МАТЕРИАЛ 6. БАЗОВЫЕ АЛГОРИТМЫ ОБРАБОТКИ ДАННЫХ В МИКРОПРОЦЕССОРНЫХ СИСТЕМАХ.

Цель - изучить базовые алгоритмы обработки данных и программирование их в интегральной среде разработки Keil Elektronik GmbH.

СБОР И ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ.

1. АЛГОРИТМ ЛИНЕЙНОГО УСРЕДНЕНИЯ.

При обработке стационарных эргодических фрагментов случайных сигналов за оценку искомой вероятностной характеристики принимают **среднее значение** реализации соответствующего фрагмента случайного сигнала при достаточном интервале наблюдения.

Алгоритм вычисления среднего значения зависит от принятого математического выражения оценки. Широкий класс составляют **линейные несмещённые оценки**, представляющие собой линейные функции дискретных выборок исследуемого случайного процесса. Наиболее распространена принадлежащая к указанному классу оценка, получаемая как **среднее арифметическое значение**, которое формируется из дискретных выборок реализации исследуемого процесса. Известно, что такой линейный алгоритм статистического усреднения представляется следующим математическим выражением:

$$S_k = \frac{1}{k}\sum_{i=1}^{k} S_i \quad (1)$$

,

где S_k - оценка измеряемой характеристики, k - число измерений, S_i - текущее мгновенное значение измеряемой характеристики.

Формулу (1) представим в виде следующего рекуррентного соотношения:

$$S_k = S_{k-1} + \frac{1}{k}(S_k - S_{k-1}) \quad (2)$$

По формуле (2) получается **оценка текущего мгновенного значения среднего** для выбранного временного окна как функция длины окна. Позволяет оценить интервал наблюдения сигнала.

2. ЭКСПОНЕЦИАЛЬНЫЙ АЛГОРИТМ УСРЕДНЕНИЯ.

Линейный алгоритм (2) можно преобразовать к **рекуррентному виду**, который в своей записи представляет **экспоненциальный алгоритм усреднения** с заданным множителем. В общем случае получается смещённая и несостоятельная оценка среднего в виде:

$$S_i = S_{i-1} + \frac{1}{b}(S_i - S_{i-1}), b = const, i = 0,1,...k \quad (3)$$

Следует отметить, что несостоятельность оценки среднего, определяемой по экспоненциальному алгоритму (3), не препятствует её практическому применению, так как при проведении измерений всегда задаётся требуемая точность получения результатов обработки. При постоянном множителе b>1 оценка среднего (3) в общем случае становится асимптотически несмещённой. В качестве величины b выбирается степень 2, что позволяет обойтись лишь сдвигами.

При практической реализации алгоритмов усреднения операция деления на произвольные k и b вызывает дополнительные затраты оборудования и времени. В специализированных цифровых устройствах k и b выбирают кратными 2, т.е. $k=2^p$, либо $b=2^p$, p=0,1,2,…, что позволяет свести операцию деления к операции сдвига числа на p разрядов. Но в этом случае линейный алгоритм усреднения позволяет наблюдать оценку измеряемой характеристики только дискретно при $k=2^p$, что неприменимо в некоторых физических задачах.

3. АЛГОРИТМ КВАЗИЛИНЕЙНОГО УСРЕДНЕНИЯ.

Устраняет указанные недостатки линейного и экспоненциального алгоритмов комбинированный или **квазилинейный алгоритм усреднения**. Для такого алгоритма параметр $b=2^p$ в (3) изменяется дискретно и при достижении текущим шагом i некоторых значений k остается постоянным в промежутках между этими шагами. Квазилинейный алгоритм позволяет получить оценку на каждом шаге усреднения. При этом, если первую выборку усредняемого процесса делить на 1, то получаемые оценки будут несмещенные.

