



УДК 004.67

© А. В. Левенец, В. В. Нильга, 2009

## СТРУКТУРНОЕ УПОРЯДОЧЕНИЕ ДАННЫХ ДЛЯ ЗАДАЧ СЖАТИЯ В ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫХ СИСТЕМАХ

*Левенец А. В.* – канд. техн. наук, доц. кафедры «Автоматика и системотехника», тел.: +79141913339, e-mail: levalvi@bk.ru; *Нильга В. В.* – асп. кафедры «Автоматика и системотехника», тел.: +79242038639, e-mail: Vasiliy.Nilga@gmail.com (ТОГУ)

Сжатие измерительного сигнала является одной из актуальных задач современных информационно-измерительных систем. Существующие методы сжатия в основном оперируют с линейным представлением измерительных данных. Предлагаемый способ позволяет реализовать алгоритмы оптимального преобразования одномерного представления измерительного сигнала в двумерную форму с целью повышения эффективности процедур сжатия.

Compressing the measuring signal is one of the urgent tasks of modern information-measurement systems. Compression methods now in use deal mainly with a linear presentation of measured data. The proposed method allows one to implement the optimum algorithms for conversion of the one-dimensional presentation of the measuring signal into two-dimensional form with a view to improving the efficiency of compression.

*Ключевые слова:* предварительная обработка, информационно-измерительные системы, сжатие, способ, эффективность.

Современное развитие информационно-измерительных систем привело к увеличению потоков измерительной информации. Данные, получаемые от источников (датчиков), весьма разнообразны и зачастую весьма объемны. В то же время существующие каналы связи, а также аппаратные возможности ИИС в области обработки измерительной информации все чаще не в состоянии справиться с растущими потоками информации. Вследствие чего необходимо разрабатывать новые подходы к обработке измерительной информации еще до ее передачи по каналам связи.

Существующие подходы к сжатию измерительной информации предполагают работу с каждым источником в отдельности, точнее на одиночный поток данных линейной структуры. Для такой модели исходных данных сейчас разработаны эффективные алгоритмы сжатия, такие как арифметическое

кодирование, методы Хаффмана, LZW и др. Тем не менее, для задач реального времени на настоящий момент практически актуальным является только разностный метод с последующим кодированием кодами переменной длины, так как остальные методы дают ощутимое запаздывание и/или требовательны к вычислительным затратам. Вследствие такого подхода не учитываются корреляционные связи между источниками, обусловленные не только физическими взаимосвязями, но и технической реализацией самой измерительной системы, особенностями методов кодирования, используемых в ИИС и т. п.

С этой точки зрения логичнее разрабатывать алгоритмы кодирования, работающие с кадрами определенной структуры. Для дальнейшего развития методов сжатия измерительных данных необходимо исследовать статистические характеристики поля источников информации и разработать способы его преобразования в более структурированное представление [1], что позволит повысить эффективность работы даже существующих алгоритмов сжатия.

Любую линейную структуру можно разбить на блоки равной длины и составить из них  $n$ -мерный массив, к примеру, двумерный. Данный способ преобразования данных кратко можно охарактеризовать как «строка  $\rightarrow$  массив». Предложенный в [2; 3; 4] способ основан на поиске корреляционных зависимостей в одиночном потоке измерительных данных и анализе максимумов полученной автокорреляционной функции с целью выявления оптимального интервала разбиения исходной последовательности.

Вследствие нетривиальности пути нахождения оптимальной длины блока возникает необходимость оценки качества предварительной обработки. Для чего предполагается ввести понятие «дерево». Под термином «дерево» понимается совокупность расположенных в непосредственной близости (по вертикали или горизонтали) одинаковых элементов. На рис. 1 показан пример разбиения исходного массива двоичных данных на деревья, которые выделены пунктирными линиями.

