УДК 519.172:519.683

И. В. КОМАРОВСКИЙ

ЭКСТРЕМАЛЬНЫЕ СВОЙСТВА α-ВЗВЕШЕННЫХ ДВОИЧНЫХ ДЕРЕВЬЕВ

В современном программировании деревья являются одними из наиболее интенсивно применяемых структур, причем они используются не только как способ представления структур данных, но и как средство анализа поведения определенных алгоритмов (например, различных стратегий слияния в задачах внешней сортировки [1]). В этой связи возникает потребность в количественных измерениях различных характеристик деревьев. В [1—3] описаны такие характеристики деревьев, как длины внешнего, взвешенного внешнего и степенного путей, а также построены соответствующие деревья, имеющие минимальные длины таких путей. Использование этих характеристик для оценки различных стратегий слияния участков в задачах внешней сортировки предполагает, что в результате слияния длина результирующего участка равна сумме длин исходных участков.

В задачах совместной обработки файлов [4, 5] длина результирующего файла, как правило, меньше суммы длин исходных файлов. Для оценки различных стратегий совместной обработки полезными оказываются двоичные деревья с α-взвешенной длиной внешнего пути (α-взвешенные двоичные деревья), исследуемые в настоящей работе.

Ориентированным двоичным деревом назовем ориентированное дерево [6], в котором полустепень исхода любой вершины равна 0 или 2. В статье рассматриваются только такие деревья, поэтому слова «ориентированное двоичное» будут опускаться. Вершины с полустепенью исхода, отличной от нуля, назовем внутренними, а с полустепенью исхода, равной нулю, — листьями. Листья при изображении дерева будем обозначать квадратом. Уровень вершины есть длина пути из корня в эту вершину. Высота дерева — это длина самого длинного пути из корня дерева в какой-нибудь лист. Множество всех деревьев с $n \geqslant 1$ листьями обозначим $\mathbf{D}_n$.

Терминологию, относящуюся к лесам и деревьям, будем использовать из [7].

На вершинах дерева $D_n \in \mathbf{D}_n$ определим рекурсивно функции μ_α и σ'_α:

$$\mu_\alpha(x) = \begin{cases} 1, & \text{если } x \text{ — лист,} \\ \alpha(\mu_\alpha(x_1) + \mu_\alpha(x_2)) & \text{в противном случае,} \end{cases}$$

$$\sigma'_\alpha(x) = \begin{cases} \mu_\alpha(x), & \text{если } x \text{ — лист,} \\ \mu_\alpha(x) + \sigma'_\alpha(x_1) + \sigma'_\alpha(x_2) & \text{в противном случае,} \end{cases}$$

где x_1 и x_2 — вершины, в которые заходят дуги из x, $\alpha > 0$. Значения функций μ_α и σ'_α в корне дерева (поддерева) $D_m \in \mathbf{D}_m$ будем обозначать соответственно $\mu_\alpha(D_m)$ и $\sigma'_\alpha(D_m)$, $0 < m \leqslant n$.

Пусть $\sigma_\alpha(D_m) = \sigma'_\alpha(D_m) - \mu_\alpha(D_m)$, $0 < m \leqslant n$.

Для фиксированного α назовем оптимальным в $\mathbf{D}_n$ дерево D_n^*, для которого $\sigma_\alpha(D_n^*)$ минимально. Нашей задачей является нахождение структуры оптимальных деревьев для различных α и вычисление для них значений функций σ_α.

Каждой дуге дерева D_n, заходящей в вершину x, сопоставим значение $\mu_\alpha(x)$, которое будет весом этой дуги. Внешним путем дерева назовем совокупность всех его путей от корня в лист, а α-взвешенной длиной внешнего пути — сумму весов всех дуг дерева.

Заметим, что 1-взвешенная длина внешнего пути дерева есть длина внешнего пути в [2].

Лемма 1. α-Взвешенная длина внешнего пути дерева D_n есть $\sigma_\alpha(D_n)$.

