

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ,  
МОЛОДЕЖИ И СПОРТА УКРАИНЫ

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ  
УНИВЕРСИТЕТ РАДИОЭЛЕКТРОНИКИ

ISSN 0135-1710

# АВТОМАТИЗИРОВАННЫЕ СИСТЕМЫ УПРАВЛЕНИЯ И ПРИБОРЫ АВТОМАТИКИ

Всеукраинский межведомственный  
научно-технический сборник

Основан в 1965 г.

Выпуск 161

Харьков

2012

В сборнике представлены результаты исследований, касающихся компьютерной инженерии, управления, технической диагностики, автоматизации проектирования, оптимизированного использования компьютерных сетей и создания интеллектуальных экспертных систем. Предложены новые подходы, алгоритмы и их программная реализация в области автоматического управления сложными системами, оригинальные информационные технологии в науке, образовании, медицине.

Для преподавателей университетов, научных работников, специалистов, аспирантов.

У збірнику наведено результати досліджень, що стосуються комп'ютерної інженерії, управління, технічної діагностики, автоматизації проектування, оптимізованого використання комп'ютерних мереж і створення інтелектуальних експертних систем. Запропоновано нові підходи, алгоритми та їх програмна реалізація в області автоматичного управління складними системами, оригінальні інформаційні технології в науці, освіті, медицині.

Для викладачів університетів, науковців, фахівців, аспірантів.

Редакционная коллегия:

*В.В. Семенец, д-р техн. наук, проф. (гл. ред.); М.Ф. Бондаренко, д-р техн. наук, проф.; И.Д. Горбенко, д-р техн. наук, проф.; Е.П. Пуятин, д-р техн. наук, проф.; В.П. Тарасенко, д-р техн. наук, проф.; Г.И. Загарий, д-р техн. наук, проф.; Г.Ф. Кривуля, д-р техн. наук, проф.; Чумаченко С.В., д-р техн. наук, проф.; В.А. Филатов, д-р техн. наук, проф.; Е.В. Бодянский, д-р техн. наук, проф.; Э.Г. Петров, д-р техн. наук, проф.; В.Ф. Шостак, д-р техн. наук, проф.; В.М. Левыкин, д-р техн. наук, проф.; Е.И. Литвинова, д-р техн. наук, проф.; В.И. Хаханов, д-р техн. наук, проф. (отв. ред.).*

Свидетельство о государственной регистрации  
печатного средства массовой информации

КВ № 12073-944ПР от 07.12.2006 г.

*Адрес редакционной коллегии:* Украина, 61166, Харьков, просп. Ленина, 14, Харьковский национальный университет радиоэлектроники, комн. 321, тел. 70-21-326

