TEORIA TRANSMITERII ȘI CODIFICĂRII INFORMAȚIEI

TEORIA TRANSMITERII ŞI CODIFICĂRII INFORMAŢIEI

Cursul "*Teoria transmiterii şi codificării informaţiei*" este o disciplină care îngoblează într-o formă unitară concepte din teoria codurilor, teoria semnalelor aleatoare şi teoria deciziilor statistice şi reprezintă una din disciplinele de pregătire care, pentru profilul INFORMATICĂ, este necesară pentru pregătirea studenţilor şi pentru obţinerea creditelor transferabile prin procedurile de evaluare. Modul de prezentare a acestui material are în vedere particularităţile învăţământului la distanţă, la care studiul individual este determinant. Pentru orice nelămuriri faţă de acest material vă rugăm să contactaţi tutorele de disciplină care are datoria să vă ajute oferindu-vă toate explicaţiile necesare.

Disciplina "*Teoria transmiterii şi codificării informaţiei*" îşi propune următoarele obiective specifice:

- Însuşirea noţiunilor fundamentale din domeniul *Teoriei transmiterii şi codificării informaţiei*.
- Formarea deprinderilor de modelare matematică şi de transpunere în programare a unor probleme de natură tehnică, socială sau economică, cu utilizarea cunoştinţelor însuşite.
- Formarea şi dezvoltarea bazei matematice a studenţilor pentru disciplinele fundamentale şi de specialitate din anii superiori;
- Formarea şi dezvoltarea aptitudinilor şi deprinderilor de analiză logică, formulare corectă şi argumentare fundamentată, în rezolvarea problemelor tehnico-economice şi de specialitate;
- O comparaţie critică a metodelor de rezolvare evidenţiind, eventual, calea optimă de soluţionare.

Vă precizăm de asemenea că, din punct de vedere al verificărilor şi al notării, cu adevărat importantă este capacitatea pe care trebuie să o dobândiţi şi să o probaţi de a rezolva toată tipologia de probleme aplicative aferente materialului teoretic prezentat în continuare. De aceea vă recomandăm să parcurgeţi cu atenţie toate aplicaţiile rezolvate, să rezolvaţi aplicaţiile propuse prin testele de autoevaluare şi temele de control; fiţi convinşi că examenul final apelează la tipurile de aplicaţii prezente în secţiunile menţionate anterior

Coordonator disciplină: Prof. univ. dr.ing. RĂCUCIU CIPRIAN
Tutori: Lector univ. dr.ing. GRECU DAN

MODULUL 1

MĂSURA CANTITATIVĂ A INFORMAŢIEI

În acest modul sunt prezentate principalele noţiuni cu care operează teoria informaţiei. Notiunea de informatie a aparut mult mai tarziu decat notiunea de energie, iar legile dupa care informatia apare, se transforma, se pastreaza, se prelucreaza si se foloseste sunt inca insuficient studiate; abia in zilele noastre se stabilesc bazele intelegerii lor, se elucideaza metodele de studiu si investigare.

Stabilirea notiunii generalizate de informatie pentru caracterizarea proceselor de conducere dintr-un punct de vedere unitar,a fost un moment important in stiinta. Intocmai cum introducerea notiunii de energie a permis sa se analizeze toate fenomenele naturii dintr-un punct de vedere unic, independent de substratul lor fizic, tot asa,introducerea notiunii de informafie a permis studierea dintr-un punct de vedere comun a celor mai diferite procese de comanda din natura.

Se numeste informatie orice stire care poarta in sine urma unui fapt, eveniment sau proces oarecare.

Informatia este comunicarea (mesajul) ce aduce stiri despre fapte, evenimente, obiecte, procese.In intelesul mai larg, in nofiunea de informatie se pot cuprinde toate stirile despre mediul care ne inconjoara sau, mai bine zis, care se obtin, in interactiunea omului cu mediul inconjurator. A obtine o informatie inseamna a afla lucruri ce nu se cunosteau mai inainte sau a obtine noi cunostinte asupra unui lucru, fapt etc., despre care s-a stiut mai putin inainte

Timpul mediu necesar însuşirii noţiunilor teoretice, formării deprinderilor de calcul şi utilizării metodelor de rezolvare a problemelor specifice teoriei informaţiei este estimat la aproximativ 6-8 ore pentru fiecare modul, într-un ritm de 2-3 ore pe zi.

MODULUL 1
CAPITOLUL 1

ELEMENTE DE TEORIA TRANSMITERII INFORMAŢIEI

1.1. Informatia – generalităţi.

În procesele de comanda, procesele energetice care insotesc transmiterea informatiei joaca un rol secundar. Cantitatea de informatie si cu atat mai mult efectul ei,nu sunt determinate de cantitatea de energie folosita pentru transmiterea infornafiei. Esenta proceselor de conducere, care se desfasoara pe baza schimbului de informatie, consta tocmai in aceea ca miscarea si actiunea unor mase materiale mari sau transmiterea si transformarea unor cantitati mari de energie se dirijeaza si se controleaza cu ajutorul unor mase materiale mici si al unor cantitati reduse de energie.

În teoria informatiei, caracteristica energetica a fenomenelor trece pe plan secundar, evidentiindu-se in mod deosebit latura informafionala a sistemului.

Asadar, notiunea de informatie este foarte larga si se poate prezenta sub cele mai variate forme: aceasta constituie o proprietate de seama a informatiei. Prin mijloacele de telecomunicatii -telefon,telegraf, radio - se transmit informantii. Prin intermediul vazului, auzului, precum si al celorlalte simturi, omul primeste zilnic tot felul de informafii despre evenimentele din lumea ce il inconjoara.

Comunicari complexe, ordine si dispozitiuni se transmit cu ajutorul telefonului, telegrafului si radioului. Comunicari si mai complexe sunt cele transmise prin intermediul televiziunii, unde imaginile in miscare sunt insotite de semnale audio.La toate aceste sisteme, transmiterea informatiei este insotita de un fenomen nedorit,de adaugare la informatia transmisa a unor semnale perturbatoare ce nu au fost produse de sursele initiale de informantii; in telefonie se distorsioneaza semnalul de vorbire, in televiziune se deformeaza imaginea, in telegrafie apar greseli de imprimare.

Aceste exemple evidentiaza o alta proprietate de baza a informatiei: in nici un sistem fizic informatia nu apare intr-o forma curata ci este insotita de diferite perturbatii care pot duce la greseli. De aceea,una din problemele principale ale teoriei informatiei consta in stabilirea metodelor optime pentru extragerea informatiei din semnalele care sunt insotite de perturbatii. Notiunea de informatie a cucerit un loc sigur in stiinta numai atunci cand s-a gasit o masura adecvata pentru caracterizarea ei. O alta proprietate de seama a informatiei este aceea de a putea fi masurata. Nu este suficient insa sa se gaseasca o modalitate de masurare a informafiei: trebuie sa existe posibilitatea folosirii acestei masuri, adica sa existe siguranta ca se pastreaza obiectul masuratorii. Tot asa, informatia care ia nastere in cadrul unui sistem bine definit si se pastreaza in limitele sistemului respectiv, poate fi masurata, indiferent de natura sistemului.

Problema principala a teoriei informatiei este studierea transformarii, pastrarii si transmiterii informatiei. Analiza acestui fenomen a fost facuta pentru prima data de inginerii de telecomunicatii, care s-au ocupat cu organizarea canalelor destinate transmiterii informatiei.

În realitate, informatia se transmite prin intermediul semnalelor care poarta stirea. Tipuri de informatie :informatii numerice, informatii logice, de tip text, informatii multimedia:audio, imagine, video, semnale .

1.2. Sistemul de transmitere a informatiei.

Purtatorul material al informatiei - semnalul - isi pastreaza capacitatea sa de a transmite informatia numai in cadrul unui sistem de transmisiuni; schema bloc destul de generala a unui sistem de transmisiuni este data in fig.1.1

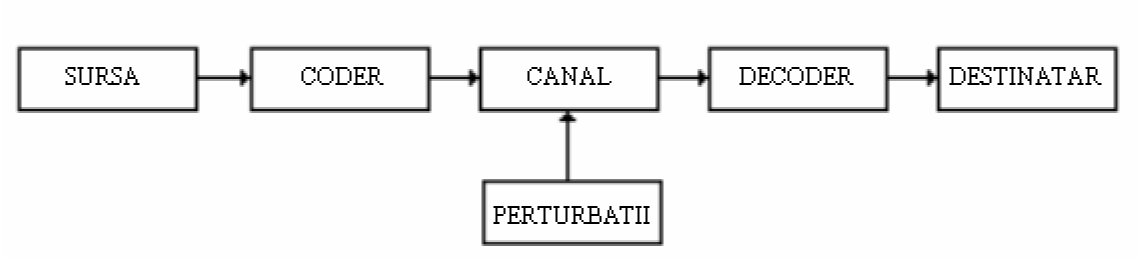


Fig.1.1

Coderul din fig.1.1 executa orice prelucrare a semnalului generat de sursa. O asemenea prelucrare poate include, de exemplu, o anumita combinatie de modulatie, comprimare de date sau introducerea unei redundante pentru lupta cu perturbatiile.

Canalul este mediul fizic utilizat pentru transmiterea semanalului:de exemplu linia telefonica, linia radio sau radioreleu, dispozitivul de memorie sau organismul uman. Asupra canalului, de regula, actioneaza diferite perturbatii care in liniile de telefonie pot apare din cauza modificarilor caracteristicii de frecventa, a convorbirilor ce se induc din alte linii, a zgomotului termic, a impulsurilor parazite, sursa carora pot fi schemele de comutare, a bruiajului intentionat al adversarului etc.

Decoderul executa prelucrarea semnalului de la iesirea canalului in scopul de a reproduce la partea de receptie o copie acceptabila iesirii sursei.

Destinatarul poate fi omul sau un dispozitiv tehnic oarecare. Pentru a simplifica analiza modelelor de surse si canale este de dorit a separa efectele legate de sursa de efectele legate de canal.

Sarcina coderului sursei este de a reprezenta iesirea sursei cu ajutorul succesiunilor de semnale binare, si una din problemele importante ce apar, consta in a stabili cate simboluri binare,in unitatea de timp sunt necesare pentru reprezentarea semnalului de la iesirea unei surse date.

Sarcina coderului si decoderului canalului consta in a reproduce cat mai sigur succesiunile binare ale datelor obtinute la iesirea decoderului canalului si una din problemele inportante ce apare este, daca acest lucru este posibil sa se faca,si cum sa se faca.

Coderul sursei transforma mesajul de la iesirea sursei intr-o succesiune de semnale binare sau care apartin unui alfabet finit, din care decoderul sursei restabileste mesajul initial cu o precizie adoptata de catre destinatar. Astfel, independent de proprietatile sursei sau destinatarului, la intrarea coderului canalului si la iesirea decoderului canalului,se formeaza o succesiune de simboluri binare sau de simboluri care apartin unui alfabet finit. Reprezentarea informatiei de transmis sub forma unei succesiuni binare intermediare da posibilitatea sa se calculeze si sa se construiasca dispozitive de codificare si decodificare de canal, independent de dispozitivele corespunzatoare care se refera la sursa. Sarcina sistemului de transmisiuni, este de a transmite mesajul de la sursa la destinatar, adica de a reproduce mesajul de la iesirea sursei la locul indicat de destinatar. Cand spunem “reproduce" nu intelegem o reproducere absolut fidela ci o reproducere care corespunde anumitor scopuri specifice.Criteriul acceptabilitatii depinde de scopul transmisiuni. În descrierea “obiectului" transmis prin sistemul de transmisie trebuie inclus si criteriul acceptabilitatii. Astfel, obiectul ce se transmite nu determina numai proprietatile sursei, ci caracterizeaza proprietatile cuplulul “sursa-destinatar". Vom numi acest obiect “informatia transmisa".

