

Кодування Рівнів Квантування для Ущільнення Інформації

Юрій Ляш
кафедра інформатики
Прикарпатський національний університет
Івано-Франківськ, Україна
yurchukil@cym.org

Data Compression by Encoding Quantization Levels

Yurii Liash
dept. of Computer Science
Precarpathian National University
Ivano-Frankovsk, Ukraine
yurchukil@cym.org

Анотація— подано результати дослідження кодування рівнів квантування рекурсивними кодовими послідовностями, встановлено граничні межі ефективності використання запропонованого методу зменшення надлишковості даних.

Abstract— the paper presents the results of the research of the encoding of quantization levels by the recursive sequence codes, and sets the limits of the efficiency of the proposed method of data compression.

Ключові слова— рекурсивні кодові послідовності, адаптивні методи зменшення надлишковості інформації.

Keywords— recursive code sequences, adaptive data compression methods.

I. ВСТУП

Велика частина повідомлень, які передаються по каналах зв'язку є малоінформативними, тобто несуть надлишкові або несуттєві відомості. В результаті виникла задача стиснення об'єму даних, які передаються або запам'ятовуються, шляхом усунення природної надлишковості дискретних та неперервних повідомлень. Для зменшення надлишковості інформаційних потоків розроблена величезна кількість алгоритмів, призначених для того чи іншого типу вхідних даних і для того чи іншого споживача інформації.

В роботі пропонується спосіб формування службової інформації в адаптивних апертурних методах зменшення надлишковості інформаційних повідомлень, шляхом

кодування, елементами рекурентних кодових послідовностей, рівнів квантування.

II. ПОРІВНЯННЯ МЕТОДІВ ЗМЕНШЕННЯ НАДЛИШКОВОСТІ ДАНИХ

Для порівняння методів зменшення надлишковості між собою необхідно встановити систему єдиних числових характеристик. Такими характеристиками можуть бути: коефіцієнт стиснення, показник ефективності представлення повідомлень, коефіцієнт складності, час затримки повідомлення в кодері джерела, показник завадостійкості, порівняння з еталонною процедурою кодування, складність алгоритму декодування, час декодування.

Ефективність процедури обробки даних можна оцінити шляхом порівняння із заданою еталонною процедурою при заданій похибці відновлення повідомлення. В якості еталона можна вибрати:

оптимальний метод кодування джерела, який дозволяє реалізувати потенційні характеристики системи обробки,

стандартний метод, при якому повідомлення $f(t)$ рівномірно дискретизують в часі і в кожному відліку f_i співставляють кодове слово фіксованої довжини.

В першому випадку можна оцінити якість алгоритму або підсистеми зменшення надлишковості інформації по відношенню до максимального ефекту, який визначається ε -ентропією джерела.



Але цей метод оцінки має і суттєвий недолік: при цьому важко отримати максимальну швидкість передачі повідомлення для будь-якого джерела при заданій допустимій похибці ϵ . Необхідно також відзначити, що цей метод кодування потребує повної статистичної інформації про джерело, чого в реальній ситуації не буває. Це можна обійти, якщо в якості оптимальної процедури використовувати універсально-оптимальний або адаптивний кодер джерела з обмеженою затримкою.

В другому випадку ці недоліки відсутні. Коефіцієнт зменшення надлишковості повинен характеризувати зміну об'єму повідомлення оскільки зменшення надлишковості інформаційного потоку використовується для розвантаження каналів зв'язку і запам'ятовуючих пристроїв. Природно, що оцінку коефіцієнта стиснення доцільно проводити шляхом порівняння об'єму повідомлень, отриманого в результаті компресії, тобто при вибраному способі представлення (кодування) повідомлення, з об'ємом, що отримується при деякому початковому або еталонному способі формування повідомлень, тобто

$$K_1 = \frac{J_0}{J}$$

де J_0 - об'єм повідомлення вибраного методу представлення $f(t)$ на інтервалі $(0, T)$; J - об'єм повідомлення, що передається.

Коефіцієнт K_1 задовольняє основним вимогам, які ставляться до показника якості системи; він залежить від процесу функціонування системи та порівняно просто обчислюється на основі математичної моделі.

