SHENNON AXBOROT NAZARIYASI VA AXBOROT ENTROPIYASI. AXBOROT ENTROPIYASINING XOSSALARI VA AMALIY QO‘LLANILISHI

O’runboyeva Mohinur Amonboy qizi

Namangan davlat universiteti, Fizika-matematika fakulteti, Amaliy matematika yo‘nalishi 1-kurs talabasi

E-mail: mohinurorunboyeva@gmail.com,ORCID:0009-0000-8158-2054

Tel.: +998 94 254 54 08

Dehqanova Gulnoza Alijon qizi

Namangan davlat universiteti, Fizika-matematika fakulteti, Amaliy matematika yo‘nalishi 1-kurs talabasi

E-mail: gulnozadehqanova1501@gmail.com,ORCID:0009-0008-7903-8111

Tel.: +998 50 211 50 08

Ubaydullayeva Dilorom Maxmudjon qizi

Namangan davlat universiteti o’qituvchisi

E-mail:ubaydullayevadilorom578@gmail.com,ORCID:0009-0003-5363-3549

Annotatsiya. Maqolada Klod Shennonning axborot nazariyasi asoslari va axborot entropiyasi tushunchasi ko‘rib chiqiladi. Entropiyaning ta’rifi, uning asosiy xossalari (manfiymaslik, maksimumga erishish sharti, additivlik, botiqlik), shartli entropiya va o‘zaro axborot tushunchalari aniq sonli misollar bilan yoritiladi. Manba kodlash teoremasi Xaffman kodi misolida, kanal sig‘imi esa ikkilik simmetrik kanal misolida tushuntiriladi. Shuningdek, entropiyaning ma’lumotlarni siqish, aloqa tizimlari, mashinaviy o‘rganish va axborot xavfsizligidagi amaliy qo‘llanilishi ko‘rsatib beriladi.

Kalit so‘zlar: *axborot nazariyasi, Shennon entropiyasi, bit, shartli entropiya, o‘zaro axborot, kodlash, kanal sig‘imi, ma’lumotlarni siqish.*

Аннотация. В статье рассматриваются основы теории информации Клода Шеннона и понятие информационной энтропии. На конкретных числовых примерах раскрываются определение энтропии, её основные свойства (неотрицательность, условие достижения

максимума, аддитивность, вогнутость), а также понятия условной энтропии и взаимной информации. Теорема о кодировании источника поясняется на примере кода Хаффмана, а пропускная способность канала — на примере двоичного симметричного канала. Показано практическое применение энтропии в сжатии данных, системах связи, машинном обучении и информационной безопасности.

Ключевые слова: *теория информации, энтропия Шеннона, бит, условная энтропия, взаимная информация, кодирование, пропускная способность канала, сжатие данных.*

Abstract. The article examines the foundations of Claude Shannon's information theory and the concept of information entropy. The definition of entropy, its main properties (non-negativity, the condition for attaining the maximum, additivity, concavity), as well as the notions of conditional entropy and mutual information are explained with concrete numerical examples. The source coding theorem is illustrated by a Huffman code, and channel capacity by the binary symmetric channel. The practical use of entropy in data compression, communication systems, machine learning and information security is also demonstrated.

Keywords: *information theory, Shannon entropy, bit, conditional entropy, mutual information, coding, channel capacity, data compression.*

Kirish

Har kuni telefonimizdan xabar yuboramiz, rasm almashamiz, video koʻramiz va kamdan-kam hollarda bularning orqasida qanday matematik gʻoya turganini oʻylaymiz. Vaholanki, «axborot» soʻzi kundalik hayotda juda koʻp ishlatilsa-da, bir paytlar uning aniq ta'rifi ham, oʻlchov birligi ham boʻlmagan. «Bu xabar ikkinchisidan ikki marta koʻproq axborot tashiydi» degan gap ilmiy jihatdan hech narsani anglatmagan. Bu holat 1948-yilda oʻzgardi: Klod Shennon «Aloqaning matematik nazariyasi» nomli maqolasida axborotni ehtimollar nazariyasi tili bilan oʻlchash usulini taklif qildi [1]. Aynan shu ish bugungi raqamli dunyoning nazariy poydevoriga aylandi.

Bu maqolani tayyorlash jarayonida biz oʻzimiz uchun ham kutilmagan narsani payqadik: entropiya formulasi juda qisqa, bir qatorga sigʻadi, lekin uning orqasida ma'lumotlarni siqish,

xatolarni tuzatish, hatto parollarning mustahkamligigacha bo‘lgan keng soha yotibdi. Mavzuning dolzarbligi ham shunda. Amaliy matematika yo‘nalishi talabalari uchun entropiya ehtimollar nazariyasi, diskret matematika va dasturlash bilimlarini birlashtiruvchi tushuncha bo‘lib, kelgusida mashinaviy o‘rganish, kriptografiya va ma’lumotlar tahlili fanlarida albatta uchraydi.

