

УДК 004.67

# ПРЕДСТАВЛЕНИЕ ТЕЛЕМЕХАНИЧЕСКИХ ДАННЫХ ОДНОРОДНЫМИ $n$ -МЕРНЫМИ СТРУКТУРАМИ КАК ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА В ЗАДАЧАХ СЖАТИЯ

**Чье Ен Ун,**

доктор техн. наук, профессор

**А. В. Левенец,**

канд. техн. наук, доцент

**В. В. Нильга,**

аспирант

Тихоокеанский государственный университет

Предложен способ предварительной обработки измерительных данных, основанный на их представлении однородными  $n$ -мерными структурами и позволяющий повысить эффективность сжатия таких данных. Разработан критерий оценки эффективности предварительной обработки и проведены исследования нового способа предварительной обработки на случайных процессах и данных телемеханики.

**Ключевые слова** — предварительная обработка, сжатие, однородная  $n$ -мерная структура, информационно-измерительная система, телемеханика.

## Введение

Существующая на настоящий момент тенденция широкого использования распределенных информационно-измерительных систем при решении самых разнообразных задач привела к значительно повышенным требованиям к пропускной способности применяемых каналов связи, что связано как с увеличением объемов передаваемой информации, так и с необходимостью обеспечения защиты этой информации от помех и несанкционированного доступа. Следствием этого является завышенная стоимость каналов связи из-за индивидуального подхода при их проектировании. Сокращение объема передаваемой информации является самым распространенным решением данной проблемы, при этом следует отметить, что основным требованием, предъявляемым при сжатии данных распределенных информационно-измерительных систем, является обратимость сжатия или сжатие без потерь. Здесь следует уточнить, что речь идет о цифровых данных, т. е. таких данных, для которых принципиально возможно осуществить процедуру обратимости.

## Предварительная обработка в сжатии данных

Широко известные универсальные алгоритмы сжатия без потерь используют простые статистические модели, поэтому либо не могут обеспечить требуемый уровень сжатия, либо имеют сложные для технической реализации алгоритмы кодирования и декодирования, что существенно сокращает область их применения, особенно при работе с телемеханическими данными, обычно поступающими в режиме реального времени.

В общем случае процесс сжатия телемеханических данных можно разделить на следующие три основных этапа [1].

1. Сегментирование или выделение квазистационарных участков, для которых статистические свойства считают неизменными. В результате непрерывный поток данных разбивается на блоки отсчетов, что позволяет упростить обработку данных и повысить эффективность алгоритмов сжатия.

2. Предварительная обработка, необходимая для приведения исходных данных к модели, лежащей в основе последующего метода кодирова-

ния. Наиболее широко развиты линейные предсказания (авторегрессионная модель) и линейные преобразования, например дискретно-косинусное.

3. Кодирование. Лидирующие позиции здесь удерживают методы энтропийного кодирования, например арифметическое кодирование, кодирование Хаффмана.

Следует отметить, что этап предварительной обработки данных играет весьма важную роль с точки зрения эффективности процедуры сжатия в целом. Учитывая достаточно хорошую на текущий момент проработку методов сегментирования и кодирования, следует связать дальнейшее повышение эффективности алгоритмов сжатия с разработкой методов и способов предварительной обработки данных.

По способу обработки входной последовательности предварительную обработку можно разделить на два вида:

1) обработка потока — последовательности с неизвестными границами и последовательным доступом;

2) обработка блоков — конечных последовательностей с произвольным доступом.

В результате предварительной обработки возможно формирование нескольких потоков (блоков), причем даже если их суммарная длина больше длины исходной последовательности, структура сформированных потоков (блоков) значительно улучшает последующее сжатие как по коэффициенту сжатия, так и по времени.

### Способ предварительной обработки данных

Предлагаемый способ предварительной обработки измерительных данных относится к блочным методам. Суть способа заключается в трансформации блоков измерительных данных в целях увеличения их однородности.

*Однородность* выходного потока определяется наличием серий, состоящих только из нулей или только из единиц, длина которых стремится к максимально возможной. Идеальный вариант преобразования данных даст всего две серии, одна из которых будет состоять из нулей, а другая — из единиц. Длина серии в этом случае равна количеству всех нулей (всех единиц) во входном потоке и является максимально возможной.

