Қазақстан Республикасы Тәуелсіздігінің 30 жылдығына арналған «Сейфуллин оқулары – 17: «Қазіргі аграрлық ғылым: цифрлық трансформация» атты халықаралық ғылыми – тәжірибелік конференцияға материалдар = Материалы международной научно – теоретической конференции «Сейфуллинские чтения – 17: «Современная аграрная наука: цифровая трансформация», посвященной 30 – летию Независимости Республики Казахстан.- 2021.- Т.1, Ч.2 - С.134-136

ОБ АЛГОРИТМЕ КОДИРОВАНИЯ ИНФОРМАЦИИ С ПОМОЩЬЮ КОДА ХЭММИНГА

Филонцева Дарья
студент 1-го курса,
Казахский агротехнический университет им.С.Сейфуллина, г.Нур-Султан

Коды возникли в глубокой древности фактически с появлением системы знаков для записи звуков, слов, информации, которые позднее развились в различные языки. Каждый язык представляет собой сложную систему кодирования, включая в свою конструкцию алфавит, слова, грамматику. Язык позволяет в окружающем шуме передавать информацию по возможности быстро, надежно, с достаточно высокой степенью избыточности.

Позднее появились (еще до нашей эры) криптограммы. Такими кодами пользовались для засекречивания сообщений. Уже в V в. до н. э. знаменитый греческий историк Геродот приводил примеры писем-криптограмм, понятных только одному адресату. Спартанцы имели специальный механический прибор, при помощи которого записывались сообщения–криптограммы, позволяющие сохранить тайну. Собственную секретную азбуку имел Юлий Цезарь. В Средние века и эпоху Возрождения над изобретением тайных шифров работали многие выдающиеся умы, в том числе философ Фрэнсис Бэкон, математики Франсуа Виет, ДжероламоКардано. Криптографией занимались в монастырях, при дворах королей. Вместе с искусством шифрования сообщений развивалось и искусство их дешифрования. Многие оптимистично полагали, что вряд ли существует такая криптограмма, которую нельзя разгадать. И только в прошлом веке Клод Шеннон (1949 г.) показал, что существует совершенно секретный шифр – шифр Вернама, называемый также лентой однократного действия или шифром-блокнотом. ([1]).

В настоящее время теория кодирования имеет важное широкое практическое применение как средство удобной, быстрой и экономной, а также надежной передачи сообщений по линиям связи с различного вида шумами(телефон,телеграф,радио, телевидение, компьютерная, космическая связи и т. д.).

Пусть мы имеем множество всех двоичных слов длины $\boldsymbol{m}$. Эти слова передаются по каналу связи, в котором действует источник помех. Этот источник помех при передаче двоичного слова длины $\boldsymbol{m}$ может выдавать ошибки не более чем в $\boldsymbol{p}$ символах.

Это означает, что если исходное слово передавать без предварительного кодирования, то установить на выходе истинное сообщение практически невозможно. Поэтому возникает задача построения по исходному, любому слову

$a_1a_2a_3 \dots a_m$ его самокорректирующегося кода $b_1b_2 \dots b_l$ $(l > m)$, позволяющему по полученному на выходе канала кода $b_1'b_2' \dots b_l'$ однозначно восстановить передаваемый код $b_1b_2 \dots b_l$, а значит, и исходное сообщение $a_1a_2a_3 \dots a_m$. При передаче кода $b_1b_2 \dots b_l$ по каналу связи код, возможно, исказился и, следовательно, на выходе канала будет $b_1'b_2' \dots b_l'$, вообще говоря, отличающиеся от $b_1b_2 \dots b_l$ не более чем в $\boldsymbol{p}$ позициях.

Коды, обладающие вышеуказанными свойствами, называют самокорректирующимися кодамиотносительно источника помех или кодами, исправляющими $\boldsymbol{p}$ ошибок. Американский ученый Ричард Хэмминг в 1950 году опубликовал способ, который известен как код Хэмминга Хэмминг был первым, кто предложил конструктивный метод построения кодов с избыточностью и простым декодированием. Его труд предопределил направление большинства работ в этой области, последовавших позже. ([2]). Коды Хэмминга являются самоконтролирующимися кодами, то есть кодами, позволяющими автоматически обнаруживать ошибки при передаче данных. Код Хэмминга исправляет одиночные ошибки. Он состоит из комбинации из $\boldsymbol{m}$ информационных и $\boldsymbol{k}$ проверочных символов.

