_{j_t} = f'_{j_t}. \end{array} \right.$$

Взяв в этой системе первые s уравнений, получим

$$\left\{ \begin{array}{c} \sum\limits_{i=1}^{m} q_i \cdot i^{j_1-1} = \sum\limits_{i=1}^{m} q'_i \cdot i^{j_1-1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sum\limits_{i=1}^{m} q_i \cdot i^{j_s-1} = \sum\limits_{i=1}^{m} q'_i \cdot i^{j_s-1}. \end{array} \right.$$

$$\left\{ \begin{array}{c} \sum\limits_{l=1}^{s} (q_{i_l} - q'_{i_l}) \cdot {i_l}^{j_1-1} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sum\limits_{l=1}^{s} (q_{i_l} - q'_{i_l}) \cdot {i_l}^{j_s-1} = 0. \end{array} \right.$$

Определитель этой системы

$$\left| \begin{array}{cccc} {i_1}^{j_1-1} & {i_2}^{j_1-1} & \ldots & {i_s}^{j_1-1} \\ {i_1}^{j_2-1} & {i_2}^{j_2-1} & \ldots & {i_s}^{j_2-1} \\ \vdots & \vdots & \ddots & \vdots \\ {i_1}^{j_s-1} & {i_2}^{j_s-1} & \ldots & {i_s}^{j_s-1} \end{array} \right|$$

есть минор определителя Вандермонда и не равен 0, поскольку $i_1, i_2, \ldots, i_s$ — положительные числа. Это довольно простой факт и встречается в качестве упражнения, например, в [7, стр. 54]. Поэтому эта система имеет единственное решение

$$\left\{ \begin{array}{c} q_{i_1} = q'_{i_1}, \\ \ldots\ldots \\ q_{i_s} = q'_{i_s}. \end{array} \right.$$

Но согласно предположению

$$\left\{ \begin{array}{c} q_{i_1} \neq q'_{i_1}, \\ \ldots\ldots \\ q_{i_s} \neq q'_{i_s}. \end{array} \right.$$

Противоречие. Значит, $t < s$ и, следовательно, любая пара записей из V имеет не более, чем $m-1$ общую переменную.

Рассмотрим такую библиотеку V', что $V' \subset V$ и $|V'| = k$. Пусть $I = \langle B^n, V', \succeq \rangle$ — соответствующая этой библиотеке ЗИП типа S_b. Пусть U' — произвольная древовидная сеть из $\mathcal{U}_d(I, \mathcal{F})$.

Возьмем произвольную запись $y \in V'$. Главной цепью записи y назовем цепь, ведущую из корня в какой-либо лист, которому приписана запись y, такую, что проводимость этой цепи равна характеристической функции записи y. Как было показано в [2], для любой записи y из V' в сети U' существует главная цепь, которую обозначим C_y. Для каждой записи из V' выберем по главной цепи и далее будем говорить только об этих главных цепях. Длина цепи C_y не меньше, чем $m + r$ и каждому ребру этой цепи приписана некоторая переменная из характеристической функции записи y. Пусть $y = x_{i_1} x_{i_2} \ldots x_{i_{m+r}}$, где переменные упорядочены в порядке следования вдоль цепи C_y (если некоторая переменная встречается в цепи дважды и более, то упорядочиваем по первому вхождению). Сопоставим каждой переменной x_{i_j} $(j \in \{\overline{1, m+r}\})$ первое ребро, которому приписана эта переменная. Сложность этого ребра не меньше, чем

$$\mathbf{P}(N_{x_{i_1} \& x_{i_2} \& \ldots \& x_{i_{j-1}}}) = 2^{1-j}.$$

Если ребро цепи C_y, соответствующее переменной x_{i_j}, принадлежит также некоторой другой главной цепи $C_{y'}$, то все ребра, соответствующие переменным $x_{i_1}, \ldots, x_{i_{j-1}}$, также принадлежат цепи $C_{y'}$, так как сеть U' — древовидная. Отсюда, учитывая, что любая пара записей из V' имеет не более, чем $m-1$ общую переменную, получаем, что все ребра, соответствующие переменным $x_{i_m}, \ldots, x_{i_{m+r}}$, принадлежат только одной главной цепи C_y. Таким образом, каждой записи из V' мы можем сопоставить цепочку из $r+1$ ребра, причем цепочки, соответствующие различным записям, не пересекаются. Суммарная сложность каждой такой цепочки не меньше, чем

