

12.TÉTEL

Változó hosszúságú kódok. Szabad félcsoporth. Cayley gráf. Kód. Feltételek felbonthatóságra. Egyenletes, vesszős prefixmentes kódok. Kraft-McMillan egyenlőtlenség. Fano-Huffman kód. Fixhosszúságú kódok, hibajelzés, hibajavítás. Hamming távolság. Lineáris kód. Generátor és távolságellenőrző mátrix. Hamming és BCH kódok.

KÓDOLÁS

1. változó hosszúságú kódok
2. hibajavító kódok
3. kriptográfia

1. Változó hosszúságú kódok (pl.: Morse kód)

Def:

A véges nem üres ábécé (forrás ábécé)

B véges nem üres ábécé (csatorna ábécé)

$B^* = \{ \text{a B-beli jelekkel képezhető összes véges szó} \}$

$f: A \rightarrow B^*$

- Ekkor f -et egy kódnak nevezzük
- Van aki az f -et nevezi kódnak
- B^* egy véges részhalmaza az egy kód

Változó hosszúságú kódot akkor alkalmazzuk, ha a forrás-ABC betűi különböző előfordulási valószínűséget mutatnak. Fajtái:

1. *irreducibilis kódok:* az egyértelmű dekódolhatóság nem szükséges, de elégséges feltétele, hogy a kódrendszerben használt egyes kódszavak kiegészítése további kódjelekkel ne eredményezzen újabb kódszavakat.

2. *optimális hosszúságú kódok szerkesztése:* - a nagyobb valószínűségű betű kódja rövidebb, mint a kisebb valószínűségűé, - a kisebb hosszúságú kód minden variációs lehetősége ki van használva, - a kódfa ágain a megengedett csatorna jelek számával megegyező elágazás van, kivéve az utolsó csomópontokat.

Szabad félcsoporth

A egy nem üres véges ábécé ($0 \leq |A| \leq \infty$). Képezzük A^* -t (félcsoporth elemeinek halmaza). A művelet a szavak egymásután írása. Amit így kaptunk az egy félcsoporth. Ez egy szabad félcsoporth.

Legyen A egy halmaz, az úgynevezett ábécé, elemei a betűk. Képezzünk a betűkből (véges hosszú) szavakat, ezek közötti művelet legyen a konkatenáció, az egymásutánírás, amelyet asszociatívnak tekintünk. Kaptuk az úgynevezett szabad félcsoporthot.

Elméleti területen a félcsoporthok vizsgálata, gyakorlati felhasználása a kódoláselméletben a betűnkénti kódolásoknál.

Biztosan nincs kör, tehát a félcsoporthoz szabad. Az ekvivalenciaosztályok egy félcsoporthoz alkotnak.

A matematikában azon gráfokat nevezik **Cayley-gráfoknak**, amelyek egy csoport struktúráját reprezentálják. A Cayley-gráfok központi szerepet játszanak a kombinatorikában és a geometriai csoportelméletben. Arthur Cayley brit matematikus nevét őrzi az elnevezés.
Tulajdonságai:

- Egy adott csoporthoz tartozó Cayley-gráf nem egyértelmű, mert egy adott csoport generátorhalmaza sem egyértelmű.
- Ha a generátorhalmaz n elemű, akkor minden csúcsból pontosan n él indul ki és pontosan n él érkezik minden csúcsba.
- Ha a generátorhalmaz tartalmazza az egységelemet, akkor minden csúcs rendelkezik egy hurokéllel, ezért gyakran már a definícióban megkövetelik, hogy a generátorhalmaz ne tartalmazza az egységelemet.

Felbontható kódok

A egy véges, nem üres ábécé.

A^* az A feletti véges szavak halmaza

$C \subseteq A^*$

C véges és nem üres

C egy felbontható kód

DEF 1:

$u_1, u_2, \dots, u_r, v_1, v_2, \dots, v_s \in C$

$u_1, u_2, \dots, u_r = v_1, v_2, \dots, v_s$

1. $r = s$

2. $u_i = v_i$

DEF 2:

Nézzük a C-t egy ábécének. C^* a C feletti szavak halmaza. (C^* ,konkatenáció) félcsoporthoz.

Ha (C^* ,k) egy szabad félcsoporthoz, akkor C felbontható.

FELBONTHATÓSÁG FELTÉTELEI:

1. prefix-mentes kódok
2. postfix-mentes kódok
3. vesszős kódok
4. állandó hosszúságú kódok

EGYENLETES KÓDOK

$C : S \rightarrow A^n$ injektív függvény (az összes kódszó n hosszúságú), ahol S az elsődleges szimbólumok halmaza.

