



Math-Net.Ru

Общероссийский математический портал

А. Д. Витол, Н. А. Жукова, А. Б. Тристанов, Алгоритм сегментации кусочно-постоянных телеметрических параметров, *Вестник КРАУНЦ. Физ.-мат. науки*, 2010, выпуск 1(1), 46–53

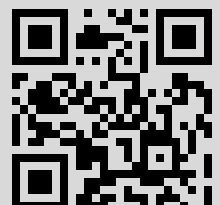
DOI: <http://dx.doi.org/10.18454/2079-6641-2010-1-1-46-53>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 77.82.199.43

20 июля 2016 г., 04:27:15



DOI: 10.18454/2079-6641-2010-1-1-46-53

Информационные и вычислительные технологии
УДК 519.688

АЛГОРИТМ СЕГМЕНТАЦИИ КУСОЧНО-ПОСТОЯННЫХ ТЕЛЕМЕТРИЧЕСКИХ ПАРАМЕТРОВ

А.Д. Витол¹, Н.А. Жукова¹, А.Б. Тристанов²

¹ Санкт-Петербургский государственный электротехнический университет имени В.И. Ульянова (Ленина), 197376, г. Санкт-Петербург, ул. проф. Попова, 5

² Институт космофизических исследований и распространения радиоволн ДВО РАН, 684034, Камчатский край, с. Паратунка, ул. Мирная, 7

E-mail: alextristanov@mail.ru

В статье рассматривается принцип модификации алгоритма оптимального разбиения и его применение в обработке телеметрических сигналов.

Ключевые слова: интеллектуальный анализ временных рядов, сегментация, телеметрия

© Витол А.Д., Жукова Н.А., Тристанов А.Б., 2010

Informational and computational technologies
MSC 62N86

THE ALGORITHM OF PIECEWISE CONSTANT TELEMETRIC PARAMETERS SEGMENTATION

A.D. Vitol¹, N.A. Zhukova¹, A.B. Tristanov¹

¹ Saint Petersburg Electrotechnical University «LETI», 197376, St.-Petersburg, Popova st., 5, Russia

² Institute of Cosmophysical Researches and Radio Wave Propagation Far-Eastern Branch, Russian Academy of Sciences, 684034, Kamchatskiy Kray, Paratunka, Mirnaya st., 7, Russia

E-mail: alextristanov@mail.ru

This paper is devoted to the modification of optimal segmentation algorithm and using it to telemetric signals processing.

Key words: time series data mining, segmentation, telemetry

© Vitol A.D., Zhukova N.A., Tristanov A.B., 2010

Введение

В процессе испытаний (эксплуатации) сложных динамических объектов контроль и оценка состояния объекта осуществляются на основе анализа передаваемой с объекта телеметрической информации. Передаваемые данные содержат информацию, полученную от датчиков, установленных на объекте, и представляют собой структурированный бинарный поток, получивший название группового телеметрического сигнала (ГТС) [1]. Извлечение телеметрических параметров из бинарного потока (декоммутиация параметров) осуществляется в соответствии с формуляром, содержащим детальное описание структуры потока. Декоммутированные параметры группируются в соответствии с их типом, при этом, как правило, используется следующее структурное деление: константные параметры, счетчики, контактно-кодовые параметры, меандры, функциональные параметры.

Необходимость группировки определяется использованием различных технологий для обработки каждого типа параметров. Функциональные параметры представляют собой информационно-ценные параметры, отражающие поведение отдельных систем, по результатам анализа которых экспертом принимается решение о состоянии сложного динамического объекта (СДО). Для применения современных средств анализа данных к функциональным параметрам необходимо представить параметры в виде последовательности стационарных процессов (сегментов), каждый из которых может быть описан набором характеристик.

