

На правах рукописи



Скворцова Татьяна Сергеевна

**МОДЕЛИ ПРОГНОЗИРОВАНИЯ ПРОЦЕССОВ,
ПРЕДСТАВЛЕННЫХ ВРЕМЕННЫМИ РЯДАМИ
С КОРОТКОЙ АКТУАЛЬНОЙ ЧАСТЬЮ**

Специальность 05.13.01 –
«Системный анализ, управление и обработка информации
(технические системы)»

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Рязань 2011

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего профессионального образования «Рязанский государственный радиотехнический университет» на кафедре вычислительной и прикладной математики.

Научный руководитель: доктор технических наук, доцент
ДЕМИДОВА Лилия Анатольевна

Официальные оппоненты: доктор технических наук, профессор
МАЙКОВ Константин Анатольевич

кандидат технических наук
НОВИКОВ Геннадий Александрович

Ведущая организация: Московский государственный институт
электроники и математики (технический
университет)

Защита диссертации состоится 14 декабря 2011 г. в 12 часов на заседании диссертационного совета Д 212.211.01 в ФГБОУ ВПО «Рязанский государственный радиотехнический университет» по адресу: 390005, г. Рязань, ул. Гагарина, 59/1.

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВПО «Рязанский государственный радиотехнический университет».

Автореферат разослан « 5 » ноября 2011 г.

Ученый секретарь
диссертационного совета
канд. техн. наук, доцент



В.Н. Пржегорлинский

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Прогнозирование – это один из ключевых моментов при принятии управленческих решений в постиндустриальном обществе. Регулярное прогнозирование процессов позволяет не только принимать эффективные управленческие решения, но и накапливать опыт, позволяющий повысить точность и надёжность прогнозов, улучшить методы и алгоритмы прогнозирования.

Методы прогнозирования, лежащие в основе упреждающего управления, призваны выявить общие перспективы и тенденции развития прогнозируемых процессов, обеспечить сбалансированность краткосрочных и долгосрочных программ. При этом важным является установление всей совокупности факторов, влияющих на развитие тех или иных процессов. Изменчивость производственно-экономических отношений, неустойчивость, концептуальная непоследовательность и неполнота законодательной базы, регулирующей эти отношения, необходимость принятия адекватных управленческих решений в новых сферах человеческой деятельности и т.п. приводят к тому, что временные ряды (ВР), представляющие многие производственные, экономические и социально-экономические процессы, имеют короткую актуальную часть. Кроме того, многие новые процессы производственной и социальной сферы представляются физически короткими ВР, поскольку ранее не являлись предметом статистического учета.

Традиционные подходы к проблеме прогнозирования достаточно полно изложены в научных трудах Айвазяна С.А., Андерсена Т., Бендита Дж., Бокса Дж., Бриллинджера Д., Вучкова И., Демиденко Е.З., Дженкинса Г., Джонсона М., Кендэла М., Монтгомери Б., Пирсола А., Степанова В.С., Стьюарта А., Уолкера Г., Чатфилда К. Наиболее известными пакетами программ, реализующими традиционные методы, модели и алгоритмы прогнозирования, являются BMDP, CART, CSS, Deductor, Forecast Expert, MVSP, Predictor, SAS, S-plus, SPSS, STADIA, STATISTICA, STATGRAPHICS, SYSTAT, КЛАСС-МАСТЕР, МЕЗОЗАВР, ОЛИМП: СтатЭксперт, ЭВРИСТА, Статистик-Консультант.

Однако эти программные пакеты не могут быть эффективно использованы для прогнозирования процессов, представленных малым объемом статистического материала, поскольку в своём большинстве не содержат элементы адаптивности и самообучения и, следовательно, предлагают пользователям самостоятельно конструировать алгоритмы прогнозирования, что требует глубоких теоретических знаний в области статистики и эконометрики. При этом многие алгоритмы прогнозирования используют при построении прогнозов предварительные оценки автокорреляционной функции, расчет которых является достаточно сложной задачей, так как предполагает выбор адекватного порядка модели прогнозирования на основе ВР с короткой актуальной частью. Кроме того, для подавляющего большинства реальных процессов, в силу их неустойчивости, невозможно построить эффективную процедуру автоматического выделения тренда из ВР.

