5.1 Markovljevi izvori I reda

Kod izvora I reda broj simbola (n) jednak je broju stanja i stacionarne vjerovatnoće simbola jednake su stacionarnim vjerovatnoćama stanja. Ovaj izvor I reda matematički je opisan listom simbola $S = \{S_1, S_2 ... S_n\}$ i skupom *uslovnih vjerovatnoća* emitovanja jednog simbola kada se zna prethodno emitovani simbol, tj: $P(S_j | S_i)$ $(j = 1,2....n)$ $(i = 1,2...n)$.

Same uslovne vjerovatnoće se vrlo često upisuju u okviru jedne matrice, ***tranzicione matrice.*** U opštem slučaju tranziciona matrica ima oblik:

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ . & & & \\ . & & & \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix}$$

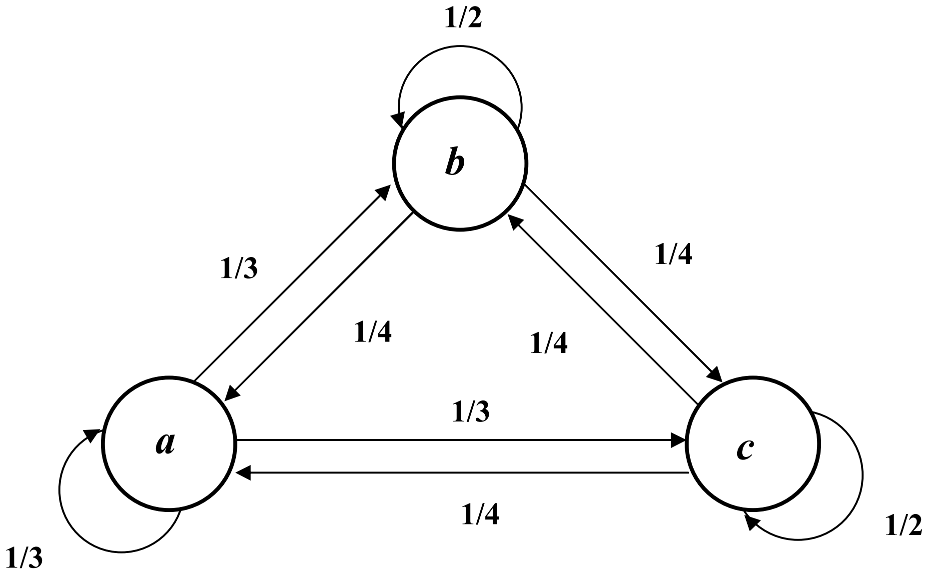
gdje su P_{ij} vjerovatnoće prelaska iz stanja S_i u stanje S_j. Zbir elemenata u svakoj vrsti (ili koloni) ove matrice mora da bude jednak jedinici: $\sum_{j=1}^{n} P_{ij} = 1 \quad (i = 1,2...n)$. Matrice čiji su elementi nenegativni i zadovoljavaju prethodne relacije zovu se *stohastičke matrice*.

Primjer:

Neka je zadat skup simbola koji emituje Markovljev izvor I reda $S = \{a, b, c\}$ i neka su date uslovne vjerovatnoće (broj uslovnih vjerovatnoća je $n^{m+1} = 3^{1+1} = 9$).

$$P(a|a) = \frac{1}{3}; P(b|b) = \frac{1}{2}; P(c|c) = \frac{1}{2}$$
$$P(b|a) = \frac{1}{3}; P(a|b) = \frac{1}{4}; P(a|c) = \frac{1}{4}$$
$$P(c|a) = \frac{1}{3}; P(c|b) = \frac{1}{4}; P(b|c) = \frac{1}{4}$$

Dijagram stanja je:

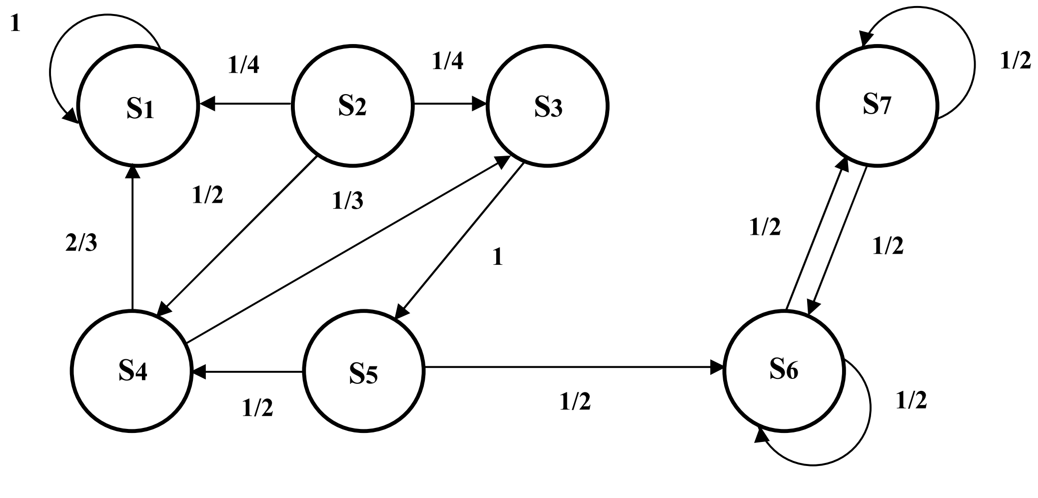


Matrica tranzicije je:

$$P = \begin{bmatrix} P(a|a) & P(a|b) & P(a|c) \\ P(b|a) & P(b|b) & P(b|c) \\ P(c|a) & P(c|b) & P(c|c) \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

Ako elementi tranzicione matrice ne zavise od vremena (to je jedan od uslova da imamo ergodičan ili stacionaran proces) tada je ovaj izvor ***homogen.***

Radi lakšeg objašnjenja pojmova koji će biti uvedeni na sledećoj slici prikazan je dijagram stanja jednog Markovljevog slučajnog procesa.



Za stanje S_i kažemo da je ***tranzijentno*** (nepovratno) ako iz bar jednog stanja u koje se može doći iz S_i više se nikad ne može vratiti u S_i. Tranzijentna stanja na slici su: S_2, S_3, S_4, S_5.

Stanje S_i je ***rekurentno*** (povratno) ako iz svakog stanja u koje se može doći iz S_i može se vratiti u S_i. Rekurentna stanja na slici su: S_1, S_6, S_7.

Stanje S_i je ***periodično*** ako postoji cio broj d ($d > 1$) takav da se poslije napuštanja toga stanja izvor može vratiti u isto stanje poslije kd koraka. Periodična stanja na slici su: S_3, S_4, S_5.

Stanje S_i je ***apsorbujuće*** ako izvor po dolasku u to stanje više ne može se da ga napušti. Apsorbujuće stanje na slici je: S_1.

Markovljev proces je ***regularan*** ako za neko konačno k_o matrica P^{k_o} nema nultih elemenata. Tada i sve matrice P^k $(k > k_o)$ neće imati nultih elemenata. Naravno, ako već tranziciona matrica P nema nijedan nulti elemenat, tada se već u prvom koraku može doći iz svakog stanja u svako stanje (ergodičan proces).

Zaključak: Markovljev proces je ergodičan ako matrica P^k (k je konačan broj) ima bar jednu nenultu kolonu.

Ovakvi procesi imaju stacionarne vjerovatnoće stanja (kao što smo već ranije računali) do kojih se može doći rješavanjem matrične jednačine:

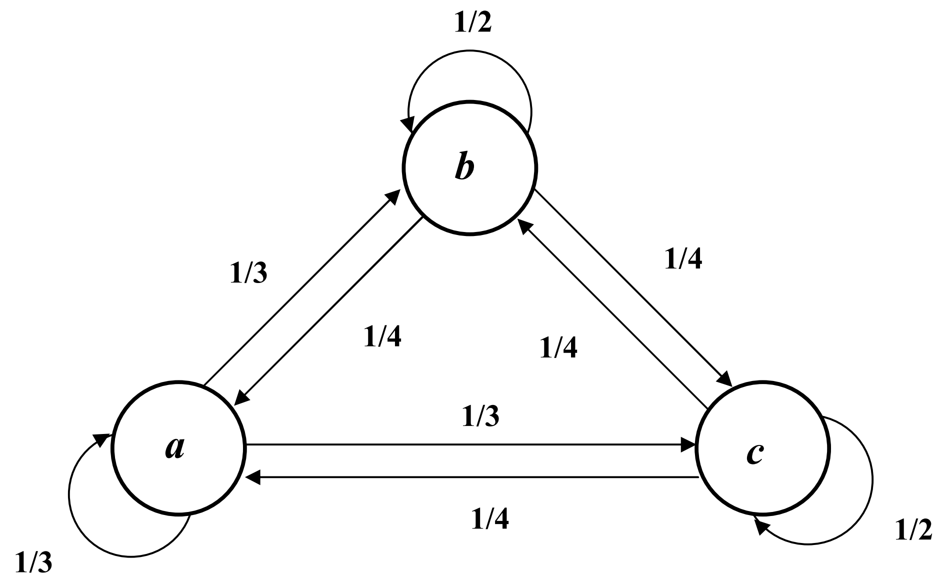
$$p^T = P\,p^T$$

gdje je p matrica dimenzije $1 \times n$ čiji su elementi jednaki vjerovatnoćama pojavljivanja pojedinih simbola $p = \left[P(a)\ P(b)\ P(c)\right]$. Ako je p^o matrica koja sadrži vjerovatnoće stanja u početnom trenutku, tada se tekuće vjerovatnoće stanja (poslije M trenutaka) mogu dobiti kao:

$$p^M = P\,p^o$$

Primjer:

Neka je dat dijagram stanja Markovljevog izvora I reda i neka su vjerovatnoće početnih stanja $P(a) = \frac{1}{2}$ $P(b) = \frac{1}{4}$ $P(c) = \frac{1}{4}$. Odrediti vjerovatnoće u stacionarnom stanju.



$$\begin{bmatrix} P(a) \\ P(b) \\ P(c) \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{14}{48} \\ \frac{17}{48} \\ \frac{17}{48} \end{bmatrix}$$

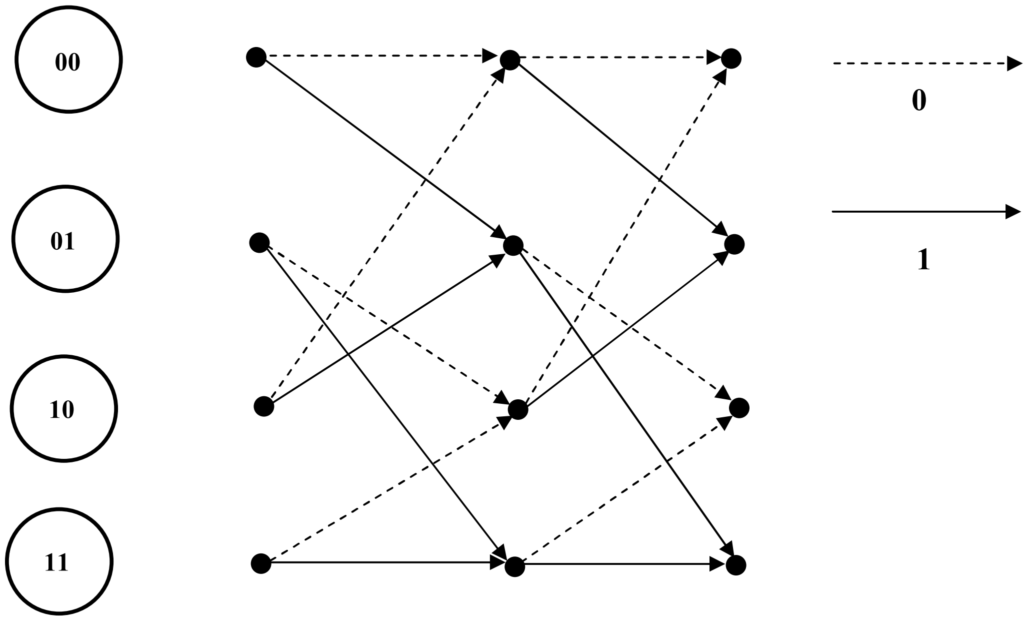
Provjera: $P(a) + P(b) + P(c) = \frac{14}{48} + \frac{17}{48} + \frac{17}{48} = 1$

