

На правах рукописи

Федяев Александр Юрьевич

**АЛГОРИТМЫ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ
ДЛЯ ЗАДАЧ СЖАТИЯ ДАННЫХ
В ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫХ СИСТЕМАХ**

05.11.16 – Информационно-измерительные и управляющие
системы (техника и технологии)

**АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук**

28 Ноя 2013



005541261

Хабаровск - 2013

Работа выполнена в ФГБО ВПО «Тихоокеанский государственный университет».

Научный руководитель: доктор технических наук, профессор Чье Ен Ун

Официальные оппоненты:

Корчевский Вячеслав Владимирович –
доктор физико-математических наук, доцент,
ФГБОУ ВПО «Тихоокеанский государственный
университет», кафедра «Технологическая
информатика и информационные системы»,
профессор

Бахрушин Александр Петрович –
кандидат технических наук, доцент, ФГБОУ ВПО
«Приамурский государственный университет
имени Шолом-Алейхема», кафедра «Информатика
и вычислительная техника», доцент

Ведущая организация:

Дальневосточный филиал федерального
государственного унитарного предприятия
«Всероссийский научно-исследовательский
институт физико-технических и радиотехнических
измерений», г. Хабаровск

Защита диссертации состоится «25» декабря 2013 г. в 16:00 на заседании
диссертационного совета Д 212.294.05 в ФГБОУ ВПО «Тихоокеанский
государственный университет» по адресу: 680035, г. Хабаровск,
ул. Тихоокеанская, 136, ауд. 315л.

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВПО «Тихо-
океанский государственный университет».

Автореферат разослан «15» ноября 2013 г.

Ученый секретарь
диссертационного совета

Бурдинский Игорь Николаевич

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Следствием быстрых темпов развития систем телеметрии для решения задач контроля сложных объектов и вычислительных комплексов, управления и их испытания стало увеличение объемов передаваемой по каналам связи информации, что в результате привело к достижению возможных границ по расширению полосы частот каналов связи, а также объемов памяти устройств регистрации. Вместе с этим передаваемые данные зачастую обладают значительной избыточностью, что послужило причиной создания большого количества способов и алгоритмов предварительной обработки и сжатия информации. Значительный вклад в разработку подобных алгоритмов внесли такие учёные, как Б. Я. Авдеев, Е. М. Антонюк, Д. С. Ватолин, В. А. Виттих, Т. Велч, К. Гольдбах, А. Н. Дядюнов, Дж. Зив, Ф. Катц, В. И. Левенштейн, А. Лемпель, Л. Рабинер, Д. Саломон, Р. Фано, Д. Хаффман, С. Эвен, П. Элиас и др.

На настоящий момент рост производительности вычислительных средств позволяет не просто выделять существенную информацию, но и проводить её анализ, выявляя скрытые закономерности и связи, с целью более компактного описания. При этом наиболее актуальной задачей, связанной с проблемой сжатия данных, является приведение данных к некоторой оптимальной модели, заложенной в алгоритме сжатия, что обеспечивает их наилучшее сжатие. Данной тематике посвящены работы М. Барроуза, С. Голомба, Б. Я. Рябко, Д. Уиллера и др. Однако, несмотря на существование целого ряда способов предварительного преобразования, ориентированных на улучшение работы существующих методов сжатия, лишь немногие из них ориентированы на работу с измерительными данными. Этот факт существенно ограничивает применение существующих способов предварительной обработки в информационно-измерительных системах. Таким образом, актуальной является задача поиска способов предварительной обработки, способных эффективно работать с измерительными данными.

Еще одной причиной, определяющей необходимость использования средств сжатия и предварительной обработки данных является все больший рост количества автоматизированных систем. Использование производительных микропроцессоров позволяет реализовывать системы сбора и обработки данных, которые являются автономными и функционируют без участия оператора. Как следствие, в результате длительного времени наблюдения накапливается большой объем данных, причем для снижения требований к объему носителей информации, очевидно, следует широко использовать сжатие и, как следствие, предварительную обработку накопленных данных.

Целью данной работы является исследование и разработка способов предварительного преобразования измерительных данных для увеличения эффективности их сжатия в информационно-измерительных системах (ИИС).

Основные задачи исследования:

1. Обзор и оценка возможностей существующих методов сжатия сигналов, используемых в ИИС. Выделение методов сжатия, эффективность которых может быть существенно повышена за счет предварительного преобразования входного потока данных;

2. Разработка алгоритмов предварительного преобразования, повышающих эффективность алгоритмов сжатия используемых в ИИС;

3. Разработка программных средств на базе предложенных алгоритмов.

Основные методы исследования базируются на математическом аппарате теории случайных процессов, теории кодирования, теории цифровой обработки сигналов, теории алгоритмов, теории групп и математической статистике. Для экспериментального исследования применены методы математического моделирования, средства объектно-ориентированного программирования, язык объектно-ориентированного программирования Java.

