ГЛАВА 6. ТЕОРИЯ ИНФОРМАЦИИ

Теорией информации называется раздел теории вероятностей, в котором изучаются количественные закономерности, связанные:

1)с передачей; 2)приемом; 3)обработкой; 4)хранением информации.

Одной из задач теории информации является отыскание экономных методов кодирования. Другая задача состоит в оптимизации пропускной способности канала связи. Ряд задач теории информации относится к определению объема запоминающих устройств, к способам ввода информации в них и вывода ее для непосредственного использования.

Для решения подобных задач требуется определение количественной меры объема информации. Основным понятием теории информации является *информационная энтропия*, являющаяся мерой степени неопределенности состояния физической системы.

§1. Информационная энтропия

Любое сообщение в теории информации описывает состояние какой-то физической системы, которая **случайным** образом может оказаться в том или ином состоянии.

Поэтому в качестве объекта, о котором передается информация, в теории информации рассматривается физическая система, которой присуща **степень неопределенности**. Степень неопределенности физической системы определяется ***числом*** ее возможных состояний и ***вероятностями*** состояний. В качестве меры априорной неопределенности системы в теории информации применяется характеристика, называемая энтропией.

Рассмотрим физическую систему X с конечным множеством состояний. Каждому состоянию источника X ставиться в соответствие условное обозначение в виде знака. Пусть $x_1, x_2, \ldots x_n$ – конечное множество состояний, в которых она может находиться; $(X \approx x_i)$ – событие, состоящее в том, что система X находится в состоянии x_i; $p_i \equiv P(X \approx x_i)$ – вероятность события $(X \approx x_i)$; $\sum_{i=1}^{n} p_i = 1$. Совокупность знаков $x_1, x_2, \ldots x_n$ соответствующих всем n возможным состояниям источника называют его ***алфавитом***, а количество состояний n ***объемом алфавита***. Система X характеризуется полной совокупностью состояний x_i с вероятностями их появления p_i, составляющими в сумме 1, т.е. таблицей вида

x_i	x_1	x_2	....	x_n
p_i	p_1	p_2	....	p_n

В отличие от ряда распределения для описания степени неопределенности системы неважно, какие значения x_i записаны в верхней строке таблицы; важны только ***количество*** x_i (объем алфавита n) и их ***вероятности*** p_i.

Определение. Энтропией $H(X)$ системы X называется величина, равная

$$H(X) = -\sum_{i=1}^{n} p_i \log(p_i), \qquad (1)$$

где чаще всего в качестве основания логарифма используют число 2.

Энтропия $H(X)$ обладает рядом свойств, которые оправдывают выбор ее в качестве характеристики степени неопределенности.

Свойство 1. Энтропия системы с равновозможными состояниями $p_i = \frac{1}{n}$ равна логарифму числа состояний n.

Доказательство. По условию таблица состояний системы имеет вид

x_i	x_1	x_2	....	x_n
p_i	$\frac{1}{n}$	$\frac{1}{n}$	....	$\frac{1}{n}$

Из определения (1) с учетом таблицы имеем:

$$H(X) = -\sum_{i=1}^{n} p_i \log(p_i) = -n\frac{1}{n}\log\left(\frac{1}{n}\right) = -\log(1) + \log(n)$$

или

$$H(X) = \log(n) \qquad (2)$$

Свойство 2. Если состояние системы в точности известно заранее, то ее энтропия равна нулю.

Доказательство. В этом случае все вероятности $p_1 = p_2 = ... = p_{k-1} = p_{k+1} = ... = p_n = 0$, кроме одной – например $p_k = 1$. Так как $p_k \log(p_k) = \log(1) = 0$, а для остальных членов энтропии $\lim\limits_{p \to 0}[p\log(p)] = 0$,то

$$H(x) = 0. \qquad (3)$$

Свойство 3. Максимальное значение энтропии физической системы с конечным числом состояний равно логарифму числа состояний и достигается, когда все состояния равновероятны.

Доказательство. Пользуясь методом неопределенных множителей Лагранжа, найдем экстремум функции Лагранжа при условии $\sum_{i=1}^{n} p_i = 1$

$$L = -\sum_{i=1}^{n} p_i \log(p_i) + \lambda \sum_{i=1}^{n} p_i. \qquad (4)$$

Дифференцируя (4) по p_i и приравнивая производные нулю, получим

$$\log(p_i) = -\lambda - \log(e), \ (i = 1,2,3,....n).$$

Откуда следует, что $p_i = const$. При этом с учетом условия $\sum_{i=1}^{n} p_i = 1$

$$p_1 = p_2 = ... = p_n = \frac{1}{n},$$

а максимальная энтропия системы равна:

$$H_{\max} = \log(n). \qquad (5)$$

Свойство 4. При объединении **независимых** систем их энтропии складываются.