Оптимальная функция $b=b(i)$, обеспечивающая минимальную дисперсию оценки при дополнительном условии $b(i) \in \left\{2^p\right\}, p = 0,1,2...$, для квазилинейного алгоритма

$$S_i = S_{i-1} + \frac{1}{b(i)}(S_i - S_{i-1}) \quad (4)$$

приведена в таблице:. (s - значение ***i***, после которого удваивается ***b***)

i	1	2	3-5	6-10	11-22	23-44	45-88	89-177	178-354	355-709	710-1418	1419-2837
b	1	2	4	8	16	32	64	128	256	512	1024	2048
s	0	1	2	5	10	22	44	88	177	354	709	1418

4. ИЗМЕРЕНИЕ МОМЕНТОВ.

Несмещенные оценки четырех центральных моментов плотности распределения вероятности реализации случайного процесса $x(iT), i = 0,1,...N-1$, T - интервал дискретизации) имеют следующий вид:

Среднее значение

$$M = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{5}$$

Дисперсия

$$D = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - M)^2 \tag{6}$$

Среднеквадратическое отклонение

$$\sigma = \sqrt{D} \tag{7}$$

Коэффициент асимметрии

$$a = \frac{1}{N-2}\sqrt{N(N-1)}\,\frac{1}{N\sigma^2}\sum_{i=1}^{N} (x_i - M)^3 \tag{8}$$

Коэффициент эксцесса

$$E = \frac{1}{(N-2)(N-3)}\left[(N+1)\frac{1}{N\sigma^4}\sum_{i=1}^{N} (x_i - M)^4 - 3N + 3\right] \tag{9}$$

Первые моменты (среднее значение и дисперсия) непрерывно оцениваются по мере поступления отсчётов x_i измеряемого процесса длинной N. Таким образом на каждой реализации прослеживается изменение среднего значения и дисперсии.

Оценки текущих значений первого и второго моментов имеют вид

$$m_i = m_{i-1} + \frac{x_i - m_{i-1}}{i}, m_0 = 0, \tag{10}$$

$$D_i = \frac{i-2}{i-1} D_{i-1} + \frac{\left(x_i - m_{i-1}\right)^2}{i}, D_0 = 0, i = 1,2,...N. \tag{11}$$

Используя квазилинейный алгоритм статистического усреднения для реализации этих формул, окончательно получаем:

$$m_i = m_{i-1} + \frac{x_i - m_{i-1}}{b(i)}, m_0 = 0, \tag{12}$$

$$D_i = k(i) D_{i-1} + \frac{\left(x_i - m_{i-1}\right)^2}{b(i)}, D_0 = 0, i = 1,2,...N, \tag{13}$$

где $k(i)$ - массив поправочных коэффициентов $\frac{i-2}{i-1}, i = 2,3,...$ (0.5; 0.667; 0.75; 0.8; 0.833; 0.867; 0.875; 0.889; 0.9; 0.909; 0.916; 0.923; 0.929; 0.933; 0.9375; 0.941; 0.944; 0.947). При таком выборе $k(i)$ ошибка дисперсии не превышает 5%.

5. ПОДГОНКА И УДАЛЕНИЕ СРЕДНЕГО.

Подгонка и удаление среднего - это самый частный случай полиномиальной фильтрации и удаления полиномиального тренда. Само вычисление простое: вычисляется среднее по формуле (1), полученное значение вычитается из отсчётов исходной последовательности.

Важно. При большой длине последовательности происходит потеря значащих разрядов при вычислении среднего. Последовательность следует разбить на небольшие фрагменты; суммировать частями, не прибегая к многозначной арифметике. Самая лучшая точность получается при суммировании соседних пар отсчётов, итерационно, последовательность каждый раз сокращается в два раза пока не получится одно значение результирующей суммы.

6. ПЕРЕХОД К ФИЗИЧЕСКИМ ЕДИНИЦАМ

Здесь рассмотрим шаговую и синусоидальную калибровки для процедур перевода цифровых отсчётов в физические единицы. В обоих случаях предлагается перед записью данных исключить из системы датчик. Это связано с тем, что датчик, как правило, не может обеспечить эталонной точности.

