SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Ivan Zovko

TEOREM O KANALU SA ŠUMOM

Diplomski rad

Voditelj rada:
Prof. dr. sc. Siniša Slijepčević

Zagreb, rujan 2014

Ovaj diplomski rad obranjen je dana _____ pred ispitnim povjerenstvom u sastavu:

1. _____, predsjednik
2. _____, član
3. _____, član

Povjerenstvo je rad ocijenilo ocjenom _____.

Potpisi članova povjerenstva:

1. _____
2. _____
3. _____

Sadržaj

Uvod

Teorem o kanalu sa šumom je jedan od ključnih rezultata koje je dokazao Claude Elwood Shannon koji nam govori o mogućnostima komunikacije u neidealnim uvjetima. Važnost teorema je iznimna, pogotovo u Teoriji informacija. No objasnimo prvo malo pozadinu teorema. Opišimo dvije situacije kako bi dobili bolju sliku o problemu koji nastojimo riješiti.

Primjer 0.0.1. *Letjelica Galileo se nalazi u orbiti planeta Jupitera i na dnevnoj bazi obrađuje podatke. Zanstvenici na Zemlji imaju pristup komunikaciji sa Galileom i pri istraživanjima šalju upite, a Galileo im odgovara šaljući informacije.*

Primjer 0.0.2. *Prilikom obrade nekih podataka želimo sačuvati svoje rezultate na neko trajno mjesto i u nekom budućem vremenu ih pročitati.*

U oba ova primjera primjećujemo jednostavan koncept. Uočavamo kako imamo neku ulaznu informaciju (informacije koje prikuplja Galileo ili rezultate nekog istraživanja), neki način transfera informacija (radio valovi i zapisivanje podataka na tvrdi disk) te imamo još izlaznu informaciju (podaci koje dobivaju znanstvenici na Zemlji i čitanje rezultata istraživanja nakon što je prošlo neko vrijeme).

Prilikom komunikacije u ovakvim sustavima može doći do nekih neželjenih svojstava. U ovom radu ćemo pretpostavljati kako se komunikacija odvija u bitovima (Galileo šalje poruku kodiranu bitovima, obrađeni podaci zapisuju rezultate u bitovima na tvrdom disku). Kada izvor šalje ovu ulaznu informaciju tokom transfera može doći do problema koji se očituje u promjeni sadržaja izvorne poruke (npr. prilikom učitavanja podataka s Galilea pojavljuje se pozadinska radijacija koja promijeni neke bitove poruke ili pri zapisu podataka na tvrdi disk se pojavi neki nepoželjni magnetski utjecaj i poremeti podatke).

Prva ideja koja nam pada na pamet je fizički poboljšati performace ovih sustava, no ona ima jedu veliku manu. Galilea je gotovo nemoguće naknadno modificirati (ali i konstrukcija novog satelita koje bi ga mijenjao se ne čini kao jednostavno rješeneje). Usavršavanje

sustava za spremanje podataka iziskuje također nove troškove. A također, provođenje ovakih rješenja nam opet ne garantira kako će ovaj problem biti sigurno riješen.

Za razliku od fizičkih rješenja, Teorija informacija i teorija kodiranja nude malo drukčiji pogled na ovaj problem. Umjesto fizičkih poboljšanja performasi, prihvatimo kako će nam transfer informacija biti pod nekim vanjskim utjecajem. Zato ćemo dodati samo jedan sustav kojih će nam na pametan način kodirati podatke te ih slati i pri tome na neki način svesti količinu grešaka na neku zanemarivu vrijednost.

Sve ove rezultate dugujemo Cloudu Shannonu koji je u svojim prvim radovima zadao temelje Teoriji informacija. Njegova je zasluga u tome što je pokazao mnogima kako su bili u zabludi mislivši kako se određeni pomak u smanjenju greške pri transferu informacija treba kompenzirati sa recirpočnim povećanjem količine poslanih podataka.

Poglavlje 1

Teorija informacija

Teorija informacija je grana primjenjene matematike i računarstva koja se bavi kvantifikacijom informacija i rješavanjem problema vezanih u transfere infomracija. U ovom poglavlju ćemo uvesti neke osnovne pojmove koji su nam potrebni za dokazivanje konačnog rezultata ovog rada, a to je dokaz teorema o kanalu sa šumom (teorem 3.0.5).

1.1 Kanali komunikacije

Pri komunikaciji između dva objekta postoji neki put kojim putuju informacije koje zanimaju objekte koji komuniciraju. Te puteve nazivamo kanali, a oni su kao i sve na svijetu podložni utjecaju nekih faktora koji su izvan sustava komuniciranja. Definirajmo prvo osnovni pojam koji nas zanima.

Definicija 1.1.1 (Kanal sa šumom)**.** *Svaki kanal informacija u kojem može nastati neka vrsta promjene nad podacima se naziva kanal sa šumom i označavati ćemo ga slovom Q.*

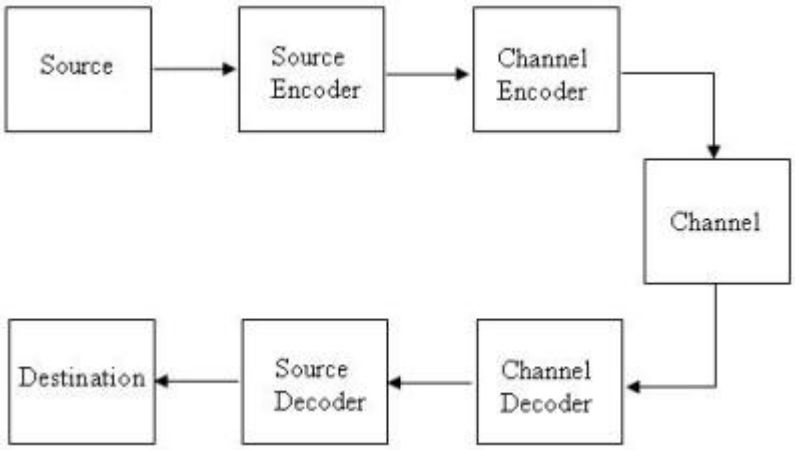
U ovom radu bazirati ćemo se na kanale koji komuniciraju bitovima te ćemo u skladu s tim i precizirati Kanal sa šumom kao kanal u kojem pri slanju poruke postoji neka smetnja koja uz neku vjerojatnost mijenja pojedini bit poruke i time čini promjenu na podacima koje smo poslali.

Nadalje, u samom procesu transfera informacija pojavljuju se posebne klase podataka, koje možemo razlikovati na određene načine (mp3.pjesma, video zapis, tekstualni dokument). Štoviše na razini bitova te klase se razlikuju po nekim oblicima ponavljanja određenih nizova bitova. Tako je dosta pametno te podatke komprimirati na neki pametan način u cilju smanjnja količine podataka koje šaljemo. U ovom radu pretpostavljamo kako su podaci dobro komprimirani i kako komprimiranje ne stvara nikakve probleme, te

nećemo ulaziti u dublju analizu. Detaljniji opis se nalazi na [1, str. 146].

Pogledajmo shemu na Slici 1.1. koja nam slikovito opisuje kanal sa šumom i shematski prikaz toka informacije koji promatramo pri dokazu teorema 3.0.5.

Slika 1.1: Tok informacije u kanalu sa šumom



Proces kojim ćemo promatrati je kodiranje i dekodiranje podataka te pokušati odediti neke njegove značajke. Znači ono što ćemo mi raditi je umjesto poboljšavanja fizičkih svojstava kanala, je to da ćemo pokušati na neki *pametan* način *zapisati* i posalti podatke te ih naknadno (mi ili netko drugi) *pročitati*. Ono što nas zanima u ovom radu je to da ćemo pokušati kvantificirati mogućnosti koje nam nude softverska rješenja problema sa nepouzdanim kanalom za komunikaciju. Dakle naš sustav koji promatramo se sastoji od:

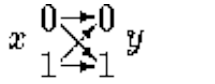
- **kompresor**- se brine kako su podaci koje želimo poslati optimalno spremljeni kako nebismo nepotrebno slali velike količine podataka,
- **kodera podataka**- kodira podatke na način da osigura što sigurniji transfer podataka kanalom te ih šalje,
- **kanal sa šumom**- kanal za razmijenu informacija u kojem se pojavlju fenomen koji pojedine podatke mijenja,
- **dekoder podataka**- na osnovu zadanog koda, dekodira podatke pristigle kanalom,
- **dekompresor**- pobrinuti će se kako pristigle podatke vratiti u prvobitnu formu zapisa podataka.

Pogledajmo sada jedan jako jednostavan primjer ovog sustava kroz primjer 1.1.2

Primjer 1.1.2 (Binarni simetrični kanal)**.** *Binarni simetrični kanal je kanal kojim šaljemo podatke bit po bit i pri tome se pojavljuje šum koji mijenja podatke sa vjerojatnšću q , tj. podatci se prenose sa vjerojatnošću točnosti od* $(1 - q)$*.*

Na Slici 1.2. se nalazi prikaz binarnog simetričnog kanala.

