ОБЪЕДИНЕННЫЙ
ИНСТИТУТ
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ
ДУБНА

Р11 - 9103

В.М.Сумароков

НЕКОТОРЫЕ КОМБИНАТОРНЫЕ СВОЙСТВА ОДНОГО КЛАССА АССОЦИАТИВНЫХ ПРИЗНАКОВЫХ СТРУКТУР

1975

P11 - 9103

В.М.Сумароков

НЕКОТОРЫЕ КОМБИНАТОРНЫЕ СВОЙСТВА ОДНОГО КЛАССА АССОЦИАТИВНЫХ ПРИЗНАКОВЫХ СТРУКТУР

Направлено в журнал "Программирование"

I. Предварительные замечания и результаты

Среди ассоциативных списковых структур, предназначенных для организации в памяти ЭВМ машинных образов некоторых объектов или их описаний, идентифицированных их признаками, наиболее эффективными являются древовидные структуры. Они обеспечивают высокое быстродействие доступа к данным не только при их поиске по ключу, но также при включении новой и исключении устаревшей информации. Это определило их широкое использование в информационных системах с развитыми динамическими свойствами банков данных. В настоящей работе сообщаются результаты, дополняющие исследования по позиционным и наращиваемым признаковым деревьям.

Будем следовать в основном терминологии, разработанной в теории ассоциативного программирования /I/, а также используемой в /2,3/. Определим некоторые понятия.

Дерево - это связный граф, не содержащий циклов. В каждый узел дерева, кроме корня, входит одно ребро и из каждого узла исходит несколько ребер. Совокупность ребер, исходящих из одного узла, образует пучок, а количество таких ребер называется размером пучка. Номер уровня, которому принадлежит узел, определяется числом ребер, соединяющих узел с корнем. В исследуемых деревьях узлы, принадлежащие одному уровню, имеют одинаковые размеры пучков.

Поиск объекта в древовидной признаковой структуре осуществляется в процессе прослеживания узлов дерева на его последовательных уровнях сверху вниз. Траектория, описываемая при этом поисковой точкой, зависит от алгоритма поиска и называется поисковым треком. Узел дерева, которому поставлен в соответствие объект или некоторое его описание, называется объектным. Узел, не содержащий объектной информации, является признаковым.

Структура исследуемых деревьев однозначно определяется кортежем $\langle m, n_0, n_1, \dots, n_{m-1} \rangle$, где m – количество уровней дерева, n_i – размер пучка i-го уровня, $i=\overline{0,m-1}$. В качестве критерия выбора этих параметров следует принять минимальную трудоемкость поиска, характеристикой которой может служить длина поискового трека искомого объектного узла (цели поиска). Но эта величина сильно зависит от положения цели поиска в ряду последовательных объектных узлов дерева, пилообразно изменяясь в широких пределах. Поэтому в качестве характеристики трудоемкости поиска принимается сумма Φ длин поисковых треков всех объектных узлов, отнесенная к их общему числу N.

Для позиционного дерева /I/, в котором

$$N=\prod_{i=0}^{m-1} n_i, \qquad \text{(I.I)}$$

в /3/ с помощью предложенной зонной модели получено

$$\Phi=\frac{1}{2}\prod_{i=0}^{m-1} n_i \sum_{j=0}^{m-1}(n_j+1). \qquad \text{(I.2)}$$

Для приведенной длины поискового трека с учетом (I.I) и (I.2) справедлива следующая цепочка тождественных преобразований $(n_i \in [2,N])$:

$$\varphi=\frac{1}{2}\left[\frac{n_0+1}{\ln n_0}\ln n_0+\sum_{i=1}^{m-1}\frac{n_i+1}{\ln n_i}\ln n_i\right]=\frac{1}{2}\left[H(n_0)\left(\ln N-\sum_{i=1}^{m-1}\ln n_i\right)+\sum_{i=1}^{m-1}H(n_i)\ln n_i\right]=\frac{1}{2}\left\{H(n_0)\ln N+\sum_{i=1}^{m-1}\ln n_i\left[H(n_i)-H(n_0)\right]\right\}, \qquad \text{(I.3)}$$

