# Лекция 1: Введение в теорию информации. Кодирование источника.

Алексей Фролов

al.frolov@skoltech.ru

Сколковский институт науки и технологий (Сколтех)


Современные методы теории информации, оптимизации и управления

Сочи, Россия

2–23 августа, 2020

Skoltech
Skolkovo Institute of Science and Technology

# Outline



Skoltech
Skolkovo Institute of Science and Technology

# Содержание



Skoltech
Skolkovo Institute of Science and Technology

# Telegram канал

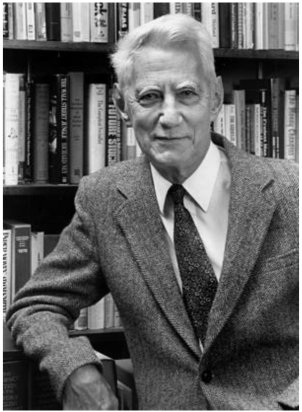
https://t.me/infotheory

Skoltech
Skolkovo Institute of Science and Technology

# Клод Шеннон

*Shannon C. E.* A mathematical theory of communication. — Bell Syst. Tech. J., 1948, v. 27, p. 379–423 (Part I), p. 623–656 (Part II).

[Рус. пер.: *Шеннон К.* Работы по теории информации и кибернетике. М.: ИЛ, 1963, с. 243–332.]



30.04.1916 — 26.02.2001

Skoltech
Skolkovo Institute of Science and Technology

Случайная величина $X$ – функция $X : \Omega \to \mathcal{X}$, $(\Omega, \mathcal{F}, \mathbb{P})$ – вероятностное пространство.

**Определение (Дискретные случайные величины)**

*Случайная величина $X$ является дискретной, если существует не более, чем счетное множество $\mathcal{X} = \{x_j, j = 1, \ldots\}$, такое что*

$$\sum_{j=1}^{\infty} P_X(x_j) = 1.$$

Далее:

- $\mathcal{X}$ – алфавит
- $x \in \mathcal{X}$ – атомы
- $P_X$ – распределение.

Будем писать $P(x)$ вместо $P_X(x)$.

Через $\mathbb{E}[X]$ обозначим мат. ожидание $X$.

Skoltech
Skolkovo Institute of Science and Technology

Рассмотрим две случайные величины $X$ и $Y$ с алфавитами $\mathcal{X}$ и $\mathcal{Y}$. Через $P(x, y)$ (или $P_{X,Y}(x, y)$) обозначим *совместное* распределение $x$ и $y$.

*Маргинальное* распределения $P(x)$ можно получить суммированием:

$$P(x) = \sum_{y \in \mathcal{Y}} P(x, y).$$

Условная вероятность:

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

Независимость:

$$X \perp Y \text{ iff } P(x,y) = P(x)P(y).$$

Цепное правило

$$P(x,y) = P(x|y)P(y).$$

Правило полной вероятности

$$P(x) = \sum_{y \in \mathcal{Y}} P(x|y)P(y).$$

Skoltech
Skolkovo Institute of Science and Technology

# Теорема Байеса

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

Skoltech
Skolkovo Institute of Science and Technology

# Содержание



Skoltech
Skolkovo Institute of Science and Technology

# Энтропия

Как измерить случайность случайной величины $X$?

Энтропия Шеннона $H(p_1, p_2, \ldots, p_n)$. Пусть $p_i = P(x_i)$.
Условия:

- **Непрерывность.** Изменение значений вероятностей на малую величину должно изменить энтропию на небольшую величину.
- **Симметричность.** Например, $H(p_1, p_2, \ldots, p_n) = H(p_2, p_1, \ldots, p_n)$.
- **Максимальное значение.** Максимальное значение должно соответствовать случаю, когда все исходы равновероятны.
- **Аддитивность.** Энтропия не должна зависеть о того, как процесс делится на части

$$H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = H\left(\frac{b_1}{n}, \ldots, \frac{b_k}{n}\right) + \sum_{i=1}^{k} \frac{b_i}{n} H\left(\frac{1}{b_i}, \ldots, \frac{1}{b_i}\right)$$

Skoltech
Skolkovo Institute of Science and Technology

Любая функция, удовлетворяющая данным условиям, имеет вид

$$H(X) = -c \sum_{x \in \mathcal{X}} P(x) \log P(x) = c\mathbb{E}[\log(1/P(X))],$$

где $c$ – это константа. Далее мы полагаем $0 \log 0 = 0$.