Ishning maqsadi Shennon entropiyasining mohiyatini tushuntirish, uning asosiy xossalarini keltirish va amaliy qo‘llanilish sohalarini ko‘rsatishdan iborat. Buning uchun quyidagi vazifalar qo‘yildi: axborot miqdorini o‘lchash g‘oyasining paydo bo‘lishini ko‘rib chiqish; entropiya ta’rifini keltirish va sodda misollarda tushuntirish; uning xossalarini isbotlash yoki asoslash; shartli entropiya va o‘zaro axborot tushunchalarini kiritish; manba kodlash va kanal sig‘imi haqidagi natijalarni misollarda ko‘rsatish; entropiyani hisoblaydigan dastur tuzish. Ishda adabiyotlar tahlili, matematik hisoblash va dasturiy modellashtirish usullaridan foydalanildi.

1. Axborotni o‘lchash muammosi

Axborotni o‘lchash haqidagi birinchi jiddiy qadam Ralf Xartli nomi bilan bog‘liq. U 1928-yilda aloqa kanali orqali uzatiladigan xabarlar sonini ko‘rib chiqib, N ta teng imkoniyatli variantdan bittasini tanlash axborot miqdorini log N ga teng deb olishni taklif qildi [2]. Agar logarifm asosi 2 bo‘lsa, birlik sifatida bit olinadi. Masalan, 8 ta teng imkoniyatli variantdan bittasini ko‘rsatish uchun $\log_2 8 = 3$ bit kerak. Bu hisobni boshqacha ko‘rish ham mumkin: to‘g‘ri javobni topish uchun «ha/yo‘q» turidagi uchta savol yetarli.

Xartli formulasi bitta muhim faraz ustiga qurilgan: hamma variantlar bir xil ehtimolli. Real hayotda esa bunday emas. Ingliz tilidagi matnda «e» harfi «z» harfiga qaraganda ancha tez-tez uchraydi, shuning uchun biror joyda «e» ni ko‘rish bizni «z» ni ko‘rishdan kam hayratga soladi. Shennon aynan shu kuzatishni matematik tilga o‘girdi: kamroq kutilgan hodisa ko‘proq axborot tashiydi. Agar x hodisaning ehtimolligi p(x) bo‘lsa, uning xususiy axborot miqdori quyidagicha aniqlanadi:

$$I(x) = -\log_2 p(x) = \log_2(1/p(x)) \qquad (1)$$

Bu ta’rif sezgiga mos keladi. Aniq sodir bo‘ladigan hodisa (p = 1) uchun I = 0, ya’ni biz allaqachon bilgan narsani eshitish yangi narsa bermaydi. Tanga tashlanganda gerb tushishi (p = 1/2) bizga 1 bit axborot beradi. O‘yin kubigining tushgan yog‘i (p = 1/6) esa $\log_2 6 \approx 2{,}585$ bit axborot tashiydi. Mustaqil hodisalar uchun axborot miqdorlari qo‘shiladi: ikki marta tanga tashlash 2 bit beradi. Mana shu qo‘shiluvchanlik talabi logarifmik bog‘lanishni deyarli yagona mumkin variantga aylantiradi.

2. Shennon entropiyasi

Xususiy axborot bitta hodisani tavsiflaydi. Biroq aloqa nazariyasida ko‘pincha butun manba, ya’ni ehtimollari ma’lum bo‘lgan hodisalar to‘plami qiziqtiradi. Shu sababli xususiy axborotning o‘rtacha qiymati kiritiladi va u entropiya deb ataladi. X diskret tasodifiy miqdor $x_1, x_2, \ldots, x_n$ qiymatlarni mos ravishda $p_1, p_2, \ldots, p_n$ ehtimollar bilan qabul qilsin. Uning Shennon entropiyasi:

$$H(X) = -\Sigma\, p_i \log_2 p_i \qquad (2)$$

Bu yerda yig‘indi i = 1 dan n gacha olinadi, $p_i = 0$ bo‘lgan hadlar uchun esa $0 \cdot \log 0 = 0$ deb qabul qilinadi (bu limit bilan asoslanadi). Logarifm asosi 2 bo‘lsa, entropiya bitlarda, natural logarifm olinsa natlarda, o‘nli logarifm olinsa xartlilarda o‘lchanadi. Entropiyani «tasodifiy tajriba natijasi bizni o‘rtacha qanchalik hayron qoldiradi» degan savolga javob sifatida ham, «natijani yozib olish uchun o‘rtacha nechta bit kerak» degan savolga javob sifatida ham tushunish mumkin. Shennonning asosiy yutug‘i aynan shuki, bu ikki talqin bir-biriga teng ekani keyinchalik isbotlandi (5-bo‘limga qarang).

