



УДК 004.627

© А. В. Левенец, 2009

КЛАССИФИКАЦИЯ ТЕЛЕМЕХАНИЧЕСКИХ ДАННЫХ И ИХ РАЗНОСТНЫХ РЯДОВ С ТОЧКИ ЗРЕНИЯ ЗАДАЧИ СЖАТИЯ

Левенец А. В. – канд. техн. наук, доц. кафедры «Автоматика и системотехника», тел.: 8-914-191-33-39, e-mail: levalvi@mail.ru (ТОГУ)

Современная классификация данных телемеханики использует функциональный подход, в то время как для задачи сжатия более полезной может служить классификация таких данных по их разностным рядам. Такой подход позволяет уже на этапе классификации грубо оценить коэффициент сжатия и выбрать оптимальный алгоритм сокращения избыточности.

Current classification of telemechanics data utilizes a functional approach whereas for the compression the classification of data according to their difference series may be of greater utility. Such an approach makes it possible even at an initial stage to roughly estimate the compression coefficient and choose an optimal algorithm of data compression.

Ключевые слова: телемеханическая система, телеизмерение, телесигнализация, классификация, сжатие данных, разностный ряд.

Телемеханика в энергетике является быстро развивающейся отраслью техники сбора, передачи, обработки и отображения информации, необходимой для оперативного централизованного контроля процессами производства и распределения электроэнергии.

На современном этапе развития телеметрических систем отмечается рост объемов передаваемой информации, а также расширение типов передаваемой информации. Так, современные системы телемеханики должны передавать не только собственно телемеханические данные (телеизмерения, телесигнализация, телеуправление), но также «срезы» данных за заданный период времени, информацию автоматизированных систем контроля и учёта, файлы оперативно-информационных комплексов и другую информацию [1]. Учитывая, что часть данных, например информация АСКУЭ, передаётся в символьном виде, требования к пропускной способности канала связи значительно повышаются.

Снижение загруженности канала связи можно добиться введением глубокой обработки данных до их поступления в канал связи и, в частности, широким использованием процедур сжатия. Следует отметить, что сжатие данных АСКУиЭ, файлов ОИК и некоторых других данных не представляет собой сложной задачи, т. к. зачастую для таких данных можно использовать уже существующие алгоритмы и программы архивации – в силу отсутствия жёстких временных рамок для передачи такой информации. В то же время сжатие собственно телемеханических данных является задачей более сложной, для которой следует учитывать необходимость работы в режиме реального времени (в лучшем случае, «мягкого» реального времени).

Очевидно, что алгоритмы сжатия телемеханических данных должны быть адаптивными и учитывать их изменяющиеся свойства. Для этого было бы полезно иметь возможность оперативной классификации данных для выбора оптимальных параметров процедур сжатия.

Существующая классификация по функциональному признаку подразделяет телемеханические данные на три основных типа: телеизмерения (ТИ), телесигнализация (ТС) и телерегулирование [1]. Однако с точки зрения сжатия данных такое разделение не является оптимальным, т. к. не учитывает статистические свойства данных и не позволяет делать предположения о параметрах процедур сжатия априори.

В основу большинства алгоритмов сжатия измерительной информации, работающих в режиме реального времени, положен разностный принцип, предполагающий передачу не всего отсчета, а разности между текущим и предыдущим отсчетом. Таким образом, для получения более полного представления о свойствах данных телемеханики, следует провести исследования как временных рядов таких данных, так и их разностных рядов. Такие исследования были проведены на данных телемеханики ряда энергетических объектов Дальнего Востока, причем исследованию подвергались наборы данных ТИ и ТС.