$\log_2 \leftrightarrow$ биты
$\ln \leftrightarrow$ наты
$\log_{256} \leftrightarrow$ байты

$$H(X) \geq 0$$

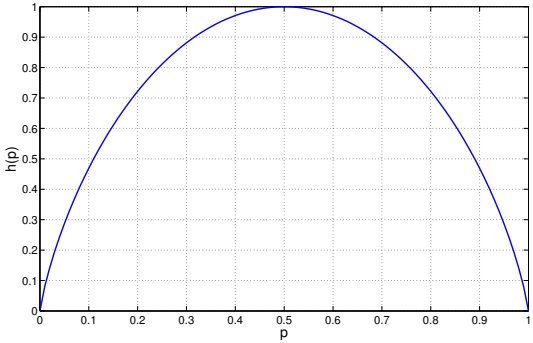
Skoltech
Skolkovo Institute of Science and Technology

# Энтропия

Пусть $X \sim \text{Bern}(p)$. Тогда

$$H(X) = h(p) \triangleq -p \log p - (1-p) \log(1-p),$$

где $h(p)$ – двоичная функция энтропии.




Skoltech
Skolkovo Institute of Science and Technology

**Пример**

$X : \mathcal{X} = \{a, b, c, d\}, P_X = \{1/2, 1/4, 1/8, 1/8\}.$

$$\begin{aligned} H(X) &= -1/2 \log(1/2) - 1/4 \log(1/4) - 2/8 \log(1/8) \\ &= 1.75 \ \textit{бит} \end{aligned}$$

Положим, что мы хотим определить значение $X$ с помощью минимального числа двоичных вопросов. Среднее число таких вопросов – 1.75. Минимальное число вопросов заключено между $H(X)$ и $H(X) + 1$.

Skoltech
Skolkovo Institute of Science and Technology

# Совместная энтропия

**Определение**

*Совместная энтропия $H(X, Y)$ определяется следующим образом*

$$H(X, Y) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P(x, y) \log P(x, y) = -\mathbb{E}[\log P(x, y)].$$

Skoltech
Skolkovo Institute of Science and Technology

**Определение**

*Условная энтропия $H(Y|X)$ определяется следующим образом*

$$\begin{aligned} H(Y|X) &= \sum_{x\in\mathcal{X}} P(x)H(Y|x) \\ &= -\sum_{x\in\mathcal{X}} P(x)\sum_{y\in\mathcal{Y}} P(y|x)\log P(y|x) \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log P(y|x) \\ &= -\mathbb{E}[\log P(Y|X)]. \end{aligned}$$

Skoltech
Skolkovo Institute of Science and Technology

# Цепное правило

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

Skoltech
Skolkovo Institute of Science and Technology

# Относительная энтропия

Относительная энтропия – расстояние междураспределениями $P(x)$ и $Q(x)$.

**Определение**

*Относительная энтропия или расстояние Кульбака–Лейблера между распределениями $P(x)$ и $Q(x)$ определяется как*

$$D(P||Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} = \mathbb{E}\left[\frac{P(X)}{Q(X)}\right].$$

Условимся, что $0 \log \frac{0}{q} = 0$ и $p \log \frac{p}{0} = \infty$.

Является ли данная величина метрикой?

