



УДК 004.627

© 2015 г. **И.В. Богачев,**
А.В. Левенец, канд. техн. наук,
Чье Ен Ун, д-р техн. наук
(Тихоокеанский государственный университет, Хабаровск)

ГЕОМЕТРИЧЕСКИЙ ПОДХОД К СЖАТИЮ ДАННЫХ ТЕЛЕМЕТРИЧЕСКИХ СИСТЕМ

Предлагается алгоритм сжатия телеметрических данных, в основе которого лежит отображение исходных данных, представленных в битовом виде, на поверхность тора с последующим поиском однородных областей. Показано, что алгоритм обеспечивает хорошую степень сжатия телеметрических данных различного типа.

Ключевые слова: телеметрические данные, преобразование данных, алгоритм Кадана, сжатие данных.

Введение

В настоящее время, несмотря на все технические достижения, перед разработчиками информационно-измерительных систем (ИИС) остро продолжает стоять вопрос о балансе между экономической и функциональной эффективностью. В качестве мер снижения итоговой стоимости ИИС обычно предлагается использовать низкоскоростные каналы связи, снижать объемы информационных хранилищ, выбирать менее быстродействующие вычислительные устройства и т.п. Такие меры не всегда приемлемы, особенно в случае телеметрических систем, для которых характерны большие объемы передаваемой информации и жесткие требования к времени ее передачи. Выходом из такой ситуации стало развитие специальных алгоритмов сжатия данных, которые позволяют представить исходные данные в более компактном виде, учитывая ее избыточность [1].

Все существующие на настоящий момент алгоритмы сжатия можно разбить на две группы: универсальные и специализированные. Алгоритмы первой группы не привязаны к типу сжимаемых данных, обеспечивая некоторое среднее значение степени сжатия. К универсальным алгоритмам можно отнести классический алгоритм Хаффмана, коды переменной длины, арифметическое сжатие и др. В свою очередь, специализированные алгоритмы разработаны для эффективного сжатия конкретного типа данных, однако для данных других типов их применение зачастую весьма неэффективно. Так, в работе [2] показано, что использование

ряда наиболее распространенных алгоритмов сжатия изображений для сокращения избыточности телеметрических данных, где основным требованием являются отсутствие потерь в данных и высокая средняя степень сжатия, в целом оказывается малоэффективным, хотя и дает заметный прирост степени сжатия для некоторых типов данных.

В современных телеметрических системах наиболее часто используются широко известные универсальные алгоритмы сжатия данных, готовых к передаче по цифровому каналу связи [3]. Очевидно, что в ряде случаев прибегают к специализированным алгоритмам, информация о которых, однако, практически всегда является закрытой. Таким образом, существует практическая необходимость в разработке высокоэффективных методов сжатия телеметрических данных.

Постановка задачи

К особенностям телеметрических данных сложных технических систем следует отнести наличие корреляции не только между отсчетами одного датчика в соседних кадрах, но и между отсчетами внутри кадра [4]. Важно отметить, что существующие открытые алгоритмы сжатия не учитывают взаимосвязь данных внутри одного кадра, хотя можно предположить, что такой учет позволил бы повысить эффективность сжатия.

Для телеметрических данных справедливо предположить, что в штатном режиме работы системы в большинстве случаев арифметическая разность между одним и тем же отсчетом в двух соседних кадрах будет составлять не более четырех единиц (без учета знака). Такое свойство предлагается называть «стационарностью» потока кадров и далее в статье термин будет использоваться именно в этом смысле.