У випадку, коли об'єм даних J , які формуються, передаються або реєструються, розраховують без врахування службової інформації $J=J_a$, а кількість рівнів квантування на кожний інформативний елемент сталий і рівний m , коефіцієнт стиснення визначається:

$$K_2 = \frac{J_0}{J_a} = \frac{N' \cdot \log_2 m}{N \cdot \log_2 m} = \frac{N'}{N} \quad (1)$$

де N' - кількість елементів f_i масиву $\{f_i\}$, що представляє повідомлення $f(t)$ на інтервалі T ; N - кількість істотних відліків, отриманих на виході пристрою стиснення даних.

Достить часто цей коефіцієнт називається коефіцієнтом за кількістю відліків [1].

Використання адаптивних апертурних методів зменшення надлишковості дозволяє скоротити об'єми інформаційного потоку за рахунок вилучення елементів, тобто тих компонентів інформаційних потоків, які можуть

бути відновлені в границях заданої апертури. Недоліком існуючих методів адаптивного стиснення інформації є формування великих об'ємів службової інформації. Використання рекурентних кодових послідовностей дозволяє значно підвищити ефективність адаптивних методів та зменшити об'єми службової інформації.

Для визначення ефективності використання методів зменшення надлишковості інформаційних потоків важливим є визначення об'ємів формованих повідомлень, за якими можна визначити коефіцієнт стиснення.

При рівномірній дискретизації та квантуванню об'єм повідомлень можна визначити враховуючи кількість відліків та рівнів квантування (визначення об'єму за двійковими значеннями)

$$V = nE^{\wedge}[\log_2 m] \quad (2)$$

де n - кількість відліків,

m - кількість рівнів квантування,

$E^{\wedge}[\dots]$ - функція округлення до більшого цілого числа.

Аналогічно можна визначити об'єми інформації після компресії. Але потрібно враховувати службову інформацію, яка вноситься в інформаційний потік. Під службовою інформацією розуміється номер істотного відліку, або час його появи. Відповідно формула знаходження об'єму набуде вигляду:

$$V_{cm}^I = n'(E^{\wedge}[\log_2 m] + E^{\wedge}[\log_2 n]), \quad (3)$$

де n' - кількість істотних відліків.

Нехай маємо деякий цифровий сигнал:

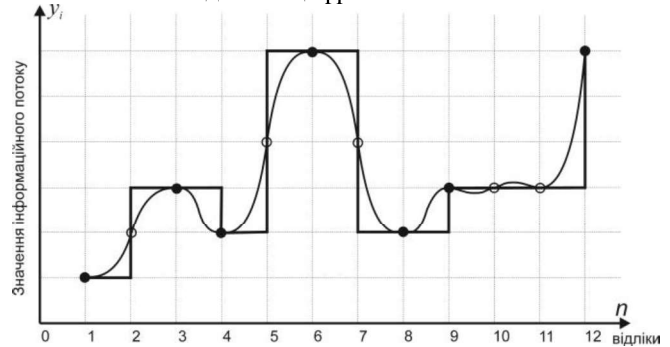


Рис. 1. Цифровий сигнал.

За допомогою першого метода представлення службової інформації отримаємо наступну бітову послідовність:

$$[\log_2 n]y_{12}[\log_2 n]y_9[\log_2 n]y_8[\log_2 n]y_6[\log_2 n]y_4[\log_2 n]y_3[\log_2 n]y_1$$

Відомі також і інші методи представлення службової інформації. Часто в методах зменшення надлишковості

інформації [2] використовують властивості РКП для прив'язки дискретів до часу їх появи. Ідея методу полягає



в наступному: кожен відлік представляється одним бітом РКП, при чому, якщо даний відлік істотний, то до бітової послідовності записується інвертований біт РКП і відповідне йому значення, якщо відлік неістотний, то передається прямий біт РКП.

Обчислення об'єму стиснутих повідомлень при прив'язці бітів РКП до дискретів буде здійснюється за формулою:

$$V_{cm}^{II} = n' E^{\wedge}[\log_2 m] + n \quad (4)$$

де n' - кількість істотних відліків,
 n - кількість дискретів,
 m - кількість рівнів квантування.

