

УДК: 629.783

Вероятностный и структурный метод сжатия данных глобальных навигационных спутниковых систем

А. Д. Борискин

Аннотация

Статья посвящена двум способам сжатия данных глобальных навигационных спутниковых систем (ГНСС). Первый способ – вероятностный, когда более вероятные значения кодируются коротким кодом, а менее вероятные – более длинным. Второй способ – структурный, позволяет более компактно представить информацию о номерах спутников и сигналов в виде маски.

Ключевые слова:

сжатие нормального процесса; сжатие по хатфману; маска спутников и сигналов; мультисигнальные сообщения.

Сокращения:

ГНСС – Глобальные Навигационные Спутниковые Системы.

RTCM3 - Radio Technical Commission for Maritime Services (международный формат записи дифференциальных данных версии 3).

CMR – Compact Measurement Record (формат записи дифференциальных данных фирмы TRIMBLE).

Введение

Задачи сжатия являются ключевыми в системах хранения, обработки и передачи ГНСС данных. Наибольшее распространение сжатие получило в дифференциальной навигации, когда требуется передача измерений привязываемого приемника на базовый через каналы с ограниченной пропускной способностью. Не менее востребовано сжатие и

при записи «сырых» измерений приемника (особенно высокочастотных) на флеш память для их последующей пост обработки.

Основным способом сжатия ГНСС измерений является уменьшение их динамического диапазона. Существующие методы, позволяют сузить диапазон кодовых измерений до $\pm 300[m]$ или $16[бит]$ (при шаге квантования в $2[см]$), а фазовых до $\pm 2048[u]$ (RTCM3, CMR) или $20[бит]$ (при шаге квантования в $1/256[u]$). Для более эффективного сжатия, совместно с существующими методами, можно использовать и другие. Среди них – вероятностное и структурное сжатие, речь о которых пойдет в данной статье.

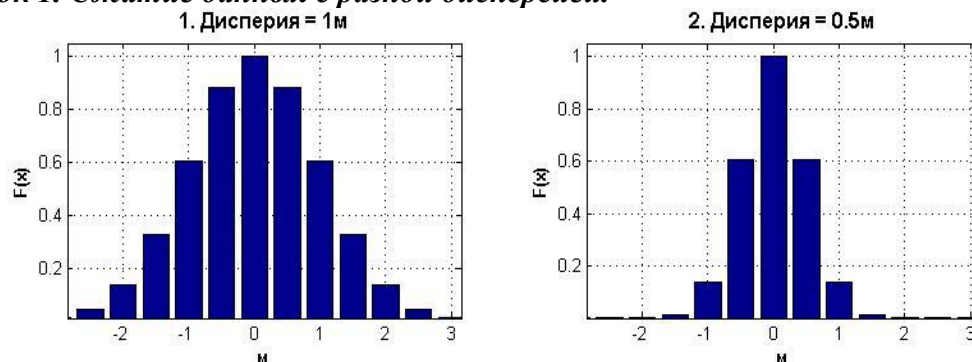
Вероятностное сжатие.

Вероятностное сжатие широко используется при работе с различными типами данных. Оно эффективно при условии, что функция распределения кодируемой величины отлична от равномерной. Тогда, для записи наиболее вероятных значений используют короткий двоичный код, а для записи менее вероятных – более длинный код. Как результат, суммарный объем информации уменьшается по сравнению с обычным битовым уплотнением.

Часто вероятностное сжатие используется с величинами, распределенными по нормальному закону. В этом случае, эффективность сжатия зависит от 2-х параметров:

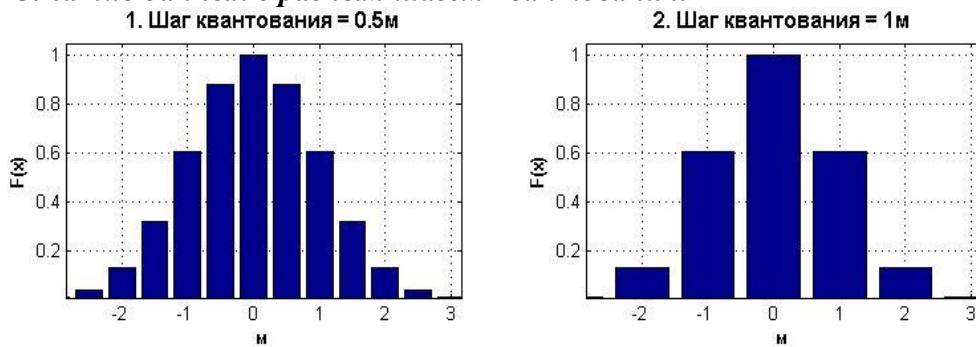
1. Дисперсия данных: На Рис. 1 представлены 2 процесса с различными дисперсиями ($\sigma=1[m]$ и $\sigma=0.5[m]$) и одинаковым шагом квантования ($\Delta=0.5[m]$). Очевидно, что вероятностное сжатие 2-ого процесса будет эффективнее, т.к. оно имеет меньшую дисперсию.

