

Лекции по теории информации

А.А. Соловьев

Оглавление

1. Введение	2
2. Количественные информационные характеристики дискретных источников сообщений	4
2.1 Энтропия и ее свойства	4
2.2 Условная информация. Условная энтропия.	8
2.3 Кодирование дискретных источников неравномерными кодами	10
2.4 Оптимальные неравномерные коды	16
3. Теоремы кодирования для каналов связи	19
3.1 Средняя взаимная информация между источниками	19
3.2 Постановка задачи кодирования в дискретном канале	24
3.3 Информационная емкость дискретных каналов без памяти	27
3.4 Методы декодирования	29
3.5 Помехоустойчивое кодирование в ДСК	32
3.6 Прямая и обратная теорема кодирования для дискретного канала без памяти	34
3.7 Теорема Шеннона для ДСК канала	35
4. Конспект лекций по теории кодирования	39
4.1 Линейные коды	39
4.2 Циклические коды	42

Глава 1

Введение

Информация, наряду с материей и энергией, является первичным понятием и в строгом смысле не может быть определена. В повседневной жизни под информацией обычно понимают совокупность сведений об окружающем мире, являющихся объектом хранения, передачи и преобразования.

Знаки и сигналы, организованные в последовательности, несут информацию в силу однозначного соответствия с объектами и понятиями реального мира, например: предметы и слова их обозначающие.

Информация, основанная на однозначной связи знаков и сигналов с объектами реального мира, называется *семантической или смысловой*.

Информация, заключенная в характере следования знаков (порядке и взаимосвязи) называется *синтаксической*.

В курсе теории информации изучаются проблемы синтаксического уровня, касающиеся создания теоретических основ построения *систем связи*, основные показатели функционирования которых были бы близки к предельно возможным. Рассмотрению подлежат вопросы доставки получателю информации как совокупности знаков. При этом полностью игнорируется смысловое и прагматическое содержание информации. Синтаксическая информация также имеет практическую ценность потому, что интересующая в конечном итоге семантическая информация заключена в доставляемой получателю последовательности знаков или сигналов.

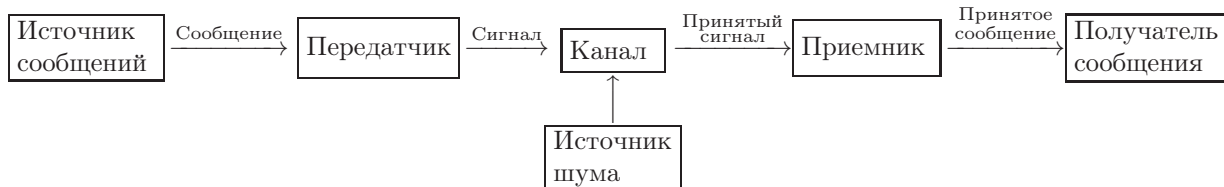
Введем некоторые понятия и определения. Информация, представленная в какой-либо форме называется *сообщением*. Для того, чтобы сообщение можно было передать получателю, необходимо воспользоваться некоторым физическим процессом, способного с той или иной скоростью распространяться от источника к получателю сообщения. Изменяющийся во времени физический процесс, отражающий передаваемое сообщение, называется *сигналом*. Сигнал является функцией времени и их делают на четыре типа:

- 1) непрерывный или аналоговый сигнал (т.е. аналогичный порожденному процессу);
- 2) дискретный по времени сигнал или последовательность отсчетов (временной интервал между соседними отсчетами $\Delta t = t_{k+1} - t_k$ называется шагом дискретизации);
- 3) дискретный по уровню или квантованный сигнал (принимает лишь разрешенные значения уровня, отделенные друг от друга шагом квантования $\Delta x = x_{k+1} - x_k$);
- 4) дискретный по уровню и по времени.

Дискретная информация удобней для обработки, но непрерывная информация встречается чаще. Как например модем, который переводит цифровые данные в звуковой сигнал и наоборот.

При дискретизации сигнала часть информации, как правило, теряется. Теорема об отсчетах Найквиста – Шеннона – Котельникова гласит, что для точной дискретизации сигнала частота дискретизации должна быть не менее, чем в два раза выше наибольшей частоты сигнала.

Совокупность технических средств, используемых для передачи сообщений от источника к потребителю информации называется *системой связи*. Приведем пример системы связи:



1. Сообщения могут быть разных типов: последовательностью букв или цифр, а также одной или более функцией времени.

2. Передатчик перерабатывает некоторым образом сообщения в сигналы определенного типа.

3. Канал – это комплекс технических средств, обеспечивающих передачу сигналов от передатчика к приемнику по *линии связи*. Линией связи называется среда, используемая для передачи сигнала от приемника к передатчику (пара проводов, коаксиальный кабель, световод, область распространения радиоволн). Если сигнал на входе и выходе канала непрерывен по уровню (типа 1) или 2)), то канал называется непрерывным. Канал называется дискретным, если на его входе и выходе присутствуют сигналы, дискретные по уровню (типа 3) или 4)). В общем случае в процессе передачи сигнал искажается шумом, что соответствует наличию источника шума.

4. Приемник восстанавливает сообщение по принимаемому сигналу.

Процесс преобразования сообщения в сигнал, осуществляющийся в передатчике, называется *кодированием* и обратный ему процесс, реализуемый в приемнике, – *декодированием*.

Теория информации (ТИ) исследует методы кодирования для экономного представления сообщений различных источников сообщений и для надежной передачи сообщений по каналам связи с шумом.

В основе ТИ лежит *статистическое* описание источников сообщений и понятие *количества информации*, содержащейся в сообщении. Теория информации является разделом статистической теории связи.

На основе ТИ можно ответить на вопросы о предельных возможностях реальных систем и определить в какой мере проектируемая система уступает теоретически возможной.

Датой рождения ТИ является 1948 г. В этот год вышла основополагающая статья Клода Шеннона "Математическая теория связи". Начиная с этого времени ТИ интенсивно развивалась в немалой степени благодаря работам и наших соотечественников Колмогорова, Добрушина, Харкевича, Хинчина и других. При подготовке лекций были использованы источники [1- 5].

Глава 2

Количественные информационные характеристики дискретных источников сообщений

2.1 Энтропия и ее свойства

Источник сообщений может в каждую единицу времени случайным образом принять одно из возможных состояний. Каждому состоянию источника ставится в соответствие условное обозначение в виде знака. Совокупность знаков u_1, u_2, \dots, u_N соответствующих всем N состояниям источника называется его *алфавитом*, а количество состояний N *объемом алфавита*. Под *элементарным дискретным сообщением* будем понимать символ u_j , генерируемый источником. В течение времени T источник порождает *дискретное сообщение* в виде последовательность символов. Отдельные состояния источника выбираются им чаще, другие реже. Поэтому каждое состояние u_j принимается дискретным источником с определенной вероятностью $p(u_j)$.

Определение 1. Дискретным источником сообщений будем называть конечное множество U вместе с заданным на нем распределением вероятностей $p(u)$, $x \in U$ и будем обозначать его символом $\{U, p(u)\}$. То есть, под дискретным источником сообщений понимается конечное дискретное вероятностное пространство.

Пусть $X = \{x_1, \dots, x_M\}$ и $Y = \{y_1, \dots, y_N\}$ – два конечных множества. Символом XY будем обозначать декартово произведение множеств X и Y , элементами которого являются упорядоченные пары (x_i, y_j) , $x_i \in X$, $y_j \in Y$, $i = 1, \dots, M$, $j = 1, \dots, N$. Если $X = Y$ то произведение XY будем обозначать через X^2 . Аналогичным образом определяются произведения более чем двух множеств. В частности, X^n – это множество всех последовательностей длины n элементов множества X .

Пусть на множестве XY задано совместное распределение вероятностей $p(x, y)$, которое каждой паре (x_i, y_j) , $x_i \in X$, $y_j \in Y$, сопоставляет вероятность $p(x_i, y_j)$. Соотношения

$$p_1(x_i) = \sum_{y_j \in Y} p(x_i, y_j), \quad i = 1, \dots, M,$$
$$p_2(y_j) = \sum_{x_i \in X} p(x_i, y_j), \quad j = 1, \dots, N,$$

задают распределения вероятностей $p_1(x)$ и $p_2(y)$ на множествах X и Y . Таким образом, при задании источника $\{XY, p(x, y)\}$ фактически задаются еще два источника $\{X, p_1(x)\}$ и $\{Y, p_2(y)\}$. Источники $\{X, p_1(x)\}$ и $\{Y, p_2(y)\}$ будем называть совместно заданными источником $\{XY, p(x, y)\}$.

Если распределение вероятностей на произведении двух множеств X и Y удовлетворяют условию

$$p(x_i, y_j) = p_1(x_i)p_2(y_j) \quad \text{для всех } x_i \in X, y_j \in Y,$$

то источники $\{X, p_1(x)\}$ и $\{Y, p_2(y)\}$ называются *статистически независимыми*. В противном случае, говорят, что эти источники *статистически зависимы*.

В каждом элементарном сообщении содержится для получателя информация о состоянии источника сообщений. При определении количественной меры информации не учитывается ее смысловое содержание. Количество информации, содержащейся в дискретном сообщении измеряется величиной исчезнувшей в ходе эксперимента неопределенности. Поэтому меру неопределенности можно

рассматривать как количественную меру информации содержащейся в сообщении. Определение меры неопределенности обсудим на примере источника U с равновероятными состояниями.

Мера должна удовлетворять ряду естественных условий. С увеличением объема выбора, то есть объема алфавита источника, мера неопределенности должна возрасти. Кроме того, вводимая мера неопределенности должна обладать свойством аддитивности: если два независимых источника X и Y с объемами алфавитов M и N объединены в один источник, реализующий пары состояний (x_i, y_j) , то неопределенность объединенного источника должна быть равной сумме неопределенностей исходных источников. Мера неопределенности в случае равновероятности состояний является функцией объема источника и поскольку объем алфавита объединенного источника равен MN , то искомая функция должна удовлетворять условию

$$f(MN) = f(M) + f(N).$$

Функцией, удовлетворяющей этому соотношению, является логарифмическая функция. Перечисленные требования выполняются, если в качестве меры неопределенности источника с равновероятными состояниями принять логарифм объема алфавита источника с основанием большим единицы

$$H(U) = \log N.$$

Ясно, что

- а) с ростом N величина $H(U)$ монотонно возрастает;
- б) если объем алфавита источника равен $N = 1$, то $H(U) = \log 1 = 0$, то есть неопределенность отсутствует;
- в) величина $H(U)$ обладает свойством аддитивности

$$\log MN = \log M + \log N.$$

Основание логарифма определяет единицу количества информации. Если основание равно 2, то единица количества информации называется битом и представляет собой информацию, содержащуюся в одном дискретном сообщении источника равновероятных сообщений с объемом алфавита, равным двум. Если основание равно 10, то получаем единицу, называемую дитом. С основанием e единица информации называется натом.

Данная мера неопределенности была предложена Хартли в 1928 году.

В общем случае, когда вероятности различных состояний источника $\{U, p(u)\}$ с объемом N не одинаковы, степень неопределенности конкретного состояния зависит не только от объема алфавита источника, но и от вероятности этого состояния. В таком случае количество информации, содержащейся в одном дискретном сообщении u_k , имеет смысл определить как функцию вероятности $p(u_k)$ появления этого дискретного сообщения

$$I(u_k) = -\log p(u_k) = \log \frac{1}{p(u_k)}.$$

Знак $(-)$ выбирается с тем, чтобы $I(u_k) \geq 0$. В случае достоверного сообщения, когда $p(u_k) = 1$, имеем $I(u_k) = 0$.

Количество информации, содержащейся в дискретном сообщении источника является случайной величиной, так как зависит от степени неожиданности (вероятности) реализуемого источником сообщения. Среднее количество информации, содержащееся в отдельном сообщении, называется *энтропией* источника

$$H(U) = M \left\{ \log \frac{1}{p(u)} \right\} = \sum_{i=1}^N p(u_i) \log \frac{1}{p(u_i)}. \quad (2.1)$$

Чем больше энтропия источника, тем больше степень неопределенности реализуемых им сообщений в среднем, то есть более неопределенным является ожидание сообщений. Впервые мера (2.1)

была предложена Клодом Шенноном в его фундаментальной работе "Математические основы теории связи" опубликованной в 1948 году. Название "энтропия" не случайно, так как соотношение (2.1) совпадает с выражением для энтропии Больцмана термодинамической системы.

Рассмотрим теперь свойства энтропии:

1. Энтропия любого дискретного источника неотрицательна, $H(U) \geq 0$. Равенство возможно лишь в том случае, когда источник генерирует одно единственное сообщение с вероятностью, равной единице.

2. Пусть N – объем алфавита дискретного источника. Тогда $H(U) \leq \log N$. Причем равенство имеет место только в том случае, когда все сообщения равновероятны.

$$H(U) - \log N = \sum_{k=1}^N p(u_k) \log \frac{1}{p(u_k)} - \log N \sum_{k=1}^N p(u_k) = \sum_{k=1}^N p(u_k) \log \frac{1}{p(u_k)N}.$$

Так как $\ln x < x - 1$ при $x > 0$ и $\ln x = \frac{\log x}{\log e}$, то

$$\begin{aligned} H(U) - \log N &= \log e \sum_{k=1}^N p(u_k) \ln \frac{1}{Np(u_k)} \leq \log e \sum_{k=1}^N p(u_k) \left[\frac{1}{Np(u_k)} - 1 \right] = \\ &= \log e \sum_{k=1}^N \left[\frac{1}{N} - p(u_k) \right] = \log e(1 - 1) = 0. \end{aligned}$$

то есть $H(U) \leq \log N$.

3. Свойство аддитивности – энтропия нескольких совместно заданных статистических дискретных источников сообщений равна сумме энтропий исходных источников.

Энтропия совместного источника $\{XY, p(x, y)\}$ равна

$$\begin{aligned} H(XY) &= \sum_{i=1}^M \sum_{j=1}^N p(x_i)p(y_j) \log \frac{1}{p(x_i)p(y_j)} = \\ &= \sum_{j=1}^N p(y_j) \sum_{i=1}^M p(x_i) \log \frac{1}{p(x_i)} + \sum_{i=1}^M p(x_i) \sum_{j=1}^N p(y_j) \log \frac{1}{p(y_j)} = H(X) + H(Y). \end{aligned}$$

Предложение 2.1. Для любых двух вероятностных распределений $p(u)$ и $q(u)$ на алфавите $U = \{u_1, \dots, u_N\}$ справедливо неравенство

$$\sum_{i=1}^N p(u_i) \log \frac{1}{p(u_i)} \leq \sum_{i=1}^N p(u_i) \log \frac{1}{q(u_i)},$$

которое переходит в равенство тогда и только тогда, когда $p(u_i) = q(u_i)$ для всех $u_i \in U$.

Доказательство.

$$\begin{aligned} \sum_{i=1}^N p(u_i) \log \frac{1}{p(u_i)} - \sum_{i=1}^N p(u_i) \log \frac{1}{q(u_i)} &= \sum_{i=1}^N p(u_i) \log \frac{q(u_i)}{p(u_i)} = \log e \sum_{i=1}^N p(u_i) \ln \frac{q(u_i)}{p(u_i)} \leq \\ &\leq \log e \sum_{i=1}^N p(u_i) \left[\frac{q(u_i)}{p(u_i)} - 1 \right] = \log e \left[\sum_{i=1}^N q(u_i) - \log e \left[\sum_{i=1}^N p(u_i) \right] \right] = \log e(1 - 1) = 0. \end{aligned}$$

□

Следствием этого предложения является, в частности, свойство 2.

Избыточностью источника дискретных сообщений с энтропией H и объемом алфавита N называется величина, равная

$$1 - \frac{H}{\log N},$$

где $\log N$ – максимально возможное значение энтропии при данном объеме алфавита. Избыточность показывает, какая доля возможной при заданном объеме алфавита неопределенности (энтропии) не используется источником. В частности, избыточность английского текста составляет 50%, избыточность русского текста – 70%.

Пример 1. Энтропия двоичного источника

$$U = \{0, 1\}, \quad P(0) = p, \quad P(1) = 1 - p$$

равна

$$H(U) = p \log_2 \frac{1}{p} + (1 - p) \log_2 \frac{1}{1 - p} = h(p).$$

Функция $h(p)$ называется *двоичной энтропией*. Здесь $0 \leq h(p) \leq 1$ и переходит в равенство при $p = \frac{1}{2}$. В последнем случае источник называется *двоичным симметричным источником* (ДСИ) и каждый символ на выходе ДСИ содержит один бит информации.

Пример 2. Некто задумал целое число в интервале от 0 до 3. Опыт состоит в угадывании этого числа. На наши вопросы Некто может отвечать только "Да" или "Нет". Сколько вопросов мы должны задать, чтобы узнать задуманное число, или иначе, какое количество информации мы должны получить, чтобы полностью снять начальную неопределенность.

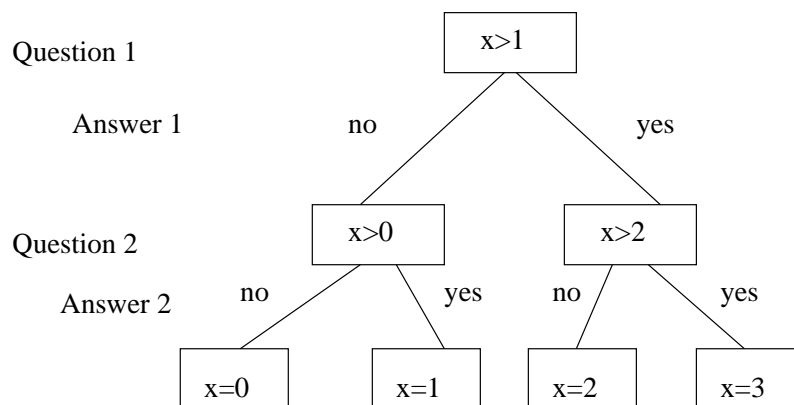
Решение. Исходами в данном случае являются:

A_1 ="задуман 0"; A_2 ="задуман 1"; A_3 ="задуман 2"; A_4 ="задуман 3".

Естественно предположить, что вероятности "задумать число" у всех чисел одинаковы: $N = 4$, следовательно, $p(A_i) = 1/4$, $\log_2 p(A_i) = -2$ и $H = 2$ битам. Для полного снятия неопределенности опыта (угадывания задуманного числа) нам необходимы 2 бита информации, то есть ответы на 2 вопроса с двумя возможными вариантами ответов (да – нет).