где

$$H(x)=\frac{1+x}{\ln x}.$$

Зависимость $H(x)$ изображена на рис.I, причем $H_{min}\simeq H(3{,}592)$ и $[H(3)\simeq 3{,}641]>[H(4)\simeq 3{,}607]$. Поэтому из (I.3) можно видеть, что оптимальным является позиционное дерево, размеры пучков на уровнях которого одинаковы и равны 4. Действительно, φ минимально, когда $n_0=4$, а $n_i=n_0$, $i=\overline{1,m-1}$; $\varphi(4)\simeq 1{,}8\ln N$.

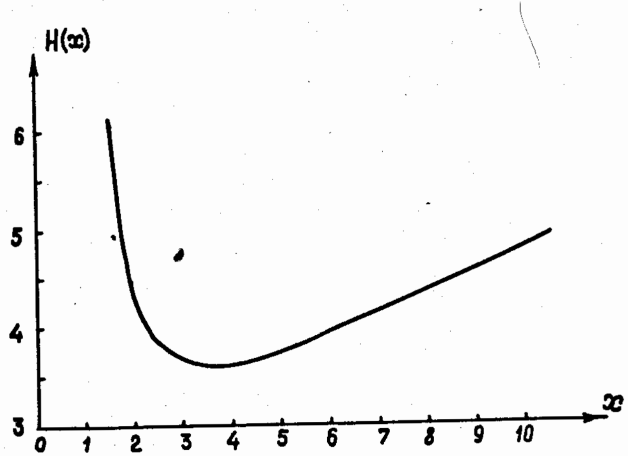

Рис.I. Зависимость $H(x)=(1+x)/\ln x$.

В позиционном дереве все объектные узлы расположены на терминальном уровне. Доступ к каждому из них осуществляется в результате проверки от I до n_i узлов в пучках вышележащих уровней, и длина l всякого поискового трека

$$m \le l \le \sum_{i=0}^{m-1} n_i.$$

Лемма. Энумератор для последовательности всевозможных значений длин поисковых треков позиционного признакового дерева есть

$$F(t)=\prod_{j=0}^{m-1}\sum_{i=1}^{n_j} t^i=\sum_{l=m}^{\lambda} C_l t^l , \qquad (I.4)$$

где t - свободный параметр,

C_l - количество треков длины l ,

$\lambda=\sum_{i=0}^{m-1} n_i$.

Доказательство. Соотношение (I.4) имеет силу в качестве перечисляющей производящей функции, или энумератора, для комбинаторной композиции /4/ с точно m частями, каждая из которых не превышает n_i, $i=\overline{0,m-1}$. Поэтому доказательство сводится к установлению адекватности моделей алгоритма поиска в позиционном дереве и данной композиции. Действительно, композиция как упорядоченный набор целых положительных чисел при данной их сумме характеризуется образующими ее частями, характеристикой (их суммой) и величинами частей. Если этим параметрам поставить в соответствие число уровней дерева, длину поискового трека и размеры пучков, то достигается полнота аналогии.

Количество треков длины l в позиционном дереве будет

$$C_l=\frac{1}{l!}\,\frac{\partial^{(l)}F(t)}{\partial t^{(l)}}\bigg|_{t=0} .$$

В частном случае оптимальной структуры

$$F(t)=(t+t^2+t^3+t^4)^m=\sum_{l=m}^{4m} C_l t^l .$$

Ее реализация соответствует наибольшему смещению симметричной относительно оси $l=(m+\lambda)/2$ кривой энумератора в область коротких длин треков по сравнению с соответствующими кривыми для неоптимальных равноемких деревьев.

