Лекция 9

Теория кодирования

Рассмотрим один из примеров приложения теории групп – кодирование.

Обозначения

A – множество сообщений.

B – множество кодовых слов.

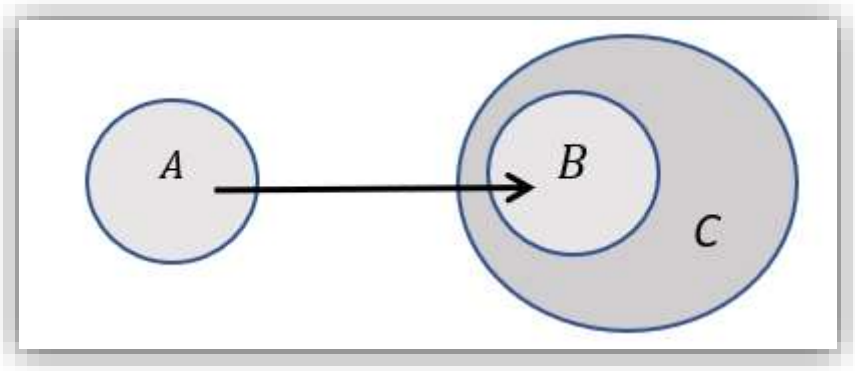
(m, n)-код, m – длина сообщения, n – длина кодового слова. $m < n$

C – множество двоичных слов длины n.

Слова из рассматриваемых множеств будем записывать в алфавите $\{0; 1\}$.

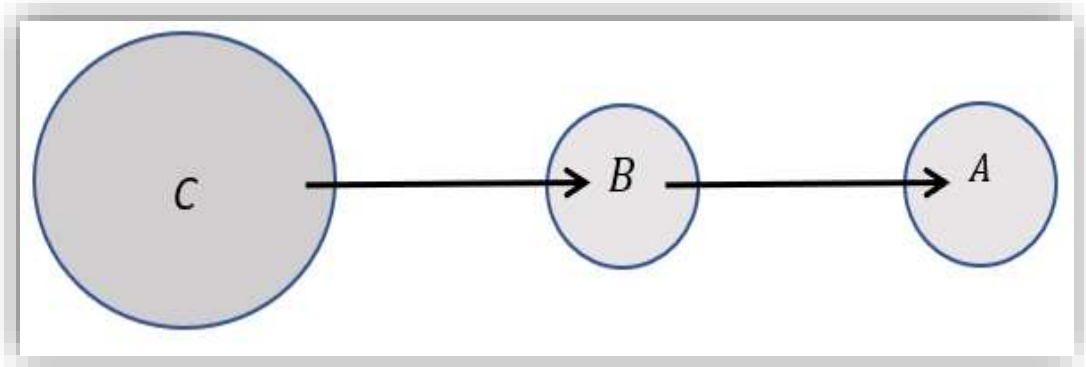
Кодирование

Обычно сообщениям из множества A по определенному правилу приписываются дополнительные символы. В результате получаем кодовые слова B.



Декодирование

В принятом слове из C распознают и исправляют ошибки, получают кодовое слово из B, а затем исходное сообщение из A.



Пример 1. Сообщения длины 2 – множество двоичных слов длины 2.

$$a^1 = 00, \quad a^2 = 01, \quad a^3 = 10, \quad a^4 = 11$$

1. Закодируем проверкой четности, добавив символ – сумма цифр по mod 2:

   $$b^1 = 000, \ b^2 = 011, \ b^3 = 101, \ b^4 = 110$$

2. Закодируем двукратным повторением и проверкой четности:

   $$b^1 = 00000, b^2 = 01011, b^3 = 10101, b^4 = 11110$$

Расстояние Хэмминга

Определение 1. **Расстоянием Хэмминга** между словами b^1 и b^2: $d(b^1, b^2)$ называется число поразрядных несовпадений символов b^1, b^2.

Такое расстояние удовлетворяет аксиомам метрики.