Доказательство. Под объединением двух систем X и Y с возможными состояниями

$x_1, x_2, ...x_n$, $y_1, y_2, ...y_m$ понимается сложная система (X,Y), стояния которой (x_i, y_j) представляют собой все возможные комбинации состояний x_i, y_j систем X и Y. Таблица состояний сложной системы представляется в виде

y_j	x_1	x_2	....	x_n
y_1	P_{11}	P_{21}	....	P_{n1}
y_2	P_{12}	P_{22}	....	P_{n2}
....	....	....	....	....
y_m	P_{1m}	P_{2m}	....	P_{nm}

Здесь $P_{ij} \equiv P\left[(X \approx x_i)\cdot(Y \approx y_j)\right]$ – вероятность того, что сложная система (X,Y) будет находиться в состоянии (x_i, y_j).

По определению энтропия сложной системы равна:

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij}\log(P_{ij}). \qquad (6)$$

Представим энтропию (6) в форме математического ожидания

$$H(X,Y) = -M\left[\log(P(X,Y))\right]. \qquad (7)$$

По теореме умножения вероятностей для независимых событий

$$P(X,Y) = P(X)P(Y). \qquad (8)$$

Подставляя (8) в (7), с учетом свойств математического ожидания получим

$$H(X,Y) = -M\left[\log(P(X)) + \log(P(Y))\right] = H(X) + H(Y), \qquad (9)$$

что и требовалось показать.

Свойство сложения энтропий может быть обобщено по методу индукции на произвольное число независимых систем:

$$H(X_1, X_2 ....X_s) = \sum_{k=1}^{s} H(X_k).$$

Свойство 5. Если две произвольные системы X и Y объединяются в одну, то энтропия сложной системы (X,Y) равна:

$$H(X,Y) = H(X) + H(Y/X), \qquad (10)$$

где $H(Y/X)$ – условная энтропия второй части Y относительно первой X.

Доказательство. Для доказательства представим энтропию $H(X,Y)$ в форме математического ожидания

$$H(X,Y) = -M\left[\log(P(X,Y))\right]. \qquad (11)$$

По теореме умножения вероятностей

$$P(X,Y) = P(X)P(Y/X),$$

следовательно,

$$H(X,Y) = -M\left[\log(P(X))\right] - M\left[\log(P(Y/X))\right] = H(X) + H(Y/X), \qquad (12)$$

где

$$H(Y/X) = -M[\log(P(Y/X))] = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij}\log(P(y_j/x_i)) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij}\log\left(\frac{P_{ij}}{p_i}\right), \quad (13)$$

$P(y_j/x_i)$ – условная вероятность того, что система Y находится в состоянии y_j при условии, что система X находится в состоянии x_i; по теореме умножения $P(y_j/x_i) = \frac{P_{ij}}{p_i}$, $p_i = P(X \approx x_i)$.

Свойство можно распространить по индукции на любое число объединяемых систем:

$$H(X_1, X_2, ...X_s) = H(X_1) + H(X_2/X_1) + H(X_3/X_1X_2) + ... + H(X_s/X_1X_2.....X_{s-1}),$$

где энтропия каждой последующей системы вычисляется при условии, что состояние всех предыдущих известно.

Свойство 6. Полная условная энтропия системы $H(Y/X)$ не превосходит ее безусловной энтропии $H(Y)$.

Доказательство. Для доказательства рассмотрим величину $H(Y) - H(Y/X)$.

Согласно (1) и (13)

$$H(Y) - H(Y/X) = -\sum_{j=1}^{m} p_j \log(p_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij}\log\left(\frac{P_{ij}}{p_i}\right) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij}\log(p_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij}\log\left(\frac{P_{ij}}{p_i}\right) =$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij}\log\left(\frac{P_{ij}}{p_i p_j}\right) \quad (14)$$

Так как при любом $x > 0$

$$\ln(x) \le x - 1,$$

то

$$\log\left(\frac{P_{ij}}{p_i p_j}\right) = \frac{\ln\left(\frac{P_{ij}}{p_i p_j}\right)}{\ln 2} \ge \frac{1}{\ln 2}\left(1 - \frac{p_i p_j}{P_{ij}}\right). \quad (15)$$

С учетом (15) и $\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij} = \sum_{i=1}^{n} p_i = \sum_{j=1}^{m} p_j = 1$ из (14) имеем:

$$H(Y) - H(Y/X) = \sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij}\log\left(\frac{P_{ij}}{p_i p_j}\right) \ge \frac{1}{\ln 2}\left[\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij} - \sum_{i=1}^{n} p_i \sum_{j=1}^{m} p_j\right] = 0,$$

так что

$$H(Y/X) \le H(Y), \quad (16)$$

или

$$H(X,Y) \le H(X) + H(Y), \quad (17)$$

что и требовалось показать.

