

РАЗДЕЛ 1. МОДЕЛИРОВАНИЕ, РАСЧЕТ И ОБРАБОТКА ДАННЫХ В АВТОМАТИЗИРОВАННЫХ СИСТЕМАХ

2. Ефремов, М.С. Моделирование напряженного состояния зуба после реставрации кариозных полостей [Текст] / М.С. Ефремов [и др.] // Ползуновский альманах. – Барнаул: АлтГТУ, 2008. – №2 – С. 162-164.
3. Зенкевич, О. Метод конечных элементов в технике [Текст] / О. Зенкевич. – М.: Мир, 1975. – 539с.
4. Наймушин, Ю.Н. Моделирование напряженно-деформированного состояния зуба и пломбы с учетом их эмпирических деформационных характеристик [Текст] / Ю.Н. Наймушин, [и др.] // Вестник ИжГТУ. – 2010. – №2. – С.79-83.
5. Ремизов, С.М. Микромеханические характеристики реставрационных стоматологических материалов, эмали и дентина зубов человека [Текст] / С.М. Ремизов // Стоматология. – 2001 – №4 – С. 28-32.

Д.т.н., зав. лаб. Шелковников Е.Ю., м.н.с. Кириллов А.И. – Институт механики УрО РАН, iit@udman.ru; к.т.н., доц. Ефремов С.М. – Ижевский Государственный технический университет, efsemi@mail.ru; д.м.н., проф. Рединова Т.Л., аспирант Тимофеев А.А., аспирант Метелева Т.Ю. – Ижевская медицинская академия, (3412)425488, dimitu@deikin.ru

УДК 519.722

ЭФФЕКТИВНЫЙ МЕТОД УНИВЕРСАЛЬНОГО КОМБИНАТОРНОГО КОДИРОВАНИЯ

М.П. Бакулина

В данной работе рассматривается задача построения и реализации эффективного метода комбинаторного универсального кодирования, т.е. метода кодирования при неизвестной статистике сообщений. Приводятся экспериментальные результаты работы данного алгоритма, как для построенного префикса, так и для префикса, обеспечивающего минимальную избыточность.

Ключевые слова: универсальное кодирование, комбинаторный код, избыточность, энтропия

Введение

Задача эффективного кодирования источников является одной из центральных в теории информации, что объясняется ее теоретической важностью и многочисленными практическими приложениями к сжатию данных самой различной природы.

Комбинаторное кодирование, названное так по аналогии с арифметическим кодированием, использует статистику исходных данных, и является разновидностью статистических (энтропийных) методов сжатия информации. К таким методам также относятся арифметическое кодирование, и ряд методов Хаффмана других методов.

Суть комбинаторного метода заключается в том, что из известного алфавита и таблицы частот можно сформировать строго определенное число различных последовательностей данных. Поэтому каждой последовательности данных ставится в соответствие строго определенный (комбинаторный) номер (код) и наоборот, каждому комбинаторному номеру строго определенная последовательность данных.

Рассматриваются задачи кодирования источников с известной и неизвестной статистикой. Методы, предназначенные для кодирования источников с известной статистической структурой, называются статистическими (неадаптивными) методами. Однако

наиболее важной и сложной является задача построения кодов с неизвестной статистикой источника. Такие коды называются универсальными кодами.

Возможность построения универсальной схемы комбинаторного кодирования и доказательство простоты ее реализации были впервые описаны В.Ф. Бабкиным [1]. Исследованию универсальных комбинаторных кодов посвящены также работы [2], [3].

В данной работе рассматривается задача построения и реализации эффективного метода комбинаторного универсального кодирования, т.е. метода кодирования при неизвестной статистике сообщений. Приводятся экспериментальные результаты работы данного алгоритма.

Алгоритм комбинаторного универсального кодирования. Предположим, что алфавит состоит из двух букв a_1 и a_2 , появляющихся независимо с вероятностями $p, q = 1 - p$, но величина p заранее неизвестна. Требуется построить код, для которого среднее число символов 0 и 1 на одну букву алфавита приближалось бы к энтропии источника при любом $p < 1$.