Skoltech
Skolkovo Institute of Science and Technology

**Пример**

$X \sim Bern(r), Y \sim Bern(s)$

$$D(P||Q) = (1-r)\log\frac{1-r}{1-s} + r\log\frac{r}{s}$$

*и*

$$D(Q||P) = (1-s)\log\frac{1-s}{1-r} + s\log\frac{s}{r}$$

*Если* $r = 0.5$ *и* $s = 0.25$, *тогда* $D(P||Q) = 0.2075$ *бит and* $D(Q||P) = 0.1887$ *бит*

Skoltech
Skolkovo Institute of Science and Technology

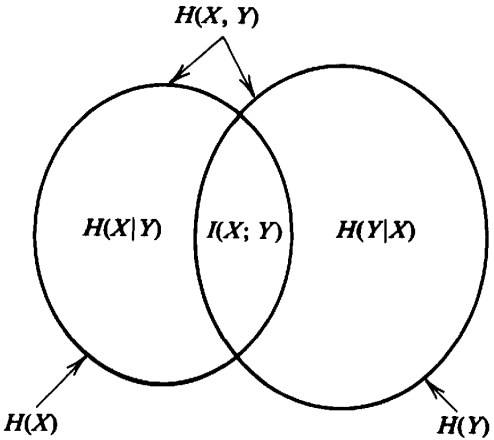
**Определение**

$$
\begin{aligned}
I(X;Y) &= \sum_{x \in cX} \sum_{y \in cY} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \\
&= D(P_{X,Y} || P_X P_Y) \\
&= \mathbb{E}\left[ \log \frac{P(X,Y)}{P(X)P(Y)} \right]
\end{aligned}
$$

Skoltech
Skolkovo Institute of Science and Technology

# Взаимная информация и энтропия

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$




Skoltech
Skolkovo Institute of Science and Technology

# Цепные правила

Энтропия

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i | X_1, \ldots, X_{i-1})$$

Взаимная информация

$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y | X_1, \ldots, X_{i-1})$$

Skoltech
Skolkovo Institute of Science and Technology

# Неравенство Йенсена

**Определение**

*Функция $f(x)$ называется выпуклой на интервале $(a, b)$, если $\forall x_1, x_2 \in (a, b)$ и $\lambda \in [0, 1]$*

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

*Строго выпуклой, если равенство имеет место только в случае $\lambda = 0$ или $\lambda = 1$.*

**Определение**

*$f(x)$ является вогнутой, если $-f(x)$ является выпуклой.*

Skoltech
Skolkovo Institute of Science and Technology

# Неравенство Йенсена

**Теорема**

*Если $f(x)$ – выпуклая функция и $X$ – случайная величина, то*

$$\mathbb{E}[f(X)] \geq f\left(\mathbb{E}[X]\right).$$

Skoltech
Skolkovo Institute of Science and Technology

**Теорема**

$$D(P||Q) \geq 0$$

**Доказательство.**

$$\begin{aligned} -D(P||Q) &= -\sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)} \\ &= \sum_{x\in\mathcal{X}} P(x)\log\frac{Q(x)}{P(x)} \\ &\leq \log\sum_{x\in\mathcal{X}} P(x)\frac{Q(x)}{P(x)} = 0. \end{aligned}$$

□

**Следствие**

$$I(X;Y) \geq 0,$$

$$H(X) \leq \log |\mathcal{X}|,$$

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

$$H(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i).$$

# Неравенство обработки данных

Неравенство обработки данных может быть использовано, чтобы показать, что никакие умные манипуляции с данными не могут улучшить выводы, которые можно сделать из данных.

**Определение**

*Случайные величины $X$, $Y$ и $Z$ образуют цепь Маркова $X \to Y \to Z$, если условное распределение $Z$ зависит только от $Y$ и не зависит от $X$, т.е.*

$$P(z|y,x) = P(z|y).$$

Следствия:

- $P(x,z|y) = P(x|y)P(z|y)$
- $X \to Y \to Z$ означает, что $Z \to Y \to X$
- $Z = f(Y)$, тогда $X \to Y \to Z$

Skoltech
Skolkovo Institute of Science and Technology

# Неравенство обработки данных

**Теорема (Неравенство обработки данных)**

*Если* $X \to Y \to Z$ , *тогда*

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

**Доказательство.**

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

□

$Z = f(Y)$, $I(X;Y) \geq I(X;f(Y))$.