Количество информации должно быть равно числу вопросов с бинарными вариантами ответов, которые необходимо задать, чтобы полностью снять неопределенность задачи.

Убедимся, что два полученных ответа полностью снимают неопределенность и, тем самым, позволяют узнать задуманное число.



Таким образом, действительно, два полученных ответа решают задачу.

2.2 Условная информация. Условная энтропия.

Пусть $\{X, p(x)\}$ и $\{Y, p(y)\}$ совместно заданы источником $\{XY, p(x, y)\}$. Зафиксируем некоторое элементарное сообщение $y_j \in Y$, $p(y_j) \neq 0$ и рассмотрим условное распределение $p(x|y_j)$ на X . Для каждого сообщения $x_i \in X$ источника $\{X, p(x)\}$ определена *условная собственная информация*

$$I(x_i|y_j) = -\log p(x_i|y_j),$$

элемента сообщения x_i при фиксированном сообщении y_j . Функцию $I(x|y_j)$, $x \in X$, можно рассматривать как случайную величину на вероятностном пространстве $\{X, p(x|y_j)\}$. Ее математическое ожидание

$$H(X|y_j) = \sum_{i=1}^M p(x_i|y_j) I(x_i|y_j) = - \sum_{i=1}^M p(x_i|y_j) \log p(x_i|y_j)$$

называется условной энтропией источника $\{X, p(x)\}$ относительно сообщения $y_j \in Y$.

В свою очередь, условную энтропию $H(X|y)$, $y \in Y$, можно рассматривать как случайную величину на вероятностном пространстве $\{Y, p(y)\}$.

Определение 2. Математическое ожидание $H(X|Y)$ случайной величины $H(X|y)$, определенной на вероятностном пространстве $\{Y, p(y)\}$ называется условной энтропией источника X относительно источника Y

$$\begin{aligned} H(X|Y) &= MH(X|y) = \sum_{j=1}^N p(y_j) H(X|y_j) = \\ &= - \sum_{i=1}^M \sum_{j=1}^N p(y_j) p(x_i|y_j) \log p(x_i|y_j) = - \sum_{i=1}^M \sum_{j=1}^N p(x_i, y_j) \log p(x_i|y_j). \end{aligned}$$

Рассмотрим свойства условной энтропии.

1. $H(X|Y) \leq H(X)$, Равенство имеет место тогда и только тогда, когда источники X и Y статистически независимы.

$$\begin{aligned} H(X|Y) - H(X) &= - \sum_{i=1}^M \sum_{j=1}^N p(x_i, y_j) \log p(x_i|y_j) + \sum_{i=1}^M p(x_i) \log p(x_i) = \\ &= \sum_{i=1}^M \sum_{j=1}^N p(x_i, y_j) \left[\log \frac{1}{p(x_i|y_j)} + \log p(x_i) \right] = \sum_{i=1}^M \sum_{j=1}^N p(x_i, y_j) \log \frac{p(x_i)}{p(x_i|y_j)} \leq \\ &\leq \log e \sum_{i=1}^M \sum_{j=1}^N p(x_i, y_j) \left[\frac{p(x_i)}{p(x_i|y_j)} - 1 \right] = \log e \left[\sum_{i=1}^M \sum_{j=1}^N p(x_i) p(y_j) - \sum_{i=1}^M \sum_{j=1}^N p(x_i, y_j) \right] = \log e(1 - 1) = 0. \end{aligned}$$

Равенство возможно тогда и только тогда, когда $p(x|y) = p(x)$, то есть когда x и y независимы для всех $x \in X$ и $y \in Y$.

Таким образом, результат опыта Y может уменьшить неопределенность опыта X .

2. Имеет место соотношение,

$$H(XY) = H(Y) + H(X|Y),$$

называемое свойством аддитивности энтропии. В самом деле, с помощью равенства $p(x, y) = p(y)p(x|y)$, находим

$$H(XY) = - \sum_{i=1}^M \sum_{j=1}^N p(x_i, y_j) \log p(x_i, y_j) - \sum_{i=1}^M \sum_{j=1}^N p(x_i, y_j) \log p(y_j) = H(X|Y) + H(Y).$$

Аналогично, пользуясь соотношением $p(x, y) = p(x)p(y|x)$ можно получить равенство

$$H(XY) = H(X) + H(Y|X).$$

3. Теорема о невозрастании информации при отображении.

Теорема 2.1. Пусть задан источник $\{X, p(x)\}$ и на нем определено отображение, $\varphi : X \rightarrow Y$. Это отображение определяет источник $\{Y, p(y)\}$, где $p(y) = \sum_{x: \varphi(x)=y} p(x)$. Пусть $H(X)$ и $H(Y)$ – энтропии источников X и Y , тогда

$$H(Y) \leq H(X).$$

Знак равенства имеет место тогда и только тогда, когда отображение $\varphi(x)$ обратимо, то есть φ является взаимно однозначным отображением X на Y .

Доказательство. Совместное распределение $p(x, y)$ на произведении множеств XY задается соотношением $p(x, y) = p(x)p(y|x)$, где $p(y|x) = 1$, если $y = \varphi(x)$ и $p(x, y) = 0$, если $y \neq \varphi(x)$. Тогда либо $\log p(y|x) = 0$, либо $p(y|x) = 0$. Поэтому

$$H(Y|X) = - \sum_{i=1}^M \sum_{j=1}^N p(x_i) p(y_j|x_i) \log p(y_j|x_i) = 0.$$

Из аддитивности и неотрицательности энтропии получим, что

$$H(Y) \leq H(Y) + H(X|Y) = H(X) + H(Y|X) = H(X).$$

Энтропия сохранится тогда и только тогда, когда $H(X|Y) = 0$. Поэтому для всех $y \in Y$ имеем $H(X|y) = 0$, значит, $p(x|y)I(x|y) = -p(x|y) \log p(x|y) = 0$ для всех $x \in X$ при каждом $y \in Y$. Тогда для каждого $y \in Y$ существует единственный $x \in X$ такой, что $p(x|y) = 1$. Последнее равенство выполняется, если $\varphi(x) = y$, то есть отображение φ обратимо. \square

В случае $H(X|Y) = 0$ будем говорить, что источник Y однозначно определяет источник X .

4. Пусть $\{XYZ, p(x, y, z)\}$ вводит три совместно заданных источника X, Y, Z и пусть

$$I(x|y, z) = -\log p(x|y, z)$$

есть условная собственная информация при фиксированной паре сообщений y, z , где

$$p(x|y, z) = \frac{p(x, y, z)}{\sum_{x \in X} p(x, y, z)}.$$

Число

$$H(X|YZ) = - \sum_{\substack{x \in X \\ y \in Y \\ z \in Z}} p(x, y, z) \log p(x|y, z)$$

называется условной энтропией источника X относительно пары источников Y, Z .

С помощью Предложения 2.1 доказывается следующее неравенство

$$H(X|YZ) \leq H(X|Y).$$

Равенство выполняется в том и только в том случае, когда

$$p(x|y, z) = p(x|y) \text{ для всех } (x, y, z) \in XYZ,$$

то есть когда при данном сообщении y сообщения x статистически независят от z .

$$H(X|YZ) = \sum_{\substack{x \in X \\ y \in Y \\ z \in Z}} p(x, y, z) \log \frac{1}{p(x|y, z)} \leq \sum_{\substack{x \in X \\ y \in Y}} \sum_{z \in Z} p(x, y, z) \log \frac{1}{p(x|y)} = H(X|Y).$$

В частности, верно неравенство

$$H(X|Y) \leq H(X).$$

Это неравенство обобщается на случай n совместно заданных источников. Рассмотрим источник $\{X_1, \dots, X_n, p(x^{(1)}, \dots, x^{(n)})\}$. Тогда для любых s и m , $1 \leq s \leq m \leq i$, выполняется неравенство

$$H(X_i|X_{i-1} \dots X_{i-s}) \leq H(X_i|X_{i-1} \dots X_{i-m}).$$

5. Свойство аддитивности допускает обобщение. Если $\{X_1, \dots, X_n, p(x^{(1)}, \dots, x^{(n)})\}$ – совместно заданный источник, тогда

$$H(X_1, \dots, X_n) = H(X_1) + H(X_2|X_1) + \dots + H(X_n|X_{n-1} \dots X_1).$$

Из свойства 4 следует, что

$$H(X_1, \dots, X_n) \leq \sum_{i=1}^n H(X_i)$$

и равенство возможно тогда и только тогда, когда источники $\{X_i, p_i(x^{(i)})\}$ статистически независимы, то есть

$$p(x^{(1)}, \dots, x^{(n)}) = \prod_{i=1}^n p_i(x^{(i)}),$$

где

$$p_i(x^{(i)}) = \sum_{k \neq i} \sum_{x^{(k)} \in X_k} p(x^{(1)}, \dots, x^{(n)}).$$

Если источники $\{X_i, p_i(x^{(i)})\}$ совпадают с источником $\{X, p(x)\}$ и статистически независимы, то

$$H(X^n) = nH(X).$$

2.3 Кодирование дискретных источников неравномерными кодами

Определение 3. Дискретным источником без памяти (ДИБП) называется источник сообщений такой, что для любых $n = 1, 2, \dots$ и любой последовательности и любой последовательности $(x^{(1)}, \dots, x^{(n)})$, $x^{(i)} \in X$, имеет место равенство

$$p(x^{(1)}, \dots, x^{(n)}) = \prod_{j=1}^n p(x^{(j)}).$$

Обозначим через A некоторое множество, состоящее из D , $D > 1$, элементов: $A = \{a_1, \dots, a_D\}$. Назовем его *алфавитом кода источника*. Элементы алфавита A будем называть *кодowymi символами*. Последовательности кодовых символов будем называть *кодowymi словами*, а любое семейство кодовых слов – *кодом над алфавитом A* .

Пример 3. Пусть $A = \{0, 1\}$. Тогда множества $M = \{011, 0101, 11, 10\}$ и $M = \{00, 01, 10, 11\}$ являются двоичными кодами объема 4.

Определение 4. Код называется равномерным, если все его слова имеют одинаковую длину m . Это число называется длиной кода. Если хотя бы два кодовых слова имеют различные длины, то код называется неравномерным.

Пример 4. Количество различных D -ичных последовательностей длины m равно D^m .

Количество различных слов неравномерного кода с максимальной длиной кодовых слов m равно $D(D^m - 1)/(D - 1)$.

Определение 5. Кодированием сообщений источника X посредством кода называется отображение (необязательно взаимно однозначное) множества сообщений в множество кодовых слов.

Примером неравномерного кода является код Шеннона-Фэно. При кодировании по методу Шеннона-Фэно алфавит расположенный в порядке убывания вероятностей появления символов, разбивается на две группы таким образом, чтобы сумма вероятностей появления символов в каждой группе была приблизительно одинаковой. Каждая группа в свою очередь также разбивается на две по такому же принципу. Операция продолжается до тех пор, пока в каждой группе не останется по одному символу. Каждый символ обозначается двоичным числом, последовательные цифры которого (нули и единицы) показывают в какую группу попал данный символ при очередном разбиении.

В коде Шеннона-Фэно часто встречающиеся буквы кодируются относительно короткими двоичными символами, а редкие – длинными.

Основной характеристикой неравномерного кодирования является количество символов, затрачиваемых при кодировании одного элементарного сообщения. Обозначим через m_i длину слова, кодирующего сообщение $x_i \in X$. Пусть $p(x_i)$ – вероятность этого сообщения. Тогда

$$\bar{m}(X) = \sum_{x_i \in X} m_i p(x_i)$$

есть средняя длина кодовых слов, кодирующих источник сообщений $\{X, p(x)\}$.

Предположим, что неравномерными кодами кодируются сообщения длины n , то есть кодируется источник сообщений $\{X^n, p(\bar{x})\}$.

Определение 6. Число

$$R = \frac{\bar{m}(X^n)}{n}$$

называется *средней скоростью* неравномерного кодирования посредством двоичного кода при разбиении последовательности сообщений на блоки длины n .

Пример 5.

X_i	$p(x_i)$	Равномерный код	Неравномерный код	m_i
x_1	1/4	000	00	2
x_2	1/4	001	01	2
x_3	1/8	010	100	3
x_4	1/8	011	101	3
x_5	1/16	100	1100	4
x_6	1/16	101	1101	4
x_7	1/16	110	1110	4
x_8	1/16	111	1111	4

Оба кода осуществляют побуквенное кодирование. Энтропия источника сообщений равна 2,75. Скорость кодирования в первом случае равна 3 бит на элементарное сообщение, во втором случае – 2,75 бит на элементарное сообщение.

Пример 6. Предположим, что источник порождает сообщения x_1, x_2, x_3, x_4 и эти сообщения кодируются кодовыми словами 0, 01, 10, 011 соответственно. Кодовый алфавит состоит из двух символов 0 и 1. Пусть на выходе источника появилось следующее сообщение $x_2x_3x_2x_1$. На выходе кодера возникает последовательность 0110010. Эта последовательность допускает несколько способов декодирования. Кроме правильного декодирования возможны варианты: $x_4x_1x_2x_1$ и $x_4x_1x_1x_3$.

Коды, в которых ни одно слово не является началом другого называются *префиксными*. Префиксные коды являются *кодами со свойством однозначного декодирования*.

Пример 7. Код 0, 01, 011 не является префиксным, но является однозначно декодируемым.

Определение 7. *Скоростью создания информации* источником $(X, p(x))$ при неравномерном кодировании называется наименьшее число H такое, что для любого $R > H$ найдется n (длина кодируемых сообщений) и неравномерный код со средней скоростью кодирования R , который допускает однозначное декодирование.

Будет доказано, что скорость кодирования при неравномерном кодировании, как и при равномерном кодировании, равна энтропии источника элементарных сообщений.

Как и ранее нужно доказать прямую и обратную теоремы. Первая из них утверждает, что при всех $R > H(X)$ найдется n и однозначно декодируемый неравномерный код со скоростью кодирования R , а вторая будет утверждать, что для любого $R < H(X)$ не существует однозначно декодируемого кода ни при каком n .

Теорема 2.2. *Предположим, что однозначно декодируемый двоичный код состоит из M слов длины которых равны m_1, \dots, m_M и кодовый алфавит двоичный. Тогда*

$$\sum_{i=1}^M 2^{-m_i} \leq 1.$$

Доказательство. Пусть L – произвольное положительное число. Имеем

$$\left(\sum_{i=1}^M 2^{-m_i} \right)^L = \sum_{i_1=1}^M \dots \sum_{i_L=1}^M 2^{-(m_{i_1} + \dots + m_{i_L})}. \quad (2.2)$$

В выражении в правой части равенства каждое слагаемое соответствует каждой возможной последовательности из L кодовых слов. Сумма $m_{i_1} + \dots + m_{i_L}$ равна суммарной длине соответствующей последовательности кодовых слов. Если через A_j обозначить число последовательностей из L кодовых слов, имеющих суммарную длину j , то (2.2) можно переписать в виде

$$\left(\sum_{i=1}^M 2^{-m_i} \right)^L = \sum_{j=1}^{Lm} A_j 2^{-j},$$

где m – максимальное из чисел m_1, \dots, m_M .

Так как 2^j – максимальное количество различных последовательностей длины j , то $A_j \leq 2^j$. Поэтому

$$\left(\sum_{i=1}^M 2^{-m_i} \right)^L \leq Lm$$

для всех возможных L . Поскольку слева стоит экспоненциальная функция, а справа – линейная функция переменной L , это неравенство может выполняться тогда и только тогда, когда

$$\sum_{i=1}^M 2^{-m_i} \leq 1.$$

□

Удобное описание префиксных кодов дают специальные графы, называемые кодовыми деревьями: 2-ичным деревом называется граф, в котором нет петель и в котором из каждого узла выходит не более 2 ребер и каждый узел, кроме корня дерева, входит только одно ребро.

Каждому из ребер, выходящему из узла, сопоставляется один символ двоичного кодового алфавита. Различным ребрам, выходящим из одного узла, сопоставляются различные символы.

Узлы дерева, отстоящие от корня на i ребер, образуют ярус порядка i . Порядком дерева называется максимальный из порядков его узлов. Узел, из которого не выходит ни одного ребра, называется конечным. Наконец, код является префиксным, если кодовые слова соответствуют только конечным узлам дерева.

Теорема 2.3. (Неравенство Крафта.) Для того, чтобы существовал двоичный префиксный код с длинами кодовых слов m_1, m_2, \dots, m_M необходимо и достаточно, чтобы

$$\sum_{i=1}^M 2^{-m_i} \leq 1. \quad (2.3)$$

Доказательство. Приведем здесь доказательство необходимости, отличное от приведенного в Теореме 2. Заметим, что максимальное количество узлов на ярусе j равно 2^j . Пусть $m = \max\{m_1, \dots, m_M\}$. Рассмотрим конечный узел порядка m_i . Этот узел отстоит от яруса m на $m - m_i$ ребер и, следовательно, исключает из этого яруса 2^{m-m_i} возможных узлов. Так как количество узлов, исключаемых из яруса m всеми конечными узлами порядков m_1, \dots, m_M , не может превосходить максимального количества узлов на этом ярусе, то

$$\sum_{i=1}^M 2^{m-m_i} \leq 2^m \quad (\text{см. Рис. 2.1}).$$

После деления обеих частей неравенства на 2^m получаем (2.3).

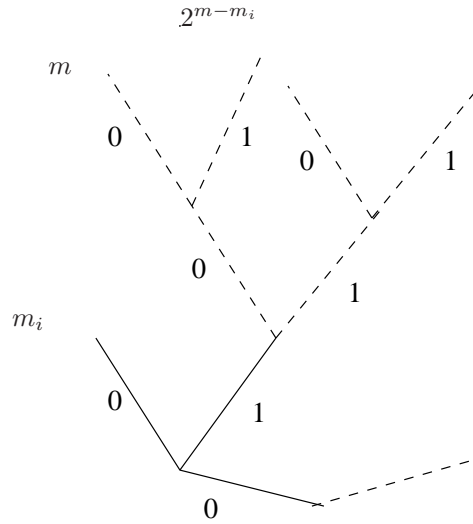


Рис. 2.1:

Достаточность. При выполнении (2.3) дерево с конечными узлами порядков m_1, \dots, m_M может быть построено. Предположим, что среди этого набора порядков число s встречается ровно α_s раз, $s = 1, \dots, m$. Тогда

$$\sum_{i=1}^M 2^{-m_i} = \sum_{s=1}^m \alpha_s 2^{-s} \leq 1.$$

Перепишем это неравенство следующим образом:

$$\sum_{s=1}^{i-1} \alpha_s 2^{-s} + \alpha_i 2^{-i} + \sum_{s=i+1}^m \alpha_s 2^{-s} \leq 1.$$

Тогда

$$\alpha_i \leq 2^i - \sum_{s=1}^{i-1} \alpha_s 2^{i-s} - \sum_{s=i+1}^m \alpha_s 2^{i-s} \leq 2^i - \sum_{s=1}^{i-1} \alpha_s 2^{i-s}. \quad (2.4)$$

Применим метод математической индукции.

