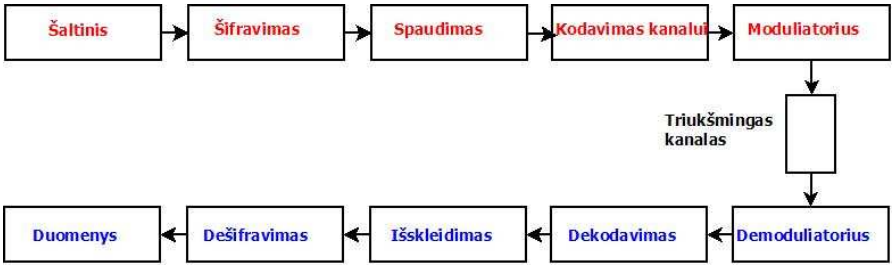
Kodavimo teorija

Vilius Stakėnas

2010 metų ruduo

Duomenų perdavimo schema



3 / 27

Klaidos ir jų taisymas

duomenų žodis $= x_1 x_2 \dots x_k$

$\xrightarrow{\text{į kanalą}} x_1 x_2 \dots x_k \xrightarrow{\text{iš kanalo}} x_1^* x_2^* \dots x_k^*$

Klaidų aptikti neįmanoma!

duomenų žodis $= x_1 x_2 \dots x_k$

$\xrightarrow{\text{į kanalą}} y_1 y_2 \dots y_k y_{k+1} \dots y_n \xrightarrow{\text{iš kanalo}} y_1^* y_2^* \dots y_n^*$

Klaidas galima aptikti, kartais - ištaisyti!

4 / 27

Klaidos ir jų taisymas

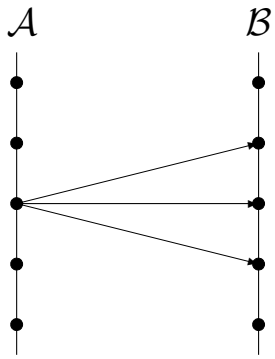
duomenų žodis $= x_1 x_2 \dots x_k$

$\xrightarrow{\text{į kanalą}} y_1 y_2 \dots y_k y_{k+1} \dots y_n \xrightarrow{\text{iš kanalo}} y_1^* y_2^* \dots y_n^*$

Kodo koeficientas

$$R = \frac{k}{n}.$$

Siuntėjai, kanalai, gavėjai



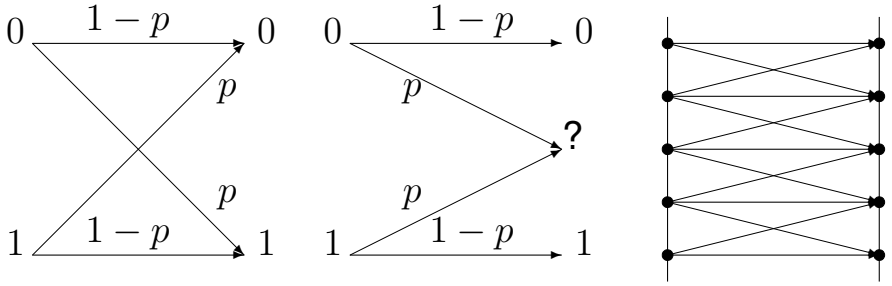
Šaltinio abėcėlė $\mathcal{A}_q$, gavėjo abėcėlė $\mathcal{B}_r$:

$$\mathcal{A}_q = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_q\}, \quad \mathcal{B}_r = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_r\}$$

Dažniausiai

$$\mathcal{A} = \mathcal{B} = \{0, 1\}$$

Kanalų įvairovė



7 / 27

Žodžiai ir perdavimo tikimybės

Žodžių aibės

$$\begin{aligned} \mathcal{A}_q^n &= \{\mathbf{a} = a_{i_1}a_{i_2}\ldots a_{i_n} : a_{i_j} \in \mathcal{A}_q\}, \\ \mathcal{B}_r^m &= \{\mathbf{b} = b_{i_1}b_{i_2}\ldots b_{i_m} : b_{i_j} \in \mathcal{B}_r\} \end{aligned}$$

$\mathbf{X}$ - siunčiamas žodis, $\mathbf{Y}$ - gaunamas žodis.

