



УДК 004.046

© А. Ю. Федяев, А. В. Левенец, Ен Ун Чье, 2013

ОЦЕНКА ЭФФЕКТИВНОСТИ АЛГОРИТМОВ ПРЕДСКАЗАНИЯ КАК ПРОЦЕДУРЫ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ПЕРЕД СЖАТИЕМ ТЕЛЕМЕХАНИЧЕСКИХ ДАННЫХ

Федяев А. Ю. – асп. кафедры «Автоматика и системотехника», e-mail: futgik@gmail.com; *Левенец А. В.* – канд. техн. наук, доц. кафедры «Автоматика и системотехника», e-mail: levalvi@bk.ru; *Чье Ен Ун* – д-р техн. наук, проф. зав. кафедрой «Автоматика и системотехника», e-mail: chye@ais.khstu.ru (ТОГУ)

Производится анализ эффективности алгоритмов предсказания значений телемеханических данных, которые могут быть применены в процедуре предварительного преобразования с целью увеличения степени их последующего сжатия. Даются рекомендации по выбору способа предсказания значений в зависимости от используемого алгоритма сжатия. Предлагаются способы оптимизации кадров телемеханических данных.

Analysis of efficiency for algorithms of prediction remote control data values which can be used in procedure of data reduction to increase compression ratio is carried out. Recommendations for choosing way to predict values depending on compression algorithm used are given. Ways for optimization remote control data frame are proposed.

Ключевые слова: информационно-измерительная система, алгоритм сжатия, Хаффман, LZW, RLE, предварительное преобразование, коэффициент сжатия.

Методы предварительного преобразования успешно применяются к данным для увеличения эффективности процедур последующего сжатия [1, 2]. Это относится, в том числе, и к измерительным данным [3, 4]. Однако при этом выбор средств существенно ограничивается в связи с налагаемыми ограничениями: необходимостью, в большинстве случаев, передачи данных в реальном режиме времени, с небольшими задержками во времени, применением ограниченного, в отношении производительности, аппаратного обеспечения и, как следствие, необходимостью применения нетребовательных к аппаратным ресурсам алгоритмов. Увеличение масштабов автоматизации, и как следствие рост объемов передаваемых данных приводит к необходимости постоянного

усовершенствования и создания новых алгоритмов и методов сжатия, процедур предварительного преобразования.

В работе рассматриваются телемеханические данные ряда энергетических объектов Дальнего Востока. Каждый кадр данных состоит из определенного, фиксированного для всех кадров, числа опрашиваемых информационно-измерительной системой каналов, каждый из них в свою очередь включает в себя ограниченное для каждого канала число датчиков. Данные, полученные от каждого канала, снабжаются информацией о времени получения значений, количестве опрашиваемых датчиков, при этом каналные данные дополняются до фиксированного системой размера подкадра (рис. 1, а).

Анализ структуры кадра показывает большое содержание в ней избыточной информации. Например, в результате анализа выборки из 10000 кадров, полученных от информационно-измерительной системы, вероятность появления нулевого отсчета составила 80%. Поэтому перед проведением процедуры предварительного преобразования предлагается оптимизировать структуру кадра. Поскольку наибольшую избыточность вносят “хвостовые” данные, не несущие информационной нагрузки, их следует исключить из кадра, вместе с последовательностью байт, дополняющих заголовок до фиксированного размера. Перенос информации о размере в начало подкадра позволит для пустых подкадров, не содержащих информации об измеряемой величине, не передавать дополнительно временной штамп (рис. 1, б).