Избыточная часть кода строится таким образом, чтобы при декодировании можно было установить не только наличие ошибки, но и ее расположение внутри кодовой комбинации. Достигается это путем многократной проверки принятой кодовой комбинации на четность. При этом число проверок всегда равно числу контрольных разрядов $\boldsymbol{k}$. При каждой проверке охватывается часть информационных символов и один из контрольных разрядов, в ходе проверке получают один проверочный символ.

Код Хэмминга состоит из двух частей. Первая часть кодирует исходное сообщение, вставляя в него в определённых местах контрольные биты (вычисленные особым образом). Вторая часть получает входящее сообщение и заново вычисляет контрольные биты (по тому же алгоритму, что и первая часть). Если все вновь вычисленные контрольные биты совпадают с полученными, то сообщение получено без ошибок. В противном случае, выводится сообщение об ошибке и при возможности ошибка исправляется. Построение кодов Хэмминга основано на принципе проверки на четность числа единичных символов: к последовательности добавляется такой элемент, чтобы число единичных символов в получившейся последовательности было четным. ([3], [4])

Сам алгоритм можно показать следующим образом. Пусть есть некоторое сообщение, которое необходимо передать без ошибок. Для этого сначала нужно наше сообщение закодировать при помощи Кода Хэмминга. Надо представить его в бинарном виде. На этом этапе стоит определиться с длиной информационного слова, то есть длиной строки из нулей и единиц, которые нужно закодировать. После этого необходимо вычислить значение каждого контрольного бита. Значение каждого контрольного бита зависит от значений информационных бит, которые этот контрольный бит контролирует. Далее записываем соответствующее полученным данным сообщение.

Рассмотрим на примере применение данного кодирования. Закодируем сообщение кодом Хэмминга 1011011. Для кодирования данного сообщения длиной m = 7 потребуется k = 4 дополнительных разряда, т.е. на выходе получим сообщение длиной n = 11 (количество дополнительных разрядов подбирали из соотношения 2^k ≥ n+1, где n – число полученных разрядов, k – число

дополнительных разрядов) [3].

Пусть закодированное сообщение имеет вид b1 b2 b3 b4 b5 b6 b7 b4 b8 b9 b10 b11, причем разряды b1, b2, b4, b8, будут контрольными, а остальные информационными.

Помещаем в информационные разряды исходного числа по порядку, т.е.

b3 =1, b5 = 0, b6 =1, b7=1, b9=0, b10 =1, b11=1.

Теперь найдем значения контрольных разрядов.

Введем для удобства следующие множества:

V1 = 1, 3, 5, 7, 9, 11 - все числа, у которых первый разряд равен 0.

V2 = 2, 3, 6, 7, 10, 11 - все числа, у которых второй разряд равен 0.

V3 = 4, 5, 6, 7 - все числа, у которых третий разряд равен 1.

V4 = 8, 9, 10, 11 - все числа, у которых четвертый разряд равен 0.

Далее под⊕будем понимать сложение по модулю 2.

Тогда b1 = b3⊕ b5⊕ b7⊕ b9⊕ b11 = 1 (все разряды из V1, кроме первого)

b2 = b3⊕ b6⊕ b7⊕ b10⊕ b11 = 1 (все разряды из V2, кроме первого)

b4 = b5⊕ b6⊕ b7 = 0 (все разряды из V3, кроме первого)

b8 = b9⊕ b10⊕ b11= 0 (все разряды из V4, кроме первого),

Таким образом, получили код 11100110011

По методу Хемминга могут быть построены коды разной длины. При этом, чем больше длина кода, тем меньше относительная избыточность. Код Хэмминга используется в некоторых прикладных программах в области хранения данных, особенно в RAID 2; кроме того, метод Хэмминга давно применяется в памяти типа ECC и позволяет «на лету» исправлять однократные и обнаруживать двукратные ошибки.

Список использованной литературы:

1. Соловьева Ф.И. «Введение в теорию кодирования»/НГУ, Новосибирск/2006, 126 с.
2. Питерсон У., Уэлдон Э. «Коды, исправляющие ошибки»: Пер. с англ. /М.: Мир./ 1976, 594 с.
3. Пенин П. Е., Филиппов Л. Н. «Радиотехнические системы передачи информации»./ М.: Радио и Связь / 1984, 256 с.
4. Fish,W., Key, J.D., Mwambene, E., Rodrigues, B.G. «Hamming graphs and special LCD codes.»/ Journal of Applied Mathematics and Computing/61(1-2), 2019 с. 461-479

Руководитель: Аскарова А.Ж.