Математические структуры и моделирование
2018. № 1(45). С. 69–80

УДК 519.688
DOI: 10.25513/2222-8772.2018.1.69-80

АЛГОРИТМ РАСЧЁТА БЛОЧНОГО КРИТЕРИЯ ИНТЕРВАЛЬНОЙ ПРОГНОЗИРУЕМОСТИ ДИНАМИЧЕСКОГО ПОКАЗАТЕЛЯ НА ОСНОВЕ КОЭФФИЦИЕНТА ТАРСИТАНО-ЛОМБАРДО

Ю.М. Краковский[1]
д.т.н., профессор, e-mail: kum@stranzit.ru
А.Н. Лузгин[2]
к.т.н., преподаватель, e-mail: alexln@mail.ru

[1]Иркутский государственный университет путей сообщения, Иркутск, Россия
[2]Иркутский государственный университет, Иркутск, Россия

Аннотация. Предложен и протестирован алгоритм расчёта блочного критерия интервальной прогнозируемости динамического показателя на основе непараметрического коэффициента корреляции Тарситано-Ломбардо. Экспериментально показано, что значение блочного критерия интервальной прогнозируемости, рассчитанное по предыстории динамического показателя, позволяет количественно оценить целесообразность проведения интервального прогнозирования динамического показателя, исходя из его статистических свойств. Интервальное прогнозирование заключается в определении интервала (из двух заранее заданных), в котором будет находиться будущее значение показателя на основе оценок вероятностей этих событий. При этом разделительная граница интервалов задаётся расчётным способом, исходя из статистических свойств динамического показателя. Предложенный алгоритм расчёта блочного критерия интервальной прогнозируемости динамического показателя реализован на языке программирования «R».

Ключевые слова: интервальное прогнозирование, оценка интервальной прогнозируемости, динамический показатель, непараметрическая корреляция, точность интервального прогнозирования.

Введение

В последнее время все большее внимание исследователей уделяется технологиям прогнозирования различных динамических показателей (ДП), как важнейшему фактору, направленному на повышение эффективности принятия управленческих решений в организациях и предприятиях, осуществляющих свою деятельность в условиях неопределённости. Кроме того, наблюдается возрастающий интерес к вероятностным методам прогнозирования [1, 2]. Прежде всего, это объясняется тем, что вероятностные прогнозы позволяют получить количественную оценку неопределённости самого прогноза (этой оценкой является оценка вероятности будущего события). Как правило, такая задача решается с помощью машинного обучения (machine learning) [3].

Частным и достаточно распространённым случаем вероятностного прогнозирования является вероятностное прогнозирование бинарных исходов, когда в будущий момент времени может произойти только одно из двух возможных событий. Необходимость бинарного прогнозирования часто возникает во многих прикладных областях: метеорологической, энергетической, экономической, социальной, геологической. Так, например, в работе [4] используются данные по суммарному количеству осадков за один день в городе Рединге (Великобритания). В работе строится модель, на основе которой оцениваются вероятности того, будет или нет следующий день дождливым. Из двух оценок вероятностей выбирается максимальная и делается прогноз. Работа [5] посвящена вопросам бинарного прогнозирования цен на электроэнергию на материковой части Испании, а работа [6] — прогнозированию ценовых пиков на рынке электроэнергии Австралии. Применение бинарного прогнозирования в сфере экономики рассматривается в работе [7]. Более подробный обзор прикладных аспектов применения бинарного прогнозирования можно найти в работе [8].

Задачу бинарного прогнозирования можно решать с помощью различных методов прогнозирования. Например, методом логистической регрессии, методом Байеса, методом опорных векторов или случайных лесов [9–11].

Разновидностью бинарного прогнозирования является интервальное прогнозирование (ИП), описанное авторами в работах [12, 13]. Суть ИП заключается в определении интервала из двух заранее заданных интервалов, в котором будет находиться будущее значение ДП, на основе оценок вероятностей этих событий. При этом разделительная граница интервалов задаётся расчётным способом, исходя из статистических свойств ДП.

