T E O R I J A I N F O R M A C I J A I K O D O V A

1. Nastanak teorije informacija i kodova

Same komunikacije su stare koliko i čovječanstvo, ali u ovom izlaganju biće dovoljno da se krene od početka dvadesetog vijeka. U tom periodu alfanumerički simboli i ljudski glas bili su prenošeni putem električnih signala. Odgovarajuća teorija je bila deterministička po prirodi. U okviru ovoga pristupa signali su predstavljani kao sinusoide ili zbirovi sinusoida, tj. kao periodični signali, dok su aperiodični signali tretirani pomoću Furijeove transformacije. Rezultati dobijeni na taj način korišćeni su za projektovanje klasičnih analognih telekomunikacionih sistema (potreban opseg, snaga, itd.). Ovi sistemi su radili dobro u "normalnim" uslovima. Međutim, u "nepovoljnim" uslovima radili su loše ili nijesu uopšte radili. Nepovoljni uslovi se srijeću i za vrijeme ratova, kada je potrebno obezbijediti tajnost prenešenih poruka.

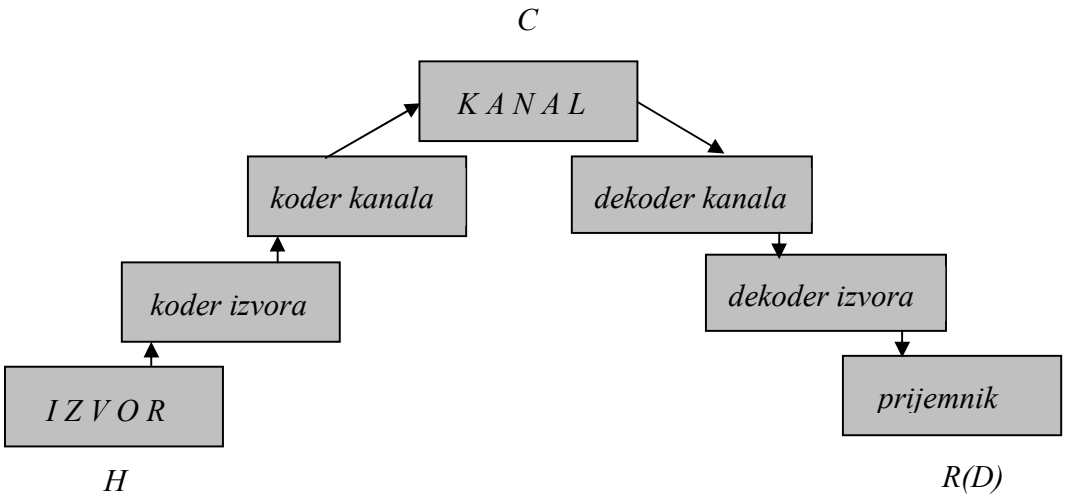
Teorija informacija je "rođena" u okviru telekomunikacija i prvenstveno za primjenu u samim telekomunikacijama. Kasnije je našla primjenu i u drugim oblastima kao što su biologija, genetika, ekonomija, itd.

Polovinom dvadesetog vijeka razvijena je odgovarajuća teorija - Teorija informacija. Sve fundamentalne rezultate Teorije informacija dao je praktično jedan čovjek Klod Šenon. Njegov osnovni članak "Matematička teorija komunikacija" je kamen temeljac današnje Teorije informacija. U sledećih desetak godina, sa još nekoliko članaka, Šenon je praktično zaokružio osnove svega onoga što danas proučava u Teoriji informacija. On je teoriju vjerovatnoće koristio kao osnovni matematički aparat. Šenonova teorija se pojavila praktično na početku informacione ere, pa su zbog toga mnoga njena rješenja morala čekati odgovarajući razvoj tehnologije da bi bila primijenjena u praksi. Iako je Teorija informacija stvorena praktično samim Šenonom, treba pomenuti i doprinose Nikvista i Hartlija. Nikvist je odredio minimalan potreban opseg učestanosti za prenos nezavisnih diskretnih signala datom brzinom, dok je Hartli predložio korišćenje

logaritamske mjere za količinu informacija. Šenon je u svojoj teoriji razmatrao informacije, a ne signale. Informacije se predstavljaju (koduju) signalima-nosiocima informacija. Šenon je definisao količinu informacija koju emituje informacioni izvor i pokušao da predstavi te informacije što efikasnije signalima, tako da one ostanu neoštećene čak i ako se signali pri prenosu izobliče. On je takođe definisao granice do kojih se može ići u komuniciranju, imajući u vidu brzinu kojom izvor emituje informacije, kao i karakteristike kanala (frekvencijski opseg i odnos signal-šum).

