

Лабораторна робота № 1

ЕКСПЕРИМЕНТАЛЬНЕ ВИЗНАЧЕННЯ ЕНТРОПІЇ ПОВІДОМЛЕННЯ

Мета роботи: вивчення властивостей ентропії як кількісної міри інформації.

1.1. Теоретичні відомості

У 1946 американський вчений-статистик Джон Тьюкі запропонував назву БІТ (BIT - аббревіатура від Binary digiT), одне з головних понять XX століття. Тьюкі обрав біт для позначення одного двійкового розряду, здатного приймати значення 0 або 1. Шенноном запропоновано використовувати біт як одиницю виміру інформації, а мірою кількості інформації - функцію, названу ним ентропією.

Представлення інформації потребує набору символів, які творять алфавіт. В залежності від ситуації алфавіт може бути і множиною числових значень, отриманих при вимірюванні фізичної величини, і множиною цифр десяткової чи іншої системи числення, і буквами, наприклад, українського алфавіту, тощо. Для того щоб абстрагуватися від конкретного змісту інформації, тобто її смислового значення, і отримати саме загальне визначення кількості інформації, кількісну міру інформації визначають без урахування її смислового змісту, а також цінності і корисності для одержувача.

Стосовно алфавіту мови, то Шеннон зауважив, що при передачі різних букв ми передаємо різну кількість інформації. Якщо ми передаємо букви, що часто зустрічаються, то інформації менше; при передачі рідкісних букв - більше. Зв'язок між частотою (інакше ймовірністю) появи букви в повідомленні і кількістю інформації в ньому за Шенноном виражається ентропією.

Ентропія $H_r(X)$ (за основою r) джерела інформації обчислюється за наступним виразом:

$$H_r(X) = - \sum_{i=1}^K p(x_i) \log(p(x_i)) \quad , \quad (1.1)$$

де $p(x_i)$ — ймовірність отримання символу x_i , $X = \{x_i\}$ — повідомлення джерела інформації, K — кількість символів в алфавіті.

Чим вища ентропія повідомлення, тим більша кількість інформації в ньому закладена, тим важче її запам'ятати (записати) або передати каналом зв'язку.

Ентропія як кількісна міра інформаційності джерела має такі властивості:

- 1) ентропія дорівнює нулю, якщо хоча б одне з повідомлень достовірне;
- 2) ентропія завжди більша або дорівнює нулю, є величиною дійсною і обмеженою;

- 3) ентропія джерела з двома альтернативними подіями може змінюватися від 0 до 1;
- 4) ентропія - величина адитивна: ентропія джерела, повідомлення якого складаються з повідомлень декількох статистично незалежних джерел, дорівнює сумі ентропій цих джерел;
- 5) ентропія максимальна, якщо всі повідомлення мають однакову імовірність, тобто:

$$H_{\max}(X) = \log_2 k. \quad (1.2)$$

Вираз (1.2) називається формулою Хартлі. Її легко вивести з виразу Шеннона (1.1), припустивши, що $p_i = 1/k$, де $i = 1 \dots k$.

Приклад 1.1. Найчастіше за основу вибирається число 2. В такому випадку інформаційна ентропія збігається з числом бітів, якими можна закодувати інформацію. Наприклад, якщо однакова ймовірність запису в комірці пам'яті одиниці або нуля (ці дві цифри складають алфавіт), тоді

$$H_2(2) = \frac{1}{2} \log_2 2 + \frac{1}{2} \log_2 2 = 1,$$

тобто інформаційна ентропія такої комірки дорівнює 1.

Приклад 1.2. Знайти ентропію дискретної випадкової величини X , заданої розподілом

$x_i \in X$	x_1	x_2	x_3	x_4	x_5	x_6
$p_i \in P$	0,4	0,3	0,15	0,1	0,03	0,02

За формулою (1.1) маємо:

$$H_2(X) = - \sum_{i=1}^6 p(x_i) \log(p(x_i)) = 0,4 \cdot 1,322 + 0,3 \cdot 1,737 + 0,15 \cdot 2,737 + 0,1 \cdot 3,22 + 0,03 \cdot 5,059 + 0,02 \cdot 5,644 = 2,0573 \text{ біт/повідомлення.}$$

Зображення також є носієм інформації – повідомленням. Цифрове зображення складається з окремих структурних елементів – пікселів, кожен з яких має сталі значення кольору. Сучасні технічні засоби формують колір шляхом змішування трьох базових кольорів (система RGB). Абревіатура утворена початковими буквами англійських слів red — червоний, green — зелений та blue — синій). В системі HLS колір задається трьома параметрами: відтінком, контрастністю та яскравістю.