Функциональные параметры представляют собой последовательность дискретных измерений $\{x_n\} : n = 1, 2, \dots, N$, где N – общее количество измерений, которым поставлена в соответствие упорядоченная последовательность временных отсчетов $\min(I) \leq t_1 \leq t_2 < \dots < t_N \leq \max(I)$, где I – интервал измерения. Удобно представить данные в виде пар (ячеек) $C_n \equiv (x_n, t_n)$. Задача сегментации подразумевает разбиение интервала I на некоторое количество сегментов: $P(I) \equiv \langle N_{blocks}, \{n_k\}, k = 1, 2, 3, \dots, N_{blocks} \rangle$, где N_{blocks} – количество сегментов, $n_k = \{(x_j, t_j)\}_{j=block_k}^{block_{k+1}}$ – сегменты.

Анализ существующих подходов к сегментации временных рядов

В работе [2] приведен детальный анализ существующих алгоритмов сегментации кусочно-постоянных временных рядов и предложен алгоритм оптимального разбиения данных (Optimal partition of the Interval). Исследование эффективности применения алгоритма оптимального разбиения для обработки телеметрических параметров проводилось на группе из 20 быстроменяющихся параметров.

В качестве основных критериев оценки качества работы алгоритма были выбраны два критерия: во-первых, правильность разбиения; во-вторых – время обработки параметра. Оценка правильности расстановки сегментов производилась на основе экспертного заключения. Алгоритм был реализован в среде MatLab. Исследования проводились на машине с процессором Intel Core 2 Duo с частотой ядра 3,0 ГГц.

На рис. 1 представлены примеры из группы обрабатываемых параметров и результаты работы алгоритма оптимального разбиения. Моменты смены значения кривой черного цвета соответствуют началу новых сегментов.