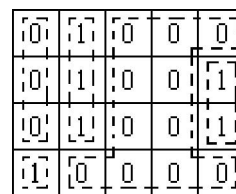


Рис. 1. Пример выделения деревьев матрицы

Очевидно, что однородность данных показатель отображает степень упорядоченности данных в результате предварительной обработки и его можно использовать для оценки эффективности предварительной обработки. Большое число деревьев отображает высокую неоднородность данных и, как следствие, низкую эффективность их сжатия. Следует отметить значительное влияние формы дерева на эффективность преобразования. Так, мощное, но



обладающее сложной структурой («разветвленное») дерево может привести к образованию значительного числа маломощных деревьев, что не может являться положительным результатом. Форма деревьев также играет важную роль в оценке эффективности предварительной обработки. Слишком большие и изогнутые деревья порождают множество мелких деревьев, что, в свою очередь, сказывается на увеличении общего числа деревьев. Также предполагается, что средняя мощность, число элементов дерева также может характеризовать предварительную обработку.

Среднюю мощность дерева можно рассчитать исходя из числа элементов  $N$  в массиве данных и числа деревьев  $k$  в том же массиве следующим образом:

$$n_{cp} = \frac{N}{k}.$$

Идеальное преобразование в случае двоичных элементов исходного массива даст два дерева: одно, состоящее из нулей, которое назовем «нулевым» и одно, состоящее из единиц, которое назовем «единичным». Таким образом, для оценки эффективности преобразования можно использовать коэффициент эффективности  $k_{\text{Э1}}$ , вычисляемый как отношение числа деревьев при идеальном преобразовании  $k_{\text{ид}}$  к числу деревьев при реальном преобразовании  $k$ :

$$k_{\text{Э1}} = \frac{k_{\text{ид}}}{k} = \frac{2}{k}.$$

Очевидно, что коэффициент  $k_{\text{Э1}}$  будет стремиться к единице, и чем выше будет значение данного коэффициента, тем выше эффективность преобразования данных.

Также, учитывая, что данные до преобразования являются предельным случаем двумерного массива (представляет собой строку), эффективность преобразования можно оценивать как отношение числа деревьев до  $k_1$  и после  $k_2$  преобразования. В этом случае можно ввести следующий коэффициент эффективности:

$$k_{\text{Э2}} = \frac{k_1}{k_2} = \frac{n_{cp2}}{n_{cp1}},$$

где  $n_{cp1}$  – средняя мощность деревьев до преобразования;

$n_{cp2}$  – средняя мощность деревьев после преобразования.

Такой коэффициент эффективности показывает степень изменения числа деревьев после предварительной обработки.

Для оценки результатов работы предложенного способа также следует оценить пределы диапазонов изменения предложенных коэффициентов эффективности. Здесь необходимо уточнить, что предложенное преобразование ведется для двоичных чисел, поэтому отбрасывается тривиальный случай однородного массива данных. В этом случае минимальное число деревьев  $k_{\text{min}} = 2$ , а максимально возможное число деревьев  $k_{\text{max}} = N$ .

Соответственно коэффициент  $k_{\mathcal{E}1}$  ограничен пределами:

$$\frac{2}{k_{\max}} \leq k_{\mathcal{E}1} \leq \frac{2}{k_{\min}} \Rightarrow \frac{2}{N} \leq k_{\mathcal{E}1} \leq 1.$$

Очевидно, что чем ближе значение  $k_{\mathcal{E}1}$  к единице, тем выше эффективность преобразования данных.

Аналогично, коэффициент  $k_{\mathcal{E}2}$  будет иметь следующие ограничения:

$$\frac{k_{\min}}{k_{\max}} \leq k_{\mathcal{E}2} \leq \frac{k_{\max}}{k_{\min}} \Rightarrow \frac{2}{N} \leq k_{\mathcal{E}2} \leq \frac{N}{2}.$$

В этом случае эффективность преобразования считается положительной, если значение коэффициента  $k_{\mathcal{E}2}$  лежит в пределах  $1 \dots N/2$ , что обуславливает неувеличение числа деревьев после преобразования данных.

Так как целью предложенной обработки данных является повышение эффективности процедур сжатия, логично предложить и второй вариант оценки эффективности преобразования. Сравнение коэффициентов сжатия для исходной последовательности  $k_{\text{сж}1}$  и преобразованных данных  $k_{\text{сж}2}$  покажет, является ли предложенный алгоритм предварительной обработки эффективным для задач сжатия. С этой целью выбран простой и распространенный алгоритм сжатия RLE. Применение алгоритма RLE для исходной битовой последовательности измерительных данных и полученной 2-мерной структуры позволяет вычислить коэффициент эффективности:

$$k_{\mathcal{E}3} = \frac{k_{\text{сж}1}}{k_{\text{сж}2}}.$$

Коэффициент эффективности  $k_{\mathcal{E}3}$  показывает количественное изменение коэффициента сжатия алгоритма RLE после предварительной обработки данных. Очевидно, что преобразование данных является эффективным, если значения  $k_{\mathcal{E}3}$  больше единицы.