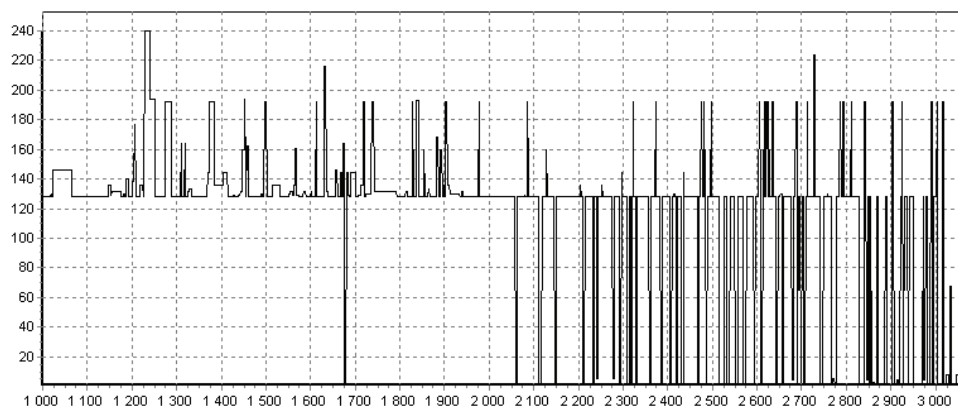
Исследованию подвергались данные от шести энергетических объектов, в том числе подстанций и генерирующих станций. В рамках данной работы были проанализированы 158 наборов данных телеизмерений и 17 наборов данных телесигнализации. Размер выборки для анализируемых данных варьировался в диапазоне от 29752 до 48652 отсчетов. Каждый отсчет представляет собой данные от одного датчика, полученные в кадре циклической передачи ТИ-ТС. Время поступления отсчетов фиксировано и составляет 5 секунд. Следует отметить также, что принимаемые данные являются 8-разрядными.

Сигналы телесигнализации представляют собой сгруппированные в один байт двоичные значения положений переключателей и тому подобное, поэтому наиболее простыми для анализа являются такие данные в стационарном режиме работы объекта в силу крайне редкой смены своих значений. Очевидно, что сигналы ТС для стационарного режима работы объекта легко сжимаются даже без использования специальных алгоритмов. Ситуация не-

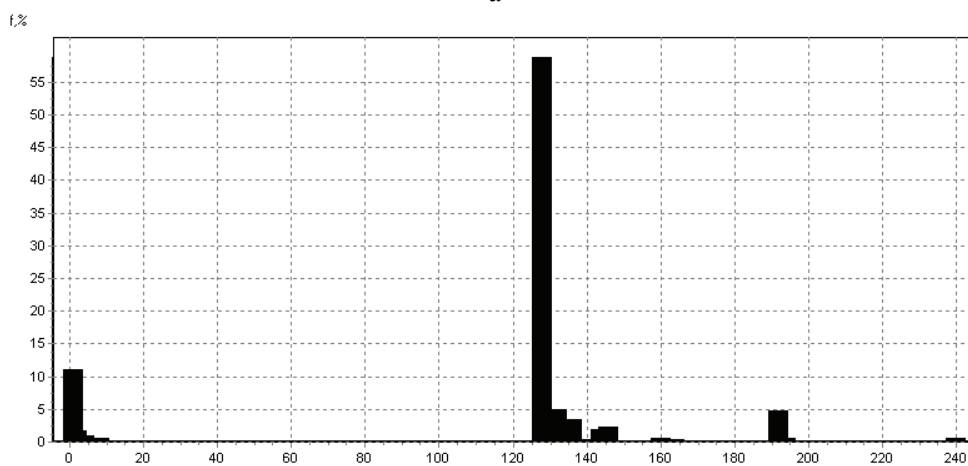


сколько ухудшается в случае перехода объекта в нестационарный режим работы, например, во время аварийной ситуации. Так, на рис. 1, а и б показан пример типичного поведения такого сигнала ТС и соответствующая ему гистограмма.

Проведенный анализ по имеющимся данным ТС показывает, что примерно 99 % всех состояний сигнала ТС в случае стационарного режима работы системы лежит в области его центрального значения. В случае же нестационарного режима работы на центральное (основное) значение сигнала приходится более 50 % всех состояний, причем остальные состояния не обязательно лежат вблизи центрального. Необходимо отметить, что такой тип данных должен хорошо сжиматься словарными методами, однако для данных телемеханики не всегда возможно применение таких методов из-за требований работы в реальном масштабе времени.



