

Оптимальная аппроксимация частотных вероятностей

Д. В. Сушко

Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, Москва, Россия

Поступила в редколлегию 12.13.2018

Аннотация—Представлено решение задачи оптимальной аппроксимации заданного дискретного распределения вероятностей с конечным числом значений в классах дискретных линейных и дискретных экспоненциальных распределений вероятностей. Мерой качества аппроксимации является кодовая избыточность приближенного распределения относительно исходного распределения, и оптимальным приближением в классе является то распределение, для которого кодовая избыточность принимает минимальное значение. Рассмотрен пример, который демонстрирует высокую эффективность метода универсального комбинаторного кодирования, основанного на оптимальной аппроксимации частотных распределений, в задачах обратимого сжатия данных посредством арифметического кодирования.

КЛЮЧЕВЫЕ СЛОВА: аппроксимация распределений, арифметическое кодирование, универсальное кодирование, обратимое сжатие данных, компьютерная томограмма.

1. ВВЕДЕНИЕ

Традиционная схема решения задач обратимого (без потерь) сжатия цифровых данных (см., например, [1]) заключается в применении к исходному массиву данных некоторого обратимого преобразования, обеспечивающего декорреляцию его отсчетов, а также уменьшение диапазона их значений, и последующего кодирования полученного таким образом массива как последовательности независимых отсчетов. В качестве метода кодирования в настоящее время все чаще используется арифметическое кодирование (см., например, [2]).

Рассмотрим конечное множество $\mathcal{A} = \{a\}$ (алфавит), состоящее из A элементов (букв), которые представляют собой целые числа из некоторого известного априори диапазона $[a_-, a_+]$:

$$\mathcal{A} = \{a_-, a_- + 1, \dots, a_+\}, \quad A = a_+ - a_- + 1.$$

Пусть $\mathbf{x} = \{x_n\}_{n=0}^{N-1}$ ($N \gg 1$) – последовательность букв алфавита \mathcal{A} , т.е. последовательность целых чисел из заданного диапазона. Арифметическое кодирование последовательности \mathbf{x} , игнорирующее возможную корреляцию ее значений, осуществляется следующим образом. Прежде всего, для данной последовательности \mathbf{x} выбирается некоторое распределение вероятностей $\mathbf{q}_{\mathbf{x}} = \mathbf{q} = \{q(a), a \in \mathcal{A}\}$, т.е. набор неотрицательных вещественных чисел с суммой, равной 1 ($\sum_{a \in \mathcal{A}} q(a) = 1$). Дополнительно предполагается, что выполнение равенства $q(a) = 0$ для некоторого a возможно лишь в том случае, если ни один элемент последовательности \mathbf{x} априори не может принять это значение a . Такое распределение называется *кодovým* распределением вероятностей. Кодирование представляет собой рекуррентную процедуру. На n -ом шаге кодер просматривает очередное значение x_n , вычисляет, используя заданное кодовое распределение, текущую кодовую вероятность (кодovую вероятность начального участка последовательности до n -ого члена включительно) по формуле

$$Q_n = q(x_n) \cdot Q_{n-1} \quad (Q_{-1} \doteq 1)$$

и формирует текущее двоичное кодовое слово длины $\lceil -\log Q_n/2 \rceil$ битов (бт). Здесь и далее $\log(\cdot) = \log_2(\cdot)$ – двоичный логарифм, $\lceil \cdot \rceil$ – результат округления вещественного числа до ближайшего целого вверх (ceil). По завершении всей процедуры кодирования мы имеем двоичное кодовое слово (результат сжатия) для всей последовательности, длина которого равна

$$L \equiv L(\mathbf{x}) = \left\lceil -\log \frac{Q_{N-1}}{2} \right\rceil = \left\lceil -\log \prod_{n=0}^{N-1} q(x_n) + \frac{1}{2} \right\rceil = \left\lceil \sum_{n=0}^{N-1} -\log q(x_n) + \frac{1}{2} \right\rceil. \quad (1)$$

Детали формирования кодового слова по кодовым вероятностям не принципиальны для нашего рассмотрения и опускаются. Для декодирования (восстановления исходной последовательности по кодовому слову) необходимо знать то кодовое распределение вероятностей, которое было использовано при кодировании. Отметим, что декодирование осуществляется последовательно и без задержки, т.е. в момент восстановления очередного значения x_n декодеру уже известны все предыдущие значения $\{x_0, \dots, x_{n-1}\}$.

Таким образом, ключевую роль в процессе арифметического кодирования играют кодовые распределения вероятностей, выбор которых определяет длину кодового слова, т.е. степень сжатия исходных данных. При этом построение кодовых распределений, обеспечивающих получение возможно более коротких кодовых слов для входных данных с неизвестной (не полностью известной) статистикой, – задача универсального кодирования. Фундаментальным трудом, в котором подробно рассмотрены как теория, так и практика применения универсального кодирования, является монография [3].

Скоростью кодирования V (средней скоростью кодирования) называется отношение длины кодового слова $L(\mathbf{x})$ к числу элементов N кодируемой последовательности; единица измерения скорости кодирования – бит/пиксель (бт/п). Из формулы (1) для скорости арифметического кодирования с точностью до малого члена $\sim 1/N$ легко имеем

$$V_A(\mathbf{x}) = \sum_{x \in \mathcal{A}} \frac{N(x)}{N} [-\log q(x)], \quad (2)$$

где $N(x)$ – число элементов последовательности, принимающих значение x , а сумма берется по всем встречающимся в данной последовательности значениям. Далее будем использовать соглашение о том, что $0 \cdot \log 0 = 0$. В таком случае сумму в (2) можно распространить на все множество значений \mathcal{A} . Действительно, в силу предъявляемых к кодовым вероятностям требований, равенство $q(x) = 0$ влечет $N(x) = 0$.

Величины $N(x)/N$, $x \in \mathcal{A}$, образуют *частотное (эмпирическое)* распределение вероятностей значений последовательности \mathbf{x} . Используя для этого частотного распределения обозначение $\mathbf{f} \equiv \mathbf{f}_\mathbf{x} = \{f(x), x \in \mathcal{A}\}$, перепишем формулу (2) для скорости кодирования в виде

$$V_A(\mathbf{x}) = H(\mathbf{x}) + R_A(\mathbf{x}),$$

где

$$H(\mathbf{x}) \doteq \sum_{x \in \mathcal{A}} f(x) [-\log f(x)], \quad R_A(\mathbf{x}) \doteq V_A(\mathbf{x}) - H(\mathbf{x}) = \sum_{x \in \mathcal{A}} f(x) \left[-\log \frac{q(x)}{f(x)} \right]. \quad (3)$$

Величина H в (3) – *квазиэнтропия* (или *эмпирическая энтропия*) последовательности. Квазиэнтропия не зависит от кодового распределения и, очевидно, является неотрицательной ($H \geq 0$). Величина R_A (3) – *избыточность* (*индивидуальная избыточность*) арифметического кодирования. Нетрудно показать, что избыточность также является неотрицательной ($R_A \geq 0$) и обращается в нуль тогда и только тогда, когда кодовое распределение вероятностей совпадает с частотным распределением $\mathbf{q} = \mathbf{f}$, т.е. $q(x) = f(x)$ для всех $x \in \mathcal{A}$.

Таким образом, минимальная скорость арифметического кодирования достигается тогда и только тогда, когда $\mathbf{q} = \mathbf{f}$, т.е. в качестве кодового распределения используется частотное распределение. Частотное распределение \mathbf{f} априори не известно, но может быть вычислено кодером по исходным данным \mathbf{x} , что позволяет использовать упрощенное комбинаторное универсальное кодирование (см., например, [4]).

