Расширенный теорминимум по теории информации и кодированию

По Григорьеву А.А. Chatgpt 5.2 Thinking, Gemini 3.0 Pro
и немного Жданова Елисея

Содержание

1 Теория информации: энтропия Шеннона, Хартли и информационная дивергенция **4**
1.1 Условная вероятность 4
1.2 Совместное, маргинальные и условные распределения 4
1.3 Формула Байеса, априорное/апостериорное, правдоподобие 4
1.4 Энтропия Хартли 5
1.5 Энтропия Шеннона 5
1.6 Двоичная функция энтропии и q-ичная энтропия 5
1.7 Свойство группировки энтропии 5
1.8 Аддитивность энтропии 5
1.9 Максимальность при равномерном распределении 6
1.10 Информационная дивергенция (дивергенция Кульбака–Лейблера) . . . 6

2 Информационная дивергенция, принцип максимума энтропии и границы Чебышёва **6**
2.1 Выпуклая функция и неравенство Йенсена 6
2.2 Фундаментальное неравенство логарифма 6
2.3 Выпуклость множества распределений вероятностей 7
2.4 Неотрицательность дивергенции (неравенство Гиббса) 7
2.5 Условие равенства нулю 7
2.6 Метод множителей Лагранжа 7
2.7 Принцип максимума энтропии 8
2.8 Неравенство Чебышёва 8

3 Энтропия и пределы сжатия данных: от Чебышёва до теоремы Шеннона **8**
3.1 Слабый закон больших чисел 8
3.2 AEP (асимптотическая равнораспределённость) 8
3.3 Типичное множество 8
3.4 Совместно типичные последовательности 9

4 Кодирование источника: энтропия, неравенство Крафта и алгоритм Хаффмана **9**
4.1 Кодирование источника и энтропия источника 9
4.2 Средняя длина и скорость кодирования 9

4.3 Префиксный и однозначно декодируемый коды. Кодовое дерево 10
4.4 Неравенство Крафта 10
4.5 Код Хаффмана и оптимальность 10
4.6 Граница Шеннона для средней длины и блоковое кодирование 10

5 Алгоритмы сжатия данных: от Хаффмана и Танстолла до арифметического и универсального кодирования Лемпеля–Зива 11
5.1 Код Танстолла 11
5.2 Арифметическое кодирование 11
5.3 Алгоритмы Лемпеля–Зива, словарное и адаптивное сжатие 11

6 Теория информации: условная энтропия и взаимная информация 11
6.1 Совместная и условная энтропия 11
6.2 Цепное правило для энтропии 12
6.3 Неравенство: условная энтропия не превосходит безусловную 12
6.4 Взаимная информация 12

7 Теория информации: взаимная информация, пропускная способность и абсолютная стойкость шифров 12
7.1 Канал, матрица переходных вероятностей и скорость передачи 12
7.2 Пропускная способность канала 13
7.3 Канал без памяти 13
7.4 Цепь Маркова и лемма об обработке информации 13
7.5 Абсолютная стойкость шифров (шифр Вернама) 13

8 Пропускная способность и фундаментальные пределы: неравенство Фано и обратная теорема 14
8.1 Неравенство Фано 14
8.2 Обратная теорема кодирования Шеннона (а это важно) 14

9 Прямая теорема Шеннона: случайное кодирование, совместная типичность и примеры каналов 14
9.1 Прямая теорема кодирования Шеннона (achievability) 14
9.2 Матрица переходных вероятностей 14
9.3 Двоичный симметричный канал (BSC) 14
9.4 Двоичный канал со стиранием (BEC) 15
9.5 q-ичный симметричный канал 15

10 Пропускная способность гауссовского канала: дифференциальная энтропия и формула Шеннона–Хартли 15
10.1 Дифференциальная энтропия 15
10.2 Гауссовский канал и AWGN 15
10.3 Максимальная энтропия при фиксированной дисперсии 16
10.4 Формула Шеннона–Хартли 16
10.5 Отношение сигнал/шум, спектральная эффективность и предел Найквиста 16
10.6 Гильбертово пространство сигналов и ортогональность 16

11 Пропускная способность гауссовского канала и фундаментальный предел Шеннона **17**
11.1 Предел Шеннона и E_b/N_0 17
11.2 Геометрическая интерпретация пропускной способности 17
11.3 Эффект концентрации меры 17

12 Основы помехоустойчивого кодирования: гауссовские каналы, мягкие решения и метрики **18**
12.1 Оптимальное решающее правило, ML и MAP 18
12.2 Мягкие и жёсткие решения 18
12.3 Евклидово расстояние и метрика для AWGN 18
12.4 Расстояние Хэмминга, сферы и шары 19

13 Блоковые коды: границы Хэмминга, Синглтона, Варшамова–Гилберта и введение в линейные коды **19**
13.1 Параметры кода и скорость 19
13.2 Граница Хэмминга (упаковочная) 19
13.3 Граница Синглтона 19
13.4 Граница Варшамова–Гилберта (существование) 20
13.5 Введение в линейные коды 20

14 Линейные коды: порождающие и проверочные матрицы, коды Хэмминга и Рида–Соломона **20**
14.1 Линейный код, вес и расстояние 20
14.2 Порождающая и проверочная матрицы 20
14.3 Код Хэмминга 21
14.4 Коды Рида–Соломона, MDS и матрица Вандермонда 21

15 Свёрточные коды: алгоритмы Витерби, BCJR и принципы турбо-кодирования **21**
15.1 Свёрточные коды и решётчатая диаграмма 21
15.2 Алгоритм Витерби (ML-декодирование по пути) 22
15.3 Алгоритм BCJR (MAP/APP по битам, forward–backward) 22
15.4 Турбо-коды и итеративное декодирование 22
15.5 Message passing (общий принцип) 22

Как читать документ

Этот файл — сборник понятий из лекций. Определения тут расписаны плохо, а алгоритмы - без примеров работы. И вообще много лишних описаний. Но понятия тут покрывают теорминимум Григорьева + Бибикова.