Главные трудности, с которыми приходится сталкиваться при анализе и прогнозировании динамики процессов, имеющих короткую актуальную часть, заключаются в следующем. Во-первых, высокая нестационарность многих процессов, представленных ВР с короткой актуальной частью, делает невозможным приме-

нение хорошо разработанных и изученных традиционных методов, моделей и алгоритмов прогнозирования. Во-вторых, традиционные методы, модели и алгоритмы прогнозирования обеспечивают получение удовлетворительного прогноза только для процессов, ВР которых содержат, как правило, не менее 50 измерений. В-третьих, для построения корректных математических моделей прогнозирования, адекватно описывающих динамику различных процессов, необходим тщательный анализ и учет факторов, влияющих на прогнозируемый процесс.

Таким образом, существует необходимость в разработке новых альтернативных подходов к проблеме прогнозирования ВР с короткой актуальной частью. В настоящее время всё большее внимание уделяется разработке гибридных подходов к решению широкого спектра прикладных задач, реализующих комплексное применение различных методов искусственного интеллекта, позволяющих сформировать новую информационную технологию, важную роль в которой играют знания предметной области конкретной прикладной задачи. К таким методам, в первую очередь, относятся методы, основанные на применении теории нечетких множеств (ТНМ), искусственных нейронных сетей (ИНС), искусственных иммунных систем (ИИС) и генетических алгоритмов (ГА).

Один из современных подходов, используемых в различных задачах принятия решений в условиях неопределенности, основан на применении инструментария ТНМ, основоположником которой является Л.А. Заде (1965 г.). Проблема неопределенности присуща всем сложным системам, в частности, техническим, экономическим и социально-экономическим системам. Эта неопределенность является систематической, так как обусловлена сложностью задач, дефицитом информации, лимитом времени на принятие решений и т.п. Неполнота и неточность исходной информации в контексте проблемы прогнозирования ВР с короткой актуальной частью могут заключаться: в принципиальной невозможности полного сбора и учета информации о прогнозируемом процессе; в некоторой недостоверности и недостаточности исходной информации о прогнозируемом процессе; в возможности проявления таких особенностей прогнозируемого процесса, существование которых не предполагалось. Процессы, представленные ВР, в реальных условиях также характеризуются неполнотой информации, которая может быть принципиальной из-за неповторимости (невоспроизводимости) явлений.

В настоящее время значительный вклад в разработку инструментария ТНМ для решения разнообразных математических и прикладных задач внесли российские и зарубежные ученые, такие как А.Н. Аверкин, А.В. Алексеев, Р. Беллман, Л.С. Бернштейн, А.Н. Борисов, Л.А. Заде, А. Кофман, О.А. Крумберг, Н.Г. Малышев, А.Н. Мелехов, С.А. Орловский, Р. Ягер, Т.Л. Саати и др. Проблема разработки базовых одно- и двухфакторных моделей краткосрочного прогнозирования на основе нечетких ВР с короткой актуальной частью рассматривается в работах Л. Х. Ванга, Л.В. Ли, К. Сонга, С.М. Чена, Б.С. Чиссома.

Анализ базовых моделей краткосрочного прогнозирования, основанных на применении ТНМ, показывает, что довольно часто они не обеспечивают получение адекватных результатов прогнозирования ввиду недостаточно обоснованного выбора параметров моделей, а поиск эффективных решений сопровождается значительными временными затратами из-за необходимости выполнения многократ-

ных реализаций базовых моделей краткосрочного прогнозирования с целью выбора оптимальных параметров.

Одним из современных бионических принципов решения широкого класса прикладных задач, трудноразрешимых классическими методами, особенно в области *NP*-полных задач оптимизации, является применение ГА – адаптивных методов поиска, реализующих эволюционные вычисления, основанные на генетических процессах биологических организмов. Общие принципы ГА были сформулированы Д.Х. Холландом (1975 г.) и описаны в работах Л.А. Гладкова, Д.И. Голдберга, В.В. Емельянова, В.М. Курейчика, В.В. Курейчика и др. Использование модификаций ГА позволяет решить проблему выбора оптимальных параметров базовых моделей краткосрочного прогнозирования на основе нечетких ВР с короткой актуальной частью при приемлемых временных затратах.