Из свойства стационарности следует, что в качестве метода сжатия первого уровня эффективным будет широко известный алгоритм дельта-кодирования, представляющий собой поэлементное вычисление операции арифметической разности между одними и теми же отсчетами в текущем и предыдущем кадрах, что позволяет существенно снижать динамический диапазон данных. Если телеметрический кадр содержит отсчеты n датчиков, то разностный кадр можно представить в виде вектор-столбца \mathbf{d} размерностью n :

$$\mathbf{d} = (d_1, d_2, \dots, d_j, \dots, d_n)^T.$$

Предполагая, что телеметрические данные имеют разрядность k , каждый элемент вектор-столбца d_i можно представить в виде битовой последовательности и записать в виде вектор-строки \mathbf{b} размерности k :

$$d_i = \mathbf{b}_i = (b_{i,1}, b_{i,2}, \dots, b_{i,j}, \dots, b_{i,k}).$$

Таким образом, вектор-столбец \mathbf{d} можно представить в виде матрицы \mathbf{B} , размерностью $n \times k$:

$$\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_j, \dots, \mathbf{b}_n)^T.$$

Такое рассмотрение сжимаемых данных позволит совместить операции уменьшения динамического диапазона данных после дельта-кодирования и обнаружения корреляций между соседними отсчетами. Действительно, рассматривая полученное битовое поле, можно выделять и кодировать однородные области

(единичные или нулевые), несмотря на то, что такая область будет состоять из бит, принадлежащих отсчетам от разных датчиков. Этот принцип предлагается использовать для разработки алгоритма сжатия телеметрических данных. Например, для кадра, состоящего из n отсчетов в предельном случае отсутствия изменений в соседних кадрах, традиционный подход будет заключаться в передаче на приемную сторону n флагов неизменности, а для предложенного принципа в таком случае будет передаваться только один флаг.

Алгоритм сжатия

Таким образом, в основе предлагаемого алгоритма лежит поиск однородных областей битового представления телеметрического кадра. После обнаружения найденные зоны предлагается кодировать двумя парами координат $[x_1, y_1]$ и $[x_2, y_2]$, которые задают положение соответственно левого верхнего и правого нижнего углов выделенной области. Подобный подход позволяет сжимать выделенные области, размер которых больше количества бит, отводимого на хранение координат области.

Для удобства представления и работы с данными предлагается отображать полученную матрицу **B** на поверхности некоторой геометрической фигуры. В качестве такой фигуры предлагается использовать тор, что позволяет в отличие от классического расположения данных на поверхности, ограниченной прямоугольной областью, более просто проводить циклический поиск однородных зон, снимая ограничения, вводимые границами прямоугольной области. Пример расположения данных на прямоугольной области и на поверхности тора показан на рис. 1, где темным цветом обозначены единичные биты, а белым – нулевые.

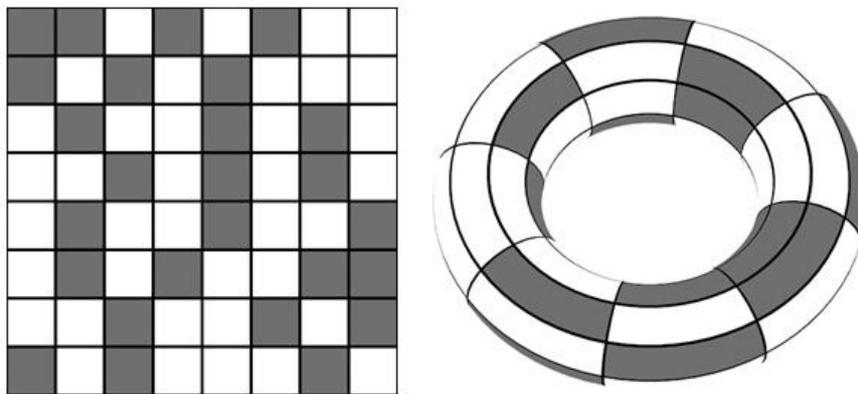


Рис. 1. Представление результата дельта-кодирования на прямоугольной плоскости и на поверхности тора.

Преобразование данных между прямоугольной областью и поверхностью тора происходит согласно следующим формулам:

$$b_{i,j} = t_{i',j'}; \quad i = i' + n \cdot \left\lfloor \frac{i}{n} \right\rfloor; \quad j = j' + k \cdot \left\lfloor \frac{j}{k} \right\rfloor,$$

где i – индекс строки, указывающий на номер отсчета в кадре данных; j – индекс, указывающий на номер разряда отсчета; i' и j' – представление индексов i и j в системе координат тора; $b_{i,j}$ – бит отсчета, представленный в системе координат прямоугольной области; $t_{i',j'}$ – бит отсчета, представленный в системе координат

тора.

