1.5 Загальні поняття теорії інформації

1.5.1 Структура інформаційного каналу

Процес управління складається в отриманні інформації про об'єкт управління і вироблення на її основі керуючої інформації. Для передачі інформації використовуються сигнали. Виникає задача оцінки кількості інформації у сигналі і визначення втрат інформації в ході формування сигналів, а також під час передачі сигналів по каналах зв'язку.

Під повідомленням зазвичай розуміють інформацію, яка виражена в певній формі і підлягає передачі. Повідомлення - це форма представлення інформації.

Повідомлення називається *дискретним,* якщо воно є послідовністю окремих елементів (букв, цифр, символів). Дискретне повідомлення, що складається з n елементів, можна розглядати як слово довжини n, елементи якого приймають значення з кінцевого алфавіту $X = \{x_1, x_2, \dots, x_m)$. Зазвичай букви алфавіту $x_1, x_2, \dots, x_m$ кодуються числами, переважно в двійковій системі числення, а відповідні ним сигнали є послідовності імпульсів певної тривалості. Використовуються *рівномірні коди,* в яких букви алфавіту представляються комбінаціями з однакового числа елементів, і *нерівномірні коди,* складені з комбінацій різної довжини. Прикладом рівномірного коду може служити телетайпний код Бодо, а нерівномірного – азбука Морзе, які для деяких українських букв мають вигляд:

Буква	**А**	**Б**	**В**	**Г**	**Д**	**Е**
Код Бодо	**10000**	**00110**	**01101**	**01010**	**11110**	**01000**
Азбука Морзе	· —	— · · ·	— —	— — ·	— · ·	·

Неперервне повідомлення, на відміну від дискретного, представляється неперервною функцією часу (наприклад, при передачі звуків або зображень). Проте на практиці спектр функцій зазвичай обмежується, тобто вважається, що спектральне розкладання не містить частот вище за деяку граничну частоту ω. Відповідно до теореми Котельникова такі функції цілком визначаються кінцевим числом значень, відлічених через інтервали часу $\Delta t=\pi/\omega$.

Таким чином, передача неперервного повідомлення в тому практично важливому випадку, коли воно може бути представлене функцією з обмеженим спектром, може бути зведена, як і у разі дискретного повідомлення, до передачі послідовності чисел.

Класична теорія інформації була розроблена Клодом Шенноном для рішення задач передачі інформації з мінімальними втратами. Згідно Шеннону, модель системи зв'язку має вигляд, представлений на рис. 1.19.

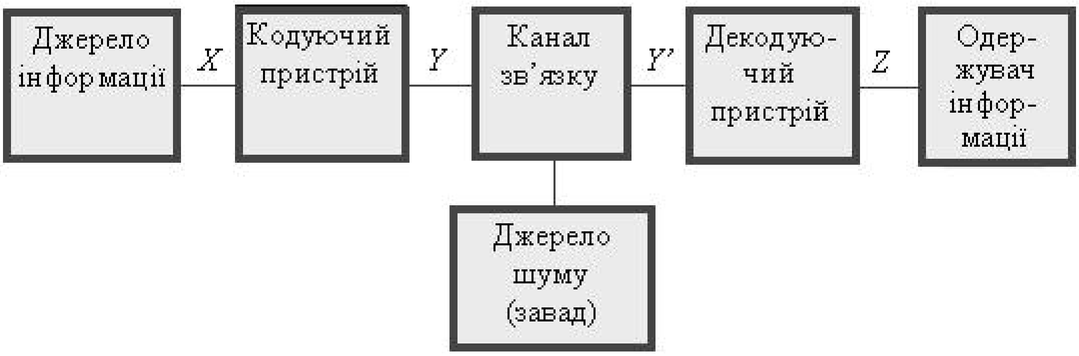

Рис. 1.19 – Структура інформаційного каналу

Модель містить такі складові частини:

а) джерело інформації, яке виробляє повідомлення або послідовність повідомлень;

б) кодуючий пристрій, який перетворює повідомлення в сигнали, придатні для передачі по каналу зв'язку; цей пристрій називають також передавачем;

в) канал передачі інформації від передавача до приймача;

г) декодуючий пристрій, що виконує функції відновлення початкового повідомлення за отриманим сигналом; цей пристрій називають також приймачем;

д) об'єкт, якому призначається повідомлення (одержувач інформації);

е) джерело шуму (завад).