В последние годы широкое применение при решении прикладных задач находят ИИС, основанные на принципах естественной иммунной системы, реализующей модель адаптивного действия на локальном уровне и эмерджентность поведения на глобальном уровне и обладающей мощными и гибкими способностями обработки информации. Результативность применения принципов и механизмов естественной иммунной системы, в частности, алгоритма клонального отбора и молекулярных механизмов супрессии антител в практических приложениях решения задач интерполяции, экстраполяции и прогнозирования, доказанная в работах зарубежных исследователей, подтверждает целесообразность применения принципов обработки информации иммунными системами для прогнозирования поведения нестационарных динамических систем.

Следует отметить, что применение ИИС при разработке моделей краткосрочного прогнозирования ВР с короткой актуальной частью не является целесообразным ввиду наличия малого объема статистических данных.

Актуальность настоящей работы определяется необходимостью разработки эффективных моделей, основанных на комплексном использовании инструментария ТНМ, ГА и ИИС, и обобщения результатов многоальтернативного прогнозирования, что позволит создать качественно новые программные средства, обеспечивающие повышение точности прогнозирования процессов, представленных ВР с короткой актуальной частью при приемлемых временных затратах.

Цель диссертационной работы состоит в повышении точности прогнозирования посредством разработки эффективных моделей краткосрочного прогнозирования процессов с преобладающей детерминированной составляющей, представленных временными рядами с короткой актуальной частью, в условиях отсутствия априорной информации о вероятностных характеристиках процесса на основе комплексного использования инструментария теории нечетких множеств, генетических алгоритмов и искусственных иммунных систем.

Задачи исследования. Для достижения поставленной цели необходимо решить следующие задачи.

1. Провести анализ существующих методов и моделей прогнозирования временных рядов с короткой актуальной частью, выявить перспективные направления исследований.

2. Исследовать возможность использования инструментария ТНМ, ГА и ИИС при разработке моделей краткосрочного прогнозирования (на 1 шаг вперед) временных рядов с короткой актуальной частью.

3. Разработать модели краткосрочного прогнозирования с использованием инструментария теории нечетких множеств и генетических алгоритмов.

4. Разработать модели краткосрочного прогнозирования с использованием инструментария искусственных иммунных систем.

5. Разработать пакет прикладных программ (ППП) для краткосрочного прогнозирования временных рядов с короткой актуальной частью на основе предложенных моделей прогнозирования.

Методы исследования. Теоретические исследования выполнены с использованием методов теории нечетких множеств, нечеткой логики, генетических алгоритмов, искусственных иммунных систем, теории вероятностей, математической статистики, математического и системного анализа, математического моделирования, модульного и объектно-ориентированного программирования.

Научная новизна. В рамках диссертационной работы были получены следующие результаты.

1. Разработаны и исследованы однофакторные и двухфакторные модели краткосрочного прогнозирования k -го порядка, которые обеспечивают повышение точности прогноза за счет:

- применения групп нечетких логических зависимостей для описания поведения временных рядов и рационального выбора параметров моделей с помощью разработанного модифицированного генетического алгоритма (МГА);
- автоматического формирования аналитической зависимости, адекватно описывающей известные значения временного ряда, с использованием разработанного модифицированного алгоритма клонального отбора (МАКО).

2. Разработана рекурсивная процедура формирования аналитических зависимостей, реализующая способ корректного кодирования антител с помощью бинарных деревьев в модифицированном алгоритме клонального отбора.

3. Разработана методика обобщения прогнозов, позволяющая компенсировать потенциальные ошибки применяемых моделей прогнозирования, обусловленные неполным восстановлением процессов частными прогнозирующими описаниями, посредством консолидации прогнозов.