Научная новизна работы:

1. Предложен способ представления данных в виде отрезков фиксированной длины для компактного описания измерительных сигналов;

2. Разработан и исследован способ преобразования данных с помощью кубических структур, позволяющий повысить степень сжатия небольших объемов измерительных данных;

3. Обоснован способ разбиения блоков измерительных данных на сегменты для их более эффективного сжатия существующими алгоритмами;

4. Разработан способ преобразования блоков измерительных данных на плоскости с помощью кодирующих функций с целью их более компактного представления;

5. Предложена классификация данных телемеханики, основанная на статистическом анализе разностных рядов, даны рекомендации по использованию алгоритмов предварительного преобразования для данных различных классов.

Практическая значимость:

1. На базе предложенных способов разработаны алгоритмы предварительной обработки измерительных данных, ориентированные на работу с рядом широко распространенных алгоритмов сжатия;

2. Разработаны программные комплексы для исследования предложенных способов предварительной обработки;

3. Предложена методика выбора оптимального алгоритма преобразования, учитывающая статистические свойства обрабатываемых данных и базирующаяся на предложенной классификации данных телемеханики.

Основные положения, выносимые на защиту:

1. Способ представления измерительных данных в виде отрезков фиксированной длины и реализованные на основе способа алгоритмы, позволяющие преобразовать данные в L-систему, обладающую потенциальной возможностью более эффективного сжатия.

2. Алгоритм кубического преобразования данных, позволяющий повысить эффективность применения словарных методов сжатия, при условии использования критерия длины повторов для данных с малой избыточностью и неоднородных данных, а также статистических методов при условии использования блочного критерия для аналогичных типов данных.

3. Классификация данных разностными рядами первого и второго порядков, позволяющая выделить семь классов измерительных типов данных, отличающихся динамическим диапазоном, что дает возможность более точно подбирать подходящие к данному классу алгоритмы обработки.

4. Алгоритмы сегментирования измерительных данных, позволяющие повысить эффективность сжатия на основе LZ-методов.

Реализация результатов работы. Исследования и разработки, выполненные в диссертационной работе, проводились в рамках федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг. по теме «Программно-аппаратная платформа для промышленно-ориентированных сенсорных сетей ISM-диапазонов» (номер гос. рег. 01201276122), в рамках аналитической ведомственной целевой программы «Развитие научного потенциала высшей школы» проект «Совершенствование аппаратно-программного обеспечения адаптивных систем передачи данных распределенных информационно-измерительных систем», 2012–2014 гг. (номер гос. рег. 01201256236). Проведенные теоретические исследования и разработки использованы в проектно-конструкторской и исследовательской деятельности ООО ГК «РосАвтоматизация ДВ», ООО НВФ «Сенсоры, модули, системы ДВ».

Апробация работы. Основные результаты диссертационной работы докладывались и обсуждались:

- на VI Всероссийской НПК студентов, аспирантов и молодых ученых "Молодежь и современные информационные технологии", г. Томск, 2008 г.;
- Региональной НПК "Молодежь и научно-технический прогресс", г. Владивосток, 2010 г.;
- Всероссийской НТК студентов, аспирантов и молодых ученых: "Научная сессия ТУСУР-2010", г. Томск, 2010 г.;
- XII Международной НПК "Фундаментальные и прикладные исследования, разработка и применение высоких технологий в промышленности", г. Санкт-Петербург, 2011 г.;
- Международной НПК "Научные исследования и их практическое применение. Современное состояние и пути развития '2011", г. Одесса, 2012 г.;
- VII Международной НПК "Перспективы развития информационных технологий", г. Новосибирск, 2012 г.;
- Международной НПК "Научные исследования и их практическое применение. Современное состояние и пути развития, 2012", г. Одесса, 2012 г.;
- XIV Международной НПК "Фундаментальные и прикладные исследования, разработка и применение высоких технологий в промышленности и экономике", г. Санкт-Петербург, 2012 г.;
- XV Международной НПК "Фундаментальные и прикладные исследования, разработка и применение высоких технологий в промышленности и экономике", г. Санкт-Петербург, 2013 г.;
- II international research and practice conference "Science, Technology and Higher Education", Canada, 2013 г.;
- Международной НПК "Информационные технологии XXI века", г. Хабаровск, 2013 г.;

Публикации. По результатам выполненных исследований опубликовано 14 работ, в том числе две работы в журналах из перечня ВАК, получены два свидетельства о государственной регистрации программ для ЭВМ. Результаты исследования представлены на 8 международных конференциях.

Структура и объем работы. Работа состоит из введения, четырех глав, заключения, списка литературы, содержащего 112 наименований, семи приложений и содержит 121 страницу основного текста, 11 таблиц и 65 рисунков.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность выбранной темы исследования, сформулированы цель работы, научная новизна, основные положения, выносимые на защиту, приведены основные методы исследования, дана информация о реализации результатов, апробации, структуре работы.

В первой главе даются необходимые сведения об ИИС. Выделены системы телемеханики, приводятся основные особенности таких систем, а также требования, предъявляемые к данным системам.

Особенности систем, заключающиеся в необходимости обеспечения высокой точности передачи, высокой надежности передачи команд управления, централизованности переработки, высокой степени автоматизации процессов сбора и использования информации, отражаются на свойствах циркулирующих в таких системах данных. По сравнению с алгоритмами, разработанными для текстовых и медиаданных, алгоритмов обработки и сжатия измерительных данных не так много, что обуславливает их актуальность.

