Sveučilište u Zagrebu
Fakultet elektrotehnike i računarstva

Međuispit iz predmeta **TEORIJA INFORMACIJE**, 2. prosinca 2021.

Pravilo bodovanja zadataka

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

Zadatak 1. Neka je Z slučajna varijabla koja poprima vrijednosti iz skupa {0, 1} i neka vrijedi:

$$P(Z=z)=\begin{cases} p, & z=0 \\ q, & z=1 \end{cases}, \quad p+q=1$$

Također, pretpostavimo da p poprima vrijednosti iz skupa {0, 0,25, 1}, pri čemu vrijedi: $P(p = 0) = P(p = 0{,}25) = P(p = 1) = 1/3$. Koliko iznosi očekivana vrijednost entropije $H(Z)$?

a) 0,3333 bit/simbol; b) 1 bit/simbol; c) 0,8113 bit/simbol; d) 0,2704 bit/simbol;

e) Ništa od navedenog.

Postupak rješavanja:

Prvo je potrebno razmotriti što se zbiva sa slučajnom varijablom Z u ovisnosti o vrijednosti slučajne varijable p.

a) za $p = 0$ vrijedi $P(Z_1=z)=\begin{cases} 0, & z=0 \\ 1, & z=1 \end{cases}$ pa je entropija $H(Z_1)$ jednaka nuli.

b) za $p = 1$ vrijedi $P(Z_2=z)=\begin{cases} 1, & z=0 \\ 0, & z=1 \end{cases}$ pa je entropija $H(Z_2)$ jednaka nuli.

c) za $p = 0{,}25$ vrijedi $P(Z_3=z)=\begin{cases} 0{,}25, & z=0 \\ 0{,}75, & z=1 \end{cases}$ pa je entropija $H(Z_3)$ jednaka

$$H(Z_3)=\frac{1}{4}\log_2 4+\frac{3}{4}\log_2\frac{4}{3}=\log_2 4-\frac{3}{4}\log_2 3=2-\frac{3}{4}\log_2 3\,[\text{bit/simbol}].$$

Prosječna entropija iznosi:

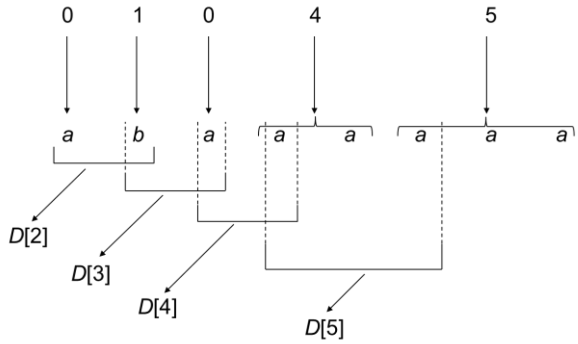
$$E[H(Z)]=\sum_{i=1}^{3}P(p=p_i)H(Z_i)=\frac{1}{3}H(Z_3)=\frac{2}{3}-\frac{1}{4}\log_2 3=0{,}2704\,[\text{bit/simbol}].$$

Zadatak 2. Temeljem polaznog rječnika $D[0] = a$ i $D[1] = b$ dekodirajte primljenu poruku 0 1 0 4 5 koristeći algoritam LZW.

a) abaaaaa (7 znakova); b) abaaaaaaa (9 znakova); c) abaaaaaa (8 znakova);

d) abaaaa (6 znakova); e) Ništa od navedenog.

Postupak rješavanja:



Prošireni rječnik: D[2] = *ab*, D[3] = *ba*, D[4] = *aa*, D[5] = *aaa*. Dekodirana poruka je: *abaaaaaa*.

Zadatak 3. Promatrajte izvor na čijem se izlazu pojavljuju dva simbola, i to: točka (•) i crtica (–). Trajanje točke iznosi 0,2 s, trajanje crtice je tri puta dulje, a trajanje stanke između simbola iznosi 0,2 s. Vjerojatnost pojavljivanja točke je dva puta veća od vjerojatnosti pojavljivanja crtice. Izračunajte prosječnu brzinu generiranja informacije izvora u jedinici bit/s.

a) 0.4897 bit/s; b) 2,7552 bit/s; c) 0,9183 bit/s; d) 1,7219 bit/s; e) Ništa od navedenog.