Цифрове зображення зберігається у вигляді файлу, в якому вказується інформація про пікселі зображення. З метою оптимізації розміру графічних файлів використовуються різні способи їх стиснення. Після збереження графічним редактором файл отримує розширення – три, інколи чотири літери після назви. За розширенням можна зробити висновок про спосіб кодування. Розглянемо деякі з цих форматів.

BMP (від англ. Bitmap Picture) — формат зберігання растрових зображень. Спочатку формат міг зберігати лише апаратно-залежні растри (англ. Device Dependent Bitmap, DDB), але з розвитком технологій відображення графічних даних формат BMP став переважно зберігати апаратно-незалежні растри (англ. Device Independent Bitmap, DIB). З форматом BMP працює величезна кількість програм, оскільки її підтримка інтегрована в операційні системи Windows і Os/2. Файли формату BMP можуть мати розширення .bmp, .dib і .rle. Крім того, дані цього формату включаються в двійкові файли ресурсів RES і в ре-файлі. Формат не використовує стиснення.

PNG (Portable Network Graphics) — растровий формат збереження графічної інформації, що використовує стиснення без втрат. PNG був створений для заміни формату GIF, графічним форматом, який не потребує ліцензії для використання. Зазвичай файли формату PNG мають розширення .png і використовують позначення MIME-типу image/png.

GIF (англ. Graphics Interchange Format) — 8-бітний растровий графічний формат, що використовує до 256 чітких кольорів із 24-бітного діапазону RGB. Формат було розроблено компанією CompuServe у 1987 році, і з того часу набув широкої популярності у всесвітній павутині завдяки своїй відносній простоті та мобільності. Одними із головних особливостей формату є підтримка анімації та прозорості. Зображення у форматі GIF можуть містити не більше 256 кольорів. Це дозволяє забезпечити швидке завантаження графіки за рахунок скорочення кількості графічних даних, що зберігаються у файлі.

Формат WMF підтримує векторну і растрову графіку. Використовує палітри в 65 тис. і 16 млн кольорів. У файлі WMF використовуються ті самі команди опису графіки, які використовує сама Windows для зображення малюнків на екрані дисплея або принтера. Ці кольори — опис об'єктів, інформація про колір, растрові і текстові дані — записуються у вигляді інструкцій Windows. Для відтворення оригінального зображення ці інструкції необхідно "програти". Зображення, яке при цьому одержується, залежить від програми, що виконує інструкції. Якщо прочитати файл WMF у векторному графічному редакторі, то одержимо векторний малюнок. Якщо той самий файл прочитати у растровому редакторі, то одержимо растрове зображення.

1.2. Завдання лабораторної роботи

1.2.1. Обчислити ентропію дискретної випадкової величини X , заданої розподілом $P(x_i)$. Значення $P(x_i)$ взяти з таблиці 2.1 згідно варіанту.

Таблиця 2.1. Розподіл дискретної випадкової величини

Вар.	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	$p(x_5)$	$p(x_6)$	$p(x_7)$	$p(x_8)$	$p(x_9)$	$p(x_{10})$
1	2	3	4	5	6	7	8	9	10	11
1	0,15	0,07	0,09	0,08	0,09	0,12	0,04	0,13	0,12	0,10
2	0,11	0,08	0,13	0,10	0,08	0,08	0,13	0,10	0,09	0,11
3	0,09	0,11	0,09	0,11	0,13	0,08	0,11	0,08	0,09	0,09
4	0,10	0,09	0,06	0,07	0,12	0,15	0,07	0,09	0,13	0,10
5	0,14	0,08	0,09	0,05	0,08	0,14	0,12	0,09	0,12	0,09
6	0,13	0,06	0,15	0,12	0,13	0,09	0,07	0,07	0,05	0,12
7	0,06	0,13	0,08	0,13	0,10	0,07	0,10	0,14	0,09	0,09
8	0,10	0,10	0,08	0,10	0,07	0,09	0,12	0,11	0,13	0,10
9	0,12	0,08	0,11	0,11	0,09	0,14	0,14	0,09	0,07	0,05
10	0,11	0,08	0,08	0,14	0,08	0,11	0,08	0,07	0,14	0,13

