

Insinöörimatematiikka 3

Mika Hirvensalo
`mikhirve@utu.fi`

Matematiikan ja tilastotieteen laitos
Turun yliopisto

2022

Esimerkki

- "Pysäyttäkää auto JBN-372"
- "Pysäyttäkää auto IPM-372"
- JBN → "Jaakko – Bertta – Niilo"
- IPM → "Ilmari – Pekka – Mika"
- Redundantti informaatio parantaa luotettavuutta

Tiedonsiirto

- Sähkömagneettinen säteily
- Ääniaallot
- Analoginen ↔ digitaalinen
- Digitaalisessa muodossa informaation suojaaminen virheiltä on helpompaa.

Audio CD-standardi

- Kaksi äänikanavaa, 44100 Hz, 16 bitin näytteet.
- $2^{16} = 65536$
- $2 \cdot 44100 \cdot 16 = 1411200 \text{ bit/s} = 172,266 \text{ KB/s}$
- 5 minuuttia vaatii $5 \cdot 60 \cdot 172,266 \text{ KB} \approx 50 \text{ MB}$
- 74 minuuttia vaatii $74 \cdot 60 \cdot 172,266 \text{ KB} \approx 746 \text{ MB}$

4K kuvaformaatti

- $3840 \times 2160 = 8294400$ pikseliä
- Joka pikseliä varten RGB-komponentit vaativat $3 \cdot 8 = 24$ bittiä.
- Yhteensä $24 \cdot 8294400 \text{ bit} \approx 23 \text{ MB}$
- 25 kuvaa sekunnissa vaatii n. 593 MB/s

Digitaalisuus

Digitaalisesti esitettävä informatio merkitsee sitä, että ainoastaan äärellinen määrä erilaisia symboleita on käytettävissä.

Minimalistisessa tapauksessa jolloin symboleja on vain kaksi, sanotaan informaation olevan binääristä. Binäärisen informaation perusyksikkö on *bitti*, jolla voi olla arvo 0 tai 1.

Laajempia symbolijoukkoja voidaan esittää bittijonoilla.

Esimerkiksi kahden bitin jonoilla voidaan esittää neljä erilaista symbolia: 00, 01, 10, ja 11.

Digitaalinen tiedonsiirto

Monenlaiset häiriötekijät voivat aiheuttaa virheitä tiedonsiirtoon, jolloin digitaalinen viesti 1110100111010111 voi muuntua esim. muotoon 1110100111010101. Virheellinen viesti pitäisi voida tunnistaa ja siitä pitäisi voida päätellä alkuperäinen.

Virheitä korjaavat koodit

Esimerkki: Pariteettibitti

Lisätään 7-pituisten bittijonojen perään 0 tai 1 siten, että ykkösten määrä tulee olemaan parillinen. Esimerkiksi $0010100 \mapsto 00101000$ ja $0101100 \mapsto 01011001$. Jos näin saaduissa jonoissa yksi bitti vaihtuu, voidaan todeta virhe mutta ei palauttaa alkuperäistä.

Esimerkki: Toistokoodi

$0 \rightarrow 000$, $1 \rightarrow 111$, jolloin esimerkiksi

$$01001 \mapsto 000111000000111.$$

Tällöin yksi virhe muuntaa viestin esim. muotoon

$$000111000000101,$$

joka voidaan dekodata viestiksi $000111000000111 \rightarrow 01001$.

Virheitä korjaavat koodit

Toistokoodi

$0 \rightarrow 000$, $1 \rightarrow 111$. Hintana kolminkertainen lähetysaika, 1 virhe korjattavissa / 3-pituinen lohko

Esimerkki

Koodaus $00 \mapsto 00000$ $01 \mapsto 00111$, $10 \mapsto 11100$, $11 \mapsto 11011$.
hintana 2,5-kertainen lähetysaika, 1 virhe korjattavissa / 5-pituinen lohko.

