Dvejetainiai grupiniai kodai

Įvadas

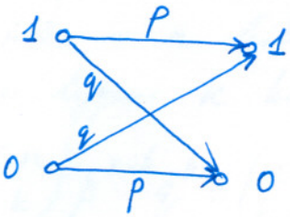
Ši tema skirta vienam informacijos perdavimo problemos sprendimo metodui: dvejetainiam kodavimui ir dekodavimui, garantuojančiam patikimą informacijos perdavimą kanalais, kuriuose informacija iškraipoma „užtriukšminama". Tipiška tokia situacija: mes norime perduoti pranešimą, kuris gali būti tam tikro baigtinio alfabeto eilutė. Alfabetu gali būti aibė $\{0,1\}$, lotynų abėcėlės didžiosios, dešimtainiai (arabiški) skaitmenys ir t.t. Pavyzdžiui, pranešimu gali būti koks nors tekstas anglų ar lietuvių kalba (tokiu atveju prie alfabeto simbolių reikėtų prijungti tarpą ir skyrybos ženklus. Pranešimais gali būti taip pat baigtinės dvejetainės simbolių sekos, pavyzdžiui, perduodant informaciją iš vieno kompiuterio į kitą ar perduodant telemetrinę informaciją (tarkim, iš kosmoso).

Vienaip ar kitaip duomenų perdavimas sutinka ėmimą tam tikro alfabeto simbolių perdavimą ryšio kanalu. Praktiškai ryšio kanalas niekada nebūna idealus, t.y. su tam tikra nenuline tikimybe q perduodamas simbolis bus priimtas neteisingai. Jei perduodami signalai ilgi, tai net ir maža klaidos tikimybė vienam simboliui, tarkime, $q = 10^{-8}$ gali būti nepriimtina. Taip yra, pavyzdžiui, kompiuterinėse tinklose.

Pavyzdžiui, vienas kompiuteris su kitu gali būti susietas palydovinio ryšio. Tokiu atveju dažniausiai naudojamas dvejetainis alfabetas $\{0,1\}$. Ryšio kanalas fiziškai realizuojamas per elektromagnetinį lauką tarp žemės paviršiaus ir palydovo. Elektromagnetiniai signalai, atitinkantys 0 ir 1, sąveikaudami su išorinio elektromagnetinio lauko, gali neatpažįstamai pasikeisti (dėl Saulės dėmių, atmosferos trikdžių ir pan.).

Dvejetainiai simetriniai kanalai:

Tegul dvejetainiai signalai 0, 1 moksliai perduodami ryšio kanalu į imtuvą.



1 pav. Perėjimo tikimybės dvejetainiame simetriniame kanale

1 pav. pavaizduota situacija, kada kiekvienas simbolis priimamas teisingai su tikimybe p ir klaidingai su tikimybe $q=1-p$. Be to daroma prielaida, kad klaidos ryšio kanale perduodant simbolius įvyksta nepriklausomai skkančio apibrėžimo prasme

Apibrėžimas. Tegul $E_1, E_2, E_3, \ldots$ – bandymų seka, o T – įvykis, kuris gali įvykti arba neįvykti atliekant bandymą E_i. Pažymėsime p_i įvykio T tikimybę, o $q_i = 1 - p_i$ – tikimybę, kad T neįvyks. Sakysime, kad bandymai E_i yra nepriklausomi atžvilgiu įvykio T, jei bet kuriems bandymų aibės poaibiams I ir J tikimybė to, kad kai $E_i \in I$ įvyks T įvyks, o kai $E_j \in J$ neįvyks, yra lygi $\prod_I p_i \prod_J q_j$.

