Компьютерная
Математика

УДК 004.3

Е.А. ТИМАШОВ

АЛГОРИТМЫ ФУНКЦИОНИРОВАНИЯ ВИРТУАЛЬНОГО ПРИБОРА ДЛЯ МЕДИЦИНСКИХ ЛЕЧЕБНО-ДИАГНОСТИЧЕСКИХ КОМПЛЕКСОВ

Приведены результаты анализа и исследования алгоритмов построения виртуального измерительного прибора для лечебно-диагностических комплексов, реализующих методы биоинформационных технологий. Описаны алгоритмы реализации.

Введение. В настоящей работе приведены результаты анализа и исследования алгоритмов функционирования виртуального измерительного прибора для лечебно-диагностических комплексов, реализующих методы биоинформационных технологий. Описаны алгоритмы реализации.

Лечебно-диагностические комплексы (ЛДК) – это автоматизированные системы для выработки путем автоматизированной диагностики и реализации лечебных воздействий на пациента, являющегося в данном случае объектом управления, на который направлены медикаментозные и процедурные воздействия в соответствии с критерием управления, выработанным врачом в процессе выявления заболевания или группы заболеваний, как с помощью ЛДК, так и с помощью амбулаторных и других методов. Исходя из этого отнесем ЛДК к одной из разновидностей автоматизированных систем управления (АСУ), которым свойственны следующие признаки, общие для всех АСУ. ЛДК – это человеко-машинная система, в которой врач играет важнейшую роль, принимая основное участие в выработке решений по диагностике и лечению [1]. Существенное место в ЛДК занимают различные датчики (сенсоры), устройство связи с пациентом (УСП), средства вычислительной техники, выполняющие операции по сбору, обработке и переработке информации, для установления диагноза и выработке стратегии и тактики лечения;

важную роль играют исполнительные устройства, т. е. устройства, осуществляющие лечебное воздействие на пациента, а также участвующие в управлении процессом получения диагностической информации. Следует иметь в виду, что ЛДК обеспечивает управление процессом лечения в целом, а его технические средства участвуют в выработке врачом решений по диагнозу и лечению [3].

Одна из задач при разработке программно-аппаратных комплексов для диагностики заболеваний по методам Фолля и Накатани – создать виртуальный прибор, который осуществляет измерения слаботочных сигналов проводимости в биологически активных точках тела пациента и через канал УНЧ-АЦП-ПЭВМ вводит цифровые результаты измерений. Отображение результатов измерений осуществляется на виртуальном стрелочном приборе, реализованном программными средствами на экране дисплея. Стрелочные приборы являются привычными для медицинского персонала в диагностических системах такого типа и их разработка выполняется по их требованиям.

Проведенные эксперименты и исследования показали, что реальные измерения сопровождаются высоким уровнем помех, порождаемых 50-ти герцовой сетью и импульсным блоком питания ПЭВМ. С учетом высокой скорости квантования сигнала и его отображения стрелка прибора на экране осуществляет «скачущие» движения, не естественные для обычного стрелочного прибора, обладающего достаточной инерционностью [2]. Для качественной имитации решение задачи осуществлялось путем создания цифрового фильтра, осуществляющего сглаживание ряда динамики с наложенной случайно составляющей, т. е. прогнозирование случайного процесса.

Для решения таких задач применяется большое количество разнообразных методов и приемов, которые разработаны и разрабатываются в различных областях науки и техники. Они отличаются различным математическим аппаратом, положенным в основу, а также другими характеристиками, в частности, объемом вычислений на шаге, точностью, количеством точек предыстории и др. Как критерий прогноза используется условие минимума среднеквадратичной ошибки

$$\overline{(y_i + y_i^*)^2} = \varepsilon^2_{0\min} \text{ при } \left| y_i - y_i^* \right| \le \delta, \qquad (1)$$

где δ – величина отклонения на каждом шаге прогноза.

Точность прогноза оценивалась по величине вариации.

$$\varepsilon = \frac{\left(y_i - y_i^*\right)^2}{\left(\overline{y_i}\right)^2 - \left(\overline{y_i^*}\right)^2} \cdot 100\,\%, \qquad (2)$$

где y_i – замеры процесса $y(t)$ в t_i, i =1, 2, y_i^* – предсказанное значение процесса в t_i.

