

ГЛАВА 3. ЭНЕРГЕТИКА И ЭНЕРГОСБЕРЕГАЮЩИЕ ТЕХНОЛОГИИ

СЖАТИЕ ДАННЫХ ДИСКРЕТНОГО ПРЕДСТАВЛЕНИЯ ОБЛАСТЕЙ РАБОТОСПОСОБНОСТИ МЕТОДОМ КОДИРОВАНИЯ ДЛИН СЕРИЙ

Назаров Д.А.

Институт автоматизации и процессов управления ДВО РАН, Владивосток
nazardim@iacp.dvo.ru

Работа посвящена проблеме уменьшения объема данных дискретного представления многомерных областей работоспособности технических систем. Рассматривается метод кодирования длин серий (RLE), как наиболее подходящий к решению задачи уменьшения избыточности данных без потери информации ввиду специфики данных, а также возможности быстрого доступа к ним без полной распаковки. Проводится экспериментальное исследование эффективности сжатия в зависимости от размера полей, кодирующих длину серии повторяющихся элементов.

Ключевые слова: область работоспособности, надежность, сжатие данных.

Compression of the data of acceptability regions discrete representation using run length encoding algorithm. Nazarov D.A., Institute of Automation and Control Processes FEB RAS.

The work is devoted to the problem of large data amount of multidimensional acceptability region discrete representation. The method of run length encoding (RLE) is considered as the most appropriate lossless data compression algorithm which provides fast data access without its full decompression. An experimental study of compression efficiency depending on the size of the fields encoding the length of a series of repeating elements is carried out.

Keywords: acceptability region, reliability, data compression.

Введение

Задача построения областей работоспособности (ОР) возникает на этапе проектирования сложных технических систем с учетом случайных параметрических возмущений и требований надежности. Под ОР понимается множество точек пространства параметров элементов системы, в которых эта система выполняет возложенные на нее функции в рамках требований спецификации (находится в работоспособном состоянии). Характеристики ОР, как правило, неизвестны, а их нахождение позволяет оценить способность системы сохранять работоспособность при отклонениях параметров ее элементов от расчетных значений, вызванных различными технологическими и эксплуатационными факторами, назначать допуски на параметры элементов, выбирать оптимальные в определенном смысле номинальные значения параметров. К основным трудностям построения и анализа ОР относятся высокая размерность пространства параметров, отсутствие априорной информации о конфигурации области, частое отсутствие явных выражений зависимостей выходных характеристик от параметров элементов в аналитическом виде [1, 2].

Среди множества известных методов построения ОР в данной работе рассматривается метод, основанный на представлении многомерной области дискретным множеством элементарных гиперпараллелепипедов, заданных узлами регулярной сетки. Основная проблема такого представления состоит в большом объеме данных, необходимых для описания области в рамках этого представления. В работе [1] описана идея применения методов сжатия данных ОР, основанных на их двоичном характере: кодирование двоичными разрядами целого числа и алгоритм кодирования длин серий повторяющихся значений (Run Lengths Encoding - RLE)[3]. Последний метод позволяет достигать существенно более высоких степеней сжатия данных при сохранении возможности доступа к определенному элементу массива за один проход без полной его распаковки.

Для эффективного использования метода кодирования длин серий необходимо определить оптимальный размер (в двоичных разрядах) элемента, кодирующего длину серии из одинаковых элементов массива. Данная работа посвящена экспериментальному исследованию влияния размера кодирующего элемента на эффективность сжатия данных дискретного представления ОР.

Дискретное представление области работоспособности

Объектом исследования является модель аналоговой технической системы, связывающая свои выходные характеристики, заданные m -вектором:

$$\mathbf{y} = (y_1, y_2, \dots, y_m) \quad (1)$$

с параметрами ее элементов (*внутренние параметры*), заданными n -вектором:

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \quad (2)$$

в виде зависимостей:

$$y_j = y_j(\mathbf{x}), \forall j = 1, 2, \dots, m \quad (3)$$

На выходные и внутренние параметры налагаются ограничения. Внутренние параметры ограничиваются производственными допусками:

$$x_{i \min} \leq x_i \leq x_{i \max}, \forall i = 1, 2, \dots, n \quad (4)$$

а выходные параметры — требованиями к выходным характеристикам системы, как правило, заданным в техническом задании:

$$y_{j \min} \leq y_j \leq y_{j \max}, \forall j = 1, 2, \dots, m \quad (5)$$

Ограничения (5) называются *условиями работоспособности* и определяют критерий корректного функционирования системы. Нарушение какого-либо из неравенств (5) квалифицируется как отказ. К такого рода отказам приводят обычно изменения внутренних параметров вследствие их дрейфа под влиянием факторов внешней среды (влажность, температура, излучения) и внутренних процессов износа и старения. Условия работоспособности (5) определяют в пространстве внутренних параметров область, которая называется областью работоспособности:

$$D_x = \{\mathbf{x} \in R^n : y_{j \min} \leq y_j(\mathbf{x}) \leq y_{j \max}, \forall j = 1, 2, \dots, m\} \quad (6)$$

Основную трудность определения характеристик ОР составляют размерность пространства параметров и отсутствие явных аналитических выражений модели (3) для большинства сложных систем ввиду того, что они часто задаются алгоритмически или в виде имитационной модели, вычисляющей выходные характеристики для заданного набора внутренних параметров [4].