Убедимся, что дерево, содержащее α_1 концевых узлов порядка 1 может быть построено. Так как из (2.3) следует, что $\alpha_1 2^{-1} \leq 1$, то $\alpha_1 \leq 2$. Максимально возможное количество концевых узлов порядка 1 равно 2 и $\alpha_1 \leq 2$. Поэтому дерево с α_1 концевыми узлами порядка 1 может быть построено.

Предположим, что дерево с α_s концевыми узлами порядка s , $s = 1, \dots, i-1$, может быть построено. Докажем, что к этому дереву можно добавить еще α_i концевых узлов порядка i . Если верно предположение индукции, то из яруса порядка i исключается $\sum_{s=1}^{i-1} \alpha_s 2^{i-s}$ возможных концевых узлов (каждый узел из яруса порядка s исключает из яруса порядка i 2^{i-s} возможных узлов). Так как максимальное количество возможных концевых узлов на этом уровне равно 2^i , то $2^i - \sum_{s=1}^{i-1} \alpha_s 2^{i-s}$ есть количество свободных узлов на ярусе i . Из (2.4) следует, что количество α_i узлов на ярусе i , которые должны быть добавлены, не превосходит количества свободных узлов. Следовательно, к дереву с α_s концевыми узлами порядка s , $s = 1, \dots, i-1$, могут быть добавлены α_i концевых узлов порядка i . \square

Перейдем к доказательству теоремы Шеннона.

Докажем два вспомогательных утверждения, относящихся к побуквенному кодированию произвольных источников. Пусть $\{X, p(x)\}$, $X = (x_1, \dots, x_M)$ – произвольный дискретный источник и пусть $\bar{m}(X) = \sum_{i=1}^M m_i p(x_i)$ – средняя длина 2-ичного кода, слова которого длиной m_i сопоставляются элементарным сообщениям x_i .

Теорема 2.4. Для любого неравномерного кода со свойством однозначного декодирования верно неравенство

$$\bar{m}(X) \geq H(X).$$

Доказательство. Среднюю длину кодовых слов $\bar{m}(X)$ представим в виде

$$\bar{m}(X) = \sum_{i=1}^M p(x_i) \log 2^{m_i}.$$

Рассмотрим разность

$$H(X) - \bar{m}(X) = - \sum_{i=1}^M p(x_i) \log p(x_i) + \sum_{i=1}^M p(x_i) \log 2^{-m_i} = \sum_{i=1}^M p(x_i) \log \frac{2^{-m_i}}{p(x_i)}.$$

Воспользуемся неравенством $\ln x < x - 1$ при $x > 0$. В результате получим, что

$$H(X) - \bar{m}(X) \leq \log e \sum_{i=1}^M p(x_i) \left(\frac{2^{-m_i}}{p(x_i)} - 1 \right) = \log e \left(\sum_{i=1}^M 2^{-m_i} - 1 \right) \leq 0. \quad (2.5)$$

(см. Теорему (3)). \square

Равенство в (2.5) возможно тогда и только тогда, когда $p(x_i) = 2^{-m_i}$, $i = 1, \dots, M$. Таким образом, если вероятности элементарных сообщений являются целыми отрицательными степенями двойки и $\sum_{i=1}^M 2^{-m_i} = 1$, то для соответствующего 2-ичного неравномерного кода имеет место равенство

$$\bar{m}(X) = H(X).$$

Коды, для которых средняя длина кодовых слов равна наименьшему возможному значению, называются *оптимальными*.

Теорема 2.5. *Существует 2-ичный неравномерный код со свойством однозначного декодирования, для которого*

$$\bar{m}(X) \leq H(X) + 1.$$

Доказательство. Пусть m_i – наименьшее целое число, удовлетворяющее неравенству $m_i' \geq I(x_i)$, где $I(x_i) = -\log p(x_i)$ – собственная информация сообщения x_i , $i = 1, \dots, M$. Ясно, что

$$I(x_i) \leq m_i' \leq I(x_i) + 1.$$

Поскольку

$$\sum_{i=1}^M 2^{-m_i'} \leq \sum_{i=1}^M 2^{-I(x_i)} = \sum_{i=1}^M p(x_i) = 1,$$

то по теореме 3 существует 2-ичное дерево с концевыми вершинами порядков m_1', \dots, m_M' . Соответствующий код будет иметь среднюю длину

$$\bar{m}(X) = \sum_{i=1}^M m_i' p(x_i) \leq H(X) + 1.$$

□

Пусть кодируются сообщения длины n ДИБП и $H(X^n)$ – энтропия источника $\{X^n, p(\bar{x})\}$.

Замечание к теореме 4. Для любого 2-ичного кода, однозначно декодирующего последовательность из X^n , среднее число символов, приходящихся на одно сообщение, удовлетворяет неравенству

$$R = \frac{\bar{m}(X^n)}{n} \geq \frac{H(X^n)}{n}.$$

Замечание к теореме 5. Существует 2-ичный код, однозначно декодирующий последовательности из X^n , для которого верно неравенство

$$R = \frac{\bar{m}(X^n)}{n} \leq \frac{H(X^n)}{n} + \frac{1}{n}.$$

Теорема 2.6. *(Обратная теорема кодирования.) Для любого кода, однозначно кодирующего последовательности сообщений длиной n ДИБП, средняя скорость кодирования R удовлетворяет неравенству $R \geq H(X)$.*

Доказательство. В самом деле, согласно Замечанию к теореме 4 имеем

$$R \geq \frac{H(X^n)}{n} = \frac{nH(X)}{n} = H(X).$$

□

Теорема 2.7. (Прямая теорема кодирования.) Пусть ε – произвольное положительное число. Существует n и однозначно декодируемый 2-ичный код, кодирующий последовательности сообщений длиной n ДИБП $\{X, p(x)\}$, для которого верно неравенство

$$R < H(X) + \varepsilon.$$

Доказательство. Так как $H(X^n) = nH(X)$, согласно Замечанию к теореме 5 существует однозначно декодируемый 2-ичный код скоростью кодирования R , не превосходящим $H(X) + 1/n$. Выберем n_0 так, чтобы $1/n_0 \leq \varepsilon$, тогда для всех $n \geq n_0$ будем иметь $R < H(X) + \varepsilon$. \square

2.4 Оптимальные неравномерные коды

Оптимальным называется код, средняя длина кодовых слов которого равна минимально возможной. В простейшем случае, когда вероятности элементарных сообщений источника $\{X, p(x)\}$, $X = \{x_1, \dots, x_M\}$ являются целыми отрицательными степенями двойки

$$p(x_i) = 2^{-m_i}, \quad i = 1, \dots, M$$

любой 2-ичный код со свойством однозначно декодируемости является оптимальным, так как средняя длина кодовых слов равна

$$\bar{m}(X) = H(X)$$

(см. Теорему 4). В таком коде сообщению x_i ставится в соответствие слово длины m_i . Всякое дерево с набором концевых вершин порядков m_1, \dots, m_M и указанным правилом соответствия дает оптимальный код.

Будем предполагать, что

$$p(x_1) \geq p(x_2) \geq \dots \geq p(x_M).$$

Ограничимся рассмотрением только префиксных кодов.

Лемма 2.1. В оптимальном коде слово, соответствующее наименее вероятному сообщению, имеет наибольшую длину.

Доказательство. Пусть m_i – длина кодового слова сообщения $x_i \in X$, и \bar{m} – средняя длина кодовых слов

$$\bar{m}(X) = \sum_{i=1}^M m_i p(x_i).$$

Предположим, что в оптимальном коде $m_i > m_M$ для некоторого $i < M$. Рассмотрим код, в котором i -е и M -е кодовые слова исходного кода заменены одно другим. Средняя длина \bar{m}' для этого кода удовлетворяет соотношению

$$\begin{aligned} \bar{m}' &= \bar{m} - p(x_i)m_i - p(x_M)m_M + p(x_i)m_M + p(x_M)m_i = \\ &= \bar{m} - (m_i - m_M)(p(x_i) - p(x_M)) < \bar{m}, \end{aligned}$$

что противоречит предположению об оптимальности кода. \square

Лемма 2.2. В оптимальном двоичном префиксном коде два наименее вероятных сообщения кодируются словами одинаковой длины, которые, можно считать, различаются только в последнем знаке, одно из них оканчивается нулем, а другое – единицей.

Доказательство. Обозначим через u_j слово, кодирующее сообщение x_j . Пусть u_M – слово наибольшей длины оптимального кода. Тогда существует по крайней мере еще одно слово, скажем u_i , такой же длины оптимального кода. В противном случае единственное слово наибольшей длины кода может быть укорочено без нарушения декодируемости и, тем самым, получим меньшую среднюю длину кодовых слов. Кодовые слова u_i и u_M должны отличаться в последнем знаке. В противном случае длины кодовых слов можно уменьшить, сохраняя однозначную декодируемость, и получить при этом меньшую среднюю длину кодовых слов. Можем считать, модифицируя кодовое дерево, что $m_M - 1$ первых символов у них совпадают. Покажем теперь, что эти слова кодируют наименее вероятные сообщения. Предположим противное, что $i < M - 1$. Тогда $m_i = m_M > m_{M-1}$. В этом случае среднюю длину кода можно было бы уменьшить, заменив слово u_i на u_{M-1} и u_{M-1} на u_i . Следовательно, это предположение не верно и наибольшую длину имеют слова u_{M-1} и u_M . \square

Рассмотрим новый источник сообщений X' , состоящий из $M - 1$ элементарных сообщений $\{x'_1, \dots, x'_{M-1}\}$ с вероятностями

$$p(x'_i) = \begin{cases} p(x_i), & i = 1, \dots, M - 2; \\ p(x_{M-1}) + p(x_M), & i = M - 1. \end{cases}$$

Любой декодируемый префиксный код для источника X' может превратиться в декодируемый код для источника X приписыванием к кодовому слову, кодирующему сообщение x'_{M-1} , символы 0 и 1 для получения слов, кодирующих сообщения x_{M-1} и x_M .

Лемма 2.3. *Если оптимален однозначно декодируемый префиксный код для источника X' , то оптимален, полученный из него префиксный код для источника X .*

Доказательство. Обозначим через \bar{m}' среднюю длину кодовых слов префиксного не обязательно оптимального кода источника X' . Тогда средняя длина \bar{m} кодовых слов для источника X

$$\begin{aligned} \bar{m} &= \sum_{i=1}^M m_i p(x_i) = \sum_{i=1}^{M-2} m_i p(x_i) + m_{M-1} p(x_{M-1}) + m_M p(x_M) = \\ &= \sum_{i=1}^{M-1} m'_i p(x'_i) - m'_{M-1} [p(x_{M-1}) + p(x_M)] + m_{M-1} p(x_{M-1}) + m_M p(x_M) = \\ &= \bar{m}' + p(x_{M-1}) [m_{M-1} - m'_{M-1}] + p(x_M) [m_M - m'_{M-1}] = \bar{m}' + p(x'_{M-1}). \end{aligned} \quad (2.6)$$

Здесь учтено, что длины m'_i , $i = 1, 2, \dots, M - 1$, кодовых слов для источника X' связаны с длинами m_i , $i = 1, \dots, M$ кодовых слов для источника X следующим соотношением

$$\begin{cases} m_i = m'_i, & i = 1, \dots, M - 2; \\ m_M = m_{M-1} = m'_{M-1} + 1. \end{cases}$$

Из (2.6) следует, что \bar{m} и \bar{m}' отличаются на константу $p(x'_{M-1})$, которая не зависит от выбора кодовых слов. Покажем, что, строя декодируемый код для источника X' с минимальным значением \bar{m}' , мы получаем декодируемый код для источника X с минимальным значением \bar{m} .

Обозначим через $\bar{m}_{\text{опт}}$ и $\bar{m}'_{\text{опт}}$ средние длины оптимальных кодов для источников $\{X, p(x)\}$ и $\{X', p(x')\}$. Для оптимального кода источника $\{X', p(x')\}$ имеем

$$\bar{m}_{\text{опт}} \leq \bar{m} = \bar{m}'_{\text{опт}} + p(x'_{M-1}). \quad (2.7)$$

Модифицируем оптимальный код источника $\{X, p(x)\}$, удаляя последние символы 0 и 1 в кодовых словах максимальной длины и объединяя соответствующие им сообщения в одно сообщение, получим код для источника $\{X', p(x')\}$, для которого верно соотношение

$$\bar{m}_{\text{опт}} = \bar{m}' + p(x'_{M-1}) \geq \bar{m}'_{\text{опт}} + p(x'_{M-1}) \quad (2.8)$$

Из (2.7) и (2.8) следует, что

$$\overline{m}_{\text{опт}} = \overline{m}'_{\text{опт}} + p(x'_{M-1}).$$

□

Таким образом, задача построения оптимального префиксного кода сводится к задаче построения оптимального префиксного кода для источника, содержащего на одно сообщение меньше. В этом источнике снова можно выделить два наименее вероятных сообщений и, объединяя их, получить новый источник, содержащий теперь уже на два сообщения меньше, чем исходный. Продолжая эту процедуру, можно прийти до источника, содержащего всего два сообщения, оптимальным кодом для которого являются 0 для одного сообщения и 1 для другого. Описанный метод построения префиксного кода называется методом Хаффмена.

Глава 3

Теоремы кодирования для каналов связи

3.1 Средняя взаимная информация между источниками

Пусть X и Y – два дискретных множества. Рассмотрим источник $\{XY, p(x, y)\}$, алфавит которого состоит из всевозможных пар $(x, y) \in XY$. Задание источника XY определяет также источники $\{X, p(x)\}$ и $\{Y, p(y)\}$, где

$$p(x) = \sum_{y \in Y} p(x, y); \quad p(y) = \sum_{x \in X} p(x, y).$$

Кроме того, для каждого из сообщений $y \in Y$ и $x \in X$, для которых $p(x) \neq 0$ и $p(y) \neq 0$, определены условные распределения вероятностей $p(x|y)$ и $p(y|x)$, а следовательно, и условные источники $\{X, p(x|y)\}$ и $\{Y, p(y|x)\}$.

Пусть

$$I(x) = -\log p(x); \quad I(y) = -\log p(y)$$

собственная информации

$$I(x|y) = -\log p(x|y); \quad I(y|x) = -\log p(y|x).$$

и условная собственная информация сообщений $x \in X$ и $y \in Y$.

Определение 8. Количеством информации о сообщении $y \in Y$, содержащейся в сообщении $x \in X$, называется величина

$$I(y; x) = \log \frac{p(y|x)}{p(y)}.$$

Так как $p(x, y) = p(x|y)p(y) = p(y|x)p(x)$, то

$$\frac{p(x|y)}{p(x)} = \frac{p(x, y)}{p(x)p(y)} = \frac{p(y|x)}{p(y)}.$$

Поэтому количество информации о сообщении $y \in Y$ в сообщении $x \in X$ равно количеству информации о сообщении $x \in X$ в сообщении $y \in Y$, или

$$I(x; y) = I(y; x) = \log \frac{p(x, y)}{p(x)p(y)}.$$

На этом основании $I(x; y)$ называют *количеством взаимной информации* между сообщениями x и y или просто *взаимной информацией* между сообщениями x и y . Отметим, что в отличие от $I(x)$ взаимная информация $I(x; y)$ может принимать и отрицательные значения в случае, если $p(x|y) < p(x)$, то $I(x; y) < 0$.

Взаимная информация между сообщениями обладает следующими свойствами:

1. Если сообщения x и y независимы, то есть $p(x, y) = p(x)p(y)$, то сообщение y не дает никакой информации о сообщении x . В этом случае $I(x; y) = 0$.

2. Если сообщение x влечет сообщение y , то есть $p(y|x) = 1$, тогда $I(y; x) = I(y)$ (количество информации о сообщении y в сообщении x равно собственной информации сообщения y).

Количество взаимной информации будем рассматривать как случайную величину на источнике $\{XY, p(x, y)\}$.

Определение 9. Математическое ожидание случайной величины $I(x; y)$ на источнике $\{XY, p(x, y)\}$ называется *средним количеством взаимной информации* или просто *средней взаимной информацией* между источниками $\{X, p(x)\}$ и $\{Y, p(y)\}$

$$I(X; Y) = \sum_{\substack{x \in X \\ y \in Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}.$$

Теорема 3.1. Средняя взаимная информация между источниками X и Y удовлетворяет соотношению

$$0 \leq I(X; Y) \leq \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \frac{p(y|x)}{\hat{p}(y)},$$

где $\hat{p}(\cdot)$ любое другое распределение вероятностей на множестве сообщений Y . В нижней границе равенство достигается тогда и только тогда, когда источники $\{X, p(x)\}$ и $\{Y, p(y)\}$ статистически независимы.

Доказательство. Нижняя граница находится с помощью неравенства $\ln x \leq x - 1$ следующим образом:

$$\begin{aligned} -I(X; Y) &= \sum_y \sum_x p(y|x)p(x) \log \frac{p(y)}{p(y|x)} = \\ &= \log e \sum_y \sum_x p(y|x)p(x) \ln \frac{p(y)}{p(y|x)} \leq \log e \sum_y \sum_x p(y|x)p(x) \left[\frac{p(y)}{p(y|x)} \right] = \\ &= \log e \left\{ \sum_y \sum_x p(x)p(y) - \sum_y \sum_x p(y|x)p(x) \right\} = 0 \end{aligned}$$

Равенство справедливо тогда и только тогда, когда $p(x, y) = p(x)p(y)$ для всех $x \in X$ и $y \in Y$.