1 Теория информации: энтропия Шеннона, Хартли и информационная дивергенция

1.1 Условная вероятность

Определение 1.1 (Условная вероятность)**.** Для событий A, B при $\mathbb{P}(A) > 0$:

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Интуиция из лекции: «после наступления A вероятностное пространство как бы сжимается до A», и мы нормируем меру на этом подмножестве.

1.2 Совместное, маргинальные и условные распределения

Пусть X и Y — дискретные случайные величины.

Определение 1.2 (Совместное распределение)**.**

$$p_{X,Y}(x, y) = \mathbb{P}(X = x,\ Y = y).$$

Определение 1.3 (Маргинальные распределения)**.**

$$p_X(x) = \sum_y p_{X,Y}(x, y), \qquad p_Y(y) = \sum_x p_{X,Y}(x, y).$$

Определение 1.4 (Условные распределения)**.** При $p_X(x) > 0$:

$$p(y \mid x) = \mathbb{P}(Y = y \mid X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

Аналогично

$$p(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} \quad \text{при } p_Y(y) > 0.$$

В матричном представлении (часто рисовали на лекции): строки соответствуют x, столбцы y, элементы — $p_{X,Y}(x, y)$. Тогда маргиналы — суммы по строкам/столбцам.

А еще удобно интерпретировать как проекции из облака p(x, y) на координатные оси p(x) и p(y). (Бибиков А.М.)

1.3 Формула Байеса, априорное/апостериорное, правдоподобие

Теорема 1.1 (Формула Байеса)**.** *При $p_Y(y) > 0$:*

$$p(x \mid y) = \frac{p(x)\, p(y \mid x)}{p(y)}, \qquad p(y) = \sum_x p(x) p(y \mid x).$$

Определение 1.5 (Априорное распределение)**.** $p(x)$ — распределение X «до наблюдения» Y.

Определение 1.6 (Функция правдоподобия)**.** $p(y \mid x)$, рассматриваемая как функция гипотезы x при фиксированном наблюдении y.

Определение 1.7 (Апостериорное распределение)**.** $p(x \mid y)$ — распределение X после наблюдения $Y = y$.

Лекционная интерпретация: «априорное» → «наблюдение через канал/датчик» → «апостериорное». Байес даёт точную формулу обновления.

1.4 Энтропия Хартли

Определение 1.8 (Энтропия Хартли)**.** Если $|\mathcal{X}| = M$ и X равномерна на $\mathcal{X}$, то

$$H_0(X) = \log_2 M.$$

Смысл: минимальное число бит, чтобы однозначно идентифицировать один из M равновероятных исходов.

1.5 Энтропия Шеннона

Определение 1.9 (Самоинформация и энтропия)**.** Самоинформация исхода x:

$$I(x) = -\log_2 p(x).$$

Энтропия Шеннона:

$$H(X) = \mathbb{E}[I(X)] = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x).$$

Смысл (как в курсе). Энтропия — «средняя трудность угадывания» или «средняя длина оптимального описания» при безошибочном сжатии (позже формализуется теоремой Шеннона и кодами Хаффмана/арифметикой).

1.6 Двоичная функция энтропии и q-ичная энтропия

Определение 1.10 (Двоичная энтропия)**.**

$$H_2(p) = -p \log_2 p - (1-p) \log_2 (1-p), \qquad p \in [0, 1].$$

1.7 Свойство группировки энтропии

Утверждение 1.1 (Группировка)**.** *Пусть исходы X сгруппированы в классы G. Тогда*

$$H(X) = H(G) + \sum_{g} \mathbb{P}(G = g)\, H(X \mid G = g).$$

Смысл: «сначала кодируем номер группы, потом внутри группы». Это фундаментальный принцип композиции описаний.

1.8 Аддитивность энтропии

Утверждение 1.2 (Цепное правило для двух)**.**

$$H(X, Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y).$$

Следствие 1.1 (Аддитивность при независимости)**.** *Если $X \perp Y$, то $H(Y \mid X) = H(Y)$ и*

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

1.9 Максимальность при равномерном распределении

Теорема 1.2 (Максимум энтропии на конечном алфавите)**.** *Если* $|\mathcal{X}| = M$*, то*

$$0 \le H(X) \le \log_2 M,$$

и $H(X) = \log_2 M$ *тогда и только тогда, когда* X *равномерна.*

Лекционный путь доказательства через логарифмическое неравенство. Обычно показывают, что для p и равномерного u:

$$D(p\|u) = \sum_x p(x) \log_2 \frac{p(x)}{1/M} = \sum_x p(x) \log_2 p(x) + \log_2 M = \log_2 M - H(X) \ge 0,$$

откуда $H(X) \le \log_2 M$. Равенство $D = 0$ означает $p = u$.

1.10 Информационная дивергенция (дивергенция Кульбака–Лейблера)

Определение 1.11 (KL-дивергенция)**.** Для распределений p, q на $\mathcal{X}$ (и условия $p(x) > 0 \Rightarrow q(x) > 0$):

$$D(p\|q) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)}.$$

Смысл. $D(p\|q)$ измеряет «насколько дорого» кодировать данные, устроенные как p, кодом, оптимальным для q. В лекциях также подчёркивалось:

- $D(p\|q) \ge 0$ (неотрицательность).
- $D(p\|q) = 0 \Leftrightarrow p = q$ (на носителе p).
- Не симметрична: $D(p\|q) \ne D(q\|p)$, не является метрикой.