Если точность не удовлетворяет, то порядок полинома увеличивается на единицу и так далее к получению малой или равной нолю вариации.

После получения аппроксимирующей функции $y^*(t)$ определяются значения интересующего нас процесса как в прошлом, так и в будущем. Однако если при этом приближение осуществляется путем построения полинома регрессии методом наименьших квадратов (а в большинстве случаев так и есть), то каждый раз при увеличении меры полинома приходится опять решать систему так называемых нормальных уравнений, которая увеличивает объем вычислений. Это увеличение еще более существенное учитывая то, что прогноз, как правило, осуществляется в «скользящем» режиме. Данный недостаток можно устранить благодаря свойству интерполяционного полинома Чебышева, которое заключается в том, что при увеличении меры полинома не нужно заново вычислять предыдущие члены ряда.

Ряд Чебышева для равноотстоящих точек, интерполирующий функцию с приближением, минимизирующим сумму

$$\sum_{i=1}^{n}\left[y(t_i)-Q_m(t_i)\right]^2$$

имеет вид

$$Q_m(t)=\sum\frac{\sum_{i=1}^{n}y_i f_\lambda(t_i)}{\sum_{i=1}^{n}f_\lambda^2(t_i)}f_\lambda(t).$$

В нем $f_\lambda(t)$ и $\sum f_\lambda^2(t)$ определяются по формулам:

$$f_\lambda(t)=(-1)^\lambda 1\cdot 2\cdot 3\cdot\ldots\cdot\lambda\cdot(n-1)(n-2)\ldots(n-\lambda)\left[1-\frac{\lambda(\lambda+1)(t-1)}{(n-1)(n-2)\cdot 1^2}+\right.$$

$$\left.+\frac{(\lambda-1)\lambda(\lambda+1)(\lambda+2)(t-1)(t-2)}{(n-1)(n-2)\cdot 1^2}-\ldots\right],$$

$$\sum_{i=1}^{n}f_\lambda^2(t_i)=\frac{(\lambda!)^2 n(n^2-2^2)\ldots(n^2-\lambda^2)}{2\lambda+1}.$$

Аппроксимирующий полином имеет два параметра: n – количество точек предыстории, m – мера полинома.

Интервал дискретности при вычислении экстраполяционных формул выбирается равным относительной единице, которая дает возможность работать с предысторией процесса независимо от абсолютной величины интервала дискретизации.

Предлагаемый алгоритм требует небольшого количества вычислений на шаге прогноза, прост в реализации и как показывает его применение, практически удобен.

При решении задачи прогнозирования с использованием указанных экстраполирующих полиномов как критерий прогноза применялось условие минимума среднеквадратичной ошибки

$\overline{(y_i+y_i^*)}=\varepsilon_{0_{\min}}^2$ при $\left|y_i-y_i^*\right|\le\delta$, где δ – величина отклонения на каждом шаге прогноза.

Точность оценивалась по величине вариации

$$\varepsilon = \frac{\overline{\left(y_i - y_i^*\right)^2}}{y_i^2 - \left(\overline{y_i}\right)^2}.$$

Выбор числа членов экстраполирующего полинома (длина предыстории наблюдения при этом n) связан с методической погрешностью представления процесса и неизменной погрешностью измерения. Поскольку с увеличением числа членов полинома первая погрешность убывает, а друга растет, то существует оптимальная в некотором смысле длина предыстории, которая удовлетворяет нужному соотношению указанных погрешностей. Эта длина оценивалась по минимальному значению ε.

Уравнение регрессии, выраженное через полиномы Чебышева, имеет вид

$$\hat{y} = b_0 P_0(x) + b_1 P_1(x) + ... + b_k P_k(x),$$

где $P_0(x)$, $P_1(x)$, …, $P_k(x)$ ортогональные полиномы Чебышева на множестве точек $x1$, $x2$,…, xn. Это означает, что для всех $u \neq j$ выполняются соотношения

$$\sum_{i=1}^{n} P_u(x_i) P_j(x_i) = 0,$$

где $P_{k+1}(x)$ зависит только от объема выборки n.