Кодовое слово комбинаторного кода состоит из двух частей. Первая часть (преамбула) содержит значения $N(x)$ для всех $x \in \mathcal{A}$, которые вычисляются в процессе кодирования. Длина преамбулы равна $(A - 1)\lceil \log N \rceil$ бт. Вторая часть (основная) – результат арифметического кодирования последовательности \mathbf{x} с помощью частотного распределения \mathbf{f} . Получив кодовое слово, декодер выделяет преамбулу, «считывает» значения $N(x)$ и реконструирует частотное распределение \mathbf{f} , использованное при кодировании. Это позволяет однозначно декодировать основную часть кодового слова и восстановить тем самым исходную последовательность.

Определим полную избыточность R процедуры кодирования как разность скорости кодирования V и квазиэнтропии H :

$$R(\mathbf{x}) \doteq V(\mathbf{x}) - H(\mathbf{x}).$$

Отметим, что избыточность любой процедуры, использующей арифметическое кодирование и преамбулу (для передачи декодеру необходимой информации), состоит из избыточности арифметического кодирования $R_A(\mathbf{x})$ (см. (3)) и избыточности *передачи* R_T , определяемой как отношение длины преамбулы к числу элементов кодируемой последовательности:

$$R(\mathbf{x}) = R_A(\mathbf{x}) + R_T. \quad (4)$$

Для описанной выше процедуры комбинаторного кодирования избыточность арифметического кодирования (первое слагаемое в (4)) обращается в нуль, а избыточность передачи (второе слагаемое в (4)) равна

$$R = R_T = \frac{(A - 1)\lceil \log N \rceil}{N}. \quad (5)$$

При большом алфавите избыточность передачи может оказаться недопустимо большой. Так в примере, который рассмотрен в разделе 3, $A = 2^{13} - 1$, $N = 2^{16}$ и для избыточности передачи из (5) имеем оценку $R_T \sim 1.99$ бт/п, при том, что квазиэнтропия данных составляет $H \sim 8.23$ бт/п.

Использование более совершенных методов универсального кодирования (см. [5], [6]) позволяет уменьшить избыточность; это, однако, не всегда решает проблему. Поэтому в работе [7] впервые был предложен метод построения кодовых распределений посредством аппроксимации частотных распределений, который может рассматриваться как некоторый вариант универсального комбинаторного кодирования. Идея метода заключается в следующем. Аппроксимируем известное на этапе кодирования частотное распределение \mathbf{f} некоторым распределением \mathbf{p} из предопределенного класса \mathcal{P} такого, что каждое распределение из класса однозначно определяется значениями некоторого небольшого числа параметров. Используемый класс распределений известен как кодеру, так и декодеру. Кодер строит аппроксимацию $\mathbf{p} \in \mathcal{P}$, использует построенное распределение в качестве кодового распределения ($\mathbf{q} = \mathbf{p}$) и включает значения соответствующих параметров в преамбулу. Декодер «считывает» значение параметров из преамбулы и строит по ним распределение $\mathbf{q} = \mathbf{p} \in \mathcal{P}$, использованное при кодировании. Это позволяет однозначно декодировать основную часть кодового слова и восстановить тем самым исходную последовательность.

Применение описанного метода позволяет кардинальным образом сократить избыточность передачи. Действительно, вместо частотных вероятностей всех возможных значений $a \in \mathcal{A}$ преамбула теперь содержит лишь значения небольшого числа параметров. В примере, который рассмотрен в разделе 3, избыточность передачи составляет менее 0.005 бт/п. Однако теперь

избыточность арифметического кодирования R_A не равна нулю и зависит от распределения, выбранного в качестве кодового распределения при аппроксимации частотного распределения. Поэтому задача уменьшения избыточности кодирования метода связана с решением следующей оптимизационной задачи: в заданном классе \mathcal{P} по заданному распределению f найти то распределение p , для которого избыточность R_A принимает минимальное значение.

Сделаем два важных замечания. Во-первых, для того чтобы решение оптимизационной задачи могло быть использовано на практике, оно должно быть конструктивным и давать «быстрый» численный алгоритм построения оптимального приближения. Во-вторых, использование оптимального распределения при кодировании само по себе еще не гарантирует приемлемой с практической точки зрения избыточности кодирования, поскольку даже минимальная избыточность может оказаться, вообще говоря, слишком большой.

Работа имеет следующую структуру. В разделе 2 приведена точная формулировка задачи оптимальной аппроксимации в заданном классе распределений и представлены решения задач линейной оптимизации (оптимизации в классе дискретных линейных распределений) и экспоненциальной оптимизации (оптимизации в классе дискретных экспоненциальных распределений). В разделе 3 приведен пример, демонстрирующий эффективность описанного выше метода универсального комбинаторного кодирования, основанного на аппроксимации частотных распределений.

2. ОПТИМАЛЬНАЯ АППРОКСИМАЦИЯ РАСПРЕДЕЛЕНИЙ

Пусть $f = \{f(k)\}_{k=0}^{K-1}$, $K \geq 2$, – распределение вероятностей, т.е. набор неотрицательных вещественных чисел, сумма которых равна 1 ($f(k) \geq 0$, $\sum_{k=0}^{K-1} f(k) = 1$). Функционал

$$r(p|f) = \sum_{k=0}^{K-1} f(k) \left[-\log \frac{p(k)}{f(k)} \right] \quad (6)$$

называется *избыточностью* распределения p относительно распределения f (другое название – *кодирование*). Используя элементарное неравенство $-\log(x) \geq 1 - x$, $x > 0$, которое обращается в равенство только в при $x = 1$, нетрудно показать, что избыточность (6) неотрицательна ($r(p|f) \geq 0$) для любых распределений f , p и обращается в нуль тогда и только тогда, когда распределения равны ($p = f$, т.е. $p(k) = f(k)$ для всех k). Отметим, кроме того, что если $p(k) = 0$ и $f(k) \neq 0$ хотя бы для одного k , то $r(p|f) = +\infty$.

Пусть \mathcal{P} – некоторый заданный класс (множество) распределений. Задача **оптимальной аппроксимации** заданного распределения вероятностей f в классе \mathcal{P} формулируется следующим образом: найти распределение $\tilde{p} \in \mathcal{P}$, для которого избыточность $r(\tilde{p}|f)$, заданная формулой (6), принимает минимальное значение.

2.1. Линейная аппроксимация

Рассмотрим распределение вероятностей $p_{0,\rho}$ следующего вида:

$$p_{0,\rho}(k) = \frac{1}{K} \left(1 + \frac{\rho - 1}{\rho + 1} - \frac{2}{K - 1} \frac{\rho - 1}{\rho + 1} k \right), \quad k = 0, 1, \dots, K - 1, \quad 1 \leq \rho \leq +\infty. \quad (7)$$

Нетрудно видеть, что любая последовательность вида (7) представляет собой арифметическую прогрессию, которая состоит из неотрицательных членов, является невозрастающей и имеет сумму, равную 1, т.е. действительно является распределением вероятностей. Верно и обратное утверждение: любая арифметическая прогрессия, обладающая перечисленными выше свойствами, представима в виде (7) со значением параметра ρ , равным отношению начального

члена прогрессии к ее конечному члену ($\rho = p(0)/p(K-1)$). В случае $\rho = 1$ имеем арифметическую прогрессию с постоянными членами, т.е. равномерное распределение. В случае $\rho = +\infty$ имеем арифметическую прогрессию, последний член которой обращается в нуль ($p(K-1)=0$).

Распределения вида (7) будем называть дискретными линейными распределениями. Для класса всех дискретных линейных распределений будем использовать обозначение \mathcal{P}_0 .

Сформулируем задачу **оптимальной линейной аппроксимации**. Задано распределение вероятностей f . Найти в классе дискретных линейных распределений \mathcal{P}_0 , т.е. в классе распределений вида (7), то распределение \tilde{p} , для которого избыточность $r(\tilde{p} | f)$, заданная формулой (6), принимает минимальное значение.

