

СТАТИСТИЧЕСКИЙ АНАЛИЗ ТЕЛЕМЕТРИЧЕСКИХ ДАННЫХ С ТОЧКИ ЗРЕНИЯ ЗАДАЧИ СЖАТИЯ

И. В. Богачев^а, аспирант

А. В. Левенец^а, канд. техн. наук, доцент

Чье Ен Ун^а, доктор техн. наук, профессор

^аТихоокеанский государственный университет, Хабаровск, РФ

Введение: задача адаптивного сжатия подразумевает необходимость оперативного анализа (классификации) сжимаемых данных для выбора наиболее эффективного алгоритма сжатия и установки его оптимальных параметров. Классификации данных по функциональному признаку не учитывают особенностей сжатия измерительной информации, поэтому более эффективной следует считать классификацию данных по статистическим свойствам. **Цель исследования:** оценка перспектив алгоритмов сжатия, учитывающих статистические и автокорреляционные свойства телеметрических данных. **Результаты:** исследования гистограмм распределения как исходных данных, так и их разностного представления показывают, что большой динамический диапазон не позволяет удовлетворительно решить задачу сжатия нестационарных данных, однако для стационарных данных можно прогнозировать достаточно высокую эффективность сжатия. Результаты проведенных исследований одномерных и двумерных автокорреляционных функций разных типов телеметрических данных позволяют сделать вывод о том, что наиболее эффективными могут оказаться алгоритмы сжатия, в основе которых лежит учет корреляций как между отсчетами отдельного кадра телеметрии, так и между кадрами потока данных. Проведенные исследования показали, что наиболее перспективными можно считать алгоритмы, в основе которых лежит представление телеметрических кадров в разностно-битовом виде, а также учет особенности объединения данных от отдельных источников в телеметрические кадры. Предложены базовые принципы способа сжатия, основанного на повторном проведении операции дельта-кодирования над элементами анализируемых данных, которые соответствуют первой и второму максимуму одномерной автокорреляционной функции. В качестве дальнейшего развития способов сжатия телеметрических данных предлагается поиск однородных структур внутри телеметрического кадра.

Ключевые слова — телеметрические данные, автокорреляционная функция, классификация данных, сжатие данных, гистограмма распределения.

Введение

Все более глубокая информатизация и развитие промышленности приводят к значительному росту потоков данных в информационно-измерительных системах промышленных предприятий, что в свою очередь влечет за собой существенное увеличение стоимости таких систем, в частности, в силу необходимости либо аренды уже существующих, либо прокладки собственных высокоскоростных линий связи. Снижение таких затрат в первую очередь связывают с процедурами сжатия информационных потоков, а также с глубокой обработкой исходных данных в целях их анализа и выявления свойственных им признаков и закономерностей, что в дальнейшем может послужить основой для разработки новых, более эффективных алгоритмов сжатия.

Во многих отраслях промышленности телеметрические системы отвечают за решение таких важных и разноплановых задач, как сбор и обработка данных, осуществление управления как отдельными элементами, так и техническим объектом в целом и т. д. Тем не менее следует отметить, что до недавнего времени не уделялось особого внимания проблеме модернизации таких систем [1–6]. Более того, принципы обработки и сжатия информации даже в новых системах часто остаются неизменными фактически в течение

нескольких десятилетий. В первую очередь такое положение связано с ограничениями существующего подхода к сжатию данных телеметрии, согласно которому сжатию подвергаются данные от каждого источника по отдельности, без учета корреляционных взаимосвязей между самими источниками.

На настоящий момент можно выделить два подхода к классификации телеметрических данных: функциональный [7] и статистический [8, 9]. Суть функционального подхода заключается в анализе данных по их функциональному признаку. Этот способ предполагает, что телеметрические данные могут быть отнесены к одной из трех групп: телеуправлению, телесигнализации и телеизмерению. У такого подхода есть существенный недостаток, так как он учитывает только природу данных, не затрагивая их свойства, что приводит к серьезному ограничению по применению такой классификации для решения задач сжатия.

Другой подход основывается на анализе данных по их статистическим характеристикам, что более рационально. При этом ввиду широкого распространения разностного принципа как средства уменьшения динамического диапазона данных были подвергнуты исследованию как сами временные ряды, так и их разностное представление.