Доказательство. Из определения σ_α и σ'_α нетрудно получить, что

$$\sigma_\alpha(D_n) = \sigma_\alpha(D') + \sigma_\alpha(D'') + \mu_\alpha(D') + \mu_\alpha(D''), \quad (1)$$

где D' и D'' — подвешенные к корню (непосредственные) поддеревья дерева D_n, $n > 1$. Теперь доказательство — элементарная индукция по высоте дерева. Лемма доказана.

Лемма 2. Пусть $D_n \in \mathbf{D}_n$ — дерево высоты h. Тогда

$$\mu_\alpha(D_n) = a_h\alpha^h + a_{h-1}\alpha^{h-1} + \ldots + a_1\alpha + a_0,$$

где $\sum_{i=0}^{h} a_i = n$ и a_j — число листьев дерева на уровне j, $j = \overline{0, h}$.

Доказательство проведем индукцией по высоте h дерева. Для $h = 0$ утверждение леммы, очевидно, верно. Пусть D_n — дерево высоты $h + 1$, $h \geqslant 0$, и D'_{n_1}, D''_{n_2} — его непосредственные поддеревья (непустые), $n_1 > 0$, $n_2 > 0$, $n_1 + n_2 = n$. Высоты h_1 и h_2 этих поддеревьев не превосходят h, причем хотя бы одна из них равна h. Следовательно, по предположению индукции имеем

$$\mu_\alpha(D'_{n_1}) = a'_{h_1+1}\alpha^{h_1} + a'_{h_1}\alpha^{h_1-1} + \ldots + a'_2\alpha + a'_1,$$

$$\mu_\alpha(D''_{n_2}) = a''_{h_2+1}\alpha^{h_2} + a''_{h_2}\alpha^{h_2-1} + \ldots + a''_2\alpha + a''_1,$$

где $\sum_{i=1}^{h_1+1} a'_i = n_1$, a'_i — число листьев в поддереве D'_{n}

на уровне j дерева D_n, $j=\overline{1, h_1+1}$; $\sum_{i=1}^{h_2+1} a''_i = n_2$, a''_j — число листьев в поддереве D''_{n_2} на уровне j дерева D_n, $j=\overline{1, h_2+1}$. Согласно определению функции μ_α имеем

$$\mu_\alpha(D_n) = \alpha(\mu_\alpha(D'_{n_1}) + \mu_\alpha(D''_{n_2})) = \\ = \sum_{i=1}^{h_1+1} a'_i\alpha^i + \sum_{i=1}^{h_2+1} a''_i\alpha^i = \sum_{i=0}^{h+1} a_i\alpha^i,$$

где $h+1=\max(h_1+1, h_2+1)$ — высота дерева D_n, $a_0=0$, $\sum_{i=1}^{h+1} a_i = n$, a_j — число листьев уровня j в дереве D_n, $j=\overline{0, h+1}$. Лемма доказана.

Полным путем дерева назовем путь от корня дерева в лист, а его α-взвешенной длиной — сумму $\alpha^{h-1}+\ldots+\alpha+1$, где h — уровень листа. Для $h=0$ α-взвешенную длину единственного полного пути положим равной 0. Так, если $(d_h, d_{h-1}), \ldots, (d_1, d_0)$ есть полный путь дерева, то в выражении для его α-взвешенной длины слагаемому α^i соответствует дуга (d_{i+1}, d_i), $i=\overline{0, h-1}$.

Лемма 3. α-Взвешенная длина внешнего пути дерева $D_n \in \mathbf{D}_n$ равна сумме по всем листьям α-взвешенных длин полных путей.