Määritelmä

- $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$ on kahden alkion kunta.
- $\bar{0} = 0 + 2\mathbb{Z}$, $\bar{1} = 1 + 2\mathbb{Z}$.
- Esim. $\bar{0} + \bar{1} = \bar{1}$, $\bar{1} + \bar{1} = \bar{0}$
- Jätetään yläviivat pois ja merkitään $\mathbb{F}_2 = \{0, 1\}$, missä $1 + 1 = 0$.
- n -ulotteinen vektoriavaruus $\mathbb{F}_2^n$ yli kunnan $\mathbb{F}_2$ koostuu n -pituisista järjestetyistä kunnan $\mathbb{F}_2$ jonoista (bittijonoista):
 $\mathbb{F}_2^n = \{(x_1, \dots, x_n) \mid x_i \in \mathbb{F}_2\}$, jossa yhteenlasku ja skalaarikertolasku on määritelty kuten reaalisisessa.

Määritelmä

Vektoreiden $\mathbf{x} = (x_1, \dots, x_n)$ ja $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{F}_2^n$

Hamming-etäisyys on

$$d_H(\mathbf{x}, \mathbf{y}) = |\{i \mid x_i \neq y_i\}|$$

Vektorin $\mathbf{x}$ Hamming-paino määritellään $w_H(\mathbf{x}) = d_H(\mathbf{x}, \mathbf{0})$.

Esimerkki

$$d_H(00000, 00111) = 3, d_H(00000, 11100) = 3,$$

$$d_H(00111, 11100) = 4, d_H(11100, 11011) = 3$$

$$d_H(00111, 11011) = 3 \text{ ja } d_H(00000, 11011) = 4$$

Huomautus

- Hamming-etäisyys toteuttaa metriikan aksioomat
- $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$
- Jos $d_H(\mathbf{x}, \mathbf{y}) \geq 3$ on $\mathbf{x}' = \mathbf{x} + \mathbf{e}_i$ edelleen lähempänä vektoria $\mathbf{x}$ kuin vektoria $\mathbf{y}$, joten virhe $\mathbf{x} \mapsto \mathbf{x}' = \mathbf{x} + \mathbf{e}_i$ voidaan korjata dekoodaamalla $\mathbf{x}'$ vektoriksi $\mathbf{x}$.
- Yleistys: Jos $d_H(\mathbf{x}, \mathbf{y}) \geq 2t + 1$, voidaan t :n virheen jälkeen saatu $\mathbf{x}'$ palauttaa vektoriksi $\mathbf{x}$, sillä $\mathbf{y}$ on silti vielä kauempana.

Virheitä korjaavat koodit

Määritelmä

$C \subset \mathbb{F}_2^n$ on t virhettä korjaava koodi, jos $d_H(\mathbf{x}, \mathbf{y}) \geq 2t + 1$ aina kun $\mathbf{x} \neq \mathbf{y} \in C$.

Periaate

Viestit $\mathbf{x} \in \mathbb{F}_2^m$ (pituus m) koodataan *koodisanoiksi* (pituus $n > m$), jotka ovat etäällä toisistaan. Mikäli tiedonsiirtokanava aiheuttaa virheitä koodisanoissa, on toivottavaa että muuttuneet viestit yhä muistuttaisivat eniten lähetettyjä koodisanoja ja voitaisiin täten dekoddata.

Määritelmä

Koodia C sanotaan (n, M, d) -koodiksi, jos koodisanojen pituus on n , $|C| = M$, ja erisuurten koodisanojen Hamming-etäisyys on vähintään d .

Periaate

Kuvataan lyhyemmät jonot (pituus m) bijektiivisesti pidemmiksi jonoiksi (pituus n),

$$\mathbb{F}_2^m \rightarrow C \subset \mathbb{F}_2^n$$

siten että kaikkien $\mathbf{x} \neq \mathbf{y} \in C$ Hamming-etäisyys on ainakin $2t + 1$. Mikäli kuvavektorin lähetys tiedonsiirtokanavassa $\mathbf{x} \mapsto \mathbf{x}' \notin C$ tuottaa korkeintaan t virhettä, voidaan alkuperäinen vektori $\mathbf{x}$ saada yksikäsitteisesti lähimpänä Hamming-etäisyyden vektorina jolle $\mathbf{x} \in C$.