Следствие. ***Из доказанных свойств следует, что энтропия может служить мерой неопределенности состояния физической системы.***

§2. Энтропия и информация

Из свойств энтропии следует, что она является мерой неопределенности состояния физической системы. Естественно поэтому количество информации измерять уменьшением энтропии системы, для уточнения состояния которой предназначены сведения. В теории информации физическую систему X, которая может в каждый момент времени случайным образом принять одно из конечного множества возможных состояний x_i, называют дискретным источником сообщений. Каждое возможное состояние x_i называют ***элементарным дискретным сообщением***. Набор элементарных сообщений называют сообщением. При этом в течение некоторого времени T источник может выдать дискретное сообщение в виде последовательности элементарных дискретных сообщений, представляющей сбой набор символов x_i (например, x_5, x_1, x_3), каждый из которых имеет длительность t_i секунд, в общем случае необязательно одинаковую для различных i. При выдаче источником сообщений в виде последовательности элементарных дискретных сообщений, полное вероятностное описание дается вероятностью совместного появления набора различных символов x_i в момент $t_1, t_2, ...t_k$

$$P\left(x_i^{t_1}, x_j^{t_2}, ...x_s^{t_k}\right),$$

где k - длина последовательности. Располагая такими сведениями об источнике можно вычислить вероятность любого отрезка сообщения длиной меньше k. Если функция $P\left(x_i^{t_1}, x_j^{t_2}, ...x_s^{t_k}\right)$ не меняется во времени, то источник называется стационарным. Если при определении вероятностных характеристик стационарного источника усреднение по ансамблю можно заменить усреднением по времени, то такой источник называется эргодическим. Вероятностные свойства эргодического источника можно оценить, рассматривая лишь одну его достаточно длинную реализацию. В каждом элементарном сообщении содержится для его получателя определенная информация совокупность сведений о состоянии дискретного источника сообщения. Определяя количественную меру этой информации, мы не будем учитывать ее смыслового содержания, так же ее значения для конкретного получателя. Очевидно, что при отсутствии сведений о состоянии источника имеется неопределенность относительно того, какое сообщение x_i из числа возможных им выбрано, а при наличии этих сведений данная неопределенность полностью исчезает. Естественно количество информации содержащейся в дискретном сообщении измерять величиной исчезнувшей неопределенности. Введем меру этой неопределенности, которую можно рассматривать и как меру количественной информации. Мера должна удовлетворять ряду естественных условий, одним из них является необходимость ее монотонного возрастания с увеличением возможности выбора, т.е. объема алфавита источника n. Кроме того, желательно, чтобы вводимая мера обладала свойством адетивности.

Определение. Если наблюдение ведется непосредственно за самой системой X, то ***полной информацией,*** приобретаемой при полном выяснении состояния системы X, называется величина

$$I_X = H(X). \qquad (18)$$

Представим формулу (18) с учетом (1) в виде

$$I_X = -\sum_{i=1}^{n} p_i I_{x_i} \ ,$$

где $I_{x_i} = \log(p_i)$ – называют частной информацией, получаемой от отдельного сообщения о том, что система X находится в состоянии x_i. Когда основание логарифма равно 2, то единица количества информации называется битом (bit), и представляет собой информацию, содержащуюся в одном дискретном состоянии x_i системы X равновероятных $p_i = \frac{1}{n}$ состояний с объемом алфавита $n = 2$. При выборе основания логарифма равным 10 единицу количества информации называют дитом. Иногда используют натуральную единицу количества информации, называемую натом.

На практике часто система X недоступна для наблюдения, и выясняется состояние не самой системы X, а системы Y, связанной с системой X.

Определение. Полной информацией о системе X, содержащейся в системе Y, называется величина

$$I_{Y \to X} = H(X) - H(X/Y). \qquad (19)$$

Информация $I_{Y \to X}$ обладает следующими свойствами.