Доказательство проведем индукцией по высоте h дерева. Для $h=0$ и $h=1$ справедливость леммы очевидна. Допустим, что утверждение леммы справедливо для всех деревьев высоты не более $h \geqslant 1$. Пусть D_n — дерево высоты $h+1$. Его непосредственные поддеревья D'_{n_1} и D''_{n_2} имеют высоты h_1 и h_2 такие, что $h=\max(h_1, h_2)$, и, следовательно, для них, по предположению индукции, утверждение леммы справедливо. Если к сумме α-взвешенных длин внешних путей поддеревьев D'_{n_1} и D''_{n_2} добавить $\mu_\alpha(D'_{n_1})+\mu_\alpha(D''_{n_2})$ (см. (1)), то получим α-взвешенную длину внешнего пути дерева D_n. В то же время согласно лемме 2

$$\mu_\alpha(D'_{n_1}) = \sum_{i=0}^{h_1} a'_{i+1}\alpha^i, \quad \mu_\alpha(D''_{n_2}) = \sum_{i=0}^{h_2} a''_{i+1}\alpha^i,$$

где $\sum_{i=1}^{h_1+1} a'_i = n_1$, $n_1>0$, a_j — число листьев в поддереве D'_{n_1} на уровне j дерева D_n, $j=\overline{1, h_1+1}$; $\sum_{i=1}^{h_2+1} a''_i = n_2$, $n_2>0$, a_j — число листьев в поддереве D''_{n_2} на уровне j дерева D_n, $j=\overline{1, h_2+1}$, $n_1+n_2 = n$. Отсюда

$$\mu_\alpha(D'_{n_1}) + \mu_\alpha(D''_{n_2}) = a_{h+1}\alpha^h + \ldots + a_2\alpha + a_1, \quad (2)$$

где $\sum_{i=1}^{h+1} a_i = n$, a_j — число листьев на уровне j в дереве D_n, $j=\overline{1, h+1}$, $h=\max(h_1, h_2)$. Все полные пути дерева D_n можно получить из полных путей поддеревьев D'_{n_1} и D''_{n_2} посредством добавления дуги, соединяющей корень дерева с корнем соответствующего поддерева. Заметим теперь, что для полного пути дерева D_n, заканчивающегося в листе уровня i, слагаемое в выражении для α-взвешенной длины этого пути, соответствующее исходящей из корня дерева D_n дуге, будет равно α^{i-1}, а остальные слагаемые составят α-взвешенную длину полного пути поддерева. Сумма этих α^{i-1} по всем листьям дерева D_n, как нетрудно видеть из (2), и есть $\mu_\alpha(D'_{n_1})+\mu_\alpha(D''_{n_2})$. Лемма доказана.

Полностью сбалансированным назовем дерево, в котором листья могут находиться на максимальном и предшествующем ему уровнях.

Лемма 4 [1]. Высота полностью сбалансированного дерева с n листьями равна $\lceil \log_2 n \rceil$ и число листьев на уровнях $\lceil \log_2 n \rceil - 1$ и $\lceil \log_2 n \rceil$ равно соответственно $2^{\lceil \log_2 n \rceil} - n$ и $2n - 2^{\lceil \log_2 n \rceil}$, где $\lceil \log_2 n \rceil$ — наименьшее целое, не меньшее $\log_2 n$.

Квазилинейным назовем дерево, все внутренние вершины которого составляют цепь.

Теорема 1. В классе $\mathbf{D}_n$ оптимальным деревом является: при $0<\alpha<0{,}5$ — квазилинейное, а при $\alpha=0{,}5$ — произвольное. В случае $\alpha>0{,}5$ дерево из $\mathbf{D}_n$ оптимально тогда и только тогда, когда оно полностью сбалансировано.

Доказательство. Пусть D^*_n — оптимальное в $\mathbf{D}_n$ дерево. Согласно лемме 1 α-взвешенная длина такого дерева минимальна.

