ÚVOD DO TEORIE INFORMACE

INFORMACE – POZNATKY O OBJEKTU, JEVU, PROCESU, ...

FORMA INFORMACE – TEXT, OBRAZ, ŘEČOVÝ SIGNÁL, ...

NOSIČ INFORMACE – ELEKTRICKÝ SIGNÁL, MAGNETIZACE, KŘÍDOVÝ PRÁŠEK NA TMAVÉ TABULI, ...

MATEMATICKÝ MODEL SDĚLOVACÍ SOUSTAVY

ZI .... ZDROJ INFORMACE

K ... KODÉR

MEDIUM ... SDĚLOVACÍ NEBO ZÁZNAMOVÉ PROSTŘEDÍ

D ... DEKODÉR

PI ... PŘÍJEMCE INFORMACE

R ... ZDROJ RUŠENÍ

$U(t), V(t), V'(t), U'(t), \xi(t), \dots$

... NÁHODNÉ PROCESY

NAPŘ. $U : \Omega \times T \to A$

„NÁHODA" ČAS ABECEDA ZDROJE

FIXUJEME-LI ČAS, ZÍSKÁME NÁHODNOU PROMĚNNOU.

FIXUJEME-LI ω, ZÍSKÁME DETERMINISTICKOU FUNKCI.

ZDROJ INFORMACE

- SPOJITÝ (ZPRÁVA JE REPREZENTOVÁNA SPOJITOU ČASOVOU FCÍ)
- DISKRÉTNÍ (ZPRÁVA JE REPREZENTOVÁNA ŘETĚZCEM PRVKŮ NAD DANOU ABECEDOU)

MÉDIUM (SDĚLOVACÍ KANÁL)

- SPOJITÝ (PŘENÁŠÍ HODNOTY Z URČITÉHO INTERVALU)
- DISKRÉTNÍ (PŘENÁŠÍ PRVKY Z DANÉ KONEČNÉ MNOŽINY)

V DALŠÍCH ÚVAHÁCH SE OMEZÍME NA DISKRÉTNÍ ZDROJ INFORMACE A DISKRÉTNÍ SDĚLOVACÍ KANÁL.

KODÉR – PŘEVÁDÍ ZPRÁVY (TJ. ŘETĚZCE PRVKŮ Z ABECEDY ZDROJE) NA ŘETĚZCE PRVKŮ Z ABECEDY KANÁLU.

DEKODÉR – PROVÁDÍ INVERZNÍ OPERACI KE KÓDOVÁNÍ.

ZDROJ RUŠENÍ – MODEL VNĚJŠÍHO OKOLÍ, NEŽÁDOUCÍ OVLIVŇOVÁNÍ PŘENÁŠENÝCH (RESP. ULOŽENÝCH) ZPRÁV.

ZÁKLADNÍ PŘEDPOKLAD TEORIE

ZDROJ MŮŽE GENEROVAT POUZE TAKOVÉ ZPRÁVY, KTERÉ MŮŽE PŘÍJEMCE VYHODNOTIT.

JINÝMI SLOVY: ZDROJ GENERUJE PRVKY Z DANÉ KONEČNÉ ABECEDY, KTEROU PŘÍJEMCE ZNÁ. NEZNALOST (NEURČITOST) PŘÍJEMCE SPOČÍVÁ POUZE V TOM, ŽE NEVÍ, JAKÝ KONKRÉTNÍ PRVEK ZDROJ V KONKRÉTNÍM ČASOVÉM OKAMŽIKU VYGENERUJE. NEURČITOST = PŘEDPOKLAD SDĚLOVÁNÍ.

DISKRÉTNÍ ZDROJ BEZ PAMĚTI – VYSÍLÁNÍ JEDNOTLIVÝCH ZNAKŮ TVOŘÍ NEZÁVISLÉ JEVY; TO, JAKÝ ZNAK JE VYSLÁN JAKO m-TÝ V POŘADÍ NEZÁVISÍ NA m-1 ZNACÍCH VYSLANÝCH PŘED NÍM.