При рассмотрении классификации измерительных данных особое внимание уделяется разделению данных в телемеханических системах, где можно выделить три основных типа данных: данные телерегулирования, данные телеизмерений, данные телесигнализации. Однако оказывается, что с точки зрения задач предварительной обработки и последующего сжатия такая классификация данных не является оптимальной, поскольку не учитывает статистические свойства данных и не позволяет сделать предположение о выборе оптимальных параметров для процедур сжатия. Поэтому, в качестве альтернативы, приводится классификация на основе разностных рядов телемеханических данных, позволяя заранее оценить степень сжатия обрабатываемых данных и выбрать оптимальные параметры последующих алгоритмов преобразования.

Представлено описание процедуры сжатия, приводится общая классификация методов сжатия, особое внимание уделено классификации обратимых и необратимых методов сжатия измерительных данных. Производится описание статистических, на основе алгоритма Хаффмана, и словарных, на основе метода Лемпела-Зива, алгоритмов сжатия, поскольку данные методы используются в работе при оценке эффективности осуществления предварительного преобразования.

Приведены формулы для определения коэффициента и степени сжатия, а также рассмотрены подходы Колмогорова-Хайтина и Шеннона для анализа количества информации.

Количество сжимаемых конкретным алгоритмом файлов и степень сжатия отдельных файлов может быть улучшена за счет применения предварительного преобразования данных. Рассматриваются основные способы предварительного преобразования данных, такие, как:

- модификация унифицированных блоков конкретных видов данных для улучшения структуры результирующей;
- использование словаря *n*-графов, преобразование заглавных букв, модификация разделителей, специальное кодирование символов конца строки; для нетекстовых данных - преобразование относительных адресов, преобразование табличных структур и др.;
- уменьшение коррелированности соседних выборок путем их взаимного дифференцирования;
- переупорядочивание блоков данных без их модификации в источнике с целью создания более коррелируемой структуры.

В главе также анализируются основные характеристики рассмотренных способов идается их краткая оценка.

Во второй главе в качестве способа первичной обработки измерительных данных предлагается способ представления данных отрезками фиксированной длины. Приводятся теоретические основы реализации способа, описания алгоритмов реализации, оценивается эффективность представления.

В основу предлагаемого способа представления измерительных данных положен принцип L-систем. Теоретически эти системы позволяют с помощью итерационного подхода создавать достаточно сложные объекты с помощью небольших объёмов данных. При описании дискретного сигнала достаточно использования трех команд L-системы: "+" – увеличить угол θ на величину α , "-" – уменьшить угол θ на величину α , "F" – переместиться вперёд на один шаг, прошивая след. Регулируя параметры шага приращения l и величины приращения по углу α , можно добиваться различной точности описания сигнала с различной выходной длиной кодового слова.

К основным способам описания значений можно отнести следующие:

1) L-система описывает значения, используя края отрезка. При этом способе края отрезков должны стремиться к максимальному приближению к дискретным значениям измерительного сигнала. Таким образом, значения сигнала могут быть получены после одинаковых групп символов "F", как это показано на рис. 1, *a*.

2) L-система описывает значения, используя середину отрезка. В этом случае отрезок состоит из чётного количества символов "F", при этом конец первой половины группы символов (т. е. середина отрезка) должен по возможности максимально приблизиться к действительному значению сигнала (рис. 1, *b*).

Приводятся способы описания сигнала, среди которых наибольшее внимание уделяется способам, использующим для представления одного отсчета сигнала четыре параметра: два из которых настраивают ориентацию РТ, а остальные

два используются для первого и второго приближения к значению сигнала соответственно. Данный механизм позволяет за счет подбора различных сочетаний параметров максимально точно приблизиться к преобразуемому отсчету сигнала.

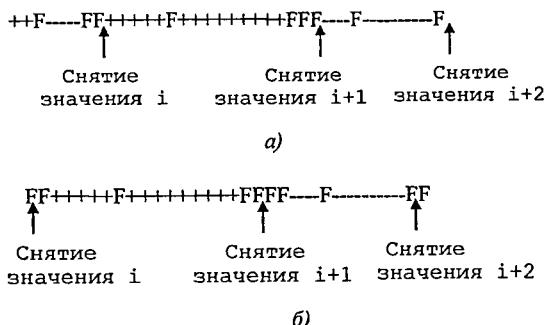


Рис. 1. Снятие значений кодового слова способом краевых отрезков (a) способом серединных отрезков (б)

Принципы двух, используемых в работе, способов описания сигналов приведены на рис. 2, 3.