1. Пусть $0<\alpha<0{,}5$. Допустим, что внутренние вершины в D^*_n не образуют цепь. Тогда в дереве существуют по крайней мере две внутренние вершины, к которым присоединены по два листа. Пусть уровни этих листьев M и m, $M \geqslant m$. Удалим два листа, которые являются братьями на уровне m, в результате чего их отец станет листом. Затем эту пару присоединим к одному из листьев пары на уровне M, который в результате такой операции станет внутренним. Описанная процедура показана на рис. 1. Такая модификация дерева сохраняет число листьев. Используя лемму 3, нетрудно вычислить величину Δ, на которую изменится α-взвешенная длина внешнего пути:

$$\Delta = 2(1+\alpha+\ldots+\alpha^{m-1}) - (1+\alpha+\ldots+\alpha^{m-2}) + \\ + (1+\alpha+\ldots+\alpha^{M-1}) - 2(1+\alpha+\ldots+\alpha^{M}) = \\ = \frac{\alpha^{m-1}}{1-\alpha}(1-\alpha^{M-m+1})(1-2\alpha).$$

Поскольку при $0<\alpha<0{,}5$ и $M \geqslant m$ величина Δ положительная, такое преобразование уменьшает α-взвешенную длину внешнего пути, что противоречит оптимальности дерева D^*_n.

2. В случае $\alpha=0{,}5$, как нетрудно видеть, все дуги имеют вес 1 для любого дерева $D \in \mathbf{D}_n$. Отсюда, α-взвешенная длина внешнего пути равна постоянной величине — числу дуг дерева, и,

следовательно, не зависит от конфигурации дерева. Это и доказывает требуемое утверждение теоремы.

3. Пусть $\alpha > 0{,}5$. Обозначим M и m соответственно максимум и минимум всех уровней, на которых появляются листья, $M \geqslant m$. Предположим, что $M \geqslant m + 2$. Удалим два листа, которые являются братьями на уровне M, в результате чего их отец станет листом. Затем эту пару присоединим к листу на уровне m, который станет внутренним (рис. 2). Такая модификация дерева сохраняет число листьев. Используя лемму 3, нетрудно вычислить величину Δ, на которую изменится α-взвешенная длина внешнего пути:

$$\Delta = 2(1 + \alpha + \ldots + \alpha^{M-1}) - (1 + \alpha + \ldots + \alpha^{M-2}) + (1 + \alpha + \ldots + \alpha^{m-1}) - 2(1 + \alpha + \ldots + \alpha^{m}) = \alpha^{m+1}(1 + \alpha + \ldots + \alpha^{M-m-2}) + \alpha^{M-1} - \alpha^{m}. \quad (3)$$

Ясно, что при $\alpha = 1$ и $M \geqslant m+2$ эта величина положительная. Для $\alpha > 0$ и $\alpha \neq 1$ из (3) получим $\Delta = \frac{\alpha^{m}}{1-\alpha}(1 - \alpha^{M-m-1})(2\alpha - 1)$. Отсюда следует, что и при $\alpha > 0{,}5$, $\alpha \neq 1$, $M \geqslant m + 2$ величина Δ положительная. Таким образом, в результате преобразования α-взвешенная длина внешнего пути дерева D_n^* будет уменьшена, что противоречит оптимальности D_n^*.

Обратно, докажем, что произвольное полностью сбалансированное дерево $D \in \mathbf{D}_n$ оптимальное. Из леммы 4 получаем, что для различных полностью сбалансированных деревьев с n листьями число листьев на уровнях $\lceil \log_2 n \rceil - 1$ и $\lceil \log_2 n \rceil$ постоянно. Отсюда, в силу леммы 3, имеем, что α-взвешенная длина внешнего пути дерева одинакова для всех полностью сбалансированных деревьев с n листьями.

Теорема доказана.