MODEL DISKRÉTNÍHO ZDROJE INFORMACE

N.V. VELIČINA X ; REALIZACE $x_1, x_2, \ldots, x_n$

P-STI $p_1, p_2, \ldots, p_n$

PŘ: OSUDÍ S DESETI LÍSTKY, NĚKTERÉ JSOU OZNAČENY „0", NĚKTERÉ „1".

VARIANTA a)

„0" 5x „1" 5x

VARIANTA b)

„0" 9x „1" 1x

VARIANTA c)

„0" 10x „1" 0x

SNAŽME SE POROVNÁVAT „MÍRU PŘEKVAPENÍ", JE LI V DANÝCH VARIANTÁCH VYLOSOVÁNA „0" (RESP. „1").

JE ZŘEJMÉ, ŽE NEJVĚTŠÍ PŘEKVAPENÍ JE VYTAŽENÍ „1" V b), NEJMENŠÍ „0" V b), „0" A „1" V a) JSOU STEJNÉ.

VARIANTA c) JE NEZAJÍMAVÁ.

„MÍRU PŘEKVAPENÍ" REPREZENTUJE VELIČINA ELEMENTÁRNÍ ENTROPIE REALIZACI x_i. ZNAČENÍ: $H(x_i)$, H_i

ZŘEJMĚ PLATÍ $H(x_i) = f(p_i)$

$$p_i < p_j \Rightarrow f(p_i) > f(p_j)$$

ROZUMNÝ POŽADAVEK:

$$f(p_i \cdot p_j) = f(p_i) + f(p_j)$$

POŽADAVKŮM VYHOVUJE

$$f(x) = -\log_2 x$$

DEF: D.N.V. X_i REALIZACE $x_1, x_2, \dots, x_n$
P-STI $p_1, p_2, \dots, p_n$

ELEMENTÁRNÍ ENTROPIE REALIZACE x_i

$$H(x_i) = -\log_2 p(x_i)$$

NENÍ DEFINOVANÁ PRO $p(x_i) = 0$

JEDNOTKA ENTROPIE – 1 BIT

ELEMENTÁRNÍ ENTROPIE JE VLASTNOSTÍ KONKRÉTNÍ REALIZACE.

STŘEDNÍ ENTROPIE NÁHODNÉ VELIČINY X

$$H(X) = E\left[H(x_i)\right] = -\sum_{i=1}^{n} p(x_i) \cdot \log_2 p(x_i)$$

V DEFINICI KLADEME

$$0 \cdot \log_2 0 = \lim_{x \to 0+} x \cdot \log_2 x = 0$$

STŘEDNÍ ENTROPIE JE VLASTNOSTÍ „CELÉ" NÁHODNÉ VELIČINY.

PŘ: HOD MINCÍ X $x_1 = "P"$ $p(x_1) = 0{,}5$

$x_2 = "O"$ $p(x_2) = 0{,}5$

$$H(X) = -(0{,}5 \cdot \log_2 0{,}5 + 0{,}5 \cdot \log_2 0{,}5) =$$

$$= -\log_2 0{,}5 = -(-1) = 1 \text{ BIT}$$

DŮLEŽITÁ VLASTNOST:

$$0 \leq H(X) \leq \log_2 m$$

KRAJNÍ HODNOTY:

$H(X) = 0$ ... X JE DETERMINISTICKÁ

$H(X) = \log_2 m$ ... VŠECHNY REALIZACE MAJÍ STEJNOU P-ST

$$p(x_i) = \frac{1}{m} \quad \forall i = 1, 2, \dots, m$$

CO JE TO INFORMACE? NĚCO, CO ZMENŠUJE NEURČITOST.

INFORMACE =

= NEURČITOST „PŘED EXP." − NEURČITOST „PO EXP."

