УДК 621.3.037.37

МЕТОД ВЫЧИСЛЕНИЯ БИНОМИАЛЬНЫХ КОЭФФИЦИЕНТОВ НА ОСНОВЕ КАНОНИЧЕСКОГО РАЗЛОЖЕНИЯ ЧИСЕЛ

И.А. Кулик, Е.М. Скордина
Сумский государственный университет, г. Сумы

В статье предлагается новый метод вычисления биномиальных коэффициентов на основе канонического разложения чисел. Данный метод позволяет существенно сократить временные и аппаратно-программные затраты при вычислении чисел сочетаний в случае больших значений их параметров. Кроме того, двоично-каноническое кодирование дает возможность сжимать двоичное представление биномиальных коэффициентов.

ПОСТАНОВКА ПРОБЛЕМЫ

Целый ряд устройств и методов нумерационного кодирования требуют подсчёта чисел сочетаний C_n^k для различных длин n двоичных последовательностей и числа k, содержащихся в них единиц. Достаточно большое распространение получили методы нумерационного кодирования на основе биномиальных систем счисления, в качестве весовых коэффициентов которых применяются числа сочетаний C_n^k [1, 2]. Эффективность методов биномиального кодирования, например, методов биномиального сжатия, прямо пропорционально зависит от значений n и k. Следовательно, актуальной является задача вычисления биномиальных коэффициентов C_n^k при больших длинах n двоичных комбинаций и чисел k, содержащихся в них единиц.

Традиционные методы подсчета биномиальных коэффициентов C_n^k при больших значениях n и k достаточно трудно реализовать на практике по следующим причинам:

1) громоздкость вычислений, связанных с подсчётом факториалов $n!$ и $k!$;

2) сложность представления конечного целочисленного результата, достигающего огромных значений (например, при $n = 500$ и $k = 250$ – $C_{500}^{250} \approx 1,17 \times 10^{149}$).

Особую остроту эти трудности приобретают при разработке специализированных устройств подсчёта C_n^k и устройств нумерационного кодирования, техническая реализация которых сталкивается с реальными ограничениями на время вычисления чисел сочетаний и объем аппаратных затрат, в частности памяти для хранения результатов подсчета.

1 СУЩЕСТВУЮЩИЕ ПОДХОДЫ К ВЫЧИСЛЕНИЮ ЧИСЕЛ СОЧЕТАНИЙ

Существующие способы подсчета чисел C_n^k сочетаний или имеют оценочный характер, или требуют больших аппаратно-программных и временных затрат [1, 3]. На сегодняшний день для вычисления чисел сочетаний получили распространение следующие методы:

1) метод с использованием факториалов параметров n, k и $(n-k)$, (назовём его классическим методом) [3]:

$$C_n^k = \frac{n!}{k!(n-k)!}; \qquad (1);$$

2) упрощенный классический метод – для $k \le n/2$

$$C_n^k = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}{k!} \qquad (2)$$

и для $k > n/2$

$$C_n^k = \frac{n \cdot (n-1) \cdot \ldots \cdot (k+1)}{(n-k)!}; \qquad (3)$$

3) рекурсивный метод, предложенный проф. Борисенко А.А. в работе [1].

При проведении сравнительного анализа следует воспользоваться тремя критериями, с помощью которых будет характеризоваться каждый из приведенных методов:

– числом операций для проведения вычислений;
– временем выполнения операций;
– аппаратно-программнми затратами.

Классический метод. С увеличением значений параметров n и k возрастает количество операций умножения по вычислению факториалов $n!$ и $k!$ (1). Следовательно возрастает время, необходимое для подсчёта C_n^k. Кроме того, в случае практической реализации данного метода при больших n и k, к примеру уже при $n > 30$ требуются значительные аппаратно-программные затраты.

Упрощённый метод. Для данного метода число операций умножения меньше, чем для классического метода при тех же значениях n и k. По предварительной оценке выражений (2, 3) количество множителей для упрощенного метода уменьшается в $2(n-k)$ раза для случая $k \le n/2$ и в $2k$ раза для случая $k > n/2$, что в целом при реализации приводит к уменьшению временных и аппаратно-программных затрат по сравнению с классическим методом, но их объем по-прежнему остается достаточно значительным.