Tikimybės

$$\begin{aligned} p(\mathbf{a}) &= P(\mathbf{X} = \mathbf{a}), \quad p(\mathbf{b}) = P(\mathbf{Y} = \mathbf{b}), \\ p(\mathbf{a}, \mathbf{b}) &= p(\mathbf{X} = \mathbf{a}, \mathbf{Y} = \mathbf{b}), \quad p(\mathbf{b}|\mathbf{a}) = p(\mathbf{Y} = \mathbf{b}|\mathbf{X} = \mathbf{a}), \\ p(\mathbf{a}|\mathbf{b}) &= p(\mathbf{X} = \mathbf{a}|\mathbf{Y} = \mathbf{b}). \end{aligned}$$

8 / 27

Kanalai be atminties

Žymėjimai: $U^{(n)} \in \mathcal{A}^n$ – siunčiamas žodis,
$V^{(n)} \in \mathcal{B}^n$ – gaunamas žodis.

Apibrėžimas. Perdavimo kanalą vadinsime kanalu be atminties, jeigu visiems žodžiams $\mathbf{u} = u_1 u_2 \dots u_n \in \mathcal{A}^n$, $\mathbf{v} = v_1 v_2 \dots v_n \in \mathcal{B}^n$ teisinga lygybė

$$P(V^{(n)} = \mathbf{v} | U^{(n)} = \mathbf{u}) = p(v_1|u_1)p(v_2|u_2)\dots p(v_n|u_n),$$

čia $p(v_j|u_j)$ reiškia tikimybę, kad pasiuntus simbolį u_j gautas simbolis v_j.

Entropija

$\mathbf{X}, \mathbf{Y}$ - siųsti ir gauti simboliai arba žodžiai
Entropija – šaltinio perduodamos informacijos kiekybinis matas:

$$\begin{aligned} H(\mathbf{X}) &= \sum_{\mathbf{a}} \log_2 \frac{1}{p(\mathbf{a})} p(\mathbf{a}) \\ H(\mathbf{Y}) &= \sum_{\mathbf{b}} \log_2 \frac{1}{p(\mathbf{b})} p(\mathbf{b}) \\ H(\mathbf{X}|\mathbf{Y} = \mathbf{b}) &= \sum_{\mathbf{a}} \log_2 \frac{1}{p(\mathbf{a}|\mathbf{b})} p(\mathbf{a}|\mathbf{b}) \\ H(\mathbf{X}|\mathbf{Y}) &= \sum_{\mathbf{Y}=b} H(\mathbf{X} = \mathbf{b}|\mathbf{b}) p(\mathbf{b}) \end{aligned}$$

Kanalo talpa

Perduodamos informacijos kiekis ir kanalo talpa

$$\begin{aligned} I(\mathbf{X}|\ \mathbf{Y}) &= H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y}), \\ \mathcal{C} &= max\{I(\mathbf{X}|\ \mathbf{Y}) : \mathbf{X} \text{ tikimybių skirstinys } \} \end{aligned}$$

Kodas ir dekodavimo taisyklė

Abėcėlės $\mathcal{A}$ n ilgio žodžių kodu vadinsime bet kokį poaibį $\mathbf{C} \subset \mathcal{A}^n$.

Apibrėžimas. Dekodavimo taisykle vadinsime funkciją

$$f : \mathcal{A}^n \to \mathbf{C}.$$

Klaidos tikimybė

Žymime: $\mathbf{a}$ siunčiamus, $\mathbf{b}$ – gaunamus žodžius.

$$\begin{aligned} P(\text{klaida}|\mathbf{a}) &= \sum_{\substack{b \\ f(b)\neq a}} p(\mathbf{b}|\mathbf{a}) = 1 - \sum_{\substack{b \\ f(b)=a}} p(\mathbf{b}|\mathbf{a}) \\ P(\text{klaida}) &= \sum_{a} P(\text{klaida}|\mathbf{a}) p(\mathbf{a}) \end{aligned}$$

Idealaus stebėtojo taisyklė

Jeigu pasirinktas kodas **C** , tai į kanalą siunčiami tik šio kodo žodžiai **c** .