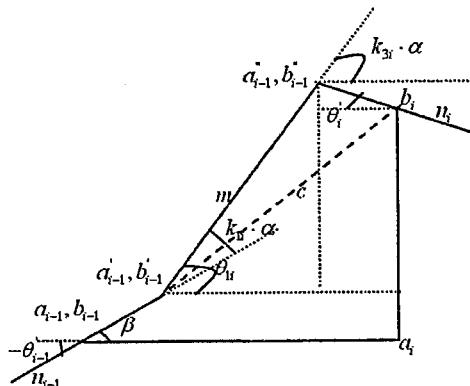


Рис. 2. Принцип преобразования способом серединных отрезков с использованием четырех параметров

Используя преобразование способом серединных отрезков с использованием четырех параметров, можно вычислить четыре необходимых параметра на основе следующих формул:

$$k_{1i} = \frac{\theta_{il} + \theta_{i-1l}}{\alpha};$$

$$k_{2i} = \frac{m}{l}; \quad (1)$$

$$k_{3i} = \frac{-\theta_{ii} - \arcsin \frac{2(m \cdot \sin \theta_{ii} + b'_{i-1} - b_i)}{n}}{\alpha};$$

$$k_{4i} = \frac{n}{l}; \quad (2)$$

где $\theta_{ii} = \arccos \left(\frac{4(m^2 + c^2) - n^2}{8mc} \right) + \arcsin \left(\frac{b_i - b'_{i-1}}{c} \right); \quad b'_{i-1} = \frac{n_{i-1}}{2} \sin \theta_{i-1} - b_{i-1};$

$c = \sqrt{(a_i - a'_{i-1})^2 + (b_i - b'_{i-1})^2}$; k_{ii}, k_{3i} – параметры, означающие, какое количество значений “+”, если $k > 0$, либо “-”, если $k < 0$, необходимо записать в поток для получения угла θ_{ii} ; k_{2i}, k_{4i} – параметры, отвечающие за то, какое количество шагов “F” необходимо сделать в первом и во втором приближении.

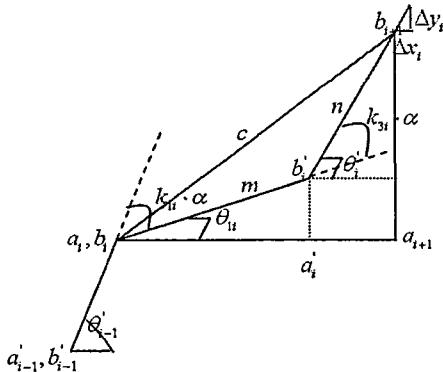


Рис. 3. Принцип преобразования способом равных краевых отрезков с использованием четырёх параметров

Искомые параметры в случае преобразования способом краевых отрезков с использованием четырех параметров можно найти используя (1), (2), а также следующие формулы:

$$k_{ii} = \frac{\theta_{ii} - \theta'_{i-1}}{\alpha};$$

$$k_{3i} = \frac{\arcsin \frac{b_{i+1} - b_i}{n} - \theta_{ii}}{\alpha};$$

где $c = \sqrt{(a_{i+1} - a_i)^2 + (b_{i+1} - b_i)^2}$; $\theta_{ii} = \arcsin \frac{b_{i+1} - b_i}{c} - \arccos \frac{m^2 + c^2 - n^2}{2mc}$.

Структуры кодовых последовательностей для обоих способов одинаковы и имеют вид представленный на рис. 4. На рисунке приняты следующие обозначения: S_i – отсчёты сигнала; $i=0\dots N$; N – количество дискретных отсчётов; $S_i = k_{ii}k_{2i}k_{3i}k_{4i}$; k_{ii} – настраивает ориентацию РТ для первого приближения к i -му

значению дискретной функции; k_{2i} – реализует первое приближение РТ к i -му значению дискретной функции; k_{3i} – настраивает ориентацию РТ для второго приближения к i -му значению дискретной функции, k_{4i} – реализует второе приближение РТ к i -му значению дискретной функции.

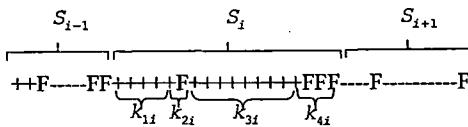


Рис. 4. Структура кодового слова

На базе представленного выше способа предложен ряд алгоритмов, реализующих разные варианты осуществления преобразования.

Первый алгоритм реализует способ серединных отрезков с четырьмя параметрами. Смысл данного алгоритма заключается в выборе таких значений m и n и подборе таких углов θ_{ii} и θ'_i , чтобы середина отрезка n находилась как можно ближе к значениям (a_i, b_i) .

Второй алгоритм реализует способ равных краевых отрезков с четырьмя параметрами. Смысл данного способа преобразования заключается в подборе таких значений m и n и углов θ_{ii} и θ'_i , чтобы максимально приблизиться к значениям a_i и b_i и минимизировать тем самым погрешности Δy_i и Δx_i .

Третий алгоритм реализует способ адаптивных краевых отрезков с четырьмя параметрами. Данный способ является модификацией способа равных краевых отрезков с четырьмя параметрами и нацелен на уменьшение погрешности преобразования. В данном способе применяется последовательное изменение значений m и n для обеспечения заданной погрешности преобразования.

Анализ избыточности преобразования предложенных алгоритмов сжатия приводится для трех базовых типов оцифрованных сигналов: синусоидального, прямоугольной формы и пилообразной формы. Производится выбор оптимальных параметров алгоритмов на основе анализа объема кодовой последовательности после преобразования, а также оценки абсолютной погрешности по оси времени и по оси значений. В результате проведенных исследований были получены оптимальные значения основных параметров:

- для алгоритма представления сигналов способом серединных отрезков: $\delta = 0,001$; $l = 0,1$; $\alpha = 1$;
- для алгоритма представления сигналов способом равных краевых отрезков: $\delta = 0,001$; $l = 0,1$; $\alpha = 2$;
- для алгоритма представления сигналов способом адаптивных краевых отрезков: $\delta = 0,0001$; $l = 0,1$; $\alpha = 5$.