а



б

Рис. 1. Типичный поток данных ТС для нестационарного режима работы объекта

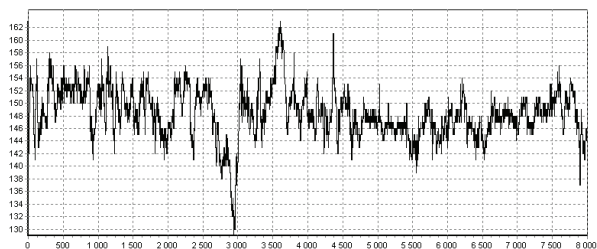
Наиболее сложные с точки зрения процедур сжатия формы имеют сигналы телеизмерений, представляющие собой цифровые данные, поступающие от различного типа датчиков. Проведенный анализ данных показал, что все ТИ можно условно разбить на два основных класса, отличающиеся формой гистограммы и характером поведения во временной области. Так, для данных класса ТИк1 при длительном наблюдении характерно группирование значений сигнала в некоторой относительно узкой полосе значений, причём не наблюдается резких переходов между существенно различающимися значениями. Условно к такому классу можно отнести данные от датчика, регистрирующего некоторый физический процесс, обладающий стационарными свойствами. В это же время для данных класса ТИк2 при достаточно большом времени наблюдения можно выделить несколько участков, на которых сигнал ведёт себя стационарно, причём разница между средними значениями данных на соседних участках различаются более чем на 10 % от полного динамического диапазона. Переход между такими участками происходит достаточно быстро, характерной величиной является 100...200 отсчётов между установившимися значениями. На рис. 2 и 3 приведены графики и гистограммы типичных сигналов ТИ для класса ТИк1 и ТИк2 соответственно.

Для данных класса ТИк1 характерен тот факт, что значения сконцентрированы в относительно небольшом диапазоне (6–8 % от полного динамического диапазона изменения данных), что позволяет прогнозировать хорошие перспективы сжатия для таких сигналов.

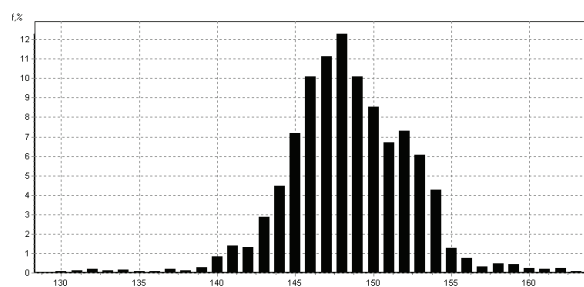
Анализ гистограмм сигналов показывает, что для класса ТИк2 характерно наличие нескольких диапазонов с ярко выраженными на них максимальными частотами повторяемости. Такие диапазоны соответствуют участкам «стационарности», при этом ширина гистограммы каждого участка составляет 3–5 % от полного динамического диапазона изменения сигнала, а полная ширина гистограммы, очевидно, зависит от количества таких участков. Этот факт следует учитывать при разработке устройств сбора данных телемеханики, например, реализуя режимы работы, соответствующие различным состояниям сигнала ТИ.

Исследование разностных рядов телемеханических данных, как было показано выше, также представляет собой практический интерес, т. к. позволяет оценить динамический диапазон сжимаемых данных и выбрать оптимальные параметры алгоритмов сжатия.

Наиболее простой вид как самого разностного ряда, так и его гистограммы имеют сигналы ТС для стационарного режима работы системы, при этом подавляющее большинство значений (более 99 %) разностного ряда приходится на нулевое значение. При нестационарном характере исходного сигнала поведение разностного сигнала меняется незначительно и на боковые лепестки гистограммы приходится не более 0,5 % от общего количества отсчётов сигнала. Таким образом, сжатие таких сигналов не представляет собой сложной задачи и может быть реализовано достаточно простым способом, причём степень сжатия будет достигать значительных величин.

