УДК 004.627 DOI: 10.30748/soi.2019.157.07

И.А. Кулик, А.И. Новгородцев, М.С. Шевченко

Сумской государственный университет, Сумы

МЕТОД ОЦЕНКИ ГРАНИЦ ПРИМЕНЕНИЯ СЖАТИЯ НА ОСНОВЕ ДВОИЧНЫХ БИНОМИАЛЬНЫХ ЧИСЕЛ

В статье рассматривается задача по обнаружению границ рационального использования метода сжатия двоичных последовательностей на основе биномиальных чисел. Предлагаемый метод оценки границ сжатия позволяет однозначно и достаточно просто выделить области эффективного использования двоичных биномиальных чисел для сжатия произвольных последовательностей, имеющих переменное число единиц. Учитывание границ биномиального сжатия позволяет повысить степень сжатия, снизить аппаратно-программные и временные затраты при преобразовании комбинаций.

Ключевые слова: *биномиальное сжатие, двоичные биномиальные числа, границы сжатия.*

Введение

Постановка проблемы. Область применения любого метода сжатия информации определяется, прежде всего, тем, что длина результирующего сжатого образа должна быть меньше, чем длина исходной информационной последовательности [1–2].

В работах [3–4] рассматривается метод биномиального сжатия f_b равновесных комбинаций Y_j, имеющих длину n и число k единиц, на основе двоичных (n,k)-биномиальных чисел X_j: $f_b : Y[n,k] \to X[n,k]$, где $Y[n,k]$ – множество равновесных комбинаций $Y_j \in Y[n,k]$, а $X[n,k]$ – множество (n,k)-биномиальных чисел $X_j \in X[n,k]$.

Одним из существенных достоинств биномиального сжатия f_b являются простота его операций, приводящая в результате к высокой скорости преобразования информации и минимуму аппаратно-программных затрат при практической реализации. При этом степень сжатия может быть достаточно высокой. Отдельно следует отметить то, что получаемые при f_b сжатые образы представляют собой числа X_j, над которыми можно проводить вычислительные операции, реализовывать отношения порядка и т.п., что особенно полезно для обработки сжатых записей в базах данных, поскольку не требуется при этом обратного восстановления информации [5–6].

Недостатком сжатия f_b на основе двоичных биномиальных чисел является его ориентирование на фиксированное значение k единиц, т.е. на конкретное множество $Y[n,k]$, что ограничивает применение биномиального сжатия.

С целью повышения эффективности применения метода сжатия f_b на основе двоичных биномиальных чисел необходимо решить задачу по определению области применения f_b для случая, когда k может принимать любые значения из заданного диапазона $0 \le k \le n$, а исходный сжимаемый массив представляет собой массив двоичных последовательностей A_j вида $A = \bigcup_{k=0}^{n} Y[n,k] = \{0,1\}^n$, $A_j \in A$.

Анализ последних достижений и публикаций. Из работы [7] следует, что длина r двоичных (n,k)-биномиальных чисел удовлетворяет неравенству $\min(k, n-k) \le r \le n-1$. При этом для полного множества $Y[n,k]$ равновесных комбинаций Y_j, $\mathrm{Card}(Y[n,k]) = C_n^k$, средняя длина L_{cp} результирующих двоичных (n,k)-биномиальных чисел, определяемая в работах [8–9], будет меньше n. Таким образом, сжатие n-разрядных двоичных последовательностей $Y_j \in Y[n,k]$ на основе (n,k)-бино–миальных чисел применимо для любого, но фиксированного значения k.

Решение же задачи − определение области сжатия на основе двоичных биномиальных чисел для случая, когда $0 \le k \le n$ − подразумевает поиск n и k, при которых выполняется неравенство вида:

$$v + r < n, \tag{1}$$

где v – количество разрядов, требуемых для хранения числа k единиц сжимаемой последовательности $A_j \in A$. В этом случае, когда $0 \le k \le n$, следует говорить об обобщенном сжатии вида f_{bg} на основе двоичных биномиальных чисел.