Свойство 1. Полная информация о системе X, содержащейся в системе Y, равна полной информации о системе Y, содержащейся в системе X

Доказательство. Согласно пятому свойству

$$H(X,Y) = H(X) + H(Y/X),$$
$$H(X,Y) = H(Y) + H(X/Y),$$

откуда

$$H(X) - H(X/Y) = H(Y) - H(Y/X),$$

или

$$I_{Y \to X} = I_{X \to Y}. \qquad (20)$$

Поэтому информацию $I_{Y \to X}$ обозначают $I_{Y \leftrightarrow X}$ и называют полной взаимной информацией.

Свойство 2. Полная взаимная информация, содержащаяся в двух системах, равна сумме энтропий составляющих систем минус энтропия объединенной энтропии.

Доказательство. По пятому свойству энтропии

$$H(X/Y) = H(X,Y) - H(Y).$$

Подставляя это выражение в формулу (19), получим:

$$I_{X \leftrightarrow Y} = H(X) + H(Y) - H(X,.Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij} \log\left(\frac{P_{ij}}{p_i p_j}\right),$$

что и требовалось показать.

Свойство 3. Полная информация, содержащаяся в **независимых** системах, равна нулю.

Доказательство. Если системы независимы, то по теореме умножения

$$P_{ij} = p_i p_j .$$

Тогда согласно формуле (13) и $\sum_{i=1}^{n} P_{ij} = p_i$

$$H(Y/X) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij} \log\left(\frac{P_{ij}}{p_i}\right) = -\sum_{j=1}^{m}\left(\sum_{i=1}^{n} P_{ij}\right)\log(p_j) = -\sum_{j=1}^{m} p_j \log(p_j) = H(Y),$$

откуда с учетом (20)

$$I_{X\leftrightarrow Y} = 0,$$

что и требовалось показать.

Свойство 4. Полная информация, содержащаяся в полностью **зависимых** системах, равна полной информации системы X, либо системы Y.

Доказательство. Если состояние системы X полностью определяет состояние системы Y и наоборот, то

$$P_{ij} = p_i = p_j ,$$

так что

$$H(Y/X) = H(X/Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij} \log\left(\frac{P_{ij}}{p_i}\right) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij} \log(1) = 0 .$$

Подставляя это выражение в формулу (19) и учитывая (18), получим:

$$I_{X\leftrightarrow Y} = H(X) = H(Y) = I_X = I_Y ,$$

что и требовалось показать.

Свойство 5. Полная взаимная информация, содержащаяся в системах, из которых одна является подчиненной, равна энтропии подчиненной системы.

Доказательство. Система, состояние которой полностью определяется состоянием другой, называется ***подчиненной*** системой. Пусть из двух систем X и Y подчиненной является система X. Тогда $P_{ij} = p_j$, откуда

$$H(X/Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij} \log\left(\frac{P_{ij}}{p_j}\right) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij} \log(1) = 0,$$

и согласно (18)

$$I_{X\leftrightarrow Y} = H(X),$$

что и требовалось показать.

В ряде случаев представляет интерес оценить частную информацию о системе X, содержащейся в событии $(Y \approx y_j)$. Обозначим эту частную информацию $I_{y_j \to X}$.

Представим полную информацию $I_{Y\to X}$ в эквивалентном виде

$$I_{X\leftrightarrow Y} = \sum_{i=1}^{n}\sum_{j=1}^{m} P_{ij} \log\left(\frac{P_{ij}}{p_i p_j}\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_j P(x_i / y_j)\log\left(\frac{P(x_i / y_j)}{p_i}\right) = \sum_{j=1}^{m} p_j I_{y_j\to X} ,$$

где $I_{y_j\to X} = \sum_{i=1}^{n} P(x_i / y_j) I_{y_j\to x_i}$ – частная информация о системе X, содержащаяся в событии $(Y \approx y_j)$, $I_{y_j\to x_i} = \log\left(\frac{P(x_i / y_j)}{p_i}\right)$ – частная информация о событии $(X \approx x_i)$, содержащаяся в событии $(Y \approx y_j)$.

Замечание. Согласно шестому свойству энтропии $I_{Y \to X} \geq 0$ и $I_{y_j \to X} \geq 0$, а частная информация $I_{y_j \to x_i}$ может быть как положительной, так и ***отрицательной***. Если условная вероятность $P(x_i / y_j) > p_i$, то $I_{y_j \to x_i} > 0$; в противном случае $I_{y_j \to x_i} < 0$.

§3. Энтропия и информация для систем с непрерывным множеством состояний

На практике часто встречаются физические системы, аналогичные непрерывным случайным величинам. Состояния таких систем нельзя перенумеровать и каждое отдельное состояние имеет вероятность, равную нулю. Такие системы называют ***непрерывными*** системами, в отличие от рассмотренных выше дискретных систем.