1.3. Modele de surse informationale.

Ideea fractionarii mesajelor posibile la iesirea sursei intr-o multine discreta de clase are o importanta fundamentala, intrucat conduce la enumerarea reprezentantilor claselor din intervalul de timp dat. Multimea claselor se numeste multimea de mesaje posibile din intervalul de timp dat, admisibile pe timpul transmisiei.Sarcina sistemului de transmisiuni consta in reproducerea mesajului de la iesirea sursei cu o precizie adoptata de catre destinatar. Existenta criteriului de precizie permite sa se grupeze toate mesajele posibile, in orice interval de timp, de la iesirea sursei in clase disjuncte. Sistemul de transmisiuni indica destinatarului clasa din care face parte mesajul respectiv.

Toate sursele in teoria informatiei se modeleaza cu ajutorul proceselor sau succesiunilor aleatorii. Cea mai simpla clasa de modele de surse este clasa surselor discrete fara memorie. La aceste surse iesirea este o succesiune (in timp) de litere, fiecare din ele alese dintr-un alfabet dat, a1, a2, ..., ak. Succesiunea la iesirea sursei este formata din aceste litere alese din alfabet statistic independent si intamplator, alegere ce are la baza o repartitie oarecare de probabilitati P(a1), ..., p(ak). În cazul unei codificari de acest tip, combinatiile de cod vor avea aceeasi lungime pentru ca sa fie posibila decodificarea lor.

Remarcam faptul ca scrierea numerelor, in cazul transmiterii mesajelor admisinile, intr-o forma binara nu are o importanta principala. Ele pot fi scrise in orice sistem de numeratie.Prezinta importanta posibilitatea insasi de transformare a mesajului admisibil intr-o succesiune de simboluri care fac parte dintr-un alfabet finit.Rationamentul prezentat permite sa se presupuna ca numarul de mesaje admisibile sau numarul de simboluri binare, necesare pentru reprezentarea fiecarui mesaj din multimea mesajelor posibile , poate fi luata ca o masura a cantitatii de informatie transmisa de sursa intr-un interval de timp. Dar, dupa cum vom vedea, aceasta presupunere este justa numai intr-un anumit caz. Numarul M de mesaje admise nu descrie complet multimea mesajelor si este necesar sa se ia in consideratie si probabilitatile cu care sunt generate aceste mesaje admisibile.

Prima teorema a lui C. Shannon - teorema fundamentala a codificarii in lipsa zgomotelor da tocmai o limita inferioara si una superioara pentru lungimea medie a combinatiilor de cod. Probabilitatile p(ai) depind de caracteristicile statistice ale sursei si de procedeul de grupare a mesajelor posibile in clase de echivalenta. In afara de aceasta, numarul mediu de sinboluri binare,generate intr-o secunda, depinde de intervalul de timp T cu care s-a lucrat la gruparea mesajelor in clase de echivalenta. Pentru a mari eficacitatea sistemulul de transmisiuni se cauta sa se minimizeze numarul mediu de simboluri binare generate intr-o secunda de coderul sursei.

Unul din rezultatele principale ale teoriei transmisiunii informatiei este: in cazul unor conditii destul de generale poate fi indicat numarul R, care, pentru fiecare cuplu sursa-destinatar, exprima viteza de generare a informatiei pentru un criteriu de precizie adoptat. Aceasta viteza se determina ca cel mai mic numar mediu de simboluri binare intr-o secunda, care trebuie sa se transmita pentru ca mesajul sa poata fi reprodus in conformitate cu criteriul de precizie adoptat.

1.4. Modelul probabilistic al semnalelor.

Avand in vedere ca pentru orice fenomen din natura sau din societate aprecierile cantitative constituie o conditie de baza a analizei stiintifice s-a cautat o modalitate pentu calculul cantitatii de informatie si s-a stabilit o unitate de masura pentru informatia continuta intr-un semnal purtator de informatie. La calculul cantitatii de informatie si la stabilirea unitatii de masura a informatiei se pleaca de la o descriere probabilistica a semnalelor si se considera semnalele ca evenimente aleatoare. Semnalele discrete care intervin in diferite sisteme informationale, destinate transmiterii si prelucrarii datelor, permit o tratare corespunzatoare probabilistica, iar semnalele continue, de asemenea, permit o descriere probabilistica dupa discretizare - ceea ce se face cu ocazia observarii cu o precizie data. Astfel, puten conchide universalitatea cantitatii de informatie obtinuta pe baza unui model probabilistic.

Unitatea de masura a informatiei se refera numai la partea cantitativa a informatiei si intotdeauna se face abstractie de continutul semantic al mesajului. Cauza acestei tratari unilaterale rezida in faptul ca aparatul matematic existent deocamdata,nu ne permite efectuarea unui studiu calitativ al informatiei. Construirea unui model probabilistic pentru semnalele discrete - utilizate in cadrul unui sistem informational dat - presupune cunoasterea probabilitatilor cu care apar aceste semnale in urma unui experiment. Aparitia unui semnal in cadrul unui experiment se considera ca un eveniment.

În cazul semnalelor discrete, totdeauna se poate realiza o corespondenta biunivoca intre semnalele posibile si intre o multime de numere naturale X - facand ca fiecarui semnal sa corespunda un numar natural $x \in X$- si astfel multimea semnalelor discrete se poate inlocui cu o multime de numere naturale X. In continuare, prin x se intelege fie un semnal oarecare, fie numarul natural care corespunde semnalului respectiv.

Modelul probabilistic al semnalelor X este dat prin urmatoarea repartitie a variabilei aleatoare:

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_M \\ P_X(x_1) & P_X(x_2) & \dots & P_X(x_M) \end{pmatrix} \tag{1,1}$$

Prin notatia $P_X(x_k)$ se intelege probabilitatea de aparitie a evenimentului $x=x_k$, adica probabilitatea de aparitie a semnalului care corespunde lui xk si care face parte din multimea X considerata. Evident:

$$\sum_{K=1}^{M} P_X(x_K) = 1 \tag{1.2}$$

Probabilitatea ca rezultatul experimentului va fi un element oarecare x, se noteaza cu Px(x) in care, prin indicele X se accentueaza ca rezultatul experimentului este un semnal din multimea semnalelor posibile X. In cazul cand nu exista ambiguitati in privinta apartenentei lui x se poate suprima indicele X. Experimentele cu mai multe rezultate simultane vor fi caracterizate prin elementele unui produs de multimi si prin probabilitatile de aparitie a acestor elemente.

MODULUL 1
CAPITOLUL 2

MĂSURI INFORMAŢIONALE

2.1. Modele de surse informationale

Informatia proprie, ca si informatia reciproca se poate considera ca o variabila aleatoare si se calculeaza valoarea sa medie. Valoarea medie a informatiei proprii pentru evenimente din multinea X se numeste entropia lui X si se calculeaza cu formula:

$$H(X)=\sum_{K=1}^{K}P_X(x_K)\log\frac{1}{P_X(x_K)} \tag{1,3}$$

sau mai simplu se scrie sub forma:

$$H(X)=\sum_{K=1}P(x)\log\frac{1}{P(x)} \tag{1.4}$$

Variatia entropiei unui sistem de evenimente format din doua evenimente in functie de repartitia de probabilitati este data in fig 1.2:

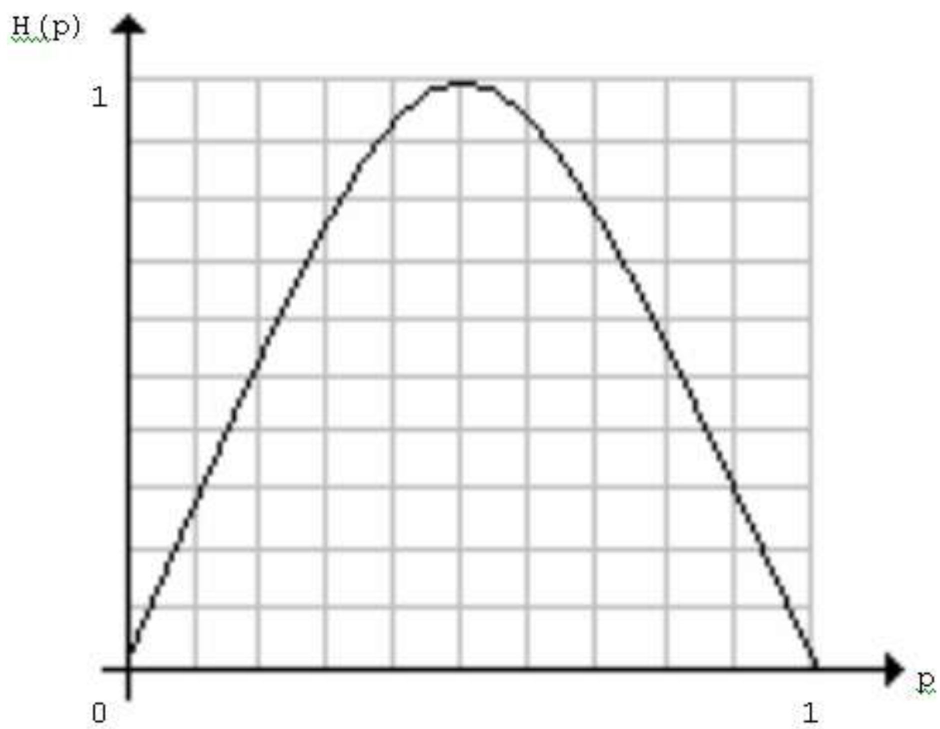


Fig.1.2 Entropia semnalului

Fie

$$X=\begin{pmatrix} x_1 & x_2 \\ p & 1-p \end{pmatrix} \tag{1.5}$$

atunci:

$$H(X) = p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p} = H(p) \tag{1.6}$$

2.2. Informatia proprie conditionata.

Cantitatea de informatie proprie conditionata a evenimentului X=Xk ,cu conditia ca a aparut evenimentul Y=Yi se defineste pe produsul cartezian XY si se calculeaza cu formula:

$$I_{\frac{X}{Y}}\left(\frac{x_k}{y_i}\right) = \log\frac{1}{P_{\frac{x}{y}}\left(\frac{x_k}{y_i}\right)} \tag{1.7}$$

sau mai simplu se scrie:

$$I_{\frac{x}{-}}\left(\frac{x}{y}\right) = \log\frac{1}{P\left(\frac{x_k}{y_i}\right)} \tag{1.8}$$

Aceasta cantitate de informatie proprie a evenimentului X=Xk, conditionata de evenimentul y=yi se poate interpreta ca informatia necesara pentru specificarea evenimentului x=xk . dtupa ce a avut loc evenimentul y =yi.

Cu ajutorul relatiilor anterioare se poate exprima cantitatea de informatie reciproca ca o diferenta intre informatia proprie si informatia proprie conditionata.

$$I(x;y) = \log\frac{P\left(\frac{x}{y}\right)}{P(x)} = \log\frac{1}{P(x)} - \log\frac{1}{P\left(\frac{x}{y}\right)} \tag{1.9}$$

de unde

$$I(x;y) = I(x) - I\left(\frac{x}{y}\right) \tag{1.10}$$

Din relatia anterioara se obtine pe baza reciprocitatii relatia:

$$I(x;y) = I(y) - I\left(\frac{y}{x}\right) \tag{1.11}$$

In mod analog I(x;y) se poate scrie sub forma:

$$I(x;y)=\log\frac{P\left(\frac{x}{y}\right)}{P(x)}=\log\frac{P(y)P\left(\frac{x}{y}\right)}{P(y)P(x)}=\log\frac{P(x,y)}{P(y)P(x)}=\log\frac{1}{P(x)}+\log\frac{1}{P(y)}-\log\frac{1}{P(x,y)} \quad (1.12)$$

deci:

$$I(x;y)=I(x)+I(y)-I(x,y) \quad (1.13)$$

unde :

$$I(x,y)=\log\frac{1}{P(x,y)} \quad (1.14)$$

reprezinta cantitatea de informatie proprie a unui eveniment (x;y) din produsul cartezian de evenimente xy.

