УДК 621

ТЕТА-ЭНТРОПИЯ

А.А. Борисенко, *д-р техн. наук, профессор,*
Сумский государственный университет, г. Сумы

Предложена новая форма энтропии, позволяющая оценивать различные неравномерные коды по своей эффективности, применительно к различным распределениям вероятностей сообщений. Из нее получена, как частный случай, обычная энтропия, которая по величине всегда меньше или равна ее общей форме.
Ключевые слова: *энтропия, коды, вероятности.*

Запропонована нова форма ентропії, яка дозволяє оцінювати різні нерівномірні коди за своєю ефективністю, відповідно до різних розподілів імовірностей повідомлень. З неї отримана як особливий випадок звичайна ентропія, яка за своєю величиною завжди буде меншою чи дорівнювати її загальній формі.
Ключові слова: *ентропія, коди, ймовірності.*

Современная теория информации, созданная в середине прошлого века Шенноном, нашла широкое распространение при решении многих информационных задач, которых в настоящее время становится все больше и больше. Однако за более чем полувековой путь своего развития основные идеи этой теории остались практически без изменений. С одной стороны, такое положение дел свидетельствует об устойчивости, эффективности и правильности данной теории, но с другой - об определенном застое в ее основах. Данная работа направлена на пусть и небольшой, но сдвиг в этом вопросе, путем некоторого обобщения основной функции теории информации - *энтропии* источника информации.

В простейшем случае эта функция имеет вид [1-5]

$$H = -\sum_{j=1}^{n} P_j \log_2 P_j \, , \tag{1}$$

где P_j - вероятность j-го *сообщения* (слова, буквы, последовательности символов), отражающего *состояние* источника информации.

При этом

$$0 \le P_j \le 1 \tag{2}$$

и

$$\sum_{j=1}^{n} P_j = 1 \, . \tag{3}$$

Доказано, например в [1-5], что максимальное значение энтропии H

$$H_{\max} = \log_2 n \tag{4}$$

в случае, когда $P_j = \frac{1}{n}$ для всех $j = 1, 2, \ldots, n$. Если же для одного из значений $j = 1, 2, \ldots, n$ $P_j = 1$ *и* $P_j = 0$ для всех остальных j, а также в случае, если для всех j без исключения $P_j = 0$, то энтропия H достигает своего минимального значения: $H = H_{\min} = 0$.

На практике энтропия используется для оценки предела возможного сжатия сообщений источником информации, генерируемых с вероятностями P_j, что позволяет достигнуть максимальной скорости передачи информации. Достигается такое сжатие с помощью специальных методов *оптимального* кодирования, в которых каждому генерируемому источником информации сообщению ставится в соответствие некоторая кодовая комбинация определенной длины [1-5]. Подобранные таким образом кодовые комбинации в совокупности образуют оптимальный код, средняя длина которых как угодно близко может приближаться к энтропии источника информации, но ни при каких условиях не может превысить ее без искажения информации. Об этом говорит теорема Шеннона о максимальной скорости передачи информации в отсутствие шумов [1]. В данном случае под средней длиной кодовых комбинаций понимается математическое ожидание, полученное на множестве всех длин кодовых комбинаций оптимального кода. В этом коде длины отдельных двоичных комбинаций не могут быть меньше величин частных энтропий $\log_2 P_j$, а в идеальном случае, когда $P_j = \frac{1}{2}, \frac{1}{4}, \ldots$, равны этой величине. Учитывая то, что вероятности P_j, как правило, имеют неравную величину, то и длины кодовых комбинаций неравны между собой. Поэтому большинство оптимальных кодов относятся к кодам с неравной длиной кодовых комбинаций – *неравномерным* кодам. Только в случае, когда вероятности генерируемых сообщений равны между собой, неравномерные коды могут преобразоваться в неизбыточные *равномерные* коды, то есть коды, в которых используются все возможные кодовые комбинации, их длины будут равны между собой, а средняя длина будет равна длине отдельно взятой комбинации.