Slika 1.2: Binarni simertrični kanal



$$P(y{=}0|x{=}0) = 1-q; \quad P(y{=}0|x{=}1) = q;$$
$$P(y{=}1|x{=}0) = q; \quad P(y{=}1|x{=}1) = 1-q.$$

Opišimo malo sustav komunikacije preko ovakog kanala. Neka je $q = 0.2$ a mi pokušavamo poslati neki podatak preko kanala sa šumom:

1. Uzmemo neki niz bitova npr. $x = 1111100000$ za koji znamo da je dobro komprimiran
2. Pošaljemo ga kroz kanal sa šumom u kojem je $q = 0.2$ (recimo da ga zapišemo na neki loši tvrdi disk)
3. Na izlazi kanala (prilikom nekog kasnijeg čitanja podataka s diska) dobijemo niz $y = 1101100001$

Uočavamo kako je $x \neq y$. Naš cilj će biti da pokušamo odrediti koje su mogućnosti ovakvog kanala. Znači, da li se moramo zadovoljiti kako će nam x i y biti ovoliko različiti? Odgovor je očito ne, primjere rješavanja ovakvih problema imamo u Poglavlju 5. Ono što će se pokazati ključnim je to kako postoji neka kompenzacija između q i količine informacija koju šaljemo. Detaljnije o ovoj vezi u narednom poglavlju.

1.2 Podaci i mjerenje

Za početak ćemo definirati ansamblm, koji će nam biti osnovni pojam na kojem ćemo definirati neke mjere.

Definicija 1.2.1 (Ansambl)**.** *Ansambl X je trojka* $(x, \mathcal{A}_X, \mathcal{P}_X)$*, gdje je x vrijednost slučajne varijable, koja se odvija na skupu događaja* $\mathcal{A}_X$*, s vjerojatnostima* $\mathcal{P}_X$ *tako da vrijedi* $P(x = a_i) = p_i$ *te* $\sum_{a_i \in \mathcal{A}_X} P(x = a_i) = 1$.

Primjer 1.2.2. *Primjer ansambla možemo vidjeti u abecedi hrvatskog jezika, točnije ansambl je slučajno odabrano slovo abecede iz nekog dokumenta.*

Ovdje odmah možemo uočiti kako će nam ansambl služiti kao opis ulaznog koda i izlaznog koda. Vidimo kako će ulazni kod i izlazni kod imati neki vid ovisnost, koju koju ćemo kasnije definirati. Nas će zanimati neka informacija o podacima koji su ušli u kanal sa šumom, na osnovu podataka koje isčitamo pri izlazu kanala sa šumom. Štoviše, ta informacija koju izlazni kod otkriva o ulaznom kodu će se ispostaviti kao ključna stvar pri karakterizaciji mogućnosti komunikacije preko kanala sa šumom.

Vratimo se kanalima. Samo razmišljanje o razmijeni informacija putem kanala sa šumom navodi nas na dvije stvari koje će nam dati dobar uvid u stanje u kom se nalazimo. Vjerojatno prva stvar koja nam pada na pamet je kako kvantificirati grešku. Nju možemo vidjeti kroz možda neki postotak informacija koji se promijeni, što nas navodi na uvođenje vjerojatnosti. Neka je $s = \{s_1, \ldots, s_n\}$ niz bitova koji šaljemo a $\hat{s} = \{\hat{s_1}, \ldots, \hat{s_n}\}$ niz bitova koji dobiva primatelj informacije. Tada možemo definirati **vjerojatnost greške bloka** bitova p_B kao:

$$p_B = P(\hat{s} \neq s) \tag{1.1}$$

Nadalje, možemo definirati i **vjerojatnost greške pojedinog bita** p_b kao:

$$p_b = \frac{1}{n} \sum_{k=1}^{n} P(\hat{s_k} \neq s_k). \tag{1.2}$$

Sljedeća stvar o kojoj razmišljamo je slučaj kada kodiramo samu informaciju. Razumno je zapitati se koliko u našem kodu ,podacim koje ćemo slati kroz kanal, treba biti velik blok bitova da bi se poslao određeni bolok bitova neke informacije (nekodirane poruke), koju želimo posalti. Sljedeća definicija 1.2.3 nam opisuje taj pojam.

Definicija 1.2.3 (Stopa transfera)**.** *Stopa transfera ,u oznaci R,je omjer količine informacija potrebnih za opis informacije koju želimo poslati i količine podataka koje ćemo poslani kroz kanal.*

Nama je u interesu da nam stopa transfera bude čim veća. Jer ćemo imati veću korisnost transfera pojedinog bita. Nadalje, željeli bi i opisat neku maksimalnu vrijednost pri kojoj možemo dobro komunicirati.

Definicija 1.2.4 (Kapacitet kanala)**.** *Kapacitet kanala definiramo kao maksimalnu stopu transfera informacija pri kojoj je moguće komunicirati uz zanemarivo malu vjerojatnost greške bita. A oznavati ćemo ga slovom C.*

Pojam koje će nas uz ovo sve zanimati jest kod. Kod će biti onaj *pametni način zapisa* podataka pris slanju. A u nastavku, kod će biti uvijek označen slovom C i treba ga razlokovati od oznake C za kapacitet kanala.

Sada kada smo se upoznali sa nekim osnovnim pojmovima iz Teorije informacja, specificirajmo i okarakterizirajmo Diskretni informacijski kanal koji ćemo u nastavku ovog rada podrazumijevati kao kanal sa šumom a definiciju 1.1.1 možemo ostaviti u službi opisa.

Definicija 1.2.5 (Diskretni informacijski kanal)**.** *Diskretni informacijski kanal Q je karakteriziran sa ulaznom abecedom $\mathcal{A}_X$, izlaznom abecedom $\mathcal{A}_Y$ i za svaki $x \in \mathcal{A}_X$ definiranom distrbucijom uvjetne vjerojatnosti $P(x|y)$.*

Dodatno zanimat će jedna posebna vrsta koda, koju ćemo uvesti radi potreba teorema 3.0.5.

Definicija 1.2.6. *(N, K) blok kod za kanal Q je lista od $S = 2^K$ kodnih riječi duljine N*

$$\{x^{(1)}, x^{(2)}, \ldots, x^{(2^K)}\}, x^{(s)} \in \mathcal{A}_X^N. \tag{1.3}$$

Koristeći ovaj kod možemo kodirati signal $s \in 1, 2, 3, \ldots, 2^K$ kao $x^{(s)}$. Ovdje uočavamo kako je stopa R za ovaj kanal dana izrazom

$$R = K/N, \tag{1.4}$$

što odgovara definiciji stpe 1.2.3.

Sada smo dobili lagani uvid u ono što želimo pokazati i trebamo upoznati neke alate i pojmove koj će nam pomoći pri dokazu teorema 3.0.5. Jedan od takvih pojmova je etropija, koja će nam označavati jednu vrstu količine informacija koju sadrži neki objekt.

Poglavlje 2

Entropija i tipičnost

Koncept entropije je uveo Clausius 1854; Shannon ga je prenio u teoriju informacija 1948, a Kolmogorov u ergodsku teoriju 1958. Svi su je definirali kao mjeru slučajnosti ili neizvjesnosti bar kako tvrdi Petersen [2, str.227]. U ovom poglavlju ćemo se pozabaviti konceptima entropije i tipičnosti.

2.1 Entropija

Što možemo reći o količini informacija koje možemo prenjeti kroz kanal. Znima nas količina podataka koju možemo kodirati i poslati koroz kanal a da pri tom promjena na izvornim podacima bude što manje. Naša ideja je da za dani anasambl X možemo mjeriti koliko informacija nam izlazni podatak otkriva o ulaznom podatku. Za početak definirajmo dva jako bitna pojma.

Definicija 2.1.1 (Shannonov sadržaj informacija ishoda)**.** *Shannonova sadržaj informacije ishoda se definira kao*

$$h(x) = log_2 \frac{1}{P(x)} \tag{2.1}$$

a mjeri se u bitovima.

Definicija 2.1.2 (Entropija ansambla)**.** *Entropiju ansambla X definiramo kao prosječnu Shannonovu informaciju događaja:*

$$H(X) \equiv \sum_{x \in \mathcal{A}_X} P(x) \log \frac{1}{P(x)} \tag{2.2}$$

uz konvenciju da za $P(x) = 0$ *vrijedi* $0 \times \log \frac{1}{0} \equiv 0$.

Entropija ima neka jako zanimljiva svojstva od kojih je možda najzanimljivije ono o dekompoziciji entropije. Detalje se nalaze u 5

Sada ćemo se vratiti malo definiciji ansambla, točnije pogledati ćemo nešto što se zove povezani ansambl. Taj pojam ćemo primjeniti na objašnjenje kanala sa šumom koji ima ulazni podatak x te izlazni podatak y. Povezani ansambl XY je ansambl u kome uređeni par (x, y) predstavlja ishod (slučajnu varijablu) tako da vrijedi $x \in \mathcal{A}_X = \{a_1, \ldots, a_n\}$ te $y \in \mathcal{A}_Y = \{b_1, \ldots, b_n\}$ te $P(x, y)$ nam definira vjerojatnost pojedinog ishoda (uređenog para).

Dalje, analogno kao u vjerojatnost, definiramo pojmove kao što su uvjetna entropija (uz danu ishod, uz dani ansambl) te entropiju povezanog ansambla. Te napominjemo kako iz definicije 2.1.2 vidimo da entropiju pojedinog ansambla definira samo skup vrijednosti vjerojatnosti ishoda.

Definicija 2.1.3 (Enropija povezanog ansambla)**.** *Za dane ansamble X, Y definiramo entropiju povezanog ansambla*

$$H(X, Y) = \sum_{xy \in \mathcal{A}_X \mathcal{A}_Y} P(x, y) \log \frac{1}{P(x, y)}. \tag{2.3}$$

Nadalje, entropija je aditivna funkcija za nezavisne slučajne varijable

$$H(X, Y) = H(X) + H(Y) \text{akko} P(x, y) = P(x)P(y). \tag{2.4}$$

U Poglavlju (5.7) se nalazi izvod ovog rezultata.