Eng oddiy misol sifatida ikkita natijali tajribani olamiz: bitta natija p ehtimol bilan, ikkinchisi 1 − p ehtimol bilan chiqadi. Bunda entropiya ikkilik entropiya funksiyasi deb ataladi:

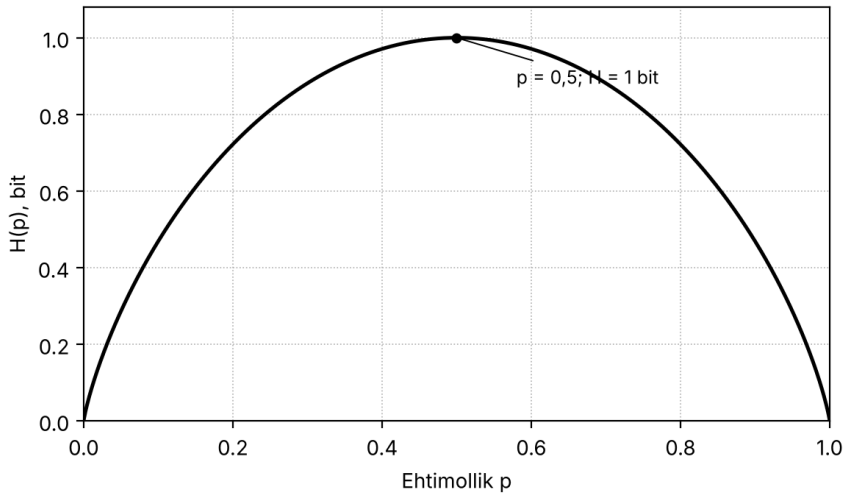
$$H(p) = -p \log_2 p - (1 - p) \log_2 (1 - p) \qquad (3)$$

Adolatli tanga uchun p = 0,5 va H = 1 bit. Noadolatli tanga, masalan p = 0,9 bo‘lsa, $H = -(0{,}9 \cdot \log_2 0{,}9 + 0{,}1 \cdot \log_2 0{,}1) \approx 0{,}136 + 0{,}332 = 0{,}469$ bit. Ya’ni natijani oldindan taxmin qilish

osonlashgani sari entropiya kamayadi. 1-jadvalda p ning bir nechta qiymati uchun H(p) hisoblab chiqilgan, 1-rasmda esa shu funksiyaning grafigi keltirilgan.

1-jadval. Ikkilik entropiya funksiyasi H(p)

p	**0**	**0,01**	**0,05**	**0,1**
H(p), bit	0	0,081	0,286	0,469
p	0,2	0,3	0,4	0,5
H(p), bit	0,722	0,881	0,971	1



1-rasm. Ikkilik entropiya funksiyasi H(p) grafigi

Grafikdan ikki narsa ko‘rinib turibdi. Birinchisi, funksiya p = 0,5 nuqtada maksimumga yetadi va bu nuqtaga nisbatan simmetrik. Ikkinchisi, uning qiymati p nolga yaqinlashganda ham, birga yaqinlashganda ham nolga intiladi. Ikkinchi holat qiziq: p = 0,01 bo‘lganda ham entropiya 0,081 bitgina, ya’ni juda kam uchraydigan hodisa mavjudligiga qaramay, manbaning o‘zi deyarli «bashorat qilinadigan» bo‘lib qoladi.

Endi to‘rt natijali manbani qaraymiz: ehtimollari 1/2, 1/4, 1/8 va 1/8. Entropiya H = 1/2·1 + 1/4·2 + 1/8·3 + 1/8·3 = 1,75 bit. Agar to‘rtala natija teng ehtimolli bo‘lganda, entropiya $\log_2 4 = 2$ bit bo‘lar edi. Demak, notekis taqsimot axborot miqdorini kamaytiradi. Aynan shu «tafovut» ma’lumotlarni siqish imkonini beradi, biz bunga 5-bo‘limda qaytamiz.

3. Entropiyaning asosiy xossalari

Entropiya funksiyasi bir qator muhim xossalarga ega. Ularning ko‘pchiligi oddiy, ammo har biri nazariyaning qaysidir joyida ishlatiladi. Quyida eng asosiylarini keltiramiz.

Manfiymaslik. Har qanday taqsimot uchun $H(X) \geq 0$. Chunki $0 < p_i \leq 1$ bo‘lganda $-\log_2 p_i \geq 0$. Tenglik faqat bitta natijaning ehtimoli 1 bo‘lgandagina, ya’ni tasodifiylik yo‘q bo‘lganda o‘rinli.

Maksimum xossasi. n ta natijali manba uchun $H(X) \leq \log_2 n$ va tenglik faqat barcha natijalar teng ehtimolli ($p_i = 1/n$) bo‘lgandagina bajariladi. Buni qisqa isbotlash mumkin. $\ln x \leq x - 1$ tengsizligidan foydalanamiz (bu tengsizlik $x > 0$ uchun o‘rinli). U holda

$$H(X) - \log_2 n = \Sigma\, p_i \log_2(1/(n\, p_i)) \leq (k/n - 1)/\ln 2 \leq 0 \qquad (4)$$

bu yerda yig‘indi $p_i > 0$ bo‘lgan k ta had bo‘yicha olinadi va $k \leq n$. Tenglik $k = n$ va barcha $n \cdot p_i = 1$ bo‘lganda, ya’ni taqsimot tekis bo‘lganda erishiladi. Natijaning amaliy ma’nosi aniq: eng ko‘p noaniqlik hamma variantlar teng imkoniyatli paytda bo‘ladi.