Приступим к решению задачи. Прежде всего, изменим параметризацию распределений класса \mathcal{P}_0 , сделав в формуле (7) взаимнооднозначную замену переменной

$$\rho \rightarrow \alpha = \alpha(\rho) = (\rho - 1)/(\rho + 1), \quad 0 \leq \alpha \leq 1 \quad (\alpha(1) = 0, \alpha(+\infty) = 1),$$

после чего перепишем формулу (6) в следующем виде:

$$r(p_{0,\alpha} | f) = r(\alpha) = \frac{1}{\ln 2} \sum_{k=0}^{K-1} f(k) \left[-\ln \frac{\frac{1}{K} \left(1 + \frac{K-1-2k}{K-1} \alpha\right)}{f(k)} \right]. \quad (8)$$

Исследуем поведение избыточности (8) как функции переменной α . Дифференцируя (8), получаем следующие формулы для первой и второй производных функции $r(\alpha)$:

$$r'(\alpha) = -\frac{1}{\ln 2} \sum_{k=0}^{K-1} f(k) \left[1 - \frac{2k}{K-1} \right] \left[1 + \frac{K-1-2k}{K-1} \alpha \right]^{-1} \quad (9)$$

и

$$r''(\alpha) = \frac{1}{\ln 2} \sum_{k=0}^{K-1} f(k) \left[1 - \frac{2k}{K-1} \right]^2 \left[1 + \frac{K-1-2k}{K-1} \alpha \right]^{-2}. \quad (10)$$

Формула (10) показывает, что вторая производная всегда неотрицательна ($r''(\alpha) \geq 0$) и обращается в нуль только в следующем вырожденном случае: K – нечетное, $f(\frac{K-1}{2}) = 1$ и все остальные $f(k)$ равны нулю. При этом из формул (8)–(10) сразу следует, что $r''(\alpha) \equiv r'(\alpha) \equiv 0$ и $r(\alpha) \equiv \log K$. Таким образом, в вырожденном случае избыточность постоянна на всем классе дискретных линейных распределений \mathcal{P}_0 , и в качестве решения задачи можно взять распределение $\tilde{p} = p_{0,\alpha}$ при $\alpha = 0$, т.е. равномерное распределение.

В невырожденном случае вторая производная $r''(\alpha)$ строго положительна ($r''(\alpha) > 0$) на всем отрезке $[0, 1]$, следовательно, первая производная $r'(\alpha)$ является строго возрастающей функцией на всем этом отрезке. Поэтому могут представиться лишь следующие три варианта:

1. $r'(0) \geq 0$ (при этом $r'(\alpha) > 0$ для всех $\alpha \in (0, 1]$).
2. $r'(1) \leq 0$ (при этом $r'(\alpha) < 0$ для всех $\alpha \in [0, 1)$).
3. $r'(0) < 0$ и $r'(1) > 0$.

В первом случае избыточность $r(\alpha)$ является возрастающей функцией на отрезке $[0, 1]$ и принимает минимальное значение в точке $\alpha = 0$. Решением задачи в этом случае является распределение $\tilde{p} = p_{0,\alpha}$ при $\alpha = 0$, т.е. равномерное распределение.

Во втором случае избыточность $r(\alpha)$ является убывающей функцией на отрезке $[0, 1]$ и принимает минимальное значение в точке $\alpha = 1$. Решением задачи в этом случае является распределение $\tilde{p} = p_{0,\alpha}$ при $\alpha = 1$, т.е. линейное распределение, у которого $p(K-1) = 0$.

Рассмотрим, наконец, третий случай. В этом случае избыточность $r(\alpha)$ принимает минимальное значение во внутренней точке $\tilde{\alpha}$ отрезка $[0, 1]$, которая является единственным корнем уравнения $r'(\alpha) = 0$. Решая это уравнение методом деления отрезка пополам, получаем

$$\tilde{\alpha} = \lim_{i \rightarrow \infty} \tilde{\alpha}_i, \quad \tilde{\alpha}_i = \frac{1}{2}(\alpha_i^0 + \alpha_i^1), \tag{11}$$

где $i = 0, 1, \dots$ – номер итерации, $\alpha_0^0 = 0, \alpha_0^1 = 1$, а значения α_i^0, α_i^1 для i -ой итерации ($i > 0$) вычисляются с помощью следующих рекуррентных соотношений:

$$\begin{cases} \alpha_i^0 = \tilde{\alpha}_{i-1}, \alpha_i^1 = \alpha_{i-1}^1, & \text{если } r'(\tilde{\alpha}_{i-1}) < 0; \\ \alpha_i^0 = \tilde{\alpha}_{i-1}, \alpha_i^1 = \tilde{\alpha}_{i-1}, & \text{если } r'(\tilde{\alpha}_{i-1}) = 0; \\ \alpha_i^0 = \alpha_{i-1}^0, \alpha_i^1 = \tilde{\alpha}_{i-1}, & \text{если } r'(\tilde{\alpha}_{i-1}) > 0. \end{cases} \tag{12}$$

Точным решением задачи в этом случае является распределение $\tilde{\mathbf{p}} = \mathbf{p}_{0, \tilde{\alpha}}$.

Как видно из формулы (12), мы получаем точное решение уравнения и соответствующее точное решение задачи за конечное число итераций только в той крайне редкой ситуации, когда на некоторой (i -ой) итерации величина $r'(\tilde{\alpha}_{i-1})$ обращается в нуль. В типичной ситуации мы имеем решение уравнения как предел последовательности приближений $\tilde{\alpha}_i$ и, кроме того, на каждой итерации имеем приближенное решение задачи $\tilde{\mathbf{p}} = \mathbf{p}_{0, \tilde{\alpha}_i}$, причем для приближения, построенного на i -ой итерации, справедлива следующая оценка:

$$r(\tilde{\alpha}_i) - r(\tilde{\alpha}) \leq 2^{-i}(\alpha_0^1 - \alpha_0^0) \max\{|r'(\alpha_i^0)|, r'(\alpha_i^1)\}. \tag{13}$$

Эта оценка используется для прерывания итерационного процесса при численном решении задачи: вычисления прекращаются, когда правая часть неравенства (13) становится меньше некоторого наперед заданного значения $\varepsilon > 0$. При этом конкретное значение ε должно выбираться исходя из практических требований.

В заключение раздела сделаем следующее замечание. Для того чтобы распределение $\mathbf{p}_{0,1}$, т.е. линейное распределение, у которого $p(K - 1) = 0$, было решением задачи оптимальной линейной аппроксимации, необходимо выполнение условия $f(K - 1) = 0$. Действительно, если $f(K - 1) > 0$, то, как легко видеть из (8)–(10), $r(1) = r'(1) = r''(1) = +\infty$.

2.2. Экспоненциальная аппроксимация

Рассмотрим распределение вероятностей $\mathbf{p}_{\nu, \rho}$ следующего вида:

$$p_{\nu, \rho}(k) = \frac{1}{\sum_{k=0}^{K-1} e^{-(\ln \rho) \left(\frac{k}{K-1}\right)^\nu}} e^{-(\ln \rho) \left(\frac{k}{K-1}\right)^\nu}, \quad k = 0, 1, \dots, K - 1, \quad 1 \leq \rho \leq +\infty, \quad \nu > 0. \tag{14}$$

Любая последовательность вида (14) является распределением вероятностей (все ее члены положительны и их сумма равна единице). Ясно также, что последовательность является невозрастающей, а отношение начального члена последовательности к ее последнему члену равно ρ ($\rho = p(0)/p(K - 1)$). Отметим, что в частном случае $\nu = 1$, последовательность представляет собой геометрическую прогрессию со знаменателем $\sqrt[K-1]{1/\rho}$. При любом ν имеем в случае $\rho = 1$ равномерное распределение, а в случае $\rho = +\infty$ – распределение следующего вида: $p(0) = 1$ и $p(k) = 0$ при $k = 1, \dots, (K - 1)$.