Pavyzdžiui, tarkime, kad vieno simbolio 0 arba 1 (vieno informacijos bito) klaidingo perdavimo tikimybė $q = 1\% = 0.01$ ir mes norime tiksliai perduoti seką iš 10 000 simbolių. Tada tiesiogiai perduodant simbolį po simbolio tokia seka teisingai bus priimta su labai maža tikimybe:

$$P_0 = (1-0.01)^{10000} \approx 10^{-4.4} < 0.004\%$$

Čia mes pritaikime Bernulio formulę. Jos bendras pavidalas yra toks:

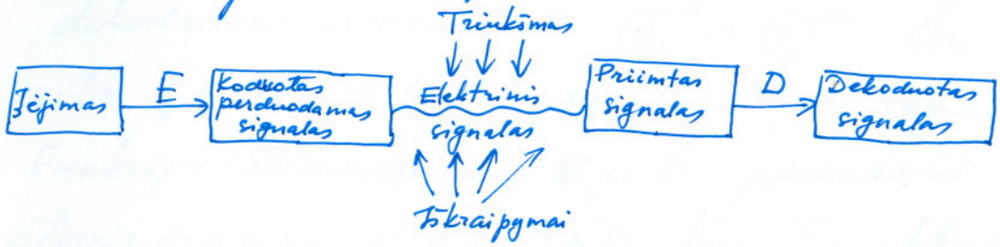
Tegul $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\dots(n-k+1)}{1\cdot 2\cdot 3\dots k}$

Teorema. Tegul simetriniu dvejetainiu kanalu perduodama seka is n leitų. Tikimybė to, kad ji bus priimta su lygiai k klaidų, yra lygi

(1) $$P_k = \binom{n}{k} p^{n-k} q^k = \binom{n}{k} p^{n-k} (1-p)^k = \binom{n}{k} (1-q)^{n-k} q^k$$

Didžiųjų skaičių dėsnis teigia, kad kai n didelis, klaidingai perduotų simbolių dalis bus artima q. Pavyzdžiui, kai $p = 0.99$, tai iš 10000 simbolių seko maždaug 100 bus priimta klaidingai ir beveik tikrai klaidų skaičius bus tarp 50 ir 100.

Bendra skaitmeninės informacijos perdavimo schema pateikta 2 pav.



2 pav. Ryšio kanalo schema

Kiekviena ryšio linija turi natūralius apribojimus: siųstuvo galia ribota, signalo slopinimas kanale, trikdžiai. Jeigu iškraipymai dideli, imtuvas priims informaciją klaidingai. Tenka arba susitaikyti su tam tikra klaidų dalimi arba sugalvoti būdus, kaip tą dalį sumažinti.

Reikia pastebėti, kad daugelio atvejų (pavyzdžiui, ryšiuose su kosminiais aparatais) klaidos gali labai brangiai kainuoti.

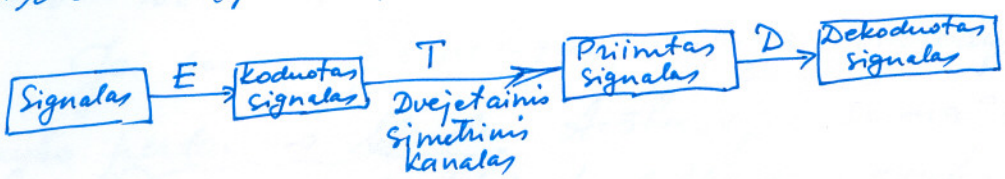
Kodavimas ir dekodavimas

Nagrinėsime taip vadinamus sisteminius kodus. Simbolių seka, kurią reikia perduoti, kodavimo įrenginyje užrašoma tų pačių simbolių sekomis (paprastai 0 ir 1) pagal tam tikrą kodavimo schemą. Imtuvas tada sugeba atpažinti ir/arba ištaisyti klaidas, atsirandančias dėl trikdžių, – tai įvykdoma analizuojant papildomą informaciją, esančią papildomuose simboliuose. Priimama ilga seka dekoduojama pagal dekodavimo schemą, t.y., grįžtama prie tos sekos, kurią iš tikrųjų reikėjo perduoti.