В данной работе рассматривается способ построения ОР на основе аппроксимации многомерной фигуры дискретным множеством *элементарных гиперпараллелепипедов* (ЭГ), заданных узлами регулярной сетки, и методе многомерного зондирования внутри этих ЭГ.

Построение ОР выполняется внутри ограничивающего бруса B , заданного интервалами:

$$B = \{[x_{i \min}, x_{i \max}], \forall i = 1, 2, \dots, n\}. \quad (7)$$

В качестве этого бруса может выступать как брус допусков (4), так и описанный вокруг ОР брус, найденный методом Монте-Карло [5].

Каждый диапазон $[x_{i \min}, x_{i \max}], i = 1, 2, \dots, n$ разбивается эквидистантно на q_i интервалов - квантов. Их пересечения формируют узлы сетки, которые являются вершинами ЭГ. Каждый ЭГ задается набором n целочисленных индексов $(k_1, k_2, \dots, k_n), k_i = 1, 2, \dots, q_i$. По этим индексам, зная диапазоны ограничивающего бруса (9) и количества квантов $q_i, i = 1, 2, \dots, n$ можно однозначно восстановить геометрические характеристики ЭГ:

$$\begin{aligned} x_{i \min}^{k_i} &= x_{i \min} + h_i \cdot (k_i - 1), \\ x_{i \max}^{k_i} &= x_{i \min}^{k_i} + h_i = x_{i \min} + h_i \cdot k_i \end{aligned} \quad (8)$$

где $h_i = (x_{i \max} - x_{i \min}) / q_i$ — шаг сетки по каждому параметру внутри бруса B .

В центре каждого ЭГ выбирается точка-представитель:

$$x_r(k_1, k_2, \dots, k_n) = \left(\frac{x_{1 \min}^{k_1} + x_{1 \max}^{k_1}}{2}, \dots, \frac{x_{n \min}^{k_n} + x_{n \max}^{k_n}}{2} \right), \quad (9)$$

в которой вычисляются выходные характеристики системы (1). Если полученные выходные характеристики удовлетворяют ограничениям (5), то ЭГ с индексами $(k_1, k_2, \dots, k_n)$ присваивается признак «1», противном случае - «0». Таким образом, для хранения признаков принадлежности ОР каждого ЭГ используется *множество индикаторов* (МИ):

$$S = (s_1, s_2, \dots, s_R), \quad (10)$$

где $R = q_1 \cdot q_2 \cdot \dots \cdot q_n$ — количество ЭГ, $s_p \in \{0, 1\}$ — результат вычисления функции принадлежности:

$$\chi(k_1, k_2, \dots, k_n) = \begin{cases} 1, & y_{j \min} \leq y_j(x_r(k_1, k_2, \dots, k_n)) \leq y_{j \max}, j = 1, 2, \dots, m \\ 0, & y_{j \min} > y_j(x_r(k_1, k_2, \dots, k_n)) \vee y_{j \max} < y_j(x_r(k_1, k_2, \dots, k_n)) \end{cases}, \quad (11)$$

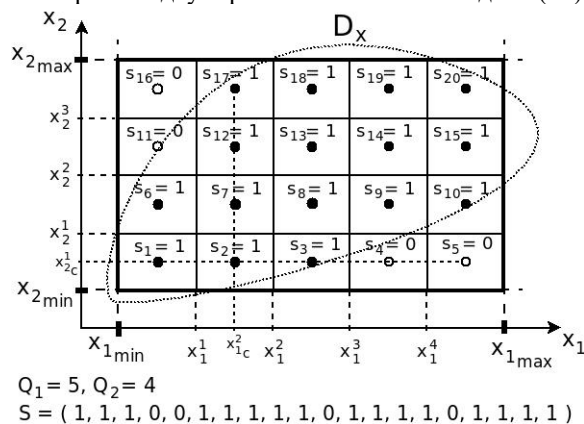
Между индексом p элемента s_p множества индикаторов S и набором индексов $(k_1, k_2, \dots, k_n)$ существует взаимнооднозначное соответствие [1]:

$$p(k_1, k_2, \dots, k_n) = k_1 + q_1 \cdot (k_2 - 1) + q_1 q_2 (k_3 - 1) + \dots + q_1 q_2 \dots q_{n-1} (k_n - 1), \quad (12)$$