Такой подход, в отличие от функционального, позволяет произвести переход от анализа и классификации непосредственно к решению задачи сжатия, но его особенностью является то, что исследованию подлежали только временные ряды, полученные от одного источника данных, тогда как наиболее эффективным может оказаться подход, в основу которого положен анализ данных, объединенных в телеметрические кадры. В таком случае появляется возможность перехода от одномерных алгоритмов сжатия, учитывающих лишь особенности каждого отдельного потока данных, к двумерному сжатию, в основу которого может быть положен учет свойств телеметрических кадров, представляющих собой совокупность множества отдельных источников.

Структура анализируемых данных

В данной работе исследованию подверглись телеметрические кадры \mathbf{d} , а также их разностные формы $\Delta\mathbf{d}$, объединяющие m отсчетов данных, полученных от n датчиков, которые можно описать вектор-столбцами следующего вида:

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

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

Предполагая, что телеметрические данные имеют разрядность k , каждый элемент d_i и Δd_i вектор-столбцов можно развернуть в виде битовых последовательностей, записав в виде вектор-строк \mathbf{b} и $\Delta\mathbf{b}$ соответственно размерности k :

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

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

Таким образом, вектор-столбцы \mathbf{d} и $\Delta\mathbf{d}$ можно записать в виде матриц \mathbf{B} и $\Delta\mathbf{B}$ соответственно размерностью $n \times k$:

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

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

Такое рассмотрение данных позволит произвести статистический анализ не только между всей совокупностью кадров, но и между отдельными отсчетами одного кадра.

Для исследования использовалось пять наборов телеметрических данных, обладающих как стационарными, так и нестационарными свойствами, объем которых составлял от десяти до двадцати тысяч кадров. Кадр данных в рассмотренном случае представлял собой набор однобайтных отсчетов, причем для разных наборов данных кадры объединяли в себя разное число датчиков (от 32 до 56). Все телеметрические потоки условно предлагается разделить на

две группы: стационарные и нестационарные. Необходимость такого разделения вызвана делением по уровню сложности алгоритмов сжатия, так как чем менее стационарен поток, тем более сложный алгоритм сжатия позволит получить на нем приемлемую степень компрессии.

Анализ гистограмм распределения

Хорошее представление о свойствах сигнала, в частности о его динамическом диапазоне, стационарности, законе распределения и т. д., дает его гистограмма, что обуславливает ее применение для анализа свойств потока телеметрических кадров. На рис. 1 приведены типовые гистограммы для стационарных (рис. 1, а) и нестационарных (рис. 1, б) потоков телеметрических кадров, а на рис. 2, а и б — их разностные формы соответственно.

Проведенный по имеющимся данным анализ показывает, что как стационарный, так и нестационарный потоки кадров имеют достаточно ши-

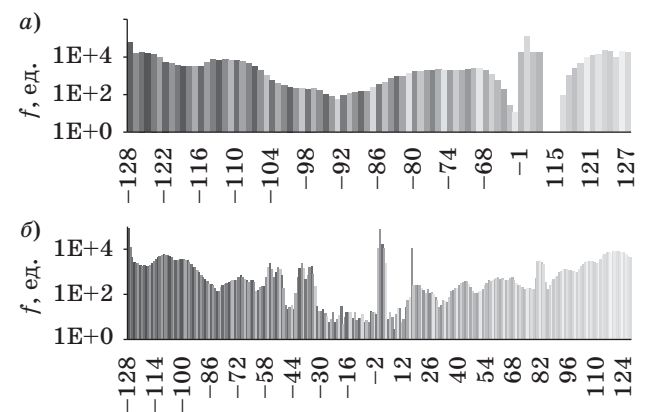


Рис. 1. Гистограммы для стационарного (а) и нестационарного (б) потоков кадров

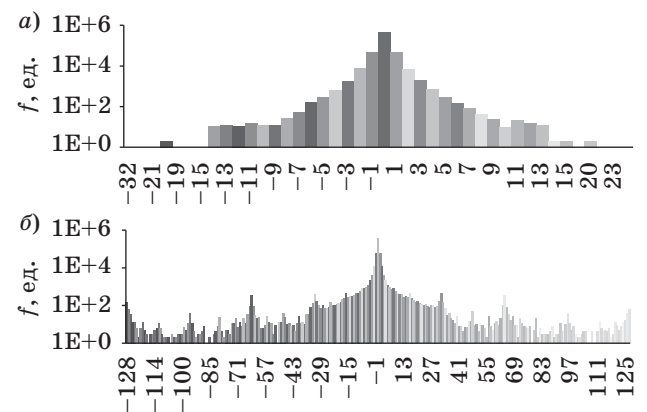


Рис. 2. Гистограммы для разностного эквивалента стационарного (а) и нестационарного (б) потоков кадров