Верхняя граница для $I(X; Y)$ вытекает из соотношения:

$$I(X; Y) = \sum_y p(y) \log \frac{1}{p(y)} - \sum_y \sum_x p(y|x)p(x) \log \frac{1}{p(y|x)} \quad (3.1)$$

и предложения 2.1., в котором утверждается, что

$$\sum_y p(y) \log \frac{1}{p(y)} \leq \sum_y p(y) \log \frac{1}{\hat{p}(y)},$$

причем равенство достигается тогда и только тогда, когда $\hat{p}(y) = p(y)$ для всех $y \in Y$. \square

Рассмотрим теперь источник $\{XYZ, p(x, y, z)\}$, который порождает различные условные и безусловные источники. Так $p(x, y) = \sum_{z \in Z} p(x, y, z)$ является безусловным распределением вероятностей на парах $(x, y) \in XY$, $p(x) = \sum_{y \in Y} p(x, y)$ – безусловное распределение вероятностей на X . Далее,

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)}, \quad p(z) \neq 0$$

– условное распределение вероятностей на XY при заданном фиксированном сообщении $z \in Z$ и

$$p(y|x) = \frac{p(x, y)}{p(x)}, \quad p(x) \neq 0, \quad p(y|xz) = \frac{p(x, y, z)}{p(x, z)}, \quad p(x, z) \neq 0$$

– условные распределения вероятностей на сообщениях $y \in Y$ при фиксированном $x \in X$ и $(x, z) \in XZ$ соответственно.

Введем условную взаимную информацию $I(y; z|x)$ между сообщениями $y \in Y$ и $z \in Z$ при данном сообщении $x \in X$:

$$I(y; z|x) = I(y|x) - I(y|x, z) = \log \frac{p(y|x, z)}{p(y|x)} \quad (3.2)$$

и взаимную информацию между парой сообщений $(y, z) \in YZ$ и сообщением $x \in X$:

$$I((y, z); x) = I(y, z) - I((y, z)|x) \quad (3.3)$$

Воспользовавшись свойством аддитивности собственной информации, получим

$$I(x, y) = -\log p(x, y) = -\log[p(x)p(y|x)] = I(x) + I(y|x)$$

и

$$I((x, y)|z) = -\log p(x, y|z) = -\log \frac{p(x, y, z)}{p(z)} = -\log \frac{p(x, z)}{p(z)} - \log \frac{p(x, y, z)}{p(x, z)} = I(x|z) + I(y|x, z).$$

Отсюда, а также из (3.2) и (3.3) находим

$$\begin{aligned} I((x, y); z) &= I(x, y) - I((x, y)|z) = I(x) + I(y|x) - I(x|z) - I(y|x, z) = I(x; z) + I(y; z|x) \\ \text{или} \\ I((x, y); z) &= I(y; z) + I(x; z|y). \end{aligned} \quad (3.4)$$

Эти соотношения называются *свойством аддитивности собственной взаимной информации*.

Определение 10. Математическое ожидание случайной величины $I(x; y|z)$ источника $\{XYZ, p(x, y, z)\}$ называется *средней взаимной информацией* между источниками X и Y относительно источника Z и обозначается через $I(X; Y|Z)$

$$I(X; Y|Z) = MI(x; y|z) = \sum_{x, y, z} p(x, y, z) \log \frac{p(x|y, z)}{p(x|z)} = \sum_z \sum_{x, y} p(x, y|z)p(z) \log \frac{p(x|yz)}{p(x|z)}.$$

Для источника $\{XYZ, p(x, y, z)\}$ определено также количество взаимной информации $I((x, y); z)$ между парой сообщений $(x, y) \in XY$ и сообщением $z \in Z$.

Определение 11. Математическое ожидание случайной величины $I((x, y); z)$ на источнике XYZ представляет собой *среднюю взаимную информацию* между источником XY и источником Z

$$I(XY; Z) = \sum_{x, y, z} p(x, y, z) \log \frac{p((x, y)|z)}{p(x, y)}.$$

Из свойства аддитивности (3.4) следует, что

$$I(XY; Z) = I(X; Z) + I(Y; Z|X) = I(Y; Z) + I(X; Z|Y),$$

а также из свойства (3.3) следует, что

$$I(XY; Z) = H(XY) - H(XY|Z) = H(Z) - H(Z|XY).$$

Одно из важнейших свойств средней взаимной информации состоит в том, что она не увеличивается при преобразовании.

Пусть $\varphi(\cdot)$ некоторое преобразование, отображающее множество X на другое множество, скажем Z , то есть $Z = \varphi(X)$. Предположим также, что задан источник $\{XY, p(x, y)\}$ и, тем самым,

определена средней взаимной информации $I(X; Y)$. Преобразование $\varphi(\cdot)$ определяет источник $\{ZY, p(z, y)\}$, для которого

$$p(z, y) = \sum_{x: \varphi(x)=z} p(x, y).$$

Поэтому средняя взаимная информация $I(Z; Y)$ определена для каждого отображения $\varphi(\cdot)$ и принимает значения, определяемое выбором $\varphi(\cdot)$.

Теорема 3.2. Для любого отображения $Z = \varphi(X)$ источника X в источник Z

$$I(X; Y) \geq I(Z; Y),$$

причем равенство имеет место всегда, когда отображение обратимо, то есть каждому элементу $z \in Z$ соответствует единственный элемент $x \in X$.

Доказательство. Рассмотрим множество XYZ . Так как при выбранном сообщении $x \in X$ сообщение $z \in Z$ однозначно определено и, следовательно, не зависит от сообщения $y \in Y$, то

$$p(z|x, y) = p(z|x) \quad (3.5)$$

или $p(x, y, z) = p(x, y)p(z|x)$ для всех $(x, y, z) \in XYZ$. При заданном $x \in X$ с вероятностью 1 имеем $z = \varphi(x)$, то есть

$$p(z|x) = \begin{cases} 1, & \text{если } z = \varphi(x), \\ 0, & \text{если } z \neq \varphi(x). \end{cases}$$

Из условия (3.5) следует, что

$$I(z; y|x) = \log \frac{p(z|x, y)}{p(z|x)} = 0$$

для всех $(x, y, z) \in XYZ$, для которых $p(x, y, z) \neq 0$, а следовательно, $I(Z; Y|X) = 0$. Отсюда следует, что

$$I(XZ; Y) = I(X; Y) + I(Z; Y|X) = I(X; Y).$$

С другой стороны, в силу неотрицательности средней взаимной информации $I(X; Y|Z)$ имеем

$$I(XZ; Y) = I(Z; Y) + I(X; Y|Z) \geq I(Z; Y).$$

Поэтому

$$I(X; Y) \geq I(Z; Y).$$

Равенство здесь имеет место только в том случае, когда $I(X; Y|Z) = 0$. Ясно, что последнее равенство выполняется, если для всех $(x, y, z) \in XYZ$ выполняется соотношение

$$p(x|yz) = p(x|z).$$

Это условие всегда выполняется, если сообщение z однозначно определяет сообщение x , то есть если сообщения x и z однозначно определяют друг друга. Иначе, если отображение $\varphi(\cdot)$ обратимо. \square

Свойство невозрастания средней взаимной информации можно трактовать следующим образом. Пусть X – множество возможных сигналов на выходе некоторого канала связи, а Y – множество различных передаваемых сообщений. Теорема утверждает, что никакая обработка наблюдений, при

которой происходит их преобразование, не может увеличить информацию об интересующем нас сообщении.

Очевидно, что теорема остается в силе в том случае, когда преобразование осуществляется над источником Y , а также в том случае, когда осуществляется преобразование как источника X , так и источника Y . Пусть $U = \varphi(X)$ и $V = \psi(Y)$ – два отображения, заданные на X и Y соответственно. Тогда

$$I(X; Y) \geq I(U; V).$$

Если оба отображения обратимы, то имеет место равенство.

Пример 1. Пусть $X = \{0, 1\}$ – множество сообщений на входе канала, $Y = \{0, 1\}$ – множество сообщений на выходе канала и переходы входных сообщений в выходные задаются с помощью графа переходов. Вероятности $p(0|1) = p(1|0)$ неправильных переходов будем считать одинаковыми и равными p . Описанный канал называется *двоичным симметричным каналом* (ДСК).

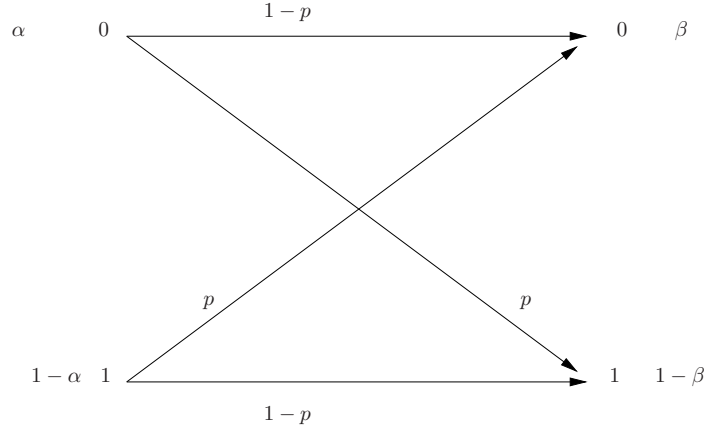


Рис. 3.1:

Вероятности входных сообщений на графе обозначены через α и $1 - \alpha$, а выходных сообщений – через β и $1 - \beta$. По формуле полной вероятности находим

$$\beta = (1 - p)\alpha + p(1 - \alpha).$$

Среднюю взаимную информацию $I(X; Y)$ будем рассматривать как функцию двух параметров α и p и записывать ее как $I(\alpha, p)$.

Имеем

$$I(\alpha, p) = H(Y) - H(Y|X),$$

где

$$H(Y) = -\beta \log \beta - (1 - \beta) \log(1 - \beta) = h(\beta)$$

и

$$H(Y|X) = \sum_i p(x_i) H(Y|x_i) = H(Y|x_1) = -p \log p - (1 - p) \log(1 - p) = h(p),$$

так как $H(Y|x_i)$ не зависит от x_i . Таким образом, имеем

$$I(\alpha, p) = h(\beta) - h(p),$$

где β определяется через α .

Предположим, что p фиксировано и будем рассматривать $I(\alpha, p)$ как функцию параметра α . Поскольку $h(\beta)$ выпуклая вверх функция, а β – линейная функция от α , то $I(\alpha, p)$ – выпуклая вверх функция.

Пусть теперь зафиксирован параметр α , а p будем изменять. При $\alpha = 1/2$ имеем $\beta = 1/2$ и $I(1/2, p) = 1 - h(p)$. Поэтому $I(1/2, p)$ – выпуклая вниз функция. Если $\alpha \neq 1/2$, то

$$\frac{d^2}{dp^2} I(\alpha, p) = \frac{1}{p(1-p)} > 0,$$

и значит, $I(\alpha, p)$ – выпуклая вниз функция параметра p .

3.2 Постановка задачи кодирования в дискретном канале

В системах связи пару "источник – кодер источника" можно рассматривать как новый источник дискретных сообщений и пару "декодер – получатель" можно рассматривать в качестве получателя сообщений. Сообщения источника на входе канала должны быть представлены в форме сигнала, то есть кодированы, а на выходе канала – декодированы.

Для теории информации физическая природа сигналов и шумов является несущественной. Поэтому так же как при кодировании источников, будем рассматривать сигналы на входе и выходе канала как элементы некоторых абстрактных множеств.

Определение 12. Канал называется *дискретным по входу (выходу)*, если множество входных выходных сигналов конечно.

Канал называется каналом с *дискретным временем*, если сигналы на входе и выходе представляют собой конечные или бесконечные последовательности элементов алфавита X на входе канала и алфавита Y на выходе канала.

Дискретный по входу и выходу канал с дискретным временем будем называть *дискретным каналом*.

Наличие шума может привести к тому, что один и тот же входной сигнал канала может перейти в различные выходные сигналы. Такие переходы в теории информации описываются с помощью условных распределений вероятностей. В случае дискретного канала трансформация входных сигналов в выходные задаются условными вероятностями $p(y|x)$. $x \in X$, $y \in Y$, получения на выходе сигнала y , если на вход был послан сигнал x .

В дальнейшем X и Y будем рассматривать как множество сигналов на входе и выходе канала, которые появляются в некоторый фиксированный момент времени. Поэтому условные вероятности $\{p(y|x)\}$ описывают процесс передачи одного сигнала. Однако по каналу, как правило, передается достаточно длинная последовательность сигналов.

Определение 13. Будем говорить, что дискретный канал задан, если для любого целого n и любых последовательностей $(x^{(1)}, \dots, x^{(n)})$ и $(y^{(1)}, \dots, y^{(n)})$ из элементов X и Y соответственно заданы условные (переходные) вероятности $p(y^{(1)}, \dots, y^{(n)} | x^{(1)}, \dots, x^{(n)})$ получения на выходе канала последовательности $(y^{(1)}, \dots, y^{(n)})$, если на входе была последовательность $(x^{(1)}, \dots, x^{(n)})$.

Определение 14. Дискретный канал называется *каналом без памяти* (ДКБП), если для любого n и любых последовательностей $(x^{(1)}, \dots, x^{(n)}) \in X^n$ и $(y^{(1)}, \dots, y^{(n)}) \in Y^n$ имеет место равенство

$$p(y^{(1)}, \dots, y^{(n)} | x^{(1)}, \dots, x^{(n)}) = \prod_{i=1}^n p(y^{(i)} | x^{(i)}).$$

Дискретный канал без памяти (ДКБП) будем обозначать через $\{XY, p(y|x)\}$, где X и Y – входные и выходные алфавиты, а $p(y|x)$, $x \in X$, $y \in Y$, – переходные вероятности канала.

Если задано некоторое входное распределение вероятностей, скажем, $p(\bar{x})$, то оно вместе с условными распределениями $p(\bar{y}|\bar{x})$ задает совместное распределение вероятностей на парах $(\bar{x}, \bar{y}) \in X^n Y^n$

$$p(\bar{x}, \bar{y}) = p(\bar{y}|\bar{x})p(\bar{x})$$

и распределения вероятностей выходных последовательностей канала

$$p(\bar{y}) = \sum_{\bar{x} \in X^n} p(\bar{x})p(\bar{y}|\bar{x}).$$

Пример 2. Пусть имеется двоичный симметричный канал (ДСК) без памяти, $X = \{0, 1\}$, $Y = \{0, 1\}$ и пусть $p(0|1) = p(1|0) = p$ – вероятность передачи сигнала с ошибкой. Если \bar{x} , \bar{y} – последовательность длины n из нулей и единиц на выходе и входе канала, то

$$p(\bar{y}|\bar{x}) = p^t (1-p)^{n-t},$$

где t количество позиций, в которых последовательности \bar{x} , \bar{y} различаются, то есть t – количество ошибок при передаче \bar{x} и получении \bar{y} . Предположим, что $p < 0.5$ и требуется передать одно из двух сообщений z_1 и z_2 . Если закодировать сообщения как $z_1 \rightarrow 0$ и $z_2 \rightarrow 1$, то вероятность неправильного приема сообщения равнялась бы p .

Рассмотрим другой способ кодирования (передачу с помощью повторений): если надо передать z_1 , то по каналу передается последовательность из n нулей, если же надо передать z_2 , то по каналу передается последовательность из n единиц. Приемник работает по следующему правилу: если в принятой последовательности количество нулей больше количества единиц, то считается, что передано z_1 , в противном случае считается, что передавалось z_2 .

Ясно, что ошибка декодирования возникает всякий раз, когда при передаче последовательности длины n число ошибок t превосходит или равно $n/2$. Поэтому вероятность неправильного приема сообщения P_{en} определяется следующим образом:

$$P_{en} = P\{t \geq n/2\} = \sum_{i \geq n/2} C_n^i p^i (1-p)^{n-i}.$$

Так как

$$\left\{ \frac{\mu_n}{n} > \frac{1}{2} \right\} = \left\{ \frac{\mu_n}{n} - p > \frac{1}{2} - p \right\} \subset \left\{ \left| \frac{\mu_n}{n} - p \right| > \frac{1}{2} - p > 0 \right\},$$

то при возрастании n по теореме Бернулли вероятность ошибки P_{en} стремится к нулю:

$$P_{en} \leq P\left\{ \left| \frac{\mu_n}{n} - p \right| > \frac{1}{2} - p \right\} \rightarrow 0 \quad \text{при} \quad n \rightarrow \infty.$$

Таким образом, вероятность неправильной передачи сообщений по каналу может быть сделана сколь угодно малой, если это сообщение передавать достаточно большое количество раз. Время передачи при таком методе кодирования пропорционально числу повторений. При этом скорость передачи, то есть количество информации, передаваемое в единицу времени, будет стремиться к нулю, так как за все время передачи будет передано одно из двух сообщений или не более одного бита информации.

Мы покажем, что произвольно малая вероятность ошибки может быть достигнута и при скоростях передачи, отличных от нуля, за счет усложнения методов кодирования и декодирования.

Определение 15. Кодом длины n и объемом M для канала называется множество из M пар $\{\bar{u}_1, A_1; \bar{u}_2, A_2; \dots; \bar{u}_M, A_M\}$, где $\bar{u}_i \in X^n$, $i = 1, \dots, M$ – последовательности длины n , образованные входными сигналами канала и называемые кодовыми словами ($\bar{u}_i \neq \bar{u}_j$ при $i \neq j$), и $A_i \subset Y^n$, $i = 1, \dots, M$, – решающие области, образованные выходными последовательностями канала, причем при $i \neq j$ множества A_i и A_j не пересекаются.

Если задан код, то задано как множество кодовых слов, так и правило, по которому приемник принимает решение о переданном кодовом слове: если на выходе канала появляется последовательность \bar{y} и $\bar{y} \in A_i$, то приемник принимает решение о том, что передавалось слово \bar{u}_i .

Определение 16. *Скоростью кода (или скоростью передачи) называется величина*

$$R = \frac{1}{n} \log M \frac{\text{бит}}{\text{симв}},$$

где M – объем кода и n – длина кода.

Скорость кода R представляет собой максимальное количество информации, которое может быть передано с помощью одного сигнала (или символа). Такое количество информации передается по каналу, если кодовые слова имеют одинаковую вероятность появления. Скорость кода измеряется в битах на символ.

Отметим различие в определениях скорости кода канала и скорость равномерного кода источника. В случае кода источника скорость определяется как отношение логарифма числа кодовых слов к длине отрезков кодируемых сообщений. В случае кода канала скорость определяется как отношение того же числа к длине кодовых слов. Код длины n со скоростью R имеет объем $M = 2^{nR}$. Такой код будем обозначать через $G(n, R)$.