<u>ELEMENTÁRNÍ INFORMACE REALIZACE</u> x_i

$$I(x_i) = H(x_i) = -\log_2 p(x_i) \quad [BIT]$$

<u>STŘEDNÍ INFORMACE D.N.V X</u>

$$I(X) = H(X) = -\sum_{i=1}^{m} p(x_i) \cdot \log_2 p(x_i)$$

<u>REDUNDANCE (NADBYTEČNOST) ZDROJE ZPRÁV</u>

$$\rho = 1 - \frac{H(X)}{\log_2 m}$$

<u>PŘ</u>: ABECEDA $\{0,1\}$, TROJNÁSOBNÉ OPAKOVÁNÍ.

x_i	$p(x_i)$
000	0.5
001	0
⋮	⋮
110	0
111	0.5

$$\rho = 1 - \frac{1}{\log_2 8} = 1 - \frac{1}{3} = \underline{\underline{\frac{2}{3}}}$$

KÓDOVÁNÍ

SMYSL (OBECNĚ):

- PŘIZPŮSOBIT PŘENÁŠENÉ ZPRÁVY ABECEDĚ KANÁLU
- ZVÝŠENÍ ODOLNOSTI PROTI RUŠENÍ (BEZPEČNOSTNÍ KÓDY)
- EFEKTIVNĚJŠÍ VYUŽITÍ MÉDIA (KÓDY S MINIMÁLNÍ STŘEDNÍ DÉLKOU, KOMPRESE)
- UTAJENÍ INFORMACE (ŠIFROVÁNÍ, KRYPTOVÁNÍ)

TEORIE KÓDOVÁNÍ JE PRAKTICKOU APLIKACÍ MATEMATICKÝCH DISCIPLÍN:

- LINEÁRNÍ ALGEBRA
- KOMBINATORIKA
- TEORIE GRUP
- TEORIE ČÍSEL.

PŘEDMĚT NAŠEHO ZÁJMU:

KÓDOVÁNÍ PRO KANÁL BEZ ŠUMU
(ODSTRAŇOVÁNÍ REDUNDANCE Z PŘENÁŠENÝCH ZPRÁV)

KÓDOVÁNÍ PRO KANÁL SE ŠUMEM
(DOPLŇOVÁNÍ REDUNDANCE DO PŘENÁŠENÝCH ZPRÁV)

KÓDOVÁNÍ PRO KANÁL BEZ ŠUMU

PŘ: BINÁRNÍ KÓDOVÁNÍ DEKADICKÝCH ČÍSLIC

0	0000
1	0001
2	0010
3	0011
4	0100
5	1011
6	1100
7	1101
8	1110
9	1111

KÓDOVÁNÍ ZPRÁVY:

7253

1101|0010|1011|0011

PŘ: KÓDOVÁNÍ INFORMACE O TEPLOTĚ VODY

ČTYŘI STAVY: L, S, V, T

VÍME: VODA JE VĚTŠINOU LEDOVÁ

PŘENOS KAŽDÉHO ZNAKU STOJÍ

NEZANEDBATELNOU ČÁSTKU

K_1:			K_2:	
	L	00	L	0
	S	01	S	01
	V	10	V	011
	T	11	T	111

(LEPŠÍ)
(SLOŽITĚJŠÍ DEKÓDOVÁNÍ)

KÓDOVÁNÍ ZPRÁVY KÓDEM K_2:

L L S T   0001111

POKUD DEKÓDOVÁNÍ ZEPŘEDU:

L L
0|0|0 1 1 1 1
L NEBO S ?