Definicija 2.1.4. *Za dani događaj $y = b_k$ i ansambl X definiramo uvjetu entropiju kao entropiju uvjetne vjerojatnosti $P(x|y = b_k)$ točnije*

$$H(X|y = b_k) \equiv \sum_{x \in \mathcal{A}_X} P(x|y = b_k) \log \frac{1}{P(x|y = b_k)}. \tag{2.5}$$

Definicija 2.1.5 (Uvjetna entropija)**.** *Uvjetna entropija X-a uz dani Y je uprosječnje (po vrijednostima od y) uvjetne entropije iz 2.1.4.*

$$H(X|Y) = \sum_{y \in \mathcal{A}_Y} \Big[\sum_{x \in \mathcal{A}_X} P(x|y) \log \frac{1}{P(x|y =)} \Big] \tag{2.6}$$

$$= \sum_{xy \in \mathcal{A}_X \mathcal{A}_Y} P(x, y) \log \frac{1}{P(x|y)}. \tag{2.7}$$

Uvjetnu entropiju koristimo kako bismo pokazali neku mjeru neizvjesnosti od x kada nam je poznat y.

Definicija 2.1.6 (Zajednička informacija od X i Y)**.** *Zajedničku informaciju od X i Y definiramo kao*

$$I(X;Y) = H(X) - H(X|Y). \tag{2.8}$$

Lako uočavamo kako vrijedi $I(X;Y) = I(Y;X)$ te $I(X;Y) \geq 0$. I nam služi kao mjera redukcije neizvjesnoti od x činjenicom kako nam je poznat y. Jednostavnije, ako se vratimo na uvod ovog poglavnja I će nam predstavljati količinu informacija koju ćemo saznati o x kada imamo y. Znači temeljna ideja je maksimitirati I. Ovdje sada precizirati definiciju za kapacitet C za dani kanal Q, koji smo uveli u definicij 1.2.4, kao

$$C(Q) = \max_{\mathcal{P}_X} I(X;Y). \tag{2.9}$$

Vjerojatnosnu distribuciju pri kojoj se podstiže masimalan vrijednost 2.9 nazivamo **optimalna distribucija unosa**. Neka o svojstvima 2.1.6 se može naći u Poglavlju 5. Sada vidimo kako bi voljeli imati neki rezultat koji bi nam pomogao okarakterizirati funkciju zajedniče informacije.

Definirajmo još i uvjetnu zajedničku informaciju.

Definicija 2.1.7. *uvjetna zajedniča infomracija između X i Y ako znamo Z je*

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) \tag{2.10}$$

Sljedeći rezultat koji nam je potreban kako bi dokazali teorem 3.0.5 jest teorem o procesuiranju podataka (Nejednakost o obradi podataka).

Teorem 2.1.8. *Nejednakost o obradi podataka Niti jedna obrada podataka ne može poboljšati zaključke koji se izvuku iz podataka.*

Uzeti ćemo u obzi kako je SPO ansambal u kojem s označava svijet, p prikupljene podatke i o neke obrađene podatke, tako da ove tri varijable tvore Markovljev lanac.

$$s \to p \to o \tag{2.11}$$

dakle, kako se vjerojatnost $P(s,p,o)$ može zapisati

$$P(s,p,o) = P(s)P(p|s)P(o|p) \tag{2.12}$$

Pokažimo sada kako prosječna informacija koju O otkriva o S ,$I(S;O)$ je manja od prosječne informacije koju P otrkica o S , $I(S;P)$.

Prvo ćemo pokazati da vrijedi lančano pravilo za zajedničku informaciju.

$$I(X;Y,Z) = H(X) - H(X|Y,Z) \tag{2.13}$$

dodamo

$$0 = H(X|Y) - H(X|Y) \tag{2.14}$$

te dobijemo

$$I(X;Y,Z) = H(X) - H(X|Y) + H(X|Y) - H(X|Y,Z) \tag{2.15}$$

iskoristimo 2.1.6 i 2.1.7 (Y i Z iz definicje trebamo zamijeniti) i dobijemo lančano pravilo za zajedničku informaciju.

$$I(X;Y,Z) = I(X;Y) + I(X;Z|Y) \tag{2.16}$$

Dokaz. Sada iskoristimo Markovljevo svojstvo ([3, str. 180]) kako bi dobili $I(S\,;O|P) = 0$, zatim iskoristimo 2.16 (za P i za O)

$$I(S\,;P,O) = I(S\,;P) \tag{2.17}$$
$$I(S\,;P,O) = I(S\,;O) + I(S\,;P|O) \tag{2.18}$$

iz čega slijedi primjenom $I(X,Y) \geq 0$

$$I(S,O) \leq I(S,P) \tag{2.19}$$

Sadaa zaključujemo kako P otkriva više informacija o S no O. Čime smo pokazali teorem. □

Sljedeći teorem samo navodimo. Bitna nam je samo interpretacija teorema koja će nam osigurati dovoljnu količinu znanja da bismo mogli dokazati teorem 3.0.5, dok se više informacija o narednom teoremu se nalazi na [1, str. 78, Teorem 4.1] te dokaz [1, str. 82, Proof of theorem 4.1].

Teorem 2.1.9 (Teorem od kodiranju izvora)**.** *Neka je X anasambl sa entropijom $H(X) = H$ bitova. Za dani $\epsilon > 0$ i $\delta \in [0,1]$ postoji pozitivan prirodan broj N_0 takav da za svaki $N > N_0$ vrijedi*

$$\left|\frac{1}{N}H_\delta(X^N) - H\right| < \epsilon \tag{2.20}$$

Ovaj teorem nam daje jedan važan rezultat *Neovisno koliku grešku ćemo tolerirati broj bitova, po simbolu, potrebnih kako bi odredili x je H bitova.*

2.2 Tipičnost

Uvedimo sada pojam pridruženo tipičnog niza, koja će nam trebati u dokazu teorema 3.0.5. Definirati ćemo $x^{(s)}$ kao kodnu riječ koja dolati iz ansambla X^N, te je uzeti u obzir slučajni kao slučajni odabir jedne riječi koda te danog dobivenog podatka y, time definirajući ansambl $(XY)^N$. Kosristiti ćemo dekoder pomoću tipične povezanosti, koji dekodira y kao s ukoliko su $x^{(s)}$ i dobiveni izlazni kod y *tipično povezani* (pojam koji ćemo uskoro definirati). Tada će se dokaz fokusirati na određivanje dviju vjerojatnosti

(a) da je prava kodna riječ unosa nije tipično povezana sa izlaznim kodom

(b) da je pogrešna kodna riječ unosa *tipično povezana* sa izlasnim kodom

Pokazati će se kako obje ove vjerojatnosti za dovoljno velike N idu prema nuli dok god postoji manje od $2^N C$ kodnih riječi te da X ima optimalnu distribuciju unosa.

Definicija 2.2.1 (Tipična povezanost)**.** *Par nizova bitova x, y duljine N je tipično povezan (uz toleranciju β) uz danu distribuciju $P(x, y)$ ukoliko*

$$\begin{aligned}
&x \text{ je tipičan uz } P(x), && t\,j., && \left|\frac{1}{N}\log\frac{1}{P(x)} - H(X)\right| < \beta,\\
&y \text{ je tipičan uz } P(y), && t\,j., && \left|\frac{1}{N}\log\frac{1}{P(y)} - H(Y)\right| < \beta,\\
&te\ x, y \text{ su tipični uz } P(x, y), && t\,j., && \left|\frac{1}{N}\log\frac{1}{P(x,y)} - H(X,Y)\right| < \beta.
\end{aligned}$$

Definicija 2.2.2. *Skup tipično povezanih parova $J_{N\beta}$ je skup svih tipično povezanih parova duljine N.*

Primjer 2.2.3. *Pogledajmo primjer jednog tipično povezanog para sa $N = 60$ za ansambl zadan vjerojatnošću P(x,y) u kojoj $P(x)$ ima $(p_0, p_1) = (0.9, 0.1)$ a P(y—x) odgovara binarnom simetričnom kanalu sa razinom šuma 0.2 (20%),*

x 001100000000000000000001100000000000000000110000000000000000

y 101100001001000100100001100010001000100001000010000000100100

Primjećujemo kako x ima 6 jedinica; dok y ima 17 jedinica ; te kako se razlikuju u 12 bitova što je tipični broj promjene bitova koju prođu koz kanal.