Рассмотрим систему X, определяемую случайной величиной X с плотностью распределения $f(x)$. По теореме Бернулли и теореме о среднем

$$f(x_i) = \frac{\nu_i}{n\Delta x}, \tag{21}$$

где n – объем выборки, Δx – длина каждого частичного интервала, ν_i – число выборочных значений в i-ом частичном интервале, x_i – средина частичного интервала. Из (21) видно, что вероятность попадания случайной точки в i–й прямоугольник равна:

$$p_i \approx \frac{\nu_i}{n} = f(x_i)\Delta x. \tag{22}$$

Тогда энтропию системы X, рассматриваемую с точностью до Δx, приближенно можно определить следующим образом

$$H(X) = -\sum_{i=1}^{n} f(x_i)\Delta x \log(f(x_i)\Delta x) = -\sum_{i=1}^{n} [f(x_i)\log(f(x_i))]\Delta x - \log(\Delta x)\sum_{i=1}^{n} f(x_i)\Delta x. \tag{23}$$

Переходя в (23) к пределу $\Delta x \to 0$, получим

$$\sum_{i=1}^{n} f(x_i)\log(f(x_i))\Delta x = \int_{-\infty}^{\infty} f(x)\log(f(x))dx,$$

$$\sum_{i=1}^{n} f(x_i)\Delta x = \int_{-\infty}^{\infty} f(x)dx = 1,$$

так что формула (23) принимает вид

$$H(X) = -\int_{-\infty}^{\infty} f(x)\log(f(x))dx - \log(\Delta x). \tag{24}$$

Частная условная энтропия определяется по аналогии

$$H(Y / x) = -\int_{-\infty}^{\infty} f(y / x)\log(f(y / x))dy - \log(\Delta y). \tag{25}$$

Тогда полная условная энтропия равна:

$$H(Y / X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)f(y / x)\log(f(y / x))dxdy - \log(\Delta y)$$

или, учитывая, что

$$f(x,y)=f(x)f(y/x),$$

$$H(Y/X)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\log(f(y/x))dxdy-\log(\Delta y). \qquad (26)$$

При этом энтропия системы (X,Y) равна:

$$H(X,Y)=H(X)+H(Y/X). \qquad (27)$$

Полная взаимная информация дается выражением

$$I_{Y\to X}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\log\left(\frac{f(x,y)}{f_1(x)f_2(y)}\right)dxdy. \qquad (28)$$

Полная взаимная информация, как и в случае дискретных систем, есть неотрицательная величина, обращающаяся в нуль только тогда, когда системы X и Y независимы.

§4. Кодирование и передача информации без искажений

При передаче сообщений по линии связи всегда приходится пользоваться кодом. Сначала рассмотрим вопросы, связанные с кодированием и передачей информации по каналу связи в идеальном случае, когда процесс передачи информации осуществляется без искажений.

Кодированием называется отображение состояния одной физической системы X с помощью состояния другой системы Y. Ограничимся случаем, когда обе системы X и Y имеют конечное число возможных состояний.

Пусть имеется физическая система X (например, буква русского алфавита), которая может случайным образом принять одно из состояний $x_1, x_2, \ldots\ldots x_n$. Будем кодировать ее с помощью системы Y, возможные состояния которой $y_1, y_2, \ldots y_m$. Если $m<n$, то нельзя каждое состояние x_i закодировать с помощью одного–единственного состояния y_j. В этом случае одно состояние x_i приходится кодировать с помощью комбинации состояний y_j. Выбор таких комбинаций и установление их соответствия передаваемым сообщениям называется кодированием в узком смысле слова.

Коды различаются по числу m. Если $m=2$, то код называется двоичным, который широко применяется на практике. Если на передачу закодированного сообщения затрачивается минимальное время, то код считается оптимальным. Когда на передачу каждого элементарного символа (например, 0 или 1) тратится одно и то же время, то оптимальным является код, при котором на передачу сообщения заданной длины будет затрачено минимальное количество элементарных символов.

Пример. Закодируем двоичным кодом буквы русского алфавита так, чтобы каждой букве соответствовала определенная комбинация элементарных символов 0 и 1, и чтобы среднее число этих символов на букву было минимальным.

Решение. Буквам, записанным в алфавитном порядке, присвоим номера от 0 до 31 (будем не различать буквы ь и ъ, е и ё; добавим знак пробела «–»).

Затем переведем нумерацию в двоичную систему исчисления. Так как число букв 32, то двоичный код должен быть пятизначным.