Задача ИП предполагает построение математических стохастических моделей. В работе [12] применялась вероятностная кластерная модель, а в работе [13] — вероятностная нейронная модель. При построении таких моделей и расчёте их параметров используется доступная предыстория значений ДП, которая непосредственно влияет на качество построенной модели и, как следствие, на точность ИП. Очевидно, что точность прогнозирования является важной характеристикой ИП.

На практике истинная точность ИП будущих значений ДП неизвестна, и для её оценки приходится применять методы кросс-валидации (cross-validation) по ретроспективным (уже известным) значениям ДП [14] (например, такой подход широко практикуется в соревнованиях Kaggle [15]). Если оценка точности ИП приемлема (неприемлема), то метод прогнозирования применять целесообразно (нецелесообразно). Формализация и определение приемлемой точности ИП даны в работе [16]. Главный недостаток здесь состоит в том, что, оценивая точность ИП по ретроспективным значениям ДП методами кросс-валидации, полностью игнорируются статистические характеристики ДП, что может приводить к неадекватным и неприемлемым результатам ИП новых значений этого ДП.

Возникает вопрос: можно ли по предыстории ДП количественно оценить целесообразность осуществления ИП, не осуществляя само ИП на основе соответствующих моделей? Решение данного вопроса подразумевает предварительное изучение (расчёт) некоторых статистических характеристик ДП и разработку некоторого критерия для оценки целесообразности применения ИП на основе соответствующих моделей. Вариант такого критерия под названием «критерий интервальной прогнозируемости» (КИП) был предложен в работе [17] на основе модифицированного рангового коэффициента корреляции Спирмена, с помощью которого оценивалась целесообразность (нецелесообразность) осуществления ИП на основе предложенных моделей [12,13]. При этом авторами была дана рекомендация осуществлять ИП ДП в текущий момент времени только в том случае, если рассчитанное по имеющейся выборке значение КИП не меньше 0,60. Однако в ходе последующих экспериментов [18] с использованием ДП с различными статистическими характеристиками на интервале тестирования были выявлены случаи, когда при значениях КИП значительно меньше 0,60 точность ИП была приемлемой (то есть значения КИП были неадекватны реальности). Для решения данной проблемы и улучшения точности оценок КИП различных ДП авторами был разработан модифицированный вариант, который получил название «блочный критерий интервальной прогнозируемости» (БКИП).

В данной работе описан алгоритм расчёта БКИП и проведено его сравнение с КИП по данным ДП с различными статистическими характеристиками на заданном интервале тестирования. При этом были заранее отобраны такие ДП, где значения КИП дают как адекватную, так и неадекватную характеристику целесообразности ИП. Для проверки точности ИП на заданном интервале тестирования использовалась модель вероятностной нейронной сети.

Учитывая, что алгоритм расчёта БКИП основан на непараметрическом коэффициенте корреляции, предложенном в 2013 году А. Тарситано и Р. Ломбардо (коэффициент корреляции Тарситано-Ломбардо) [19], рассмотрим его подробно.

1. Непараметрический коэффициент корреляции Тарситано-Ломбардо

Любой ДП формализуется как временной ряд:

$$Q = \{q_t; t \in T\}. \tag{1}$$

Здесь q_t – значения ДП, доступные в дискретные моменты времени t; $t \in T$; $T = \{0, ..., n-1\}$; n — количество доступных значений.

Выделим из Q два произвольных вектора с равным количеством элементов f и обозначим их так:

$$Q_i^f = \{q_i, ..., q_{i+f-1}\}, Q_j^f = \{q_j, ..., q_{j+f-1}\}. \tag{2}$$

Здесь $q_i, ..., q_{i+f-1}$ — элементы вектора Q_i^f; $q_j, ..., q_{j+f-1}$ — элементы вектора Q_j^f; индекс f — определяет размерность вектора; индексы $i = 0, ..., n-f$ и

$j = 0, ..., n - f$ — определяют индексы первого элемента каждого вектора в исходной выборке Q (1). При этом $i \neq j$.