Simmetriya. Entropiya ehtimollarning qaysi tartibda yozilishiga bog‘liq emas. Qiymatlarning nomi emas, faqat ehtimollar muhim: «gerb/raqam» ham, «ha/yo‘q» ham 0,5 : 0,5 taqsimotda bir xil 1 bit beradi.

Imkonsiz hodisaning ta’siri yo‘qligi. Taqsimotga ehtimoli nolga teng natija qo‘shilsa, entropiya o‘zgarmaydi. Haqiqatan ham, bunday natija hech qachon sodir bo‘lmasa, u tufayli noaniqlik ortmaydi.

Additivlik. Agar X va Y mustaqil tasodifiy miqdorlar bo‘lsa, ularning birgalikdagi entropiyasi $H(X, Y) = H(X) + H(Y)$. Misol uchun, adolatli tanga (1 bit) va adolatli kub (2,585 bit) birgalikda 3,585 bit beradi. Bu xossa logarifmning ko‘paytmani yig‘indiga aylantirishidan kelib chiqadi.

Botiqlik. Entropiya ehtimollar taqsimotining botiq (konkav) funksiyasi. Ya’ni ikkita taqsimotni aralashtirsak, aralashmaning entropiyasi entropiyalarning aralashma bo‘yicha o‘rtachasidan kam bo‘lmaydi. Bu xossa optimallashtirish masalalarida muhim: maksimumni topish botiq funksiya uchun ishonchli va yagona yechimga olib keladi.

Guruhlash xossasi. Tanlovni ketma-ket bosqichlarga bo‘lish entropiyani o‘zgartirmaydi: avval guruhni tanlash entropiyasi, keyin guruh ichida tanlash entropiyasining o‘rtacha qiymati qo‘shiladi. Masalan, 1/2, 1/4, 1/4 ehtimolli uch natijani «birinchi natija yoki qolganlari» deb ikki bosqichga bo‘lsak, H = H(1/2, 1/2) + 1/2·H(1/2, 1/2) = 1 + 0,5 = 1,5 bit. To‘g‘ridan-to‘g‘ri hisoblasak ham 1/2·1 + 2·(1/4)·2 = 1,5 chiqadi.

Shennon o‘z maqolasida muhim narsani ko‘rsatgan [1]: agar entropiyadan uzluksizlik, teng ehtimolli holatlar soni ortgani sari qiymatning ortishi va guruhlash xossasini talab qilsak, uning formulasi (2) ko‘rinishda bo‘lishi (musbat ko‘paytuvchi aniqligida) mutlaqo muqarrar bo‘lib qoladi. Keyinchalik A.Ya. Xinchin bu natijani qat’iy shaklda isbotlab, uni aniqlashtirdi [5]. Ya’ni entropiya formulasi «o‘ylab topilgan» emas, u bir necha tabiiy talabdan mantiqiy kelib chiqadi.

4. Shartli entropiya va o‘zaro axborot

Aloqa muammolarida odatda ikkita miqdor bor: yuborilgan X va qabul qilingan Y. Qabul qilingan Y ni bilgan holda X haqida qancha noaniqlik qoladi? Shu savolga shartli entropiya javob beradi:

$$H(X|Y) = -\Sigma\,\Sigma\, p(x, y)\, \log_2 p(x|y) \qquad (5)$$

Birgalikdagi entropiya uchun zanjir qoidasi H(X, Y) = H(Y) + H(X|Y) o‘rinli. Eng muhim xossa: shartlash entropiyani oshirmaydi, ya’ni H(X|Y) ≤ H(X). Boshqacha aytganda, qo‘shimcha ma’lumot bilish noaniqlikni kamaytirishi mumkin, lekin hech qachon oshirmaydi. Bu ikki miqdor orasidagi farq o‘zaro axborot deb ataladi:

$$I(X; Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y) \geq 0 \qquad (6)$$

O‘zaro axborot Y ni kuzatish orqali X haqida qancha axborot olinishini o‘lchaydi. U nolga teng bo‘lishi faqat X va Y mustaqil bo‘lgandagina mumkin. Aniq misol ko‘ramiz: X va Y ikkilik belgilar bo‘lib, ularning birgalikdagi ehtimollari p(0,0) = p(1,1) = 0,4 va p(0,1) = p(1,0) = 0,1 bo‘lsin. Bu holda H(X) = H(Y) = 1 bit. H(X, Y) = −(2·0,4·log₂0,4 + 2·0,1·log₂0,1) ≈ 1,722 bit. Demak, I(X; Y) = 1 + 1 − 1,722 ≈ 0,278 bit. Haqiqatan ham, Y

ning X dan farq qilish ehtimoli 0,2, shuning uchun $H(X|Y) = H(0,2) \approx 0,722$ va $I = 1 - 0,722 = 0,278$. Ikki usul bir xil natija berdi, bu hisoblarning to‘g‘ri ekanini ko‘rsatadi.