DEKÓDOVÁNÍ ZEZADU JE JEDNOZNAČNÉ:

0|0|0 1|1 1 1
L L S T

DEF: A... ZDROJOVÁ ABECEDA (n PRVKŮ)

B... KÓDOVÁ ABECEDA (s PRVKŮ)

$n > s$

KÓDOVÁNÍ ZNAKŮ: $K: A \to B^+$

(K JE PROSTÉ ZOBRAZENÍ)

KÓDOVÁNÍ ZPRÁV (ŘETĚZCŮ): $K^*: A^* \to B^*$

K^* JE JEDNOZNAČNĚ URČENO POMOCÍ K (JE JEHO ROZŠÍŘENÍM):

$$K^*(a_1 a_2 \dots a_n) = K(a_1) \cdot K(a_2) \cdot \ldots \cdot K(a_n)$$

PODMÍNKA JEDNOZNAČNÉ DEKÓDOVATELNOSTI:

K^* JE PROSTÝM ZOBRAZENÍM

TERMÍN PRO ZAKÓDOVANÝ ZNAK:

KÓDOVÁ ZNAČKA, KÓDOVÉ SLOVO

PŘ. NEJEDNOZNAČNÉHO KÓDOVÁNÍ:

$$K: \begin{array}{ll} a & 0 \\ b & 1 \\ c & 01 \end{array} \qquad \begin{array}{l} K^*(ab) = 01 \\ K^*(c) = 01 \end{array}$$

BLOKOVÉ KÓDOVÁNÍ - PROSTÉ KÓDOVÁNÍ, PŘI KTERÉM MAJÍ VŠECHNY KÓDOVÉ ZNAČKY STEJNOU DÉLKU (l).

$$K: A \to B^l$$

KAŽDÉ BLOKOVÉ KÓDOVÁNÍ JE JEDNOZNAČNĚ DEKÓDOVATELNÉ („ROZSEKÁNÍ NA l-TICE).

PREFIXOVÉ KÓDOVÁNÍ - PROSTÉ KÓDOVÁNÍ S NESTEJNOU DÉLKOU KÓDOVÝCH ZNAČEK, KDE ŽÁDNÁ KÓDOVÁ ZNAČKA NENÍ PREFIXEM JINÉ KÓDOVÉ ZNAČKY.

JE JEDNOZNAČNĚ DEKÓDOVATELNÉ (STAČÍ NA TO MEALYHO KA).

PŘ. TROJKOVÉHO (TERNÁRNÍHO) KÓDU:

K:	A	0
	B	1
	C	20
	D	21
	E	220
	F	221

KÓDOVÁNÍ ZPRÁVY:

BBFBFB

1122112211

PREFIXOVÉ KÓDY LZE DEKÓDOVAT „ZNAK PO ZNAKU", TJ. LZE DEKÓDOVAT UŽ V PRŮBĚHU PŘENOSU (NENÍ TŘEBA ČEKAT NA DOKONČENÍ PŘENOSU CELÉ ZPRÁVY).

PŘI KÓDOVÁNÍ m ZNAKY LZE SESTROJIT PREFIXOVÝ KÓD PRÁVĚ KDYŽ PLATÍ

$$m^{-d_1} + m^{-d_2} + \dots + m^{-d_n} \leq 1$$

KDE

d_i .... DÉLKA i-TÉ KÓDOVÉ ZNAČKY.

(KRAFTOVA NEROVNOST)

PŘ: UVAŽUJME BINÁRNÍ PREFIXOVÝ KÓD PRO KÓDOVÁNÍ CIFER 0,1,2,...,9 VHODNÝ PRO ZPRÁVY, KDE SE ČASTO VYSKYTUJÍ ZNAKY 0,1 A ŘÍDCE ZNAKY 8,9.

NÁPAD: PRO 0 A 1 SLOVA DÉLKY 2
PRO 3....7 SLOVA DÉLKY 3
PRO 8 A 9 SLOVA DÉLKY 4

0 00
1 01
2
3
4
5
6
7

MUSÍ ZAČÍNAT ZNAKEM 1 (ABY BYLO PREFIXOVÉ)

1xx POSKYTUJE JEN 4 MOŽNOSTI ⇒ SPOR!