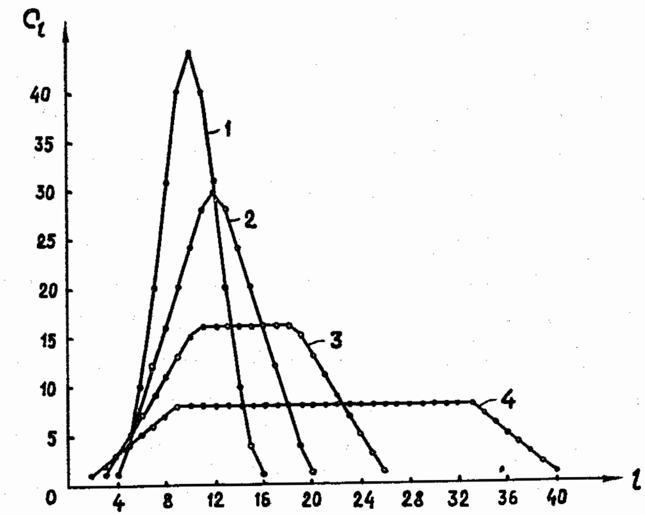

Рис.2. Функции распределения числа треков по их длинам в позиционных деревьях: I - <[illegible],4,4,4,4>; 2 - <4,8,8,2,2>; 3 - <3,I6,8,2 >; 4 - <2,32,8 >.

Из рис.2 (все деревья равноемки: N =256) видно, что по мере удаления значений параметров структуры от оптимальных число треков малых длин уменьшается, а больших длин - увеличивается, и поисковые свойства дерева ухудшаются. С уменьшением числа уровней, что в равноемких деревьях возможно только за счет увеличения размеров пучков, график функции $C_l(l)$ "уплощается", что соответствует развитию тенденции равномерного представления треков разных длин. Другим крайним случаем, по отношению к оптимальной структуре, является дерево $\langle 1,N\rangle$, фактически представляющее собой последовательный список /I/. Оно содержит по одному треку длин от I до N.

2. Анализ наращиваемого дерева

Наращиваемое дерево /I,2/ строится для одного поискового признака, который имеет количественный смысл и значения которого являются идентификаторами узлов дерева. В качестве значения такого признака можно рассматривать совокупность значений признаков поискового ключа объекта, если объект характеризуется несколькими признаками. Набор допустимых значений признака заранее не фиксируется и может произвольно изменяться в соответствии с фактическими значениями признака поступающих в систему объектов. В узлах каждого пучка дерева объекты размещаются в порядке возрастания значений признака, поэтому упорядоченность сохраняется в отношении ключей всех объектов на каждом из уровней. Пучок Π_0 делит весь диапазон изменения поискового признака на несколько интервалов. Каждый пучок последующего уровня делит соответствующие интервалы на еще более мелкие и т.д.

Если объекты поступают в систему в монотонном порядке значений поисковых ключей, то ветви такого дерева начинают расти преимущественно в одну сторону, что увеличивает трудоемкость доступа к данным. В /2/ описан алгоритм балансировки наращиваемого дерева, начинающий формирование нового уровня только после заполнения всех узлов предыдущего.

Все узлы дерева /2/, кроме узлов последнего уровня, являются признаковыми. Наибольшее значение признака в ряду узлов пучка равно значению признака в корневом узле пучка. При поиске в процессе проверки узлов дерева сверху вниз сравнивается заданное значение признака поискового ключа со значениями признака в узлах просматриваемых пучков. Переход к пучку нижележащего уровня происходит, когда значение признака проверяемого узла не меньше заданного значения признака, при условии, что в предыдущем узле этого пучка

значение признака было меньше заданного. По достижении узла с заданным значением поискового признака происходит спуск по правым узлам пучков нижележащих уровней к объектному узлу. Такие дервья, эффективно используемые для целей информационного поиска и построения транслирующих систем, впервые исследованы Вальтером Ландауэром в его докторской диссертации. Описанное им дерево используется в качестве надструктуры по отношению к многоассоциативной области памяти, состоящей из узловых объектных списков; поисковые признаки могут быть кодами дескрипторов или слов языка программирования, а содержимое объектных узлов – адресами объектных списков или соответствующих шаблонов в машинном языке.