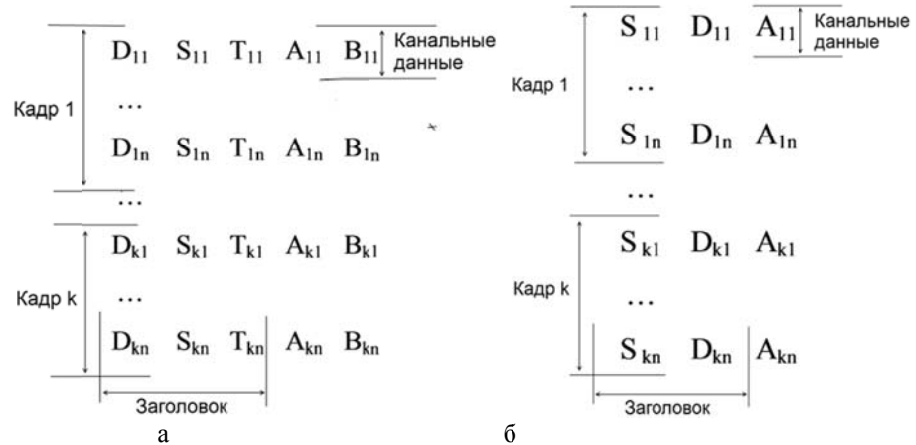


Рис. 1. Структура кадра телемеханики:

а – до оптимизации; б – после оптимизации.

D – набор байт выделяемых для передачи временной информации, S – набор байт для указания размера, T – байты дополняющие заголовок до фиксированного размера, A – набор байт для передачи информации с датчиков, B – дополнение каналных данных до фиксированного размера



Гистограмма для выборки из 1000 кадров после проведения преобразования представлена на рис. 2. Рисунок демонстрирует, что частота повторения нулевого байта после процедуры преобразования существенно снижается.

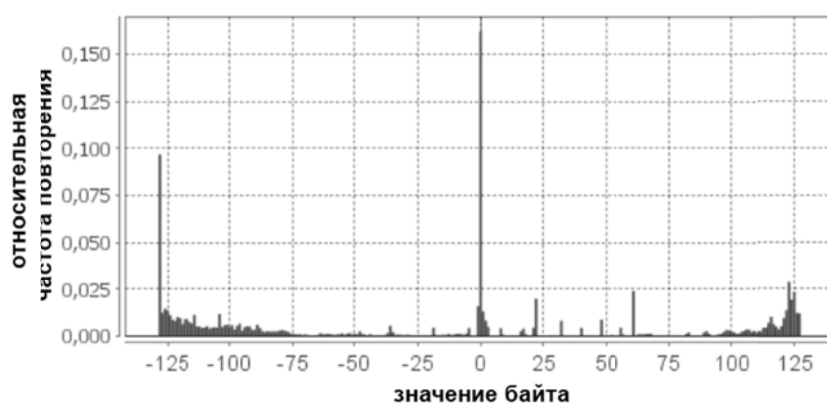


Рис. 2. Гистограмма распределения значений байт после анализа выборки из 10 000 кадров

Проведение данных шагов по оптимизации на тестовой выборке позволяет в среднем уменьшить размер кадра в 5 раз. При этом в самом худшем случае размер кадра уменьшится незначительно (за счет удаления дополнительных данных из заголовка), в лучшем случае уменьшится в 128 раз (остается один байт указывающий отсутствие данных).

В условиях передачи данных небольшими порциями, значительного внимания заслуживает механизм учета уже принятых значений. На основании анализа данных значений можно спрогнозировать, с некоторой вероятностью, какие значения будут приняты в дальнейшем. Процедура предсказания значений может повысить эффективность последующей процедуры сжатия, например, если применить алгоритм сжатия к разностному ряду образованному разностью предсказанных и действительных значений. При этом, чем точнее алгоритм предсказания повторяет изменение анализируемого параметра, тем эффективнее будет последующее сжатие. Сама процедура преобразования не изменяет структуру данных, и размер получаемой после преобразования последовательности не изменяется.

Для конкретной информационно-измерительной системы выбор используемого алгоритма предсказания может применяться отдельно к каждому опрашиваемому датчику. При этом возможно как статическое, так и динамическое задание алгоритма преобразования. Статическое задание алгоритма предполагает, что для опрашиваемого датчика известны необходимые, определяемые алгоритмом предсказания характеристики. На основании данных характеристик происходит выбор алгоритма, который сохраняется до конца процедуры опроса. При динамическом задании алгоритма характеристики опрашиваемого датчика либо не известны, либо могут значительно изменять-

ся в ходе опроса. При этом перед процедурой передачи выбирается алгоритм предсказания, который в ходе процедуры опроса может изменяться в зависимости от характера получаемых данных.