2. Model komunikacionog sistema (sistem za prenos informacija)

Na slici 1. prikazana je blok-šema komunikacionog sistema s gledišta Teorije informacija.



Slika 1. Sistem za prenos informacija

Izvor informacija - bira poruke iz skupa mogućih poruka i saopštava ih (emituje ih).
Koder izvora - informacije koje emituje izvor ekonomično predstavlja (kodira) signalima.

Koder kanala(zaštitni koder) - pošto signali (nosioci informacija) prolaze kroz kanal, pri čemu se mogu izobličiti (usled karakteristika samog kanala ili usled dejstva smetnji u kanalu) ovaj koder dodaje izvjesnu redudansu (tj. signale koji ne nose informacije) u prenošeni niz, čime se na prijemu omogućava otkrivanje i eventualno ispravljanje grešaka koje su se pojavile pri prenosu.

Kanal - je medijum preko kojeg se prenose poruke. To su na primjer: radio veze, satelitske komunikacije, kompjuterske mreže, magnetski mediji za smještaj podataka itd.

Dekoder kanala - vrši detekciju (otkrivanje) i korekciju (ispravljanje) grašaka.

Dekoder izvora - vrši operacije koje su neophodne da se dobije poruka u obliku prihvatljivom za korisnika.

3. Izvori informacija

U okviru Teorije informacija smatra se da izvorima informacija stoji na raspolaganju skup poruka. Izvor odabrane poruke emituje sukcesivno, obično konstantnom brzinom, tako što iz datog skupa bira po jednu poruku i saopštava je. Ako izvoru stoji na raspolaganju *konačan (prebrojiv)* skup poruka kaže se da je to ***diskretan izvor.*** Ako je skup poruka *neprebrojiv*, izvor se naziva ***kontinualni izvor***.

Niz poruka da izlazu diskretnog izvora čini diskretni slučajni proces, a niz poruka na izlazu kontinualnog izvora čini kontinualni slučajni proces. Kod diskretnih izvora uobičajena je dalja podjela na ***izvore bez memorije i izvore sa memorijom***. Izvori bez memorije emituju poruke statistički nezavisno, dok kod izvora sa memorijom postoji zavisnost sukcesivno emitovanih poruka.

3.1. Definicija količine informacija

Postavljanje matematičkog modela sistema za prenos informacija zahtijeva uvođenje mjere za količinu informacija. Ovo je neophodno isto tako kao što se mora uvesti pojam količine elektriciteta da bi se kvantitativno proučavale električne pojave. Da bi najjednostavnije objasnili pojam količine informacija razmotrimo sledeći primjer: neka je

korisnik zainteresovan za poruku da se neki događaj desio (S_i) i neka mu je poznata odgovarajuća vjerovatnoća tog događaja $P(S_i)$. Ukoliko je vjerovatnoća ovog događaja manja, utoliko je veća neizvjesnost u kojoj se nalazi korisnik. Dobijanjem poruke o događaju S_i korisnik razrješava tu svoju neizvjesnost. Odnosno, ako je neizvjesnost veća može se smatrati da je korisnik prijemom poruke dobio veću količinu informacija. Na osnovu ovoga možemo zaključiti da je količina informacija koju nosi neka poruka obrnuto srazmjerna njenoj vjerovatnoći i da je količina informacija koju nosi siguran događaj ($P(S_i)=1$) jednaka nuli.

Veličina koja predstavlja mjeru neizvjesnosti jednog sistema naziva se ***entropija***. U Teoriji informacija ova *veličina predstavlja srednju količinu informacija koju izvor emituje po jednom simbolu.*

-Entropija kod diskretnih izvora bez memorije

Neka je diskretni izvor bez memorije potpuno definisan listom simbola (poruka): $S=\{S_1,S_2,...,S_n\}$ i skupom odgovarajućih vjerovatnoća emitovanja tih poruka $P(S_i)$ $(i=1,2....n)$. Pošto je izvor bez memorije, emitovanje simbola predstavlja skup nezavisnih događaja čiji je zbir vjerovatnoća jednak jedinici ($\sum_{i=1}^{n} P(S_i)=1$). Srednja količina informacija koju ovakav izvor emituje po jednom simbolu je:

$$H(S)=-\sum_{i=1}^{n} P(S_i)\log_a P(S_i)$$

Izbor baze logaritma je potpuno proizvoljan i on samo određuje jedinicu mjere:

- Ako je ***a=2*** mjerna jedinica entropije je ***bit*** *(Shannon).*
- Ako je ***a=10*** mjerna jedinica entropije je ***Hartley***.
- Ako je ***a=e=2,7281(eulerov broj)*** mjerna jedinica entropije je ***Nat***.