Ongelmia

- Miten valita *koodisanojen* joukko $C \subseteq \mathbb{F}_2^n$ siten että $d_H(\mathbf{x}, \mathbf{y})$ olisi suuri aina kun $\mathbf{x} \neq \mathbf{y} \in \mathbb{F}_2^m$.
- Miten laskea koodaus $\mathbb{F}_2^m \rightarrow C$ tehokkaasti?
- Miten laskea dekodaus $\mathbb{F}_2^n \rightarrow C$ tehokkaasti?
- Miten laskea kuvaus $C \rightarrow \mathbb{F}_2^m$ tehokkaasti?

Ongelmia

- Monissa sovelluksissa dekodaus $\mathbb{F}_2^n \rightarrow C$ on hankalinta toteuttaa tehokkaasti.
- Osoittautuu, että informaatio suhde paranee lohkon pituutta kasvatettaessa, mutta laskennalliset ongelmat kasvavat myös.

Virheitä korjaavat koodit

Esimerkki: Taulukointi

Koodi: {00000, 00111, 11100, 11011}, dekooodaus

00000 → 00000

00001 → 00000

00010 → 00000

00011 → 00111

...

01001 → 00000 tai 11011

...

11110 → 11100

11111 → 11011

Bittijonon pituuden ollessa $n \geq 30$ on taulukoinnin mahdollisuus suljettu käytännössä pois: $2^{30} = 1\,073\,741\,824$

Määritelmä

Koodi $C \in \mathbb{F}_2^n$ on *lineaarinen*, jos $C \subseteq \mathbb{F}_2^n$ on aliavaruus. Tällöin $\alpha \mathbf{x} + \beta \mathbf{y} \in C$ aina, kun $\mathbf{x}, \mathbf{y} \in C$ ja $\alpha, \beta \in \mathbb{F}_2 = \{0, 1\}$.

Huomautus

Jos C on avaruuden $\mathbb{F}_2^n$ m -ulotteinen aliavaruus ($m < n$), on sillä kanta $\{\mathbf{c}_1, \dots, \mathbf{c}_m\} \subset \mathbb{F}_2^n$ ja jokainen $\mathbf{c} \in C$ voidaan esittää yksikäsitteisesti muodossa

$$\mathbf{c} = x_1 \mathbf{c}_1 + \dots x_m \mathbf{c}_m = (x_1, \dots, x_m) \begin{pmatrix} \mathbf{c}_1 \\ \dots \\ \mathbf{c}_m \end{pmatrix} = (x_1, \dots, x_m) G.$$

Matriisia $G = \begin{pmatrix} \mathbf{c}_1 \\ \dots \\ \mathbf{c}_m \end{pmatrix}$ kutsutaan koodin *generoijamatriisiksi*.

Huomautus

Lineaarisen koodin *koodausfunktio* $\mathbb{F}_2^m \rightarrow \mathbb{F}_2^n$ voidaan esittää matriisikertolaskuna

$$\mathbf{c} = \mathbf{x}G,$$

missä $\mathbf{x} \in \mathbb{F}_2^m$ on lähetettävä viesti (esitetään rivivektorina), G on $m \times n$ -matriisi yli kunnan $\mathbb{F}_2$, ja $\mathbf{c} \in C \subset \mathbb{F}_2^n$ on koodattu viesti. Tällöin kyseessä on $(n, 2^m, d)$ -koodi, jossa minimietäisyys d riippuu generoijamatriisin ominaisuuksista.

Määritelmä

Lineaarikoodia, jonka pituus on n ja dimensio m ja minimietäisyys d , sanotaan $[n, m, d]$ -koodiksi. Jos minimietäisyyttä ei ilmoiteta, sanotaan koodia $[n, m]$ -koodiksi.