Однако в силу того, что вероятности сообщений, генерируемых тем или иным источником информации, находятся статистическим путем с помощью ограниченного числа испытаний, то точное распределение вероятностей сообщений на выходе источника информации на самом деле обычно остается неизвестным. Поэтому используемые на практике оптимальные коды не решают в полной мере задачу оптимального кодирования, в силу чего большинство из них можно назвать не оптимальными, а *эффективными* кодами. В данном случае в понятие эффективности кода вкладывается смысл уменьшения средней длины его кодовых комбинаций по сравнению с неизбыточным равномерным кодом, получаемым в соответствии с выражением (4), вплоть до величины энтропии H в идеальном случае, и соответственно повышение этим кодом *скорости* передачи сообщений или уменьшение требуемой емкости памяти. К эффективным кодам можно отнести любые неравномерные коды, средняя длина которых меньше средней длины не избыточного равномерного кода. Однако здесь возникает задача оценки того, насколько эти неравномерные коды по своей эффективности отличаются от оптимальных кодов, и, как показывает практика, решить эту задачу без введения дополнительного критерия наподобие энтропии H

достаточно трудно. Задачей данной работы как раз и является получение такого критерия.

Рассмотрим случай, когда при неизменных исходных вероятностях P_j оптимальный код заменяется другим с тем же числом кодовых комбинаций, но с измененными их длинами. Очевидно, что тогда будет получен код, не являющийся оптимальным, так как двух оптимальных кодов с отличающимися длинами комбинаций при кодировании ими одних и тех же сообщений не может быть, и соответственно средняя длина кодовых комбинаций в нем увеличится. Например, для вероятностей P_j, равных 0,5, 0,25, 0,25, в оптимальном коде длины комбинаций должны быть соответственно равны 1, 2, 2. Допустим, что это будут следующие комбинации – 0, 10, 11. Соответственно средняя длина кодовых комбинаций оптимального кода будет равна 1,5. Любое изменение длины кодовых комбинаций от оптимальных приведет к увеличению средней длины комбинаций кода. Так, при длинах комбинаций равномерного кода 00, 10, 11 средняя длина его комбинаций будет равна 2. При этом в коде появится запрещенная комбинация - 01, и поэтому данный код, кроме того, что он потерял оптимальность, стал еще и избыточным. Более эффективным по сравнению с этим кодом был бы код с комбинациями 11, 10, 0, так как средняя длина его кодовых комбинаций будет равняться 1,75. Однако он все же будет хуже оптимального кода, в котором средняя длина кодовых комбинаций будет равна 1,5. В общем же случае делаем вывод, что появление вместо оптимального кода любого другого с тем же количеством кодовых комбинаций, но с другой их длиной, хотя бы в их части, увеличивает среднюю длину кодовых комбинаций и соответственно уменьшает эффективность передачи информации.

Сложнее для анализа будет случай, когда изменяется не сам оптимальный код, а исходные вероятности P_j, по отношению к которым собственно и был разработан оптимальный код. Однако в процессе эксплуатации кода вероятности сообщений P_j по каким-то причинам могут измениться, и тогда взамен их в том же количестве появляются новые вероятности сообщений φ_j. В этом случае возникает необходимость в разработке нового оптимального кода с иными длинами кодовых комбинаций, чем у исходного кода. Очевидно, что энтропия такого источника с измененными вероятностями сообщений

$$H = -\sum_{j=1}^{n} \varphi_j \log_2 \varphi_j . \qquad (5)$$

Она может быть или больше, или меньше исходной энтропии источника с вероятностями P_j, и в соответствии с этим средняя длина кодовых комбинаций нового оптимального кода также увеличится или уменьшится.

Однако для нового распределения вероятностей сообщений может быть задано условие, что для кодирования сообщений источника информации необходимо использовать старый код, ранее бывший оптимальным. Тогда возникает задача оценки его эффективности в новых условиях. Для этого в данной работе предлагается специальная функция, названная *тета-энтропией* (θ-энтропия), особенностью которой является то, что она оценивает эффективность ранее оптимального кода применительно к новому распределению вероятностей сообщений φ_j в нем:

$$\theta = -\sum_{j=1}^{n} \varphi_j \log_2 P_j\,, \tag{6}$$

где P_j - первоначальные вероятности сообщений источника информации,

φ_j - новые вероятности сообщений, для которых

$$\sum_{j=1}^{n} \varphi_j = 1. \tag{7}$$

В данном случае неизменными параметрами функции будут исходные вероятности P_j, а переменными новые вероятности φ_j.