Номер буквы «а» изобразится в виде

$$0 = 0 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0$$

и в пятизначной двоичной системе запишется в виде 00000. Номер буквы «б» изобразится в виде

$$1 = 0 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$$

и в пятизначной двоичной системе запишется в виде 00001. Для «в» – $2 = 0 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0$ и код – 00010. Тогда получим следующий код:

$$а \to 00000,\ б \to 00001,\ в \to 00010,\ г \to 00011, \ldots\ldots, я \to 11110,\ (-) \to 11111$$

Определим оптимальность такого способа кодирования. Для этого будем использовать таблицу частот букв русского алфавита

Таблица 1

Буква	Частота	Буква	Частота	Буква	Частота	Буква	Частота
«–»	0.145	р	0.041	я	0.019	х	0.009
о	0.095	в	0.039	ы	0.016	ж	0.008
е	0.074	л	0.036	з	0.015	ю	0.007
а	0.064	к	0.029	ь,ъ	0.015	ш	0.006
и	0.064	м	0.026	б	0.015	ц	0.004
т	0.056	д	0.026	г	0.014	щ	0.003
н	0.056	п	0.024	ч	0.013	э	0.003
с	0.047	у	0.021	й	0.010	ф	0.002

Найдем среднюю информацию, содержащуюся в одной в одной букве передаваемого текста, т.е. энтропию на одну букву

$$H = -\sum_{i=1}^{32} p_i \log(p_i) = 4.42 \text{ (bit)}.$$

Деля энтропию на среднее число элементарных символов на букву $n_c = 5$, получим информацию на один элементарный символ

$$I_{1c} = \frac{4.42}{5} = 0.884 \text{ (bit)}.$$

Согласно третьему свойству энтропии максимальная информация на один элементарный символ 0 или 1 равна:

$$I_{1\max} = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1 \text{ (bit)}. \qquad (29)$$

Составим более экономичный код. Из (29) следует, что для этого нужно потребовать, чтобы элементарные символы 0 и 1 в закодированном тексте встречались в среднем одинаково часто $\left(p_1 = p_2 = \frac{1}{2}\right)$. Будем использовать принцип построения кода, известный под названием «код Шеннона – Фэно». Идея принципа состоит в том, что кодируемые символы (буквы или их комбинации) разделяются на две приблизительно **равновероятные** группы: для первой группы символов на первом месте ставится 0 (это первый знак двоичного числа, изображающего символ (букву)

); для второй группы – 1. Далее каждая группа снова делится на две приблизительно **равновероятные** подгруппы; для символов (букв) первой подгруппы на втором месте ставится 0; для второй подгруппы – единица и т.д. Процесс продолжается до тех пор пока в каждом подразделении не останется ровно одна буква. Механизм построения кода представим в виде следующей таблицы

Таблица 2

<table>
<tr><th></th><th colspan="9">Двоичные знаки</th></tr>
<tr><th>Буквы</th><th>$1^й$</th><th>$2^й$</th><th>$3^й$</th><th>$4^й$</th><th>$5^й$</th><th>$6^й$</th><th>$7^й$</th><th>$8^й$</th><th>$9^й$</th></tr>
<tr><td>–</td><td rowspan="6">0</td><td rowspan="2">0</td><td>0</td><td rowspan="2"></td><td rowspan="8"></td><td rowspan="13"></td><td rowspan="23"></td><td rowspan="27"></td><td rowspan="30"></td></tr>
<tr><td>о</td><td>1</td></tr>
<tr><td>е</td><td rowspan="4">1</td><td rowspan="2">0</td><td>0</td></tr>
<tr><td>а</td><td>1</td></tr>
<tr><td>и</td><td rowspan="2">1</td><td>0</td></tr>
<tr><td>т</td><td>1</td></tr>
<tr><td>н</td><td rowspan="25">1</td><td rowspan="6">0</td><td rowspan="2">0</td><td>0</td></tr>
<tr><td>с</td><td>1</td></tr>
<tr><td>р</td><td rowspan="4">1</td><td rowspan="2">0</td><td>0</td></tr>
<tr><td>в</td><td>1</td></tr>
<tr><td>л</td><td rowspan="2">1</td><td>0</td></tr>
<tr><td>к</td><td>1</td></tr>
<tr><td>м</td><td rowspan="19">1</td><td rowspan="6">0</td><td rowspan="3">0</td><td>0</td></tr>
<tr><td>д</td><td rowspan="2">1</td><td>0</td></tr>
<tr><td>п</td><td>1</td></tr>
<tr><td>у</td><td rowspan="3">1</td><td>0</td><td></td></tr>
<tr><td>я</td><td rowspan="2">1</td><td>0</td></tr>
<tr><td>ы</td><td>1</td></tr>
<tr><td>з</td><td rowspan="13">1</td><td rowspan="4">0</td><td rowspan="2">0</td><td>0</td></tr>
<tr><td>ъ,ь</td><td>1</td></tr>
<tr><td>б</td><td rowspan="2">1</td><td>0</td></tr>
<tr><td>г</td><td>1</td></tr>
<tr><td>ч</td><td rowspan="9">1</td><td rowspan="3">0</td><td>0</td></tr>
<tr><td>й</td><td rowspan="2">1</td><td>0</td></tr>
<tr><td>х</td><td>1</td></tr>
<tr><td>ж</td><td rowspan="6">1</td><td rowspan="2">0</td><td>0</td></tr>
<tr><td>ю</td><td>1</td></tr>
<tr><td>ш</td><td rowspan="4">1</td><td rowspan="2">0</td><td>0</td></tr>
<tr><td>ц</td><td>1</td></tr>
<tr><td>щ</td><td rowspan="2">1</td><td>0</td></tr>
<tr><td>э</td><td></td><td>0</td></tr>
</table>