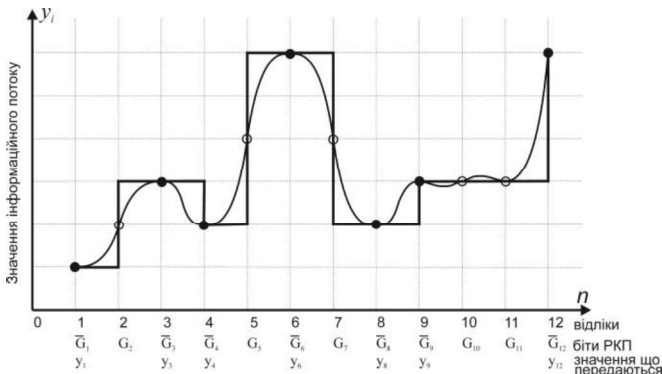


Рис. 2. Кодування відліків елементами РКП

Бітова послідовність даного способу представлення службової інформації матиме наступний вигляд

$$y_{12} \overline{G}_{12} G_{11} G_{10} y_9 \overline{G}_9 y_8 \overline{G}_8 G_7 y_6 \overline{G}_6 G_5 y_4 \overline{G}_4 y_3 \overline{G}_3 G_2 y_1 \overline{G}_1$$

Дослідивши відомі методи представлення службової інформації, зроблено висновок, що при певній частоті дискретизації, різних кількостях рівнів квантування та визначених кількостях істотних відліків дані методи не будуть ефективними. Тому актуальною задачею буде створення нового методу представлення службової інформації, який би дозволяв покращити зменшення надлишковості інформаційних потоків [3].

III. СПОСІБ ПРЕДСТАВЛЕННЯ ІНФОРМАЦІЇ ЗА ДОПОМОГОЮ КОДУВАННЯ РІВНІВ КВАНТУВАННЯ ЕЛЕМЕНТАМИ РКП

В роботі досліджується спосіб представлення службової інформації – кодування рівнів квантування, де кожен рівень квантування буде представлений бітом РКП [3].

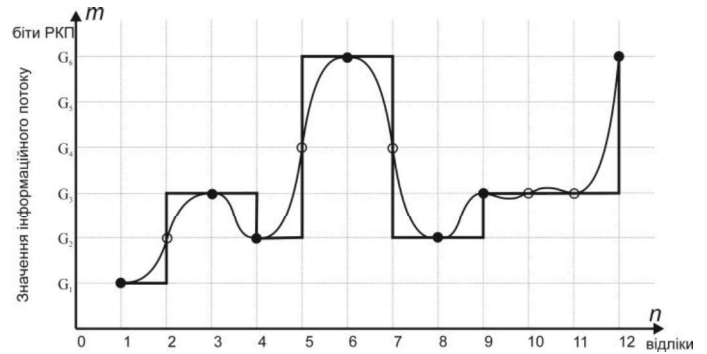


Рис. 3. Кодування рівнів квантування елементами РКП

Після знаходження всіх істотних відліків формується бітова послідовність даних. Якщо на відповідному рівні квантування є істотний відлік, то біт РКП інвертується, якщо немає істотних відліків – передається прямий біт РКП.

Якщо ж на деякому рівні квантування є кілька істотних відліків, то в потік потрібно внести додаткову службову інформаційну ознаку наявності ще одного істотного відліку на даному рівні квантування. В якості такої ознаки беремо біт РКП відповідного рівня. Якщо після істотного відліку ітиме прямий біт, то на даному рівні немає більше істотних відліків, якщо інвертований, то істотні відліки ще наявні. Після кожного інвертованого біта буде передаватись порядковий номер відповідного дискрета.

Таким чином бітова послідовність стиснутих даних набуде вигляду :

$$G_6 12 \overline{G}_6 6 \overline{G}_6 G_5 G_4 G_3 9 \overline{G}_3 3 \overline{G}_3 G_2 8 \overline{G}_2 4 \overline{G}_2 1 \overline{G}_1$$

$G_1, G_2, G_3 \dots$ - біти РКП, рівні квантування

Об'єм стиснутих таким чином даних становить:

$$V_{cm}^{III} = n' (E^{\wedge}[\log_2 n] + 1) + m,$$

або

$$V_{cm}^{III} = n' E^{\wedge}[\log_2 n] + n' + m \quad (5)$$

Визначивши об'єми повідомлень, отриманих в результаті компресії за допомогою запропонованого способу кодування повідомлення та за допомогою еталонного способу формування повідомлень, визначаємо коефіцієнт зменшення надлишковості для запропонованого методу представлення службової інформації:

$$k^{III} = \frac{V}{V^{III}} = \frac{n E^{\wedge}[\log_2 m]}{n' (E^{\wedge}[\log_2 n] + 1) + m}$$