Skoltech
Skolkovo Institute of Science and Technology

# Неравенство Фано

Предположим, что $X$ и $Y$ зависимы, и мы хотим угадать $X$, зная $Y$. Пусть $\hat{X} = g(Y)$. Получаем $X \to Y \to \hat{X}$. Через $P_e$ обозначим вероятность ошибки

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

**Теорема (Неравенство Фано)**

$$h(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|Y)$$

**Доказательство.**

Через $E$ обозначим индтикатор ошибки.

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

# Содержание



Skoltech
Skolkovo Institute of Science and Technology

**Определение**

*$F(x) = \Pr(X \leq x)$ – функция распределения. Если $F(x)$ непрерывна, то случайная величина называется непрерывной. $f(x) = F'(x)$ – плотность распределения $X$. Множество $f(x) > 0$ называется носителем $X$.*

**Определение**

*Дифференциальная энтропия $h(X)$ непрерывной случайной величины $X$ с плотностью распределения $f(x)$ определяется следующим образом*

$$h(X) = -\int_S f(x) \log f(x) dx,$$

*где $S$ – это носитель $X$.*

# Свойства

- $D(f||g) = \int f \log \frac{f}{g} \geq 0$
- $h(X|Y) \leq h(X)$
- $h(aX) = h(X) + \log |a|$
- $I(X;Y) = \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} \geq 0$

Skoltech
Skolkovo Institute of Science and Technology

**Пример**

*Let* $X \sim \phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$.

$$
\begin{aligned}
h(X) &= -\int \phi \ln \phi \\
&= -\int \phi(x) \left[-\frac{x^2}{2\sigma^2} - \ln \sqrt{2\pi\sigma^2}\right] dx \\
&= \frac{\mathbb{E}[X^2]}{2\sigma^2} + \frac{1}{2} \ln 2\pi\sigma^2 \\
&= \frac{1}{2} \ln 2\pi e\sigma^2 \ \textit{nats.}
\end{aligned}
$$

Skoltech
Skolkovo Institute of Science and Technology

**Пример**

*Пусть*

$$\mathbf{X} = [X_1, X_2, \ldots, X_n],$$

*имеет многомерное нормальное распределение со средним*

$$\mu = \mathbb{E}[\mathbf{X}] = [\mathbb{E}[X_1], \mathbb{E}[X_2], \ldots, \mathbb{E}[X_n]]$$

*и матрицей ковариации*

$$K = \mathbb{E}[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T],$$

*т.е.*

$$f(\mathbf{x}) = \frac{1}{\sqrt{2\pi}^n \sqrt{|K|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T K^{-1}(\mathbf{x} - \mu)\right).$$

Skoltech
Skolkovo Institute of Science and Technology

## Пример

$$
\begin{aligned}
h(\mathbf{X}) &= -\int f(\mathbf{x}) \left[ -\frac{1}{2}(\mathbf{x}-\mu)^T K^{-1}(\mathbf{x}-\mu) - \ln \sqrt{2\pi}^n \sqrt{|K|} \right] d\mathbf{x} \\
&= \frac{1}{2}\mathbb{E}\left[ \sum_{i,j} (x_i - \mu_i)^T K_{i,j}^{-1} (x_j - \mu_j) \right] + \frac{1}{2}\ln(2\pi)^n |K| \\
&= \frac{1}{2}\sum_{i,j} \mathbb{E}\left[ (x_i - \mu_i)^T (x_j - \mu_j) \right] K_{i,j}^{-1} + \frac{1}{2}\ln(2\pi)^n |K| \\
&= \frac{1}{2}\sum_{i,j} K_{j,i} K_{i,j}^{-1} + \frac{1}{2}\ln(2\pi)^n |K| \\
&= \frac{n}{2} + \frac{1}{2}\ln(2\pi)^n |K| = \frac{1}{2}\ln(2\pi e)^n |K| \ \ \textit{nats.}
\end{aligned}
$$