Однако алгоритм поиска в дереве /2/ может быть улучшен сообщением объектных функций тем корневым признаковым узлам пучков, в которых заданное значение признака встречается впервые при движении поисковой точки от корня дерева. В этом случае при совпадении отпадает необходимость спуска к нижнему уровню, что ускоряет поиск. В структуре, модифицированной таким образом, объектными являются узлы всех уровней, кроме правых узлов всех пучков. Такая перестройка не затрагивает основ алгоритмов включения и исключения данных. Ниже на основе зонной модели исследуются поисковые свойства модифицированного дерева Ландауэра.

Количество объектных узлов N сбалансированного дерева с завершенным нижнем уровнем равно разности между общим числом всех узлов дерева и количеством всех признаковых узлов. Так как число признаковых узлов i -го уровня равно общему количеству узлов предыдущего уровня, то с учетом корня

$$N = 1+\sum_{j=0}^{m-1}\prod_{i=0}^{j} n_i - \left(1+\sum_{j=0}^{m-2}\prod_{i=0}^{j} n_i\right) = \prod_{i=0}^{m-1} n_i ,$$

что совпадает с (I.I).

<u>Утверждение I</u>. Наибольшая скорость обследования достигается в сбалансированном наращиваемом дереве, размеры пучков на уровнях которого следуют закону:

$$n_\kappa = n_0 - \frac{2}{N}\sum_{j=0}^{\kappa-1}\prod_{i=0}^{j} n_i\,, \quad \kappa=\overline{1,m-1}, \qquad (2.I)$$

где n_i – размер пучка i-го уровня, N – емкость дерева по (I.I).

<u>Доказательство</u>. Для минимизации воспользуемся критерием, описанным в предыдущем параграфе.

Найдем сумму длин поисковых треков объектных узлов одного уровня, а затем интегрированием по всем уровням получим параметр Φ. Сумму длин треков объектных узлов κ-го уровня определим как разность соответствующих выражений для всех узлов уровня и для его признаковых узлов. Используя соотношение (I.2), можно записать уменьшаемое в виде:

$$\Phi_{1\kappa} = \frac{1}{2}\prod_{i=0}^{\kappa-1} n_i \sum_{j=0}^{\kappa-1}(n_j+1) + \prod_{i=0}^{\kappa-1} n_i\,.$$

Второе слагаемое здесь учитывает, что треки всех узлов κ-го уровня $\prod\limits_{i=0}^{\kappa-1} n_i$ в наращиваемом дереве, в отличие от позиционного, начинаются с проверки корня дерева.

Поисковые треки признаковых узлов κ-го уровня можно рассматривать состоящими из двух участков: от корня дерева до корневого узла пучка $(\kappa-1)$-го уровня, которому принадлежит признаковый узел κ-го уровня, и от него – до признакового узла. Количество признаковых узлов κ-го уровня равно числу всех узлов $(\kappa-1)$-го уровня, так как в каждом пучке имеется по одному признаковому узлу. Поэтому сумма длин первых участков треков признаковых узлов κ-го уровня будет равна

$$\frac{1}{2}\prod_{i=0}^{\kappa-2} n_i \sum_{j=0}^{\kappa-2}(n_j+1) + \prod_{i=0}^{\kappa-2} n_i\,.$$

Второй участок трека каждого признакового узла равен $n_{\kappa-1}$, так как для его достижения нужно просмотреть все узлы пучка. Сумма длин вторых участков треков признаковых узлов κ-го уровня