1. $d(b^1, b^2) \geq 0$, причем $d(b^1, b^2) = 0$ тогда и только тогда, когда $b^1 = b^2$;

2. $d(b^1, b^2) = d(b^2, b^1)$;

3. $d(b^1, b^2) + d(b^2, b^3) \geq d(b^1, b^3)$ для любых b^1, b^2, b^3

   (неравенство треугольника).

Введем на множестве двоичных слов операцию поразрядного сложения по $mod\ 2$.

И определим вес слова $w(b^1)$, как число единиц в данном слове. Тогда

$$w(b^1 + b^2) = d(b^1, b^2).$$

Пример 2.

Расстояние между словами $b^1 = 00110010$ и $b^2 = 01010010$ $d(b^1, b^2) = 2$.

$$w(b^1) = 3.$$

$$\begin{cases} b^1 = 00110010 \\ + \\ b^2 = 01010010 \\ \hline b^1 + b^2 = 01100000 \end{cases}$$

$$w(b^1 + b^2) = 2$$

$$d(b^1, b^2) = w(b^1 + b^2)$$

Приведем теоремы об обнаружении и исправлении ошибок.

Теорема 1. Для того чтобы обнаружить ошибки в k позициях кодового слова, необходимо и достаточно, чтобы минимальное расстояние между словами было не менее $k+1$.

Теорема 2. Для того чтобы исправить ошибки в k позициях кодового слова, необходимо и достаточно, чтобы минимальное расстояние между кодовыми словами было не менее $2k+1$.

Доказательство теоремы 2. (Теорема 1 доказывается аналогично).

Пусть при передаче некоторого кодового слова b^1 произошли ошибки в er позициях, и было получено двоичное слово c. Пусть b^2 произвольное кодовое слово. По условию теоремы $d(b^1, b^2) \geq 2k+1$. Из неравенства треугольника, имеем

$$d(b^1, c) + d(c, b^2) \geq d(b^1, b^2) \geq 2k+1$$

$d(b^1, c) = er$. Следовательно,

$$\underbrace{d(b^1, c)}_{er \text{ ошибок}} + \underbrace{d(\mathrm{c}, b^2)}_{er2 \text{ ошибок}} \geq 2k+1$$

Если $er = k$, то $er2 \geq k+1$. И ближайшее кодовое слово к c будет b^1, которое и выберем при декодировании.

Если расстояние между кодовыми словами меньше, хотя бы k, то

$$d(b^1, c) + d(c, b^2) \geq d(b^1, b^2) = 2k \text{ и } \underbrace{d(b^1, c)}_{er \text{ ошибок}} + \underbrace{d(\mathrm{c}, b^2)}_{er2 \text{ ошибок}} \geq 2k,$$

то при $er = k$ будет и $er2 = k$. Тогда при декодировании выбор кодового слова неоднозначен: b^1 или b^2. В этом случае не удастся правильно распознать и исправить ошибки.

Матричное кодирование

$$\boldsymbol{b = aG}$$

G – порождающая матрица кода размерности $m \times n$. В G содержится (желательно) единичная матрица E (позволяет сохранять символы исходного сообщения).

Утверждение 1. Множество двоичных слов C длины n (всего 2^n слов) с операцией поразрядного сложения по $mod\ 2$ образует группу.

Доказательство утверждения 1.

1) ассоциативность операции $(+mod\ 2)$ очевидно выполняется;

2) $e = 00\ldots00 -$ единичный элемент;

3) для любого двоичного слова существует обратный элемент:

$c = c^{-1}$, т. к. $c + c^{-1} = 00\ldots00 = e$

Утверждение 2. При матричном кодировании кодовые слова B длины n образуют подгруппу всех двоичных слов C длины n.

Доказательство утверждения 2. Напомним, что при матричном кодировании $b = aG$.

1) ассоциативность операции $(+mod\ 2)$ очевидна;

2) $e = 00\ldots00$ принадлежит множеству кодовых слов B;

3) $b^{-1} = b$, т. к. $b + b^{-1} = 00\ldots00 = e$;

4) замкнутость $b^1 + b^2 = a^1G + a^2G = (a^1 + a^2)G -$ кодовое слово. Дистрибутивность.

