УДК 519.677

В. В. Поцепаев[1](канд. техн. наук, доц.),
Р. В. Поцепаев[2] **(**канд. физ.-мат. наук, Phd**)**
[1]Донецкий национальный технический университет
[2]Технологический центр Шлюмберже, Абингдон, Великобритания
Potsepaev56@mail.ru, RPotsepaev@slb.com

ВЫБОР ФУНКЦИОНАЛЬНЫХ ЗАВИСИМОСТЕЙ И КРИТЕРИЕВ СВЯЗИ ВРЕМЕННЫХ РЯДОВ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ

Рассматривается задача прогнозирования временных рядов, предложены критерии определения функциональных зависимостей между рядами известными в области прогнозирования и прогнозируемыми рядами для семейств монотонных, унимодальных и непрерывных функций, приведена настройка параметров выбранных критериев.

прогнозирование, временной ряд, функциональная зависимость, критерии

Общая постановка задачи

Поведение временных рядов можно прогнозировать, изучив в достаточной степени их связь с совокупностью других временных рядов, которые оказывают на них основное влияние. Разделив набор рядов на те, для которых известно их поведение в будущем (т.е. в области прогнозирования) и те, значения которых необходимо предсказать, можно с успехом решать проблему прогнозирования, если удается выяснить существование и вид связей между выбранными группами рядов и доказать, что другие факторы оказывают незначительное влияние на прогнозируемые ряды. В связи с описанным подходом возникает следующая задача.

Предположим, что задан набор временных рядов

$$A_1, A_2, \dots, A_m, где\ A_i = \{a_1^i, a_2^i, \dots, a_n^i\}, a_j^i \in E_i$$

Необходимо выявить все или почти все существующие зависимости между данными рядами, т.е. зависимости вида

$$a_t^k = f\left(a_1^1, \dots, a_n^1, \dots, a_1^m, \dots, a_n^m\right) \quad (1)$$

Ограничимся рассмотрением случая

$$a_t^k = f\left(a_t^1, \dots, a_t^{k-1}, a_t^{k+1}, \dots, a_t^m\right) \quad (2)$$

То есть, предположим, что нет зависимости от предыстории, либо эта зависимость несущественна. Отметим, что задачу с зависимостью вида (1) можно свести к задаче (2) добавлением новых рядов вида

$$\left\{a_2^k,\ldots,a_{i-1}^k,\ldots,a_{n-1}^k\right\},\left\{a_3^k,\ldots,a_{i-2}^k,\ldots,a_{n-2}^k\right\},\ldots,\left\{a_s^k,\ldots,a_{i-s}^k,\ldots,a_{n-s}^k\right\},$$

учитывая тем самым зависимость от предыстории. Очевидно, что ряд A_k далеко не всегда зависит от всех исследуемых рядов, поэтому часто число аргументов в функции (2) существенно меньше. Таким образом, если ряд A_k зависит от рядов $A_1, A_2, \ldots, A_l$, то эта зависимость может быть представлена в виде

$$a_t^k = f\left(a_t^1,\ldots,a_t^l\right), \text{ где } l \le m\text{-}1 \tag{3}$$

Для многих предметных областей о зависимости (3) можно говорить лишь приблизительно, так как на значения рядов влияют многие факторы, не описываемые в виде (3), например – случайные процессы, вызывающие случайные отклонение a_t^k, непрогнозируемые события, нарушающие вид связи (3) в течение небольшого промежутка времени, изменения функциональной зависимости со временем, влияние других, неучтенных временных рядов и т.п. Если все эти факторы не слишком сильно нарушают вид связи (3), т.е.

$$a_t^k = f\left(a_t^1,\ldots,a_t^l\right) + \varepsilon_t \quad , \tag{4}$$

где ε_t не слишком велико, то можно подобрать разумный критерий, выясняющий существование функциональной зависимости вида (4) между выбранными рядами.

Выбор семейств функциональных зависимостей

Очевидно, что для правильного выбора критерия, определяющего существование функциональной зависимости вида (4), необходимо выделить семейство функциональных зависимостей (далее просто «семейство функций»), существование которых может быть найдено с помощью выбранного критерия. Например, можно говорить о критерии для определения существования линейной зависимости между рядами. Необходимо выбрать набор семейств функций и соответствующий им набор критериев для выделения как можно большего числа различных зависимостей между временными рядами, характерных для исследуемой предметной области. Описанный набор семейств функций варьируется для различных предметных областей, однако, очевидно, что он должен быть существенно уже множества всевозможных отображений

$\left(a_t^k, a_t^1, ..., a_t^l\right)$ для независимых рядов $A_k, A_1, A_2, ..., A_l$. Также он должен быть достаточно узок для того, чтобы с помощью созданных критериев можно было находить зависимости из заданного набора семейств функций при достаточно больших ε_t.

