

## ВДОСКОНАЛЕННЯ МЕТОДУ СТИСНЕННЯ ДАНИХ НА ОСНОВІ РАНДОМІЗАЦІЇ

Коваль І.С., Семеляк М.Б.

*Тернопільський національний економічний університет, магістри*

Стрімкий розвиток поширення сенсорних мереж вимагають надійних каналів зв'язку, які могли б забезпечити велику віддачу передавання інформації, високу швидкість та низьку енергозатратність на передавання даних. Зрозуміло, що виконання всіх цих вимог є практично неможливим, тому в таких випадках як правило шукають компроміси, які зазвичай покращують одні характеристики за рахунок інших, так наприклад зниження енерговикористання при передачі одного біта, можливе за рахунок, зменшення швидкості або відстані. При цьому як показують дослідження що робота процесора сенсора споживає набагато менше енергії ніж використовується при передаванні.

Отже, доцільно використовувати різноманітні методи компресії даних, що дозволяє знизити вимоги до ширини каналу та більш раціонально використовувати енергоресурси, що є актуальним при автономному живленні.

В більшості випадків багато процесів не є сильно динамічними[1], тобто значення контрольованого параметру часто або тривалий час залишається незмінним (рис.1), або змінюється на незначну величину. Тому в таких випадках доцільно використовувати стиснення інформації.

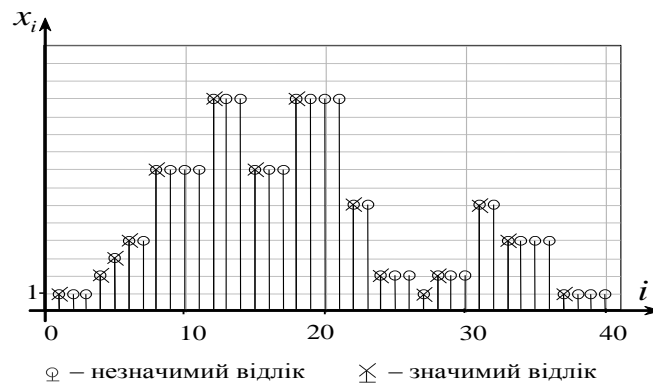


Рисунок 1 - Приклад технологічного процесу

Як видно з рисунку, в даному процесі присутні статичні ділянки, які можна стиснути, при цьому виділені значення, які є значимими, тобто змінилися, і не значимими, тобто тривалий час зберігають попереднє значення. При цьому слід відмітити, що в наведений приклад технологічного процесу, хоча й містить невеликі статичні ділянки, проте має певну динаміку, що більш точно наближає його до реальних процесів та робить класичні методи стиснення, або мало ефективними або взагалі неефективними (при збільшенні динаміки процесу). Класичний метод адаптивного стиснення полягає в тому, що кодуються тільки “значимі” відліки сигналу, а також їх номери.

Розрядність коду, при цьому, визначається за формулою ентропії Хартлі:  $E_x = \hat{E}[\log_2 A]$ , де  $A$  – діапазон квантування,  $\hat{E}$  - цілочисельна функція округлення до більшого.

В результаті такого адаптивного кодування формується наступний потік даних:  $\{i, x_i^*\}$ ,  $i = 1, 2, \dots, N$ , де  $x_i^*$  - значимий відлік,  $N$  - число відліків.

Розрядність коду номера значимого відліку визначається за формулою:  $l = \hat{E}[\log_2 N]$ .

В залежності від динамічності процесу, кількість значимих відліків  $j$  може змінюватися в діапазоні  $1 \leq j \leq N$ .

Коефіцієнт стиснення при цьому рівний:  $K_c = \frac{\hat{E}[\log_2 A] \cdot N}{j \cdot (\hat{E}[\log_2 A] + \log_2 N)}$

де  $A$  – діапазон квантування,  $N$  – кількість відліків у вибірці,  $j$  – кількість значимих відліків.

При цьому: для  $j=1$ , коефіцієнт стиснення максимальний; при  $j=N/2$ , коефіцієнт стиснення рівний 1; при  $j=N$  коефіцієнт стиснення рівний 0,5. Тобто класичний метод стиснення ефективний тільки при  $j < 0.5N$ .