Ранжируем элементы каждого вектора (2) в порядке возрастания и обозначим полученные векторы рангов так:

$$R_i^f = \{r_i, ..., r_{i+f-1}\}, R_j^f = \{r_j, ..., r_{j+f-1}\}. \tag{3}$$

Здесь $r_i, ..., r_{i+f-1}$ — ранги элементов вектора Q_i^f; $r_j, ..., r_{j+f-1}$ — ранги элементов вектора Q_j^f.

Далее создадим векторы анти-рангов элементов каждого вектора (2) так:

$$\begin{aligned} A_i^f &= \{a_i = f + 1 - r_i, ..., a_{i+f-1} = f + 1 - r_{i+f-1}\}, \\ A_j^f &= \{a_j = f + 1 - r_j, ..., a_{j+f-1} = f + 1 - r_{j+f-1}\}. \end{aligned} \tag{4}$$

Здесь a_{i+f-1} — анти-ранги элементов вектора Q_i^f; a_{j+f-1} — анти-ранги элементов вектора Q_j^f.

Введем следующую функцию:

$$\mu = \begin{cases} x, & x \geqslant 1; \\ x^{-1}, & x < 1. \end{cases} \tag{5}$$

Здесь x — положительное вещественное число.

Рассчитаем величины:

$$\begin{aligned} a(R_i^f, A_j^f) &= \frac{1}{f} \sum_{k=0}^{f-1} \mu\left(\frac{r_{i+k}}{a_{j+k}}\right), \ b(A_i^f, A_j^f) = \frac{1}{f} \sum_{k=0}^{f-1} \mu\left(\frac{a_{i+k}}{a_{j+k}}\right), \\ c(R_i^f, R_j^f) &= \frac{1}{f} \sum_{k=0}^{f-1} \mu\left(\frac{r_{i+k}}{r_{j+k}}\right), \ d(A_i^f, R_j^f) = \frac{1}{f} \sum_{k=0}^{f-1} \mu\left(\frac{a_{i+k}}{r_{j+k}}\right). \end{aligned} \tag{6}$$

Здесь r_{i+k} — элементы вектора R_i^f; r_{j+k} — элементы вектора R_j^f; a_{i+k} — элементы вектора A_i^f; a_{j+k} — элементы вектора A_j^f.

Дополнительно рассчитаем такую величину:

$$s = \frac{1}{n^2} \cdot \left(\sum_{k=1}^{[f/2]} \frac{n + 1 + k}{k} \right)^2 - 1. \tag{7}$$

Здесь $[\cdot]$ — операция взятия целой части вещественного числа.

Непараметрический коэффициент корреляции Тарситано-Ломбардо между двумя векторами Q_i^f и Q_j^f рассчитывается так:

$$r(Q_i^f, Q_j^f) = \frac{a(R_i^f, A_j^f) \cdot d(A_i^f, R_j^f) - b(A_i^f, A_j^f) \cdot c(R_i^f, R_j^f)}{s}. \tag{8}$$

Здесь $a(R_i^f, A_j^f)$, $b(A_i^f, A_j^f)$, $d(A_i^f, R_j^f)$, $c(R_i^f, R_j^f)$ — величины, рассчитанные по формуле (6); R_i^f — вектор рангов (3) элементов вектора Q_i^f; R_j^f — вектор рангов (3) элементов вектора Q_j^f; A_i^f — вектор анти-рангов (4) элементов вектора Q_i^f; A_j^f — вектор анти-рангов (4) элементов вектора Q_j^f; s — величина, рассчитанная по формуле (7).

Данный коэффициент в сравнении с другими подобными непараметрическими коэффициентами корреляции (например, Спирмена или Кенделла) обладает преимуществами, среди которых можно выделить [19]: 1) лучшую чувствительность к коррелированным выборкам различного объёма; 2) лучшую устойчивость к коррелированным выборкам при наличии выбросов. Поэтому он и был выбран авторами при расчётах БКИП.