Postupak rješavanja:

Dakle, trajanje simbola točka $t_1 = t(\bullet) = 0{,}2$ s, trajanje simbola crtica $t_2 = t(–) = 0{,}6$ s, a trajanje stanke $t_s = 0{,}2$ s. Neka je vjerojatnost pojavljivanja točke $P(\bullet) = p_1$, a vjerojatnost pojavljivanja crtice $P(–) = p_2$. S obzirom da je zadano $p_1 = 2p_2$, a mora vrijediti i jednakost $p_1 + p_2 = 1$, slijedi da je $p_2 = 1/3$, a $p_1 = 2/3$. Prosječna količina informacije po svakom simbolu određuje se proračunom entropije zadanog izvora:

$$H(X) = -\sum_{i=1}^{2} p_i \log_2 p_i = 0{,}9183 \frac{\text{bit}}{\text{simbol}}.$$

S obzirom da se iza svakog simbola generira i stanka, za prosječno trajanje generiranog simbola vrijedi:

$$T = p_1(t_1 + t_s) + p_2(t_2 + t_s) = p_1 t_1 + p_2 t_2 + t_s = 0{,}5333 \frac{\text{s}}{\text{simbol}}.$$

Konačno, prosječna brzina generiranja informacije u jedinici vremena iznosi:

$$R = \frac{H(X)}{T} = \frac{0{,}9183}{0{,}5333} = 1{,}7219 \frac{\text{bit}}{\text{s}}.$$

Zadatak 4. Koristeći algoritam LZ77 kodirajte poruku *aaaabbbbbbccccccd** (4×*a*, 6×*b*, 6×*c*, *d*, *), uzimajući pri tome da je maksimalna duljina posmičnog prozora 6, a prozora za kodiranje 5 simbola. Napomena: simbol * označava kraj poruke. Odredite broj kodiranjem dobivenih trojki oblika (pomak, duljina, sljedeći simbol).

a) 5; b) 6; c) 7; d) 4; e) ništa od navedenog.

Postupak rješavanja:

Postupak rješavanja prikazan je u zbirci zadataka „Teorija informacije i kodiranje", 3. izdanje, zadatak 2.18, stranica 98.

Zadatak 5. Instrument mjeri slučajnu veličinu čije su vrijednosti zadane skupom $X = \{ -2, -1, 0, 1, 2\}$. Sve su vrijednosti jednako vjerojatne. Pokazivač instrumenta, namijenjen brojčanom prikazu izmjerene vrijednosti, u kvaru je koji se manifestira tako da se znak za "minus" ne upali u 30%

slučajeva. Promatrajte opisani mjerni sustav kao komunikacijski kanal i odredite transinformaciju u kanalu.

a) 1,9167 bit/simbol;

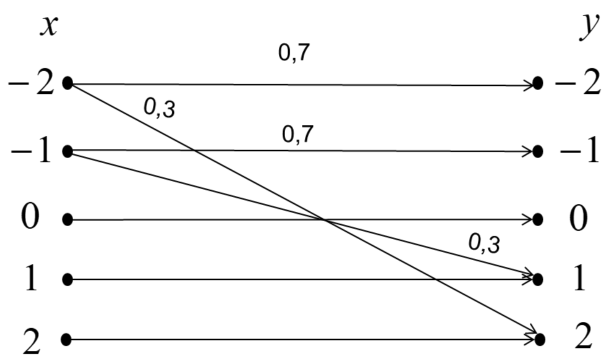
b) 0,4053 bit/simbol;

c) 2,3219 bit/simbol;

d) 2,2692 bit/simbol;

e) ništa od navedenog.

