1 Общие слова о линейных кодах

Пусть задано продмножество $\mathfrak{K} \subset V_n$ множества двоичнх векторов длины n, называемое кодом и пусть задана матрица H.

Определение: (n, k)-кодом называется код со словами длиной n в каждом из которых содержится k информационных символов.

Определение: Линейным называется код, каждый вектор которого удовлетворяет уравнению $Hc^t = 0$. Матрица H называется проверочной матрицей кода $\mathfrak{K}$.

Определение: Код называется групповым, если множество его слов образует группу.

Очевидно, что каждый линейный код является групповым (более того, множество его слов образует продпространоство в пространстве V_n).

Примеры:

1. Рассмотрим код с повторением. $\mathfrak{K} = \{(c_1, c_2, \ldots, c_n) | c_1 = c_i, i = 1 \ldots n\}$ Его проверочной матрицей будет

$$H = \begin{pmatrix} 1 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

2. Код с проверкой на четность. $\mathfrak{K} = \{(c_1, c_2, \ldots, c_n) | c_n = \sum_{i=1}^{n-1} c_i\}$ Проверочная матрица:

$$H = (1, 1, \ldots, 1)$$

Определение: Будем говорить что проверочная матрица записана в каноническом виде если $H = (A| - I)$, где I – единичная матрица

В случае когда проерочная матрица записана в каноническом виде очень просто отделить информационные и и проверочные символы: первые k – информационные, остальные – проверочные.

Пусть $u = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ – исхоное информационное слово. Тогда

$$c^t = \begin{pmatrix} I \\ -- \\ -A \end{pmatrix} u^t$$

Транспонировав равенство получим $c = uG$, где $G = (I| - A^t)$. Матрица G называется порождающей матрицей. Из полученного результата сразу вытекает следующий результат Матрицы G и H связаны соотношением $HG^t = 0$.

На самом деле матрица G иметь такой вид не обязана. Это справедливо для кодов с матрицей H в каноническом виде, то есть для систематических кодов.

Есть более общий способ получить матрицу G.

Посмотрим на равенство $Hc^t = 0$ как на однородную систему линейных уравнений. Посмотрим базис решений этой системы, пусть это будут $\{e_1, e_2, \ldots, e_k\}$. Тогда любой вектор c мы сможем записать как линейную комбинацию этих векторов $c = \sum_{i=1}^{k} \alpha_i e_i$, где $\alpha_i \in \mathbb{Z}_2$. Далее запишем найденные решения в строки матрицы G. Таким образом получим порождающую матрицу.

Заметим, что раз строки матрицы G есть ничто иное как решение уравнения $Hc^t = 0$, то получаем следующую теорему:

Теорема: Матрицы H и G связаны соотношением $HG^t = 0$.

2 Как определять и корректировать ошибки

Перед тем как идти дальше следует сделать несклько замечаний. Во-первых, имея информационное слово длины k мы не можем просто так выбрать кодирующую матрицу G.

Чтобы декодировать принятое сообщение c, формально, нам нужно найти его прообраз, а это ничто иное как решение системы уравнений $uG = c$. Это значит что $\operatorname{rank} G = k$. Аналогично, $\operatorname{rank} H = n - k$

Теорема: Пусть код $\mathfrak{K}$ имеет минимальное расстояние $\geq d \Leftrightarrow$ любые $d-1$ столбцов матрицы H линейно независимы.

Доказательство: $\square$ $\Rightarrow$ Заметим что минимальное расстояние линейного кода равно минимальному весу его ненулевого слова.

Пусть $w(\mathfrak{K}) = d$, и пусть $c = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathfrak{K}$ – кодовое слово и $H = (h_1, h_2, \ldots, h_n)$ – проверочная матрица кода (h_i – столбцы H). Из соотношения $Hc^t = 0$ имеем $\alpha_1 h_1 + \alpha_2 h_2 + \cdots + \alpha_n h_n = 0$. Пусть $w(c) = d$. Тогда существует ровно d линейно зависимых столбцов проверочной матрицы.

Пусть теперь известно что t столбцов матрицы H линейнозависимы, то есть справедливо равенство

$$\varepsilon_{i_1} h_{i_1} + \varepsilon_{i_2} h_{i_2} + \ldots \varepsilon_{i_t} h_{i_t} = 0.$$

Это значит что существует вектор x такой что на позициях $i_j, j = 1 \ldots t$ у него стоят 1, а на других 0.

Значит для такого вектора x справедливо равенство $Hx^t = 0$, тогда, с учетом того что $w(\mathfrak{K}) = d$, получаем $t \geq d$. Следовательно любые $d-1$ столбцов матрицы H линейно независимы.

$\Leftarrow$ Обратно, пусть любые $d-1$ столбцов H – линейнонезависимы, тогда если $c \in \mathfrak{K}$, то $w(c) \geq d$. $\blacksquare$

Будем считать что шум в канале просто прибавлет к нашему слову c вектор ошибки e. Пусть полученное слово будет $y = c + e$. Тогда справедлива

Теорема: Групповой код $\mathfrak{K}$ оставляет незамеченными те и только те ошибки, которые являются его элементами.

Предположим что при передаче ошибки происходят независимо друг от друга с вероятностью $q < \frac{1}{2}$. Тогда, воспользовавшись законом Бернулли, вероятность того что проихошло ровно w ошибок равна $C_n^w q^w (1-q)^{n-w}$.

Отсюда видно что при $q < 1/2$ вероятность появления шумового слова веса t меньше, чем вероятность появления шумового слова веса $t-1$.

Заметим что если происходит ошибка, то вектор ошибки будет находится в смещном классе $\mathfrak{K}+y$. Обратно, если $c'+g=y$, то $\mathfrak{K}+g=\mathfrak{K}+y$. Поэтому множество векторов ошибок в точности составлет смежный класс $\mathfrak{K}+y$.

Поэтому возникает следующаяя идея: выпишем все возможные элементы смежного класса $\mathfrak{K}+y$ и среди них выберем слово наименьшего веса, которое назовем лидером смежного класса. Справедлива

Теорема: Групповой код $\mathfrak{K}$ исправлет в точности те ошибки, которые являются лидерами смежных классов.

Сформулируем еще одну теорему о линейных кодах

Теорема: Линейны код исправлет одиночные ошибки тогда и только тогда когда все столбцы матрицы H отличны от нуля и различны.

Доказательство: $\square$ $\Rightarrow$ Пусть $H=(h_1,h_2,\ldots,h_n)$ и $e_i=(0,0,\ldots,0,1,0,\ldots,0)$, где 1 стоит на i позиции. Тогда $H(c+e_i)=Hc+He_i=h_i$. Так как нам нужно чтобы мы умели различать ошибку в позиции $i\neq j$, то столбец $h_i\neq h_j$ и $h_i\neq 0$ при всех $i=1..n$.

$\Leftarrow$ Так как все столбцы матрицы H различны и отличны от нуля, то $H(c+e_i)=He_i=h_i$, то можно построить следующее соответствие

$$\begin{array}{c} h_1\to 1\\ h_2\to 2\\ \vdots\\ h_n\to n\end{array}$$

То есть если мы получили слово y и $Hy=h_i$ это значит что при передаче произошла ошибка в позиции i. $\blacksquare$