# Многомерное нормальное распределение максимизирует дифференциальную энтропию

**Теорема**

*Пусть случайный вектор* $\mathbf{X} \in \mathbb{R}^m$ *имеет нулевое мат. ожиданиеи матрицу ковариации* $K = \mathbb{E}[\mathbf{X}\mathbf{X}^T]$, *т.е.* $K_{i,j} = \mathbb{E}[X_i X_j]$. *Тогда*

$$h(\mathbf{X}) \leq \frac{1}{2} \log(2\pi e)^n |K|$$

*с равенством тогда и только тогда, когда* $\mathbf{X} \sim N(0, K)$.

**Доказательство.**

$$\begin{aligned} 0 \leq D(g||\phi) &= \int g \log \frac{g}{\phi} = -h(g) - \int g \log \phi \\ &= -h(g) - \int \phi \log \phi = -h(g) + h(\phi). \end{aligned}$$

# Содержание



Skoltech
Skolkovo Institute of Science and Technology

Пусть $X_1, X_2, \ldots, X_n$ независимые одинаково распр. сл. величины, тогда

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X] \text{ по вероятности}$$

# Свойство асимптотического выравнивания

Пусть $X_1, X_2, \ldots, X_n$ независимые одинаково распр. сл. величины $\sim p(x)$, тогда

$$-\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) \to H(X) \text{ по вероятности}$$

Skoltech
Skolkovo Institute of Science and Technology

# Типичное множество

$$A_\varepsilon^{(n)} = \{(x_1, x_2, \ldots, x_n)\}$$

$$2^{-n(H(X)+\varepsilon)} \leq p(x_1, x_2, \ldots, x_n) \leq 2^{-n(H(X)-\varepsilon)}$$

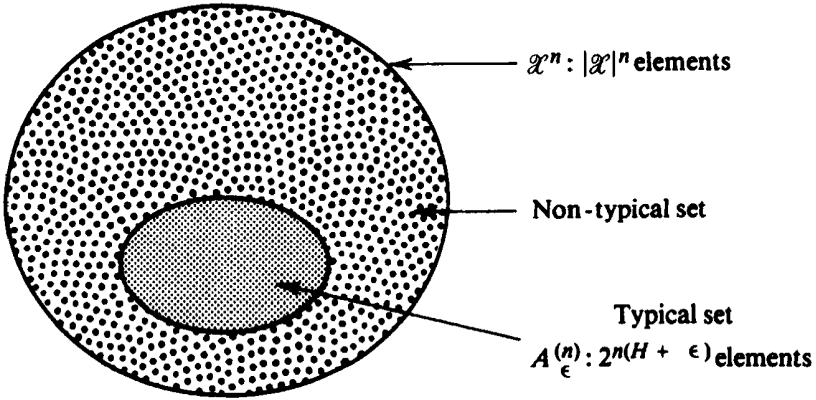
Skoltech
Skolkovo Institute of Science and Technology

# Свойства

- $\Pr(A_\varepsilon^{(n)}) > 1 - \varepsilon$ (для дост. большого $n$)
- $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$
- $|A_\varepsilon^{(n)}| \geq (1-\varepsilon)2^{n(H(X)-\varepsilon)}$ (для дост. большого $n$)

Skoltech
Skolkovo Institute of Science and Technology

# Свойство асимптотического выравнивания




Skoltech
Skolkovo Institute of Science and Technology

**Теорема**

*Пусть $X^n$ независимые одинаково распр. сл. величины $\sim p(x)$. Пусть $\varepsilon > 0$. Тогда для достаточно большого $n$ существует код*

$$\mathbb{E}[l(X^n)/n] \leq H(X) + \varepsilon$$

# Содержание



Skoltech
Skolkovo Institute of Science and Technology

# Характеристики источника

Пусть задан конечный алфавит $\mathcal{X}$ мощности $m$ и источник, который выдает символы $i \in \mathcal{X}$ с вероятностями $p_i$.