11	0,11	0,10	0,09	0,10	0,12	0,10	0,08	0,07	0,14	0,10
12	0,05	0,08	0,09	0,11	0,14	0,13	0,09	0,09	0,11	0,11
13	0,13	0,12	0,09	0,09	0,11	0,09	0,08	0,12	0,06	0,12
14	0,13	0,10	0,08	0,08	0,10	0,09	0,07	0,12	0,13	0,09
15	0,11	0,08	0,13	0,06	0,06	0,11	0,11	0,09	0,11	0,13
16	0,10	0,09	0,19	0,10	0,04	0,13	0,08	0,10	0,09	0,10
17	0,12	0,11	0,08	0,06	0,16	0,11	0,06	0,11	0,10	0,09
18	0,15	0,10	0,12	0,07	0,13	0,12	0,03	0,03	0,15	0,10
19	0,10	0,07	0,10	0,07	0,15	0,12	0,04	0,14	0,10	0,11
20	0,14	0,12	0,14	0,08	0,07	0,12	0,11	0,08	0,10	0,05
21	0,11	0,10	0,11	0,12	0,07	0,08	0,09	0,11	0,12	0,10
22	0,11	0,14	0,10	0,15	0,07	0,10	0,10	0,04	0,09	0,10
23	0,09	0,12	0,09	0,12	0,09	0,07	0,13	0,11	0,07	0,11
24	0,09	0,08	0,08	0,11	0,12	0,08	0,13	0,10	0,12	0,08
25	0,11	0,05	0,15	0,09	0,11	0,13	0,12	0,06	0,09	0,09
26	0,10	0,09	0,09	0,14	0,14	0,15	0,12	0,07	0,04	0,07
27	0,11	0,09	0,09	0,08	0,12	0,11	0,09	0,10	0,13	0,08
28	0,02	0,14	0,01	0,12	0,11	0,10	0,09	0,13	0,15	0,13
29	0,11	0,10	0,13	0,12	0,04	0,11	0,12	0,09	0,10	0,08
30	0,10	0,10	0,11	0,07	0,12	0,12	0,08	0,10	0,09	0,13

1.2.2. Обчислити ентропію дискретної випадкової величини на основі згенерованої послідовності цілочисельних значень, які вона приймає. Довжина послідовності - випадкове число в діапазоні 10 — 20. Елемент послідовності — випадкове число в діапазоні 0-10.

1.2.3. В довільному середовищі розробити програму для визначення ентропії зображення. В програмі передбачити наступні функції:

- відображення гістограми зображення;
- розрахунок ентропії інформації зображення за виразом (2.1);
- розрахунок ентропії з використанням вбудованої функції;
- відображення обчисленого значення ентропії.

1.2.4. За допомогою програми визначити ентропію для одного і того ж зображення збереженого у форматах bmp, jpeg, tiff, png.

1.2.5. Провести порівняльну оцінку одержаних результатів.

1.2.6. Оформити звіт.

Примітка: отримані при виконанні роботи програмні коди будуть потрібні вам в подальшому. Збережіть їх.

1.3. Вимоги до звіту по лабораторній роботі

Звіт по лабораторній роботі повинен містити:

- 1) індивідуальне завдання згідно варіанту, який відповідає номеру в списку групи (завдання 1.2.1);
- 2) індивідуальне завдання (завдання 1.2.2): згенерована послідовність; отриманий розподіл;
- 3) обчислену ентропію зображення (довільні 3(три), різні формати) та побудувати гістограму зображення;
- 4) обчислену ентропію зображень (із завдання 2) з використанням вбудованої функції;
- 6) висновки до роботи з аналізом отриманих результатів;
- 5) зображення, що використовувались для роботи;
- 7) текст програм.