Códigos Corretores de Erros e Cliques de Grafos

Natália Pedroza

Jayme Szwarcfiter
Paulo Eustáquio

UFRJ/UERJ

2016

Apresentação

- Códigos corretores de erros;
- Códigos Lineares;
- Generalização do código Hamming;

Códigos corretores de erros de tamanho fixo

- $A = \{a_1, \cdots, a_s\} \rightarrow$ símbolos a serem codificados;
- $F = \{0, 1\}$ alfabeto;
- $c = c_1 c_2 \cdots c_n$, com $c_i \in F \rightarrow$ *palavra*;
- *Código* $C = \{c_1, \cdots, c_M\} \rightarrow$ conjunto de palavras.

Distância Hamming entre duas palavras, $d_H(c_1, c_2) \rightarrow$ número de posições em que c_1 e c_2 diferem.

Ex.: $d_H(000, 011) = 2$

Distância Hamming entre duas palavras, $d_H(c_1, c_2) \rightarrow$ número de posições em que c_1 e c_2 diferem.

Ex.: $d_H(000, 011) = 2$

$d_H(C)$ - distância Hamming de um código $C \rightarrow$ menor distância entre duas palavras quaisquer.

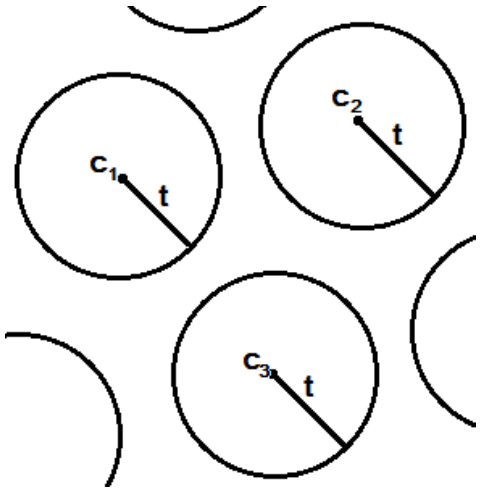
Distância Hamming entre duas palavras, $d_H(c_1, c_2) \rightarrow$ número de posições em que c_1 e c_2 diferem.

Ex.: $d_H(000, 011) = 2$

$d_H(C)$ - distância Hamming de um código $C \rightarrow$ menor distância entre duas palavras quaisquer.

Teorema

C é capaz de corrigir até t erros se $d(C) = 2t + 1$.



$A_q(n, d) \rightarrow$ o maior valor de M tal que existe um código de tamanho fixo q-ário com M palavras de n bits e distância Hamming d.

Theorem

Seja d um número ímpar. Um código binário com M palavras de tamanho n com distância d existe se, e somente se, um código binário com M palavras de tamanho $n+1$ com distância $d+1$ existe.

Corollary

Se d é ímpar, então $A_2(n+1, d+1) = A_2(n, d)$.

n \ d	3	5	7	9	11	13	15
5	4	2	1	1	1	1	1
6	8	2	1	1	1	1	1
7	16	2	2	1	1	1	1
8	20	4	2	1	1	1	1
9	40	6	2	2	1	1	1
10	72	12	2	2	1	1	1
11	144	24	4	2	2	1	1
12	256	32	4	2	2	1	1
13	512	64	8	2	2	2	1
14	1024	128	16	4	2	2	1
15	2048	256	32	4	2	2	2
16	2816-3276	256-340	36	6	2	2	2
17	5632-6552	512-673	64-72	10	4	2	2
18	10496-13104	1024-1237	128-135	20	4	2	2
19	20480-26168	2048-2279	256	40	6	2	2
20	40960-43688	2560-4096	512	42-47	8	4	2
21	81920-87333	4096-6941	1024	64-84	12	4	2
22	147456-172361	8192-13674	2048	80-150	24	4	2
23	327680-344308	16384-24106	4096	136-268	48	6	4
24	2^{19}-599184	17920-47538	4096-5421	192-466	52-55	8	4
25	2^{20}-1198368	32768-84260	4104-9275	384-836	64-96	14	4

Clique

Determinar $A_2(n, d) \equiv$ Determinar a **clique máxima** de um grafo G.

$V(G) \to 2^n$ vetores binários de tamanho n
$uv \in E(G) \Leftrightarrow d_H(u, v) \geq d$.