Пример 3. Предположим, что двоичный источник без памяти имеет энтропию $H(X) < 1$. Как известно, при кодировании сообщений такого источника можно достичь скорости близкой к $H(X)$. Это означает, что при появлении на входе кодера источника n двоичных символов, где n достаточно велико, на выходе кодера появляется примерно $nH(X)$ двоичных символов, что меньше, чем n . Если теперь рассматривать последовательности длины $nH(X)$ как входные сообщения для кодера двоичного канала, осуществляющего кодирование со скоростью $R < 1$, то длина кодовых слов будет равна $n \frac{H(X)}{R}$, что больше, чем $nH(X)$. Таким образом, кодирование источника понижает длину последовательностей сообщений, а кодирование в канале её увеличивает, то есть кодирование источника устраняет избыточность, а кодирование в канале вводит избыточность. Последовательное применение этих двух операций в большинстве случаев увеличивает эффективность по сравнению с передачей сообщений без кодирования.

Ошибка декодирования слова \bar{u}_i возникает, когда последовательность на выходе канала не принадлежит решающей области A_i . Через λ_i обозначим ошибку в декодировании слова \bar{u}_i

$$\lambda_i = \sum_{\bar{y} \in \bar{A}_i} p(\bar{y} | \bar{u}_i).$$

Мерой надежности канала является *средняя вероятность ошибки*

$$\lambda = \sum_{i=1}^M \lambda_i p(\bar{u}_i),$$

где $p(\bar{u}_i)$ – вероятность передачи i -го кодового слова. Так как распределение вероятностей $p(\bar{u}_i)$, $i = 1, \dots, n$, характеризует источник сообщений и никак не связано ни с каналом, ни с кодом, то под средней вероятностью ошибки декодирования будем иметь ввиду

$$\lambda = \frac{1}{M} \sum_{i=1}^M \lambda_i.$$

В случае оптимального кодирования источника, когда $p(\bar{u}_i) = 1/M$, $i = 1, \dots, M$, оба определения средней вероятности ошибки декодирования совпадают.

В качестве другой количественной меры надежности передачи с помощью кода $G(n, R)$ используется максимальная вероятность ошибки.

$$\Lambda = \max\{\lambda_1, \dots, \lambda_M\}.$$

Определение 17. *Пропускной способностью дискретного канала называется максимальное число C такое, что для любого сколь угодно малого δ , $\delta > 0$, и любого R , $R < C$, существует код $G(n, R)$ такой, что средняя вероятность ошибки удовлетворяет неравенству*

$$\lambda < \delta \tag{3.6}$$

Так как C является верхней гранью скоростей кодов, для которых выполняется неравенство (3.6), значит, для любого R , $R > C$, существует $\delta' > 0$ такое, что $\lambda > \delta'$ для любого n и любого кода $G(n, R)$.

3.3 Информационная емкость дискретных каналов без памяти

Пусть $p(y|x)$, $x \in X$, $y \in Y$, – переходные вероятности задающие дискретный канал без памяти. По определению такого канала

$$p(\bar{y}|\bar{x}) = \prod_{i=1}^n p(y^{(i)}|x^{(i)})$$

для любых последовательностей $\bar{x} \in X^n$ и $\bar{y} \in Y^n$, $\bar{x} = (x^{(1)}, \dots, x^{(n)})$, $\bar{y} = (y^{(1)}, \dots, y^{(n)})$.

Средняя взаимная информация между последовательностями на входе и выходе канала имеет вид

$$I(X^n; Y^n) = \sum_{\bar{x}} \sum_{\bar{y}} p(\bar{y}|\bar{x}) p(\bar{x}) \log[p(\bar{y}|\bar{x})/p(\bar{y})],$$

где $p(\bar{y}) = \sum_{\bar{x} \in X^n} p(\bar{y}|\bar{x}) p(\bar{x})$.

Для любого распределения вероятностей $p(\bar{x})$, $\bar{x} \in X^n$, на входе канала введем распределения вероятностей по каждой компоненте последовательности \bar{x} , полагая

$$p_i(x^{(i)}) = \sum_{x^{(1)}} \cdots \sum_{\substack{x^{(j)} \\ j \neq i}} \cdots \sum_{x^{(n)}} p(\bar{x}).$$

Распределения вероятностей $p_i(x)$, $x \in X$, порождают источники $X_i = \{X, p_i(x)\}$ и $Y_i = \{Y, p_i(y)\}$ на входе и выходе канала, где $p_i(y) = \sum_{x \in X} p(y|x) p_i(x)$.

Поскольку рассматривается канал без памяти средняя взаимная информация между источниками X_i и Y_i запишется в виде

$$I(X_i; Y_i) = \sum_y \sum_x p(y|x) p_i(x) \log[p(y|x)/p_i(y)].$$

Определение 18. Информационная емкость C^* дискретного канала без памяти определяется соотношением

$$C^* = \max_{\{p(x)\}} I(X, Y),$$

где максимум берется по всем входным распределениям вероятностей $p(x)$ на X .

Лемма 3.1. Для произвольного входного распределения $p(\bar{x})$, $\bar{x} \in X^n$ выполняется неравенство

$$I(X^n; Y^n) \leq \sum_{i=1}^n I(X_i; Y_i) \leq nC^*.$$

Нижняя граница в равенстве достигается тогда, когда источники X_i , $i = 1, \dots, n$, статистически независимы, а равенство в верхней границе достигается тогда и только тогда, когда для распределений вероятностей по отдельным составляющим на входе канала достигается информационная емкость канала.

Доказательство. Пусть $p(\bar{x})$ – произвольное распределение вероятностей на входе дискретного канала без памяти. Имеем

$$\begin{aligned}
I(X^n; Y^n) &= H(Y^n) - H(Y^n | X^n) = \\
&= H(Y^n) + \sum_{\bar{x} \in X^n} \sum_{\bar{y} \in Y^n} p(\bar{x}) p(\bar{y} | \bar{x}) \log \prod_{i=1}^n p(y^{(i)} | x^{(i)}) = \\
&= H(Y^n) + \sum_{i=1}^n \sum_{\bar{x} \in X^n} \sum_{\bar{y} \in Y^n} p(\bar{x}) p(\bar{y} | \bar{x}) \log p(y^{(i)} | x^{(i)}) = \\
&= H(Y^n) + \sum_{i=1}^n \sum_{x^{(i)} \in X_i} \sum_{y^{(i)} \in Y_i} p_i(x^{(i)}) p(y^{(i)} | x^{(i)}) \log p(y^{(i)} | x^{(i)}) = \\
&= H(Y^n) - \sum_{i=1}^n H(Y_i | X_i) \leq \sum_{i=1}^n H(Y_i) - \sum_{i=1}^n H(Y_i | X_i) = \sum_{i=1}^n I(X_i; Y_i). \quad (3.7)
\end{aligned}$$

В неравенстве (3.7) имеет место знак равенства, если источники Y_1, \dots, Y_n статистически независимы, то есть если

$$p(\bar{y}) = \prod_{i=1}^n p_i(y^{(i)})$$

для всех $\bar{y} \in Y^n$. Для дискретного канала без памяти это выполняется, если выбрать

$$p(\bar{x}) = \prod_{i=1}^n p_i(x^{(i)}).$$

Действительно, в этом случае

$$\begin{aligned}
p(\bar{y}) &= \sum_{\bar{x} \in X^n} p(\bar{x}) p(\bar{y} | \bar{x}) = \sum_{x^{(1)} \in X_1} \dots \sum_{x^{(n)} \in X_n} \prod_{i=1}^n p_i(x^{(i)}) p(y^{(i)} | x^{(i)}) = \\
&= \prod_{i=1}^n \sum_{x^{(i)} \in X_i} p_i(x^{(i)}) p(y^{(i)} | x^{(i)}) = \prod_{i=1}^n p_i(y^{(i)}).
\end{aligned}$$

Далее, для произвольного входного распределения $p(\bar{x})$, $\bar{x} \in X^n$, имеем

$$I(X^n; Y^n) \leq \sum_{i=1}^n \max_{\{p_i(x^{(i)})\}} I(X_i; Y_i) = n \max_{\{p(x)\}} I(X; Y) = n C^*.$$

Если $p(x)$, $x \in X$, – распределение вероятностей на входе канала такое, что на нём достигается информационная емкость канала, тогда для распределения

$$p(\bar{x}) = \prod_{i=1}^n p(x^{(i)})$$

выполняется соотношение

$$I(X^n; Y^n) = \sum_{i=1}^n I(X_i; Y_i) = n \max_{\{p(x)\}} I(X; Y) = n C^*.$$

Лемма доказана. □

Пример 4. (Информационная емкость двоичного симметричного канала.) В обозначениях примера 1 имеем, что $I(X; Y) = I(\alpha, p) = h(\beta) - h(p)$. Функция $h(\beta)$ достигает максимума, равного 1, при $\beta = 1/2$. Для ДСК распределение на выходе канала будет равномерным, если равномерно распределение на входе канала. Поэтому

$$C^* = 1 - h(p).$$

Пример 5. (Информационная емкость двоичного симметричного канала со стиранием.) Пусть $X = \{0, 1\}$ — множество элементарных сообщений на входе канала и $Y = \{0, 1, *\}$ — сигналы на выходе канала. Если полученный сигнал не поддается расшифровке, то его лучше стереть. В симметричном канале вероятность стирания символов 0 и 1 одинакова и равна q . Если стирания не произошло, то оба сигнала 0 и 1 с одинаковой вероятностью $1 - p - q$ будут правильно расшифрованы, а с вероятностью p будет иметь место ошибка. Так как

$$H(Y|0) = H(Y|1) = H(Y; X) = -(1 - p - q) \log(1 - p - q) - p \log p - q \log q,$$

средняя взаимная информация между источниками X и Y равна

$$I(X; Y) = H(Y) + (1 - p - q) \log(1 - p - q) + p \log p + q \log q.$$

Пусть $\{\alpha, 1 - \alpha\}$ распределение вероятностей на входе канала. Тогда по формуле полной вероятности находим, что на выходе канала

$$\beta_1 = p(0) = \alpha(1 - p - q) + (1 - \alpha)p \text{ и } \beta_2 = p(1) = (1 - \alpha)(1 - p - q) + \alpha p.$$

Заметим, что $\beta_1 + \beta_2 = 1 - q$.

Энтропия $H(Y) = -\beta_1 \log \beta_1 - \beta_2 \log \beta_2 - q \log q$ максимальна, если максимально выражение $-\beta_1 \log \beta_1 - \beta_2 \log \beta_2$, где $\beta_1 + \beta_2 = 1 - q$. Легко проверить, что максимум этого выражения достигается при $\beta_1 = \beta_2 = (1 - q)/2$. Можно убедиться, что соотношение $\beta_1 = \beta_2 = (1 - q)/2$ выполняется, если принять $\alpha = 1/2$. Поэтому

$$H_{\max}(Y) = -2 \frac{1 - q}{2} \log \frac{1 - q}{2} - q \log q.$$

Подставляя это выражение в $I(X; Y)$ найдем максимальное значение средней взаимной информации, равной информационной емкости канала со стиранием

$$C^* = -(1 - q) \log \frac{1 - q}{2} + (1 - p - q) \log(1 - p - q) + p \log p \text{ (бит/симв.)}.$$

В случае $p = 0$ информационная емкость канала равна $C^* = 1 - q$ (бит/симв.).

3.4 Методы декодирования

Декодирование по максимуму правдоподобия (МП-декодирование). Рассмотрим некоторый ДК-БП канал $\{X^n Y^n, p(\bar{x}, \bar{y})\}$ и обозначим через $\bar{u}_1, \dots, \bar{u}_M$, $\bar{u}_i \in X^n$, его кодовые слова. Предположим, что набор кодовых слов фиксирован. Укажем решающие области A_1, \dots, A_M , $A_i \subset Y^n$, при которых средняя вероятность ошибки декодирования минимизируется.

Правило декодирования w будем рассматривать как результат отображения источника Y^n в множество кодовых слов. При МП-декодировании заданному $\bar{y} \in Y^n$ ставится в соответствие кодовое слово с индексом j , $w_{\text{МП}}(\bar{y}) = \bar{u}_j$, наименьшим среди чисел $i = \overline{1, M}$, на котором достигается максимум

$$\max_{i \in \overline{1, M}} p(\bar{y} | \bar{u}_i) = p(\bar{y} | \bar{u}_j). \quad (3.8)$$

Для $j = \overline{1, M}$ определим

$$A_j^{\text{МП}} = \{\bar{y} : w_{\text{МП}}(\bar{y}) = \bar{u}_j\},$$

Для любого другого кода канала вероятность ошибки при передаче слова \bar{u}_i равна

$$\lambda_i = \sum_{\bar{y} \in \bar{A}_i} p(\bar{y}|\bar{u}_i).$$

Поэтому вероятность принятия правильного решения

$$1 - \lambda = 1 - \frac{1}{M} \sum_{k=1}^M \lambda_i$$

можно оценить следующим образом

$$\begin{aligned} 1 - \lambda &= \frac{1}{M} \sum_{k=1}^M \sum_{\bar{y} \in A_i} p(\bar{y}|\bar{u}_i) = \frac{1}{M} \sum_{\bar{y} \in \bigcup_{i=1}^M A_i} p(\bar{y}|w(\bar{y})) \leq \\ &\leq \frac{1}{M} \sum_{\bar{y} \in Y^n} p(\bar{y}|w_{\text{МП}}(\bar{y})) = 1 - P_{en}^{\text{МП}}. \end{aligned}$$

Из этого неравенства следует, что МП-декодирование минимизирует вероятность ошибки.

Пример 6. Рассмотрим МП-декодирование в стационарном двоичном симметричном (ДСК) канале. Предположим, что в этом канале вероятность ошибки при передаче одного сигнала $p < \frac{1}{2}$. Пусть $\bar{x} = (x^{(1)}, \dots, x^{(n)})$ и $\bar{y} = (y^{(1)}, \dots, y^{(n)})$. Тогда

$$p(\bar{y}|\bar{x}) = \prod_{i=1}^n p(\bar{y}^{(i)}|\bar{x}^{(i)}) = p^t (1-p)^{n-t},$$

где t – количество позиций, в которых последовательность \bar{x} отличается от последовательности \bar{y} . Та как

$$\frac{p^{t+1}(1-p)^{n-t-1}}{p^t(1-p)^n - 1} = \frac{p}{1-p} < 1,$$

то в случае МП-декодирования последовательность \bar{y} отображается в то слово используемого кода, которому соответствует минимальное значение t .

Количество позиций, в которых последовательность \bar{x} отличается от последовательности \bar{y} , называется расстоянием Хемминга между \bar{x} и \bar{y} . Поэтому МП-декодирование в ДСК канале отображает выходную последовательность канала в такое кодовое слово, которое находится на минимальном расстоянии Хемминга от него, то есть декодирование происходит по минимуму расстояния Хемминга.

Пример 7. Рассмотрим МП-декодирование в стационарном симметричном стирающем канале. Предположим, что в рассматриваемом канале вероятность ошибки равна p и вероятность стирания равна q , причем $q + 2p < 1$. Имеем

$$p(\bar{y}|\bar{x}) = \prod_{i=1}^n p(\bar{y}^{(i)}|\bar{x}^{(i)}) = p^t q^s (1-p-q)^{n-s-t}, \quad (3.9)$$

где s – число стираний в последовательности \bar{y} и t – количество нестертых позиций в которых последовательности \bar{x} и \bar{y} отличаются.

Число s определяется только каналом и не зависит от передаваемого кодового слова. При фиксированном s и при $q + 2p < 1$ правая часть (3.9) убывает с ростом t . Поэтому МП-декодирование в двоичном симметричном стирающем канале отображает выходную последовательность канала в такое кодовое слово, которому соответствует минимальное значение t , то есть МП-декодирование при фиксированном s определяется по минимуму расстояния Хемминга на нестертых позициях.

МП-декодирование со стиранием. При заданном $\bar{y} \in Y^n$ положим $\hat{w}_{\text{МП}}(\bar{y}) = \bar{u}_j$, если существует единственное j , $j = \overline{1, M}$, для которого достигается максимум в (3.8), и откажемся от принятия решения, если $\bar{y} \notin A_j$ для всех $j = \overline{1, M}$.

Для мягкого МП-декодирования обозначим через $\chi_m(\bar{y})$, $m = \overline{1, M}$ характеристическую функцию множества

$$\overline{A_m^{\text{МП}}} = \{\bar{y} : \hat{w}_{\text{МП}}(\bar{y}) \neq \bar{u}_m\}.$$

Можем записать, что

$$\chi_m(\bar{y}) = \begin{cases} 1, & \text{если существует } m' \neq m \text{ такое, что } p(\bar{y}|\bar{u}_{m'}) \geq p(\bar{y}|\bar{u}_m); \\ 0, & \text{если для любого } m' \neq m \text{ выполняется } p(\bar{y}|\bar{u}_{m'}) < p(\bar{y}|\bar{u}_m). \end{cases}$$

В этом случае условная вероятность ошибки примет вид

$$\lambda_m = \sum_{\bar{y}} p(\bar{y}|\bar{u}_m) \chi_m(\bar{y}).$$

Пороговое декодирование. Рассмотрим дискретный канал, задаваемый переходными вероятностями $p(\bar{y}|\bar{x})$, $\bar{x} \in X^n$, $\bar{y} \in Y^n$. Пусть $p(\bar{x})$, $\bar{x} \in X^n$, – некоторое распределение вероятностей на входных последовательностях канала и $I(\bar{x}; \bar{y})$ – взаимная информация между двумя последовательностями $\bar{x} \in X^n$ и $\bar{y} \in Y^n$, имеющая вид

$$I(\bar{x}; \bar{y}) = \log \frac{p(\bar{y}|\bar{x})}{p(\bar{y})},$$

где $p(\bar{y}) = \sum_{\bar{x} \in X^n} p(\bar{y}|\bar{x})p(\bar{x})$. Рассмотрим пороговое декодер, который работает следующим образом. Для принятой последовательности \bar{y} декодер вычисляет статистику

$$\theta(i) = I(\bar{u}_i; \bar{y})$$

для каждого подового слова \bar{u}_i , $i = 1, \dots, M$. Декодер сравнивает все значения статистики $\theta(i)$ с числом Tn , где T – некоторый фиксированный параметр (порог). Если имеется единственное значение $i = j$, для которого $\theta(j) > Tn$, то декодер принимает решение, что передавалось кодовое слово \bar{u}_j . В противном случае декодер производит стирание.

Для данного способа принятия решения при передаче кодового слова \bar{u}_m ошибочное решение (необнаруженная ошибка) принимается тогда и только тогда, когда статистика $\theta(m) \leq Tn$ и существует ровно одно значение $m' \neq m$, для которого $\theta(m') > Tn$. Правильное решение при передаче слова \bar{u}_m принимается тогда и только тогда, когда статистика $\theta(m) > Tn$ и для всех значений $m' \neq m$ статистика $\theta(m') \leq Tn$.

