Сжатие двоичной информации на основе биномиальной числовой функции

Алексей Борисенко
кафедра электроэнергетики
Сумский государственный университет
Сумы, Украина
5352008@ukr.net

Игорь Кулик
кафедра электроники и компьютерной техники
Сумский государственный университет
Сумы, Украина
i.kulyk@ekt.sumdu.edu.ua

Binary information compression on basis of binomial number function

Olexiy Borysenko
dept. of Electroenergetics
Sumy State University
Sumy, Ukraine
5352008@ukr.net

Igor Kulyk
dept. of Electronics and Computer Technics
Sumy State University
Sumy, Ukraine
i.kulyk@ekt.sumdu.edu.ua

***Аннотация*— Изложено обоснование разработки метода сжатия на основе биномиальной числовой функции. Рассмотренный метод сжатия использует биномиальные отображения, решающие задачи перечисления двоичных последовательностей и генерирования соответствующих им биномиальных чисел. В работе сформулированы теоремы, определяющие теоретические основы сжатия и восстановления на основе биномиальной числовой функции**

***Abstract*— Reasonability to developing data compression method on basis of binomial number function is stated. The considered compression method uses binomial mappings, which decides enumeration tasks for binary sequences and generation tasks for the corresponding to them binomial numbers. The theorems determining theoretic basics for binary data compression and recovery on basis of binomial number function are formulated in the paper**

***Ключевые слова*— сжатие информации; равновесный код; биномиальные отображения; двоичные биномиальные числа; биномиальня числовая функция**

***Keywords*— data compression; constant weight code; binomial mappings; binary binomial numbers; binomial number function**

I. Введение

Дальнейшим развитием биномиального сжимающего кодирования, использующего двоичные (n,k)-биномиальные числа X_j, является разработка метода сжатия двоичной информации на основе биномиальной числовой функции [1]:

$$F_j = \operatorname{dec} X_j = \sum_{i=1}^{r} x_i C_{n-i}^{k-q_i}, \qquad (1)$$

где $X_j = x_1 x_2 ... x_i ... x_r$, $x_i \in \{0,1\}$; $X_j \in X[n,k]$, $j = \overline{1,|X|}$; n и k – параметры двоичной (n,k)-биномиальной системы счисления;

$$q_i = \sum_{t=1}^{i-1} x_t, \; q_i \le k. \qquad (2)$$

Функция (2), будучи нумерационной, позволяет ставить в соответствие двоичным (n,k)-биномиальным числам X_j их десятичные эквиваленты – номера $D_j = \operatorname{Bin}(\operatorname{dec} X_j)$, которые представляются в двоичном виде.

Целесообразность разработки методов сжатия и восстановления на основе биномиальной числовой функции (1) определяется следующими утверждениями:

1) длина номеров $|D_j|$ удовлетворяет неравенствам $|D_j| < n$ и $|D_j| < L_{n,k}$, где $L_{n,k}$ – средняя длина двоичных биномиальных чисел X_j [2], что обеспечивает достаточно высокий коэффициент сжатия;

2) функциональность соответствий между множеством $X[n,k]$ двоичных биномиальных чисел X_j и множеством $Y[n,k]$ равновесных комбинаций Y_j с одной стороны и множеством $X[n,k]$ и множеством D двоичных номеров D_j с другой, что приводит к биективным биномиальным отображениям и обеспечивает взаимно однозначность кодирования и декодирования;

3) равномерность двоичных номеров D_j, что существенно снижает аппаратно-программные затраты при построении устройств (или программных модулей) сжатия и восстановления на основе биномиальной числовой функции;

4) распространенность кодовых комбинаций, в основе структуры которых лежат биномиальные числа (например, равновесные и квазиравновесные коды, комбинации с ограничениями на взаимное расположение нулей и единиц и т.д.), для представления данных в информационных системах [1, 3, 4].

Сжатие информации на основе биномиальной числовой функции (1) имеет такие положительные свойства, как достаточно высокий коэффициент сжатия, универсальность к типу сжимаемых двоичных данных, отсутствие информационных потерь при сжатии, невосприимчивость к изменению статистических характеристик источника информации.

Целью научной работы являются разработка теоретических основ сжатия и восстановления на основе биномиальной числовой функции.

В основе предлагаемого метода сжатия двоичной информации лежат биномиальные отображения вида $\varphi: X \to Y$ и φ^{-1}: $Y \to X$, где X – множество двоичных биномиальных чисел, а Y – множество двоичных последовательностей, удовлетворяющих заданному ограничению R_Y [3]. В работе рассматриваются двоичные n-разрядные последовательности, количество единиц которых постоянно и равно k, т.е. $R_Y = k$, где k может изменяться в диапазоне $0 < k < n$.