6.1. Шаговая калибровка.

При шаговой калибровке используется набор уровней напряжений, называемых шагами. Например, для датчика, напряжения которого изменяются в диапазоне от 0 до 5 В, такими уровнями могут служить 0.0, 2.5 и 5.0 В. Напряжение каждого уровня подается в течение установленного промежутка времени, равного, например, одной секунде, и преобразуется в цифровую форму. Для удобства последующих вычислений результаты следует располагать так, чтобы их нельзя было спутать.

(1). Сначала вычисляются средние значения результатов, полученных для каждого уровня. Получающиеся по одному для каждого уровня средние числа обозначим $\boldsymbol{a_1, a_2, \dots, a_N}$ (предполагается, что всего используется N шагов).

(2). После этого вводятся величины в физических единицах для каждого уровня, по одной для каждого числа $\boldsymbol{a_k}$. Эти величины обозначим $\boldsymbol{p_1, p_2, \dots, p_N}$.

(3). Допустим теперь, что в систему вводятся данные для обработки, которые поступают с АЦП в отсчётах. Они составляют последовательность $\boldsymbol{c(i)}$. С учетом всей полученной выше информации эта последовательность преобразуется в последовательность калиброванных данных $\boldsymbol{x(i)}$.

Для этой цели обычно используется линейная интерполяция, сводящаяся к следующим действиям. Сначала определяется такое K, что $\boldsymbol{a_K <= c(i) < a_{K+1}}$, а затем по формуле

$$x(i) = p_K + (p_{K+1} - p_K)\frac{c(i) - a_K}{a_{K+1} - a_K} \qquad (14)$$

получают соответствующее $\boldsymbol{c(i)}$ калиброванное значение. При этом для крайних точек вычисления производят по таким же формулам, полагая K=1, если $\boldsymbol{c(i) < a_1}$, и K=N-1, если $\boldsymbol{c(i) > a_N}$. В последних случаях интерполяционная формула будет выступать в качестве экстраполяционной.

6.2. Синусоидальная калибровка.

(1). При синусоидальной калибровке в систему вводится синусоидальный сигнал с известной амплитудой и на большом числе периодов производится запись и преобразование в цифровую форму. При этом определяется среднее значение по формуле (5) $\bar{d}$ последовательности $\boldsymbol{d(i)}$ чисел, полученных при преобразовании синусоиды в цифровую форму. Помимо этого, вычисляется выборочная дисперсия по приводившейся ранее формуле (6) s_d^2.

(2). По этим величинам определяют три точки:

$$a_3 = \bar{d} + s_d\sqrt{2}\ ,\ a_2 = \bar{d}\ ,\ a_3 = \bar{d} - s_d\sqrt{2}\ , \qquad (15)$$

которые отвечают положительному пику, нулю и отрицательному пику синусоиды соответственно (все эти величины измерены в отсчётах). Затем вводятся соответствующие физические единицы и применяются приведенные выше формулы для интерполяции и экстраполяции.

Если бы не было шума и использовалось целое число периодов синусоиды, то эта процедура была бы точной. Однако избежать шума невозможно. Нельзя также начать и остановиться в одной и той же точке периода синусоиды. Поэтому для того, чтобы не получить ошибок или по крайней мере свести их к минимуму, берётся большое число периодов.

7. РЕДАКТИРОВАНИЕ ШУМА

Большинство систем сбора информации вносит в данные ложные значения. Это может происходить по многим причинам - … . Неправдоподобные значения, возникающие в результате этих сбоев, могут вызвать значительные трудности при последующем анализе. Одно неправдоподобное значение, равное максимально представимой в АЦП величине, может повысить предельный уровень шума и, значит, привести к значительному изменению плотности спектра мощности. Близко расположенная пара таких значений может дать несколько ложных частот в ПСМ. По этим причинам лучше всего с помощью схемы предварительного преобразования предельных данных организовать поиск и удаление неправдоподобных значений. К сожалению, трудно определить точно, какие данные неправдоподобны. Общих процедур для автоматического удаления таких данных нет.