$$\sum_{j=0}^{r} 2^{1-m-j} = 2^{2-m}(1 - 2^{-r-1}).$$

Следовательно

$$T(U') \geq k \cdot 2^{2-m}(1 - 2^{-r-1}).$$

Возьмем $r(n) = [\log_{m+1}(n/p)]$. При таком выборе r

$$pm + p\sum_{i=1}^{r} m^i \leq p(m+1)^r \leq n,$$

то есть выполняется условие 3.

Оценим величину $r(n)$. Для этого, как и ранее, рассмотрим два случая: когда $m = const$, и $m \to \infty$ при $n \to \infty$.

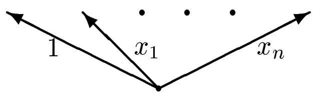

Рис. 1: Универсальный многополюсник для множества $\mathcal{K}_1^n$

В первом случае, используя (1), получим

$$r \geq \frac{1}{\log(m+1)} \cdot \log\frac{n}{p} - 1 \sim \frac{\log n - \log\alpha + \log((m-1)!)}{m\log(m+1)} \to \infty.$$

Во втором случае, используя (2), получим аналогичный результат

$$\begin{aligned} r \quad \geq \quad & \frac{1}{\log(m+1)} \cdot \log\frac{n}{p} - 1 \sim \frac{1}{\log(m+1)}(\log n - \frac{m-1}{m}\log n - \\ & -\frac{m-1}{m}\log e + \frac{m-1}{m}\log(m-1) + \\ & +\frac{1}{2m}\log(2\pi(m-1) - \frac{1}{m}\log\alpha) = \\ = \quad & \frac{\log n - \log\alpha + (m-1)\log((m-1)/l) + \log\sqrt{2\pi(m-1)}}{m\log(m+1)} \to \infty. \end{aligned}$$

Откуда следует, что для любой сети $U' \in \mathcal{U}_d(I,\mathcal{F})$

$$T(U') \gtrsim k \cdot 2^{2-m}.$$

Следовательно,

$$\inf\{T(U) : U \in \mathcal{U}_d(I,\mathcal{F})\} \gtrsim k \cdot 2^{2-m}. \tag{4}$$

Построим некоторую ИСД, которая разрешала бы задачу I по аналогии с тем как это делалось в теореме 3 из [2].

Обозначим через $\mathcal{K}_l^n$ — множество всех монотонных элементарных конъюнкций, содержащих не более l переменных (или другими словами, имеющих длину не более l).

Построим по индукции информационную сеть U_l^n, реализующую как функции фильтров каких-либо вершин все функции из $\mathcal{K}_l^n$.

Базис индукции. $l = 1$. $\mathcal{K}_1^n = \{1, x_1, \ldots, x_n\}$.

U_1^n будет иметь вид, изображенный на рисунке 1, где жирной точкой изображен корень сети, а на концах ребер реализуются функции из $\mathcal{K}_1^n$.

Шаг индукции. Пусть U_{l-1}^n — сеть, реализующая все функции из $\mathcal{K}_{l-1}^n$. Сеть U_l^n будем строить, добавляя к U_{l-1}^n вершины и ребра, чтобы реализовать функции из $\mathcal{K}_l^n \setminus \mathcal{K}_{l-1}^n$, следующим образом. Возьмем произвольную монотонную элементарную конъюнкцию длины l $x_{i_1} \& \cdots \& x_{i_l}$. В сети U_{l-1}^n

найдем вершину, на которой реализуется функция $x_{i_1}\&\cdots\&x_{i_{l-1}}$ и выпустим из нее ребро, которому припишем переменную x_{i_l}. На вершине, в которое ведет это ребро, будет реализовываться как функция фильтра функция $x_{i_1}\&\cdots\&x_{i_l}$. Перебрав все функции из $\mathcal{K}_l^n \setminus \mathcal{K}_{l-1}^n$ и проделав для каждой эту операцию, получим требуемую сеть U_l^n, реализующую все функции из $\mathcal{K}_l^n$.