C injektív: $C(s_k) = C(s_l) \rightarrow s_k = s_l$

PREFIXMENTES KÓDOK

A egy véges nem üres ábécé

$C \subseteq A^*$ C véges, nem üres

$u, v \in A^*$

u prefixe a v-nek, ha létezik $w \in A^*$, amire $uw = v$ $u|v$.

Kiderül:

1. | tranzitív
2. | reflexív
3. | antiszimmetrikus

DEF:

Ha C prefixmentes, akkor nincs C-ben 2 olyan különböző szó u, v úgy, hogy $u|v$.

Ha $\varepsilon \in C \rightarrow C$ nem lehet prefixmentes. A prefixmentességet **prefixnek** rövidítjük.

Ha C prefixmentes, akkor C felbontható.

Prefix=egyértelműen megfejthető.

U **postfixe** v-nek ha létezik $w \in A^*$, amire $wu = v$.

VESSZŐS KÓD

$C = \{a, ab, abb\}$ ez egy vesszős kód. Itt most „a” a vessző. Általában v a vessző
 $v \in A^*$

- $v \notin \mathcal{E}$
- v postfixe (prefixe) minden szónak

Egy vesszős kód mindig felbontható. Állandó hosszúságú kódok mindig felbonthatóak.

Egy felbonthatatlansági feltétel

Ha $M \leq 1$, akkor a C felbontható, ha $M > 1$, akkor C felbonthatatlan.

A véges, nem üres ábécé.

$|A|=r$

$0 < r < \infty$

$C \subseteq A^*$, C véges nem üres, $|C|=s$

$C = \{c_1, \dots, c_s\}$

$\lambda(c_1) = c_1$ hossza Ha $c_1 = aa \rightarrow \lambda(c_1) = 2$

$M = r^{-\lambda(c_1)} + \dots + r^{-\lambda(c_s)}$

HUFFMANN-KÓDOK

- Speciális prefix kód
- Nem véletlenszerű a fa
- Gyakoribb betűk rövidebbek
- Átlagos betűhosszúság: $\sum l_i \cdot p_i$, ahol l a kódhossz, p a valószínűség.

Kraft-McMillan egyenlőtlenség:

Tétel: Ha a $K = \{k_1, k_2, \dots, k_n\}$ kód felbontható $\rightarrow \sum_{i=1}^n 2^{-l_i} \leq 1$, ahol l_i a k_i kódszó hossza.

Biz: Emeljük fel az összeget a t -ik hatványra.

$$\left(\sum_{i=1}^n 2^{-l_i} \right)^t = \left(\sum_{i=1}^n 2^{-l_{i1}} \right) * \left(\sum_{i=2}^n 2^{-l_{i2}} \right) * \dots * \left(\sum_{i=t}^n 2^{-l_{it}} \right) = \sum_{i_1, i_2, \dots, i_t} 2^{-(l_{i_1} + l_{i_2} + \dots + l_{i_t})}, \text{ ahol } 1 \leq i_j \leq n \text{ és } 1 \leq j \leq t$$

Az $l_{i_1} + l_{i_2} + \dots + l_{i_t}$ hossz-sorozat nem más, mint az $\{a_{i_1}, a_{i_2}, \dots, a_{i_t}\}$ t hosszú elsődleges közlés kódolásának hossza. Jelöljük ezt $L(a_{i_1}, a_{i_2}, \dots, a_{i_t})$ -vel.

$$\dots = \sum_{a_{i_1}, a_{i_2}, \dots, a_{i_t}} 2^{-L(a_{i_1}, a_{i_2}, \dots, a_{i_t})}, \text{ vagyis az összes } t \text{ hosszú elsődleges közlésre összegzünk.}$$

- Jelölje $N(t, r)$ azoknak a t hosszúságú elsődleges közléseknek a számát, amelyekre $L(a_{i_1}, a_{i_2}, \dots, a_{i_t}) = r$, és legyen M a K belüli kódszavak hosszának a maximuma. Ekkor
- $t \leq L(a_{i_1}, a_{i_2}, \dots, a_{i_t}) \leq M \cdot t$

$$\sum_{a_{i_1}, a_{i_2}, \dots, a_{i_t}} 2^{-L(a_{i_1}, a_{i_2}, \dots, a_{i_t})} = \sum_{r=t}^{M \cdot t} N(t, r) \cdot 2^{-r} \dots$$

(áttérünk kódszó-bithossz szerinti összegzésre)

- Másrészt: $N(t, r) \leq 2^r$, hiszen 2^r -nél több szimbólum felbontható módon nem kódolható le r bittel. Ezért:

$$\dots \leq \sum_{r=t}^{M \cdot t} 2^r \cdot 2^{-r} = \sum_{r=t}^{M \cdot t} 1 = (M-1) \cdot t + 1$$

- Vagyis tetszőleges t -re: $\left(\sum_{i=1}^n 2^{-l_i} \right)^t \leq (M-1) \cdot t + 1$

- Ha $\left(\sum_{i=1}^n 2^{-li}\right) > 1$ lenne igaz, akkor eléggé nagy t -re a fenti egyenlőtlenség nem állhatna fenn.
- Tehát $\left(\sum_{i=1}^n 2^{-li}\right) = 1$ áll fenn.