Teorem 2.2.4 (Teorem o pridruženoj tipičnosti)**.** *Neka su x, y izvučeni iz anasbmla $(XY)^N$ definiranog sa vjerojatnosti*

$$P(x, y) = \prod_{n=1}^{N} P(x_n, y_n) \tag{2.21}$$

Tada vrijedi:
1. vjerojatnost da su x, y tipično povezani (uz toleranciju β) teži prema 1 kada $N \to \infty$;
2. broj tipično poveznaih nozova $|J_{N\beta}|$ je blizu $2^{NH(X,Y)}$, točnije,

$$|J_N\beta| \leq 2^{N(H(X,Y)+\beta)}; \tag{2.22}$$

3 ako x' X^N i y' Y^N, tj. x' i y' su nezavisni uzorci sa istim marginalnim distribucijama kao $P(x, y)$, tada vjerojatnost da (x', y') bude u skupu tipično povezanih je oko $2^{-NI(X;Y)}$, točnije,

$$P((x', y') \in J_{N\beta}) \leq 2^{-N(I(X;Y)-3\beta)} \tag{2.23}$$

Dokaz. Prve dvije tvrdnje su direktne posljedice primjene zakona velikih brojeva ([3, str.146, Teorem 6.10.]) i teorem o kodiranju izvora (2.1.9). S upozorenjem kako za drugi dio treba pripaziti, jer umjesto niza bitova gledamo na uređeni par dvaju nizova bitova, te koristeći distribuciju vjerojatnosti $P(x, y)$ umjesto $P(x)$.
Za treći dio vrijedi naredno,

$$\begin{aligned} P((x', y') \in J_{N\beta}) &= \sum_{(x,y)\in J_{N\beta}} P(x)P(y) && (2.24)\\ &\leq |J_{N\beta}|2^{-N(H(X)-\beta)}2^{-N(H(Y)-\beta)} && (2.25)\\ &\leq 2^{N(H(X,Y)+\beta)-N(H(X)+H(Y)-2\beta)} && (2.26)\\ &= 2^{-N(I(X;Y)-3\beta)}. && (2.27) \end{aligned}$$

□

Prije samog teorema 3.0.5 pokušati ćemo dočarati ideju koja se krije u dokazu teorema. Zamislite kako želimo pokazati kako u nekom vrtiću jedno od 100 male djece ima manje od 10kg no uhvatiti svu djecu je gotovo nemoguće. Shanonova metoda kaže da umjesto vaganja djece pojedinačno jednostavno ih sve izvagamo odjednom. Ukoliko je ukupni zbroj manji od 1000kg dobili smo odgovor. Ovim postupkom nećmo naći dijete koje ima manje od 10kg no utvrditi ćemo da takvo dijete postoji postoji.

Mi želimo pokazati kako postoji kod i dekoder koji imaju malu vjerojatnost greške. Računanje vjerojatnosti greške bilo kojeg koda je mukotrpan posao. Mi ćemo zato iskoristiti Shanonovu dosjetku: *umjesto konstrukcije idealnog kodiranjućeg i dekodirajućeg*

sustava i računanja njegove greške, bolje je izračunati prosjek vjerojatnst greške bloka svih kodova, i pokazati kako je taj prosjek zanemariv.

Ukoliko to pokažemo sigurno mora postojati barem jedan kod koji ima malu vjeroajtnost greške bloka i time ćemo dokazati teorem 3.0.5.

Poglavlje 3

Teorem o kanalu sa šumom

U ovom poglavlju rada dokazujemo teorem za koji smo napravili većinu predradnji u prethodnim poglavljima.Teorem o kanalu sa šumom govori o mogućnostima komuniciranja kada koristimo kanal as šumom.

Teorem 3.0.5 (Teorem o kanalu sa šumom)**.** *Informacijama se može komunicirati preko kanala sa šumom pri ne-nul stopi transfera informacija sa razmjerno malom razinom vjerojatnosti greške (dakle pri vjerojatnosti greške koja nam je prihvatljiva). Dakle:*

1. Za svaki diskretni kanal informacija kapacitet kanala dan izrazom

$$C = \max_{\mathcal{P}_X} I(X;Y) \tag{3.1}$$

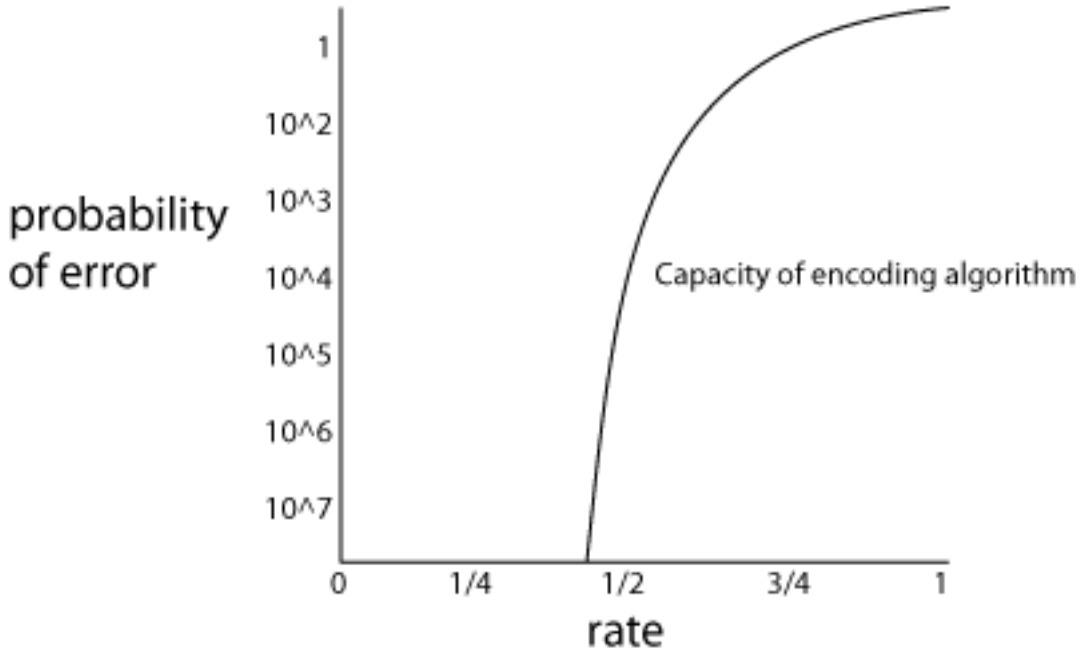
ima sljedeće svojstvo. Za svaki $\epsilon > 0$ i $R < C$, uz dovoljno veliki N , postoji kod duljine N i stopa $\geq R$ te neki dekodirajući algoritam, tako da je maksimalna vjerojatnost greške boloka bude $< \epsilon$.
2. Ukoliko je vjerojatnost greške bita p_b prihvatljiva, tada su stope do $R(p_b)$ dostižne , gdje

$$R(p_b) = C/(1 - H_2(p_b)). \tag{3.2}$$

.
3. Za bilo koji p_b, stope veće od $R(p_b)$ nisu dostižne

Promotrimo malo iskaz. Ono što smo do sada saznali je kako znamo da možda postoji kapacitet, tj. stopa transfera informacija za koju znamo da pri njoj možemo komunicirati uz zanemarivu gršku (grešku koju smo spremni prihvatiti). Mi ćemo se u ovom poglavlju pretvarati kako znamo da postoji kapacitet, a u narednom poglavlju ćemo to i dokazati.

Tvrdnja teorema se jako dobro može prikazati na sljedećoj slici 3.1. Iz razloga što se vidi način na koji kompenziramo vjerojatnost greške sa stopom transfera informacija.



Slika 3.1: Odnos između vjerojatnosti greške bita i stope transfera infprmacija koja je rezultat teorema 3.0.5

Shannon je time opovrgnuo uvjerenje kako za veliku pozdanost slanja (malu vjerojatnost greške) potrebno slati javo veliku količinu informacij. Tada su postojala razmišljanja kako vrijedi da postizanje $p_b \to 0$ povlači $R \to 0$.

3.1 Kodiranje i prvi dio teorema

Naš cilje je pokazati postojanje koda i dekodera koji imaju malu vjerojatnost greške. Procjena vjerojatnosti greške bilo kojeg sustava kodiranja i dekodiranja nije lagan zadatak. Shannonova inovacije je bila ta da je, umjesto konstukcije dobrog sustava kodiranja/dekodiranja i procijene vjerojatnosti greške istog, računao prosječnu vjerojatnost greške blok kodova. Ukoliko bi ta greška bila mala tada bi postojao kod koji zadovoljava navedene kriterije.

Promotrimo naredni sustav kodiranja, čija je stopa R'

1. Fiksirajmo vjerojatnost $P(x)$ i generiramo $S = 2^{NR'}$ kodnih riječi od $(N, NR') = (N, K)$ koda C na slučajan način uz

$$P(x) = \prod_{n=1}^{N} P(x_n). \tag{3.3}$$

2. Kod je poznat i pošiljatelju i primatelju informacije.

3. Poruka s je izabrana izabrana iz skupa $1, 2, \ldots, 2^N R'$ te je poslana poruka $x^(s)$. Primljena poruke je y, uz

$$P(y|x^(s)) = \prod -n = 1^N P(y_n|x_n^(s)). \tag{3.4}$$

4. Signal se dekodira pomoću dekodiranja tipičnim skupom.
Dekodiranje tipičnim skupom. y dekodiramo kao $\hat{s}$ ako je $(x^{(\hat{s})}, y)$ tipično spojen a ne postoji neki drugi s' tekav da vrijedi da su $(x^{(s')}, y)$ tipično spojeni. Inače vrati 0 (označava grešku).

5. Greška dekodiranja se događa ukoliko $\hat{s} \neq s$.

Pojavljuju se tri vrste grešaka koje možemo istaknuti. Prvo, postoji vjerojatnost greške bloka za neki kod C koja je

$$p_b(C) \equiv P(\hat{s} \neq s|C). \tag{3.5}$$

koju je jako teško odrediti.

Drugo, postoji prosjek po svim kodovima ove vjerojatnosti greške bloka

$$\langle p_b \rangle \equiv \sum_C P(\hat{s} \neq s|C)P(C) \tag{3.6}$$

koju je srećom puno zogodnije izračunati no 3.5.