Мы предполагаем, что следующий символ появляется независимо от предыдущего.

Skoltech
Skolkovo Institute of Science and Technology

# Главная цель

Мы хотим представить каждую букву начального алфавита как последовательность из другого алфавита (или *кодовое слово*) таким образом, чтобы разные буквы исходного алфавита могли иметь разную длину (*код переменной длины*).

Мы хотим минимизировать среднюю длину последовательности.

Очевидно, что наиболее вероятная буква должна иметь наименьшую длину, и мы должны решить задачу минимизации $\min \sum p_i l_i$. Через $l_i$ мы обозначаем длину последовательности, соответствующей букве $i$.

Skoltech
Skolkovo Institute of Science and Technology

# Код

Код $\mathcal{C}$ для случайной величины $X$ – это отображение из $\mathcal{X}$ в $D^*$.

Пусть $\mathcal{C}(x)$ – это кодовое слово, соотв $x$ и пусть $l(x)$ – это длина $\mathcal{C}(x)$.

Skoltech
Skolkovo Institute of Science and Technology

# Пример

**Пример**

$\mathcal{C}(\textit{Красный}) = 00$, $\mathcal{C}(\textit{Синий}) = 11$ – *это код для* $\mathcal{X} = \{\textit{Красный, Синий}\}$, $D = \{0, 1\}$.

Skoltech
Skolkovo Institute of Science and Technology

# Средняя длина

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

где $l(x)$ – это длина $x$.

**Пример**

$$\begin{aligned}
\Pr(X=1) &= 1/2,\ \mathcal{C}(1)=0 \\
\Pr(X=2) &= 1/4,\ \mathcal{C}(2)=10 \\
\Pr(X=3) &= 1/8,\ \mathcal{C}(3)=110 \\
\Pr(X=4) &= 1/8,\ \mathcal{C}(4)=111
\end{aligned}$$

$L(\mathcal{C}) = H(X) = 1.75$ бит.

Skoltech
Skolkovo Institute of Science and Technology

# Пример

## Пример

$$\begin{aligned} \Pr(X=1) &= 1/3,\ \mathcal{C}(1)=0 \\ \Pr(X=2) &= 1/3,\ \mathcal{C}(2)=10 \\ \Pr(X=3) &= 1/3,\ \mathcal{C}(3)=11 \end{aligned}$$

$L(\mathcal{C}) = 1.66 > H(X) = 1.58$ (бит)

Skoltech
Skolkovo Institute of Science and Technology

# Несингулярные коды

$$x_i \neq x_j \Rightarrow \mathcal{C}(x_i) \neq \mathcal{C}(x_j)$$

Skoltech
Skolkovo Institute of Science and Technology

# Однозначно декодируемые коды

Расширение $\mathcal{C}^*$ кода $\mathcal{C}$ – это отображение из множества строк конечной длины над $\mathcal{X}$ в мноежество строк конечной длины над $D$, определенное след. образом

$$\mathcal{C}(x_1 x_2 \dots x_n) = \mathcal{C}(x_1)\mathcal{C}(x_2)\dots\mathcal{C}(x_n),$$

где $\mathcal{C}(x_1)\mathcal{C}(x_2)\dots\mathcal{C}(x_n)$ – это конкатенация кодовых слов.

Код является однозначно декодируемым, если его расширение несингулярно.

Skoltech
Skolkovo Institute of Science and Technology

# Префиксные коды

Код называется префиксным, если ни одно кодовое слово не является префиксом другого.