Следует отметить, что применение алгоритма RLE также обусловлено его возможностью сжимать и двумерные матрицы, выбрав направление обхода данных. В данном случае для преобразованных данных выбран обход по вертикали, при прочих равных параметрах алгоритма RLE.

Проверка работоспособности предлагаемой предварительной обработки и предварительная оценка границ её функциональности были проведены на искусственных сигналах следующего типа:

1) Прямоугольный сигнал (меандр). В силу тривиальности сигнала исследования проводились для фиксированного числа периодов на объеме выборки. Периоды принимали значения 10 %, 30 % и 100 % от длины выборки.

2) Синусоидальный сигнал. Период сигнала принимает такие же значения, что и для меандра.

3) Случайный сигнал с равномерным распределением (ССРР) в интервале  $[-127, 127]$ . Это обусловлено восемью битами, отводимыми под кодирование одного отчета измерительного сигнала.



4) Случайный сигнал с нормальным распределением (ССНР), математическое ожидание которого равно нулю, а среднеквадратичное отклонение (СКО) принимает значения в диапазоне 1...40. Данный ряд обусловлен желанием проанализировать работу алгоритмов при различной амплитуде при фиксированной максимальной амплитуде (8 бит), так при СКО = 1 по правилу трех сигм амплитуда практически никогда не превысит 3 единиц, а при СКО = 40 – не превысит 120 единиц, что приближается к максимальной амплитуде.

Длина выборки для всех искусственных сигналов изменялась в диапазоне 512...2400 бит, в то время как длина интервала разбиения попадала в диапазон от 8 до четверти длины выборки.

Полученные значения коэффициентов эффективности приведены в табл. 1, 2, 3.

Таблица 1

**Коэффициенты эффективности для меандра**

Длина выборки	k <sub>Э1</sub>			k <sub>Э2</sub>			k <sub>Э3</sub>		
	Процентное отношение периода к длине выборки								
	10 %	30 %	100 %	10 %	30 %	100 %	10 %	30 %	100 %
512	0,133	0,081	0,168	9,24	5,44	10,72	2,566	1,910	2,864
1024	0,070	0,060	0,124	9,39	7,92	15,07	2,600	1,984	2,800
1600	0,048	0,046	0,089	9,79	9,33	17,22	2,612	2,571	2,404
2400	0,033	0,036	0,062	10,05	10,91	18,62	2,646	2,571	2,660

Таблица 2

**Коэффициенты эффективности для синусоидального сигнала**

Длина выборки	k <sub>Э1</sub>			k <sub>Э2</sub>			k <sub>Э3</sub>		
	Процентное отношение периода к длине выборки								
	10 %	30 %	100 %	10 %	30 %	100 %	10 %	30 %	100 %
512	0,069	0,034	0,047	7,550	4,320	4,890	3,194	1,410	1,242
1024	0,033	0,014	0,020	7,800	3,550	5,620	3,635	1,541	1,455
1600	0,025	0,012	0,017	9,040	4,070	6,400	3,817	1,283	1,803
2400	0,024	0,009	0,012	12,410	5,060	7,260	3,627	1,580	1,932

Таблица 3

**Коэффициенты эффективности для ССРР**

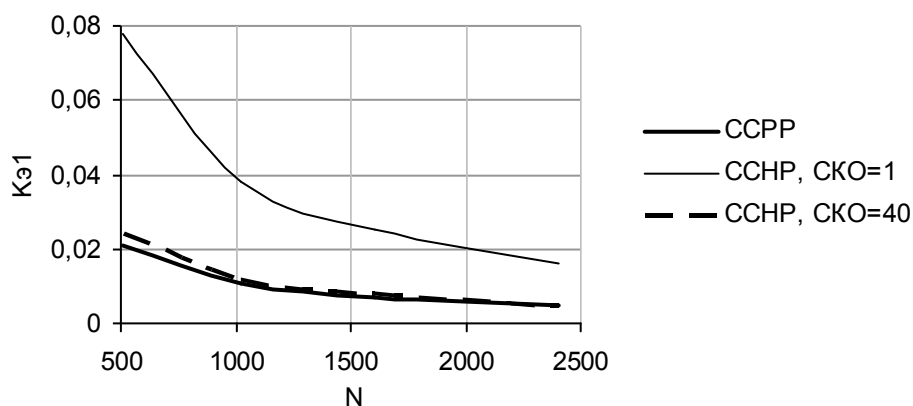
Длина выборки	k <sub>Э1</sub>	k <sub>Э2</sub>	k <sub>Э3</sub>
512	0,021	2,750	1,050
1024	0,011	2,800	1,050
1600	0,007	2,800	1,046
2400	0,005	2,820	1,041

