

# ПОДХОД К ОЦЕНКЕ КАЧЕСТВА МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ НА ОСНОВЕ СТРОГО БИНАРНЫХ ДЕРЕВЬЕВ И МОДИФИЦИРОВАННОГО АЛГОРИТМА КЛОНАЛЬНОГО ОТБОРА

**Л.А. ДЕМИДОВА**

доктор технических наук, профессор кафедры вычислительной и прикладной математики, факультет вычислительной техники,  
Рязанский государственный радиотехнический университет

Адрес: 390000, г. Рязань, ул. Гагарина, д. 59/1

E-mail: liliya.demidova@rambler.ru

Рассматриваются модели краткосрочного прогнозирования коротких временных рядов на основе строго бинарных деревьев и модифицированного алгоритма клонального отбора, обеспечивающие повышение точности прогноза посредством поиска аналитических зависимостей, формируемых на основе антител, кодирующих строго бинарные деревья и адекватно описывающих известные значения временных рядов.

Антитело представляет собой символьную строку, элементы которой выбираются из трёх предварительно заданных символьных алфавитов: алфавита арифметических операций; алфавита функционалов и алфавита терминалов. Использование трех символьных алфавитов обеспечивает при реализации модифицированного алгоритма клонального отбора корректное преобразование в аналитические зависимости случайным образом формируемых антител, структура которых может быть описана с помощью строго бинарных деревьев. При кодировании антител на основе строго бинарных деревьев осуществляется последовательная запись в символьную строку всех узлов строго бинарного дерева, начиная слева направо и снизу вверх. При формировании аналитических зависимостей на основе антител используется рекурсивная процедура интерпретирования антител.

Модифицированный алгоритм клонального отбора относится к группе эволюционных алгоритмов, реализующих возможность одновременного поиска среди нескольких альтернативных вариантов решений и выбора лучших из них. Главной отличительной особенностью модифицированного алгоритма клонального отбора является применение механизмов клональной селекции, гипермутации и супрессии в ходе смены поколений популяций антител, используемых для формирования искомым аналитических зависимостей.

Предложен и исследован новый подход к оценке качества моделей прогнозирования на основе строго бинарных деревьев и модифицированного алгоритма клонального отбора. Показана целесообразность одновременного учета значений средней относительной ошибки прогнозирования и показателя несовпадения тенденций при вычислении аффинитета антител с целью оценки качества моделей прогнозирования, определяемых с использованием аналитических зависимостей, формируемых на основе строго бинарных деревьев. Рассматриваемый подход к оценке качества моделей прогнозирования позволяет при реализации модифицированного алгоритма клонального отбора исключить из дальнейшего рассмотрения модели прогнозирования, характеризующиеся большими значениями показателя несовпадения тенденций.

Предлагаемые модели прогнозирования позволяют существенно сократить время поиска аналитической зависимости, наилучшим образом описывающей известные значения коротких временных рядов, и могут быть рекомендованы для решения задач краткосрочного прогнозирования (на 1-3 шага вперед).

**Ключевые слова:** временной ряд, модель прогнозирования, строго бинарное дерево, модифицированный алгоритм клонального отбора, средняя относительная ошибка прогнозирования, показатель несовпадения тенденций, аффинитет.

**Цитирование:** Демидова Л.А. Подход к оценке качества моделей прогнозирования на основе строго бинарных деревьев и модифицированного алгоритма клонального отбора // Бизнес-информатика. 2015. № 1 (31). С.58–68.

## 1. Введение

**А**нализ коротких временных рядов (ВР), длина которых ограничена 20–30 значениями, играет важную роль в решении многих практических задач, например, задач, связанных с прогнозированием различных социально-экономических процессов [1–4].

В последние годы все чаще при разработке моделей прогнозирования коротких ВР используются различные технологии искусственного интеллекта, применение которых позволяет обеспечить получение адекватных оценок возможных изменений в поведении ВР на основе их известных значений и принятие соответствующих управленческих решений [5–13].

При применении традиционных технологий прогнозирования ВР, основанных, например, на экстраполяционных описаниях, приходится решать непростые вопросы, связанные с выбором адекватной модели прогнозирования и оценкой свободных параметров этой модели, в частности, степени полинома или порядка авторегрессии [1, 2].

Одна из перспективных технологий, которая может быть использована при разработке моделей прогнозирования коротких ВР, реализует применение искусственных иммунных систем (ИИС), основанных на принципах естественной иммунной системы и хорошо зарекомендовавших себя при решении широкого спектра прикладных задач [14–19]. Эффективность использования аппарата ИИС, в частности, алгоритма клонального отбора и механизмов супрессии антител в решении задач интерполяции, экстраполяции и прогнозирования, доказанная в работах зарубежных исследователей [15, 18], подтверждает перспективность применения ИИС для решения задачи прогнозирования коротких ВР.

Практика использования моделей прогнозирования, определяемых с применением аналитических зависимостей, формируемых на основе строго бинарных деревьев (СБД), полученных при реализации модифицированного алгоритма клонального отбора (МАКО), для прогнозирования тенденций рынка труда в России свидетельствует о целесообразности проведения дальнейших разработок в данном направлении [16, 20–23].

В большинстве случаев качество моделей прогнозирования коротких ВР оценивается с применением какого-либо одного показателя, например, средней относительной ошибки прогнозирования, значение которой должно быть минимизировано [5, 6, 11–13, 16, 20–23]. В то же время для коротких

ВР может быть выполнен расчет и других показателей качества моделей прогнозирования, таких как коэффициент детерминации, среднеквадратическая ошибка прогнозирования, минимальное и максимальное значения ошибки прогнозирования, коэффициент несовпадения тенденций и т.п. [24–26]. Задача одновременного учета значений таких показателей при оценке качества модели прогнозирования представляет собой задачу многокритериальной оптимизации, которая с той или иной долей успеха может быть решена только высококвалифицированным экспертом-аналитиком.

Одновременный учет даже каких-либо двух показателей качества моделей прогнозирования обеспечил бы существенное повышение точности прогнозирования коротких ВР. В частности, учет наряду со значением средней относительной ошибки прогнозирования значения показателя несовпадения тенденций позволил бы исключить из дальнейшего рассмотрения модели прогнозирования, имеющие «лучшие» (малые) значения средней относительной ошибки прогнозирования, но при этом характеризующиеся «плохим» (большим) значением показателя несовпадения тенденций.

Так как при реализации МАКО для оценки качества моделей прогнозирования на основе СБД с целью отбора «лучшей» модели применяется традиционный подход, основанный на минимизации значения средней относительной ошибки прогнозирования [16, 20–23], то можно сделать вывод об актуальности разработки нового подхода к оценке качества моделей прогнозирования на основе СБД, который позволил бы обеспечить одновременный учет значений средней относительной ошибки прогнозирования и показателя несовпадения тенденций при реализации МАКО.

