

## ЛІНІЙНІ ТА МАТРИЧНІ СПОСОБИ СТИСНЕННЯ ДАНИХ ПРИ ПЕРЕДАЧІ

В роботі розглядаються лінійні та матричні способи стиснення даних при передачі в інформаційних каналах систем управління. Наведена класифікація способів стиснення при передачі. Виведені формули для середньої довжини кодової комбінації та коефіцієнтів стиснення для різних способів.

**Ключові слова:** лінійні та матричні способи стиснення, середня довжина кодової комбінації, коефіцієнт стиснення.

Розглянемо кілька найпоширеніших способів стиснення даних, які застосовуються при передачі їх. До них належать лінійні, матричні, комбіновані, каскадні та каскадно-комбіновані способи, що гарантують повне відновлення початкового стану стисненої та переданої інформації.(рис.1)

### Лінійні способи

#### Стиснення даних із використанням замість повторень додаткових символів.

Ці способи ґрунтуються на заміні повторюваних елементів деякими умовними символами. Вони є ефективними в тому разі, коли масиви інформації, які подаються у вигляді рядків або стовпців, розташованих у зростаючому порядку, мають однакові значення елементів в одних і тих самих розрядах, що характерно для техніко-економічної інформації. Таке стиснення даних дає змогу скоротити масив у кілька разів[ 1].