O‘zaro axborot tushunchasining ahamiyati shundaki, u o‘zgaruvchilar orasidagi bog‘lanishni chiziqli bog‘lanish bilan cheklanmasdan o‘lchaydi. Korrelyatsiya koeffitsiyenti nolga teng bo‘lgan, lekin bir-biriga chuqur bog‘liq miqdorlar mavjud (masalan, $Y = X^2$ bo‘lib, X nolga nisbatan simmetrik taqsimlangan holat), o‘zaro axborot esa bunday bog‘liqlikni aniqlaydi. Shu sababli u ma’lumotlarni tahlil qilishda belgilarni tanlash uchun ishlatiladi.

5. Manba kodlash teoremasi va ma’lumotlarni siqish

Endi entropiyaning amaliy ma’nosiga o‘tamiz. Faraz qilaylik, manba belgilarni berilgan ehtimollar bilan chiqaradi va biz har bir belgini ikkilik kod so‘zi bilan yozmoqchimiz. Maqsad o‘rtacha kod uzunligini iloji boricha kamaytirish. Shennonning manba kodlash teoremasiga ko‘ra, hech bir bir ma’noli dekodlanadigan kod o‘rtacha uzunligi manba entropiyasidan kichik bo‘la olmaydi, boshqa tomondan, shunday prefiks kod mavjudki, uning o‘rtacha uzunligi L uchun [1, 4]

$$H(X) \le L < H(X) + 1 \qquad (7)$$

tengsizlik o‘rinli bo‘ladi. Shunday qilib, entropiya «bir belgiga o‘rtacha nechta bit kerak» savoliga eng quyi chegarani beradi va ma’lumotlarni siqishning fundamental chegarasini belgilaydi. Shuni ham aytish kerakki, u kodlash usulini ko‘rsatib bermaydi, faqat nimaga erishish mumkin va nimaga mumkin emasligini aytadi.

Chegaraga amalda yaqinlashadigan usullardan biri Devid Xaffman 1952-yilda taklif etgan kodlashdir [3]. Uning g‘oyasi oddiy: tez-tez uchraydigan belgilarga qisqa, kam uchraydigan belgilarga uzun kodlar beriladi. Yuqorida ko‘rilgan 1/2, 1/4, 1/8, 1/8 ehtimolli manba uchun kodlar 0, 10, 110, 111 bo‘ladi va o‘rtacha uzunlik $1/2 \cdot 1 + 1/4 \cdot 2 + 2 \cdot (1/8) \cdot 3 = 1,75$ bit, ya’ni aynan entropiyaga teng. Bunday mukammal moslik faqat ehtimollar ikkining manfiy darajalari bo‘lganda chiqadi. Umumiy holda biroz farq qoladi. Ehtimollari 0,4; 0,3; 0,2; 0,1 bo‘lgan to‘rt belgili manbani ko‘ramiz. Uning entropiyasi $H \approx 0,4 \cdot 1,322 + 0,3 \cdot 1,737 + 0,2 \cdot 2,322 + 0,1 \cdot 3,322 \approx 1,846$ bit. Xaffman algoritmi bo‘yicha eng kam ehtimolli ikki belgi

(0,2 va 0,1) birlashtiriladi, hosil bo‘lgan 0,3 yana 0,3 bilan birlashadi va hokazo. Natijada 2-jadvaldagi kod hosil bo‘ladi.

2-jadval. Xaffman kodi (manba entropiyasi H ≈ 1,846 bit)

Belgi	Ehtimollik	Kod so‘zi	Uzunlik (bit)
A	0,4	0	1
B	0,3	10	2
C	0,2	110	3
D	0,1	111	3

O‘rtacha uzunlik L = 0,4·1 + 0,3·2 + 0,2·3 + 0,1·3 = 1,9 bit. Bu (7) tengsizlikka mos: 1,846 ≤ 1,9 < 2,846. Agar belgilarni teng uzunlikda, ya’ni har birini 2 bit bilan kodlasak, o‘rtacha 2 bit ketar edi, demak, siqish 5 % ga yaqin tejash beradi. Samaradorlik H/L ≈ 0,97. Belgilarni bittalab emas, guruhlab kodlasak, bu ko‘rsatkich yanada entropiyaga yaqinlashadi.

Real matnlar uchun entropiya qiymatlari ham qiziq. Ingliz alifbosidagi 26 ta harf teng ehtimolli bo‘lganda, har bir harf $\log_2 26 \approx 4{,}7$ bit tashir edi. Shennon 1951-yilda odamlar bilan o‘tkazgan tajribalar orqali, harflarning o‘zaro bog‘liqligini hisobga olganda, ingliz matnining entropiyasi bir harfga taxminan 0,6 dan 1,3 bitgacha ekanini baholadi [6]. Bu shuni anglatadiki, oddiy matnni kod jadvalidagi 8 bitdan ancha kamroq joyda saqlash mumkin va aynan shu zip kabi arxivlovchilarning ishlash asosini tashkil etadi. Matnlarning yuqori darajada «ortiqcha» ekani aslida tilning xatolardan himoyalanish xususiyati ham: harfning bir qismi tushib qolsa ham, biz gapni o‘qib ola olamiz.