Для рассматриваемого порогового декодирования область декодирования $A_m^{(T)}$, $m = \overline{1, M}$, при использовании порога $T > 0$ задается следующим образом

$$A_m^{(T)} = \{\bar{y} : \theta(m) > Tn, \text{ и для всех } m' \neq m \text{ выполняется неравенство } \theta(m') \leq Tn\}.$$

Пример 8. Пороговое декодирование для ДСК с вероятностью ошибки p , $0 < p < 1/2$. В качестве распределения вероятностей выберем равномерное распределение на входе канала. Тогда распределение вероятностей на выходе канала также будет равномерным. В частности, $p(\bar{y}) = 2^{-n}$. Поэтому

$$I(\bar{x}; \bar{y}) = n(1 + \log(1 - p)) + d(\bar{x}, \bar{y}) \log \frac{p}{1 - p}, \quad 0 < p < 1/2.$$

Выберем порог T и положим

$$T^* = \frac{1 + \log(1 - p) - T}{\log \frac{1-p}{p}},$$

Тогда решающая область будет иметь вид

$$A_m^{(T)} = \{\bar{y} : d(\bar{u}_m, \bar{y}) < T^*n, \text{ для других } m' \neq m \text{ выполняется неравенство } d(\bar{u}_{m'}, \bar{y}) \geq T^*n\}.$$

3.5 Помехоустойчивое кодирование в ДСК

Рассмотрим двоичный симметричный канал (ДСК) без памяти с алфавитом $X = \{0, 1\}$ на входе канала и алфавитом $Y = \{0, 1\}$ на выходе канала и пусть $p(0|1) = p(1|0) = p < 1/2$ — вероятность передачи сигнала с ошибкой.

Определение 19. Код длиной n и объемом M называется избыточным, если $M < 2^n$.

При использовании кода без избыточности появление ошибки в любом из принятых слов остается незамеченным, поскольку изменение символа хотя бы в одном разряде приводит к одному из кодовых слов. Возможность выявления ошибки в принимаемых кодовых словах появляется только в случае избыточности кода. Будем называть используемые кодовые последовательности (блоки) *разрешенными*, а остальные $(N - M)$ блоков — *запрещенными*. Если последовательность на выходе ДСК оказывается запрещенной, то это свидетельствует об ошибке при приеме. Существует ненулевая вероятность, что принятая последовательность является другим разрешенным кодовым словом. При выборе помехоустойчивого кода стремятся к тому, чтобы вероятность такого события была как можно меньше.

Введем код $\{\bar{u}_1, A_1; \dots; \bar{u}_M, A_M\}$ длиной n и объемом M , и функцию принятия решения $w(\bar{y})$, $\bar{y} \in Y^n$.

При декодировании принятые запрещенные кодовые слова преобразуются декодером в разрешенные по определенному правилу: если $\bar{y} \in A_i$, то $w(\bar{y}) = \bar{u}_i$. Выбрав подходящим образом решающие области, при декодировании появится возможность исправить ошибки, допущенные при передаче кодовых слов по каналу с шумом.

Если через \hat{u}_m обозначим кодовое слово \bar{u}_m как элемент Y^n , то при декодировании по максимуму правдоподобия $\hat{u}_m \in A_m$, так как при $p < 1/2$ для всех $t = 1, \dots, n$ выполняется неравенство $p(\hat{u}_m | \bar{u}_m) = (1-p)^n > p^t(1-p)^{n-t}$. Поэтому A_m состоит из последовательности \hat{u}_m и соответствующих кодовому слову запрещенных блоков, которые декодируются в кодовое слово \bar{u}_m . Может случиться так, что $A_m \cap A_{m'} \neq \emptyset$, то есть найдется \bar{y} минимально равноудаленный по Хеммингу от \hat{u}_m и $\hat{u}_{m'}$. Поэтому в некоторых случаях МП-декодер не в состоянии однозначно распознать переданное кодовое слово.

Введем понятие кодового расстояния.

Определение 20. Наименьшее из расстояний между любыми парами используемых кодовых слов кода $G = G(n, R)$ называется *кодовым расстоянием* и обозначается через $d(G)$.

При МП-декодировании в ДСК исправляющая способность кода характеризуется теоремой Хемминга.

Теорема 3.3. Код в ДСК при декодировании по наименьшему расстоянию Хемминга исправляет любые t и менее ошибок в каждом принятом кодовом слове тогда, когда кодовое расстояние $d(G)$ удовлетворяет неравенству $d(G) \geq 2t + 1$.

Доказательство. Если $d(G) \geq 2t + 1$, то для любых кодовых слов \bar{u}_i и \bar{u}_j имеем $d(\hat{u}_i, \hat{u}_j) \geq 2t + 1$. Пусть при передаче некоторого кодового слова \bar{u}_k произошло $r \leq t$ ошибок, в результате чего было принято слово \bar{y} . Тогда $d(\hat{u}_k, \bar{y}) = r \leq t$ и в то же время расстояние до любой другой последовательности \bar{u}_i больше t . Последнее вытекает из неравенства треугольника

$$d(\hat{u}_k, \bar{y}) + d(\hat{u}_i, \bar{y}) \geq d(\hat{u}_k, \hat{u}_i) \geq 2t + 1.$$

Значит для правильного декодирования принятого слова \bar{y} необходимо найти кодовое слово $\bar{u} \in G$, ближайшее в смысле расстояния Хемминга, если число ошибок в принятом слове действительно не превосходит t . \square

Пусть условие $d(G) > 2t$ нарушается и найдутся кодовые слова \bar{u}_i и \bar{u}_j расстояние между которыми $d(\bar{u}_i, \bar{u}_j) = 2t$. Допустим, что было передано слово \bar{u}_i . Заменим в слове \bar{u}_i t разрядов на соответствующие разряды из \bar{u}_j , в которых кодовые слова \bar{u}_i и \bar{u}_j различаются. Получим последовательность \bar{y} , удаленную от \hat{u}_i на расстояние t , $d(\hat{u}_i, \bar{y}) = t$. Тогда $d(\hat{u}_j, \bar{y}) = d(\bar{u}_i, \bar{y}) = t$ и при

декодировании слова \bar{u} может быть также отождествлено с кодовым словом \bar{u}_j . Поэтому в этом случае нельзя определить какое на самом деле слово было передано.

Поскольку вероятность ошибки кратности t в ДСК определяется биномиальным распределением

$$P_{en}(t) = C_n^t p^t (1-p)^{n-t},$$

вероятность ошибки декодирования

$$P_{en} \leq \sum_{t=[d(G)/2]}^n P_{en}(t) = \sum_{t=[d(G)/2]}^n C_n^t p^t (1-p)^{n-t}.$$

Декодирование по минимуму расстояния Хэмминга невозможно, если получено кодовое слово, не совпадающее с посланным кодовым словом. Пусть A_i – число кодовых слов (n, k) кода C веса i , тогда вероятность необнаруженной ошибки P_r равно

$$P_r = \sum_{t=d(G)}^n A_t p^t (1-p)^{n-t}.$$

Пример 9. Пусть код G состоит из четырех слов 00000, 01011, 10110 и 11101, так что каждые два слова отличаются не менее чем в трех разрядах, $d(G) = 3$. Согласно теореме декодер может исправить одиночную ошибку в любом разряде. При декодировании по наименьшему расстоянию Хемминга каждому из 28 неразрешенных блоков нужно поставить в соответствие наиболее близкое кодовое слово.

В таблице под каждым кодовым словом выписываются все возможные блоки, отличающиеся от кодового слова в одном разряде.

00000	01011	10110	11101
10000	11011	00110	01101
01000	00011	11110	10101
00100	01111	10010	11001
00010	01001	10100	11100
00001	01010	10111	11100
.....
10001	11010	00111	01100
11000	10011	01110	00101

Оставшиеся 8 неразрешенных блоков отличаются от каждого кодового слова не менее чем в двух разрядах. Однозначно их в таблице разместить нельзя. Так блок 10011 находится во 2-м столбце таблицы и в 3-м столбце строки: 00101 01110 10011 11000.

При декодировании по наименьшему расстоянию нужно найти столбец, в котором содержится принятый блок и выбрать кодовое слово, находящееся в верхней строке этого столбца.

Поэтому задача помехоустойчивого кодирования состоит в поиске кода, обладающего максимальным кодовым расстоянием $d(G)$ при заданной длине n и числе M кодовых слов.

В общем виде задача помехоустойчивого кодирования решения не имеет. Рассмотренный табличный метод декодирования даже при умеренных n на практике не реализуем. Поэтому основным направлением современной теории кодирования является поиск кодов, для которых кодирование и декодирование осуществляются на основе алгебраических принципов, без перебора. К числу таких кодов относятся линейные коды, в частности, циклические коды.

3.6 Прямая и обратная теорема кодирования для дискретного канала без памяти

Пусть задан дискретный канал, то есть заданы множества входных X и выходных Y сигналов, а также при всех $n = 1, 2, \dots$ заданы условные распределения вероятностей $p(\bar{y}|\bar{x})$, $\bar{y} \in Y^n$, $\bar{x} \in X^n$. Предположим, что для передачи по каналу используется код $G(n, R) = \{\bar{u}_1, A_1; \bar{u}_2, A_2; \dots; \bar{u}_M, A_M\}$ длины n и объема $M = 2^{nR}$, где A_1, \dots, A_M — решающие области. Введем в рассмотрение среднюю вероятность ошибки $\lambda(G)$:

$$\lambda(G) = \frac{1}{M} \sum_{i=1}^M \sum_{\bar{y} \in A_i^c} p(\bar{y}|\bar{u}_i), \quad (3.10)$$

где A_i^c — дополнение A_i .

Обозначим через \bar{p} вероятностный вектор $(p(x_1), \dots, p(x_L))$ и положим

$$E_0(\rho, \bar{p}) = -\log \sum_{y \in Y} \left[\sum_{x \in X} p(x) p(y|x)^{1/(1+\rho)} \right]^{1+\rho}.$$

Функция $E_0(\rho, \bar{p})$ называется *функцией Галлагера*. Введем функцию

$$E(R) = \max_{\rho, \bar{p}} (-\rho R + E_0(\rho, \bar{p})),$$

где максимум разыскивается по всем ρ , $0 \leq \rho \leq 1$ и по всем распределениям $\{p(x)\}$ на X .

Теорема 3.4. (Прямая теорема кодирования.) Для произвольного дискретного канала без памяти $\{XY, p(y|x)\}$ существует код со скоростью R , для которого средняя вероятность ошибки декодирования удовлетворяет неравенству

$$\lambda \leq 2^{-nE(R)}, \quad (3.11)$$

где n — длина кода, а экспонента случайного кодирования зависит только от матрицы переходных вероятностей канала и от скорости кода, причем $E(R) > 0$ при всех R , $0 \leq R \leq C^*$, где C^* — информационная емкость канала.

Теорема 3.5. (Обратная теорема кодирования для каналов без памяти.) Пусть C^* — информационная ёмкость дискретного канала и $R = C^* + \varepsilon$, где ε — произвольное положительное число. Тогда существует положительное число δ , зависящее от R , такое, что для всякого кода $G(n, R)$ средняя вероятность ошибки λ удовлетворяет неравенству

$$\lambda \geq \delta.$$

Следствием двух последних теорем кодирования является теорема кодирования Шеннона для каналов без памяти.

Теорема 3.6. Пусть C — пропускная способность канала без памяти, то есть такое число, что для каждого $R < C$ существует код $G(n, R)$ со средней вероятностью ошибки, меньшей чем заданное наперед произвольное положительное число, и что для любого $R > C$ не существует кода с таким свойством. Пусть C^* — информационная емкость канала, равная

$$C^* = \max_{\bar{p}} I(X; Y).$$

Тогда $C = C^*$.

3.7 Теорема Шеннона для ДСК канала

Рассмотрим двоичный симметричный канал. Пусть $0 < p < 1/2$ есть вероятность неверной передачи символа по каналу связи. Пусть C - двоичный код с M равновероятными кодовыми словами $\bar{u}_1, \dots, \bar{u}_M$ длины n . Пусть λ_i - вероятность неправильного декодирования кодового слова \bar{u}_i . Тогда средняя вероятность ошибки декодирования λ определим как

$$\lambda = \lambda_C(p) = \frac{1}{M} \sum_{i=1}^M \lambda_i,$$

где λ_i зависит от p . Рассмотрим совокупность \mathcal{L} всех двоичных кодов длины n мощности M и определим

$$\lambda^*(M, n, p) = \min_{C \in \mathcal{L}} \{\lambda_C\}.$$

Как нам известно, скорость кода R длины n мощности M определяется как $(\log M)/n$, а информационная емкость двоичного симметричного канала с вероятностью ошибки p равна

$$C^* = C^*(p) = 1 - h(p) = 1 - p \log p - (1 - p) \log(1 - p)$$

Теорема 3.7. Для любой сколь угодно малой величины ε и любого $0 < R < C^*(p)$ существует двоичный код C длины n мощности M и скорости R такой, что средняя вероятность ошибки декодирования $\lambda_C < \varepsilon$.

Иначе говоря, для достаточно больших n существует хороший код длины n со скоростью сколь угодно близкой к пропускной способности канала связи.

Нам потребуется несколько вспомогательных утверждений.

При передаче информации по двоичному симметричному каналу число ошибок в полученном слове является биномиально распределенной случайной величиной ν , принимающей значения $0, 1, \dots, n$ с математическим ожиданием $M\nu = np$ и дисперсией $D\nu = np(1 - p)$. Если в кодовом слове произошло t ошибок, то вероятность получить вектор ошибок e веса t равна $p^t(1 - p)^{n-t}$.

Выберем произвольное $\varepsilon > 0$. Для случайной величины ν введем следующую величину

$$b = \left(\frac{D\nu}{\varepsilon/2} \right)^{1/2}.$$

Согласно неравенству Чебышева, имеем

$$P\{|\nu - M\nu| \geq b\} \leq \frac{D\nu}{b^2} = \frac{\varepsilon}{2}.$$

Отсюда следует

$$P\{\nu > \rho\} \leq \frac{\varepsilon}{2}, \quad (3.12)$$

где

$$\rho = [M\nu + b] = \left[np + \left(\frac{np(1 - p)}{\varepsilon/2} \right)^{1/2} \right].$$

Таким образом, вероятность того, что в результате ν ошибок полученное на приеме слово \bar{y} находится от переданного кодового слова \bar{u} на расстояние большее, чем ρ , мала.

При фиксированном $\varepsilon > 0$ для достаточно больших n величина ρ не превосходит $n/2$, поскольку $p < 1/2$.

Рассмотрим шар радиуса $[pn]$ с центром в некоторой точке $\bar{u} \in F_2^n$:

$$B_{[pn]}(x) = \{\bar{y} \in F_2^n \mid d_H(\bar{u}, \bar{y}) \leq [pn]\}.$$

Оценим объем шара.

Лемма 3.2. Пусть $0 \leq p \leq \frac{1}{2}$. Тогда верна оценка

$$\sum_{i=0}^{[np]} C_n^i \leq 2^{nh(p)}.$$

Доказательство. Имеем

$$\begin{aligned} 1 &= (p + (1-p))^n \geq \sum_{i=0}^{[np]} C_n^i p^i (1-p)^{n-i} \geq \sum_{i=0}^{[np]} C_n^i p^{np} (1-p)^{n-pn} = \\ &= \sum_{i=0}^{[np]} C_n^i 2^{\log[(1-p)^n \left(\frac{p}{1-p}\right)^{np}]} = \sum_{i=0}^{[np]} C_n^i 2^{n \log(1-p) + pn \log\left(\frac{p}{1-p}\right)} = \\ &= \sum_{i=0}^{[np]} C_n^i 2^{n(p \log p + (1-p) \log(1-p))} = 2^{-nh(p)} \sum_{i=0}^{[np]} C_n^i. \end{aligned}$$

Отсюда следует

$$\sum_{i=0}^{[np]} C_n^i \leq 2^{nh(p)}.$$

□

Оценим теперь объём шара радиуса $\rho = [M\nu + b]$ с центром в некоторой точке, используя функцию энтропии $h(p)$.

Лемма 3.3. Пусть $0 \leq p \leq \frac{1}{2}$ и $\rho = [M\nu + b]$, где $b = \left(\frac{D\nu}{\varepsilon/2}\right)^{1/2}$. Тогда

$$\frac{1}{n} \log |B_\rho(\bar{u})| \leq h(p) + O\left(\frac{1}{\sqrt{n}}\right) \text{ при } n \rightarrow \infty.$$

Доказательство. По предыдущей лемме имеем

$$\begin{aligned} \frac{1}{n} \log |B_\rho(\bar{u})| &\leq h\left(\frac{\rho}{n}\right) = -\frac{\rho}{n} \log \frac{\rho}{n} - \left(1 - \frac{\rho}{n}\right) \log \left(1 - \frac{\rho}{n}\right) = \\ &= -\frac{[np+b]}{n} \log \frac{[np+b]}{n} - \left(1 - \frac{[np+b]}{n}\right) \log \left(1 - \frac{[np+b]}{n}\right) = \\ &= -p \log p - (1-p) \log(1-p) + O\left(\frac{b}{n}\right) = h(p) + O\left(\frac{1}{\sqrt{n}}\right) \text{ при } n \rightarrow \infty. \end{aligned}$$

что доказывает лемму.

□

Перейдем к доказательству теоремы Шеннона для кодирования в двоичном симметричном зашумленном канале.

Введем функцию $f(\bar{y}, \bar{x})$. Пусть $\bar{y}, \bar{x} \in F_2^n$, тогда

$$f(\bar{y}, \bar{x}) = \begin{cases} 0, & d(\bar{y}, \bar{x}) > \rho \\ 1, & d(\bar{y}, \bar{x}) \leq \rho. \end{cases} \quad (3.13)$$

Функция $f(\bar{y}, \bar{x})$ – характеристическая функция принадлежности вектора \bar{y} шару $B_\rho(\bar{x})$ с центром в точке \bar{x} .

Доказательство. Выберем сколь угодно малую величину $\varepsilon > 0$. Рассмотрим случайный двоичный код длины n мощности M , то есть выберем случайным образом кодовые слова $\bar{u}_1, \dots, \bar{u}_M$. Декодируем полученный вектор y следующим образом: если существует в точности одно кодовое слово \bar{u}_i такое, что

$$d(\bar{u}_i, \bar{y}) \leq \rho,$$

то \bar{y} декодируем в \bar{u}_i , в противном случае регистрируем ошибку или, если должны декодировать в любом случае, всегда декодируем в \bar{u}_1 .

