Занятие 10. Коды, исправляющие ошибки. Линейные коды. Коды Хэмминга.

Селезнева Светлана Николаевна
selezn@cs.msu.ru

факультет ВМК МГУ имени М.В. Ломоносова

Страница курса на сайте http://mk.cs.msu.ru

Для разбора домашнего задания

Ошибки

Пусть φ — кодирование из A в B, $\alpha \in A^*$ и $\beta = \varphi(\alpha) \in B^*$.

Предположим, что закодированное сообщение β хранится в памяти компьютера или передается по каналу связи.

При этом в нем могут произойти ошибки, и слово β перейдет в слово β'.

Можно ли по слову β' установить, что оно с ошибками?

А можно ли по слову β' восстановить слово β?

Ошибки замещения

Пусть $A = B = \{0, 1\}$. Будем рассматривать ошибки замещения, т. е. если 0 может заменяться на 1, а 1 — на 0.

Пусть задано число t, $t \geqslant 1$. Можно ли построить разделимый код, устойчивый к t ошибкам замещения?

Код, обнаруживающий t ошибок

Код $C_\varphi \subseteq B^*$ назовем **обнаруживающим t ошибок**, если для любого слова $\beta \in C_\varphi$ выполняется следующее условие:

если в слове β произойдет не более t ошибок замещения и при этом оно перейдет в слово β', то по неправильному слову β' можно установить, что ошибки были.

Код, исправляющий t ошибок

Код $C_\varphi \subseteq B^*$ назовем **исправляющим t ошибок**, если для любого слова $\beta \in C_\varphi$ выполняется следующее условие:

если в слове β произойдет не более t ошибок замещения и при этом оно перейдет в слово β', то по неправильному слову β' можно:

1) установить, что ошибки были;

2) в случае, когда ошибки были, восстановить правильное слово β.

Слова и наборы

Ошибки замещения не меняют длину слов, поэтому будем рассматривать коды $C_\varphi \subseteq B^*$, в которых всех слова одинаковой длины n.

Далее иногда будем взаимозаменять понятия слова длины n в алфавите B и набора длины n из B^n.

Поэтому код C_φ можно рассматривать как подмножество множества B^n, т. е. $C_\varphi \subseteq B^n$.

Равномерные коды

Пусть $B = \{0, 1\}$, $n \geqslant 1$ и $C \subseteq B^n$.

Множество C назовем **равномерным кодом**.

Если $\beta \in C$, то β назовем **кодовым словом**.

Шары в множестве B^n

Расстоянием $\rho(\alpha, \beta)$ между наборами $\alpha, \beta \in B^n$ называют число разрядов, в которых они отличаются.

Шаром радиуса r, $r \geqslant 0$, с центром в точке $\alpha \in B^n$ называется множество:

$$S_r(\alpha) = \{\beta \in B^n \mid \rho(\alpha, \beta) \leqslant r\}.$$

Т. е. шар $S_r(\alpha)$ содержит в точности все такие наборы $\beta \in B^n$, которые от набора α находятся на расстоянии не более r.

Если $S_r(n)$ обозначает число наборов в шаре радиуса r в B^n, то

$$S_r(n) = \sum_{k=0}^{r} C_n^k,$$

где C_n^k — биномиальный коэффициент из n по k.

Кодовое расстояние

Пусть $B = \{0, 1\}$, $n \geqslant 1$ и $C \subseteq B^n$ — равномерный код.

Кодовым расстоянием кода C назовем величину

$$d_C = \min_{\beta_1, \beta_2 \in C, \beta_1 \neq \beta_2} \rho(\beta_1, \beta_2).$$

Т. е. кодовое расстояние кода C равно наименьшему расстоянию между различными его кодовыми словами.

Коды, обнаруживающие и исправляющие t ошибок

Теорема. *Пусть* $B = \{0, 1\}$ *и* $C \subseteq B^n$ *— равномерный код,* $n \geqslant 1$. *Код* C *обнаруживает* t *ошибок замещения тогда и только тогда, когда* $d_C \geqslant t + 1$.

Теорема. *Пусть* $B = \{0, 1\}$ *и* $C \subseteq B^n$ *— равномерный код,* $n \geqslant 1$. *Код* C *исправляет* t *ошибок замещения тогда и только тогда, когда* $d_C \geqslant 2t + 1$.