Рассмотрим вначале одномерный случай $l=1$, $a_t = f(b_t)$, т.е. выберем набор функциональных зависимостей для случая функции одной переменной. Через произвольный набор точек (b_t, a_t) всегда можно провести ломанную, поэтому всегда существует непрерывная функция, проходящая через все точки и, следовательно, невозможно отличить зависимость вида (4) как непрерывную функцию от случайного набора точек, т.е. нельзя сказать, существует ли непрерывная зависимость между рядами *A* и *B*. Это затрудняет поиск непрерывных, тем более кусочно-непрерывных зависимостей в общем случае. Поэтому набор семейств функций должен обладать какими-то дополнительными свойствами. Можно выделить набор семейств функций, который достаточно широк - зависимости из этого набора наблюдаются во многих предметных областях и в тоже время достаточно узок в указанном выше смысле.

Рассмотрим следующий набор Ф = $\{C,M,U\}$, где

С - семейство непрерывных функций с ограниченным числом экстремумов, причем точки экстремума находятся друг от друга на достаточно большом расстоянии и отношение суммарной площади областей, ограниченных сверху частью кривой функции, содержащей один экстремум, а снизу горизонтальным отрезком длины Δ к общей площади графика меньше некоторой величины Σ (Δ, Σ – некоторые константы);

M - семейство монотонных функций;

U - семейство унимодальных функций (т.е. функций с одним экстремумом).

Заметим, что из монотонности и унимодальности ограниченных функций на отрезке следует их кусочная непрерывность. Для произвольной меры непрерывную функцию можно сколь угодно мало приблизить кусочно-непрерывной функцией, поэтому, для удобства будем считать, что все функции из Ф непрерывны. Вообще, если величина ε_t имеет случайных характер, то при определении существования зависимости, невозможно основываться на локальных свойствах функции, таких как непрерывность или гладкость. Все функции из M и почти все из U принадлежат семейству C, однако в связи с их простой конструкцией и тем фактом, что они достаточно часто встречаются в различных предметных областях, имеет смысл выделить их в отдельное семейство, что позволит в дальнейшем находить зависимости из семейств M и U при больших значениях ε_t.

Рассмотрим теперь функцию многих переменных $a_t^k = f\left(a_t^1,...,a_t^l \right)$.

В данном случае рассмотрим набор, состоящий из семейства функций, для которых выполнено следующее условие: пусть Ф = $\{\varphi_1,\varphi_2,...,\varphi_s\}$ - выбранный набор семейств функций одной переменной. Если для любого $r \in 1...l$ существует $i \in 1...s$ такое, что для любого произвольного фиксированного набора $a_t^1,...,a_t^{r-1},a_t^{r+1},...,a_t^l$, равного C^1, функция $\grave{a}_t^k = f\left(a_t^r\right)_{a_t^1,...a_t^{r-1},a_t^{r+1},...,a_t^l = C^1}$ принадлежит семейству F_i из выбранного набора семейств Ф для случая функции одной переменной, то функция $f\left(a_t^1,...,a_t^l \right)$ принадлежит заданному семейству, т.е.

$$\forall r \ \exists i F_i \in \Phi \forall C^1 \ f\left(a_t^r \right)_{a_t^1,...a_t^{r-1},a_t^{r+1},...,a_t^l = C^1} \in F_i \qquad (5)$$

В данном случае Ф = $\{C,M,U\}$, причем предположим, что направление монотонности и унимодальности сохраняется, если функции принадлежат соответствующим семействам.

Критерии для случая функций одной переменной

Ниже будут описаны критерии для семейств из набора Ф=$\{C,M,U\}$ для функций одной переменной, то есть ограничимся пока случаем, когда ряды связаны попарно. Предположим, что заданы два временных ряда *A* и *B*, для которых необходимо установить существование зависимости из Ф. Ответ на вопрос о существовании зависимости неоднозначен в связи с существованием ε_t – случайного отклонения от истинного значения либо отклонения связанного с причинами, природа которых не выяснена. Задача состоит в том, чтобы для каждого семейства функций из набора создать критерий (другими словами алгоритм), позволяющий в большинстве случаев давать правильный ответ на вопрос о существовании зависимости между рядами из выбранного семейства, если ε_t не слишком велико. Для случая двух рядов, существование зависимости можно определить, например, зрительно, если рассмотреть график зависимости a_t, от b_t, . Например, на рис. 1. вероятно наблюдается некоторая зависимость. Нижеописанные критерии для определения существования функциональной зависимости того или иного типа заключаются в вычислении некоторой величины λ_F, определяющей степень отсутствия зависимости выбранного типа, например, степень немонотонности функции. Далее, выбрав некоторое пороговое значение μ_F, критерий можно представить в виде параметризированного предиката

$$K_{\mu}^{F}(A,B)=\{1,если\,\lambda_F<\mu_F;0,если\,\lambda_F\geq\mu_F\} \quad (6)$$

где F - семейство функциональных зависимостей, для которых данный критерий предназначен.