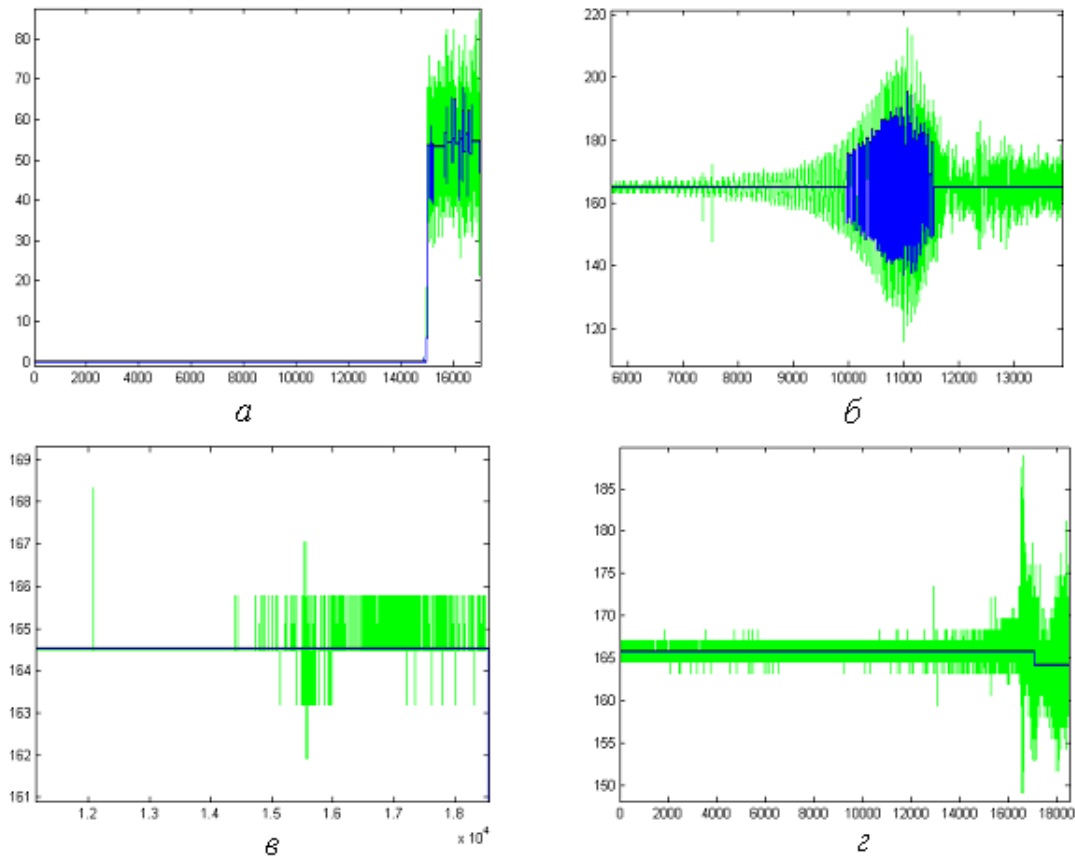


Рис. 1. Примеры результатов обработки параметров

Для первого и второго параметров (рис. 1, а, б) видно, что расставлены избыточные границы сегментов. Для третьего параметра (рис. 1, в) не определено ни одной границы, что не соответствует поведению параметра. При определении границ для четвертого параметра (рис. 1, г) правильные границы пропущены. Время обработки составило соответственно 37,563; 44,483; 27,359 и 51,766 с, что пропорционально длине параметров.

Проведенный анализ выявил необходимость модификации алгоритма оптимального разбиения для обработки быстроменяющихся телеметрических параметров.

Модификация алгоритма оптимального разбиения

Алгоритм оптимального разбиения позволяет сформировать модель временного ряда в виде последовательности блоков, при этом любой параметр модели постоянен в пределах одного блока, но испытывает резкие скачки на точках переходов, отмечая края блока. Блоки формируются из ячеек данных, каждый блок представляет собой набор из соседних ячеек. На рис. 2, б показана последовательность из 32 ячеек данных, разделенная на пять блоков. Для блоков будем использовать обозначение

$$B(n, m) \equiv \{C_n, C_{n+1}, \dots, C_m\},$$

приняв, что блок содержит $m - n + 1$ ячейку, начиная с n -й. Случай $m = n$ соответствует блоку, состоящему из одной ячейки, как в последнем блоке на рис. 2, б.

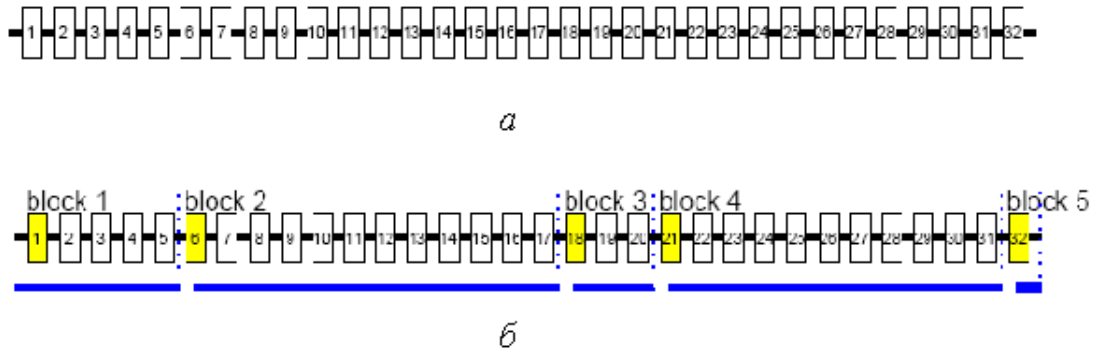


Рис. 2. Схематичное представление модели временного ряда

Разбиение интервала I – это набор непересекающихся блоков, которые вместе составляют весь интервал I . Разбиение может быть описано количеством блоков (элементов разбиения) N_{blocks} и набором граничных ячеек n_k :

$$P(I) \equiv \{N_{blocks}, n_k, k = 1, 2, 3, \dots, N_{blocks}\}.$$

Количество точек перехода на одну меньше, чем количество блоков, так как по соглашению первый блок начинается с первой ячейки данных – n_1 . Для получения оптимального разбиения осуществляется поиск такого разбиения среди всех возможных, которое максимизирует заданную функцию. Общее количество вариантов разбиения $N_{partitions} = 2^{N-1}$. Алгоритм позволяет найти оптимальное разбиение на этом пространстве за время, пропорциональное N^2 .