равна $n_{\kappa-1}\prod\limits_{i=0}^{\kappa-2} n_i$. Следовательно, сумма длин треков всех признаковых узлов κ-го уровня

$$\Phi_{2\kappa} = \frac{1}{2}\prod_{i=0}^{\kappa-2} n_i \sum_{j=0}^{\kappa-2}(n_j+1) + \prod_{i=0}^{\kappa-2} n_i + n_{\kappa-1}\prod_{i=0}^{\kappa-2} n_i\,.$$

Сумма длин треков объектных узлов κ-го уровня

$$\Phi_\kappa = \Phi_{1\kappa} - \Phi_{2\kappa} = \frac{1}{2}\prod_{i=0}^{\kappa-1} n_i \sum_{j=0}^{\kappa-1}(n_j+1) - \frac{1}{2}\prod_{i=0}^{\kappa-2} n_i \sum_{j=0}^{\kappa-2}(n_j+1) - \prod_{i=0}^{\kappa-2} n_i =$$

$$= \lambda_\kappa - \lambda_{\kappa-1} - \prod_{i=0}^{\kappa-2} n_i\,,$$

где

$$\lambda_\kappa = \frac{1}{2}\prod_{i=0}^{\kappa-1} n_i \sum_{j=0}^{\kappa-1}(n_j+1).$$

Сумма длин треков объектных узлов всего дерева

$$\Phi = \sum_{\kappa=1}^{m}\Phi_\kappa = \lambda_m - \lambda_0 - \sum_{\kappa=1}^{m}\prod_{i=0}^{\kappa-2} n_i = \frac{1}{2}\prod_{i=0}^{m-1} n_i \sum_{j=0}^{m-1}(n_j+1) - \sum_{j=0}^{m-2}\prod_{i=0}^{j} n_i\,.$$

Сумма длин поисковых треков объектных узлов всего дерева, отнесенная к их числу,

$$\varphi = \frac{\Phi}{N} = \frac{1}{2}\sum_{j=0}^{m-1}(n_j+1) - \frac{1}{N}\sum_{j=0}^{m-2}\prod_{i=0}^{j} n_i \qquad (2.2)$$

характеризует трудоемкость поиска в сбалансированном наращиваемом дереве. Считая m параметром, исследуем функцию $\varphi(n_0, n_1, \ldots, n_{m-1})$ на экстремум при условии (I.2). Провомерна такая запись:

$$\varphi = \frac{m}{2} + 1 + \frac{N}{2}\Big(\prod_{i=1}^{m-1}\Big)^{-1} + \frac{1}{2}\sum_{i=1}^{m-1} n_i - \Big(\prod_{i=1}^{m-1} n_i\Big)^{-1}\Big(1 + \sum_{j=1}^{m-1}\prod_{i=1}^{j} n_i\Big).$$

Дифференцируя по n_κ, $\kappa=\overline{1,m-1}$, имеем:

$$\frac{\partial\varphi}{\partial n_\kappa} = \Big(\frac{N}{2}-1\Big)\frac{1}{\prod\limits_{i=1}^{\kappa-1} n_i \prod\limits_{i=\kappa+1}^{m-1} n_i}\Big(-\frac{1}{n_\kappa^2}\Big) + \frac{1}{2} - \frac{1}{\Big(\prod\limits_{i=1}^{m-1} n_i\Big)^2}\Big[\prod_{i=1}^{m-1} n_i \sum_{j=1}^{m-1}\frac{\partial}{\partial n_\kappa}\prod_{i=1}^{j} n_i -$$

$$- \prod_{i=1}^{\kappa-1} n_i \prod_{i=\kappa+1}^{m-1} n_i \sum_{j=1}^{m-1}\prod_{i=1}^{j} n_i\Big] = \Big(\frac{N}{2}-1\Big)\frac{1}{\prod\limits_{i=1}^{m-1} n_i}\Big(-\frac{1}{n_\kappa}\Big) + \frac{1}{2} - \frac{1}{\prod\limits_{i=1}^{m-1} n_i}\Big[\frac{1}{n_\kappa}\sum_{j=\kappa}^{m-1}\prod_{i=1}^{j} n_i -$$

$$- \frac{1}{n_\kappa}\sum_{j=1}^{m-1}\prod_{i=1}^{j} n_i\Big] = \Big(1-\frac{N}{2}\Big)\frac{n_0}{n_\kappa}\frac{1}{N} + \frac{1}{2} + \frac{n_0}{n_\kappa}\frac{1}{N}\sum_{j=1}^{\kappa-1}\prod_{i=1}^{j} n_i\,.$$