5.2 Dinamički dijagram stanja-trelis

Koncept dijagrama stanja pogodan je za opšti uvid u ponašanje izvora. Međutim, ako bi se pratilo konkretno ponašanje izvora, tj. njegov "put" kroz dijagram stanja, vrlo brzo bi situacija bila nepregledna. Zbog toga je puno pogodnije posmatrati neku vrstu "dinamičkog dijagrama stanja", tkz. ***trelis*** (eng. rešetka). U okviru crtanja trelisa za svaki korak (emitovani simbol) crta se novi dijagram stanja, a u njemu se označavaju odgovarajući prelazi. Uz ove prelaze mogu se prikazati simboli koje emituje izvor, a mogu se dati i odgovarajuće uslovne vjerovatnoće.

Primjer:

Nacrtati trelis za binarni izvor drugog reda ($n^m = 2^2 = 4$ stanja).



Sa slike se uočava da je trelis periodična struktura i da ga je dovoljno definisati za samo jedan korak, a zatim se cijela slika periodično ponavlja. Gornja putanja se može nazvati "putanjom svih nula", a donja "putanjom svih jedinica". Trelisom se mogu predstaviti i neergodični izvori, ali kod njih trelis neće imati periodičnu strukturu.

5.3 Entropija izvora sa memorijom

Neka je dat izvor sa memorijom *m-tog* reda sa *n* simbola $S = \{S_1, S_2, ..., S_n\}$ i odgovarajućim skupom uslovnih vjerovatnoća emitovanja jednog simbola: $P(S_j | S_{i1}, S_{i2}...S_{ik}...S_{im})$ $(j = 1,2....n;\ i = 1,2....n;\ k = 1,2....m)$. Već ranije smo rekli da ovaj izvor ima n^m stanja. Srednja količina informacija koju emituje izvor po simbolu u posmatranom stanju zove se ***parcijalna entropija*** i iznosi:

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

U slučaju Markovljevog izvora I reda:

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

Entropija izvora se dobija usrednjavanjem parcijalne entropije po svim mogućim stanjima. Njena vrijednost za slučaj Markovljevog izvora I reda data je formulom:

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

5.4 Tipične sekvence i asimptotska ekviparticiona osobina

Tipične sekvence su sekvence čija vjerovatnoća ima sledeću osobinu:

$$2^{-n(H(S)+\varepsilon)} \le P(S_1, S_2 ... S_n) \le 2^{-n(H(S)-\varepsilon)}$$

gdje je n - dužina sekvence koju emituje izvor.

Sa porastom dužine sekvence n sve tipične sekvence su skoro jednako vjerovatne i njihova vjerovatnoća je $2^{-nH(S)}$, dok je njihov broj približno jednak $2^{nH(S)}$. Prema tome, skup svih mogućih sekvenci dužine n koje emituje neki ergodični izvor može se podijeliti na dva podskupa: skup tipičnih sekvenci i skup ostalih sekvenci. ***Kada n neograničeno raste, vjerovatnoća da će emitovana sekvenca pripadati skupu tipičnih sekvenci teži jedinici.*** Ova osobina je veoma važna za statističko kodovanje (kodovanje izvora). Osobina važi kako za diskretne izvore bez memorije, tako i za diskretne izvore sa memorijom.

