УДК 621.327

СПОСОБ БЫСТРОГО КОДИРОВАНИЯ ДВОИЧНЫХ ДАННЫХ

проф. А.В. Королёв

Выводятся выражения для быстрого вычисления весовых коэффициентов при кодировании двоичных последовательностей по числу серий. Проводится сравнительная оценка затрат количества операций для стандартного и быстрого способов кодирования.

Введение. Из анализа выражений, лежащих в основе компактного представления двоичных данных [1], следует, что наибольшее количество операций при вычислении значения кода по числу серий затрачивается для нахождения целочисленного ряда r_ξ, $\xi=\overline{1,n}$. При этом максимальное количество операций $\nu_{кчс}$, отводимых на получение кодового представления для факториальной схемы, равно

$$\nu_{кчс}= 2n^2+9n+7 \ . \qquad (1)$$

Выражение (1) позволило оценить удельные затраты машинных операций для определения r_ξ относительно других вычислений, которые составляют приблизительно $\dfrac{2n+6}{2n+9}\times 100\%$. Например, для $n=16$ удельные затраты достигают $93\ \%$. Поэтому необходимо разработать быстрый способ кодирования числа серий, основанный на учете структурных и статистических особенностей обрабатываемых данных.

Разработка быстрого способа кодирования двоичных данных. Сформулируем определение, которое будем использовать при разработке быстрого способа кодирования.

Определение. Пусть задана двоичная последовательность L_j, состоящая из n элементов. Назовем *сечением* этой последовательности подпоследовательность, содержащую $(n-\xi)$ элементов, начинающуюся с элемента $\ell_{j\xi}$ и заканчивающуюся элементом ℓ_{jn}.

Сформулируем утверждение, лежащее в основе разработки быстрого способа кодирования по числу серий.

Утверждение. Факториальную схему [1, 2] вычисления весовых коэффициентов r_ξ двоичных элементов $\ell_{j\xi}$ последовательности L_j длиной n элементов на каждом ее ξ -м сечении можно заменить системой рекурсивных выражений

$$r_{\xi} = \begin{cases} r_{\xi-1} \dfrac{\left(I_{1\xi}+1\right)}{\left(I_{0\xi}+1\right)}, \text{ если } \ell_{j,\xi-1} = \ell_{j,\xi}; \\ r_{\xi-1} \dfrac{\left(I_{2\xi}+1\right)}{\left(I_{0\xi}+1\right)}, \text{ если } \ell_{j,\xi-1} \neq \ell_{j,\xi}. \end{cases} \qquad (2)$$

В выражении (2) используются следующие обозначения:

Обозначение 1. Факториал $(n-\xi+1)! = I_{0\xi}$, где ξ – индекс сечения последовательности L_j.

Обозначение 2. Факториал $\left(n-\xi+1-\beta_{j\xi}\right)! = I_{1\xi}$, где $\beta_{j\xi}$ – число серий на ξ -м сечении последовательности L_j.

Обозначение 3. Факториал $\left(\beta_{j\xi}\right)! = I_{2\xi}$, зависящий от величины $\beta_{j\xi}$.

Доказательство. Рассмотрим случай, когда $\ell_{j,\xi-1} = \ell_{j,\xi}$. Тогда выполняется равенство $\beta_{j\xi} = \beta_{j,\xi-1}$ и, следовательно

$$r_{\xi} = \frac{(n-\xi+1)!}{\left(\beta_{j,\xi-1}\right)! \left(n-\xi+1-\beta_{j,\xi-1}\right)!}. \qquad (3)$$

Запишем выражение для весового коэффициента $r_{\xi-1}$ последовательности L_j на $(n-1)$ сечении

$$r_{\xi-1} = \frac{(n-(\xi-1)+1)!}{\left(\beta_{j,\xi-1}\right)! \left(n-(\xi-1)+1-\beta_{j,\xi-1}\right)!}. \qquad (4)$$

По определению факториала $(n-\xi+2)! = (n-\xi+1)!\,(n-\xi+2)$. С учетом этого разобьем выражение (4) на два сомножителя

$$r_{\xi-1} = \frac{(n-\xi+1)!}{\left(\beta_{j,\xi-1}\right)! \left(n-\xi+1-\beta_{j,\xi-1}\right)!} \, \frac{(n-\xi+2)}{\left(n-\xi+2-\beta_{j,\xi-1}\right)}. \qquad (5)$$

Замечая, что первый сомножитель (5) равен r_{ξ}, получим

$$r_{\xi-1} = r_{\xi} \, \frac{(n-\xi+2)}{\left(n-\xi+2-\beta_{j,\xi-1}\right)}.$$

Откуда

$$r_\xi = r_{\xi-1} \frac{\left(n-\xi+2-\beta_{j,\xi-1}\right)}{\left(n-\xi+2\right)}.$$

С учетом введенных обозначений получим рекурсивную формулу для вычисления весовых коэффициентов r_ξ:

$$r_\xi = r_{\xi-1} \frac{\left(I_{1\xi}+1\right)}{\left(I_{0\xi}+1\right)}.$$

Таким образом, доказана первая часть выражения (2) для $\ell_{j,\xi-1} = \ell_{j,\xi}$.