Решая уравнение $\partial\varphi/\partial n_\kappa = 0$ относительно n_κ, $\kappa=\overline{1,m-1}$, получим (2.I). Так как

$$\frac{\partial^2\varphi}{\partial n_\kappa^2} = \frac{n_0}{n_\kappa^2}\frac{1}{N}\Big(\frac{N}{2} - 1 - \sum_{j=1}^{\kappa-1}\prod_{i=1}^{j} n_i\Big) > 0\,,$$

то обследование дерева, построенного в соответствии с правилом (2.I), обеспечивается при наименьшем количестве проверок. Утверждение доказано.

Из рекуррентного соотношения (2.I) видно, что в оптимальном дереве с увеличением номера уровня размер пучка редуцируется. Уменьшение размера пучка K -го уровня по отношению к размеру пучка нулевого уровня будем характеризовать вектором редукции $\vec{R}_K$, который направлен от корня дерева к его нижнему уровню и модуль которого

$$R_K = n_0 - n_K = \Big(\sum_{j=0}^{K-1} \prod_{i=0}^{j} n_i\Big)\frac{2}{N}$$

является монотонно возрастающей функцией от номера уровня.

Строго говоря, следовало бы на основе (2.I) получить зависимость $n_K = f(n_0)$ и после подстановки в (2.2) исследовать функцию $\varphi(n_0)$ на экстремум. Но реализация этого плана затруднительна, т.к. зависимость $n_K = f(n_0)$ имеет сложный характер. Поэтому рассмотрим предварительно некоторые свойства дерева и на их основе найдем оптимальное значение параметра n_0.

Оценим величину вектора редукции размера нижнего пучка дерева

$$R_{m-1} = n_0 - n_{m-1} = \frac{2}{N}\sum_{j=0}^{m-1}\prod_{i=0}^{j} n_i$$

Имеет место следующее неравенство: $\sum_{j=0}^{m-2}\prod_{i=0}^{j} n_i < N$, - так как при $n_i \geqslant 2$

$$\prod_{i=0}^{m-1} n_i - \sum_{j=0}^{m-2}\prod_{i=0}^{j} n_i = n_0 \Big\{ n_1 \big\{ n_2 \ldots \big[n_{m-2}(n_{m-1}-1)-1 \big] \ldots -1 \big\} -1 \Big\} > 0 .$$

Поэтому $R_{m-1} < 2$, т.е. в оптимальном дереве размеры верхнего и нижнего пучков отличаются не больше, чем на два узла. Следовательно, при построении дерева, структура которого близка к оптимальной, необходимо использовать размеры пучков из трехэлементного множества $M\{n_0, n_0-1, n_0-2\}$.

Так как модуль вектора R_K, а, следовательно, и размер пучка n_K, являются, вообще говоря, дробными величинами, то строгое соблюдение (2.I) при построении дерева не представляется возможным. Реально могут иметь место лишь целочисленные n_K. Минимальный размер пучка, реализацию которого можно обеспечить, не дискредитируя самой идеи древовидной структуры, равен 2, максимальный

- равен N, при этом дерево вырождается в один пучок. Пусть в дереве имеется S пучков размера (n_0-1), t пучков размера (n_0-2) и $(m-s-t)$ пучков размера n_0. В этом случае

$$N = n_0^{m-s-t}(n_0-1)^s(n_0-2)^t$$

и после подстановки соответствующих n_K в (2.2) получим

$$\varphi = \frac{s+t}{2} n_0 - \frac{t}{2} + \frac{n_0+1}{2\ln n_0} \ln\frac{N}{(n_0-1)^s(n_0-2)^t} + \frac{1}{(n_0-1)^{s+1}(n_0-2)^{t+1}} + \frac{1}{N}\frac{n_0}{n_0-1} + \frac{1}{n_0-3}\Big[\frac{1}{(n_0-2)^{t+1}} - 1\Big].$$