Пусть λ_i , как и выше, вероятность того, что на выходе декодера получено слово, отличное от переданного слова \bar{u}_i . Для λ_i имеем следующую оценку сверху:

$$\begin{aligned} \lambda_i = \sum_{\bar{y}: d(\bar{u}_i, \bar{y}) > \rho} P(\bar{y}|\bar{u}_i) &\leq \sum_{\bar{y} \in F_2^n} P(\bar{y}|\bar{u}_i) [1 - f(\bar{y}, \bar{u}_i) + \sum_{j \neq i} f(\bar{y}, \bar{u}_j)] = \\ &= \sum_{\bar{y} \in F_2^n} P(\bar{y}|\bar{u}_i) (1 - f(\bar{y}, \bar{u}_i)) + \sum_{\bar{y} \in F_2^n} \sum_{j \neq i} P(\bar{y}|\bar{u}_i) f(\bar{y}, \bar{u}_j), \end{aligned}$$

здесь выражение $[1 - f(\bar{y}, \bar{u}_i) + \sum_{j \neq i} f(\bar{y}, \bar{u}_j)]$ равно нулю тогда и только тогда, когда найдется единственное кодовое слово \bar{u}_i такое, что

$$d(\bar{u}_i, \bar{y}) \leq \rho,$$

в противном случае $[1 - f(\bar{y}, \bar{u}_i) + \sum_{j \neq i} f(\bar{y}, \bar{u}_j)] \geq 1$.

Первая сумма в предыдущем неравенстве равна вероятности того, что полученное на выходе слово не входит на расстоянии большем ρ от переданного кодового слова \bar{u}_i . Согласно неравенству (3.12) вероятность не превышает $\varepsilon/2$. Таким образом,

$$\lambda_C(p) \leq \frac{\varepsilon}{2} + \frac{1}{M} \sum_{i=1}^M \sum_{\bar{y} \in F_2^n} \sum_{j \neq i} P(\bar{y}|\bar{u}_i) f(\bar{y}, \bar{u}_j).$$

Основная идея дальнейшего доказательства состоит в том, что величина $\lambda^*(M, n, p)$ меньше математического ожидания λ_C над ансамблем \mathcal{L} всех возможных кодов C длины n и мощности M взятых случайно. Отсюда имеем

$$\begin{aligned} \lambda^*(M, n, p) &\leq \frac{\varepsilon}{2} + \frac{1}{M} \sum_{i=1}^M \sum_{\bar{y} \in F_2^n} \sum_{j \neq i} M(P(\bar{y}|\bar{u}_j)) M(f(\bar{y}, \bar{u}_j)) = \\ &= \frac{\varepsilon}{2} + \frac{1}{M} \sum_{i=1}^M \sum_{\bar{y} \in F_2^n} \sum_{j \neq i} \frac{|B_\rho|}{2^n} M(P(\bar{y}|\bar{u}_j)) = \\ &= \frac{\varepsilon}{2} + \frac{|B_\rho|}{M \cdot 2^n} \sum_{i=1}^M \sum_{\bar{y} \in F_2^n} \sum_{j=1, j \neq i}^M M(P(\bar{y}|\bar{u}_j)) = \\ &= \frac{\varepsilon}{2} + \frac{|B_\rho|}{M \cdot 2^n} \sum_{i=1}^M \sum_{j=1, j \neq i}^M M \left(\sum_{\bar{y} \in F_2^n} P(\bar{y}|\bar{u}_j) \right) = \\ &= \frac{\varepsilon}{2} + \frac{|B_\rho| \cdot B \cdot (M-1)}{M \cdot 2^n} \leq \frac{\varepsilon}{2} + M \frac{|B_\rho|}{2^n}. \end{aligned}$$

Таким образом, $\lambda^*(M, n, p) - \frac{\varepsilon}{2} \leq M \cdot |B_\rho|/2^n$. Логарифмируя обе части, применяя последнюю лемму и деля на n получаем

$$\frac{1}{n} \log \left(\lambda^*(M, n, p) - \frac{\varepsilon}{2} \right) \leq \frac{1}{n} \log M - (1 - h(p)) + O\left(\frac{1}{\sqrt{n}}\right).$$

Подставляя $M = 2^{\lfloor R \cdot n \rfloor}$ в правую часть (вспомним, что по условию число R сколь угодно близко к пропускной способности канала $C(p) = 1 - h(p)$), получаем

$$\frac{1}{n} \log \left(\lambda^*(M, n, p) - \frac{\varepsilon}{2} \right) < -\beta < 0,$$

где β константа, равная $C(p) - R$. Отсюда $\lambda^*(M, n, p) < \frac{\varepsilon}{2} + 2^{-\beta n}$. Начиная с некоторого n будет выполняться $2^{-\beta n} < \frac{\varepsilon}{2}$, и, следовательно, $\lambda^*(M, n, p) < \varepsilon$. Таким образом,

$$\lambda^*(M, n, p) \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Теорема доказана. □

Глава 3

Конспект лекций по теории кодирования

4.1 Линейные коды

I. Пусть F_q , (q — простое) — конечное поле, $k < n$ и $F_q^k \rightarrow F_q^n$ — инъективное линейное отображение векторного пространства F_q^k в пространство F_q^n . Вектора $\bar{a} \in F_q^k$ будем называть информационными словами (векторами). Символом C , будем обозначать образ пространства F_q^k при этом отображении. C является линейным подпространством в пространстве F_q^n . Будем называть его (n, k) -пространством кодовых слов. Матрицу G из компонент базисных векторов кодового пространства будем называть порождающей матрицей. Фиксируем в F_q^k базис и представим информационный вектор \bar{a} в виде $\bar{a} = a^1, \dots, a^k$. В матричном представлении кодовое отображение запишется в виде

$$C \ni \bar{c} = \bar{a} \cdot G, \quad G - k \times n - \text{матрица}$$

II. В пространстве F_q^n введено скалярное “произведение” $(\bar{a}, \bar{b}) = \sum_{i=1}^n a_i b_i$. Для введенного произведения не выполняется первая аксиома скалярного произведения: ненулевой вектор \bar{a} может быть ортогонален самому себе. Через C^\perp обозначим ортогональное дополнение пространства C в F_q^n . Из компонент базисных векторов в C^\perp составим $(n - k) \times n$ -матрицу H , называемую проверочной матрицей. Для любого вектора $\bar{c} \in C$ имеем $\bar{c} \perp \bar{b} \in C^\perp$. Поэтому $H\bar{c}^t = \bar{0}$. Более того, для любого $\bar{c} = \bar{a}G$ выполняется равенство $HG^t\bar{a}^t = 0$. Поэтому $HG^t = GH^t = \bar{0}$. Порождающая и проверочная матрица линейного кода определяются неоднозначно.

Если G — порождающая матрица (n, k) кода C в F_q^n и H — проверочная матрица, то H является порождающей матрицей двойственного $(n, n-k)$ кода C^\perp с проверочной матрицей G .

III. Перестановка строк и сложение строки с другой строкой в порождающей матрице не изменяют кодового пространства. Перестановка столбцов в порождающей матрице приводит к перестановке компонент кодовых векторов. Коды, полученные фиксированной перестановкой компонент кодовых слов будем называть эквивалентными, а перестановку строк, сложение строк и перестановку столбцов порождающей матрицы будем называть элементарными преобразованиями.

Элементарными преобразованиями порождающая матрица приводится к систематическому виду

$$G = [E_{k \times k} : P_{k \times (n-k)}]$$

Для порождающей матрицы в систематическом виде несложно построить одну из проверочных матриц

$$H = [-P_{(n-k) \times n}^t : I_{(n-k) \times (n-k)}]$$

Если G представлена в систематическом виде, то первые k символов кодового слова являются информационными символами.

Теорема 3.8. *Каждый линейный код эквивалентен систематическому линейному коду.*

IV. Вес Хэмминга $w(\bar{x})$ вектора \bar{x} равен числу ненулевых компонент. Расстояние Хэмминга $d_H(\bar{x}, \bar{y})$ равно числу различающихся компонент векторов \bar{x} и \bar{y} . В линейном пространстве вес Хэмминга и расстояние Хэмминга связаны соотношениями

$$d_H(\bar{x}, \bar{y}) = w(\bar{x} - \bar{y}), \quad w(\bar{x}) = d_H(\bar{x}, \bar{0}).$$

Расстояние Хэмминга является метрикой

V. Предположим, что на вход канала подан информационный вектор \bar{c} и на выходе канала получен вектор \bar{y} . Вектор $\bar{e} = \bar{y} - \bar{c}$ называют вектором ошибки. Пусть $t \in \mathbb{N}$.

Определение. Код C исправляет t ошибок, если для любого вектора $\bar{y} \in F_q^n$ существует не более одного кодового вектора \bar{c} , для которого $d_H(\bar{c}, \bar{y}) \leq t$.

Другими словами, если при передачи по каналу кодового слова \bar{c} допущено t ошибок, то \bar{c} будет ближайшим к полученному слову.

Введем минимальное кодовое расстояние как

$$d_C = \min_{\substack{\bar{u}, \bar{v} \in C \\ \bar{u} \neq \bar{v}}} d_H(\bar{u}, \bar{v}) = \min_{\substack{\bar{c} \in C \\ \bar{c} \neq \bar{0}}} w(\bar{c}).$$

Теорема 3.9. (Хэмминг.) Код с $d_C \geq 2t + 1$ может исправить не менее t ошибок.

VI. Сформулируем лемму о проверочной матрице.

Лемма 3.4. Для того, чтобы линейный код C с проверочной матрицей H имел кодовое расстояние $d_C \geq s + 1$ необходимо и достаточно, чтобы любые s столбцов матрицы H были линейно независимы.

Доказательство. Пусть s столбцов матрицы H линейно зависимы. Тогда найдется вектор \bar{c} такой, что $H\bar{c}^t = \bar{0}$. Это означает, что $\bar{c} \in C$ и $w(\bar{c}) \leq s$. Это означает, что $d_C \leq s$.

Обратно. Если любые s столбцов матрицы H линейно независимы, то не существует вектора $\bar{c} \in C$, $\bar{c} \neq \bar{0}$ такого, что $w(\bar{c}) \leq s$ такого, что $d_C \leq s$. \square

VII. Пусть C есть (n, k) код и F_q^n / C - фактор-пространство. Пространство F_q^n распадается на классы смежности. В каждом классе смежности выберем представителя $\bar{a}^{(i)}$ с минимальным весом, $\bar{a}^{(0)} = \bar{0}$. Тогда

$$F_q^n = (\bar{a}^{(0)} + C) \cup \dots \cup (\bar{a}^{(q^n - k - 1)} + C).$$

Пусть \bar{y} - принятое сообщение, $\bar{y} \in \bar{a}^{(i)} + C$ и $\bar{e} = \bar{y} - \bar{c}$ - вектор ошибок. Значит, $\bar{e} \in \bar{a}^{(i)} + C$. При декодировании по минимуму расстояния Хэмминга вектор ошибок должен иметь минимальный вес, то есть $\bar{e} = \bar{a}^{(i)}$. Элемент минимального веса в смежном классе называется лидером класса. Таким образом, если $\bar{y} \in \bar{a}^{(i)} + C$, то получатель считает вектором ошибок лидера смежного класса и декодирует вектор \bar{y} в кодовое слово $\bar{c}^{(i)} = \bar{y} - \bar{a}^{(i)}$.

VIII. Пусть H - проверочная матрица линейного (n, k) кода C и пусть $\bar{y} \in F_q^n$. Вектор $S(\bar{y}) = H \cdot \bar{y}^t$ длины $(n - k)$ будем называть синдромом вектора \bar{y} .

Теорема 3.10. Если $\bar{y}, \bar{z} \in F_q^n$, то

$$\begin{aligned} (i) \quad S(\bar{y}) &= \bar{0} \text{ if and only if } \bar{y} \in C \\ (ii) \quad S(\bar{y}) &= S(\bar{z}) \text{ if and only if } \bar{y} - \bar{z} \in C \end{aligned}$$

Таким образом, процедура декодирования сводится к построению таблицы лидеров и вычислению соответствующих синдромов.

IX. Границы линейных кодов.

Теорема 3.11. (Синглтон) Пусть C - (n, k) код над F_q . Тогда

$$d_C \leq n - k + 1.$$

Доказательство. Теорема является следствием леммы о проверочной матрицы. \square

Следствие. $d_C = n - k + 1$ тогда и только тогда, когда любые $(n - k)$ столбцов проверочной матрицы линейно независимы.

Теорема 3.12. (Граница Хэмминга) Пусть C – (n, k) код над F_q , исправляющий t ошибок. Тогда

$$\sum_{r=0}^t C_n^r (q-1)^r \leq q^{n-k}.$$

Доказательство. Имеется ровно $C_n^m (q-1)^m$ векторов над F_q длины n и веса n в шаре $B_t(O) = \{\bar{x} \in F_q^n : w(\bar{x}) \leq t\}$. Шары радиуса t с центрами в кодовых словах не пересекаются и каждый содержит

$$\sum_{r=0}^t C_n^r (q-1)^r$$

векторов из F_q^n . Общее количество векторов в этих шарах не превосходит числа векторов в F_q^n , поэтому верно неравенство

$$q^k \sum_{r=0}^t C_n^r (q-1)^r \leq q^n.$$

Отсюда следует утверждение теоремы. □

Определение 21. Код называется совершенным или плотно упакованным, если

$$\sum_{r=0}^t C_n^r (q-1)^r = q^{n-k},$$

то есть имеет место плотная упаковка F_2^n шарами радиуса $[(d_C - 1)/2]$.

Теорема 3.13. (Гилберт – Варшавов) Если

$$q^{n-k} > \sum_{i=0}^{d-2} C_{n-1}^i (q-1)^i,$$

то над F_q можно построить линейный (n, k) код с минимальным расстоянием не меньшим d .

Доказательство. Построим проверочную $(n-k) \times n$ матрицу H такую, что любые $d-1$ столбцов линейно независимы. Будем считать, что построено m столбцов таких, что любые $d-1$ столбцов линейно независимы, $m \leq n-1$. Добавить можно столбец, который не является линейной комбинацией не более $d-2$ столбцов из m построенных. Подсчитаем количество таких линейных комбинаций.

Количество линейных комбинаций из i столбцов с ненулевыми коэффициентами будет равно

$$C_m^i (q-1)^i < C_{n-1}^i (q-1)^i.$$

Поэтому количество столбцов, которые не могут быть добавленными, не превышает

$$\sum_{i=0}^{d-2} C_{n-1}^i (q-1)^i < q^{n-k}$$

(q^{n-k} – общее число столбцов длины $n-k$).

Значит найдется столбец, который не является линейной комбинацией не более чем $d-2$ из уже выбранных столбцов. Его можно взять в качестве $(m+1)$ столбца. □

Теорема 3.14. (Плоткин) Для линейного (n, k) кода C с кодовым расстоянием d_C выполняется неравенство

$$d_C \leq \frac{nq^{k-1}(q-1)}{q^k - 1}.$$

Доказательство. Запишем все кодовые слова построчно в таблице. Через \bar{w} обозначим средний вес всех кодовых слов в $C \setminus \bar{0}$. Ясно, что $d_C \leq \bar{w}$. Общее число кодовых слов равно $|C| = q^k$ и $q^k - 1$ – количество ненулевых кодовых слов. Через C' обозначим множество кодовых слов с нулевой i -ой компонентой. Тогда C' подпространство в F_q^n и $|C'|$ равняется q^{k-1} . Поэтому количество кодовых слов с ненулевой i -ой компонентой равно $|C| - |C'| = q^{k-1}(q - 1)$, значит, вклад i -ой компоненты в суммарный вес кодовых слов равен $q^{k-1}(q - 1)$. Таким образом суммарный вес кодовых слов равен $nq^{k-1}(q - 1)$, а средний вес \bar{w} равен

$$\bar{w} = \frac{nq^{k-1}(q - 1)}{q^k - 1}.$$

Поэтому

$$d_C \leq \bar{w} = \frac{nq^{k-1}(q - 1)}{q^k - 1}.$$

Теорема доказана. \square

Х. Код Хэмминга и его свойства.

I. Определим код над полем F_2 посредством проверочной матрицы, столбцами которой являются все ненулевые векторы длины m . Очевидно, что любые два столбца этой матрицы линейно независимы и найдутся три линейно зависимых столбца, следовательно, по лемме о проверочной матрице кодовое расстояние равно 3 и значит код исправляет одну ошибку. Этот код называется кодом Хэмминга. Длина кодовых слов кода Хэмминга равна $n = 2^m - 1$, длина информационных слов равна $k = n - m = 2^m - m - 1$.

Согласно теореме Хэмминга код Хэмминга является совершенным кодом, исправляющим одну ошибку.

Код Хэмминга допускает простое декодирование. Представим проверочную матрицу кода Хэмминга столбцы которой записаны в лексикографическом порядке

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} = [B(1), B(2), \dots, B(n)],$$

здесь $B(i)$ – двоичное представление числа i . Пусть в канале при передаче вектора \bar{x} произошла одна ошибка в i -й координате и получен вектор $\bar{y} = \bar{x} + \bar{e}_i$. Здесь \bar{e}_i – двоичный вектор длины n с единицей только в i -ой компоненте. Найдем синдром вектора \bar{y} :

$$S(\bar{y}) = H\bar{y}^t = H\bar{x}^t + H\bar{e}_i^t = H\bar{e}_i^t = B(i).$$

Таким образом, вектором ошибки является i -ой столбец проверочной матрицы в лексикографическом виде.

4.2 Циклические коды

Линейный код $C \in F_q^n$ называется циклическим кодом, если из условия $\bar{c} = (c_0, c_1, \dots, c_{n-1}) \in C$ следует $\bar{c}' = (c_1, \dots, c_{n-1}, c_0) \in C$.

Примером является $(7, 4)$ код Хэмминга с проверочной матрицей

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}$$

Обозначим через $F_q[x]$ кольцо всех многочленов от переменной x с коэффициентами из поля F_q . Оно ассоциативно, коммутативно и содержит единицу. В кольце $F_q[x]$ рассмотрим фактор множество $F_q[x]/(x^n - 1)$, состоящее из классов вычетов кольца $F_q[x]$ по модулю многочлена $x^n - 1$. Множество $F_q[x]/(x^n - 1)$ замкнуто относительно операций сложения $(+)$ и умножения \bullet и, значит, является

кольцом, но не является полем. Кольцо $F_q[x]/(x^n - 1)$ изоморфно n -мерному векторному пространству над F_q :

$$c(x) = c_0 + c_1x + c_2x^2 + \dots + c_{n-1}x^{n-1} \longleftrightarrow \bar{c} = (c_0, c_1, c_2, \dots, c_{n-1})$$

Определение 22. Идеалом I кольца $F_q[x]/(x^n - 1)$ называется такое его линейное подпространство, что для любых многочленов $r(x) \in F_q[x]/(x^n - 1)$ и $c(x) \in I$ многочлен $r(x) \bullet c(x)$ принадлежит I .