рокий динамический диапазон данных, который все же меньше у стационарного потока. Для обоих случаев не характерна концентрация данных относительно какого-либо диапазона. Указанные свойства сохраняются и у разностного представления нестационарного потока кадров. Также у разностного ряда нестационарных данных может наблюдаться несколько локальных максимумов, хотя в целом соблюдается симметричность относительно некоторого центрального значения. В отличие от разностного представления нестационарного потока кадров, у разностного представления стационарного потока кадров динамический диапазон значительно уже, чем у исходного представления, и наблюдается концентрация данных относительно центрального значения.

Таким образом, анализ полученных результатов показывает, что ввиду большого динамического диапазона, приближенно сопоставимого с равномерным распределением, решение задачи сжатия нестационарных потоков не принесет особой пользы, что нельзя сказать относительно стационарного потока, для которого можно прогнозировать достаточно высокие коэффициенты сжатия.

Анализ автокорреляционных функций

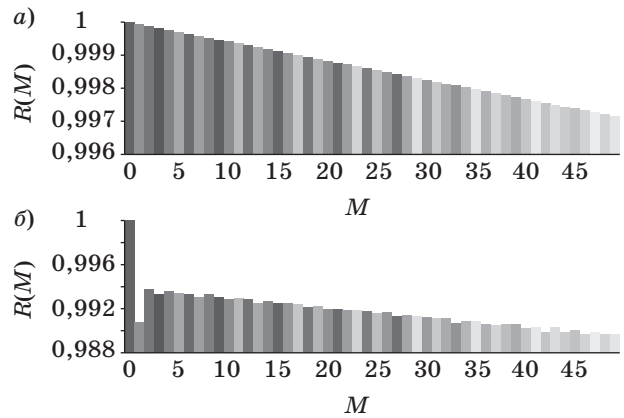
В общем случае, анализируя автокорреляционные функции (АКФ), можно выявить скрытые корреляции как между одними и теми же отсчетами в соседних кадрах (одномерная АКФ), так и между отсчетами внутри кадра (двумерная АКФ), что в свою очередь позволит не только проводить предварительную оценку возможности сжатия кадров, но и поможет выработать специальные методы сжатия, опирающиеся на корреляционные свойства данных.

Под одномерной автокорреляционной функцией (ОАКФ) понимается такая функция, которая определяет зависимость дискретного сигнала $d_j(h + M)$ в момент времени $h + M$ от того же сигнала $d_j(h)$ в момент времени h и рассчитывается согласно формуле [5]

$$R(M) = \sum_{h=1}^m d(h)d(h+M),$$

где M — аргумент ОАКФ, по которому проводится расчет АКФ по времени (в данном случае величина M не превышала значения 50).

Типовые ОАКФ для разностных представлений стационарных и нестационарных сигналов показаны на рис. 3, а и б. При этом исследовались также и исходные (не разностные) сигналы, но так как их ОАКФ имела тривиальный вид и по ней было невозможно определить принадлеж-



■ Рис. 3. ОАКФ для разностного эквивалента стационарного (а) и нестационарного (б) сигналов

ность к какому-либо классу, то их рассмотрение опущено.

Исходя из анализа представленных на рис. 3 результатов, можно сделать выводы о том, что независимо от стационарности и формы представления сигнала показатель автокорреляции носит убывающий характер, но при этом в случае разностного представления сигнала его значение в меньшей степени меняется с увеличением аргумента ОАКФ. Этот факт, наравне с тем, что в случае как стационарного, так и нестационарного сигнала в ходе их преобразования к разностному виду удается значительно увеличить показатель автокорреляции, доказывает эффективность применения разностного подхода как метода сжатия первого уровня, под которым понимается способ предварительной обработки данных, используемый в целях приведения их к виду, оптимальному с точки зрения сжатия тем или иным методом.