Рисунок 1. Сжатие данных с разной дисперсией.



2. Шаг квантования данных: На Рис. 2 представлены 2 процесса с различным шагом квантования, или, что равнозначно, с различной длиной бинарного слова. Вопрос, где сжатие более эффективно не столь очевиден и требует дополнительных расчетов.

Рис. 2. Сжатие данных с разным шагом квантования.

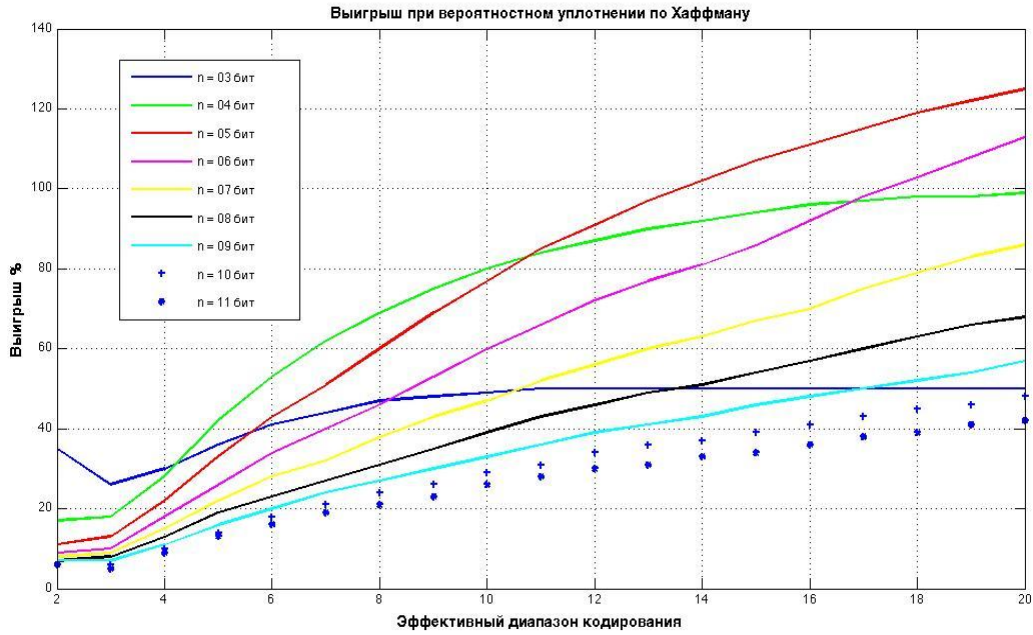


С помощью компьютерного моделирования для различной длины бинарных слов (n) и эффективного диапазона ($N\sigma$) были получены серии кривых, определяющих эффективность ($L=F(N\sigma, n)$) вероятного сжатия по Хаффману (Рис. 3). Эффективность для заданных характеристик n и N , рассчитывалась по формуле

$$L = (N / N_c - 1) \cdot 100\% \quad (1)$$

- N – количество бит при битовом уплотнении
- N_c – среднее количество бит при вероятностном сжатии

Рис. 3. Коэффициент сжатия



Для объяснения графика рассмотрим пример: пусть необходимо оценить эффективность сжатия нормального случайного процесса в диапазоне $T=\pm 5[m]$ с дисперсией $\sigma=0.5[m]$ и шагом квантования в $2[см]$. При обычном битовом уплотнении понадобится $\log_2(10/0.02)=9[бит]$ информации. По графику находим пересечение линии, соответствующей $n=9[бит]$, и эффективного диапазона кодирования $N\sigma=T/\sigma=5/0.5=10$ и получаем порядка 30% или $9/1.3\approx 7[бит]$.