Зная многочлены Чебышева $P_{k+1}(x)$, при каждом увеличении степени уравнения регрессии необходимо вычислять только коэффициент b_{k+1}. Многочлены Чебышева определяются по формулам:

$$P_0(x) = 1,$$

$$P_1(x) = x - \frac{n+1}{2},$$

$$P_{k+1}(x) = P_1(X) \cdot P_k(x) - \frac{k^2(n^2 - k^2)}{4(4k^2 - 1)} P_{k-1}(x).$$

Например,

$$P_2(x) = x^2 - (n+1)x + \frac{(n+1)(n+2)}{6},$$

$$P_3(x) = x^3 - \frac{3(n+1)}{2} x^2 + \frac{6n^2 + 15 + 11}{10} x - \frac{(n+1)(n+2)(n+3)}{20},$$

$$P_4(x) = x^4 - 2(n+1)x^3 + \frac{9n^2 + 21n + 4}{7} x^2 - \frac{(n+1)(2n^2 + 7n + 10)}{7} x -$$
$$- \frac{(n+1)(n+2)(n+3)(n+4)}{80}.$$

Определяя коэффициенты b_0, $b_1, \ldots, b_k$ уравнения регрессии (IV.66) по методу наименьших квадратов, получаем

$$b_0 = \frac{\sum_{i=1}^{n} y_i}{n},$$

$$b_1 = \frac{\sum_{i=1}^{n} y_i P_1(x_i)}{\sum_{i=1}^{n} P_k^2(x_i)},$$

....

$$b_k = \frac{\sum_{i=1}^{n} y_i P_k(x_i)}{\sum_{i=1}^{n} P_k^2(x_i)}.$$

Вычисленные по формулам коэффициенты b_i не зависят от того, каков порядок определяемого уравнения регрессии. При нахождении уравнения регрессии методом последовательных уточнений используются все ранее найденные b_i. Повышение порядка уравнения регрессии на 1 приводит к определению только одного коэффициента.

При этом удобными получаются формулы для расчета остаточной дисперсии для уравнения регрессии k-го порядка:

$$S_{k\,ост}^2 = \frac{SS_k}{n-k-1},$$

где сумма квадратов отклонений ss_k определяются по рекуррентной формуле

$$SS_k = SS_{k-1} - b_k^2 \sum_{i=1}^{n} P_k^2(x_i).$$

Необходимо только заранее подсчитать SS_0:

$$SS_0 = \sum_{i=1}^{n} [y_i - b_0 P_0(x_i)]^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}.$$

При равноотстоящих значениях аргумента

$$x_2 = x_1 + n;\ x_3 = x_1 + 2h;\ \ldots\ x_n = x_1 + (n-1)h,$$

где h – шаг интерполяции, вычисления коэффициентов облегчаются. Сделаем замену переменных

$$z = \frac{x - x_1}{h} + 1.$$

Тогда каждое значение x_i заменится своим номером, т. е. $z_i = i$.

Будем искать уравнение регрессии в виде

$$\hat{y} = a_0 P_0(z) + a_1 P_1(z) + \ldots + a_k P_k(z),$$

где

$$a_0 = \frac{\sum_{i=1}^{n} y_i}{n},$$

$$a_1 = \frac{\sum_{i=1}^{n} y_i P_1(i)}{\sum_{i=1}^{n} P_1^2(i)},$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a_1 = \frac{\sum_{i=1}^{n} y_i P_k(i)}{\sum_{i=1}^{n} y_i P_k^2(i)}.$$

Суммы, стоящие в знаменателе, можно определить по сокращенной формуле:

$$\sum_{i=1}^{n} P_k^2(i) = \frac{(k!)^2 n(n^2-1)(n^2-4)\ldots(n^2-k^2)}{[(2k-1)!!]^2 2^{2k}(2k+1)},$$

где $(2k-1)!!$ – произведение всех нечетных чисел от 1 до $2k$ – 1 включительно.

В частности,

$$\sum_{i=1}^{n} P_1^2(i) = \frac{n(n^2-1)}{12},$$

$$\sum_{i=1}^{n} P_2^2(i) = \frac{n(n^2-1)(n^2-4)}{180},$$

$$\sum_{i=1}^{n} P_3^2(i) = \frac{n(n^2-1)(n^2-4)(n^2-9)}{2800},$$

$$\sum_{i=1}^{n} P_4^2(i) = \frac{n(n^2-1)(n^2-4)(n^2-9)(n^2-16)}{44100}.$$

Эти суммы используются и для вычисления сумм SS_k, нужных для определения остаточной дисперсии:

$$SS_k = SS_{k-1} - a_k^2 \sum_{i=1}^{n} P_k^2(i).$$

После получения уравнения регрессии переменную z опять заменяют первоначальной переменной x.