Распределения вида (14) будем называть дискретными экспоненциальными распределениями. Для класса, состоящего из дискретных экспоненциальных распределений с фиксированным значением ν , будем использовать обозначение \mathcal{P}_ν .

Сформулируем задачу **оптимальной экспоненциальной аппроксимации**. Задано распределение вероятностей f . Найти в классе дискретных экспоненциальных распределений \mathcal{P}_ν , т.е. в классе распределений вида (14) с фиксированным значением ν ($\nu > 0$), то распределение \tilde{p} , для которого избыточность $r(\tilde{p} | f)$, заданная формулой (6), принимает минимальное значение.

Приступим к решению задачи. Изменим параметризацию распределений класса \mathcal{P}_ν , сделав в формуле (14) взаимнооднозначную замену переменной

$$\rho \rightarrow \alpha = \alpha(\rho) = \ln \rho, \quad 0 \leq \alpha \leq +\infty \quad (\alpha(1) = 0, \alpha(+\infty) = +\infty),$$

и введем следующие обозначения:

$$\lambda_k = \lambda_k(\nu, K) = \left(\frac{k}{K-1} \right)^\nu, \quad k = 0, 1, \dots, K-1 \quad (0 = \lambda_0 < \lambda_1 < \dots < \lambda_{K-1} = 1). \quad (15)$$

Теперь формулу (6) можно переписать в виде

$$r(p_{\nu, \alpha} | f) = r(\alpha) = \frac{1}{\ln 2} \sum_{k=0}^{K-1} f(k)[S(\alpha) + \lambda_k \alpha] + \sum_{k=0}^{K-1} f(k) \log f(k), \quad (16)$$

где

$$S(\alpha) = S(\alpha | \nu, K) = \ln \left(\sum_{k=0}^{K-1} e^{-\alpha \lambda_k} \right). \quad (17)$$

Исследуем поведение избыточности (16) как функции переменной α . Дифференцируя (16), получаем следующие формулы для первой и второй производных функции $r(\alpha)$:

$$r'(\alpha) = -\frac{1}{\ln 2} e^{-S(\alpha)} \sum_{k=0}^{K-1} \lambda_k e^{-\alpha \lambda_k} + \frac{1}{\ln 2} \sum_{k=0}^{K-1} f(k) \lambda_k \quad (18)$$

и

$$r''(\alpha) = \frac{1}{\ln 2} e^{-2S(\alpha)} \left[\sum_{k=0}^{K-1} \lambda_k^2 e^{-\alpha \lambda_k} \times \sum_{k=0}^{K-1} e^{-\alpha \lambda_k} - \sum_{k=0}^{K-1} \lambda_k e^{-\alpha \lambda_k} \times \sum_{k=0}^{K-1} \lambda_k e^{-\alpha \lambda_k} \right]. \quad (19)$$

Заметим, что вторая производная избыточности $r''(\alpha)$ не зависит от распределения f , которое подлежит аппроксимации.

Рассмотрим выражение в квадратных скобках формулы (19). Это выражение представляет собой сумму отдельных слагаемых, пропорциональных $e^{-\alpha(\lambda_i + \lambda_j)}$ ($0 \leq i \leq j \leq K-1$). Приведем подобные члены, т.е. соберем вместе члены с одинаковым показателем в экспоненте и вычислим коэффициент $c_{i,j}$ перед экспонентой. Для показателей вида $-2\alpha\lambda_i$ ($i = j$) имеем $c_{i,j} = \lambda_i^2 - \lambda_i^2 = 0$. Для показателей вида $-\alpha(\lambda_i + \lambda_j)$ ($i < j$) имеем $c_{i,j} = \lambda_i^2 + \lambda_j^2 - 2\lambda_i\lambda_j > 0$, поскольку $\lambda_i < \lambda_j$. Таким образом, показано, что при любом конечном значении α ($0 \leq \alpha < +\infty$) рассматриваемое выражение представимо в виде суммы положительных слагаемых и, следовательно, положительно. Переходя к пределу $\alpha \rightarrow +\infty$, непосредственно убеждаемся, что при $\alpha = +\infty$ выражение обращается в нуль.

Из полученного выше результата немедленно вытекает, что вторая производная избыточности $r''(\alpha)$ неотрицательна и обращается в нуль только при $\alpha = +\infty$. Это, в свою очередь, означает, что первая производная избыточности $r'(\alpha)$ является строго возрастающей функцией переменной α ; при этом из (18) следует, что

$$r'(0) = -\frac{1}{\ln 2} \frac{1}{K} \sum_{k=0}^{K-1} \lambda_k + \frac{1}{\ln 2} \sum_{k=0}^{K-1} f(k) \lambda_k, \quad r'(+\infty) = +\frac{1}{\ln 2} \sum_{k=0}^{K-1} f(k) \lambda_k. \quad (20)$$

В зависимости от распределения f , которое подлежит аппроксимации, могут представиться следующие три случая, которые должны быть рассмотрены по-отдельности:

1. $\sum_{k=0}^{K-1} f(k)\lambda_k = 0$ (при этом $\sum_{k=0}^{K-1} f(k)\lambda_k < K^{-1} \sum_{k=0}^{K-1} \lambda_k$).
2. $\sum_{k=0}^{K-1} f(k)\lambda_k > 0$ и $K^{-1} \sum_{k=0}^{K-1} \lambda_k \leq \sum_{k=0}^{K-1} f(k)\lambda_k$.
3. $\sum_{k=0}^{K-1} f(k)\lambda_k > 0$ и $K^{-1} \sum_{k=0}^{K-1} \lambda_k > \sum_{k=0}^{K-1} f(k)\lambda_k$.

Первый случай является вырожденным. Действительно, равенство $\sum_{k=0}^{K-1} f(k)\lambda_k = 0$ имеет место лишь в случае вырожденного распределения, у которого $f(0) = 1$ и все остальные $f(k)$ равны нулю. Решением задачи для такого распределения, очевидным образом, является совпадающее с ним распределение $\tilde{\mathbf{p}} = \mathbf{p}_{\nu, \alpha}$ при $\alpha = +\infty$.

Рассмотрим второй случай. Если $K^{-1} \sum_{k=0}^{K-1} \lambda_k \leq \sum_{k=0}^{K-1} f(k)\lambda_k$, то, как видно из (20), $r'(0) \geq 0$. Следовательно, возрастающая функция $r'(\alpha)$ положительна везде, кроме, быть может, точки нуля. Это, в свою очередь, означает, что избыточность $r(\alpha)$ возрастает, принимая минимальное значение в точке $\alpha = 0$. Решением задачи в этом случае является распределение $\tilde{\mathbf{p}} = \mathbf{p}_{\nu, \alpha}$ при $\alpha = 0$, т.е. равномерное распределение.