В качестве модели интервала берутся все разбиения интервала с известной статистической моделью для каждого из блоков разбиения. Если погрешности наблюдений в различное время являются независимыми, то справедливо соотношение

$$F[P(I)] = \sum_{k=1}^{N_{blocks}} f(B_k),$$

где $F[P(I)]$ – степень соответствия формируемой модели исходным данным на всем интервале; $f(B_k)$ – степень соответствия для k -го блока. Лучшая модель находится путем нахождения разбиения интервала, при котором F максимальна.

Временная отметка, разделяющая два блока, называется точкой перехода. Предполагается, что перемещение точки перехода, лежащей между двумя ячейками данных, на новое место между теми же ячейками, существенно не изменяет представления модели данных. Это упрощение снижает поиск в бесконечном пространстве до конечной задачи оптимизации. Принимается также, что точка перехода – первая ячейка в последующем блоке (а не последняя ячейка предыдущего блока). Соответственно, учитывая, что самое маленькое разбиение состоит из одного блока, содержащего все ячейки данных, первая ячейка данных – всегда точка перехода. Если последняя ячейка является точкой перехода, она разграничивает блок, состоящий из одной ячейки, как и на рис. 2, б.

Рассмотрим основные этапы модифицированного алгоритма оптимального разбиения (рис. 3).