Построение кода начнем с разбиения множества всех блоков длины n в алфавите A на группы, которые имеют одинаковые

вероятности при любом p . Число таких групп равно $n+1$. В нулевой группе отсутствует буква a_2 , при этом группа состоит из единственного блока $a_1 a_1 \dots a_1$, вероятность появления которого равна p^n . Первая группа состоит из всех блоков длины n , содержащих одну букву a_2 . Эта группа состоит из $C_n^1 = n$ блоков, вероятность каждого из которых равна $p^{n-1}q$. Группа с номером k состоит из всех блоков длины n , содержащих k букв a_2 . Эта группа содержит C_n^k блоков, вероятность каждого из которых равна $p^{n-k}q^k$.

Универсальный код для k -й группы состоит из двух частей: префикса и суффикса. В качестве префикса для кода слова x выбирается двоичная запись числа $r(x)$ единиц в этом слове. Эта двоичная запись для всех слов имеет одну и ту же длину $\lceil \log_2(n+1) \rceil$. Префикс указывает, к какой группе сообщений принадлежит кодируемый блок, суффикс же содержит $\lceil \log_2(C_n^k) \rceil$ двоичных символов и указывает номер блока в группе. Построенный таким образом код будет однозначно дешифруем. На приемном конце первоначально по $\lceil \log_2(n+1) \rceil$ элементам кода определяют, к какой группе принадлежит переданное сообщение, а затем по следующим $\lceil \log_2(C_n^k) \rceil$ элементам определяют, какое именно сообщение передавалось.

Префикс каждой из групп при комбинаторном кодировании содержит ровно $\lceil \log_2(n+1) \rceil$ символов 0 и 1. Однако более эффективным является кодирование префикса неравномерным кодом.

Из приведенного выше описания метода кодирования видно, что наиболее трудоемкой частью кодирования является нахождение суффикса. Опишем алгоритм его нахождения. Номер $N(x)$ слова x , $|x|=n$, имеющего $r_1(x) = r$, $r > 0$, вхождений единицы, во множестве C_n^r слов, содержащих r единиц, равен

$$N(x) = \sum_{i=1}^r C_{n_i-1}^r, \quad (1)$$

где n_i - номер i -й по счету единицы в слове x , $i = 1, \dots, r$.

Вычисления ведутся следующим образом. Если i -я буква слова x равна 1, то вычисляем величину

$$C_{i+1}^{k+1} = C_i^k \cdot \frac{i+1}{i-k+1} \quad (2)$$

Если же i -я буква равна 0, то вычисляем величину

$$C_{i+1}^k = C_i^k \cdot \frac{i+1}{i-k} \quad (3)$$

Прибавляя найденное значение (2) или (3) к сумматору, где вначале находится 0, после обработки последней n -й буквы x получим номер $N(x)$ этого слова согласно (1). При этом необходимо осуществить $(n-1)$ умножений и делений по формулам (2) и (3), а также n сложений. Величина первого множителя в формулах (2) и (3) не превышает 2^n , а величины второго множителя и делителя не превышают n . Как следует из [4], число операций при умножении и делении a -разрядного числа на b -разрядное равно $O(a \cdot \log b \cdot \log \log b)$, если $a \gg b$. Получаем следующее утверждение

Утверждение. Для вычисления номера слова x , $|x|=n$, в совокупности всех слов, содержащих заданное число r единиц, $r = r(x)$, достаточно затратить $O(n^2 \cdot \log \log n \cdot \log \log \log n)$ операций над одноразрядными числами.

Номер $N(x)$ можно вычислить, пользуясь таблицей биномиальных коэффициентов (треугольником Паскаля). Элементы этой таблицы вычисляются по мере надобности либо размещаются в памяти кодирующего устройства. Ниже приведен фрагмент такой таблицы. На пересечении i -й строки и j -го столбца стоит C_{i-1}^j .

8	7	21	35	35	21	7	1
7	6	15	20	15	6	1	
6	5	10	10	5	1		
5	4	6	4	1			
4	3	3	10				
3	2	1					
2	1						
1	0						

РАЗДЕЛ 1. МОДЕЛИРОВАНИЕ, РАСЧЕТ И ОБРАБОТКА ДАННЫХ В АВТОМАТИЗИРОВАННЫХ СИСТЕМАХ

Рассмотрим следующий пример вычисления номера $N(x)$. Пусть $n=8$, $x=01101101$. Здесь номера вхождений единицы в слово x равны 2, 3, 5, 6, 8. Соответствующие слагаемые из формулы (1) выделены в таблице жирным шрифтом. Суммируя их, находим, что номер $N(x)=21+5+4+1+1=32$.