В теперешній час на основі класичної теорії інформації створена інформаційна теорія управління.

1.5.2 Поняття кількості інформації

Кількість інформації визначається як математична міра зменшення невизначеності знань про який-небудь об'єкт в процесі пізнання. Якщо H_1 - початкова (апріорна) невизначеність знання з даного питання, а H_2 - залишкова (апостеріорна) невизначеність, що характеризує стан знання після отримання повідомлення, то кількість інформації, що міститься в цьому повідомленні визначається різницею:

$$I = H_1 - H_2. \tag{1.50}$$

Таким чином, повідомлення несе тим більше інформації, чим в більшому степені зменшується невизначеність про об'єкт, що досліджується.

Виникає задача числової оцінки невизначеності. Інтуїтивно зрозуміло, що невизначеність змінюється від нуля до нескінченності: 0 – ситуація повної визначеності; ∞ – ситуація повної невизначеності. Такими властивостями володіє функція $-\log P$, де P – ймовірність того чи іншого стану досліджуваного об’єкта.

Якщо через x_1, x_2, …, x_i, …, x_n позначити можливі варіанти повідомлення X, при чому відомі ймовірності цих варіантів $P(x_\mathrm{i})$, то невизначеність i-го варіанту можна оцінити величиною

$$H(x_i) = -\log_\mathrm{a} P(x_i).$$

Для всієї сукупності варіантів повідомлення невизначеність оцінюється усередненням за ймовірностями $P(x_\mathrm{i})$ варіантів:

$$H(X) = \sum_{i=1}^{n} P(x_i) H(x_i) = -\sum_{i=1}^{n} P(x_i) \log_a P(x_i). \tag{1.51}$$

Така міра невизначеності отримала назву інформаційної ентропії. Вибір основи логарифмів a неістотний. У теорії інформації використовуються двійкові логарифми і кількість ентропії визначається у бітах (англ. bit – binary digit).

Інформаційна ентропія оцінює апріорну невизначеність знань про досліджуваний об'єкт.

1.5.3 Властивості ентропії Шеннона

Якщо всі варіанти повідомлення рівноймовірні, тобто $P(x_1)=P(x_2)=\ldots=P(x_n)=1/n$, то невизначеність максимальна. Згідно формули (1.51)

$$H(X) = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = \log n \ . \qquad (1.52)$$

В цьому випадку ентропія визначається виключно числом різних можливих варіантів повідомлень n і, по суті, є характеристикою алфавіту повідомлень. Якщо ж повідомлення різноімовірні, то алфавіт можна розглядати як дискретну випадкову величину, задану статистичним розподілом частот n_i або ймовірностей $p_i = n_i/n$.

Чим менш ймовірне повідомлення, тим воно більш інформативне для одержувача. Якщо формулу (1.51) переписати у вигляді

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i) = \sum_{i=1}^{n} P(x_i) \log \frac{1}{P(x_i)}$$

то величину $\log[1/P(x_i)]$ можна розглядати як *часткову ентропію,* що характеризує інформативність повідомлення x_i, а ентропію H – як середнє значення часткових ентропій. При малих $P(x_i)$ часткова ентропія велика, а з наближенням $P(x_i)$ до одиниці вона прагне до нуля (рис. 1.20).

