Модели сжатия и восстановления данных на основе двоичных биномиальных чисел

Борисенко А.А.[1], Кулик И.А.[2], Онориукпе Аджири[3]

[1]Проф., д.т.н., заведующий кафедры электроэнергетики, Сумский государственный университет
ул. Римского-Корсакова 2, г. Сумы, Украина, 5352008@ukr.net
[2]Доц., к.т.н., кафедра электроники и компьютерной техники, Сумский государственный университет
ул. Римского-Корсакова 2, г. Сумы, Украина, i.kulyk@ekt.sumdu.edu.ua
[3]Студент, кафедра электроники и компьютерной техники, Сумский государственный университет
ул. Римского-Корсакова 2, г. Сумы, Украина, electron@sumdu.edu.ua

***Аннотация* — Обоснована целесообразность разработки методов сжатия и восстановления данных на основе двоичных биномиальных чисел. При разработке указанных методов предлагается использовать биномиальные отображения, решающие задачи перечисления двоичных последовательностей и генерирования соответствующих им биномиальных чисел. В качестве исходных кодовых последовательностей на данном этапе в работе рассматриваются равновесные комбинации. В рамках построения биномиальных отображений разработаны модели процессов сжатия и восстановления на основе двоичных биномиальных чисел, которые обладают простотой реализации и высоким быстродействием.**

Ключевые слова: модели процессов сжатия и восстановления, биномиальные отображения, двоичные биномиальные числа.

Data compression and recovery models on basis of binary binomial numbers

Borysenko O.A.[1], Kulyk I.A.[2], Onoriukpe Adjiri[3]

[1]Prof., Head of Department of Electroenergetics, Sumy State University
Rimskogo-Korsakova str., 2, Sumy, Ukraine, 5352008@ukr.net
[2]Associate Prof., Department of Electronics and Computer Technics, Sumy State University
Rimskogo-Korsakova str., 2, Sumy, Ukraine, i.kulyk@ekt.sumdu.edu.ua
[3]Student, Department of Electronics and Computer Technics, Sumy State University
Rimskogo-Korsakova str., 2, Sumy, Ukraine, electron@sumdu.edu.ua

***Abstract* — Reasonability of data compression and recovery methods on basis of binary binomial numbers is grounded. When developing the pointed methods it proposes to use binomial mappings, which decides an enumeration task for binary sequences and an generation task for the corresponding to them binomial numbers. The constant weight code combinations are under review as initial ones at this stage in the paper. Within the framework of the binomial mappings models of compression and recovery processes on basis of binary binomial numbers are developed. They possess simplicity when realizing and high speed of operation.**

Keywords: models of data compression and recovery proccesses, binomial mappings, binary binomial numbers.

I. Введение

Обоснованием целесообразности разработки методов сжатия информации на основе двоичных биномиальных чисел, генерируемых биномиальными системами счисления [1], является:

1) неравномерность биномиальных чисел, длина r которых меньше длины n исходных сжимаемых кодовых комбинаций, что обеспечивает коэффициент сжатия больше единицы [1, 2];

2) функциональность соответствий между множествами двоичных биномиальных чисел и кодовсочетаний, что приводит к биективным биномиальным отображениям и обеспечивает взаимно однозначность кодирования и декодирования;

3) префиксность двоичных биномиальных чисел [1], что позволяет без дополнительных аппаратных (и/или программных) и временных затрат на разделители проводить сжимающее кодирование и восстановление двоичных последовательностей;

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 информационно-управляющих системах [2, 3].

Метод сжатия данных на основе двоичных биномиальных чисел имеют такие положительные свойства, как:

– высокое быстродействие, связанное с использованием простых операций;
– универсальность к типу сжимаемой двоичной информации;
– достаточно простую аппаратную и/или программную реализацию;
– отсутствие информационных потерь при сжатии.

Кроме того, важным достоинством метода является то, что сжимаемые последовательности наделяются числовыми характеристиками, которые выражаются соответствующими им биномиальными числами.

