Лекция 16. Коды, исправляющие одну ошибку. Коды Хэмминга и их свойства. Мощность кода, исправляющего одну ошибку.

Лектор — Селезнева Светлана Николаевна
selezn@cs.msu.ru

факультет ВМК МГУ имени М.В. Ломоносова

Лекции на сайте https://mk.cs.msu.ru

Коды Хэмминга

Рассмотрим один вид кодов, исправляющих одну ошибку.

Они основаны на свойствах представления натуральных чисел в позиционной системе счисления.

Такие коды называются **кодами Хэмминга**.

Вспомогательные множества

Пусть $n \geqslant 3$, $2^{k-1} < n < 2^k$, где $k \in \mathbb{N}$, и $N_n = \{1, 2, \ldots, n\}$.

Любое число $s \in N_n$ можно представить в двоичной системе счисления с k разрядами:

$$s_{k-1} \ldots s_1 s_0,$$

где $s_{k-1}, \ldots, s_1, s_0 \in B$ и $s = \sum_{i=0}^{k-1} s_i \cdot 2^i$.

Для каждого $i = 0, 1, \ldots, k-1$ положим:

$$D_i = \{s \in N_n \mid s_i = 1\}.$$

Другими словами, в D_i содержатся все натуральные числа, не превосходящие n, в двоичном представлении которых i-й разряд равен 1.

Вспомогательные множества

Например, пусть $n = 5$. Тогда $2^2 < n < 2^3$ и

$$N_5 = \{1, 2, 3, 4, 5\}.$$

Рассмотрим представление чисел из N_5 в двоичной системе счисления:

$s \in N_5$	s_2	s_1	s_0
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1

Получаем:

$$\begin{aligned} D_0 &= \{1, 3, 5\}, \\ D_1 &= \{2, 3\}, \\ D_2 &= \{4, 5\}. \end{aligned}$$

Свойства вспомогательных множеств

Предложение 16.1. *Пусть* $D_0, D_1, \ldots, D_{k-1}$ — *введенные выше множества. Тогда*:

1) $2^i \in D_i$ *и* $2^j \notin D_i$ *при* $j \neq i$;

2) $\min\limits_{s \in D_i} s = 2^i$.

Код Хэмминга

Пусть $n \geqslant 3$ и $2^{k-1} < n < 2^k$, где $k \in \mathbb{N}$.

Множество $H \subseteq B^n$ называется **кодом Хэмминга** *порядка* n, если для любого набора $\beta \in H$ верна система уравнений:

$$\begin{cases} \bigoplus\limits_{j \in D_0} \beta_j = 0, \\ \bigoplus\limits_{j \in D_1} \beta_j = 0, \\ \dots \\ \bigoplus\limits_{j \in D_{k-1}} \beta_j = 0, \end{cases} \qquad (*)$$

и, кроме того, H содержит все наборы из B^n, для которых эта система $(*)$ верна.

Код Хэмминга

Пример. Проверим, принадлежат ли коду Хэмминга H порядка 5 слова $\beta_1 = 11100$ и $\beta_2 = 00111$?

Мы уже построили множества D_0, D_1 и D_2:

$$D_0 = \{1, 3, 5\}, \quad D_1 = \{2, 3\}, \quad D_2 = \{4, 5\}.$$

1. Для слова $\beta_1 = 11100$:

$$\begin{aligned} 11100 &: \quad 1 \oplus 1 \oplus 0 = 0, \\ 11100 &: \quad 1 \oplus 1 = 0, \\ 11100 &: \quad 0 \oplus 0 = 0. \end{aligned}$$

Значит, $\beta_1 \in H$.

2. Для слова $\beta_2 = 00111$:

$$\begin{aligned} 00111 &: \quad 0 \oplus 1 \oplus 1 = 0, \\ 00111 &: \quad 0 \oplus 1 = 1, \\ 00111 &: \quad 1 \oplus 1 = 0. \end{aligned}$$

Значит, $\beta_2 \notin H$.

Код Хэмминга

Теорема 16.1. *Пусть* $n \geqslant 3$ *и* $2^{k-1} < n < 2^k$. *Код Хэмминга порядка* n *содержит* 2^{n-k} *слов и исправляет одну ошибку*.

Код Хэмминга

Доказательство. Пусть $H \subseteq B^n$ — код Хэмминга.