Пример 3.

Закодируем сообщения длины три с помощью заданной порождающей матрицы кода G (первые три столбца – единичная матрица, сохраняет сообщение)

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

По формуле $b = aG$ получим подгруппу B всех кодовых слов длины шесть.

$$\boldsymbol{a^1 = 000 \to b^1 = 000000}$$

$$\boldsymbol{a^2 = 001 \to b^2 = 001111}$$

$$\boldsymbol{a^3 = 010 \to b^3 = 010101}$$

$$\boldsymbol{a^4 = 011 \to b^4 = 011010}$$

$$\boldsymbol{a^5 = 100 \to b^5 = 100110}$$

$$\boldsymbol{a^6 = 101 \to b^6 = 101001}$$

$$\boldsymbol{a^7 = 110 \to b^7 = 110011}$$

$$\boldsymbol{a^8 = 111 \to b^8 = 111100}$$

Групповые коды

Определение 2. Код называется групповым, если множество кодовых слов B образует подгруппу всех двоичных слов C длины n.

При матричном кодировании $b = aG$ получаем групповой код (утверждение 2).

Докажем следующую очень важную теорему.

Теорема 3. В групповом коде наименьшее расстояние d_{min} между различными кодовыми словами равно наименьшему весу w_{min} ненулевого кодового слова.

Доказательство теоремы 3.

Пусть $d_{min} = d(b^1, b^2)$. Так как $d(b^1, b^2) = w(b^1 + b^2)$, то $d_{min} = w(b^1 + b^2)$. d_{min} равно весу некоторого кодового слова, значит $d_{min} \geq w_{min}$.

Пусть w_{min} равно весу некоторого кодового слова b: $w_{min} = w(b)$. Так как $w(b) = d(b, 0)$, то $w_{min} = d(b, 0)$, т.е. расстоянию между кодовыми словами. Тогда $w_{min} \geq d_{min}$.

Окончательно получаем $d_{min} = w_{min}$. Ч.т.д.

Рассмотрим групповой код, в котором процесс декодирования происходит с помощью нахождения смежных классов.

Алгоритм декодирования при групповом кодировании

Пусть получено сообщение $c = b^j + er^i$, где b^j – кодовое слово, er^i – ошибка.

1. Строим таблицу смежных классов группы C по подгруппе B. В каждой строке таблицы смежные классы $er^i + B$, первый элемент er^i ошибка (лидер) – слово, имеющее наименьший вес.
2. Находим в таблице полученное слово $c = b^j + er^i$ и выбираем соответствующее ему кодовое слово b^j из данного столбца и первой строки таблицы.

Рассмотрим множество кодовых слов из примера 3. Всего двоичных слов длины 6 будет $2^6 = 64$. Подгруппа кодовых слов содержит 8 элементов. Следовательно, число смежных классов $i = \frac{|C|}{|B|} = \frac{64}{8} = 8$. Составим таблицу смежных классов и выделим в них ошибку – двоичное слово с одной единицей. На последний класс придется взять слово с двумя единицами, поэтому процесс декодирования для этих слов может быть неверным. Групповой код может исправить одну ошибку, т.к. расстояние между кодовыми словами в

данном случае (иначе надо добавлять столбцы в порождающую матрицу кода) равно 3.

ТАБЛИЦА смежных классов

$er \setminus B$	000000	001111	010101	011010	100110	101001	110011	111100
000000	000000	001111	010101	011010	100110	101001	110011	111100
000001	000001	001110	010100	011011	100111	101000	110010	111101
000010	000010	001101	010111	011000	100100	101011	110001	111110
000100	000100	001011	010001	011110	100010	101101	110111	111000
001000	001000	000111	011101	010010	101110	100001	111011	110100
010000	010000	011111	000101	001010	110110	111001	100011	101100
100000	100000	101111	110101	111010	000110	001001	010011	011100
110000	110000	111111	100101	101010	010110	011001	000011	001100

Пусть получено слово $c = 011101$, тогда $c = b^j + er^i = 010101 + 001000$. Ошибка в третьем символе. Кодовое слово $b^j = 010101$, исходное сообщение $a^j = 010$.