На рис. 4 подана залежність величини коефіцієнта k^{III} від кількості істотних відліків



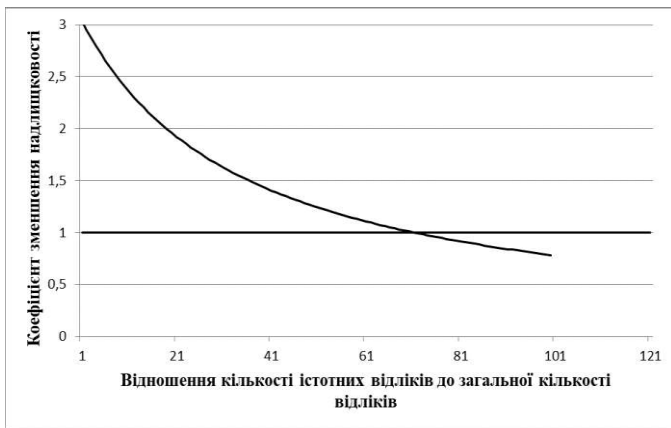


Рис. 4. Величина коефіцієнта зменшення надлишковості в адаптивних системах зменшення надлишковості інформації з використанням РКП для прив'язки істотних відліків до рівнів квантування

Для того, щоб запропонований спосіб представлення інформації був ефективніший за відомі способи представляється службової інформації, необхідно порівняти різницю об'ємів стиснутих повідомлень різними способами. Якщо різниці $V_{cm}^{III} - V_{cm}^I$ і $V_{cm}^{III} - V_{cm}^{II}$ будуть від'ємними, то запропонований новий метод можна вважати ефективнішим.

Порівняємо ефективність запропонованого методу представлення службової інформації з методом, в якому службова інформація представляється номерами істотних відліків. Розглянемо різницю об'ємів стиснутих повідомлень: $V_{cm}^{III} - V_{cm}^I < 0$

$$\begin{aligned} V_{cm}^{III} - V_{cm}^I &= n' E^{\wedge}[\log_2 n] + n' + m - n' E^{\wedge}[\log_2 m] - \\ &- n' E^{\wedge}[\log_2 n] = n' + m - n' E^{\wedge}[\log_2 m] = \\ &= n'(1 - E^{\wedge}[\log_2 m]) + m < 0 \end{aligned}$$

$$n'(1 - E^{\wedge}[\log_2 m]) < -m$$

Оскільки $1 - E^{\wedge}[\log_2 m] < 0$ то

$$n' > \frac{-m}{1 - E^{\wedge}[\log_2 m]}$$

Отже, якщо кількість істотних відліків перевищуватиме $\frac{m}{E^{\wedge}[\log_2 m] - 1} / n * 100\%$ від загальної

кількості відліків, то при використанні нового способу формування службової інформації при зменшенні надлишковості інформаційних потоків адаптивними апертурними методами з використанням рекурентних

кодових послідовностей можна отримати вищі коефіцієнти стиснення.

Варто зауважити, що в даному випадку для ефективності запропонованого методу кількість істотних відліків не залежить від загальної кількості відліків. Розрахуємо мінімальні кількості істотних відліків при різній кількості рівнів квантування.

ТАБЛИЦЯ І. МІНІМАЛЬНА КІЛЬКІСТЬ ІСТОТНИХ ВІДЛІКІВ НЕОБХІДНИХ ДЛЯ ЕФЕКТИВНОСТІ МЕТОДУ

m	$(-m)/(1 - E^{\wedge}[\log_2 m])$
4	4
8	4
16	5
32	8
64	13
128	20
256	37
512	64
1024	114
2048	205
4096	372

З таблиці можна побачити, наприклад, що при кількості рівнів квантування 256 кількість істотних відліків повинна складати не менше 37 або $(37/n) \cdot 100\%$ від загальної кількості відліків.

