Application of Coset in Error Correction

Soumik Guha Roy

Junior Research Fellow in Computer Science
Machine Intelligence Unit

20-April-2024

Presentation Overview

Coding of Binary Information and Error Detection

- The basic unit of information called a **message** which is finite sequence of characters from a finite alphabet.Here ,our alphabet set is $B = 0, 1$.
- Basic unit of information,called **word**, is a sequence of m 0's or 1's.
- The set $B = \{0, 1\}$ is forming a **group** under the operation '+' defined as:

+	0	1
0	0	1
1	1	0

- $B^m = B \times B \times B \cdots \times B$ is group under the operation $\oplus$ defined as $(x_1, x_2, \ldots, x_m) \oplus (y_1, y_2, \ldots, y_m) = (x_1 + y_1, x_2 + y_2, \ldots, x_m + y_m)$ and the **identity element** of $B^m = (0, 0, \ldots, 0) = \bar{0}$
- Elements of B^m will be written as $b_1, b_1, \ldots, b_m$
- Sender sends $x \in B^n$ and the receiver recives $x_t \in B^n$. Due to **Noise** $x \neq x_t$.Hence,if noise exists then x_t can be any element in B^n.
- **Basic task** is to reduce the likelihood of receiving data.**How?**

Definition

- **Encoding function (m,n)** Choose $m, n \in \mathbb{Z}$ such that $n > m$ and a **one to one** function $e : B^m \mapsto B^n$
- **Decoding function (n,m)** associated with an encoding fucntion e is an **onto** function $d : B^n \mapsto B^m$

Definition

- $(B^n, \oplus)$ is a group
- An (m,n) Encoding function $e : B^m \mapsto B^n$ is called **Group Code** if $e(b^m) = \{e(b) | b \in B^m\} = Range(e)$ is a subgroup of $(B^n, \oplus)$

Generating Group Code

- **Minimum distance of a group code**
  Theorem 1: Given $e : B^m \mapsto B^n$ is a group code. The ***minimum distance of e*** is the minimum weight of the nonzero code word.
- **Theorem 2:** Let $\mathbf{D}$ and $\mathbf{E}$ be $m \times p$ Boolean matrices, and let $\mathbf{F}$ be a $p \times n$ Boolean matrix. Then $(\mathbf{D} \oplus \mathbf{E}) \star \mathbf{F} = (\mathbf{D} \star \mathbf{F}) \oplus (\mathbf{E} \star \mathbf{F})$
- **Theorem 3:** Let $m, n \in \mathbb{N}$ with $m < n$ and $r = n - m$ and let $\mathbf{H}$ be an $n \times r$ Boolean matrix. Then the function $f_H : B^n \mapsto B^r$ defined by $f_H(x) = x \star \mathbf{H}$ where $x \in B^n$ is a homomorphism from the group B^n to B^r.
  - **Corollary 3.1:** $N = \{x | x \in B^n, x \star \mathbf{H} = 0\}$ is a normal subgroup of $(B^n, \oplus)$.

Parity Check Matrix and Encoding function

- **Parity check matrix** An $n \times r$ Boolean matrix defined as $\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1r} \\ h_{21} & h_{22} & \dots & h_{2r} \\ \dots \\ h_{m1} & h_{m2} & \dots & h_{mr} \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$ where the last r rows is a identity matrix I_r.
- **Encoding function using parity check matrix:** $e_H : B^m \mapsto B^n$.Let $b = b_1 b_2 \dots b_m$ and $x = e_H(b) = b_1 b_2 \dots b_m x_1 x_2 \dots x_r$ where

$$\begin{aligned} x_1 &= b_1 . h_{11} + b_2 . h_{21} + \dots + bm . h_{m1} \\ x_2 &= b_1 . h_{12} + b_2 . h_{22} + \dots + bm . h_{m2} \\ \dots \\ x_r &= b_1 . h_{1r} + b_2 . h_{2r} + \dots + bm . hmr \end{aligned} \tag{1}$$

- **Theorem** Let $x = y_1 y_2 \dots y_m x_1 x_2 \dots x_r \in B^n$. Then $x \star H = 0 \leftrightarrow x = e_H(b), b \in B^m$
  - **Corollary** $e_H(B^m) = \{e_H(b) | b \in B^m\}$ is a sub group of B^n

