

Кодирование и декодирование текстов. Равномерные и неравномерные коды

Равномерные (фиксированной длины) и неравномерные (переменной длины) коды, уникальная декодируемость, префиксные коды, неравенство Крафта—Макмиллана, идея кода Хаффмана.

Цели урока

- Различать равномерные и неравномерные коды, понимать их назначение.
- Проверять префиксность и уникальную декодируемость кодов.
- Оценивать среднюю длину кода и сравнивать с энтропией.
- Декодировать поток по словарю; понимать идею кода Хаффмана.

Сценарий видео (7–11 минут)

- 0:00–0:40 Мотивация: почему одинаково кодировать не всегда выгодно; частые буквы vs редкие.
- 0:40–3:00 Базовые определения
 - **Равномерный код** — фиксированная длина слова: $m = \lceil \log_2 N \rceil$.
 - **Неравномерный код** — переменная длина, выгодно для частотных символов.
 - **Префиксный код** — ни одно слово не является префиксом другого → мгновенная декодировка.
 - **Крафт—Макмиллан**: для префиксного кода $\sum 2^{-l_i} \leq 1$.
 - **Средняя длина**: $L_{\text{avg}} = \sum p_i l_i$, l_i ; нижняя граница $H(X) \leq L_{\text{avg}} < H(X) + 1$ для Хаффмана.
- 3:00–5:30 Префиксность и декодирование
 - Как выявить конфликт: префикс, двусмысленность при декодировании.
 - Алгоритм декодирования: идти по дереву; при листе — выводить символ.
- 5:30–7:30 Идея Хаффмана (интуитивно)
 - Сливаем две наименее вероятные вершины, строим дерево, назначаем 0/1 по ребрам.
 - Выигрыш: частым — короткие коды, редким — длинные.

Быстрые примеры

Пример 1: Равномерный код

$N = 40$ символов $\rightarrow m = \lceil \log_2 40 \rceil = 6$ бит

Сообщение 200 символов: $I = 200 \times 6 = 1200$ бит = 150 В.

Пример 2: Неравномерный (префиксный) код

Словарь: $A \rightarrow 0$, $B \rightarrow 10$, $C \rightarrow 110$, $D \rightarrow 111$

Строка 0110111 декодируется как A C D: 0 | 110 | 111.

Пример 3: Крафт—Макмиллан

Длины: 1,2,3,3

Сумма: $2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} = 1 \rightarrow$ возможен префиксный код.

Закрепление: задачи с подробными решениями**1. Равномерный код: объём сообщения**

Условие: алфавит 50 символов, длина 180. Найти объём.

Решение: $m = \lceil \log_2 50 \rceil = 6$; $I = 180 \times 6 = 1080$ бит = 135 В.

Ответ: 135 В.

2. Проверка префиксности

Условие: $A \rightarrow 0$, $B \rightarrow 01$, $C \rightarrow 011$, $D \rightarrow 111$. Является ли код префиксным?

Решение: слово B (01) имеет префикс A (0) \rightarrow **не префиксный**, декодирование неоднозначно.

Крафт-сумма может быть ≤ 1 , но префиксность не гарантируется — это лишь необходимое условие для существования префиксного кода с такими длинами.

3. Декодирование по словарю

Условие: $A \rightarrow 0$, $B \rightarrow 10$, $C \rightarrow 110$, $D \rightarrow 111$. Декодировать 01011110110.

Решение: 0 | 10 | 111 | 10 | 110 \rightarrow A B D B C.

Ответ: A B D B C.

4. Средняя длина и выигрыш

Условие: $p(A, B, C, D) = (0.4, 0.3, 0.2, 0.1)$; код $A \rightarrow 0$, $B \rightarrow 10$, $C \rightarrow 110$, $D \rightarrow 111$.

Решение: $L_{\{avg\}} = 0.4 \cdot 1 + 0.3 \cdot 2 + 0.2 \cdot 3 + 0.1 \cdot 3 = 1.9$ бита.

Равномерный: $m = \lceil \log_2 4 \rceil = 2 \rightarrow$ выигрыш $2 - 1.9 = 0.1$ б/симв.

5. Идея Хаффмана (мини-пример)

Условие: $p(A, B, C, D) = (0.4, 0.2, 0.2, 0.2)$. Построить коды.

Решение (один из вариантов): слить три по 0.2 последовательно; можно получить $A \rightarrow 0$, $B \rightarrow 10$, $C \rightarrow 110$, $D \rightarrow 111$. Коды Хаффмана не единственны, но $L_{\{avg\}}$ минимальна среди префиксных бин. кодов.

Мини-викторина

- Что такое префиксный код? → Ни одно слово не является префиксом другого.
- Зачем нужны неравномерные коды? → Сократить среднюю длину при неравных вероятностях.
- Крафт-сумма для длин (1,2,3,3) равна? → 1.
- Равномерный код для $N=20$ символов имеет длину? → 5 бит.
- Как декодировать поток в префиксном коде? → Идти по дереву до листа.

Конспект (коротко)

- **Равномерный:** $m = \lceil \log_2 N \rceil$ бит/символ.
- **Неравномерный:** переменные длины; эффективен при разных вероятностях.
- **Префиксный:** мгновенная декодируемость; проверка префикса.
- **Крафт—Макмиллан:** $\sum 2^{-l_i} \leq 1$ — существование префиксного кода.
- **Средняя длина:** $L_{\{avg\}} = \sum p_i l_i$, для Хаффмана близка к энтропии.

Домашнее задание (самопроверка)

1. **Задача А:** $N=64$, длина текста 300. Найти объём (равномерный).

Ответ: $m=6$; $I=300 \times 6=1800$ бит = 225 В.

2. **Задача В:** Длины (1,2,3,3,4). Выполнима ли Крафт-сумма?

Решение: $1/2+1/4+1/8+1/8+1/16=1.0625 > 1 \rightarrow$ нет префиксного кода.

3. **Задача С:** Декодировать 0110110111 при словаре $A \rightarrow 0$, $B \rightarrow 10$, $C \rightarrow 110$, $D \rightarrow 111$.

Ответ: A C B C D.

Визуальные подсказки

- **Дерево кода** с листьями A,B,C,D.
- **Таблица префиксности:** примеры допустимых/недопустимых наборов.
- **Гистограмма частот** → короткие/длинные коды.

Подготовлено для урока «Кодирование и декодирование текстов» · Печать: Ctrl/Cmd + P