ÁLLANDÓ HOSSZÚSÁGÚ KÓDOK

Eddig a kódok $f: A \rightarrow B^*$ függvények voltak, ahol f injektív.

A: forrásábécé

B: csatornaábécé

A helyzet változik, mert itt $f: T^n \rightarrow T^k$, ahol $n < k$ $T = \{0, 1\}$

T^n : forrásábécé

T^k : csatornaábécé

Mod 2 adunk össze és szorzunk.

Hibajelzésre és hibajavításra használják.

Szerkesszünk egy G gráfot!

- G csúcsai T^k elemei
- Két különböző pontot u, v összekötünk, ha $u \in H_v$ és $v \notin H_u$
- Keresünk egy klikket G -ben (bármely pont bármely másikkal össze van kötve)
- Legyen C egy klikk, ekkor C csúcsai alkalmasak egy hibajelző kódnak.

HIBAJELZÉS

- $f: T^n \rightarrow H^k$ injektív
- $T = \{0, 1\}$
- $k > n$
- $|T^k| = 2^k$
- $|T^n| = 2^n$

T^k -nak sokkal több eleme van, mint T^n -nek. Nem használjuk a teljes T^k -t csak annak egy C részhalmazát.

$u \in T^k$ (u a hibatartomány) (t sugarú)
 $H_u = \{v: v \in T^k \text{ és } v \text{ legfeljebb } t \text{ jegyben tér el } n\text{-től}\}$

DEF:

- I. $C \subseteq T^k$

Azt mondjuk, hogy C egy t -hibajelző kód, ha

1. $c_1 \in C, c_2 \in T^k, c_2 \in H_{c_1}$
2. Ha $c_1 \neq c_2 \Rightarrow c_2 \notin C$

Hogyan jelez ez hibát?

- Jön egy $u \in T^k \rightarrow H$ ismerjük C -t
- Feltesszük, hogy t -nél több jegy nem változik meg a küldés során
- Ha $u \notin C$, akkor biztosan hiba történt

II. $C \subseteq T^k$ azt mondjuk, hogy C egy t -hibajelző kód, ha:

- $c_1, c_2 \in C$
- $c_1 \neq c_2 \Rightarrow c_1 \approx c_2$

A két definíció egyenértékű.

HAMMING TÁVOLSÁG

$u(u_1, \dots, u_n) \in T^n$

$v(v_1, \dots, v_n) \in T^n$

Egy távolságot értelmezünk $d(u, v)$ = azon pozíciók száma, ahol az u és a v eltér.

d tényleg egy távolság.

$f: T^n \rightarrow T^k$ $T = \{0,1\}$ $k > 0$

$C = \text{im} f$ ezt **kódnak** hívjuk.

Van olyan eset, amikor a C zárt az összeadásra (és kivonásra). Ilyenkor C lineáris altere T^k -nak $\rightarrow k$ dimenziós lineáris tér T felett.

C -nek sok pontja van. Ha C dimenziója s , akkor $|C| = 2^s$. G egy generátormátrixa C -nek, ha G oszlopai bázist alkotnak C -ben.

HAMMING KÓD

- Tétel: egy K lineáris kóddal pontosan akkor lehet t hibát kijavítani, ha a K ellenőrző mátrixában tetszőlegesen kiválasztva $2t$ oszlopot, ezek lineárisan függetlenek lesznek.
- Legyen a K kód blokkmérete $n = 2^l - 1$. Álljon a K kód ellenőrző mátrixa az összes l hosszú, > 0 vektorból. Ilyenkor semelyik 2 oszlop nem lineárisan összefüggő, vagyis a kódtávolság $d(K) > 2$, vagyis legalább 1 hibát javítani tudunk. Az ilyen kódokat Hamming kódoknak nevezzük.
- A Hamming kód dimenziója: $n - l = 2^l - l - 1$.
- A Hamming kód sűrűsége: $(2^l - l - 1) / (2^l - 1) = 1 - (l + 1) / 2^l \approx 1 - \log(n) / n$ Ez lényegesen jobb, mint a korábbi példa

A hibajelzés tesztelhető távolságszámítással:

$C \subseteq T^k$

Kiszámoljuk $L = \min\{d(c_1, c_2) : c_1, c_2 \in C \text{ és } c_1 \neq c_2\}$. L a C halmaz min belső távolsága.