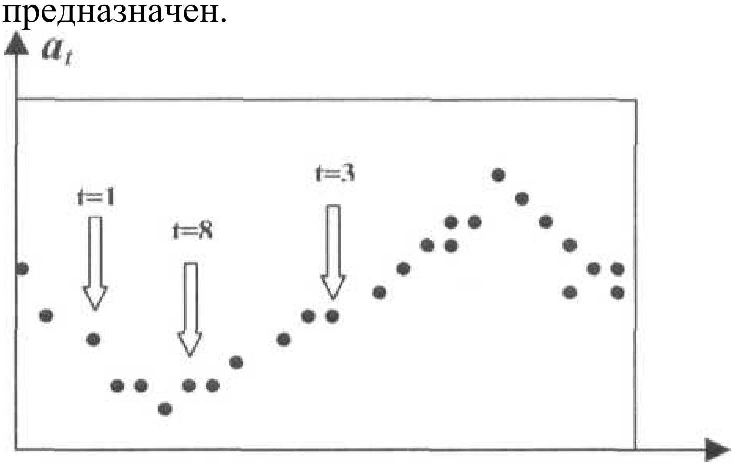

Рисунок 1 – Зависимость a_t от b_t

1. Непрерывные функции.

Выберем максимальное и минимальное значение из последовательностей a_t и из последовательности b_t. Рассмотрим прямоугольную область, образованную из этих четырех точек. Очевидно, что все остальные точки (b_t, a_t) лежат внутри этой области. Пусть S_0- площадь указанной области. Выберем некоторое значение Δ. Из множества точек (b_t,a_t) выберем два подмножества следующим образом – в первое подмножество входят точки (b_t, a_t), для которых известно, что в прямоугольной области $\{(b_t,a_t),(b_t+\Delta,a_t),(b_t,0),(b_t+\Delta,0)\}$ либо в области $\{(b_t,a_t),(b_t-\Delta,a_t),(b_t,0),(b_t-\Delta,0)\}$ нет других точек (b_t,a_t), см. рис. 2. Во второе подмножество входят точки, для которых известно, что в прямоугольной области $\left\{(b_t,a_t),(b_t+\Delta,a_t),\left(b_t,a_t^{max}\right),\left(b_t+\Delta,a_t^{max}\right)\right\}$ либо в области $\left\{(b_t,a_t),(b_t-\Delta,a_t),\left(b_t,a_t^{max}\right),\left(b_t-\Delta,a_t^{max}\right)\right\}$ нет других точек.

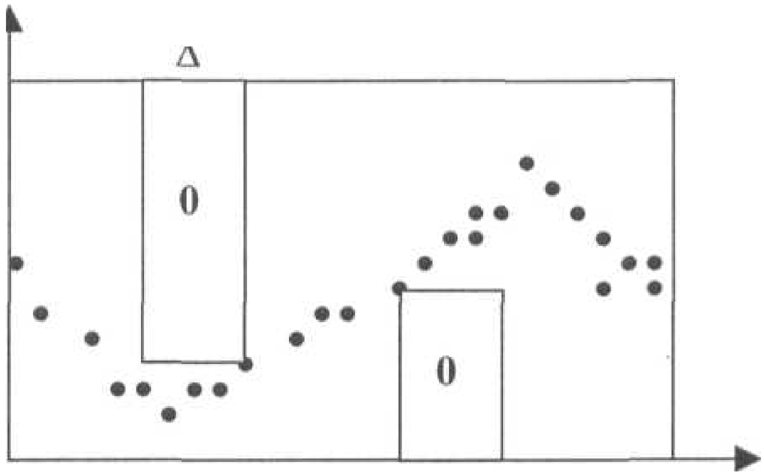

Рисунок 2– К определению меры отсутствия зависимости рядов

Для каждого полученного подмножества рассмотрим ломаную, полученную последовательным соединением точек. Каждая ломанная делит область S_0 на две части. Пусть ломанная, образованная точками первого подмножества, ограничивает снизу область с площадью S_1, ломаная, образованная точками второго подмножества, ограничивает снизу область с площадью S_2. Мерой отсутствия зависимости из С назовем отношение

$$\lambda_C = \frac{S_2 - S_1}{S_0} \qquad (7)$$

Легко показать, что ломанные не пересекаются, но могут частично совпадать, следовательно, величина $S_2 - S_1$ неотрицательна, и λ_C принимает значения из отрезка [0,1].