Цель статьи – повышение эффективности обобщенного метода сжатия информации на основе двоичных биномиальных чисел за счет определения границ рационального применения биномиальных чисел в зависимости от длины n сжимаемых комбинаций и переменного значения k единиц в них. Это позволит, с одной стороны, увеличить степень сжатия двоичных последовательностей на основе двоичных биномиальных чисел, а, с другой, не расходовать временные ресурсы на обработку последовательностей, сжатие которых приводит к обратному нежелательному эффекту – увеличению длины результирующей комбинации.

Изложение основного материала

Определение границ сжатия на основе двоичных биномиальных чисел

Сумма в левой части отношения (1) представляет собой длину $l_j = v + r$ результирующей сжатой последовательности после f_{bg} для случая, когда k есть переменная величина из диапазона $0 \le k \le n$.

Поскольку количество различных значений k равно $n+1$, то число двоичных разрядов v для представления k, хранение которого необходимо для однозначного восстановления A_j из сжатого образа X_j, составляет $v = \lceil \log_2(n+1) \rceil$.

Тогда общий вид для длины l_j сжатой последовательности

$$l_j = \lceil \log_2(n+1) \rceil + r, \quad (2)$$

а неравенство (1) для определения области сжатия f_{bg} преобразуется к виду:

$$\lceil \log_2(n+1) \rceil + r < n. \quad (3)$$

Теорема 1. При отображении f_{bg} максимальное значение l_j:

$$L_{\max} = \max_{0 \le k \le n} l_j = n + \lceil \log_2(n+1) \rceil - 1, \quad (4)$$

а минимальное –

$$L_{\min} = \min_{0 \le k \le n} l_j = \min(k, n-k) + \lceil \log_2(n+1) \rceil. \quad (5)$$

Доказательство. Максимум $L_{\max}$ и минимум $L_{\min}$ длины l_j (2) определяются максимальным $r_{\max}$ и минимальным $r_{\min}$ значениями количества разрядов двоичного (n,k)-биномиального числа. Согласно [7] соответственно имеем $r_{\max} = n-1$, а $r_{\min} = \min(k, n-k)$. Отсюда следуют равенства (4) и (5). **Теорема доказана.**

Теорема 2. Для любых целых $n \ge 2$ выполняется неравенство:

$$L_{\max} = n + \lceil \log_2(n+1) \rceil - 1 > n, \quad (6)$$

а для любых целых $n \ge 10$ и $0 \le k \le n$ имеет место

$$L_{\min} = \min(k, n-k) + \lceil \log_2(n+1) \rceil < n. \quad (7)$$

Доказательство. Докажем вначале справедливость $L_{\max} > n$ при $n \ge 2$. Перейдем от (6) к неравенству вида

$$n + \log_2(n+1) - 1 > n, \quad (8)$$

доказательство которого однозначно будет определять истинность (6), поскольку имеем истинным $\lceil \log_2(n+1) \rceil \ge \log_2(n+1)$. После преобразований в левой и правой частях неравенства (8), используя свойство логарифмов [10], получаем $\log_2(n+1) > 1$ или $\log_2(n+1) > \log_2 2$.

Исходя из того, что логарифмическая функция есть функция строго монотонно возрастающая [8–9], можем заключить, что $n+1 > 2$ или $n \ge 2$. Таким образом, истинность неравенства (6) при $n \ge 2$ доказана.

Теперь приведем обоснование $L_{\min} < n$ при любых целых $n \ge 10$ и $0 \le k \le n$. Пусть n − четное число и $k = n/2$, что предопределяет доказательство $L_{\min} < n$ для наихудшего случая, означая тем самым, что при n − нечетное и $k \ne n/2$ неравенство (7) тем более будет выполняться. Таким образом,

$$\frac{n}{2} + \lceil \log_2(n+1) \rceil < n.$$

Воспользуемся равенством

$$\lceil \log_2(n+1) \rceil = \lfloor \log_2(n+1) \rfloor + 1$$

и перейдем к неравенству следующего вида

$$\frac{n}{2} + \log_2(n+1) + 1 < n,$$

обоснование которого однозначно будет определять и усиливать истинность (7), поскольку $\lfloor \log_2(n+1) \rfloor \le \log_2(n+1)$.