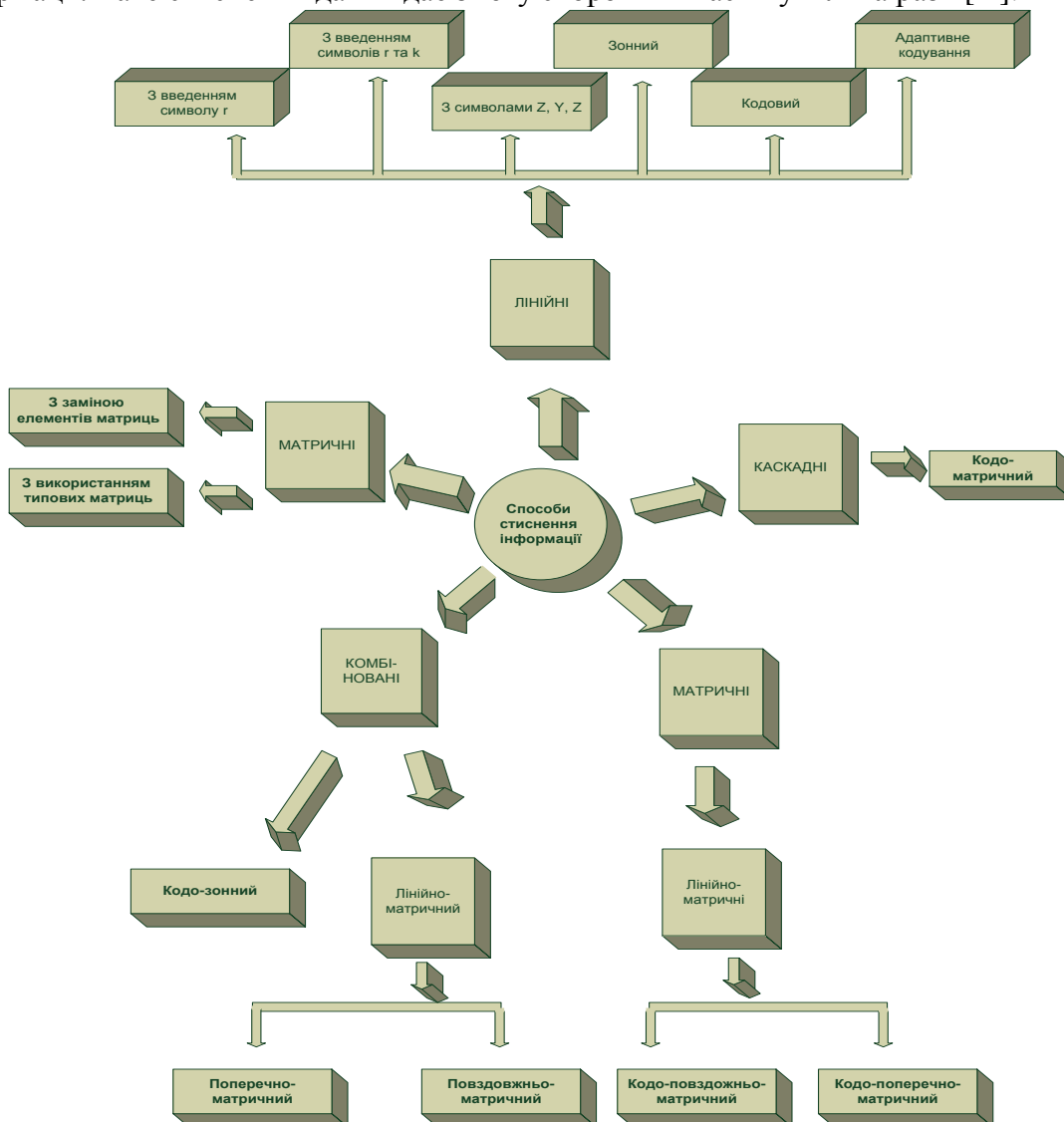


Рис.1 Способи стиснення інформації

Для згорнання масивів, у яких в одному рядку (стовпці) є тільки одна повторювана ділянка, можна використати спосіб з уведенням додаткового символу К (кінець рядка, стовпця). При цьому розгортання масиву ведеться від К до К. В разі фіксованої довжини рядка (стовпця) всі розряди, які знаходяться між К, разом із пропущеними розрядами мають утворювати повний рядок (стовпець) [2]. Так, початковий і згорнений по рядках масиви матимуть вигляд

4 7 5 3 6 8 9 1	4 7 5 3 6 8 9 1
4 7 5 3 6 4 3 2	К р 4 3 2 К 3 5
3 5 5 3 6 6 9 0	р 6 9 0 К 6 9 р
6 9 5 3 6 2 4 1	2 4 1 К 7 р 1 0
7 9 5 3 6 2 1 0	К 8 р 8 0 2 К 1
8 9 5 3 6 8 0 2	6 р К 6 3 р К 4
1 6 5 3 6 8 0 2	6 9 р 5 К р 3 4
6 3 5 3 6 8 0 2	8 К
4 6 9 3 6 8 0 5	
4 6 9 3 6 3 4 8	

Поновлення масиву може виконуватися з початку або з кінця. За наявності в рядку (стовпці) кількох повторюваних ділянок замість р вводять спеціальні символи, що вказують необхідну кількість пропусків. При цьому необхідність у символі К, який визначає кінець рядка (стовпця), відпадає [3]. Так, якщо для позначення заданої кількості пропусків у рядку ввести відповідно символи X-2, Y-3, Z-5, то початковий і згорнений масиви десяткових знаків матимуть такий вигляд:

2 3 4 5 6 7 8 5	2 3 4 5 6 7 8 5
6 3 4 5 6 7 9 8	6 Z 9 8 X 7 X X
6 3 7 5 6 7 9 2	2 2 4 Y Y 8 9 3
2 4 7 5 6 7 9 2	X 2 3 1 1 Z X Z
8 9 3 5 6 2 3 1	4 8 9 3 8 Y 5 X
1 9 3 5 6 2 3 1	2 Y 7 Y 1 7 2 X
1 9 3 5 6 4 8 9	X 0
3 8 3 5 6 5 8 9	
2 8 3 5 7 5 8 9	
1 7 2 5 7 5 8 0	

2 3 4 5 6 7 8 5
6 . . . . . 9 8
. . 7 . . . . . 2
2 4 . . . . .
8 9 3 . . 2 3 1
1 . . . . .
. . . . . 4 8 9
3 8 . . . 5 . .
2 . . . 7 . . .
1 7 2 . . . . . 0