Postupak rješavanja:



Informacijski kanal

Na gornjoj slici x određuje mjerenu veličinu, a y prikazanu. One se ponekad međusobno razlikuju zbog kvara pokazivača. Temeljem skice kanala moguće je odrediti matricu uvjetnih vjerojatnosti prijelaza i matricu združenih vjerojatnosti:

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

Nadalje, s obzirom da su sve vrijednosti mjerene veličine međusobno jednako vjerojatne, vrijedi $P(x_i) = 0,2$, $i = 1, \dots, 5$. Koristeći matricu kanala i apriorne vjerojatnosti mjerene veličine moguće je odrediti matricu parova vjerojatnosti (x_i, y_j) koje čine mjerena i prikazana veličina:

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

Zbrajanjem po stupcima matrice [$P(x_i$, $y_j)$] dobivamo vjerojatnosti pojave izmjerene veličine na pokazivaču, $P(y_j) = [0,14 \;\; 0,14 \;\; 0,2 \;\; 0,26 \;\; 0,26]$, $j = 1, \dots, 5$. Transinformaciju u kanalu moguće je odrediti koristeći izraz:

$$I\left(X;Y\right)=\sum_{i=1}^{5}\sum_{j=1}^{5}P\left(x_i,y_j\right)\log\left(\frac{P\left(x_i,y_j\right)}{P\left(x_i\right)\cdot P\left(y_j\right)}\right).$$

Uvrštavanjem otprije poznatih vrijednosti dobivamo $I(X; Y) = 1,9167$ bit/simbol.

Zadatak 6. Binarni izvor generira dva simbola iz abecede $X_1 = \{x_1, x_2\}$ s pripadajućim vjerojatnostima pojavljivanja $P(x_1) = 2/3$ i $P(x_2) = 1/3$. Nadalje, pretpostavimo da isti izvor kombinira simbole x_1 i x_2 u združene simbole abecede $X_2 = \{x_1x_1, x_1x_2, x_2x_1, x_2x_2\}$, $P(x_i,x_j) = P(x_i)\cdot P(x_j)$, $\forall\ i, j \in \{1, 2\}$. Odredite omjer efikasnosti kôda, ako se Huffmanov kôd primijeni nad proširenom abecedom X_2 u odnosu na njegovu primjenu na početnu abecedu X_1.

a) 2;

b) 18/17;

c) 17/9;

d) 36/17;

e) ništa od navedenog.

Postupak rješavanja:

Ako se Huffmanov kôd primijeni na simbole iz abecede X, tada se x_1 kodira binarnim simbolom 1, a x_2 binarnim simbolom 0. Srednja duljina kodne riječi iznosi 1 bit/simbol, a entropija $H(X)$ iznosi –[2/3 + $\log_2(1/3)$] bit/simbol. Ako primijenimo prošireni Huffmanov kôd, dobivamo četiri združena simbola:

Združeni simbol	Vjerojatnost	Huffmanov kôd
$x_1\, x_1$	4/9	0
$x_1\, x_2$	2/9	10
$x_2\, x_1$	2/9	111
$x_2\, x_2$	1/9	110

Srednja duljina kodne riječi iznosi 17/9 bit/združeni simbol, a entropija združenih simbola, $H_2(X)$, jednaka je $2\cdot H(X)$ [bit/združeni simbol]. Dakle, omjer učinkovitosti Huffmanova koda nad proširenim skupom simbola prema učinkovitosti nad izbornim skupom od dva simbola iznosi:

$$\frac{\varepsilon_2}{\varepsilon_1} = \frac{\dfrac{2H(X)}{17/9}}{\dfrac{H(X)}{1}} = \frac{18}{17}$$

Zadatak 7. Zadan je skup X koji sadrži dva simbola, x_1 i x_2. Simbol x_1 ima stvarnu (izmjerenu) vjerojatnost nastupa $P(x_1) = 0{,}8$. Promatrač eksperimenta greškom zamijeni vjerojatnosti nastupa simbola x_i, $i = 1, 2$, jednu s drugom. Odredite relativnu entropiju stvarne prema pogrešnoj razdiobi vjerojatnosti simbola.

a) 1,2 bit/simbol; b) 1,6 bit/simbol; c) -0,4 bit/simbol; d) 0,117 bit/simbol;

e) Ništa od navedenog.