## 2. Основные определения.

### Разработка модели прогнозирования

В работах [16, 20–23] для решения задачи прогнозирования коротких ВР  $d(t)$  предлагается использовать модели краткосрочного прогнозирования  $k$ -го порядка, получаемые при реализации МАКО, который позволяет при приемлемых временных затратах сформировать на основе СБД аналитическую зависимость, наилучшим образом описывающую известные значения ВР и обеспечивающую получение минимального значения аффинитета  $Aff$  (*affinity*) – средней относительной ошибки прогнозирования  $AFER$  (*Average Forecasting Error Rate*):

$$AFER = \frac{\sum_{t=k+1}^m |(f(t) - d(t))/d(t)|}{m - k} \cdot 100\%, \quad (1)$$

где  $f(t)$  и  $d(t)$  – предсказанное и реальное значения ВР для  $t$ -го отсчета времени;  $m$  – количество значений ВР (количество отсчетов времени).

При реализации МАКО возможные варианты аналитических зависимостей кодируются в виде антител  $Ab$ , которые должны осуществлять распознавание антигенов  $Ag$  – известных значений ВР. При этом в качестве «лучшего» антитела выбирается антитело  $Ab$ , обеспечивающее минимальное значение аффинитета  $Aff$  [16, 20–23].

Антитело  $Ab$  представляет собой символьную строку, элементы которой выбираются из трех предварительно заданных символьных алфавитов [20–23]: алфавита арифметических операций  $Operation = \{ '+', '-', '*', '/' \}$ , то есть операций сложения, вычитания, умножения и деления; алфавита функционалов  $Functional = \{ 'S', 'C', 'Q', 'L', 'E', '_' \}$ , в котором символы 'S', 'C', 'Q', 'L', 'E' соответствуют математическим функциям «синус», «косинус», «квадратный корень», «натуральный логарифм», «экспонента», а символ '\_' определяет отсутствие какой-либо математической функции; алфавита терминалов  $Terminal = \{ 'a', 'b', \dots, 'z', '@' \}$ , в котором символ '@' определяет некоторую константу, а символы 'a', 'b', ..., 'z' соответствуют аргументам искомой аналитической функции. Предполагается, что арифметические операции являются двухместными, а применение функционала должно предшествовать применению арифметической операции [20–23].

Использование трех символьных алфавитов обеспечивает при реализации МАКО корректное преобразование в аналитические зависимости случайным образом формируемых антител, структура которых может быть описана с помощью СБД [20–23], примеры которых приведены на рис. 1, где рядом с узлами показаны значения их уровней. При использовании почти полных СБД (ППСБД) [20, 22] (рис. 1б) удастся сформировать более сложные аналитические зависимости, чем в случае применения просто СБД (рис. 1а) [21, 23], и, как следствие, обеспечить получение меньших значений средней относительной ошибки прогнозирования  $AFER_L$  [22]. По определению, **строго бинарное дерево** (СБД) уровня  $L$  состоит только из узлов уровня  $l$  ( $l \leq L$ ), у которых степень (количество выходящих из узла ветвей) равна двум или нулю [27]. В контексте решаемой задачи использовались СБД уровня  $L$ , у которых каж-

дый правый узел  $l$ -го ( $l < L$ ) уровня является листом (концевым узлом, то есть узлом, чья степень равна нулю), а каждый левый узел  $l$ -го ( $l < L$ ) уровня имеет двух потомков (при этом левый и правый узлы  $L$ -го уровня являются концевыми). По определению, **почти полное СБД** (ППСБД) – это СБД, для которого существует неотрицательное целое  $l$  такое, что каждый лист в дереве имеет уровень  $l$  или  $l + 1$  и, если узел дерева имеет правого потомка уровня  $l + 1$ , тогда все его левые потомки, являющиеся листьями, также имеют уровень  $l + 1$  [27]. В контексте решаемой задачи использовались ППСБД уровня 3 (рис. 1б, рис. 3), которые применялись либо непосредственно для формирования аналитических зависимостей на основе антител длиной 18 символов ( $2 \cdot r$ , где  $r$  – количество узлов в ППСБД уровня 3;  $r = 9$ ), либо – для построения СБД более высокого уровня с целью последующего формирования аналитических зависимостей на основе антител большей длины. В последнем случае построение СБД осуществлялось посредством итерационного присоединения уже имеющегося СБД к новому узлу в качестве «левого» поддерева и присоединения к этому же узлу в качестве «правого» поддерева – СБД уровня 1. Сформированные таким образом СБД в дальнейшем условно рассматриваются как ППСБД.

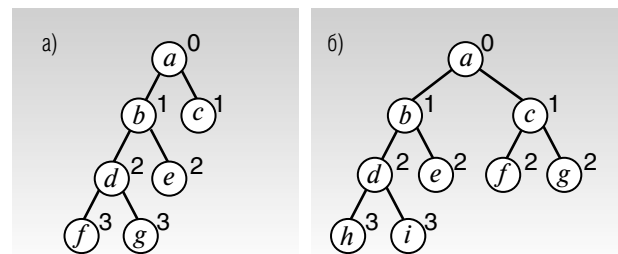


Рис. 1. Примеры бинарных деревьев:  
а) – строгое бинарное дерево уровня 3;  
б) – почти полное строгое бинарное дерево уровня 3

Количество терминальных позиций  $Term$  в антителе определяется максимально возможным порядком модели прогнозирования [20–23]. Если максимально возможный порядок равен  $K$ , то это означает, что при прогнозировании значения ВР  $d(t)$  на момент времени  $t$  могут использоваться  $K$  значений ВР:  $d(t - K), \dots, d(t - 2), d(t - 1)$ . При этом реальный и максимальный порядки модели прогнозирования  $k$  и  $K$  соответственно удовлетворяют условию:  $k \leq K$  (ввиду возможного кратного вхождения некоторых терминальных символов в запись аналитической зависимости и возможного наличия константы). Для антител на основе СБД (рис. 1а) при

формировании аналитических зависимостей произвольного максимального возможного порядка  $K$  может быть использована рекурсивная процедура интерпретирования антител [22], заключающаяся в рекурсивном преобразовании СБД в символьную строку посредством последовательной записи всех узлов, начиная слева направо и снизу вверх. Терминальные узлы (листья) СБД могут содержать только символы из алфавита терминалов *Terminal*, а остальные узлы формируются из символов алфавита арифметических операций *Operation* и алфавита функционалов *Functional*. При этом минимальный и максимальный номера позиций антитела, в которых стоят терминальные символы, равны  $2 \cdot Term$  и  $4 \cdot Term - 2$  соответственно, а длина антитела определяется максимальным номером позиции, в которой стоит терминальный символ.