Важно отметить, что согласно полученным результатам для случая разностного представления нестационарного сигнала характерно то, что $R(1)$ не является вторым максимумом. Опираясь на это свойство, можно разработать методы сжатия, основанные на повторном проведении операции дельта-кодирования над разностным представлением первого и второго максимума ОАКФ анализируемых данных. Но такой метод пригоден лишь для архивирования данных в целях их длительного хранения и не подходит для сжатия данных, передаваемых по каналу связи, ввиду требования работы телеметрической системы в режиме реального или «мягкого» реального времени.

Относительно потока телеметрических кадров под двумерной автокорреляционной функцией (ДАКФ) понимается такая функция, которая определяет зависимость дискретного сигнала $d_j(h + M)$ не только в момент времени $h + M$ от того

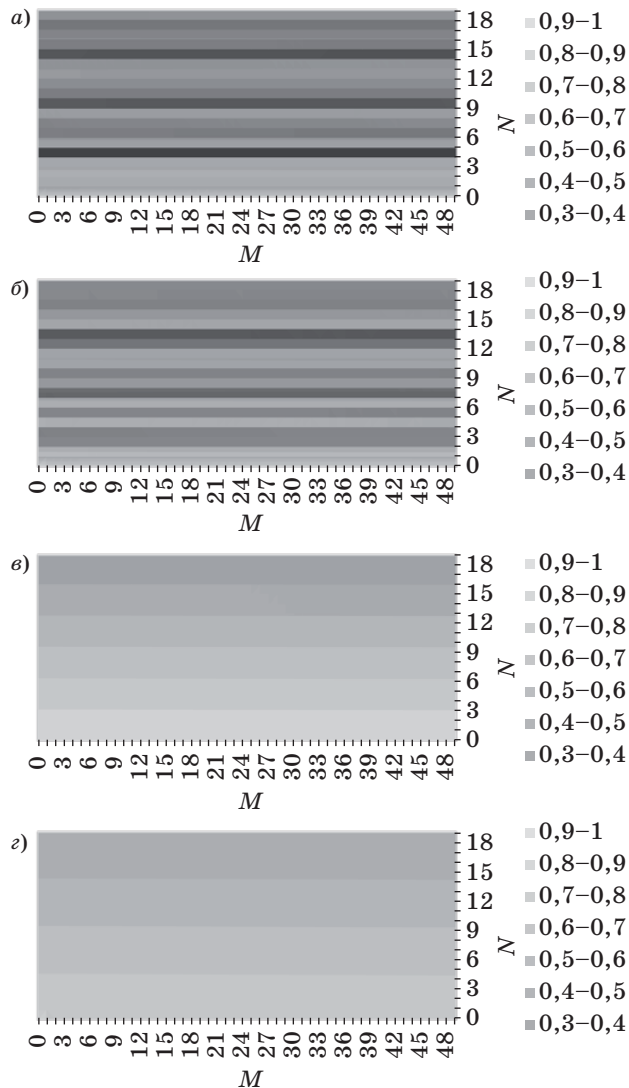
же сигнала $d_j(h)$ в момент времени h , но и между отсчетами, находящимися внутри одного кадра, и рассчитывается согласно формуле [10]

$$R(N, M) = \sum_{h=1}^n \sum_{l=1}^m d(h, l)d(h+N, l+M),$$

где N — аргумент ДАКФ, по которому проводился расчет АКФ по датчикам (N было ограничено числом датчиков); M — аргумент ДАКФ, по которому проводился расчет АКФ по времени (максимальное значение M , как и выше, не превышало 50).

Типовые ДАКФ, построенные относительно потока телеметрических кадров для стационарных и нестационарных сигналов, а также для их разностных представлений, показаны на рис. 4, *a–г*.