2 Информационная дивергенция, принцип максимума энтропии и границы Чебышёва

2.1 Выпуклая функция и неравенство Йенсена

Определение 2.1 (Выпуклость)**.** Функция f выпукла на интервале, если

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y), \qquad \lambda \in [0,1].$$

Теорема 2.1 (Йенсен)**.** *Если* f *выпукла и* $\mathbb{E}|X| < \infty$*, то*

$$f(\mathbb{E}X) \le \mathbb{E} f(X).$$

Ключевой пример из курса. Функция $-\log x$ выпукла на $x > 0$. Это лежит в основе доказательства Гиббса (неотрицательности KL).

2.2 Фундаментальное неравенство логарифма

Утверждение 2.1 (Неравенство логарифма)**.** *Для* $x > 0$*:*

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

равенство только при $x = 1$*.*

Часто используется как «локально лучшая линейная аппроксимация логарифма».

2.3 Выпуклость множества распределений вероятностей

Утверждение 2.2. *Множество всех распределений на конечном алфавите $\mathcal{X}$ выпукло: если p, q — распределения и $\lambda \in [0,1]$, то $r = \lambda p + (1-\lambda)q$ — распределение.*

2.4 Неотрицательность дивергенции (неравенство Гиббса)

Теорема 2.2 (Гиббс)**.**

$$D(p\|q) \geq 0.$$

Идея доказательства (как на лекции через Йенсена). Пишем

$$D(p\|q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)} = -\sum_x p(x) \log_2 \frac{q(x)}{p(x)}.$$

Перейдём к натуральному логу (константа $\log_2 e$ роли не играет):

$$-\sum_x p(x) \ln \frac{q(x)}{p(x)} = \mathbb{E}_p\left[-\ln\left(\frac{q(X)}{p(X)}\right)\right].$$

Функция $-\ln(\cdot)$ выпукла, значит по Йенсену:

$$\mathbb{E}_p[-\ln Z] \geq -\ln \mathbb{E}_p[Z], \qquad Z = \frac{q(X)}{p(X)}.$$

Но

$$\mathbb{E}_p[Z] = \sum_x p(x)\frac{q(x)}{p(x)} = \sum_x q(x) = 1,$$

значит $\mathbb{E}_p[-\ln Z] \geq -\ln 1 = 0$, откуда $D(p\|q) \geq 0$.

2.5 Условие равенства нулю

Утверждение 2.3.

$$D(p\|q) = 0 \iff p(x) = q(x) \text{ для всех } x, \text{ где } p(x) > 0.$$

2.6 Метод множителей Лагранжа

Определение 2.2 (Лагранжиан)**.** Для задачи $\max f(x)$ при ограничениях $g_i(x) = c_i$:

$$L(x, \lambda) = f(x) - \sum_i \lambda_i \left(g_i(x) - c_i\right).$$

Определение 2.3 (Условие стационарности)**.** Необходимое условие экстремума:

$$\nabla_x L(x, \lambda) = 0, \qquad g_i(x) = c_i.$$

Геометрическая интерпретация. В точке оптимума градиент цели лежит в линейной оболочке градиентов ограничений: «уровень цели касается поверхности ограничений».

2.7 Принцип максимума энтропии

Теорема 2.3 (Максимум энтропии при фиксированных ограничениях)**.** *В реальном мире наиболее часто встречающимся распределением будет то, которое максимизирует энтропию*

$$p^* = \arg\max_{p \in \mathcal{P}} H(p).$$

В лекциях это мотивировалось комбинаторно: число реализаций макросостояния $\propto 2^{nH}$ (через формулу Стирлинга), поэтому максимум энтропии соответствует «наиболее многочисленному» макросостоянию.

2.8 Неравенство Чебышёва

Теорема 2.4 (Чебышёв)**.** *Для случайной величины X с $\mathbb{E}X = m$, $\mathrm{Var}(X) = \sigma^2$:*

$$\mathbb{P}(|X - m| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Применение из курса: к среднему $\overline{X}_n$ независимых одинаково распределённых величин, чтобы получить слабый ЗБЧ.

3 Энтропия и пределы сжатия данных: от Чебышёва до теоремы Шеннона

3.1 Слабый закон больших чисел

Теорема 3.1 (Слабый ЗБЧ)**.** *Если $X_1, \ldots, X_n$ i.i.d., $\mathbb{E}X_i = m$, $\mathrm{Var}(X_i) = \sigma^2$, то*

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} m.$$

Доказательство в лекциях: Чебышёв к $\overline{X}_n$, так как $\mathrm{Var}(\overline{X}_n) = \sigma^2/n \to 0$.

3.2 AEP (асимптотическая равнораспределённость)

Теорема 3.2 (AEP, формулировка)**.** *Для i.i.d. источника X^n:*

$$-\frac{1}{n} \log_2 p(X^n) \xrightarrow{\mathbb{P}} H(X).$$

Интуиция: почти вся вероятность сосредоточена на множестве $\approx 2^{nH}$ последовательностей, каждая из которых имеет вероятность порядка 2^{-nH}.

3.3 Типичное множество

Определение 3.1 (ε-типичное множество)**.**

$$T_\varepsilon^{(n)} = \left\{ x^n : \left| -\frac{1}{n} \log_2 p(x^n) - H(X) \right| < \varepsilon \right\}.$$

Утверждение 3.1 (Свойства типичности). *Для i.i.d. источника:*

- $\mathbb{P}(X^n \in T_\varepsilon^{(n)}) \to 1$ *при* $n \to \infty$.
- *Для* $x^n \in T_\varepsilon^{(n)}$*:*
$$2^{-n(H+\varepsilon)} \le p(x^n) \le 2^{-n(H-\varepsilon)}.$$
- *Следовательно, мощность типичного множества порядка* 2^{nH}*:*
$$(1-o(1))\,2^{n(H-\varepsilon)} \;\lesssim\; \left|T_\varepsilon^{(n)}\right| \;\lesssim\; 2^{n(H+\varepsilon)}.$$

3.4 Совместно типичные последовательности

Определение 3.2 (Совместная типичность). Пара (x^n, y^n) совместно типична, если одновременно типична по $p(x)$, $p(y)$ и $p(x,y)$ (эквивалентно: $-\frac{1}{n}\log p(x^n, y^n) \approx H(X,Y)$ и согласованы маргиналы/условные энтропии).