8
9

CO ŘÍKÁ KRAFTOVA NEROVNOST ?

$$2\cdot 2^{-2} + 6\cdot 2^{-3} + 2\cdot 2^{-4} = \frac{8+12+2}{16} = \frac{22}{16} > 1$$

DOBRÉ ŘEŠENÍ:

0	00	6	1110
1	01	7	1101
2	1000	8	1011
3	1100	9	1111
4	1010		
5	1001		

KRAFTOVA NEROVNOST:

$$2 \cdot 2^{-2} + 8 \cdot 2^{-4} = \frac{8+8}{16} = \frac{16}{16} = 1$$

LEPŠÍ ŘEŠENÍ

0	00	5	1100
1	01	6	1101
2	100	7	1110
3	1010	8	11110
4	1011	9	11111

$$2 \cdot 2^{-2} + 1 \cdot 2^{-3} + 5 \cdot 2^{-4} + 2 \cdot 2^{-5} = \frac{16+4+10+2}{32} = 1$$

MC MILLANOVA VĚTA:

PRO KAŽDÉ JEDNOZNAČNĚ DEKÓDOVATELNÉ KÓDOVÁNÍ PLATÍ KRAFTOVA NEROVNOST.

DOSAVADNÍ ÚVAHY O PREFIXOVÝCH KÓDECH ZKOUMALY POUZE SOUVISLOST EXISTENCE KÓDU S DÉLKAMI KÓDOVÝCH ZNAČEK, NEVĚNOVALY SE „KVALITĚ KÓDU".

ROZUMNÝM KRITÉRIEM KVALITY JE STŘEDNÍ DÉLKA KÓDOVÉ ZNAČKY:

$$\bar{d} = \sum_{i=1}^{n} \underbrace{p(a_i)} \cdot \left|\underbrace{K(a_i)}\right|$$

STŘEDNÍ DÉLKA KÓDOVÉ ZNAČKY

P-ST JEHO VÝSKYTU VE ZPRÁVÁCH

ZNAK ZDROJOVÉ ABECEDY

KÓDOVÁ ZNAČKA

DÉLKA KÓDOVÉ ZNAČKY

ÚLOHA: PRO DANOU ZDROJOVOU ABECEDU SE ZADANÝM ROZLOŽENÍM P-STÍ A PRO DANOU ABECEDU KÓDU URČIT KÓD S MINIMÁLNÍ STŘEDNÍ DÉLKOU KÓDOVÉ ZNAČKY.

HUFFMANOVA KONSTRUKCE KÓDU S MINIMÁLNÍ STŘEDNÍ DÉLKOU KÓDOVÉ ZNAČKY

VSTUP: $A, p(A), B$ — R PRVKŮ / A PRVKŮ

VÝSTUP: $K: A \to B^+$, $\bar{d}(K)$ JE MINIMÁLNÍ

ALGORITMUS:

1) PRVKY ABECEDY A SEŘADÍME PODLE P-STÍ DO NEROSTOUCÍ POSLOUPNOSTI.

2) SEŘAZENÉ PRVKY ABECEDY A ROZDĚLÍME DO SKUPIN.

ZAČÍNÁME OD PRVKŮ S NEJVYŠŠÍMI P-STMI. SKUPINY BUDOU MÍT $\Delta-1$ PRVKŮ. VÝJIMKOU MŮŽE BÝT POSLEDNÍ SKUPINA, KTERÁ MŮŽE MÍT Δ PRVKŮ.

3) SDRUŽÍME PRVKY V POSLEDNÍ SKUPINĚ A SKUPINU ZAŘADÍME PODLE JEJÍ SOUČTOVÉ PRAVDĚPODOBNOSTI DO POSLOUPNOSTI.

4) BODY 2) A 3) OPAKUJEME TAK DLOUHO, DOKUD NEZÍSKÁME Δ-PRVKOVOU SKUPINU SE SOUČTEM P-STÍ 1.