Рекурсивный метод основан на использовании биномиального прямоугольника Борисенко [1]. Биномиальный прямоугольник составляют элементы следующего вида:

$$A_{ij} = \sum_{\alpha=0}^{j} A_{(i-1)\alpha} = C_{i+j}^{j}; \qquad (4)$$

$$i = 1, 2, \ldots, k; \; A_{0j} = 1; \; A_{i0} = 1; \; A_{1j} = A_{1(j-1)} + 1, \; j = 1, 2, \ldots, k;$$

$$A_{i1} = A_{(i-1)1} + 1, \; i = 1, 2, \ldots, n-k.$$

После задания элементов нулевой строки и нулевого столбца с помощью единиц остальные элементы прямоугольника могут быть получены с помощью равенства (4). Биномиальный коэффициент C_n^k может быть определен из прямоугольника суммированием всех элементов предпоследней $(n-k-1)$-й строки

$$C_n^k = A_{(n-k-1)k} + A_{(n-k-1)(k-1)} + \ldots + A_{(n-k-1)1} + A_{(n-k-1)0}\,.$$

В данном методе вычисления числа сочетаний C_n^k арифметические операции ограничиваются только операцией суммирования. Так как операции умножения и деления отсутствуют, то положительным свойством рекурсивного метода по сравнению с предыдущими является существенное снижение аппаратно-программных затрат, но значительными при рекурсии остаются временные затраты.

Как показывает предварительный анализ методов подсчёта C_n^k, основные сложности в их практической реализации заключаются в многочисленных операциях умножения и деления, на что требуются достаточно большие временные ресурсы. Кроме того, существуют трудности в представлении промежуточных результатов вычисления и результирующего значения числа C_n^k сочетаний на практике, поскольку при значительных n и k происходит переполнение разрядной сетки представления чисел. В связи с этим для решения указанных трудностей необходимо решить следующие задачи:

1) при вычислении C_n^k уменьшить число операций умножения и деления и по возможности свести их к операциям сложения и вычитания;

2) разработать экономичную форму представления промежуточных значений и результирующих чисел C_n^k сочетаний с точки зрения количества используемых двоичных разрядов.

Для решения поставленных задач в данной работе предлагается рассмотреть новый метод вычисления чисел сочетаний на основе канонической формы разложения чисел n, k и C_n^k, начальные идеи которого были изложены еще в работе [4].

2 АРИФМЕТИЧЕСКИЕ СВОЙСТВА ЧИСЕЛ СОЧЕТАНИЙ

Разработка эффективного метода вычисления чисел C_n^k сочетаний с точки зрения снижения затрат при реализации на практике связана с изучением свойств самой функции C_n^k и особенностей ее получения.

Известно, что любое целое число z может быть представлено как произведение простых чисел [5]:

$$z = q_1^{a_1} \times q_2^{a_2} \times \ldots \times q_j^{a_j} \times \ldots \times q_m^{a_m}\,, \tag{5}$$

где q_j – j-й элемент ряда простых чисел, $j = 1, 2, \ldots, m$;

a_j – показатель степени простого числа q_j;

m – количество простых чисел, необходимых для представления числа z.

Такая форма записи числа z в соответствии с основной теоремой арифметики [5] является единственной и позволяет однозначно представлять любое целое число с точностью до порядка множителей.

Таким образом, существует взаимооднозначное отображение f_k между множеством Z целых чисел и множеством соответствующих векторов показателей степени $V = \{A_1, A_2, \dots, A_i, \dots\}$, где $A_i = (a_{1i}, a_{2i}, \dots, a_{mi})$: $f_k : Z \to V$.

Представление любых целых чисел z_i и z_j из множества Z в виде векторов A_i и A_j позволяет трудоемкие операции умножения и деления над числами z_i и z_j заменить простыми операциями сложения и вычитания над их соответствующими векторами A_i и A_j.