2. Монотонные функции.

По определению функция называется монотонно возрастающей, если для любых x, y, таких что x<y, выполнено условие $f(x) \le f(y)$. Исходя из определения, можно определить следующий простой критерий монотонности. Пусть N_1 – количество пар точек $\{(b_t, a_t),(b_s, a_s)\}$, для которых выполнено следующее условие: $b_t < b_s \wedge a_t > a_s$; N_2 – количество пар точек, для которых выполнено следующее условие: $b_t < b_s \wedge a_t < a_s$, тогда мерой немонотонности функции будем называть величину

$$\lambda_M = \frac{min(N_1, N_2)}{C_N^2}, \qquad (8)$$

где C_N^2 - общее количество пар $\{(b_t, a_t),(b_s, a_s)\}$. Заметим, что мера (8) применима как к возрастающим, так и к убывающим функциям.

3. Унимодальные функции.

Пусть (b_t, a_t) - экстремум унимодальной функции, d – количество точек (b_s, a_s), для которых $b_s < b_t$. Тогда для точек слева и справа от (b_t, a_t) можно вычислить меру немонотонности функции. Складывая эти величины, получим меру отсутствия унимодальности. Так как экстремум заранее не известен, то можно выбрать точку (b_t, a_t) так, чтобы мера принимала наименьшее значение. Таким образом, мерой отсутствия унимодальности функции будем называть величину

$$\lambda_U = \min_d \min\left\{ \frac{N_1^d}{C_d^2} + \frac{N_2^{N-d}}{C_{N-d}^2}, \frac{N_2^d}{C_d^2} + \frac{N_1^{N-d}}{C_{N-d}^2} \right\} \tag{9}$$

где N_1^d, N_2^d - количество пар точек $\{(b_t, a_t), (b_s, a_s)\}$, для которых выполнено условие $b_t < b_s \wedge a_t > a_s$ (соответственно $b_t < b_s \wedge a_t < a_s$) и которые лежат левее экстремума, аналогично N_2^{N-d}, N_1^{N-d} - количество пар точек с указанными условиями, лежащие правее экстремума.

Общий критерий поиска зависимостей для функции многих переменных

Определим меру зависимости ряда A_k от рядов $A_1, A_2, ..., A_l$ для функций многих переменных $a_t^k = f\left(a_t^1, ..., a_t^l\right)$, для которых выполнено условие (5). Рассмотрим вначале набор Ф = *{M,U}*. Так как в этом случае при произвольных фиксированных $a_t^1, ..., a_t^{r-1}, a_t^{r+1}, ..., a_t^l$ функция $a_t^k = f\left(a_t^r\right)\Big|_{a_t^1, ... a_t^{r-1}, a_t^{r+1}, ..., a_t^l = const}$ является либо монотонной, либо унимодальной, то, выбрав несколько наборов $a_t^1, ..., a_t^{r-1}, a_t^{r+1}, ..., a_t^l$, можно предварительно определить вид зависимости $a_t^k = f\left(a_t^r\right)$, которая предположительно может существовать. То есть для большинства выбранных наборов необходимо, чтобы функция $a_t^k = f\left(a_t^r\right)$ была одновременно либо возрастающая, либо убывающая, либо унимодальная с одним максимумом, либо унимодальная с одним минимумом. Если это не так, то можно утверждать, что нет зависимости ряда A_k от ряда A_r, или же вид зависимости нам неизвестен. В обоих случаях будем считать, что нет зависимости ряда A_k от рядов $A_1, A_2, ..., A_l$ в виде функции многих переменных с условием (5). Рассмотрим ряды A_r, для которых функция $a_t^k = f\left(a_t^r\right)$ является унимодальной. Допустим, что рассмотрев несколько наборов $a_t^1, ..., a_t^{r-1}, a_t^{r+1}, ..., a_t^l$ (обозначим эти наборы C^1, C^2 ,...,C^p), мы пришли к выводу, что возможна унимодальная зависимость, причем положение экстремума (например, максимума) в рассмотренных случаях было в точках $\left(a_t^1, ..., a_t^l\right)_1, \left(a_t^1, ..., a_t^l\right)_2, ..., \left(a_t^1, ..., a_t^l\right)_p$. По указанным точкам можно построить процедуру вычисления положения точки экстремума функции $a_t^k = f\left(a_t^r\right)$ для произвольного набора $a_t^1, ..., a_t^{r-1}, a_t^{r+1}, ..., a_t^l$ - например, можно исходить из предположения, что экстремум мало меняется при изменении набора $a_t^1, ..., a_t^{r-1}, a_t^{r+1}, ..., a_t^l$, в таком случае можно принять за точку экстремума среднее значение точек $\left(a_t^1, ..., a_t^l\right)_1, \left(a_t^1, ..., a_t^l\right)_2, ..., \left(a_t^1, ..., a_t^l\right)_p$. Другой способ заключается в интерполяции функции положения экстремума по точкам $\left(a_t^1, ..., a_t^l\right)_1, \left(a_t^1, ..., a_t^l\right)_2, ..., \left(a_t^1, ..., a_t^l\right)_p$.

