объема выборки может потребоваться, для обеспечения вычислимости C_n^k, увеличивать значение S или r, а возможно S и r одновременно, причем увеличение S более эффективно по затратам времени на вычисление.

Сущность алгоритма состоит в следующем:

1. Ввод данных n и k.

2. Установка выбранных значений S и r.

3. Проверка ограничительных условий $2 \le S \le M$; $0 \le r < M$; $n < M$; $k < M$ и сброс на нуль значений $l_j, j=0,r$.

4. Проверка ситуации $k > n$. Если она имеет место, то расчет считается оконченным. В противном случае - переход к п.5.

5. Присвоение значения $l_0=1$, вычисление параметра m, проверка условия $m < l$. Если оно имеет место, то расчет считается оконченным. В противном случае - переход к п.6.

6. Проверка вычислимости C_i при заданных значениях M,n,r,S . При обнаружении невычислимости процесс прерывается, что свидетельствует о необходимости увеличения заданных значений S и r. В противном случае - переход к п.7.

7. Выполнение m циклов последовательного приближения значения C_n^k по формуле (3). В каждом из циклов получается очередное значение C_i в форме (1) с целочисленными значениями разрядных коэффициентов l_j.

Для задачи кодирования сообщений в системах передачи информации предлагаемый алгоритм представляет интерес в случаях передачи слов большой длины. К его достоинствам относятся возможность вычисления целочисленных значений C_n^k при больших значениях n (порядка 1000) с представлением результата в позиционной системе счисления с основанием $S \le M$. Кроме того, предложенный алгоритм удобно принять за основу при разработке аппаратного вычислителя, создание и применение которого может оказаться целесообразным для сокращения времени вычислений.

SUMMARY

The precise algorithm, computing the binomial coefficient for the large size of the general sampling is proposed in this article. The efficiency of computing procedure is increased due to the possibility of using the positional systems with random radix. It is convinient to put practice the proposed algorithm.

СПИСОК ЛИТЕРАТУРЫ

1. Борисенко А.А., Губарев С.И., Куно Г.В. Биномиальные системы счисления с двоичным алфавитом.-АСУ и приборы автоматики: Респ. межвед. научн.-техн.сб.- Харьков: Вища школа, ХГУ, 1985, вып.76, с. 78-92.

2. Липский В. Комбинаторика для программистов.-М.:Мир,1988.

Поступила в редколлегию 7 февраля 1996 г.

УДК 621.391

ОБ ИЗБЫТОЧНОСТИ АДРЕСНО-ВЕКТОРНОГО КОДИРОВАНИЯ

Кулик И.А., асс.

С точки зрения сжатия показателем качества кодирования является его избыточность, представляющая собой разность между стоимостью кода и энтропией кодируемого источника информации [1].

В статье [2] рассматривался вопрос о рациональном кодировании двоичных последовательностей, который заключается в переходе от векторного к адресному методу кодирования в зависимости от числа k логических единиц. Такой переход осуществляется согласно неравенству

$$\frac{n}{\lceil \log_2 n \rceil} > k > n - \frac{n}{\lceil \log_2 n \rceil}, \qquad (1)$$

при выполнении которого двоичная последовательность кодируется адресами k логических единиц (или нулей) или, в противном случае, остается без изменений. Кодирующий метод и получаемый в результате код будем называть адресно-векторным.

Утверждение 1. Пусть S_k-источник класса эквивалентности по числу k логических единиц n-разрядной двоичной записи генерируемого символа, k=0,...,n. Относительная избыточность адресно-векторного кодирования такого источника

$$r \le 1 - \frac{\log_2 C_n^{\alpha}}{n}, \qquad (2)$$

где $\alpha = \left\lfloor \frac{n}{\lceil \log_2 n \rceil} \right\rfloor$.

Доказательство. Известно[2], что относительная избыточнность векторного r_v и адресного r_a методов кодирования соответственно имеют вид:

$$r_v = 1 - \frac{\log_2 C_n^k}{n}, \qquad (3)$$

$$r_a = \begin{cases} 1 - \dfrac{\log_2 C_n^k}{k\lceil \log_2 n \rceil}, & 0<k\le \left\lceil \dfrac{n}{2} \right\rceil \\ 1 - \dfrac{\log_2 C_n^k}{(n-k)\lceil \log_2 n \rceil}, & \left\lceil \dfrac{n}{2} \right\rceil <k<n. \end{cases} \qquad (4)$$

Как уже отмечалось, адресно-векторный метод кодирования включает процедуру выбора между кодированием адресами и векторным кодированием. Критерием выбора является меньшее значение избыточности, зависящее от значения k, а условием выбора для источника одного эквивалентного класса -

$$\begin{cases} 0<k<\alpha-1 \\ n-\alpha+1<k<n \end{cases} \qquad (5)$$

Очевидно из неравенства (5) следует, что для граничных значений к=α и k=n-α значение избыточности для рассматриваемого метода будет максимальным. Таким образом, учитывая $C_n^{\alpha}=C_n^{n-\alpha}$, из выражений (3,4) следует

$$r_{av} \le 1 - \frac{\log_2 C_n^{\alpha}}{n},$$

что и требовалось доказать.