Postupak rješavanja:

Relativna entropija jedne prema drugoj razdiobi vjerojatnosti nastupa simbola, npr. $P(x_i)$ prema $Q(x_i)$, izračunava se izrazom:

$$D(P\|Q) = \sum_{i=1}^{n} P(x_i)\log_2\frac{P(x_i)}{Q(x_i)}.$$

Dakle, u ovom konkretnom slučaju opisanom u tekstu zadatka vrijedi:

$$D\left(P\|Q\right)=\sum_{i=1}^{2}P\left(x_i\right)\log_2\frac{P\left(x_i\right)}{Q\left(x_i\right)}=0,8\cdot\log_2\frac{0,8}{0,2}+0,2\cdot\log_2\frac{0,2}{0,8}=1,2\,\frac{\text{bit}}{\text{simbol}}\,.$$

Zadatak 8. Zadan je informacijski izvor s memorijom modeliran markovljevim lancem s dva stanja. Taj je lanac opisan matricom prijelaznih vjerojatnosti:

$$\begin{bmatrix}5/8 & 3/8\\ 3/4 & 1/4\end{bmatrix}.$$

Odredite entropiju izvora.

a) 0,2704 bit/simbol; b) 0,8115 bit/simbol; c) 0,9067 bit/simbol; d) 0,6363 bit/simbol;

e) Ništa od navedenog.

Postupak rješavanja:

Stacionarne vjerojatnosti lanca $\boldsymbol{\pi} = [\pi_0, \pi_1]$ moguće je odrediti temeljem jednakosti $\boldsymbol{\Pi}^{\mathrm{T}}\boldsymbol{\pi} = \boldsymbol{\pi}$:

$$\begin{bmatrix}5/8 & 3/8\\ 3/4 & 1/4\end{bmatrix}^{\mathrm{T}}\cdot\begin{bmatrix}\pi_0\\ \pi_1\end{bmatrix}=\begin{bmatrix}\pi_0\\ \pi_1\end{bmatrix}.$$

Pomoćni izraz je $\pi_0 + \pi_1 = 1$. Dakle, $\pi_0 = 2/3$, $\pi_1 = 1/3$. Sada je moguće odrediti entropiju izvora:

$$H=-\sum_{i=0}^{1}\pi_i\sum_{j=0}^{1}p_{ij}\log\left(p_{ij}\right)=-\frac{2}{3}\left(\frac{5}{8}\log_2\frac{5}{8}+\frac{3}{8}\log_2\frac{3}{8}\right)-\frac{1}{3}\left(\frac{3}{4}\log_2\frac{3}{4}+\frac{1}{4}\log_2\frac{1}{4}\right)=0,9067\,\frac{\text{bit}}{\text{simbol}}\,.$$

Zadatak 9. Razmatrajte izvor koji generira četiri simbola iz skupa $X = \{x_1, x_2, x_3, x_4\}$ s odgovarajućim vjerojatnostima pojavljivanja za koje vrijedi:

$$1 > p(x_1) = p_1 > p(x_2) = p_2 > p(x_3) = p_3 > p(x_4) = p_4 > 0 \text{ i } \sum_{i=1}^{4}p_i=1.$$

Svi su simboli potpuno neovisni jedni o drugima. Nadalje, izvor je spojen s koderom informacije koji navedene simbole kodira binarnim simbolima sukladno algoritmu Shannon-Fano, a rezultat toga je prefiksni kôd. Kodne riječi na izlazu kodera informacije, $C(x_i)$, ovise o razdiobi vjerojatnosti simbola $x_i \in X$. Zadane su vjerojatnosti $p_3 = 0,19$ i $p_4 = 0,15$. Neka izvor informacije generira poruku duljine 9 simbola x_2. Sukladno pretpostavci da $C(x_1)$ mora imati duljinu jedan bit, odredite koliko može iznositi najveći sadržaj informacije prenijet porukom sastavljenom od 9 simbola x_2.

a) I < 16,441 bit;

b) I < 21,563 bit;

c) I < 23,959 bit;

d) I < 18 bit;

e) ništa od navedenog.