Приступим к рассмотрению третьего случая. Поскольку $K^{-1} \sum_{k=0}^{K-1} \lambda_k > \sum_{k=0}^{K-1} f(k)\lambda_k > 0$, то, как видно из (20), $r'(0) < 0$ и $r'(+\infty) > 0$. Следовательно, возрастающая функция $r'(\alpha)$ принимает значение нуль в единственной точке $\tilde{\alpha}$ внутри области определения ($0 < \alpha < +\infty$), которая является корнем уравнения $r'(\alpha) = 0$. Рассмотрим величину $r'(\tilde{\alpha})$, где

$$\tilde{\alpha} = (K - 1)^\nu \ln \frac{K - 1}{\sum_{k=0}^{K-1} f(k)\lambda_k}. \quad (21)$$

Заметим, что поскольку в рассматриваемом случае $\sum_{k=0}^{K-1} f(k)\lambda_k > 0$ и, кроме того, очевидно, что $\sum_{k=0}^{K-1} f(k)\lambda_k < 1$, то $\tilde{\alpha} > 1$. Далее, используя формулы (18), (17), (15) и (21), убеждаемся, что имеет место следующая цепочка соотношений:

$$\begin{aligned} \ln 2 \cdot r'(\tilde{\alpha}) &= \sum_{k=0}^{K-1} f(k)\lambda_k - e^{-S(\tilde{\alpha})} \sum_{k=0}^{K-1} \lambda_k e^{-\tilde{\alpha}\lambda_k} > \sum_{k=0}^{K-1} f(k)\lambda_k - (K - 1)e^{-\tilde{\alpha}\lambda_1} = \\ &= \sum_{k=0}^{K-1} f(k)\lambda_k - (K - 1)e^{-(K-1)^\nu \cdot \ln \frac{K-1}{\sum_{k=0}^{K-1} f(k)\lambda_k} \cdot \left(\frac{1}{K-1}\right)^\nu} = \sum_{k=0}^{K-1} f(k)\lambda_k - \sum_{k=0}^{K-1} f(k)\lambda_k = 0. \end{aligned}$$

Таким образом, $r'(\tilde{\alpha}) > 0$ и мы можем решать уравнение $r'(\alpha) = 0$ методом деления отрезка пополам, используя значение $\tilde{\alpha}$ из формулы (21) в качестве правого конца отрезка ($\alpha_0^1 = \tilde{\alpha}$) на нулевой итерации. Процедура решения ничем не отличается от соответствующей рекуррентной процедуры в линейном случае (см. раздел 2.1). Рекуррентные соотношения имеют вид (12), корень $\tilde{\alpha}$ и последовательность приближений $\tilde{\alpha}_i$ задаются формулой (11). Точным решением задачи является распределение $\tilde{\mathbf{p}} = \mathbf{p}_{\nu, \tilde{\alpha}}$, и имеет место оценка (13).

Итак, задача минимизации избыточности (6) в классе \mathcal{P}_ν дискретных экспоненциальных распределений вида (14) с фиксированным значением ν полностью решена. Решение является конструктивным и содержит эффективный (не связанный с перебором) алгоритм построения распределения, которое минимизирует избыточность с любой наперед заданной точностью.

В заключение рассмотрим задачу минимизации избыточности (6) в классе распределений $\mathcal{P}_{[\nu_0, \nu_1]} = \bigcup_{\nu_0 \leq \nu \leq \nu_1} \mathcal{P}_\nu$ ($0 < \nu_0 < \nu_1 < +\infty$). Поскольку

$$\min_{\mathbf{p} \in \mathcal{P}_{[\nu_0, \nu_1]}} r(\mathbf{p} | \mathbf{f}) = \min_{\nu_0 \leq \nu \leq \nu_1} \left[\min_{\mathbf{p} \in \mathcal{P}_\nu} r(\mathbf{p} | \mathbf{f}) \right],$$

то такая задача сводится к нахождению минимума функции $r(\nu)$ на отрезке $[\nu_0, \nu_1]$, причем функцию $r(\nu) = r(\nu, \tilde{\alpha}(\nu)) = \min_{\mathbf{p} \in \mathcal{P}_\nu} r(\mathbf{p} | \mathbf{f})$ можно считать известной (соответствующая задача уже решена). Функция $r(\nu)$ имеет, вообще говоря, сложное нерегулярное поведение, о чем

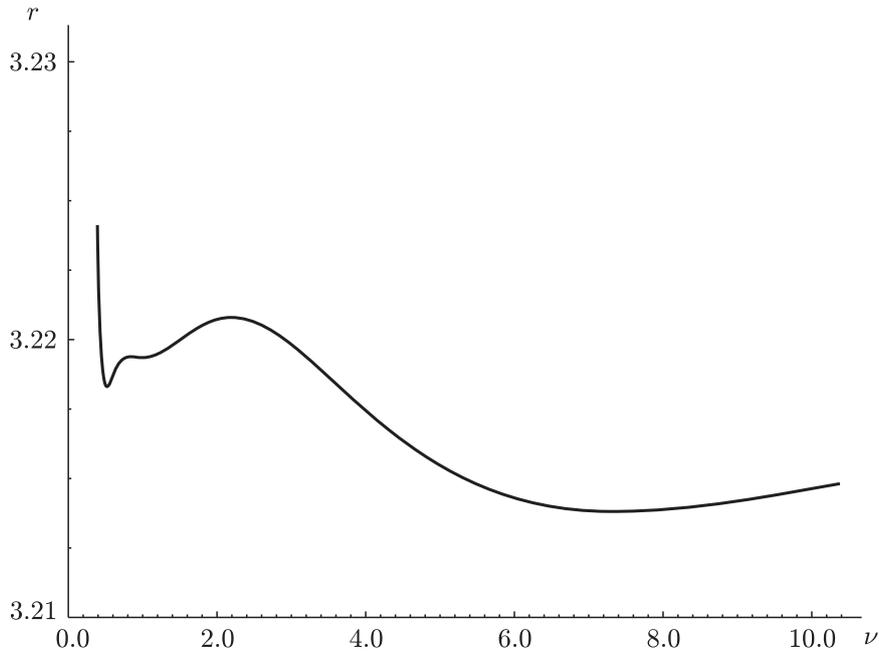


Рис. 1. Поведение функции $r(\nu)$ на отрезке $0,2 \leq \nu \leq 10,2$

свидетельствует приведенный далее пример, и найти «аналитически» ее абсолютный минимум в общем случае (для любого f) не удастся. Поэтому для решения задачи на практике приходится прибегать к прямому перебору, т.е. к сравнению значений $r(\nu_i)$ для всех ν_i на выбранной равномерной сетке.

Приведем анонсированный выше пример, для чего рассмотрим следующее распределение вероятностей f^{200} ($K = 200$):

$$f^{200}(k) = \begin{cases} 1/17 & \text{при } k = 0, 6, 11, 12, 14, 27, 32, 47, 48, 55, 85, 89, 100, 110, 115, 136, 152; \\ 0 & \text{в остальных случаях.} \end{cases}$$

Вычислим значения функции $r(\nu)$ на отрезке $0,2 \leq \nu \leq 10,2$ в узлах равномерной сетки с шагом 0,01. Результат вычислений представлен на рис. 1 и показывает, что функция $r(\nu)$ имеет три локальных минимума, а абсолютный минимум достигается в точке $\nu = 7,24$.

3. ПРИМЕР КОМБИНАТОРНОГО КОДИРОВАНИЯ

В этом разделе рассматривается пример применения метода универсального комбинаторного кодирования, основанного на аппроксимации частотных распределений. В качестве объекта кодирования (сжатия) в примере будут использованы данные компьютерной (рентгеновской) томографии. Цель примера – продемонстрировать эффективность метода аппроксимации, при этом задача получения минимально возможной скорости кодирования не ставится. Отметим, что решению последней задачи посвящена работа [8].