Ключевой факт для доказательства прямой теоремы Шеннона: если X^n и Y^n независимы, то вероятность «случайно оказаться совместно типичными» экспоненциально мала $\approx 2^{-nI(X;Y)}$.

4 Кодирование источника: энтропия, неравенство Крафта и алгоритм Хаффмана

4.1 Кодирование источника и энтропия источника

Определение 4.1 (Источник). Дискретный источник задаётся распределением $p(x)$ на алфавите $\mathcal{X}$. Энтропия источника $H(X)$ измеряет среднюю «информацию на символ».

Определение 4.2 (Кодирование источника). Кодирование сопоставляет каждому символу $x \in \mathcal{X}$ кодовое слово над кодовым алфавитом (обычно $\{0,1\}$). Длина слова $l(x)$ — число бит.

4.2 Средняя длина и скорость кодирования

Определение 4.3 (Средняя длина).

$$L = \sum_{x \in \mathcal{X}} p(x)\, l(x).$$

Определение 4.4 (Скорость кодирования). В безошибочном кодировании обычно сравнивают L с $H(X)$. Для блоков длины n скорость часто пишут как «бит/символ»:

$$R = \frac{\text{число бит на блок}}{n}.$$

Для фиксированного кода по символам $R = L$.

4.3 Префиксный и однозначно декодируемый коды. Кодовое дерево

Определение 4.5 (Префиксный код)**.** Код префиксный, если ни одно кодовое слово не является префиксом другого.

Определение 4.6 (Однозначно декодируемый код)**.** Любая конкатенация кодовых слов декодируется единственным образом.

Факт из курса: префиксный ⇒ однозначно декодируемый; удобство префиксных — «декодирование на лету» без разделителей.

Кодовое дерево. Префиксный двоичный код соответствует двоичному дереву: левое/правое ребро — 0/1, лист — кодовое слово. Префиксность означает: все слова соответствуют листьям.

4.4 Неравенство Крафта

Теорема 4.1 (Крафт)**.** *Для 2-ичного префиксного кода длины* $\{l(x)\}$ *удовлетворяют*

$$\sum_{x\in\mathcal{X}} 2^{-l(x)} \le 1.$$

Теорема 4.2 (Крафт–Макмиллан (существование))**.** *Если задан набор целых длин* $\{l(x)\}$*, удовлетворяющий* $\sum_x 2^{-l(x)} \le 1$*, то существует префиксный 2-ичный код с этими длинами.*

Смысл. Крафт — «необходимое и достаточное» условие реализуемости длин как префиксного кода.

4.5 Код Хаффмана и оптимальность

Определение 4.7 (Код Хаффмана)**.** Алгоритм Хаффмана строит оптимальное префиксное дерево, итеративно объединяя два наименее вероятных символа (листа) в один узел с суммарной вероятностью, пока не останется один корень.

Теорема 4.3 (Оптимальность Хаффмана)**.** *Код Хаффмана минимизирует среднюю длину* L *среди всех префиксных кодов для данного распределения* $p(x)$*.*

4.6 Граница Шеннона для средней длины и блоковое кодирование

Теорема 4.4 (Граница Шеннона на среднюю длину)**.** *Для любого префиксного (однозначно декодируемого) двоичного кода:*

$$H(X) \le L.$$

Существует код (например, Шеннона–Фано или Хаффмана), такой что

$$L < H(X) + 1.$$

Блоковое кодирование. Если кодировать блоки длины n как символы супералфавита, то энтропия блока $H(X^n) = nH(X)$, а средняя длина на символ может приближаться к $H(X)$ сколь угодно близко при больших n.

5 Алгоритмы сжатия данных: от Хаффмана и Танстолла до арифметического и универсального кодирования Лемпеля–Зива

5.1 Код Танстолла

Определение 5.1 (Код Танстолла)**.** Код Танстолла строит словарь фраз переменной длины (по входу) фиксированной мощности, чтобы максимизировать среднюю длину фразы при фиксированном числе выходных бит.

Лучши загптшить пример работы кода и посмотреть на него

Смысл, который часто спрашивают: «у Хаффмана переменная длина кода, фиксированная длина входного символа; у Танстолла наоборот — переменная длина входной фразы и фиксированная длина выходного индекса».

5.2 Арифметическое кодирование

Определение 5.2 (Арифметическое кодирование)**.** Сообщение (последовательность символов) отображается в подотрезок $[0, 1)$, длина которого равна вероятности сообщения по выбранной модели. Для передачи выбирается двоичный цилиндр (интервал двоичного разбиения), целиком лежащий в этом подотрезке; его двоичная запись и есть код.

Лучши загптшить пример работы кода и посмотреть на него

5.3 Алгоритмы Лемпеля–Зива, словарное и адаптивное сжатие

Определение 5.3 (Словарное сжатие)**.** Сжатие описывает поток как последовательность ссылок на ранее встречавшиеся подстроки/фразы (явный или неявный словарь).

Определение 5.4 (LZ77 (идея))**.** Скользящее окно: ищем самое длинное совпадение текущего префикса буфера с подстрокой в окне и кодируем тройкой $(\text{offset}, \text{length}, \text{next})$.

Лучши загптшить пример работы кода и посмотреть на него

Определение 5.5 (Адаптивное кодирование)**.** Модель/словарь обновляются онлайн по мере чтения данных (статистика не фиксирована заранее).