В качестве алгоритма поиска однородных областей целесообразно реализовать такой метод, при котором в процессе поиска происходила бы одновременная оценка влияния найденной на текущем шаге области на общий коэффициент сжатия. В связи с этим было принято решение использовать модернизированный для работы с трехмерной циклической поверхностью алгоритм, разработанный Д. Каданом, в исходном виде предназначенный для поиска подотрезка вектор-строки с максимальной суммой [5].

Условно модифицированную версию алгоритма можно описать как последовательность из следующих шагов:

- 1) матрица **V** отображается на поверхность тора;
- 2) левые и правые столбцы последовательно фиксируются внутри матрицы **V**, причем расстояние, на которое могут быть удалены зафиксированные столбцы, не превышает k ;
- 3) последовательно фиксируются нижние и верхние строки внутри матрицы **V**, причем расстояние, на которое могут быть удалены зафиксированные строки, не превышает n ;
- 4) для промежуточных вычислений используется вектор-строка **t**, с размерностью n ;
- 5) в каждый столбец вектор-строки **t** записывается сумма всех столбцов i -й строки;
- 6) вектор-строка **t** анализируется алгоритмом одномерного поиска подотрезка вектора-строки с максимальной суммой;
- 7) если сумма на текущем этапе работы, превышает ранее рассчитанную, то границы текущей зоны поиска фиксируются и поиск продолжается дальше. Если же сумма на текущем этапе работы не превышает ранее рассчитанную, то поиск продолжается без фиксации границ;
- 8) поиск ведется до тех пор, пока циклически не будет обработана вся матрица **V**.

Необходимо отметить, что в своем исходном виде матрица **V** не применима для работы с модифицированным алгоритмом Кадане в силу того, что сумма всех элементов в однородной зоне всегда будет равна нулю. С целью устранить возникшие сложности была введена дополнительная процедура, названная взвешиванием матрицы, основанная на формуле, служащей для определения степени влияния выделенной области на общий коэффициент сжатия [6]:

$$W = (x_2 - x_1)(y_2 - y_1) - \sum_{i=x_1}^{x_2} \sum_{j=y_1}^{y_2} b_{ij}z,$$

где W – оценочная характеристика, показывающая степень влияния выделенной области на коэффициент сжатия; z – количество бит, необходимое на кодирование координаты одной ячейки; x_1, y_1, x_2, y_2 – координаты области.

Из формулы следует, что оценка W равна общему числу бит в выделенной области, за вычетом количества бит, необходимого на кодирование позиции единичных бит. Из этого следует, что процедура взвешивания должна происходить следующим образом: для бит, принимающих значение единицы, присваивается

вес, равный количеству бит данных, необходимых на кодирование их позиции в матрице, взятый с отрицательным знаком, а для бит, принимающих значение нуля, присваивается вес равный единице. В процессе работы алгоритма вес каждой ячейки, вошедшей в найденную область, становится равен произведению величин n и k , взятому с отрицательным знаком.

Работу алгоритма можно продемонстрировать на следующем примере. Пусть исходная матрица \mathbf{B} с размерностью 4×8 имеет следующий вид:

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

В этом случае для кодирования координат однородной области $[x_1, y_1]$ и $[x_2, y_2]$ потребуется десять бит. Взвешенная матрица \mathbf{B}_v , все единичные биты которой приняли вес равный минус пяти (так как на кодирование позиции бита в матрице необходимо количество бит равное пяти), а все нулевые приняли единичное значение, будет иметь следующий вид:

$$\mathbf{B}_v = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & -5 \\ 1 & 1 & 1 & 1 & 1 & 1 & -5 & -5 \\ 1 & 1 & 1 & 1 & 1 & 1 & -5 & 1 \\ 1 & 1 & 1 & 1 & -5 & 1 & 1 & -5 \end{bmatrix}.$$

Для полученной взвешенной матрицы наибольшую сумму (в размере 18 единиц) имеет область, координаты которой принимают следующие значения: $x_1 = 1, y_1 = 1, x_2 = 6, y_2 = 4$. Таким образом, можно определить все однородные области в исходной матрице \mathbf{B} .