Основные принципы вероятностного сжатия могут быть применимы и к ГНСС измерениям. Для его эффективного использования необходимо найти такую линейную комбинацию измерений, функция распределения которой, будет иметь минимальную дисперсию. Существующие комбинации ГНСС измерений, такие как кодовые и фазовые коррекции (RTCM2), разностные измерения (RTCM3, CMR) плохо подходят. Они имеют широкий диапазон (порядка $200-300[m]$) и их дисперсия и математическое ожидание меняются во времени и сильно зависят от угла возвышения спутника. Для увеличения степени сжатия необходимо воспользоваться свойством сильной корреляции измерений во времени.

Рассмотрим приращение основных измерений во времени.

$$\begin{aligned}
 \Delta\varphi_1 &= (\Delta D - \Delta I_1 + \Delta O + \Delta T) / \lambda_1 + \Delta t + \Delta\eta_{\varphi,1} \\
 \Delta\rho_1 &= \Delta D + \Delta I_1 + \Delta O + \Delta T + \Delta t + \Delta\eta_{\rho,1} \\
 \Delta\varphi_2 &= (\Delta D - \Delta I_2 + \Delta O + \Delta T) / \lambda_2 + \Delta t + \Delta\eta_{\varphi,2}
 \end{aligned}
 \tag{2}$$

- $\Delta\varphi_1 [\text{ц}], \Delta\rho_1 [\text{м}]$ – приращение фазовых и кодовых измерений
- $\Delta t [\text{м}]$ – приращение ухода часов приемника относительно системного времени.
- $\Delta I_1, \Delta O, \Delta T [\text{м}]$ – приращение ионосферной, орбитальной, тропосферной ошибок измерений.
- $\Delta\eta_{\varphi,1}, \Delta\eta_{\rho,1}, \Delta\eta_{\varphi,2} [\text{м}]$ – приращение шумовых и других ошибок измерений.

Введем понятие разностного измерений, тогда приращение разностных измерений можно выразить через приращение фазового измерения на первой частоте.

$$\begin{aligned}\Delta\rho_1 &= \Delta\varphi_1 + \zeta\rho \\ \Delta\varphi_2 &= \Delta\varphi_1 + \zeta\varphi\end{aligned}\quad (3)$$

- $\zeta\rho, \zeta\varphi$ - ошибки восстановления, представляющие собой случайные величины.

При отсутствии срывов слежений, функция разностного фазового измерения будет определяться изменением ионосферы, скорость которой ограничена и не превышает $3[\text{см/сек}]$, а для кодового разностного измерения – шумовой ошибкой и не превышает $3[\text{м}]$. Приращение разностного фазового измерения можно рассматривать как нормальный случайный процесс с усредненной среднесуточной дисперсией $\sigma_{\zeta\varphi} < 3[\text{мм/сек}]$. Зная начальное значение разностного фазового измерения и его приращения, можно безошибочно восстановить полное разностное фазовое измерение. При шаге квантования в $0.5[\text{мм}]$ для передачи фазового измерения ($\Delta\varphi_2$) с частотой $1[\text{Гц}]$ понадобится $\log_2(3[\text{см}]/0.05[\text{см}])=6[\text{бит}]$, а при использовании вероятностного сжатия (Рис. 3: $n=6, N=3/0.3=10 \rightarrow L=60\%$) – $5/1.6 \approx 3.1[\text{бит}]$.

Структурное сжатие.

ГНСС данные состоят не только из измерений, в них также присутствуют дополнительные идентификаторы измерений, такие как: номера спутников и сигналов, а также их количество. В зависимости от того, в какой последовательности и как записана эта информация, можно осуществить дополнительное сжатие.

Рассмотрим пример: пусть необходимо записать последовательность целых чисел $32,1,6,7,13,3,15$. Известно, что диапазон любого числа в последовательности находится в промежутке от 1 до 32. Существует 2 способа записи приведенной последовательности.

Первый, и наиболее распространенный – последовательная запись, когда целые числа (точнее их бинарный код) записываются друг за другом. В этом случае, суммарный объем данных будет равен $5 \cdot 7 = 35 [\text{бит}]$ (5 – количество бит необходимое для записи 1-ого числа, 7 – количество чисел в последовательности).