Skoltech
Skolkovo Institute of Science and Technology

# Связь между префиксными и одн. декодируемыми кодами

Префиксный код $\Rightarrow$ одн. декодируемый код

Одн. декодируемый код $\not\Rightarrow$ Префиксный код

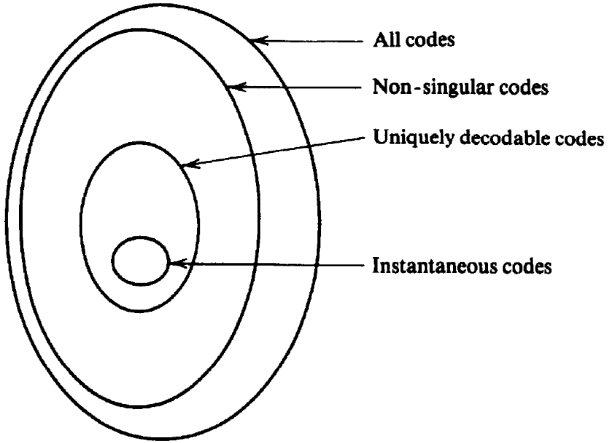
**Пример**

*Код* $C = \{z_1 = 0, z_2 = 10, z_3 = 110, z_4 = 111\}$ *является префиксным и одн. декодируемым.*

*Код* $C = \{z_1 = 0, z_2 = 01, z_3 = 011, z_4 = 111\}$ *не явл префиксным, но явл одн. декодируемым*

Skoltech
Skolkovo Institute of Science and Technology

# Связь между префиксными и одн. декодируемыми кодами




Skoltech
Skolkovo Institute of Science and Technology

# Связь между префиксными и одн. декодируемыми кодами

| X | Singular | Non-singular, but not uniquely decodable | Uniquely decodable, but not instantaneous | Instantaneous |
|---|---|---|---|---|
| 1 | 0 | 0 | 10 | 0 |
| 2 | 0 | 010 | 00 | 10 |
| 3 | 0 | 01 | 11 | 110 |
| 4 | 0 | 10 | 110 | 111 |

Skoltech
Skolkovo Institute of Science and Technology

# Содержание



Skoltech
Skolkovo Institute of Science and Technology

**Теорема**

*Любой префиксный код над алфавитом $D$ с кодовыми словами с длинами $l_1, \ldots, l_m$ соответствует неравенству*

$$\sum_{i=1}^{m} D^{-l_i} \leq 1.$$

*И наоборот, пусть даны длины кодовых слов, соотв. неравенству Крафта, тогда существует префиксный код с этими длинами слов.*

Skoltech
Skolkovo Institute of Science and Technology

**Теорема**

*Любой одн. декодируемый код над алфавитом $D$ с кодовыми словами с длинами $l_1, \ldots, l_m$ соответствует неравенству*

$$\sum_{i=1}^{m} D^{-l_i} \leq 1.$$

*И наоборот, пусть даны длины кодовых слов, соотв. неравенству Крафта, тогда существует одн. декодируемый код с этими длинами слов.*

# Границы для оптимальной средней длины

Оптимизационная задача: минимизировать ($L = \sum p_i l_i$) при условии, что $l_1, l_2, \ldots, l_m$ целые числа и $\sum D^{-l_i} \leq 1$.

**Теорема**

*Пусть $l_1^*, l_2^*, \ldots, l_m^*$ – оптимальрные длины кодовых слов для распределения $P_X$ и пусть $L^* = \sum p_i l_i^*$. Тогда*

$$\frac{H(X)}{\log D} \leq L^* < \frac{H(X)}{\log D} + 1$$

Skoltech
Skolkovo Institute of Science and Technology

**Доказательство.**

Релаксация: пусть $l_i$ вещественные.