Пример формирования антитела на основе СБД для случая, когда  $K = 3$ , что соответствует 3 концевым узлам (листьям) и алфавит терминалов имеет вид:  $Terminal = \{ 'a', 'b', 'c', '@' \}$ , приведен на рис. 2: в позициях 6, 8, 10 стоят символы из алфавита терминалов, в позициях 1, 3, 5, 7, 9 – из алфавита функционалов, а в позициях 2, 4 – из алфавита арифметических операций. Антитело  $Ab$ , сопоставляемое СБД, изображенному на рис. 2, и порожденная им аналитическая зависимость  $F(b, c)$  имеют соответственно вид:

$$Ab = ('L' '*' 'C' '-' '.' '@' 'E' 'b' 'S' 'c');$$

$$F(b, c) = \ln(\cos(\sin(c) - \exp(b)) * const),$$

где  $const$  – константа, соответствующая символу '@' и определяемая случайным образом.

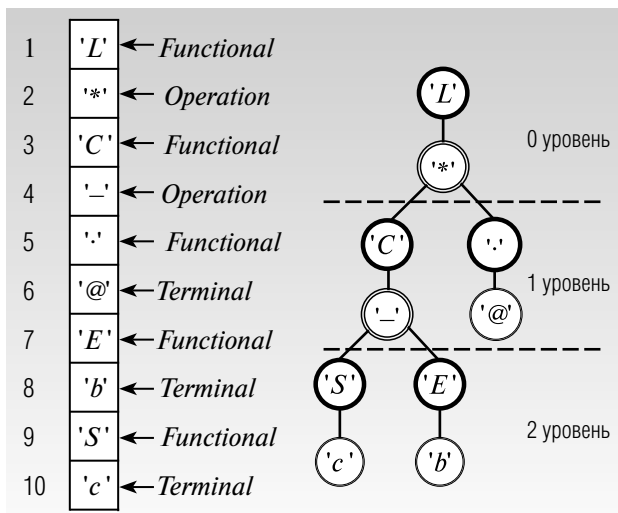


Рис. 2. Пример формирования антитела на основе СБД

Формирование антитела, соответствующего ППСБД (рис. 1б), может быть реализовано различными способами, обеспечивающими корректный результат интерпретирования антитела. В частности, могут быть использованы [20]: способ, основанный на применении стандартного подхода к обходу вершин бинарного дерева и реализующий формирование упорядоченного списка символов, находящихся в вершинах ППСБД; способ, основанный на разбиении ППСБД на поддеревья и реализующий сначала с применением стандартного подхода к обходу вершин бинарного дерева для каждого поддерева формирование упорядоченных списков символов, находящихся в его вершинах, а затем – последовательное объединение этих списков.

В [22] показано, что применение при реализации МАКО способа формирования антител, основанного на разбиении ППСБД на поддеревья, обеспечивает минимизацию временных затрат на интерпретацию антител в аналитические зависимости. Для антител на основе ППСБД может быть использована рекурсивная процедура интерпретирования антител [22], но ее применение в этом случае имеет ряд особенностей [20].

Так как для антител на основе ППСБД максимально возможный порядок  $K$  (количество концевых узлов – листьев) всегда является нечетным числом, то само ППСБД может быть представлено в виде композиции одного базового «левого» поддерева максимально возможного порядка  $K_L = 3$  (рис. 3) и некоторого количества  $n$  ( $n \geq 1$ ) базовых «правых»  $i$ -х ( $i = \overline{1, n}$ ) поддеревьев максимально возможного порядка  $K_n^i = 2$  (рис. 3). При этом

$$K_L + \sum_{i=1}^n K_n^i = K.$$

Базовое «левое» поддерево – это СБД максимально возможного порядка  $K_L = 3$ , являющееся одновременно СБД уровня 2 (рис. 3). Базовое «правое» поддерево – это СБД максимально возможного порядка  $K_n^i = 2$ , являющееся одновременно СБД уровня 1 (рис. 3). Формирование ППСБД максимально возможного порядка  $K$  осуществляется посредством итерационного (до достижения порядка  $K$ ) присоединения уже имеющегося СБД к новому узлу в качестве «левого» поддерева и присоединения к этому же узлу в качестве «правого» поддерева – базового «правого» поддерева (СБД уровня 1). При этом только на первом шаге итерационного формирования ППСБД «левое» поддерево является базовым, а на всех остальных шагах в качестве «лево-

го» используется СБД, полученное на предыдущем шаге.

Количество терминальных позиций в базовом «левом» и базовых «правых» поддеревьях равно соответственно  $Term_{л} = 3$  и  $Term_{п}^i = 2$  ( $i = 1, n$ ), а длина антитела равна  $10 + 8 \cdot n$  [22]. Процедура формирования антитела на основе ППСБД может быть реализована циклически в виде рекурсивного формирования частей антитела на основе соответствующих поддеревьев с последующей их композицией (слева направо и снизу вверх).

На рис. 3 приведен пример ППСБД для случая, когда  $K = 5$ , что соответствует 5 концевым узлам (листьям), и алфавит терминалов имеет вид:  $Terminal = \{ 'a', 'b', 'c', 'd', 'e', '@' \}$ . Длина антитела в этом случае находится как сумма длин «левого» поддерева ( $4 \cdot Term_{л} - 2 = 10$ ), «правого» поддерева ( $4 \cdot Term_{п} - 2 = 6$ ) и количества символов композиции, равного 2, то есть равна 18.

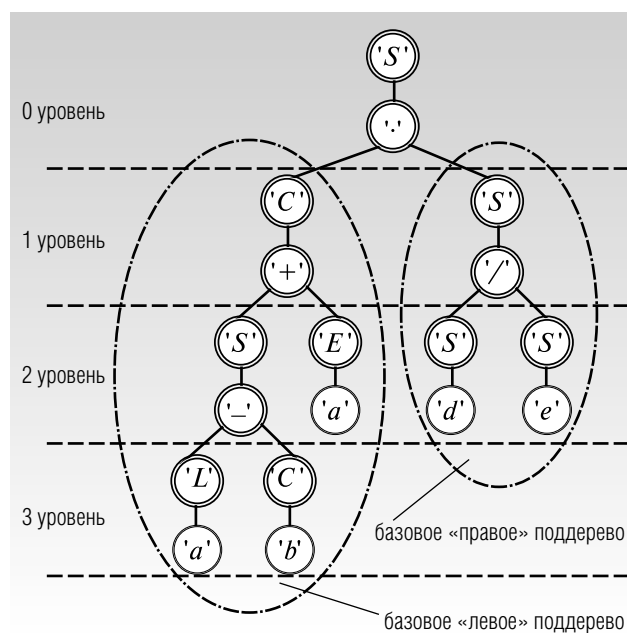


Рис. 3. Пример ППСБД, используемого для формирования антитела

Как уже было отмечено выше, качество антител и соответствующих им моделей прогнозирования, получаемых при реализации МАКО, оценивается по значению аффинитета  $Aff$ , который должен быть минимизирован. При этом обычно в качестве аффинитета  $Aff$  используется средняя относительная ошибка прогнозирования  $AFER$ , вычисляемая по формуле (1).