В случае если в последовательности нет повторяющихся чисел и важны только их значения, но не порядок следования, тогда возможна их запись в виде маски. Маска представляет собой последовательность бит размером в их диапазон (32), где единицы стоят на позициях, соответствующих значениям чисел. Для рассматриваемого примера, маска будет иметь вид $10100110000010100000000000000001$. Суммарный объем данных будет равен 32 бит независимо от количества чисел в последовательности.

Второй способ требует меньше бит и потому может рассматриваться как способ сжатия. Сжатие с использованием маски имеет смысл, когда количество чисел в последовательности больше некоторого предельного значения (N), которое может быть рассчитано как:

$$N > K / \log_2(K) \tag{4}$$

- K – диапазон чисел в последовательности.

Для ГНСС данных также можно использовать масочный подход для сжатия. С помощью них можно передавать информацию о номерах спутников и сигналах.

Известно, что один спутник может излучать множество сигналов, причем несколько на одной частоте. В статье будут использоваться сокращения названий сигналов, используемых в стандарте RINEX. Рассмотрим пример: пусть приемник следит за спутниками $1,3,6,7,13,15,32$ и сигналами $1C(2)$, $1P(4)$, $2P(10)$, $2M(15)$. В таблице слежения (Таблица 1) буквой X обозначены те сигналы, из всех возможных, за которыми следит в данный момент приемник. Диапазон в 40 спутников выбран с запасом, и является максимальным для всех существующих и планируемых ГНСС систем, диапазон в 24 сигнала, также выбран с запасом. В Таблице 2 приводится соответствие между номером сигнала и его общепринятым обозначением.

Таблица 1. Таблица слежения приемника

Спутник	1	2	3	4	5	6	7	8	..	1	1	1	1	...	3	...	4	Маска
Сигнал									.	2	3	4	5		2		0	сигналов

1																	0
2	X		X			X	X				X		X		X		1
3																	0
4	X		X			X					X				X		1
5																	0
6																	0
7																	0
8																	0
9																	0
10	X		X			X					X				X		1
11																	0
12																	0
13																	0
14																	0
15	X						X					X		X			1
16																	0
...																	...
24																	0
Маска спут-ов	1	0	1	0	0	1	1	0	0	0	1	0	1	...	1	...	0

Под сигналом подразумевается не только физические характеристики сигнала, такие как частота, тип модуляции и т.д., но и способы слежения за ним. Серым цветом в *Таблице 2* выделены наиболее распространенные в настоящее время сигналы.

Таблица 2. Идентификация сигналов

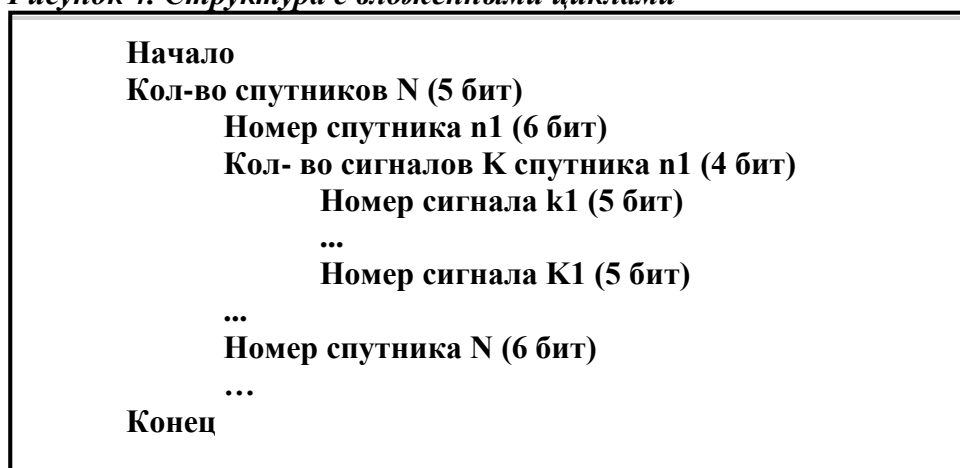
Номер сигнала	GPS, RINEX обозначение	SBAS, RINEX обозначение	ГЛОНАСС, RINEX обозначение	Galileo, RINEX обозначение
1				
2	1C	1C	1C	1C
3	1P		1P	1A
4	1W			1B
5	1Y			1X
6				1Z
7				
8	2C		2C	6C
9	2P		2P	6A
10	2W			6B
11	2Y			6X
12				6Z
13				
14				7I
15	2S			7Q
16	2L			7X
17	2X			
18				8I

19				8Q
20				8X
21				
22	5I	5I		5I
23	5Q	5Q		5Q
24	5X	5X		5X

Существует несколько способов записи информации о номерах следимых спутников и сигналов.