Оптимальные длины (метод множителей Лагранжа)

$$l_i = log_D \frac{1}{p_i}$$

и $L^* \geq \sum p_i log_D \frac{1}{p_i} = \frac{H(X)}{\log D}$

Так как необходимо найти целочисленное решение

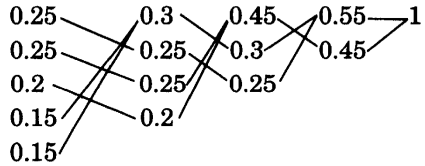
$$l_i = \left\lceil log_D \frac{1}{p_i} \right\rceil$$

и $L^* \leq \sum p_i \left( log_D \frac{1}{p_i} + 1 \right) = \frac{H(X)}{\log D} + 1.$ □

# Код Хаффмана

Ниже мы иллюстрируем алгоритм Хаффмана. Основная идея состоит в том, чтобы объединить наименее вероятные буквы в одну виртуальную букву и построить многоуровневое дерево, которое соответствует коду

| Codeword length | Codeword | $X$ | Probability |
|---|---|---|---|
| 2 | 01 | 1 | 0.25 |
| 2 | 10 | 2 | 0.25 |
| 2 | 11 | 3 | 0.2 |
| 3 | 000 | 4 | 0.15 |
| 3 | 001 | 5 | 0.15 |




This code has average length 2.3 bits.

**Лемма**

*Для любого распределения существует оптимальный префиксный код, который соответствует свойствам:*

- *Если $p_j > p_k$, то $l_j \leq l_k$*
- *Два самых длинных кодовых слова имеют одинаковую длину*
- *Два самых длинных кодовых слова отличаются только последним битом и соответствуют двум наименее вероятным символам*

**Теорема**

*Код Хаффмана оптимален, т.е. если $\mathcal{C}^*$ – это код Хаффмана, а $\mathcal{C}'$ какой-то другой код, то $L(\mathcal{C}^*) \leq L(\mathcal{C}')$.*

Skoltech
Skolkovo Institute of Science and Technology

# Содержание



Skoltech
Skolkovo Institute of Science and Technology

# Practical aspects

Несмотря на оптимальность, код Хаффмана на практике широко не используется. Причина заключается в следующем: для построения кода необходимо знать точные значения вероятностей $p_i$.

Skoltech
Skolkovo Institute of Science and Technology

# Алгоритм Лемпеля-Зива (1977)

Пусть задана строка $1011010100010\ldots$, мы составляем словарь подстрок: $\lambda$, 1, 0, 11, 01, 010, 00, 10, $\ldots$

Skoltech
Skolkovo Institute of Science and Technology

# Алгоритм Лемпеля-Зива (1977)

$n = 13$, вход $= 1011010100010$.

Encoding:

| подстроки | $\lambda$ | 1 | 0 | 11 | 01 | 010 | 00 | 10 |
|---|---|---|---|---|---|---|---|---|
| m | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| pointer | 000 | 001 | 010 | 011 | 100 | 101 | 110 | 111 |
| кодовые слова | | (000, 1) | (000, 0) | (001, 1) | (010, 1) | (100, 0) | (010, 0) | (001, 0) |

Число подстрок $c(n) = 7$
Выход $= 000100000011010110000010000010$.
Длина выхода $c(n)\,(\lceil \log_2 c(n) \rceil + 1)$.

Skoltech
Skolkovo Institute of Science and Technology

# Алгоритм Лемпеля-Зива (1977)

Декодирование:

- Вычислить $c(n)$.
- Прочитать кодовое слово слева направо и обработать пары (указатель, бит).

Skoltech
Skolkovo Institute of Science and Technology

# Универсальное кодирование

$$\frac{1}{n}\mathbb{E}[L(X^n)] \leq H(X) + \frac{\log_2 n}{n}$$

Обратите внимание, что длина полученной последовательности здесь больше по сравнению с оптимальным исходным кодом ($H(X) + \frac{\log_2 n}{n}$ против $H(X) + \frac{1}{n}$). Но в асимптотическом режиме разница незначительна.

Skoltech
Skolkovo Institute of Science and Technology

# Проекты

- Теоретико-информационный анализ нейросетей
- Оценка энтропии многоразмерного датасета
- Сжатие изображений с помощью глубоких нейронных сетей

Skoltech
Skolkovo Institute of Science and Technology

Спасибо за внимание!

Skoltech
Skolkovo Institute of Science and Technology