Порівняємо ефективність розробленого методу представлення службової інформації з методом, в якому службова інформація формується за допомогою бітів РКП. Розглянемо різницю об'ємів стиснутих повідомлень:

$$V_{cm}^{III} - V_{cm}^{II} < 0$$

$$\begin{aligned} V_{cm}^{III} - V_{cm}^{II} &= n' E^{\wedge}[\log_2 n] + n' + m - n' E^{\wedge}[\log_2 m] - n = \\ &= n'(E^{\wedge}[\log_2 n] - E^{\wedge}[\log_2 m] + 1) + m - n < 0 \\ &n'(E^{\wedge}[\log_2 n] - E^{\wedge}[\log_2 m] + 1) < n - m \end{aligned}$$

$$n' < \frac{n - m}{E^{\wedge}[\log_2 n] - E^{\wedge}[\log_2 m] + 1}, \text{ при } n > m$$

$$n' > \frac{n - m}{E^{\wedge}[\log_2 n] - E^{\wedge}[\log_2 m] + 1}, \text{ при } n < m$$

Згідно отриманих нерівностей побудовані дві таблиці відповідно для двох випадків.

Якщо частота дискретизації менша від кількості рівнів квантування, то кодування рівнів квантування недоречно, оскільки завжди буде отриманий програш в об'ємі стиснутих даних.



ТАБЛИЦЯ II. МІНІМАЛЬНА КІЛЬКІСТЬ ІСТОТНИХ ВІДЛІКІВ ПРИ $m > n$

n \ m	2	4	8	16	32	64	128	256	512	1024	2048	4096
2			6	7	10	16	25	42	73	128	228	409
4				12	14	20	31	50	85	146	256	455
8					24	28	40	62	101	169	291	511
16						48	56	80	124	202	339	583
32							96	112	160	248	403	677
64								192	224	320	496	806
128									384	448	640	992
256										768	896	1280
512											1536	1792
1024												3072
2048												
4096												

ТАБЛИЦЯ III. МАКСИМАЛЬНА КІЛЬКІСТЬ ІСТОТНИХ ВІДЛІКІВ ПРИ $m < n$

n \ m	2	4	8	16	32	64	128	256	512	1024	2048	4096
2												
4	1											
8	2	2										
16	4	4	4									
32	6	7	8	8								
64	10	12	14	16	16							
128	18	21	24	28	32	32						
256	32	36	41	48	56	64	64					
512	57	64	72	83	96	112	128	128				
1024	102	113	127	144	165	192	224	256	256			
2048	186	204	227	254	288	331	384	448	512	512		
4096	341	372	409	453	508	576	661	768	896	1024	1024	

За допомогою табл. 3 можна вирахувати, що при кількості рівнів квантування 256 і кількості відліків 512 кількість істотних відліків не повинна перевищувати 128 відліків або ж 25% від загальної кількості відліків.

ВИСНОВКИ

Аналіз отриманих результатів показує, що якщо кількість рівнів квантування менша ніж кількість дискретів, то при кількості істотних відліків, яка не

перевищує величини $(n - m) / \left(\log_2 \frac{n}{m} + 1 \right)$ доцільніше

кодувати рівні квантування. Тобто, запропонований у роботі метод буде ефективнішим, коли кількість рівнів квантування менша ніж кількість дискретів і кількість істотних відліків не перевищує

$$(n - m) / \left(\log_2 \frac{n}{m} + 1 \right) \cdot 100\%$$

від загальної кількості відліків.

ЛІТЕРАТУРА REFERENCES

- [1] Ю.Ю.Ляш Критерії та оцінка ефективності зменшення надлишковості інформаційних потоків/ Ляш Ю.Ю. // «Восточно-Европейский журнал передовых технологий» № 1/2(55).2012, 64 с. – с.54–58.
- [2] Y. Iliash Data compression based on THE recurrent sequences of Galois In monography Information Technology in Selected Areas of Management. / Sc.Ed. Lyubomyr Petryshyn. Published by AGH University of Science and Technology Press. KU 0651. Krakow 2016. (152 p.) – P. 91-101.
- [3] Ляш Ю.Ю. Формування стиснутих повідомлень шляхом кодування рівнів квантування / Ляш Ю.Ю. // Інформаційні технології та комп'ютерна інженерія: матеріали статей п'ятої міжнар. наук.-практ. конф., м. Івано-Франківськ, 27-29 травня 2015р. – Івано-Франківськ: Супрун В.П., 2015. – С.100-101.