Apibrėžimas. Dekodavimo taisyklę $f : \mathcal{A}^n \to \mathrm{C}$ vadinsime idealaus stebėtojo taisykle, jeigu kiekvienam $\mathbf{b} \in \mathcal{A}^n$ tenkinama sąlyga

$$p(f(\mathbf{b})|\mathbf{b}) = \max\{p(\mathbf{c}|\mathbf{b}) : \mathbf{c} \in \mathbf{C}\}.$$

Didžiausio tikėtinumo taisyklė

Apibrėžimas. Dekodavimo taisyklę $f : \mathcal{A}^n \to \mathrm{C}$ vadinsime didžiausio tikėtinumo taisykle, jeigu kiekvienam $\mathbf{b} \in \mathcal{A}^n$ tenkinama sąlyga

$$p(\mathbf{b}|f(\mathbf{b})) = \max\{p(\mathbf{b}|\mathbf{c}) : \mathbf{c} \in \mathbf{C}\}.$$

Minimalaus atstumo taisyklė

Tegu $\mathbf{x}, \mathbf{y} \in \mathcal{A}^n, \mathbf{x} = x_1x_2 \dots x_n, \mathbf{y} = y_1y_2 \dots y_n,$

$$h(\mathbf{x}, \mathbf{y}) = \sum_{\substack{i=1,\dots,n \\ x_i \neq y_i}} 1.$$

Apibrėžimas. Dekodavimo taisyklę $f : \mathcal{A}^n \to \mathrm{C}$ vadinsime minimalaus atstumo taisykle, jeigu kiekvienam $\mathbf{b} \in \mathcal{A}^n$ tenkinama sąlyga

$$h(\mathbf{b}, f(\mathbf{b})) = \min\{h(\mathbf{b}, \mathbf{c}) : \mathbf{c} \in \mathbf{C}\}.$$

Dvinaris simetrinis kanalas

$$\begin{aligned} \mathcal{A}_q &= \mathcal{B}_r = \mathcal{B} = \{0,1\}, \quad p(0|1) = p(1|0) = p, \\ \mathcal{C} &= 1 - p\log_2 p - (1-p)\log_2(1-p) \end{aligned}$$

Klaidingo dekodavimo tikimybė

Tarkime, informacija yra užrašyta abėcėlės $\mathcal{B} = \{0,1\}$ abėcėlės žodžiais, o ją reikia perduoti simetriniu be atminties kanalu, kuris su tikimybe p $(0 < p < 1)$ kiekvieną simbolį iškreipia.
Į kanalą siunčiami kodo $\mathbf{C} = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}, \mathbf{x}_i \in \mathcal{B}^n$, dekodavimui naudojama minimalaus atstumo taisyklė.

Vidutinė klaidingo dekodavimo tikimybė:

$$P_C = \frac{1}{M}\sum_{i=1}^{M} P(\text{klaidingai dekoduota}|\mathbf{x}_i).$$

Shannono teorema

Teorema. (Claude Shannon, 1948) Tegu skaičius R tenkina nelygybę

$$0 < R < 1 - p\log_2\frac{1}{p} - q\log_2\frac{1}{q}, \quad q = 1-p.$$

Tada egzistuoja kodai $\mathbf{C}_n = \{x_1, \ldots, x_{M_n}\} \subset \{0;1\}^n$, $M_n = 2^{[nR]}$, kad $P_{C_n} \to 0$, kai $n \to +\infty$.

Kodo koeficientas

Apibrėžimas. Kodo $\mathbf{C} \subset \mathcal{B}^n$ koeficientu vadinamas skaičius

$$R(\mathbf{C}) = \frac{\log_2 |\mathbf{C}|}{n}.$$

Kartojimo kodai

$$x \in \{0, 1\}, \quad x \to \underbrace{xx \ldots x}_{2n+1}$$

$$R = \frac{1}{2n+1}$$

Stačiakampių kodai

$St(n,m), \quad n=4, m=3$

$$x_1x_2\ldots x_{nm} \to \begin{matrix} x_1 & x_2 & x_3 & x_n & y_1 \\ x_5 & x_6 & x_7 & x_{2n} & y_2 \\ x_9 & x_{10} & x_{11} & x_{mn} & y_m \\ z_1 & z_2 & z_3 & z_n & z \end{matrix}$$

$$x_1+x_2+x_3+x_4+y_1=0, \quad x_1+x_5+x_9+z_1=0,$$
$$x_5+x_6+x_7+x_8+y_2=0, \quad x_2+x_6+x_{10}+z_2=0,$$
$$x_9+x_{10}+x_{11}+x_{12}+y_3=0, \quad x_3+x_7+x_{11}+z_3=0,$$
$$z_1+z_2+z_3+z_4+z=0$$

$$y_1+y_2+y_3+z_1+z_2+z_3+z_4+z=0$$

$St(m,n)$

$$m=3, \quad n=4,$$
$$x_1x_2\ldots x_{12} \quad \to \quad x_1x_2\ldots x_{12}y_1y_2y_3z_1z_2z_3z_4z,$$
$$R=\frac{3}{5}$$

Bet kurie du $St(m,n)$ žodžiai skiriasi bent 3 bitais!
Vieno bito klaida visada ištaisoma!