Tinand seama ca:

$$P(x,y)=P(x)P\left(\frac{y}{x}\right)=P(y)P\left(\frac{x}{y}\right) \quad (1.15)$$

rezulta ca, cantitatea de informatie proprie a unui eveniment (x,y) se poate scrie sub forma:

$$I(x,y)=I(x)+I\left(\frac{y}{x}\right)$$
$$I(x,y)=I(y)+I\left(\frac{x}{y}\right) \quad (1.16)$$

Luand mediile expresiilor din relatiile anterioare se obtine:

$$I(x;y)=H(x)-H\left(\frac{x}{y}\right)$$
$$I(x;y)=H(y)-H\left(\frac{y}{x}\right)$$
$$I(x;y)=H(x)+H(y)-H(x,y)$$
$$H(x,y)=H(x)+H\left(\frac{y}{x}\right)$$
$$H(x,y)=H(y)+H\left(\frac{x}{y}\right) \quad (1.17)$$

Prima relatie din sistemul de mai sus permite interpretarea lui I(X;Y). Deoarece H(X) este nedeterminarea lui X,iar H(X/Y) este nedeterminarea lui X dupa receptionarea lui Y, rezulta ca diferenta H(X) - H(X/Y) arata cu cat s-a micsorat nedeterminarea lui X prin observarea lui y, adica ce cantitate de informatie se transmite despre X prin observarea lui Y. Iata motivul pentru care cantitatea de informatie medie I(X;Y) este numita si informatia transmisa sau pe scurt transinformatia.

Din formula entropiei, data de catre C. Shannon in anul 1948 in lucrarea sa, rezulta ca entropia U(X) a multimii X depinde numai de probabilitatile de aparitie a elementelor x€X. Evident daca multimile X si Y au aceeasi repartitie de probabilitati atunci H(X)=H(y),insa invers,din egalitatea entropiilor nu rezulta identitatea repartitiilor.

Proprietatea 1. H(X)>0

In adevar,suma H(X) contine termeni de forma P(x)log(1/P(x)),care sunt mai mari sau egali cu zero pentru 0<P(x)<1. Daca P(xk)=1, repartitia lui X este degenerata si este de forma:

$$X = \begin{pmatrix} x_1 & \dots & x_k & \dots & x_K \\ 0 & \dots & 1 & \dots & 0 \end{pmatrix} \tag{1.18}$$

si H(x)=0.Daca repartitia lui Xnu este degenerate,atunci H(x)>0.

Inainte de a trata celelalte prorprietati ale entropiei se da o inegalitate importanta care a fost pusa in evidenta de Jensen.

Fie x,y,Psi Q numere positive si P+Q=1,atunci avem:

$$\log(Px + Qy) \geq P\log x + Q\log y \tag{1.19}$$

in care avem egalitate daca si numai daca x=y.

Aceasta inegalitate scoate in evidenta proprietatea de convexitate a functiei logaritmice,o proprietate care de asemenea joaca un rol important in teoria informatiei.

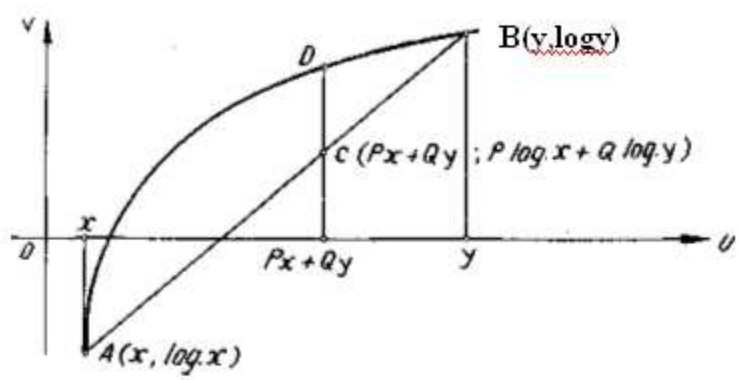


Fig.1.3.

Proprietatea 2.

Fie X multimea formata din K semnale ,atunci:

$$H(x) \leq \log K \tag{1.20}$$

unde avem egalitate,daca si numai daca repartitia lui X este uniforma adica:

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_k \\ \frac{1}{k} & \frac{1}{k} & \dots & \frac{1}{k} \end{pmatrix} \tag{1.21}$$

Proprietetea 3.

Pentru doua multimi de semnale X si Y ,avem:

$$H(x,y) \le H(x) + H(y) \tag{1.22}$$

in care avem egalitate daca si numai daca X si Y sunt statistic independente.

Exerciţii rezolvate.

Conceptia Shannon pleaca de la premiza ca orice informatie cu privire la un eveniment este utilizata in scopul reducerii gradului de incertitudine asupra realizarii acelui eveniment. Din punctul de vedere al destinatarului, comunicatia este o variabila aleatoare, continutul informational fiind cu atat mai mare cu cat el se asteapta mai putin la realizarea acelui eveniment.

Fie o sursa discreta care emite unul dintre cele q mesaje $m_1, m_2, \ldots, m_q$ cu probabilitatile de aparitie $p_1, p_2, \ldots, p_q$. Este evident ca probabilitatile satisfac relatia $p_1 + p_2 + \ldots + p_q = 1$. Continutul informational al mesajului k este notat cu $I(m_k)$. Pentru ca acest continut sa fie o masura adecvata a cantitatii de informatie este necesara satisfacerea urmatoarelor proprietati :

(i) daca $p_k < p_j \quad \Rightarrow \quad I(m_k) > I(m_j)$

(ii) daca $p_k \to 1 \quad \Rightarrow \quad I(m_k) \to 0$

(iii) daca $0 \le p_k \le 1 \quad \Rightarrow \quad I(m_k) \ge 0$

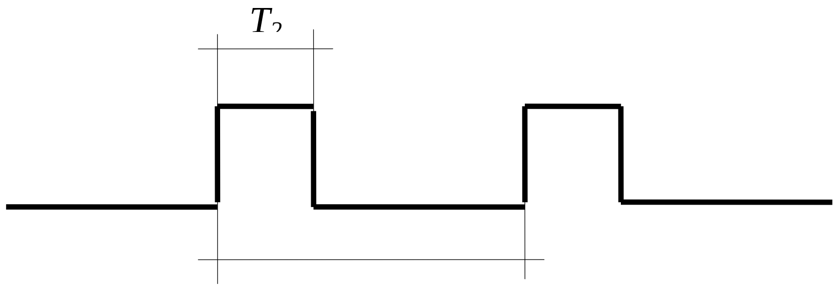
(iv) (aditivitatea) daca m_k si m_j sunt mesaje independente $\Rightarrow \quad I(m_k \text{ si } m_j) = I(m_k) + I(m_j)$

O functie continua de variabila p_k care satisface proprietatile (i)-(iv) este functia logaritmica.

$$I(m_k) = \log \frac{1}{p_k} = -\log p_k$$

Daca baza logaritmului este 2, unitatile de masura sunt biti (informationali).

Pentru cazul simplu al unei transmiteri seriale asincrone



se definesc

(a) rata de biti=(durata unui bit)$^{-1}$ $=1/T_2$ exprimata in biti/secunda (abreviat bps).

(b) rata de bauds=(durata minima intre doua modificari ale semnalului) $=1/T_1$ exprimata in bauds.

Problema 1

Se considera o trasmisie fax : $2{,}25\cdot 10^6$ pixeli cu 12 tonuri de gri, echiprobabile. Care este cantitatea de informatie transmisa ?

Solutie

I=nr.elemente · informatie per element=

$$= 2{,}25\cdot 10^6 \cdot \left[-\log_2 \frac{1}{12}\right] = 2{,}25\cdot 10^6 \cdot \log_2 2^2 \cdot 3 = 2{,}25\cdot 10^6 (2+\log_2 3) \quad [\text{biti}]$$

Problema 2

Un display monocolor cu	24 linii
	80 caractere/linie
	128 puncte/caracter
	3 tonuri de gri/punct

(a) Care este cantitatea de informatie pe pixel, caracter, ecran ?

(b) Care este debitul de informatie stiind ca frecventa cadrelor este de 24 cadre/secunda ?

Solutie

(a) $I = 24\cdot 80\cdot 128\cdot \log_2 3 \quad [\text{biti}]$

(b) $\tau = \dfrac{1}{f_c}$

$$R = \frac{I}{\tau} = I\cdot f_c \quad [\text{bps}]$$

Problema 3

Un echipament de teletransmisie genereaza cuvinte constituite dintr-un grup de 4 impulsuri de tensiune care pot avea nivelurile 0,1,2 sau 3 volti (echiprobabile) urmate de un impuls de pauza de nivel -1 volt. Durata tuturor impusurilor este de 1 ms.

(a) Care este debitul de informatie ?

(b) Care este rata de bauds ?

Solutie

(a) $R = \frac{I}{\tau} = \frac{4\log_2 4}{5\cdot 10^{-3}} = 1{,}6 \quad [\text{kbps}]$

(b) $r_{\text{bauds}} = \frac{1}{10^{-3}} = 10^3\,[\text{baud}] = 1\,[\text{kbaud}]$

Problema 4

Fie 12 monede dintre care una este falsa (mai usoara sau mai grea decat celelalte). Se cere sa se deterrmine numarul minim de cantariri necesar depistarii monedei false si precizarii daca ea este mai usoara sau mai grea. Se foloseste pentru cantariri o balanta fara mase marcate.

Solutie

- cantitatea de informatie necesara determinarii monedei false este $I_1 = \log_2 \frac{1}{\frac{1}{12}} = \log_2 12$
- cantitatea de informatie necesara pentru a decide daca moneda este mai grea sau mai usoara este $I_2 = \log_2 \frac{1}{\frac{1}{2}} = \log_2 2$
- cantitatea de informatie totala necesara a fi determinata $I = I_1 + I_2 = \log_2 24$
- cantitatea de informatie furnizata de o cantarire (exista 3 stari ale balantei) $I_3 = \log_2 \frac{1}{\frac{1}{3}} = \log_2 3 \quad \Rightarrow$ numarul minim de cantariri $I \le kI_3 \quad \Rightarrow \quad 24 \le 3^k \quad \Rightarrow \quad k = 3.$
- sa se propuna un algoritm de depistare.

Problema 5

Sa se reia problema # 3 daca probabilitatile de aparitie a nivelurilor sunt
nivel 0: 1/2
nivel 1:1/4
nivel 2:1/8
nivel 3:1/8

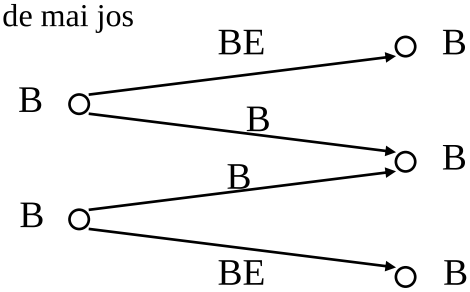
Solutie

$$R = r\cdot H = \frac{1}{5\cdot 10^{-3}}\cdot\left[-\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - 2\cdot\frac{1}{8}\log\frac{1}{8}\right]$$

Se reaminteste ca entropia unei surse reprezinta numarul mediu de biti/simbol si ca entropia este maxima pentru o sursa care genereaza simboluri echiprobabile $H_{\max} = \log_2 n$.

Problema 6

Sa se determine capacitatea unui canal binar cu zona de anulare avand graful asociat matricei de tranzitie din figura de mai jos



Solutie

Acest model descrie un canal care perturba simbolurile unei surse binare in masura in care la receptie sa poata fi interpretate ca fiind incerte.

