

УДК 621.327:681.5

В.В. Баранник¹, А.В. Хаханова²¹Харьковский университет Воздушных Сил им. И. Кожедуба, Харьков²Харьковский национальный университет радиоэлектроники, Харьков

МЕТОД ОДНОМЕРНОГО ПЛАВАЮЩЕГО СТРУКТУРНОГО КОДИРОВАНИЯ ДВОИЧНЫХ ДАННЫХ

Разрабатывается метод кодирования одномерных плавающих структурных чисел в двоичном пространстве. При этом учитывается режим, когда код-номер формируется для переменного количества элементов, а количество разрядов под кодограмму отводится по равномерному принципу. Данный режим обеспечивает за счет использования правил отбора допустимых сочетаний, составленных из двоичных последовательностей. Для снижения времени кодирования данных процесс обработки реализуется на основе рекуррентных выражений. В этом случае весовые коэффициенты двоичных элементов определяются на основе значений весового коэффициента и элементов на предыдущем шаге обработки. Сжатие данных достигается за счет сокращения комбинаторной избыточности, обусловленной ограниченным количеством серий единиц.

Ключевые слова: одномерное плавающее структурное кодирование, рекуррентная обработка, избыточность.

Введение

Постановка проблемы и анализ литературы.

Особенностью функционирования современных распределенных информационно-вычислительных систем является обработка и передача больших объемов телеметрических и диагностических данных. Это приводит к увеличению времени доведения информации и к снижению ее достоверности [1; 2]. Поэтому совершенствование информационного обеспечения, состоящего в разработке методов компактного представления двоичных данных без внесения погрешности, является **актуальным** направлением научно-прикладных исследований [3 – 5].

Среди различных подходов относительно сокращения избыточности в двоичных данных наиболее эффективным является подход, базирующийся на устранении структурной и комбинаторной видов избыточности [4, 5]. В этом случае выявляются закономерности, которые присущи любым комбинациям двоичных элементов [4, 5]. Поэтому **цель статьи** состоит в создании метода компактного представления двоичных данных на основе одномерного плавающего структурного кодирования.

Изложение основного материала

Процесс формирования компактного представления одномерных структурных чисел, образованных на базе двоичных матриц, включающий в себя: определение размеров и позиционирования элементов плавающего структурного числа на основе рекуррентного вычисления количества допустимых двоичных последовательностей с ограниченным числом серий единиц в направлении столбцов матрицы знаков; рекуррентную систему вычисления кодовых номеров плавающих структурных чисел - организуется на основе выполнения следующих этапов:

Этап 1. Формирование структурного числа начинается с компонент первого столбца матрицы знаков. На начальном этапе $v=1$ анализируется элемент g_{11} . В предположении того, что $g_{01} = 0$ находится величина V_{1,η_1} , равная количеству двоичных комбинаций составленных из одного элемента с числом серий единиц равным η_1 :

$$V_{v,\eta} = V_{2,\eta_1} = (2) ! / (2\eta_1) ! (2 - 2\eta_1) !. \quad (1)$$

После чего проверяется неравенство

$$V_{2,\eta_1} \leq 2^M - 1. \quad (2)$$

Если неравенство (2) выполняется, то анализируется возможность добавить к текущему структурному числу элемент g_{21} .

Этап ξ . На ξ -м шаге обработки анализируется возможность добавления к текущему структурному числу, содержащему $(\xi-1)$ элементов с числом серий равным $\eta_{\xi-1}$, элемент $g_{\xi\gamma}$. При этом данный элемент может находиться на произвольной позиции в матрице знаков. Для проверки такой возможности проверяется неравенство:

– если $\eta_\xi = \eta_{\xi-1}$, то

$$V_{\xi-1,\eta_{\xi-1}} (\xi+1) / (\xi+1 - 2\eta_{\xi-1}) \leq 2^M - 1; \quad (3)$$

– если $\eta_\xi = \eta_{\xi-1} + 1$, то

$$V_{\xi-1,\eta_{\xi-1}} \frac{(\xi+1) (\xi - 2\eta_{\xi-1})}{(2\eta_{\xi-1} + 1) (2\eta_{\xi-1} + 2)} \leq 2^M - 1. \quad (4)$$

В случае, когда соответствующее неравенство (3) или (4) не выполняется, то:

1. Элемент $g_{\xi\gamma}$ считается первым элементом очередного структурного числа;

2. Для структурного числа, состоящего из $(\xi-1)$ элементов с числом серий равным $\eta_{\xi-1}$,

формируется значение кода $C_{\xi-1}$. Для этого используются следующие соотношения:

$$1) \text{ на начальном шаге кодирования} \\ \beta_{0,\ell} = 2\eta_{\xi-1}, g_{0,\ell} = 0 \quad \text{и} \\ V_{\xi-1, \eta_{\xi-1}} = ((\xi-1)+1)! / (\beta_{0,\ell})! ((\xi-1)+1-\beta_{0,\ell})!, \quad (5)$$

2) определение весового коэффициента $p_{1\ell}$ на первом шаге обработки для:

$$- |g_{0,\ell} - g_{1,\ell}| = 1 : \\ p_{1,\ell} = V_{\xi-1, \eta_{\xi-1}} ((\xi-1)+1-2\eta) / ((\xi-1)+1); \quad (6) \\ - |g_{0\ell} - g_{1\ell}| = 0 : \\ p_{1,\ell} = 2\eta_{\xi-1} V_{\xi-1, \eta_{\xi-1}} / ((\xi-1)+1). \quad (7)$$

3) формирование весовых коэффициентов на последующих шагах:

$$- |g_{k-2,\ell} - g_{k-1,\ell}| = 1 \text{ и } |g_{k-1,\ell} - g_{k\ell}| = 1 : \\ p_{k\ell} = p_{k-1,\ell} (\beta_{k-1,\ell} + 1) / ((\xi-1) - (k-1) + 1); \quad (8) \\ - |g_{k-2,\ell} - g_{k-1,\ell}| = 1 \text{ и } |g_{k-1,\ell} - g_{k\ell}| = 0 : \\ p_{k-1,\ell} = p_{k-1,\ell} \times \\ \times \left(\frac{(\beta_{k-1,\ell} + 1)\beta_{k,\ell}}{((\xi-1) - (k-2) - \beta_{k-1,\ell})(\xi-1) - k + 2} \right); \quad (9) \\ - |g_{k-2,\ell} - g_{k-1,\ell}| = 0 \text{ и } |g_{k-1,\ell} - g_{k\ell}| = 1 : \\ p_{k,\ell} = p_{k-1,\ell} \times \\ \times \left(\frac{((\xi-1) - k - \beta_{k-1,\ell} + 3)((\xi-1) - k + 2 - \beta_{k-1,\ell})}{(\beta_{k-1,\ell})(\xi-1) - k + 2} \right); \quad (10)$$

$- |g_{k-2,\ell} - g_{k-1,\ell}| = 0 \text{ и } |g_{k-1,\ell} - g_{k\ell}| = 0 :$
 $p_{k\ell} = p_{k-1,\ell} ((\xi-1) - k - \beta_{k-1,\ell} + 3) / ((\xi-1) - k + 2), \quad (11)$
 где $(\xi-1)$ – длина обрабатываемой двоичной последовательности.

Если соответствующее неравенство (3) или (4) выполняется, то элемент $g_{\xi\gamma}$ добавляется к текущему одномерному структурному числу и повторяется ξ -й этап обработки.

Процесс обработки завершается когда проанализирован $(n; n)$ -й элемент матрицы знаков.

Пример вычисления кода для ℓ -й двоичной последовательности $G^{(\ell)} = \{1; 0; 0; 0\}$, т.е. $g_{1,\ell} = 1; g_{2,\ell} = 0; g_{3,\ell} = 0; g_{4,\ell} = 0$. Длина кодограммы равна $M = 4$ бита. Проверяем неравенство (2) для $\eta_1 = 1$:

$$V_{2,1} = (2)! / (2)! (0)! = 1 \leq 2^4 - 1.$$

Неравенство (2) выполняется, то анализируется возможность добавить к текущему структурному числу элемент $g_{2,\ell}$. При этом $\eta_1 = \eta_2$, тогда величина $V_{3,1}$ определяется по формуле:

$$V_{3,1} = V_{2,1} (3) / (3-2) = 3 \leq 2^M - 1;$$

Поскольку неравенство (2) выполняется, то проверяется возможность добавления к текущему одномерному структурному числу элемента $g_{3,\ell}$. Для этого проверяем следующее условие, учитывая, что $\eta_2 = \eta_3$: $V_{4,1} = V_{3,1} (4) / (4-2) = 6 \leq 2^M - 1$;