В тоже время особое внимание при отборе «лучших» антител и соответствующих им моделей прогнозирования следует уделять анализу значений

показателя несовпадения тенденций  $Tendency$ , который должен быть минимизирован:

$$Tendency = \frac{h}{m - k - 1}, \quad (2)$$

где  $h$  – количество отрицательных произведений  $(f(t-1) - f(t)) \cdot (d(t-1) - d(t))$  при  $t = k + 2, m$ ;  $f(t)$  и  $d(t)$  – предсказанное и реальное значения ВР для  $t$ -го отсчета времени;  $m$  – количество значений ВР (количество отсчетов времени);  $k$  – порядок модели;  $m - k - 1$  – общее количество произведений  $(f(t-1) - f(t)) \cdot (d(t-1) - d(t))$ .

Для одновременного учета при оценке качества антител и соответствующих им моделей прогнозирования наряду со значением средней относительной ошибки прогнозирования  $AFER$  значения показателя несовпадения тенденций  $Tendency$  предлагается вычислять аффинитет модели прогнозирования следующим образом:

$$Aff = AFER \cdot (1 + Tendency). \quad (3)$$

Представление формулы для вычисления аффинитета  $Aff$  в таком виде позволит обеспечить при реализации МАКО совместную одновременную минимизацию значений средней относительной ошибки прогнозирования  $AFER$  и показателя несовпадения тенденций  $Tendency$ .

Если для некоторой модели прогнозирования окажется, что  $AFER = 0$ , то этот факт будет свидетельствовать не только о том, значение средней относительной ошибки прогнозирования равно нулю, но и о полном совпадении тенденций ВР, при котором  $Tendency = 0$ . При этом значение аффинитета  $Aff$ , вычисленное по формуле (3), окажется минимально возможным. Если для некоторой модели прогнозирования окажется, что  $Tendency = 0$ , то этот факт будет свидетельствовать лишь о полном совпадении тенденций ВР. При этом значение средней относительной ошибки прогнозирования  $AFER$  может оказаться существенно далеким от нулевого. В связи с этим при вычислении аффинитета  $Aff$  антитела в формуле (3) учет значения показателя несовпадения тенденций  $Tendency$  реализован в виде множителя, представленного с помощью суммы  $(1 + Tendency)$ , что позволит при реализации МАКО в случае, когда  $Tendency = 0$ , продолжить поиск лучших вариантов моделей прогнозирования, имеющих меньшие значения средней относительной ошибки прогнозирования  $AFER$ .

Пусть модифицированный показатель несовпадения тенденций имеет вид:

$$Tendency_M = 1 + Tendency \quad (4)$$

В ходе отбора «лучших» антител и соответствующих им моделей прогнозирования при реализации МАКО предлагается считать, что антитело  $Ab_1$  лучше антитела  $Ab_2$ , если наряду с условием:

$$Aff_1 < Aff_2; \quad (5)$$

выполняются условия:

$$AFER_1 \leq AFER_2 \text{ и } Tendency_1 \leq Tendency_2. \quad (6)$$

При использовании такого подхода при реализации МАКО в качестве «лучших» антител в новом поколении МАКО будут полагаться антитела, имеющие значение аффинитета  $Aff$ , не больше, чем у «лучшего» антитела предыдущего поколения МАКО, и, кроме того, характеризующиеся неувеличением значения показателя несовпадения тенденций  $Tendency$  при одновременном неувеличении значения средней относительной ошибки прогнозирования  $AFER$ .

Для описанных выше вариантов формирования антител (на основе СБД и ППСБД) реализация МАКО осуществляется аналогичным образом. При реализации МАКО искомая аналитическая зависимость  $F$  кодируется в виде антитела  $Ab$ , которое должно обеспечивать распознавание антигенов  $Ag$  – элементов прогнозируемого ВР  $d(t)$  ( $t = 1, m$ ) – таким образом, чтобы осуществлялась минимизация аффинитета  $Aff(3)$  при соблюдении условий (5) и (6).

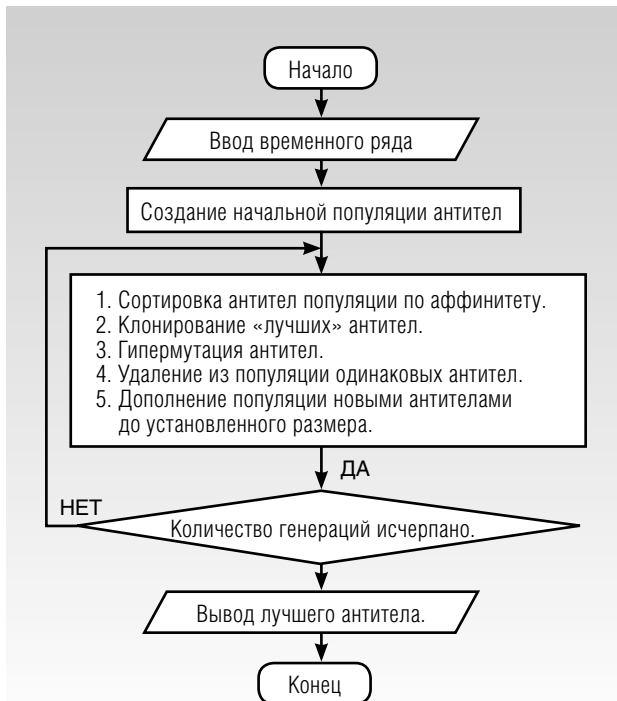


Рис. 4. Укрупненная схема МАКО

Укрупненная схема МАКО, используемого при разработке моделей прогнозирования  $k$ -го порядка, приведена на рис. 4.

МАКО относится к группе эволюционных алгоритмов [14], реализующих возможность одновременного поиска среди нескольких альтернативных вариантов решений и выбора лучших из них. Главной отличительной особенностью МАКО является использование механизмов клональной селекции (воспроизводства), гипермутации (изменения) и супрессии антител в ходе смены поколений популяций антител [18, 22]. При этом именно механизм супрессии, обеспечивающий удаление и уничтожение «похожих» антител, позволяет обходить локальные минимумы целевой функции (аффинитета) и расширять, при необходимости, область поиска решения.