Apibrėžimas. Dvejetainių (m, n)-kodu vadinama pora, susidedanti iš kodavimo schemos

$$E : ②^m \to ②^n$$ ir

dekodavimo schemos $D : ②^n \to ②^m$. Čia ② – tai aibė $\{0, 1\}$, o $②^n$ – visų dvejetainių ilgio n sekų aibė. Funkcijos (atvaizdavimai) E ir D parenkami taip, kad atvaizdavimas $H = E \circ T \circ D$, kur T – „klaidų funkcija", su tikimybe artima vienetui būtų tapatingas. Kadangi kodavimas ir dekodavimas vykdomas kontroliuojamuose sąlygose, galime laikyti, kad tos operacijos atliekamos be klaidų. Tokiu būdu ryšio sistemos matematinį modelį galima pavaizduoti tokia blokine schema (pav. 3.), kur E ir D – determinuotos funkcijos.



3 pav. Ryšio sistemos modelis

Kodai dalinami į dvi dideles klases. Klaidas taisantys kodai skirti tam, kad su tikimybe, artima 1 atstatytų (rekonstruotų) pasiųstą signalą. Klaidas aptinkantys kodai turi tikslą su artima vienetui tikimybe aptikti klaidas. Iliustracijai patikrinime poza pavyzdžių.

Pvz. 1. Paprastas klaidas aptinkantis kodas paremtas lyginumo tikrinimo schema. Ji taikoma signalui $a = (a_1, a_2, \dots a_m) = a_1 a_2 \dots a_m$, t.y. bet kurio baigtinio m ilgio signalui. Kodavimo schema apibrėžiama taip:

$$E : (a_1, a_2, \dots a_m) = a_1 a_2 \dots a_m = a \to b = b_1 b_2 \dots b_{m+1},$$

čia

(2) $b_i = a_i$, kai $i = 1, \dots, m$

(2') $b_{m+1} = \begin{cases} 0, \text{ jeigu } \sum_{i=1}^{m} a_i \text{ lyginis} \\ 1, \text{ jeigu } \sum_{i=1}^{m} a_i \text{ nelyginis} \end{cases}$

Tarkime, kai $m = 2$, E apibrėžiamas prisilyginais: $00 \to 000$, $01 \to 011$, $10 \to 101$, $11 \to 110$. Iš apibrėžimų (2) ir (2') matyti, kad kiekvienu atveju iškoduoto signalo $b = Ea$ skaitmenų suma bus lyginė.

Atitinkama dekodavimo schema tokia:

$$D : b \to c,$$

(2'') kur $b_i = c_i$, kai $i = 1, \dots, m$.

Jei suma $\sum_{i=1}^{m+1} b_i$ nelyginė, intuityvus nurodys, kad tuo atveju įvyko perdavimo klaida. Aišku, kai suma $\sum b_i$ lyginė, mes negalime būti tikri, kad klaida neįvyko.

Tarkime, kai $m=2$ ir klaidos tikimybė q, klaidingai priimtų signalų dalis bus

$$q^3+3q^2p+3pq^2 \quad (\text{trys, dvi arba viena klaida}).$$

To jų, taikant šį schemą, liks nepastebėtos klaidos lygiai dviejuose simboliuose. Todėl nepastebėtų klaidų dalis visoje klaidų aibėje bus lygi

$$\frac{3q^2p}{q^3+3q^2p+3qp^2} < \frac{q}{q+p} < q$$

Taigi, klaidos praleidimo tikimybė bus $< q$.

Dabar detaliai panagrinėsime klaidas taisantį kodą. Paprasčiausias tokio kodo pavyzdys – tai tiesiog signalo kartojimas. Šis būdas neefektyvus (yra geresnių).