Первый – это структура с внутренними циклами, которая в том или ином виде используется в протоколах RTCM3 и CMR. Информация о номерах спутника передается в первом цикле, а информация о номере сигнала во вложенном в него втором цикле. На *Рис 4* обозначена описанная выше схема.

Рисунок 4. Структура с вложенными циклами



Рассчитаем объем информации необходимый для записи номеров спутников и сигналов по схеме, изображенной на *Рисунке 4*. Будем исходить из предположения, что количество спутников, за которыми одновременно следит приемник, не превышает 32, а количество сигналов одного спутника не превышает 16.

$$L = 5 + (6 + 4) \cdot N + 5 \cdot S = 5 + 10 \cdot 7 + 5 \cdot 21 = 180[\text{бит}] \quad (5)$$

- N – количество спутников.
- S – количество сигналов.

Другим способом записи является запись с использованием масок. Таблицу слежения можно представить в виде 3-ех отдельных масок:

- Маска сигналов (серая вертикальная строка в Таблице 1)
- Маска спутников (серая горизонтальная строка в Таблице 1)
- Композитная таблица масок (Таблица 3)

Таблица 3. Композитная таблица масок

Спутники Сигналы	1	3	6	7	13	15	32
2	1	1	1	1	1	1	1
4	1	1	1	0	1	0	1
10	1	1	1	0	1	0	1
15	1	0	0	1	0	1	1

Композитная таблица масок получается из «таблицы слежения» путем удаления нулевых столбцов и строк. Ее размер равен $N_{sat} \cdot N_{sig}$, где N_{sat} - это количество спутников (=количеству единиц в маске спутников), а N_{sig} – это количество сигналов (=количеству единиц в маске сигналов).

Рисунок 5. Структура без вложенных циклов

<p>Начало. Маска сигналов (24 бит) Маска спутников (40 бит) Композитная маска ($N_{sat} \cdot N_{sig}$) Конец</p>

Рассчитаем объем информации необходимый для записи номеров спутников и сигналов по схеме, изображенной на рисунке 5.

$$L = 24 + 40 + N_{sat}N_{sig} = 64 + 28 = 72[\text{бит}] \quad (6)$$

Из представленного примера видно, что запись информации о номерах спутников и сигналов в виде масок, позволяет сжать этот тип информации более чем в 2 раза.

Заключение

В статье изучена возможность использования вероятностного сжатия при работе с ГНСС данными. Оценена эффективности сжатия по Хаффману величины, распределенной по нормальному закону. Предложены конкретные примеры для ГНСС данных, где вероятностное сжатие имеет целесообразное применение.

Наряду с вероятностным, рассмотрено структурное сжатие, при котором информация о номерах спутников и сигналов передается в виде масок. Представление данных в виде маски не только позволяет произвести дополнительное сжатие, но и решает многие проблемы существующих форматов хранения данных, таких как RTCM3 (в котором не существует возможности сохранять более 2-ух измерений одного спутника одновременно). На основе предложенного автором метода, в RTCM3 комитете разработаны новые сообщения (MSM – мультисигнальные сообщения), для записи сырых измерений приемника и полностью совместимых с RINEX форматом.

Библиографический список

1. William J Kellar. Transmission of RTK Corrections and Measurements using Optimal Coding. Journal of Global Positioning System, 2006
2. Leick. A. GPS satellite surveying. Second Edition. John Wiley&Sons Inc., 1995. 584 стр.
3. Тепляков. И.М. и др. Радиосистемы передачи информации. - М. Радио и Связь, 1982. 264 стр.
4. Bradford W. Parkinson. Global Positioning System: Theory and Applications. Volume 1. American Institute of Aeronautics, 1996. 793 стр.

Сведения об авторах

Борискин Алексей Дмитриевич, аспирант Московского авиационного института (государственного технического университета).

МАИ, Волоколамское ш., 4, Москва, А-80, ГСП-3, 125993;
тел:1582518; e-mail: adbtor@mail.com