МАКО является модификацией алгоритма клонального отбора, предложенного к применению для разработки моделей прогнозирования в работе [16], поскольку реализует корректное преобразование в аналитические зависимости антител, формируемых случайным образом на основе СБД, благодаря предложенному в [21, 22] и описанному выше подходу, основанному на использовании трёх символьных алфавитов для представления узлов СБД.

МАКО включает в себя подготовительную часть, реализующую формирование начальной популяции антител размером  $P$ , и итерационную часть, состоящую из следующих шагов: упорядочение антител по возрастанию значений аффинитета  $Aff$ ; отбор и клонирование определенной доли «лучших» антител, имеющих наименьшие значения аффинитета  $Aff$ ; гипермутация клонов антител; самоуничтожение клонов антител, «похожих» на другие клоны и антитела текущей популяции; вычисление аффинитета клонов антител и формирование новой популяции антител; супрессия полученной популяции; генерация новых антител и добавление их к текущей популяции до получения ее исходного размера  $P$ ; проверка условия достижения заданного количества поколений  $G$  и завершение работы МАКО при выполнении этого условия, а в противном случае – повтор итерационной части. Подробное описание МАКО дано в работе [22].

### 3. Экспериментальные исследования

Предлагаемый подход к оценке качества моделей прогнозирования на основе СБД, реализующий оценку аффинитета антител с помощью формулы (3) при соблюдении условий (5) и (6), был исполь-

зован при реализации МАКО в ходе решения задач краткосрочного прогнозирования тенденций рынка труда в России, в частности, для прогнозирования ВР, описывающего «экономически активное население» (в тыс. чел.). Одновременно в ходе сравнительного анализа было реализовано построение моделей прогнозирования на основе СБД и МАКО с применением традиционно используемой для оценки аффинитета антител средней относительной ошибки прогнозирования *AFER* (1).

При разработке моделей прогнозирования использовались 18 известных значений ВР  $d(t)$  ( $t = \overline{1,18}$ ) с февраля 1999 года по август 2003 года для отсчетов времени по месяцам: февраль, май, август и ноябрь [11], а оценка качества прогнозирования выполнялась для 3 значений ВР (ноябрь 2003 года, февраль 2004 года и май 2004 года), то есть на 3 шага вперед.

На рис. 5 приведены результаты прогнозирования на основе модели 4-го порядка, полученной при использовании предлагаемого (по формуле (3)) подхода к оценке качества моделей прогнозирования в процессе смены 2000 поколений антител. Значения средних относительных ошибок прогнозирования *AFER* и на три шага вперед составили соответственно 0,261 и 0,389, количество несовпадающих тенденций равно 0 как для 13 тенденций, используемых при построении модели прогнозирования, так и для тенденций при прогнозировании на три шага вперед, что свидетельствует о применимости разработанной модели для выполнения краткосрочного прогнозирования данного ВР.

При этом «лучшее» антитело имеет вид:

$$- + \_ * S - Q * Q / C - \_ c Q @ S f S d C g C f S c ,$$

а аналитическая зависимость, определяющая модель прогнозирования 4-го порядка, записывается соответственно как:

$$F(d(t-1), d(t-2), d(t-3), d(t-4)) = \sin(\sqrt{\cos(\sin(d(t-4))) - \cos(d(t-2))}) / \cos(d(t-1)) * \sin(d(t-3)) - \sin(d(t-2)) * \sqrt{1071900} + d(t-4).$$

Аналогичные результаты прогнозирования для рассматриваемого примера могут быть получены и при использовании традиционного (по формуле (1)) подхода к оценке качества моделей прогнозирования, однако их получение может сопровождаться хаотичными изменениями значений показателя несовпадения тенденций *Tendency* (2), поскольку поиск будет вестись только в направлении убывания значений средней относительной ошибки прогно-

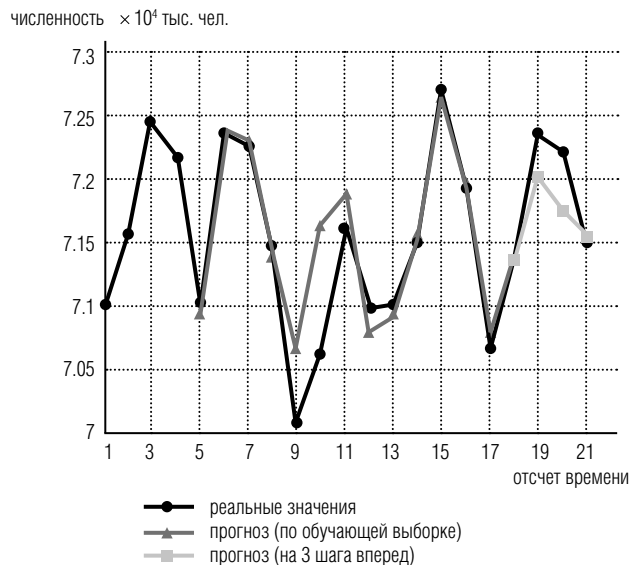


Рис. 5. Результаты прогнозирования

зирования *AFER* (1). Использование предлагаемого подхода в ходе смены поколений МАКО позволяет сузить область поиска искомого решения.

Так как использование МАКО не гарантирует нахождение оптимальной модели прогнозирования, то особое внимание должно уделяться вопросу признания модели прогнозирования в качестве «приемлемой» («субоптимальной»).

На рис. 6 и 7 приведены соответственно графические зависимости для средней относительной ошибки прогнозирования *AFER* (1) и модифицированного показателя несовпадения тенденций *Tendency<sub>M</sub>* (4), полученные при использовании предлагаемого (по формуле (3)) и традиционного (по формуле (1)) подходов к оценке качества моделей прогнозирования в процессе смены 1000 поколений одной реализации МАКО.

Как видно из рис. 6, в ходе реализации МАКО при использовании предлагаемого подхода к оценке качества моделей прогнозирования наблюдается постепенное уменьшение значений показателя *Tendency<sub>M</sub>* и, соответственно, количества несовпадающих тенденций «лучшего» антитела, в то время как при использовании традиционного подхода к оценке качества моделей прогнозирования возможно скачкообразное уменьшение и увеличение значений показателя *Tendency<sub>M</sub>* и, соответственно, количества несовпадающих тенденций «лучшего» антитела (рис. 7). При этом в обоих случаях наблюдается постепенное уменьшение значений средней относительной ошибки прогнозирования *AFER* «лучшего» антитела.