В вышеприведенном неравенстве переносим единицу и слагаемое $n/2$ из левой части в правую для удобства потенцирования:

$$\log_2(n+1) < \frac{n}{2} - 1.$$

Далее, потенцируя обе части данного неравенства по степени 2, получаем

$$n+1 < 2^{(n-2)/2} \text{ или } 2n+2 < 2^{n/2}.$$

Известно [11–12], что числа сочетаний $C_{n/2}^{0} = C_{n/2}^{n/2} = 1$ и $C_{n/2}^{1} = C_{n/2}^{(n-2)/2} = n/2$. Правую часть неравенства представим в следующем виде:

$$2^{n/2} = \sum_{k=0}^{n/2} C_{n/2}^{k} = C_{n/2}^{0} + C_{n/2}^{1} + C_{n/2}^{(n-2)/2} + C_{n/2}^{n/2} + \\ + \sum_{k=2}^{(n-4)/2} C_{n/2}^{k} = (n+2) + \sum_{k=2}^{(n-4)/2} C_{n/2}^{k}.$$

Отсюда следует, что

$$2n+2<(n+2)+\sum_{k=2}^{(n-4)/2} C_{n/2}^{k} \text{ или } n<\sum_{k=2}^{(n-4)/2} C_{n/2}^{k} .$$

Левая часть неравенства: $n=n/2+n/2=C_{n/2}^{1}+C_{n/2}^{1}$. Отсюда,

$$C_{n/2}^{1}+C_{n/2}^{1}<\sum_{k=2}^{(n-4)/2} C_{n/2}^{k} .$$

Но в правой части неравенства любой из биномиальных коэффициентов $C_{n/2}^{k}$ при $k=2,\ldots,(n-4)/2$ является большим по значению, чем $C_{n/2}^{1}$, поскольку $k>1$. При этом количество коэффициентов $C_{n/2}^{k}$ должно быть не менее двух и, соответственно, должно выполняться условие $((n-4)/2)>2$, т.е. $n\geq 10$. Учитывая, что доказательство проводилось для наихудшего случая, когда n – четное число и $k=n/2$, то неравенство (7) будет истинным и для нечетного n, $k\leq\lfloor n/2\rfloor$ или $k\geq\lceil n/2\rceil$. Таким образом, неравенство (7) выполняется для любых целых $n\geq 10$ и $0\leq k\leq n$. **Теорема доказана.**

С точки зрения практики применения сжатия f_{bg} на основе двоичных (n,k)-биномиальных чисел X_j для случая исходных последовательностей $A_j\in A=\{0,1\}^n$, $j=\overline{1,2^n}$, теорема 2 позволяет заключить, что:

1) условия для сжатия n-разрядных A_j появляются, если $n\geq 10$;

2) существует область значений r, для которых выполняется неравенство (3).

Очевидно, необходимость использования двоичного представления $\operatorname{Bin}k$ числа k единиц для однозначного восстановления сжатых комбинаций при f_{bg}, когда значение k может изменяться в пределах $0\leq k\leq n$, что, в свою очередь, будет приводить к снижению степени сжатия. Минимизировать негативный эффект от применения $\operatorname{Bin}k$, связанный с увеличением количества разрядов сжатой комбинации, возможно за счет использования процедуры переключения между кодированием f_b и векторным кодированием f_v, т.е. тождественным преобразованием исходной последовательности в саму себя, что позволит увеличить скорость сжатия и восстановления данных. В качестве основы для выбора между f_b и f_v можно предложить неравенство вида (3), но тогда вместе с $\operatorname{Bin}k$ для однозначного декодирования возникает необходимость использовать двоичные значения $\operatorname{Bin}q$ или $\operatorname{Bin}l$, поскольку длина r есть функция чисел q единиц или l нулей в самом биномиальном числе [7]. Как следствие, это еще в большей мере понизит степень сжатия. Поэтому значение r в неравенстве (3) предлагается заменить значением средней длины L_{cp} двоичных (n,k)-биномиальных чисел:

$$L(f_{bg})=\lceil\log_2(n+1)\rceil+L_{cp}<n, \tag{9}$$

которое, являясь функцией известного n и вычисленного k, дает возможность ограничиться для восстановления исходной последовательности из сжатой комбинации служебным словом вида $\operatorname{Bin}k$.

