Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

Međuispit iz predmeta **TEORIJA INFORMACIJE**, 21. studenog 2018.

Pravilo bodovanja zadataka

Svaki točno odgovoreni zadatak donosi 5 bodova, netočno odgovoreni 2 negativna boda, a neodgovoreni 0 bodova.

Zadatak 1. Neki diskretni informacijski izvor bez memorije generira simbole iz abecede X. Ta abeceda sadrži n simbola koje označavamo kao x_i, $1 \le i \le n$, pri čemu vrijedi $i, n \in \mathbf{N}$, $n > 1$. Svakom simbolu pridružena je vjerojatnost pojavljivanja $P(x_i)$, $0 < P(x_i) < 1$, $\sum_{i=1}^{n} P(x_i) = 1$. Svakom simbolu pridružena je i količina informacije $I(x_i)$, sukladno vjerojatnosti pojave tog simbola. Odredite standardnu devijaciju slučajne varijable $I(X)$, čija je razdioba definirana izrazom:

$$I(X) = \begin{pmatrix} I(x_1) & \dots & I(x_n) \\ P(x_1) & \dots & P(x_n) \end{pmatrix} = \begin{pmatrix} -\log_2 P(x_1) & \dots & -\log_2 P(x_n) \\ P(x_1) & \dots & P(x_n) \end{pmatrix},$$

za slučaj kad vrijedi: $I(x_i) = \log_2(n)\ \forall i$, $1 \le i \le n$.

a) 0 bit/simbol

b) $\log_2(n)$ bit/simbol

c) $[\log_2(n)]^2$ bit/simbol

d) $[\log_2(n)]^{1/2}$ bit/simbol

e) ništa od navedenog

Postupak rješavanja:

Disperzija slučajne varijable $I(X)$ određena je izrazom:

$$D[I(X)] = Var[I(X)] = E\left\{[I(X)]^2\right\} - \left\{E[I(X)]\right\}^2 = \sum_{i=1}^{n} P(x_i)[-\log_2 P(x_i)]^2 - [H(X)]^2$$

Ako vrijedi da je $I(x_i) = \log(n)\ \forall i$, $1 \le i \le n$. tada je $p(x_i) = 1/n$, $\forall\ 1 \le i \le n$, te slijedi:

$$D[I(X)] = \sum_{i=1}^{n} \frac{1}{n}[\log_2 n]^2 - (\log_2 n)^2 = (\log_2 n)^2 - (\log_2 n)^2 = 0$$

Dakle, standardna devijacija od $I(X)$ iznosi 0 bit/simbol.

Zadatak 2. Neki diskretni izvor informacije generira 4 simbola, m_1, m_2, m_3 i m_4. Vjerojatnosti pojavljivanja simbola zadane su na sljedeći način: $P(m_1) = 2P(m_2) = 4P(m_3)$ i $P(m_3) = P(m_4)$, $\sum_{i=1}^{4} P(m_i) = 1$. Koder informacije kodira svaki simbol m_i, $i = 1, \ldots, 4$, s dva binarna simbola po simbolu m_i prema sljedećem pravilu: $m_1 \rightarrow 00$, $m_2 \rightarrow 01$, $m_3 \rightarrow 10$ i $m_4 \rightarrow 11$. Analiziranjem jako dugačkog slijeda binarnih simbola na izlazu kodera odredite količinu informacije koju u sebi nosi binarna poruka 111.

a) 2,688 bit/poruka

b) 1,622 bit/poruka

c) 5,034 bit/poruka

d) 3 bit/poruka

e) ništa od navedenog

Postupak rješavanja:

Temeljem poznavanja vjerojatnosti pojavljivanja simbola m_i i binarnog koda moguće je odrediti vjerojatnost pojavljivanja simbola 1 na izlazu kodera:

$$P(1)=\frac{\sum_{i=1}^{4} l_{i1}N_i}{\sum_{i=1}^{4} l_i N_i}=\frac{\frac{1}{M}\sum_{i=1}^{4} l_{i1}N_i}{\frac{1}{M}\sum_{i=1}^{4} l_i N_i}=\frac{\sum_{i=1}^{4} l_{i1}P(m_i)}{\sum_{i=1}^{4} l_i P(m_i)}=\frac{\frac{1}{2}\cdot 0+\frac{1}{4}\cdot 1+\frac{1}{8}\cdot 1+\frac{1}{8}\cdot 2}{2\sum_{i=1}^{4} P(m_i)}=\frac{\frac{5}{8}}{2}=\frac{5}{16}$$

Dakle, svaka binarna jedinica će na izlazu kodera sadržavati količinu informacije koja odgovara vjerojatnosti njena pojavljivanja, $I(1) = -\log_2(5/16) = 1{,}678$ bita, a tri uzastopne jedinice nose u sebi 5,034 bita informacije.