Неравенство (2) выполняется. Следовательно, элемент $g_{3,\ell}$ принадлежит к текущему структурному числу. Анализируем последний элемент $g_{4,\ell}$. Так как $g_{4,\ell} = 0$, то $\eta_3 = \eta_4$. Поэтому проверяется неравенство

$$V_{5,1} = V_{4,1} (5) / (5-2) = 6 \times \frac{5}{3} = 10 \leq 2^M - 1;$$

В связи с тем, что неравенство выполняется, то одномерное структурное число состоит из четырех элементов. Проведем формирование кода-номера для построенного структурного числа.

Для заданной последовательности соответствуют следующие начальные параметры: длина двоичной последовательности $m = 4$; значение нулевого элемента по умолчанию равно $g_{0,j} = 0$; число серийных перепадов $2\eta = \beta_{0,\ell} = 2$.

Этап 1. Вычисление значения весового коэффициента $p_{1,\ell}$ на первом шаге $k=1$. Поскольку $|g_{0,\ell} - g_{1,\ell}| = 1$, то $p_{1,\ell} = V_{4,1} (v+1-2\eta) / (v+1)$, где $V_{4,1}$ – суммарное количество четырех элементов двоичных последовательностей с количеством серийных перепадов, равным $\eta=1$;

$$p_{1,\ell} = V_{4,1} (v+1-2\eta) / (v+1) = \\ = 10 (4+1-2) / (4+1) = 30/5 = 6.$$

Этап 2. Нахождение значения весового коэффициента $p_{2,\ell}$ на втором шаге $k=2$. Поскольку $|g_{0,\ell} - g_{1,\ell}| = 1$, а $|g_{1,\ell} - g_{2,\ell}| = 1$, то величина $p_{2,\ell}$ определяется по формуле для значения

$$\beta_{1,\ell} = \beta_{0,\ell} - |g_{0,\ell} - g_{1,\ell}| = 2-1=1 : \\ p_{2,\ell} = p_{1,\ell} (\beta_{1,\ell} + 1) / (v - (k-1) + 1) = \\ = 6(1+1) / (4 + (2-1) + 1) = 12 / 4 = 3.$$

Этап 3. Определение весового коэффициента $p_{3,\ell}$ на третьем шаге $k=3$. Поскольку $|g_{1,\ell} - g_{2,\ell}| = 1$, а $|g_{2,\ell} - g_{3,\ell}| = 0$, то величина $p_{3,\ell}$ определяется по формуле для $\beta_{2,\ell} = \beta_{1,\ell} - |g_{1,\ell} - g_{2,\ell}| = 1-1=0$:

$$p_{3,\ell} = p_{2,\ell} \left(\frac{(\beta_{2,\ell} + 1)\beta_{3,\ell}}{(v - (k-2) - \beta_{k-1,\ell})(v - k + 2)} \right) = \\ = 3 \times 0 / (4-1-0)(4-3+2) = 0.$$

Этап 4. Формирование значения кода в соответствии с формулой:

$$C_v = \sum_{k=1}^v \ell_{k\ell} = \sum_{k=1}^v g_{k\ell} P_{k\ell}.$$

Подставив в последнюю численные значения, можно получить следующую оценку кода матрицы:

$$C_v = \sum_{k=1}^v g_{k\ell} P_{k\ell} = 1 \times 6 + 0 \times 3 + 0 \times 0 + 0 \times 0 = 6.$$

Значит, создан метод сжатия на основе одномерного структурного кодирования двоичных последовательностей переменной длины. При этом исключаются возможности переполнения машинного слова.

Оценка степени сжатия

Значение коэффициента сжатия $k_{сж}$ для разработанного метода кодирования относительно исходного представления определяется на основе выражения $k_{сж} = v / \log_2 C_v$, где v – количество разрядов, отводимое на исходное представление двоичной последовательности (до выявления закономерностей).