Таблица 4

		Коэффициент $k_{Э1}$ эффективности для ССНР								
	Длина выборки	Среднеквадратичное отклонение								
		1	5	10	15	20	25	30	35	40
$k_{Э1}$	512	0,078	0,034	0,031	0,029	0,028	0,027	0,026	0,025	0,024
	1024	0,038	0,017	0,015	0,015	0,014	0,013	0,013	0,013	0,012
	1600	0,025	0,011	0,01	0,009	0,009	0,006	0,008	0,008	0,008
	2400	0,016	0,007	0,006	0,006	0,006	0,006	0,006	0,005	0,005
$k_{Э2}$	512	1,436	2,17	2,506	2,606	2,707	2,721	2,749	2,762	2,72
	1024	1,443	2,18	2,447	2,621	2,669	2,722	2,749	2,756	2,77
	1600	1,453	2,18	2,434	2,586	2,664	2,704	2,748	2,748	2,77
	2400	1,451	2,18	2,443	2,578	2,68	2,721	2,755	2,775	2,78
$k_{Э3}$	512	0,555	0,641	0,709	0,773	0,819	0,859	0,889	0,914	0,987
	1024	0,553	0,636	0,703	0,768	0,812	0,853	0,882	0,907	0,978
	1600	0,559	0,643	0,705	0,764	0,81	0,848	0,878	0,903	0,969
	2400	0,555	0,645	0,701	0,762	0,806	0,844	0,872	0,899	0,965

Как можно наблюдать из табл. 1 и табл. 2, все три коэффициента показывают положительный результат, не зависимо от длины выборки и периода сигнала.

Для анализа поведения коэффициентов эффективности построены графики для случайных процессов с нормальным распределением, как наиболее интересных для измерительных систем. На рис. 2 приведены графики поведения коэффициента  $k_{Э1}$  для различных типов входных сигналов в зависимости от длины выборки.


 Рис. 2. Коэффициент эффективности  $k_{Э1}$  для случайных сигналов



Результаты проведенных исследований показывают, что увеличение длины выборки приводит к уменьшению коэффициента эффективности  $k_{Э1}$ . Такой результат можно объяснить усложнением структуры данных и, как следствие, ростом числа деревьев. Рост числа деревьев происходит в прямой пропорциональности к росту длины выборки, что объясняет гиперболичность графиков вследствие обратной пропорциональности  $k_{Э1}$  от числа деревьев.

На рис. 3 приведены графики поведения коэффициента  $k_{Э2}$  в зависимости от значения СКО. График построен для длины выборки входного сигнала, равного 512, поскольку, как видно из таблицы, длина выборки практически не оказывает влияния на поведение коэффициента. Как видно из графиков, увеличение СКО способствует улучшению коэффициента  $k_{Э2}$ . Такое поведение может быть объяснено только неполным заполнением разрядной сетки, т. е. амплитуда случайного сигнала не достигает максимально возможной для восьми разрядов, для СКО, равного 1, амплитуда равна 3 единицам, а для СКО, равного 40, амплитуда равна 120. Несколько лучшие результаты коэффициента  $k_{Э2}$ , как это можно наблюдать из табл. 3, достигаются при равномерном распределении случайной величины.