В качестве алгоритмов предсказания значений в данной работе рассматривались алгоритмы экспоненциального сглаживания, экстраполяции нулевого порядка, линейной регрессии, простого скользящего среднего. Каждый из этих алгоритмов прогнозирует следующий отсчет на основе набора предыдущих отсчетов и задаваемых параметров. Для экспоненциального сглаживания задается коэффициент сглаживания, увеличение значения которого позволяет уменьшить влияние ранних отсчетов в пользу более поздних, тем самым эффективнее реагировать на изменение параметров. Для алгоритма экстраполяции нулевого порядка задается параметр допустимого отклонения, влияющий на решение о смене зафиксированного ранее значения, используемого в качестве прогнозируемого на текущем шаге, на вновь полученное значение. Алгоритмы линейной регрессии и скользящего среднего используют параметр ширины окна, т.е. количества предыдущих отсчетов по которым вычисляется прогнозируемое значение. Алгоритм линейной регрессии отдает большее предпочтение более поздним значениям параметра, в то время как алгоритм скользящего среднего в одинаковой мере учитывает все накопленные значения.

Анализ эффективности осуществлялся на основе временных выборок из 10000 значений, полученных от датчиков для различных каналов. Выборки были классифицированы по группам. Группировка основывается на работе [5], в которой выделяются четыре класса телемеханических данных - PP1, PP2, PP3, PP4 в зависимости от концентрации значений их разностных рядов. PP1 включает ряды с весом центрального значения гистограммы более 95%. Класс PP2 включает ряды, для которых характерно распределение 95% основных значений в диапазоне ± 1 . PP3 ряды с распределением 95% основных значений в диапазоне ± 4 . PP4 включает ряды с распределением 95% основных значений в диапазоне свыше ± 4 .

Оценки эффективности сжатия производится на основе трех алгоритмов сжатия: RLE, LZW и Хаффмана. Выбор алгоритмов обосновывается тем, что алгоритм RLE потоковый и не требовательный к ресурсам, поэтому может использоваться как отдельно, так и в совокупности с другими алгоритмами сжатия. Алгоритм LZW универсальный и принадлежит к довольно распространенной, широко используемой группе словарных методов сжатия. Алгоритм Хаффмана является классическим представителем не менее распространенной, статистической группы алгоритмов, некоторые его модификации просты в реализации [1].

Временные выборки подвергаются предварительному преобразованию с использованием исследуемых алгоритмов прогнозирования и последующим сжатием с использованием перечисленных алгоритмов сжатия. На основании полученных результатов делается вывод об относительной эффективности процедур преобразования.



На рис. 3 приведены значения коэффициентов сжатия ($K_{сж}$) для исходных и преобразованных данных. Распределения коэффициентов сжатия показывает, что при возрастании класса данных от PP1 до PP4 значение коэффициента сжатия увеличивается, соответственно усложняется структура данных.