Практическая ценность работы состоит в том, что разработанные модели прогнозирования позволяют реализовать новый подход к решению задачи прогнозирования ВР с короткой актуальной частью и обеспечивают:

- высокую точность прогнозирования процессов, представленных малыми объемами статистического материала;
- минимизацию временных и финансовых затрат, связанных как с необходимостью сбора и учета точных и полных исходных данных (что может быть принципиально невозможным), так и с необходимостью многократной реализации базовых моделей краткосрочного прогнозирования, основанных на применении технологий искусственного интеллекта, с целью выбора соответствующих параметров моделей, обеспечивающих принятие адекватных решений.

Достоверность приведенных в диссертационной работе научных положений, результатов, выводов и рекомендаций подтверждается:

- корректным использованием понятий ТНМ, теории ГА и теории ИИС;
- результатами математического моделирования предложенных моделей прогнозирования на ЭВМ;
- результатами сравнительного анализа предложенных и классических моделей прогнозирования;
- разработкой прикладных программных средств, подтвержденных свидетельствами об официальной регистрации;
- апробацией разработанных моделей прогнозирования на конкретных практических примерах;
- наличием актов внедрения результатов диссертационной работы в практику разработки прогнозов производственных и экономических процессов, представленных ВР с короткой актуальной частью.

На защиту выносятся:

1) однофакторные и двухфакторные модели краткосрочного прогнозирования k -го порядка, которые обеспечивают повышение точности прогноза за счет:

- применения групп нечетких логических зависимостей для описания поведения временных рядов и рационального выбора параметров моделей с помощью разработанного модифицированного генетического алгоритма;
- автоматического формирования аналитической зависимости, адекватно описывающей известные значения временного ряда, с использованием разработанного модифицированного алгоритма клонального отбора;

2) рекурсивная процедура формирования аналитических зависимостей, реализующая способ корректного кодирования антител с помощью бинарных деревьев в модифицированном алгоритме клонального отбора;

3) методика обобщения прогнозов, позволяющая компенсировать потенциальные ошибки применяемых моделей прогнозирования, обусловленные неполным восстановлением процессов частными прогнозирующими описаниями, посредством консолидации прогнозов;

4) ППП для краткосрочного прогнозирования временных рядов с короткой актуальной частью на основе разработанных моделей прогнозирования.

Внедрение результатов. Исследования по тематике диссертационной работы проводились в рамках госбюджетных НИР 9-07Г «Разработка математических моделей, методов и алгоритмов обработки больших потоков информации в сложно организованных вычислительных структурах» (2007-2008 гг.) и НИР 7-09Г «Разработка математических методов и алгоритмов передачи и обработки цифровой информации для поддержки интеллектуальных систем управления» (2009-2011 гг.).

Результаты диссертационной работы внедрены в деятельность ООО «Компьютерные системы и технологии» (г. Москва) при разработке информационно-аналитических автоматизированных систем, в частности, сервиса «Прогноз» информационной системы, входящей в состав промышленной АСУТП для прогно-

зирования и оценки параметров высоконагруженного технологического оборудования, функционирующего в условиях автоматизированного производства, используются в работе ООО «Независимый центр оценки и экспертиз» (г. Рязань) для прогнозирования основных составляющих денежного потока при расчетах стоимости объектов в рамках доходного подхода к оценке, а также в работе ООО «М-Ком» (г. Рязань) для прогнозирования фактической загрузки каналов связи при организации широкополосного доступа в сеть Интернет.

Результаты полученных в диссертации теоретических, прикладных и экспериментальных исследований используются в учебном процессе Рязанского государственного радиотехнического университета при обучении студентов специальности 230105 «Программное обеспечение вычислительной техники и автоматизированных систем» в курсе «Проектирование искусственного интеллекта», специальности 080801 «Прикладная информатика в экономике» в курсах «Информационные технологии» и «Элементы теории нечетких множеств».

Использование результатов диссертационной работы на практике подтверждено соответствующими актами о внедрении.