Primjer:

Binarni izvor ($S = \{0,1\}$) emituje poruku od 10 bita. Vjerovatnoća emitovanja 1 je $P(1) = p = 0.3$, a vjerovatnoća emitovanja 0 je $P(0) = q = 0.7$. Odrediti tipičnu sekvencu ovog izvora.

sekvenca sadrži	**broj kombinacija**	**vjerovatnoća pojedinačne kombinacije**	**totalna vjerovatnoća (br.kom x vjer. poj. kom.)**
0 jedinica	1	$q^{10} = 0.0282$	0.0282
1 jedinica	10	$q^9 p = 0.0121$	0.121
2 jedinice	45	$q^8 p^2 = 0.005$	0.225

3 jedinice	120	$q^7 p^3 = 0.0022$	0.264
4 jedinice	210	$q^6 p^4$	0.1995
5 jedinica	252	$q^5 p^5$	0.1008
6 jedinica	210	$q^4 p^6$	0.036
7 jedinica	120	$q^3 p^7$	0.009
8 jedinica	45	$q^2 p^8$	0.00144
9 jedinica	10	$q^1 p^9$	0.000138
10 jedinica	1	p^{10}	0.0000059

Broj kombinacija se računa $\binom{n}{k} = \frac{n\cdot(n-1)...\cdot(n-k+1)}{1\cdot 2\cdot 3...\cdot k}$, gdje je n - dužina sekvence i k - broj jedinica u sekvenci.

Entropija pojedinačnog bita je:

$$H(S) = -p\log p - (1-p)\log(1-p) = 0.8813 bit/simb$$

Pošto se sekvenca sastoji od 10 bita entropija proširenog izvora je:

$$n\cdot H(S) = 10\cdot 0.8813 bit/simb = 8.813 bit/simb$$

Vjerovatnoća pojavljivanja tipične sekvence približno je jednaka:

$$2^{-n\cdot H(S)} = 2^{-8.813} = 0.00223$$

Iz tabele slijedi da je to sekvenca koja sadrži 3 jedinice (ima ih 120). Obratiti pažnju da je totalna vjerovatnoća ovih sekvenci najveća.

6. Statističko kodovanje

Zadatak statističkog kodovanja, tj. kodovanja izvora, je da informacije koje emituje izvor ekonomično predstavi (koduje) simbolima koji će ih prenositi kroz telekomunikacioni kanal. Određeni broj autora ovaj postupak naziva *kompresija podataka*.

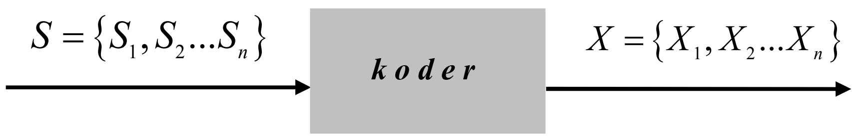
Postoje dvije mogućnosti pri kodovanju simbola izvora:

- prenijeti sve informacije koje izvor emituje

- prenijeti samo jedan njihov dio

Kod diskretnih izvora mogu se primijeniti oba tipa kompresije. Kodovi dobijeni statističkim kodovanjem su pogodni za "kanale bez šumova", gdje nema potrebe za kodovima za kontrolu grešaka. U opštem slučaju, prvo se izvrši statističko kodovanje, a zatim se dodaju redudantni biti (simboli) koji služe za otkrivanje i eventualno ispravljanje greški.

6.1 Definicija koda



Blok šema kodera

Posmatrajmo diskretni izvor sa brojem simbola *n*.

Lista izvornih simbola (S) se još naziva *alfabet (azbuka) izvora*.

Lista kodnih simbola (X)- kodna lista je takođe konačna.

Definicija koda bi mogla da bude sledeća:

Kod (kodovanje) je preslikavanje sekvenci simbola liste izvora u sekvence simbola kodne liste.

Blok kod je preslikavanje simbola liste izvora u sekvence simbola kodne liste. Ove sekvence se zovu ***kodne riječi***. Broj kodnih simbola u kodnoj riječi naziva se ***dužina kodne riječi.*** Postoje:

- blok kodovi sa fiksnom dužinom kodne riječi
- blok kodovi sa promjenljivom dužinom kodne riječi

Primjer:

Neka je dat binarni kod za izvor sa *n=4* simbola.