Анализируя данные, представленные на рис. 4, можно сделать вывод о том, что независимо от



■ Рис. 4. Типовые ДАКФ для стационарного (*a*) и нестационарного (*б*) потоков кадров и их разностных эквивалентов (*в*, *г*) соответственно

стационарности и формы представления потока телеметрических кадров практически полностью отсутствуют корреляции между отсчетами одного и того же кадра. При этом наблюдаются ярко выраженные корреляции между соседними отсчетами всего потока кадров, что делает невозможным построение алгоритма с использованием дельта-кодирования между отсчетами одного кадра. Исходя из вышесказанного, можно ввести оценку сжимаемости потока кадров на косвенном расчете его стационарности согласно следующей формуле:

$$Q = \left(\frac{\sum_{i=1}^n R_i(2)}{R_i(1)} \right) \frac{1}{n},$$

где Q — оценочная характеристика, косвенно показывающая степень стационарности кадра; n — число датчиков; $R_i(1)$ и $R_i(2)$ — значения соответственно первого и второго аргумента ОАКФ, полученные для i -го датчика.

Так, чем больше характеристика Q , тем в большей степени стационарен кадр, а, следовательно, наибольшую эффективность могут показать методы, изначально адаптированные для сжатия стационарных данных, и напротив, чем меньше характеристика Q , тем более эффективными могут оказаться методы сжатия, основанные на учете особенностей нестационарных кадров.

В результате такой оценки для предлагаемых наборов данных (НД) получены следующие значения: НД1 — 0,999; НД2 — 0,996; НД3 — 0,999; НД4 — 0,998; НД5 — 0,999. Видно, что наибольшее значение характеристика Q принимает для наборов НД1, НД3 и НД5, следовательно, на этих наборах может быть достигнут наибольший коэффициент сжатия с использованием алгоритмов, предназначенных для сжатия стационарных данных.

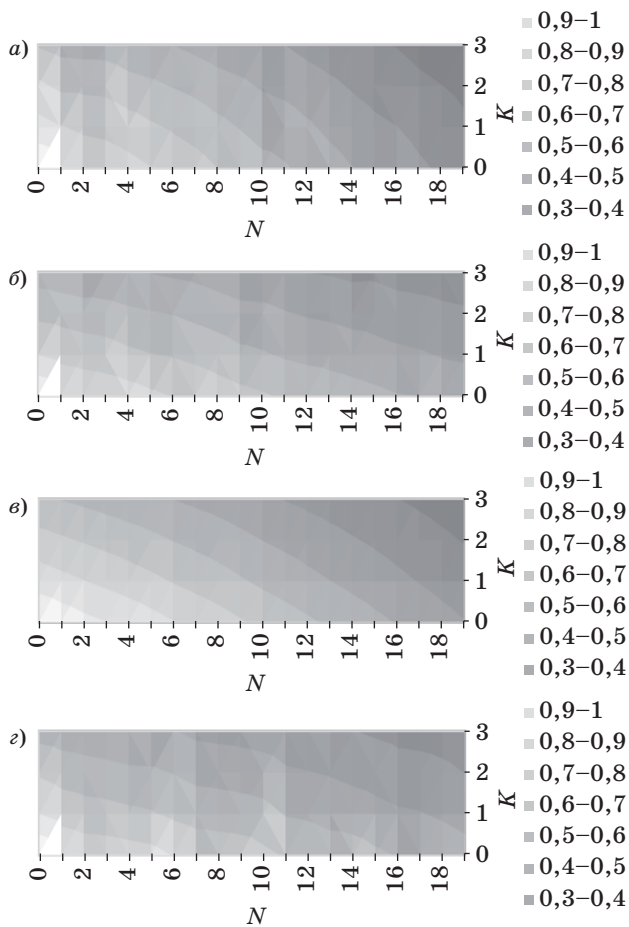
Относительно телеметрических кадров, представленных в битовом виде, под ДАКФ понимается такая функция, которая определяет зависимость между битами отсчетов внутри кадра и рассчитывается согласно следующей формуле [10]:

$$R(N, K) = \sum_{h=1}^n \sum_{l=1}^k B(h, l)B(h+N, l+K),$$

где K — аргумент ДАКФ, по которому проводился расчет АКФ по разрядам (K было ограничено разрядностью данных).

Типовые ДАКФ, построенные относительно телеметрических кадров, представленных в битовом виде для стационарных и нестационарных сигналов, а также для их разностных представлений, показаны на рис. 5, *a–г*.