Апробация работы. Основные положения диссертации докладывались на следующих конференциях: 15-я Международная научно-техническая конференция «Проблемы передачи и обработки информации в сетях и системах телекоммуникаций» (февраль 2008 г., Рязань); VIII Всероссийская научно-техническая конференция «Проблемы информатики в образовании, управлении, экономике и технике» (ноябрь 2008 г., Пенза); 17-я Международная научная конференция «Методы и алгоритмы принятия эффективных решений» (ноябрь 2009 г., Таганрог); XIII и XIV Всероссийские научно-технические конференции студентов, молодых ученых и специалистов «Новые информационные технологии в научных исследованиях и в образовании» (май 2008 г. и ноябрь 2009 г., Рязань); V Международная научно-практическая конференция «Тенденции развития современных информационных технологий, моделей экономических, правовых и управленческих систем» (апрель 2010 г., Рязань); XXVI Международная научно-техническая конференция «Математические методы и информационные технологии в экономике, социологии и образовании» (декабрь 2010 г., Пенза); XVI Международная открытая научная конференция «Современные проблемы информатизации в экономике и обеспечении безопасности» (декабрь 2010 г. – январь 2011 г., Воронеж).

Публикации. По теме диссертации опубликовано 17 печатных работ: 3 статьи в изданиях из перечня ВАК, 4 статьи в межвузовских сборниках, 8 докладов на Международных и Всероссийских конференциях, 2 свидетельства об официальной регистрации программ для ЭВМ в РОСПАТЕНТ.

Структура и объем диссертации. Диссертационная работа состоит из введения, четырех глав, списка литературы и приложений. Содержит 255 страниц, 45 таблиц, 48 рисунков. Список литературы включает 153 наименования. В приложениях приведены список сокращений, копии документов о внедрении результатов и копии свидетельств о регистрации программ для ЭВМ.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обосновывается актуальность выбора темы диссертационной работы, формулируются цель и задачи исследований, научная новизна и практическая ценность основных результатов диссертационной работы.

В первой главе выполнен обзор классических методов и моделей прогнозирования временных рядов с короткой актуальной частью, выявлены их достоинства, недостатки и особенности применения. Приведены основные понятия и определения, связанные с проблемой прогнозирования процессов, представленных ВР. Исследована возможность использования инструментария ТНМ, ГА и ИИС при разработке моделей краткосрочного прогнозирования ВР. Сформулирована задача разработки нового подхода к прогнозированию временных рядов с короткой актуальной частью с использованием моделей прогнозирования на основе НМ и МГА и моделей прогнозирования на основе МАКО.

Вторая глава диссертационной работы посвящена разработке *моделей прогнозирования ВР на основе НМ и МГА*, реализующих краткосрочное прогнозирование на 1 шаг вперед на основе известных значений ВР.

Так как большинство реальных процессов (прогнозируемых факторов) характеризуются некоторой неопределенностью, то описывающие их ВР можно рассматривать как нечеткие, а каждому наблюдению нечеткого ВР можно поставить в соответствие нечеткую переменную с некоторой функцией принадлежности.

В классической базовой однофакторной модели прогнозирования на основе НМ – модели С.М. Чена – реализовано представление ВР непосредственно с помощью значений прогнозируемого фактора. В данной работе предложено для представления ВР использовать значения приращений прогнозируемого фактора, что позволяет повысить точность прогнозирования (в смысле минимизации показателя точности прогнозирования – средней относительной ошибки прогнозирования *AFER* (Average Forecasting Error Rate)).

Пусть $d(t)$ ($t = \overline{0, n}$) – ВР на основе реальных значений некоторого фактора. Тогда ВР на основе значений приращений фактора представляется как

$$\Delta d(t) = d(t) - d(t-1) \quad (t = \overline{1, n}).$$

Значение нечеткого ВР представляется с помощью НМ A , определенного на универсуме X : $A = u_A(x_1)/x_1 + u_A(x_2)/x_2 + \dots + u_A(x_n)/x_n$, где $u_A(x)$ – функция принадлежности НМ A , $u_A(x): X \rightarrow [0, 1]$, $u_A(x_r)$ ($r = \overline{1, n}$) определяет степень принадлежности элемента x_r .

При построении однофакторной модели прогнозирования на основе НМ для прогнозируемого ВР $d(t)$ ($t = \overline{0, n}$) (или $\Delta d(t)$ ($t = \overline{1, n}$)) определяется универсум X : $X = [D_{\min} - D_1, D_{\max} + D_2]$, где $D_{\min} = \min_{t=0, n}(d(t))$ и $D_{\max} = \max_{t=0, n}(d(t))$ (или соответственно $D_{\min} = \min_{t=1, n}(\Delta d(t))$ и $D_{\max} = \max_{t=1, n}(\Delta d(t))$); D_1 и D_2 – два действительных числа, использование которых позволяет разбить универсум X на p интервалов x_1, x_2, \dots, x_p равной длины.