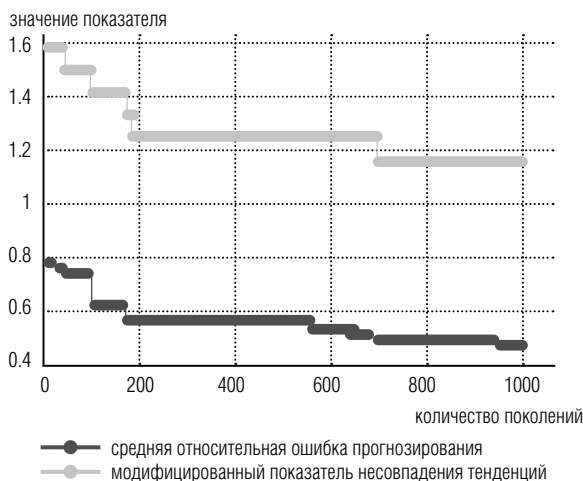


Рис. 6. Графические зависимости для показателей оценивания качества моделей прогнозирования при использовании предлагаемого подхода

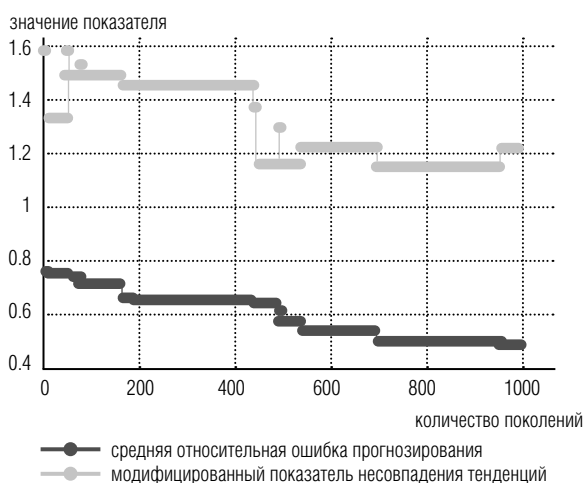


Рис. 7. Графические зависимости для показателей оценивания качества моделей прогнозирования при использовании традиционного подхода

В результате при применении традиционного подхода к оценке качества моделей прогнозирования «приемлемой» моделью может быть признана та, которая обладает меньшим значением средней относительной ошибки прогнозирования *AFER* (1), но при этом «плохим» значением показателя несовпадения тенденций *Tendency* (2).

Использование предлагаемого подхода к оценке качества моделей прогнозирования позволит сбалансировать значения средней относительной ошибки прогнозирования *AFER* (1) и показателя несовпадения тенденций *Tendency* (2). При этом удастся в ходе смены поколений МАКО исключить из дальнейшего рассмотрения антитела и соответствующие им модели прогнозирования с худшими значениями показателя несовпадения тенденций *Tendency* (2), поскольку в процессе смены поколений МАКО для антител будет запрещено увеличение значения показателя несовпадения тенденций *Tendency* (2).

#### 4. Заключение

Предложенный подход к оценке качества моделей прогнозирования на основе СБД и МАКО обеспечивает совместный одновременный учет «лучших» значений средней относительной ошибки прогнозирования *AFER* (1) и показателя несовпадения тенденций *Tendency* (2), позволяя при реализации МАКО исключить из дальнейшего рассмотрения модели прогнозирования, характеризующиеся большими значениями показателя несовпадения тенденций *Tendency*.

Получаемая на основе СБД и МАКО функциональная зависимость не является аналитической зависимостью исследуемого ВР, а представляет собой искусственно подобранную с использованием МАКО функцию, которая минимизирует введенный критерий качества (3), но, при этом, не претендует на содержательную интерпретацию по отношению к источнику ВР.

Модели прогнозирования на основе СБД и МАКО позволяют существенно сократить время поиска функциональной зависимости, наилучшим образом (с точки зрения одновременной минимизации средней относительной ошибки прогнозирования *AFER* и показателя несовпадения тенденций *Tendency*) описывающей известные значения коротких ВР, и могут быть рекомендованы для решения задач краткосрочного прогнозирования (на 1-3 шага вперед). ■

#### Литература

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: Юнити, 1998. 1022 с.
2. Андерсен Т. Статистический анализ временных рядов. М.: Мир, 1976. 756 с.
3. Mills T.C., Markellos R.N. The econometric modelling of financial time series. Cambridge: Cambridge University Press, 2008. 472 p.
4. Woodward W.A., Gray H.L., Elliott A.C. Applied time series analysis. Statistics: A series of textbooks and monographs. CRC Press, 2011. 564 p.



