лон B длины $l+2$ следующего вида: $\sigma_1 \underbrace{*\ldots*}_{l} \sigma_2$, где $\sigma_1, \sigma_2 \in \{0,1\}$. И пусть $Z_{l+2} = (z_t, \ldots, z_{t+l+1})$ и $Y_{l+2} = (y_1, \ldots, y_{l+2})$, где y_i – независимые случайные величины Бернулли со следующим распределением вероятностей: $P\{y_i = 1\} = \frac{1}{2}$, $P\{y_i = 0\} = \frac{1}{2}$. Используя теорему 2, получим:

$$|E\{template_B(Z_{l+2})\} - E\{template_B(Y_{l+2})\}| =$$

$$= |P\{z_t = \sigma_1, z_{t+l+1} = \sigma_2\} - P\{y_1 = \sigma_1, y_{l+2} = \sigma_2\}| = |P\{z_t = \sigma_1, z_{t+l+1} = \sigma_2\} - \frac{1}{4}| = O\left(\frac{l}{2^{|A|}}\right).$$

Запишем формулу корреляции между z_t и z_{t+l+1} следующим образом:

$$corr(z_t, z_{t+l+1}) = 2P\{z_t = z_{t+l+1}\} - 1 =$$

$$= 2(P\{z_t = 0, z_{t+l+1} = 0\} + P\{z_t = 1, z_{t+l+1} = 1\}) - 1 = 2\left(\frac{1}{4} + \frac{1}{4} + O\left(\frac{l}{2^{|A|}}\right)\right) - 1 = O\left(\frac{l}{2^{|A|}}\right). \blacksquare$$

Следствие 2. *Пусть P – произвольная бинарная k-грамма. И пусть Z_k – k последовательных бит последовательности z, порождаемой самосжимающимся генератором. Тогда* $|P\{Z_k = P\} - 2^{-k}| = O\left(\frac{k}{2^{|A|}}\right)$.

Доказательство. Пусть $Y_k = (y_1, \ldots, y_k)$, где y_i – независимые случайные величины Бернулли со следующим распределением вероятностей: $P\{y_i = 1\} = \frac{1}{2}$, $P\{y_i = 0\} = \frac{1}{2}$. Так как произвольную бинарную k-грамму P можно рассматривать как шаблон $B = (\sigma_1, \ldots, \sigma_k)$ размера k, где $\sigma_i \in \{0,1\}$, из теоремы 2, получим: $|E\{template_B(Z_k)\} - E\{template_B(Y_k)\}| = |P\{Z_k = P\} - \frac{1}{2^k}| = O\left(\frac{k}{2^{|A|}}\right)$. $\blacksquare$

Литература. 1. Meier W., Staffelbach O. The Self-Shrinking Generator // Eurocrypt'94. Proceedings. Springer-Verlag. 1998. P. 205-214. **2.** Coppersmith D., Krawchuk Y., Mansour Y. The shrinking generator // Advanced in Cryptology: Proceedings of Crypto 93, LNCS 773. 1994. P. 22-39.

ВОССТАНОВЛЕНИЕ ФУНКЦИОНАЛЬНОЙ ЗАВИСИМОСТИ В ЗАДАЧАХ ТЕПЛОПЕРЕНОСА

Чехменок С.Л., Иванов С.Н., БГУ, Минск

Для методологии теории тепло- и массопереноса характерно то, что она сочетает методы расчета температурных и массовых полей с помощью дифференциальных уравнений переноса и экспериментальные методы определения характеристик тепло- и массопереноса. В настоящее время для решения задач в

данной области проблемой являются сложности, связанные с информационным обеспечением математических моделей. Для информационного обеспечения моделей тепло- и массопереноса необходимо с помощью экспериментальных исследований получить эмпирические зависимости всех необходимых параметров и характеристик.

Характеристики тепло- и массопереноса, входящие в дифференциальные уравнения, как правило, находятся с помощью экспериментальных методов, в основе которых лежат либо решения обратных задач тепло- и массопереноса, либо восстановление зависимостей с помощью статистических методов, опираясь на систематизированные экспериментальные данные. Информационное обеспечение в виде эмпирических формул, с помощью которых аппроксимированы характеристики переноса, можно использовать для широкого класса задач в данной области и удобно вводить в программы для расчета процессов.

С учетом специфики для восстановления функциональной зависимости рассмотрим следующий алгоритм.