Все машинные программы, предназначенные для этих целей, действуют следующим образом. Сначала для поиска неправдоподобных значений просматриваются все данные. Если такие значения обнаружены, то они выводятся на печать в виде таблицы или графика. После просмотра этих распечаток исследователь вводит очередную программу, чтобы удалить плохие значения, заменить их или вовсе не использовать. Последнее лучше всего, **если, правда, это можно сделать.**

Для устранения неправдоподобных значений предложено несколько схем автоматического редактирования, но ни одна из них не является полностью удовлетворительной. Рассмотрим две такие схемы, чтобы пояснить, какого рода операции в этом случае производятся.

7.1. Общий подход.

В первой из этих схем, представленной на рис. 1, приведена общая идея.

(1). Используются два цифровых RC-фильтра для сглаживания. Они производят низкополосную фильтрацию исходной функции и вырабатывают сглаженные оценки. При рассмотрении обеих схем неявно предполагается, что нужные данные имеют «плавный» характер, а неправдоподобные значения-«резкий». Определяются две последовательности $[\bar{x}(i)]^2$ и $\overline{x^2(i)}$ как последовательности сглаженных данных с последующим возведением их в квадраты и сглаженных данных, предварительно возведенных в квадраты, соответственно.

Задача той части схемы, которая на рисунке ограничена пунктирной линией, состоит в выработке постоянно обновляемого значения выборочной дисперсии $s^2(i)$. Это значение определяется по формуле

$$s^2(i) = \overline{x^2(i)} - [\bar{x}(i)]^2 . \qquad (16)$$

(2). Извлекая из него квадратный корень, получают стандартное отклонение.

(3). На следующем шаге проверяют очередное значение ряда $\boldsymbol{x(i+1)}$. Это значение считается хорошим, если

$$\bar{x}(i) - ks(i) < x(i+1) < \bar{x}(i) + ks(i) . \qquad (17)$$

Подходящий для данных параметр ***k***. выбирается самим исследователем. Как правило, он колеблется в пределах от 3 до 9, но для начала лучше выбирать его равным 6.

(4). Плохие значения ***x(i+1)*** могут заменяться значениями оценки, вычисленными по формуле

$$\hat{x}(i+1) = 2\,x(i) - x(i-1)\,, \tag{18}$$

т. е. полученными, по сути дела, линейной экстраполяцией.