В соответствии с (3,4) и условием рационального кодирования (5) относительная избыточность источника S_k

$$r_{av}=\begin{cases}1-\dfrac{\log_2 C_n^k}{k\lceil\log_2 n\rceil}, & 0<k<\alpha-1\\[2ex] 1-\dfrac{\log_2 C_n^k}{n}, & \alpha-1\le k\le n-\alpha+1\\[2ex] 1-\dfrac{\log_2 C_n^k}{(n-k)\lceil\log_2 n\rceil}, & n-\alpha+1<k<n.\end{cases} \qquad (6)$$

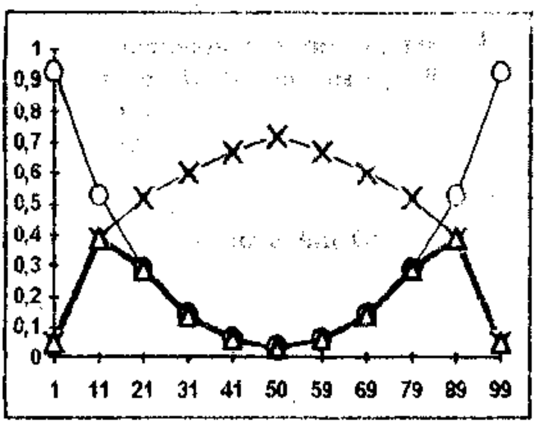

Рис.1

На рис.1 представлены при n=100:

Δ-график зависимости $r_{av}=f(k)$ для адресно-векторного метода кодирования;

о-график зависимости $r_v=f(k)$ для векторного кодирования;

x-график зависимости $r_\alpha=f(k)$ для адресного кодирования.

Кодирование отдельного класса эквивалентности представляет собой частный случай общей задачи, на практике наиболее встречающейся, когда кодируется источник, объединяющий несколько эквивалентных классов. Пусть $S=\{\alpha_i;|\alpha_i|=n,\ i=\overline{1,2^n}\}$ - информационный бернуллиевский источник, порождающий множество слов длины n, состоящих из букв алфавита B={0,1}. Символы 0 и 1 появляются независимо друг от друга с вероятностями p и 1-p соответственно. Для множества S рассмотрим разбиение на (n+1) классов по числу k логических единиц.

В результате адресно-векторного кодирования $f_{\alpha v}$ длины кодовых элементов из разных классов будут иметь разную длину, зависимую от значения k. Поэтому в целях дешифруемости кода, представляющего собой конкатенацию таких кодовых элементов, необходимо использовать служебное слово, несущее информацию о числе единиц в исходном векторе. Тогда с учетом служебного слова условие рационального кодирования и выражения для числа I разрядов адресно-векторного кода примет вид [2]:

$$\begin{cases}0\le k<\alpha-1\\ n-\alpha+1<k\le n,\end{cases} \qquad (7)$$

$$I=\begin{cases}(k+1)\lceil\log_2 n\rceil, & 0\le k<\alpha-1\\ n+\lceil\log_2 n\rceil, & \alpha-1\le k\le n-\alpha+1\\ (n-k+1)\lceil\log_2 n\rceil, & n-\alpha+1<k\le n.\end{cases} \qquad (8)$$

Таким образом, для стоимости адресно-векторного кодирования $f_{\alpha v}$ бернуллиевского источника S можно записать

$$C(f_{\alpha v},S) = \sum_{k=0}^{n} P_k l = \sum_{k=0}^{\alpha-2} P_k(k+1)\lceil \log_2 n \rceil + \sum_{k=\alpha-1}^{n-\alpha+1} P_k(n+\lceil \log_2 n \rceil) + \sum_{k=n-\alpha+2}^{n} P_k(n-k+1)\lceil \log_2 n \rceil, \tag{9}$$

где $P_k = C_n^k p^k (1-p)^{(n-k)}$ -вероятность появления двоичного слова a_i с числом единиц k или вероятность появления признака эквивалентности. Рассматривая энтропию H(S) источника Бернулли как сумму энтропий источников класса и признака эквивалентности [3], избыточность адресно-векторного кодирования будет равна

$$R(f_{\alpha v},S) = C(f_{\alpha v},S) - H(S) = \sum_{k=0}^{\alpha-2} P_k(k+1)\lceil \log_2 n \rceil + \sum_{k=\alpha-1}^{n-\alpha+1} P_k(n+\lceil \log_2 n \rceil) + \sum_{k=n-\alpha+2}^{n} P_k(n-k+1)\lceil \log_2 n \rceil - \left(\sum_{k=0}^{n} P_k \log_2 \frac{1}{P_k} + \sum_{k=0}^{n} P_k \log_2 C_n^k \right). \tag{10}$$

Аналитические выражения (7,10) определяют математическую модель и алгоритм работы адресно-векторного кодирующего устройства, используя которые можно провести анализ и оценку получаемого кода с точки зрения степени сжатия и сложностных характеристик алгоритма.

SUMMARY

The redundancy of the address-vector coding based on the reciprocal transitions from the vector coding method to the address coding one is analysed in this article. The two cases are under review: the first is when the infomation source generates binary chains with the equal number of the units, the second is when the number of the units is varied. The computed analytic expression determines the mathematical model of the address-vector coding device.

СПИСОК ЛИТЕРАТУРЫ

1. Кричевский Р.Е. Сжатие и поиск информации.-Москва: Радио и связь.- 1989-168 с.
2. Соловей В.А., Кулик И.А., Арбузов В.В. Алгоритм быстрого сжатия двоичных последовательностей // Вестн. Сумс. гос. ун-та.- 1995.- № 2. - С.68-70.
3. Борисенко А.А. О разложении бернуллиевых источников //Вестн. Сум. гос. ун-та.- 1995.- №3.- С.57-59.

Поступила в редколлегию 7 февраля 1996 г.