6 Теория информации: условная энтропия и взаимная информация

6.1 Совместная и условная энтропия

Определение 6.1 (Совместная энтропия)**.**

$$H(X, Y) = -\sum_{x,y} p(x, y) \log_2 p(x, y).$$

Определение 6.2 (Условная энтропия)**.**

$$H(X \mid Y) = \sum_{y} p(y) H(X \mid Y = y) = -\sum_{x,y} p(x, y) \log_2 p(x \mid y).$$

6.2 Цепное правило для энтропии

Теорема 6.1 (Цепное правило)**.**

$$H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y).$$

Более общо:

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i \mid X_1, \ldots, X_{i-1}).$$

6.3 Неравенство: условная энтропия не превосходит безусловную

Теорема 6.2.

$$0 \le H(X \mid Y) \le H(X).$$

Критерии равенств (важно для устного ответа):
- $H(X \mid Y) = H(X) \iff X \perp Y$ (наблюдение Y не даёт информации о X).
- $H(X \mid Y) = 0 \iff X = f(Y)$ (по Y X восстанавливается однозначно).

6.4 Взаимная информация

Определение 6.3 (Взаимная информация)**.**

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) = H(X) + H(Y) - H(X,Y).$$

Утверждение 6.1 (Через KL-дивергенцию)**.**

$$I(X;Y) = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} = D\big(p_{X,Y} \,\big\|\, p_X p_Y\big) \ge 0.$$

Интерпретация (как в лекции про каналы). До наблюдения Y неопределённость о X равна $H(X)$. После наблюдения Y остаётся $H(X \mid Y)$. Разница $I(X;Y)$ — среднее «улучшение знания о X».

7 Теория информации: взаимная информация, пропускная способность и абсолютная стойкость шифров

7.1 Канал, матрица переходных вероятностей и скорость передачи

Определение 7.1 (Дискретный канал)**.** Канал задаётся условными вероятностями $p(y \mid x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. При выборе входного распределения $p(x)$ получаем совместное:

$$p(x,y) = p(x)p(y \mid x),$$

и выходное

$$p(y) = \sum_{x} p(x)p(y \mid x).$$

Определение 7.2 (Скорость передачи информации)**.** В блочном кодировании: передаём k информационных бит за n использований канала:

$$R = \frac{k}{n} \quad \text{(бит/использование канала)}.$$

7.2 Пропускная способность канала

Определение 7.3 (Пропускная способность)**.**

$$C = \max_{p(x)} I(X;Y).$$

Лекционный смысл: C — предельная достижимая скорость надёжной передачи (при $R < C$ возможно $P_e \to 0$; при $R > C$ — невозможно).

7.3 Канал без памяти

Определение 7.4 (Канал без памяти)**.** При n последовательных использованиях:

$$p(y^n \mid x^n) = \prod_{i=1}^{n} p(y_i \mid x_i).$$

7.4 Цепь Маркова и лемма об обработке информации

Определение 7.5 (Марковская цепь)**.** $X \to Y \to Z$ — марковская цепь, если

$$p(z \mid x, y) = p(z \mid y).$$

Теорема 7.1 (Лемма об обработке информации (DPI))**.** *Если $X \to Y \to Z$, то*

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

Интуиция из лекции: «дополнительная обработка/пропуск через следующий канал не может увеличить информацию о первоначальном входе».

7.5 Абсолютная стойкость шифров (шифр Вернама)

Определение 7.6 (Шифр Вернама)**.**

$$C = M \oplus K,$$

где ключ K независим от сообщения M и имеет ту же длину.

Утверждение 7.1 (Абсолютная стойкость)**.** *Если K равномерный и независим от M, то*

$$I(M;C) = 0.$$

Замечание 7.1. Классическое необходимое условие «энтропия ключа не меньше энтропии сообщения»:

$$H(K) \ge H(M)$$

для абсолютной стойкости (в разных формулировках).

8 Пропускная способность и фундаментальные пределы: неравенство Фано и обратная теорема

8.1 Неравенство Фано

Теорема 8.1 (Фано). . ***Чето сложное, мб не обязательно учить***
Пусть $\hat{X}$ — оценка X, $P_e = \mathbb{P}(\hat{X} \neq X)$, $|\mathcal{X}| < \infty$. Тогда

$$H(X \mid \hat{X}) \leq H_2(P_e) + P_e \log_2(|\mathcal{X}| - 1).$$

Как используется. Если P_e мало, то $H(X \mid \hat{X})$ тоже мало: по оценке мы почти восстанавливаем X.

8.2 Обратная теорема кодирования Шеннона (а это важно)

Теорема 8.2 (Обратная теорема (содержательная формулировка)). *Если $R > C$, то не существует последовательности кодов, обеспечивающей $P_e^{(n)} \to 0$ при $n \to \infty$. Иначе говоря, надёжная передача со скоростью выше C принципиально невозможна.*

9 Прямая теорема Шеннона: случайное кодирование, совместная типичность и примеры каналов

9.1 Прямая теорема кодирования Шеннона (achievability)

Теорема 9.1 (Прямая теорема). *Для канала без памяти при любой скорости $R < C$ существуют коды длины n, такие что вероятность ошибки $P_e^{(n)} \to 0$ при $n \to \infty$.*

9.2 Матрица переходных вероятностей

Определение 9.1 (Матрица канала). Для конечных алфавитов канал задаётся таблицей $p(y \mid x)$ (строки x, столбцы y). Для входного $p(x)$ совместное $p(x, y) = p(x)p(y \mid x)$, а выходное $p(y) = \sum_x p(x)p(y \mid x)$.

9.3 Двоичный симметричный канал (BSC)

Определение 9.2 (BSC). $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. Ошибка бита с вероятностью p:

$$p(y \mid x) = \begin{cases} 1 - p, & y = x, \\ p, & y \neq x. \end{cases}$$

Утверждение 9.1 (Пропускная способность BSC).

$$C = 1 - H_2(p).$$

Ключевые шаги (по лекции): $H(Y \mid X) = H_2(p)$ одинаково для всех x, а равномерный вход даёт равномерный выход $H(Y) = 1$.