Для различных типов сигналов производится исследование зависимостей коэффициента сжатия алгоритма от степени зашумления, изменения параметров сигналов и объема словаря. В качестве алгоритма сжатия используется словарный метод сжатия LZW. Выбор данного алгоритма преобразования обусловлен

тем, что словарные методы демонстрируют наибольшую степень сжатия для кодового слова среди прочих рассматриваемых подходов осуществления сжатия. Исследования показывают, что существующие алгоритмы демонстрируют хорошие потенциальные возможности сжатия кодового слова, что было подтверждено использованием LZW-метода.

В третьей главе для осуществления предварительной обработки данных предлагаются способы кубического преобразования, кодирования на плоскости, сегментирования данных и реализованные на их базе соответствующие алгоритмы. Для каждого способа вводятся необходимые теоретические сведения, приводятся описания реализованных алгоритмов, выбираются оптимальные параметры алгоритмов для осуществления процедуры преобразования измерительных данных.

Способ кубического преобразования предлагает подход, заключающийся в представлении входных данных некоторым геометрическим объектом, сжатие которого будет производиться более эффективно. В общем виде предварительное преобразование может быть описано с помощью следующей последовательности шагов:

Шаг 1. Данные переводятся в удобную для преобразования систему счисления;

Шаг 2. Выбирается структура (геометрическое тело с возможностью отображения на него массива данных и изменения данного массива путем трансформации тела) с определенной на ней группой операций модификации, которые позволяют изменять состояние структуры;

Шаг 3. Данные отображаются на эту структуру;

Шаг 4. Происходит их преобразование с использованием последовательности операций;

Шаг 5. Подбирается такая последовательность операций модификации (манипуляция h), при которой свойства преобразованных данных максимально приближены к некоторому оптимальному для сжатия варианту.

В качестве формирующей структуры для рассматриваемого способа предлагается куб в связи с простотой использования.

Для осуществления преобразования используется группа «Кубик Рубика», которая содержит операции, позволяющие осуществить как прямое, так и обратное преобразование. Вводятся понятия операции модификации как одной базовой операции, понятия манипуляции как совокупности операций модификации. Примеры обозначений и реализаций различных операций модификации приведены на рис. 5.

Приводится функция предварительного преобразования для последовательности l , которая может быть определена следующим образом:

$$f(l, h) = \sum_{i=1}^k f_i(l_i, h_i) = l,$$

где $l = \sum_i^k l_i$ – преобразуемая последовательность; $h = \sum_i^k h_i$ – манипуляция, используемая для преобразования; l_i – i -й блок данных; h_i – манипуляция для i -го

блока данных; $k = n/6$ – количество блоков данных; $f_i(l_i, h_i) = l_{ti}$ – функция преобразования одного блока данных; $P_{i=1}^k$ – правило получения l_i из совокупности l_{ti} .

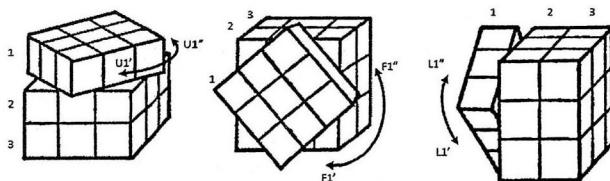


Рис. 5. Примеры обозначения операций модификации

На основе способа предлагаются алгоритмы предварительной обработки измерительных данных с использованием кубических структур. Приводятся описания исходного, а также оптимизированного алгоритмов, осуществляется их анализ. Оценка алгоритма включает список основных действий, используемых алгоритмами, и расчет количества выполняемых действий. Так, оптимизированный алгоритм включает действия по инициализации блока данных, копированию, трансформации, расчету критерия, вставке пары значений в буфер выборки с учетом порядка, смене буфера выборки, взятию строки из буфера выборки, считыванию блока данных, очистке буфера выборки.

Особое внимание уделено выбору критерия оценки оптимальности преобразования, поскольку от него существенно зависит эффективность использования алгоритма сжатия. Так, для методов RLE необходимо учитывать порядок следования символов, а для статистических методов важным является количество одинаковых символов во всей исходной последовательности данных.

Для оценки степени приближения свойств преобразованных данных к теоретически наиболее эффективным с точки зрения процедур сжатия предлагается использование коэффициента подобия K_p :

$$K_p = \frac{c_t}{c_p},$$

где c_t – максимально возможная (теоретическая) оценка критерия, означающая наилучшее сжатие последовательности алгоритмом сжатия; c_p – оценка критерия текущей последовательности (т. е. практическая оценка).

Для каждого критерия приводятся способы расчета теоретических и практических оценок.

Для различных типов данных производится выбор параметров алгоритма кубического преобразования, к которым относится способ проецирования блока на куб, определения числа внешних и внутренних итераций, размер буфера выборки, а также список используемых операций модификаций.