Согласно [6] объем временных затрат, необходимый для умножения и сложения двух r-разрядных чисел, пропорционален r^2 и r соответственно. Отсюда следует, что и время, необходимое для суммирования векторов степени A_i и A_j, будет меньше в r раз времени, необходимого для умножения соответствующих им чисел z_i и z_j, что значительно снижает временные затраты. Например, при умножении чисел z_i и z_j, каждое из которых равно C_{100}^{50}, и требующих в этом случае для своего двоичного представления $r = 97$ разрядов, временные затраты при использовании канонической формы записи на операцию умножения

$$z_i \cdot z_j = (A_i + A_j) = ((a_{1i} + a_{1j}), (a_{2i} + a_{2j}), \dots, (a_{mi} + a_{mj}))$$

будут приблизительно в $r = 97$ раз меньше, чем при обычном двоичном умножении чисел z_i и z_j.

Кроме того, значительная часть аппаратных затрат при построении устройств вычисления чисел C_n^k сочетаний при достаточно больших n и k связана с их двоичным представлением. Как уже отмечалось, при $n = 100$ и $k = 50$ для хранения целочисленного значения числа сочетаний C_{100}^{50} понадобилось бы $\log_2 C_{100}^{50} \approx 97$ двоичных разрядов. Использование особенностей подсчета C_n^k и свойств самого числа сочетаний позволяет не только существенно снизить время вычислений, но и перейти к его более рациональному виду, что демонстрируют сформулированные ниже теоремы 1 и 2.

Теорема 1. Для представления числа сочетаний в канонической форме $C_n^k = q_1^{a_1} \times q_2^{a_2} \times \dots \times q_j^{a_j} \times \dots \times q_m^{a_m}$ достаточен набор простых чисел q_j, значения которых не больше n, т.е. $q_j \le n$ при $j = 1, 2, \dots, m$.

Доказательство. Данное утверждение может считаться доказанным, если удастся показать, что процедура получения числа C_n^k определяется функцией $C_n^k = f(n, k)$, характер операции с числами n и k которой не приводит к получению в выражении (5) простых чисел $q_j > n$.

Допустим $q_1 < q_2 < \dots < q_j < \dots < q_m$. Из определения [5], где простыми числами являются целые положительные числа $q_j > 1$, делители которых исчерпываются числами q_j и 1, следует, что невозможно из множества $\{q_1, q_2, \dots, q_j, \dots, q_m\}$ с помощью только операций умножения и деления получить хоть одно простое число $q_{m+1} > q_m$.

Анализ выражения для подсчета числа сочетаний

$$C_n^k = f(n,k) = \frac{n!}{k!(n-k)!} = \frac{q_1^{a_1} \cdot q_2^{a_2} \cdot \ldots \cdot q_m^{a_m}}{(q_1^{b_1} \cdot q_2^{b_2} \cdot \ldots \cdot q_m^{b_m}) \cdot (q_1^{c_1} \cdot q_2^{c_2} \cdot \ldots \cdot q_m^{c_m})} = = q_1^{a_1-b_1-c_1} \cdot q_2^{a_2-b_2-c_2} \cdot \ldots \cdot q_m^{a_m-b_m-c_m} \quad (6)$$

показывает, что перечень операций с числами в функции (6) для получения каждого элемента $q_j^{a_j}$ ограничивается только операциями умножения и делении над числами меньше n, что не может привести к появлению в канонической форме записи простых чисел $q_j > n$, что и требовалось доказать.

Теорема 2. Число C_n^k сочетаний в канонической форме записи содержит весь ряд простых чисел q_j:

$$\begin{cases} q_j \geq n-k+1, \; k \leq n/2 \\ q_j \geq k+1, \; k > n/2 \end{cases},$$

для каждого из которых показатель степени будет равен единице $a_j = 1, \; j = 1, 2, \ldots, m$.