В роботі [1] запропоновано використання методу рандомізації, для впорядкування початкової вибірки по зростанню, а потім застосуванню стиснення (рис.2). З рис. 1 видно, що в початковому масиві є однакові значимі відліки (наприклад  $x_1$  та  $x_{27}$ ,  $x_4$  та  $x_{24}$ ). Тому, вводиться поняття “групового значимого” – це відлік, значення якого відмінне від наступного, і окрім того його значення повинно бути рівним одному з “значимих” попередніх відліків. Таким чином, на рис. 1 є 6 значимих і 8 групових значимих значень.

Після впорядкування по зростанню кодувати потрібно амплітуду і номер значимих відліків, тільки номер значимих групових і не потрібно кодувати незначимих. Позначивши через  $x_a$  – значимі,  $x_g$  – групові значимі відліки, послідовність  $X$  можна записати в наступному вигляді:

$$x_a, N_{x_a}, 0, N_{x_g}, 0, N_{x_g}, 0, N_{x_g}, 0, \dots, 1, x_a, N_{x_a}, 0, N_{x_g}, 0, N_{x_g}, 0, N_{x_g}, 0, \dots, 1, x_a, N_{x_a}, 0, N_{x_g}, \dots$$

де  $N_{x_g}$ ,  $N_{x_a}$  – номери відповідно групових і значимих відліків у початковому неупорядкованому масиві  $X$ ,

«0» – означає що за ним слідує номер відліку з амплітудою, значення якої слідує після кожної нової «1». Стиснена послідовність наведена на рис.3.

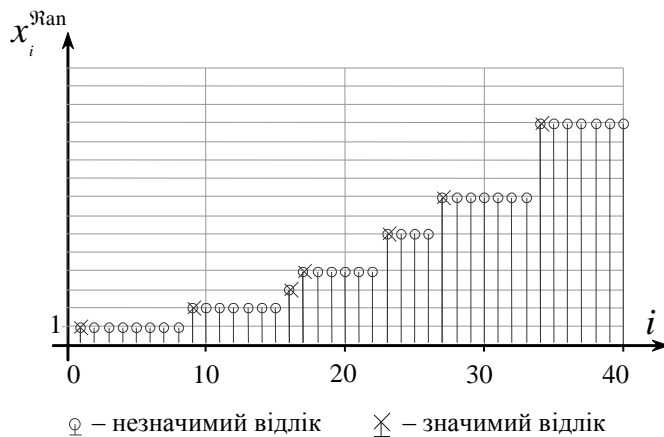


Рисунок 2 - Впорядкований масив X

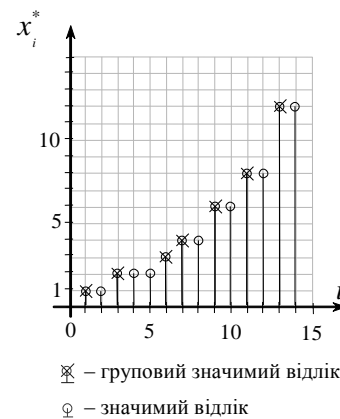


Рисунок 3 - Стиснений масив X

При цьому, як видно з рис.2-3 дані в масиві впорядковані по зростанню та відрізняються на незначну величину, звідки видно, що кодувати кожне значиме значення повним значенням амплітуди є недоцільним, ефективно при цьому кодувати лише прирости від попереднього значення, що дозволить додатково підвищити ефективність кодування в порівнянні з відомим методом описаним в [1]. При цьому ефективність такого підходу зростатиме зі збільшенням розрядності, що використовується для представлення аналогового сигналу.

Застосування рандомізації типу впорядкування по зростанню при використанні адаптивного методу стиснення даних дозволяє суттєво підвищити коефіцієнт стиснення. Якщо класичний метод стиснення ефективний при кількості значимих відліків  $j < 0.5N$ , то при застосуванні рандомізації метод ефективний при  $j < (0.7 \div 0.9)N$  в залежності від співвідношення “значимих” і “групових значимих” відліків. Ефективність методу зменшується при зростанні числа “значимих” (тому що треба реєструвати їх амплітуду і номер), і збільшується при зростанні числа групових значимих (для яких фіксується тільки номер). При збільшенні розрядності діапазону квантування ( $E_x$ ) ефективність методу також підвищується. Використання методу кодування значимих значень на основі приростів дозволяє знизити надлишковість та підвищити ефективність стиснення даних.

### Список використаних джерел

1. Lazarowych I.M., Nikolaychuk J.M. Method of randomization and its application for adaptive data compression // Proc. of Second IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Application IDAACS'2003 – Lviv, Ukraine, 2003. P.362-364.