Metoda scalara

$$C = \max_{p(x_i)}\left[H(Y) - H\left(Y|X\right)\right]$$

$$H(Y) = -\sum_{i=1}^{3} p(y_i)\log p(y_i)$$

$$p(y_1) = p\left(y_1|x_1\right)p(x_1) + p\left(y_1|x_2\right)p(x_2) = p\left(y_1|x_1\right)p(x_1) = (1-q)\,p(x_1)$$

S-a utilizat formula probabilitatii totale, evenimentele x_1, x_2 fiind mutual exclusive.
Analog se poate scrie

$$p(y_2) = (1-q)\,p(x_2)$$

$$p(y_3) = p\left(y_3|x_1\right)p(x_1) + p\left(y_3|x_2\right)p(x_2) = q\left(p(x_1) + p(x_2)\right) = q\cdot 1 = q \quad \Rightarrow$$

$$\begin{aligned} H(Y) &= -\left[(1-q)\,p(x_1)\log(1-q)\,p(x_1) + q\log q + (1-q)\,p(x_2)\log(1-q)\,p(x_2)\right] = \\ &= -(1-q)\left[p(x_1)\log p(x_1) + p(x_2)\log p(x_2)\right] - (1-q)\log(1-q) - q\log q = \\ &= (1-q)H(X) - (1-q)\log(1-q) - q\log q \end{aligned}$$

$$H\left(Y|X\right) = -\sum_{i=1}^{2}\sum_{j=1}^{3} p(x_i, y_j)\log p\left(y_j|x_i\right)$$

Conform regulii de inlantuire

$$p(x_i, y_j) = p\left(y_j|x_i\right)p(x_i)$$

Din graf se deduce ca

$$p\left(y_1|x_1\right) = 1-q \qquad\qquad p\left(y_1|x_2\right) = 0$$

$p\left(y_2 \middle| x_1\right) = 0 \qquad\qquad p\left(y_2 \middle| x_2\right) = q$

$p\left(y_3 \middle| x_1\right) = q \qquad\qquad p\left(y_3 \middle| x_2\right) = 1 - q$

Se obtine

$H\left(Y \middle| X\right) = -(1-q)\log(1-q) - q\log q$

$$C = (1-q)\max_{p(x_i)} H(X) = (1-q)\cdot 1 = 1-q$$

pentru setul optim de probabilitate la intrare $p_0(x_1) = p_0(x_2) = 1$.

Metoda matriciala

Se considera sursa care genereaza un alfabet de simboluri x_i , $i = 1,\ldots,n$ si la destinatie se receptioneaza un alfabet de simboluri y_j , $j = 1,\ldots,m$.

Se pot scrie urmatoarele relatii:

$$P(Y) = P(X) \cdot P\left(Y \middle| X\right)$$

unde $P(X)$ este matricea linie a sursei $(1 \times n)$;

$P(Y)$ este matricea linie a destinatiei $(1 \times m)$;

$P\left(Y \middle| X\right)$ este matricea dreptunghiulara de zgomot $(n \times m)$.

Observatie

In matricea de tranzitie (zgomot) liniile sunt asociate intrarilor iar coloanele sunt asociate iesirilor.

Matricea campurilor reunite (joint) este

$$P(X,Y) = \begin{bmatrix} p(x_1) & 0 & \ldots & 0 \\ 0 & p(x_2) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & \ldots & p(x_n) \end{bmatrix} P\left(Y \middle| X\right)$$

Matricea de zgomot $P\left(Y \middle| X\right)$ se poate obtine prin impartirea fiecarei linii *i* prin $p(x_i)$.

Matricea de echivocatie $P\left(X \middle| Y\right)$ se poate obtine prin impartirea fiecarei coloane *j* prin $p(y_j)$.

Problema 7

Fie matricea de tranzitie $P\left(Y \middle| X\right) = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix}$ si $p(x_1)$=3/4 si $p(x_2)$=1/4.

Se cere sa se calculeze

(a) entropia campului de intrare

(b) entropia campului de iesire
(c) entropia campurilor reunite
(d) eroarea medie
(e) echivocatia
(f) transinformatia
(g) capacitatea canalului si setul optim la intrare
(h)eficienta si redundanta relativa a canalului

Solutie

(a) $$H(X) = -\sum_{i=1}^{2} p(x_i)\log p(x_i) = -\left(\frac{3}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{1}{4}\right) = 2 - \frac{3}{4}\log 3 \cong 0{,}81 \text{ bit / simbol}$$

(b) $$H(Y) = -\sum_{j=1}^{2} p(y_j)\log p(y_j)$$

$$P(Y) = \begin{bmatrix} 3/4 & 1/4 \end{bmatrix}\begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix} = \begin{bmatrix} 7/12 & 5/12 \end{bmatrix}$$

$$H(Y) = -\left(\frac{7}{12}\log\frac{7}{12} + \frac{5}{12}\log\frac{5}{12}\right) = 0{,}98 \text{ bit / simbol}$$

(c) $$H\left(X|Y\right) = -\sum_{i=1}^{2}\sum_{j=1}^{2} p(x_i, y_j)\log p\left(x_i|y_j\right)$$

$$P(X,Y) = \begin{bmatrix} 3/4 & 0 \\ 0 & 1/4 \end{bmatrix}\begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/4 \\ 1/12 & 1/6 \end{bmatrix}$$

$$H(X,Y) = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{12}\log\frac{1}{12} + \frac{1}{6}\log\frac{1}{6}\right) = 1{,}73 \text{ bit / simbol}$$

(d) $$H\left(Y|X\right) = -\sum_{i=1}^{2}\sum_{j=1}^{2} p(x_i, y_j)\log p\left(y_j|x_i\right) = 0{,}92 \text{ bit / simbol}$$

sau $$H\left(Y|X\right) = H(X,Y) - H(X)$$

(e) $$H\left(X|Y\right) = -\sum_{i=1}^{2}\sum_{j=1}^{2} p(x_i, y_j)\log p\left(x_i|y_j\right)$$

unde

$$P\left(X|Y\right)=\begin{bmatrix}\frac{\frac{1}{2}}{\frac{7}{12}} & \frac{\frac{1}{4}}{\frac{5}{12}}\\ \frac{\frac{1}{12}}{\frac{7}{12}} & \frac{\frac{1}{6}}{\frac{5}{12}}\end{bmatrix}=\begin{bmatrix}\frac{6}{7} & \frac{3}{5}\\ \frac{1}{7} & \frac{2}{5}\end{bmatrix}$$

si $\quad H\left(X|Y\right)=-\left(\frac{1}{2}\log\frac{6}{7}+\frac{1}{4}\log\frac{3}{5}+\frac{1}{12}\log\frac{1}{7}+\frac{1}{6}\log\frac{2}{5}\right)=0{,}75\text{ bit / simbol}$

sau $\quad H\left(X|Y\right)=H(X,Y)-H(Y)$

(f) $\quad I(X,Y)=H(X)+H(Y)-H(X,Y)=0{,}06\text{ bit / simbol}$

(g) canalul fiind dublu uniform

$$C=\log 2+\left(1-\frac{1}{3}\right)\log\left(1-\frac{1}{3}\right)+(2-1)\frac{1}{3}\log\frac{1}{3}=0{,}082\text{ bit / simbol}$$

Setul optim $p_0(x_1)$, $p_0(x_2)$ se obtine din

$$C=\max_{p_0(x_i)}\left[H(Y)-H\left(Y|X\right)\right]$$

$$p(y_1)=\frac{2}{3}p_0(x_1)+\frac{1}{3}p_0(x_2)$$

$$p(y_2)=\frac{1}{3}p_0(x_1)+\frac{2}{3}p_0(x_2)$$

$$H\left(Y|X\right)=-\sum_{i=1}^{2}\sum_{j=1}^{2}p(x_i,y_j)\log p\left(y_j|x_i\right)=-\sum_{i=1}^{2}\sum_{j=1}^{2}p\left(y_j|x_i\right)p_0(x_i)\log p\left(y_j|x_i\right)=$$
$$=-\left[\frac{2}{3}\log\frac{2}{3}+\frac{1}{3}\log\frac{1}{3}\right]\left(p_0(x_1)+p_0(x_2)\right)=-\left[\frac{2}{3}\log\frac{2}{3}+\frac{1}{3}\log\frac{1}{3}\right]$$

deci $H\left(Y|X\right)$ nu depinde de $P(X)\Rightarrow\ C=\max\limits_{p_0(x_i)}H(Y)+\left[\frac{2}{3}\log\frac{2}{3}+\frac{1}{3}\log\frac{1}{3}\right]$

dar $\max_{p_0(x_i)} H(Y) = \log 2 = 1 \Leftrightarrow p(y_1) = p(y_2) = \frac{1}{2} \quad \Rightarrow$

$$\begin{cases} p(y_1) = p(y_1|x_1)p_0(x_1) + p(y_1|x_2)p_0(x_2) \\ p(y_2) = p(y_2|x_1)p_0(x_1) + p(y_2|x_2)p_0(x_2) \end{cases}$$

$\Rightarrow \quad p_0(x_1) = p_0(x_2) = \frac{1}{2}$.

Observatie

Se mai poate utiliza si relatia $p_0(x_1) + p_0(x_2) = 1$

(h) $\eta(C) = \frac{I(X,Y)}{C} = 0{,}73$

$R(C) = C - I(X,Y) = 0{,}022$ bit / simbol

Problema 8

Fie un canal binar simetric avand matricea campurilor reunite $P(X,Y) = \begin{bmatrix} 0{,}4 & ? \\ ? & 0{,}4 \end{bmatrix}$ si pentru care sursa genereaza simboluri echiprobabile.

(a) Calculati matricea $P(X,Y)$.
(b) Calculati matricea de zgomot.
(c) Calculati transinformatia.

Solutie

(a) canalul fiind simetric $\Rightarrow p(x_2,y_1) = p(x_1,y_2)$ si folosind proprietatea (iii) a evenimentelor mutual exclusive se obtine $p(x_1,y_1) + 2p(x_1,y_2) + p(x_2,y_2) = 1 \quad \Rightarrow \quad p(x_1,y_2) = 0{,}1$ si $P(X,Y) = \begin{bmatrix} 0{,}4 & 0{,}1 \\ 0{,}1 & 0{,}4 \end{bmatrix}$.

(b) $$P(Y|X) = \begin{bmatrix} \frac{0{,}4}{1/2} & \frac{0{,}1}{1/2} \\ \frac{0{,}1}{1/2} & \frac{0{,}4}{1/2} \end{bmatrix} = \begin{bmatrix} 0{,}8 & 0{,}2 \\ 0{,}2 & 0{,}8 \end{bmatrix}$$

TESTE DE AUTOEVALUARE ŞI TEME DE CONTROL

Testul nr. 1

Fie un alfabet format din literele A,B,C. Se cere să se calculeze:

a. numărul maxim de mesaje de lungime 3 ce se pot forma cu acest alfabet;
b. cantitatea de informaţie conţinută de un asemenea mesaj.

Testul nr. 2

Să se calculeze cantitatea de informaţii necesară pentru precizarea poziţiei unei figuri pe tabla de şah.

Temă de control

Matricea probabilităţilor reunite intrare-ieşire asociată unui canal de transmisie este de forma:

$$P(Y|X) = \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix}$$

Se cere să se calculeze:

a. entropia câmpului de la intrare;
b. entropia câmpului de la ieşire;
c. entropia câmpurilor reunite;
d. eroarea medie şi echivocaţia;
e. transinformaţia şi capacitatea canalului.

BIBLIOGRAFIE RECOMANDATĂ LA MODULUL 1:

[1] A. Spătaru: Teoria Transmisiunii Informaţiei, Ed. Didactică şi Pedagogică, Bu- cureşti, 1983.

[2] A. T. Murgan, I. Spânu, I. Gavăt, I. Sztojanov, V. E. Neagoe, A. Vlad: Teoria Transmisiunii Informaţiei - probleme, Ed. Didactică şi Pedagogică, Bucureşti, 1983

[3] I. Angheloiu, Teoria codurilor, Ed. Militară, Bucureşti, 1972.

[4] J.C. Moreira, P.G. Farrell, ESSENTIALS OF ERROR-CONTROL CODING, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England,2006.

MODULUL 2
CODAREA SURSELOR INFORMAŢIONALE

Compresia datelor este un proces de re prin care se urmăreşte micşorarea redundanţei mesajului generat de o sursă, pentru a reduce resursele necesare memorării sau transmiterii acestui mesaj. Deoarece, de regulă, pentru memorarea sau transmiterea unui mesaj se foloseşte, eventual într-o formă intermediară, reprezentarea prin simboluri binare a acestuia, şi pentru că cele mai multe metode de compresie folosesc metode de codare binar - binar, putem spune că obiectivul compresiei constă în reducerea numărului de simboluri binare necesar pentru reprezentarea mesajului.