Zadatak 3. Diskretni izvor informacije generira 3 simbola iz abecede M, m_1, m_2 i m_3, s pripadajućom razdiobom vjerojatnosti pojavljivanja simbola

$$M=\begin{pmatrix} m_1 & m_2 & m_3 \\ 0{,}55 & 0{,}3 & 0{,}15 \end{pmatrix}$$

Nadalje, simboli m_i dolaze na ulaz kodera informacije. Razmatrajte dva scenarija: a) koder informacije u predajniku koristi Huffmanovo kodiranje kvaternarnim simbolima iz abecede {0, 1, 2, 3} i b) koder informacije u predajniku koristi Huffmanovo kodiranje binarnim simbolima, abeceda {0, 1}. Odredite omjer efikasnosti koda u scenariju b prema efikasnosti koda u scenariju a, $\varepsilon_b/\varepsilon_a$.

a) 0,69

b) 1,45

c) 0,725

d) 1,38

e) ništa od navedenog

Postupak rješavanja:

Koder informacije kodira simbole m_i kodnim riječima duljine l_i, i = 1, 2, 3.

a) Prilikom kodiranja simbola m_i Huffmanovim kodom koji koristi kvaternarne simbole, svakom simbolu m_i bit će pridijeljen točno jedan kvaternarni simbol (l_i = 1, i = 1, 2, 3). Stoga će srednja duljina kodne riječi na izlazu kodera informacije iznositi L_4 = 1 kv.sim./simbol (oznaka se odnosi na kvaternarni simbol po simbolu m_i). Istovremeno će entropija na izlazu kodera informacije iznositi:

$$H_4(M) = -\sum_{i=1}^{3} P(m_i)\log_4 P(m_i) = -\sum_{i=1}^{3} P(m_i)\frac{\log_2 P(m_i)}{\log_2 4} = -\frac{1}{2}\sum_{i=1}^{3} P(m_i)\log_2 P(m_i) = \frac{H_2(M)}{2}\left[\frac{\text{kv.sim.}}{\text{simbol}}\right]$$

b) Prilikom kodiranja simbola m_i Huffmanovim kodom koji koristi binarne simbole, bit će ostvarene sljedeće duljine kodnih riječi: l_1 = 1 bit/simbol, $l_2 = l_3$ = 2 bit/simbol (odnosi se na broj bita po simbolu m_i).

Dakle,

$$\frac{\varepsilon_b}{\varepsilon_a} = \frac{\dfrac{H_2(M)}{\overline{L_2}}}{\dfrac{H_4(M)}{\overline{L_4}}} = 2\frac{\overline{L_4}}{\overline{L_2}} = 2\frac{1}{0,55\cdot 1 + 0,3\cdot 2 + 0,15\cdot 2} = \frac{2}{1,45} = 1,38$$

Zadatak 4. Promatrajte kanal kojeg karakterizira svojstvo da su mu reci matrice kanala, [$P(Y|X)$], permutacije jedan drugog, a zbroj članova matrice po svakom stupcu međusobno je jednak. Pri tome X predstavlja skup simbola na ulazu, a Y skup simbola na izlazu kanala. Matrica kanala zadana je sljedećim izrazom:

$$\left[P(Y|X)\right] = \begin{bmatrix} 1/3 & 1/3 & 1/6 & a \\ b & c & d & e \end{bmatrix}, 0 < a, b, c, d, e < 1$$

Odredite kapacitet kanala. Napomena: Permutacija brojeva q_1, q_2, q_3, q_4 (brojevi q_i predstavljaju prvi redak matrice kanala) je svaka uređena četvorka oblika (r_1, r_2, r_3, r_4) u kojoj se svaki od brojeva q_1, q_2, q_3, q_4 javlja točno jedanput. Brojevi r_i predstavljaju drugi redak matrice kanala.

a) 2 bit/simbol

b) 0,082 bit/simbol

c) 1,918 bit/simbol

d) 1,585 bit/simbol

e) ništa od navedenog

Postupak rješavanja:

S obzirom da zbroj elemenata po retku matrice kanala mora iznositi 1, slijedi da je a = 1/6. Pod uvjetom da je drugi redak permutacija prvog retka (dakle, sadrži dvije vjerojatnosti 1/3 i dvije vjerojatnosti 1/6) te uz zadani uvjet da je zbroj elemenata matrice kanala po svakom stupcu međusobno jednak, postoji samo jedno moguće rješenje, a to je:

$$P(Y|X) = \begin{bmatrix} 1/3 & 1/3 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{bmatrix}$$

Očito se radi o djelomično simetričnom kanalu (engl. *weakly symmetric channel*) čiji se kapacitet računa prema izrazu:

$$C = \log\left[\text{card}(Y)\right] - H(Y|x)$$

pri čemu je

$$H\left(Y\middle|x\right)=\sum_{j=1}^{4}P\left(y_j\middle|x_i\right)\log\left(\frac{1}{P\left(y_j\middle|x_i\right)}\right), i\in\{1,2\}$$

Dakle, za proračun kapaciteta kanala dovoljno je izračunati entropiju $H(Y|x)$ za jedan redak matrice kanala. S obzirom da skup Y ima 4 člana, vrijedi log[card (Y)] = 2 te:

$$H\left(Y\middle|x\right)=-2\frac{1}{3}\log_2\frac{1}{3}-2\frac{1}{6}\log_2\frac{1}{6}=\frac{2}{3}\log_2 3+\frac{1}{3}\log_2 6=\frac{1}{3}+\log_2 3$$

pa je kapacitet kanala jednak $C = 2 - 1/3 - \log_2(3) = 5/3 - \log_2(3) = 0{,}082$ bit/simbol.

Zadatak 5. Promatrajte eksperiment bacanja ispravnog novčića (vjerojatnost da ispadne "pismo" jednaka je vjerojatnosti da ispadne "glava"). Neka slučajna varijabla X poprima vrijednost koja je jednaka broju bacanja novčića do prve pojave "pisma" (napomena: u taj je broj uključeno i bacanje u kojem se prvi put javlja "pismo"). Odredite entropiju slučajne varijable X. Pri određivanju entropije možete koristiti sljedeće izraze:

$$\sum_{n=1}^{\infty}q^n=\frac{q}{1-q},\ \sum_{n=1}^{\infty}nq^n=\frac{q}{(1-q)^2}, 0<q<1$$

a) 2 bit/simbol

b) 1 bit/simbol

c) 1,5 bit/simbol

d) 0,5 bit/simbol

e) ništa od navedenog

Postupak rješavanja:

Neka je općenito vjerojatnost pojave glave jednaka q, a pisma $p = 1 - q$. Broj bacanja do prve pojave pisma ima geometrijsku razdiobu: $P(X = n) = pq^{n-1}$. Sukladno tome

$$H(X)=-\sum_{n=1}^{\infty}pq^{n-1}\log_2\left(pq^{n-1}\right)=-\sum_{n=1}^{\infty}pq^{n-1}\log_2 p-\sum_{n=1}^{\infty}pq^{n-1}\log_2 q^{n-1}=$$

$$=-p\log_2 p\sum_{n=1}^{\infty}q^{n-1}-p\sum_{n=1}^{\infty}(n-1)q^{n-1}\log_2 q=-p\log_2 p\sum_{k=0}^{\infty}q^k-p\log_2 q\sum_{k=0}^{\infty}kq^k=$$

$$\left|\sum_{k=0}^{\infty}q^k=1+\sum_{k=1}^{\infty}q^k=\frac{1}{1-q},\quad \sum_{k=0}^{\infty}kq^k=\sum_{k=1}^{\infty}kq^k=\frac{q}{(1-q)^2}\right|$$

$$=-p\frac{1}{1-q}\log_2 p-p\frac{q}{(1-q)^2}\log_2 q=-\frac{p^2\log_2 p+pq\log_2 q}{p^2}=\frac{H(p)}{p}\left[\frac{\text{bit}}{\text{simbol}}\right]$$

pri čemu je $H(p) = -p\log_2 p - (1 - p)\log_2(1 - p)$. Uz $p = q = 1/2$, vrijedi $H(p) = 1$ bit/simbol pa je $H(X) = 2$ bit/simbol.