Допустим, что для рядов $A_1, ..., A_{r_1}$ предположительно наблюдается монотонно возрастающая функция $a_t^k = f\left(a_t^r\right)$, для рядов $A_{r_1+1}, ..., A_{r_2}$ - монотонно убывающая функция, для рядов $A_{r_2+1}, ..., A_{r_3}$ - унимодальная с максимумом, и для рядов $A_{r_3+1}, ..., A_{r_l}$ - унимодальная с минимумом. Рассмотрим пару точек $\left(a_i^1, ..., a_i^l\right)$ и $\left(a_j^1, ..., a_j^l\right)$, для которых выполнены следующие условия:

$$
\begin{array}{l}
a_i^k < a_j^k \\
a_i^r \ge a_j^r \; r \le r_1 \\
a_i^r \le a_j^r \; r_1 < r \le r_2 \\
a_i^r \ge a_j^r \wedge a_i^r < a_{d_i}^r \wedge a_j^r < a_{d_j}^r \vee a_i^r \le a_j^r \wedge a_i^r > a_{d_i}^r \wedge a_j^r > a_{d_j}^r \quad r_2 < r \le r_3 \\
a_i^r \le a_j^r \wedge a_i^r < a_{d_i}^r \wedge a_j^r < a_{d_j}^r \vee a_i^r \ge a_j^r \wedge a_i^r > a_{d_i}^r \wedge a_j^r > a_{d_j}^r \quad r_3 < r \le r_l
\end{array}
, \tag{10}
$$

где $a_{d_i}^r, a_{d_j}^r$ - значения экстремума унимодальной функции $a_t^k = f\left(a_t^r\right)$ при $a_t^1, ..., a_t^{r-1}, a_t^{r+1}, ..., a_t^l$ равном $a_i^1, ..., a_i^{r-1}, a_i^{r+1}, ..., a_i^l$ (соответственно $a_j^1, ..., a_j^{r-1}, a_j^{r+1}, ..., a_j^l$).

Описанное выше условия выделяет те пары точек, для которых не выполнено сделанное предположение о монотонности и унимодальности рядов. Также можно указать условие, описывающее те точки, для которых предположение полностью выполнено:

$$
\begin{array}{l}
a_i^k < a_j^k \\
a_i^r < a_j^r \; r \le r_1 \\
a_i^r > a_j^r \; r_1 < r \le r_2 \\
a_i^r < a_j^r \wedge a_i^r < a_{d_i}^r \wedge a_j^r < a_{d_j}^r \vee a_i^r > a_j^r \wedge a_i^r > a_{d_i}^r \wedge a_j^r > a_{d_j}^r \quad r_2 < r \le r_3 \\
a_i^r > a_j^r \wedge a_i^r < a_{d_i}^r \wedge a_j^r < a_{d_j}^r \vee a_i^r > a_j^r \wedge a_i^r > a_{d_i}^r \wedge a_j^r > a_{d_j}^r \quad r_3 < r \le r_l
\end{array}
, \tag{11}
$$

Мерой отсутствия зависимости назовем отношение количества пар точек n, для которых условие (10) выполнено к количеству пар точек m, для которых выполнено условие (11)

$$\lambda_n = \frac{n}{m} \tag{12}$$

Рассмотрим теперь семейство Ф = *{C}*. Критерий для семейства *C* для случая одной переменной можно обобщить на случай функции двух переменных следующим образом. Мерой отсутствия зависимости назовем величину $\lambda_C = \frac{V_2 - V_1}{V_0}$, где V_0 – общий объем, V_1, V_2 – объемы, ограничиваемые снизу поверхностями, образованными точками, для которых хотя бы в одном из четырех параллелепипедов (построенных аналогично прямоугольникам для одномерного случая) нет ни одной точки (c_t, b_t, a_t).

Подобное обобщение неприменимо для случая l>2 в связи с вычислительной сложностью численного интегрирования. Рассмотрим семейство функций одной переменной *D*, которое имеет более простую конструкцию, чем семейство *C*. Рассмотрим семейство непрерывных функций с ограниченной в точках существования производной.