Primjer:

Neka izvor informacija emituje slučajni niz sa samo dva moguća simbola $S=\{S_1,S_2\}$ i neka su oba simbola jednako vjerovatna, tj. $P(S_1)=P(S_2)=0,5$, tada je entropija jednaka:

$$H(S)=-\sum_{i=1}^{2}P(S_i)\log_2 P(S_i)=-(P(S_1)\log_2 P(S_1)+P(S_2)\log_2 P(S_2))=$$

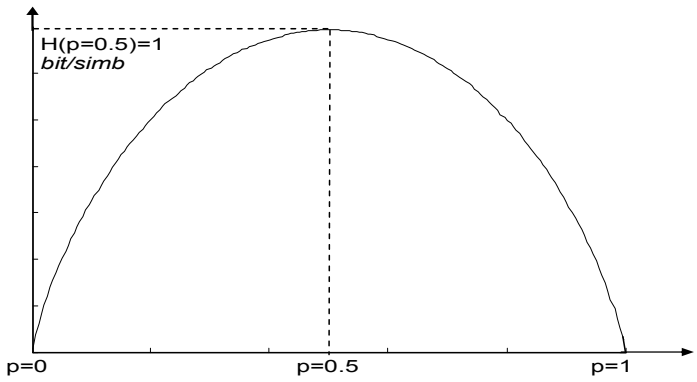
$$-\left(\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{2}\log_2\frac{1}{2}\right)=-\frac{1}{2}\cdot(-2)=1\ bit/simb$$

Primjer:*entropija binarnog izvora informacija*

Da bi jedan izvor mogao da emituje informacije mora imati na raspolaganju makar dva simbola. Takav izvor se naziva *binarni izvor informacija*. Uobičajeno je da se simboli ovakvog izvora obilježavaju sa 0 i 1 i nazivaju se biti ($S=\{0,1\}$).

Neka je vjerovatnoća emitovanja simbola 0 $P(0)=p$, tada je vjerovatnoća pojavljivanja simbola 1 jednaka $P(1)=1-p$.

Entropija je: $H(S)=-p\log_2 p-(1-p)\log_2(1-p)$



Ova entropija je maksimalna kada su vjerovatnoće emitovanja oba simbola jednake ($p=1-p=0,5$) i tada iznosi *1bit/simb*.

3.2. Osobine entropije

1. Entropija je sigurno veća ili jednaka nuli ($H(S) \geq 0$).

2. Ako jedan simbol ima vjerovatnoću jednaku jedinici, tada su vjerovatnoće svih ostalih simbola jednake nuli, pa je i entropija jednaka nuli.

$$S = \{S_1, S_2 .... S_i ... S_n\}$$

$$P(S) = \{0, 0 ... 1 ... 0\}$$

$$H(S) = 1 \cdot \log 1 = 0$$

Napomena: po konvenciji se usvaja da je $0 \cdot \log 0 = 0$

3. Entropija je ograničena sa gornje strane i ne može biti veća od logaritma broja simbola

$$0 \leq H(S) \leq \log N, \ N - broj\ simbola$$

Primjer: Binarni sistem ima dva simbola ($N = 2$) i

za njega važi $0 \leq H(S) \leq \log 2 = 1$

4. Maksimalna vrijednost entropije $(\log N)$ postiže se kada su svi simboli jednako vjerovatni.

Primjer: Skicirati entropiju ternarnog izvora informacija. Kada je ova entropija maksimalna i koliko iznosi?

$$S = \{S_1, S_2, S_3\}, \ P(S) = \{p, q, 1 - p - q\}$$

$$H(S) = -p \log_2 p - q \log_2 q - (1 - p - q) \log_2 (1 - p - q)$$

Nacrtati u Matlabu ovu funkciju. Određivanjem izvoda funkcije $H(S)$ po p i q i njihovim izjednačavanjem sa nulom dobija se maksimalna entropija $H(S) = 1,585 bita / simb$ u slučaju kada su vjerovatnoće $p = q = 1 - p - q = \frac{1}{3}$.