1. Сначала найдем число слов в коде H. По определению кода Хэмминга для любого $\beta \in H$ верна система:

$$\begin{cases} \bigoplus\limits_{j \in D_0} \beta_j = 0, \\ \bigoplus\limits_{j \in D_1} \beta_j = 0, \\ \dots \\ \bigoplus\limits_{j \in D_{k-1}} \beta_j = 0. \end{cases} \qquad (*)$$

Код Хэмминга

Доказательство. Перепишем систему $(*)$ в виде:

$$\begin{cases} \beta_{2^0} = \bigoplus\limits_{j\in D_0, j\neq 2^0} \beta_j, \\ \beta_{2^1} = \bigoplus\limits_{j\in D_1, j\neq 2^1} \beta_j, \\ \dots \\ \beta_{2^{k-1}} = \bigoplus\limits_{j\in D_{k-1}, j\neq 2^{k-1}} \beta_j. \end{cases} \qquad (**)$$

Отметим, что $\beta_{2^0}, \beta_{2^1}, \dots, \beta_{2^{k-1}}$ в правых частях системы $(**)$ не встречаются.

Поэтому если заданы $\beta_j \in B$ при $j \in N_n$, $j \neq 2^0, 2^1, \dots, 2^{k-1}$, то $\beta_{2^0}, \beta_{2^1}, \dots, \beta_{2^{k-1}}$ однозначно определяются из системы $(**)$.

Код Хэмминга

Доказательство. Число возможностей задать $\beta_j \in B$ при $j \in N_n$, $j \neq 2^0, 2^1, \ldots, 2^{k-1}$ равно 2^{n-k}.

Каждая из них определяет одно слово из H, а все они — все слова из H.

Значит, $|H| = 2^{n-k}$.

Код Хэмминга

Доказательство. 2. Теперь покажем, что код H исправляет одну ошибку замещения.

Пусть $\beta \in H$, в слове β произошла ошибка в s-м разряде и оно перешло в слово $\beta' \in B^n$.

Отметим, что

$$\beta'_i = \begin{cases} \beta_i, & i \neq s, \\ \bar{\beta}_i & i = s. \end{cases}$$

Пусть в двоичной системе счисления число s, $1 \leqslant s \leqslant n$, записывается как $s_{k-1} \dots s_1 s_0$, где $s_{k-1}, \dots, s_1, s_0 \in B$ и $s = \sum_{i=0}^{k-1} s_i \cdot 2^i$.

Код Хэмминга

Доказательство. Для каждого $i = 0, 1, \ldots, k-1$ рассмотрим *проверочную* сумму:

$$\bigoplus_{j \in D_i} \beta'_j.$$

Возможны два случая.

Код Хэмминга

Доказательство. 1) Если $s_i = 0$, то $s \notin D_i$.

Поэтому $\beta'_j = \beta_j$ для всех $j \in D_i$.

Получаем:

$$\bigoplus_{j \in D_i} \beta'_j = \bigoplus_{j \in D_i} \beta_j = 0.$$

Значит, в этом случае верно:

$$\bigoplus_{j \in D_i} \beta'_j = s_i.$$

Код Хэмминга

Доказательство. 2) Если $s_i = 1$, то $s \in D_i$.

Поэтому $\beta'_j = \beta_j$ для всех $j \in D_i$, $j \neq s$, $\beta'_s = \bar{\beta}_s = \beta_s \oplus 1$.

Получаем:

$$\bigoplus_{j \in D_i} \beta'_j = \left(\bigoplus_{j \in D_i, j \neq s} \beta_j \right) \oplus \beta'_s = \left(\bigoplus_{j \in D_i, j \neq s} \beta_j \right) \oplus (\beta_s \oplus 1) =$$

$$= \left(\bigoplus_{j \in D_i} \beta_j \right) \oplus 1 = 1.$$

Значит, и в этом случае верно:

$$\bigoplus_{j \in D_i} \beta'_j = s_i.$$

Код Хэмминга

Доказательство. Итак, верны равенства:

$$\begin{cases} s_0 = \bigoplus\limits_{j \in D_0} \beta'_j, \\ s_1 = \bigoplus\limits_{j \in D_1} \beta'_j, \\ \dots \\ s_{k-1} = \bigoplus\limits_{j \in D_{k-1}} \beta'_j. \end{cases}$$

Значит, по неправильному слову β' можно найти все $s_0, s_1, \dots, s_{k-1}$ и разряд $s = \sum\limits_{i=0}^{k-1} s_i \cdot 2^i$, в котором произошла ошибка в слове β.