9.4 Двоичный канал со стиранием (BEC)

Определение 9.3 (BEC)**.** $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, \perp\}$. С вероятностью p стирание $\perp$, иначе передаётся без ошибок.

Утверждение 9.2 (Пропускная способность BEC)**.**

$$C = 1 - p.$$

Интерпретация из лекции: стирание «честно сообщает незнание», поэтому при той же «частоте проблем» BEC может иметь большую ёмкость, чем BSC.

9.5 q-ичный симметричный канал

Определение 9.4 (q-ичный симметричный канал)**.** Алфавит размера q. Символ сохраняется с вероятностью $1 - p$, иначе равновероятно превращается в один из $q - 1$ других (вероятность $p/(q - 1)$).

Утверждение 9.3 (Ёмкость q-ичного симметричного канала)**.**

$$C = \log_2 q - H(\textit{распределение строки}).$$

Эквивалентно часто пишут $C = \log_2 q - H_q(p)$ (с нужной конвенцией логарифма).

10 Пропускная способность гауссовского канала: дифференциальная энтропия и формула Шеннона–Хартли

10.1 Дифференциальная энтропия

Определение 10.1 (Дифференциальная энтропия)**.** Для непрерывной случайной величины с плотностью f:

$$h(X) = -\int_{\mathbb{R}} f(x) \log_2 f(x)\, dx.$$

Замечание 10.1. Дифференциальная энтропия может быть отрицательной и зависит от масштаба (измерения). Это нормально: смысл появляется в разностях вида

$$I(X; Y) = h(Y) - h(Y \mid X).$$

10.2 Гауссовский канал и AWGN

Определение 10.2 (Вещественный AWGN-канал)**.**

$$Y = X + N, \qquad N \sim \mathcal{N}(0, \sigma^2), \qquad \mathbb{E}[X^2] \le P.$$

Определение 10.3 (Комплексный AWGN-канал (как в лекциях про I/Q))**.**

$$Y = C + W, \qquad W \sim \mathcal{CN}(0, N_0), \qquad \mathbb{E}|C|^2 \le E_s.$$

10.3 Максимальная энтропия при фиксированной дисперсии

Теорема 10.1 (Гауссовское распределение максимизирует h)**.** *Среди всех распределений с фиксированной дисперсией* $\mathrm{Var}(X) = \sigma^2$ *дифференциальная энтропия* $h(X)$ *максимальна у гауссовского, и*

$$h(\mathcal{N}(0, \sigma^2)) = \frac{1}{2} \log_2(2\pi e \sigma^2).$$

Это ключевой шаг для вывода ёмкости AWGN.

10.4 Формула Шеннона–Хартли

Теорема 10.2 (Шеннон–Хартли)**.** *Для AWGN-канала в полосе* B *при мощности сигнала* P *и спектральной плотности шума* $N_0/2$*:*

$$C = B \log_2\left(1 + \frac{P}{N_0 B}\right) \quad \textit{бит/с.}$$

В лекциях также появлялась «на символ» форма:

$$C_{\mathrm{sym}} = \log_2(1 + \mathrm{SNR}), \qquad \mathrm{SNR} = \frac{E_s}{N_0}.$$

10.5 Отношение сигнал/шум, спектральная эффективность и предел Найквиста

Определение 10.4 (SNR)**.** Типичные формы:

$$\mathrm{SNR} = \frac{P}{N_0 B} \text{ (по мощности)}, \qquad \mathrm{SNR} = \frac{E_s}{N_0} \text{ (на символ)}.$$

Определение 10.5 (Спектральная эффективность)**.**

$$\eta = \frac{R}{B} \quad \text{бит/с/Гц}.$$

Определение 10.6 (Предел Найквиста (смысл))**.** Число независимых степеней свободы сигнала в полосе B за время T порядка $2BT$ (вещественных). Это даёт фундаментальное ограничение «символов в секунду на герц», которое обсуждалось при переходе к L_2-модели сигналов.

10.6 Гильбертово пространство сигналов и ортогональность

Определение 10.7 (Пространство L_2)**.** Сигналы $x(t)$ рассматриваются как элементы L_2 с внутренним произведением

$$\langle x, y \rangle = \int x(t) y(t)\, dt.$$

Норма $\|x\| = \sqrt{\langle x, x \rangle}$ соответствует корню из энергии.

Определение 10.8 (Ортогональность)**.** Сигналы x, y ортогональны, если $\langle x, y \rangle = 0$.

Ортонормированный базис $\{p_k(t)\}$ позволяет разлагать сигналы

$$x(t) = \sum_k c_k p_k(t),$$

а коэффициенты на приёме получаются согласованной фильтрацией:

$$\hat{c}_k = \langle y, p_k \rangle .$$

Лекционный вывод: физический канал с AWGN сводится к набору независимых гауссовских шумов в координатах базиса.

11 Пропускная способность гауссовского канала и фундаментальный предел Шеннона

11.1 Предел Шеннона и E_b/N_0

Определение 11.1 (Энергия на бит)**.** Если скорость R (бит/с), мощность P, то

$$E_b = \frac{P}{R}.$$

Связь с лекциями: при малых скоростях/малой спектральной эффективности фундаментальный минимум

$$\left(\frac{E_b}{N_0}\right)_{\min} = \ln 2 \approx 0.693 \ (\approx -1.59 \text{ dB}).$$

Это «энергетическая граница Шеннона».

11.2 Геометрическая интерпретация пропускной способности

Идея «упаковки сфер». Кодовые слова длины n в AWGN можно рассматривать как точки в $\mathbb{R}^n$ (или $\mathbb{C}^n$). Шум добавляет случайный вектор; правильное декодирование означает, что шум не «выбил» точку из её области решения. При ML-декодировании области напоминают «шары» вокруг кодовых слов. Ограничение на число кодовых слов связано с тем, сколько маленьких шумовых шаров можно поместить в большой шар допустимых выходов.