Example of Encoding

- Given the group code $e_H : B^2 \mapsto B^5$ then $m = 2, n = 5$ and the parity check matrix
$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
- $B^2 = \{00, 01, 10, 11\}$.
- Then $e(00) = 00x_1x_2x_3$ where $b_1 = b_2 = 0$ and x_1, x_2, x_3 can be obtained from the **H** matrix $.x_1 = x_2 = x_3 = 0$.Hence, $e(00) = 00000$
- Similarly,$e(10) = 10x_1x_2x_3$ where $b_1 = 1, b_2 = 0$ and $x_1 = x_2 = 1, x_3 = 0$.Hence,$e(01) = 10110$
- In the same way, $e(01) = 01x_1x_2x_3$ where $b_1 = 0, b_2 = 1$ and $x_1 = 0, x_2 = x_3 = 1$.Hence,$e(01) = 01011$
- Finally, $e(11) = 11x_1x_2x_3$ where $b_1 = 1, b_2 = 1$ and $x_1 = 1, x_2 = 0, x_3 = 1$.Hence,$e(01) = 11101$
- Minimum distance of this group code (2,5) is 3 (**Why this distance?**)

Definition

- **Maximum likelihood decoding function:** Given $e_H : B^m \mapsto B^n$. Let us list the code words in a fixed order: $x^{(1)}, x^{(2)}, \ldots, x^{(2^m)}$
- Let x_t be the received word and compute $\delta(x^{(i)}, x_t) \forall i = 1$ to 2^m and choose the first code word , $x^{(s)}$ such that min $\delta(x^{(i)}, x_t) = \delta(x^{(s)}, x_t)$ $\forall i = 1$ to 2^m. Hence,$x^{(s)}$ is the closest code to x_t and the first in the list $x^{(1)}, x^{(2)}, \ldots, x^{(2^m)}$
- Let $x^{(s)} = e(b)$.Then **maximum likelihood decoding function d** associated with e by $d(x_t) = b$ where x_t is the received word.
- The maximum likelihood decoding function d depends on the order $x^{(1)}, x^{(2)}, \ldots, x^{(2^m)}$.
- **Theorem:** Given that e is an (m, n) encoding function and d is the maximum likelihood decoding function associated with e. Then (e,d) can correct k or fewer erros if and only if the minimum distance of e is atleast $2k + 1$.

Coset Leader

- Let $e : B^m \mapsto B^n$ be an (m, n) encoding function.N is set of code words in B^n such that $N = \{x^{(1)}, x^{(2)}, \ldots, x^{(2^m)}\}$
- Let $x = e(b)$ where $b \in B^m$ is transmitted and received as $x_t \in B^n$. Left coset of N is $x_t + N = \{x_t + x^{(1)}, x_t + x^{(2)}, \ldots, x_t + x^{(2^{(m)})}\} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_{2^{(m)}}\}$ where $\epsilon_{2^{(i)}} = x_t \oplus x^{(i)}$.
  - Distance between the received code word x_t and $x^{(i)}$ is $|\epsilon_i|$
  - ϵ_j is a coset member with smallest weight, then $x^{(j)}$ must be the code word that is closest to x_t.Here, $x^j = \bar{0} \oplus x^j = x_t \oplus x_t \oplus x^j = x_t \oplus \epsilon_j$
  - **Coset Leader** An element ϵ_j having the smallest weight,called the *Coset leader*.
    - *Coset leader may not be unique*

Basic Decoding function

- Given $e : B^m \mapsto B^n$ is a group code and sender sends the data b encoded as $x = e(b)$ to the receiver.
- **Step 1:** Determine all the left cosets of $N = e(B^m)$
- **Step 2:** For each coset, find the coset header(a word with smallest weight)
- **Step 3:** Determine in which coset of N, x_t belongs.[As N is normal subgroup of B^n ,due to partition of N, x_t will be in exactly one coset among 2^{n-m}]
- **Step 4:** Let ϵ be the coset leader as determined in Step 3.Compute $x = x_t \oplus \epsilon$. If $e(b) = x$, then $d(x_t) = b$.Hence, receiver decodes x_t as b.
- The **main problem of this algorithm** is the calculation of the entire table containing all the coset elements.

Syndrome of a code word

- **Theorem:** Given $m, n, r = n - m$ and $f_H : B^n \mapsto B^m$ and defined as $f_H(x) = x \star H$, then f_H is onto function.
- **Syndrome of x:** B^r and B^n/N are isomorphic where $N = ker(f_H) = e_H(B^m)$ under the homomorphism $g : B^n/N \mapsto B^r$ defined by $g(xN) = f_H(x) = x \star H$.Here , the element $x \star H$ called the *Syndrome of* x.
- **Theorem:** Let $x, y \in B^n$.Then x and y are same left coset of N in B^n if and only if $f_H(x) = f_H(y)$, that is if and only if x and y have the same syndrome.