Pvz 2. Tarkime, turime anksčiau aprašytą dvejetainės informacijos perdavimo kanalą. Nagrinėsime $(m, 3m)$ kodą su trigubu kartojimu. Tuo tikslu bet kuris signalas suskaidomas į blokus po m simbolių kiekviename ir kiekvienas blokas perduodamas tris kartus: taip apibrėžiama funkcija E. Funkciją D atrodo taip: priimtas signalas skaidomas į blokus, kurių ilgiai $3m$. Jeigu eilinis blokas susideda iš trijų vienodų ilgio m signalų, tai toks signalas ir yra dekodavimo rezultatas. Bendru atveju pagal simbolių trejetę c_i, c_{i+m}, c_{i+2m} tame bloke atstatomas simbolis, dažniausiai (du ar tris kartus) sutinkamas tame trejete ir užrašomas į i-tąjį vietoj dekoduotame bloke.

Tikimybė įvykio, kad simbolis dvejetoje pozicijoje bus teisingai priimtas lygi p^3. Vienos klaidos tikimybė yra $3p^2q$. Todėl teisingo priėmimo tikimybė (simbolio esančio dvejetoje pozicijoje) yra lygi p^3+3p^2q, o klaidos tikimybė yra $3pq^2+q^3$. Tarkime, $q=0.1$. Tada kitoje pozicijoje simbolis bus priimtas teisingai su tikimybe 0.729 ir dukart teisingai su tikimybe 0.243. Jis bus priimtas dukart neteisingai su tikimybe 0.027 ir trisdart neteisingai su tikimybe 0.001. Tokiu būdu, mišrus kodas sumažina klaidos perdavimo tikimybę nuo 10% iki $\approx 2.8\%$.

Jei signalą kartosime penkiskart ir dekoduosime vėl „pagal balsų daugumą", tai klaidos tikimybė

$$q^5+5q^4p+10q^3p^2=0.00856,$$

t.y. bus mažesnė 1%.

Jei perduosime signalą ilgio 10 (10 bitų signalą), tai teisingo perdavimo tikimybė nekartojant bus $(0.9)^{10}\approx 35\%$, kartojant tris kartus $\approx 74\%$, kartojant 5 kartus $\approx 91{,}5\%$.

Trigubas kartojimas tris kartus pailgina perdavimo laiką. Tolesnė parodysime, kad vieną klaidą su tokiu pat patikimumu ištaiso (3, 6) kodas, kuris perdavimo laiką tik padvigubina.

Blokiniai kodai

Aukščiau aprašyti pavyzdžiai priklauso taip vadinamų blokinių kodų klasei. Pagal apibrėžimą blokinis kodas keičia kiekvieną bloką iš m simbolių tam tikru ilgesniu bloku iš n simbolių, kurį po perdavimo reikia dekoduoti. (Kartais taikomi ir „nuoseklieji" kodai, kada perduodamo signalo simboliai „persiriti kontroliniais ir eilinio simbolio reikšmė priklauso nuo viso prieš tai buvusio signalo fragmento).

Daugumą ryšio sistemų konstruojama dvejetainių signalų perdavimui. Kaip jau buvo minėta, blokinis (m, n) kodas apibrėžiamas dviem funkcijomis:

(3) $E : (2)^m \to (2)^n$, $D : (2)^n \to (2)^m$, $m \le n$.

Turi būti įvykdyta sąlyga $E \circ D = D \circ E = 1$, t.y., signalas turi būti priimtas teisingai, kai nėra trikdžių. Tokioje formuluotėje klaidas ištaisančių kodų optimizacijos uždavinys gali būti suprantamas taip: esant duotiems m ir n rasti tokius D ir E, kad klaidingo signalo priėmimo tikimybė būtų minimali.