Код Хэмминга

Доказательство. Теперь если $s = 0$, то ошибки не было, а если $s \neq 0$, то можно восстановить правильное слово β:

$$\beta_i = \begin{cases} \beta'_i, & i \neq s, \\ \bar{\beta}'_i & i = s. \end{cases}$$

□

Информационные и проверочные разряды

При рассмотрении кодов Хэмминга обычно в словах разряды с номерами, являющимися степенями двойки, называют **проверочными**, а остальные разряды — **информационными**.

Построение кода Хэмминга

В доказательстве теоремы 16.1 показано, как для заданного n найти код Хэмминга H порядка n.

Построение кода Хэмминга

Пример. Найдем код Хэмминга H порядка 5. Мы знаем:

$$D_0 = \{1, 3, 5\}, \quad D_1 = \{2, 3\}, \quad D_2 = \{4, 5\}.$$

Поэтому:

$$\begin{cases} \beta_1 = \beta_3 \oplus \beta_5, \\ \beta_2 = \beta_3, \\ \beta_4 = \beta_5. \end{cases}$$

Получаем:

β_3	β_5	$\beta \in H$
0	0	00000
0	1	10011
1	0	11100
1	1	01111

Значит,

$$H = \{00000, 10011, 11100, 01111\}.$$

Алгоритмы

Из доказательства теоремы 16.1 можно извлечь алгоритмы кодирования, исправления ошибки и декодирования в коде Хэмминга.

Код Хэмминга

Пусть $n \geqslant 3$, $2^{k-1} < n < 2^k$ и $m = n - k$.

Если H — код Хэмминга порядка n, то H содержит 2^m слов и исправляет одну ошибку.

Поэтому найдется такое разделимое кодирование

$$\varphi_H : A^m \to B^n,$$

что $C_{\varphi_H} = H$.

Кодирование в коде Хэмминга

Опишем *алгоритм кодирования в коде Хэмминга*.

Вход: слово $\alpha \in A^m$, где $m = n - k$, $n \geqslant 3$, $2^{k-1} < n < 2^k$.

Выход: слово $\beta = \varphi_H(\alpha) \in H$, где $\beta \in B^n$.

Кодирование в коде Хэмминга

Описание алгоритма.

1. *Заполнение* **информационных разрядов**.
Для всех $j \in N_n, j \neq 2^0, 2^1, \ldots, 2^{k-1}$ положить:

$$\beta_j = \alpha_{j-\lceil \log_2 j \rceil}.$$

2. *Заполнение* **проверочных разрядов**.
Для всех $i = 0, 1, \ldots, k-1$ положить:

$$\beta_{2^i} = \bigoplus_{j \in D_i, j \neq 2^i} \beta_j.$$

3. Выдать $\beta \in H$.

Окончание описания алгоритма.

Кодирование в коде Хэмминга

Пример. Закодируем в коде Хэмминга слово $\alpha = 0011$.

1. *Заполняем информационные разряды*:

$$\beta_3 = 0,\ \beta_5 = 0, \beta_6 = 1, \beta_7 = 1.$$

Значит, $n = 7$ и $k = 7 - 4 = 3$.

2. *Заполняем проверочные разряды*:

$$\begin{aligned} \beta_1 &= \beta_3 \oplus \beta_5 \oplus \beta_7 = 0 \oplus 0 \oplus 1 = 1, \\ \beta_2 &= \beta_3 \oplus \beta_6 \oplus \beta_7 = 0 \oplus 1 \oplus 1 = 0, \\ \beta_4 &= \beta_5 \oplus \beta_6 \oplus \beta_7 = 0 \oplus 1 \oplus 1 = 0. \end{aligned}$$

3. *Выдаем*: $\beta = 1000011 \in H$.

Исправление ошибки в коде Хэмминга

Опишем *алгоритм исправления ошибки в коде Хэмминга*

Вход: слово $\beta' \in B^n$, полученнное из некоторого слова $\beta \in H$, в котором могла произойти одна ошибка замещения, где $n \geqslant 3$, $2^{k-1} < n < 2^k$.

Выход: слово $\beta \in H$, где $\beta \in B^n$.

Исправление ошибки в коде Хэмминга

Описание алгоритма.

1. *Вычисление* **проверочных сумм**.
Для всех $i = 0, 1, \ldots, k-1$ найти:

$$s_i = \bigoplus_{j \in D_i} \beta'_j,$$

затем положить: $s = \sum_{i=0}^{k-1} s_i \cdot 2^i$.