II. Теоретические основы сжатия на основе биномиальной числовой функции

Обоснованием сжатия $f_e : Y[n,k] \to D$ на основе биномиальной числовой функции (1) является нижеследующая теорема 1, которая приводится без доказательства. При этом условимся операцию декатенации обозначать символом вида "/".

Теорема 1. Всякой двоичной последовательности $Y_j = y_1 y_2 ... y_i ... y_n$, $Y_j \in Y[n,k]$, $j = \overline{1, C_n^k}$, из n разрядов y_i, сумма значений которых равна k, можно поставить в соответствие единственное целое положительное число $D_j \in D = \left\{ \mathrm{Bin}\left(0,...,C_n^k - 1\right) \right\}$ с помощью отображения f_e в два этапа:

1) переход от $Y_j = y_1 y_2 ... y_i ... y_n$ к двоичному (n,k)-биномиальному числу $X_j = x_1 x_2 ... x_i ... x_r$, $X_j \in X[n,k]$, $r < n$, с помощью функции $X_j = f_b(Y_j)$ вида:

$$X_j = x_1 x_2 ... x_i ... x_r = \begin{bmatrix} y_1 y_2 ... y_i ... y_{n-1} 0/00...0 \\ y_1 y_2 ... y_i ... y_{n-1} 1/11...1 \end{bmatrix};$$

2) вычисление номера $D_j = \mathrm{Bin}\left(\mathrm{dec}\, X_j\right)$ двоичного (n,k)-биномиального числа X_j в соответствии с числовой функцией (1).

Обоснованием восстановления $f_e^{-1} : D \to Y[n,k]$ двоичных последовательностей на основе биномиальной числовой функции (1) является теорема 2, которая также приводится без доказательства. При этом условимся операцию конкатенации обозначать символом вида "++".

Теорема 2. Всякому двоичному номеру $D_j \in D = \left\{ \mathrm{Bin}\left(0,...,C_n^k - 1\right) \right\}$ можно поставить во взаимно однозначное соответствие единственную двоичную последовательность $Y_j = y_1 y_2 ... y_i ... y_n$, $Y_j \in Y[n,k]$, $j = \overline{1, C_n^k}$, из n разрядов y_i, сумма значений которых равна k, с помощью отображения f_e^{-1} в два этапа:

1) вычисление значений разрядов $x_i \in \{0,1\}$ неравномерного (n,k)-биномиального числа $X_j = x_1 x_2 ... x_i ... x_r$, $X_j \in X[n,k]$, $r < n$, в соответствии с рекуррентным соотношением:

$$x_i = \mathrm{sign}\left(\mathrm{sign}\left(F(i) - \rho_i\right) + 1\right),$$

где $F(1) = F_j$, $F(i+1) = F(i) - x_i \rho_i$, $\rho_i = C_{n-i}^{k-q_i}$, $F_j = \mathrm{dec}\, D_j$;

2) переход от двоичного (n,k)-биномиального числа $X_j = x_1 x_2 ... x_i ... x_r$ к двоичной последовательности $Y_j = y_1 y_2 ... y_i ... y_n$ с помощью функции $Y_j = f_b^{-1}(X_j)$ вида:

$$Y_j = y_1 y_2 ... y_i ... y_n = \begin{bmatrix} x_1 x_2 ... x_i ... x_{r-1} 0{+}{+}11...1 \\ x_1 x_2 ... x_i ... x_{r-1} 1{+}{+}00...0 \end{bmatrix}.$$

III. Пример сжатия на основе биномиальной числовой функции

Рассмотрим сжатие двоичной 8-разрядной последовательности вида $Y_j = 10000100$, используя биномиальную числовую функцию (1).

1. Производится вначале вычисление числа k двоичных единиц в исходной 8-разрядной последовательности $Y_j = 10000100$:

$$k = \sum_{i=1}^{8} y_i = 2 .$$

Тем самым определяется класс равновесных комбинаций $Y_j \in Y[8,2]$, а также параметры $n = 8$ и $k = 2$ двоичной биномиальной системы счисления, которая генерирует соответствующее биномиальное число $X_j \in X[8,2]$.

2. Определяется в 8-разрядной равновесной комбинации $Y_j = 10000100$, имеющей $k = 2$ единиц, значение последнего разряда равное $y_8 = 0$.

