Лабораторна робота № 3

Стиснення даних

Мета роботи: застосування методів кодування повідомлень для стиснення їх даних.

1.1. Теоретичні відомості. Кодування Хаффмана.

Один з перших алгоритмів ефективного кодування інформації був запропонований Д. А. Хаффманом в 1952 році. Ідея алгоритму полягає в наступному: знаючи ймовірності символів у повідомленні, можна описати процедуру побудови кодів змінної довжини, що складаються з цілої кількості бітів. Символам з більшою ймовірністю ставляться у відповідність більш короткі коди. Коди Хаффмана володіють властивістю префіксності (тобто жодне кодове слово не є префіксом іншого), що дозволяє однозначно їх декодувати.

Класичний алгоритм Хаффмана на вході отримує таблицю частот з якими зустрічаються символи у повідомленні. Далі на підставі цієї таблиці будується дерево кодування Хаффмана .

1. Символи вхідного алфавіту утворюють список вільних вузлів. Кожен лист має вагу, яка може бути рівною або ймовірності, або кількості входжень символу у стиснене повідомлення.
2. Вибираються два вільних вузла дерева з найменшими вагами.
3. Створюється новий вузол з вагою, рівною їх сумарній вазі.
4. Новий вузол додається в список вільних вузлів, а два його нащадка видаляються з цього списку.
5. Кроки, починаючи з другого, повторюються до тих пір, поки в списку вільних вузлів не залишиться тільки один вільний вузол. Він і буде вважатися коренем дерева.

Для формування кодових слів проводиться обхід дерева від кореня до листків при цьому одній дузі, котра виходить з вузла “батька”, ставиться у відповідність біт 1, інший - біт 0 (бінарне дерево).

Приклад 4.1.

Довжина	Кодове слово	X	Ймовірність				
2	01	1	0.25	0.3	0.45	0.55	1
2	10	2	0.25	0.25	0.3	0.45	
2	11	3	0.2	0.25	0.25		
3	000	4	0.15	0.2			
3	001	5	0.15				

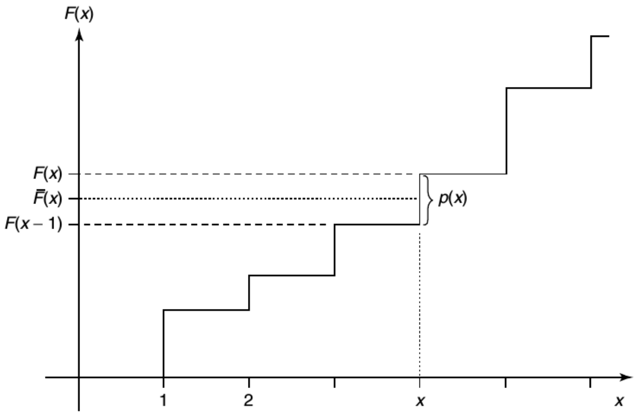
Код Шенона-Фано-Еліаса

Для побудови кодів **Шенона-Фано-Еліаса** використовується кумулятивна функція розподілу випадкової величини F(x). Якщо x — випадкова величина, а p(x) — ймовірність з якою вона приймає деяке значення, то F(x) визначається як:

$$F(x) = \sum_{a \le x} p(a).$$

Визначимо величину $\bar{F}(x)$ як:

$$\overline{F}(x) = \sum_{a<x} p(a) + \frac{1}{2} p(x),$$



Кумулятивна функція розподілу.

Властивості величини $\bar{F}(x)$ дозволяють використати її для побудови кодових слів для випадкової величини x. В якості кодового слова використовується представлення дробової частини $\bar{F}(x)$ у двійковій системі числення.

Приклад.

x	$p(x)$	$F(x)$	$\overline{F}(x)$	$\overline{F}(x)$ бінарне	$l(x) = \left\lceil \log \frac{1}{p(x)} \right\rceil + 1$	Кодове слово
1	0.25	0.25	0.125	0.001	3	00$\boxed{1}$
2	0.5	0.75	0.5	0.10	2	1$\boxed{0}$
3	0.125	0.875	0.8125	0.1101	4	110$\boxed{1}$
4	0.125	1.0	0.9375	0.1111	4	111$\boxed{1}$

Якщо ймовірності розподілу випадкової величини не є діадними (у спадній послідовності ймовірностей кожна наступна вдвічі менша за попередню), тоді бінарне представлення $\bar{F}(x)$ може містити нескінчене число бітів. Позначимо 0.01010101 як $0.\overline{01}$

x	$p(x)$	$F(x)$	$\overline{F}(x)$	$\overline{F}(x)$ бінарне	$l(x) = \left\lceil \log \frac{1}{p(x)} \right\rceil + 1$	Кодове слово
1	0.25	0.25	0.125	0.001	3	001
2	0.25	0.5	0.375	0.011	3	011
3	0.2	0.7	0.6	$0.1\overline{0011}$	4	1001
4	0.15	0.85	0.775	$0.110\overline{0011}$	4	1100
5	0.15	1.0	0.925	$0.111\overline{0110}$	4	1110

1.2. Завдання.

1.2.1. Побудувати бінарні коди Хаффмана та Шенона для:

а) розподілу випадкової величини заданої у табл.1.1 , № варіанту відповідно до журналу групи.

б) послідовності символів із власного П.І.П.

1.2.2. Роздрукувати отримані кодові таблиці у вигляді:

Символ	Ймовірність	Кодове слово	Довжина кодового слова

Символ	F(x)	$\bar{F}(x)$	$\bar{F}(x)$ бінарне	Кодове слово	Довжина кодового слова

1.2.3. Перевірити виконання нерівності Крафта для отриманої послідовності кодових слів.

1.2.4. Обчислити ентропію вихідної послідовності і порівняти із очікуваною довжиною коду.

1.3. Зміст звіту по лабораторній роботі

Звіт по лабораторній роботі повинен містити:

1) Схематичне дерево для коду Хафмана;
2) Графіки кумулятивної функції розподілу та величини $\bar{F}(x)$ для коду Шенона.
3) Кодові таблиці.
4) Значення ентропії та очікуваної довжини.
5) **висновки до роботи з аналізом отриманих результатів.**

Таблиця 1.1. 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