Простота конструкции семейства *D* позволяет создать следующий критерий – определим меру зависимости ряда A_k от рядов $A_1, A_2, ..., A_l$ следующим образом: рассмотрим пары точек $\left(a_i^1, ..., a_i^l\right)$ и $\left(a_j^1, ..., a_j^l\right)$, расстояние между которыми меньше некоторой константы δ. Мерой отсутствия зависимости назовем отношение количества $k_{\leq}$ тех пар, для которых $a_i^k - a_j^k \leq \varepsilon$ к количеству $k_{>}$ тех пар, для которых $a_i^k - a_j^k > \varepsilon$.

Недостатком выбора семейства *D* можно считать то, что в него входят функции с произвольно большим количеством локальных экстремумов, однако не входят, например, монотонные функции с отрезками быстрого возрастания.

Настойка пороговых значений μ_F для выбранных критериев

Пусть Φ = $\{_1, \varphi_2, ..., \varphi_s\}$ – выбранный набор семейств функций. Сопоставим каждому из семейств набора соответствующий критерий из множества К = $\{K^{\phi_1}, K^{\phi_2}, ..., K^{\phi_s}\}$.

Введем предикат F*(A, B)*, принимающий истинное значение тогда и только тогда, когда ряд *A* зависит от ряда *B*, т.е. b_t входит как аргумент в (3). Очевидно, что если ряд *A* зависит от ряда *B* , ряд *B* зависит от *C*, то ряд *A* зависит от ряда *C,* т.е.

$$\forall A \forall B \forall C (F(A,B) \cap F(B,C) \to F(A,C)) \qquad (13)$$

Другими словами, формула (6) означает, что если

$$a_t = f(b_t, ...) \text{ и } b_t = g(c_t, ...), \text{ то } a_t = f \circ g(c_t, ...) \qquad (14)$$

Если для любых $f \in \phi_i, g \in \phi_j$ известно, что $f \circ g$ принадлежит одному из семейств набора Φ, например, $f \circ g \in \phi_k$, то семейства ϕ_i, ϕ_j, ϕ_k имеют преимущества перед семействами, для которых указанное свойство не выполняется. Дело в том, что если $f \circ g$ не принадлежит ни одному из семейств, то для этой функции у нас нет критерия, проверяющего зависимость непосредственно, и сделать вывод о том, что ряд A зависит от ряда C можно лишь по косвенной информации. Если же указанное свойство выполнено, то для критериев K_i, K_j, K_k становиться необходимым следующее групповое условие: если с помощью критерия K_i установлено существование зависимости φ_i между рядами A и B и с помощью критерия K_j установлено существование зависимости φ_j между рядами B и C, то необходимо, чтобы с помощью критерия K_k была выявлена зависимость φ_k между рядами A и C.

$$\begin{gathered}\left(\forall f \in \phi_i \ \forall g \in \phi_i \ f \circ g \in \phi_k\right) \rightarrow \\ \rightarrow \forall A \forall B \forall C\left(K_{\mu}^{\phi_i}(A,B,\ldots) \cap K_{\mu}^{\phi_j}(B,C,\ldots) \rightarrow K_{\mu}^{\phi_k}(A,C,\ldots)\right),\end{gathered} \tag{15}$$

где $K_{\mu}^{\phi_i}(A,B,\ldots)$ обозначает тот факт, что с помощью критерия установлено, что ряд A зависит от ряда B.

Если $a_t = f(b_t,\ldots)$ и $a_t = f \circ g(c_t,\ldots)$, причем C_t не входит в f как параметр и b_t не входит в $f \circ g$ как параметр, то необходимо $b_t = g(c_t,\ldots)$. Для критериев сформулируем условие, аналогичное описанному выше.

Если с помощью критерия K_i установлено существование зависимости φ_1 между рядами А и B, причем C не входит в эту зависимость, и с помощью критерия K_k установлено существование зависимости φ_k между рядами A и C , причем B не входит в эту зависимость, то необходимо, чтобы с помощью критерия K_j была выявлена зависимость φ_j между рядами B и C.

$$\begin{gathered}\left(\forall f \in \phi_i \ \forall g \in \phi_i \ f \circ g \in \phi_k\right) \rightarrow \\ \rightarrow \forall A \forall B \forall C(K_{\mu}^{\phi_i}(A,B,\ldots) \cap K_{\mu}^{\phi_k}(A,C,\ldots) \cap \\ \cap \neg K_{\mu}^{\phi_i}(A,B,C,\ldots) \cap \neg K_{\mu}^{\phi_k}(A,B,C,\ldots) \rightarrow K_{\mu}^{\phi_j}(B,C,\ldots))\end{gathered} \tag{16}$$

При выполнении этих условий мы располагаем дополнительной информацией для настройки параметров критериев и проверки правильности их работы.