Предлагается способ линейного кодирования на плоскости, общая схема которого представлена на рис. 6. Задачей является интерпретация каждого интервала как одного из двух возможных состояний «0» или «1». Кодирование произ-

водится при помощи некоторой функции, которую предлагается называть кодирующей (КФ). При этом рассматриваются два способа детектирования нулевого или единичного разряда: способ детектирования по вхождению и способ детектирования по пересечению. В связи со значительной сложностью расчетов для способа детектирования по вхождению для проведения исследований используется способ детектирования по пересечению.

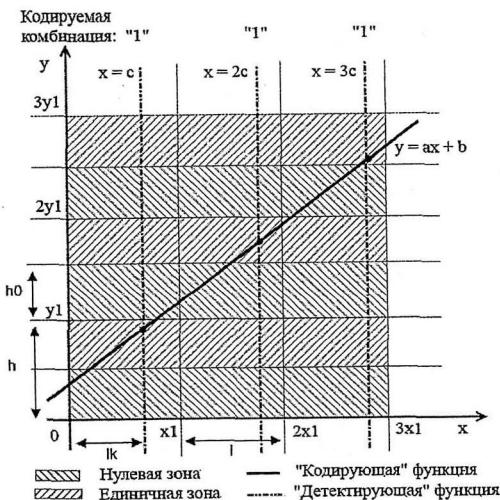


Рис. 6. Пример кодирования на плоскости

Исследование параметров для способа кодирования включает результаты выбора кодирующей функции, среди которых рассматриваются линейная функция $ax + b$, степенные функции $ax^n + bx$, где $n=2, \dots, 7$, функция синуса $b \cdot \sin(ax)$, функция логарифма $a \cdot \lg(bx)$.

Для выбора КФ вводится понятие коэффициента покрытия возможных бинарных комбинаций заданной длины (K_p), который прямо пропорционален числу кодируемым функцией разрядов:

$$|K_p| = n_{\text{pcf}} / m_{\text{tcf}},$$

где n_{pcf} – количество последовательностей, для которых были найдены параметры кодирующей функции; $m_{\text{tcf}} = 2^l$ – количество бинарных комбинаций длины l .

Предлагаемый способ сегментирования включает описание, которое заключается в том, что блок преобразуемых данных разбивается на сегменты, состоящие из символов алфавита фиксированной длины. Используемые символы алфавита выносятся в начало сегмента, образуя словарь. Данным символам присваиваются определенные условные значения, используемые для замены соответствующих символов в сегменте.

Приводятся теоретические оценки минимального и максимального коэффициентов сжатия.

Большое внимание уделено задаче выделения сегментов в преобразуемом блоке для минимизации размера таблицы символов, используемых в сегменте, уменьшению объема преобразованного сегмента и сокращению общего объема преобразуемого блока. Данную задачу предлагается решать с помощью трех алгоритмов: сканирования, усреднения и усреднения с приближением. Алгоритм сканирования является наиболее простым в реализации. При его использовании преобразуемая последовательность проходит ряд итераций, в ходе каждой из которых происходит поиск сегмента с минимальным КС и процедура разбиения для подготовки к следующей итерации. В алгоритме усреднения исходная последовательность в ходе преобразования разбивается на подблоки, каждый из которых представляется в виде совокупности сегментов с одинаковым размером слова. Следует отметить, что при этом КС выбранного подблока – наименьший среди альтернативных вариантов представления. Алгоритм усреднения с приближением дополнительно к алгоритму усреднения анализирует группы сегментов, объединяя или разбивая некоторые из них для получения меньшего значения КС.

Приводится структура преобразованного сегмента данных и осуществляется выбор, произведенный на основе анализа телемеханических данных, параметров алгоритмов, который включает в себя определение размера элементов заголовка сегмента.

Четвертая глава содержит результаты оценки возможностей предложенных способов на основе преобразования телемеханических данных системы сбора и обработки информации для контроля параметров технологического оборудования. В рамках данной работы были проанализированы 349 наборов данных телемеханических измерений и 32 набора данных телесигнализации.

Для проведения оценки все данные были классифицированы. Приводятся принципы проведения классификации, а также примеры распределения данных от разных датчиков, даются оценки по распределению данных по классам. Классификация включает статистический анализ разностных рядов первого и второго порядков. Результаты классификации приведены в табл. 1, 2.

Таблица 1
Классификация данных на основе анализа гистограммы распределения

Класс данных	Признаки класса
PP1	Характеристика ряда со значительным (более 95 % от всех значений) весом центрального значения гистограммы
PP2	Характеристика ряда с узкой гистограммой, с распределением 95 % основных значений в диапазоне ± 1
PP3	Характеристика ряда со средней гистограммой, с распределением 95 % основных значений в диапазоне ± 4
PP4	Характеристика ряда с широкой гистограммой, с распределением 95 % основных значений в диапазоне свыше ± 4

Приводятся результаты по оценке эффективности для алгоритмов предсказания для каждого класса данных. Среди алгоритмов предсказания рассматриваются алгоритмы экспоненциального сглаживания, экстраполяции нулевого порядка, линейной регрессии и простого скользящего среднего.