6. Shovqinli kanal va kanal sig‘imi

Shennon nazariyasining ikkinchi buyuk natijasi shovqinli kanallarga tegishli. Real aloqa kanalida signal buziladi: Wi-Fi tarmog‘ida uzatilgan bit qabul qilinganda boshqacha bo‘lishi mumkin. Intuitiv ravishda, xatosiz aloqa uchun ma’lumotni qayta-qayta takrorlash kerak, bu esa tezlikni nolga yaqinlashtiradi, deb o‘ylash mumkin. Shennon buning aksini isbotladi: agar

uzatish tezligi R kanal sig‘imi C dan kichik bo‘lsa, xatolik ehtimoli ixtiyoriy kichik bo‘lgan kodlar mavjud [1, 4]. C dan yuqori tezlikda esa ishonchli uzatish mumkin emas. Kanal sig‘imi o‘zaro axborotning maksimumi sifatida aniqlanadi:

$$C = max\ I(X; Y) \qquad (8)$$

Maksimum kirish belgilarining barcha mumkin bo‘lgan taqsimotlari bo‘yicha olinadi. Eng sodda model, ikkilik simmetrik kanal: har bir uzatilgan bit p ehtimol bilan teskarisiga aylanadi. Bunda sig‘im C = 1 − H(p) formula bilan topiladi. Agar p = 0,1 bo‘lsa, C = 1 − 0,469 = 0,531 bit (kanalning bir marta ishlatilishiga). Ya’ni har 10 bitning bittasi buzilsa ham, kanalning taxminan yarmidan ko‘pi foydali ma’lumot uzatishga yaroqli bo‘ladi. p ≈ 0,11 da sig‘im taxminan 0,5 bitga tushadi, p = 0,5 bo‘lsa C = 0, chunki bunda chiqish kirishga umuman bog‘liq emas: qabul qilingan bit tanga tashlashdan farq qilmaydi.

Uzluksiz signallar uchun Shennon–Xartli formulasi ma’lum: $C = B \cdot \log_2(1 + S/N)$, bunda B kanalning o‘tkazish polosasi (Gs), S/N signalning shovqinga nisbati [1]. Misol uchun, B = 3000 Gs va S/N = 1000 (30 dB) bo‘lsa, $C = 3000 \cdot \log_2 1001 \approx 29\ 900$ bit/s, ya’ni taxminan 30 kbit/s. Aynan shu qiymat o‘tgan asrning oddiy telefon liniyalari orqali modem tezligining nazariy chegarasini belgilab bergan. Formula shuni ham ko‘rsatadi: polosani kengaytirish tezlikni chiziqli oshiradi, signal quvvatini oshirish esa faqat logarifmik ravishda.

Shennon teoremasi mavjudlikni isbotlaydi, lekin yaxshi kodlarni qurish usulini ko‘rsatmaydi. Amaliy xatolarni tuzatuvchi kodlar R. Xemming kashfiyotidan boshlandi [7]. Xemming kodi, masalan, 4 ta ma’lumot bitiga 3 ta nazorat bit qo‘shib, istalgan bitta xatoni topadi va tuzatadi. Keyinchalik Rid–Solomon kodlari, turbo-kodlar va LDPC kodlari paydo bo‘ldi. Bugun ular CD va DVD disklarda, QR-kodlarda, mobil aloqa va sun’iy yo‘ldosh tizimlarida ishlatiladi.

7. Entropiyaning amaliy qo‘llanilishi

Ma’lumotlarni siqish. Entropiya siqish chegarasini aniqlaydi. Xaffman kodlash JPEG va MP3 kabi formatlarning yakuniy bosqichida ishlatiladi. Lempel va Ziv 1977-yilda taklif qilgan algoritm [8] esa manba ehtimollarini oldindan bilishni talab qilmaydi va ZIP, PNG, GIF kabi

formatlarning asosini tashkil etadi. Bunday usullar «universal» deyiladi, chunki ular har qanday manba uchun entropiyaga yaqinlashadi.

Mashinaviy o‘rganish. Qaror daraxtlarini qurishda (ID3, C4.5 algoritmlari) har bir qadamda eng ko‘p axborot beradigan belgi tanlanadi [9]. Axborot yutug‘i $Gain(S, A) = H(S) - \Sigma\,(|S_v|/|S|)\cdot H(S_v)$ formulasi bilan hisoblanadi. Klassik misolda 14 ta namunaning 9 tasi «ha», 5 tasi «yo‘q» bo‘lsa, boshlang‘ich entropiya $H = 0{,}940$ bit chiqadi, va har bir belgi shu noaniqlikni qanchalik kamaytirishiga qarab baholanadi. Bundan tashqari, neyron tarmoqlarni o‘rgatishda keng qo‘llaniladigan kross-entropiya yo‘qotish funksiyasi va Kullbak–Leybler masofasi [10] ham entropiyaning bevosita avlodlaridir.