Вначале рассмотрим случай одной переменной. В этом случае второе условие для функций и для критериев можно записать в виде:

$$\begin{gathered}\forall A \forall B \forall C\left(F(A,B) \cap F(A,C) \rightarrow F(B,C)\right) \\ \left(\forall f \in \phi_i \ \forall g \in \phi_i \ f \circ g \in \phi_k\right) \rightarrow \\ \rightarrow \forall A \forall B \forall C\left(K_{\mu}^{\phi_i}(A,B) \cap K_{\mu}^{\phi_k}(A,C) \rightarrow K_{\mu}^{\phi_j}(B,C)\right)\end{gathered} \tag{17}$$

В общем случае для произвольного набор критериев $K = \{K_1, K_2,...,K_s\}$, для которых должны быть выполнены вышеописанные условия (15)-(16), настройка критериев заключается в выборе пороговых значений μ_{ϕ_i}, минимизирующих количество несоответствий данным условиям. Допустим, что получены оценки $\lambda_{\phi_i}(A,B), \lambda_{\phi_j}(B,C), \lambda_{\phi_k}(A,C)$, несоответствие может возникнуть в случае, если пороговые значения $\mu_{\phi_1}, \mu_{\phi_2},..., \mu_{\phi_s}$ для критериев выбраны так, что $\lambda_{\phi_i}(A,B) < \mu_{\phi_i} \wedge \lambda_{\phi_j}(B,C) < \mu_{\phi_j} \wedge \lambda_{\phi_k}(A,C) > \mu_{\phi_k}$ - несоответствие условию (15), либо $\lambda_{\phi_i}(A,B) < \mu_{\phi_i} \wedge \lambda_{\phi_k}(A,C) < \mu_{\phi_k} \wedge \lambda_{\phi_j}(B,C) > \mu_{\phi_j}$ - несоответствие условию (17).

Известно, что суперпозиция монотонных функций является монотонной функцией, суперпозиция монотонной и унимодальной функции является унимодальной функцией, однако суперпозиция унимодальных функций, вообще говоря, не входит ни в одно из выбранных семейств. Для набора Ф = $\{M, U\}$ для случая функции одной переменной можно составить таблицу

	M	**U**
M	M	U
U	U	?

Введем предикаты M (A, B), U (A, B), принимающие истинное значение тогда и только тогда, когда ряд A монотонно (унимодально) зависит от ряда B, т.е. b_t входит как аргумент в (3). Из вышесказанного следует, что

$$\forall A \forall B \forall C (M(A,B) \cap M(B,C) \rightarrow M(A,C)) \quad (18а)$$

$$\forall A \forall B \forall C (M(A,B) \cap U(B,C) \rightarrow U(A,C)) \quad (18б)$$

$$\forall A \forall B \forall C (U(A,B) \cap M(B,C) \rightarrow U(A,C)) \quad (18в)$$

$$\forall A \forall B \forall C (U(A,B) \cap U(A,C) \rightarrow M(B,C)) \quad (18г)$$

$$\forall A \forall B \forall C (M(A,B) \cap U(A,C) \rightarrow U(B,C)) \quad (18д)$$

$$\forall A \forall B \forall C (M(A,B) \cap M(A,C) \rightarrow M(B,C)) \quad (18е)$$

Условия (15) - (16) для критериев для определения монотонности и унимодальности, описанных в пункте 3, можно записать в виде

$$\forall A \forall B \forall C \left(K_{\mu}^{M}(A,B) \cap K_{\mu}^{M}(B,C) \rightarrow K_{\mu}^{M}(A,C) \right) \quad (19а)$$

$$\forall A \forall B \forall C \left(K_{\mu}^{M}(A,B) \cap K_{\mu}^{U}(B,C) \rightarrow K_{\mu}^{U}(A,C) \right) \quad (19б)$$

$$\forall A \forall B \forall C \left(K_{\mu}^{U}(A,B) \cap K_{\mu}^{M}(B,C) \rightarrow K_{\mu}^{U}(A,C) \right) \quad (19в)$$

$$\forall A \forall B \forall C\left(K_{\mu}^{U}(A,B) \cap K_{\mu}^{U}(A,C) \rightarrow K_{\mu}^{M}(B,C)\right) \quad (19г)$$

$$\forall A \forall B \forall C\left(K_{\mu}^{M}(A,B) \cap K_{\mu}^{U}(A,C) \rightarrow K_{\mu}^{U}(B,C)\right) \quad (19д)$$

$$\forall A \forall B \forall C\left(K_{\mu}^{M}(A,B) \cap K_{\mu}^{M}(A,C) \rightarrow K_{\mu}^{M}(B,C)\right) \quad (19е)$$