ф								1	1

С помощью таблицы 2 сам код представим в виде таблицы

Таблица 3

Буква	Двоичное число	Буква	Двоичное число	Буква	Двоичное число	Буква	Двоичное число
«–»	000	р	10100	я	110110	х	1111011
о	001	в	10101	ы	110111	ж	1111100
е	0100	л	10110	з	111000	ю	1111101
а	0101	к	10111	ь,ъ	111001	ш	11111100
и	0110	м	11000	б	111010	ц	11111101
т	0111	д	110010	г	111011	щ	11111110
н	1000	п	110011	ч	111100	э	111111110
с	1001	у	110100	й	1111010	ф	111111111

С помощью таблицы 3 можно закодировать и декодировать любое сообщение.

Согласно таблицам 1 и 3 среднее число элементарных символов на одну букву равно:

$$n_c = 3 \cdot 0.145 + 3 \cdot 0.095 + 4 \cdot 0.074 + ... + 9 \cdot 0.003 + 9 \cdot 0.002 = 4.45,$$

так что информация на один элементарный символ

$$I_{1c} = \frac{4.42}{4.45} = 0.994 \text{ (bit)}.$$

Таким образом, код Шеннона – Фэно близок к своему верхнему пределу 1. Оставаясь в пределах кодирования по буквам лучшего результата получить нельзя. Более экономный код можно построить, если кодировать не отдельные буквы, а блоки из букв. Кодирование по блокам осуществляется по тому же принципу.

Обычно источники передают сообщения с некоторой скоростью, затрачивая в среднем время T на передачу одного сообщения. ***Производительностью*** источника $H(X)$ назовем суммарную энтропию сообщений переданных за единицу времени

$$H'(X) = \frac{1}{T} H(X)$$

Производительность измеряется в битах на секунду. Если сообщение может быть представлено в виде последовательности элементарных дискретных сообщений x_i источника с энтропией $H(X)$ следующих со скоростью $v_c = \frac{1}{T}$ элементов в секунду, то

$$H'(X) = v_c H(X).$$

§5. Кодирование и передача информации с искажениями

Канал передачи, в котором, в котором возможны искажения, называется ***каналом с помехами.*** Очевидно, что наличие помех приводит к потере информации. Для борьбы с помехами приходится принимать специальные меры. Одной из таких мер является введение ***избыточности*** в передаваемое сообщение.

Мерой избыточности языка служит величина

$$U = 1 - \frac{H'}{H_{\max}}, \qquad (30)$$

где H' – средняя фактическая энтропия, приходящаяся на один передаваемый символ (букву), рассчитанная для достаточно длинных отрывков текста, с учетом зависимости между символами, n – число применяемых символов (букв), $H_{\max} = \log(n)$ – максимальная энтропия на один передаваемый символ, которая была бы, если бы все символы были **равновероятны и независимы**. Избыточность языка может оказаться как чрезмерной, так и недостаточной. С помощью методов теории информации можно для каждого уровня помех найти нужную степень избыточности источника информации.

Рассмотрим систему, состоящую из источника информации X, канала связи K и приемника Y. Источник информации представляет собой физическую систему X, которая имеет n возможных состояний

$$x_1, x_2, \ldots . x_n$$

с вероятностями

$$p_1 . p_2, \ldots . p_n .$$

Будем рассматривать эти состояния как элементарные символы, которые может передавать источник X через канал K к приемнику Y. Тогда при наличии искажений количество информации на один символ равно:

$$I_{Y \leftrightarrow X} = I_X - H(X/Y), \qquad (31)$$

где $I_X = H(X)$, $H(X)$ – энтропия системы X, приходящаяся на один передаваемый символ; условная энтропия $H(X/Y)$ характеризует ***потерю информации*** на один элементарный символ, связанную с наличием помех.