Нижеследующая теорема подтверждает необходимость поиска границ применения обобщенного метода сжатия f_{bg} двоичных последовательностей на основе двоичных биномиальных чисел.

Теорема 3. Для любых целых $n\geq 2$ выполняется неравенство:

$$L_{\max}(f_{bg})=\lceil\log_2(n+1)\rceil+\frac{n^2}{n+2}>n,$$

если n – четное число, или

$$L_{\max}(f_{bg})=\lceil\log_2(n+1)\rceil+\frac{(n-1)(n+2)}{n+3}>n,$$

если n – нечетное число, а для любых целых $n\geq 4$ имеет место

$$L_{\min}(f_{bg})=\lceil\log_2(n+1)\rceil<n.$$

Доказательство данной теоремы строится на тех же подходах, что и доказательство предыдущей теоремы.

Из теоремы 3 можно сделать вывод, что при $n\geq 4$ функция $L(f_{bg})$ имеет точки пересечения с линией $f(k)=n$, параллельной оси абсцисс. Учитывая, что функция L_{cp} является симметричной относительно линии $k=n/2$ (допустив n – нечетное, значение $k=n/2$ есть число рациональное), параллельной оси ординат, можно заключить, что таких точек пересечения будет две – α_b и β_b, где $\beta_b=n-\alpha_b$. При этом

$$\alpha_b-1=\max_{k\in L_b} k \text{ и } \beta_b+1=\min_{k\in H_b} k$$

представляют собой границы множеств – нижней и верхней областей значений k двоичных единиц соответственно

$$L_b=\{1,2,\ldots,\alpha_b-1\} \text{ и } H_b=\{\beta_b+1,\beta_b+2,\ldots,n-1\},$$

для которых применение метода сжатия f_b на основе двоичных (n,k)-биномиальных чисел является целесообразным. И, напротив, существует множество $M_v=\{\alpha_b,\alpha_b+1,\ldots,\beta\}$ – область значений k двоичных единиц, расположенная на числовой оси между L_b и H_b, для которой рассматриваемый метод f_b дает увеличение средней длины $L\left(f_{bg}\right)$ результирующих комбинаций. При $(k=0)\vee(k=n)$ рассматривается только $\text{Bin}\,k$.

Графики функций $L\left(f_b\right)=\varphi(k)$ и $\varphi(k)=n$ при $n=64$ и $0\le k\le 64$ представлены на рис. 1. Их пересечение имеет место для значений $\alpha_b=7$ и $\beta_b=57$ чисел единиц (при округлении до ближайшего целого), которые располагаются симметрично относительно вертикальной линии $k=n/2=32$.

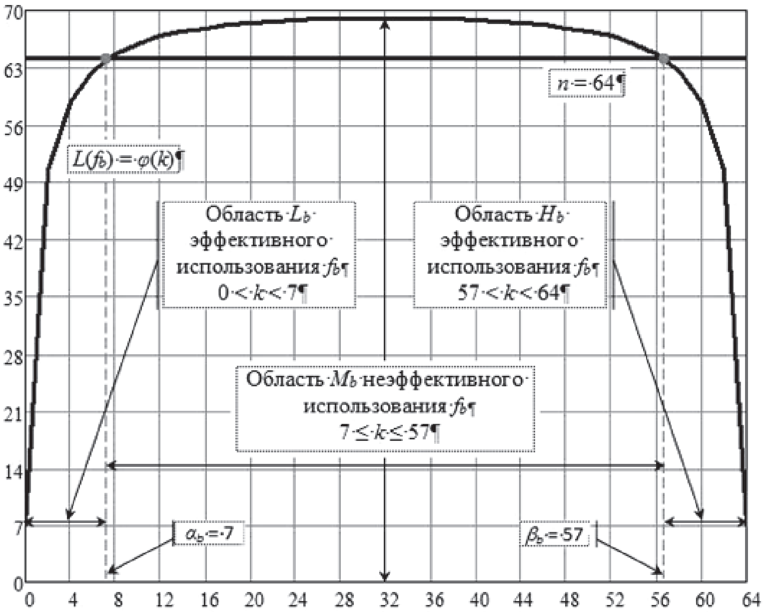

Рис. 1. Графики $L\left(f_b\right)=\varphi(k)$ и $\omega(k)=64$, области L_b и H_b эффективного и M_v неэффективного использования f_b при заданном $n=64$

Очевидно, что диапазоны $0<k<7$ и $57<k<64$ значений k единиц при заданном $n=64$ составляют соответственно области L_b и H_b использования сжатия f_b на основе двоичных биномиальных чисел, когда $L\left(f_{bg}\right)<64$.