Покажем, что для описанных критериев для монотонной и унимодальной зависимости условия (19) не всегда выполняются. Рассмотрим, например, условие (19а) – суперпозицию $a_t = f \circ g(c_t)$ двух функций $a_t = f(b_t), b_t = g(c_t)$, к которым применяется критерий монотонности. Допустим, что количество пар точек, для которых выполняется условие немонотонности ($b_t<b_s \wedge a_t>a_s$ – для возрастающей функции, $b_t<b_s \wedge a_t<a_s$ – для убывающей) для рядов A и B равно n, а для рядов B и C равно k. Предположим, что обе зависимости возрастающие, т.е. $min\ (N_1, N_2) = N_1$ для формулы (8). Тогда количество пар точек (t, s), таких, что $b_t<b_s \wedge a_t<a_s$ равно $C_N^2 - n$. Среди n точек $b_t>b_s \wedge a_t<a_s$, может оказаться от 0 до $min(k,n)$ точек, для которых $b_t>b_s \wedge a_t<a_s \wedge c_t>c_s$. Среди $C_N^2 - n$ точек $b_t<b_s \wedge a_t<a_s$ может оказаться от 0 до $min(k,\ C_N^2 - n)$ точек, для которых $b_t<b_s \wedge a_t<a_s \wedge c_t>c_s$. Таким образом, общее число точек, таких, что $a_t<a_s \wedge c_t>c_s$ варьируется от 0 до $min\ (k,\ C_N^2 - n)+min\ (k,n)$, что больше k и больше n. Если выбрать значение μ_M так, что $\mu_M > \frac{k}{C_N^2}$ и $\mu_M > \frac{n}{C_N^2}$, но $\mu_M < \frac{min\left(k, C_N^2 - n\right) + min(k,n)}{C_N^2}$, то (19а) не выполняется. Аналогичное доказательство можно провести для остальных условий из (19).

Выводы

Для поиска зависимостей прогнозируемых рядов предложены виды семейств функциональных зависимостей: непрерывные функции с ограниченным числом экстремумов, монотонные функции и унимодальные. Для рассмотренных семейств предложены критерии поиска зависимостей прогнозируемых рядов. Также показан метод настройки параметров выбранных критериев.

Выбранные критерии распространены на случай функции многих переменных, т.е. случай зависимости между несколькими временными рядами.

Список литературы

1. Андерсон Т. Статистический анализ временных рядов. - М: Мир, 1976. - 523 с.
2. Брусиловский П.М., Гаев Л.В. Коллективные решения в задаче прогнозирования бинарных временных рядов // Изв. АН СССР. Сер. техн. киберн. - 1987. - 3. - С. 75-81.
3. Вапник В.Н. Восстановление зависимостей по эмпирическим данным. - М.: Наука, 1979. - 415 с.
4. Вайну Я.Я.-Ф. Корреляция рядов динамики. - М.: Статистика, 1977. - 119 с.
5. Newbold R., Grander C.W.J. Experience with Forecasting Univaiate Time Series and the Combination of Forecasts // J. R. Statist. Soc. A., - 1994. - V. 137, part 2. - P. 131-145.
6. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы - М.: Финансы и статистика, 1998 г.

Поступила в редакцію 18.10.2009 р. Рецензент: ктн., доц. Красічков О.О.

В.В. Поцепаєв[1], Р.В. Поцепаєв[2]
[1]Донецький національний технічний університет, [1]Технологічний центр Шлюмберже, Абінгдон, Великобританія

Вибір функціональних залежностей і критеріїв зв'язку часових рядів в задачах прогнозування. Розглядається задача прогнозування часових рядів, запропоновані критерії визначення функціональних залежностей між рядами відомими в області прогнозування і рядами, які прогнозуються для сімейств монотонних, унімодальних та нерозривних функцій, наведене налаштування параметрів обраних критеріїв.
прогнозування, часовий ряд, функціональна залежність, критерії

V.V. Potsepaev[1], R.V. Potsepaev[1]
[1]Donetsk National Technical University, [2]Technology center Shlumberge, Abingdon, United Kingdom

Selection of functional dependences and criteria for time series relationships in forecasting problems. The problem of time series forecasting is discussed; criteria for determining functional relationships between series known in the interval of forecasting and series subject to prediction are suggested for monotone, unimodal and continuous functions; the algorithm for the adjustment of parameters of selected criteria is proposed.
forecasting, time series, functional dependence, criteria