Прогнозирование и сглаживание с использованием модели – метод гармонических весов.

$x\{1{:}n\}$ – реализация процесса; n – длина предыстории.

Вычисляется тренд $\hat{x}[1:n]$, одним из существующих методов.

1. Вычисляется вес в каждой точке:

$$A_i = \frac{\hat{x}_n - \hat{x}_i}{n-1}, \qquad (I = 1, 2, \ldots, n).$$

Вычисляем средний вес:

$$W = \frac{1}{n-1}\sum_{i=1}^{n} A_i.$$

Тогда значение в (n+1) точке будет

$$x^*_{n+1} = W \cdot t + \hat{x}_n,$$

где t – длина интервала предсказания.

К предыстории прибавляется предсказанное значение и следующее предсказание делается по расширенной предыстории, пока

$$n \le n_0 + T,$$

где T – горизонт прогноза, n_0 – начальная предыстория.

2. Прогнозирование без сглаживания.

Веса вычисляются по формуле

$$A_i = \frac{x_n - x_i}{n-1}, \quad W' = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{x_n - x_i}{n-1},$$

при этом вес вычисляется как

$$W' = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{x_n - x_i}{n-1}.$$

Прогноз в (n+1) точке определяется как

$$x^*_{n+1} = W'\cdot t + x_n.$$

Далее предыстория расширяется, как и в п. 1.

2.1. Прогнозирование с частичным сглаживанием (без вычисления тренда)

Вычисляем среднее значение $\overline{x}$ по последним 5 точкам предыстории:

$$\overline{x} = \frac{1}{5}\sum_{n-5}^{n} x_m; \quad m = \frac{1}{5}\sum_{n-5}^{n} m$$

и определяем коэффициент корреляции в виде

$$b = \frac{\sum_{n-5}^{n}(m - \overline{n})\cdot(x_m - \overline{x})}{\sum_{n-5}^{n}(m - \overline{n})^2}.$$

При этом сглаженные значения предыстории получаем в виде:

$$\hat{x}_n = \overline{x} - b(\overline{n} - m),$$

далее вычисляем веса и их среднее значение: A_i, W' вводим уклонение истинного значения x_n от сглаженного $\hat{x}_n$

$$\Im = x_n - \hat{x}_n.$$

Тогда прогнозированное значение

$$x^{*}_{n+1} = W\,'\cdot t + \hat{x}_n + a\mathfrak{I}_n,$$

где вычисляется следующим образом:

$$a = \frac{1}{n-1}\left(\sum \frac{1}{n-k-1}\right)\frac{t-n+3}{4}.$$

Метод простого скользящего среднего состоит в том, что расчет показателя на прогнозируемый момент времени строится путем усреднения значений этого показателя за несколько предшествующих моментов времени.

При составлении прогноза этим методом часто приходится наблюдать, что влияние используемых при расчете реальных показателей оказывается неодинаковым, при этом обычно более свежие данные имеют больший вес.

Вклад различных моментов времени можно учесть, вводя вес для каждого значения показателя в скользящем интервале. В результате приходим к методу взвешенного скользящего среднего:

$$f_k = \frac{\sum_{i=1}^{N} W_{k-1} x_{k-1}}{\sum_{i=1}^{N} W_{k-1}},$$

где W_{k-1} – вес, с которым используется показатель x_{k-1} при расчете. Вес – это всегда положительное число. В случае, когда все веса одинаковы, мы возвращаемся к обычному методу скользящего среднего.

Методам скользящего среднего недостает адаптивности (приспособляемости). Прогнозы в них зависят от предыдущих значений показателя, но не от качества предыдущих прогнозов.

Метод экспоненциального сглаживания. При расчете прогноза методом экспоненциального сглаживания учитывается отклонение предыдущего прогноза от реального показателя, а сам расчет проводится по следующей формуле:

$$f_k = f_{k-1} + \alpha\left(x_{k-i} - f_{k-1}\right),$$

где α – постоянная сглаживания $\left(0 \le \alpha \le 1\right)$.