S	***X_i***
S_1	0
S_2	01

S_3	11
S_4	01

Kod je ***nesingularan*** ako su sve kodne riječi međusobno različite. Kod iz primjera je singularan, jer su S_2 i S_4 jednako kodovani.

Primjer:

Nesingularan kod dobijen je iz prethodnog primjera zamjenom kodne riječi 01 kodnom riječi 00 za simbol S_4.

S	***Xi***
S_1	0
S_2	01
S_3	11
S_4	00

Međutim, ni ovo nije dovoljno za praktičan rad, jer ako je primljena sekvenca 0011 dekoder ima na raspolaganju dvije sekvence izvornih simbola koje mogu dati datu sekvencu. To su $S_1S_1S_3$ i S_4S_3. Na osnovu ovoga zaključujemo da kod mora biti ***jednoznačno dekodibilan***.

Primjer:

Sledeći kodovi su jednoznačno dekodibilni:

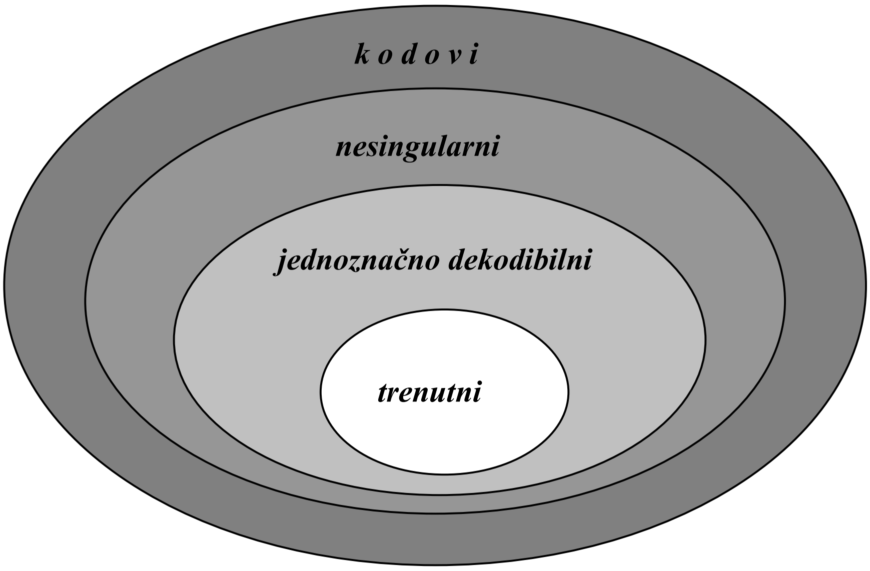
S	***kod(a)***	***kod(b)***	***kod(c)***
S_1	00	0	0
S_2	01	10	01
S_3	10	110	011
S_4	11	1110	0111

Kod (a) ima kodne riječi jednake dužine i nesingularan je, pa je i jednoznačno dekodibilan, jer se na prijemu prihvataju prispjeli kodni simboli u parovima.

Kod (b) je nesingularan i 0 kod njega igra ulogu zareza, jer pomaže da se tačno odredi završetak kodne riječi.

Kod (c) je takođe dekodibilan, ali se kodna riječ ne može dekodovati dok ne stigne sledeći simbol.

Jednoznačno dekodibilni kod je ***trenutan*** ako se svaka kodna riječ u sekvenci može dekodovati bez oslanjanja na sledeće kodne simbole.



GRAYOV KOD

Grayov kod pripada klasi specijalnih kodova. Grayov kod se dobija tako što prvi bit ostaje isti kao kod standardnog koda, a svaki sledeći se dobija kao $a_i \oplus a_{i-1}$.

Standardni kod	0000	0001	0010	0011	0100	0101
Grayov kod	0000	0001	0011	0010	0110	0111
Standardni kod	0110	0111	1000	1001	1010	1011
Grayov kod	0101	0100	1100	1101	1111	1110
Standardni kod	1100	1101	1110	1111		
Grayov kod	1010	1011	1001	1000		