Целями научной работы являются разработка:

1) теоретических основ сжатия на основе двоичных биномиальных чисел;

2) моделей сжатия и восстановления двоичной информации, которые отличались бы быстродействием и простотой практической реализации.

В основе метода сжатия на основе двоичных биномиальных чисел лежат биномиальные отображения вида φ: $X \to Y$ 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 двоичных биномиальных чисел

Реализация биномиального отображения φ^{-1}:$Y \to X$ в рамках построения математической модели перечисления исходных двоичных n-разрядных последовательностей $Y_j \in Y$ с заданным ограничением $R_Y = k$ представляет собой сжатие

$$f_b : Y\left[n,k\right] \to X\left[n,k\right] \qquad (1)$$

равновесных комбинаций $Y_j \in Y[n,k]$ на основе двоичных (n,k)-биномиальных чисел $X_j \in X[n,k]$. В свою очередь, реализация биномиального отображения φ: $X \to Y$ в рамках построения математической модели генерирования для $Y_j \in Y$ с заданным ограничением $R_Y = k$ означает восстановление

$$f_b^{-1} : X\left[n,k\right] \to Y\left[n,k\right] \qquad (2)$$

исходных равновесных комбинаций $Y_j \in Y[n,k]$ на основе двоичных (n,k)-биномиальных чисел $X_j \in X[n,k]$, где $Y[n,k]$ – множество равновесных n-разрядных комбинаций Y_j, имеющих k единиц, а $X[n,k]$ – множество двоичных (n,k)-биномиальных чисел, соответствующих Y_j.

Нижеследующие теоремы 1, 2 и 3, которые приводятся без доказательства, отражают свойства отображений (1), (2) и способы их практической реализации. Условимся операцию декатенации далее обозначать символом вида "/".

Теорема 1. Всякой двоичной последовательности $Y_j = y_1y_2\ldots y_i\ldots y_n$, составленной из n разрядов y_i, сумма значений которых равна k, можно поставить в соответствие единственное двоичное (n,k)-биномиальное число $X_j = x_1x_2\ldots x_i\ldots x_r$ с помощью функции $X_j = f_b(Y_j)$ вида:

$$X_j = x_1x_2...x_i...x_r = \begin{bmatrix} y_1y_2...y_i...y_{n-1}0/00...0 \\ y_1y_2...y_i...y_{n-1}1/11...1 \end{bmatrix}, \qquad (3)$$

где $Y_j \in Y[n,k]$, $X_j \in X[n,k]$, $r < n$, $j = 0,1,\ldots,C_n^k-1$.

Теорема 2. Всякому двоичному (n,k)-биномиальному числу $X_j = x_1x_2\ldots x_i\ldots x_r$ можно поставить в соответствие единственную двоичную последовательность $Y_j = y_1y_2\ldots y_i\ldots y_n$, составленную из n разрядов y_i, сумма значений которых равна k, с помощью функции $Y_j = f^{1}{}_b(X_j)$ вида:

$$Y_j = y_1y_2...y_i...y_n = \begin{bmatrix} x_1x_2...x_i...x_{r-1}0++11...1 \\ x_1x_2...x_i...x_{r-1}1++00...0 \end{bmatrix}, \qquad (4)$$

где $X_j \in X[n,k]$, $Y_j \in Y[n,k]$, $r < n$, $j = 0,1,\ldots,C_n^k-1$.

Следствие. Функция $Y_j = f^{1}{}_b(X_j)$ может также иметь вид:

$$Y_j = y_1y_2...y_i...y_n = \begin{bmatrix} x_1x_2...x_i...x_r ++11...1, \text{ если } 0 \le q < k \\ x_1x_2...x_i...x_r ++00...0, \text{ если } q = k\,. \end{bmatrix} \qquad (5)$$

Теорема 3. Отображение

$$f_b : Y\left[n,k\right] \to X\left[n,k\right]$$

является биективным.