Исходным массивом данных в рассматриваемом примере является фрагмент томограммы легких, который представлен на рис. 2 а. Фрагмент представляет собой полутоновое изображение размером 256×256 , пиксели которого содержат значения рентгеновской плотности тканей, выраженные в единицах шкалы Хаунсфилда. Шкала состоит из целых чисел в диапазоне $[-1024, 3071]$, ширина диапазона – 12 бт. Рентгеновская плотность воды при нормальных условиях принята за нуль, рентгеновская плотность воздуха при нормальных условиях по опреде-

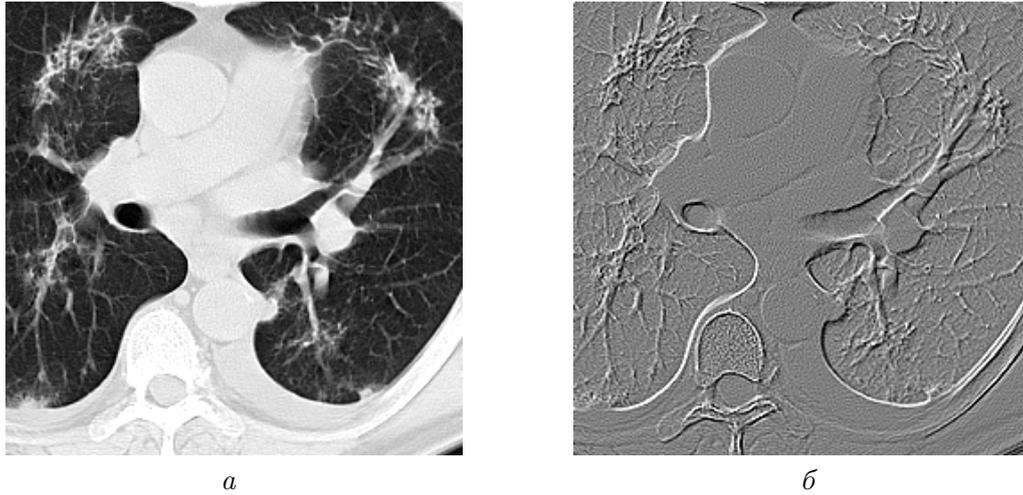


Рис. 2. Изображение фрагмента томограммы (а) и его ошибок предсказания (б)

лению считается равной -1000 . Для некоторого материала с линейным коэффициентом поглощения μ значение рентгеновской плотности по шкале Хаунсфилда равно $1000(\mu - \mu_0)/\mu_0$, где μ_0 – линейный коэффициент поглощения воды при той же (эффективной) энергии, а значения округляются до ближайшего целого. Для визуализации фрагмента томограммы на рис. 2 а использовано окно визуализации $[-1000, 150]$. Напомним, что визуализация изображения в окне $[a_1, a_2]$ предполагает преобразование значений яркости, при котором диапазон $[a_1, a_2]$ линейно отображается на стандартный диапазон $[0, 255]$, значения $a < a_1$ отображаются в значение 0, значения $a > a_2$ – в значение 255, и вывод полученного таким образом изображения на экран монитора или устройство печати. Окно визуализации томограммы выбирается рентгенологом исходя из диагностических задач.

В качестве обратимого преобразования, обеспечивающего декорреляцию отсчетов исходного массива данных (изображения фрагмента томограммы), используем переход к ошибкам предсказания. Пусть T – некоторый элемент исходного изображения, U_T и L_T – соседние к нему сверху и слева элементы. Если T – элемент первой строки и/или первого столбца, т.е. верхний и/или левый соседний элемент отсутствует, то будем полагать $U_T = 0$ и/или $L_T = 0$ по определению. Величина $(U_T + L_T)/2$ (деление является целочисленным) – это простейшее предсказание нулевого порядка для текущего элемента T , а величина $X = T - (U_T + L_T)/2$ – соответствующая ошибка предсказания. Ясно, что ошибки предсказания могут принимать целые значения в диапазоне $[-4095, 4095]$. Вся совокупность ошибок предсказания образует изображение \mathbf{X} размером 256×256 , которое представлено на рис. 2 б в окне визуализации $[-300, 300]$. Последовательность \mathbf{x} элементов данного изображения, упорядоченных естественным образом (построчно слева направо и сверху вниз), и является объектом кодирования в рассматриваемом примере.

Итак, мы имеем последовательность ошибок предсказания $\mathbf{x} = \{x_n\}_{n=0}^{N-1}$ ($N = 2^{16}$), т.е. последовательность целых чисел из диапазона $\mathcal{A} = [-4095, 4095]$ ($A = 2^{13} - 1$). Минимальное и максимальное значения ошибок предсказания равны $a_{\min} = -1030$ и $a_{\max} = 800$, соответственно, т.е. диапазон встречающихся значений заметно уже полного диапазона. На рис. 3 представлены частотные вероятности $f(x)$ значений ошибок предсказания из диапазона $[-450, 450]$. Суммарная вероятность попадания значений в этот диапазон составляет 0,9958 и, таким образом, нетривиальная часть частотного распределения вероятностей представлена на рисунке практически полностью. Квазиэнтропия последовательности ошибок предсказания, вычисленная по формуле (3), равна $H(\mathbf{x}) = 8,2255$ бт/п.

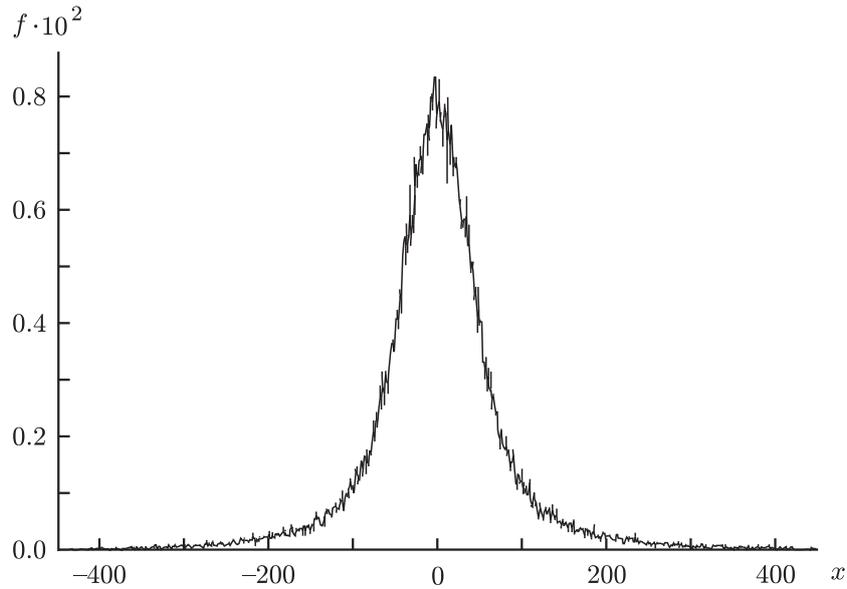


Рис. 3. Частотное распределение значений ошибок предсказания

Оценим избыточность описанной во введении процедуры упрощенного комбинаторного универсального кодирования. Из формулы (5) сразу имеем $R = 1,9995$ бт/п. Используя то обстоятельство, что все значения ошибок предсказания попадают в диапазон, который значительно уже полного диапазона \mathcal{A} , нетрудно предложить такую модификацию процедуры комбинаторного кодирования, которая приводит к меньшей избыточности. Действительно, вместо того чтобы передавать в преамбуле значения $N(x)$ для всех $x \in \mathcal{A}$, достаточно передать нижнюю и верхнюю границы диапазона $\mathcal{A}^* = [a_{\min}, a_{\max}]$ реально встречающихся значений и значения $N(x)$ для всех $x \in \mathcal{A}^*$; $N(x)$ и соответствующие частотные вероятности остальных значений $x \in \{\mathcal{A} \setminus \mathcal{A}^*\}$ заведомо равны нулю. Оценим избыточность такой модифицированной процедуры. Поскольку для передачи одной границы требуется $\lceil \log A \rceil$ бт, а для передачи одного значения $N(x) - \lceil \log N \rceil$ бт, легко имеем

$$R = R_T = \frac{2\lceil \log A \rceil}{N} + \frac{(a_{\max} - a_{\min})\lceil \log N \rceil}{N} = 0,4472 \text{ бт/п.}$$

Альтернативной по отношению к упрощенному комбинаторному кодированию является стандартная процедура универсального кодирования, которая предполагает использование на каждом текущем шаге арифметического кодирования/декодирования кодового распределения, формируемого кодером/декодером исходя из значений элементов последовательности, встретившихся на предыдущих шагах. Отметим, что поскольку декодирование осуществляется без задержки, то в момент декодирования текущего элемента значения всех предыдущих элементов уже известны декодеру. В работе [6] показано, что для источников без памяти (последовательности статистически независимых отсчетов) почти оптимальным является использование на n -ом шаге арифметического кодирования/декодирования следующего частотного распределения:

$$q_n(x) = \frac{1 + 2\theta_n(x)}{A + 2n}, \quad n = 0, 1, \dots, (N - 1),$$

где $\theta_n(x)$ – число вхождений буквы x (т.е. число элементов, принимающих значение x) в начальном участке последовательности \mathbf{x} до $(n - 1)$ -го члена включительно.