Таблица 2

Классификация данных на основе разностных рядов первого и второго порядка

Класс гистограммы для разностного ряда	PP1_1	PP1_2	PP2_2	PP2_3	PP3_3	PP3_4	PP4_4
1-го порядка	PP1	PP1	PP2	PP2	PP3	PP3	PP4
2-го порядка	PP1	PP2	PP2	PP3	PP3	PP4	PP4

Полученные данные показывают, что наибольшую степень сжатия в большинстве случаев обеспечивает использование алгоритма LZW. При этом эффективность данного алгоритма можно значительно повысить, используя алгоритм предсказания на основе разностного ряда. Алгоритм Хаффмана показывает наилучшие результаты с использованием алгоритма экспоненциального сглаживания.

Для исследования эффективности представления измерительных данных кубическими структурами приводятся результаты преобразования для следующих форм измерительных данных: однородных данных, однородных данных с большой избыточностью, данных с малой избыточностью, неоднородных данных.

Каждый тип измерительных данных был представлен выборкой из 20 кадров, которые и подвергались преобразованию с использованием двух критериев. Наибольшую эффективность алгоритмы показали для данных с малой избыточностью и неоднородных данных.

Оценка эффективности алгоритмов сегментирования включает описание результатов работы алгоритмов сканирования, усреднения, усреднения с приближением при преобразовании последовательностей различных классов. Производится непосредственное сравнение алгоритмов преобразования с алгоритмами RLE, Хаффмана, LZW, а также оценивается эффективность использования алгоритмов сжатия совместно с алгоритмами сегментирования.

Наибольшую эффективность алгоритмы показывают для классов данных PP1_1, PP1_2, PP2_2. При этом для данных классов использование алгоритма сжатия LZW совместно с алгоритмами сегментирования позволяет повысить эффективность сжатия по сравнению с применением алгоритма к исходным данным.

В приложении приводятся результаты экспериментов, не вошедшие в основное содержание, документы о внедрении результатов работы, свидетельства о регистрации программ для ЭВМ.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

1. В работе предложены способы предварительного преобразования, обеспечивающие повышение эффективности сжатия измерительных данных. На базе предложенных способов разработан ряд алгоритмов, позволяющих повысить эффективность некоторых словарных и статистических алгоритмов сжатия.

2. Предложенные на основе способа представления данных отрезками фиксированной длины алгоритмы позволяют учесть различные требования к преобразованию и различные характеристики преобразуемого сигнала. Так, алгоритм краевых отрезков позволяет получать более короткие последовательности, а алгоритм серединных отрезков устойчив к изменениям сигнала. В результате проведенного исследования на тестовых данных было найдено наилучшее соотношение размера выходной последовательности к погрешности преобразования для ряда типовых сигналов.

3. Исследование алгоритма на основе способа кубического преобразования позволяет рекомендовать использование данного алгоритма для данных с малой избыточностью, а также для неоднородных данных. В качестве основных критериев эффективности были предложены критерий длины повторов и блочный критерий, причем первый критерий предназначен для совместного использования со словарными алгоритмами сжатия, а второй – для статистических алгоритмов сжатия.

4. Предложенная классификация измерительных данных на основе разностных рядов первого и второго порядков, позволяет оценить возможность использования способов предварительного преобразования для различных типов данных и более точно выбрать алгоритм, обеспечивающий наибольшую эффективность сжатия.

5. Исследование алгоритмов сегментирования на основе предложенной классификации показывает, что наибольшую эффективность алгоритмы сегментирования демонстрируют при совместном использовании с алгоритмом сжатия LZW для классов данных PP1_1, PP1_2, PP2_2. Поскольку для исследуемой системы доля данных такого класса составляет примерно 60 %, использование алгоритмов сегментирования может быть оправданным.

6. Для типов данных PP2, PP3, PP4 использование алгоритмов экспонциального сглаживания, линейной регрессии и скользящего среднего может быть рекомендовано при совместном использовании с алгоритмом Хаффмана, поскольку именно для такой связки наблюдается увеличение эффективности по сравнению с разностным способом. В отличие от статистических алгоритмов, для словарных методов сжатия большая эффективность достигается при совместном использовании с разностным способом.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях, рекомендованных ВАК РФ

1. **Федяев, А. Ю.** Способ кубического преобразования для повышения эффективности сжатия данных / А.В. Левенец, А.Ю. Федяев, Е.У. Чье // Вестник Тихоокеанского государственного университета. – 2010. – № 3 (18). – С. 43–52.
2. **Федяев, А. Ю.** Оценка эффективности алгоритмов предсказания как процедуры предварительной обработки перед сжатием телемеханических данных / А.В. Левенец, А.Ю. Федяев, Е.У. Чье // Вестник Тихоокеанского государственного университета. – 2013. – № 2 (29). – С. 59–68.

Свидетельства о государственной регистрации программ для ЭВМ

3. Свидетельство о государственной регистрации программы для ЭВМ № 2102610271. Предварительное преобразование данных способом линейного кодирования на плоскости [Текст] / А.Ю. Федяев, А.В. Левенец. – Зарегистрировано в Реестре программ для ЭВМ 10.01.2012 (РФ).

4. Свидетельство о государственной регистрации программы для ЭВМ № 2013611458. Программа предварительной обработки телемеханических данных с возможностью предсказания значений [Текст] / А.Ю. Федяев, А.В. Левенец. – Зарегистрировано в Реестре программ для ЭВМ 21.01.2013 (РФ).