Розгортання масивів при цьому способі стиснення виконується з їх початку або з кінця, а заповнення відповідної кількості пропусків замість додаткових символів X, Y, Z — перенесенням відповідної кількості символів, які знаходяться на однойменних розрядах попереднього рядка:

**Кодове стиснення.** При кодовому стисненні виконується перетворення початкового (вхідного) інформаційного потоку, передаваного алфавітом  $q_1$ , в послідовність елементів вторинного (вихідного потоку) алфавіту  $q_2$ . При цьому  $h$  ( $h = \log_2 q_2$  елементів первинного потоку алфавіту  $q_1$  ставиться у відповідність елементи вторинного потоку алфавіту  $q_2$ .

Таким чином, коефіцієнт стиснення вхідного інформаційного потоку  $K_{ст}$  можливо визначити з формули:

$$K_{ст} = \frac{\log_2 q_2}{\log_2 q_1} \quad (1)$$

Для зручності реалізації коефіцієнт  $K_{ст}$  повинен мати цілочисленне значення  $K_{ст}$  повинен мати цілочисленне ( $K_{ст} = 2, 3, \dots$ ). Це тягне за собою необхідність використання в якості  $q_1$ , і  $q_2$  величин рівних  $2^h$ , де  $h = 1, 2, 3, \dots$ , при умові, що  $q_1 > q_2$ .

**Стиснення інформації використанням адаптивного кодування.** Вимоги до вірогідності інформації можуть коливатися в широких межах залежно від характеру повідомлень. Так, при передачі статистичної інформації про кількість виробленої продукції помилка в молодших розрядах повідомлення (десятки та одиниці) менше впливатиме на правильність переданої інформації, ніж спотворення старших розрядів (мільйони та тисячі). При передачі текстових повідомлень вплив помилок на передачу ще менший.

Таким чином, доцільно передавати повідомлення з заданою вірогідністю у відповідних межах, використовуючи адаптивне кодування відносно джерела повідомлень, що дає змогу забезпечити вірогідність інформації, яка передається, залежно від вимог джерела повідомлень [4]. Для цього повідомлення необхідно поділити за категоріями з заданою вірогідністю передачі. На початку кожної категорії в інформацію вводяться додаткові службові комбінації, які при надходженні в кодер (під час передавання) та декодер (під час приймання) системи передачі перебудовують їх. Перебудова кодера може виконуватися також за сигналами від аналізатора інформації, якщо він установлений на вході системи передачі. Аналізатор обробляє інформацію, що надходить від джерела повідомлень, і поділяє її на категорії.

Для передачі доцільно використовувати три категорії інформації з різною вірогідністю (з різною ймовірністю помилок):  $10^{-3} \dots 10^{-4}$  - для передачі текстів;  $10^{-5} \dots 10^{-6}$  - для передачі цифрових повідомлень від одиниць до тисячі;  $< 10^{-6}$  — для передачі цифрових повідомлень більше тисячі та будь-яких дуже важливих цифрових і текстових повідомлень. Принципи поділу інформації на категорії можуть змінюватися залежно від конкретних умов і вимог [5].

Передачу інформації з різною вірогідністю можна забезпечити зміною способу кодування. Так, для передачі інформації по провідних каналах з вірогідністю  $10^{-3} \dots 10^{-4}$  досить застосувати двійковий первинний код без уведення будь-якого захисту інформації; з вірогідністю  $10^{-5} \dots 10^{-6}$  — двійковий код, що виявляє помилки, а з вірогідністю  $< 10^{-6}$  — код, який виправляє відповідну кількість помилок. Тоді для передачі, наприклад, 32 кодових комбінацій їх довжина в першому випадку становитиме  $n_1 = 5$ , у другому —  $n_2 = 6 \dots 7$ , у третьому —  $n_3 = 8 \dots 12$  елементів.