Axborot xavfsizligi. Parol yoki kalitning mustahkamligi uning entropiyasi bilan o‘lchanadi. Agar parol uzunligi n bo‘lib, har bir belgi N ta belgidan teng ehtimollik bilan tasodifiy tanlangan bo‘lsa, entropiya $n\cdot\log_2 N$ bit bo‘ladi. 3-jadvalda bir nechta misol keltirilgan. Shennon bu g‘oyalarni 1949-yildayoq kriptografiyaga tatbiq etgan [11].

3-jadval. Tasodifiy tanlangan parollarning entropiyasi

Parol turi	Alifbo hajmi	Uzunlik	Entropiya, bit
Faqat raqamlar	10	6	19,9
Kichik lotin harflari	26	8	37,6
Harflar va raqamlar	62	8	47,6
Barcha bosma belgilar	94	12	78,7
2048 so‘zlik lug‘atdan so‘zlar	2048	4 so‘z	44,0

Jadvaldan ko‘rinib turibdiki, uzunlik alifbo hajmidan ko‘ra kuchliroq ta’sir qiladi: oddiy lotin harflaridan iborat 12 belgili parol (≈ 56 bit) murakkab belgili, lekin 8 belgili paroldan (≈ 52 bit) kuchliroq bo‘lishi mumkin. Albatta, bu hisob parol tasodifiy tanlangan bo‘lsagina

to‘g‘ri. Inson o‘ylab topgan parollar, masalan «Ali2005!», alifbosi katta bo‘lsa ham, kam entropiyaga ega, chunki ularni tanlashda ham qonuniyat bor.

Fizika va statistika. Entropiya g‘oyasi Shennondan oldin Boltsman va Gibbs ishlarida termodinamika uchun paydo bo‘lgan va formulalarning o‘xshashligi tasodif emas. E. Jeyns 1957-yilda axborot entropiyasini statistik mexanikaning asosi sifatida ko‘rishni taklif qildi [12]. Uning maksimal entropiya prinsipi shunday deydi: noma’lum taqsimotni tanlashda ma’lum cheklovlarga zid kelmaydigan va eng ko‘p entropiyali taqsimotni tanlash kerak, chunki u hech qanday ortiqcha farazni kiritmaydi.

Biologiya. DNK to‘rt xil nukleotiddan (A, C, G, T) tuzilgan, shuning uchun har bir pozitsiya ko‘pi bilan $\log_2 4 = 2$ bit axborot saqlaydi. Genom ketma-ketliklarining haqiqiy entropiyasi bundan kamroq, chunki ular takrorlanuvchi qismlarga ega. Entropiya mutatsiyalar sodir bo‘ladigan va saqlanib qoladigan joylarni topishda, shuningdek, DNK ketma-ketligini siqishda foydalaniladi.

8. Hisoblash misoli: entropiyani dasturda topish

Nazariyani amalda sinab ko‘rish uchun oddiy Python dasturini tuzdik. U berilgan matndagi har bir belgining chastotasini sanaydi, ehtimollarni topadi va (2) formula bo‘yicha entropiyani hisoblaydi:

```
import math
from collections import Counter


def entropiya(matn):
    n = len(matn)
    h = 0.0
    for k in Counter(matn).values():
        p = k / n
        h -= p * math.log2(p)
    return h
```

```
print(entropiya('aaaa'))        # 0.0
print(entropiya('abab'))        # 1.0
print(entropiya('abracadabra'))  # 2.0403...
```

Natijalar nazariyaga mos keladi. «aaaa» qatorida bitta belgi 1 ehtimol bilan uchraydi va entropiya nolga teng. «abab» da ikki belgi teng ehtimolli, shuning uchun 1 bit. «abracadabra» so‘zida 11 ta belgi bor: a — 5 marta, b va r — 2 martadan, c va d — 1 martadan. Hisob entropiyani taxminan 2,040 bit deb beradi, bu 5 ta turli belgi uchun mumkin bo‘lgan maksimum $\log_2 5 \approx 2{,}322$ bitdan kam. Demak, nazariy jihatdan bu so‘zni belgiga o‘rtacha 2,04 bit bilan yozish mumkin, 8 bitli standart kodlashda esa belgiga 8 bit ketadi. Dasturni o‘zbek tilidagi matnlarga qo‘llab, ularning entropiyasini turli tillar bilan solishtirish ham qiziqarli mustaqil ish bo‘la oladi.

9. Entropiya haqidagi keng tarqalgan yanglish tasavvurlar

Entropiya o‘rganilganda bir necha tushunmovchilik tez-tez uchraydi va ularni oldindan bilib olgan ma’qul. Birinchisi, entropiya tayyor xabarning emas, manbaning (ehtimollar taqsimotining) xususiyati. «abracadabra» so‘zining o‘ziga «entropiya» tayinlab bo‘lmaydi, chunki ehtimollarni biz chastotalardan taxmin qilyapmiz. Boshqa taxminlar boshqa natija beradi.