Так как производная

$$\frac{\partial\varphi}{\partial n_0} = \frac{s+t}{2} + \frac{\ln n_0 - 1 - 1/n_0}{2\ln^2 n_0}\ln\frac{N}{(n_0-1)^s(n_0-2)^t} - \frac{n_0+1}{2\ln n_0}\Big(\frac{s}{n_0-1} + \frac{t}{n_0-2}\Big) - \frac{1}{(n_0-1)^{s+1}(n_0-2)^{t+1}}\Big(\frac{s+1}{n_0-1} + \frac{t+1}{n_0-2}\Big) - \frac{1}{N(n_0-1)^2} + \frac{1}{(n_0-3)^2}\Big[1 - \frac{2n_0-5+t(n_0-3)}{(n_0-2)^{t+2}}\Big]$$

зависит от N, то выбор оптимального n_0 определяется числом объектов, организуемых в дерево. Анализ уравнения $\partial\varphi/\partial n_0 = 0$ показывает, что зависимость $n_{0_{опт}} = F(N)$ является монотонно возрастающей функцией, асимптотически приближающейся к $n_0 \simeq 3{,}59$, т.к., если N достаточно велико $(N \to \infty)$, то корень уравнения $\frac{\partial\varphi}{\partial n_0} = 0$ определится из трансцендентного уравнения

$$\ln n_0 - 1 - 1/n_0 = 0$$

которое имеет единственное решение при $n_0 \simeq 3{,}592$.

В действительности модуль вектора редукции R_K может принимать только целочисленные значения (в том числе и нуль), поэтому $M\{4,3,2\}$, а параметры s, t, и m должны быть выбраны из условия минимума функции (2.2) при условии, что

$$L \leq N = 4^{m-s-t}\, 3^{s} \cdot 2^{t},$$

где L - число объектов, организованных в дерево.

Размеры пучков на уровнях дерева будут удовлетворять соотношению:

$$n_i = 4 - E\Big(\frac{i}{m-t}\Big) - E\Big(\frac{i}{m-s-t}\Big),$$

где

$$i = \overline{0, m-1},$$

$$E(i/g) = \begin{cases} 0, & \text{если } i < g, \\ 1, & \text{если } i \geqslant g. \end{cases}$$

Получим еще некоторые соотношения, позволяющие выяснить смысл требования редукции размеров пучков на последовательных уровнях оптимального дерева.

<u>Утверждение</u> 2. Энумератор для последовательности всевозможных значений длин поисковых треков объектных узлов наращиваемого сбалансированного дерева есть

$$F(t)=t\Big[1+\sum_{i=1}^{n_0-1}t^i+\sum_{k=2}^{m}\sum_{i=1}^{n_{k-1}-1}t^i\Big(\prod_{j=0}^{k-2}\sum_{i=1}^{n_j}t^i\Big)\Big]=\sum_{l=1}^{\lambda}c_l t^l, \qquad (2.3)$$

где t - свободный параметр, c_l - количество треков длины l,

$$\lambda=\sum_{l=0}^{m-1}n_l .$$

<u>Доказательство</u>: В рассматриваемом дереве длины треков изменяются от I, для объекта в корне дерева, до λ, для объекта с наибольшим поисковым путем. Для объектных узлов нулевого уровня (такой узел один - корень дерева) энумератор $F_0(t)=t$, т.е. имеется один трек - единичной длины. Для объектных узлов первого уровня энумератор будет

$$F_1(t)=t\sum_{i=1}^{n_0-1}t^i$$

т.е. имеется по одному треку длины от двух до n_0.