$d(u, v) \to$ tamanho do caminho mínimo entre u e v.

Grafo potência d de um grafo $G = (V, E) \to$
grafo $G^d = (V, E^d)$ tal que $uv \in E^d \Leftrightarrow d(u, v) \leq d$.

n-cubo $= Q_n \to d(u, v) = d_H(u, v)$.

Arestas de $Q_n^{d-1} \to$ Arestas correspondentes às distâncias
$1, 2, \cdots, d-1$ de K_n.

Problema: **Máximo de palavras**

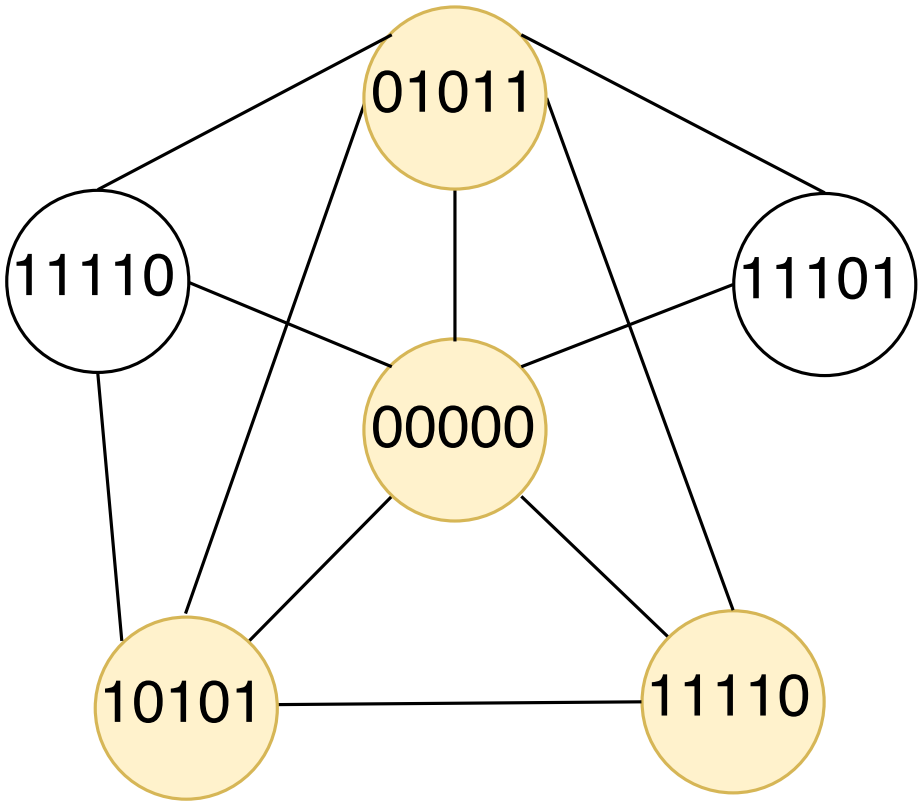
Entrada: Um grafo $G = K_{2^n} - E[Q_n^{d-1}]$.
Propriedade: O conjunto de vértices C forma uma **clique máxima**.

Grafo regular
Grau de v:

$$d(v) = 2^n - \sum_{i=0}^{d-1} \binom{n}{i}.$$

Cliques

Parte do grafo para $n = 5$ e $d = 3$.



Códigos Lineares

Seja F_q um corpo de q elementos. Um código C q-ário é dito $[n, k]$-código linear se é um subespaço vetorial de dimensão k de F_q^n. Isto é:

1. $u + v \in C, \ \forall \ u$ e $v \in C$ e
2. $au \in C, \ \forall \ u \in C, \ a \in F_q$.

Em particular, um código binário é linear se, e somente se, a soma de duas palavras quaisquer do código é também uma palavra do código.

Peso w de uma palavra - número de bits não nulos da palavra.

Peso de um código C - o menor dos pesos das palavras não nulas de C.

Theorem

Se C é um código linear, então $d(C) = w(C)$.

Demonstração: Existem x e $y \in C$ tais que $d(C) = d(x, y) = w(x - y)$.
$\Rightarrow d(C) = w(x - y) \geq w(C)$, pois $x - y$ é uma palavra do código.

Por outro lado, para alguma palavra $x \in C$,
$w(C) = w(x) = d(x, 0) \geq d(C)$.