Ikkinchisi, entropiya xabarning ma’nosi yoki foydaliligini o‘lchamaydi. Tasodifiy belgilar ketma-ketligi eng yuqori entropiyaga ega, lekin hech qanday ma’no tashimaydi. Shennonning o‘zi ham maqolasida axborotning semantik jihati muhandislik masalasiga aloqasi yo‘qligini ta’kidlagan [1]. Uchinchisi, entropiyani har doim «tartibsizlik» deb tushunish. Bu o‘xshatma fizikada qisman ishlaydi, ammo axborot nazariyasida aniqroq talqin kerak: entropiya — noaniqlik o‘lchovi, ya’ni natijani bilmasligimiz darajasi. Ko‘p manbalarda aytilishicha, Shennon bu so‘zni fon Neyman maslahati bilan tanlagan, chunki formulaning fizikadagi entropiyaga o‘xshashligi bor edi. Bu rivoyatning aniqligi bahsli, lekin o‘xshashlikning o‘zi haqiqiy.

Xulosa

Maqolada Shennon axborot nazariyasining asosiy tushunchalari ko‘rib chiqildi va quyidagi xulosalarga kelindi. Birinchidan, axborot miqdori hodisaning ehtimolligi bilan bog‘liq: kam uchraydigan hodisa ko‘proq axborot beradi, entropiya esa bu miqdorning o‘rtacha qiymati sifatida manbaning noaniqligini bitlarda o‘lchaydi. Ikkinchidan, entropiya manfiy emas, teng ehtimolli taqsimotda maksimumga erishadi ($H \le \log_2 n$), mustaqil manbalar uchun additiv va ehtimollar bo‘yicha botiq. Bu xossalar uning formulasini tabiiy talablardan mantiqan keltirib chiqaradi.

Uchinchidan, entropiya amaliy chegaralarni belgilaydi: manba kodlash teoremasi ma’lumotni siqishning, kanal sig‘imi haqidagi teorema esa ishonchli uzatishning nazariy chegarasini beradi. Aniq misollar (1,75 bit uchun mukammal Xaffman kodi, 0,4–0,3–0,2–0,1 taqsimot uchun 1,9 bitlik kod, p = 0,1 bo‘lgan kanalning 0,531 bit sig‘imi) bu chegaralar hisoblanadigan va amalda erishiladigan ekanini ko‘rsatadi. To‘rtinchidan, entropiya tushunchasi faqat aloqa bilan cheklanmaydi: u siqishda, mashinaviy o‘rganishda, kriptografiyada, fizika va biologiyada ishlatiladi.

Biz uchun, amaliy matematika yo‘nalishining birinchi kurs talabalari uchun, eng qimmatli xulosa shu bo‘ldi: ehtimollar nazariyasidagi oddiy formula tabiat va texnika haqidagi chuqur savollarga javob bera oladi. Kelgusida bu mavzuni kodlash nazariyasi va statistik o‘rganish nuqtai nazaridan chuqurroq o‘rganish, shuningdek, o‘zbek tili matnlarining entropiyasini tajribada baholash rejalashtirilmoqda.

Foydalanilgan adabiyotlar

1. Shannon C.E. A Mathematical Theory of Communication // Bell System Technical Journal, 1948, Vol. 27, No. 3, pp. 379–423; No. 4, pp. 623–656.
2. Hartley R.V.L. Transmission of Information // Bell System Technical Journal, 1928, Vol. 7, No. 3, pp. 535–563.
3. Huffman D.A. A Method for the Construction of Minimum-Redundancy Codes // Proceedings of the IRE, 1952, Vol. 40, No. 9, pp. 1098–1101.

4. Cover T.M., Thomas J.A. Elements of Information Theory. 2nd ed. Hoboken, NJ: Wiley-Interscience, 2006.

5. Khinchin A.I. Mathematical Foundations of Information Theory. New York: Dover Publications, 1957.

6. Shannon C.E. Prediction and Entropy of Printed English // Bell System Technical Journal, 1951, Vol. 30, No. 1, pp. 50–64.

7. Hamming R.W. Error Detecting and Error Correcting Codes // Bell System Technical Journal, 1950, Vol. 29, No. 2, pp. 147–160.

8. Ziv J., Lempel A. A Universal Algorithm for Sequential Data Compression // IEEE Transactions on Information Theory, 1977, Vol. 23, No. 3, pp. 337–343.

9. Quinlan J.R. Induction of Decision Trees // Machine Learning, 1986, Vol. 1, No. 1, pp. 81–106.

10. Kullback S., Leibler R.A. On Information and Sufficiency // Annals of Mathematical Statistics, 1951, Vol. 22, No. 1, pp. 79–86.

11. Shannon C.E. Communication Theory of Secrecy Systems // Bell System Technical Journal, 1949, Vol. 28, No. 4, pp. 656–715.

12. Jaynes E.T. Information Theory and Statistical Mechanics // Physical Review, 1957, Vol. 106, No. 4, pp. 620–630.

13. MacKay D.J.C. Information Theory, Inference, and Learning Algorithms. Cambridge: Cambridge University Press, 2003.