Применительно к рассматриваемому примеру выгодно модифицировать описанную выше процедуру так, чтобы учесть равенство нулю частотных вероятностей всех значений x из

множества $\{\mathcal{A} \setminus \mathcal{A}^*\}$. Именно, будем использовать на n -ом шаге арифметического кодирования/декодирования частотное распределение

$$q_n(x) = \begin{cases} \frac{1 + 2\theta_n(x)}{a_{\max} - a_{\min} + 1 + 2n} & \text{при } x \in [a_{\min}, a_{\max}]; \\ 0 & \text{в противном случае.} \end{cases}$$

Значения границ a_{\min} , a_{\max} должны быть, разумеется, переданы декодеру в преамбуле. При этом для полной избыточности кодирования в рассматриваемом примере имеем

$$R = R_A + R_T = \frac{L}{N} + \frac{2\lceil \log A \rceil}{N} = 0,0990 \text{ бт/п},$$

где L – длина кодового слова арифметического кодирования – вычислена непосредственно по формуле, аналогичной формуле (1).

Приступим к рассмотрению предложенного в [7] метода универсального комбинаторного кодирования. Начнем рассмотрение с выбора класса аппроксимирующих распределений \mathcal{P} . Пусть параметр ν принимает значения из множества $\{\nu\} = \{0, 0,5, 0,6, \dots, 3,5\}$ (всего 32 значения), а множество $\{\rho\}$ значений параметра ρ включает те вещественные числа, которые в нормализованном десятичном представлении имеют мантиссу, состоящую из трех цифр, и порядок от 0 до 7. Для аппроксимации частотных распределений будет использоваться класс $\mathcal{P} = \{\rho_{\nu,\rho}\}$, состоящий из дискретных линейных распределений вида (7) ($\nu = 0$) и дискретных экспоненциальных распределений вида (14) (прочие значения ν). Отметим, что все распределения класса \mathcal{P} являются убывающими (невозрастающими). Нетрудно видеть, что для передачи декодеру значений одной пары параметров (ν, ρ) , определяющих при заданном числе членов распределения K конкретное распределение из класса \mathcal{P} , требуется $5 + (10 + 3) = 18$ бт.

Наша задача заключается в том, чтобы построить аппроксимацию представленного на рис. 3 частотного распределения $f = \{f(x), x \in \mathcal{A}^* = [a_{\min}, a_{\max}]\}$, используя распределения выбранного класса. Как видно из рисунка, в области значений $x < 0$ распределение f имеет общую тенденцию не убывать, а в области значений $x \geq 0$ – не возрастать. Следовательно, одно (неубывающее) распределение класса \mathcal{P} заведомо не обеспечит приемлемой точности аппроксимации на всем множестве значений \mathcal{A}^* . Поэтому диапазон \mathcal{A}^* необходимо разбить на несколько отдельных непересекающихся интервалов с тем, чтобы использовать в каждом интервале отдельное аппроксимирующее распределение.

Рассмотрим некоторое множество интервалов $\mathcal{J} = \{\mathcal{I}_i\}$, $i = \pm 1, \dots, \pm I$, вида $\mathcal{I}_i = [a_i^B, a_i^E]$ ($a_i^B < a_i^E$), где $a_{-I}^B = a_{\min}$, $a_{-1}^E = -1$, $a_1^B = 0$, $a_I^E = a_{\max}$ и $a_i^B = a_{i-1}^E + 1$ при $i \neq -I, 1$. Интервалы, очевидно, не пересекаются и образуют разбиение диапазона значений $\mathcal{A}^* = \bigcup_{i=\pm 1}^{\pm I} (\mathcal{I}_i)$. Число значений, входящих в интервал \mathcal{I}_i , равно $K_i = a_i^E - a_i^B + 1 \geq 2$. Число интервалов равно $2I$. Кроме того, интервалы разбиения с отрицательными номерами ($i < 0$) содержат отрицательные значения, а интервалы с положительными номерами ($i > 0$) – положительные значения.

Пусть $N(\mathcal{I})$ – число элементов кодируемой последовательности, значения которых попадают в интервал \mathcal{I} известного разбиения \mathcal{J} . Величина $f(\mathcal{I}) = N(\mathcal{I})/N$ – это частотная вероятность попадания значения x в интервал \mathcal{I} , а величина $f(x|\mathcal{I}) = N(x)/N(\mathcal{I})$ при $x \in \mathcal{I}$ – условная частотная вероятность значения x в данном интервале. Величины $f(\mathcal{I})$, $\mathcal{I} \in \mathcal{J}$, образуют частотное распределение вероятностей интервалов $f_{\mathcal{J}}$, а величины $f(x|\mathcal{I})$, $x \in \mathcal{I}$, – условное частотное распределение вероятностей значений $f_{x|\mathcal{I}}$ в данном интервале. Отметим, что величины $N(\mathcal{I})$ для всех $\mathcal{I} \in \mathcal{J}$ и, следовательно, частотные распределения $f_{\mathcal{J}}$, $f_{x|\mathcal{I}}$ могут быть вычислены кодером.

Построим аппроксимацию условного частотного распределения вероятностей $f_{\mathbf{x}|\mathcal{I}}$ в каждом из интервалов \mathcal{I}_i , $i = \pm 1, \dots, \pm I$, разбиения \mathfrak{J} ; эта аппроксимация определит условное кодовое распределение вероятностей $\hat{q}_{\mathbf{x}|\mathcal{I}}$ значений в соответствующем интервале. Для этого решим задачу минимизации избыточности (6) в классе \mathcal{P} , где

$$f(k) = \begin{cases} f(k + a_i^B | \mathcal{I}_i) & \text{при } i > 0; \\ f(a_i^E - k | \mathcal{I}_i) & \text{при } i < 0, \end{cases}$$

$k = 0, \dots, (K_i - 1)$, $K_i = a_i^E - a_i^B + 1$. Задача решается следующим образом:

1. С помощью алгоритмов, представленных в разделе 2, для каждого $\nu \in \{\nu\}$ вычисляется $\tilde{\rho}(\nu) = \rho(\tilde{\alpha}(\nu))$ такое, что распределение (7) (при $\nu = 0$) или (14) (при $\nu \neq 0$) с $\rho = \tilde{\rho}(\nu)$ является решением задачи оптимальной линейной или экспоненциальной аппроксимации.
2. Полученные значения $\tilde{\rho}(\nu)$ округляются до ближайших значений $\hat{\rho}(\nu)$ из множества $\{\rho\}$, строятся распределения $\mathbf{p}_{\nu, \hat{\rho}(\nu)}$ из класса \mathcal{P} и для каждого распределения вычисляется избыточность $r(\mathbf{p}_{\nu, \hat{\rho}(\nu)} | \mathbf{f})$.
3. Перебором $\nu \in \{\nu\}$ находится то значение $\hat{\nu}$, для которого значение избыточности минимально.

