

УДК 004.67

© 2010 г. **А.В. Левенец**, канд. техн. наук,
В.В. Нильга,
Чье Ен Ун, д-р техн. наук
(Тихоокеанский государственный университет, Хабаровск)

АЛГОРИТМ СТРУКТУРНОГО УПОРЯДОЧИВАНИЯ ИЗМЕРИТЕЛЬНЫХ ДАННЫХ¹

Одним из “узких мест” современных информационно-измерительных систем являются каналы связи. Для устранения этой проблемы активно используется сжатие передаваемой информации. Предлагаемый способ позволяет реализовать алгоритмы упорядочивания измерительных данных, основываясь на анализе скрытых корреляций, с целью повышения эффективности процедур сжатия.

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

Введение

Увеличение потока измерительных данных является неотъемлемым следствием развития информационно-измерительных систем. И зачастую каналы связи является “узким местом” во всей системе. Для снижения требований к пропускной способности канала активно используется сжатие передаваемой информации. В настоящий момент существует значительное количество различных методов сжатия цифровой информации, ориентированных на тот или иной тип сжимаемых данных и обладающих приемлемыми коэффициентами сжатия [1]. Но эти методы сжатия измерительной информации предполагают работу с каждым источником в отдельности, точнее, на одиночный поток данных линейной структуры. Для повышения эффективности алгоритма сжатия часто используется предварительная (препроцессорная) обработка сжимаемых данных, которая видоизменяет входной поток таким образом, чтобы коэффициент сжатия преобразованных данных был в среднем выше коэффициента сжатия исходных (непреобразованных). Модули препроцессор-постпроцессор обычно применяются в сочетании с различными кодерами и архиваторами, повышая их степень сжатия и, возможно, скорость работы.

¹ Работа выполнена при поддержке федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009-2013 гг., в рамках реализации мероприятия № 1.3.2. Проведение научных исследований целевыми аспирантами (ГК № П2029).

Тем не менее для задач реального времени в 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. Но не следует забывать о корреляционных связях между источниками, обусловленные не только физическими взаимосвязями, но и технической реализацией самой измерительной системы, особенностями методов кодирования, используемых в ИИС, и т.п. Для дальнейшего развития методов сжатия измерительных данных необходимо исследовать статистические характеристики поля источников информации и разработать способы его преобразования в более структурированное представление, что позволит повысить эффективность работы даже существующих алгоритмов сжатия.

Алгоритм преобразования

Наиболее простой способ – это представление линейной последовательности измерительных данных некой структурой, – например, двумерным массивом. Пусть дана выборка измерительного сигнала, полученная с четырехразрядного АЦП и представленная в двоичной системе счисления. Данные такого вида можно представить, например, в виде массивов (рис. 1) с длиной строки, равной пяти.

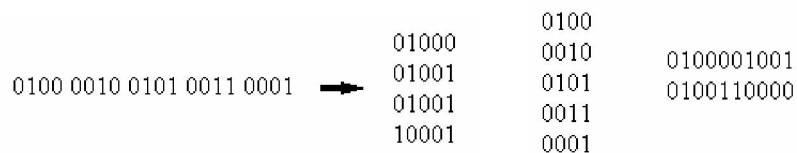


Рис. 1. Варианты преобразования измерительных данных.

Как видно из представленных рисунков, длина строки влияет на упорядоченность данных, т. е. при правильной длине строки получаем структуру с однородными областями нулей и единиц. Группировка способствует образованию однородных областей, что повышает эффективность алгоритмов сжатия. Главной задачей при выполнении такого преобразования является поиск оптимальной длины строки массива, т. е. оптимального интервала разбиения исходных данных. Также следует заметить, что преобразование формально не ограничивается двумерностью структуры, в общем случае можно получить n -мерные структуры.

Предлагаемый алгоритм является развитием алгоритма, предложенного в работе [2], который основан на определении автокорреляционной функции (АКФ) измерительных данных. Учитывая, что анализируются двоичные последовательности, АКФ предложено рассчитывать следующим образом:

$$R_c(n) = \frac{1}{N-1} \sum_{i=0}^{N-1-n} S[i] \leftrightarrow S[i+n], \quad S[i+n] = 0 \text{ при } i+n > N-1, \quad (1)$$

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

Операция логической эквивалентности применяется вследствие работы с двоичными последовательностями.