© Харківський національний університет  
радіоелектроніки, 2012

## СОДЕРЖАНИЕ

<b>ХАХАНОВА И.В.</b> КВАНТОВЫЙ ПРОЦЕССОР ОПТИМАЛЬНОГО ПОКРЫТИЯ.....	4
<b>ХАХАНОВ В.И., ANDERS CARLSSON, ЧУМАЧЕНКО С.В., БУТЕНКО С.А.</b> МОДЕЛИ УПРАВЛЕНИЯ УЯЗВИМОСТЬЮ.....	10
<b>МИЗЬ В.А., ХАХАНОВА А.В.</b> АНАЛИЗ СИСТЕМ АВТОМАТИЗИРОВАННОГО МОНИТОРИНГА АВТОМОБИЛЬНОГО ТРАНСПОРТА И УПРАВЛЕНИЯ ДОРОЖНЫМ ДВИЖЕНИЕМ.....	25
<b>БОГОМОЛОВ В.А., ИЕВЛЕВА С.Н., РАЗНИЦЫН И.Л., СИДОРОВ М.В.</b> ЧИСЛЕННЫЙ АНАЛИЗ НАПРЯЖЕННО-ДЕФОРМИРОВАННОГО СОСТОЯНИЯ СЛОЯ ДОРОЖНОЙ ОДЕЖДЫ КАК ЛИНЕЙНОЙ И ГЕОМЕТРИЧЕСКИ-НЕЛИНЕЙНОЙ ВЯЗКОУПРУГОЙ СРЕДЫ НА ОСНОВАНИИ РЕОЛОГИЧЕСКОЙ ОДНОЭЛЕМЕНТНОЙ МОДЕЛИ КЕЛЬВИНА.....	31
<b>ЛЯШЕНКО С.А., ЛЯШЕНКО А.С.</b> ПОСТРОЕНИЕ ЛИНЕАРИЗИРОВАННЫХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ РАБОТЫ ВАКУУМ-АППАРАТОВ КРИСТАЛЛИЗАЦИОННОГО ОТДЕЛЕНИЯ САХАРНОГО ЗАВОДА.....	38
<b>БОЖИНСКИЙ И.А.</b> ОБ ОДНОМ ПОДХОДЕ К МОДУЛЬНОМУ ПРОЕКТИРОВАНИЮ ИНФОРМАЦИОННЫХ СИСТЕМ.....	41
<b>КОРАБЛЁВ Н.М., ФОМИЧЁВ А.А.</b> ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ИММУННЫХ МЕТОДОВ КЛАССИФИКАЦИИ ОБЪЕКТОВ С ПОМОЩЬЮ ЦЕЛЕВОГО КЛОНАЛЬНОГО ОТБОРА.....	45
<b>БЕРЗЛЕВ А.Ю.</b> РАЗРАБОТКА КОМБИНИРОВАННЫХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ С КЛАСТЕРИЗАЦИЕЙ ВРЕМЕННЫХ РЯДОВ ПО МЕТОДУ БЛИЖАЙШЕГО СОСЕДА.....	51
<b>НОВИКОВ Ю.С.</b> СТРУКТУРИЗАЦИЯ СОСТАВНЫХ ОБЪЕКТОВ ПРИ ФОРМИРОВАНИИ ПРОЦЕССНОГО ПРЕДСТАВЛЕНИЯ ЗНАНИЙ.....	59
<b>ЛИТВИН В.В., ГОПЯК М.Я., ДЕМЧУК А.Б.</b> МЕТОД АВТОМАТИЗОВАНОЇ РОЗБУДОВИ ТА ОЦІНЮВАННЯ ЯКОСТІ ОНТОЛОГІЙ БАЗ ЗНАНЬ.....	62
<b>ГВОЗДИНСКИЙ А.Н., ЯНОВ Д.М.</b> ИССЛЕДОВАНИЕ ЗАДАЧ ПРИНЯТИЯ РЕШЕНИЙ В СИСТЕМАХ УПРАВЛЕНИЯ ОГРАНИЧЕННЫМИ РЕСУРСАМИ.....	69
<b>ГУРИН В.Н., ФИРСОВ А.Г., ГУРИН Д.В.</b> МАТЕМАТИЧЕСКИЕ МОДЕЛИ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА НАПЫЛЕНИЯ НАНОСТРУКТУРИРОВАННЫХ ДИЭЛЕКТРИЧЕСКИХ ПЛЕНОК, ПОЛУЧЕННЫХ МЕТОДОМ ИОННО-ПЛАЗМЕННОГО РАСПЫЛЕНИЯ.....	74
<b>САКАЛО Е.С.</b> ИССЛЕДОВАНИЕ И ПОСТРОЕНИЕ ОБЛАЧНОЙ МОДЕЛИ УСЛУГ В CLOUD-СРЕДЕ.....	77
<b>ГУБНИЦКАЯ Ю.С.</b> ИНФОРМАЦИОННЫЕ МОДЕЛИ ДАННЫХ И КРИТЕРИИ ОЦЕНКИ КАЧЕСТВА КОМПОНОВКИ СТАТЕЙ НА ПОЛОСЕ ПРИ ДОПЕЧАТНОЙ ПОДГОТОВКЕ ИЗДАНИЙ.....	81
<b>ІЄВЛЄВА С.М.</b> ІНТЕРАКТИВНА СИСТЕМА МОДЕЛЮВАННЯ ТА ОПТИМІЗАЦІЇ РЕЖИМІВ РОБОТИ КОМПРЕСОРНОГО ЦЕХУ З УРАХУВАННЯМ РОБОТИ СИСТЕМИ АВТОМАТИЧНОГО УПРАВЛІННЯ.....	89
<b>ЧАЛЫЙ С.Ф., БОГАТОВ Е.О., МЕЛЕШКО Д.Г.</b> ТЕХНОЛОГИЯ ПРЕДВАРИТЕЛЬНОЙ СТРУКТУРИЗАЦИИ ЖУРНАЛОВ РЕГИСТРАЦИИ СОБЫТИЙ СЛАБОСТРУКТУРИРОВАННЫХ БИЗНЕС-ПРОЦЕССОВ.....	98
<b>ЧАЛЫЙ С.Ф., БУЦУКИНА И.Б.</b> МОДЕЛЬ БИЗНЕС-ПРОЦЕССА С ИЗМЕНЯЕМОЙ НА ОСНОВЕ ПРАВИЛ СТРУКТУРОЙ.....	103
<b>ЧАЛЫЙ С.Ф., АЛЬШЕЙХ АЛИ ДЖАМИЛЬ.</b> СЕРВИС-ОРИЕНТИРОВАННАЯ МОДЕЛЬ БИЗНЕС-ПРОЦЕССА С ИЗМЕНЯЕМОЙ СТРУКТУРОЙ.....	106
<b>ГВОЗДИНСКИЙ А.Н., ЯКИМОВА Н.А., ГУБИН В.А.</b> БИНАРНЫЕ ПРЕДИКАТЫ ПРИ ОПИСАНИИ БУЛЕВЫХ ЛОГИЧЕСКИХ ПРОСТРАНСТВ.....	108
<b>ИЕВЛЕВ Е.С.</b> О ВЫБОРЕ ЗАКОНА РАСПРЕДЕЛЕНИЯ ПРОДОЛЖИТЕЛЬНОСТИ ПЕРЕДАЧИ ПАКЕТИРОВАННЫХ ДАННЫХ В КОРПОРАТИВНЫХ КОМПЬЮТЕРНЫХ СЕТЯХ.....	113
<b>КИРИЧЕНКО Л.О., РАДИВИЛОВА Т.А., КАЙАЛИ Э.</b> РАСЧЕТ СТОИМОСТИ МАРШРУТИЗАЦИИ В СЕТИ MPLS С УЧЕТОМ ФРАКТАЛЬНЫХ СВОЙСТВ ТРАФИКА.....	116
<b>ОКСАНИЧ А.П., ПРИТЧИН С.Э., ТЕРБАН В.А.</b> УСОВЕРШЕНСТВОВАНИЕ МЕТОДА ИЗМЕРЕНИЯ ОСТАТОЧНЫХ НАПРЯЖЕНИЙ В ПОДЛОЖКАХ АРСЕНИДА ГАЛЛИЯ.....	122
<b>РЕФЕРАТЫ.....</b>	129
<b>ПРАВИЛА ОФОРМЛЕНИЯ СТАТЕЙ ДЛЯ АВТОРОВ НАУЧНО-ТЕХНИЧЕСКОГО СБОРНИКА.....</b>	136

## РАЗРАБОТКА КОМБИНИРОВАННЫХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ С КЛАСТЕРИЗАЦИЕЙ ВРЕМЕННЫХ РЯДОВ ПО МЕТОДУ БЛИЖАЙШЕГО СОСЕДА

Рассматривается проблема разработки комбинированных моделей прогнозирования временных рядов. Предлагаются комбинированные модели прогнозирования селективного и гибридного типов с предварительной кластеризацией временных рядов по методу ближайшего соседа. Предложенные модели могут быть составными информационных прогнозных систем, а также использоваться для повышения точности и эффективности управления процессом прогнозирования в части учета преимуществ конкурирующих прогнозов.

### 1. Введение

Принято считать, что на сегодняшний день развитие экономических систем характеризуется высокой степенью изменчивости, которое происходит в условиях неопределенности, неустойчивости и риска. Предпосылкой этого являются тенденции глобализации, усложнение хозяйственных взаимосвязей, рост темпов развития национальных рынков и т.д. В условиях нестабильности экономических рынков применение классических статистических прогнозных моделей и соответствующих методов прогнозирования временных рядов, которые представляют собой экономические показатели, довольно ограничено. Это объясняется также тем, что подавляющее большинство временных рядов не только экономической природы, для которых возникает задача прогнозирования, характеризуется нелинейностью, нестационарностью и неустойчивостью относительно среднего уровня. Многие прогнозные модели, в частности экспоненциальные [1,2], линейные регрессионные [3,4], авторегрессионные типа ARIMA [5] не предназначены для прогнозирования таких временных рядов, так как ограничены требованиями стационарности, нормальности и независимости остатков рядов. Модификации приведенных моделей для прогнозирования нелинейных процессов также имеют недостатки. Нелинейная регрессионная модель характеризуется сложностью идентификации функциональных зависимостей, а применение модели ARIMAX ограничено сложностью расчета многочисленных параметров. Все это обуславливает необходимость разработки новых и модификации известных методов и моделей прогнозирования временных рядов. Для их прогнозирования на финансовом рынке такие модели должны характеризоваться необходимой точностью, гибкостью и прозрачностью в оценке параметров, должны экономить время и преодолевать ограничения, наложенные на классические статистические модели, а также своевременно реагировать на изменения в динамике развития рынка.