Zadatak 6. Ponašanje promatranog izvora s memorijom moguće je opisati homogenim markovljevim lancem prvog reda. Taj lanac predstavlja niz diskretnih slučajnih varijabli X_0, X_1, X_2, …, a ima dva stanja, 0 i 1. Razdioba slučajne varijable X_1 je $\mathbf{p}(1) = [P(X_1 = 0), P(X_1 = 1)] = [3/4, 1/4]$, slučajne varijable X_2 je $\mathbf{p}(2) = [P(X_2 = 0), P(X_2 = 1)] = [5/8, 3/8]$ i slučajne varijable X_3 je $\mathbf{p}(3) = [P(X_3 = 0), P(X_3 = 1)] = [9/16, 7/16]$. Odredite entropiju izvora.

a) 1 bit/simbol

b) 0,406 bit/simbol

c) 1,622 bit/simbol

d) 0,811 bit/simbol

e) ništa od navedenog

Postupak rješavanja:

Za razdiobe slučajnih varijabli X_1, X_2 i X_3 vrijedi:

$$\mathbf{p}(2)=\mathbf{p}(1)\mathbf{\Pi}$$
$$\mathbf{p}(3)=\mathbf{p}(2)\mathbf{\Pi}$$

pri čemu je matrica prijelaznih vjerojatnosti markovljevog lanca. Pretpostavimo da ta matrica ima oblik

$$\mathbf{\Pi}=\begin{bmatrix}\alpha & 1-\alpha\\ 1-\beta & \beta\end{bmatrix}$$

Dakle:

$$\begin{bmatrix}\frac{3}{4},\frac{1}{4}\end{bmatrix}\cdot\begin{bmatrix}\alpha & 1-\alpha\\ 1-\beta & \beta\end{bmatrix}=\begin{bmatrix}\frac{5}{8},\frac{3}{8}\end{bmatrix}$$
$$\begin{bmatrix}\frac{5}{8},\frac{3}{8}\end{bmatrix}\cdot\begin{bmatrix}\alpha & 1-\alpha\\ 1-\beta & \beta\end{bmatrix}=\begin{bmatrix}\frac{9}{16},\frac{7}{16}\end{bmatrix}$$

Iz tog sustava jednadžbi odredimo α i β: α = 3/4 = β. Sad je moguće odrediti stacionarne vjerojatnosti lanca: $\boldsymbol{\pi}=[\pi_0, \pi_1]$ temeljem jednakosti $\mathbf{\Pi}^{\mathrm{T}}\boldsymbol{\pi}=\boldsymbol{\pi}$:

$$\begin{bmatrix}3/4 & 1/4\\ 1/4 & 3/4\end{bmatrix}^{\mathrm{T}}\cdot\begin{bmatrix}\pi_0\\ \pi_1\end{bmatrix}=\begin{bmatrix}\pi_0\\ \pi_1\end{bmatrix}$$

Dakle, $\pi_0=\pi_1=1/2$. Sada je moguće odrediti entropiju izvora

$$H=-\sum_{i=0}^{1}\pi_i\sum_{j=0}^{1}p_{ij}\log\left(p_{ij}\right)=-\frac{1}{2}2\left[\frac{3}{4}\log_2\frac{3}{4}+\frac{1}{4}\log_2\frac{1}{4}\right]=\frac{3}{4}\log_2\frac{4}{3}+\frac{1}{4}\log_2 4=$$
$$=\frac{3}{4}\left[2-\log_2 3\right]+\frac{1}{4}2=\frac{6}{4}+\frac{2}{4}-\frac{3}{4}\log_2 3=2-\frac{3}{4}\log_2 3=0{,}811\frac{\text{bit}}{\text{simbol}}$$

Zadatak 7: Poruka „aaaaaaaaaaa*" (navodnici nisu dio poruke, a * je oznaka kraja slijeda) kodira se algoritmom LZ77 tako da je maksimalna duljina posmičnog prozora 1 simbol, a prozora za kodiranje 10 simbola. Koliko uređenih trojki (*pomak, duljina, sljedeći simbol*) generira navedeni algoritam kako bi kodirao poruku?

a) jednu
b) dvije
c) tri
d) četiri
e) ništa od navedenog

Rješenje:

U postupku kodiranja dobiju se tri trojke: (0,0,a), (1,9,a) i (0,0,*).