Obs.: 0 pertence a todo código linear.

Matriz geradora - matriz cujas linhas formam uma base de um $[n, k]$-código.

Exemplo

$$[5,2,3]\text{-}código = \begin{cases} 00000 \\ 01101 \\ 10110 \\ 11011 \end{cases} \qquad G = \begin{bmatrix} 01101 \\ 10110 \end{bmatrix}$$

Matriz geradora - matriz cujas linhas formam uma base de um $[n, k]$-código.

Exemplo

$$[5,2,3]\text{-}código = \begin{cases} 00000 \\ 01101 \\ 10110 \\ 11011 \end{cases} \qquad G = \begin{bmatrix} 01101 \\ 10110 \end{bmatrix}$$

Codificação

- $M = 2^k$;
- Identificamos as mensagens com $u \in F^k$;
- Codificamos as mensagens multiplicando u por G;
- $uG \in C$ é uma combinação linear das linhas da matriz geradora.

Matriz geradora na **forma padrão**: $G = [I_k | A_{k\times(n-k)}]$.

Exemplo

A forma padrão de uma matriz geradora do $[7, 4]$-código é dada por:

$$G = \begin{bmatrix} 1000101 \\ 0100111 \\ 0010110 \\ 0001011 \end{bmatrix}$$

O vetor mensagem 1000 é codificado como 1000101.

Código dual de C: $C^{\perp} = \{u \in F^n \mid \langle c, u \rangle = 0, \ \forall c \in C\}$.

Lemma

Se C é um $[n,k]$-código linear com matriz geradora G, então

$$u \in C^{\perp} \Leftrightarrow uG^T = 0.$$

Theorem

Se C é um $[n,k]$-código linear, então $C^{\perp}$ é $[n, n-k]$-código linear.

Exemplo

$$C = \left\{ \begin{array}{l} 000 \\ 110 \\ 011 \\ 101 \end{array} \right. \rightarrow \ C^{\perp} = \left\{ \begin{array}{l} 000 \\ 111 \end{array} \right.$$

H **matriz de paridade** de C: matriz geradora de $C^{\perp}$.

H satisfaz $GH^T = 0$.

$$C = \{c \in F^n |\ cH^T = 0\}.$$

Theorem

Se $G = [I_k|A]$ é uma matriz geradora na forma padrão de um $[n, k]$-código C, então a matriz de paridade de C é $H = [-A^T|I_{n-k}]$.

Exemplo

$$G = \left[\begin{array}{c|c} I_4 & \begin{matrix}101\\111\\110\\011\end{matrix} \end{array}\right] \qquad H = \left[\begin{array}{c|c} \begin{matrix}1110\\0111\\1101\end{matrix} & I_3 \end{array}\right]$$

Códigos Hamming

$Ham(r)$ - matriz de paridade $H_{r\times(2^r-1)}$ cujas colunas são os vetores não nulos de F^r.

Exemplo

$Ham(3)$

$$H = \begin{bmatrix} 0001111 \\ 0110011 \\ 1010101 \end{bmatrix}$$

Theorem

Um código Hamming, $Ham(r)$, para $r \geq 2$,

1. *é um $[2^r - 1, 2^r - 1 - r]$-código;*
2. *tem distância mínima 3;*
3. *é um código perfeito.*

Uma generalização dos Códigos Hamming

$$Gham(n) = \{c_1^n, c_2^n, \cdots, c_M^n\}$$

- Código linear;
- Distância Hamming 3;
- $M = 2^k$ palavras.

$$c_i^n = \underbrace{b(c_i^n)}_{k \text{ bits}} \quad \underbrace{p(c_i^n)}_{r \text{ bits}}$$

(informação ↓ $b(c_i^n)$; paridade ↓ $p(c_i^n)$)

$$n = r + k, \text{ onde } r = \lceil \log_2(n+1) \rceil.$$

Recursão

Criar $Gham(n) = \{c_1^n, \cdots, c_{M'}^n\}$

A partir de $Gham(n-1) = \{c_1^{n-1}, \cdots, c_M^{n-1}\}$ com $M = 2^{n-1-\lceil \log_2(n) \rceil}$ palavras.

Base: $Gham(3) = \{000, 111\}$.