Форматы сжатых данных

В результате работы алгоритма необходимо сформировать пакет, в который будет запакован исходный кадр с целью его дальнейшей передачи по каналу связи. При формировании пакета могут возникнуть следующие ситуации.

1. Исходный кадр поддается сжатию и при этом значения всех отсчетов не изменились (нулевой кадр). В таком случае пакет состоит из двух бит: F_c – флага сжимаемости кадра (устанавливается если кадр поддается сжатию) и F_{const} – флага неизменности значений отсчетов кадра (устанавливается, если результат дельта-кодирования над кадром не равен нулю).

2. Исходный кадр не поддается сжатию. Пакет состоит из нулевого флага сжимаемости F_c и n бит несжатых данных B_d .

3. Кадр данных стационарный (в приведенном выше смысле), поддается сжатию и при этом значения хотя бы некоторых отсчетов изменились. В таком случае заголовок пакета состоит из шести бит флагов $F_c, F_{\text{const}}, F_{\text{st}}, F_s, F_{\text{FSB}}$ и F_{hc} , после которых идут поля описания сжимаемых областей: Q_0, C_0, Q_1, C_1 и B_d .

Флаг F_{st} фиксирует стационарность кадра (устанавливается, если кадр стационарен), флаг F_s показывает наличие отрицательных отсчетов (устанавливается

если результат дельта-кодирования над кадром содержит отрицательные отсчеты), флаг F_{FSB} фиксирует использование младшего разряда (устанавливается если разряд используется), а флаг F_{hc} определяет ситуацию неполного сжатия (устанавливается, если в кадре остались несжатые данные).

Поля описания имеют следующий смысл: поля Q_0 и Q_1 описывают количество нулевых и единичных областей соответственно, поля C_0 и C_1 задают координаты нулевых и единичных областей соответственно, поле B_d включает в себя биты несжатых данных.

4. Кадр данных нестационарный, поддается сжатию, но при этом значения хотя бы некоторых отсчетов изменились. В таком случае заголовок пакета состоит из пяти бит флагов F_c , F_{const} , F_{st} , F_s и F_{hc} , после которых идут поля описания сжимаемых областей: Q_0 , C_0 , Q_1 , C_1 и B_d .

Тестирование алгоритма сжатия

Предложенный алгоритм тестировался с помощью нескольких наборов телеметрических данных, полученных от разных объектов. Для исследования использовались данные как со стационарными свойствами, так и нестационарного характера. Объем наборов данных составлял 10 – 20 тысяч кадров. Кадр данных в рассмотренном случае представлял собой набор однобайтных отсчетов датчиков, причем для разных наборов данных кадры включали разное число датчиков (от 32 до 56).

На рис. 2 представлены типичные графики поведения степени сжатия для стационарных и нестационарных данных. На графиках по оси ординат используется логарифмический масштаб, а по оси абсцисс отложен порядковый номер сжимаемого кадра.