5. Chen S.M. Forecasting enrollments based on high-order fuzzy time series // *Cybernetic Systems*. 2002. Vol. 33. № 1. P. 1–16.
6. Hwang J.-R., Chen S.-M., Lee C.-H. Handling forecasting problems using fuzzy time series // *Fuzzy Sets and Systems*. 1998. Vol. 100. P. 217–228.
7. Cao L.J., Tay Francis E.H. Support vector machine with adaptive parameters in financial time series forecasting // *IEEE Transaction on Neural Networks*. Vol. 14. № 6. November 2003. P. 1506–1518.
8. Raicharoen T., Lursinsap C., Sanguanbhoki P. Application of critical support vector machine to time series prediction // *Circuits and Systems*, 2003. ISCAS '03. Proceedings of the 2003 International Symposium. 2003. Vol. 5. P. V-741–V-744.
9. Sah M., Degtiarev K.Y. Forecasting enrollment model based on first-order fuzzy time series // *Proceeding of International Conference on Computer Intelligence*. 2004. Vol. 1. P. 375–378.
10. Su S.F., Li S.H. Neural network based fusion of global and local information in prediction time series // *Proceedings of the 2003 IEEE International Joint Conference on Systems, Man and Cybernetics*. 2003. Vol. 5. P. 4445–4450.
11. Демидова Л.А. Разработка однофакторных нечетких моделей для анализа тенденций временных рядов с использованием генетического алгоритма // *Научно-технические ведомости СПбГПУ*. 2007. № 52 (2). С. 156–164.
12. Демидова Л.А. Прогнозирование тенденций временных рядов на основе однофакторной нечеткой модели с использованием дискретных нечетких множеств второго типа и генетического алгоритма // *Бизнес-информатика*. 2008. № 4 (6). С. 46–53.
13. Демидова Л.А., Скворцова Т.С. Двухфакторная модель прогнозирования временных рядов с короткой длиной актуальной части с использованием генетического алгоритма // *Системы управления и информационные технологии*. 2010. № 1 (39). С. 7–12.
14. Джонс М.Т. Программирование искусственного интеллекта в приложениях / Пер. с англ. А.И.Осипов. М.: ДМК Пресс, 2004. 312 с.
15. De Castro L.N., Von Zuben F.J. The clonal selection algorithm with engineering applications // *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'00), Workshop on Artificial Immune Systems and Their Applications*. Las Vegas, USA, July 2000. p. 36–37.
16. Бидюк П.И., Баклан И.В., Литвиненко В.И., Фефелов А.А. Алгоритм клонального отбора для прогнозирования нестационарных динамических систем // *Штучний інтелект*, 2004. № 4. С. 89–99.
17. Демидова Л.А., Титов С.Б. Исследование влияния основных параметров алгоритма функционирования искусственной иммунной сети на качество кластеризации объектов // *Вестник Рязанского государственного радиотехнического университета*. 2012. № 40. С. 54–60.
18. Искусственные иммунные системы и их применение / Под ред. Д.Дасгупты. М.: Физматлит, 2006. 344 с.
19. Кравец О.Я. Перспективные подходы к исследованию технологий менеджмента в сложных системах управления // *Экономика и менеджмент систем управления*. 2014. Т. 1, № 1 (11). С. 93–100.
20. Астахова Н.Н., Демидова Л.А. Использование почти полных строго бинарных деревьев и модифицированного алгоритма клонального отбора при разработке моделей прогнозирования временных рядов с короткой актуальной частью // *Вестник Рязанского государственного радиотехнического университета*. 2013. № 4-2 (46). С. 89–96.
21. Демидова Л.А., Корячко А.В., Скворцова Т.С. Модифицированный алгоритм клонального отбора для анализа временных рядов с короткой длиной актуальной части // *Системы управления и информационные технологии*. 2010. Т. 42, № 4.1. С. 131–136.
22. Демидова Л.А. Модели прогнозирования временных рядов с короткой актуальной частью на основе модифицированного алгоритма клонального отбора // *Вестник Рязанского государственного радиотехнического университета*. 2012. № 39-2. С. 64–71.
23. Демидова Л.А., Пылькин А.Н., Скворцов С.В., Скворцова Т.С. Гибридные модели прогнозирования коротких временных рядов. М.: Горячая линия – Телеком, 2012. 208 с.
24. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. СПб.: Питер, 2013. 704 с.
25. Белов В.В. Проблемы факторного прогнозирования социально-экономических показателей // *Вестник Московского государственного университета приборостроения и информатики*. 2005. № 2. С. 116.
26. Терехов А.А. Идентификация статистического материала и консолидация временных рядов // *Вестник Рязанского государственного радиотехнического университета*, 2009. № 27. С. 62–70.
27. Калмыкова О.В., Грибанов В.П., Сорока Р.И. Основы алгоритмизации и программирования: Учебное пособие. М.: МЭСИ, 2004. 220 с.

## AN APPROACH TO EVALUATION OF FORECASTING MODELS QUALITY USING STRICTLY BINARY TREES AND MODIFIED CLONAL SELECTION ALGORITHM

**Liliya DEMIDOVA**

Professor, Department of Computational and Applied Mathematics,  
Faculty of Computer Engineering,  
Ryazan State Radio Engineering University

Address: 59/1, Gagarina street, Ryazan, 390005, Russian Federation

E-mail: liliya.demidova@rambler.ru

Models of short-term forecasting of short-time series on the base of strictly binary trees and modified clonal selection algorithm are considered. These enable to increase forecast accuracy by selecting analytical dependences to be formed on the antibodies base, coding strictly binary trees and adequately describing known values of time series.

The antibody constitutes a symbolical line, which elements are selected from three preset symbolical alphabets: alphabet of arithmetic operations; alphabet of functionalities and alphabet of terminals. When implementing the modified clonal selection algorithm the use of three symbolical alphabets ensures correct transformation to analytical dependences of antibodies formed in a random way, which structure can be described by means of strictly binary trees.

When antibodies are coded on the base of strictly binary trees all knots of strictly binary tree are consecutively recorded in a symbolical line, beginning from left to right and from bottom to top. When analytical dependences are formed on the base of antibodies the recursive procedure of antibodies interpretation is applied.

The modified clonal selection algorithm belongs to a group of evolutionary algorithms, which enable to carry out simultaneous search among several decision alternatives to make the best choice. The main distinctive feature of the modified clonal selection algorithm is use of mechanisms of clonal selection, hypermutation and suppression during alternation of generations of antibodies populations, used to form required analytical dependences.

A new approach to quality estimation of forecasting models on the base of strictly binary trees and modified clonal selection algorithm has been offered and investigated. The paper has shown the expediency of simultaneous accounting of mean relative forecast error rate and tendencies discrepancy indicator in antibodies affinity calculations for the purpose of forecasting models quality estimation to be defined by involving analytical dependences, formed on the base of strictly binary trees. When applying the modified clonal selection algorithm the considered approach to forecasting models quality estimation enables to exclude from further consideration forecasting models, which are characterized by great values of tendencies discrepancy indicator.

The offered forecasting models enable to reduce significantly time needed to retrieve an analytical dependence, which gives the best description of short time series known values, and can be recommended to address short-term forecasting tasks (for 1–3 steps forward).

**Keywords:** time series, forecasting model, strictly binary tree, modified clonal selection algorithm, mean relative forecast error rate, tendencies discrepancy indicator, affinity.

**Citation:** Demidova L.A. (2015) Podhod k ocenke kachestva modelej prognozirovaniya na osnove strogo binarnykh derev'ev i modifitsirovannogo algoritma klonal'nogo otbora [An approach to evaluation of forecasting models quality using strictly binary trees and modified clonal selection algorithm]. *Business Informatics*, no. 1 (31), pp. 58–68 (in Russian).

### References

1. Ayvazyan S.A., Mkhitarian V.S. (1998) *Prikladnaya statistika i osnovy ekonometriki* [Applied statistics and fundamentals of econometrics]. Moscow: Uniti. (in Russian)
2. Andersen T. (1976) *Statisticheskii analiz vremennykh ryadov* [Time series' statistical analysis]. Moscow: Mir. (in Russian)
3. Mills T.C., Markellos R.N. (2008) *The econometric modelling of financial time series*. Cambridge: Cambridge University Press.
4. Woodward W.A., Gray H.L., Elliott A.C. (2011) *Applied time series analysis*. Statistics: A series of textbooks and monographs. CRC Press.