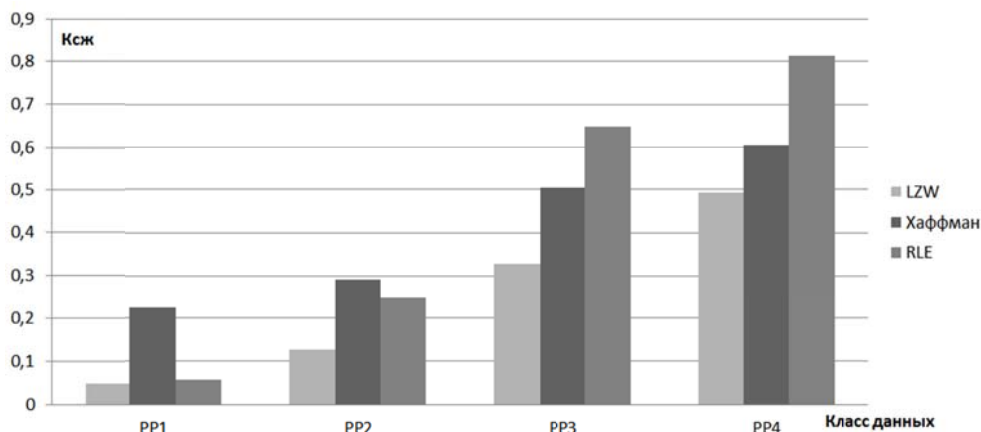


Рис. 3. Коэффициенты сжатия исходных данных различных классов

Анализ алгоритмов предсказания показывает, что алгоритмы экспоненциального сглаживания, экстраполяции нулевого порядка, линейной регрессии, скользящего среднего влияют на распределение статистических параметров преобразуемой последовательности. Вследствие этого алгоритм Хаффмана для последовательности класса PP2, PP3 показал более низкие коэффициенты сжатия, чем для последовательности класса PP1. В то же время использование контекстных алгоритмов, таких как LZW, RLE показывает, что исследуемые преобразования в равной степени влияют на структуру преобразуемых данных. С возрастанием класса преобразуемых данных, сложность структуры так же возрастает, и при применении контекстных методов коэффициенты сжатия увеличиваются при переходе от класса PP1 к PP4. Это говорит об универсальности алгоритмов преобразования и о примерно одинаковой эффективности при применении к любому из классов данных.

Среди рассматриваемых алгоритмов предсказания присутствует два алгоритма использующих параметр ширины окна, что позволяет произвести непосредственное сравнение алгоритмов между собой. Для классов данных с относительно не продолжительными во временной области всплесками (класс PP1), эффективность применения алгоритма скользящего среднего для последующего сжатия будет выше в связи с менее динамичной реакцией на изменения значения и фильтрацией частых сложно прогнозируемых всплесков (рис. 4, а).