Treće, maksimalna vjerojatnost greške bolka koda C

$$p_{BM}(C) \equiv max_s P(\hat{s} \neq |s, C), \tag{3.7}$$

je vjerojatnost koja nas zanima. Želimo pokazati kako postoji kod C sa traženom stopom čija je maksimalna vrijednost vjerojatnosti greške bloka mala. Do tog rezultata ćemo doći računajući prosječnu vjerojatnost greške bloka $\langle p_b \rangle$. Jednom kada pokažemo kako ova veličina može biti manja od dovoljno malog broja, momentalno ćemo zaključiti kako mora

postojati barem jedan kod C čija vjerojatnost greške bloka je također manja od tog broja. Konačno, pokazati ćemo kako ovaj kod, čija je vjerojatnost greške bloka zadvoljavajuće mala ali čija je maksimalna vjerojatnost greške bloka nepoznata, možemo modificirati tako da od njega dobijemo neznatno manje "stope" čija maksimalna vjerojatnost greške bloka sigurno manja od unaprijed zadane. Kod modificiramo izbacujući iz njega 50% lošijih riječi koda. Prebacimo se sada na pronalaženje prosječne vjerojatnosti greške bloka.

Kada koristimo dekoder pomoću tipičnog skupa događaju se dvije vrste(izvora) grešaka:

(a) y nije tipično spojen sa poslanom kodnom riječi x^s, ili

(b) postoji neka druga kodna riječ u kodu C koja je tipično spojena sa y.

Po simetriji konstrukcije koda, prosječna vjerojatnost greške koda usrednjena po svim kodovima ne ovisi o izboru vrijednosti s. Pretpostavimo BSO kako je $s = 1$.

(a) Vjerojatnost da x^1 i y nisu tipično spojeni, po teoremu 2.2.4. Neka je δ gornja ograda ove vjerojatnosti, koja zadovoljava $\delta \to 0$ kada $N \to \infty$; Za bilo koji δ, možemo naći duljinu bloka $N(\delta)$ koja daje $P((x^1, y) \notin J_{N\beta}) \leq \delta$.

(b) Vjerojatnost da $x^{(s')}$ i y su tipično spojeni, za dani $s' \neq 1$ je $2^{-N(I(X;Y)-3\beta)}$, po trećem dijelu teorema 2.2.4. Nadalje, postoji $2^{NR'} - 1$ suprotnih vrijednosti od s' koje ne odgovaraju. Stoga vjerojatnost greške $\langle p_b \rangle$ zadovoljava :

$$\langle p_b \rangle \leq \delta + \sum_{s'=2}^{2^{NR'}} 2^{-N(I(X;Y)-3\beta)} \tag{3.8}$$

$$\leq \delta + 2^{-N(I(X;Y)-R'-3\beta)} \tag{3.9}$$

Vjerojatnost greške 3.9 može se učiniti manjom od $< 2\delta$ povećavajući N dok god vrijedi

$$R' < I(X;Y) - 3\beta. \tag{3.10}$$

1. Biramo $P(x)$ u dokazu tako da bude optimalna distribucija unosa kanala. Tada uvjet 3.10 postaje $R' < C - 3\beta$

2. Kako je prosječna vjerojatnost greške svih kodova manja od 2δ, mora postojati kod sa srednjom vjerojatnosti greške bloka $p_b(C) < 2delta$.

3. Da bi pokazali kako možemo ne samo prosječnu nego i maksimalnu vjerojatnost greške smanjiti, modificiramo kod tako da izuzmemo pola lošijih kodnih riječi (one koje će najvjerojatnije učiniti grešku). Ove koje su preostale moraju imati uvjetnu vjerojatnosti greške koda manju od 4δ. Dakle ove preostale uzmemo za definiranje koda. Ovaj novi kod ima $2^(NR'-1)$ riječi, tj smanjili smo stopu sa R' na $R'-1/N$ (što je zanemariva korekcija za velike N) a postigli $p_B M < 4\delta$. Dobiveni rezultat možda i nije najbolje rješenje uz ovu stopu i dužinu koda, ali je dovoljan kako bi pokazali teorem o kanalu sa šumom.

Dokaz. Sada možemo konstruirati kod sa stopom $R' - 1/N$, gdje $R' < C - 3\beta$, sa maksimalnom vjerojatnošću greške 4δ. Teorem dokazujemo postavljanjem $R' = (R + C)/2$, $\delta = \epsilon/4$, $\beta < (C - R')/3$, te N-om dovoljnov velikim za održavanje ostalih uvjeta. Čime smo dokazali prvu tvrdnju teorema.

□

3.2 Drugi dio teorema

Za dokaz drugog dijela teorema trebamo razmotriti komunikaciju kanalom iznad kapaciteta. Dokazali smo kako za svaki diskretni kanal postoje dostižne točke na ravnini (R, p_b) -grafa kao što je na slici 3.1 . Ovaj rezultat nam govori kako kanal sa šumom možemo pretvoriti u gotovo bešuman kanal sa kapaciteom C po ciklusu izmjene informacija. Sada želimo kvantificirati ovaj rezultat, na način da ćemo odrediti desnu granicu prostora dsotižnih točaka. Ona nas zanima jer će nam ona dati opis mogućnosti transfera informacija za pojedinu razinu greške.

Ovdje koristimo novi trik. Kako znamo da možemo kanal sa šumom pretvorit u perfektan kanal nauštrb gubitka stope prenosa informacija, dovoljno je uzeti u obzir komunikaciju sa greškama preko kanala bez šuma. Koliko brzo možemo komunicirati preko kanala bez šuma , ako je dozvoljeno praviti greške?

Uzmimo u obzir binarni bešumni kanal, i pretpostavmo kako potičemo komunikaciju na stopama preko njegovog kapaiciteta koji iznosi 1 bit (Npr.ako tražimo da pošiljatelj pokuša komunicirati na 2bita po ciklusu tada on efektivno mora odbaciti pola informacija). Koji je najbolji način da to izvedemo a pri tome dosegnemo najmanju moguću vjerojatnost greške bita? Jedna jednostavna strategija je komunicirati fragmet 1/R od izvornih bitova i ignorirati ostatak. Primatelj pogađa ostatak djela poruke , točnije 1-1/R dio poruke, na slučajan način. te je vjerojatnost greške bita dana

$$p_b = \frac{1}{2}(1 - \frac{1}{R}). \tag{3.11}$$

No, ova granica je pre jednostavna i suviše neprecizna za rezultat koji mi trebamo. Ono što ćemo učiniti jest iskoristiti alate koje već imamo. Ideja je da podijelimo rizik promjene bitova jednako po svim bitovima.

Dokaz. Alat koji trebamo je (N, K) kod za kanal sa šumom te ga primjenio inverzno kako bi dekoderom odredili nespretni kompresor. Specijalno, uzmemo odlični (N, K) kod za binarni simetrični kanal. Pretpostavimo da takav kod ima stopu $R' = N/K$ te da je sposoban korigirati greške koje proizvede kanal čija je vjerojatnost protoka inforacije q. Asimptotski, kodovi stope R' postoje koji imaju R jednaku približno $1-H_2(q)$. Prisjetimo se, ukoliko dodamo jedan od ovih kodova dužine N koji postižu kapacitet u binarni simetrični kanal tada (a) distribucija vjerojatnosti izlaza kanala je približno uniformna, pošto je entropija izlaza jednaka entropiji izvora (NR') uz enropiju šuma ($NH_2(q)$), te (b) optimalni dekoder koda, u ovoj situaciji, tipično preslikava primljeni vektor dužine N u poslani vektor koji se razlikuje u qN bitova od primljenog vektora.

Uzmemo signal koji želimo poslati , i isječemo ga u dijelove duljine N bitova. Provedemo svaki blok kroz dekoder i zadžimo kraći signal dužine K bitova, koji pošaljemo preko bašumnog kanala. Za dekodiranje slanja, proslijedimo K bitnu poruku koderu originalnog koda. Rekonstruirana poruka će se sada razlikovati, uoko prilike, qN bitova od originalne poruke. Tako će vjerojatnost greške bita biti na $p_b = q$. Stopa ovog problematičnog kompresora je $R = N/K = 1/R' = 1/(1 - H_2(p_b))$.

Dodajući ovaj problematični kompresor našem kapacitetu C komunikacije bez greške, dokazali smo dostižnost komunikacije do vrijendsti krivulje na ravnini (pb, R) (slika 3.1) definirane:

$$R = \frac{C}{1 - H_2(p_b)} \tag{3.12}$$

□

3.3 Nedostižne tokče ravnine

Dokaz. Izvor, koder, kanal sa šumom i dekoder definiraju Markovljev lanac:

$$P(s, x, y, s') = P(s)P(x|s)P(y|x)P(s'|y) \tag{3.13}$$

Nejednakost koja slijedi iz teorema 2.1.8 možea se primjeniti na ovaj lanac: $I(s; s') <= I(x; y)$. Štoviše, po definiciji kapaciteta kanala $I(x; y) \leq NC$ pa vrijedi $I(s; s') \leq NC$. Pretpostavimo kako sustav postiže stopu R i vjerojatnost greške bita p_b : tada zajedniča

informacija $I(s; s')$ je $\geq NR(1 - H_2(p_b))$. No $I(s; s') > NC$ nije dostižna, pa ni $R > C/(1 - H_2(p_b))$ nije dostižna. □

Ovime smo dokazali sve tri tvrdnje teorema. Preciznije, ukoliko smo primorani komunicirati preko nekog kanala u kojem će nam dio podataka biti promijenjen ipak možemo komunicirati na svim stopama transfera informacija do kapaciteta kanala, no za stope koje su veće od kapaciteta to nije moguće.