Zadatak 8: Dan je skup simbola $\{s_1, s_2, \ldots, s_m\}$ s pripadajućim vjerojatnostima pojavljivanja: $p(s_i) = p_i$, $i = 1, \ldots, m$. Simboli su jednoznačno kodirani prefiksnim kodom. Ako je $m = 6$ i ako su duljine kodnih riječi zadane kao $l_1 = l_2 = 1$, $l_3 = l_4 = l_5 = 2$, $l_6 = 3$, odredite najmanji broj simbola abecede prefiksnog koda.

a) 3

b) 2

c) 4

d) 1

e) ništa od navedenog

Postupak rješavanja:

$$\sum_{i=1}^{m} d^{-l_i} \le 1 \rightarrow \frac{1}{d}+\frac{1}{d}+\frac{1}{d^2}+\frac{1}{d^2}+\frac{1}{d^2}+\frac{1}{d^3} \le 1$$

$$2d^2 + 3d - d^3 \le -1$$

Očito za $d = 1$ kod nema smisla. Za $d = 2$ i $d = 3$ nejednakost nije zadovoljena, a za $d = 4$ je. Dakle, najmanji broj simbola abecede prefiksnog koda je 4.

Zadatak 9: Instrumentom očitavamo vrijednosti iz skupa simbola $X = \{-2,-1, 0, 1, 2\}$. Sve vrijednosti su jednako vjerojatne. Na pokazniku instrumenta pokvaren je indikator za "minus" koji se pri očitavanju negativnih vrijednosti ne upali u 30% slučajeva. Ako sustav promatramo kao komunikacijski kanal, izračunajte transinformaciju u ovom sustavu.

a) 2,3219 bit/simbol
b) 0,4053 bit/simbol
c) 1,1474 bit/simbol
d) 1,9167 bit/simbol
e) ništa od navedenog

Postupak rješavanja:

Indikator za „minus" na pokazniku se ne upali u 30% slučajeva, dakle vjerojatnost da će pokaznik prikazati vrijednost "-2" kao "-2" je 0.7, a kao "2" je 0.3. Analogno zaključujemo i za "-1", dok se ostale vrijednosti prikazuju ispravno te je:

$$\left[p(y_j|x_i)\right]=\begin{bmatrix}0,7 & 0 & 0 & 0 & 0,3\\ 0 & 0,7 & 0 & 0,3 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 1\end{bmatrix}$$

$$p(x_i)=0,2,\quad i=1,\ldots,5$$

$$\left[p(x_i,y_j)\right]=\left[p(x_i)p(y_j|x_i)\right]=\begin{bmatrix}0,14 & 0 & 0 & 0 & 0,06\\ 0 & 0,14 & 0 & 0,06 & 0\\ 0 & 0 & 0,2 & 0 & 0\\ 0 & 0 & 0 & 0,2 & 0\\ 0 & 0 & 0 & 0 & 0,2\end{bmatrix}$$

Iz matrice združenih vjerojatnosti zbrajanjem po stupcima dobivamo vrijednosti pojave pojedine vrijednosti na indikatoru instrumenta:

$$\left[p(y_j)\right]=\left[0,14 \quad 0,14 \quad 0,2 \quad 0,26 \quad 0,26\right]$$

Transinformaciju možemo dobiti prema sljedećem izrazu:

$$I(X;Y)=-\sum_{j=1}^{5}\sum_{i=1}^{5}p(x_i,y_j)\log\frac{p(x_i,y_j)}{p(x_i)\,p(y_j)}$$

$$I(X;Y)=\ldots=1{,}9167\text{ bit/simbol}$$

Zadatak 10: Odredite količinu informacije sadržane u slici koja se sastoji od 500 redaka i 500 točaka u svakom retku. Intenzitet svjetline svake točke može imati 8 stupnjeva. Pojave različitih gradacija svjetline su jednako vjerojatne, a svjetline pojedinih točaka nisu u međusobnoj ovisnosti.

a) 0,25 Mbit/slici
b) 0,75 Mbit/slici
c) 2 Mbit/slici
d) 1,5 Mbit/slici
e) ništa od navedenog

Postupak rješavanja

Srednji sadržaj informacije svake točke iznosi $H(P)$, a ukupna količina informacije u slici je $H(S)$.

$$H(P) = \log_2 8 = 3 \frac{\text{bit}}{\text{točka}}$$

$$H(S) = M \cdot H(P) = 500 \cdot 500 \cdot 3 = 0,75 \frac{\text{Mbit}}{\text{slika}}$$