Условие обеспечения максимальных длин серий можно использовать для оценки эффективности предварительной обработки. В качестве критерия можно предложить коэффициент эффективности  $k_{эф}$ , определяемый как отношение средних длин серий после ( $l_{после}$ ) и до ( $l_{до}$ ) предварительной обработки:  $k_{эф} = l_{после} / l_{до}$ .

Очевидно, что эффективность преобразования декларируется при  $k_{эф} > 1$ , т. е. в том случае, ког-

да происходит увеличение средней длины серии. Учитывая, что средняя длина — это отношение длины блока  $N$  к числу серий  $n$ :  $\bar{l} = N/n$ ,  $k_{эф}$  можно представить следующим образом:  $k_{эф} = n_{до} / n_{после}$ .

Способ предварительной обработки измерительных данных, предлагаемый в настоящей работе, является дальнейшим развитием алгоритмов, подробно рассмотренных в публикациях [2, 3]. Для предлагаемого способа характерно, что входная последовательность бит делится на равные блоки, длина которых определяется структурой кадров телемеханики. Предварительная обработка каждого блока ведется независимо от других, что налагает определенные ограничения при использовании энтропийных методов сжатия.

Суть способа предварительной обработки заключается в вычислении автокорреляционной функции (АКФ) битовой последовательности, которая позволяет обнаружить скрытые зависимости как в различных каналах в кадре телемеханики, так и в самих кадрах. Учитывая, что обрабатываются двоичные последовательности, АКФ предложено рассчитывать следующим образом:

$$\left\{ \begin{aligned} R(n) &= \frac{1}{N} \left( \sum_{i=0}^{N-1} S[i] \leftrightarrow S[i+n] - \sum_{i=0}^{N-1} S[i] \leftrightarrow S[i+n] \right), \\ S[i+n] &= S[(i+n) - N] \text{ при } (i+n) > N-1 \end{aligned} \right.$$

где  $S[i]$  —  $i$ -й бит исходной битовой последовательности  $S$ ;  $i \in [0, N-1]$  — параметр битовой последовательности;  $N$  — число бит в исходной последовательности;  $n \in [0, N-1]$  — параметр АКФ;  $\leftrightarrow$  — логическая операция эквивалентности.

Из рассчитанных значений АКФ делаются выборки с интервалом  $m$  [4,  $N/4$ ]. Среди полученных выборок определяется единственная, которая имеет максимальное математическое ожидание при минимальной дисперсии. Соответствующий данной выборке интервал предлагается назвать основным и обозначать  $m_{осн}$ . Из входной последовательности, разбитой на блоки по  $m_{осн}$  бит, можно сформировать двумерный массив размером  $m_{осн} \times \text{ceil}(N / m_{осн})$ , где  $\text{ceil}()$  — функция округления в большую сторону до ближайшего целого. Для получения большей размерности структуры производится поиск дополнительных интервалов  $m_i$ , обладающих максимальным математическим ожиданием при минимальной дисперсии. Единственным дополнительным условием является то, что каждый последующий интервал должен быть кратен интервалу, найденному на предыдущем шаге. Диапазон интервалов лежит в пределах от четырех до  $N/4$ . Количество найденных интервалов задает размерность струк-

туры. Таким образом, в результате формируется  $M$ -мерная структура с размерами  $m_0 \times m_1/m_0 \times \dots \times m_i/m_{i-1} \times \dots \times \text{ceil}(N/m_{M-2})$ .

### Результаты исследований

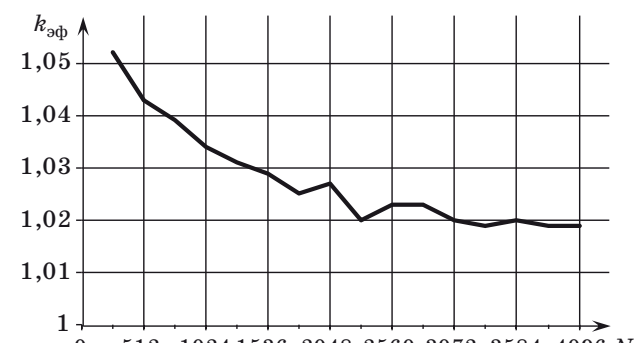
Исследование возможностей предварительной обработки было произведено на случайных последовательностях данных, а также на данных телемеханики, полученных от ряда энергетических объектов Дальневосточного региона.