В последнее время в области разработки систем прогнозирования сформировались два основных направления исследований:

- создание адаптивных комбинированных и многоуровневых моделей прогнозирования [4-6];
- разработка и использование методов интеллектуального анализа временных рядов [8-10].

Комбинированные модели применяются для повышения точности прогнозирования. Целью управления процессом прогнозирования комбинированных моделей является учет в оценке прогноза особых преимуществ и «полезных» характеристик каждой из прогнозных моделей, которые составляют их основное множество, т.е. отобранных для построения прогноза на определенном шаге. Недостатками комбинированных моделей являются сложность разработки и необходимость компенсации недостатков каждой из моделей основного множества, при этом не теряя их преимуществ.

На сегодняшний день также пользуется популярностью интеллектуальный анализ временных рядов или Time-Series Data Mining, который представляет собой набор инновацион-



ных методов, предназначенных для идентификации скрытых или ассоциативных правил в данных временных рядов, методы нелинейной динамики [11], оптимизации на основе генетических алгоритмов и программирования генетических выражений [12] и т.д. С момента возникновения интеллектуального анализа временных рядов разработано множество специальных методов, которые касаются задач индексации, кластеризации, классификации [8-10], идентификации выбросов и т.п. В последние годы появляется понимание того, что для многих задач интеллектуального анализа данных необходимо применять не быстрый и приблизительный, а медленный, но точный и подробный поиск [13], предлагается использовать методы, которые могут быть максимально легко практически применимы.

Автором предложен подход, который базируется на применении в комплексе адаптивных комбинированных моделей прогнозирования и кластеризации временных рядов по методу ближайшего соседа.

Актуальность этих исследований имеет не только практическое значение для прогнозистов, инвесторов и аналитиков, но и теоретическое: разработка концепций и математического инструментария, которые могут составить основу для дальнейших научных исследований.

Целью является рассмотрение основных понятий кластеризации временных рядов, способов представлений кластеров и формул расчета мер близости между ними, а также разработка адаптивных комбинированных моделей прогнозирования, механизмы селекции и гибридизации в которых корректируются на основе предварительной кластеризации временных рядов по методу ближайшего соседа.

## 2. Постановка задачи прогнозирования временных рядов

Пусть  $S$  — дискретное множество. Дискретным временным рядом

$$\{z_i\}_{i=1}^n = \{z_1, z_2, \dots, z_n\} = \{z(t_1), z(t_2), \dots, z(t_n)\}$$

будем называть конечную последовательность измерений, проводимых в дискретные моменты времени  $t_i \in S$ ,  $i = \overline{1, n}$  и фиксирующих определенные характеристики изменения состояний исследуемого процесса или явления,  $t_1$  — начальный момент времени. Будем считать, что измерения проводятся через определенные временные интервалы: минута, час, день, неделя, месяц, квартал и т.д. Результаты измерений  $z_i$  действительны и могут представлять собой однородные экономические показатели (уровень инфляции, объем выпущенной продукции), финансовые показатели и т.д. В эконометрике часто рассматривают временной ряд как частичную реализацию стохастического процесса, который представляется бесконечной случайной последовательностью. Но в отличие от элементов случайной выборки, значения временного ряда не являются независимыми и одинаково распределенными.

Выделяют следующие основные задачи прогнозирования временных рядов: прогнозирование будущих значений, прогнозирования знаков приростов временных рядов, идентификация моментов локальных экстремумов. В данной работе будем ограничиваться первой задачей. Сформулируем ее так: на основе ретроспективных значений  $z_n, z_{n-1}, z_{n-m+1}$ ,  $m \leq n$  ряда  $\{z_i\}_{i=1}^n$ , наиболее точно оценить его поведение в будущем в моменты  $t_{n+1}, t_{n+2}, \dots, t_{n+\theta}$ , т.е. построить последовательность прогнозных значений

$$\{z_i\}_{i=n+1}^{n+\theta} = \{z_{n+1}, z_{n+2}, \dots, z_{n+\theta}\},$$

где  $\theta$  — горизонт прогнозирования, а  $m$  — объем ретроспективной выборки. Обозначим через  $\hat{z}_\tau(n)$  — прогноз, который рассчитывается в момент  $t_n$  (в точке  $n$ ) на  $\tau$  точек вперед,  $\tau = \overline{1, \theta}$ .

Такая функциональная зависимость  $f$ , которая позволяет описать поведение временного ряда, называется моделью прогнозирования. Прогноз временного ряда  $\{z_i\}_{i=1}^n$  на одну точку вперед можно формально записать в виде  $\hat{z}_{n+1} = \hat{z}_1(n) = f(z_{n-m+1}, z_{n-m+2}, \dots, z_n)$ . В случае прогнозирования с горизонтом  $\theta > 1$  применяется итерационный подход:

$$\hat{z}_{n+1} = \hat{z}_1(n) = f(z_{n-m+1}, z_{n-m+2}, \dots, z_n),$$

$$z_{n+2} = \hat{z}_2(n) = f(z_{n-m+2}, z_{n-m+3}, \dots, z_n, z_{n+1}),$$

⋮

$$z_{n+\theta} = \hat{z}_\theta(n) = f(z_{n-m+\theta}, \dots, z_{n+\theta-2}, z_{n+\theta-1}), \quad m \leq n.$$

Необходимо отметить, что разные модели прогнозирования могут иметь различные механизмы расчета прогнозов, к примеру, в них могут учитываться внешние факторы, которые определенным образом влияют на моделируемый процесс. Кроме того, модель может иметь ряд параметров, требующих отдельного оценивания. К таким моделям относятся регрессионные, авторегрессионные, экспоненциальные, адаптивные полиномиальные модели [3,4]. Еще одной важной характеристикой каждой из моделей является объем ретроспективной выборки, используемой для реализации прогнозов.

Построить наиболее точный прогноз означает построить такую модель, которая удовлетворяет соответствующие критерии оценки качества прогнозирования. Для данной задачи такими критериями могут быть среднее абсолютное отклонение, средняя квадратичная погрешность, стандартное отклонение, относительная погрешность и т.д.