Результаты, представленные на рис. 5, позволяют сделать вывод о том, что независимо от стационарности и формы представления кадров для



■ **Рис. 5.** Типовые ДАКФ для стационарного (а) и нестационарного (б) кадров в битовом виде и их разностных эквивалентов (в, з) соответственно

них характерна ярко выраженная автокорреляция как между разрядами одного отсчета, так и между разрядами всех отсчетов, объединенных

внутри кадра. При этом кадры, представленные в разностном виде, показывают наилучшую автокорреляцию. Это, в свою очередь, может стать основой для построения алгоритмов сжатия, учитывающих такую особенность. Как и в случае с анализом ДАКФ для потока кадров, наблюдается более яркая корреляционная зависимость для потока стационарных данных и для их разностных рядов.

В таком случае в виде простейшего метода сжатия можно предложить способ, основанный на поиске однородных зон внутри кадра. Под такими зонами понимаются скопления бит, значения которых принимают только одно из двух возможных состояний. В простейшем случае такие зоны можно ограничить прямоугольной областью. Эффективность использования такого подхода показана в работе [11].

Заключение

Проведенное исследование статистических свойств телеметрических данных показывает, что наибольшие перспективы с точки зрения задачи сжатия будут иметь те алгоритмы, в основе которых соблюдаются следующие условия:

- объединение разрозненных данных, полученных от отдельных источников, в телеметрические кадры;
- учет свойств одномерных и двумерных автокорреляционных функций;
- представление телеметрических кадров в разностно-битовом виде.

Выполнение этих трех условий может позволить разработать специализированные алгоритмы сжатия, преимуществом которых будет учет всех возможных корреляций как в потоке телеметрических кадров, так и между отсчетами каждого отдельного кадра.

Литература

1. Чичев С. И., Калинин В. Ф., Глинкин Е. И. Информационно-измерительная система электросетевой компании. — М.: Спектр, 2011. — 156 с.
2. Иванен Н. Т. Автоматизированные системы контроля и учета энергоресурсов на примере газовой отрасли: дис. ... канд. техн. наук. — СПб.: СПбГУ, 2005. — 126 с.
3. Нечистяк М. М. Применение данных телеметрических систем для оптимизации затрат производства // Автоматизация, телемеханизация и связь в нефтяной промышленности. 2011. № 5. С. 32–34.
4. Макарихин К. Б. и др. Система приема и обработки телеметрической информации, система контроля технологических параметров/ К. Б. Макарихин, А. А. Поляков, М. Г. Скрипкин, Н. В. Ушаков //

- Вестник НПО им. С. А. Лавочкина. 2014. № 1. С. 59–63.
5. Влацкая И. В., Голубенко И. В., Голубенко Д. В. Автоматизированная система распознавания и классификации декодированных данных с глубоких блоков телеметрических систем // Перспективы развития информационных технологий. 2011. № 4. С. 109–113.
6. Лоскутов А. И., Патраков С. С., Шестопалова О. Л. Интеллектуальная информационно-диагностическая система оценивания технического состояния бортовой аппаратуры космических аппаратов при подготовке их к запуску // Информационно-управляющие системы. 2014. № 2. С. 18–24.
7. Митюшкин К. Г. Телеконтроль и телеуправление в энергосистемах. — М.: Энергоатомиздат, 1990. — 286 с.

8. Юлин С. С., Паламарь И. Н. Метод классификации сигналов на основе спектрального анализа графа кластерного разбиения // Информационно-управляющие системы. 2015. № 2. С. 23–29. doi:10.15217/issn1684-8853.2015.2.23
9. Левенец А. В. Классификация телемеханических данных и их разностных рядов с точки зрения задачи сжатия // Вестник ТОГУ. 2009. № 4. С. 71–80.
10. Липкин И. А. Статистическая радиотехника. Теория информации и кодирования. — М.: Вузовская книга, 2002. — 216 с.
11. Богачев И. В., Левенец А. В., Чье Ен Ун. Геометрический подход к сжатию данных телеметрических систем // Информатика и системы управления. 2015. № 4. С. 16–22.