Решение уравнения

$$\left\lceil \log_2(n+1)\right\rceil+\frac{k(n-k)(n+2)}{(k+1)(n-k+1)}=n, \qquad (10)$$

которое следует из (9), относительно $k=\alpha_b$ и $k=\beta_b$, где $\beta_b=n-\alpha_b$, для поиска L_b и H_b позволит эффективно использовать обобщенное сжатие f_{bg} на основе двоичных биномиальных чисел.

Путем группирования слагаемых, имеющих одинаковые степени k^2 и k, от равенства (10) перейдем к полному квадратному уравнению вида:

$$(\lambda+2)k^2-n(\lambda+2)k+(n-\lambda)(n+1)=0, \qquad (11)$$

где $\lambda=\left\lceil \log_2(n+1)\right\rceil$. Нахождение корней α_b, β_b уравнения (11) производится в соответствии с формулой [10–11]:

$$\alpha_b,\beta_b=\left|\frac{-n(\lambda+2)\pm\sqrt{D}}{2(\lambda+2)}\right|,$$

где $D=n^2(\lambda+2)^2-4(\lambda+2)(n-\lambda)(n+1)$ – дискриминант уравнения (11).

Некоторые значения α_b и β_b, полученные в результате программного моделирования в системе компьютерной алгебры Mathcad для различных n, представлены ниже в табл. 1.

Таблица 1

Значения α_b в зависимости от n *

n	α_b	n	α_b
10	1,3	128	13,5
16	1,9	136	14,4
24	3,3	144	15,3
32	3,8	152	16,2
40	5	160	17,1
48	6,1	168	18
56	7,3	176	18,9
64	7,3	184	19,8
72	8,3	192	20,7
80	9,3	200	21,6
88	10,3	208	22,5
96	11,3	216	23,4
104	12,3	224	24,3
112	13,4	232	25,2
120	14,4	240	26,1

* – β_b вычисляется как $\beta_b=n-\alpha_b$

Выводы

Обнаруженные границы α_b и β_b обобщенного сжатия на основе двоичных биномиальных чисел определяют область рационального использования данного метода, что особенно значимо, когда сжимаются двоичные последовательности с переменным числом единиц.

Учитывание значений α_b и β_b при аппаратной или программной реализации метода сжатия на основе двоичных биномиальных чисел позволит:

1) повысить результирующий коэффициент сжатия для произвольных двоичных комбинаций длины n, но имеющих различное число k единиц;

2) сэкономить аппаратно-программные и временные ресурсы для случая обработки двоичных последовательностей, сжатие которых не приводит к сокращению информационной избыточности.

Вместе с вычисленными значениями границ применения метода сжатия на основе двоичных биномиальных чисел практическое значение имеет и обнаруженное минимальное значение длины двоичных последовательностей, сжатие которых является целесообразным.

Список литературы

1. Кричевский Р.Е. Сжатие и поиск информации / Р.Е. Кричевский. – М.: Радио и связь, 1989. – 168 с.

2. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео / Д. Ватолин, А. Ратушняк, М. Смирнов, В. Юкин. – М.: ДИАЛОГ-МИФИ, 2003. – 384 с.

3. Кулик И.А. Модели сжатия и восстановления данных на основе двоичных биномиальных чисел / И.А. Кулик, А.А. Борисенко, Аджири Онориукпе // V Міжнародна наук.-практ. конференція “Методи та засоби кодування, захисту й ущільнення інформації”, 19-21 квітня 2016 р.: тез. доп. / Вінницький нац. техн. ун-т. – Вінниця, 2016. – С. 101-105.

4. Kulyk I.A. Compression of Binary Data Based on Binary Binomial Numbers / I.A. Kulyk, Ajiri Onoriukpe, Kehinde Ologunloluwa // П'ята Міжнародна наук.-техн. конференція “Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління”, 21-22 квітня 2016: матеріали / Полтавський нац. техніч. ун-т. – Полтава, 2016. – С. 63.