Lause

Jos $C \subseteq \mathbb{F}_2^n$, on $[n, m]$ -koodi, on olemassa sellainen $(n - m) \times n$ -matriisi H , että

$$C = \{\mathbf{x} \in \mathbb{F}_2^n \mid H\mathbf{x}^T = 0\}.$$

Matriisia H kutsutaan koodin C tarkistusmatriisiksi. Jos G on koodin generoijamatriisi, on $GH^T = O$ nollamatriisi.

Huomautus

Tarkistusmatriisi voidaan muodostaa generoijamatriisista ja päinvastoin.

Syndromit

- Jos $\mathbf{x} \in \mathbb{F}_2^n$ toteuttaa $H\mathbf{x}^T = \mathbf{0}$, on $\mathbf{x} \in C$ koodisana, eikä virhettä tulkita tapahtuneen.
- Jos $H\mathbf{x}^T \neq \mathbf{0}$, tulkitaan virhe tapahtuneeksi.

Lineaariset koodit

Jos $\mathbf{c} \in C$ ja $\mathbf{x} = \mathbf{c} + \mathbf{e}$ on "virheellinen", on kuitenkin $H\mathbf{x}^T = H(\mathbf{c} + \mathbf{e})^T = H\mathbf{c}^T + H\mathbf{e}^T = H\mathbf{e}^T$. Virheenkorjaus perustuu vektoreihin $H\mathbf{e}^T$. näitä kutsutaan *syndromeiksi*.
Ns. Syndromipankki muodostetaan tallentamalla muistiin Hamming-painon suhteen kaikkein kevyimmät syndromin tuottavat vektorit $\mathbf{x} \in \mathbb{F}_2^n$. Dekoodaus perustuu näihin, *johtajiksi* kutsuttuihin vektoreihin.

Dekoodaus

- Vastaanotettaessa (mahdollisesti) virheellinen vektori $\mathbf{x} \in \mathbb{F}_2^n$, lasketaan $\mathbf{e} = H\mathbf{x}^T$ ja valitaan syndromipankista vektori $\mathbf{s}$ joka toteuttaa ehdon $H\mathbf{s}^T = \mathbf{e}$.
- Dekoodataan vektori $\mathbf{x}$ vektoriksi $\mathbf{x} - \mathbf{s}$.

Esimerkki: Hammingin [7, 4]-koodi

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

on [7, 4]-koodin tarkistusmatriisi (Huomio että H :n sarakkeet ovat lukujen 1–7 binääriesitykset). Generoijamatriisi voidaan muodostaa ehdon $G^T H = O$ perusteella, mutta dekooodaus edellyttää syndromipankin luomista. Voidaan valita

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Syndromit

$$H(0000000)^T = (000)^T$$

$$H(0000001)^T = (001)^T$$

$$H(0000010)^T = (010)^T$$

...

$$H(1000000)^T = (110)^T$$

Esimerkki: Hammingin koodi

$$(1, 0, 1, 1) \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} = (1, 0, 1, 1, 0, 1, 0).$$

Virhe lähetyksessä

$$(1, 0, 1, 1, 0, 1, 0) \mapsto (1, 0, 1, 1, 0, 0, 0)$$

Esimerkki

$$H(1, 0, 1, 1, 0, 0, 0)^T = (0, 1, 0),$$

jota vastaava syndromipankin alkio on $(0, 0, 0, 0, 0, 1, 0)$.

Dekoodattu vektori on siis $(1, 0, 1, 1, 0, 1, 0)$.

Hammingin koodi

- $[7, 4, 3]$ -koodi
- Hintana $7/4 = 1.75$ -kertainen lähetysaika
- 1 virhe korjattavissa / 7- pituinen lohko

Hammingin koodit

Jokaista $r \geq 1$ kohti on olemassa $[2^r - 1, 2^r - r - 1, 3]$ -koodi, jonka tarkistusmatriisin sarakkeet muodostetaan kaikista r -bittisten lukujen $\neq 0$ binääriesityksistä.

Huomautus

Hammingin koodit korjaavat lähtökohtaisesti vain yhden virheen, mutta osoittavat, että informaationsuhde voidaan saada korkeaksi: Hintana $(2^r - 1)/(2^r - r - 1)$ -pituisen lähetysaika