- Se $n \neq 2^m$, então:

$$c_i^n = \begin{cases} \mathbf{0c_i^{n-1}}, \text{ para } 1 \leq i \leq M \\ \mathbf{1[c_i^{n-1} \oplus n]}, \text{ para } M+1 \leq i \leq 2M \end{cases}$$

- Se $n = 2^m$, então:

$$c_i^n = \mathbf{c_{i_1}^{n-1} \cdots c_{i_k}^{n-1} 0 c_{i_{k+1}}^{n-1} \cdots c_{i_{n-1}}^{n-1}}, \text{ para } 1 \leq i \leq M.$$

Exemplo

Gham(3)
000 111

Gham(4)
0000 1011

$$r = \lceil \log_2(n+1) \rceil = 2 \text{ e } k = 1.$$

Exemplo

Gham(3)
000 111

Gham(4)
0000 1011

Gham(5)
00000
01011

Exemplo

Gham(3)	
000	111

Gham(4)	
0000	1011

Gham(5)	
00000	1
01011	1

$$\begin{array}{rr} & 0000 \\ \oplus & 0101 \\ \hline & 0101 \end{array}$$

Exemplo

Gham(3)	
000	111

Gham(4)	
0000	1011

Gham(5)	
00000	10101
01011	11100

$$\begin{array}{rr} & 1011 \\ \oplus & 0101 \\ \hline & 1100 \end{array}$$

Propriedade

O código Gham(n) é gerado ordenadamente e $b(c_i^n)$ corresponde à representação binária, usando k bits, de todos os inteiros de 0 a 2^{k-1}.

Propriedade

O código Gham(n) é gerado ordenadamente e $b(c_i^n)$ corresponde à representação binária, usando k bits, de todos os inteiros de 0 a 2^{k-1}.

Tabela 1: *Gham*(*n*) para *n* de 3 a 8

n = 3		n = 4		n = 5	
000	111	0000	1011	00000	10101
				01011	11110

n = 6		n = 7		n = 8	
000000	100110	0000000	1000111	00000000	10000111
001011	101101	0001011	1001100	00010011	10010100
010101	110011	0010101	1010010	00100101	10100010
011110	111000	0011110	1011001	00110110	10110001
		0100110	1100001	01000110	11000001
		0101101	1101010	01010101	11010010
		0110011	1110100	01100011	11100100
		0111000	1111111	01110000	11110111

Propriedade

Os códigos Gham(n) são códigos lineares, cujas matrizes geradoras podem ser escritas na forma $G = [I_{k\times k}|A_{k\times n-k}]$, *onde as linhas de* $A_{k\times n-k}$ *são formadas pelas representações binárias dos números de 3 a n que não são potências de 2, em ordem decrescente, usando* $n-k$ *bits.*

Algoritmo de codificação

Tomar $b(c_i^n) = u_i^k = u_{i_k} u_{i_{k-1}} \cdots u_{i_1}$.

Calcular $p(c_i^n)$: se $u_{i_j} = 1$, fazer a operação xor de $\underbrace{0 \cdots 0}_{\text{bits}}$ com $A(j)$.

Algoritmo 1: Obtenção de c_i^n:

---

início

$A \leftarrow [3, 5, 6, 7, \ldots, n]$; $\quad p(c_i) \leftarrow \mathbf{0}$;

para j *de* k *a* *1* **:**

se $(u_{i_j} = 1)$ **então** $p(c_i) \leftarrow p(c_i) \oplus A[k - j + 1]$;

fp

$c_i^n \leftarrow u_i p(c_i)$;

retorna c_i^n;

fim

Exemplo

- Determinar c_{13}^8
- Tome $b(c_{13}^8) = u_{13}^4 = 1101$.
- Bits 1 da direita para a esquerda: $1^{\underline{o}}, 3^{\underline{o}}$ e $4^{\underline{o}}$.

$$
\begin{array}{rrl}
 & 0000 & \\
 & \oplus\ 0011 & \leftarrow A[1] = 3 \\
\hline
 & 0011 & \\
 & \oplus\ 0110 & \leftarrow A[3] = 6 \\
\hline
 & 0101 & \\
 & \oplus\ 0111 & \leftarrow A[4] = 7 \\
\hline
p(c_{13}^8) \rightarrow & 0010 &
\end{array}
$$

$$
\Rightarrow c_{13}^8 = \underbrace{1101}_{b(c_{13}^8)}\,\underbrace{0010}_{p(c_{13}^8)}.
$$