5. Кулик И.А. Биномиальная модель векторного представления базы данных с колоночной структурой / И.А. Кулик, А.И. Новгородцев, Е.М. Скордина // Системи обробки інформації. – 2016. – № 4(141). – С. 50-56.

6. Агальцов В.П. Базы данных. В 2-х книгах. Книга 2: Распределенные и удаленные базы данных / В.П. Агальцов. – М.: Форум, 2016. – 271 с.

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

8. Кулик И.А. О средней длине двоичных биномиальных чисел / И.А. Кулик // Вісник СумДУ, Серія “Технічні науки”. – 2004. – № 12(71). – С. 106-112.

9. Кулик И.А. Средняя длина двоичных биномиальных чисел произвольного диапазона / И.А. Кулик // Вісник СумДУ, Серія “Технічні науки”. – 2007. – № 2. – С. 130-139.

10. Клепко В.Ю. Вища математика в прикладах і задачах / В.Ю. Клепко, В.Л. Голець. – К.: ЦУЛ, 2017. – 592 с.

11. Кнут Д. Искусство программирования. Т 1. Основные алгоритмы: Пер. с англ. / Д. Кнут. – М: Издательский дом “Вильямс”, 2017. – 720 с.

12. Андерсон Джеймс А. Дискретная математика и комбинаторика: Пер. с англ. / Джеймс А. Андерсон. – М: Издательский дом “Вильямс”, 2016. – 960 с.

References

1. Krichevskij, R.E. (1989), “*Szhatie i poisk informacii*” [*Compression and searching of information*], Radio i svjaz', Moscow, 168 p.

2. Vatolin, D., Ratushnjak, A., Smirnov, M. and Jukin, V. (2203), “*Metody szhatija dannyh. Ustrojstvo arhivatorov, szhatie izobrazhenij i video*” [*Data compression methods. Archivers arranging, images and movies compression*], DIALOG-MIFI, Moscow, 384 p.

3. Kulyk, I.A., Borysenko, O.A. and Onoriukpe, Adzhiri (2016), “Modeli szhatija i vosstanovlenija dannyh na osnove dvoichnyh binomial'nyh chisel” [Models for data compression and decompression on basis of binary binomial numbers], *5th International Scientific Practical Conference “Methods of information coding, defence and compression”*, April 19-21, Vinnitsya, pp. 101-105.

4. Kulyk, I.A., Onoriukpe, Ajiri and Ologunloluwa, Kehinde (2016), Compression of Binary Data Based on Binary Binomial Numbers, *5th International Scientific Technical Conference “Contemporary development trends of information communication technologies and management facility”*, April 21-22, Poltava national technical university, Poltava, pp. 63.

5. Kulyk, I.A., Novgorodcev, A.I. and Skordina, E.M. (2016), “Binomial'naja model' vektornogo predstavlenija bazy dannyh s kolonochnoj strukturoj” [Binomial model for data base vector representation with columnar structure], *Information Processing Systems*, No. 4 (141), pp. 50-56.

6. Agal'cov, V.P. (2016), “*Bazy dannyh. Raspredelennye i udalennye bazy dannyh, Kniga 2*” [*Data bases. Distributed and remote data bases, Vol. 2*], Forum, Moscow, 271 p.

7. Borysenko, O.A. and Kulyk, I.A. (2010), “*Binomial'noe kodirovanie: monografija*” [*Binomial coding*], Sumy state university, Sumy, 206 p.

8. Kulyk, I.A. (2004), "O srednej dline dvoichnyh binomial'nyh chisel" [About the average length of binary binomial numbers], *Bulletin of Sumy state university, Series "Tehnichni nauki"*, No. 12 (71), Sumy state university, pp. 106-112.

9. Kulyk, I.A. (2007), "Srednjaja dlina dvoichnyh binomial'nyh chisel proizvol'nogo diapazona" [The average length of binary binomial numbers for an arbitrary range], *Bulletin of Sumy state university, Series "Tehnichni nauki"*, No. 2, Sumy state university, pp. 130-139.