1. Определяем, являются ли данные зависимы. Для этого используем метод серий, основанный на медиане, и метод восходящих и нисходящих серий [1]. И если хотя бы один метод дает положительный результат, то данные считаются зависимыми и производится переход ко второму пункту. Если оба метода говорят нам о том, что данные являются независимыми, то на этом останавливаемся, поскольку выделять больше нечего.

2. Если у нас одна независимая переменная, то проверяем данные на стационарность методом серий и методом превышений [1]. Если хотя бы один из методов дает положительный результат, то тренда у нас уже нет и можно выделить только периодические компоненты. Для этого используем метод, основанный на построении периодограммы [2]. Если у нас несколько переменных, то сразу переходим к пункту 3.

3. Итерационно подбираем зависимость следующим образом.

а) На начальном этапе предполагаем, что зависимость $f(x)$ - линейная, то есть $f(x) = a_0 + a_1 x_1 + ... + a_n x_n$.

b) Определяем коэффициенты $a_i, i=\overline{1,n}$ с помощью метода наименьших квадратов.

c) Вычисляем среднеквадратическое отклонение.

d) Усложняем зависимость $f(x)$, и предполагаем, что у нас появляется новое слагаемое $a_{n+1}g(x)$, где $g(x)$ принадлежит множеству элементарных функций. Повторяем процедуру пересчета, начиная с пункта (b) и, после вычисления среднеквадратического отклонения, сравниваем с отклонением, полученным на предыдущем шаге. Если оно уменьшилась, то мы добавили нужное слагаемое в функциональную зависимость. Если среднеквадратичное отклонение увеличилась, то слагаемое не верно. Мы данное слагаемое отбрасываем, добавляем новое, и пересчитываем заново.

С целью проверки и использования вышеприведенного алгоритма, были рассмотрены стандартные статистические пакеты (Statistica 6.0, SPSS 11.0 for Windows, Statit Professional QC 5.2). Производилось сравнение работы пакетов как на модельных, так и на реальных данных. При их использовании были сделаны следующие выводы.

1. Все статистические пакеты поддерживают стандартные форматы представления данных и позволяют обрабатывать большие объемы данных (выборки более 5 000 000 элементов).

2. Ни одна из рассмотренных программ не поддерживает динамического обновления данных. Следовательно, для восстановления функциональной зависимости по незначительно измененным входным данным, нужно проделать все этапы восстановления сначала.

3. На простых моделях скорость и точность вычисления приблизительно одинаковы для всех пакетов. Время вычисления зависит от объема выборки. Например, при выборке 1000 элементов точность вычисления коэффициентов равна 10^{-16}. В случае, если модель простая, то объем выборки не должен быть большим.

4. При восстановлении зависимостей, которые содержат несколько независимых переменных, нужно чтобы данные были распределены равномерно в той области, в которой идет восстановление. Если такое условие не соблюдается, то исследователь должен убедиться, что восстанавливаемая им зависимость будет однозначна.

5. Все статистические пакеты используют начальные задания параметров. В зависимости от того, на сколько удачно они выбраны, зависит время работы программы. Это справедливо как для моделей без ошибок, так и с ошибками.

6. Точность построения оценок резко уменьшается, если в моделях наблюдаются ошибки. Если в качестве ошибки использовать случайные величины, у которых есть конечная дисперсия, то точность вычислении зависит от объема выборки и величины дисперсии. При уменьшении дисперсии уменьшается и ошибка (при постоянном объеме выборки). При постоянной дисперсии и увеличении объема выборки увеличивается и точность.

Дисперсия	Объем выборки	Точность вычисления
1	1000	10^{-5}
0.5	1000	10^{-7}
1	2000	10^{-7}
0.5	2000	10^{-9}

7. При использовании в качестве ошибки устойчивых случайных величин, ошибки на два порядка больше чем в случае использования случайных величин с конечной дисперсией. Чем меньше индекс устойчивости, тем больше ошибка. Это связано с тем, что устойчивые законы не имеют конечную дисперсию и используемые в статистических пакетах алгоритмы не работают в данном случае.

Индекс устойчивости	Объем выборки	Точность вычисления
1.7	1000	10^{-3}
1.3	1000	10^{-2}
1.7	2000	10^{-5}
1.3	2000	10^{-4}

Литература. 1. Бендат Дж., Пирсол А., Прикладной анализ случайных данных. // М., Мир, 1989. **2.** Андерсен Т., Статистический анализ временных рядов. // М., Мир, 1978.