Atstumas tarp žodžių. Kodavimo teorijoje viena iš svarbiausių sąvokų yra atstumas tarp dvejetainių žodžių. Kiekvienas dvejetainis n-ilgio žodis

$$a = (a_1, a_2, \dots, a_n)$$

gali būti suprantamas kaip visų dvejetainių žodžių grafo - hiperkubo viršūnė. Atstumu tarp žodžių a ir b tada bus pozicijų, kuriose $a_i \neq b_i$, skaičius. Atstumas minimalus ir lygus vienetui, jei žodžiai skiriasi (lygiai) vienoje pozicijoje.

Galima laikyti n ilgio dvejetainį žodį sutapatinti su Abelio grupės Z_2^n elementais. Toje grupėje grupinė operacija apibrėžta kaip pakoordinatinė sudėtis modulio 2, pvz., kai n=5, turėsime

$$10110 + 01101 = 11011$$
$$11011 + 01101 = 10110 \quad \text{ir t.t.}$$

Apibrėžimas. Žodžio $a = (a_1, a_2, .., a_n)$ svoriu $w(a)$ vadinamas vienetų skaičius jo koordinatėse. Atstumu tarp dviejų vienodo ilgio žodžių vadinamas jų sumos svoris, t.y., $d(a, b) = w(a+b)$.

Akivaizdu, kad šis apibrėžimas sutampa su prieš tai duotuoju. Pavyzdžiui,

$w(0101) = 2$, $w(1101) = 3$, $d(1011, 1111) = 1$,

$d(0000, 0011) = 2$, $d(1101, 1101) = 0$.

Pastebėsime, kad poslinkis $x \to x+c$ nekeičia atstumų:

$$d(a+c, b+c) = w(a+b+c+c) = w(a+b) = d(a, b).$$

Priimant signalą lygiai k klaidų tikimybė lygi $\binom{n}{k} p^{n-k} q^k$, o tikimybė, kad bus $\le l$ klaidų, lygi

$$(4) \quad p^n + \binom{n}{1} p^{n-1} q + \binom{n}{2} p^{n-2} q^2 + \dots + \binom{n}{l} p^{n-l} q^l$$

Pasinaudosime atstumo funkcija $d(b, b^*)$ klaidos tikimybių išraiškai. Tikimybė to, kad perduotas žodis b bus priimtas kaip b^*, lygi $p^{n-d(b,b^*)} \cdot q^{d(b,b^*)}$, pavyzdžiui, 0011 bus priimtas kaip 1011 su tikimybe $p^3 q$.

Pastebėsime, kad tam, kad būtų galimybė aptikti klaidą vienoje pozicijoje, minimalus atstumas tarp kodinių žodžių turi būti lygus 2. Kitaip klaida vienoje pozicijoje gali vieną kodinį žodį paversti kitu, ir ji liks nepastebėta.

Teorema 2. Tam, kad kodas leistų aptikti visas klaidas visose $\le k$ pozicijose, būtina ir pakankama, kad minimalus atstumas tarp dviejų kodinių žodžių būtų $k+1$.

Tokiam kodui tikimybė to, kad klaidos žodyje liks nepastebėtos, lygi

$$(5)\quad q^n + \binom{n}{1} p q^{n-1} + \dots + \binom{n}{k+1} p^{n-k-1} q^{k+1}$$

kai q - mažas skaičius, o k - nelabai didelis, tai svarbiausias narys čia bus

$$\binom{n}{k+1} p^{n-k-1} q^{k+1}.$$

Teorema 3. Tam, kad kodas suteiktų galimybę ištaisyti visas klaidas $\le k$ pozicijose, būtina ir pakankama, kad mažiausias atstumas tarp dviejų kodinių žodžių būtų $2k+1$.

Jei šis reikalavimas funkcijai E išpildytas, tai funkciją D atrodys taip:

$a \to$ (artimiausias žodis iš E vaizdo). Pavyzdžiui,

(1,3) kodui $E: 0 \to 000$, $1 \to 111$, o D - tai funkcija

$000 \to 0$	$111 \to 1$
$001 \to 0$	$011 \to 1$
$010 \to 0$	$101 \to 1$
$100 \to 0$	$110 \to 1$

Ši schema taiso klaidas vienoje pozicijoje.