Ono što nam ostaje za kraj je kako izračunati kapacitet kanala (točnije diksrento memorijskog kanala) i postoji li. Ono što nam rezultatati iz prethodnih poglavlja sugeriraju je pronalazak optimalne distribucije kanala. Prva ideja koja nam pada je jednostavno derivacijom 2.1.6. U narednom poglavlju ćemo pokazati da je funkcija zajedničke informacije I konkavna. Pa pronalaskom stacionarne točke ove funkcije možemo odrediti točku maksimuma. No postoji problem, znamo kako vjerojatnosti pojedinog unosa može biti nula (razumno je kako u nekom kodovima nećemo koristiti određene kodne riječi jer ih možemo izbaciti sa onih 50%). Ono što će ovdje donijeti rješenje je svojstvo simetričnosti koja se ispostavlja kao jako bitan koncept, a ono na neki način slijedi iz konkavnosti. Detalji se nalazi u [1, str 169., Poglavlje 10.6]

Poglavlje 4

Kapacitet

U ovom poglavlju ćemo pokazati kako postoji kapacitet kanala sa šumom, tako što ćemo pokazati da je funkcija zajedniček informacije I konkavna u varijabli vjerojatnosne distribuceje podataka iz skupa $\mathcal{P}_X$. BSO ćemo sve rezultate promatrati na slučajnim varijablama umjesto na ansamblima. Svi rezultati se mogu poopćiti na ansambl.

4.1 Konkavnost

Definirajmo konveksnost i konkavnost

Definicija 4.1.1 (Konveksnost)**.** *Za neku funkciju $f(x)$ kažemo kako je konveksna na intervalu (a, b) ukoliko za svaki $x_1, x_2 \in (a, b)$ te $\lambda \in [0, 1]$ vrijedi,*

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \tag{4.1}$$

*Nadalje za f kažemo kako je **striktno** konveksna ukoliko jednakost vrijedi za $\lambda = 1$ ili $\lambda = 1$.*

Definicija 4.1.2 (Konkavnost)**.** *Za funkciju f kažemo da je konkavna ukoliko je -f konveksna.*

Sada ćemo iskazati jedna poznati nam rezultat, čiji dokaz možemo naći u [4, str. 27, Teorem 2.6.2].

Teorem 4.1.3 (Jensenova nejednakost)**.** *Ako je f konveksna funkcija i X slučajna varijabla, vrijedi*

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]) \tag{4.2}$$

Teorem 4.1.4 (Log sum nejednakost)**.** *Za nenegativne brojeve $a_1, a_2, \ldots, a_n$ i $b_1, b_2, \ldots, b_n$ vrijedi*

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq (\sum_{i=1}^{n} a_i) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \tag{4.3}$$

a jednakosti samo ukoliko $\frac{a_i}{b_i}$ =const.

Dokaz. BSO pretpostavmo kako vrijedi $a_i b_i > 0$ Funkcija $f(x) = x \log x$ je striktno konveksna, $f''(x) > 0$ za sve $x > 0$. Postavimo sada

$$\alpha_i = \frac{b_i}{\sum_{j=1}^{n} b_j} \qquad , t_i = \frac{a_i}{b_i} \tag{4.4}$$

Po Jensenovoj nejednakosti vrijedi

$$\sum \alpha_i f(t_i) \geq f(\sum \alpha_i t_i) \tag{4.5}$$

za svaki $\alpha_i \geq 0, \sum_i \alpha_i = 1$. Sada samo prepravimo nejednakost 4.3.

$$\sum_{i=1}^{n} \frac{a_i}{\sum_{j=1}^{n} b_j} \log \frac{a_i}{b_i} \geq \sum_{i=1}^{n} \frac{a_i}{\sum_{j=1}^{n} b_j} \log \sum_{i=1}^{n} \frac{a_i}{\sum_{j=1}^{n} b_j} \tag{4.6}$$

Odakle slijedi tražena tvrdnja.

□

Ovime smo napravili mali uvod sa konveksnošću koja će nam trebati kako bismo dokazali konkavnost funkcije zajedničke informacije I.

4.2 Općenitije o entropiji

Uvedimo pojam **relativne entropije** ili Kullback-Leiblerove udaljenosti.

Definicija 4.2.1. *Relativna entropija između dvije funkcije gustoće $p(x)$ i $q(x)$ se definira kao*

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \tag{4.7}$$

$$= \mathbb{E}_p \log \frac{p(X)}{q(X)} \tag{4.8}$$

Definirajmo sada neku novu informaciju I' preko relativne entropije.

Definicija 4.2.2. *Neka su X, Y slučajne varijable sa zajedničkom funkcijom gustoće $p(x, y)$ i graničnim funkcijama gustoće $p(x)$ i $p(y)$. Zajednička informacija se tada definira kao*

$$I'(X;Y) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{4.9}$$

$$= D(p(x,y)||p(x)p(y)). \tag{4.10}$$

Pokažimo kako se 2.1.6 i 4.2.2 ne razlikuju, tj. kako je $I = I'$. Dakle vrijedi

$$I'(X;Y) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{4.11}$$

$$= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log\frac{p(x|y)}{p(x)} \tag{4.12}$$

$$= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log\frac{1}{p(x)} + \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log p(x|y) \tag{4.13}$$

$$= \sum_{x\in\mathcal{X}} p(x)\log\frac{1}{p(x)} - \sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log\frac{1}{p(x|y)} \tag{4.14}$$

$$= H(X) - H(X|Y) = I(X;Y) \tag{4.15}$$

Znači da za I možemo koristiti obe karakterizacije zajedničke informacije. U poglavlju 5 su opisane neke karakteriskike funkcije I.

Vratimo se sada ponovo na relativnu entropiju te pokažimo jedno bitno svojstvo.

Lema 4.2.3 (Nenegativnost relativne entropije)**.** *Neka su $p(x)$, $q(x)$ dvije funkcije gustoće definirane na istom prostoru. Vrijedi,*

$$D(p||q) \geq 0. \tag{4.16}$$

sa jednakosti ukoliko je $p \equiv q$.

Dokaz. Uzmimo kako je $A = \{x : p(x) > 0\}$ nosač od $p(x)$ Tada vrijedi

$$-D(p\|q) = -\sum_{x\in A} p(x)\log\frac{p(x)}{q(x)} \tag{4.17}$$

$$= \sum_{x\in A} p(x)\log\frac{q(x)}{p(x)} \tag{4.18}$$

$$\leq \log\sum_{x\in A} p(x)\frac{q(x)}{p(x)} \tag{4.19}$$

$$= \log\sum_{x\in A} q(x) \tag{4.20}$$

$$\leq \log\sum_{x\in \mathcal{X}} q(x) \tag{4.21}$$

$$= \log 1 = 0 \tag{4.22}$$

Nejednakost 4.19 proizlati iz Jensenove nejednakosti. Zbog striktne konkavnosti od $\log t$ u t, imamo kako za 4.19 vrijedi akko $q(x)/p(x)$ je konstantna svugdje to znači kako bi $q = cp$ gjde je c konstanta. Stoga vrijedi $\sum_{x\in A} q(x) = c\sum_{x\in A} p(x) = C$. Tako u 4.21 imamo jednakost $\sum_{x\in A} q(x) = \sum_{x\in\mathcal{X}} q(x) = 1$ što sugerira $c = 1$.

Znači vrijedi $D(p\|q) = 0$ akko $p(x) \equiv q(x)$ □

4.3 Postojanje kapaciteta

Teorem 4.3.1 (Konveksnost relativne entropije)**.** *$D(p\|q)$ je konveksna u paru (p, q); ukoliko su (p_1, q_1) i (p_2, q_2) su dva para vjerojatnosnih distribucija, tada za $\lambda \in [0, 1]$ vrijedi*

$$D(\lambda p_1 + (1-\lambda)p_2)\|\lambda q_1 + (1-\lambda)q_2) \leq \lambda D(p_1\|q_1) + (1-\lambda)D(p_2\|q_2) \tag{4.23}$$

Dokaz. Za dokaz ovog teorema koristiti ćemo teorem 4.1.4 na desni izraz u 4.23

$$(\lambda p_1(x)+(1-\lambda)p_2(x))\log\frac{(\lambda p_1(x)+(1-\lambda)p_2(x))}{(\lambda q_1(x)+(1-\lambda)q_2(x))} \leq \lambda p_1(x)\log\frac{\lambda p_1(x)}{\lambda q_1(x)}+(1-\lambda)p_2(x)\log\frac{(1-\lambda)p_2(x)}{(1-\lambda)q_2(x)} \tag{4.24}$$

Na prethodnu formulu djelujemo sumiranjem po svim x-evima čime se dobiav tražena nejednakost. □

Lema 4.3.2. *$H(X) \leq \log|\mathcal{X}|$ s tim da jednakost vrijedi samo u slučaku kada je X U.*

Dokaz. Neka je $u(x) = \frac{1}{|\mathcal{X}|}$ funkcija gustoće uniformne slučajne varijable na $\mathcal{X}$. Uzmimo sada da je p(x) neka praoizvolja funkcija gusotoće od X. Vrijedi

$$D(p||u) = \sum p(x) \log \frac{p(x)}{u(x)} = \log|\mathcal{X}| - H(X) \tag{4.25}$$

Sada iskoristimo neneagativnost od relativne entropije

$$0 \leq D(p||q) = \log|\mathcal{X}| - H(X) \tag{4.26}$$

Time smo pokazali ovu kratku lemu. □

Za sljedeči dokaz trebamo se podsjetiti karakterizacije konkavnosti. Znamo da je f konakvna ukoliko je $-f$ konveksna, također vrijedi kako je f konkavna ukoliko je $M - f$ konveksna , za neku konstatu M.

