

Лекция 1. Введение в кодирование и линейные коды

Все данные, которые копируются или передаются по сети, кодируются, чтобы убрать искажения.

Упрощённая модель цифровой системы связи

Математическая модель:

- Дискретные последовательности.
- Цифровая последовательность — это элементы поля $GF(2) = \{0, 1\}$.
- Канал — двоичный симметричный канал (ДСК).

Для двоичного симметричного канала:

$$3p = 10^{-3}.$$

Как уменьшить вероятность ошибки p ?

1) Дублирование данных (повторный код). Пример:

$$0011 \longrightarrow 000\ 000\ 111\ 111,$$

то есть

$$0 \longrightarrow 000, \quad 1 \longrightarrow 111.$$

Такой код называется **линейным**: мод 2 сумма кодовых слов снова является кодовым словом.

Для $p = 10^{-3}$ посчитаем вероятность ошибки в *кодowych* символах.

При передаче 000 возможны приёмы:

$$000 \rightarrow 000, 001, 010, 100 \quad (\text{правильно решаем})$$

и

$$000 \rightarrow 011, 101, 110, 111 \quad (\text{ошибка}).$$

Вероятность ошибки при декодировании (для одного кодового блока длины 3):

$$P_{\text{ош}} = 3p^2(1 - p) + p^3 = 3p^2 - 2p^3 \approx 10^{-6}.$$

В данном случае код исправляет **одну ошибку**.

2) Другой пример кода. Рассмотрим кодирование пар битов:

$$00 \longrightarrow 00000,$$

$$01 \longrightarrow 10110,$$

$$10 \longrightarrow 01011,$$

$$11 \longrightarrow 11101.$$

Этот код тоже линейный. Для оценки $P_{\text{ош}}$ достаточно рассмотреть кодовое слово 00000 (остальные получаются сложением с ним).

Примеры декодирования по принципу минимального числа отличий (минимального расстояния Хэмминга):

- Пришло 00001 — ближайшее кодовое слово 00000 (1 ошибка).
- Пришло 01001 — ближайшее кодовое слово 01011 (2 ошибки и т.д.).

Реально такой код также исправляет **одну ошибку**.

Скорость кода

Пусть:

- k — число *информационных* символов,
- n — число *кодových* символов.

Тогда **скорость кода**:

$$R = \frac{k}{n}.$$

В задаче кодирования от помех $R \leq 1$.

$$\text{В примере 1): } R = \frac{1}{2}, \quad \text{в примере 2): } R = \frac{2}{5}.$$

Принцип “меньше отличий” означает меньшее расстояние между кодовыми словами (формальное определение ниже).

Пропускная способность двоичного симметричного канала

(Клод Шеннон, 1948 г.)

Для ДСК с переходной вероятностью ошибки p вводится **пропускная способность**:

$$C = 1 - h(p),$$

где

$$h(x) = -x \log_2 x - (1 - x) \log_2 (1 - x)$$

— энтропия двоичного ансамбля.

Оказывается:

- При $R < C$ может быть обеспечена сколь угодно малая вероятность ошибки декодирования за счёт увеличения длины используемых кодов.
- Если $R > C$, то надёжная передача невозможна.

Вес и расстояние Хэмминга

Пусть x — кодовое слово. Тогда

$\omega(x)$ — вес Хэмминга = число ненулевых элементов в x .

$d(x, y)$ — расстояние Хэмминга = число позиций, в которых x и y различаются.

Примеры:

$$\omega(001101) = 3, \quad d(001101, 101001) = 2.$$

Свойства:

- $d(x, y)$ — метрика.
- Для бинарного случая $d(x, y) = \omega(x + y)$, где сложение поразрядное по модулю 2.
- В частности, $d(x, 0) = \omega(x)$.

Пример кода длины 5

Рассмотрим код:

$$\begin{aligned}00 &\longrightarrow 00000, \\01 &\longrightarrow 10110, \\10 &\longrightarrow 01011, \\11 &\longrightarrow 11101.\end{aligned}$$

Построим матрицу расстояний:

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0 | 3 | 3 | 4 |
| 2 | 3 | 0 | 4 | 3 |
| 3 | 3 | 4 | 0 | 3 |
| 4 | 4 | 3 | 3 | 0 |

Минимальное расстояние кода:

$$d_{\min} = \min_{x \neq y} d(x, y) = 3.$$

Код исправляет ошибки кратности

$$t = \left\lfloor \frac{d_{\min} - 1}{2} \right\rfloor.$$

Определение линейного кода

Линейный код — это код, в котором сумма (по модулю 2) любых двух кодовых слов является кодовым словом:

$$x, y \in C \Rightarrow x + y \in C.$$

Тогда:

$$d(x, y) = \omega(x + y),$$

и, в частности,

$$d_{\min} = \min_{x \neq y} d(x, y) = \min_{x \neq 0} d(x, 0) = \min_{x \neq 0} \omega(x).$$

Линейный q -ичный (n, k) -код C — это любое k -мерное подпространство пространства F^n всех векторов длины n над полем $GF(q)$.

Пример: $q = 3$, $k = 2$, $n = 5$.

$$GF(3) = \{0, 1, 2\}.$$

Из $k = 2$ следует, что имеется $q^k = 3^2 = 9$ различных кодовых слов. Из всех 3^5 возможных векторов длины 5 выбираются 9 векторов, образующих линейный код.

Обычно мы рассматриваем случай $q = 2$.

Базис и порождающая матрица

Для приведённого бинарного кода длины 5:

00000
10110
01011
11101

Одним из возможных базисов будут векторы

$$10110 \quad \text{и} \quad 01011.$$

Порождающая матрица (n, k) -кода — это матрица размера $k \times n$, строки которой являются базисными кодовыми словами. Она обозначается G . Любое кодовое слово является линейной комбинацией строк G .

Пусть

- m — информационное слово (вектор длины k),
- c — кодовое слово (вектор длины n).

Тогда

$$c = mG,$$

где произведение берётся по модулю 2 (или по модулю q в общем случае).

Проверочная матрица

Предположим, что существует вектор

$$h = (h_1, h_2, \dots, h_n),$$

такой что для любого кодового слова c выполняется

$$(c, h) = 0,$$

где (\cdot, \cdot) — скалярное произведение по модулю 2.

Вектор h ортогонален всем кодовым словам и является **проверкой** (проверяющим вектором).

Если собрать $(n - k)$ линейно независимых проверяющих векторов в строки матрицы H размера $(n - k) \times n$, то

$$GH^T = 0,$$

и H называется **проверочной матрицей** кода.

Всего существует $(n - k)$ независимых проверок.