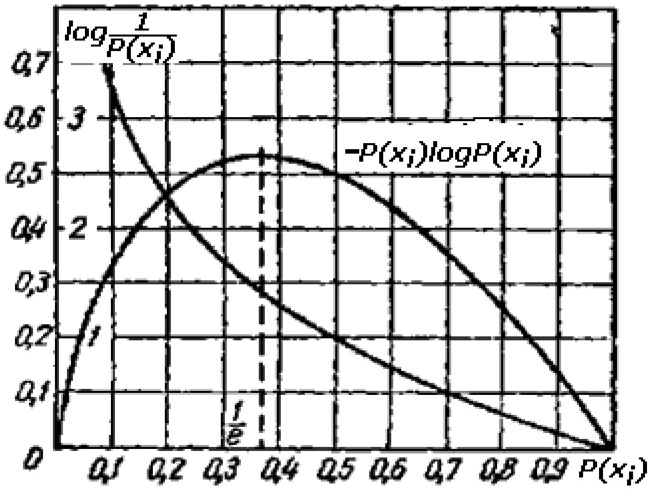

Рис. 1.20

Функція $-P(x_i)\log P(x_i)$ відображає внесок повідомлення x_i, в ентропію H. Як видно, при $P(x_i) = 1$ ця функція дорівнює нулю, потім зростає до свого максимуму і при зменшенні $P(x_i)$ прагне до нуля. Щоб знайти максимум, прирівняємо до нуля похідну:

$$0 = \frac{\partial}{\partial P(x_i)}[-P(x_i)\log P(x_i)] = -\log P(x_i) - \log e = \log[P(x_i)e]$$

Звідси отримуємо $P(x_i)e=1$, де e основа натурального логарифму. Таким чином, функція $-P(x_i)\log P(x_i)$ при $P(x_i)=1/e=0{,}37$ має максимум

$$\frac{1}{e}\log e = 0{,}531.$$

Ентропія H – величина дійсна, ненегативна і обмежена, тобто $H \geq 0$ (ця властивість виходить з того, що такими ж якостями володіють всі її доданки $P(x_i)\log[1/P(x_i)]$. Ентропія дорівнює нулю, якщо повідомлення відоме наперед (в цьому випадку один елемент повідомлення має ймовірність, рівну одиниці, а ймовірності решти є рівними нулю). Можна також показати, що ентропія максимальна, якщо всі повідомлення рівноімовірні, тобто $H_{\max} = \log m$. Особливий інтерес представляють *бінарні* (*двійкові*) *повідомлення,* що використовують двохбуквений алфавіт (цифри 0 та 1). Оскільки при $n = 2$ ймовірність букв алфавіту $P(x_1) + P(x_2) = 1$, то можна покласти $P(x_1) = P$ і $P(x_2) = 1 - P$. Тоді ентропія визначається співвідношенням:

$$H = -P\log P - (1-P)\log(1-P).$$

Графік цього співвідношення є сумою двох графіків, як це показано на рис. 1.21. Як бачимо, ентропія двійкових повідомлень досягає максимального значення, рівного 1 біту, при $P = 0{,}5$ і її графік симетричний щодо цього значення.