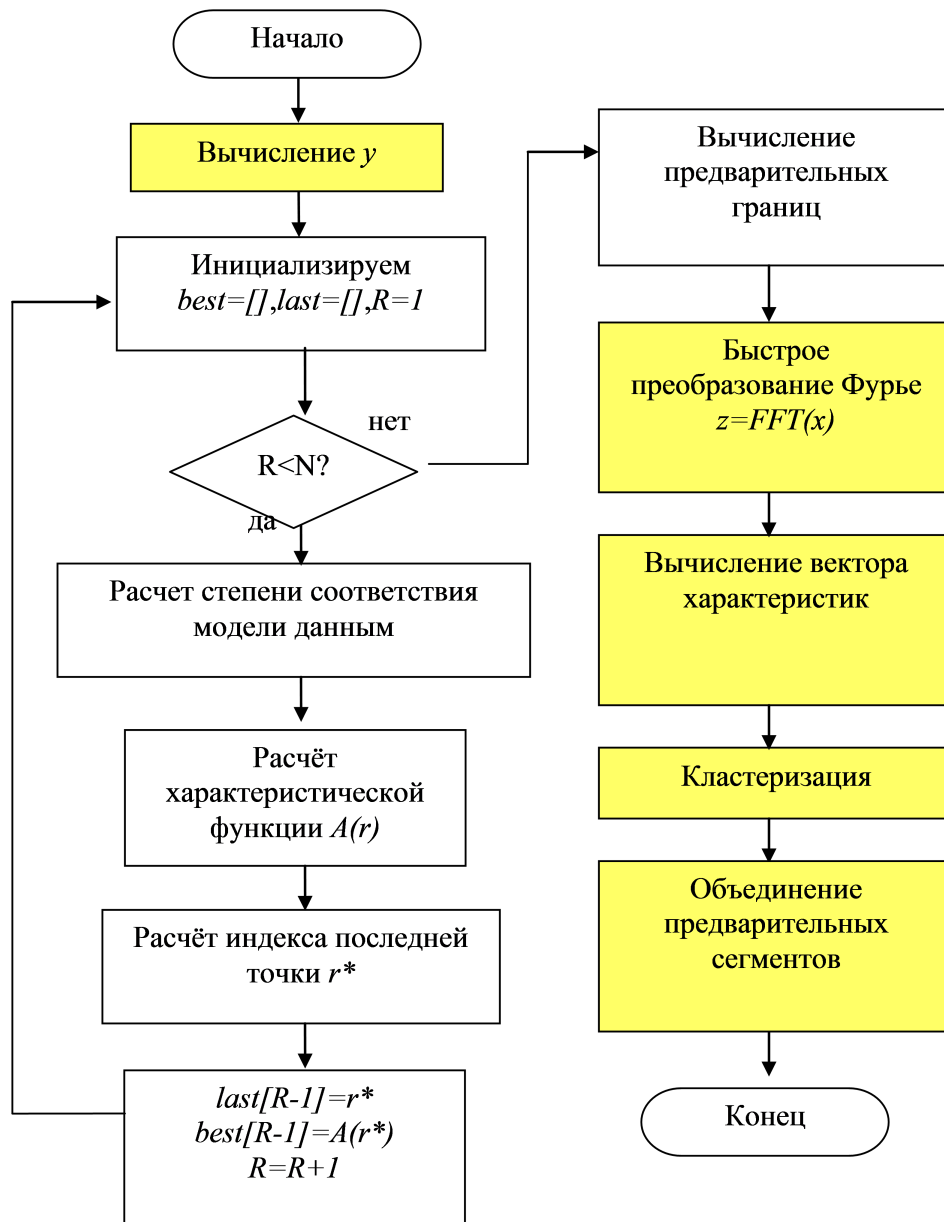


Рис. 3. Блок-схема алгоритма оптимального разбиения

1. *Предварительная обработка.* На стадии предобработки исходный сигнал преобразуется следующим образом: от сигнала вычисляется длина кривой для каждой точки сигнала и от нее берется производная. Таким образом, формируется массив y обрабатываемых данных.

2. *Определение предварительных границ блоков с применением алгоритма оптимального разбиения.* Алгоритм начинает обработку данных с первой ячейки, добавляя по одной на каждом шаге, пока весь интервал не будет обработан.

На шаге R алгоритм находит оптимальное разбиение отрезка, составленного из ячеек данных $I_R \equiv \{C_1, C_2, \dots, C_R\}$. Чтобы проанализировать все данные, принимается $R = 1, 2, \dots, N$. Случай $R = 1$ тривиален: есть только одна ячейка и единственный возможный способ разбиения, который является оптимальным.

Рис. 4. Сегменты, полученные в результате работы алгоритма

Для их устранения применяется алгоритм кластерного анализа. Учитывая специфику сигналов, для формирования описания сегментов целесообразно использовать быстрое преобразование Фурье. При этом в качестве альтернативы может быть использовано вейвлет-преобразование.

Вектор характеристик каждого сегмента составляется из вариации $V(x) = \max_i (|x_i - \bar{x}|)_i$, энергии $E(x) = \sum_i x_i^2$ и мощности $M(x) = \sum_i x_i^2 / \max(i)$ – вещественной части преобразованного сигнала.

Объединение сегментов осуществляется с использованием алгоритмов кластерного анализа [3], в частности алгоритма EM с автоматическим определением числа кластеров. На основе результатов кластерного анализа осуществляется объединение последовательно идущих сегментов, попавших в один кластер:

$$c'_i = \{c_i \cup c_j\}; |i - j| = 1, c_i, c_j \in q_l,$$

где q_l – l -й кластер.

Исследование модификации алгоритма оптимального разбиения

Оценка модификации алгоритма проводилась на тех же параметрах и по тем же критериям, по которым оценивалась работа оригинального алгоритма. Алгоритм был реализован в среде MatLab. Исследования проводились на машине с процессором Intel Core 2 Duo с частотой ядра 3,0 ГГц. Оценка правильности расстановки сегментов производилась на основе экспертного заключения. На рис. 5 представлены результаты обработки тех же параметров, что и на рис. 1.

Время обработки составило соответственно 43,359; 50,480; 33,311; 57,446 с, что пропорционально длине параметров. Время обработки возросло в среднем на 15 %. На первом, втором и четвертом параметрах модифицированный алгоритм расставил границы правильно. На третьем параметре расстановка границ, по мнению эксперта, могла быть и другой, но основные процессы выделены верно.