2. Формализация блочного критерия интервальной прогнозируемости

Преобразуем исходную выборку ДП Q (1) методом конечных разностей первого порядка:

$$s_t = q_t - q_{t-1}, \tag{9}$$

где s_t — преобразованные значения ДП, q_t — исходные значения ДП (1), $t \in [1; n-1]$. После проведения процедуры преобразования обозначим полученный ДП так:

$$S = \{s_t; t \in \dot{T}\}. \tag{10}$$

Здесь t принимает значения из множества $\dot{T} = 1, ..., \dot{n}$; $\dot{n} = n - 1$ — объём выборки показателя S (10).

Выделим из S (10) следующий вектор размерностью f и назовём его базовым:

$$S^f_{\dot{n}-f+1} = \{s_{\dot{n}-f+1}, ..., s_{\dot{n}}\}. \tag{11}$$

Сформируем из S (10) новые выборки (блоки) объёмом m. Объём этих блоков определяется по формуле:

$$m = \left[\frac{\dot{n}}{[\dot{n}/z] + sign(\dot{n} - z \cdot [\dot{n}/z]}\right] - 1. \tag{12}$$

Здесь z — параметр, ограничивающий количество элементов в блоках так, что $m \leqslant z$; $\dot{n}$ — объём выборки S (10); $sign(\cdot)$– функция знака вещественного числа; $[\cdot]$ — операция взятия целой части вещественного числа.

Нами экспериментально было установлено, что лучшее значение для параметра z находится в диапазоне $450 - 550$ значений. В данной работе это значение было задано равным 500 и не менялось на протяжении всех дальнейших экспериментов.

Непосредственно формирование из исходной выборки S (10) блоков с количеством элементов m осуществляется так:

$$S_1 = \{s_1, ..., s_m\},\ S_2 = \{s_{1+g}, ..., s_{m+g}\}, ...,\ S_k = \{s_{1+(k-1)\cdot g}, ..., s_{m+(k-1)\cdot g}\}. \tag{13}$$

Здесь m — объём выборок, определённый по формуле (12); $g = [m/2]$; $k = 1 + [(\dot{n} - m - 1)/g]$ — число выборок; $[\cdot]$ — операция взятия целой части вещественного числа.

Следует отметить, что при объёме выборки S (10) $\dot{n} \leqslant z$ будет сформирована единственная выборка $S_1 = s_1, ..., s_m$, в которой $m = \dot{n} - 1$.

Произвольный вектор размерностью f, взятый из какой-либо выборки i (13), обозначим так:

$$S_{i,j}^f = \{s_{i,j}, ..., s_{i,j+f-1}\}. \qquad (14)$$

Здесь $s_{i,j}, ...s_{i,j+f-1}$ — элементы вектора $S_{i,j}^f$; индекс f — определяет размерность вектора; индекс $i = 1, ..., k$ – номер выборки (13); индекс $j = 1, ..., m-f+1$ — номер первого элемента вектора в выборке i.

Применение БКИП основывается на следующей предпосылке: если количественно оценить множественную взаимосвязь между базовым вектором $S_{\dot{n}-f+1}^f$ (11) различной размерности f и векторами $S_{i,j}^f$ (14) блоков (13) на основе предыстории S (10), то можно дать рекомендацию о целесообразности или нецелесообразности ИП показателя Q (1). Отсюда происходит и название критерия.

Взаимосвязь между базовым вектором $S_{\dot{n}-f+1}^f$ (11) и вектором $S_{i,j}^f$ (14) предлагается определять так:

$$\tilde{r}(S_{\dot{n}-f+1}^f, S_{i,j}^f) = \begin{cases} |r(S_{\dot{n}-f+1}^f, S_{i,j}^f)|, & f \neq 1; \\ 1, & f = 1. \end{cases} \qquad (15)$$

Здесь $r(S_{\dot{n}-f+1}^f, S_{i,j}^f)$ — непараметрический коэффициент корреляции Тарситано-Ломбардо (8).

БКИП рассчитывается следующим образом.