Ha C t -hibajelző, akkor $L \geq t + 1$

Ha $L \geq t + 1$, akkor C t -hibajelző

Azaz C t -hibajelző $\Leftrightarrow L \geq t + 1$

LINEÁRIS KÓD

Lineáris kódok alatt azokat a kódokat értjük, amelyek kódvektorai egy lineáris térnek az elemei. Talán pontosabban fogalmazva: a kódszavak egy lineáris tér lineáris alterét alkotják. De mit is jelent az, hogy egy tér lineáris? Ez szavakban kimondva azt jelenti, hogy a tér akárhány vektorát összeadva, (koordinátánkénti összegzést végezve!) az eredő vektor bennmarad a térben.

Van még egy alapvető fogalom a kódok jellemzésére, ez pedig a kódszavak súlya. Egy vektor súlya alatt a koordinátaiban szereplő nem nulla elemek számát értik. Egy kódszókészlet, vagy röviden egy kód minimális súlyán pedig a legkönnyebb, de nem nulla vektor súlyát értik.

Egy lineáris kód esetén a kódtávolság megegyezik a kód minimális súlyával.

Elemi példánkban a nem nulla súlyú kódszavak mind kettő súlyúak, és ezzel egyenlő a kódtávolság is.

A lineáris térről elmondottak már szinte tökéletesen elegendőek a lineáris kód előállításához.

Megkeressük az N dimenziós bináris tér egyik bázisát $(\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_K)$, és ezzel bármelyik \mathbf{u} üzenetvektorhoz generáljuk a megfelelő \mathbf{c} kódszót a lineáris tér bázisára vonatkozó alábbi összefüggés alapján:

$$\mathbf{c} = \sum_{i=1}^K u_i \mathbf{g}_i.$$

Ezt az egyenletet mátrixalakban is felírhatjuk:

$$\mathbf{c} = \mathbf{u} \cdot \mathbf{G}.$$

GENERÁTORMÁTRIX

Egy kód generátormátrixán azt a mátrixot értjük, amelyet balról beszorozva tetszőleges vektorral a kód egy kódszavát kapjuk.

A \mathbf{G} mátrixot generátormátrixnak hívják, és sorai az N dimenziós tér bázisvektorai, pontosabban azok közül K darab. Ilyen mátrix több is lehet, amelyek közül kitéüntetett szerepe van az ún. szisztematikus generátormátrixnak, amely szisztematikus kódot generál.

Szisztematikusnak nevezik a kódot, ha a kódszó első (vagy utolsó) K eleme megegyezik az üzenettel. Könnyű belátni, hogy ezt olyan \mathbf{G} mátrix hozza létre, amelynek első (vagy utolsó) K oszlopa egységmátrix.

BCH KÓD

BCH=Bose-Chaudhuri-Hocquenghem

D BCH kód

Az n kódszóhosszú, $n = q^m - 1$, $GF(q)$ feletti, t hibát javító kódot BCH kódnek nevezzük, ha $g(x)$ generátor polinomjának gyökei az $\alpha^i \in GF(q^m)$, $i = 1, 2, \dots, 2t$ testelemek.

T Ha az n kódszóhosszú, $GF(q)$ feletti C ciklikus kód generátorpolinomjának az $\alpha \in GF(q^m)$ $d-1$ elem egymás utáni (különböző) hatványa gyöke, azaz valamely $i_0 \geq 0$, $d > 1$ esetén.

$$g(\alpha^{i_0}) = g(\alpha^{i_0+1}) = \dots = g(\alpha^{i_0+d-2}) = 0,$$

akkor a kód minimális távolsága legalább d .

Generálás

Egy t hibát javító BCH kód generátorpolinomját úgy konstruálhatjuk meg, hogy megkeressük az $\alpha^i \in GF(q^m)$ gyökök különböző $GF(q)$ feletti minimálpolinomjait, s azokat összeszorozzuk.

Tehát a BCH kód generátorpolinomja:

$$g(x) = \Phi_1(x) \Phi_3(x) \dots \Phi_{2t-1}(x)$$

Ha $q=2$ akkor bináris BCH-kódot kapunk, míg $m=1$ esetben Reed-Solomon-kódra jutunk. A $t=1$ $q=2$ választással kapjuk a bináris Hamming-kódokat. Mivel a $GF(q)$ feletti polinomok között az $\alpha \in GF(q)$ testelem az $x - \alpha$ elsőfokú polinom gyöke, így a t hibát javító RS kód generátorpolinomja lehet pl a $g(x) = (x-1)(x-\alpha)(x-\alpha^2) \dots (x-\alpha^{2t-1})$ polinom, ahol α rendje n .