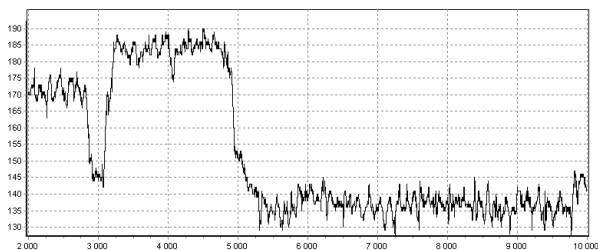


а

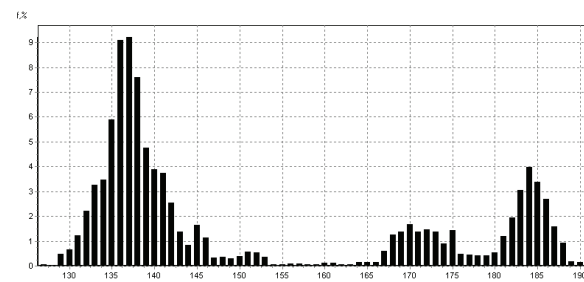


б

Рис. 2. Пример данных класса ТИк1



а



б

Рис. 3. Пример данных класса ТИк2

Необходимо отметить, что разностные ряды, построенные для различных классов данных ТИ, зачастую имеют аналогичные статистические характеристики. Этот факт говорит о необходимости введения собственной классификации для разностных рядов данных телемеханики, при этом классификационным признаком следует выбрать ширину гистограммы. Выбор такого классификационного признака связан с его принципиально важным влиянием на перспективы сжатия анализируемого набора данных, т. к. фактически определяет динамический диапазон изменения этого набора.

В табл. 1 приведён пример распределения данных от разных источников одному объекту энергетики. Размер выборки для данного случая составил 48652 отсчёта. Анализ данных показывает, что в значительном числе случаев (около 80 %) соседние отсчёты данных телемеханики не изменяются. Тем не менее, на изменяемые данные приходится до 20 % всех отсчётов. Так, для приведённого набора данных, по усреднённым данным, примерно 15,5 % приходится на диапазон изменений ± 1 , в то время как диапазон изменений ± 4 занимает 2,5 %. Довольно значительная часть всех значений разностных рядов приходится на диапазон изменений более чем на ± 4 , что следует учитывать при разработке алгоритмов сжатия телемеханических данных. Проведённые исследования показали, что данные по другим объектам имеют аналогичные свойства.

Таким образом, разностные ряды телемеханических данных с точки зрения процедур сжатия целесообразно разделить на четыре класса:

- класс PP1 включает ряды со значительным (более 95 % от всех значений) весом центрального значения гистограммы;
- класс PP2 включает ряды с узкой гистограммой, для которых характерно распределение 95 % основных значений в диапазоне ± 1 ;
- класс PP3 включает ряды со средней гистограммой, с распределением 95 % основных значений в диапазоне ± 4 ;
- класс PP4 включает ряды с широкой гистограммой, с распределением 95 % основных значений в диапазоне свыше ± 4 .

Характерные гистограммы для предлагаемых классов показаны на рис. 4–6, причём гистограмма класса PP1 не приводится в силу своей тривиальности. Следует отметить хорошую симметричность гистограмм, характерную для всех классов разностных рядов телемеханических данных.

В табл. 2 представлено распределение исследованных данных по предлагаемым классам. Как и следовало ожидать, все данные ТС относятся к классу PP1, в то время как данные ТИ распределены между всеми классами.

Характерным свойством данных телемеханики, по крайней мере для рассмотренных объектов, можно считать тот факт, что более 65 % всех наборов ТИ относятся к классу PP2. Можно также отметить незначительное количество наборов данных, относящихся к классу PP4, несмотря на тот факт, что среди наборов данных примерно 15 % относились к аварийным режимам работы объектов.



Таблица 1

Пример распределения данных телемеханики для одного объекта

Диапазон изменения						Тип данных
0	±1	±2	±3	±4	±5 и более	
42668	5757	217	8	2	0	ТИ
43578	4046	459	254	127	188	ТИ
43193	3930	540	355	126	108	ТИ
44403	3757	305	132	30	25	ТИ
44654	3829	124	37	5	3	ТИ
44887	3749	13	1	0	2	ТИ
41331	7074	227	15	3	2	ТИ
47781	867	1	1	0	2	ТИ
44665	3983	0	1	0	1	ТИ
31108	13307	3162	658	215	202	ТИ
36344	10533	1353	274	81	67	ТИ
40768	7534	293	39	12	6	ТИ
40907	7603	126	12	1	3	ТИ
42166	6345	127	9	3	2	ТИ
40677	7085	526	206	106	52	ТИ
31864	11685	2836	965	591	711	ТИ
31063	11321	3268	1217	618	1165	ТИ
30929	13423	3267	601	220	212	ТИ
40506	7926	207	9	0	4	ТИ
32828	11401	2849	568	333	325	ТИ
41346	6313	570	197	109	118	ТИ
46383	2197	0	0	0	72	ТИ
34523	12094	1596	273	70	96	ТИ
36230	11458	779	97	16	72	ТИ
48380	246	0	3	1	22	ТИ
38723	9741	104	6	0	78	ТИ
42129	6413	27	4	1	78	ТИ
28506	16557	1328	441	205	194	ТИ
40465	8115	0	2	0	70	ТИ
48544	58	26	0	11	13	ТС
48579	0	0	0	0	73	ТС
48578	68	2	0	0	4	ТС
48644	0	0	0	0	8	ТС
48650	0	0	0	0	2	ТС
48650	0	0	0	0	2	ТС