Случайные последовательности были представлены реализациями случайного процесса с равномерным распределением (СПРР). Исследования проводились для различных длин выборки. Для обеспечения статистической репрезентативности эксперимент для каждой длины выборки проводился 500 раз, а в качестве результатов приводятся усредненные значения.

Поскольку имеющиеся в распоряжении данные телемеханики были восьмиразрядными, то исследования проводились на длинах, кратных восьми. Объем выборки обрабатываемых данных  $N$  менялся в диапазоне от 256 до 4096 бит с шагом 256 бит. Для каждого объема выборки определялся коэффициент эффективности, график поведения которого представлен на рис. 1.

Анализ приведенных данных показывает, что предлагаемый способ обработки эффективен даже в случае его применения к случайным данным. Уменьшение значения  $k_{эф}$  при увеличении длины выборки может быть объяснено снижением корреляционной зависимости обрабатываемых данных, что приводит к увеличению количества локальных максимумов корреляционной функции и, в свою очередь, к неэффективному выбору основного интервала.

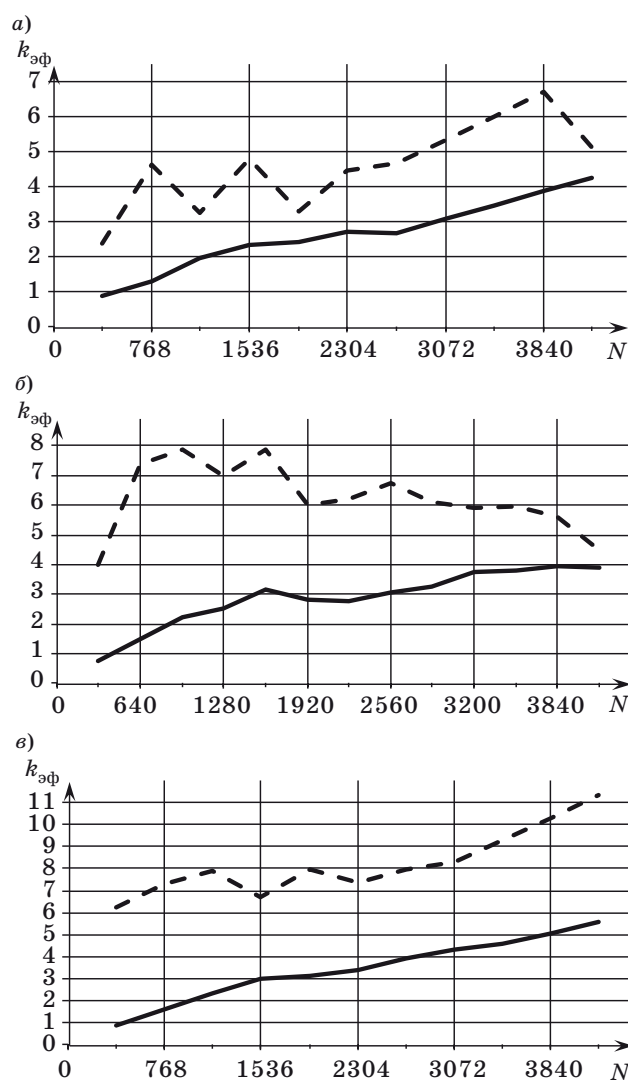
Данные телемеханики представляли собой ряд отдельных потоков, поступающих от разных объектов. Каждый поток имеет свою длину кадра, обусловленную числом каналов (датчиков). Коэффициент эффективности рассчитывался для каждого потока в отдельности.



■ Рис. 1. Влияние длины выборки на  $k_{эф}$  для СПРР

Кроме того, исследовались разностные потоки телемеханических данных, каждый кадр которых составлялся из разностей значений двух соседних кадров исходного потока. Здесь необходимо отметить, что поскольку исходный кадр обрабатываемой последовательности передается только в начале процесса обработки и в дальнейшем работа ведется уже с разностями, в проведенных исследованиях объем исходного кадра не учитывался.

Типичное поведение коэффициента эффективности при изменении длины выборки представлено на примерах трех потоков (рис. 2, а—в). На рисунках сплошной линией показаны результаты предварительной обработки исходных данных, а пунктирной — результаты предварительной обработки разностных рядов. Каждый график представляет линейную интерполяцию значений коэффициента эффективности предварительной обработки данных.