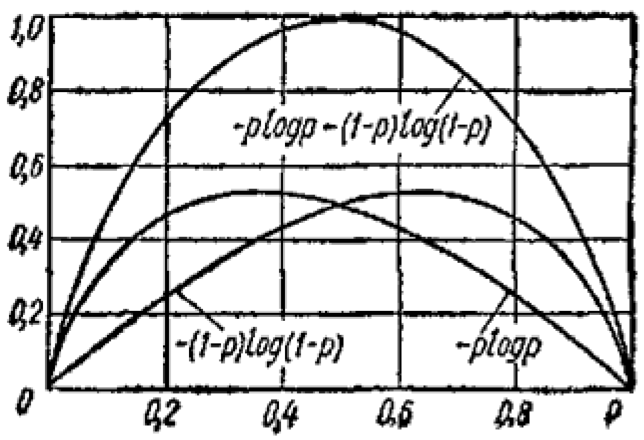

Рис. 1.21

1.5.4 Ентропія неперервного сигналу

Для оцінки ентропії неперервного сигналу застосування формули (2.60) в загальному випадку неможливо. З цієї формули видно видно, що ентропія є сумою N доданків, які мають вигляд $H_j = -P_j \log P_j$. Якщо сигнал безперервний, то в будь-якому заданому діапазоні $[X_{\min}, X_{\max}]$ він має нескінченне число значень. Сума нескінченного числа складових H_j приводить до ентропії, рівної нескінченності. Однак така постановка питання не має сенсу. Будь-яке вимірювання супроводиться погрішністю. Тому більш правильною буде така постановка задачі: визначити ентропію знання про значення неперервного сигналу, що підлягає вимірюванню з кінцевою похибкою Δ . Тоді діапазон значень сигналу можна розбити на $N=(X_{\max}-X_{\min})/\Delta$ дільниць. Якщо задана густина імовірності $f(X)$ значень величини X, то імовірність того, що внаслідок вимірювання вийде j-е значення, приблизно дорівнює: $f(X_j)\times\Delta = P_j$ (рис. 1.22).

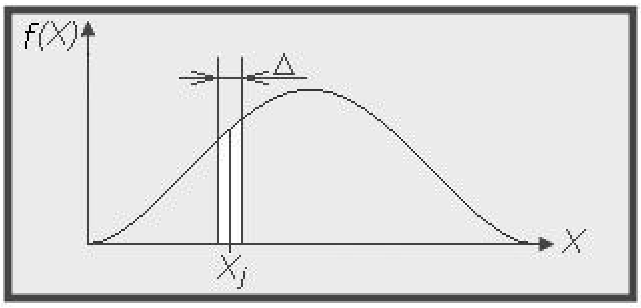

Рис. 1.22

Тоді ентропію можна знайти, підставляючи в її загальну формулу значення P_j (j=1,…,N).

$$H(X) = -\sum_{j=1}^{N} f(X_j)\Delta \log[f(X_j)\Delta] . \tag{1.53}$$

Переходячи від суми до інтеграла, отримуємо

$$H(X) = \lim_{\Delta x \to 0}\left(-\sum_i f(X_i)\Delta X_i \log f(X_i)\Delta X_i\right) = -\int_{-\infty}^{\infty} f(X)\log f(X)dX - \int_{-\infty}^{\infty} f(X)dX \log \Delta X \tag{1.54}$$

Враховуючи, що $\int_{-\infty}^{\infty} f(X)dX = 1$, отримаємо

$$H(X) = -\int_{-\infty}^{\infty} f(X)\log f(X)dX - \log\Delta X = h(X) - \log\Delta X$$

Перший доданок в отриманому виразенні

$$h(X) = -\int_{-\infty}^{\infty} f(X)\log f(X)dX \qquad (1.55)$$

називається диференціальною ентропією. Вона повністю визначається щільністю ймовірностей значень сигналу і не залежить від похибки розрізнення цих значень. Тому всі сигнали з однаковим законом розподілу характеризуються однією і тією ж диференціальною ентропією.

У теорії зв'язку велике значення має рішення питання про те, при якому розподілі забезпечується максимальна ентропія $H(x)$. Можна показати, що при заданій дисперсії σ^2 = const найбільшою інформативністю повідомлення володіє тоді, коли стани елементів розподілені по нормальному закону:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{X^2}{2\sigma^2}}, \qquad (1.56)$$

Оскільки дисперсія визначає середню потужність сигналу, то звідси слідують практично важливі висновки. Передача найбільшої кількості інформації при заданій потужності сигналу (або найбільш економічна передача даної кількості інформації) досягається при такій обробці сигналу, яка наближає розподіл до нормального. В той же час, приписуючи нормальний розподіл заваді, забезпечують її найбільшу «інформативність», тобто враховують її згубну дію на проходження сигналів в самому гіршому випадку.