Библиографический список

1. НАЗАРОВ А.В. Современная телеметрия в теории и на практике / А.В. Назаров, Г.И. Козырев, И.В. Шитов и др. СПб.: Наука и техника, 2007.
2. SCARGLE J.D. AN ALGORITHM for optimal partitioning of data on an interval // Signal Processing Letters. IEEE. 2005. P. 105–108.

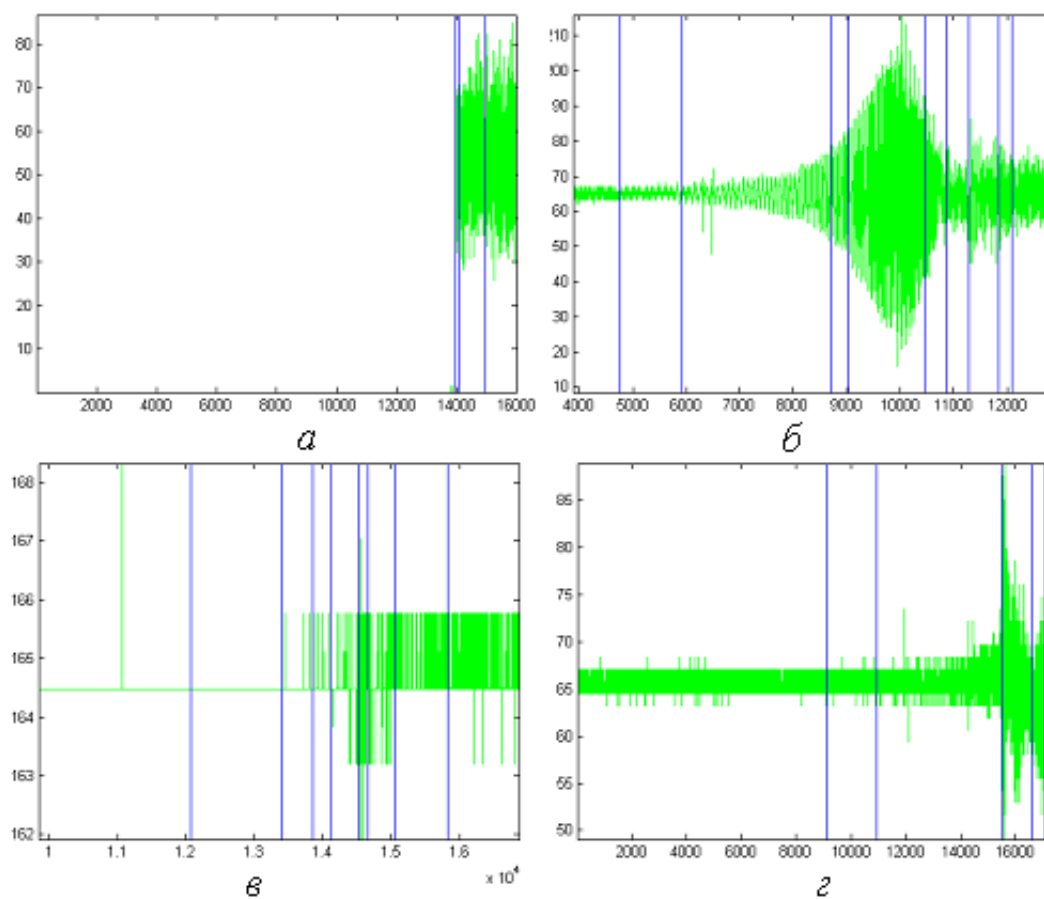


Рис. 5. Пример использования модифицированного алгоритма

3. HAN J., KAMBER M. Data Mining: concepts and techniques. Morgan Kaufman Publishers, 2005.

Поступила в редакцию / Original article submitted: 11.10.10