Поскольку для заданных значений длины обрабатываемой последовательности m , вектора ограничений на число серий единиц значение кода номера C_v ограничено сверху значением величины $V_{v,\eta}$, то минимальное значение коэффициента сжатия k_{min} находится как $k_{min} = v / [\log_2 V_{v,\eta}] + 1$.

Оценка величин k_{min} и $k_{сж}$ показала, что для произвольной комбинации двоичных элементов они принимают значения соответственно **1,4** и **8 раз**.

Выводы

1. Создан метод компактного представления двоичных данных без внесения погрешности на основе кодирования одномерных плавающих структурных чисел. Данный метод базируется на:

– системе правил, обеспечивающих отбор двоичных элементов и позиционирование плавающих структурных чисел в массивах;

– систем рекуррентных выражений, позволяющих вычислить весовой коэффициент текущего элемента на основе известного значения весового коэффициента предыдущего элемента.

Это позволяет:

- 1) сократить количество комбинаторной и структурной избыточности, обусловленной ограниченным значением числа серий единиц;
- 2) снизить временные затраты на обработку;
- 3) уменьшить количество не значимых разрядов в кодограмме;
- 4) исключить потери информации из-за переполнения кодограммы.

2. Получены аналитические выражения для оценки максимальной и минимальной степени сжатия двоичных данных. Оценка эффективности разработанного кодирования показала, что степень сжатия изменяется в среднем от **1,4** до **8 раз**.

Список литературы

1. Кульгин М.Б. Технологии корпоративных сетей. – СПб.: Питер, 2000. – 704 с.
2. Прэтт У. Цифровая обработка изображений: Том 1, 2. – М.: Мир, 1985. – 736 с.
3. Ватолин В.И., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройства архиваторов, сжатие изображений и видео. – М.: ДИАЛОГ – МИФИ, 2002. – 384 с.
4. Королев 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.
5. Баранник В.В., Юдин А.К. Рекуррентное двухпризнаковое двоичное полиадическое кодирование // Открытые информационные и компьютерные интегрированные технологии. – Х.в: НАУ «ХАИ», 2006. – Вып. 33. – С. 22-28.
6. Баранник В.В., Юдин А.К. Усеченное представление двоичных данных с ограниченным числом серий в полиадическом пространстве // Авиационно-космическая техника и технология. – 2006. – № 2. – С. 87-92.

Поступила в редколлегию 9.06.2008

Рецензент: д-р тех. наук, проф. Ю.В. Стасев, Харьковский университет Воздушных Сил им. И. Кожедуба, Харьков.

МЕТОД ОДНОВИМІРНОГО ПЛАВАЮЧОГО СТРУКТУРНОГО КОДУВАННЯ ДВІЙКОВИХ ДАНИХ

В.В. Баранник, А.В. Хаханова

Розробляється метод кодування одновимірних плаваючих структурних чисел в двійковому просторі. При цьому враховується режим, коли код-номер формується для змінної кількості елементів, а кількість розрядів під кодограму відводиться за рівномірним принципом. Даний режим забезпечується за рахунок використання правил відбору допустимих поєднань, складених з двійкових послідовностей. Для зниження часу кодування даних процес обробки реалізується на основі рекуррентних виразів. В цьому випадку вагові коефіцієнти двійкових елементів визначаються на основі значень вагового коефіцієнта і елементів на попередньому кроці обробки. Стиснення даних досягається за рахунок скорочення комбінаторної надмірності, обумовленої обмеженою кількістю серій одиниць.

Ключові слова: одновимірне плаваюче структурне кодування, рекуррентна обробка, надмірність

METHOD OF UNIDIMENSIONAL FLOATING STRUCTURAL ENCODING OF BINARY DATA

V.V. Barannik, A.N. Hahanova

The method of encoding of unidimensional floating structural numbers is developed in binary space. The mode is thus taken into account, when a code-number is formed for the variable amount of elements, and the amount of digits under a codegram is taken on even principle. This mode is provided due to the use of rules of selection of possible combinations, made from binary sequences. For the decline of time of encoding of data the process of treatment will be realized on the basis of recurrent expressions. In this case the gravimetric coefficients of binary elements are determined on the basis of values of gravimetric coefficient and elements on the previous step of treatment. The compression of data is arrived at due to reduction of combinatorial surplus, conditioned the limited amount of cerouss of units.

Keywords: unidimensional floating structural encoding, recurrent treatment, surplus.