După cum şirul simbolurilor emise de sursă este împărţit, pentru codare, în subşiruri de aceeaşi lungime sau de lungime variabilă, şi după lungimea, constantă sau variabilă, a cuvintelor de cod, codurile de compresie se clasifică în bloc- bloc, bloc – variabil, variabil – bloc şi variabil – variabil, bloc – bloc indicând aceeaşi lungime pentru subşirurile sursei şi cuvinte de cod de lungime fixă, iar variabil – variabil corespunzând unor lungimi variabile ale subşirurilor şi ale cuvintelor de cod. Din punct de vedere al măsurii în care mesajul refăcut prin decompresie se aseamănă cu cel original, asupra căruia s-a acţionat prin procesul de compresie, distingem două categorii de algoritmi de compresie: fără pierderi şi cu pierderi.

Algoritmii de compresie fără pierderi sunt reversibili, prin decompresie obţinându-se mesajul original întocmai. Algoritmii de compresie cu pierderi au ca rezultat diferenţe relativ mici între mesajul rezultat prin decompresie şi cel original şi sunt utilizaţi în compresia imaginilor şi a mesajelor video şi audio.

În acest modul vor fi prezentate aspecte privind codificarea surselor discrete şi vor fi analizaţi algoritmii Huffman şi Shannon-Fano.

MODULUL 2
CAPITOLUL 1

CODAREA SURSELOR INFORMAŢIONALE

1.1 Codificarea surselor discrete.

In general,trecerea de la un sistem de semnale primare la un sistem de semnale secundare se numeste codificare. Cerinta principala care se pune pentru orice codificare practic utilizabila este ca codificarea sa fie efectuata in asa fel ca revenirea de la sistemul de semnale secundare la cele initiale, adica decodificarea sa fie unica.

Pentru a prezenta problematicile legate de codificarea semnalelor prinare cu ajutorul semnalelor secundare este necesar sa se precizeze forma concreta a semnalelor primare si secundare. Se considera ca semnalele prinare sunt semnale generate de catre o sursa discreta si fara memorie, adica semnalele apar sub o forma discreta si probabilitatea de aparitie a unui simbol nu depinde de celelalte sinboluri anterioare.

Alfabetul sursei discrete se noteaza cu $A=\{a_1,a_2....,a_L\}$,deci sursa produce succesiuni de litere $u_1,u_2,....,u_n$ formate cu ajutorul literelor din alfabetul sursei, deci $u_i \in A$ pentru i=1,2,….

Faptul ca sursa este fara memorie se exprima prin interdependenta statistica a semnalelor,adica probabilitatea unei succesiuni date $\alpha^n=(\alpha_1,\alpha_2,...,\alpha_n)$ de n litere de sursa este egala cu produsul probabilitatilor literelor de sursa ,deci:

$$P(\alpha_1,\alpha_2,....,\alpha_n)=\prod_{i=1}^{n}P(\alpha_i) \tag{2.1}$$

1.2 Codificarea neuniforma.

Presupunem ca alfabetul unei surse discrete si fara memorie contine M semnale primare sau litere notate cu $a_1,a_2,....,a_M$ care apar cu probabilitatile $P(a_1),P(a_2),..,P(a_M)$. Fiecare litera a sursei trebuie sa fie codificata printr-o combinatie de semnale secundare luate din alfabetul codului care contine D semnale $X=\{x_1,x_2,..,x_D\}$.Aceste combinatii se numesc si cuvinte de cod;cuvintele de cod care corespund semnalelor primare $a_1,a_2,..,a_M$ se noteaza cu $c_1,c_2,..,c_M$,iar totalitatea lor formeaza un cod, $C=\{c_1,c_2,..,c_M\}$.Numarul semnalelor secundare din cuvantul de cod c_k care,dupa cum s-a vazut,corespunde lui a_k ,se noteaza cu n_k .

Dintre toate, codurile unic decodabile prezinta un interes ,din punct de vedere economic,acel cod care conduce la un $\bar{n}$ -numarul mediu de litere de cod pe litere de sursa-cat mai mic posibil,unde $\bar{n}$ se mai numeste si lungimea mediea combinatiilor de cod:

$$\bar{n} = \sum_{K=1}^{M} P(a_k) n_k \tag{2.2}$$

Codul in care doua semnale primare distincte sunt codificate printr-o singura combinatie de cod se numeste cod singular si in mod corespunzator codul in care toate combinatiile de cod atribuite semnalelor primare sunt distincte se numeste cod nesingular.

Codul $C = \{c_1, c_2, .., c_k\}$ se numeste unic decodabile daca succesiunile cuvintelor de cod,corespunzatoare diferitelor succesiuni de lungime finite a sursei,sunt distincte.

Fie un cuvant de cod $c_i = x_{i1}, x_{i2}, ..., x_{im}$ .Sirul de litere $x_{i1}, x_{i2}, ..., x_{im}$,unde $k \le m$ se numeste prefixul lui c_i .

Cu ajutorul notiunii de prefix putem defini subclasa codurilor unic decodabile care prezinta un interes deosebit din punct de vedere practice.

Codul in care nici un cuvant de cod nu este prefixul unui alt cuvant de cod se numeste cod cu proprietate de prefix.Din aceasta definitie rezulta ca in cazul unui cod cu proprietate de prefix,dintr-un cuvant de cod mai lung nu se poate obtine un cuvant de cod mai scurt prin reducerea (suprimarea) ultinelor simboluri,motiv pentru care codurile cu proprietate de prefix se numesc si coduri ireductibile.

Codurile cu proprietatea de prefix se mai numesc uneori si coduri instantanee, deoarece o combinatie de cod se poate recunoaste fara nici o referinta la urmatoarea combinatie de cod. La decodificarea unei succesiuni de cuvinte de cod dintr-un cod cu proprietatea de prefix se inainteaza in succesiunea semnnlelor pana la identificarea primului cuvant de cod si apoi lasand la o parte succesiunea identificata se trece la identificarea urmatoarei succesiuni si a.m.d. Momentul cand se obtine o succesiune cu sens, arata si sfarsitul unui cuvant de cod, dat fiind faptul ca nici un cuvant de cod nu este prefixul unui alt cuvant de cod.Relatia care exista intre codurile cu proprietate de prefix si codurile unic decodabile se pune in evidenta prin urmatoarea teorema:orice cod cu conditia de prefix este un cod unic decodabil.

Codurile cu proprietate de prefix permit o reprezentare geometrica foarte intuitiva si convenabila,cu ajutorul unui graf arborescent.Fiecarui cuvant de cod ii corespunde un drum simplu,sau nodul final de la extremitatea drumului simplu corespunzator

Este important ca in cazul unui cod cu conditia de prefix exista o corespondenta biunivoca intre cuvintele de cod si intre nodurile finale, deci fiecarui cuvant de cod ii corespunde un nod final si nu un nod internediar si invers, fiecarui nod final ii corespunde o combinatie de cod.

De asenenea si codurile care nu au proprietatea de prefix pot sa fie reprezentate grafic cu ajutorul unui graf arborescent, dar in acest caz cel putin unui cuvant de cod ii corespunde un nod internediar.

Se observa ca in cazul unui cod D-nar,constructia grafului arborescent corespunzator este asemanatoare, tinand cont ca din fiecare nod pomesc D arce.

1.3. Codarea si decodarea pe canale fara perturbatii.

In cadrul modulului 1 s-a aratat ca un sistem digital de comunicatie presupune un codor/decodor al sursei. Rolul acestuia este de a mari eficienta transmiterii prin utilizarea unor mesaje cât mai scurte pentru a transmite aceiasi cantitate de informatie. Aceastâ operatie numita generic „ compresie de date ”

Definirea unui cod. Fie o sursa discreta fara memorie având alfabetul

$$S=\{s_1,s_2,......,s_N\} \tag{2.3}$$

cu probabilitatea de aparitie $p(s_i)=p_i$

$$P=\{p_1,p_2,......,p_N\} \tag{2.4}$$

Fie alfabetul canalului

$$X=\{x_1,x_2,......,x_q\} \tag{2.5}$$

constituit dintr-un numar finit de semne (litere, caractere)

Se considera reuniunea secventelor finite de litere din alfabetul canalului:

$$X^*=\bigcup_{n\geq} X^n \tag{2.6}$$

Orice aplicatie $S\rightarrow X^*$ se numeste codarea alfabetului *S* prin alfabetul *X*

Un element $s_i^*\in X^*$ si care corespunde lui s_i este un cuvânt de cod. Lungimea cuvântului de cod, notatâ $n\left(s_i^*\right)=n_i$ este numarul de litere cale Il formeaza. Totalitatea cuvintelor de cod constitue codul lui *S* cu mentiunea ca X^* poate contine si combinatii care nu apartin codului, numite cuvinte fara sens.

Astfel, un text constituit din secvente de mesaje:

$$m_j=s_{i1},s_{i2},......,s_{ik} \tag{2.7}$$

este codat prin secventele de cuvinte de cod (cu sens)

$$m_j=s_{i1}^*,s_{i2}^*,......,s_{ik}^* \tag{2.8}$$

Decodarea implica posibilitatea de a repara cuvintele de cod în mod unic (aplicatia $S \to X^*$ sa fie injectiva). Un cod cu aceasta probabilitate se numeste regulat (nesingular).

Regularitatea este o conditie necesara dar nu suficienta pentru decodare. fie de exemplu s_1^*, $s_2^* = 10$ si $s_3 = 01$. Codul 010 poate fi interpretat fie s_1^*, s_2^* fie s_3^*, s_1^*.

Pentru a distinge fara ambiguitati un text trebuie ca fiecarui succesiune de cuvinte sa-i corespunda o succesiune unica de litere, adica aplicatia $\coprod_{k \geq 1}^{n} S^k \to X^*$ sa fie si ea injectiva.

Un cod de acest tip este un cod unic decodabil. Conditii suficiente care sa asigure aceasta proprietate sunt:

(a) utilizarea cuvintelor de cod de aceiasi lungime (bloc)

(b) utilizarea unui semn distinct Intre cuvintel (separator)

Exista însa si coduri care nu necesita utilizarea unui mijloc suplimentar pentru a asigura proprietate de unic decodabil. Aceste coduri se numesc separabile.

Alcatuirea unui cod. Teorema Kraft

Conditia necesara si suficienta pentru existenta unui cod ireductibil de N cuvinte de lungime $n_1, n_2, ..., n_N$ este ca

$$\sum_{i=1}^{N} q^{-n_i} \leq 1 \qquad (2.9)$$

Observatii

1. Se reaminteste ca $q = card(X)$ este numarul de litere din alfabetul canalului.

2. Daca numarul cuvintelor de lungime *k* este r_k atunci conditia (1.9) devine

$$\sum_{k=1}^{n} r_k \cdot q^{-k} \leq 1 \qquad (2.10)$$

cu $N = \sum_{k=1}^{n} r_k$

Teorema Mac Millan

Un cod este ireductibil daca si numai daca

$$n_1 \leq n_2 \leq ...... \leq n_N \qquad (2.11)$$

Criterii de apreciere a unui cod.

Transmiterea mesajelor presupune un cost care creste linear cu timpul. Un criteriu convenabil de apreciere a unui cod este lungimea medie a unui cuvânt

$$\bar{n} = \sum_{i=1}^{n} p_i \cdot n_i \tag{2.12}$$

unde p_i sunt definite prin (1.4) si n_i este numarul de litere din cuvântul de cod cu indicele *i*.