Поиск оптимальной длины блока будем проводить следующим образом:
исходный измерительный сигнал преобразуется в битовую последовательность (прямой код, дополнительный и т. п.);

осуществляется выборка N последовательных значений для дальнейшей обработки;

находится автокорреляционная функция измерительного сигнала по формуле (1). Полученная АКФ также является последовательностью длиной N ;

вычисленная АКФ разбивается на блоки по M элементов (бит) ;

в каждом блоке АКФ производится поиск максимального значения и запоминается его место в блоке, места максимумов объединяются в массив, первый максимум игнорируется в силу тривиальности своего расположения;

находятся математическое ожидание (МО) и дисперсия полученных значений, причем математическое ожидание округляется до целого значения, так как МО и будет являться оптимальным интервалом разбиения;

шаги с 4 по 6 повторяются для различных M , значения M изменяются в пределах от 4 до $N/4$ (это обусловлено малой вероятностью появления существенных значений АКФ вне этого диапазона);

формируется таблица значений дисперсий и МО при различных M ;

производится сортировка столбцов таблицы по возрастанию значений дисперсии;

отбираются несколько первых значений МО (определяется глубина поиска), а для каждого значения МО строится своя матрица значений, имеющая число столбцов, равное округленному до целого МО, и последовательно заполненная двоичными значениями исходного измерительного сигнала;

для каждой матрицы подсчитывается число деревьев;

выбирается оптимальная длина блока, которому соответствует наименьшее число деревьев.

Полученный в результате предварительной обработки массив характеризуется числом деревьев, которое отображает степень упорядоченности данных в результате предварительной обработки. Большое число деревьев отображает высокую неоднородность, что не может являться положительным результатом. Форма деревьев также играет важную роль в оценке работы алгоритма предварительной обработки. Слишком большие и изогнутые деревья порождают множество мелких деревьев, что в свою очередь сказывается на увеличении общего числа деревьев. Для первичной оценки проведенного преобразования необходимо ввести некоторую оценочную величину с учетом числа деревьев.

Эффективность преобразования можно оценивать как отношение числа деревьев (k_1) до и числа деревьев (k_2) после преобразования. В этом случае оценочную величину может рассчитываться как k_1/k_2 . В этом случае эффективность преобразования считается положительной, если значение k_2 превышает единицу, что характеризует уменьшение числа деревьев после преобразования данных.

Измерительные сигналы по своей природе происхождения весьма разнообразны. Их форма и динамика развития изменяются от слабо меняющихся (суточное изменение температуры воздуха) до резких скачков (регистрация толчков

земной коры). Для исследования работы алгоритма предложено воспользоваться тестовыми сигналами.

В качестве тестовых сигналов были выбраны: меандр, синусоида, случайный сигнал с равномерным распределением. Каждый из видов тестовых сигналов задан рядом параметров, которые могут изменяться, создавая различные вариации сигналов. Все тестовые последовательности являются биполярными, восьмиразрядными. При их преобразовании в двоичную систему исчисления использовался дополнительный код.

При проведении экспериментов все возможные варианты параметров тестовых сигналов принимали различные значения, но наиболее характерные результаты были получены при следующих параметрах: меандр и синусоидальный сигнал: период принимает значения 10%, 30% и 100% от длины выборки, отношение сигнал/шум (ОСШ) принимает значение 70 дБ; случайный сигнал с равномерным распределением: интервал распределения задается разрядностью и не превышает максимально возможной амплитуды.

Длина выборки всех искусственных сигналов принимается равной 256, 512, 1024, 1600, 2400, 3000 и 4000 бит. Данные параметры также были выбраны с целью более полного исследования работы алгоритма предварительной обработки измерительных данных на различных тестовых сигналах.

Результаты экспериментальных и теоретических исследований

Результаты предварительной обработки сигналов типа меандр приведены в табл. 1, а на рис. 2 эти данные представлены в графическом виде.