Фиксируется номер выборки i. Для фиксированного значения f и всех значений $j \in [1; m-f+1]$ выборки i рассчитывается множество значений $\tilde{r}(S_{\dot{n}-f+1}^f, S_{i,j}^f)$ (15), среди которых находится максимальное:

$$r_{i,max}^f = \max(R_i^f). \qquad (16)$$

Здесь $R_i^f = \{\tilde{r}(S_{\dot{n}-f+1}^f, S_{i,1}^f), ..., \tilde{r}(S_{\dot{n}-f+1}^f, S_{i,m-f+1}^f)\}$ — множество значений $\tilde{r}(S_{\dot{n}-f+1}^f, S_{i,j}^f)$ (15) для фиксированных f и i.

Величины (16) для различных $f = 1, .., f_{max}$ образуют множество $R_{i,max} = \{r_{i,max}^1, ..., r_{i,max}^{f_{max}}\}$, по которому находится медиана:

$$r_{i,med} = med(R_{i,max}). \qquad (17)$$

Здесь $med(\cdot)$ — медиана по множеству значений.

Значение f_{max} ограничивает максимальную размерность векторов, используемых при расчётах. Предварительно проведённые эксперименты показали, что для расчёта БКИП по выборкам объёма $m \geqslant 20$ достаточно использовать векторы с максимальной размерностью f_{max} равной:

$$f_{max} = [\pi \cdot \sqrt{m} + ln(m)], \qquad (18)$$

где m — объём блоков (13), $[\cdot]$ — операция взятия целой части вещественного числа.

Рассчитав множество значений $r_{i,med}$ для всех $i = 1, ..., k$, БКИП находится так:

$$u_{n-1} = med(R_{med}). \tag{19}$$

Здесь $R_{med} = r_{1,med}, ..., r_{k,med}$ — множество значений $r_{i,med}$ (17) для заданных значений i.

3. Алгоритм расчёта блочного критерия интервальной прогнозируемости

С учётом выше изложенного алгоритм БКИП в общем виде содержит следующие этапы:

1) подготовка исходных данных Q (1);

2) преобразование Q (1) методом конечных разностей (9) для получения S (10);

3) задание параметра z: 450<z<550;

4) вычисление количества элементов в блоках m (12) в зависимости от заданного z;

5) формирование блоков из исходной выборки S (10) с количеством элементов m;

6) вычисление максимальной допустимой размерности векторов f_{max} (18);

7) задание номера выборки $i = 1$ и пустого множества $R_{med} = \{\}$;

8) задание размерности векторов $f = 1$ и пустого множества $R_{i,max} = \{\}$;

9) выделение из S (10) базового вектора $S^f_{\dot{n}-f+1}$ (11);

10) задание $j = 1$ и пустого множества $R^f_i = \{\}$;

11) выделение из блока i вектора $S^f_{i,j}$ (14);

12) вычисление $\tilde{r}(S^f_{\dot{n}-f+1}, S^f_{i,j})$ (15) и добавление его во множество R^f_i;

13) если $j < m - f - 1$, то $j = j + 1$ и возвращаемся на шаг 11; иначе шаг 14;

14) вычисление $r_{i,max},^f$ (16) и добавление его во множество $R_{i,max}$;

15) если $f < f_{max}$, то $f = f + 1$ и возвращаемся на шаг 9; иначе шаг 16;

16) вычисление $r_{i,med}$ (17) и добавление его во множество R_{med};

17) если $i < k$, то $i = i + 1$ и возвращаемся на шаг 8; иначе шаг 18;

18) вычисление значения БКИП u_{n-1} (19).

4. Тестирование блочного критерия интервальной прогнозируемости

Алгоритмы расчёта КИП и БКИП реализованы на свободно распространяемом языке программирования для статистической обработки данных с открытым исходным кодом «R» [20]. Для проверки точности ИП на основе вероятностной нейронной сети был использован программный комплекс «Интервальное прогнозирование нестационарных динамических показателей» [21].

Были выбраны следующие ДП с различными статистическими характеристиками:

1) ежедневное количество новорождённых детей (CN) в городе Квебеке (Канада) [22];

2) ежедневное среднее значение давления над уровнем моря (SP) в городе Мадрасе (Индия)[23];

3) ежедневное среднее значение «возвратного» индекса (IR) акций американской корпорации IBM [24];

4) ежедневное среднее значение температуры (MT) окружающей среды озера Люцерн (Швейцария) [25];

5) значения, полученные с помощью генератора псевдослучайных чисел с нормальным законом распределения вероятностей при математическом ожидании равном 0 и дисперсии равной 1 (NR);

6) последовательность кумулятивных сумм значений показателя NR (CNR).