Este evident ca se cauta ca $\bar{n}$ sa fie cât mai mic dar trebuie avut In vedere ca el este limitat inferior de entropia informationala pe simbol a alfabetului de cod

$$\bar{n} \geq \frac{H}{\log_2 q} \tag{2.13}$$

unde H este entropia sursei. In aceste conditii, eficienta unui cod este

$$\eta = \frac{H}{\bar{n} \cdot \log_2 q} \leq 1 \tag{2.14}$$

iar redundanta codului este

$$\rho = 1 - \eta \tag{2.15}$$

Codurile cu o eficienta egala cu unitatea si deci care au lungimea medie minima se numesc coduri absolut optimale

Prima teorema a lui Shannon. Pentru orice sursa omogena exista un cod ireductibil pentru care lungimea medie a cuvintelor este orcât de apropiata de marginea sa inferioara.

Aceasta teorema se mai numeste si teorema codarii pe canale neperturbate. Prima teorema a lui Shannon se refera la problema codarii cu o lungime medie posibila cât mai mica, pe un canal neperturbat de capacitate data.

Codurile care asigura cea mai mica lungime medie posibila se numesc cvasioptimale sau compacte.

1.4. Coduri Huffman de dispersie minimă.

Acest procedeu se bazează pe ideea de a partiţiona mulţimea mesajelor sursei $S = \{s1, s2, ..., sN\}$ în submulţimile S_0 şi S_1 , astfel încât suma probabilităţilor mesajelor incluse în S_0 să fie cât mai apropiată de suma probabilităţilor mesajelor incluse în S_1.

La rândul lor, submulţimile S_0 şi S_1 pot fi partiţionate în submulţimile S_{00} şi S_{01}, respectiv S_{10} şi S_{11} astfel încât suma probabilităţilor mesajelor incluse în cele patru submulţimi să fie cât mai apropiate posibil. Procedeul se continuă în mod similar până când se obţin submulţimi ce conţin un singur mesaj.

În felul acesta, pentru orice distribuţie a sursei S ce urmează a fi codată se va obţine un cod compact, adică lungimi medii ale cuvintelor de cod ce nu mai pot fi micşorate prin nici un alt procedeu de codare.

Pentru ca partiţiile să satisfacă condiţiile menţionate, se procedează astfel:

1). Se ordonează mulţimea mesajelor sursei S în ordinea descrescătoare a probabilităţilor, obţinându-se astfel mulţimea ordonată R_0=$\{s_1, s_2, \dots, s_N\}$, cu p(s_1)$\geq$ p(s_2)$\geq$... p(s_N), cu schimbarea eventuală a indicilor mesajelor pentru realizarea ordonării respective;

2). Se reunesc ultimele două mesaje (de probabilităţile cele mai mici) într-un nou mesaj, notat cu r_1, căruia i se alocă o probabilitate egală cu suma probabilităţilor mesajelor componente. Se ordonează din nou mesajele în ordinea descrescătoare a probabilităţilor, formându-se astfel prima sursă restrânsă R_1=$\{s_1, s_2, \dots, r_1, \dots\}$ cu p(s_1)$\geq$ p(s_2)$\geq$... p(r_1)$\geq$... .

3). Se reunesc ultimele două mesaje din sursa restrânsă R_1 într-un nou mesaj r_2, de probabilitate egală cu suma probabilităţilor mesajelor componente. Se ordonează mesajele în ordine descrescătoare, formându-se astfel sursa restrânsă R_2. În mod analog, din R_2 se formează sursa restrânsă R_3 şi aşa mai departe, până când se obţine o sursă restrânsă formată numai din două mesaje, $R_n = \{r_n, r_{n-1}\}$, cu p(r_n) $\geq p(r_{n-1})$. De fapt, r_n va fi S_0 şi r_{n-1} va fi S_1 sau invers.

Din modul de formare a surselor restrânse R_i, rezultă că mulţimea *S* a mesajelor poate fi partiţionată în două submulţimi r_n şi r_{n-1} astfel încât probabilităţile p(r_n) şi $p(r_{n-1})$ sunt cele mai apropiate posibil. La rândul lor, submulţimile r_n şi r_{n-1} , pot fi partiţionate în alte două submulţimi, de probabilităţile cele mai apropiate posibil. Partiţionările se continuă până se obţin submulţimi care conţin un singur mesaj.

4). Cuvintele de cod corespunzătoare fiecărui mesaj se obţin astfel:

- submulţimii r_n i se alocă simbolul "0" (sau "1");
- submulţimii r_{n-1} , i se alocă simbolul "1" (sau "0");
- la fiecare partiţionare se alocă arbitrar celor două submulţimi "0" sau "1", operaţia continuându-se până se obţin submulţimi ce conţin un singur mesaj s_k, k=$\{1 \dots N\}$.

Deoarece alocarea lui "0" şi "1" este arbitrară la fiecare partiţionare, rezultă că unei surse *S* i se pot ataşa o multitudine de coduri instantanee, toate, însă, având aceeaşi lungime medie a cuvintelor de cod, care nu mai poate fi micşorată prin nici un alt procedeu de codare a mesajelor luate individual.

Dacă sursa primară *S* poate furniza *N* mesaje, atunci submulţimea restrânsă R_1 , va avea *N*-1 mesaje, submulţimea restrânsă R_2 va conţine *N*-2 mesaje şi aşa mai departe, ultima submulţime restrânsă R_n va conţine $N - n$ mesaje, care sunt r_nşi r_{n-1} , adică se poate scrie:

$$N-n=2 \Rightarrow n=N-2 \tag{2.16}$$

Dacă submulţimii r_n i se alocă simbolul "0" şi submulţimii r_{n-1} simbolul "1", celor *N*-2 partiţionări putându-li-se aloca arbitrar "0" sau "1", rezultă un total de 2^{N-2} posibilităţi de codare. Dacă, însă, submulţimii r_n i se alocă simbolul "1", iar submulţimii r_{n-1} simbolul "0", mai rezultă 2^{N-2} posibilităţi de codare. Rezultă, deci, că prin acest procedeu de codare se pot realiza $2^{N-2} + 2^{N-2} = 2^{N-1}$ coduri instantanee, toate având toate aceeaşi lungime medie a cuvintelor de cod.

Prin definiţie, se numeşte cod compact, codul care realizează lungimea medie minimă a cuvintelor de cod. Deoarece prin procedeul de codare Huffman se obţine cea mai mică lungime medie a cuvintelor de cod, înseamnă că prin acest procedeu se obţin coduri instantanee compacte. Evident, un cod absolut optimal este şi compact, reciproca nefiind totdeauna valabilă.

Exemplul 3.1.

Se presupune sursa discretă de informaţie caracterizată de distribuţia:

$$S:\begin{pmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 \\ 0{,}1 & 0{,}2 & 0{,}3 & 0{,}15 & 0{,}05 & 0{,}2 \end{pmatrix}$$

Codarea binară Huffman a acestei surse se poate realiza astfel:

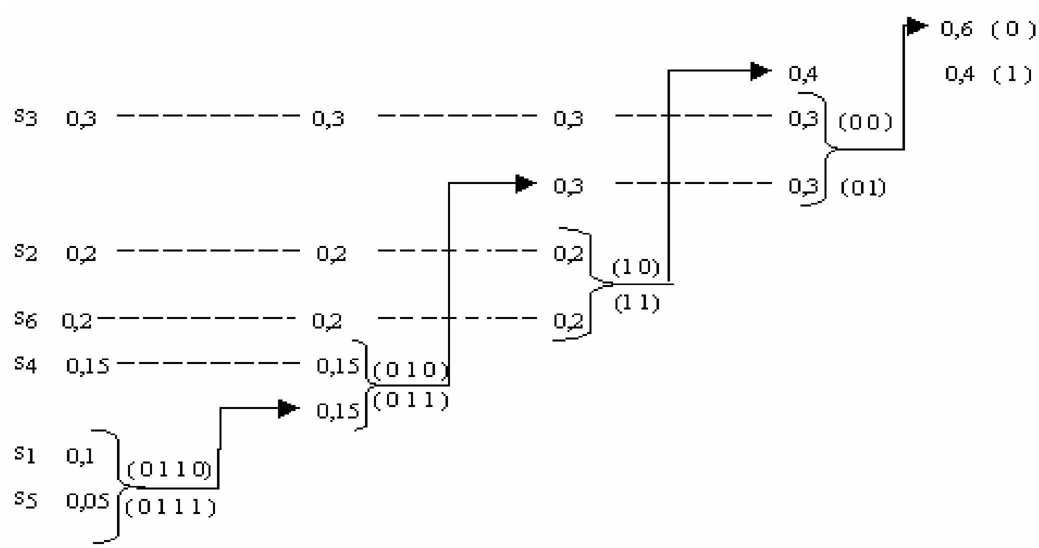


Fig. 3.2. Schema de codare pentru exemplul 3.1

Graful şi cuvintele de cod corespunzătoare codării efectuate sunt date în Fig. 3.3.

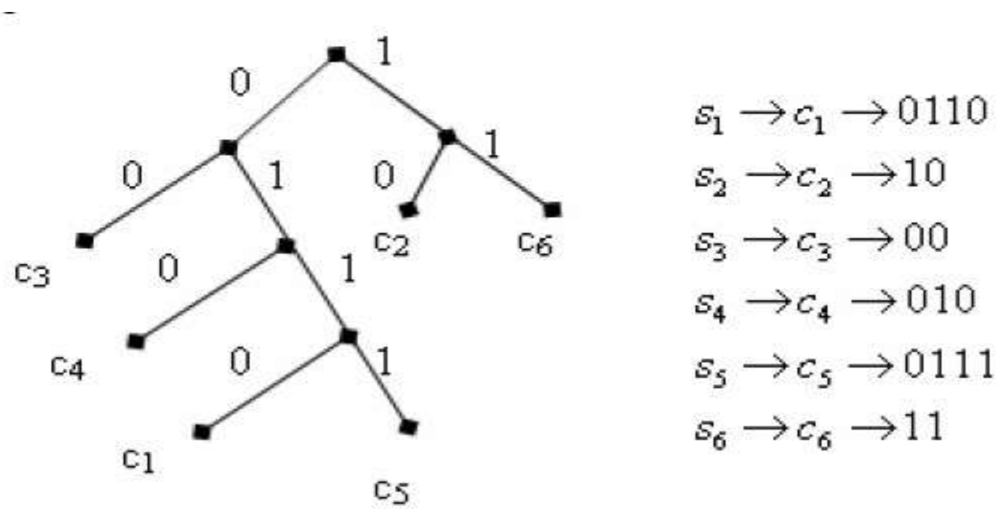


Fig.3.3. Graful corespunzător codului

Codurile Huffman de dispersie minimă se obţin când la reordonarea sursei restrânse, simbolul compus se plasează pe poziţia cea mai de sus posibil în sursa restrânsă. În felul acesta cuvântul de cod atribuit simbolului compus va avea cea mai mică lungime posibilă. Cum acest cuvânt va deveni prefix pentru simbolurile constituente, cuvintele de cod corespunzătoare acestora vor avea o lungime cu o unitate mai mare decât lungimea prefixului, deci şi acestea vor rezulta de lungime minimă. Ca urmare, diferenţele dintre lungimile cuvintelor de cod devin minime, ceea ce va conduce, evident, şi la o dispersie minimă.

Pentru fixarea ideilor, se presupune sursa discretă de informaţie caracterizată de distribuţia:

$$S:\begin{pmatrix} s_1 & s_2 & s_3 & s_4 & s_5 \\ 0{,}2 & 0{,}4 & 0{,}2 & 0{,}1 & 0{,}1 \end{pmatrix}$$

Pentru această sursă se efectuează codarea Huffman, plasând întâi mesajele sursei restrânse pe poziţiile cele mai jos posibile în listă şi apoi pe poziţiile cele mai de sus posibile. În primul caz rezultă schema de codare din Fig. 3.4, iar graful şi cuvintele de cod ca în Fig. 3.5.