### 3. Основные понятия кластеризации временных рядов

Пусть задан временной ряд  $\{z_i\}_{i=1}^n$  как конечная последовательность действительных чисел. Кластером длины  $m$  временного ряда  $\{z_i\}_{i=1}^n$  будем называть подпоследовательность  $\{z_{k_j}\}_{j=1}^m$  данного ряда с  $m$  элементов,  $n > m$ ,  $k_{j+1} = k_j + 1$  для  $j = \overline{1, m-1}$ . Иными словами, кластеры любого временного ряда представляют собой его ретроспективные подпоследовательности с  $m$  элементов, причем порядок следования элементов в них такой, как и во временном ряду. Для задания кластера необходимо указать индекс начального элемента ряда и длину. Будем обозначать кластеры через  $z_{(m)}^N$ , где  $N$  – индекс начального элемента ряда, включенного в кластер, или порядковый номер, а  $m$  – длина. Тогда кластеры временного ряда  $\{z_i\}_{i=1}^n$  можно определить таким образом:

$$(1, m) - \text{кластер} - z_{(m)}^1 = \{z_{k_1^1}, z_{k_2^1}, \dots, z_{k_m^1}\} = \{z_{k_j^1}\}_{j=1}^m, \quad k_{j+1}^1 = k_j^1 + 1, \quad j = \overline{1, m-1}, \quad z_{k_1^1} = z_1,$$

$$(2, m) - \text{кластер} - z_{(m)}^2 = \{z_{k_1^2}, z_{k_2^2}, \dots, z_{k_m^2}\} = \{z_{k_j^2}\}_{j=1}^m, \quad k_{j+1}^2 = k_j^2 + 1, \quad j = \overline{1, m-1}, \quad z_{k_1^2} = z_2,$$

⋮

$$(n-m-1, m) - \text{кластер} - z_{(m)}^{n-m-1} = \{z_{k_1^{n-m-1}}, z_{k_2^{n-m-1}}, \dots, z_{k_m^{n-m-1}}\} = \{z_{k_j^{n-m-1}}\}_{j=1}^m, \\ k_{j+1}^{n-m-1} = k_j^{n-m-1} + 1, \quad j = \overline{1, m-1}, \quad z_{k_1^{n-m-1}} = z_{n-m-1},$$

$$(n-m, m) - \text{кластер} - z_{(m)}^{n-m} = \{z_{k_1^{n-m}}, z_{k_2^{n-m}}, \dots, z_{k_m^{n-m}}\} = \{z_{k_j^{n-m}}\}_{j=1}^m, \quad k_{j+1}^{n-m} = k_j^{n-m} + 1, \\ j = \overline{1, m-1}, \quad z_{k_1^{n-m}} = z_{n-m}, \text{ т.е. } (s, m) - \text{кластер для } s = \overline{1, n-m} \text{ в целом определяется так:}$$

$$z_{(m)}^s = \{z_{k_1^s}, z_{k_2^s}, \dots, z_{k_m^s}\} = \{z_{k_j^s}\}_{j=1}^m, \quad k_{j+1}^s = k_j^s + 1, \quad j = \overline{1, m-1}, \quad z_{k_1^s} = z_s, \quad (1)$$

$(n-m, m)$  – кластер ряда  $\{z_i\}_{i=1}^n$  будем называть опорным. Все другие кластеры будем называть неопорными.

Число неопорных кластеров длины  $m$ , построенных на основе временного ряда с  $n$  элементов, равно  $n-m$ .

Следует отметить, что существует некоторая неопределенность в терминологии. Термин кластер (pattern) временного ряда, которым в дальнейшем будем пользоваться, применяется в работе [9]. В работе [10] используют термин vector, встречается также термин кусочки (pieces), set и т.д. Для определения опорного кластера применяют также термины последний пригодный вектор (last available vector), история ряда.

Введем следующие представления кластеров:

1. Кластеры могут представляться непосредственно как подпоследовательности элементов входного временного ряда (1).

2. Представление кластеров с помощью знаковой последовательности. На основе ряда  $\{z_i\}_{i=1}^n$  построим знаковый ряд  $\{\chi_i\}_{i=2}^n$ , где  $\chi_i = \text{sgn}(z_i - z_{i-1})$ ,  $i = \overline{2, n}$ . Тогда знаковые неопорные кластеры будут иметь вид:

$$\chi_{(m)}^2 = \{\chi_{k_1^2}, \chi_{k_2^2}, \dots, \chi_{k_m^2}\} = \{\chi_{k_j^2}\}_{j=1}^m, \quad k_{j+1}^2 = k_j^2 + 1, \quad j = \overline{1, m-1}, \quad \chi_{k_1^2} = \chi_2,$$

$$\chi_{(m)}^3 = \{\chi_{k_1^3}, \chi_{k_2^3}, \dots, \chi_{k_m^3}\} = \{\chi_{k_j^3}\}_{j=1}^m, \quad k_{j+1}^3 = k_j^3 + 1, \quad j = \overline{1, m-1}, \quad \chi_{k_1^3} = \chi_3,$$

$\vdots$

$$\chi_{(m)}^{n-m-1} = \{\chi_{k_1^{n-m-1}}, \chi_{k_2^{n-m-1}}, \dots, \chi_{k_m^{n-m-1}}\} = \{\chi_{k_j^{n-m-1}}\}_{j=1}^m, \quad k_{j+1}^{n-m-1} = k_j^{n-m-1} + 1, \quad j = \overline{1, m-1},$$

$$\chi_{k_1^{n-m-1}} = \chi_{n-m-1}.$$

Знаковый опорный кластер будет иметь вид

$$\chi_{(m)}^{n-m} = \{\chi_{k_1^{n-m}}, \chi_{k_2^{n-m}}, \dots, \chi_{k_m^{n-m}}\} = \{\chi_{k_j^{n-m}}\}_{j=1}^m, \quad k_{j+1}^{n-m} = k_j^{n-m} + 1, \quad j = \overline{1, m-1}, \quad \chi_{k_1^{n-m}} = \chi_{n-m}.$$

Очевидно, что число знаковых неопорных кластеров длины  $m$ , построенных на основе временного ряда с  $n$  элементов, равно  $n - m - 1$ .