2. *Исправление ошибки*.
Если $s = 0$, то *ошибки нет*, положить:

$$\beta_j = \beta'_j \text{ при } j = 1, \ldots, n.$$

Если $s \neq 0$, то *ошибка в s-м разряде*, положить:

$$\beta_j = \beta'_j \text{ при } j = 1, \ldots, n, j \neq s \text{ и } \beta_s = \bar{\beta}'_s.$$

3. Выдать $\beta \in H$.

Окончание описания алгоритма.

Исправление ошибки в коде Хэмминга

Пример. Исправим ошибку в слове $\beta' = 1010011$.

1. *Вычисляем проверочные суммы*:

$$\begin{aligned} s_0 &= \beta'_1 \oplus \beta'_3 \oplus \beta'_5 \oplus \beta'_7 = 1 \oplus 1 \oplus 0 \oplus 1 = 1, \\ s_1 &= \beta'_2 \oplus \beta'_3 \oplus \beta'_6 \oplus \beta'_7 = 0 \oplus 1 \oplus 1 \oplus 1 = 1, \\ s_2 &= \beta'_4 \oplus \beta'_5 \oplus \beta'_6 \oplus \beta'_7 = 0 \oplus 0 \oplus 1 \oplus 1 = 0, \end{aligned}$$

поэтому $s = 1 \cdot 2^0 + 1 \cdot 2^1 + 0 \cdot 2^2 = 1 + 2 = 3$.

2. *Исправляем ошибку*: ошибка в 3-м разряде, значит,

$$\beta_3 = \bar{\beta}'_3 = \bar{1} = 0.$$

3. *Выдаем*: $\beta = 1000011 \in H$.

Декодирование в коде Хэмминга

Опишем *алгоритм декодирования в коде Хэмминга*.

Вход: слово $\beta \in H$, где $\beta \in B^n$, $n \geqslant 3$, $2^{k-1} < n < 2^k$.

Выход: слово $\alpha \in A^m$, где $\beta = \varphi_H(\alpha)$, $m = n - k$.

Декодирование в коде Хэмминга

Описание алгоритма.

1. *Вычеркивание* **проверочных разрядов**.
Вычеркнуть в слове β разряды β_j для всех $j = 2^0, 2^1, \ldots, 2^{k-1}$, затем оставшееся слово обозначить α.

2. Выдать $\alpha \in A^m$.

Окончание описания алгоритма.

Декодирование в коде Хэмминга

Пример. Декодируем слово $\beta = 1000011 \in H$.

1. *Вычеркиваем проверочные разряды*:

$$\alpha = \not{1}\,\not{0}0\,\not{0}011.$$

2. *Выдаем*: $\alpha = 0011$.

Мощность кода, исправляющего t ошибок

Напомним, что $M_t(n)$ обозначает наибольшее число кодовых слов в коде C, $C \subseteq B^n$, исправляющем t ошибок замещения.

Мы показали, что

$$\frac{2^n}{S_{2t}(n)} \leqslant M_t(n) \leqslant \frac{2^n}{S_t(n)},$$

где $S_r(n)$ обозначает число наборов в шаре радиуса r из B^n.

Уточним оценку $M_1(n)$ для кодов, исправляющих одну ошибку.

Мощность кода, исправляющего 1 ошибку

Теорема 16.2. *При $n \geqslant 1$ справедливы следующие неравенства:*

$$\frac{2^n}{2n} \leqslant M_1(n) \leqslant \frac{2^n}{n+1}.$$

Мощность кода, исправляющего 1 ошибку

Доказательство. 1. *Верхняя оценка*. Известно, что

$$M_1(n) \leqslant \frac{2^n}{S_1(n)}.$$

Заметим, что $S_1(n) = n + 1$.

Поэтому

$$M_1(n) \leqslant \frac{2^n}{n+1}.$$

Мощность кода, исправляющего 1 ошибку

Доказательство. 2. *Нижняя оценка*. Если $n \leqslant 2$, то оценка верна. Поэтому пусть $n \geqslant 3$.

Сначала пусть $2^{k-1} < n < 2^k$, тогда $k = \lceil \log_2 n \rceil$.

Рассмотрим код Хэмминга H порядка n.
Он содержит 2^{n-k} слов и исправляет одну ошибку.