UDC 004.627

doi:10.15217/issn1684-8853.2017.1.11

Statistical Analysis of Telemetry Data for CompressionBogachev I. V.^a, Post-Graduate Student, ilya.bogachev@yahoo.comLevenets A. V.^a, PhD, Tech., Associate Professor, levalvi@bk.ruChye En Un^a, Dr. Sc., Tech., Professor, chye@ais.khstu.ru^aPacific National University, 136, Tikhookeanskaia St., 680035, Khabarovsk, Russian Federation

Introduction: The problem of adaptive compression implies the need for operational analysis (classification) of the compressed data in order to select the most efficient compression algorithm and set its optimal parameters. Classifying data by functional features does not take into account the compression of the measurement information, so classification on the base of statistical properties should be considered more appropriate. **Purpose:** We estimate the prospects of compression algorithms which take into account statistical and autocorrelation properties of telemetry data. **Results:** Studying the distribution histograms of the raw data and its differential view has shown that a large dynamic range does not allow you to efficiently solve the problem of non-stationary data compression, but for stationary data you can predict a sufficiently high compression efficiency. The studies of one-dimensional and two-dimensional autocorrelation functions of various telemetry data types suggest that the most efficient compression algorithms can be those which take into account the correlations between references of a single telemetry frame and between different frames of a data flow. The studies have shown that the most promising algorithms represent a telemetry frame in a difference-bit form, taking into account the way in which data from separate sources merge into telemetry frames. We proposed basic principles of a compression method based on repeated delta encoding of the analyzed data elements which correspond to the first and second maximum of a one-dimensional autocorrelation function. For further development of telemetry data compression methods, we suggest to search for homogeneous structures inside a telemetry frame.

Keywords — Telemetry Data, Autocorrelation Function, Data Classification, Data Compression, Distribution Histogram.

Reference

- Chichev S. I., Kalinin V. F., Glinkin E. I. *Informatsionno-izmeritel'naiia sistema elektrossetvoi kompanii* [Information-Measuring System in the Power Grid Companies]. Moscow, Spektr Publ., 2011. 156 p. (In Russian).
- Ivanen N. T. *Avtomatizirovannye sistemy kontroliia i ucheta energoresursov na primere gazovoi otrasli*. Dis. kand. tekhn. nauk [Automated System of Control and Metering of Energy Resources on The Example of the Gas Industry. PhD tech. sci. diss.]. Saint-Petersburg, SPbGU Publ., 2005. 126 p. (In Russian).
- Nechistyak M. M. Application of Telemetric Systems Data to Optimize Productions Costs. *Avtomatizatsiia, telemekhanizatsiia i sviaz' v nefianoi promyshlennosti*, 2011, no. 5, pp. 32–34 (In Russian).
- Makarikhin K. B., Polyakov A. A., Skripkin M. G., Ushakov N. V. The System of Receiving and Processing of Telemetry Data, the System of Technological Parameters Control. *Vestnik NPO im. S. A. Lavochkina*, 2014, no. 1, pp. 59–63 (In Russian).
- Vlatskaia I. V., Golubenko I. V., Golubenko D. V. Automated System of Recognition and Classification of the Decoded Data from Depth Units of Telemetric Systems. *Perspektivy razvitiia informatsionnykh tekhnologii*, 2011, no. 4, pp. 109–113 (In Russian).
- Loskutov A. I., Patrakov S. S., Schestopalova O. L. Intellectual Information-Diagnostic System for Evaluating Technical Condition of Space Vehicles Onboard Equipment during Pre-Launching Activities. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2014, no. 2, pp. 18–24 (In Russian).
- Mitiushkin K. G. *Telekontrol' i teleupravlenie v energosistemakh* [Telemonitoring and Telecontrol in Power Systems]. Moscow, Energoatomizdat Publ., 1990. 286 p. (In Russian).
- Yulin S. S., Palamar I. N. The Method of Time-Series Classification Based on Spectral Analysis of Self-Organizing Map. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 2, pp. 23–29 (In Russian). doi:10.15217/issn1684-8853.2015.2.23
- Levenets A. V. Classification of Telemechanical Data and Their Difference Rows from the Standpoint of Compression. *Vestnik TOGU*, 2009, no. 4, pp. 71–80 (In Russian).
- Lipkin I. A. *Staticheskaiia radiotekhnika. Teoriia informatsii i kodirovaniia* [Statistical Radio Engineering. The Theory of the Information and Coding]. Moscow, Vuzovskaia kniga Publ., 2002. 216 p. (In Russian).
- Bogachev I. V., Levenets A. V., Chye E. U. Geometrical Approach to Data Compression of Telemetry Systems. *Informatika i sistemy upravleniia*, 2015, no. 4, pp. 16–22 (In Russian).