3. Третий тип представления основан на введении расстояния между элементами кластеров. Пусть построены кластеры первого типа. В каждом кластере вычислим минимальное и максимальное значения и построим нормированные кластеры по правилу:

$$z_{\max}^s = \max\{z_{k_j^s}\}_{j=1}^m, \quad z_{\min}^s = \min\{z_{k_j^s}\}_{j=1}^m, \quad \text{тогда} \quad \delta_{k_j^s} = \frac{z_{k_j^s} - z_{\min}^s}{z_{\max}^s - z_{\min}^s}, \quad s = \overline{1, n-m}, \quad \text{причем} \\ \delta_{k_j^s} \in [0, 1].$$

Таким образом, для каждого  $s = \overline{1, n-m}$  получим нормированные неопорные кластеры и нормированный опорный кластер:

$$\delta_{(m)}^s = \{\delta_{k_1^s}, \delta_{k_2^s}, \dots, \delta_{k_m^s}\} = \{\delta_{k_j^s}\}_{j=1}^m, \quad k_{j+1}^s = k_j^s + 1, \quad j = \overline{1, m-1}, \quad \delta_{k_1^s} = \frac{z_s - z_{\min}^s}{z_{\max}^s - z_{\min}^s}.$$

Число нормированных неопорных кластеров длины  $m$ , построенных на основе временного ряда с  $n$  элементов, равно  $n - m$ .

Введем понятие меры близости между кластерами, что в случае количественной оценки выражается на основе определенной метрики. Подобие между двумя кластерами определяется на основе метрического расстояния между ними. Каждый кластер может быть представлен точкой в  $m$ -мерном пространстве. В зависимости от представления кластеров можно выделить различные метрические расстояния или меры близости. В данной работе ограничимся первым типом представления (1). Пусть заданы множество  $z_{(m)}^v$ ,  $v = \overline{1, n-m-1}$  и опорный кластер  $z_{(m)}^{n-m}$  ряда  $\{z_i\}_{i=1}^n$ , тогда меры близости между опорным и всеми неопорными кластерами могут быть определены по известным формулам:

$$1. \quad d_0(z_{(m)}^{n-m}, z_{(m)}^v) = \sqrt{\sum_{j=1}^m \left( z_{k_j^{n-m}} - z_{k_j^v} \right)^2} \quad - \text{расстояние Эвклида.}$$

$$2. \quad d_1(z_{(m)}^{n-m}, z_{(m)}^v) = \sum_{j=1}^m |z_{k_j^{n-m}} - z_{k_j^v}| \quad - \text{манхэттенская (городская) метрика.}$$

$$3. \quad d_2(z_{(m)}^{n-m}, z_{(m)}^v) = \left( \sum_{j=1}^m |z_{k_j^{n-m}} - z_{k_j^v}|^p \right)^{\frac{1}{p}} \quad - \text{расстояние Минковского.}$$

$$4. d_3(z_{(m)}^{n-m}, z_{(m)}^v) = \sqrt{\left( z_{k_j}^{n-m} - z_{k_j}^v \right)^T S_i^{-1} \left( z_{k_j}^{n-m} - z_{k_j}^v \right)}, \quad S_i = \text{cov} \left( z_{k_j}^{n-m} - z_{k_j}^v \right) - \text{расстояние Махаланобиса.}$$

$$5. d_4(z_{(m)}^{n-m}, z_{(m)}^v) = \sum_{j=1}^m I_v^j, \quad \text{где } I_v^j = \begin{cases} 1, & |z_{k_j}^{n-m} - z_{k_j}^v| < \varepsilon \\ 0, & |z_{k_j}^{n-m} - z_{k_j}^v| \geq \varepsilon \end{cases} - \text{расстояние Журавлева.}$$

Моделирование временных рядов на основе кластеризации [8, 10] базируется на том, что любой временной ряд может быть представлен как конечная последовательность кластеров одинаковой длины, использующихся в целях прогнозирования и анализа динамики. Для прогнозирования временного ряда сначала на основе некоторой степени близости находят кластер, подобный (closest match) опорному, последний элемент которого представляет собой точку, в которой рассчитывается прогноз. Далее предполагается, что поведение временного ряда, которое наблюдалось после подобного кластера, будет повторяться и после опорного кластера. Недостатком методики является неоднозначность зависимости результатов прогнозирования от выбора меры близости. Кроме того, данный подход для большинства временных рядов, в том числе тех, которые представляют собой экономические процессы, не отмечается высокой точностью.

#### 4. Адаптивные комбинированные модели прогнозирования с кластеризацией временного ряда по методу ближайшего соседа

**Постановка задачи.** Пусть задано программное множество  $\mathfrak{Z}_{PS}$  моделей прогнозирования  $f_1, f_2, \dots, f_K$ , на основе которых для ряда  $\{z_i\}_{i=1}^n$  в точке  $n$  могут быть построены оценки будущих элементов временного ряда  $\{z_{n+1}^p, z_{n+2}^p, \dots, z_{n+\theta}^p\}$ . На основе множества  $\mathfrak{Z}_{PS}$  и ретроспективных значений временного ряда  $\{z_i\}_{i=1}^n$  построить наиболее точную последовательность прогнозных значений  $\hat{Z} = \{z_i\}_{i=n+1}^{n+\theta} = \{z_{n+1}, z_{n+2}, \dots, z_{n+\theta}\}$ .

**Решение задачи.** Пусть для временного ряда  $\{z_i\}_{i=1}^n$  по правилу (1) построено множество неопорных кластеров одинаковой длины  $m$ :

$$z_{(m)}^v = \{z_{k_1}^v, z_{k_2}^v, \dots, z_{k_m}^v\} = \{z_{k_j}^v\}_{j=1}^m, \quad k_{j+1}^v = k_j^v + 1, \quad j = \overline{1, m-1}, \quad v = \overline{1, n-m-1}, \quad z_{k_1}^v = z_v. \quad (2)$$

Обозначим его через  $\mathfrak{K}$ . Прогноз, который реализуется на основе каждой модели  $f_p$  из программного множества  $\mathfrak{Z}_{PS}$ ,  $p = \overline{1, K}$  в точках ряда  $m, m+1, \dots, n-1$ ,  $m < n$ , соответствующих последним элементам неопорных кластеров  $z_{(m)}^v$  (2),  $v = \overline{1, n-m-1}$  длины  $m$  на  $\tau = \overline{1, \theta}$  точек вперед, обозначим через:

$$z_{k_{w+1}}^p = f_p \left( \{z_{k_j}^v\}_{j=1}^m \right), \quad w = v + m - 1, \quad k_{j+1}^v = k_j^v + 1, \quad j = \overline{1, m-1}, \quad z_{k_1}^v = z_v, \quad v = \overline{1, n-m-1},$$

$$z_{k_{w+2}}^p = f_p \left( \{z_{k_j}^v\}_{j=1}^m, z_{k_{w+1}}^p \right), \quad w = v + m - 1, \quad k_{j+1}^v = k_j^v + 1, \quad j = \overline{1, m-1}, \quad z_{k_1}^v = z_v, \quad v = \overline{2, n-m-1},$$

$$\vdots$$

$$z_{k_{w+\theta}}^p = f_p \left( \{z_{k_j}^v\}_{j=1}^m, z_{k_{w+1}}^p, z_{k_{w+2}}^p, \dots, z_{k_{w+\theta-1}}^p \right), \quad w = v + m - 1, \quad k_{j+1}^v = k_j^v + 1, \quad j = \overline{1, m-1},$$

$$z_{k_1}^v = z_v, \quad v = \overline{\theta, n-m-1},$$



где  $z_{k_{w+\tau}}^p$  — прогнозные значения элемента  $w + \tau$   $v$ -го кластера, полученное на основе  $p$ -й