Гл. 7, 3.20

Гл. 7, 3.20. Сколько ошибок может обнаружить и сколько ошибок может исправить код

$$C = \{00000,\ 10011,\ 11100,\ 01111\}?$$

Гл. 7, 3.20

Гл. 7, 3.20. Сколько ошибок может обнаружить и сколько ошибок может исправить код

$$C = \{00000,\ 10011,\ 11100,\ 01111\}?$$

Решение. Найдем кодовое расстояние кода C:

$$d_C = \min_{\beta_1,\beta_2 \in C, \beta_1 \neq \beta_2} \rho(\beta_1, \beta_2) = 3.$$

Гл. 7, 3.20

Гл. 7, 3.20. Сколько ошибок может обнаружить и сколько ошибок может исправить код

$$C = \{00000,\ 10011,\ 11100,\ 01111\}?$$

Решение. Найдем кодовое расстояние кода C:

$$d_C = \min_{\beta_1,\beta_2 \in C, \beta_1 \neq \beta_2} \rho(\beta_1, \beta_2) = 3.$$

1. Если код C обнаруживает t ошибок, то

$$d_C = 3 \geqslant t + 1.$$

Поэтому $t \leqslant 2$, т. е. код C может обнаружить не более двух ошибок.

Гл. 7, 3.20

Гл. 7, 3.20. Сколько ошибок может обнаружить и сколько ошибок может исправить код

$$C = \{00000,\ 10011,\ 11100,\ 01111\}?$$

Решение. Найдем кодовое расстояние кода C:

$$d_C = \min_{\beta_1,\beta_2 \in C, \beta_1 \neq \beta_2} \rho(\beta_1, \beta_2) = 3.$$

1. Если код C обнаруживает t ошибок, то

$$d_C = 3 \geqslant t + 1.$$

Поэтому $t \leqslant 2$, т. е. код C может обнаружить не более двух ошибок.

2. Если код C исправляет t ошибок, то

$$d_C = 3 \geqslant 2t + 1.$$

Поэтому $t \leqslant 1$, т. е. код C может исправить не более одной ошибки.

Для самостоятельного разбора: гл. 7, 3.20

Гл. 7, 3.20(1, 2). Сколько ошибок может обнаружить и сколько ошибок может исправить код $C = N_f \subseteq B^n$:

1) $f(x_1, \ldots, x_n) = x_1 \oplus \ldots \oplus x_n$;

2) $f(x_1, \ldots, x_n) = \bar{x}_1 \bar{x}_2 \ldots \bar{x}_n \vee x_1 x_2 \ldots x_n$.

Для решения задач

Линейный код

Пусть $n \geqslant 1$ и $C \subseteq B^n$ — равномерный код.

Код C называется **линейным**, если множество C является линейным пространством.

Кодовое расстояние линейного кода

Для набора $\beta \in B^n$ его **весом** $|\beta|$ называется число разрядов, равных единице.

Теорема. *Если $C \subseteq B^n$ — линейный код, $n \geqslant 1$, то для его кодового расстояния d_C верно равенство:*

$$d_C = \min_{\beta \in C, \beta \neq (0,\ldots,0)} |\beta|.$$

Линейный код, порожденный матрицей

Пусть H — матрица размера $k \times n$ из нулей и единиц со строками $\beta_1, \ldots, \beta_k \in B^n$.

Кодом, порожденным матрицей H назовем линейный код $C(H)$, где

$$C(H) = \{c_1\beta_1 \oplus \ldots \oplus c_k\beta_k \in B^n \mid c_1, \ldots, c_k \in B\}.$$

Для самостоятельного разбора: гл. 7, 4.7

Гл. 7, 4.7(б, г). Найти кодовое расстояние d_C линейного кода $C = C(H) \subseteq B^n$:

б)

$$H = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix};$$

г)

$$H = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

Для решения задач

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

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:

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

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

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

Информационные и проверочные разряды

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

Кодирование в коде Хэмминга

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

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

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*: $\alpha = 0011$.

Для самостоятельного разбра: гл. 7, 3.21, 3.22

Гл. 7, 3.21(1, 3, 5). Закодировать слово $\alpha \in B^m$ в коде Хэмминга:

1) $\alpha = 010$;

3) $\alpha = 1001$;

5) $\alpha = 10101011$.

Гл. 7, 3.22(1, 5, 8). Исправить ошибку в слове $\beta' \in B^n$ и декодировать слово $\beta \in B^n$:

1) $\beta' = 110$;

5) $\beta' = 0101101$;

8) $\beta' = 11011100110$.

Для решения задач

Домашнее задание

По задачнику: Гаврилов Г. П., Сапоженко А. А. Задачи и упражнения по дискретной математике. М.: Физматлит, 2004.

Гл. 7: 3.20(3, 4), 3.21(2, 4, 6), 3.22(2, 6, 9), 3.24(1), 4.7(в, д).