Значит,

$$M_1(n) \geqslant |H| = 2^{n-k}.$$

Получаем:

$$M_1(n) \geqslant 2^{n-k} = \frac{2^n}{2^{\lceil \log_2 n \rceil}} \geqslant \frac{2^n}{2^{1+\log_2 n}} = \frac{2^n}{2n}.$$

Мощность кода, исправляющего 1 ошибку

Доказательство. 2. *Нижняя оценка*.

Теперь пусть $n = 2^k$, тогда $k = \log_2 n$.

Рассмотрим код Хэмминга H порядка $(n-1)$ и построим по нему код $C \subseteq B^n$, добавив к каждому кодовому слову из H в конце 0.
Код C содержит 2^{n-1-k} слов и исправляет одну ошибку.

Значит,

$$M_1(n) \geqslant |C| = 2^{n-1-k} = \frac{2^n}{2^{1+\log_2 n}} = \frac{2^n}{2n}.$$

□

Увеличение длины слов в коде Хэмминга

Пусть $n \geqslant 3$, $2^{k-1} < n < 2^k$, где $k = \lceil \log_2 n \rceil$, и $m = n - k$.

Закодируем слово $\alpha \in A^m$ в коде Хэмминга: $\beta = \varphi_H(\alpha) \in H$, где $\beta \in B^n$.

Посмотрим, насколько увеличивается длина слова при кодировании в коде Хэмминга:

$$\frac{|\beta|}{|\alpha|} = \frac{n}{m} = 1 + O\left(\frac{\log_2 n}{n}\right).$$

Т. е. при росте n отношение длины кода слова к длине слова стремится к единице.

Линейность кода Хэмминга

Предложение 16.2. *Код Хэмминга $H \subseteq B^n$ является линейным кодом, $n \geqslant 3$.*

Доказательство. Если H — код Хэмминга, то он содержит в точности наборы из B^n, являющиеся решениями однородной системы линейных уравнений $(*)$.

Множество решений любой однородной системы линейных уравнений над полем является линейным пространством над этим полем.

Следовательно, H — линейное пространство, а значит, H — линейный код.

□

Коды, исправляющие ошибки

Коды Хэмминга применяются, в частности, в некоторых типах цифровой памяти и при хранении данных.

Рассматриваются и применяются другие виды кодов, исправляющих ошибки, например:

– коды Рида-Маллера, которые основаны *на представлении функций алгебры логики в виде полиномов Жегалкина*;

– коды Боуза-Чоудхури-Хоквингема (БЧХ-коды), которые основаны *на свойствах конечных полей*;

а также другие.

Задачи для самостоятельного решения

1. Покажите, что множество $D_i \subseteq N_n$ можно построить следующим образом:

начать с числа 2^i и, пока не закончится множество N_n, повторить: 2^i последовательных чисел включить в D_i, затем 2^i последовательных чисел пропустить.

Например, пусть $n = 10$ и $N_{10} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

Для D_1 начинаем с числа $2^1 = 2$ и далее включаем по два последовательных числа и пропускаем по два последовательных числа, пока не достигнем числа 10:

$$D_1 = \{2, 3, 6, 7, 10\}.$$

Задачи для самостоятельного решения

Проверочной матрицей кода Хэмминга порядка n называется матрица H_n из нулей и единиц размера $k \times n$, в которой столбцами являются числа от 1 до n, записанные сверху вниз в двоичной системе счисления:

$$H_n = (1 \mid 2 \mid \ldots \mid n)\,.$$

Например, проверочная матрица H_5:

$$H_5 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Задачи для самостоятельного решения

2. Покажите, как можно найти множества $D_0, D_1, \ldots, D_{k-1}$ по проверочной матрице H_n.

3. Опишите алгоритм вычисления проверочных разрядов $\beta_{2^0}, \beta_{2^1}, \ldots, \beta_{2^{k-1}}$ с помощью проверочной матрицы H_n.

4. Опишите алгоритм исправления ошибки в слове β' с помощью проверочной матрицы H_n.

Литература к лекции

1. Алексеев В. Б. Лекции по дискретной математике. М.: Инфра-М, 2012. С. 58–60.

2. Яблонский С. В. Введение в дискретную математику. М.: Высшая школа, 2001. С. 288–296.

3. Гаврилов Г. П., Сапоженко А. А. Задачи и упражнения по дискретной математике. М.: Физматлит, 2004. Гл. VII 3.21, 3.22, 3.23.