Таблица 1

| Длина выборки N | Математическое ожидание | | | | | |
|--------------------|--|------|-------|--------------------------------|-------|-------|
| | Оптимальной длины | | | k ₁ /k ₂ | | |
| | процентное отношение периода к длине выборки | | | | | |
| | 10% | 30% | 100% | 10% | 30% | 100% |
| 256 | 24 | 7,72 | 8,4 | 7,62 | 3,6 | 7,57 |
| 512 | 48 | 8 | 10,88 | 9,24 | 5,44 | 10,72 |
| 1024 | 104 | 8 | 14,72 | 9,39 | 7,92 | 15,07 |
| 1600 | 160 | 8 | 14,81 | 9,79 | 9,33 | 17,22 |
| 2400 | 240 | 8 | 14,24 | 10,05 | 10,91 | 18,62 |
| 3200 | 320 | 16 | 12,56 | 10,09 | 15,54 | 18,14 |
| 4000 | 400 | 16 | 11,16 | 10,55 | 17,09 | 17,4 |

Анализ полученных данных показывает, что увеличение длины выборки приводит к улучшению структурированности данных. Такой эффект объясняется увеличением объема анализируемых данных при определении оптимальной длины строки. При низком уровне шума лучшие результаты получены при длине периода, равной длине выборки, в то время как при других периодах мы также имеем приемлемые результаты. В общем случае структурированность данных, характеризующая оценочной величиной, в 3,6...17,4 раз превышала исходную.

Синусоидальный сигнал представляет собой более сложный вариант реализации измерительных данных. Результаты предварительной обработки синусоидальных сигналов приведены в табл. 2 (рис. 3).

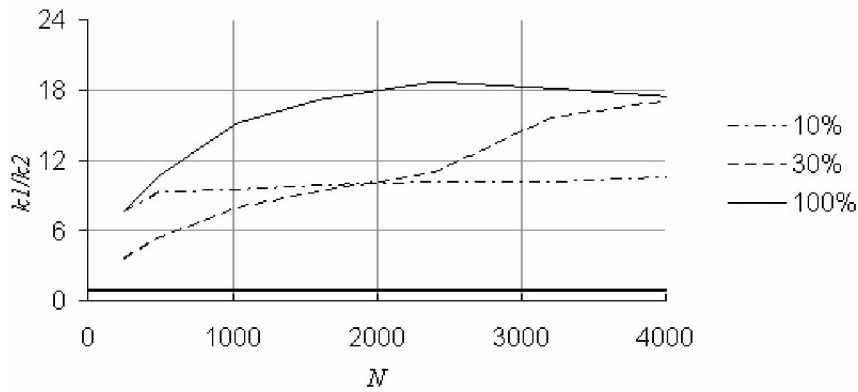


Рис. 2. Оценка работы алгоритма для сигналов, типа меандр.

Таблица 2

| Длина выборки N | Математическое ожидание | | | | | |
|-----------------|--|------|------|--------------------------------|-------|-------|
| | Оптимальной длины | | | k ₁ /k ₂ | | |
| | процентное отношение периода к длине выборки | | | | | |
| | 10% | 30% | 100% | 10% | 30% | 100% |
| 256 | 24 | 7,43 | 8 | 3,010 | 2,760 | 3,460 |
| 512 | 48 | 8 | 8 | 3,210 | 3,370 | 4,220 |
| 1024 | 104 | 7,18 | 8 | 3,360 | 4,020 | 4,740 |
| 1600 | 160 | 8 | 8 | 4,100 | 4,140 | 4,920 |
| 2400 | 240 | 8 | 8 | 4,160 | 4,500 | 4,940 |
| 3200 | 320 | 8 | 8 | 4,360 | 4,630 | 5,010 |
| 4000 | 400 | 8 | 8 | 4,780 | 4,750 | 5,010 |

Анализ результатов показывает, что увеличение длины выборки приводит к улучшению структурированности данных, что объясняется увеличением объема анализируемых данных при определении оптимальной длины строки. При низком уровне шума лучшие результаты получены при наименьшем из рассматриваемых периоде сигнала, в то время как при других периодах наблюдаются приемлемые результаты. В общем случае структурированность данных, характеризующая оценочной величиной, в 2,8...5,0 раз превышала исходную.