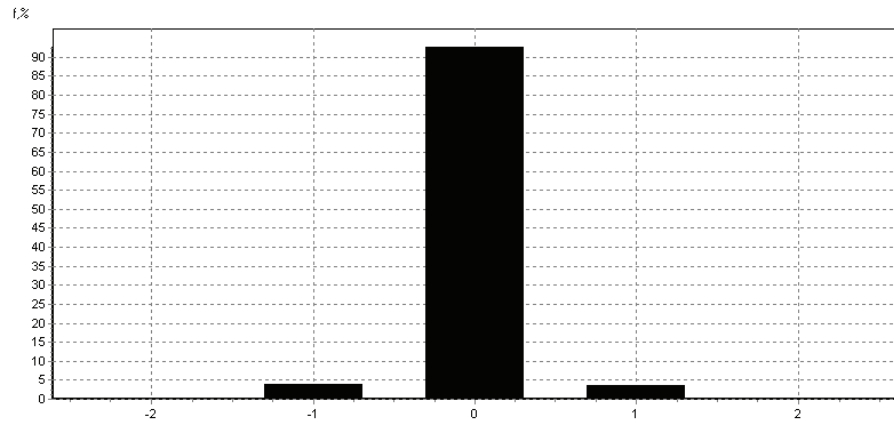


Рис. 4. Характерная гистограмма разностного ряда данных класса PP2

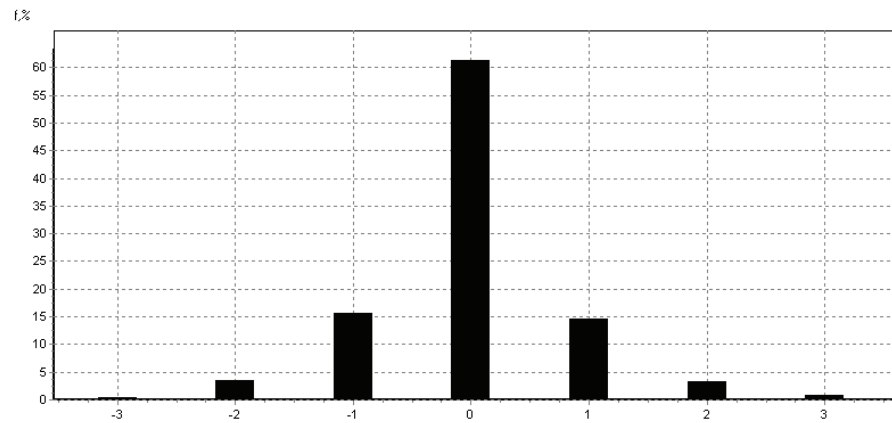


Рис. 5. Характерная гистограмма разностного ряда данных класса PP3

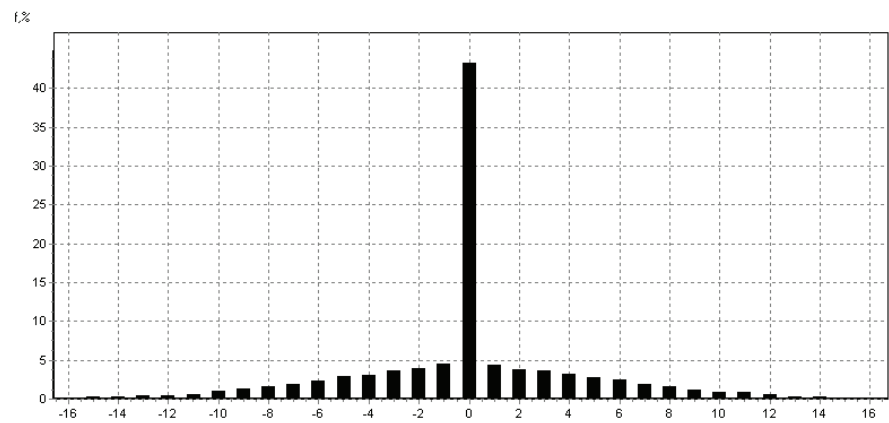


Рис. 6. Характерная гистограмма разностного ряда класса PP4



Таблица 2

Распределение телемеханических данных по классам

Тип сигнала	Класс разностного ряда			
	PP1	PP2	PP3	PP4
ТС	17	0	0	0
ТИ	16	115	23	4
Всего, %	33 (18,9)	115 (65,7)	23 (13,1)	4 (2,3)