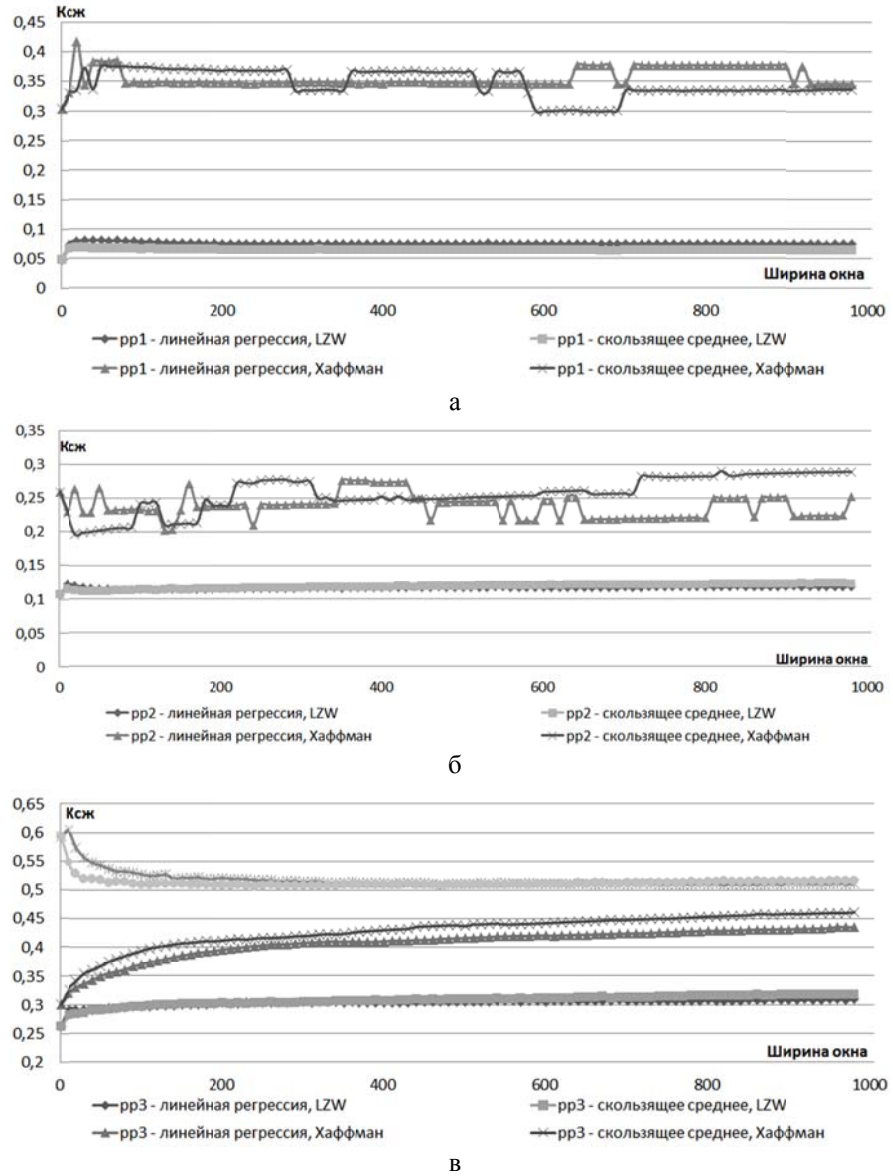


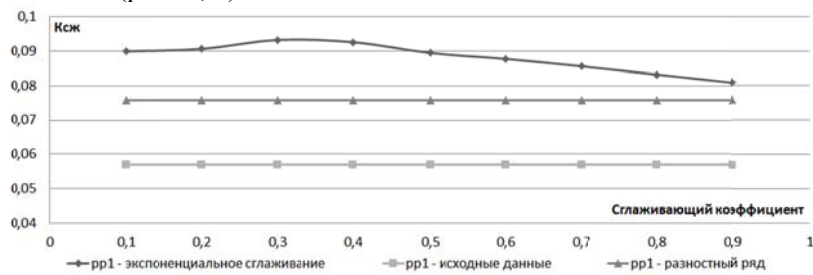
Рис. 4. Сравнение эффективности преобразования алгоритмами линейной регрессии и скользящего среднего для типа данных: а - PP1; б - PP2; в - PP3

Для классов сигналов с широкими нестационарными во временной области участками (сигналы классов PP2, PP3), эффективность применения алгоритма линейной регрессии будет выше в связи с более быстрой реакцией на изменения статистических характеристик и более быстрыми переходами в стационарное состояние (рис. 4, б, в).

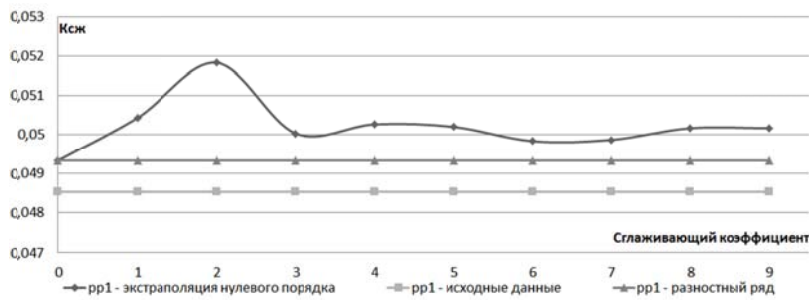


Таким образом, алгоритм скользящего среднего приобретает преимущество для классов сигналов PP1, PP4, алгоритм линейной регрессии для классов сигналов PP2, PP3.

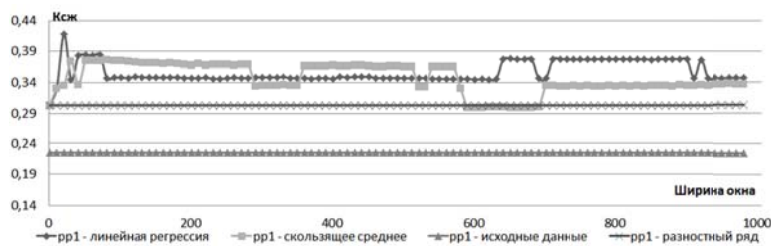
Оценка эффективности алгоритмов преобразования между собой осуществлялась с помощью коэффициентов сжатия исходной последовательности и разностного ряда полученного от исходной последовательности. Результаты показывают, что для типов данных PP1, использование предварительного преобразования, как правило, оказывается не эффективным (рис. 5). Аналогичным образом дело обстоит с алгоритмом RLE при обработке данных всех классов (рис. 5, а).