Середня довжина кодової комбінації при використанні адаптивного кодування залежатиме від процентного співвідношення повідомлень різної категорії і загалом може бути визначена як

$$n_{сеп} = \frac{n_1 N_1 + n_2 N_2 + n_3 N_3}{N} \quad (2)$$

де  $n_1, n_2, n_3$  — довжини кодових комбінацій при передачі повідомлень першої, другої та третьої категорій;  $N_1, N_2, N_3$  — відповідна кількість цих комбінацій;  $N$  — загальна кількість їх.

Коефіцієнт стиснення інформації можна визначити як відношення максимальної довжини  $n$  кодової комбінації до її середньої довжини  $n_{сеп}$ :

$$K_{ст} = n/n_{сеп} \quad (3)$$

**Стиснення інформації збільшенням основи коду.** Ґрунтується воно на перекодуванні кодованих послідовностей символів системи числення з меншою основою в систему з більшою основою. Так, якщо інформаційний масив, який складається з двійкових елементів, подати у вісімковій або шістнадцятковій системі числення (тобто перекодувати його), то це приведе до значного зменшення кількості елементів масиву (в кілька разів).

При цьому коефіцієнти стиснення первинного масиву двійкових елементів у вісімкову і шістнадцяткову системи числення становитимуть

$$K_{ст8}=N_1/N_2=192/64=3;$$

$$K_{ст16}=N_1/N_3=192/48=4,$$

тобто залежать від кількості бітів, що несе один елемент інформаційного масиву (у вісімковій системі  $8 = 2^3 \rightarrow 3$  біт/елемент, у шістнадцятковій —  $16 = 2^4 \rightarrow 4$  біт/елемент).

**Матричні способи.** Значний ефект стиснення дає матричний спосіб із заміною елементів, що повторюються на деякій площі інформаційного масиву, одиничними елементами, які несуть ознаки цих ділянок масиву. Так, якщо ввести символи, що відбивають деякі обмежені площі інформаційного масиву, в яких елементи повторюються порівняно з однойменними розрядами попереднього рядка (наприклад, R-2x2, X-2x5, Y-3x2, 2-5x3), то інформаційний масив

1	2	3	4	5	6	7	8	9	0
2	2	3	4	8	6	7	5	9	0
3	2	3	4	7	6	7	3	9	0
3	2	3	4	7	6	7	3	9	0
3	2	3	4	7	6	7	3	9	0
3	2	3	4	7	8	9	1	9	0

після згортання матиме вигляд

1	2	3	4	5	6	7	8	9	0
2	Y	8	R	5	X	3	7	3	Z
Y	8	9	1						

Розгортка виконується з початку масиву:

1	2	3	4	5	6	7	8	9	0
2	.	.	.	8	.	.	5	.	.
3	.	.	.	7	.	.	3	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	8	9	1	.	.

Коефіцієнт стиснення  $K_{cm} = 60/24 = 2,5$ .

**Стиснення інформації з типовими матрицями** застосовується головним чином для відносно великих масивів. При цьому в пам'яті ЕОМ зберігаються матриці деяких найпоширеніших наборів символів, які зустрічаються в інформаційному масиві, що стискається. Для цього заздалегідь аналізують інформаційний масив, утворюють типові матриці і при стисненні замість наборів символів, які відповідають цій типовій матриці, в послідовність символів інформаційного масиву вводять адреси цих матриць.

## ЛІТЕРАТУРА

1. Кричевский Р. Е. Сжатие и поиск информации. — М.: Радио и связь, 1989. — 168с.
2. Цымбал В. П. Теория информации и кодирование: Учебник - 4 е изд., перераб. и доп. — К.: Вища школа, 1992— 263 с.
3. Кохманюк Д. Сжатие данных: как это делается. — Index Pro, 1992, №1, с.18-29; 1993, №2, с.30-49.
4. Жураковский Ю.П., Полторак В.П., Теория информации та кодування. — К.: Вища школа 2001. — 255с., іл.

Надійшла: 17.10.2013р.

д.т.н., професор Толюпа С.В.