Определение 1. *Методом сжатия f_b на основе двоичных (n,k)-биномиальных чисел (или биномиальным сжатием) называется отображение*

$$f_b : Y\left[n,k\right] \to X\left[n,k\right],$$

которое задается функцией вида

$$X_j = f_b\left(Y_j\right) = \begin{bmatrix} Y_j/00...0,\ y_n = 0 \\ Y_j/11...1,\ y_n = 1 \end{bmatrix} = x_1x_2...x_i...x_r, \qquad (6)$$

где $Y[n,k]$ − множество исходных двоичных n-разрядных последовательностей $Y_j = y_1y_2\ldots y_i\ldots y_n$, имеющих число k единиц, $Y_j \in Y[n,k]$:

$$Y\left[n,k\right] = \left\{ Y_j = y_1y_2...y_i...y_n \,/\, \sum_{i=1}^{n} y_i = k, y_i \in \left\{0,1\right\} \right\};$$

$X[n,k]$ − множество двоичных (n,k)-биномиальных чисел $X_j \in X[n,k]$:

$$X\left[n,k\right] = \left\{ X_j = x_1x_2...x_i...x_r \,/ \begin{bmatrix} \left(l = n-k\right) \wedge \left(x_r = 0\right) \\ \left(q = k\right) \wedge \left(x_r = 1\right) \end{bmatrix} \right\}.$$

III. Модели процессов сжатия и восстановления на основе двоичных биномиальных чисел

Моделирование сжатия f_b двоичных последовательностей $Y_j = y_1y_2\ldots y_i\ldots y_n$ на основе двоичных (n,k)-биномиальных чисел X_j, используя теорему 2 и функцию (3), состоит из следующих этапов.

Этап 1. Определяется в n-разрядной равновесной комбинации $Y_j = y_1y_2\ldots y_i\ldots y_n$, имеющей число k единиц, значение последнего разряда y_n.

Этап 2. Если $y_n = 0$, то

$$X_j = Y_j/00...0 = y_1y_2...y_i...y_{n-1}0/00...0 =$$
$$= x_1x_2...x_i...x_{r-1}1,$$

т.е. от комбинации $Y_j = y_1y_2\dots y_i\dots y_{n-1}0$ отбрасываются все нулевые разряды, начиная с $y_n = 0$, до появления первой двоичной единицы $y_r = 1$, которая будет представлять последний разряд $x_r = y_r = 1$ искомого (n,k)-биномиального числа $X_j = x_1x_2\dots x_i\dots x_{r-1}1$. В противном случае

$$X_j = Y_j / 11...1 = y_1y_2...y_i...y_{n-1}1/11...1 =$$
$$= x_1x_2...x_i...x_{r-1}0\,,$$

т.е. от комбинации $Y_j = y_1y_2\dots y_i\dots y_{n-1}1$ отбрасываются все единичные разряды, начиная с $y_n = 1$, до появления первого двоичного нуля $y_r = 0$, который будет представлять значение последнего разряда $x_r = y_r = 0$ искомого (n,k)-биномиального числа $X_j = x_1x_2\dots x_i\dots x_{r-1}0$. При этом в обоих случаях значения остальных разрядов остаются без изменений, т.е. $x_1 = y_1$, $x_2 = y_2, \dots, x_{r-1} = y_{r-1}$.

Введем теперь в рассмотрение операцию конкатенации, которую обозначим как "++". Данное действие является обратным по отношению к операции декатенации.

Моделирование восстановления $f^1{}_b$ двоичных последовательностей $Y_j = y_1y_2\dots y_i\dots y_n$ на основе двоичных (n,k)-биномиальных чисел X_j, используя теорему 2 и функцию (4), состоит из этапов.

Этап 1. Определяется в двоичном (n,k)-биномиальном r-разрядном числе $X_j = x_1x_2\dots x_i\dots x_r$, $X_j \in X[n,k]$, $r < n$, значение последнего разряда x_r.