а



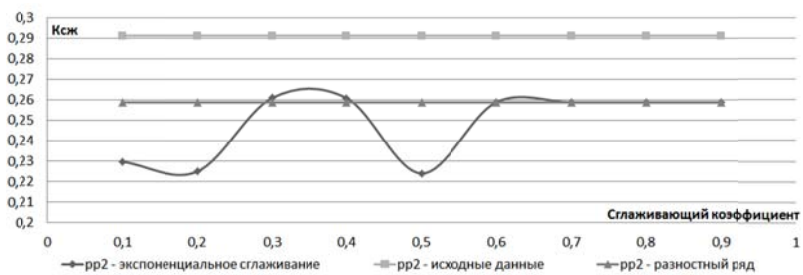
б



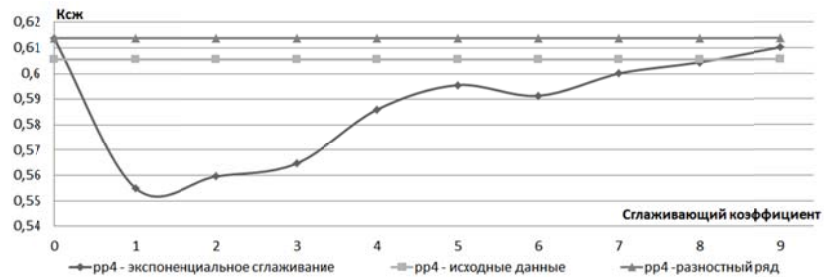
в

Рис. 5. Сравнение эффективности сжатия для данных типа PP1 в результате применения алгоритмов предсказания с исходными данными и разностным рядом: а - алгоритм экстраполяции нулевого порядка, с использованием алгоритма сжатия RLE; б - алгоритм экстраполяции нулевого порядка, с использованием алгоритма сжатия LZW; в - алгоритмы линейной регрессии, скользящего среднего, с использованием алгоритма Хаффмана

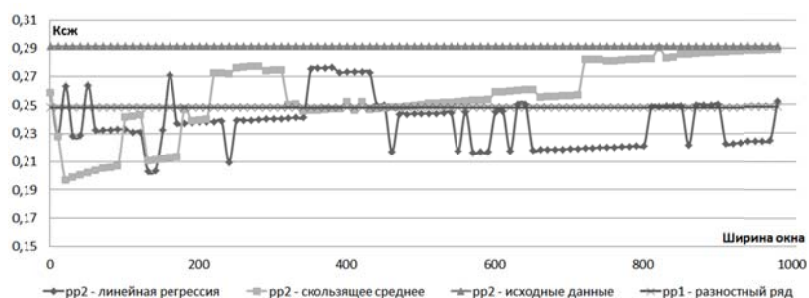
Для классов PP2, PP3, PP4 при использовании контекстных алгоритмов сжатия LZW, RLE с использованием алгоритмов экспоненциального сглаживания, алгоритмов линейной регрессии и скользящего среднего коэффициенты сжатия не превышают коэффициенты, полученные для исходной последовательности, однако превышают коэффициенты, полученные для разностного ряда. При преобразовании алгоритмом экстраполяции нулевого порядка, коэффициенты сжатия начинают превосходить коэффициенты сжатия, полученные не только для разностного ряда, но и, при достижении определенного значения ширины окна, коэффициенты, полученные для исходной последовательности.



а



б



в

Рис. 6. Сравнение эффективности сжатия для данных типа PP2 в результате применения алгоритмов предсказания с исходными данными и разностным рядом с использованием алгоритма Хаффмана:
а - экспоненциальное сглаживание; б - алгоритм экстраполяции нулевого порядка; в - алгоритм линейной регрессии



Однако для всех исследуемых типов преобразования при использовании алгоритма Хаффмана в некоторых ситуациях удается достичь уменьшения коэффициента сжатия по сравнению с исходными данными и разностным рядом (рис. 6). Это связано с перераспределением символов (байтов) после проведения преобразования.