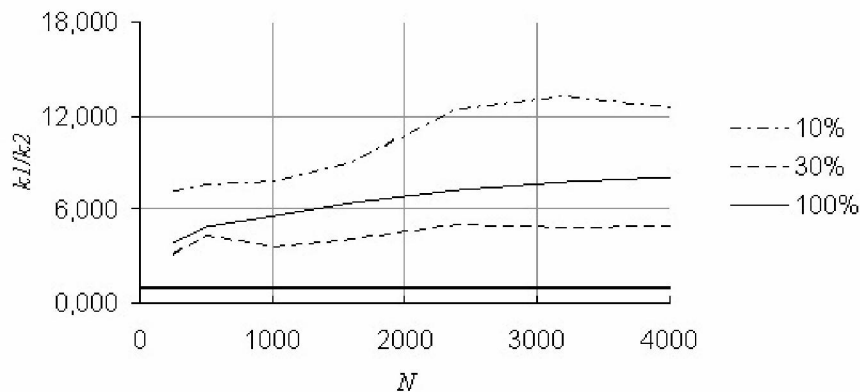


Рис. 3. Оценка работы алгоритма для синусоидальных сигналов.

Случайные процессы являются наиболее сложным типом измерительного сигнала. Результаты предварительной обработки случайных сигналов приведены в табл. 3, которые в графическом виде даны на рис. 4.

| Длина выборки N | Математическое ожидание | |
|-----------------|-------------------------|-----------|
| | Оптимальной длины | k_1/k_2 |
| 256 | 20,96 | 2,671 |
| 512 | 40,14 | 2,667 |
| 1024 | 60,24 | 2,747 |
| 1600 | 51,52 | 2,883 |
| 2400 | 40,07 | 2,954 |
| 3200 | 20,60 | 3,002 |
| 4000 | 18,27 | 3,003 |

Анализ полученных данных позволяет выявить, что и для случайных процессов с равномерным распределением мы получаем тенденции, характерные для первых двух видов измерительных сигналов.

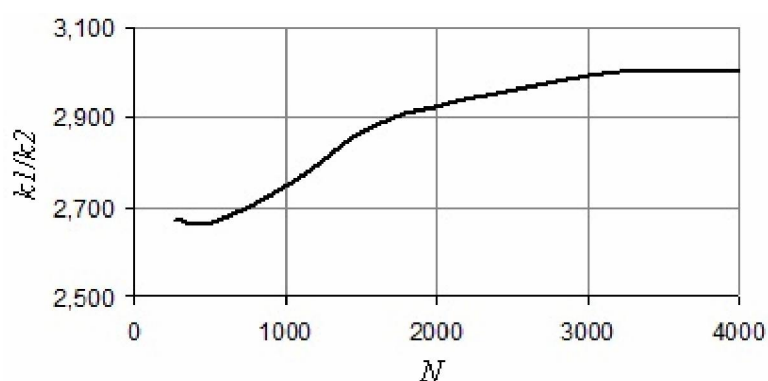


Рис. 4. Оценка работы алгоритма для случайных сигналов.

В общем случае структурированность данных, характеризующая оценочной величиной, в 2,7...3,0 раза превышала исходную.

Заключение

В качестве предварительной обработки измерительных данных предложен способ преобразования исходных данных, позволяющий повысить эффективность их сжатия за счет более полного учета внутренних корреляционных зависимостей данных. Способ основан на трансформации представления данных, из последовательности бит, байт и т.п. в двумерные структуры. Результаты проведенных исследований показали, что для различных измерительных сигналов структурированность данных превышала исходную от 2,7 до 7,4 раза.

ЛИТЕРАТУРА

1. Ватолин Д. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. – М., 2003.
2. Левенец А.В., Нильга В.В. Алгоритм двумерного представления измерительных данных для задач сжатия // Информационные и управляющие системы. – Хабаровск: Изд-во ТОГУ. 2008. – С. 68-74.

E-mail:

Нильга В.В. – Vasiliy.Nilga@gmail.com.