Iš 1-os teoremos išplaukia

Teorema 4. Jei kodas taiso $\le k$ klaidų, tai dvejetainiam signalui ilgio n, ~~dekoduotas~~ įvykio, kad dekoduotas signalas nesutaps su pasiųstu yra nedidesnė, negu

$$\binom{n}{k+1} p^{n-k-1} q^{k+1} + \dots + \binom{n}{1} p q^{n-1} + q^n$$

Atitinkamai, teisingo priėmimo tikimybė ne mažesnė, negu

$$p^n + \binom{n}{1} p^{n-1} q + \dots + \binom{n}{k} p^{n-k} q^k$$

Norint taikyti grupių teoriją patogu naginėti klaidų žodžius (eilutes). Duotas signalas (žodis) $a = a_1 a_2 \dots a_m$ užkoduojamas kodiniu žodžiu $b = b_1 b_2 \dots b_n$. Ryšio kanalas perdavimo metu prideda prie jo klaidų žodį (eilutę) $e = e_1 e_2 \dots e_n$. Tokiu būdu, imtuvas priima signalą $r = r_1 r_2 \dots r_n$, kur $r_i = b_i + e_i$. Taisanti klaidas sistema pervers žodį $r_1 r_2 \dots r_n$ į artimiausią kodinį žodį $b_1 b_2 \dots b_n$. Sistema, tiktai aptinkanti klaidas, stebi, ar priimtas žodis yra kodinis ir išduoda pranešimą apie klaidą, kai ji įvyksta.

Tegul, pavyzdžiui, perduodamas žodis $a = 01$ koduojamas žodžiu $b = 0110$, o klaidų žodis $e = 0010$. Tada bus priimtas žodis $r = 0100$. Taisanti klaidas sistema jį pervers į 0110, o po to atstatys persiųstą žodį 01.

Jei sistema tik aptinka klaidas, tai bet kuris ~~kodas~~ klaidų žodis e su vieninteliu vienetu sukurs žodį $b^* = b + e$, kuris nebus kodinių žodžių. Pavyzdžiui, nagrinėkime (2,3) kodą su lyginumo tikrinimu:

$E: 00 \to 000 \quad 10 \to 101, \quad 01 \to 011, \quad 11 \to 110$

Kodinių žodžių aibė yra 000, 011, 101, 110.

Nė vieną iš klaidų žodžių 001, 010, 100, 111 neperveda vieno kodinio žodžio į kitą. Todėl vienkartinė (o taip pat trigubą) klaidą šis kodas pastebėtų.

Kodas, aptinkantis dvi klaidas - tai toks kodas, kad nė vienas ~~kodinis~~ klaidų žodis su vienu ar dviem vienetais neperveda vieno kodinio žodžio į kitą.

Pvz.3. Kodas (2,5), kurio koduojanti funkcija E apibrėžta taip

$$00 \to 00000 = b^1 \qquad 01 \to 01011 = b^2$$
$$10 \to 10101 = b^3 \qquad 11 \to 11110 = b^4$$

aptinka dvi klaidas. Taip pat ši schema gali ištaisyti vienkartinę klaidą, todėl, kad bet kurie du kodiniai žodžiai nesutampa mažiausiai trijose pozicijose. Iš to, kad $d(b^i, b^j) \geq 3$ kai $b^i \neq b^j$ išplaukia, kad vienkartinė klaida sukurs žodį, kuris nutolęs atstumu 1 nuo vieninteliu kodinio žodžio, kurio ji buvo perdeotas.

Taigi, dekodavimo schema, perveduanti priimtą žodį į jam artimiausią kodinį, ištaisys vienkartinę klaidą. Vieno bloko teisingo perdavimo tikimybė šiuo atveju nemažesnė negu $p^5 + 5p^4q$.