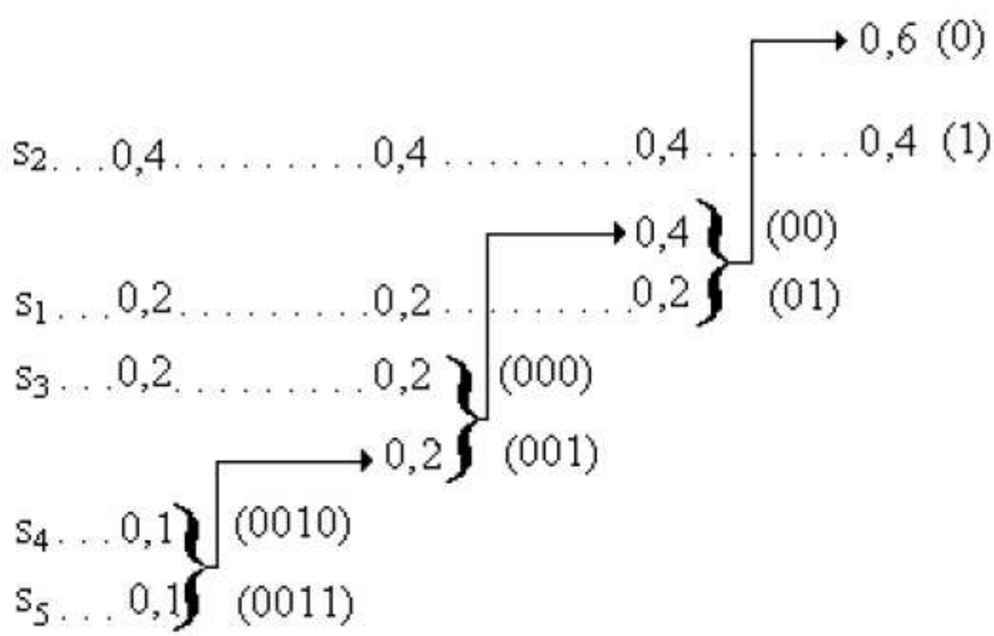


Fig. 3.4. Schema de codare

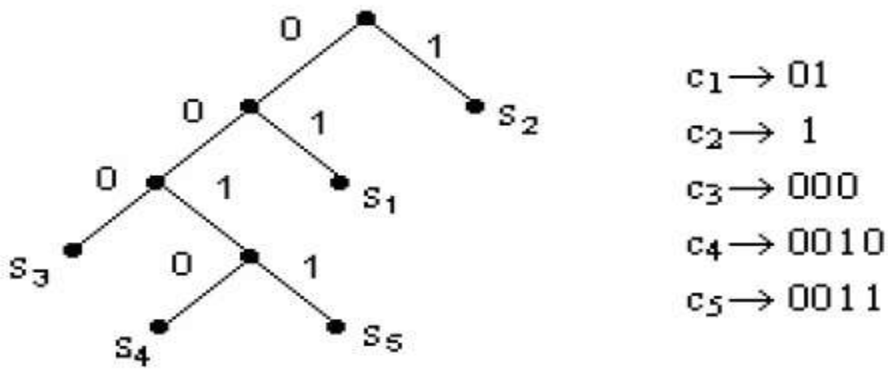


Fig. 3.5. Graful şi cuvintele de cod

Pentru acest cod, lungimea medie şi dispersia sunt:

$$\bar{l} = 0{,}2 * 2 + 0{,}4 * 1 + 0{,}2 * 3 + 0{,}1 * 3 + 0{,}1 * 4\ 2{,}2\ biți/mesaj$$

$$\sigma_1^2 = \sum_{i=1}^{5} (l_i - \bar{l})^2 = 0{,}2^2 + 1{,}2^2 + 0{,}8^2 + 1{,}8^2 + 1{,}8^2 = 8{,}6$$

Pentru cazul în care în codarea Huffman mesajele sursei restrânse se plasează pe poziţiile cele mai de sus în listă, se obţine schema de codare din Fig. 3.6 şi graful şi cuvintele de cod ca în Fig. 3.7.

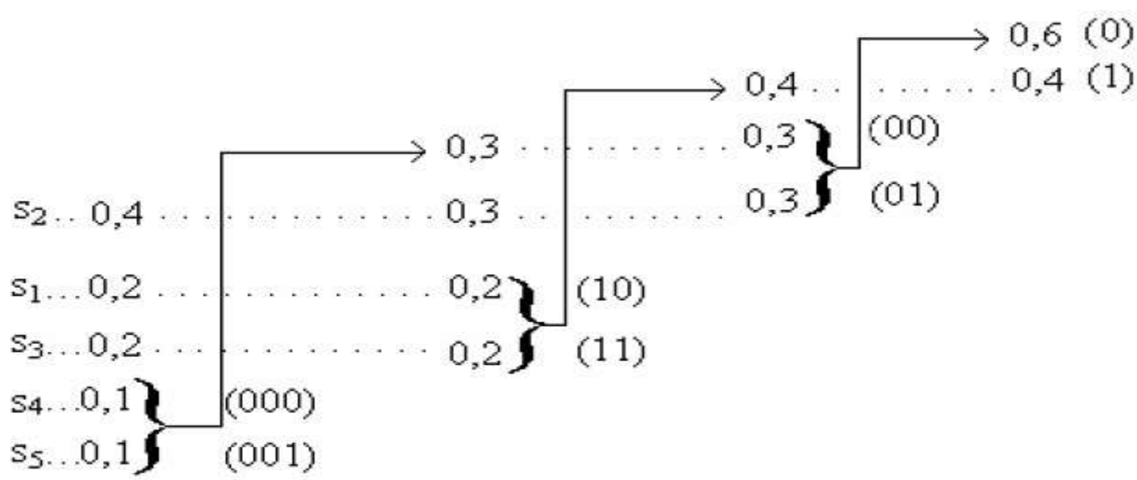


Fig. 3.6. Schema de codare

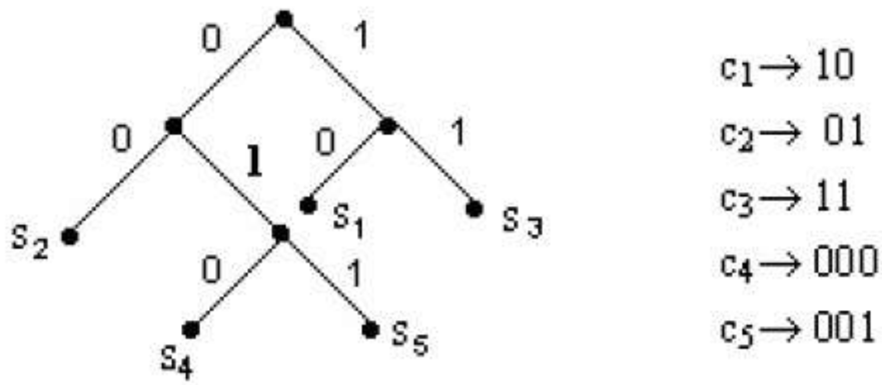


Fig. 3.7. Graful şi cuvintele de cod

Pentru acest cod, lungimea medie este, evident, aceeaşi, în timp ce dispersia devine:

$$\sigma_2^2 = \sum_{i=1}^{5} (l_i - \bar{l})^2 = 0{,}2^2 + 0{,}2^2 + 0{,}2^2 + 0{,}8^2 + 0{,}8^2 = 1{,}4$$

Deşi din punct de vedere informaţional, cele două coduri sunt identice, în practică se preferă folosirea celor de dispersie minimă, din motive de transmisie. De exemplu, dacă se doreşte să se transmită mesaje ale sursei cu o viteză de 10.000 mesaje/sec., este necesar un canal cu capacitatea de 22.000 biţi/sec. Deoarece viteza de generare a biţilor oscilează în jurul valorii de 22.000 biţi/sec., funcţie de succesiunea de mesaje furnizate la un moment dat, ieşirea sursei este încărcată într-un buffer.

Dacă, de exemplu, sursa generează la un moment dat şiruri de mesaje S_4 şi S_5 mai multe secunde, pentru primul cod se generează 40.000 biţi/sec. şi în fiecare secundă ar trebui un buffer de capacitate de 18.000 biţi. Cu al doilea cod se generează 30.000 biţi /sec. şi bufferul ar trebui să aibă capacitatea de 8.000 biţi. Dacă se transmit şiruri de mesaje s_2, cu primul cod se generează 10.000 biţi/sec. şi canalul nu e folosit la capacitatea sa, rămânând un deficit de 12.000 biţi/sec, pe când cu al doilea cod se generează 20.000biţi/sec, deficitul existent în exploatarea canalului fiind numai de 2.000 biţi/sec. Aşadar, din motive de transmisie este mai rezonabil a se alege al doilea cod decât primul.

1.5. Procedeul de codare binară Shannon – Fano.

Acest procedeu se aplică de obicei în cazurile particulare în care probabilităţile de furnizare ale mesajelor sunt puteri întregi pozitive ale lui (1/2), adică, de forma:

$$p(S_k) = \left(\frac{1}{2}\right)^{l_k} = 2^{-l_k}, \quad (\forall)\ k=\overline{1, N} \qquad (2.17)$$

unde l_k este un număr întreg pozitiv. Dacă relaţia (3.48) este satisfăcută, mulţimea *S* = {*s*1, *s*2 ,..., *sN*} a mesajelor sursei discrete de informaţie ce urmează a fi codată poate fi partiţionată în două submulţimi S_0 şi S_1, astfel încât suma probabilităţilor mesajelor incluse în S_0 , notată cu p(S_0), să fie egală cu suma probabilităţilor mesajelor incluse în S_1,, notată cu p(S_1). Sursa *S* fiind totdeauna completă, se poate scrie:

$$\left.\begin{array}{l} p(S_0) = p(S_1) \\ p(S_0) + p(S_1) = 1 \end{array}\right\} \rightarrow p(S_0) = p(S_1) = \frac{1}{2} = 2^{-1} \qquad (2.18)$$

Submulţimile S_0 şi S_1 se pot partiţiona la rândul lor în S_{00} şi S_{01}, respectiv în S_{10} şi S_{11}, astfel încât suma probabilităţilor mesajelor incluse în cele patru submulţimi să fie aceeaşi, adică se poate scrie relaţia:

$$p(S_{00}) = p(S_{01}) = p(S_{10}) = p(S_{11}) = \frac{1}{2}^2 = 2^{-2} \qquad (2.19)$$

Se procedează în mod analog până se obţin submulţimi care conţin un singur mesaj. Se observă că fiecare submulţime are suma probabilităţilor mesajelor incluse egală cu o putere întreagă a lui (1/2). Puterea întreagă este egală cu numărul indicilor submulţimii respective Dacă submulţimea conţine un singur mesaj, s_k, şi are un număr de indici egal cu l_k, atunci se poate scrie:

$$p(s_k) = \left(\frac{1}{2}\right)^{l_k} = 2^{-l_k} \qquad (3.51)$$

de unde rezultă necesitatea ca sursa *S* ce urmează a fi codată să-şi furnizeze mesajele cu probabilităţi egale cu ½ la o putere întreagă, pozitivă.

Sursa fiind completă, se poate scrie relaţia (1.21):

$$\sum_{k=1}^{N} p(s_k) = 1 \tag{2.21}$$

Înlocuind (1.20) în (1.21), rezultă relaţia (1.22):

$$\sum_{k=1}^{N} 2^{-l_k} = 1 \tag{2.22}$$

ceea ce înseamnă că inegalitatea lui Kraft devine în acest caz egalitate.

Cuvintele de cod se vor obţine, atunci, astfel:

1. Se atribuie simbolul "0" submulţimii S_0 şi simbolul "1" submulţimii S_1, (sau invers), astfel că toate cuvintele corespunzătoare mesajelor incluse în S_0 vor începe cu "0" şi toate cuvintele corespunzătoare mesajelor incluse în S_1, vor
începe cu "1" (sau invers);

2. Se alocă submulţimilor S_{00} şi S_{10} ca al doilea mesaj "0", iar submulţimilor S_{01} şi S_{11} ca al doilea mesaj "1" (sau invers). În felul acesta, cuvintele de cod corespunzătoare mesajelor incluse în S_{00} vor începe cu 00, cuvintele de cod corespunzătoare mesajelor incluse în S_{10} vor începe cu 10 şi aşa mai departe, cuvintele de cod corespunzătoare mesajelor induse în S_{11} vor începe cu 11.

