Рецензия на доклад

Marked Ancestor Problems (часть 2)

1 Важное из первой части

Лемма 5. Пусть распределение на $\mathcal{U}$ равномерно относительно F. Тогда для всякого слоя вершин W и $x > 0$ выполнено

$$P\left(|z(u,x,W)| \geq \frac{\left|F^W\right|}{2^{Cx+1}}\right) \geq \frac{1}{2}$$

2 Амортизированная оценка

Определение. t_u и t_q являются *амортизированной стоимостью* обновления и запроса соответственно, если $\forall m, m' > n$ на выполнение m операций обновления и m' операций запроса требуется не более $mt_u + m't_q$ времени.

Теорема. Для любого $m > n$ существует дерево и последовательность из запросов вида *mark*, *unmark*, *firstmarked* на нём такое, что потребуется не менее $\Omega\left(\frac{m \log n}{\log(t_u b \log n)}\right)$ операций в cell-probe модели с ячейками в b битов.

Доказательство. Предположим, что n – степень двойки. Построим сильно ветвящееся дерево: $\delta(n) = ct_uC\log^2 n, c \geq 8$. Разобьём слои W_i на подмножества W_{ij} размера $\log n$. Через F обозначим множество способов расставить метки по вершинам так, чтобы в каждом подмножестве была отмечена только одна вершина.

Для бесконечных последовательностей $A_1, A_2, \ldots, A_k$ определим $merge(A_1, A_2, \ldots, A_k) = (a_1^1, a_2^1, \ldots, a_k^1, a_1^2, \ldots)$; воспользуемся этим, чтобы определить для вершин дерева операцию T следующим образом:

- если вершина – лист L, то $T = (L, L, \ldots)$;
- если у вершины есть потомки со значениями $T_1, T_2, \ldots, T_k$, то $T = merge(T_1, T_2, \ldots, T_k)$.

Благодаря одинаковым степеням вершин дерева получившееся T от корня проходит по листам равномерно.

Рассмотрим последовательности запросов вида $(u_1, q_1, \ldots, u_{2m}, q_{2m})$, где:

- Рассмотрим множество всех W_{ij}, пересекающих путь из T_i в корень. Тогда u_i – набор операций *unmark* на отмеченной вершине в каждом из этих множеств и *mark* на некоторой вершине в каждом из них. Такой набор операций переводит состояние из F в состояние из F.
- q_i – произвольный запрос вида $firstmarked$ к листу дерева.

Результат на выходе таких запросов не зависит от того, какое состояние было на входе, так что такое множество $\mathcal{U}$ задаёт равномерное распределение. Значит, можно будет применять лемму 5.

Для раскраски e обозначим через e_j результат применения запросов $u_1, \ldots, u_j$. Обозначим $s(j,t) = u_{j-|W_t|} \ldots u_j$; заметим, что $e_j|_{\cup_{t' \leq t} W_{t'}}$ зависит только от e_j в силу построения T. Пусть $x(s,t)$ – время выполнения $s(j,t)$.

Лемма. Пусть $S = \{(j,t) | x(s(j,t-1)) \leq 4 t_u \log n \, |W_{t-1}|\}$. Тогда $|S| \geq \frac{1}{2} m(h-1)$.

Доказательство получается подсчётом числа запросов.

Определение. Регистр q *принадлежит поколению* t, если $x(s(j,t-1)) \leq y \leq x(s(j,t))$, где y – время после последнего обновления q.

Несложно заметить, что про состояние вершин до слоя $t-1$ включительно q ничего не говорит (за счёт определения операции merge).

Утверждение. С вероятностью хотя бы $\frac{1}{8}$ (возможно, чуть меньше, главное – ограничено константой снизу) при ответе на запрос q_j будет использован регистр поколения t.

Данное утверждение даёт необходимую оценку на число операций.

С вероятностью $\frac{1}{2}$ данная пара (j,t) будет из S, и $x(s(j,t-1)) \leq 4t_u \log n |W_{t-1}| \leq [\delta(n) = ct_u C \log^2 n, c \geq 8] \leq \frac{1}{2}\frac{|W_t|}{\log n \cdot C}$. Из леммы 5 можно вывести, что $P(\exists e_1, e_2 \in Z(u_1 \ldots u_j, x(s(j,t-1)), W_t) | e_1(v) \neq e_2(v)) \geq \frac{1}{4}$ (взятием логарифма от обеих частей выражения). Но состояние регистров до поколения $t-1$ включительно в вычислениях для этих раскрасок совпадает; если не использовать регистры уровня t, то e_1 и e_2 не получится отличить. При этом с большой вероятностью на пути от v до выбранного листа отметок нет.

3 ART-представление

Определение. *Разбиением* дерева на микро- и макродеревья называется деление вершин на подмножества (микродеревья). Дерево на этих подмножествах при этом называется макродеревом.

Определение. *Тяжёлая* вершина – вершина хотя бы с двумя детьми.

Определение. *Уровнем* микродерева называется максимальное расстояние до листа в соответствующем поддереве макродерева.

Определение. *ART-представление* – разбиение дерева на микро- и макродеревья, обладающее следующими свойствами:

- в каждом микродереве не более $O(\log n)$ тяжёлых вершин;
- максимальный уровень микродерева не более $O\left(\frac{\log n}{\log \log n}\right)$.

Лемма. *ART-представление* существует и может быть построено за линейное время.

Строим так: выделяем вершины, в поддеревьях которых не более $\log n$ тяжёлых вершин, а в поддеревьях их родителей – больше. Всё выкидываем и повторяем. Количество вершин при этом быстро уменьшается, так как у всякого листа в дереве после выкидываний был потомок – микродерево, их не более $\frac{n}{\log n}$, тогда в оставшемся дереве не более $\frac{n}{\log n}$ тяжёлых вершин. Продолжая, получаем, что уровней не более $\log_{\log n} n = \frac{\log n}{\log \log n}$.

Это может быть сделано одной динамикой по дереву от листьев.

Используя ART-представление, можно быстро выполнять операции: для *firstmarked* проходимся по микродеревьям на пути до корня и отвечаем на вопрос там.