

Веремчук В.А. (Veremchuk V.A.), студент,
Черепков В.В. (Cherepkov V.V.), студент
Руководитель – Тютякин А.В. (Tiutiakin A.V.), доцент
ФГБОУ ВПО «Госуниверситет – УНПК»

Сравнительный анализ алгоритмов и профилей сжатия сигналов в компьютеризированных системах контроля и диагностики технических объектов

Одним из распространенных классов интеллектуальных систем и сетей являются компьютеризированные системы контроля и диагностики (КСКД) технических объектов. Контрольно-диагностические решения в них принимаются на основании анализа (во временной или в частотной области) выходных сигналов датчиков физических величин, характеризующих состояние объекта контроля. Для массивов отсчетов указанных сигналов как исходных данных КСКД характерны следующие основные особенности [1]:

- значительные объемы каждого из файлов отсчетов (порядка сотен килобайт – мегабайта) при количестве файлов, подлежащих хранению в архиве документации каждого из объектов контроля, от нескольких десятков до нескольких тысяч;
- обязательность протоколирования и хранения исходных файлов отсчетов в электронных архивах в течение длительного времени (не менее нескольких лет);
- недопустимость каких-либо потерь информации, содержащейся в файлах, при их хранении в архивах.

Вышеуказанные особенности в большинстве практических случаев обуславливают необходимость **сжатия** указанных файлов. Сжатие должно производиться **без потерь информации**. Поэтому его рационально осуществлять способами и алгоритмами, предназначенными для сжатия дискретных сообщений, алфавит которых представляет собой конечное множество символов (цифры, служебные символы и т. п.), т. к. сжатие без потерь массивов отсчетов сигналов способами и алгоритмами,

ориентированными на компрессию мультимедийных данных, характеризуется заведомо меньшими коэффициентами сжатия [2, 3].

Известно достаточно много алгоритмов сжатия дискретных сообщений, наиболее распространенными из которых являются алгоритмы групп RPM, BZip, Deflate и LZ. Их краткие характеристики приведены в таблице 1.

Таблица 1

Распространенные алгоритмы сжатия дискретных сообщений

Алгоритм	Применяемые способы кодирования дискретных сообщений
RPMd	Статистические. Определение статистики сообщения – методом контекстного моделирования с предсказанием. Кодирование символов – арифметическое
LZMA	Словарное кодирование по модифицированному алгоритму LZ77
BZip2	Префиксное неравномерное кодирование Хаффмана с предварительными последовательными преобразованиями Барроуза – Уиллера и MTF
Deflate	Комбинация словарного кодирования по алгоритму LZ77 и префиксного неравномерного кодирования Хаффмана

Однако до настоящего времени в доступной литературе отсутствует систематизированный сравнительный анализ возможностей данных алгоритмов по сжатию массивов отсчетов сигналов. Целью настоящей статьи, является частичное восполнение данного пробела.

В качестве тестовых файлов были выбраны представленные в формате *.csv следующие массивы экспериментальных данных.

1. Массивы отсчетов выходного сигнала резистивного датчика состояния опоры качения при значениях овальности 5, 10 и 50мкм (файлы *ОКЧ_5мкм.csv*, *ОКЧ_10мкм.csv* и *ОКЧ_50мкм.csv* соответственно), использованные в как типовые примеры данных, анализируемых в частотной области.

2. Массивы отсчетов выходного сигнала фотометрического детектора жидкостного хроматографа при минимальных шуме и дрейфе, при повышенном дрейфе и при повышенных шуме и дрейфе (файлы *ХрК.csv*, *ХрД.csv* и *ХрШД.csv* соответственно). Использовались в качестве типового примера данных, анализируемых во временной области.

3. Объемы всех исследованных файлов – порядка нескольких сотен килобайт.

Были исследованы все алгоритмы сжатия дискретных сообщений, приведенные в таблице 1, и все профили данных алгоритмов, предоставляемые архиватором 7Zip.

Коэффициенты сжатия, обеспечиваемые различными профилями перечисленных в таблице 1 алгоритмов для тестовых файлов, представлены в таблице 2. Полужирным шрифтом выделены максимальные коэффициенты сжатия, полученные для каждого из файлов каждым из алгоритмов.