3. Так как $y_8 = 0$, то соответствующее $(8,2)$-биномиальное число $X_j \in X[8,2]$ имеет вид

$$X_j = 10000100/00 = 100001 ,$$

т.е. от $Y_j = 10000100$ отбрасываются все нулевые разряды, начиная с $y_8 = 0$, до появления первой двоичной единицы $y_6 = 1$, которая будет представлять значение последнего разряда $x_6 = y_6 = 1$ $(8,2)$-биномиального числа $X_j = 100001$. При этом значения остальных разрядов остаются без изменений: $x_5 = y_5 = 0$, $x_4 = y_4 = 0$, $x_3 = y_3 = 0$, $x_2 = y_2 = 0$, $x_1 = y_1 = 1$.

4. Вычисляется количественный эквивалент $\operatorname{dec} X_j$ двоичного $(8,2)$-биномиального числа $X_j = 100001$ в соответствии с числовой функцией (1) и выражением (2):

$$\operatorname{dec}(100001) = \sum_{i=1}^{6} x_i C_{8-i}^{2-q_i} =$$

$$= C_{8-1}^{2-q_1} + C_{8-6}^{2-q_6} = C_7^2 + C_2^1 = 21 + 2 = 23 .$$

5. Выполняется преобразование количественного эквивалента $\operatorname{dec}(100001) = 23$ к его двоичному виду:

$$D_j = \operatorname{Bin}(\operatorname{dec}(100001)) = \operatorname{Bin} 23 = 10111 ,$$

тем самым получая искомый номер из множества $D = \{\operatorname{Bin}(0,1,...,27)\}$ – сжатый образ исходной 8-разрядной последовательности $Y_j = 10000100$.

Полученный двоичный номер можно использовать для хранения в памяти, передачи по каналу связи или вычислительной обработки. Таким образом, 8-разрядная последовательность $Y_j = 10000100$ преобразуется в двоичное $(8,2)$-биномиальное 6-разрядное число $X_j = 100001$ и двоичный 5-разрядный номер $D_j = 10111$, длины которых меньше соответственно на два и три разряда от исходной комбинации.

Результаты сжатия $f_e : Y[8,2] \to D$ для всех равновесных комбинаций при $n = 8$ и $k = 2$ на основе биномиальной числовой функции (1) приведены в табл. 1.

ТАБЛИЦА 1. СООТВЕТСТВИЕ КОДОВ ПРИ $n = 8$ И $k = 2$

№	Равновесный код *Y*[8,2]	Множество *X*[8,2]	Двоичный номер
0	00000011	000000	00000
1	00000101	0000010	00001
2	00000110	0000011	00010
3	00001001	0000100	00011
4	00001010	0000101	00100
5	00001100	000011	00101
6	00010001	0001000	00110
7	00010010	0001001	00111
8	00010100	000101	01000
9	00011000	00011	01001
10	00100001	0010000	01010
11	00100010	0010001	01011
12	00100100	001001	01100
13	00101000	00101	01101
14	00110000	0011	01110
15	01000001	0100000	01111
16	01000010	0100001	10000
17	01000100	010001	10001
18	01001000	01001	10010
19	01010000	0101	10011
20	01100000	011	10100
21	10000001	1000000	10101
22	10000010	1000001	10110
23	10000100	100001	10111
24	10001000	10001	11000
25	10010000	1001	11001
26	10100000	101	11010
27	11000000	11	11011
	$n = 8$	$L_{n,k} = 5,71$	$\mid D_j \mid = 5$

IV. Выводы

Теоретические положения, изложенные в работе, являются фундаментом для дальнейшего развития методов сжатия двоичной информации на основе биномиальной числовой функции. К положительным качествам предлагаемого в работе сжатия следует отнести достаточно высокий коэффициент сжатия, универсальность к типу сжимаемых двоичных данных, невосприимчивость к изменению статистических характеристик источника информации и устойчивость к возникновению цепочек ошибок при декодировании.

Література References

[1] А. А. Борисенко и И. А. Кулик, *Биномиальное кодирование,* Сумы: Изд-во СумГУ, 2010.

[2] И. А. Кулик, "О средней длине двоичных биномиальных чисел", *Вісник СумДУ, Серія "Технічні науки"*, № 12 (71), С. 106–112, 2004.

[3] В. А. Амелькин, *Перечислительные задачи серийных последовательностей,* Новосибирск: Изд. ИВМиМГ СО РАН, 2008.

[4] И. А. Кулик, Е. М. Скордина, С. В. Костель, "Генерирование кодов-сочетаний для решения информационных задач ИУС", *АСУ и приборы автоматики. Всеукраин. межведомст. сборник,* № 155, С. 15–23, 2011.