Якщо дисперсія σ^2 не обмежена, то, як показує аналіз, ентропія максимальна за умови, що значення сигналу усередині інтервалу його існування $a \le X \le b$ розподілені за рівномірним законом, тобто

$$f(X) = \begin{cases} \dfrac{1}{b-a} \text{ при} \le X \le b; \\ 0 \text{ при } X < a \text{ або } X > b. \end{cases}$$

Знайдемо значення ентропії в розглянутих двох випадках. При нормальному законі розподілу

$$h(X) = -\int_{-\infty}^{\infty} f(X)\log f(X)dx = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{X^2}{2\sigma^2}} \log\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{X^2}{2\sigma^2}})dX = \log(\sqrt{2\pi e}\sigma).$$

Для рівномірного закону

$$h(X) = -\int_{-\infty}^{\infty} f(X)\log f(\tilde{O})dX = -\int_{a}^{b} \frac{1}{b-a}\log(\frac{1}{b-a})dX = -\frac{1}{b-a}\log(\frac{1}{b-a})X\Big|_{a}^{b} = \log(b-a),$$

З курсу теорії ймовірностей відомо, що дисперсія рівномірного розподілу $\sigma_p^2 = \frac{(b-a)^2}{12}$. Звідси $b - a = 2\sqrt{3}\sigma_p$, отже $h(X) = \log(2\sqrt{3}\sigma_p)$.

Порівнюючи між собою повідомлення з рівномірним і нормальним розподілом ймовірностей отримуємо співвідношення

$$\sigma_p^2 = \frac{\pi \cdot e}{6}\sigma^2 \approx 1{,}42\sigma^2.$$

Це означає, що при однаковій інформативності повідомлень середня потужність сигналів для рівноймовірного розподілу їх амплітуд повинна бути на 42% більше, ніж при нормальному розподілі.

1.5.5 Методи визначення кількості інформації

Внаслідок одержання повідомлення апріорна невизначеність знання про джерело сигналу не може бути знижена до нуля, оскільки завжди є завади в каналі передачі.

Нехай сигнал несе інформацію про виміряне значення технологічного параметру. На рис. 1.23 показаний апріорний розподіл *p(X)* значень вимірюваної величини і апостеріорний *p'(X)*, який отримуємо після вимірювання з похибкою ±Δ/2.

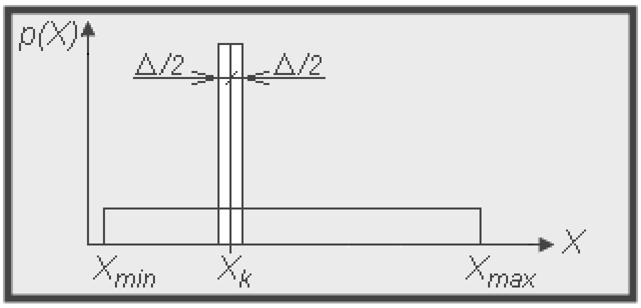

Рис. 1.23

Внаслідок вимірювання область невизначеності вужчає від $[X_{\min},X_{\max}]$ до інтервалу $[X_к–\Delta/2\ ,X_к+\Delta/2]$ довжини Δ, розташованого симетрично відносно отриманого результату $X_к$.

У дискретному випадку, сигнал може приймати тільки фіксовані значення, число яких в заданому діапазоні кінцеве, залишкова невизначеність оцінюється формулою (2.60), в яку підставляються апостеріорні значення ймовірностей сигналу. Визначимо через *P(i/j)* апостеріорну імовірність того, що одержувач отримає *j*-те значення сигналу, в той час як дійсне значення *i*-те. Оцінимо залишкову невизначеність знання про джерело сигналу *X*, якщо отриманий результат z_j:

$$H(X/z_j) = -\sum_{i=1}^{N} P(i/j)\log P(i/j).$$

Величина $H(X/z_j)$ називається частковою умовною ентропією. Середня умовна ентропія

$$H(X/Z) = -\sum_{j} P_j \sum_{i=1}^{N} P(i/j)\log P(i/j).$$

характеризує апостеріорну невизначенність досліджуваного об'єкта, стан якого визначається отриманим сигналом Z. Тоді, згідно (2.1), кількість інформації, що несе сигнал Z:

$$I(X/Z) = H(X) - H(X/Z).$$

Найпростішим випадком є вибір альтернативи з двох подій. Тому за одиницю інформації доцільно прийняти кількість інформації, що міститься у виборі однієї з двох рівноймовірних подій. Ця одиниця називається двійковою одиницею, або бітом (binary digit, bit). Отже, при будь-якій невизначеності звуження області вибору удвічі дає один біт інформації.