Заметим, что декодировать любое слово в этом примере мы не сможем. По теореме 3 минимальное расстояние между кодовыми словами в этом примере равно 3 ($d_{min} = w_{min}$, минимальный вес кодового слова), т.е. можно исправить одну ошибку. Если получено двоичное слово из последнего класса, то распознать ошибки нельзя. Последовательность слов в этом классе не определена, т.к. в качестве лидера можно взять и другое слово. Да и ошибок не менее двух.

Если при кодировании сообщений длины 4 мы возьмем порождающую матрицу кода размерности 4×7, то кодовых слов будет $2^4 = 16$, а двоичных слов длины 7 будет $2^7 = 128$. Следовательно, число смежных классов $i = \frac{|C|}{|B|} = \frac{2^7}{2^4} = 8$. А это как раз число слов-лидеров с одной 1 плюс нулевое слово. Таким образом, если $i = \frac{|C|}{|B|} = \frac{2^n}{2^m} = 2^{n-m} = n + 1$, то групповой код исправляет ровно одну ошибку.

Код Хэмминга

Матричный $(\boldsymbol{m}, \boldsymbol{n})$-код, удовлетворяющий условиям

$$\boldsymbol{m = 2^r - r - 1, n = 2^r - 1,\ r > 1}$$

Например, рассмотрим $(4, 7)$-код: $r = 3 \quad m = 4 \quad n = 7$

или $(11, 15)$-код $r = 4 \quad m = 11 \quad n = 15.$

Заметим, что если бы мы декодировали с помощью смежных классов, то

$$i = \frac{|C|}{|B|} = \frac{2^n}{2^m} = \frac{2^{2^r-1}}{2^{2^r-r-1}} = 2^r = n + 1.$$

Такой групповой год исправляет ровно одну ошибку. По теореме 2 это означает, что минимальное расстояние между кодовыми словами в таком коде равно 3 (в частности в коде Хэмминга).

Код Хэмминга устроен так, что выдает номер позиции, в которой произошла ошибка. Это влияет на однозначный вид порождающей матрицы кода.

Рассмотрим $(4, 7)$-код. Такой код имеет 3 контрольных символа.

Сообщение: $\boldsymbol{a = a_1 a_2 a_3 a_4}$

Кодирование

Кодовое слово: $\boldsymbol{b = b_1 b_2 b_3 b_4 b_5 b_6 b_7}$

Индексы, равные степеням двойки имеют контрольные символы, остальные – последовательность символов сообщения.

b_1, b_2, b_4 – контрольные символы;

$b_3 = a_1, b_5 = a_2, b_6 = a_3, b_7 = a_4.$

Кодовое слово должно удовлетворять условию bM^T=0.

Матрица $\boldsymbol{M_{r \times n}}$ **– вспомогательная матрица**, по столбцам которой стоят числа от 1 до n в двоичной системе. (Не путать с порождающей матрицей кода!)

Из условия $\boldsymbol{bM^T}$**=0** найдем контрольные символы b_1, b_2, b_4.

Декодирование

$(b + e)M^T = bM^T + eM^T = eM^T$ (номер ошибки в двоичной системе).

С помощью кода Хэмминга можно исправить одну ошибку – минимальное

расстояние между кодовыми словами – три.

Пример 4. Рассмотрим $(4,7)-$ код, $r=3$. Вспомогательная матрица

$$\boldsymbol{M_{3\times 7}}=\begin{pmatrix}\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1}\\ \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1}\\ \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1}\\ \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} & \mathbf{7}\end{pmatrix}$$

Пусть сообщение $\boldsymbol{a}=\mathbf{0101}$, тогда кодовое слово

$b=b_1b_2b_3b_4b_5b_6b_7=b_1b_20b_4101.$

Система $\boldsymbol{bM^T}=\mathbf{0}$ примет вид:

$$\begin{cases}b_4+b_5+b_6+b_7=0\\ b_2+b_3+b_6+b_7=0\\ b_1+b_3+b_5+b_7=0\end{cases}$$

Подставим символы сообщения

$$\begin{cases}b_4+1+0+1=0\\ b_2+0+0+1=0\\ b_1+0+1+1=0\end{cases}$$

Найдем: $b_4=0, b_2=1, b_1=0.$

$\boldsymbol{b=0100101.}$

Найдем порождающую матрицу кода $(4,7)$.