Доказательство. Рассмотрим доказательство теоремы для случая $k \leq n/2$. При этом воспользуемся выражением для вычисления C_n^k согласно упрощенному классическому способу:

$$C_n^k = \frac{(n-k+1)(n-k+2) \times \ldots \times n}{1 \cdot 2 \times \ldots \times k}. \quad (7)$$

Представим, что множество $R = \{r_\nu : n-k+1 \leq r_\nu \leq n, \nu = 1, 2, \ldots, k\}$ сомножителей числителя состоит из двух подмножеств, элементы которых подчинены отношению строгого порядка. Первое подмножество $Q = \{q_j : n-k+1 \leq q_j \leq n\}$ содержит только простые числа q_j. Все остальные элементы множества R, не входящие в Q, образуют множество S, каждый элемент которого является целым положительным числом, который можно разложить на простые сомножители в виде (5). Очевидно, что S является разностью R/Q. Совокупность сомножителей знаменателя выражения (7) образует множество $Y = \{1, 2, \ldots, k\}$.

Для того чтобы утверждать, что, начиная с определенного простого числа q_j, в разложении (5) участвуют все простые числа от $n-k+1$ до n включительно, необходимо показать, что в процессе преобразования выражения (7) к виду (5) не происходит взаимного сокращения элементов множеств Q и Y, множества Q и Y не содержат общих элементов и являются непересекающимися, т.е. $Q \cap Y = \varnothing$. Для выполнения этого требовании достаточно, чтобы $\min Q > \max Y$, т.е. $n-k+1 > k$. После преобразования получим $(n+1)/2 > k$. Данное неравенство соблюдается при изменении k от 1 до $n/2$. Учитывая, что доказательство теоремы производится для $k \le n/2$, можно считать доказанной первую часть теоремы 2 о полноте ряда простых чисел $q_j \in Q$.

При доказательстве второй части теоремы, что каждое простое число $q_j \in Q$ в разложении (5) имеет показатель степени $a_j = 1$, достаточно доказать, что показатель степени a_j не может быть больше 1, так как в процессе преобразования выражения (7) к виду (5) в множестве S не найдется такого элемента, в разложении которого участвует хоть один элемент $q_j \in Q$, т.е. в его разложении участвуют простые числа $q_j < \min Q$. Если при разложении числа $\tilde{s} = \max S$, это требование удастся доказать для двух сомножителей с $a_j = 1$, один из которых является наименьшим простым числом $q_1 = 2$, тем более это будет справедливо для всех остальных элементов множества S, для большего числа сомножителей, наименьший из которых может быть больше двух, и для $a_j > 1$.

Допустим $\max S = q_1 q_j = 2q_j$. Так как справедливо неравенство $\max S \le n$, то $q_j \le n/2$. Исходя из того, что $\min Q = n - k + 1$, то для любого $k = 1, 2, \ldots, n/2$ всегда $\min Q \le (n/2) + 1$. Таким образом, всегда $q_j \le \min Q$, что и требовалось доказать.

Аналогично производится доказательство теоремы 2 и для случая $k > n/2$.

2 ПРЕДСТАВЛЕНИЕ ЧИСЕЛ СОЧЕТАНИЙ В ДВОИЧНО-КАНОНИЧЕСКОЙ ФОРМЕ

Приведенные арифметические свойства чисел C_n^k сочетаний, доказанные в теоремах 1 и 2, позволяют разработать алгоритмы и устройства, сводящие к минимуму общее количество арифметических операций при подсчёте C_n^k, и обеспечить эффективное с точки зрения экономии двоичных разрядов кодирование.

Как уже отмечалось, для больших значений параметров n и k значения чисел сочетаний получаются достаточно огромными. В связи с этим возникает проблема их хранения и проведения над ними арифметических операций. Представление в обычном двоичном виде требует большого числа разрядов, которое определяется как $\lceil \log_2 C_n^k \rceil$.

