

ТЕХНОЛОГИИ ОБРАБОТКИ ИНФОРМАЦИИ

ЛЕКЦИЯ 5

ТЕОРЕМА ШЕННОНА ДЛЯ КАНАЛА С ПОМЕХАМИ

Теорема Шеннона. Пусть источник характеризуется д.с.в. X . Рассматривается канал с шумом, т.е. для каждого передаваемого сообщения задана вероятность ε его искажения в процессе передачи (вероятность ошибки). Тогда существует такая скорость передачи u , зависящая только от X , что $\forall \varepsilon > 0 \exists u' < u$ сколь угодно близкая к u такая, что существует способ передавать значения X со скоростью u' и с вероятностью ошибки меньшей ε , причем $u = C/HX$. Упомянутый способ образует помехоустойчивый код.

ПОМЕХОЗАЩИТНОЕ КОДИРОВАНИЕ

Есть несколько типов помехозащитных кодов

1. Обнаруживающие ошибки
2. Обнаруживающие и исправляющие ошибки

Простейший код для борьбы с шумом — это контроль четности, он, в частности, широко используется в модемах. Кодирование заключается в добавлении к каждому байту девятого бита таким образом, чтобы дополнить количество единиц в байте до заранее выбранного для кода четного (even) или нечетного (odd) значения. Используя этот код, можно лишь обнаруживать большинство ошибок.

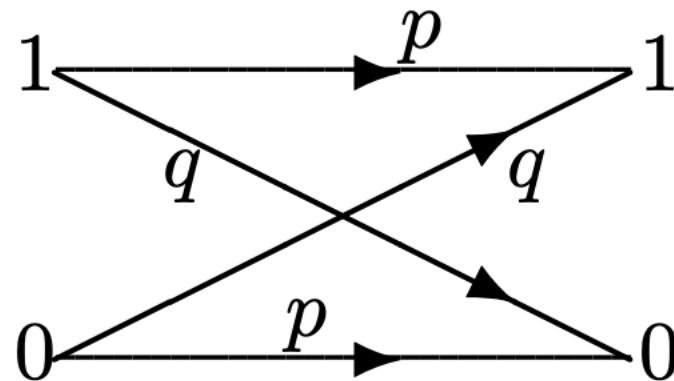
ДВОИЧНЫЙ СИММЕТРИЧНЫЙ КАНАЛ

Двоичный симметричный канал реализует схему Бернулли, поэтому вероятность передачи n бит по двоичному симметричному каналу с k ошибками равна $P_n(k) = C_n^k p^{n-k} q^k$.

Пример. Вероятность передачи одного бита информации с ошибкой равна $q = 0.01$ и нас интересует вероятность безошибочной передачи 1000 бит (125 байт).

Искомую вероятность можно подсчитать по формуле

$P_{1000}(0) = C_{1000}^0 p^{1000} q^0 = 0.99^{1000} \approx 4.32 \cdot 10^{-5}$, т.е. она ничтожно мала.



МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СИСТЕМЫ СВЯЗИ

Система кодирования

$$E(a_1 \dots a_m) = a_1 \dots a_m a_{m+1},$$

$$a_{m+1} = \begin{cases} 0, & \text{если } \sum_{i=1}^m a_i \text{ — четная;} \\ 1, & \text{если } \sum_{i=1}^m a_i \text{ — нечетная.} \end{cases}$$

Система декодирования

$$D(a_1 \dots a_m a_{m+1}) = \begin{cases} a_1 \dots a_m, & \text{если } \sum_{i=1}^{m+1} a_i \text{ — четна;} \\ \langle \text{ошибка} \rangle, & \text{если } \sum_{i=1}^{m+1} a_i \text{ — нечетна.} \end{cases}$$

КОД С ПОВТОРЕНИЕМ

Рассмотрим $(m, 3m)$ -код с тройным повторением.

Любое сообщение разбивается на блоки длиной m каждое и каждый блок передается трижды — это определяет функцию E

Функция D определяется следующим образом. Принятая строка разбивается на блоки длиной $3m$. Бит с номером i ($1 \leq i \leq m$) в декодированном блоке получается из анализа битов с номерами $i, i+m, i+2m$ в полученном блоке: берется тот бит из трех, который встречается не менее двух раз.

Вероятность того, что бит в данной позиции будет принят трижды правильно равна p^3 . Вероятность одной ошибки в тройке равна $3p^2q$. Поэтому вероятность правильного приема одного бита равна $p^3 + 3p^2q$. Аналогичным образом получается, что вероятность приема ошибочного бита равна $q^3 + 3pq^2$.