Отметим, что на первом ярусе сети U_l^n находится $n+1 = \binom{n}{1}+1$ ребер, и сложность каждого из них равна 1 (первый ярус — это ребра, исходящие из корня), а на i-ом ярусе ($i \geq 2$) находится $\binom{n}{i}$ ребер, и сложность каждого из них равна 2^{1-i} (i-ый ярус — это ребра, исходящие из концов ребер $(i-1)$-го яруса). Таким образом сложность сети U_l^n равна

$$T(U_l^n) = 1 + \sum_{i=1}^{l} \binom{n}{i} 2^{1-i}.$$

Вернемся теперь к нашей задаче I.

Информационную сеть U, разрешающую задачу I, будем строить следующим образом. Возьмем сеть U_{m-1}^n, описанную выше. Возьмем произвольную запись $y \in V$. Пусть она равна $x_{i_1}x_{i_2}\ldots x_{i_{m+r}}$. Возьмем в сети U_{m-1}^n вершину, на которой реализуется конъюнкция $x_{i_1}x_{i_2}\ldots x_{i_{m-1}}$ и выпустим из нее цепочку ребер, которым приписаны переменные $x_{i_m}, x_{i_{m+1}}, \ldots, x_{i_{m+r}}$. И так проделаем для каждой записи $y \in V$. Понятно, что полученная сеть U над базовым множеством $\mathcal{F}$ будет разрешать задачу I.

Подсчитаем сложность сети U. Поскольку она состоит из сети U_{m-1}^n и k цепочек, каждая из которых содержит $r+1$ ребер, и сложность которой равна, как мы отмечали выше, $2^{2-m}(1-2^{-r-1})$, то

$$T(U) = 1 + \sum_{i=1}^{m-1} 2^{1-i}\binom{n}{i} + 2^{2-m}(1-2^{-r-1}) = k\cdot 2^{2-m}(1+\bar{o}(1)).$$

Отсюда, учитывая неравенство 4, получаем

$$\inf\{T(U) : U \in \mathcal{U}_d(I,\mathcal{F})\} = k\cdot 2^{2-m}(1+\bar{o}(1)).$$

Для завершения доказательства теоремы осталось заметить, что

$$\left(\sum_{y\in V'} \mathbf{P}(O(y,\succeq))\right)/(k\cdot 2^{2-m}) = (k\cdot 2^{-m-r})/(k\cdot 2^{2-m}) = 2^{-2-r} = \bar{o}(1).$$

Тем самым, теорема доказана.

Список литературы

[1] Гасанов Э. Э. Об одной математической модели информационного поиска. *Дискретная математика* (1991) **3**, N 2, 69-76.

[2] Гасанов Э. Э. Нижняя оценка сложности информационных сетей для одного отношения частичного порядка. *Дискретная математика* (1996) **8**, N 4, 108-122.

[3] Гасанов Э. Э., Косолапов А. В. К вопросу о древовидности оптимальных информационных сетей включающего поиска. *Интеллектуальные системы* (1997) **2**, ...

[4] Гасанов Э. Э. Об одномерной задаче интервального поиска. *Дискретная математика* (1995) **7**, N 2, 40-60.

[5] Селтон Г. *Автоматическая обработка, хранение и поиск информации.* Сов. радио, Москва, 1973.

[6] Гаврилов Г. П., Сапоженко А. А. *Задачи и упражнения по курсу дискретной математики.* Наука, Москва, 1992.

[7] Полиа Г., Сеге Г. *Задачи и теоремы из анализа.* Наука, Москва, 1978.