Iskorostimo sada rezultat prehodne leme 4.3.2 i zapišemo entropiju od X (koja ima p funkciju gustoće) kao

$$H(p) = \log|\mathcal{X}| - D(p||u) \tag{4.27}$$

Sada iskoristimo teorem 4.3.1 koji kaže kako je funkcije D konveksna u p time zaključujemo kako je funkcija $H(p)$ konkavna u varijabli p. Time smo dokazali naredni teorem.

Teorem 4.3.3. *$H(p)$ je konkavna funkcija u varijabli p.*

Sada nam još preostaje dokazati najvažniji teorem ovog poglavlja. Kako bismo pokazali da kapacitet postoji dovoljno nam je pokazazi kako je funkcija zajedničke informacije konkavna u varijabli p time bismo pokazali kako se maksimum postiže, dakle postoji C.

Teorem 4.3.4. *Neka slučajni vektor (X, Y) $p(x, y) = p(x)p(y|x)$. Zajednička informacija $I(X; Y)$ je konkavna funkcija od $p(x)$ za fiksnu $p(y|x)$, ali je konveksna u varijabli $p(y|x)$ za fiksnu $p(x)$.*

Dokaz. Kako bismo dokazali prvi dio teorema raspisati ćemo funkciju zajedničke informacije

$$I(X; Y) = H(Y) - H(Y|X) = H(Y) - \sum_{\mathcal{X}} p(x)H(Y|X = x). \tag{4.28}$$

Kako je $p(x|y)$ fiksirano, tada je $p(y)$ linearna funkcija od $p(x)$. Stoga entropija $H(Y)$, koja je konkavna funkcija od $p(y)$ je i konkavan funkcija od $p(x)$. Nadalje drugi izraz je zapravo

linearna funkcija od $p(x)$. Tako da je čitav izraz razlika konkavne funkcije i linearne funkcije pa je i sama dakle konkavna. Time smo pokazali prvu tvrdnju.

Za drugu tvrdnju fiksiramo $p(x)$ i uzmemo u obzir dvije različite uvijetne distribucije $p_1(y|x)$ i $p_2(y|x)$. Sada to iskoristimo i napišemo odgovarajuće združene dstribucije $p_1(x, y)$ i $p_2(x, y)$ te povezane granične distribucije $p(x)$, $p_1(y)$ te $p(x)$ i $p_2(y)$. Definirajmo odgovarajuće konveksne kombinacije uz dani $\lambda \in [0, 1]$.

$$p_\lambda(x|y) = \lambda p_1(x|y) + (1 - \lambda)p_2(x|y), \tag{4.29}$$
$$p_\lambda(x, y) = \lambda p_1(x, y) + (1 - \lambda)p_2(x, y), \tag{4.30}$$
$$p_\lambda(y) = \lambda p_1(y) + (1 - \lambda)p_2(y). \tag{4.31}$$

Ukoliko definiramo $q_\lambda(x, y) = p(x)p_\lambda(y)$ kao produkt graničnih distribucija , imamo

$$q_\lambda(x, y) = \lambda q_1(x, y) + (1 - \lambda)q_2(x, y) \tag{4.32}$$

Kako je zajednička informacija relativna entropija pridruženih distribucija i umnožak graničnih,

$$I(X; Y) = D(p_\lambda(x, y)||q_\lambda(x, y)) \tag{4.33}$$

Znamk kako je $D(p||q)$ konveksna funkcija od (p, q) slijedi kako je zajednička informacija konveksna funkcija u varijabli uvijetne distribucije.

Čime smo dokazali i drugu tvrdnju teorema. □

Sada kako imamo rezultat o konkavnosti funkcije zajedničke informacije, čime smo pokazali i zadnji rezultat koji nam osigurava tvrdnju teorema 3.0.5

Poglavlje 5

Dodaci

U nastavku ćemo nastojati opisati dva koda kao primjer kodiranja za uvod u Teoriju informacija koju smo napravili u Poglavlju 1. Naredna slika 5.1 pokazuje dijagram koji opisuje način komunikacije preko kanala sa šumom.

Za početak ćemo definirati par pojmova koje ćemo koristiti u opisu ovih kodova:

s označava poruku koju nastojimo poslati

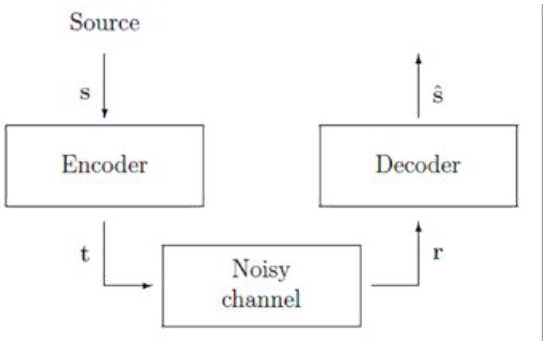
t označava poruku koju kodira koder i šalje kanlom

n označava niz bitova koji je svugdje nula osim na mjestima gdje kanal mijenja poruku koju šaljemo (p ili t)

r označava poruku koja izlazi iz kanala

$\hat{s}$ označava poruku koju dekodira dekoder

Slika 5.1: Komunikacija sustavom kodiranja/dekodiranja preko kanala sa šumom



Nadalje, pretpostavljamo kako naš kanal radi deformacije na bitovima sa vjerojatnošću 10%,tj. pri slanju ne kodirane poruke vrijedi $p_b = 0.1$.

5.1 Ponavljajući kodovi

Najjednostavniji način kodiranja koji možemo smisliti je taj da jednostavno svaki bit podatka koji želimo poslati repliciramo k-puta i takav kod pošaljemo kroz kanal. Promotrimo slučaj kada je $k = 3$ (takav kod označavamo oznakom R_3). Sada pretpostavimo kako želimo poslati poruku s gdje je

$$s = 0\,0\,1\,0\,1\,1\,0 \tag{5.1}$$

Sada primjenimo operaciju pretvorbe izvornog koda s u kod za slanje t na način da svaki bit ponovimo tri puta. Na slici 5.2

Slika 5.2: Pretvorba iz izvornog koda s informacije u kod za slanje t

Source sequence **s**	Transmitted sequence **t**
0	000
1	111

Sada našu poruku s kodiramo da dobijemo kod t, te taj kod šaljemo kroz kanal sa šumom u kom nastaju promjene na bitovima. Promjene su označene jedinicama u nizu bitova n te za rezultat dobivamo izlazni kod r koji se dekodira u poruku $\hat{s}$ koju interpretiramo. Na slici 5.3 su je ovaj proces vizualiziran.

Slika 5.3: Izgled bitova u procesu slanja poruke kanalom 5.1

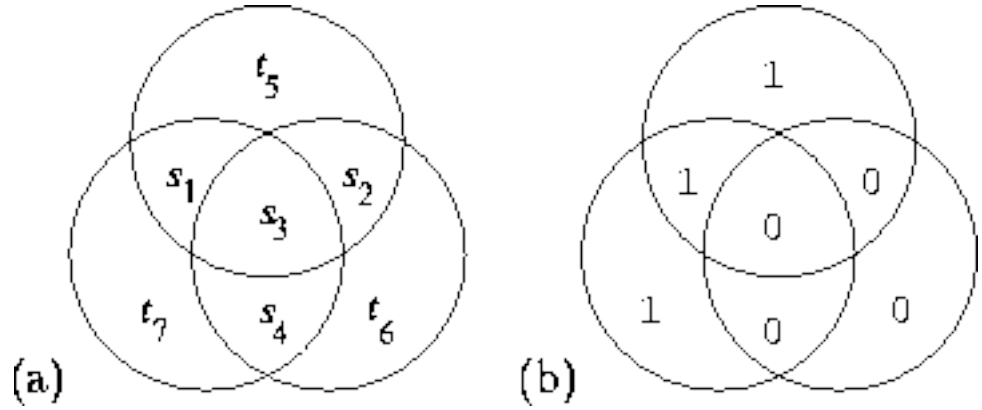
s	0	0	1	0	1	1	0
t	000	000	111	000	111	111	000
n	000	001	000	000	101	000	000
r	000	001	111	000	010	111	000
ŝ	0	0	1	0	0	1	0
corrected errors		⋆					
undetected errors					⋆		

Laganim računom pokažemo kako je u slučaju koda R_3 vjerojatnost greške bita svedena na $p_b \simeq 0.03$ no stopa transfera informacija $R = \frac{1}{3}$. Odmah uočavamo problem ovakog koda, jer ako se vratimo na 0.0.2 iz Uvoda vidimo kako bi ovakvim kodom za spremanje količine podataka od $1GB$ trebali tvrdi disk od $3GB$ što ovakav način kodiranja svodi na akademski primjer. Za detalje pogledati [1, str 8].

5.2 Hamming kodovi

Za razliku od prethodnog primjera koda naredni je malo *pametniji*. Hamming kodovi su poseban slučaj blok kodova, u oznaci (N, K)Hamming kod. Blok kod bi možda bolje definirali kao pravilo za konverziju podataka na način da se izvorni kod s, duljine K, prevodi u kod za slanje t duljine N bitova . Uočavamo da nas zanima slučaj $K < N$. Tako na svoj početni kod dodajemo $N - K$, bitova koji su linearna kombinacija ovih izvornih K bitova (oni nam služe kao provjera). Sljdeća slika 5.4 ilustrira primjer kodiranja pomoću $(7, 4)$Hamming koda. Na slici je shema spremanja i primjer jednog koda.