Энумератор для последовательности значений длин треков объектных узлов k-го уровня получим как разность энумераторов для всех узлов k-го уровня и для его признаковых узлов. Первая составляющая равна

$$F_{k1}(t)=t\prod_{j=0}^{k-1}\sum_{i=1}^{n_j}t^i .$$

Присутствие здесь сомножителя t объясняется тем, что по сравнению с позиционным деревом (I.4) и в отличие от него в наращиваемом дереве всякий трек начинается с проверки признака объекта, помещенного в корне дерева. Вторую компоненту определим с помощью энумератора всех узлов ($k-1$)-го уровня, т.к. количество признаковых узлов k-го уровня совпадает с числом всех узлов предыдущего уровня. Поэтому

$$F_{k2}(t)=\Big(t\prod_{j=0}^{k-2}\sum_{i=1}^{n_j}t^i\Big)t^{n_{k-1}} .$$

Здесь сомножитель $t^{n_{k-1}}$ объясняется тем, что длина трека признакового узла k-го уровня на n_{k-1} больше длины трека корневого узла пучка, которому принадлежит рассматриваемый признаковый узел. Таким образом, энумератор для совокупности значений длин треков объектных узлов k-го уровня будет

$$F_k(t)=t\prod_{j=0}^{k-2}\sum_{i=1}^{n_j}t^i\sum_{i=1}^{n_{k-1}-1}t^i .$$

Суммируя по всем уровням, получим (2.3)

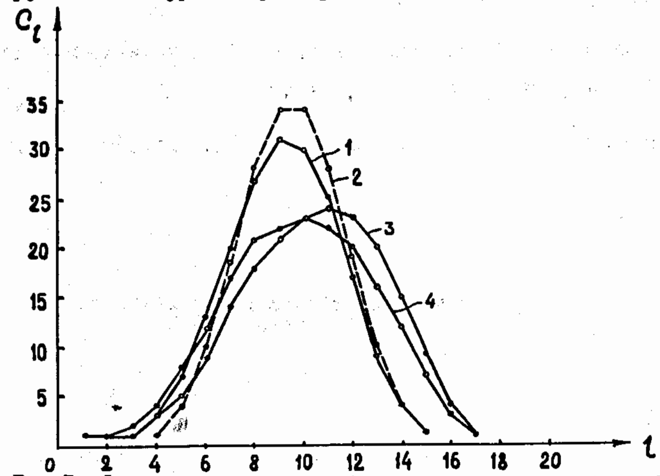

Рис.3. Функции распределения числа треков по их длинам в наращиваемых деревьях: I - <4,4,4,4,3>; 2 -<4,3,4,4,4>; 3 - <4,2,3,4,8>; 4 - <4,8,4,32>.

Соотношение (2.3) позволяет построить функцию распределения числа поисковых треков по их длинам для любого дерева. Анализ соответствующих зависимостей для равноемких деревьев показывает, что в редуцированном дереве (2.I) доступ к объектным узлам ускоряется за счет большего содержания треков малых длин. На рис.3 представлено семейство равноемких деревьев (N =192). Различие между кривыми I и 2 иллюстрирует эффект оптимальной модифицированной структуры перед деревом Ландауэра, а между кривыми 3 и 4 - влияние фазы вектора редукции на поисковые свойства дерева.

ЛИТЕРАТУРА

I. А.И.Китов. "Программирование экономических и управленческих задач". М., Советское радио, 1971.

2. W.I.Landauer.The Balanced Tree and its Utilization in Information Retrieval.-IEEE Transactions on EC,v.EC-12, 1963,December,N6,863-871.

3. В.М.Сумароков. "Оптимальные деревья для организации и поиска информации". УСиМ, 1974, № 3, 43-47.

4. Дж.Риордан. "Введение в комбинаторный анализ". М., ИЛ,1963.

Рукопись поступила в издательский отдел
31 июля 1975 г.