Если известен номер $N(x)$ и известно, что число единиц $r(x)$ в слове x равно 5, то для нахождения x вначале определяем максимальное число пятого столбца таблицы, не превосходящее $N(x)$. Это $21=C_{8-1}^5$. Значит, $n_5=8$, т. е. на восьмом месте слова x стоит 1. Так как $32-21=11$, то находим максимальное число четвертого столбца, не превосходящее 11. Это $5=C_{6-1}^4$, т. е. $n_4=6$. Продолжая таким же образом, находим $n_3=5$, $n_2=3$, $n_1=2$. Трудоемкость этой процедуры кодирования совпадает по порядку с трудоемкостью кодирования.

Отметим, что максимальная избыточность данного метода на множестве двоичных источников равна в пересчете на букву асимптотически $\frac{\log_2 n}{n}$, где n – длина слова. Минимальное значение этой избыточности равно $\frac{\log_2 n}{2n}$.

Известно (см. [4]), что асимптотически наилучший универсальный код, т.е. код, обеспечивающий минимальную избыточность, получается, если длину префикса слова выбрать равной

$$\log_2 n + \frac{1}{2} \cdot \left(\log_2 \frac{r_1(x)}{n} + \log_2 \frac{r_0(x)}{n} \right) + C, \quad (4)$$

где $r_1(x)$ и $r_0(x)$ – число вхождений 1 и 0 соответственно в слово длины n , а C – некоторая константа, выбираемая так, чтобы выполнялось неравенство Крафта.

Экспериментальные результаты. В приведенной ниже таблице 1 представлены экспериментальные результаты работы описанного выше алгоритма универсального комбинаторного кодирования. В качестве тестов были взяты различные файлы, в третьем

и четвертом столбцах таблицы указаны время работы алгоритма и коэффициент сжатия файлов, определяемый как отношение исходного объема файла к объему, полученному при кодировании.

Таблица 1 - результаты работы алгоритма универсального комбинаторного кодирования

№	Объем файла (кб)	Время (с)	Коэффициент сжатия (%)
1	3,58	1,3	38.5
2	65	3.2	41.6
3	112	4.1	43.2
4	220	6.5	39.3
5	461	9.7	45.9
6	804	15.3	41.7
7	1347	24.9	44.1
8	2524	45.6	42.8

В таблице 2 приведены результаты работы алгоритма с длиной префикса, определяемой формулой (4) и дающий асимптотически оптимальный код.

Таблица 2

№	Объем файла (кб)	Время (с)	Коэффициент сжатия (%)
1	3,58	0.2	33.2
2	65	1.5	35.6
3	112	2.1	36.1
4	220	4.0	34.5
5	461	8.1	38.9
6	804	13.7	36.3
7	1347	22.7	39.2
8	2524	43.3	37.7

Сравнение таблицы 1 и таблицы 2 показывает, что асимптотически оптимальный комбинаторный код дает лучший коэффициент сжатия и лучшее время кодирования, чем обычный комбинаторный код.

Заключение. Эффективной областью применения алгоритма универсального комбинаторного кодирования может быть сжатие данных с неизвестной статистикой, в том числе, различных потоков, таких как видео, звук, телеметрия.

СПИСОК ЛИТЕРАТУРЫ

1. Бабкин, В. Ф. Метод универсального кодирования независимых сообщений неэкспоненциальной трудоемкости [Текст] / В. Ф. Бабкин // Проблемы передачи информации. – 1971, т. 7, № 4. – С. 13-21.
2. Davisson, L. Efficient Universal Noiseless Codes / L. Davisson, R. McEliece, M. Pursley, M. Wallace. // IEEE Trans. Inf. Th. – 1981, v. 27, № 3. – P. 269-279.
3. Krichevsky, R. E. The Performance of Universal Encoding / R. E. Krichevsky, V. K. Trofimov //

КОДИРОВАНИЕ НЕИЗВЕСТНОГО СТАЦИОНАРНОГО ИСТОЧНИКА СИМВОЛАМИ НЕРАВНОЙ ДЛИТЕЛЬНОСТИ

IEEE Trans. Inf. Th. – 1981, v. 27, № 2. – P.199-207.