5) ZPĚTNÝM CHODEM PO VĚTVÍCH STROMU VYTVOŘENÉHO V KROCÍCH 2) AŽ 4) PŘIŘADÍME KÓDOVÉ ZNAČKY LISTŮM STROMU, TJ. PRVKŮM ZDROJOVÉ ABECEDY.

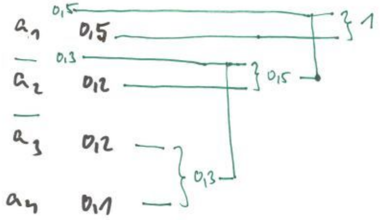
KÓDEM K S MINIMÁLNÍ STŘEDNÍ DÉLKOU KÓDOVÉ ZNAČKY $\bar{d}$ ROZUMÍME TAKOVÝ KÓD, ŽE NEEXISTUJE JINÝ KÓD X RŮZNÝ OD K, PRO KTERÝ BY PLATILO

$\bar{d}(X) < \bar{d}(K)$. MŮŽE ALE PLATIT

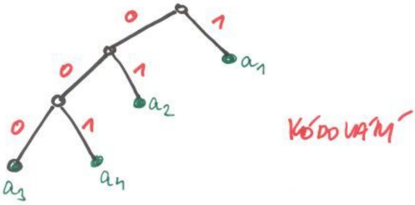
$\bar{d}(X) = \bar{d}(K)$.

PŘ: $A = \{a_1, a_2, a_3, a_4\}$ $B = \{0, 1\}$

$P(A) = (0{,}5; 0{,}2; 0{,}2; 0{,}1)$



JINÉ ZNÁZORNĚNÍ STROMU:



a_i	$K(a_i)$
a_1	1
a_2	01
a_3	000
a_4	001

ZPŮSOB KONSTRUKCE KÓDOVÉ ZNAČKY (TJ. ODHODNOCENÍ CESTY OD KOŘENE K LISTU) ZAJIŠŤUJE TO, ŽE JE KÓD PREFIXOVÝ, TEDY JEDNOZNAČNĚ DEKÓDOVATELNÝ.

$$\bar{d}(K) = 0{,}5\cdot 1 + 0{,}2\cdot 2 + 0{,}2\cdot 3 + 0{,}1\cdot 3 =$$
$$= 0{,}5 + 0{,}4 + 0{,}6 + 0{,}3 = 1{,}8$$

$$H(A) = -\left(0{,}5\cdot \lg_2 0{,}5 + 0{,}2\cdot \lg_2 0{,}2 + 0{,}2\cdot \lg_2 0{,}2 + 0{,}1\cdot \lg_2 0{,}1\right) =$$
$$= \underline{1{,}76}$$

REDUNDANCE $\rho(A) = 1 - \dfrac{1{,}76}{2} = \underline{0{,}12}$

PO ZAKÓDOVÁNÍ:

$\bar{N}(0) = (0{,}2\cdot 1 + 0{,}2\cdot 3 + 0{,}1\cdot 2) = 1$ — STŘEDNÍ POČET 0 VE ZNAČCE

$\bar{N}(1) = (0{,}5\cdot 1 + 0{,}2\cdot 1 + 0{,}1\cdot 1) = 0{,}8$ — STŘEDNÍ POČET 1 VE ZNAČCE

$p(0) = \dfrac{1}{1{,}8} = 0{,}\bar{5}$ $\quad p(1) = \dfrac{0{,}8}{1{,}8} = 0{,}\bar{4}$

P-STI VÝSKYTU VE ZPRÁVÁCH

$$H(B) = -\left(0{,}\bar{5}\cdot \lg_2 0{,}\bar{5} + 0{,}\bar{4}\cdot \lg_2 0{,}\bar{4}\right) = 0{,}99$$
$$\rho(B) = 1 - \frac{0{,}99}{1} = \underline{0{,}01}$$