В результате мы имеем значения пары параметров $\hat{\nu} \in \{\nu\}$ и $\hat{\rho} \in \{\rho\}$, а также соответствующее распределение $\hat{\mathbf{p}} = \mathbf{p}_{\hat{\nu}, \hat{\rho}}$, которое минимизирует избыточность (6) в классе \mathcal{P} . Искомым условным кодовым распределением вероятностей $\hat{q}_{\mathbf{x}|\mathcal{I}}$ (аппроксимацией распределения $f_{\mathbf{x}|\mathcal{I}}$) в интервале \mathcal{I}_i ($x \in \mathcal{I}_i$) является следующее распределение:

$$\hat{q}(x | \mathcal{I}_i) = \begin{cases} \hat{p}(x - a_i^B | \mathcal{I}_i) & \text{при } i > 0; \\ \hat{p}(a_i^E - x | \mathcal{I}_i) & \text{при } i < 0. \end{cases}$$

Условные кодовые распределения вероятностей $\hat{q}_{\mathbf{x}|\mathcal{I}}$, построенные путем аппроксимации для всех интервалов $\mathcal{I} \in \mathfrak{J}$, и частотное распределение вероятностей интервалов $f_{\mathfrak{J}}$ определяют общее кодовое распределение \hat{q} :

$$\hat{q}(x) = f(\mathcal{I})\hat{q}(x | \mathcal{I}), \quad x \in \mathcal{I} \in \mathfrak{J}.$$

Для избыточности арифметического кодирования с использованием этого распределения из формулы (3) имеем

$$R_A(\mathbf{x}) = \sum_{\mathcal{I} \in \mathfrak{J}} f(\mathcal{I})R_A(\mathbf{x} | \mathcal{I}), \quad R_A(\mathbf{x} | \mathcal{I}) = \sum_{x \in \mathcal{I}} f(x | \mathcal{I}) \left[-\log \frac{\hat{q}(x | \mathcal{I})}{f(x | \mathcal{I})} \right]. \quad (22)$$

Величины $R_A(\mathbf{x} | \mathcal{I})$ в формуле (22) – избыточность арифметического кодирования отдельных интервалов.

Результаты аппроксимации частотного распределения вероятностей для ошибок предсказания рассматриваемого примера, полученные описанным выше методом, приведены в Таблице. В первом столбце приведены интервалы использованного разбиения. Во втором столбце приведены частотные вероятности $f(\mathcal{I})$ этих интервалов. В третьем и четвертом столбцах приведены значения параметров $\hat{\rho}$ и $\hat{\nu}$, которые определяют условные кодовые распределения вероятностей $\hat{q}_{\mathbf{x}|\mathcal{I}}$ для соответствующих интервалов. В пятом столбце приведены значения избыточности арифметического кодирования $R_A(\mathbf{x} | \mathcal{I})$ для каждого из интервалов.

Общая избыточность арифметического кодирования, вычисленная на основе данных Таблицы по формуле (22), составляет $R_A = 0,0167$ бт/п.

Таблица. Результаты аппроксимации

\mathcal{I}	$f(\mathcal{I})$	$\hat{\rho}$	$\hat{\nu}$	$R_A(\mathbf{x} \mathcal{I})$
$[-1030, -144]$	0,0384	$5,69 \times 10^3$	0,8	0,1706
$[-143, -66]$	0,0948	$6,05 \times 10^0$	0,7	0,0086
$[-65, -1]$	0,3625	$2,99 \times 10^0$	1,8	0,0021
$[0, 73]$	0,3850	$3,73 \times 10^0$	1,8	0,0021
$[74, 140]$	0,0743	$4,02 \times 10^0$	0,9	0,0107
$[141, 800]$	0,0450	$3,93 \times 10^2$	0,8	0,1547

Посчитаем избыточность передачи R_T , связанную с необходимостью передавать декодеру в преамбуле значения параметров, которые вычисляются кодером и определяют используемое кодовое распределение. Такими параметрами являются границы, значения $\hat{\nu}$, $\hat{\rho}$ и частотные вероятности интервалов разбиения. Ошибки предсказания априори могут принимать значения в диапазоне $\mathcal{A} = [-4095, 4095]$, всего $A = 2^{13} - 1$ значений, поэтому для передачи одной границы требуется $\lceil \log A \rceil = 13$ бт. Для передачи значений одной пары параметров $\hat{\nu}$ и $\hat{\rho}$ требуется $b = 18$ бт. Количество ошибок предсказания равно $N = 256 \times 256 = 2^{16}$, поэтому для передачи одной частотной вероятности требуется $\lceil \log N \rceil = 16$ бт. Число интервалов разбиения $-2I = 6$. Для каждого интервала необходимо передать одну границу (начало/конец для интервала в области отрицательных/положительных значений), и одну пару значений параметров $\hat{\nu}$ и $\hat{\rho}$. Для всех интервалов, кроме одного, необходимо передать их частотные вероятности. Таким образом, для избыточности передачи имеем

$$R_T = \frac{2I(\lceil \log A \rceil + b) + (2I - 1)\lceil \log N \rceil}{N} = 0,0041 \text{ бт/п},$$

а полная избыточность кодирования в рассматриваемом примере равна

$$R = R_A + R_T = 0,0208 \text{ бт/п},$$

что существенно меньше избыточности кодирования с использованием рассмотренных в начале раздела стандартных методов.

В заключение отметим, что за рамками обсуждения остался вопрос о выборе разбиения на интервалы по известному частотному распределению, который рассматривался в работе [8].

СПИСОК ЛИТЕРАТУРЫ

1. Гонсалес Р., Вудс Р. *Цифровая обработка изображений*. М.: Техносфера, 2012, 3. (Gonzalez R., Woods R. *Digital Image Processing*. Upper Saddle River: Prentice-Hall, Inc., 2008, 3)
2. Witten I. H., Neal R. M., Cleary J. G. Arithmetic Coding for Data Compression. *Commun. of the ACM*, 1987. vol. 30. no. 6. pp. 520–540.
3. Штарьков Ю. М. *Универсальное кодирование. Теория и алгоритмы*. М.: ФИЗМАТЛИТ, 2013.
4. Shtarkov Yu. M., Babkin V. F. Combinatorial Encoding for Discrete Stationary Sources. *2nd Internat. Symp. on Inform. Theory. Tsahkadsor, Armenia, USSR, 1971*. Budapest: Akademiai Kiado, 1973, pp. 249–256.
5. Krichevsky R. E., Trofimov V. K. The Performance of Universal Encoding. *IEEE Trans. Inform. Theory*, New York, 1981, vol. IT-27, no. 2. pp. 199–207.

6. Штарьков Ю. М. Универсальное последовательное кодирование отдельных сообщений. *Проблемы передачи информации*, 1987, т. 23, № 3. стр. 3–17.
7. Сушко Д. В., Штарьков Ю. М. О сжатии томографических данных. *Информационные процессы*, 2008, т. 8, № 4. стр. 240–255.
8. Стефанович А. И. Сушко Д. В. Обратимое сжатие данных посредством универсального арифметического кодирования. *Информатика и ее применения*, 2017, т. 11, вып. 1. стр. 20–45.

Статью представил к публикации член редколлегии Б. М. Миллер

Optimal approximation of frequency probabilities

Sushko D. V

We present the solution of the problem of optimal approximation of a given discrete (finite) probability distribution in the classes of discrete linear and exponential probability distributions. The code redundancy of the approximate distribution with respect to a given distribution measures the quality of the approximation, and the optimal approximation in the class is the distribution that minimizes code redundancy. We consider an example demonstrating the high efficiency of a universal combinatorial coding method based on the optimal approximation of frequency probabilities in tasks of data lossless compression via arithmetic coding.

KEYWORDS: approximation of distributions, arithmetic coding, universal coding, lossless data compression, computerized tomogram.