Modified Decoding function

- Given $e : B^m \mapsto B^n$ is a group code and sender sends the data b encoded as $x = e(b)$ to the receiver.
- **Step 1:** Determine all the left cosets of $N = e_H(B^m)$ in B^n.
- **Step 2:** For each coset find the coset leader and find the syndrome of all coset leaders.
- **Step 3:** If x_t is the received, computer the syndrome of x and find the coset leader ϵ having the same syndrome. Then $x_t \oplus \epsilon = x$ is a code word $e_H(b)$ and $d(x_t) = b$.
- Here, we donot need to keep the entire table of cosets.

Example 1

Given the (3, 6) group $e_H : B^3 \mapsto B^6$ and consider the parity matrix $\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

The encoding function

$e(x)$	$e(000)$	$e(001)$	$e(010)$	$e(011)$	$e(100)$	$e(101)$	$e(110)$	$e(111)$
$x \in N$	000000	001011	010101	011110	100110	101101	110011	111000

The Syndrome Coset Leader table

Syndrome of Coset Leader ($x \star \mathbf{H}$)	000	001	010	011	100	101	110	111
Coset Leader (ϵ_i)	000000	000001	000010	001000	00100	010000	100000	001100

- Sender send the data $b = 011 \in B^3$ encoded as $x = e(011) = 011110$
- Receiver receives the data $x_t = 011110$
  - Step 1: Calculate the syndrome of x_t as $f_H(x_t) = x_t \star H = 011110 \star H = 101$
  - Step 2: Using the Coset Leader table, the coset header is $\epsilon = 010000$
  - Step 3: Finally ,compute $x = x_t \oplus \epsilon = 001110 \oplus 010000 = 011110$ and the data is $b = e^{-1}(011110) = 011$

Example 1

Given the (3, 6) group $e_H : B^3 \mapsto B^6$ and consider the parity matrix $\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

The encoding function

$e(000)$	$e(001)$	$e(010)$	$e(011)$	$e(100)$	$e(101)$	$e(110)$	$e(111)$
000000	001011	010101	011110	100110	101101	110011	111000

The Syndrome Coset Leader table

Syndrome of Coset Leader ($x \star \mathbf{H}$)	000	001	010	011	100	101	110	111
Coset Leader (ϵ_i)	000000	000001	000010	001000	00100	010000	100000	001100

- Sender send the data $b = 001 \in B^3$ encoded as $x = e(001) = 001011$
- Receiver receives the data $x_t = 011011$
  - Step 1: Calculate the syndrome of x_t as $f_H(x_t) = x_t \star H = 011011 \star H = 101$
  - Step 2: Using the Coset Leader table, the coset header is $\epsilon = 010000$
  - Step 3: Finally ,compute $x = x_t \oplus \epsilon = 011011 \oplus 010000 = 001011$ and the data is $b = e^{-1}(001011) = 001$

1 bit error message corrected by Receiver

Example 3

Given the $(3,6)$ group $e_H : B^3 \mapsto B^6$ and consider the parity matrix $\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

The encoding function

$e(000)$	$e(001)$	$e(010)$	$e(011)$	$e(100)$	$e(101)$	$e(110)$	$e(111)$
000000	001011	010101	011110	100110	101101	110011	111000

The Syndrome Coset Leader table

Syndrome of Coset Leader ($x \star \mathbf{H}$)	000	001	010	011	100	101	110	111
Coset Leader (ϵ_i)	000000	000001	000010	001000	00100	010000	100000	001100

- Sender send the data $b = 010 \in B^3$ encoded as $x = e(010) = 010101$
- Receiver receives the data $x_t = 011111 \in B^6$
  - Step 1: Calculate the syndrome of x_t as $f_H(x_t) = x_t \star H = 011111 \star H = 001$
  - Step 2: Using the Coset Leader table, the coset header is $\epsilon = 000001$
  - Step 3: Finally ,compute $x = x_t \oplus \epsilon = 001110 \oplus 010000 = 011110$ and the data is $b = e^{-1}(011110) = 011$.

Wrong data accepted by Receiver.But why?

References

- Bernard Kolman,Robert C.Busby,Sharon Ross *Discrete Mathematical Structures*, Prentics Hall of India.
- S.K.Mapa *Higher Algebra*, Levant Books,India, 4th ed.

Thanks for your attention