Аналогичная функция, представленная в виде

$$\theta = -\sum_{j=1}^{n} P_j \log_2 \varphi_j\,, \tag{8}$$

описывает уже рассмотренный выше случай, когда изменяются длины кодовых комбинаций при неизменных значениях исходных вероятностей P_j. Очевидно, что в соответствии с вышесказанным эта функция достигает минимума при величине $\varphi_j = P_j$.

Разность между значениями θ-энтропии и исходной энтропией H в выражении (6)

$$\Delta = \theta - H = \sum_{j=1}^{n} (\varphi_j - P_j)\log_2 P_j \tag{9}$$

определит критерий *эффективности* кода с новым распределением вероятностей φ_j. Очевидно, что этот критерий может принимать, в зависимости от соотношений φ_j *и* P_j, как положительные, так и отрицательные значения, а при значении всех $\varphi_j = P_j$ он становится равным нулю. В последнем случае θ - энтропия переходит в обычную энтропию H и характеризует среднюю длину кодовых комбинаций рассматриваемого кода применительно к распределению вероятностей P_j. Из сказанного следует, что при измененных вероятностях можно получить большую эффективность исходного кода, чем в случае, когда он был оптимальным. Хотя для этих вероятностей можно построить новый оптимальный код, который будет более эффективным кодом, чем коды, рассмотренные выше.

Рассмотрим условия распределения вероятностей комбинаций кода, при которых средняя длина его кодовых комбинаций будет больше средней длины комбинаций этого же кода с оптимальным распределением вероятностей.

Тогда имеем неравенство

$$\theta \geq H \tag{9}$$

или

$$\sum_{j=1}^{n}\varphi_j \log_2 P_j \geq \sum_{j=1}^{n} P_j \log_2 P_j . \tag{10}$$

Соответственно критерий эффективности кода

$$\Delta = \theta - H = \sum_{j=1}^{n}\varphi_j \log_2 P_j - \sum_{j=1}^{n} P_j \log_2 P_j = \\ = \sum_{j=1}^{n}(\varphi_j - P_j)\log_2 P_j \geq 0. \tag{11}$$

Если записать $\varphi_j - P_j = a_j$, то тогда

$$\Delta = \sum_{j=1}^{n} a_j \log_2 P_j \geq 0 , \tag{12}$$

где разность вероятностей a_j может принимать как отрицательное, так и положительное значение. При этом обратим внимание, что, вследствие неравенства

$$0 \leq \varphi_j = (a_j + P_j) \leq 1 , \tag{13}$$

$$0 \leq a_j \leq 1 . \tag{14}$$

Действительно, если $P_j = 1$, то из (13) следует, что $a_j = 0$, а если $P_j = 0$, то $0 \leq a_j \leq 1$.

Теорема 1

$$\sum_{j=1}^{n} a_j = 0 . \tag{15}$$

Доказательство. Так как

$$\sum_{j=1}^{n} P_j = 1 \text{ и } \sum_{j=1}^{n}(a_j + P_j) = \sum_{j=1}^{n} a_j + \sum_{j=1}^{n} P_j = 1 ,$$

то

$$\sum_{j=1}^{n} a_j = 0.$$

Теорема доказана.

Следствие 1. Значения $a_j = \varphi_j - P_j$ должны быть как положительными, так и отрицательными, за исключением случая, когда все $a_j = 0$. Следует из того, что в противном случае левая часть равенства теоремы 1 не сможет равняться нулю.

Следствие 2. Сумма пар вероятностей φ_j и P_j, образующих разности с отрицательным знаком, равняется сумме пар разностей этих же вероятностей с положительным знаком. Следует из тех же соображений, что и для следствия 1.

Убедиться в том, что данная теорема и ее следствия верны, можно из примеров, приведенных в таблице 1.

Особый случай представляет условие, при котором вероятности P_j равны между собой. Тогда выражение (6) будет иметь следующий вид:

$$\theta = -\sum_{j=n}^{n} \varphi_j \log_2 \frac{1}{n} = -\log_2 \frac{1}{n} = \log_2 n. \qquad (18)$$

Это значит, что получить более неэффективный код, чем данный, невозможно в принципе. Любое другое неравномерное распределение вероятностей не может ухудшить данный код, так как этот код в смысле средней длины кодовых комбинаций самый плохой код. Однако этот же код будет и самым простым, так как он обладает равной длиной кодовых комбинаций. Поэтому он, несмотря на свою предельную неэффективность в смысле скорости передачи информации, является на практике наиболее употребляемым кодом. Ведь его использование не требует проводить статистических испытаний и определять вероятности появления тех или иных сообщений.