В качестве примера рассмотрим представление числа C_{50}^{17} с помощью обычного двоичного кодирования и в предлагаемой двоично-канонической форме. Для записи числа C_{50}^{17} в случае двоичного кодирования необходимо использовать

$$\lceil \log_2 C_{50}^{17} \rceil \approx \lceil 45{,}16 \rceil = 46 \text{ разрядов.}$$

Чтобы применить двоично-каноническую форму записи для числа C_{50}^{17}, разложим его на простые множители, используя теоремы 1 и 2:

$$C_{50}^{17} = 2^{16} \cdot 3^7 \cdot 5^5 \cdot 7^4 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 37 \cdot 41 \cdot 43 \cdot 47.$$

Закодируем показатели степени двоичным алфавитом, представляя при этом отсутствующие простые числа двоичными нулями. В результате двоично-каноническое представление числа сочетаний C_{50}^{17} имеет вид (для наглядности показатели степени разделены точками):

$$1111.111.101.100.1.1.1.1.1.1.0.0.1.1.1.1.1.$$

В этом случае для записи числа C_{50}^{17} необходимо 24 разряда в отличие от обычного двоичного кодирования, требующего 46 разрядов. Таким образом, экономия двоичных разрядов, необходимых для представления и хранения числа сочетания C_{50}^{17} в двоично-каноническом виде, составляет $46/24 \approx 1,9$ раза.

Для однозначности представления в двоично-канонической форме степени простых множителей предлагается записывать в префиксном коде, например биномиальном [2], тем самым легко определяются начало и конец степени простого числа. Кроме того, использование двоично-канонической формы записи чисел позволяет свести операции произведения и деления к сложению и вычитанию степеней соответствующих простых чисел.

ВЫВОДЫ

В заключение данной научной работы можно сделать следующий вывод, что предлагаемый метод вычисления биномиальных коэффициентов C_n^k на основе канонической формы записи чисел позволяет:

1) существенно уменьшить время вычисления чисел сочетаний за счет замены операций умножения и деления более простыми операциями сложения и вычитания;

2) сократить объём памяти для хранения промежуточных и конечных значений при вычислении чисел C_n^k;

3) позволяет упростить построение специализированных устройств вычисления биномиальных коэффициентов C_n^k и нумерационного кодирования на основе биномиальных систем счисления.

При этом целесообразным является дальнейшее исследование арифметических свойств числа сочетаний и его разложения на простые числа, а также особенностей процедур его вычисления с тем, чтобы найти эффективные алгоритмы и схемотехнические решения для вычисления биномиальных коэффициентов при больших параметрах n и k.

SUMMARY

METHOD FOR CALCULATON OF BINOMIAL COEFFICIENTS ON BASIS OF CANONICAL DECOMPOSITION OF NUMBERS

Kulik I.A.; Skordina E.M.
Sumy State University, 2 R-Korsakova Str., Sumy, 40007

In the paper a new method is proposed for computing binomial coefficients on basis of a canonical decomposition of numbers. The method allows us to decrease time and hardware costs essentially when binomial coefficients are computed for its large parameters. Moreover, binary-canonical coding of numbers give a possibility to compress binary representations of binomial coefficients.

СПИСОК ЛИТЕРАТУРЫ

1. Борисенко А.А. Введение в теорию биномиального счета: Монография. – Сумы: ИТД "Университетская книга", 2004. – 88 с.
2. Борисенко А.А. Биномиальный счет. Теория и практика: Монография. – Сумы: ИТД "Университетская книга", 2004. – 170 с.
3. Кнут Д. Искусство программирования для ЭВМ. – Т1: Основные алгоритмы. – М.: Изд-во "Мир", 1976. – 736 с.
4. Кулик И.А., Арбузов В.В., Бережная В.В. К вычислению числа сочетаний // Вісник СумДУ. – 1995. – №3. – С. 66-70.
5. Виноградов И.М. Основы теории чисел. – Москва, 1962. – 179 с.
6. Самофалов К.Г. и др. Прикладная теория цифровых автоматов. – Киев: Изд-во "Высшая школа", 1987. – 375 с.

Кулик И.А., канд. техн. наук, доцент;
Скордина Е.М., студентка

Поступила в редакцию 23 мая 2008 г.