Публикации в других изданиях

5. **Федяев, А. Ю.** Применение кубического преобразования для повышения эффективности сжатия измерительных данных / А.Ю. Федяев // Материалы региональной научно-практической конференции «Молодежь и научно-технический прогресс», Владивосток, апрель-июль 2010. – Владивосток : Изд-во ДВГТУ, 2010. – Ч. 1. – С. 158–160.

6. **Федяев, А. Ю.** Простой способ предварительного преобразования телемеханических данных для повышения эффективности их сжатия / А.Ю. Федяев // Научная сессия ТУСУР – 2010: Материалы докладов Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых, Томск, 4–7 мая 2010 г. – Томск : В-Спектр, 2010. – Ч.5. – С. 209–212.

7. **Федяев, А. Ю.** Исследование способа линейного кодирования на плоскости для предварительного преобразования измерительных данных / А.В. Левенец, А.Ю. Федяев // Высокие технологии, фундаментальные исследования, экономика. Т. 2 : сборник статей Двенадцатой международной научно-практической конференции «Фундаментальные и прикладные исследования, разработка и применение высоких технологий в промышленности», Санкт-Петербург, 08–10 декабря 2011 года / под ред. А.П. Кудинова. – СПб : Изд-во Политехн. ун-та, 2011. – С. 119–123.

8. **Федяев, А. Ю.** Способ линейного кодирования на плоскости для предварительного преобразования данных / А.В. Левенец, А.Ю. Федяев // Сборник научных трудов SWorld. По материалам Международной научно-

практической конференции «Научные исследования и их практическое применение. Современное состояние и пути развития '2011». Том 4. Технические науки. – Одесса: Черноморье, 2011. – Т. 4. Технические науки. – Вып. 30. – С. 3–8.

9. **Федяев, А. Ю.** Алгоритм линейного преобразования измерительных данных с рандомизированным выбором оптимальных параметров / А.В. Левенец, А.Ю. Федяев // Перспективы развития информационных технологий : сборник материалов VII Международной научно-практической конференции / Под общ. ред. С.С. Чернова. – Новосибирск : Изд-во НГТУ, 2012. – С. 286–291.

10. **Федяев, А. Ю.** Классификация данных телемеханики на основе использования алгоритмов предсказания, для задач сжатия / А.В. Левенец, А.Ю. Федяев // Высокие технологии, исследования, образование, экономика. Т. 1 : сборник статей Четырнадцатой международной научно-практической конференции «Фундаментальные и прикладные исследования, разработка и применение высоких технологий в промышленности и экономике», Санкт-Петербург, 4–5 декабря 2012 года / под ред. А.П. Кудинова. – СПб. : Изд-во Политехн. ун-та, 2012. – С. 74–76.

11. **Федяев, А. Ю.** Оценка эффективности алгоритмов предсказания как процедуры предварительной обработки перед сжатием телемеханических данных / А.В. Левенец, А.Ю. Федяев // Сборник научных трудов SWorld : материалы международной научно-практической конференции «Научные исследования и их практическое применение. Современное состояние и пути развития '2012». – Вып. 3. – Т. 6. – Одесса : КУПРИЕНКО, 2012. – С. 7–11.

12. **Федяев, А. Ю.** Сравнение эффективности использования функций кодирования при осуществлении преобразования на плоскости / А.В. Левенец, А.Ю. Федяев // Высокие технологии, исследования, финансы. Т. 2: сборник статей Пятнадцатой международной научно-практической конференции «Фундаментальные и прикладные исследования, разработка и применение высоких технологий в промышленности и экономике», Санкт-Петербург, 25–26 апреля 2013 года / под ред. А.П. Кудинова. – СПб. : Изд-во Политехн. ун-та, 2013. – С. 58–60.

13. **Fedyaev, A. Y.** Selection Of Coding Function for Preliminary Data Preprocessing Using Coding on Plane Method / E.U. Chy, A.Y. Fedyaev, A.V. Levenets // Science, Technology and Higher Education: materials of the II international research and practice conference, Vol. II, Westwood, April 17th, 2013 – Westwood : Canada, 2013. – P. 106–111

14. **Федяев, А. Ю.** Способ кодирования на плоскости как процедура предварительного преобразования данных перед сжатием / А.В. Левенец, А.Ю. Федяев // Информационные технологии XXI века : материалы Международной научной конференции, Хабаровск, 20-24 мая 2013 г. – Хабаровск : Изд-во Тихоокеан. гос. ун-та, 2013. – С. 164–170.

Федяев Александр Юрьевич

**АЛГОРИТМЫ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ
ДЛЯ ЗАДАЧ СЖАТИЯ ДАННЫХ
В ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫХ СИСТЕМАХ**

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата технических наук

Подписано в печать 11.11.13. Формат 60x84 1/16.
Бумага писчая. Гарнитура «Таймс». Печать цифровая.
Усл. печ. л. 1,1. Тираж 120 экз. Заказ 258.

Отдел оперативной полиграфии
издательства Тихоокеанского государственного университета.
680035, Хабаровск, ул. Тихоокеанская, 136.