1.5.6 Пропускна спроможність каналу передачі

Якщо на відрізку часу *T* передаються *n* повідомлень в моменти часу t_i, для яких значення сигналу X_i статистично не залежать одне від одного (не корельовані), то загальна ентропія дорівнює сумі ентропій

$$H_T(X) = \sum_{i=1}^{n} H_i(X) = nH_i(X) \tag{1.57}$$

Максимальна кількість інформації, що міститься в і-му повідомлені,

$$I_i = H_i(X). \tag{1.58}$$

а загальна інформація, яка може бути отримана у n повідомленнях

$$I_T = H_T(X) = nH_i(X). \tag{1.59}$$

Відношення

$$R = \frac{I_T}{T} = \frac{nH_i(X)}{T}, \tag{1.60}$$

що характеризує швидкість отримання інформації, називається продуктивністю джерела сигналу.

При малих завадах невизначеність значень сигналу знімається в момент вимірювання майже повністю. Потім невизначеність зростає, і через час Δt стає рівною $H_i(X)$. Якщо повідомлення формуються через рівні проміжки часу Δt, то їх кількість за час T:

$$n = \frac{T}{\Delta t}. \tag{1.61}$$

Тоді (1.60), з урахуванням (1.61), отримує вигляд:

$$R = \frac{nH_i(X)}{T} = \frac{T}{\Delta t}\frac{H_i(X)}{T} = \frac{H_i(X)}{\Delta t}. \tag{1.62}$$

Таким чином, продуктивність джерела сигналу визначається швидкістю зростання ентропії.

Продуктивність джерела, як і ентропія, залежать від закону розподілу ймовірностей сигналу.

Очевидно, що швидкісні характеристики каналу повинні відповідати продуктивності джерела сигналу. Середня швидкість створення інформації визначається виразом (1.60). Ця величина залежить від T, тому як міра швидкості використовується межа $\lim\limits_{T\to\infty}\frac{I_T}{T}$.

Верхня границя швидкості може бути знайдена, якщо згадати властивість ентропії приймати максимальне значення, коли всі повідомлення рівноймовірні:

$$C = \sup\left\{ \lim_{T\to\infty} \frac{I_T}{T} \right\} = \lim_{T\to\infty} \frac{\log n}{T}, \; . \tag{1.63}$$

Така швидкісна характеристика називається пропускною спроможність каналу.

Користуючись (1.60), інформацію, що отримується від джерела за час T, можна виразити через його продуктивність

$$I_T = RT \; . \tag{1.64}$$

Значення інформації, яке може бути отримане за час T максимальне, якщо продуктивність джерела дорівнює його пропускній спроможності

$$I_{\mathrm{Tmax}} = CT. \tag{1.65}$$

Надмірністю джерела інформації називається відношення

$$\gamma = \frac{I_{T\max} - I_T}{I_{T\max}} = \frac{CT - RT}{CT} = 1 - \frac{R}{C}. \tag{1.66}$$

Якщо джерело інформації, продуктивність якого дорівнює пропускній спроможності каналу (R=C), у n_0 повідомленнях дає кількість інформації I, то джерело з надмірністю, для якого R<C, може забезпечити таку інформацію у n>n_0 повідомленнях, тобто треба створити n-n_0 надмірних повідомлень. Тоді надмірність можна інакше визначити як

$$\mathrm{g} = (n - n_0)/n = 1 - n_0/n \; . \tag{1.67}$$

Надмірність виникає не тільки при відхиленні закону розподілу сигналу $P(X)$ від оптимального, але і у випадку, якщо між значеннями сигналу X в моменти створення t_i існує статистична залежність, оскільки при цьому зменшується приріст ентропії $H_i(X)$.