11.3 Эффект концентрации меры

Содержание из лекций. В больших размерностях:

- длина гауссовского шумового вектора концентрируется около $\sqrt{n\sigma^2}$;
- объём шара концентрируется в тонком слое у поверхности;
- «почти все» случайные векторы лежат близко к сфере фиксированного радиуса.

Эта геометрия и даёт «твёрдую» асимптотику границы $\log(1 + \text{SNR})$.

12 Основы помехоустойчивого кодирования: гауссовские каналы, мягкие решения и метрики

12.1 Оптимальное решающее правило, ML и MAP

Определение 12.1 (Средняя вероятность ошибки)**.** Для решающего устройства $\hat{X}$ по наблюдению Y:

$$P_e = \mathbb{P}(\hat{X} \neq X).$$

Теорема 12.1 (MAP-правило оптимально по P_e)**.** *Решение*

$$\hat{x}(y) = \arg\max_x p(x \mid y)$$

минимизирует вероятность ошибки P_e.

Следствие 12.1 (ML-правило)**.** *Если априорное $p(x)$ равномерно, то MAP эквивалентно*

$$\hat{x}(y) = \arg\max_x p(y \mid x)$$

(максимум правдоподобия, ML).

12.2 Мягкие и жёсткие решения

Определение 12.2 (Мягкие решения)**.** Декодер использует «насколько вероятно 0/1» (например, LLR), а не только квантованный бит.

Определение 12.3 (Жёсткие решения)**.** Сначала наблюдение Y сводится к дискретному $\tilde{Y} \in \{0, 1\}$ (посимвольный детектор), далее код декодируется в метрике Хэмминга.

Лекционный акцент: мягкие решения обычно дают заметный выигрыш по требуемому E_b/N_0.

12.3 Евклидово расстояние и метрика для AWGN

Для AWGN:

$$p(y \mid c) \propto \exp\left(-\frac{\|y - c\|^2}{2\sigma^2}\right),$$

поэтому ML/MAP (при равномерном априоре по словам) сводится к

$$\hat{c} = \arg\min_{c \in \mathcal{C}} \|y - c\|^2.$$

Это объясняет роль евклидовой метрики для «созвездий» и кодовых слов в гауссовском шуме.

12.4 Расстояние Хэмминга, сферы и шары

Определение 12.4 (Расстояние Хэмминга)**.** $d_H(u, v)$ — число позиций, в которых u и v различны.

Определение 12.5 (Минимальное расстояние кода)**.**

$$d = \min_{c \neq c'} d_H(c, c').$$

Определение 12.6 (Сфера/шар Хэмминга)**.** Число слов на расстоянии ровно i от фиксированного:

$$\binom{n}{i}(q-1)^i.$$

Объём шара радиуса t:

$$V_q(n, t) = \sum_{i=0}^{t} \binom{n}{i}(q-1)^i.$$

Ключевой факт: код с расстоянием d гарантированно исправляет $t = \lfloor (d-1)/2 \rfloor$ ошибок (при ближайшем соседе в Хэмминге).

13 Блоковые коды: границы Хэмминга, Синглтона, Варшамова–Гилберта и введение в линейные коды

13.1 Параметры кода и скорость

Определение 13.1 (Блоковый код и параметры)**.** Код $\mathcal{C} \subset \mathcal{A}^n$ над алфавитом размера q. Параметры (n, M, d, q): длина n, мощность $M = |\mathcal{C}|$, минимальное расстояние d, размер алфавита q.

Определение 13.2 (Скорость и относительное расстояние)**.**

$$R = \frac{1}{n}\log_q M, \qquad \delta = \frac{d}{n}.$$

13.2 Граница Хэмминга (упаковочная)

Теорема 13.1 (Граница Хэмминга)**.** *Пусть код исправляет $t = \lfloor (d-1)/2 \rfloor$ ошибок. Тогда*

$$M \cdot V_q(n, t) \leq q^n.$$

Смысл: шары радиуса t вокруг кодовых слов не пересекаются и лежат в $\mathcal{A}^n$ размера q^n.

13.3 Граница Синглтона

Теорема 13.2 (Синглтон)**.** *Если $M = q^k$, то*

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

Интуиция: удалим $d-1$ координат; разные кодовые слова не могут слиться, иначе расстояние было бы $\leq d-1$. Значит $M \leq q^{n-(d-1)}$.

13.4 Граница Варшамова–Гилберта (существование)

Теорема 13.3 (Варшамов–Гилберт, одна из форм)**.** *Существует q-ичный код длины n и расстояния d мощности порядка*

$$M \gtrsim \frac{q^n}{V_q(n, d-1)}.$$

Смысл: жадно выбираем кодовые слова, каждый раз выкидывая шар радиуса $d-1$; пока место есть, можно выбирать новые слова.

13.5 Введение в линейные коды

Лекционный переход: для конструктивности и эффективного кодирования/декодирования рассматривают линейные коды, где $\mathcal{C}$ — подпространство над $\mathbb{F}_q$.

14 Линейные коды: порождающие и проверочные матрицы, коды Хэмминга и Рида–Соломона

14.1 Линейный код, вес и расстояние

Определение 14.1 (Линейный код)**.** $\mathcal{C} \subset \mathbb{F}_q^n$ — линейное подпространство размерности k. Тогда $|\mathcal{C}| = q^k$.

Определение 14.2 (Вес Хэмминга)**.** $w_H(c) = d_H(c, 0)$. Для линейного кода:

$$d = \min_{c \neq 0} w_H(c).$$

14.2 Порождающая и проверочная матрицы

Определение 14.3 (Порождающая матрица G)**.** Матрица $G \in \mathbb{F}_q^{k \times n}$ такая, что

$$\mathcal{C} = \{uG : \ u \in \mathbb{F}_q^k\}.$$

Определение 14.4 (Проверочная матрица H)**.** Матрица $H \in \mathbb{F}_q^{(n-k) \times n}$ такая, что

$$\mathcal{C} = \{c \in \mathbb{F}_q^n : \ Hc^\top = 0\}.$$

Определение 14.5 (Синдром)**.** Для принятого $r = c + e$:

$$s = Hr^\top = H(c+e)^\top = He^\top,$$

так как $Hc^\top = 0$. Синдром зависит только от ошибки.