Проведенные исследования показывают, что для алгоритмов сжатия LZW, RLE наиболее эффективным в большинстве случаев является использование разностного алгоритма предварительного преобразования. В этом случае удается достичь минимальных коэффициентов сжатия. Использование исследованных алгоритмов предсказания может показать свою эффективность при применении статистических методов сжатия, таких как метод Хаффмана и его разновидности. Так для исследуемой системы доля данных относящихся к классу PP2 составляло большинство (60%) [5], а именно для этой группы использование способов предварительного преобразования с алгоритмом Хаффмана показало свою эффективность.

В табл. 1 приведены сводные результаты по оптимизации объема передаваемых данных для исследуемой информационно-измерительной системы. В таблице представлены степени сжатия исходных данных с неизменной структурой кадра, с оптимизированной структурой кадра, а так же результаты сжатия в результате применения разностного метода и алгоритма экспоненциального сглаживания с коэффициентом сглаживания равным 0,1 для всех датчиков системы. Результаты построены по усреднению выборок из 10000 кадров каждая.

Таблица 1

Результаты по уменьшению объема передаваемых информационно-измерительной системой данных

| | Алгоритм сжатия | | |
|--|-----------------|---------|-------|
| | LZW | Хаффман | RLE |
| Исходные данные | 5,3 | 1,79 | 13,16 |
| Данные с оптимизированной структурой кадра | 5,56 | 6,38 | 4,94 |
| Исходные данные с разностным методом преобразования | 33,52 | 3,34 | 14,11 |
| Исходные данные с экспоненциальным сглаживанием | 12 | 3,42 | 6,65 |
| Данные с оптимизированной структурой кадра с разностным методом преобразования | 43,24 | 18,29 | 16,99 |
| Данные с оптимизированной структурой кадра с экспоненциальным сглаживанием | 13,16 | 18,31 | 7,01 |

Полученные данные демонстрируют, что наибольшую степень сжатия в большинстве случаев показывает использование алгоритма LZW. При этом эффективность данного алгоритма можно значительно повысить, используя алгоритм предварительного преобразования, наиболее подходящим из которых является разностный метод. Алгоритм Хаффмана показывает наилучшие

результаты с использованием алгоритма экспоненциального сглаживания, что говорит о корректных предположениях, сделанных при исследовании эффективности алгоритмов предсказания по отдельным датчикам. Применение оптимизации кадра позволяет дополнительно улучшить результаты сжатия для алгоритмов LZW и RLE в среднем в 1,1 раза, для алгоритма Хаффмана в 4,8 раза.

В результате оптимизации структуры кадра и использованию разностного метода сжатия для исследуемой системы удается достичь значительных коэффициентов сжатия. Исследуемые алгоритмы предсказания показывают результаты при применении статистических алгоритмов, таких как алгоритм Хаффмана.

В качестве возможных шагов для дальнейшего увеличения эффективности применения процедур сжатия следует рассмотреть корреляционные зависимости внутри самого кадра, а так же поиск алгоритмов преобразования, основанных на корреляционных зависимостях, которые дают более точные прогнозы и позволят после своего применения понизить класс данных, повысив тем самым степень сжатия.

Библиографические ссылки

1. *Ватолин Д.* Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео / Д. Ватолин, А. Ратушняк, М. Смирнов, В. Юкин - М.: ДИАЛОГ-МИФИ, 2003. - 384 с.
2. *Salomon D.* Data compression / David Salomon. New York: Springer, 2004. – 900 pages.
3. *Левенец А. В.* Способ кубического преобразования для повышения эффективности сжатия данных / А. В. Левенец, А. Ю. Федяев, Чье Ен Ун // Вестник Тихоокеанского государственного университета. – 2010. - №3 (18). - С.43-52
4. *Чье Ен Ун.* Представление телемеханических данных однородными параметрами структурами как предварительная обработка в задачах сжатия / А. В. Левенец, В. В. Нильга // Информационно-управляющие системы. - 2011. - №6. - С. 7-10
5. *Левенец А. В.* Классификация телемеханических данных и их разностных рядов с точки зрения задачи сжатия / А.В. Левенец // Вестник Тихоокеанского государственного университета. – 2009. - №4(15). - С.71-80.