4. 4. Кнут Д. Искусство программирования для ЭВМ, сортировка и поиск [Текст] / Д. Кнут. - М.: Мир. – 1978.

Научный сотрудник, к.ф.-м.н. **Бакулина М.П.**
тел. 8-961-215-79-36, marina@rav.sssc.ru - Институт Вычислительной Математики и Математической Геофизики СО РАН

УДК 621.391

КОДИРОВАНИЕ НЕИЗВЕСТНОГО СТАЦИОНАРНОГО ИСТОЧНИКА СИМВОЛАМИ НЕРАВНОЙ ДЛИТЕЛЬНОСТИ

В.К. Трофимов, Т.В. Храмова

Предложен метод слабо универсального кодирования множества стационарных источников при условии, что символы кодового алфавита имеют неравные длительности.

Ключевые слова: кодирование, избыточность кодирования, энтропия

Изучение методов кодирования информации, с целью уменьшения занимаемого ею объема является актуальной задачей информатики. Данная работа посвящена изучению оптимальных методов сжатия информации, порожденной неизвестным дискретным стационарным источником.

Рассмотрим источник сообщений Θ , генерирующий последовательность из букв некоторого конечного алфавита

$$X = \{x_1, x_2, \dots, x_k\}.$$

Последовательность должна быть закодирована и передана по каналу связи. Для решения нашей задачи оптимального кодирования, разобьем исходящую последовательность букв источника на блоки (слова) фиксированной длины N . Процедура кодирования источника заключается в том, что каждому блоку $w \in X^N$ ставится в соответствие некоторое кодовое слово $\varphi(w) \in Y^*$ из букв кодового алфавита

$$Y = \{y_1, y_2, \dots, y_m\}$$

(здесь Y^* обозначает множество всевозможных последовательностей из элементов множества Y). Кодирование, при котором блокам источника ставится в соответствие кодовое слов нефиксированной длины называется *равномерным по входу*. Разные буквы кодового алфавита имеют разную *длительность* или, другими словами, *стоимость передачи*:

$t_j = t(y_j)$, $j = \overline{1, m}$. Таким образом, каждому кодовому алфавиту можно поставить в соответствие *вектор длительностей* букв $\bar{t} = (t_1, t_2, \dots, t_m)$, ($t_j = t(y_j)$, $j = \overline{1, m}$).

В частности, длительности кодовых символов могут быть одинаковы, в этом случае соответствующий алфавиту вектор обозначим $\bar{t}_1 = (\underbrace{1, 1, \dots, 1}_m)$.

Самым популярным примером кода с неравными длительностями является код Морзе, актуальность которого не теряется и в наше время — на основе кода Морзе созданы широко используемые штрих-коды.

Длительностью кодового слова будем считать величину, равную сумме длительностей входящих в слово букв:

$$l(\varphi(w), \bar{t}) = \sum_{y \in \varphi(w)} t(y).$$

Стоимость кодирования определяется как отношение средней длительности кодового слова к средней длительности слова источника, и, в рассматриваемом случае, принимает вид:

$$L(N, \Theta, \varphi, \bar{t}) = \frac{1}{N} \sum_{w \in X^N} p_{\Theta}(w) l(\varphi(w), \bar{t}). \quad (1)$$

Эффективность метода кодирования $\varphi: X^N \mapsto Y$ определяется *избыточностью*

$$R(N, \Theta, \varphi, \bar{t}) = L(N, \Theta, \varphi, \bar{t}) - H(\Theta) / C(\bar{t}), \quad (2)$$

где $H(\Theta)$ — *энтропия источника*, определяемая законом распределения вероятностей появления букв алфавита X на выходе источника Θ , $C(\bar{t})$ — *пропускная способность* канала передачи информации, зависящая только от кодового алфавита Y .

Изучению эффективных методов кодирования посвящено множество работ. В случае известного источника, при равных