5. Chen S.M. (2002) Forecasting enrollments based on high-order fuzzy time series. *Cybernetic Systems*, vol. 33, no. 1, pp. 1–16.
6. Hwang J.-R., Chen S.-M., Lee C.-H. (1998) Handling forecasting problems using fuzzy time series. *Fuzzy Sets and Systems*, vol. 100, pp. 217–228.
7. Cao L.J., Tay Francis E.H. (2003) Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transaction on Neural Networks*, vol. 14, no. 6, pp. 1506–1518.
8. Raicharoen T., Lursinsap C., Sanguanbhoki P. (2003) Application of critical support vector machine to time series prediction. *Circuits and Systems*, 2003. ISCAS '03. Proceedings of the 2003 International Symposium, vol. 5, pp. V-741–V-744.
9. Sah M., Degtiarev K.Y. (2004) Forecasting enrollment model based on first-order fuzzy time series. *Proceeding of International Conference on Computer Intelligence*, vol. 1, pp. 375–378.
10. Su S.F., Li S.H. (2003) Neural network based fusion of global and local information in prediction time series. *Proceedings of the 2003 IEEE International Joint Conference on Systems, Man and Cybernetics*, vol. 5, pp. 4445–4450.
11. Demidova L.A. (2007) Razrabotka odnofaktornyh nechetkih modelei dlya analiza tendencii vremennyh ryadov s ispol'zovaniem geneticheskogo algoritma [Development of one-factorial fuzzy models for the analysis of time series' tendencies by means of genetic algorithm]. *Nauchno-tekhnicheskie vedomosti SPbGPU. Informatica. Upravlenie. Telekommunikatsii*, no. 52 (2), pp. 156–164. (in Russian)
12. Demidova L.A. (2008) Prognozirovaniye tendencii vremennyh ryadov na osnove odnofaktornoi nechetkoi modeli s ispol'zovaniem diskretnykh mnogestv vtorigo tipa i geneticheskogo algoritma [Tendencies' Forecasting of time series on the base of one-factor fuzzy model by means of discrete type-2 fuzzy sets and genetic algorithm]. *Business Informatics*, no. 4 (6), pp. 46–53. (in Russian)
13. Demidova L.A., Skvortsova T.S. (2010) Dvuhfaktornaya model' prognozirovaniya vremennyh ryadov s korotkoi dlinoi aktual'noi chasti s ispol'zovaniem geneticheskogo algoritma [Two-factorial forecasting model of time series with a short length of actual part by means of genetic algorithm]. *Sistemy upravleniya i informacionnye tehnologii*, no. 1 (39), pp. 7–12. (in Russian)
14. Dzhons M.T. (2004) *Programmirovaniye iskusstvennogo intellekta v prilozheniyah* [Programming of artificial intelligence in applications]. Moscow: DMK Press. (in Russian)
15. De Castro L.N., Von Zuben F.J. (2000) The clonal selection algorithm with engineering applications. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'00), Workshop on Artificial Immune Systems and Their Applications, Las Vegas, USA, July 2000*, pp. 36–37.
16. Biduk P.I., Baklan I.V., Litvinenko V.I., Fefelov A.A. (2004) Algoritm klonal'nogo otbora dlya prognozirovaniya nestacionarnykh dinamicheskikh sistem [Clonal selection algorithm for forecasting of non-stationary dynamic systems]. *Shtuchnii intellekt*, no. 4, pp. 89–99. (in Russian)
17. Demidova L.A., Titov S.B. (2012) Issledovaniye vliyaniya osnovnykh parametrov algoritma funkcionirovaniya iskusstvennoi immunnnoi seti na kachestvo klasterizatsii ob'ektov [Key parameters' influence research of an artificial immune network functioning algorithm on the objects' clusterization quality]. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, no. 40, pp. 54–60. (in Russian)
18. Dasgupta D., ed. (2006) *Iskusstvennyye immunnnyye seti i ih primeneniye* [Artificial immune systems and their application]. Moscow: Fizmatlit. (in Russian)
19. Kravets O. Ya. (2014) Perspektivnyye podhody k issledovaniyu tehnologii menedjmenta v slozhnykh sistemah upravleniya [Perspective approaches to research of management technologies in difficult control systems]. *Ekonomika i menedjment system upravleniya*, vol. 1, no. 1 (11), pp. 93–100. (in Russian)
20. Astakhova N. N., Demidova L.A. (2013) Ispol'zovanie pochty polnykh strogo binarnykh derev'ev i modifitsirovannogo algoritma klonal'nogo otbora pri razrabotke modelei prognozirovaniya vremennyh ryadov s korotkoi aktual'noi chast'yu [Using of the almost complete strictly binary trees and the modified clonal selection algorithm for the forecasting models' development of the time series with short actual part]. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, no. 4-2 (46), pp. 89–96. (in Russian)
21. Demidova L.A., Koryachko A.V., Skvortsova T.S. (2010) Modifitsirovannyi algoritm klonal'nogo otbora dlya analiza vremennyh ryadov s korotkoi dlinoi aktual'noi chasti [The modified clonal selection algorithm for the analysis of time series with a short length of actual part]. *Sistemy upravleniya i informacionnye tehnologii*, vol. 42, no. 4.1, pp. 131–136. (in Russian)
22. Demidova L.A. (2012) Modeli prognozirovaniya vremennyh ryadov s korotkoi aktual'noi chast'yu na osnove modifitsirovannogo algoritma klonal'nogo otbora [Forecasting models for time series with a short actual part on the base of modified clonal selection algorithm]. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, no. 39-2, pp. 64–71. (in Russian)
23. Demidova L.A., Pylkin A.N., Skvortsov S.V., Skvortsova T.S. (2012) *Gibridnyye modeli prognozirovaniya korotkikh vremennyh ryadov* [Hybrid forecasting models of short time series]. Moscow: Goryachaya liniya – Telekom. (in Russian)
24. Paklin N.B., Oreshkov V.I. (2013) *Bisnes-analitika: ot dannykh k znaniyam* [Business analytics: from data to knowledge]. St. Petersburg: Piter. (in Russian)
25. Belov V.V. (2005) Problemy faktornogo prognozirovaniya social'no-ekonomicheskikh pokazatelei [Factorial forecasting problems of socio-economic indexes]. *Vestnik Moskovskogo gosudarstvennogo universiteta priborostroeniya i informatiki*, no. 2, pp. 116. (in Russian)
26. Terekhov A.A. (2009) Identifikatsiya statisticheskogo materiala i konsolidatsiya vremennyh ryadov [Identification of a statistical material and time series' consolidation]. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*, no. 27, pp. 62–70. (in Russian)
27. Kalmykova O.V., Gribov V.P., Soroka P.I. (2004) *Osnovy algoritimizatsii i programmirovaniya: uchebnoe posobie* [Bases of algorithmization and programming: manual]. Moscow: MESI. (in Russian)