Исходный объём выборки каждого ДП был различен и равен $n = 500, 1000, 1500, 2000, 2500$ значений.

Точность ИП проверялась для w последних значений каждого ДП, при $w = [0.1 \cdot n]$ (интервал тестирования). Здесь $[\cdot]$ — операция взятия целой части вещественного числа. То есть для тестирования использовалось 10% значений от всего объёма взятой выборки. Для этих же самых значений рассчитывались значения КИП и БКИП. При этом предыдущие значения ДП в объёме равном $n - j, j = w, ..., 1$ использовались для построения и обучения вероятностной нейронной сети, в расчётах КИП и БКИП.

Для характеристики точности ИП использовался критерий:

$$pl = l/(l + e). \tag{20}$$

Здесь pl — доля оправдавшихся прогнозов; l — число оправдавшихся прогнозов; e — число ошибочных прогнозов; $0 \leqslant pl \leqslant 1$. pl — это мера точности ИП. Чем больше это значение — тем лучше. Мы будем считать [16], что точность ИП приемлема, если $pl > 0,6$ (21).

Для тестирования адекватности оценок КИП и БКИП использовались их средние значения на интервале тестирования:

$$\overline{k} = \left(\sum_{i=1}^{w} k_{n-w-1+i} \right) /w, \ \overline{u} = \left(\sum_{i=1}^{w} u_{n-w-1+i} \right) /w. \tag{21}$$

Здесь $k_{n-w-1+i}$ — значения КИП на интервале тестирования; $\overline{k}$ — среднее значение на интервале тестирования; $u_{n-w-1+i}$ — значения БКИП на интервале тестирования; $\overline{u}$ — среднее значение БКИП на интервале тестирования.

Напомним, что при осуществлении ИП на основе вероятностной нейронной сети необходимо задать следующие параметры: f — размерность обучающих векторов вероятностной нейронной сети; p — время упреждения; α — параметр, влияющий на границу интервалов при ИП. Эти параметры были заданы такими: $f = 5, \ p = 1, \ \alpha = 0.0$. Для алгоритма расчёта БКИП был установлен параметр $z = 500$.

В таблице 1 приведены средние значения КИП ($\overline{k}$), БКИП ($\overline{u}$) и значения pl (21).

Таблица 1. Результаты тестирования точности ИП, КИП и БКИП

ДП	n	500	1000	1500	2000	2500
CN	pl	0.75	0.75	0.77	0.75	0.75
	$\overline{k}$	0.62	0.64	0.63	0.62	0.61
	$\overline{u}$	0.84	0.84	0.83	0.81	0.81
SP	pl	0.74	0.78	0.79	0.80	0.80
	$\overline{k}$	0.81	0.79	0.79	0.78	0.78
	$\overline{u}$	0.87	0.87	0.87	0.86	0.85
IR	pl	0.60	0.62	0.66	0.67	0.69
	$\boldsymbol{\overline{k}}$	**0.23**	**0.15**	**0.15**	**0.13**	**0.12**
	$\overline{u}$	0.65	0.63	0.66	0.66	0.66
MT	pl	0.62	0.72	0.71	0.68	0.66
	$\boldsymbol{\overline{k}}$	**0.37**	**0.32**	**0.31**	**0.31**	**0.31**
	$\overline{u}$	0.64	0.65	0.64	0.63	0.64
NR	pl	0.66	0.68	0.67	0.66	0.66
	$\boldsymbol{\overline{k}}$	**0.19**	**0.17**	**0.14**	**0.12**	**0.10**
	$\overline{u}$	0.65	0.62	0.63	0.60	0.60
CNR	pl	0.40	0.44	0.45	0.45	0.48
	$\overline{k}$	0.21	0.14	0.12	0.10	0.10
	$\overline{u}$	0.56	0.57	0.55	0.57	0.57