Decodificação

Se um vetor y_i^n é recebido, calculamos $p(y_i^n)$ e fazemos a operação xor dos r bits finais de y_i^n com $p(y_i^n)$. Se o resultado:

- Contém apenas 0's, então não ocorreu erro;
- Contém apenas um bit 1, então o erro está nos r bits finais de y_i^n, na posição correspondente ao 1;
- É o i-ésimo elemento do vetor A, então o erro está posição i;
- Caso contrário, houve mais de um erro e a mensagem não é decodificada.

Propriedade

A distância Hamming de $Gham(n)$ é 3.

Demonstração.

Provar: $w(Gham(n)) = 3$.

Seja $c_i^n \in Gham(n)$.

1. $w(b(c_i^n)) = 1 \Rightarrow p(c_i^n) = \mathbf{0} \oplus n_1 \Rightarrow w(p(c_i^n)) \geq 2$.
   Pois $n_1 \neq 2^m$ e portanto $w(n_1)$.

2. $w(b(c_i^n)) = 2 \Rightarrow p(c_i^n) = \mathbf{0} \oplus n_1 \oplus n_2 \Rightarrow w(p(c_i^n)) \geq 1$.
   Pois como $n_1 \neq n_2$, temos que $n_1 \oplus n_2 \neq \mathbf{0}$.

3. $w(b(c_i^n)) \geq 3 \Rightarrow w(c_i^n) \geq 3$.

□

Lemma

A matriz de paridade H do código $Gham(n)$ é formada pelas representações binárias dos números de 1 a n escritos em coluna utilizando $n-k$ bits.

$$G = [I_k | A_{k \times n-k}] \Rightarrow H = [A^T_{n-k \times k} | I_{n-k}].$$

$\rightarrow$ As colunas de $A^T_{n-k \times k}$ são formadas pelas representações binárias dos números de 3 a n que não são potências de 2, em ordem decrescente, usando $n-k$ bits.

$\rightarrow$ As linhas de I_{n-k} são formadas pelas representações binárias dos números de 1 a n que são potências de 2, em ordem decrescente, usando $n-k$ bits.

Theorem

Os códigos de Hamming $Ham(r)$ são um caso particular dos códigos $Gham(n)$.

O código $Ham(r)$ tem matriz de paridade H cujas colunas são formadas pelos vetores não nulos de F^r que são exatamente as representações binárias dos números de 1 a n utilizando $r = n - k$ bits.

Assim, a matriz de paridade do código $Ham(r)$ é equivalente a matriz de paridade do código $Gham(n)$, para $n = 2^r - 1$.

Theorem

Um código linear ótimo de tamanho n e distância Hamming 3 tem dimensão $k = n - \lceil \log_2(n+1) \rceil$.

Corollary

Os códigos $Gham(n)$ são códigos lineares ótimos.

Obtenção gulosa de $Gham(n)$

Propriedade

O código $Gham(n)$ pode ser obtido por um algoritmo guloso.

Inicialmente coloca-se $0 \in F^k$ no código e, em seguida, para cada $u \in F^k \setminus \{0\}$, concatena-se a menor paridade $p \in F^r$ possível, tal que a palavra resultante tenha distância 3 para as palavras anteriores já definidas.

Com pequenas alterações no algoritmo, conseguimos obter os códigos ótimos para n de 8 a 11, que são códigos não lineares.
Ao invés de se concatenar uma única paridade $v \in F^r$ a cada elemento $u \in F^k \setminus 0$ pode-se ter nenhuma ou mais de uma concatenação. E acrescenta-se ao código além de $0 \in F^k$ algumas palavras $v \in F^n$.

Conclusão

- Desenvolvemos uma generalização para os códigos de Hamming para todo inteiro $n \geq 3$.
- Tais códigos são lineares, têm capacidade de corrigir um erro e são facilmente implementáveis.

Trabalhos futuros

- Estender os códigos $Gham(n)$ para distâncias Hamming $d > 3$.
- Generalizar a construção feita para os códigos $Gham(n)$ para códigos não lineares buscando obter códigos não lineares ótimos.

Obrigada

nataliaps@cos.ufrj.br