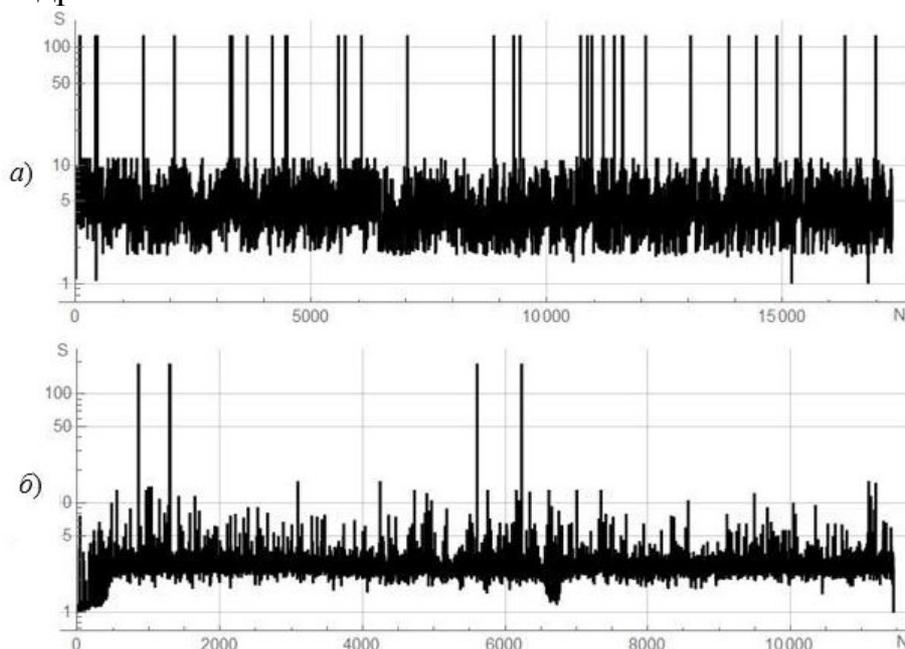


Рис. 2. Поведение степени сжатия S при работе алгоритма со стационарными (а) и нестационарными (б) данными.

Следует отметить, что при работе с нестационарными данными для алгоритма возможны переходные периоды, в течение которых происходит спад, а за-

тем достаточно плавный выход среднего значения степени сжатия на некоторый постоянный уровень. Этот факт можно связать с внезапными существенными изменениями свойств потока сжимаемых данных.

Анализ результатов сжатия стационарных потоков кадров показывает довольно частое появления максимально возможной для алгоритма степени сжатия, что еще раз показывает корректность приведенных выше рассуждений.

В целом, согласно проведенным исследованиям, для стационарных данных алгоритм показывает среднюю степень сжатия 4,87, а для нестационарных – 3,0.

Заключение

Результаты проведенных исследований показали, что алгоритм достаточно эффективно справляется с поставленной задачей, – следовательно, предложенный подход к учету корреляции между соседними отсчетами корректен.

Анализируя полученные данные, можно отметить, что при сжатии стационарных наборов тестовых данных фиксируется достаточно большое число кадров, для которых обеспечивается максимально возможная степень сжатия. Также необходимо отметить, что выход на рабочее значение средней степени сжатия при работе с нестационарными наборами тестовых данных в ряде случаев происходит с некоторой задержкой, равной примерно 600 кадрам, что следует учитывать при реализации предложенного алгоритма.

Дальнейшее развитие предложенного алгоритма можно связать с оптимизацией описания однородных областей, в частности с уменьшением объема координат, задающих размеры описываемой области.

Таким образом, полученные результаты исследований показывают, что алгоритм эффективно работает с данными стационарной группы, при этом обеспечивает также работу с нестационарными данными, показывая для них достаточно высокую степень сжатия.

ЛИТЕРАТУРА

1. *Salomon D.* Data compression. – New York: Springer, 2004.
2. *Чье Ен Ун, Левенец А.В., Токарев Р.Е.* Применение популярных алгоритмов компрессии изображений для сжатия измерительных данных // Вестник ТОГУ. – 2012. – № 4(27). – С. 125-132.
3. *Назаров А.В., Козырев Г.И., Шитов И.В. и др.* Современная телеметрия в теории и на практике. – СПб.: Наука и техника, 2007.
4. *Левенец А.В., Нильга В.В.* Структурное упорядочение данных для задач сжатия в информационно-измерительных системах // Вестник ТОГУ. – 2009. – № 2(13). – С. 45-52.
5. *Takaoka T.* Efficient algorithms for the maximum subarray problem by distance matrix multiplication // Electronic Notes in Theoretical Computer Science. – 2002. – №61. – С. 191-200.
6. *Богачев И.В., Левенец А.В.* Алгоритм сжатия телеметрических данных // Межотраслевой институт "Наука и образование". – 2015. – №3(10). – С. 12-15.

E-mail:

Богачев Илья Владимирович – ilya.bogachev@yahoo.com;

Левенец Алексей Викторович – levalvi@bk.ru;

Чье Ен Ун – chye@ais.khstu.ru.