прогнозной модели. Запись  $f_p \left( \{z_{k_j^v}\}_{j=1}^m \right)$  означает, что для построения прогноза по модели

$f_p$  использован соответствующий  $v$ -й неопорный кластер. Для упрощения записей будем считать, что  $z_{k_{w+\tau}}^p = z_{w+\tau}^p = z_t^p(w)$ , где  $z_t^p(w)$  — прогноз, рассчитываемый в точках  $w$  на  $\tau$  точек вперед по  $p$ -й модели, которая для построения прогнозного значения использует кластер  $z_{(m)}^v$ ,  $v = \overline{\tau, n-m-1}$  из множества неопорных кластеров  $\mathfrak{R}$ .

Построим опорный кластер

$$z_{(m)}^{n-m} = \{z_{k_1^{n-m}}, z_{k_2^{n-m}}, \dots, z_{k_m^{n-m}}\} = \{z_{k_j^{n-m}}\}_{j=1}^m, \\ k_{j+1}^{n-m} = k_j^{n-m} + 1, j = \overline{1, m-1}, z_{k_1^{n-m}} = z_{n-m} \quad (3)$$

и на основе некоторой меры близости, например, с помощью расстояния Евклида по методу ближайшего соседа найдем из множества неопорных кластеров  $\mathfrak{R}$  подобный опорному.

Неопорный кластер  $z_{(m)}^x \in \mathfrak{R}$  называется подобным опорному кластеру  $z_{(m)}^{n-m}$ , если не существует других неопорных кластеров  $z_{(m)}^k$ ,  $k \in [1, n-m-1]$ ,  $k \neq x$ , для которых  $d(z_{(m)}^{n-m}, z_{(m)}^k) < d(z_{(m)}^{n-m}, z_{(m)}^x)$ .

Согласно принципам построения комбинированных моделей есть два принципа расчета прогнозов: селективный и гибридный.

Селективный подход заключается в отборе для каждого значения  $\tau$  из программного множества  $\mathfrak{Z}_{PS}$  единственной модели, которая обеспечивает высокую точность прогнозирования по определенному критерию селекции: В- [4], R-критерий [7]. Параметры критериев селекции, как правило, имеют адаптивный характер. Кроме того, часто для повышения точности прогнозирования критерии отбора применяют не к программному, а к так называемому основному множеству  $\mathfrak{Z}_{BS}^t$ . Это множество состоит из моделей, дающих наиболее точные прогнозы на определенном участке временного ряда,  $\mathfrak{Z}_{BS}^t \subset \mathfrak{Z}_{PS}$ . Отбор моделей в основное множество может осуществляться, например, на основе D-критерия [4].

Прогноз по гибриднему подходу рассчитывается как взвешенная сумма прогнозов по всем моделям, составляющим основное множество  $\mathfrak{Z}_{BS}^t$ .

Построим прогноз на основе этих подходов, используя результаты кластеризации временного ряда. Пусть на основе определенной меры близости был определен кластер  $z_{(m)}^{x-m+1} \in \mathfrak{R}$ ,  $x \in [m, n-\theta-1]$ , подобный опорному кластеру. Последним элементом кластера  $z_{(m)}^{x-m+1}$  будет элемент  $z_x$ .

Построим основное множество моделей прогнозирования, используя D-критерий. Значение D-критерия в момент  $t_x$  можно рассчитать по формуле:

$$D_p(\tau) = \frac{1}{x-c+1} \sum_{j=0}^{x-c} \left( z_t^p(x-\tau-j) - z_{x-j} \right)^2, \quad (4)$$

где  $\tau$  — период прогноза;  $c$  — период предыстории;  $z_t^p(x-\tau-j)$  — прогноз, который рассчитывается в момент  $t_{x-\tau-j}$  (в точке  $x-\tau-j$ ) на  $\tau$  точек вперед за моделью  $f_p$ ,  $p = \overline{1, K}$ . Тогда основное множество моделей для фиксированного  $\tau$  определяется так:

$$\mathfrak{Z}_{BS}^t = \left\{ f_p \in \mathfrak{Z}_{PS} \mid D_p(\tau) \leq \lambda D_{\min}(\tau), p = \overline{1, K} \right\}, D_{\min}(\tau) = \min_{p=\overline{1, K}} D_p(\tau), \quad (5)$$

здесь  $\lambda$  – действительный параметр, который определяется экспериментальным путем.

Обозначим модели, включенные в множество  $\mathfrak{Z}_{BS}^{\tau}$ , через  $f_1^{\tau}, f_2^{\tau}, \dots, f_{L_{\tau}}^{\tau}$ ,  $L_{\tau}$  – количество моделей в множестве  $\mathfrak{Z}_{BS}^{\tau}$ ,  $\tau = \overline{1, \theta}$ ,  $L_{\tau} \leq K$ .