Схема редактора шума

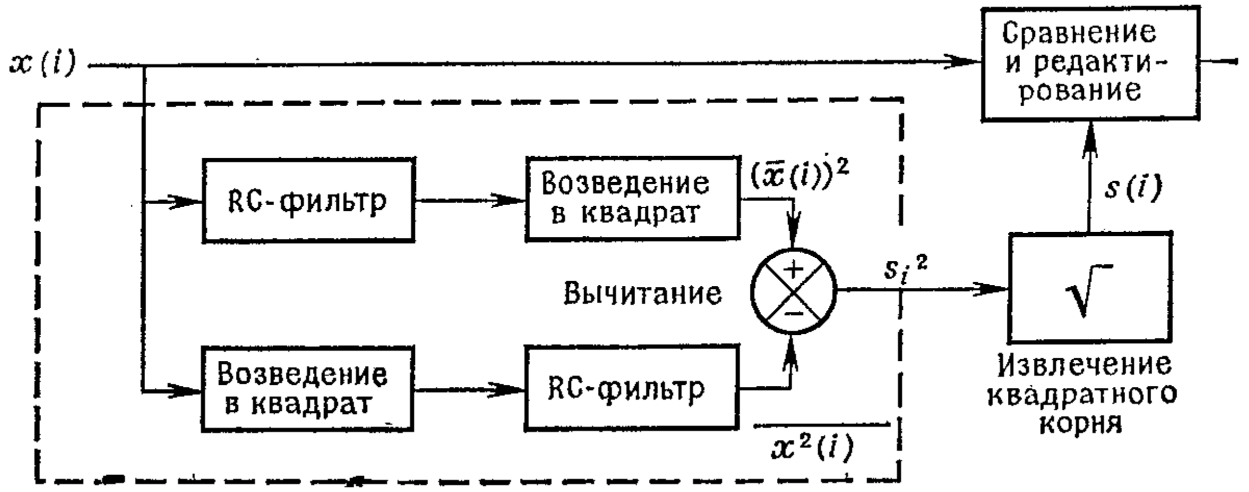

Схема редактирования данных с неправдоподобными значениями.

Рис. 1

Эта процедура требует некоторых дополнительных уточнений. Для числа последовательных экстраполяций следует с самого начала установить предел, чтобы избежать постоянной экстраполяции. Дело в том, что последовательная экстраполяция для ряда плохих точек может достаточно далеко увести полосу (11), и в результате в неё не попадут те хорошие данные, которые в конце концов появятся.

7.2. Алгоритм Тьюки.

Вторую схему редактирования данных с неправдоподобными значениями, частично принадлежащую Дж. В. Тьюки, называют процедурой **«Тьюки 53Х».** Основная идея этой процедуры состоит в том, чтобы получать оценку гладкого куска кривой, которую затем вычитают из данных; после удаления гладкой составляющей или тренда выделить неправдоподобные значения гораздо легче.

В этой процедуре используется тот факт, что медиана есть робастная оценка среднего. Если все данные упорядочить в порядке возрастания, то медианой будет просто то значение, которое в этом ряде данных займет место посередине. При вычислении оценки гладкой составляющей в процедуре Тьюки 53Х это свойство робастности медианы используется дважды. Сама процедура выглядит следующим образом:

(1). По последовательности $\boldsymbol{x(i)}$ строится новая последовательность $\boldsymbol{x^1(i)}$. Сначала определяется медиана данных $\boldsymbol{x(1)}, \ldots, \boldsymbol{x(5)}$. Это значение становится в новой последовательности элементом $\boldsymbol{x^1(3)}$. Теперь из пяти первых данных убирается $\boldsymbol{x(1)}$ и добавляется $\boldsymbol{x(6)}$. Медиана этого нового множества есть $\boldsymbol{x^1(4)}$. Так продолжается скользящая медианная фильтрация до тех пор, пока все данные не будут исчерпаны. По сравнению с последовательностью $\boldsymbol{x(i)}$ новая последовательность будет короче на четыре элемента. Медиана всегда выбирается из группы пяти последовательных значений.

(2). Почти таким же способом по последовательности $\boldsymbol{x^1(i)}$ строится последовательность $\boldsymbol{x^2(i)}$. Единственное различие в том, что теперь медиана выбирается из троек последовательных значений.

(3). На последнем этапе строится последовательность $\boldsymbol{x^3(i)}$ по правилу

$$x^3(i) = \tfrac{1}{4}\,x^2(i-1) + \tfrac{1}{2}\,x^2(i) + \tfrac{1}{4}\,x^2(i+1)\,, \tag{19}$$

отвечающему сглаживающему фильтру Хэннинга.

(4). После этого, как и раньше, для элементов последовательности $\boldsymbol{x(i) - x^3(i)}$ проверяется условие $|\boldsymbol{x(i) - x^3(i)}| >$ **k,** где **k** - заранее выбранное число. Если оно выполняется, то соответствующую величину заменяют интерполированным значением.

8. УДАЛЕНИЕ ТРЕНДА (ПОЛИНОМИАЛЬНОГО)

Иногда из некоторых временных рядов нужно удалить линейный или медленно меняющийся тренд. Такого рода тренды наблюдаются в рядах, например, при суммировании одной или нескольких компонент, приводящем к ошибкам двух типов. Во-первых, при неправильной калибровке нулевой точки в каждый момент отбора данных будет возникать небольшая ошибка. После суммирования эта постоянная величина даст прямую линию. Такой линейный тренд может привести к большим ошибкам при определении плотности спектра мощности и в связанных с этим вычислениях.