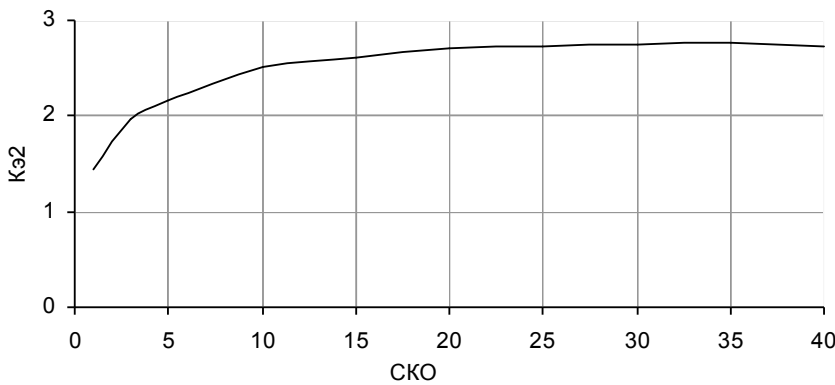


Рис. 3. Коэффициент эффективности  $k_{Э2}$  для случайных сигналов ( $N = 512$ )

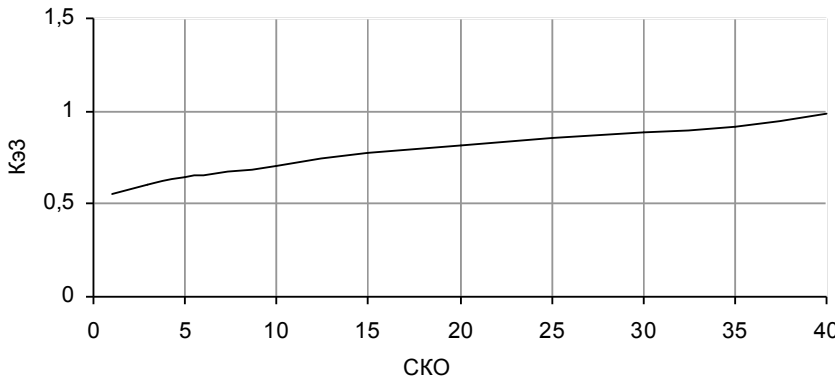


Рис. 4. Коэффициент эффективности  $k_{Э3}$  для случайных сигналов ( $N = 512$ )

На рис. 4 приведены графики поведения коэффициента  $k_{Э3}$  в зависимости от значения СКО. График построен для длины выборки входного сигнала, равного 512, поскольку, как видно из таблицы, длина выборки практически не отказывает влияния на поведение коэффициента. При нормальном распределении коэффициент  $k_{Э3}$  показывает неэффективность предварительной обработки, причем результат ухудшается при уменьшении значения СКО. Несколько лучшие результаты, как это можно наблюдать из табл. 3, достигаются при равномерном распределении случайной величины

В итоге, основываясь на значениях коэффициентов эффективности, видим, что применение предложенного способа предварительной обработки в большинстве случаев имеет положительные результаты. Так, для тривиальных сигналов все три коэффициента показывают положительный результат, независимо от длины выборки и периода сигнала. В случае же случайного сигнала коэффициенты  $k_{Э1}$  и  $k_{Э2}$  показывают эффективность предварительной обработки как для нормального, так и для равномерного распределения, в то время как коэффициент  $k_{Э3}$  показывает неэффективность предварительной обработки для нормального распределения случайной величины, причем увеличение среднеквадратичного отклонения положительно сказывается на данном коэффициенте. В случае с равномерным распределением коэффициент  $k_{Э3}$  показывает хоть и незначительное, но все же положительное влияние предварительной обработки.

Следует отметить, что предложенный метод можно распространить на структуры более высокой размерности, однако это требует дополнительного объема исследований.

### Библиографические ссылки

1. *Ватолин Д., Ратушняк А., Смирнов М.* Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. М., 2003.
2. *Левенец А. В., Нильга В. В.* Алгоритм предварительной обработки измерительных данных для повышения эффективности процедур сжатия // Перспективы развития информационных технологий: сборник материалов I Всероссийской научно-практической конференции. Новосибирск, 2008.
3. *Левенец А. В., Нильга В. В.* Алгоритм двумерного представления измерительных данных для задач сжатия // Информационные и управляющие системы : сб. науч. тр. Хабаровск, 2008.
4. *Nilga V. V.* Transformation of one-dimensional representation measuring signal in the bidimensional form // Material of the Seventh International Young Scholars' Forum of the Asia-Pacific Region Countries. Vladivostok, 2007.