Рассмотрим вариант, когда $\ell_{j,\xi-1} \neq \ell_{j,\xi}$. В этом случае получим, что $\beta_{j\xi} = \beta_{j,\ \xi-1} - 1$. Тогда значение величины r_ξ равно

$$r_\xi = \frac{\left(n-\xi+1\right)!}{\left(\beta_{j,\xi-1}-1\right)!\left(n-\ \xi+1-\left(\beta_{j,\xi-1}-1\right)\right)!}. \tag{6}$$

Разделим числитель и знаменатель выражения (6) на $\beta_{j,\xi-1}$ и $\left(n-\xi+2\right)$:

$$r_\xi = \frac{\left(n-\xi+2\right)!}{\left(\beta_{j,\xi-1}-1\right)!\left(n-\ \xi+2-\beta_{j\xi}\right)!} \frac{\beta_{j,\xi-1}}{\left(n-\xi+2\right)}. \tag{7}$$

Из сравнения (3) и (7) следует, что первый множитель выражения (7) равен $r_{\xi-1}$. Поэтому с учетом введенных обозначений формула (7) примет вид

$$r_\xi = r_{\xi-1} \frac{\left(I_{2\xi}+1\right)}{\left(I_{0\xi}+1\right)},$$

т.е. доказана вторая часть выражения (2) для случая $\ell_{j,\xi-1} \neq \ell_{j,\xi}$.

Утверждение доказано полностью.

Из доказанного утверждения следует, что для вычисления величины r_ξ на ξ-м сечении надо знать значение $r_{\xi-1}$ на ($\xi-1$)-м сечении и параметры $I_{0\xi}$, $I_{1\xi}$, $I_{2\xi}$. Начальные значения этих параметров ($\xi = 0$) равны:

$$I_{00} = \left(n+1\right)!;\quad I_{10} = \left(n+1-2\vartheta\right)!;\quad I_{20} = \left(2\vartheta\right)!;\quad r_0 = \frac{I_{00}}{I_{10}\,I_{20}}.$$

Определим количество операций ν_{r_0}, необходимых для вычисления начального значения величины r_0. Для нахождения отношения $\frac{I_{00}}{I_{20}}$ затрачивается $(n-2\vartheta)$ операции умножения. Для вычисления параметра I_{10} также необходимо выполнить $(n-2\vartheta)$ операции умножения. Тогда с учетом одной операции деления величины $\dfrac{I_{00}}{I_{20}}$ на I_{10} количество операций ν_{r_0} равно

$$\nu_{r_0} = 2(n-2\vartheta) + 1. \tag{8}$$

Из анализа (8) следует, что наибольшее количество операций на вычисление величины r_0 для числа серий, равного $\vartheta=0$, будет порядка

$$\nu_{r_0} = 2n.$$

Чтобы вычислить значение одной величины r_ξ, $\xi=\overline{1,n}$ для формирования одного кода по выражению (2), необходимо затратить 4 операции умножения и сложения и одну операцию сравнения. Тогда максимальные затраты операций для нахождения всех r_ξ, $\xi=\overline{0,n}$ равны

$$\nu_{r_\xi} = (2n+5n)=7n. \tag{9}$$

Из (8) следует, что количество операций на получение начальных параметров составляет $30\ \%$ от общего количества операций, затрачиваемых на формирование кодового представления двоичных данных.

С целью уменьшения числа операций для определения начальных параметров рекурсии (2) предлагается использовать табличный способ нахождения величины r_0. На вычисление всех величин r_ξ, $\xi=\overline{0,n}$ с учетом табличного способа нахождения параметра r_0 затрачивается ν_{2n} машинных операций

$$\nu_{r_\xi} = \frac{n}{2} + 5n. \tag{10}$$

Сравнивая (9) и (10) делаем вывод, что количество операций с учетом табличного определения начальных параметров в 1.4 раза меньше, чем для факториальной схемы вычисления начальных параметров.