Для каждой модели из основного множества  $\mathfrak{Z}_{BS}^{\tau}$  рассчитаем значения В-критерия:

$$B_{x,\tau}^{q\tau} = (1 - \alpha_B) B_{x-1,\tau}^{q\tau} e_t^{q\tau}(x - \tau), \quad (6)$$

где  $0 < \alpha_B \leq 1$  – параметр сглаживания; а  $e_t^{q\tau}(x - \tau) = |z_t^{q\tau}(x - \tau) - z_x|$  – абсолютная ошибка прогноза, который рассчитывается в момент  $t_{x-\tau}$  на  $\tau$  точек вперед по моделям  $f_{q\tau}^{\tau}$ ,  $\tau = \overline{1, \theta}$ ,  $q_{\tau} = \overline{1, L_{\tau}}$ . Тогда самой точной моделью для фиксированного  $\tau$  по В-критерию будет считаться такая модель  $f^{\tau*}$ , для которой обеспечивается минимальное значение критерия,  $\min_{q_{\tau}=\overline{1, L_{\tau}}} B_{x,\tau}^{q\tau}$ . Прогнозы, рассчитываемые на основе моделей  $f^{\tau*}$  в точке  $n$  на

$\tau$  точек вперед, обозначим через  $z_{\tau}^*(n)$ . Тогда прогноз по комбинированной модели селективного типа по В-критерию селекции (6) с кластеризацией временного ряда по методу ближайшего соседа будет рассчитываться по формуле:

$$z_{\tau}(n) = \alpha z_{\tau}^*(n) + (1 - \alpha) z_{x+\tau}, \quad (7)$$

где  $z_{x+\tau}$  – значение временного ряда, которое следует после кластера  $z_{(m)}^{x-m+1}$ , подобного опорному;  $\alpha \in [0, 1]$  – параметр, указывающий на важность учета прогнозных значений отобранной модели в прогнозе.

Применим гибридный подход. Пусть после проведенной кластеризации в точке  $x$  для каждого  $\tau$  были сформированы основные множества  $\mathfrak{Z}_{BS}^{\tau}$  и рассчитаны значения В-критериев  $B_{x,\tau}^{q\tau}$ ,  $q_{\tau} = \overline{1, L_{\tau}}$ . Обозначим через  $z_{\tau}^{q\tau}(n)$  – прогноз, который рассчитывается в точке  $n$  на  $\tau$  точек вперед по моделям  $f_{q\tau}^{\tau}$  из основного множества  $\mathfrak{Z}_{BS}^{\tau}$ ,  $q_{\tau} = \overline{1, L_{\tau}}$ ,  $\tau = \overline{1, \theta}$ . Тогда прогноз по комбинированной модели гибридного типа с кластеризацией по методу ближайшего соседа определяется по формуле:

$$z_{\tau}(n) = \alpha \sum_{q_{\tau}=1}^{L_{\tau}} \omega_{\tau}^{q\tau} z_{\tau}^{q\tau}(n) + (1 - \alpha) z_{x+\tau}, \quad (8)$$

где  $\alpha \in [0, 1]$ , веса  $\omega_{\tau}^{q\tau}$  определяются на основе В-критерия (6) с учетом коэффициента пропорциональности, находящегося из равенства суммы всех весов единице,  $\sum_{q_{\tau}=1}^{L_{\tau}} \omega_{\tau}^{q\tau} = 1$ , например, как в работе [4]:

$$\text{для } L_{\tau} = 2 \quad \omega_{\tau}^1 = \frac{B_{x,\tau}^2}{B_{x,\tau}^1 + B_{x,\tau}^2}, \quad \omega_{\tau}^2 = \frac{B_{x,\tau}^1}{B_{x,\tau}^1 + B_{x,\tau}^2},$$

$$\text{для } L_{\tau} = 3 \quad \omega_{\tau}^1 = \frac{B_{x,\tau}^2 B_{x,\tau}^3}{B_{x,\tau}^1 B_{x,\tau}^2 + B_{x,\tau}^1 B_{x,\tau}^3 + B_{x,\tau}^2 B_{x,\tau}^3},$$

$$\omega_{\tau}^2 = \frac{B_{x,\tau}^1 B_{x,\tau}^3}{B_{x,\tau}^1 B_{x,\tau}^2 + B_{x,\tau}^1 B_{x,\tau}^3 + B_{x,\tau}^2 B_{x,\tau}^3}, \quad \omega_{\tau}^3 = \frac{B_{x,\tau}^1 B_{x,\tau}^2}{B_{x,\tau}^1 B_{x,\tau}^2 + B_{x,\tau}^1 B_{x,\tau}^3 + B_{x,\tau}^2 B_{x,\tau}^3} \text{ и т.д.}$$

После расчета прогноза прогнозная точка  $z_1(n)$  используется для построения нового опорного кластера  $z_{(m)}^{n-m+1} = \{z_{k_1}^{n-m+1}, z_{k_2}^{n-m+1}, \dots, z_{k_{m-1}}^{n-m+1}, z_1(n)\}$ ,  $z_{k_1}^{n-m+1} = z_{n-m+1}$ , а ста-

рый опорный кластер  $z_{(m)}^{n-m}$  становится неопорным, т.е. включается в множество  $\mathcal{K}$ , и процесс расчета начитается сначала: находится на основе определенной меры близости подобный опорному кластер, формируются множества  $\mathcal{Z}_{BS}^t$  для каждого  $t$ , рассчитываются значения В-критерия для каждой прогнозной модели, которая использует подобную неопорную последовательность в качестве ретроспективной информации. Далее строится прогноз согласно селективному или гибридному принципу.

## 6. Выводы и численные результаты

Описанная методика была реализована в программной среде. На основе программного множества из простых экспоненциальных моделей сглаживания (SESM) или адаптивных полиномиальных моделей Брауна 0-го, 1-го и 2-го порядков с параметрами сглаживания 0,2 были построены обычные адаптивные комбинированные модели гибридного типа и селективного типа по В- (6) и R-критериям [7], а также построены модификации данных моделей с предыдущей кластеризацией временного ряда по методу ближайшего соседа. В качестве меры близости было выбрано расстояние Евклида. В каждой точке для каждого  $t$  автоматически формировалось основное множество моделей  $\mathcal{Z}_{BS}^t$  (5) по D-критерию (4) с параметром  $\lambda = 1,9$ . Параметр сглаживания в В-критерии (6)  $\alpha_B = 0,6$ , период предыстории для построения значений В-критерия равен 10. Для тестирования моделей выбраны ежедневные временные ряды цен на сырье: алюминий, бензин, серебро и т.д. за последние 3 года. Всего по 700 измерений. Были рассчитаны средние ошибки прогнозов на  $t = \overline{1,10}$  точек вперед по описанным методикам (7,8). Точность комбинированных моделей определяется тем, на сколько точнее их прогноз в сравнении с моделью-лидером, т.е. такой моделью из программного множества, для которой в среднем для фиксированного  $t$  получается минимальная ошибка, в данном случае относительная. Так, простая адаптивная комбинированная селективная модель по В-критерию улучшила самую точную модель из программного множества (модель-лидер) только для  $t = \overline{1,3}$  в среднем на 8,52%, а предложенная модификация данной модели с предыдущей кластеризацией временного ряда ( $\alpha = 0,95$ ,  $m = 20$ ) улучшила модель-лидер только для  $t = \overline{8,10}$  в среднем на 2,59 %. Подобные результаты получаются и в случае использования R-критерия. Простая адаптивная гибридная модель улучшает модель-лидер для всех  $t = \overline{1,10}$ : для  $t = \overline{1,5}$  в среднем на 4,71% по сравнению с результатом модели-лидера, для  $t > 5$  в среднем на 1,62%. Адаптивная гибридная модель с кластеризацией ( $\alpha = 0,95$ ) улучшает модель-лидер для  $t > 5$  в среднем на 1,97% и простую гибридную модель в среднем на 0,45%.