Slika 5.4: Primjer kodiranja korištenjem (7,4) Hamming koda



Koristeći sliku 5.4 sada nam je lakše objasniti postupak prevođenja izvorne posuke s u kod za slanje t. Izvornu poruku s (koja u ovom slučaju ima 4 bita) prevodimo tako što ta četiri bita poruke s prepisujemo u prva četiri bita t a ostala tri() formiramo na način:

t_4 = 0 ukoliko je $s_1 + s_2 + s_3$ paran broj, inače 1

t_5 = 0 ukoliko je $s_2 + s_3 + s_4$ paran broj, inače 1

t_6 = 0 ukoliko je $s_3 + s_4 + s_1$ paran broj, inače 1

Što vidimo na primjeru poruke $s = 1000$ koju kod prevodi u $t = 1000101$ Dekodiranje ovog koda je malo složenije no što je dekodiranje kod prethodne vrste koda. Zato što je čitav kod podložan utjecaju šuma pa može nastati greška i na dijelu koda za provjeru isto kao i na dijelu koda koji prenosi informaciju. Detaljno o dekodiranju ovog koda možete naći u [1, str.9]

5.3 Izvodi

Entropija zadovoljava svojstvo rekurzivnosti koje nam može jako pomoći pri računanju entropije. Možda bolje prvo na jednostavnom primjeru. Uzmimo kao primjer slučajnu varijablu X koja postiže vrijesnost $x \in \{0, 1, 2\}$, koja se postiže na način da bacamo simtrični novčić (BSO pretpostaviti ćemo kako su na novčiću na jednoj strani broj 0 a na drugoj broj 1) i očitamo što je palo. Ukoliko se na novčiću pojavi 0 tada je $x = 0$, a ako se pojavi broj 1 novčić bacamo ponovo te ovisno o ishodu zapišemo 1 ili 2.

$$P(X = 0) = \frac{1}{2}; \qquad P(X = 1) = \frac{1}{4}; \qquad P(X = 2) = \frac{1}{4} \tag{5.2}$$

Sada nas zanima kolika je entropija od X Izračunajmo ju preko formule:

$$H(X) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4, \tag{5.3}$$

$$= 1.5 bitova; \tag{5.4}$$

Ali pogledajmo sada alternativni pristup pomoću rekurzije. Koristimo dekompoziciju u kojoj postupno otkrivamo informaciju o X-u. Promotrimo prvo *učenje o varijabli* u kojem nastojimo saznati je li $x = 0$ a zatim *učimo* o kojoj ne-nul vrijednosti se radi. Sazanje o prvom *učenju* možemo opisati binarnom slučajnom varijablom s vjerojatnostima $1/2, 1/2$. Ovo *otkrivanje* varijable ima entropiju $H1 = \frac{1}{2}\log 2 + \frac{1}{2}\log 2 = 1$ bit. Ukoliko ne *otkrijemo* nula, onda *otkrivamo* nenul vrijednost koja ima istu distribuciju dakle $H2 = 1$ bit. No ovu drugu informaciju ćemo dobiti samo u pola bacanja stoga možemo zapisati entropiju od X kao:

$$H(X) = H1 + \frac{1}{2}H2 = 1.5 \tag{5.5}$$

Ovaj proces sada generaliziramo primjenjujući je na vjerojatnosnu distribuciju $p = p_1, p_2, \ldots, p_I$. Pošto je vrijednost entropije definirana vjerojatnosnom distribucijom opravdana je notacija $H(Y) = H(p_Y)$ koju koristimo u nastavku.

$$H(p) = H(p_1, 1 - p_1) + (1 - p_1)H(\frac{p_2}{1 - p_1}, \frac{p_3}{1 - p_1}, \ldots, \frac{p_I}{1 - p_1}) \tag{5.6}$$

Iako se možda formula ne čini jako zgodno za upotrebu, ipak je jako korisna jer pri računanju entropije računalom drastično ubrzava i olakšava proces.

Pokažimo aditivnost entropije para ansambla, kada imamo $P(x, y) = P(x)P(y)$

$$H(X,Y) = \sum_{xy\in\mathcal{A}_X\mathcal{A}_Y} P(x,y)\log\frac{1}{P(x,y)} \tag{5.7}$$

$$= \sum_{x\in\mathcal{A}_X, y\in\mathcal{A}_Y} P(x)P(y)\log\frac{1}{P(x)P(y)} \tag{5.8}$$

$$= \sum_{x\in\mathcal{A}_X, y\in\mathcal{A}_Y} P(x)P(y)\log\frac{1}{P(x)} + \log\frac{1}{P(y)} \tag{5.9}$$

$$= \sum_{xy\in\mathcal{A}_X\mathcal{A}_Y} P(x)P(y)\log\frac{1}{P(x)} + \sum_{xy\in\mathcal{A}_X\mathcal{A}_Y} P(x)P(y)\log\frac{1}{P(y)} \tag{5.10}$$

$$= \sum_{x\in\mathcal{A}_X} P(x)\log\frac{1}{P(x)} + \sum_{y\in\mathcal{A}_Y} P(y)\log\frac{1}{P(y)} \tag{5.11}$$

$$= H(X) + H(Y) \tag{5.12}$$

Nadalje pokažimo kako vrijedi i $I(X;Y) = I(Y;X)$. Koristeći definiciju 2.1.3

$$H(X,Y) = \sum_{xy\in\mathcal{A}_X\mathcal{A}_Y} P(x,y)\log\frac{1}{P(x,y)} \tag{5.13}$$

$$= \sum_{x\in\mathcal{A}_X, y\in\mathcal{A}_Y} P(x)P(y|x)\log\frac{1}{P(x)P(y|x)} \tag{5.14}$$

$$= \sum_{x\in\mathcal{A}_X} P(x)\log\frac{1}{P(x)} + \sum_{xy\in\mathcal{A}_X\mathcal{A}_Y} P(y,x)\log\frac{1}{P(y|x)} \tag{5.15}$$

$$= H(X) + H(Y|X) \tag{5.16}$$

Analogno se pokaže vrijedi i $H(X,Y) = H(Y)+H(X|Y)$. Sada znamo kako vrijedi jednakost

$$H(X) + H(Y|X) = H(Y) + H(X|Y) \tag{5.17}$$

Te jednosavnom promjenom strana pokažemo kako vrijedi

$$H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{5.18}$$

Čime smo pokazali danu jednakost.

Bibliografija

[1] David J.C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2005.

[2] Karl Petersen, *Ergodic theory*, Cambridge University Press, 1981.

[3] Nikola Sarapa, *Teorija Vjerojatnosti*, Udžbenici Sveučilišta u Zagrebu, Školska knjiga, 2002.

[4] Joy A. Thomas Thomas M.Cover, *Elements of Information Theory*, John Wiley & sons, inc., 2006.

Sažetak

Teorija informacija nudi nam jedan alternativni pogled na problem komunciranja preko kanala koji imaju određene vrste problema, dakle kvare izvornu poruku koje se njima pokušava poslati. Cilj ovog rada je bio opisati mogućnosti komunikacije preko takvih kanala uvođenjem neki novih pogleda na način komuniciranja kanalom.

Teorem o kanalu sa šumom je glavni rezultat koji smo opisali u ovme radu. On nam određuje kako kapacitet kanala određuje mjeru informacija koju kanal može prenijeti, što na početku nije toliko očito. Kapacitet kanala u stvari određuje stopu transfera informacija kojom je moguće komunicirati kanalom uz relativno mali utjecaj greške koja proizlazi iz šuma kanala. Nadalje, teorem određuje i funkcijksu povezanost između vjerojatnosti greške bita ,koju želimo minimizirati, te stope transfera informacija ,koju želimo maksimiziati. Točnije, definira funkciju koja za danu razinu vjerojatnosti greške bita (te poznavanje kapaciteta kanala) odeđuje stopu transfera informacija pri kojoj je komunikacija moguća.

Summary

Information theory gives us one alternative look on communication over information channels which have a sort of disturbance, they usually disrupt part of source message. Goal of this work was to describe possibility's of communication over this types of channels by introduction new looks on channel communicating.

Noisy channel-coding theorem is the main result which is described in this work. It gives us the way in which channel capacity defines measure of information that can be transmitted trough channel, which is not obvious. Channel capacity actually defines transfer rate of information by which it is possible to communicate over the noisy channel causing neglect error. Further, theorem determines function relation between probability of bit error ,we want to reduce, and the transfer rate of information, we want to maximize. Precisely, theorem gives us a function which for given value of bit error probability (knowing channel capacity) defines information transfer rate at which is the communication possible.

Životopis

Rođen sam 06.listopada 1990. u Mostaru. Osnovnu školu sam završio u Širokom Brijegu te u istom gradu upisao prirodoslovno-matematički smjer u gimnaziji fra Dominika Mandića. Maturiaro sam 2009.godine. Logičan slijed je bio upisivanje Preddiplomskog sveučilišnog studija Matematike na Matematičkom odsjeku Prirodoslovno-matematičkog fakulteta u Zagrebu. Godine 2012. sam u pisao Diplomski sveučilišni studij Financijske i poslovne matematike. Od strane fakulteta sam nagrađen za iznimna postignjuća u izvannastavnim aktivnostima.