10. Klepko, V.Ju. and Golec', V.L. (2017), "*Vyshha matematyka v prykladah i zadachah*" [*Higher mathematics in examples and tasks*], CUL, Kyiv, 592 p.

11. Knut, D. (2017), "*Iskusstvo programmirovanija. Tom 1. Osnovnye algoritmy*" [*The Art of Computer Programming. Volume 1. Fundamental Algorithms*], Publishing house "Vil'jams", Moscow, 720 p.

12. Anderson, James A. (2016), "*Diskretnaja matematika i kombinatorika*" [*Discrete mathematics and combinatorics*], Publishing house "Vil'jams", Moscow, 960 p.

Поступила в редколлегию 11.03.2019
Одобрена к печати 23.04.2019

Відомості про авторів:

Кулик Ігор Анатолійович
кандидат технічних наук доцент
Сумського державного університету,
Суми, Україна
https://orcid.org/0000-0003-2403-8671

Новгородцев Анатолій Іванович
кандидат технічних наук доцент
Сумського державного університету,
Суми, Україна
https://orcid.org/0000-0003-4598-5598

Шевченко Марина Сергіївна
здобувач наукового ступеня кандидата наук
Сумського державного університету,
Суми, Україна
https://orcid.org/0000-0002-1434-5996

Information about the authors:

Igor Kulyk
Candidate of Technical Sciences Associate Professor
of Sumy State University,
Sumy, Ukraine
https://orcid.org/0000-0003-2403-8671

Anatoliy Novhorodtsev
Candidate of Technical Sciences Associate Professor
of Sumy State University,
Sumy, Ukraine
https://orcid.org/0000-0003-4598-5598

Marina Shevchenko
Applicant for Candidate's Degree
of Sumy State University,
Sumy, Ukraine
https://orcid.org/0000-0002-1434-5996

МЕТОД ОЦІНКИ МЕЖ ЗАСТОСУВАННЯ СТИСНЕННЯ НА ОСНОВІ ДВІЙКОВИХ БІНОМІАЛЬНИХ ЧИСЕЛ

І.А. Кулик, А.І. Новгородцев, М.С. Шевченко

В статті розглядається задача щодо виявлення меж раціонального застосування методу стиснення двійкових послідовностей на основі біноміальних чисел. Запропонований метод оцінки меж стиснення дозволяє однозначно і достатньо просто виділити області ефективного використання двійкових біноміальних чисел для стиснення довільних послідовностей, які мають змінну кількість одиниць. Врахування меж біноміального стиснення дозволяє підвищити ступінь стиску, знизити апаратно-програмні та часові витрати під час перетворення комбінацій.

Ключові слова: *біноміальне стиснення, двійкові біноміальні числа, межи стиснення.*

METHOD FOR BORDERS ESTIMATION OF COMPRESSION ON BASIS OF BINARY BINOMIAL NUMBERS

I. Kulyk, A. Novhorodtsev, M. Shevchenko

In the paper the problem on searching the borders of the reasonable application for the compression method, using binomial numbers, is under review. The compression method would be used effectively for constant weight code, that is, when the amount of units is fixed. In case of the variable amount of units it is necessary to include the number of units by itself into the compressed form. As a result, the compression ratio is decreased, as well the hardware and program resources are spent to no effect. Therefore the searching of the reasonable application borders for the compression on basis of binary binomial numbers is a topical and useful task. The proposed method for the borders estimation allows us to separate out definitely and enough simply the areas of the efficient application of binary binomial numbers in order to compress arbitrary sequences, having the variable amount of units. The considered method for the borders estimation consist in seeking values of the units' amount and length of digits for the sequences to be compressed, when the resultant form is less than the initial sequence as to the quantity of digits. Taking into account the borders for the compression method, using binary binomial numbers, gives possibilities to improve its compression ratio and increase its compression and decompression speed. The formulated and proved theorems demonstrate the actuality and possibility to search the borders of the reasonable application for the compression method, as well they point to the minimal values of sequences length, at which the borders of the compression should be applied. The graphic and table of the found borders values, which demonstrate the areas for amounts of units to compress binary sequence, are given in the paper.

Keywords: *binomial compression, binary binomial numbers, compression borders.*