*Научная новизна.* В целях повышения точности прогнозирования временных рядов разработана адаптивная комбинированная модель прогнозирования с предыдущей кластеризацией временного ряда по методу ближайшего соседа. Отличие обыкновенной адаптированной комбинированной модели от предложенной отличается тем, что в последней отбор лучших моделей для расчета прогноза по селективному или гибридному принципу осуществляется не по результатам прогнозирования на участке ряда, предшествующего точке прогноза, так называемом опорном кластере, а по результатам прогнозирования моделей на другом участке (кластере) ряда, который подобный опорному, т.е. метрическое расстояние которого к опорному кластеру минимально.

*Практическая ценность* работы в том, что предложенные модели могут использоваться в качестве составляющих информационных прогнозных систем, для повышения точности прогнозирования комбинированных моделей с периодом  $t > 5$  и для повышения эффективности управления процессом прогнозирования в части учета преимуществ конкурирующих прогнозов. Проведенный сравнительный анализ результатов прогнозирования временных рядов цен на сырье с помощью обыкновенных адаптивных комбинированных моделей и комбинированных моделей с предварительной кластеризацией ряда позволяет сделать вывод, что предложенный подход дает возможность добиться более высокой точности в случае прогнозирования с периодом  $t > 5$ .

**Список литературы:** 1. *Brown Robert G.* Statistical forecasting for inventory control. US: McGraw-Hill Inc., 1959. 223 p. 2. *Holt Charles C.* Forecasting trends and seasonal by exponentially weighted averages // International Journal of Forecasting. 1957. Vol.20, no.1.P.5-10. 3. *Vercellis Carlo* Business intelligence: data mining and optimization for decision making. John Wiley & Sons, Ltd., Publication, 2009. 417 p. 4. *Лукашин Ю.П.* Адаптивные методы краткосрочного прогнозирования временных рядов: Учеб. пособие. М.: Финансы и статистика, 2003. 416 с. 5. *Box G.E.P., Jenkins G.M.* Time series analysis: forecasting and control. San Francisco: Holden-Day, 1976. 575 p. 6. *Берзлев О.Ю., Маляр М.М., Ніколенко В.В.* Адаптивні комбіновані моделі прогнозування біржових показників // Вісник Черкаського держ. технолог. ун-ту. Серія: технічні науки. 2011. № 1. С. 50-54. 7. *Берзлев А.Ю.* Оценка эффективности прогнозирования и принятия решений на финансовом рынке // «Problems of Computer Intellectualization», V.M. Glushkov Institute of Cybernetics of NAS of Ukraine. Kyiv-Sofia: ITHEA, 2012. С. 249-257. 8. *Keogh E., Pazzani M.* An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback // 4th Int'l Conference on Knowledge Discovery and Data Mining. 1998 Aug 27-31. New York. P. 239-241. 9. *Singh S.* Pattern Modeling in Time-Series Forecasting // Cybernetics and Systems. An International Journal. 2000. Vol. 31, no. 1. P. 49-65. 10. *Fernández-Rodríguez F., Sosvilla-Rivero S., Andrada-Félix J.* Nearest-Neighbour Predictions in Foreign Exchange Markets // Fundacion de Estudios de Economía Aplicada. 2002. no.5. 36 p. 11. *Peters E. E.* Fractal market analysis: applying chaos theory to investment and economics. John Wiley & Sons, Inc, 1994. 336 p. 12. *Снитюк В.Є.* Прогнозування. Моделі. Методи. Алгоритми: Навчальний посібник. К.: Маклайт, 2008. 364 с. 13. *Chang C.L.E., Garcia-Molina H., Wiederhold G.* Clustering for approximate similarity search in high-dimensional spaces // IEEE Transactions on Knowledge and Data Engineering. 2002. Vol 14, no.4. P. 792 -808.

Поступила в редколлегию 12.11.2012

**Берзлев Александр Юрьевич**, аспирант каф. кибернетики и прикладной математики математического ф-та ГВУЗ «УжНУ». Научные интересы: прогнозирование и интеллектуальный анализ временных рядов. Адрес: Украина, Ужгород, ул. Университетская, 14, к. 313. E-mail: [berzlev@gmail.com](mailto:berzlev@gmail.com)

УДК 519.7

**Ю.С. НОВИКОВ**

## **СТРУКТУРИЗАЦИЯ СОСТАВНЫХ ОБЪЕКТОВ ПРИ ФОРМИРОВАНИИ ПРОЦЕССНОГО ПРЕДСТАВЛЕНИЯ ЗНАНИЙ**

Предлагается подход к выполнению структуризации составных объектов как элементов процессного представления знаний. При структуризации выполняется последовательное сравнение простых атрибутов объектов и их совокупности на основе их типов и значений, сравнения уникальных атрибутов на основе специализированных алгоритмов, сравнения вложенных объектов.

### **1. Актуальность**

Процессное представление знаний формируется на основе анализа наборов последовательностей событий, имеющих временные метки, и выявления на основе такого анализа структурированных объектов и последовательностей действий, составляющих процесс, а также причинно-следственных зависимостей между такими действиями. Для решения указанных задач анализа традиционно используются методы data mining и, в последние годы, методы бурно развивающегося направления process mining (интеллектуального анализа процессов). Исходными данными при решении таких задач являются наборы последовательностей событий, фиксирующих поведение некоторых объектов и связанных с ними процессов. Интеллектуальный анализ процессов направлен на выявление последовательности взаимодействия таких объектов с учетом скрытых причинно-следственных связей между ними. Результатом данного анализа является процессное представление знаний, отражающее не просто возможные наборы действий в предметной области, но и зависимости между ними, что создает возможность организации логического вывода на процессных знаниях. В то же время выявленные причинно-следственные связи между действиями процесса отражают лишь «внешнюю» природу процессных знаний. Данные зависимости в значительной степени зависят от природы и структуры объектов, с которыми оперирует процесс. Следовательно, структура таких объектов отражает «внутренние» закономерности