Пример. Предположим $q = 0.1$. Тогда вероятность ошибки при передаче одного бита — 0.028, т.е. этот код снижает вероятность ошибки с 10% до 2.8%. Подобным образом организованная передача с пятикратным повторением даст вероятность ошибки на бит.

$$q^5 + 5pq^4 + 10p^2q^3 = 0.00856 = 0.856\%,$$

т.е. менее 1%

ПРИМЕР КОДА С КОНТРОЛЬНОЙ СУММОЙ

(2048, 2313)-код, используемый
при записи данных на магнитофонную
ленту компьютерами Apple II
CS – check sum
PB – parity bit

	1	2	3	4	5	6	256 CS	1
1					p_1			
2	p_2						p_3	
3			p_4					
4								
5				p_5				p_6
6								
7			p_7					
8								
PB	p_8						p_9	

КОДЫ ХЭММИНГА

Пример. Рассмотрим $(1, 3)$ -код, состоящий из E , задающей отображение

$0 \rightarrow 000$

$1 \rightarrow 111$

и D , задающей отображение

$000 \rightarrow 0$

$001 \rightarrow 0$

$010 \rightarrow 0$

$011 \rightarrow 1$

$100 \rightarrow 0$

$101 \rightarrow 1$

$110 \rightarrow 1$

$111 \rightarrow 1$

Этот код исправляет ошибки в одной позиции, т.к. минимальное расстояние между словами кода равно 3.

КОДЫ ХЭММИНГА

$$a_1 = 00 \rightarrow 00000 = b_1, \quad a_2 = 01 \rightarrow 01011 = b_2,$$

$$a_3 = 10 \rightarrow 10101 = b_3, \quad a_4 = 11 \rightarrow 11110 = b_4.$$

Этот код способен исправлять однократную ошибку, потому что любые два кодовых слова отличаются по меньшей мере в трех позициях.

МАТРИЧНОЕ КОДИРОВАНИЕ

Пусть E матрица размерности $m \times n$, состоящая из элементов e_{ij} , где i — это номер строки, а j — номер столбца. Каждый из элементов матрицы e_{ij} может быть либо 0, либо 1. Кодирование реализуется операцией $b = aE$ или $b_j = a_1 e_{1j} + a_2 e_{2j} + \dots + a_m e_{mj}$, где кодовые слова рассматриваются как векторы, т.е как матрицы-строки размера $1 \times n$.

$$E = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Кодирование задается отображениями:

$000 \rightarrow 000000$, $001 \rightarrow 001111$, $010 \rightarrow 010011$, $011 \rightarrow 011100$,
 $100 \rightarrow 100110$, $101 \rightarrow 101001$, $110 \rightarrow 110101$, $111 \rightarrow 111010$.

ПОМЕХОУСТОЙЧИВЫЙ КОД ХЭММИНГА

Кодируем сообщение 0100010000111101

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	1	0	1	
X		X		X		X		X		X		X		X		X		X		X	1
	X	X			X	X			X	X			X	X			X	X			2
			X	X	X	X					X	X	X	X					X	X	4
							X	X	X	X	X	X	X	X							8
															X	X	X	X	X	X	16

100110000100001011101

ПОМЕХОУСТОЙЧИВЫЙ КОД ХЭММИНГА – ДЕКОДИРОВАНИЕ

Декодирование сообщения **100110001100001011101**

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	1	0	1	
X		X		X		X		X		X		X		X		X		X		X	1
	X	X			X	X			X	X			X	X			X	X			2
			X	X	X	X					X	X	X	X					X	X	4
							X	X	X	X	X	X	X	X							8
															X	X	X	X	X	X	16

Значения помехозащитных бит – **0,0,1,1,0**

Старые значения – **1,0,1,0,0**

Несовпадающие биты – **1,0,0,1,0** → $0,1,0,0,1_{\text{bin}} == 9_{\text{dec}}$

КОДЫ РИДА–СОЛОМОНА

- Используются для кодирования информации на популярных форматах –
- CD и DVD.
- Избыточность: для записи 700 Мб данных в реальности записывается 1943 Мб
- Однако коды Рида-Соломона очень устойчивы к глубоким повреждениям дисков.
- Это достигается за счёт применения кодов Рида-Соломона и разнесению различных блоков данных по разным частям носителя