■ Рис. 2. Графики  $k_{эф}$  для первого (а), второго (б) и третьего (в) потоков данных телемеханики

■ *Влияние предварительной обработки на эффективность сжатия*

Сжимаемая последовательность	Значение коэффициента сжатия		
	худшее	усредненное	лучшее
Исходная	0,51	0,53	0,54
С предварительной обработкой	0,45	1,48	2,15
С предварительной обработкой разностных рядов	0,95	3,74	5,33

Как следует из полученных результатов, при обработке исходных данных наблюдается общий тренд увеличения значений коэффициента эффективности с увеличением числа кадров, находящихся в предварительной обработке, что обусловлено наличием связей как между кадрами, так и между каналами в кадре. В отличие от случайного процесса, увеличение объема данных, участвующих в предварительной обработке, позволяет повысить вероятность обнаружения этих связей. Негативный результат проявляется при обработке одного кадра, что позволяет определить нижнюю границу эффективности применения предлагаемого алгоритма.

При работе с разностными рядами наблюдается существенное повышение эффективности предварительной обработки, что связано, прежде всего, с низкой динамичностью телемеханических данных, в силу чего разность соседних кадров позволяет получить более однородную последовательность для дальнейшей обработки и, как следствие, более высокую эффективность преобразования. Очевидно, что эффективность предварительной обработки будет максимальной при обработке стационарных данных со спектральной массой, сосредоточенной в области низких частот.

Необходимо отметить следующее. Во-первых, выявить общий тренд в полученных данных для разностных потоков не удается, что можно объяснить индивидуальным характером динамики в каждом потоке телемеханических данных. Во-вторых, размерность получаемых структур не превышает трех измерений, причем доля двумерных структур для данных в виде реализации СПРР составила 0,5 %, а для данных телемеханики — 19 %, в то время как доля трехмерных структур составляет 95,5 и 81 % соответственно.

В качестве предварительного анализа работы предлагаемого способа предварительной обработки проводилось сжатие исходной последовательности и полученной структуры кодированием длин повторов (Run-Length Encoding — RLE). Поскольку данный алгоритм предназначен для работы с линейными данными, получение выходной линейной последовательности осуществляет-

ся путем обхода  $n$ -мерной структуры по жестко заданному алгоритму для всех обрабатываемых данных. Полученные результаты для данных телемеханики отображены в таблице.

Таким образом, применение предварительной обработки позволяет увеличить коэффициент сжатия алгоритмом RLE в среднем в 2,8 раза, а с применением разностного метода — в 7 раз. Как и предполагалось, худшие результаты получены при обработке одного кадра.

### Заключение

Таким образом, предлагаемый способ предварительной обработки позволяет потенциально повысить эффективность сжатия за счет получения более однородной последовательности данных. На случайных процессах, не имеющих сильных корреляционных зависимостей, эффективность способа существенно ниже и обратно пропорциональна длине выборки. В случае же работы с данными телемеханики средняя длина серий увеличивается до 6 раз, причем эффективность прямо пропорционально зависит от объема обрабатываемых данных. Следует также отметить существенное (до 2 раз) повышение эффективности предложенного способа предварительной обработки при работе с разностными рядами данных телемеханики.

Очевидно, что эффективность предлагаемого способа предварительной обработки для реальных информационно-измерительных систем будет существенно зависеть от применяемого алгоритма сжатия. Наиболее очевидным выбором может стать алгоритм RLE, работающий с сериями повторяющихся символов (бит). Однако такой алгоритм слабо приспособлен к сжатию многомерных структур, поэтому для получения максимального эффекта следует разрабатывать специализированные алгоритмы, учитывающие особенности предлагаемого способа предварительной обработки.

### Литература

1. Хан J., Камбер M. Data Mining: Concepts and Techniques. — Elsevier Inc., 2006. — 800 p.
2. Левенец А. В., Нильга В. В. Структурное упорядочение данных для задач сжатия в информационно-измерительных системах // Вестник ТОГУ. 2009. № 2(13). С. 45–52.
3. Чье Ен Ун, Левенец А. В., Нильга В. В. Алгоритм структурного упорядочивания измерительных данных // Информатика и системы управления. 2010. № 3(25). С. 82–87.