Kodai $T(r)$

$$r = 4, \quad x_1x_2\ldots x_{10} \to \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & y_1 \\ x_5 & x_6 & x_7 & y_2 & \\ x_8 & x_9 & y_3 & & \\ x_{10} & y_4 & & & \\ y_5 & & & & \end{array}$$

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + y_1 &= 0, \\ x_4 + x_5 + x_6 + x_7 + y_2 &= 0, \\ x_3 + x_7 + x_8 + x_9 + y_3 &= 0, \\ x_2 + x_6 + x_9 + x_{10} + y_4 &= 0, \\ x_1 + x_5 + x_8 + x_{10} + y_5 &= 0. \end{aligned}$$

Hammingo kodas H(3)

Sudarysime kodą $\mathrm{C} \subset \mathcal{B}^7$ keturiems bitams koduoti, taisantį vieną klaidą

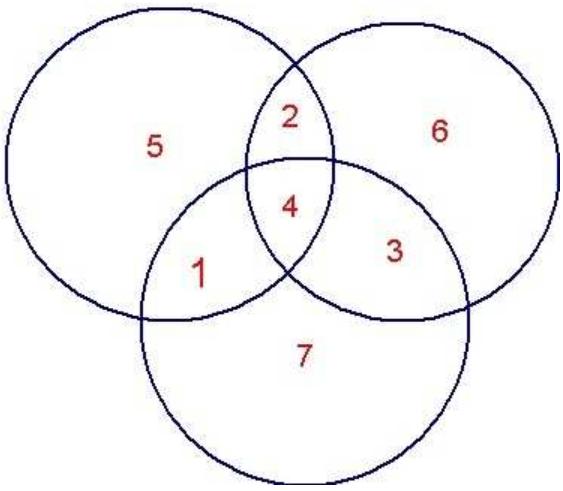
duomenų bitai $x_1x_2x_3x_4 \to y_1y_2y_3y_4y_5y_6y_7$ kodo žodis

$$y_3 = x_1,\ y_5 = x_2,\ y_6 = x_3,\ y_7 = x_4$$

$$\begin{aligned} 3 &= 1\cdot 1 + 1\cdot 2 + 0\cdot 4, & y_1 &= y_3 + y_5 + y_7, \\ 5 &= 1\cdot 1 + 0\cdot 2 + 1\cdot 4, & y_2 &= y_3 + y_6 + y_7, \\ 6 &= 0\cdot 1 + 1\cdot 2 + 1\cdot 4, & y_4 &= y_5 + y_6 + y_7. \\ 7 &= 1\cdot 1 + 1\cdot 2 + 1\cdot 4, \end{aligned}$$

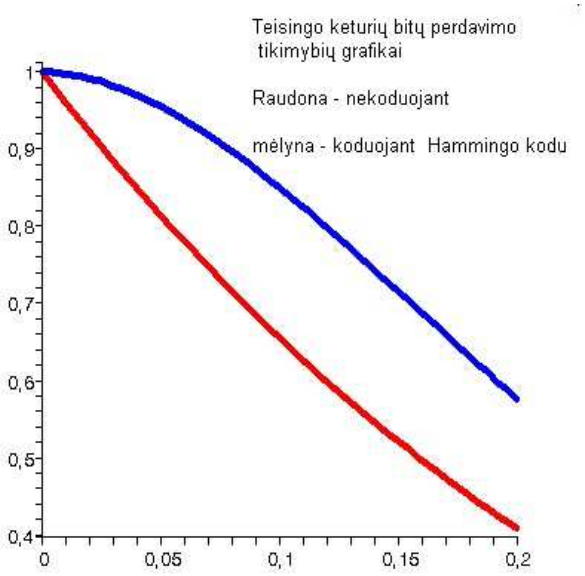
Analogiškai konstruojame kitus Hammingo kodus **H**(r), r yra kontrolinių bitų skaičius.

Hammingo kodas H(3)



26 / 27

Teisingo dekodavimo tikimybė



27 / 27