Определение 14.6 (Систематический код)**.** Код систематический, если кодовое слово содержит исходные k символов как подблок:

$$G = [I_k \ P].$$

14.3 Код Хэмминга

Определение 14.7 (Двоичный код Хэмминга)**.** Параметры:

$$[2^m - 1,\ 2^m - 1 - m,\ 3]_2.$$

Исправляет 1 ошибку.

Строительство через H. Проверочная матрица H размера $m \times (2^m - 1)$ состоит из всех ненулевых двоичных столбцов длины m. Тогда:

- одиночная ошибка в позиции i даёт синдром, равный i-му столбцу H, то есть её можно однозначно найти;
- минимальное расстояние $d = 3$.

14.4 Коды Рида–Соломона, MDS и матрица Вандермонда

Определение 14.8 (RS-код (идея))**.** Берём поле $\mathbb{F}_q$, выбираем n различных точек $\alpha_1, \ldots, \alpha_n$. Информация — коэффициенты многочлена f степени $< k$. Кодовое слово:

$$c = (f(\alpha_1), \ldots, f(\alpha_n)) \in \mathbb{F}_q^n.$$

Утверждение 14.1 (Параметры RS)**.** *RS-код имеет параметры*

$$[n, k, n - k + 1]_q,$$

то есть является MDS и достигает границы Синглтона.

Определение 14.9 (Матрица Вандермонда)**.** Матрица вида $V_{ij} = \alpha_j^{i-1}$. Порождающая матрица RS-кода может быть записана как (укороченная) матрица Вандермонда по точкам α_j.

15 Свёрточные коды: алгоритмы Витерби, BCJR и принципы турбо-кодирования

15.1 Свёрточные коды и решётчатая диаграмма

Определение 15.1 (Свёрточный код)**.** Кодер имеет внутреннее состояние (память), обычно задаётся сдвиговым регистром длины ν. На каждом шаге принимает входной бит и выдаёт несколько выходных бит (скорость, например, 1/2).

Определение 15.2 (Решётчатая диаграмма (trellis))**.** Граф по времени: вершины — состояния на каждом шаге, рёбра — возможные переходы при входных битах, рёбра маркируются выходными символами и получают метрику/вероятность по наблюдениям канала.

Ключевое отличие из лекции: для свёрточных кодов trellis регулярна и число состояний ограничено 2^ν, что делает ML/MAP-декодирование вычислимым.

15.2 Алгоритм Витерби (ML-декодирование по пути)

Теорема 15.1 (Витерби, рекурсия динамического программирования)**.** *Пусть метрика пути — сумма метрик рёбер. Тогда минимальную метрику до состояния s на шаге t можно считать рекурсивно:*

$$M_{t+1}(s') = \min_{s \to s'} \bigl\{ M_t(s) + m_t(s \to s') \bigr\}.$$

Для каждого s' сохраняется «выживший» предок, затем по backtracking восстанавливается оптимальный путь.

Сложность. Пропорциональна $n \cdot (\#\text{состояний}) \cdot (\#\text{входящих рёбер})$. Поэтому работает для ограниченного числа состояний, но становится нереалистичной для произвольных блочных кодов с огромной решёткой (отмечалось в лекции).

15.3 Алгоритм BCJR (MAP/APP по битам, forward–backward)

Определение 15.3 (BCJR: факторы α, β, γ)**.** Обычно вводят:

- $\alpha_t(s)$: вероятность прийти в состояние s и наблюдать прошлые отсчёты;
- $\beta_t(s)$: вероятность наблюдать будущие отсчёты при условии текущего состояния s;
- $\gamma_t(s \to s')$: вероятность перехода $s \to s'$ (априорная) $\times$ правдоподобие наблюдения на шаге t.

Утверждение 15.1 (BCJR-рекурсии)**.**

$$\alpha_{t+1}(s') = \sum_s \alpha_t(s)\, \gamma_t(s \to s'), \qquad \beta_t(s) = \sum_{s'} \gamma_t(s \to s')\, \beta_{t+1}(s').$$

Оценка бита. Апостериорная вероятность бита b_t получается суммированием по тем рёбрам trellis на шаге t, которые помечены $b_t = 0$ или $b_t = 1$, с весами $\alpha_t(s)\gamma_t(s \to s')\beta_{t+1}(s')$. В лекциях это объяснялось как «message passing» вдоль решётки.

15.4 Турбо-коды и итеративное декодирование

Определение 15.4 (Турбо-код)**.** Состоит из двух (или более) относительно простых свёрточных кодеров, соединённых через интерливинг (перестановку) входных бит.

Определение 15.5 (Итеративное декодирование)**.** Два MAP-декодера (BCJR) по очереди уточняют априорные вероятности битов, обмениваясь «экстиринзик»-информацией; процесс повторяется несколько итераций.

Интуиция (как в лекции). Интерливинг «разрывает» локальные зависимости: биты, соседние в одном кодере, становятся далёкими в другом. Это помогает итеративному message passing лучше приближаться к глобальному MAP.

15.5 Message passing (общий принцип)

Смысл. BCJR — частный случай алгоритма «sum-product» на фактор-графе: локальные сообщения комбинируются и распространяются, что даёт эффективные маргиналы. На деревьях это точно, на графах с циклами (как у турбо/LDPC) — приближённо, но часто очень эффективно.

Финальные замечания(автора)

- Удачи, А.М. Бибиков любит вопросы на понимание, но они простые, +- написать теорию и че-то понять прокатит