Postupak rješavanja:

Način kodiranja algoritmom Shannon-Fano ovisi o razdiobi vjerojatnosti $p(x_i)$. Pri tome je važno kako se simboli x_i, ovisno o $p(x_i)$, grupiraju. Bit je algoritma da prilikom podjele simbola u dvije grupe razlika zbroja vjerojatnosti simbola u jednoj i drugoj grupi bude minimalna. U slučaju zadanih simbola x_i i adekvatne razdiobe vjerojatnosti $p(x_i)$, konačan rezultat kodiranja algoritmom Shannon-Fano može biti:

1) $C(x_1) = 00$, $C(x_2) = 01$, $C(x_3) = 10$, $C(x_4) = 11$, ili

2) $C(x_1) = 0$, $C(x_2) = 10$, $C(x_3) = 110$, $C(x_4) = 111$.

Dakle, samo u drugom ishodu kodiranja moguće je ostvariti da $C(x_1)$ ima duljinu jednog bita. Da bi se simboli x_i dijelili u grupe na način koji odgovara binarnom kodu kreiranom u ishodu 2, mora vrijediti:

$| p_1 - (p_2 + p_3 + p_4)| \leq | (p_1 + p_2) - (p_3 + p_4)|$, tj. s obzirom da je $p_3 + p_4 = 0{,}19 + 0{,}15 = 0{,}34$

$| p_1 - p_2 - 0{,}34 | \leq | p_1 + p_2 - 0{,}34|$

Desna strana nejednakosti uvijek je jednaka $p_1 + p_2 - 0{,}34$ zbog uvjeta $1 > p(x_1) > p(x_2) > p(x_3) > p(x_4) > 0$. Lijeva strana nejednakosti može polučiti sljedeće rezultate:

1. za $p_1 \geq p_2 + p_3 + p_4$ vrijedi: $p_1 - p_2 - p_3 - p_4 \leq p_1 + p_2 - p_3 - p_4$, što daje: $2p_2 \geq 0$, a to uvijek vrijedi;

Međutim, iz uvjeta $p_1 \geq p_2 + p_3 + p_4$, tj. $p_1 \geq p_2 + 0{,}34$, te uz $p_2 = 1 - p_1 - (p_3 + p_4) = 0{,}66 - p_1$ mora vrijediti: $2p_1 \geq 1$, tj. $\boldsymbol{p_1 \geq 0{,}5}$; istovremeno, zbog uvjeta $p_2 > p_3$, tj. $p_2 > 0{,}19$, te zbog jednakosti $p_1 + p_2 = 1 - (p_3 + p_4)$, slijedi $p_1 < 0{,}66 - 0{,}19$, tj. $\boldsymbol{p_1 < 0{,}47}$. S obzirom da je ova dva uvjeta za p_1 nemoguće istovremeno zadovoljiti, $p_1 \geq p_2 + p_3 + p_4$ nije opcija koja pogoduje rješenju;

2. za $p_1 \leq p_2 + p_3 + p_4$ vrijedi: $-p_1 + p_2 + p_3 + p_4 \leq p_1 + p_2 - p_3 - p_4$, tj.

$-p_1 + p_2 + 0{,}34 \leq p_1 + p_2 - 0{,}34$,

i konačno: $\boldsymbol{p_1 \geq 0{,}34}$

Kao što je već ranije rečeno, zbog jednakosti $p_1 + p_2 = 1 - (p_3 + p_4)$, slijedi $p_1 < 0{,}66 - 0{,}19$, tj. $\boldsymbol{p_1 < 0{,}47}$. Ova dva uvjeta u opciji 2 moguće je istovremeno zadovoljiti pa je konačno rješenje: $p_1 \in [0{,}34, 0{,}47)$.

Sukladno proračunatom te zbog $p_2 = 1 - (p_1 + p_3 + p_4)$, mora vrijediti: $p_2 \in (0{,}19, 0{,}32]$. Sadržaj informacije sadržan u jednom simbolu x_2 iznosi $I(x_2) = -\log_2(p_2)$ bita. Dakle, maksimalan sadržaj informacije kojeg može prenositi simbol x_2 uz ograničenje u zadatku iznosi $I(x_2) < -\log_2(0{,}19) = 2{,}396$ bita. Konačno, sadržaj informacije u poruci duljine 9 uzastopnih simbola x_2 mora zadovoljavati uvjet:

$$I\left(\underbrace{x_2 \ldots x_2}_{9\,\text{puta}}\right) < 21{,}563\,[\text{bit}]$$

Zadatak 10. 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:

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

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:

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

pri čemu je

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

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:

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

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