Пропускная способность канала с помехами равна:

$$C = \nu_k \cdot \max[I_{Y \leftrightarrow X}], \qquad (32)$$

ν_k – число элементарных символов, которые канал может передавать в единицу времени; $\max[I_{Y \leftrightarrow X}]$ – максимальное количество информации на один символ, которую может передавать канал при наличии искажений.

Пропускная способность канала удовлетворяет системе неравенств

$$0 \le C \le \nu_k \log(n)$$

Причем $C = 0$ при независимых входе и выходе канала, т.е. когда $H(X/Y) = H(X)$ (обрыв канала или сильные помехи). Значение

$$C = \nu_k \log(n)$$

наблюдается в том случае, когда помех в канале нет $H(X/Y) = H(Y/X) = 0$ и $H(X) = H(Y) = I_{Y \leftrightarrow X}$. Таким образом, пропускная способность дискретного канала без шума определяется равенством $C = \nu_k \log(n)$. При наличии шума

$$C < \nu_k \log(n).$$

Пример. Пусть канал связи K передает от источника информации X к приемнику Y элементарные символы 0 и 1 в количестве k символов в единицу времени. В процессе передачи каждый символ, независимо от других, с вероятностью μ может быть искажен (заменен противоположным). Найдем пропускную способность канала.

Решение. Пусть источник производит символы 0 и 1 с вероятностями p и $1-p$. Тогда энтропия источника будет

$$H(X) = -p\log(p) - (1-p)\log(1-p).$$

Чтобы найти полную условную энтропию $H(Y/X)$, найдем сначала частные условные энтропии: $H(Y/x_1)$ и $H(Y/x_2)$. Предположим, $x_1 = 0$. Тогда условные вероятности равны:

$$P(y_1 = 0/x_1 = 0) = 1-\mu,\; P(y_2 = 1/x_1 = 0) = \mu,$$

так что условная энтропия

$$H(Y/x_1 = 0) = -\sum_{i=1}^{2} P(y_i/x_1)\log(P(y_i/x_1)) = -[(1-\mu)\log(1-\mu) + \mu\log(\mu)].$$

Аналогично,

$$P(y_1 = 0/x_2 = 1) = \mu,\; P(y_2 = 1/x_2 = 1) = 1-\mu,$$

откуда

$$H(Y/x_2 = 1) = -[(1-\mu)\log(1-\mu) + \mu\log(\mu)].$$

Таким образом,

$$H(Y/X) = pH(Y/x_1 = 0) + (1-p)H(Y/x_2 = 1) = -[\mu\log(\mu) + (1-\mu)\log(1-\mu)].$$

Пусть q – вероятность того, что на выходе появится символ 0, $1-q$ – вероятность того, что появится символ 1. Тогда информация, передаваемая одним символом равна:

$$I_{Y\leftrightarrow X} = H(Y) - H(Y/X) = -[q\log(q) + (1-q)\log(1=q)] + [\mu\log(\mu) + (1-\mu)\log(1-\mu)].$$

Так как $\max[H(Y)] = \log(2) = 1$, то

$$\max[I_{Y\leftrightarrow X}] = 1 + [\mu\log(\mu) + (1-\mu)\log(1-\mu)],$$

и пропускная способность канала связи будет равна

$$C_1 = \nu_k \cdot \{1 + [\mu\log(\mu) + (1-\mu)\log(1-\mu)]\}.$$

С помощью аналогичных расчетов может быть определена пропускная способность канала и в более сложных случаях.

Предельные возможности статистического кодирования раскрываются в первой теореме Шеннона для канала без шума, которая является одним из основных положений теории передачи информации.

Первая теорема Шеннона

Если $C > H'(X)$, то всегда можно закодировать достаточно длинное сообщение так, чтобы оно передавалось каналом связи без **задержек**. Если же $C < H'(X)$, то передача информации без **задержек** невозможна.

В случае передачи информации с искажениями справедлива вторая теорема Шеннона.

Вторая теорема Шеннона

Если $C > H'(X)$, то всегда можно закодировать достаточно длинное сообщение так, чтобы оно передавалось каналом связи без задержек и искажений с вероятностью, сколь угодно близкой к единице. Если же $C < H'(X)$, то передача информации без **задержек и искажений** невозможна.