Полученные данные можно использовать для прогноза коэффициента сжатия данных, основываясь на некотором простейшем алгоритме сжатия. Тогда среднюю величину «бит на отсчёт» K_{60} можно оценить через относительные частоты f_0 появления тех или иных значений сигнала и числа бит, необходимых для их кодирования n_k :

$$K_{60} = \sum_{i=1}^m f_{oi} \cdot n_{ki}$$

где m – число элементов гистограммы.

Для получения некоторого предельного значения такой величины можно воспользоваться значениями f_{oi} , усреднёнными по всем исследованным наборам данных. Такие значения приведены в табл. 3 вместе с одним из вариантов кодирования разностей, который был реализован в экспериментальной системе совместной передачи данных от энергетических объектов, разработанной для ОДУ Востока [2].

Из приведённых табличных данных величина K_{60} оценивается следующим образом:

$$K_{60} = 0,833 \cdot 1 + 0,137 \cdot 3 + 0,024 \cdot 5 + 0,006 \cdot 12 \approx 1,4.$$

Полученная оценка даёт с учётом 8-разрядности исходных несжатых данных степень сжатия $\approx 5,7$. Аналогичные величины оценок характерны и для отдельных наборов данных телемеханики.

Следует отметить, что для системы, описанной в [2], было получено среднее значение степени сжатия около четырёх, что заметно ниже полученной оценки. Для повышения степени сжатия следует использовать адаптивные алгоритмы, учитывающие текущие статистические параметры сжимаемых данных. Так, можно предложить накопление статистики по разностным рядам сжимаемых данных с последующим выбором способа кодирования. Очевидно, что следует вести как долговременную статистику, определяющую общие параметры алгоритма сжатия, так и короткопериодическую статистику, позволяющую оперативно классифицировать поступающие данные и выбирать локально-оптимальные параметры сжатия. Такой подход потенциально способен поднять общую эффективность алгоритмов сжатия, однако

наилучший результат можно прогнозировать для специализированного алгоритма.

Таблица 3

Усреднённая гистограмма и вариант кодирования разностей

	Диапазон разностей			
	0	± 1	$\pm(2...4)$	$\pm(5...)$
Относительные частоты f_0	0,833	0,137	0,024	0,006
Битовая последовательность	0	10x	11xxx	1111xxxxxxxx
Число бит n_k	1	3	5	12

Также можно предложить предварительную обработку данных перед процедурой сжатия, априори учитывающую статистические характеристики потока сжимаемых данных и адаптивно отслеживающую их в процессе работы. Один из вариантов такой процедуры предложен в работе [3], где была предпринята попытка определения границ применимости предложенной предварительной обработки к исходным данным, имеющим случайный характер.

Таким образом, проведённые исследования показывают, что данные телемеханики энергетических объектов с точки зрения процедур сокращения избыточности информации следует классифицировать по статистическим свойствам их разностных рядов. Предложенная классификация позволяет априори оценивать степень сжатия обрабатываемых данных и выбирать оптимальные параметры алгоритма сжатия и формат кадра телемеханики. Кроме того, классификацию данных можно осуществлять оперативно, производя накопление статистической информации в процессе передачи данных, что наиболее актуально для систем с нефиксированным размером и составом кадра данных.

Библиографические ссылки

1. Митюшкин К. Г. Телеконтроль и телеуправление в энергосистемах. М., 1990.
2. Алгоритмы сжатия данных АСКУЭ и телемеханики для системы совместной передачи данных / С. В. Головизин, А. В. Левенец, С. Р. Симаков // Энергетика: управление, качество и эффективность использования энергоресурсов: сб. трудов третьей Всероссийской НТК с международным участием. Т. 1. Благовещенск, 2003.
3. Левенец А. В. Структурное упорядочение данных для задач сжатия в информационно-измерительных системах // Вестник ТОГУ. 2009. № 2(13).