Таким образом, модель дискретного представления ОР описывается четверкой:

$$G_R = (n, B, Q, S), \quad (13)$$

где, n — размерность пространства внутренних параметров, B — ограничивающий брус (7), $Q = (q_1, q_2, \dots, q_n)$ — вектор с количеством квантов по каждому параметру, S — множество индикаторов принадлежности каждого ЭГ в соответствии с функцией (11). Построение ОР на основе этой модели состоит в вычислении всех элементов множества S индикаторов принадлежности. На рисунке 1 проиллюстрирован принцип построения двумерной ОР на основе модели (13).



Как можно видеть, большинство серий имеют длину менее 255, а также имеются единичные серии длиной в десятки тысяч элементов. Если их разбить на серии по 255 элементов, сократив размер, то возникнет избыточность, однако, как показано в таблице 1, такая избыточность незначительна и именно такой способ кодирования длин серий показывает наилучшую степень сжатия. В таблице 1 приведены результаты эксперимента применения алгоритма RLE к МИ данных ОР моделей ТТЛ [5] и делителя напряжения [6] при квантовании области построения на 40, 60, 80 и 100 квантов по каждому параметру.

Таблица 1 – Эффективность сжатия МИ методом кодирования длин серий повторяющихся элементов

Кол-во квантов по всем параметрам	Ширина поля длины серии, байт	Эффективность сжатия МИ, крат.	
		Модель ТТЛ	Модель делителя напряжения
40	1 байт	12,58	28,84
	2 байта	8,50	22,75
	4 байта	5,10	13,65
60	1 байт	18,59	39,44
	2 байта	12,74	33,45
	4 байта	7,46	20,07
80	1 байт	24,46	48,69
	2 байта	16,96	44,18
	4 байта	10,17	26,51
100	1 байт	30,24	56,83
	2 байта	21,18	54,88
	4 байта	12,71	32,93

Заключение

Рассмотрена проблема уменьшения объема данных дискретного представления ОР технических систем. Наибольший объем данных в таком представлении приходится на множество индикаторов принадлежности элементов ОР. Исходя из специфики этих данных, в них присутствует большое количество серий из одинаковых элементов, что позволяет применить алгоритм кодирования длин серий (RLE), который также позволяет считывать исходные данные без полной распаковки всего массива за один проход. Экспериментально на двух моделях технических систем показано, что наибольшая эффективность сжатия данных алгоритмом RLE достигается при кодировании длины каждой серии повторяющихся элементов одним байтом, несмотря на наличие существенно более длинных серий, которые приходится разбивать на фрагменты по 255 элементов.

Литература

1. Назаров Д.А. Использование областей работоспособности для оптимального выбора номиналов параметров // Информатика и системы управления. – 2011. – № 2(28). – С. 59 – 69.
2. Абрамов О.В., Назаров Д.А. Программно-алгоритмический комплекс построения, анализа и использования областей работоспособности // Информационные технологии и вычислительные системы. – 2015. – №2. С. 3 – 13
3. Salomon D. Data Compression: the complete reference. Vol. 10. – Springer, 2007.
4. Норенков И.П. Основы автоматизированного проектирования / И.П. Норенков. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2002. – 336 с
5. Катуева Я.В., Назаров Д.А. Аппроксимация и построение областей работоспособности в задаче параметрического синтеза // Труды международного симпозиума «Надежность и качество». – Пенза: ПГУ. – 2005. – 1 т. С. 130 – 134
6. Саушев А.В. Области работоспособности электротехнических систем. – СПб.: Политехника, 2013. – 414 с.: 68 ил.

МОДЕЛИРОВАНИЕ ПРОЦЕССОВ ЗАЩИТЫ ПЕРЕДАЧИ ТЕХНОЛОГИЧЕСКОЙ ИНФОРМАЦИИ ПО РЕЗЕРВНОМУ КАНАЛУ НА ОСНОВЕ АНАЛИЗА СЕТЕВОГО ТРАФИКА

Абрамова Т.В., Аралбаев Т.З., Манжосов А.В.
ФГБОУ ВО «Оренбургский государственный университет»
+7 (922) 856-51-22, taya357@gmail.com

В работе представлены результаты моделирования процессов защиты технологической информации в условиях воздействия пространственно-временных угроз на основе мониторинга сетевых информационных потоков. Приводятся результаты разработки ряда моделей, предназначенных для