Порождающая матрица кода G: **по строкам – ФСР системы $\boldsymbol{bM^T}=\mathbf{0}$.**

$$\begin{cases}b_4+b_5+b_6+b_7=0\\ b_2+b_3+b_6+b_7=0\\ b_1+b_3+b_5+b_7=0\end{cases}$$

b_3, b_5, b_6, b_7 – столбцы единичной матрицы, соответствующие **свободным переменным** – символы сообщения. Свободные переменные, как и при решении линейных систем в алгебре берем равными: одна 1, остальные 0. Находим базисные переменные b_1, b_2, b_4 (частное решение системы).

Первая строка: $b_3=1, b_5=0, b_6=0, b_7=0$. Подставляя в систему, получаем:

$$b_1=1, b_2=1, b_4=0.$$

Вторая строка: $b_3=0, b_5=1, b_6=0, b_7=0$. Подставляя в систему, получаем:

$$b_1=1, b_2=0, b_4=1.$$

Третья строка: $b_3=0, b_5=0, b_6=1, b_7=0$. Подставляя в систему, получаем:

$$b_1 = 0, b_2 = 1, b_4 = 1.$$

Четвертая строка: $b_3 = 0, b_5 = 0, b_6 = 0, b_7 = 1$. Подставляя в систему, получаем:

$$b_1 = 1, b_2 = 1, b_4 = 1.$$

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{b_1} & \boldsymbol{b_2} & \boldsymbol{b_3} & \boldsymbol{b_4} & \boldsymbol{b_5} & \boldsymbol{b_6} & \boldsymbol{b_7} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix}$$

Порождающую матрицу кода целесообразно использовать при кодировании сообщений $b = aG$, чтобы для каждого сообщения отдельно не решать систему.

Пример 5. РГР № 4

а) Закодировать сообщение $a = 0101$.

Закодировать сообщение можно двумя способами.

1. **Строим вспомогательную матрицу**

$$M_{3\times 7} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

По условию $a = 0101$, тогда $b = b_1 b_2 0 b_4 101$.

$\boldsymbol{bM^T = 0}$:

$\begin{cases} b_4 + b_5 + b_6 + b_7 = 0 \\ b_2 + b_3 + b_6 + b_7 = 0 \\ b_1 + b_2 + b_5 + b_7 = 0 \end{cases}$. Следовательно, $\begin{cases} b_4 + 1 + 0 + 1 = 0 \\ b_2 + 0 + 0 + 1 = 0 \\ b_1 + 0 + 1 + 1 = 0 \end{cases}$

Получим $b_4 = 0$, $b_2 = 1$, $b_1 = 0$.

$\boldsymbol{b = 0100101}$

2. **Закодируем сообщение с помощью порождающей матрицы кода: $\boldsymbol{b = aG}$.**

$$b = (0101)\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} = (0100101)$$

$\boldsymbol{b = 0100101}$

б) **Дано слово. Найти и исправить ошибку.**

$\boldsymbol{c = 1101001}$

Вычислим $cM^T = (b + e)M^T = bM^T + eM^T = eM^T$

$$cM^T = (1101001)\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} = (000)$$ – ошибки нет.

$b = c = 1101001$. Исходное сообщение $a = 0001$ (удалили контрольные символы: 1,2,4).

в) **Дано слово. Найти и исправить ошибку.**

$\boldsymbol{c = 1101011}$

$$cM^T = (1101011)\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} = (110)$$ – номер ошибки в двоичной системе.

Следовательно, ошибка в позиции 6. Исправим ее, получим кодовое слово $\boldsymbol{b = 1101001}$. Тогда исходное сообщение $\boldsymbol{a = 0001}$ (удалили контрольные символы: 1,2,4).