3. Operaţia se continuă în acelaşi mod, până când în fiecare submulţime rămâne un singur mesaj, căruia îi va corespunde cuvântul de cod format din şirul de indici ai submulţimii respective. Deoarece la fiecare partiţionare în două submulţimi atribuirea mesajelor "0" şi "1" este arbitrară, rezultă că prin acest procedeu se pot obţine o multitudine de coduri instantanee, dar toate absolut optimale.

În principiu, procedeul de codare descris s-ar putea aplica în general, adică şi atunci când relaţia (1.21) nu este satisfăcută. În acest caz, partiţionările în submulţimi trebuie efectuate astfel încât suma probabilităţilor mesajelor incluse în submulţimile respective să fie cât mai apropiate. Atribuind simbolurile "0" şi "1" ca în procedeul descris, se obţin totdeauna coduri instantanee.

Cu cât sumele probabilităţilor mesajelor componente ale submulţimilor respective vor fi mai apropiate, cu atât lungimea medie a cuvintelor de cod va fi mai mică.

Exemplul 3.2.

Se consideră sursa discretă de informaţie caracterizată de distribuţia:

$$S : \begin{pmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 \\ 2^{-2} & 2^{-2} & 2^{-3} & 2^{-3} & 2^{-4} & 2^{-4} & 2^{-4} & 2^{-4} \end{pmatrix}$$

Procedeul de codare binară Shannon - Fano este sintetizat în tabelul de mai jos:

Mesaje	**Probabilități**	**Partiții**				**Cuvânt de cod**
s_1	2^{-2}	0	0			0 0
s_2	2^{-2}		1			0 1
s_3	2^{-3}	1	0	0		1 0 0
s_4	2^{-3}			1		1 0 1
s_5	2^{-4}		1	0	0	1 1 0 0
s_6	2^{-4}				1	1 1 0 1
s_7	2^{-4}			1	0	1 1 1 0
s_8	2^{-4}				1	1 1 1 1

Graful arborescent ataşat codului astfel obţinut este reprezentat în Fig. 3.8

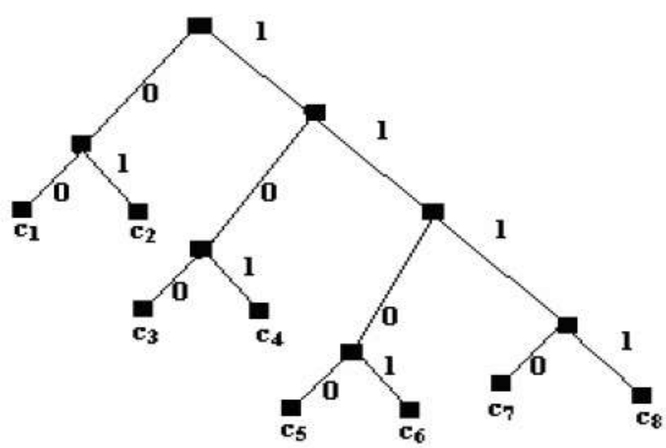
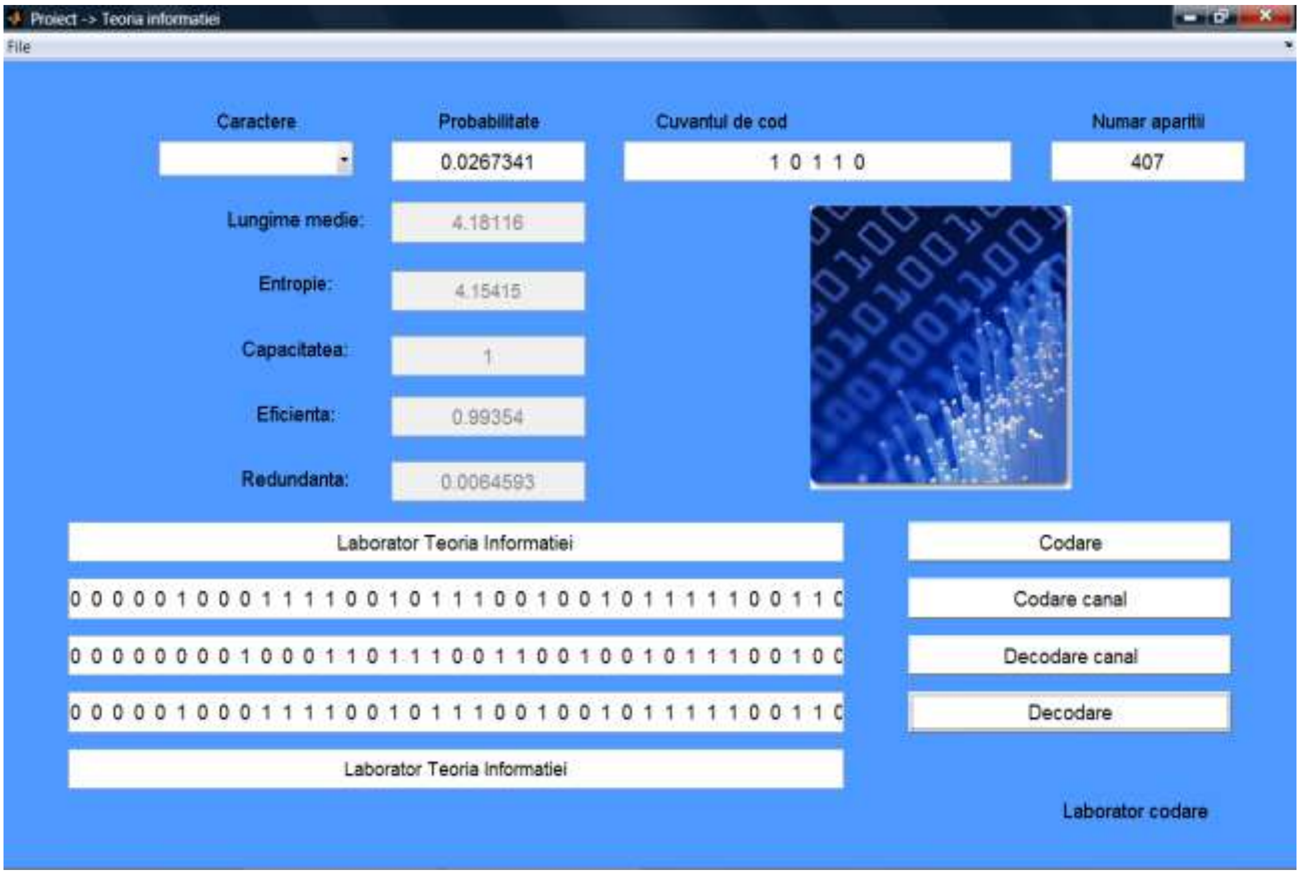


Fig. 3.8. Graful arborescent ataşat codului din tabel

Aplicaţie

Simularea cu ajutorul programului MATLAB a algoritmului de compresie Shannon-Fano

În cadrul lucrărilor de laborator va fi pusă la dispoziţie o aplicaţie software care implemenează algoritmul de compresie Shannon-Fano. Ecranul aplicaţiei este prezentat în figura următoare.



<u>Exerciţii rezolvate.</u>

Exemplu

În tabelul de mai jos sunt prezentate patru coduri separabile,

mesaj	A	B	C	D
s_0	00	0	0	0
s_1	01	10	01	10
s_2	10	110	011	110
s_3	11	1110	0111	111

Folosind codul B, succesiunile $s_3 s_1 s_0 s_2$ se codifica 1110100110. Dupa receptinarea primelor sase biti se poate determina ca s-a receptionat $s_3 s_1$. Daca Insa se folosec codul C, succesiunea $s_3 s_1 s_0 s_2$ se codifica 011010011. Dupa receptionarea primelor sase biti conduce la decodarea s_3 dar secventa 01 poate fi interpretata la acel moment fie care s_1 fie ca, s_2 fie ca s_3, ambiguitatea rezolvându-se abia dupa receptia urmatorilor biti. Un cod de tip C se numeste cod instantaneu.

Conditia necesara si suficienta ca un cod sa fie instantaneu este ca nici un cuvânt de cod sa nu fie prefix al altui cuvânt de cod (conditia de prefix).

Se considera sursa care genereaza simbolurile :

- s0 cu probabilitatea p1 = 0,5
- s1 cu probabilitatea p2 = 0,25
- s2 cu probabilitatea p3 = 0,125
- s3 cu probabilitatea p4 = 0,125

Se cere sa se determine eficienta codurilor A, B, C si D

Solutie

Entropia sursei este

$$H = -\sum_{i=1}^{4} p_i \cdot \log_2 p_i = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - 2\cdot\frac{1}{8}\log\frac{1}{8} = \frac{7}{4} \text{ biti}$$

Pentru codul A lungimea medie a codului este $\bar{n}_A = 2$ si $\eta_A = \dfrac{\frac{7}{4}}{2\log_2 2} = \dfrac{7}{8}$ $\quad iar \rho = \dfrac{1}{2}$

Codurile B si C cu aceiasi lungime medie

$$\bar{n}_B = \bar{n}_C = 0{,}5\cdot 1 + 0{,}25\cdot 2 + 0{,}125\cdot 3 + 0{,}125\cdot 4 = 1{,}875$$

$$\eta_B = \eta_C = \frac{1{,}75}{1{,}875} = \frac{14}{15}$$

$$\rho_B = \rho_C = \frac{1}{15}$$

Codul D are $\bar{n}_D = 0{,}5\cdot 1 + 0{,}25\cdot 2 + 0{,}125\cdot 3 = 1{,}75$ si

$$\eta_D = \frac{1{,}75}{1{,}75\log_2 2} = 1 \qquad \rho_D = 0$$

TESTE DE AUTOEVALUARE ŞI TEME DE CONTROL

Testul nr. 1

1. Se dă sursa A prin urmatoarea repartiţie de probabilităţi:

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 \\ 0.48 & 0.14 & 0.14 & 0.07 & 0.07 & 0.04 & 0.02 & 0.02 & 0.02 \end{pmatrix}$$

Să se codifice sursa A utilizând metoda de codificare Huffman şi să se calculeze lungimea medie a cuvintelor de cod.

Testul nr. 2

2. Se dă sursa A prin urmatoarea repartiţie de probabilităţi:

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 \\ 0.48 & 0.14 & 0.14 & 0.07 & 0.07 & 0.04 & 0.02 & 0.02 & 0.02 \end{pmatrix}$$

Să se codifice sursa A utilizând metoda de codificare Shannon-Fano şi să se calculeze lungimea medie a cuvintelor de cod.

Temă de control

Se consideră o sursă cu alfabetul $[S] = [s_1, s_2, s_3, s_4, s_5, s_6]$ şi probabilităţile $[P] = [0.05, 0.1, 0.3, 0.25, 0.1, 0.2]$:

a. să se determine un cod compact folosind algoritmul de codare Huffman, dacă alfabetul codului este $[X] = [0,1]$ şi dacă alfabetul codului este $[X] = [0,1,2]$;
b. pentru cele două cazuri să se calculeze lungimea medie a cuvintelor de cod şi eficienţa codului.

BIBLIOGRAFIE RECOMANDATĂ LA MODULUL 2:

[1] A. Spătaru: Teoria Transmisiunii Informaţiei, Ed. Didactică şi Pedagogică, Bu- cureşti, 1983.

[2] A. T. Murgan, I. Spânu, I. Gavăt, I. Sztojanov, V. E. Neagoe, A. Vlad: Teoria Transmisiunii Informaţiei - probleme, Ed. Didactică şi Pedagogică, Bucureşti, 1983

[3] I. Angheloiu, Teoria codurilor, Ed. Militară, Bucureşti, 1972.

[4] J.C. Moreira, P.G. Farrell, ESSENTIALS OF ERROR-CONTROL CODING, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England,2006.
[5] V. Munteanu, Transmiterea şi codificarea informaţiei, Note de curs.