Как видно из приведённой таблицы для показателей IR, MT и NR точность прогнозирования приемлема, однако значение КИП характеризует эти показатели как неприемлемые (значения выделены жирным шрифтом) с точки зрения осуществления ИП. Напротив, во всех случаях БКИП даёт адекватные оценки целесообразности прогнозирования ДП: там, где точность ИП менее $0,6$, значение БКИП также менее $0,6$ и наоборот. Также по данным таблицы 1 видно, что с увеличением среднего значения БКИП точность ИП повышается. Это позволяет предложить БКИП вместо КИП для проверки целесообразности или нецелесообразности ИП ДП как более точный и адекватный критерий. При этом как и для КИП рекомендация остаётся прежней и заключается в том, что значение БКИП должно быть не менее $0,6$. В противном случае проводить ИП нецелесообразно.

Заключение

В работе был предложен блочный критерий интервальной прогнозируемости динамического показателя на основе непараметрического коэффициента корреляции Тарситано-Ломбардо и соответствующий алгоритм его расчёта. Экспериментально показано, что значение блочного критерия интервальной прогнозируемости, рассчитанное по предыстории динамического показателя, позволяет количественно оценить, целесообразно или нет проводить интервальное прогнозирование динамического показателя, исходя из его статистических свойств. Рекомендовано осуществлять интервальное прогнозирование динамического показателя в текущий момент времени только в том случае, если рассчитанное по имеющейся выборке значение блочного критерия интервальной прогнозируемости не меньше 0,6. Рекомендовано на практике вместо обычного критерия интервальной прогнозируемости применять блочный критерий интервальной прогнозируемости как более точный и адекватный в своих оценках для различных динамических показателей.

Литература

1. Elliot G., Lieli R.P. Predicting binary outcomes // J Econom. 2013. No. 174. P. 15–26.
2. Lahiri K., Yang L. Forecasting binary outcomes // Handb Econ Forecast. 2013. No. 2. P. 1025–1106.
3. James G., Witten D., Hastie T., Tibshirani T. An Introduction to Statistical Learning with Applications in R. Springer Texts in Statistics, 2013. 426 p.
4. Little M.A. Generalized Linear Models for Site-Specific Density Forecasting of UK Daily Rainfall // Monthly Weather Review. 2009. No. 37(3). P. 1029–1045.
5. Anbazhagan S., Kumarappan N. Binary classification of day-ahead deregulated electricity market prices using neural networks // Power India Conference. 2013.
6. Eichler M., Grothe O., Manner H., Turk D. Models for short-term forecasting of spike occurrences in Australian electricity markets: a comparative study // Journal of Energy market. 2014. V. 7, No. 1. P. 55–81.