Теорема 3.15. Подпространство кольца $F_q[x]/(x^n - 1)$ является циклическим кодом тогда и только тогда, когда оно образует идеал.

Существенным моментом в доказательстве теоремы является тот факт, что умножение многочлена на x соответствует циклическому сдвигу вектора в пространстве F_q^n .

$$x \bullet c(x) = x \bullet (c_0 + c_1x + c_2x^2 + \dots + c_{n-1}x^{n-1}) = c_{n-1} + c_0x + c_1x^2 + \dots + c_{n-2}x^{n-1}.$$

II. В циклическом коде C выберем многочлен наименьшей степени. Умножим его на подходящий элемент поля F_q такой, что коэффициент при старшей степени многочлена равнялся 1. Обозначим этот приведенный многочлен через $g(x)$.

Предложение 3.1. Циклический код содержит единственный ненулевой приведенный многочлен наименьшей степени.

Этот ненулевой приведенный многочлен наименьшей степени называется порождающим многочленом кода.

Теорема 3.16. Циклический код состоит из всех многочленов вида

$$f(x) \bullet g(x),$$

где $g(x)$ – порождающий многочлен кода степени r , степень многочлена $f(x)$ меньше $(n-r)$.

Доказательство. Кодовый многочлен поделим на порождающий многочлен с остатком

$$c(x) = q(x)g(x) + s(x), \quad \deg s(x) < \deg g(x).$$

Так как $s(x) \in C(x)$ и $\deg s(x) < r$, то $g(x) = 0$. □

Теорема 3.17. Циклический код длины n с порождающим многочленом $g(x)$ существует тогда и только тогда, когда $g(x)$ делит $x^n - 1$.

Доказательство. Поделим многочлен $x^n - 1$ на порождающий многочлен $g(x)$ с остатком

$$x^n - 1 = h(x)g(x) + s(x).$$

Поэтому $h(x) \bullet g(x) = -s(x)$, $s(x) \in C(x)$ и $\deg s(x) < r$. Отсюда следует, что $s(x) = 0$. □

Таким образом,

$$x^n - 1 = h(x)g(x).$$

Многочлен $h(x)$ называется проверочным многочленом. Основанием служит следующее рассуждение. Для любого кодового многочлена $c(x) \in C(x)$ имеем

$$h(x)c(x) = h(x)g(x)a(x) = (x^n - 1)a(x).$$

Поэтому $c(x) \bullet h(x) = 0$.

III. Систематическое и несистематическое кодирование.

Несистематическое кодирование осуществляется следующим образом.

$$c(x) = i(x)g(x),$$

где $\deg i(x) \leq k - 1 = n - r - 1$. Многочлен $i(x)$ называется информационным многочленом.

При систематическом кодировании для информационного многочлена $i(x)$, $\deg i(x) \leq k - 1$ выберем многочлен $t(x)$ так, чтобы многочлен

$$x^{n-k}i(x) + t(x) = c(x)$$

был кодовым многочленом. Так как остаток от деления $c(x)$ на $g(x)$ должен быть равен нулю, имеем

$$x^{n-k}i(x) \% g(x) + t(x) \% g(x) = 0$$

Отсюда находим

$$t(x) = -x^{n-k}i(x) \% g(x).$$

Процедуры систематического и несистематического кодирования дают одно и то же множество кодовых слов.

IV. Пусть $c(x) \in C(x)$ переданный кодовый многочлен и $v(x) \in F_q[x]/(x^n - 1)$ принятый многочлен. Многочлен $e(x) = v(x) - c(x)$ называется многочленом ошибок.

Многочлен $s(x)$, равный остатку от деления принятого многочлена на порождающий многочлен $s(x) = v(x) \% g(x) = e(x) \% g(x)$, определяется многочленом ошибок. Будем называть его синдромным многочленом.

Теорема 3.18. Если d_C – минимальный вес циклического кода C , то каждому многочлену ошибок веса $d_C/2$ соответствует единственный синдромный многочлен.

Доказательство. Пусть $e_1(x)$ и $e_2(x)$ многочлены ошибок веса меньше $d_C/2$ и пусть $e_1(x) \% g(x) = s(x)$ и $e_2(x) \% g(x) = s(x)$. Тогда $(e_1(x) - e_2(x)) \% g(x) = 0$. Это означает, что $e_1(x) - e_2(x)$ является кодовым многочленом и вес этого многочлена меньше d_C . Что невозможно. \square

V. Порождающая и проверочная матрица циклического кода. Начнем с несистематического кодера.

Пусть $g(x) = g_0 + g_1(x) + \dots + g_{n-k}x^{n-k}$ – порождающий многочлен. Ясно, что многочлены $g(x), xg(x), x^2g(x), \dots, x^{k-1}g(x)$ являются кодовыми линейно независимыми многочленами. Поэтому $(k \times n)$ матрица

$$G = \begin{pmatrix} g_0 & g_1 & \dots & g_{n-k} & 0 & \dots & 0 \\ 0 & g_0 & g_1 & \dots & g_{n-k} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & g_0 & \dots & g_{n-k} \end{pmatrix}$$

является порождающей матрицей циклического кода C .

Пусть $v(x) = a(x)g(x)$, $\deg a(x) \leq k - 1$. Так как $x^n - 1 = h(x)g(x)$, то

$$v(x) \cdot h(x) = a(x) \cdot g(x) \cdot h(x) = a(x)[x^n - 1] = a(x)x^n - a(x).$$

Это означает, что в правой части отсутствуют слагаемые с $x^k, x^{k+1}, x^{k+2}, \dots, x^{n-1}$, то есть

$$\begin{aligned} v_0 h_k &+ v_1 h_{k-1} + \dots + v_k h_0 &= 0 \\ v_1 h_k &+ v_2 h_{k-1} + \dots + v_{k+1} h_0 &= 0 \\ \dots &\dots &\dots \\ v_{n-k-1} h_k &+ v_{n-k} h_{k-1} + \dots + v_{n-1} h_0 &= 0 \end{aligned}$$

В матричной форме соотношения записываются в виде $H \cdot \bar{v}^t = \bar{0}^t$ для любого кодового вектора, где матрица $H = H_{(n-k) \times n}$ имеет вид

$$\begin{pmatrix} h_k & h_{k-1} & \dots & h_0 & 0 & \dots & 0 \\ 0 & h_k & h_{k-1} & \dots & h_0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & h_k & \dots & h_0 \end{pmatrix}$$

Так как строки матрицы H ортогональны кодовым векторам, то матрица H является проверочной матрицей.

Многочлен $x^k h(x^{-1}) = h_k + h_{k-1}x + \dots h_0 x^k$ является порождающим многочленом $(n, n-k)$ кода C^\perp – двойственного кода C .

Построим порождающую матрицу для систематического кодера.

Пусть $i(x)$ – информационный многочлен, $\deg i(x) \leq k-1$. Процедура систематического кодирования сводится к нахождению многочлена $t(x)$ такого, что $c(x) = x^{n-k} \cdot i(x) + t(x)$ является кодовым многочленом. Таким многочленом является

$$t(x) = -x^{n-k} \bullet i(x).$$

Для $i = 1, \dots, k$ представим x^{n-i} в виде

$$x^{n-i} = q_i(x) \cdot g(x) + s_i(x), \quad s_i(x) = \sum_{j=0}^{n-k-1} s_{ji} x^j.$$

Тогда

$$x^{n-i} - s_i(x) = q_i(x) \cdot g(x), \quad i = 1, \dots, k$$

являются кодовыми линейно независимыми многочленами. Поэтому порождающая матрица систематического кодера запишется в виде

$$G = \begin{pmatrix} -s_{0,k} & -s_{1,k} & \dots & -s_{(n-k-1),k} & 1 & 0 & \dots & 0 \\ -s_{0,k-1} & -s_{1,k-1} & \dots & -s_{(n-k-1),k-1} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -s_{0,1} & -s_{1,1} & \dots & -s_{(n-k-1),1} & 0 & 0 & \dots & 1 \end{pmatrix}$$

Тогда проверочная матрица систематического кодера примет вид

$$H = \begin{pmatrix} 1 & 0 & \dots & 0 & s_{0,k} & s_{0,k-1} & \dots & s_{0,1} \\ 0 & 1 & \dots & 0 & s_{1,k} & s_{1,k-1} & \dots & s_{1,1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & s_{(n-k-1),k} & s_{(n-k-1),k-1} & \dots & s_{(n-k-1),1} \end{pmatrix}$$

VI. Циклический код Хемминга. Неприводимый многочлен $p(x)$ степени m называется примитивным, если наименьшая степень n , при котором $x^n - 1$ делится на $p(x)$ без остатка равна $n = 2^m - 1$.

Теорема 3.19. Любой циклический код Хэмминга длины $2^m - 1$ с $m \geq 3$ может быть построен с помощью некоторого примитивного многочлена степени m . И обратно, любому примитивному многочлену степени m соответствует некоторый код Хэмминга длины $2^m - 1$.

Рассмотрим поле F_8 , для построения поля используем примитивный многочлен $p(x) = x^3 + x + 1$ над F_2 , $\alpha(x) = x$ является примитивным элементом поля. Перечислим элементиты поля

$$\begin{aligned} \alpha^0 &= (001), \alpha^1 = (010), \alpha^2 = (100), \alpha^3 = (011), \\ \alpha^4 &= (110), \alpha^5 = (111), \alpha^6 = (101), \alpha^7 = (001). \end{aligned}$$

Если многочлен $c(x)$ является кодовым многочленом, то

$$\bar{c} \cdot H^t = \bar{c} \cdot [\alpha^0, \alpha^1, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6] = \sum_{i=0}^6 c_i \alpha^i = \bar{0}.$$

То есть \bar{c} – кодовое слово тогда и только тогда, когда α корень многочлена $c(x)$. Значит, $c(x)$ делится на минимальный многочлен элемента α . Поэтому многочлен $g(x) = (x - \alpha)(x - \alpha^2)(x - \alpha^4) = x^3 + x + 1$ является порождающим элементом (7,4) кода Хэмминга.

Код, двойственный к $(2^m - 1, 2^m - 1 - m)$ коду Хэмминга H_m является симплектическим кодом. Все кодовые слова двойственного кода имеют одинаковый вес.

Теорема 3.20. Если S_m – код, двойственный $(2^m - 1, 2^m - 1 - m)$ коду Хэмминга в $F_{2^m}^{2^m-1}$, то вес всех кодовых слов равен 2^{m-1} .

Доказательство. По индукции. Пусть $m = 2$ и

$$H_2 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

проверочная матрица $(3, 1)$ кода Хэмминга. Тогда кодовое пространство S^2 состоит из векторов

$$(000), (101), (011), (110).$$

Вес всех ненулевых кодовых векторов равен $2 = 2^{2-1}$.

Пусть для $m = k$ теорема верна, то есть для любого $\bar{x} \in S_k$, $\bar{x} \neq \bar{0}$, имеем $w(\bar{x}) = 2^{k-1}$. И пусть $m = k + 1$. Все кодовые слова являются линейными комбинациями строк матрицы H_{k+1} . Пусть

$$\bar{x} = \alpha_1 \bar{h}_1 + \dots + \alpha_{k+1} \bar{h}_{k+1} \in S_{k+1},$$

где $\bar{h}_1, \dots, \bar{h}_{k+1}$ – строки матрицы H_{k+1} и $\alpha_i \in F_2$. Рассмотрим два случая: $\alpha_{k+1} = 0$ и $\alpha_{k+1} = 1$. Переставляя столбцы матрицы H_{k+1} , можем эту матрицу представить в виде

$$H_{k+1} = \begin{pmatrix} H_k & \bar{0}^t & H_k \\ 0 \dots 0 & 1 & 1 \dots 1 \end{pmatrix}$$

Здесь $\bar{0}$ – нулевой вектор длины k . В первом случае кодовый вектор \bar{x} будет иметь вид

$$\bar{x} = (x_1 \dots x_{2^k-1} 0 \ x_1 \dots x_{2^k-1}).$$

Тогда $w(\bar{x}) = 2w(\tilde{x})$, $w(\tilde{x}) = 2^{k-1}$ и $w(\bar{x}) = 2^k$.

Во втором случае

$$\bar{x} = \alpha_1 \bar{h}_1 + \dots + \alpha_k \bar{h}_k + \bar{h}_{k+1} = \bar{x}' + \bar{h}_{k+1},$$

где $\bar{x}' = (x'_1 \dots x'_{2^k-1} 0 \ x'_1 \dots x'_{2^k-1})$. Можем считать, что $\bar{h}_{k+1} = (\underbrace{0 \dots 0}_{2^k-1} \underbrace{1 \dots 1}_{2^k})$.

Пусть $\tilde{x}' = (x'_1 \dots x'_{2^k-1}) \in H_k$.

Если $w(\tilde{x}') \neq 0$, то $w(\tilde{x}') = 2^{k-1}$ и $w(\bar{x}') = 2^k$. Отсюда получаем $w(\bar{x}) = w(\tilde{x}') + 1 + 2^k - 1 - w(\tilde{x}') = 2^k$.

Если же $w(\tilde{x}') = 0$, то $w(\bar{x}') = 0$. Поэтому $w(\bar{x}) = w(\bar{h}_{k+1}) = 2^k$.

VII. Двоичный код Голея. Можно заметить, что

$$(C_{23}^0 + C_{23}^1 + C_{23}^2 + C_{23}^3)2^{12} = 2^{23}.$$

Это равенство представляет собой необходимое условие существования совершенного двоичного кода, исправляющего три ошибки.

Такой код действительно существует и назван был кодом Голея. В основе конструкции лежит равенство

$$x^{23} - 1 = (x - 1)g_1(x)g_2(x)$$

где

$$g_1(x) = 1 + x^2 + x^4 + x^5 + x^6 + x^{10} + x^{11}$$

и

$$g^2(x) = 1 + x + x^5 + x^6 + x^7 + x^9 + x^{11}$$

В качестве порождающего многочлена можно использовать как $g_1(x)$, так и $g_2(x)$. Шары с центрами в кодовых словах упаковывают пространство F_2^{23} . Поэтому кодовое расстояние не может быть больше 7. И можно доказать, что оно не меньше 7. Поэтому заключаем, что кодовое расстояние кода Голея равно 7.

Помимо двоичного кода Голея существует совершенный троичный $(11, 6)$ код с кодовым расстоянием, равным 5. Других линейных совершенных кодов, исправляющих более одной ошибки, не существует. \square

VIII. Декодирование по Меггитта. Алгоритм декодирования опирается на следующее утверждение.

Предложение 3.2. Пусть $s(x)$ является синдромом принятого из канала слова $r(x)$ некоторого циклического (n, k) кода. Обозначим через $s_1(x)$ остаток от деления многочлена $x \cdot s(x)$ на порождающий многочлен $g(x)$. Тогда $s_1(x)$ является синдромом $r_1(x)$, то есть остатком от деления циклического сдвига на многочлен $g(x)$.

Доказательство. Пусть

$$r(x) = r_0 + r_1x + \dots + r_{N-1}x^{n-1}, \quad x \cdot r(x) = r_0x + r_1x^2 + \dots + r_{N-1}x^n \quad \text{и} \quad r^{(1)}(x) = x \bullet r(x).$$

Тогда

$$r^{(1)}(x) = r_{n-1} + r_0x + \dots + r_{n-2}x^{n-1} = r_{n-1}[x^n - 1] + x \cdot r(x).$$

Положим

$$r^{(1)}(x) = a(x)g(x) + \tilde{s}(x), \quad r(x) = b(x)g(x) + s(x) \quad \text{и} \quad x^n - 1 = g(x) \cdot h(x).$$

Тогда

$$\begin{aligned} r^{(1)}(x) &= a(x)g(x) + \tilde{s}(x) = r_{n-1} + r_0x + \dots + r_{n-2}x^{n-1} = r_{n-1}[x^n - 1] + x \cdot r(x) = \\ &= r_{n-1}h(x)g(x) + x[b(x)g(x) + s(x)] \end{aligned}$$

Отсюда следует, что

$$x \cdot s(x) = [a(x) + r_{n-1}h(x) + x b(x)] \cdot g(x) + \tilde{s}(x).$$

Предложение доказано. \square

В результате,

1. Между множеством всех исправляемых ошибок и множеством соответствующих синдромов существует взаимно однозначное соответствие.

2. Если $s(x)$ – синдром, соответствующий многочлену ошибок $e(x)$, то $xs(x) \bmod g(x)$ – синдром, соответствующий $xe(x) \bmod (x^n - 1)$.

3. Пусть $s(x)$ – синдром принятого слова $y(x)$ некоторого циклического (n, k) кода. Пусть $s_1(x) = xs(x) \% g(x)$. Тогда $s_1(x)$ является синдромом циклического сдвига принятого слова, то есть $xy(x) \% g(x)$.

Из вышесказанного следует, что множество всех ошибок можно разбить на классы эквивалентности таким образом, чтобы каждый класс состоял из циклических сдвигов одной комбинации ошибок и сохранять в памяти только синдромы одного из представителей каждого класса эквивалентности. Для определения принадлежности ошибок данному классу нужно выполнить операцию $xs(x) \% g(x)$, не более n раз, и сравнить результат с содержимым памяти. При обнаружении такого соответствия ошибки сдвинутого многочлена $y(x)$ исправляются и обратным сдвигом кодовое слово восстанавливается.

Используемая литература

1. Колесник В.Д., Полтырев Г.Ш. Курс теории информации. М.: Наука, 1982.
2. Самсонов Б.Б., Плохов Е.М., Филоненков А.И., Кречет Т.В. Теория информации и кодирование. Ростов на Дону: Феникс, 2002.
3. Котоусов А.С. Теория информации. М.: Радио и связь, 2003.
4. Вернер М. Основы кодирования. М.: Техносфера. 2004.
5. Соловьева Ф.И. Введение в теорию кодирования. Учебное пособие. Новосибирск. 2011.
6. Блейхут Р. Теория и практика кодов, контролирующих ошибки. М.: Мир. 1986.