Следствие 1. Пусть D_n^* — оптимальное дерево в классе $\mathbf{D}_n$. Тогда

$$\sigma_\alpha(D_n^*) = \begin{cases} \dfrac{n - \alpha^{n-1}}{1-\alpha} + \dfrac{\alpha(\alpha^{n-1} - 1)}{(1-\alpha)^2}, & 0 < \alpha < 0{,}5, \\ 2n - 2, & \alpha = 0{,}5, \\ n\,\dfrac{1 - \alpha^{k}}{1-\alpha} - \alpha^{k-1}(2^{k} - n), & \alpha > 0{,}5, \quad \alpha \neq 1, \\ (k+1)n - 2^{k}, & \alpha = 1, \end{cases}$$

где $k = \lceil \log_2 n \rceil$.

Доказательство. Непосредственными вычислениями убеждаемся в справедливости следствия при $n = 1$. Далее полагаем $n > 1$.

Пусть $0 < \alpha < 0{,}5$. Согласно теореме 1 для такого α в оптимальном дереве внутренние вершины образуют цепь. Так как число внутренних вершин двоичного дерева с n листьями равно $n -$

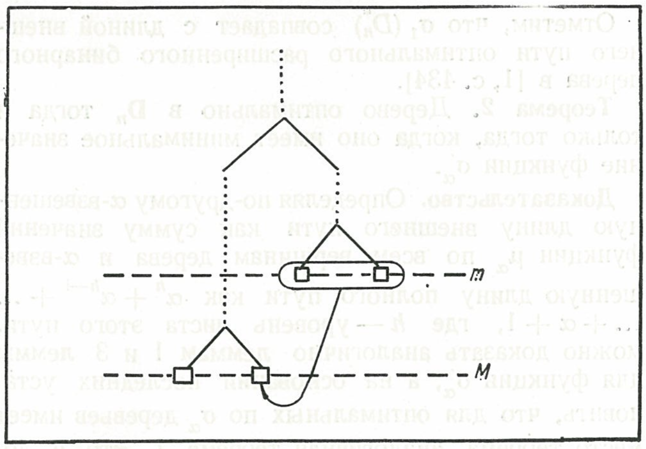

Рис. 1

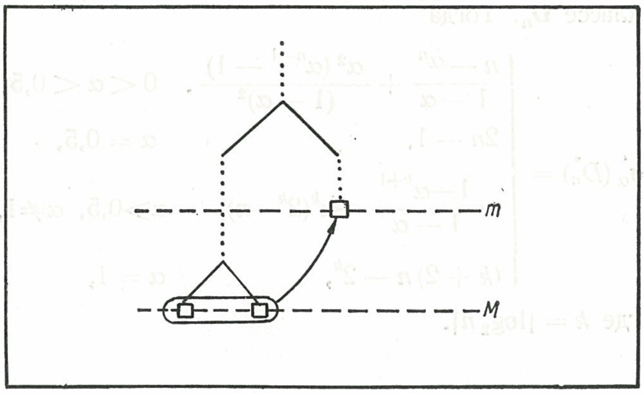

Рис. 2

$- 1$, длина цепи будет $n - 2$. Отсюда высота дерева есть $n - 1$. Теперь, используя леммы 1 и 3, нетрудно вычислить α-взвешенную длину внешнего пути:

$$\sigma_\alpha(D_n^*) = 2(1 + \alpha + \ldots + \alpha^{n-2}) + (1 + \alpha + \ldots + \alpha^{n-3}) + \ldots + 1 = \frac{n - \alpha^{n-1}}{1-\alpha} + \frac{\alpha(\alpha^{n-1} - 1)}{(1-\alpha)^2}.$$

Пусть $\alpha \geqslant 0{,}5$. Согласно теореме 1 полностью сбалансированное дерево является оптимальным (это верно и для $\alpha = 0{,}5$). По лемме 4 высота этого дерева k, причем число листьев на уровнях $k - 1$ и k равно соответственно $2^k - n$ и $2n - 2^k$. Отсюда, в силу лемм 1 и 3,

$$\sigma_\alpha(D_n^*) = (1 + \alpha + \ldots + \alpha^{k-1})(2n - 2^k) + (1 + \alpha + \ldots + \alpha^{k-2})(2^k - n).$$

При $\alpha = 1$ получим $\sigma_1(D_n^*) = k(2n - 2^k) + (k - 1)(2^k - n) = (k+1)n - 2^k$. Для $\alpha \neq 1$ имеем $\sigma_\alpha(D_n^*) = n\,\frac{1-\alpha^k}{1-\alpha} - \alpha^{k-1}(2^k - n)$, причем если $\alpha = 0{,}5$, то $\sigma_{0,5}(D_n^*) = 2n - 2$. Следствие доказано.