Рассмотрим модель экспоненциального сглаживания более детально. Для начала перепишем формулу в виде

$$f_k = \alpha x_{k-1} + (1-\alpha) f_{k-1}.$$

Для $k > 2$, k можно заменить на $k - 1$.

$$f_{k-1} = \alpha x_{k-2} + (1-\alpha) f_{k-2}$$

и подставим полученное выражение

$$f_k = \alpha x_{k-1} + \alpha(1-\alpha)x_{k-2} + (1-\alpha)^2 f_{k-2}.$$

Выполняя последовательно аналогичные подстановки, получаем следующее выражение для f_k:

$$f_k = \alpha x_{k-1} + \alpha(1-\alpha)x_{k-2} + \alpha(1-\alpha)^2 x_{k-3} + ... + \alpha(1-\alpha)^{k-2} x_1 + (1-\alpha)^{k-1} f_1.$$

Поскольку из неравенства $0 \le \alpha \le 1$ следует, что $0 \le 1-\alpha \le 1$, то $\alpha > \alpha(1-\alpha) > \alpha(1-\alpha)^2...$. Другими словами, коэффициенты при x_k убывают при уменьшении номера k: наблюдение x_{k-1} имеет больший вес, чем наблюдение x_{k-2}, которое, в свою очередь, имеет больший вес, чем x_{k-3}.

Видно, что значением f_k является взвешенная сумма всех предыдущих наблюдений, включая последнее наблюдение x_{k-1}. Последнее слагаемое суммы является не статистическим наблюдением, а «предположением» f_1. Очевидно, что с ростом k влияние f_1, на прогноз уменьшается, и в определенный момент им можно будет пренебречь. Даже если значение α достаточно малое (такое, что (1 – α) приблизительно равно 1), значение $(1-\alpha)^t$ будет быстро убывать.

Оценить точность сделанных прогнозов можно, в частности, используя среднее абсолютных отклонений (САО)

$$CAO = \frac{\sum_N |x_i - f_i|}{N}$$

или среднее относительных ошибок, в процентах (СООП)

$$COO\Pi = \frac{\sum_N \frac{|x_i - f_i|}{x_i} \times 100\%}{N},$$

где N – количество прогнозов.

Выводы. Выбор варианта прогнозирования осуществляется таким образом: сначала 3 методами прогнозируется последняя (известна точка) предыстории и с уклонения прогноза от реализации выбирается min, после чего дальнейшая работа идет по варианту соответствующему min уклонению.

Є.О. Тимашов

АЛГОРИТМИ ФУНКЦІОНУВАННЯ ВІРТУАЛЬНОГО ПРИЛАДУ ДЛЯ МЕДИЧНИХ ЛІКУВАЛЬНО-ДІАГНОСТИЧНИХ КОМПЛЕКСІВ

Наведено результати аналізу та дослідження алгоритмів функціонування віртуального вимірювального приладу для лікувально-діагностичних комплексів, що реалізують методи біоінформаційних технологій. Описано алгоритми реалізації.

E.O. Timashov

ALGORITHMS OF FUNCTIONING OF VIRTUAL DEVICES FOR MEDICAL-DIAGNOSTIC COMPLEXES

The results of analysis and research of the algorithms of virtual instrumentation functioning for medical-diagnostic systems that implement the methods of bioinformatics technologies are presented. The implementation algorithms are described.

1. *Тимашов Е.А.* Системный анализ компьютерных лечебно-диагностических комплексов. *Комп'ютерні засоби, мережі та системи*. 2004. С. 156 – 162.
2. *Тимашов Е.А.* Синтез структури програмного забезпечення біоінформаціних систем діагностики та лікування.Там само. 2005. С. 145 – 152.
3. *Тимашов Е.А.* Функциональные основы и алгоритмы работы распределенных биоинформационных систем диагностики и лечения. Збірник наукових праць Інституту кібернетики імені В.М. Глушкова НАН України *Нові комп'ютерні засоби, обчислювальні машини та мережі*. 2001. Т. 2. С. 119 – 126.

Получено 20.09.2016

Об авторе:

Тимашов Евгений Александрович,
научный сотрудник
Института кибернетики имени В.М. Глушкова НАН Украины.