7. Nyberg H. Studies on binary time series models with applications to empirical macroeconomics and finance. 2010. URL: `https://helda.helsinki.fi/bitstream/handle/10138/23519/studieso.pdf?seguence=1` (дата обращения: 17.01.2017).
8. Краковский Ю.М., Лузгин А.Н. Прикладные аспекты применения интервального прогнозирования в системном анализе // Современные технологии. Системный анализ. Моделирование. 2017. N. 2(54). С. 115-121.
9. Crawford M. M., Ham J., Chen Y., Ghosh, J. Random forests of binary hierarchical classifiers for analysis of hyperspectral data // IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data. 2003. P. 337–345.
10. Ng A.Y., Jordan M. I. On Discriminative vs. Generative classifiers: A comparison of logistic regression and naive Bayes // Advances In Neural Information Processing Systems. 2002. URL: `https://ai.stanford.edu/~ang/papers/nips01-discriminativegenerative.pdf` (дата обращения: 06.05.2017).
11. Platt C.J. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. 1999. URL: `citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.1639&rep=rep1&type=pdf`(дата обращения: 12.03.2017).
12. Краковский Ю.М., Лузгин А.Н. Алгоритм интервального прогнозирования динамических показателей на основе робастной вероятностной кластерной модели // Наука и образование. 2016. No. 11. С. 113–126. URL: `http://technomag.neicon.ru/doc/849839.html` (дата обращения: 20.08.2017).
13. Краковский Ю.М., Лузгин А.Н. Алгоритм интервального прогнозирования динамических показателей на основе вероятностной нейросетевой модели // Современные технологии. Системный анализ. Моделирование. 2016. № 4(50). С. 126–132.
14. Arlot S., Celisse A. A survey of cross-validation procedures for model selection // Statistics Surveys. 2010. No. 4. P. 40–79.
15. Kaggle: Your Home for Data Science. URL: `https://www.kaggle.com/competitions` (дата обращения: 11.09.2017)..
16. Краковский Ю.М., Лузгин А.Н. Стохастический критерий оценки приемлемой точности вероятностного бинарного прогнозирования динамических показателей // Вестник ВГУ, серия: Системный анализ и информационные технологии. 2017. № 2. С. 98–104. URL: `http://www.vestnik.vsu.ru/pdf/analiz/2017/02/2017-02-15.pdf` (дата обращения: 17.09.2017).
17. Краковский Ю.М., Лузгин А.Н. Оценка прогнозируемости динамических показателей на основе коэффициента ранговой корреляции // Наука и образование. 2016. № 9. С. 60–73. URL: `http://technomag.neicon.ru/doc/845015.html` (дата обращения: 22.04.2017).
18. Лузгин А.Н. Предпосылки модификации коэффициента интервальной прогнозируемости // Actualscience. 2017. Т. 3, № 3. С. 162–163.
19. Tarsitano A., Lombardo R. A Coefficient of Correlation Based on Ratios of Ranks and Anti-ranks // J Econ and Statis. 2013. V. 233, No. 2. P. 206–224.
20. The R project of statistical computing. URL: `http://www.r-project.org` (дата обращения: 05.05.2017).
21. Лузгин А.Н. Программный комплекс «Интервальное прогнозирование нестационарных динамических показателей» // Свидетельство о государственной регистрации программы для ЭВМ №2015617751 от 22.07.2015.

22. DataMarket. URL: `https://datamarket.com/data/` (дата обращения: 15.04.2017).
23. Madras Monthly Sea Level, CRU. URL: `http://www.comp-engine.org/timeseries/time-series_data/data-11114/` (дата обращения: 15.04.2017).
24. Center for Research in Security Prices. URL: `http://www.crsp.com` (дата обращения: 15.04.2017).
25. Air Quality Data of Switzerland. URL: `https://cran.r-project.org/web/packages/SwissAir/` (дата обращения: 15.04.2017).

CALCULATION ALGORITHM OF A BLOCK CRITERION OF THE DYNAMIC INDICATOR INTERVAL FORECASTABILITY BASED ON TARSITANO-LOMBARDO'S COEFFICIENT

Y.M. Krakovsky[1]
Dr.Sc. (Eng.), Professor, e-mail: kum@stranzit.ru
A.N. Luzgin[2]
Ph.D. (Eng.), Instructor, e-mail: alexln@mail.ru

[1]Irkutsk State University of Railway Transport, Irkutsk, Russia
[2]Irkutsk State University, Irkutsk, Russia

Abstract. Calculation algorithm of a block criterion of the interval forecastability for dynamic indicator based on Tarsitano-Lombardo's nonparametric correlation coefficient was proposed and tested. It was experimentally shown that the value of the block criterion of the interval forecastability, which was calculated from prehistory of dynamic indicator values, makes it possible to quantitatively estimate the appropriateness of performing interval forecasting of the dynamic indicator based on its statistical properties. The interval forecasting is to determine the interval from two predetermined intervals in which the future value of the indicator will be located. The forecasting is based on probability estimates of these events. In this case, the separation division of the intervals is set by the calculation method based on statistical characteristics of the dynamic indicator. The proposed calculation algorithm of the block criterion of the dynamic indicator interval forecastability was implemented in the programming language "R".

Keywords: interval forecasting, interval predictability estimating, dynamic indicator, nonparametric correlation, interval forecasting accuracy.

Дата поступления в редакцию: 05.10.2017