Отметим, что $\sigma_1(D_n^*)$ совпадает с длиной внешнего пути оптимального расширенного бинарного дерева в [1, с. 434].

Теорема 2. Дерево оптимально в $\mathbf{D}_n$ тогда и только тогда, когда оно имеет минимальное значение функции σ'_α.

Доказательство. Определяя по-другому α-взвешенную длину внешнего пути как сумму значений функции μ_α по всем вершинам дерева и α-взвешенную длину полного пути как $\alpha^h + \alpha^{h-1} + \ldots$ $\ldots + \alpha + 1$, где h — уровень листа этого пути, можно доказать аналогично леммам 1 и 3 леммы для функции σ'_α, а на основании последних установить, что для оптимальных по σ'_α деревьев имеет место теорема, аналогичная теореме 1. Это и доказывает теорему 2.

Следствие 2. Пусть D_n^* — оптимальное дерево в классе $\mathbf{D}_n$. Тогда

$$\sigma'_\alpha(D_n^*) = \begin{cases} \dfrac{n-\alpha^n}{1-\alpha} + \dfrac{\alpha^2(\alpha^{n-1}-1)}{(1-\alpha)^2}, & 0<\alpha<0{,}5, \\ 2n-1, & \alpha = 0{,}5, \\ n\dfrac{1-\alpha^{k+1}}{1-\alpha} - \alpha^k(2^k - n), & \alpha>0{,}5,\ \alpha\neq 1, \\ (k+2)n - 2^k, & \alpha = 1, \end{cases}$$

где $k = \lceil \log_2 n \rceil$.

Доказательство. Из определения σ_α следует $\sigma'_\alpha(D_n^*) = \sigma_\alpha(D_n^*) + \mu_\alpha(D_n^*)$. Согласно теоремам 2 и 1 достаточно рассмотреть три случая: $0<\alpha<0{,}5$, $\alpha = 0{,}5$, $\alpha > 0{,}5$, и для них, используя лемму 2 и следствие 1, вычислить $\sigma'_\alpha(D_n^*)$.

СПИСОК ЛИТЕРАТУРЫ

1. Кнут Д. Искусство программирования. Сортировка и поиск.— М. : Мир, 1978.— Т. 3.— 848 с.
2. Кнут Д. Искусство программирования. Основные алгоритмы.— М. : Мир, 1976.— Т. 1.— 736 с.
3. Schlumberger M., Vuillemin J. Optimal disk merge petterns // Acta Informatica.— 1973.— 3.— P. 25—36.
4. Дробушевич Г. А., Комаровский И. В. Об одной модели задачи совместной обработки файлов // Вестн. Белорус. гос. ун-та. Сер. 1. Физика, математика и механика.— 1983.— № 1.— С. 37—40.
5. Комаровский И. В. Автоматизация программирования совместной обработки файлов // Всесоюз. науч.-техн. конф. «Программные средства как продукция производственно-технического назначения». Секция «Технология разработки программных средств» : Тез. докл.— Калинин, 1985.— С. 82—85.
6. Кристофидес Н. Теория графов. Алгоритмический подход.— М. : Мир, 1978.— 432 с.
7. Ахо А., Хопкрофт Дж., Ульман Дж. Построение и анализ вычислительных алгоритмов.— М. : Мир, 1979.— 536 с.

Поступила 09.07.85