Ошибка второго типа возникает из-за возрастания в процессе суммирования мощности, соответствующей низкочастотному шуму. Как правило, такой шум в данных всегда есть. При суммировании он обретает форму случайного, но медленно меняющегося тренда. Насколько быстро меняется такой тренд, до некоторой степени зависит от интервала квантования.

Одним из способов удаления тренда служит применение высокочастотных фильтров.

Полиномиальный тренд можно удалять методом наименьших квадратов. Это основывается на методе подгонки. Подгонка к $\boldsymbol{x(n\ \Delta t)}$ степенного многочлена с помощью метода наименьших квадратов - это параметрический метод определения детерминированной составляющей $\boldsymbol{a(n\ \Delta t)}$ в нестационарном процессе. В этом случае исходный ряд выборок реализации сигнала приближается многочленом степени K:

$$\hat{x}(n\Delta t) = \sum_{i=0}^{K} c_i\ (n\Delta t)^i \quad , \tag{20}$$

где n=0,1,…,(N-1); N - длина реализации в отсчётах. Для метода наименьших квадратов подгонка осуществляется путём минимизации квадрата разностей исходной последовательности и значений получаемого многочлена:

$$Q = \sum_{n=0}^{N-1} \left(x(n\ \Delta t) - \hat{x}(n\ \Delta t) \right)^2 = \sum_{n=0}^{N-1} \left(x(n\ \Delta t) - \sum_{i=0}^{K} c_i\ (n\ \Delta t)^i \right)^2 \quad . \tag{21}$$

Искомая последовательность коэффициентов $\mathbf{c_i}$ (i = [0,1,…,K]) получается путём приравнивания нулю частных производных функции Q по переменным $\mathbf{c_i}$ в выражении (21). Последнее даёт систему из K+1 линейных алгебраических уравнений, которая легко решается численными методами и аналитически.

Например, таким решением в случае **K=0** будет коэффициент $\mathbf{c_0}$, вычисляемый как обычное среднее значение.

В случае **K=1** приходим к задаче линейного регрессионного анализа и решение дают формулы:

$$c_0 = \frac{2(2N-1)\sum_{i=0}^{N-1} x(i) - 6\sum_{i=0}^{N-1} i\,x(i)}{N(N+1)} \, , \tag{22}$$

$$c_1 = \frac{12\left(\sum_{i=0}^{N-1} i\,x(i) - ((N-1)/2)\sum_{i=0}^{N-1} x(i) \right)}{N(N^2-1)\Delta t} \, . \tag{23}$$

Вычисления значительно упрощаются, если N - нечетное число. Разумеется, и при чётном N последнюю точку можно удалить. Оказывается, что при нечетном числе точек и интервале (-N/2, N/2) изменения независимой переменной **i** многие члены обращаются при вычислениях в нуль, в частности все те, которые имеют противоположные значения. Это сокращает число необходимых вычислений и с арифметической точки зрения улучшает их.

ЛАБОРАТОРНАЯ РАБОТА 5. «ИЗУЧЕНИЕ ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ ПРОГРАММИРОВАНИЯ MCS-51 НА ЯЗЫКЕ СИ (Keil Elektronik), ОБРАБОТКА ДАННЫХ»

Часть 1.

1). Написать и отладить программу, выполняющую различные варианты усреднения измерительных данных. Поварьировать моделями памяти.

2). Написать и отладить программу, вычисляющую основные моменты измерительных данных. 1-ый и 2-ой моменты получить текущие значения в зависимости от интервала наблюдения.

3). Написать и отладить программу, выполняющую калибровку измерительных данных.

***** Измерительные данные по этим пунктам скомпоновать в программу отдельно, из файла для данных.

Часть 2.

1). Написать и отладить программу редактора шума.

2). Написать и отладить программу удаления постоянного и линейного тренда из измерительных данных.