Для кодирования двоичных данных по числу серий необходимо вычислять для ξ-го элемента величину p_ξ, равную [1]:

$$p_\xi = r_{\xi-1} - r_\xi. \tag{11}$$

Поэтому для быстрого кодирования двоичных данных требуется преобразовать выражение (11) в соответствии с рекурсивной формулой (2). Для этого подставим выражение (2) для $r_{\xi-1}, r_\xi$ при $\ell_{j,\xi-1}=\ell_{j,\xi}$ в (8) и получим

$$p_{\xi} = r_{\xi-1} - r_{\xi-1}\left(\frac{I_{1\xi}+1}{I_{0\xi}+1}\right) = r_{\xi-1}\left(\frac{I_{0\xi}-I_{1\xi}}{I_{0\xi}+1}\right). \qquad (12)$$

Для случая $\ell_{j,\xi-1} \neq \ell_{j,\xi}$ значение p_{ξ} равно

$$p_{\xi} = r_{\xi-1} - r_{\xi-1}\left(\frac{I_{2\xi}+1}{I_{0\xi}+1}\right) = r_{\xi-1}\left(\frac{I_{0\xi}-I_{2\xi}}{I_{0\xi}+1}\right). \qquad (13)$$

Объединим теперь выражения (12) и (13) в одно

$$p_{\xi} = \begin{cases} r_{\xi-1}\left(\dfrac{I_{0\xi}-I_{1\xi}}{I_{0\xi}+1}\right), \text{ если } \ell_{j,\xi-1} = \ell_{j,\xi}; \\ r_{\xi-1}\left(\dfrac{I_{0\xi}-I_{2\xi}}{I_{0\xi}+1}\right), \text{ если } \ell_{j,\xi-1} \neq \ell_{j,\xi}. \end{cases} \qquad (14)$$

Выражение (14) позволяет осуществлять быстрое кодирование двоичных данных по числу серий с учетом рекурсивного способа вычисления весовых коэффициентов r_{ξ}.

Сравнительная оценка количества операций для факториальной и рекурсивной схем вычисления коэффициентов r_{ξ}. Из анализа выражения (14) и с учетом формул (9), (10) следует, что максимальное количество операций, необходимых для вычисления кодового представления последовательности двоичных данных, равно

$$\nu_p = \frac{13n}{2} - 1. \qquad (15)$$

Проведем сравнение затрат машинных операций $\nu_{кчс}$ и ν_p соответственно для факториального и рекурсивного способов вычисления кодов по числу серий в зависимости от количества двоичных элементов n и числа ϑ двоичных серий (рис. 1). Расчеты будем проводить по формулам (1) и (15) для $2 \leq n \leq 20$.

Из анализа графиков на рис. 1 следует, что:

1. За счет рекурсивного нахождения величин r_{ξ} обеспечивается уменьшение количества операций в среднем в 6 раз по сравнению с количеством операций, затрачиваемых при факториальной схеме нахождения величин r_{ξ}.

2. В общем случае выигрыш рекурсивной обработки относительно

факториальной обработки равен $\left(\frac{2n^2+9n+7}{6{,}5n-1}\right)$ раз.

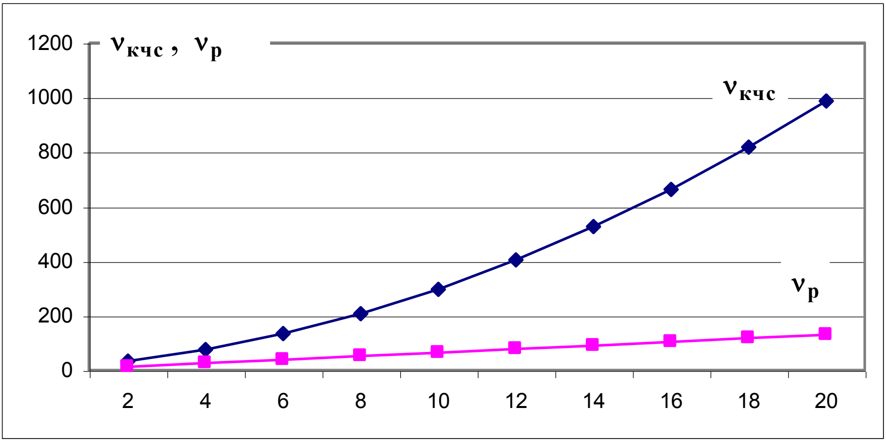

Рис. 1. График зависимости количества операций $\nu_{кчс}$ и ν_p от n

Выводы. Из рассмотренного материала можно сделать следующие выводы:

1. Разработан быстрый способ кодирования двоичных данных, основанный на переходе от факториальной схемы вычисления весовых коэффициентов r_ξ к рекурсивной.

2. В общем случае выигрыш рекурсивной обработки относительно факториальной обработки равен $\left(\frac{2n^2+9n+7}{6{,}5n-1}\right)$ раз. Это дает возможность обрабатывать изображения в реальном масштабе времени.

ЛИТЕРАТУРА

1. *Королёв А.В., Баранник В.В. Оценка количества информации изображения по числу серий одинаковых элементов // Системи обробки інформації. – Х.: НАНУ, ПАНМ, ХВУ. – 2002. – Вип. 2(18). – С. 43 – 46.*
2. *Корн Г. Корн Т. Справочник по математике. – М.: Наука, 1973. – 831 с.*

Поступила 10.09.2002

***КОРОЛЁВ Анатолий Викторович**, канд. техн. наук, профессор, профессор кафедры ХВУ. В 1969 году окончил Харьковское ВКИУ. Область научных интересов – обработка и передача информации.*