Этап 2. Если $x_r = 0$, то

$$Y_j = X_j ++11...1 =$$
$$= x_1x_2...x_i...x_{r-1}0 ++ 11...1 = y_1y_2...y_i...y_{n-1}1\,,$$

т.е. к двоичному биномиальному числу $X_j = x_1x_2\dots x_i\dots x_{r-1}0$ присоединяются единичные разряды $11\dots1$: $y_{r+1} = y_{r+2} = \dots = y_n = 1$ так, чтобы общее количество разрядов искомой двоичной равновесной комбинации Y_j составило n, $Y_j \in Y[n,k]$. В противном случае

$$Y_j = X_j ++00...0 =$$
$$= x_1x_2...x_i...x_{r-1}1 ++ 00...0 = y_1y_2...y_i...y_{n-1}0\,,$$

т.е. к двоичному биномиальному числу $X_j = x_1x_2\dots x_i\dots x_{r-1}1$ присоединяются нулевые разряды $00\dots0$: $y_{r+1} = y_{r+2} = \dots = y_n = 0$ так, чтобы общее количество разрядов искомой двоичной равновесной комбинации Y_j составило n, $Y_j \in Y[n,k]$. При этом в обоих случаях значения остальных разрядов остаются без изменений: $y_1 = x_1$, $y_2 = x_2, \dots, y_{r-1} = x_{r-1}$.

В случаях, когда имеются ограничения на длину разрядной сетки для хранения текущего количества разрядов Y_j, то более выгодным является использование функции $Y_j = f^1{}_b(X_j)$ вида (5), которая требует вычисление $q < n$ или $l < n$.

Результаты отображения (1) и, следовательно, отображения (2) для случая $n = 8$ и $k = 2$ приведены в таблице 1, в которой затененные ячейки означают отбрасываемые разряды равновесных комбинаций Y_j в соответствии с (3) или прибавляемые разряды к двоичным биномиальным числам X_j в соответствии с (4), $j = 0,\dots, 27$.

Таблица 1 – Соответствие между Y_j и X_j при $n = 8$ и $k = 2$

№	Равновесный код *Y*[8,2]								Двоичные биномиальные числа *X*[8,2]						
0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	
4	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1
5	0	0	0	0	1	1	0	0	0	0	0	0	1	1	
6	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
9	0	0	0	1	1	0	0	0	0	0	0	1	1		
13	0	0	1	0	1	0	0	0	0	0	1	0	1		
14	0	0	1	1	0	0	0	0	0	0	1	1			
16	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1
19	0	1	0	1	0	0	0	0	0	1	0	1			
20	0	1	1	0	0	0	0	0	0	1	1				
21	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0
25	1	0	0	1	0	0	0	0	1	0	0	1			
27	1	1	0	0	0	0	0	0	1	1					

Для совокупности представленных в таблице 1 двоичных комбинаций $Y_j \in Y[8,2]$ коэффициент сжатия изменяется от 1,14 до 4, а средний коэффициент сжатия для всей комбинаций данной таблицы составляет приблизительно 1,78.

Выводы

Приведенные теоретические положения относительно преобразования данных с использованием двоичных биномиальных чисел являются фундаментом для разработки эффективных с точки зрения быстродействия методов сжатия и восстановления двоичных последовательностей. В качестве основы данных методов предлагаются разработанные модели сжатия и восстановления информация на основе двоичных биномиальных чисел. Указанные модели обладают простой практической реализацией и высоким быстродействием, что объясняется использованием простых вычислительных и строчных операций. Дальнейшее перспективное развитие методов сжатия и восстановления на основе двоичных биномиальных чисел предполагает его разработку для случая, когда значение числа единиц в двоичных последовательностях, принадлежащих сжимаемому массиву, является переменным.

[1] Борисенко А.А. Биномиальное кодирование: монография / А.А. Борисенко, И.А. Кулик. − Сумы: Изд-во СумГУ, 2010. – 206 с.

[2] Амелькин В.А. Перечислительные задачи серийных последовательностей / В.А. Амелькин. – Новосибирск: Изд. ИВМиМГ СО РАН, 2008. – 317 с.

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