4. Proširenje diskretnog izvora

Neka je dat diskretni izvor S sa N simbola, čija je entropija $H(S)$. Ako se umjesto pojedinih simbola posmatraju sekvence od po *2, 3* ili više *(n)* sukcesivnih simbola, tada se kaže da se posmatra drugo, treće ili *n-to* proširenje izvora. Ovo proširenje izvora se obilježava sa S^n, a broj njegovih simbola je N^n. Drugim riječima, *n-to* proširenje izvora je izvor koji emituje sekvence od po *n* simbola prvobitnog izvora. Entropija proširenog izvora je data relacijom:

$$H(S^n) = n \cdot H(S)$$

Primjer:

a.) Neka je dat diskretni izvor bez memorije:

$$S_i : \; S_1\; S_2\; S_3 \quad N = 3 \; (brojsimbola)$$
$$P(S_i): \; \frac{1}{2}\; \frac{1}{4}\; \frac{1}{4}$$

Entropija ovog izvora je:

$$H(S) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{4}\log_2\frac{1}{4} =$$
$$H(S) = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 =$$
$$H(S) = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2} = 1.5 \; bit/simb$$

Napomena: $\log\frac{1}{a} = \log a^{-1} = -\log a, \quad \log_2 a = x \Rightarrow \; a = 2^x$

b.) Neka je izvršeno drugo *(n=2)* proširenje izvora iz prethodnog primjera (broj simbola je $N^n = 3^2 = 9$). Sada je izvor opisan sledećim vjerovatnoćama:

$$S_i, S_j : \; S_1,S_1 ; \; S_1,S_2 ; \; S_1,S_3 ; \; S_2,S_1 ; \; S_2,S_2 ; \; S_2,S_3 ; \; S_3,S_1 ; \; S_3,S_2 ; \; S_3,S_3;$$
$$P(S_i, S_j): \quad \frac{1}{4} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{16} \quad \frac{1}{16} \quad \frac{1}{8} \quad \frac{1}{16} \quad \frac{1}{16}$$

Napomena: Ako su S_i i S_j nezavisni događaji združena vjerovatnoća $P(S_i, S_j)$ se dobija množenjem vjerovatnoća događaja S_i i S_j , tj. $P(S_i, S_j) = P(S_i) \cdot P(S_j)$

Entropija izvora je:

$$H(S^2) = -\frac{1}{4}\log_2\frac{1}{4} - 4\cdot\frac{1}{8}\log_2\frac{1}{8} - 4\cdot\frac{1}{16}\log_2\frac{1}{16} = 3\ bit/simb$$

$$H(S^2) = 2\cdot H(S)$$

5. Diskretni izvori sa memorijom

Kod izvora sa memorijom pojavljivanje (emitovanje) jednog simbola zavisi od sekvence prethodno emitovanih simbola (tj. "stanja" izvora). Ako je dužina ove sekvence konačna i iznosi ***m***, tada se govori o ***izvoru s memorijom m-tog reda***. Neki autori ove izvore nazivaju ***Markovljevi izvori m-tog reda***. Izvor bez memorije je Markovljev izvor nultog reda. Markovljevi izvori m-tog reda se matematički opisuju listom simbola $S = \{S_1, S_2, ..., S_n\}$ i skupom uslovnih vjerovatnoća emitovanja jednog simbola, kada se zna prethodno emitovana sekvenca dužine ***m***, tj. :

$$P(S_j \,|\, S_{i1}, S_{i2}...S_{ik}...S_{im})\ (j = 1, 2....N)\ (k = 1, 2....m)$$

gdje je S_{i1} najstariji, a S_{im} najmlađi simbol poslije čega slijedi simbol S_j . Konkretna prethodno emitovana sekvenca dužine *m* naziva se *stanje izvora.* Ovih stanja ima n^m. Pošto se iz svakog stanja mora emitovati neki simbol, to za uslovne vjerovatnoće važi:

$$\sum_{j=1}^{n} P(S_j \,|\, S_{i1}, S_{i2}...S_{ik}...S_{im}) = 1$$

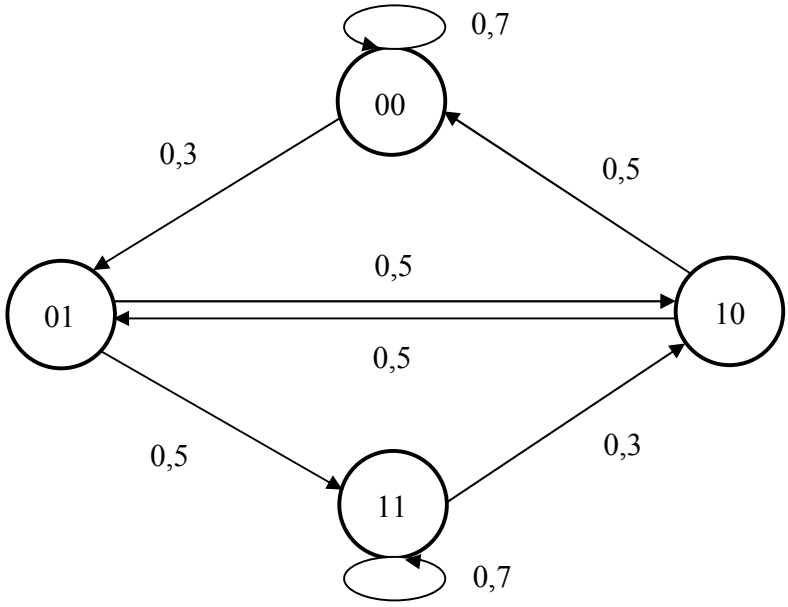
Pošto se iz svakog stanja može emitovati bilo koji od *n* simbola to je broj uslovnih vjerovatnoća jednak n^{m+1}. Kada broj stanja nije pretjerano velik, lakši uvid u prirodu procesa se može dobiti iz ***dijagrama stanja***. Na ovom dijagramu u krugove, koji predstavljaju stanja, upisuju se prethodne sekvence simbola, a uz grane grafa se daju odgovarajuće uslovne vjerovatnoće.

Primjer:

Neka je dat binarni *(n=2)* izvor drugog reda *(m=2)*. Izvor je okarakterisan sa $n^{m+1} = 2^3 = 8$ prelaznih uslovnih vjerovatnoća. Neka su ove vjerovatnoće poznate i date:

$$P(0|00) = P(1|11) = 0,7$$
$$P(1|00) = P(0|11) = 0,3$$
$$P(0|01) = P(0|10) = P(1|01) = P(1|10) = 0,5$$

Stanja su 00, 01, 10, 11.



Izvor ne može preći iz svakog stanja u svako stanje. Na primjer, ako je izvor bio u stanju 00 tada može preći u stanje 00 ako se poslije 00 pojavi 0 ($00 \cdot 0 \to 0 \cdot 00$) i u stanje 01 kada se poslije 00 pojavi 1 ($00 \cdot 1 \to 0 \cdot 01$).

Markovljev izvor je ***ergodičan*** ako se iz bilo kog stanja može preći u bilo koje stanje poslije dovoljno dugo vremena. Izvor iz prethodnog primjera je ergodičan, ali bi se mogao pretvoriti u neergodičan ako bi jedna od uslovnih vjerovatnoća bila na primjer $P(0|00) = 1$ ($P(1|00) = 0$). Tada bi sistem poslije izvjesnog vremena ušao u stanje 00 i tu ostao do kraja.

Za ergodične izvore mogu se izračunati ***stacionarne vjerovatnoće stanja***. U prethodnom primjeru te vjerovatnoće su:

$$P(00) = P(00)\cdot P(0|00) + P(10)\cdot P(0|10) = P(00)\cdot 0,7 + P(10)\cdot 0,5$$
$$P(01) = P(00)\cdot P(1|00) + P(10)\cdot P(0|10) = P(00)\cdot 0,3 + P(10)\cdot 0,5$$
$$P(10) = P(01)\cdot P(0|01) + P(11)\cdot P(0|11) = P(01)\cdot 0,5 + P(11)\cdot 0,3$$
$$P(11) = P(01)\cdot P(1|01) + P(11)\cdot P(1|11) = P(01)\cdot 0,5 + P(11)\cdot 0,7$$
$$P(00) + P(01) + P(10) + P(11) = 1$$

Rješavanjem ovog sistema dobija se:

$$P(00) = P(11) = \frac{5}{16}$$
$$P(01) = P(10) = \frac{3}{16}$$