Таблица 2

Коэффициенты сжатия тестовых файлов, обеспечиваемые распространенными алгоритмами компрессии дискретных данных

Алгоритм	Профиль		Коэффициенты сжатия файлов					
	Размер словаря, Кбайт	Разрядность слова, бит	<i>OKЧ_5мкм.csv</i>	<i>OKЧ_10мкм.csv</i>	<i>OKЧ_50мкм.csv</i>	<i>ХрД.csv</i>	<i>ХрК.csv</i>	<i>ХрШД.csv</i>
1	2	3	4	5	6	7	8	9
PPMd	Все профили, доступные в архиваторе 7zip		14,7	15,3	24,2	26,9	3,33	4,71
BZip2	100	Не является выбираемым элементом профиля	13,4	15,3	21,2	22,3	3,04	3,76
	200		13,8		22,1			
	300-900		14,3		22,8			
Deflate32	32	8	8,13	8,35	11,9	14	3,05	4,02
		12	9,23	9,52	13,4	14,7	3,05	4,04
		16	9,73	10,1	14,4	15,6	3,05	4,02
		24	10,2	10,6	15,7	16	3,05	4,02
		32	10,3	10,7	16,3	16,5	3,06	4,02
		48	10,3	10,8	16,8	17	3,06	4,02
		64-258	10,3	10,8	16,9	17,2	3,06	4,02

Окончание таблицы 2

1	2	3	4	5	6	7	8	9
Deflate64	64	8	8,1	8,34	12	14	3,05	4,02
		12	9,2	9,5	13,4	14,7	3,05	4,04
		16	9,8	10,2	14,4	15,6	3,05	4,02
		24	10,4	10,7	15,9	16	3,05	4,02
		32	10,5	10,8	16,5	16,5	3,06	4,02
		48	10,5	10,9	16,8	17	3,06	4,02
		64-257	10,5	11	17,3	17,2	3,06	4,02
LZMA	64 КБ – 64 МБ	8	11,4	11,7	17,0	18,4	3,8	4,4
		16	12,4	12,8	19,0	19,3	3,8	4,4
		32	13,0	13,4	20,7	19,9	3,8	4,4
		64	13,0	13,6	21,6	20,2	3,8	4,4
		128	13,0	13,6	21,7	20,3	3,8	4,4
		256	13,0	13,5	21,8	20,4	3,8	4,4
		273	14,7	13,6	21,7	18,4	3,8	4,4

На основании данных, представленных в таблице 2, можно сделать следующие выводы.

1. Максимальные коэффициенты сжатия для всех типов данных, за исключением хроматограммы с минимальным уровнем шума и дрейфа, обеспечиваются алгоритмом RPPMd.

2. Коэффициент сжатия, обеспечиваемый алгоритмом RPPMd для файлов исследованных типов и объемов, не зависит от профиля алгоритма.

3. Максимальный коэффициент сжатия хроматограммы с минимальным уровнем шума и дрейфа обеспечивается алгоритмом LZMA (основанном на словарном методе сжатия). Это объясняется высокой повторяемостью и предсказуемостью данного массива отсчетов. Однако, при этом выигрыш алгоритма LZMA по сравнению с RPPMd незначителен (порядка 14%).

4. Коэффициенты сжатия, обеспечиваемые алгоритмом LZMA для файлов исследованных типов и объемов, практически не зависят от объема словаря, но наблюдается их зависимость от разрядности слова, имеющая, в общем случае, немонотонный характер (см. таблицу 2).

Таким образом, в целом, из распространенных алгоритмов сжатия дискретных данных (см. таблицу 1) алгоритм RPPMd наиболее предпочтителен для сжатия массивов отсчетов сигналов независимо от области их представления (частотной или временной).

Список литературы

1. Неразрушающий контроль и диагностика: Справочник / Под ред. В. В. Ключева. – М.: Машиностроение, 2003. – 656 с.
2. Сэломон, Д. Сжатие данных, изображений и звука. – М.: Техносфера, 2006. – 368 с.
3. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео / Д. Ватолин, А. Ратушняк, М. Смирнов, В. Юкин. – М.: Диалог-МИФИ, 2003. – 384 с.