Проиллюстрируем полученные выше результаты с помощью таблиц 1 и 2. В таблице 1 в каждой строке заданы вероятности P_j и φ_j четырех сообщений и для них вычислены значения энтропии H и θ - энтропии.

Таблица 1 - Распределение вероятностей сообщений и значений энтропий

P_1	P_2	P_3	P_4	φ_1	φ_2	φ_3	φ_4	H	θ
0,5	0,25	0,125	0,125	0,25	0,25	0,25	0,25	1,75	2,25
0,5	0,25	0,25	0	0,25	0,5	0,25	0	1,5	1,75
0,25	0,25	0,25	0,25	0,25	0,5	0,125	0,125	2,0	2,0
0,5	0,25	0,25	0	0,25	0,25	0,5	0	1,5	1,75
0,25	0,5	0,25	0	0,5	0,25	0,5	0	1,5	1,75
0,25	0,5	0,125	0,125	0,5	0,25	0,125	0,125	1,75	2,25
0,125	0,5	0,125	0,25	0,5	0,25	0,125	0,125	1,75	2,375

Как видим из данной таблицы, θ - энтропия для всех ее строк больше или равна энтропии H, что приводит к положительному значению критерия эффективности и соответственно к снижению эффективности кода в новых условиях. Очевидно, что можно получить и отрицательное значение критерия эффективности кода, когда θ - энтропия будет меньше энтропии источника H. Для примера рассмотрим случай, когда в первой строке распределения вероятностей P_j в таблице 1 остаются прежними, а вероятности φ_j будут соответственно равны - 0,9, 0,034, 0,033, 0,033. Тогда θ = 0,9*1 + 0,034*2 + 0,033*3 + 0,033*3 = 0,9 + 0,068 + 0,099 + +0,099 = 1,166 бита. Критерий эффективности кода для приведенного распределения вероятностей φ_j примет отрицательное значение $\Delta = \theta - H$ = 1,66 − 1,75 = - 0,09 бита. Это значит, что применительно к распределению вероятностей P_j исходный код при заданном в примере распределении вероятностей φ_j станет более эффективным. Средняя длина его кодовых комбинаций уменьшилась на величину 0,09 бита по сравнению с исходной средней длиной кода 1,75 бита.

В таблице 2 приведены соответствующие таблице 1 длины l_j кодовых комбинаций оптимального кода и их среднее значение для исходного распределения вероятностей и измененного.

Таблица 2 – Длины отдельных кодовых комбинаций и их средние значения для оптимального и неэффективного кода

l_1	l_2	l_3	l_4	$l_{cp}(H)$	$l_{cp}(\theta)$
1	2	3	3	1,75	2,25
1	2	2	0	1,5	1,75
2	2	2	2	2,0	2,0
1	2	2	0	1,5	1,75
2	1	2	0	1,5	1,75
2	1	3	3	1,75	2,25
3	1	3	2	1,75	2,375

Таким образом, предложенное обобщение энтропии в виде тета-энтропии, позволяет оценивать эффективность кодов по скорости передачи информации для различных распределений вероятностей сообщений и сравнивать эти коды между собой.

SUMMARY

TETA – ENTROPY

A.A. Borysenko,
Sumy State University, Sumy

A new form of entropy is supposed. It allows to estimate the efficiency of different run-length codes in reference to different distributions of probabilities of messages. Besides, the ordinary form of entropy, as special case, which value is always equal or less then its general form is got.

Key words: *entropy, codes, probability.*

СПИСОК ЛИТЕРАТУРЫ

1. Шеннон К. Работы по теории информации и кибернетике: пер. с англ. – М.: Изд-во иностр. лит., 1963. – 830с.
2. Фано Р. Передача информации. Статистическая теория информации: пер. с англ. – М.: Мир, 1965. - 438с.
3. Галлагер Р. Теория информации и надежная связь: пер. с англ. – М.: Советское радио, 1974. - 720с.
4. Цымбал В.П. Теория информации и кодирования: учебник. – 4-е изд. – К.: Вища шк., 1992. – 263с.
5. Темников Ф.Е. и др. Теоретические основы информационной техники. – М.: Энергия, 1971. - 424с.

Поступила в редакцию 5 февраля 2010 г.