Лингвистические термы A_i ($i = \overline{1, p}$), представленные НМ, могут быть записаны в виде: $A_1 = 1/x_1 + 0,5/x_2$; $A_i = 0,5/x_{i-1} + 1/x_i + 0,5/x_{i+1}$ ($i = \overline{2, p-1}$); $A_p = 0,5/x_{p-1} + 1/x_p$.

При фаззификации данных для каждого значения ВР определяется интервал принадлежности x_i ($i = \overline{1, p}$) и соответствующее нечеткое значение ВР:

$$V_i = \begin{cases} 1/A_1 + 0,5/A_2, & \text{если } i = 1; \\ 0,5/A_{i-1} + 1/A_i + 0,5/A_{i+1}, & \text{если } 2 \leq i \leq p-1; \\ 0,5/A_{p-1} + 1/A_p, & \text{если } i = p. \end{cases}$$

Для всех известных значений ВР находятся нечеткие логические зависимости (НЛЗ) k -го порядка вида $(V_{j_k}, V_{j_{(k-1)}}, \dots, V_{j_1}) \rightarrow V_l$, а затем НЛЗ, имеющие одинаковую левую часть, объединяются в одну группу НЛЗ (ГНЛЗ):

$$(V_{j_k}, V_{j_{(k-1)}}, \dots, V_{j_1}) \rightarrow V_{l_1}, V_{l_2}, \dots, V_{l_g}. \quad (1)$$

Для обеспечения возможности прогнозирования на 1 шаг вперед правые части всех ГНЛЗ не должны быть пустыми.

Искомое значение прогнозируемой величины $f(t+1)$ для $(t+1)$ -го отсчета времени ($t = \overline{k, n}$) находится как дефаззифицированное значение величины $z(t+1)$ для $(t+1)$ -го отсчета времени ($t = \overline{k, n}$) в случае, когда ВР представлен значениями фактора; и как сумма значения $d(t)$, определяемого с помощью ВР на основе значений фактора для t -го отсчета времени, и дефаззифицированного значения величины $z(t+1)$ для $(t+1)$ -го отсчета времени ($t = \overline{k+1, n}$) в случае, когда ВР представлен значениями приращений фактора: $f(t+1) = d(t) + z(t+1)$.

Дефаззифицированное значение величины $z(t+1)$ для $(t+1)$ -го отсчета времени определяется по методу центра тяжести для одноточечных множеств:

$$z(t+1) = \frac{\sum_{i=1}^p v_i(t+1) \cdot \chi_i}{\sum_{i=1}^p v_i(t+1)}, \quad (2)$$

где p – количество интервалов x_i ($i = \overline{1, p}$); χ_i – средняя точка i -го интервала x_i ; $v_i(t+1)$ – значение степени принадлежности для i -го интервала результирующего НМ для $(t+1)$ -го отсчета времени в правой части ГНЛЗ вида (1), где функция принадлежности вычисляется по формуле $u_i(x_t) = \max(u_{i_1}(x_t), u_{i_2}(x_t), \dots, u_{i_g}(x_t))$.

Учет повторяющихся НМ в правых частях ГНЛЗ позволяет уменьшить значение средней относительной ошибки прогнозирования $AFER$. При этом дефаззифицированное значение величины $z(t+1)$ для $(t+1)$ -го отсчета времени может быть вычислено по формуле:

$$z(t+1) = \frac{\sum_{i=1}^p r_i(t+1) \cdot \tau_i}{\sum_{i=1}^p r_i(t+1)},$$

где p – количество интервалов x_i ($i = \overline{1, p}$); τ_i – центроид i -го интервала, полученный по формуле (2); $r_i(t+1)$ – количество повторений НМ V_i в правой части ГНЛЗ, описывающей прогноз для $(t+1)$ -го отсчета времени.

Для оценки качества модели прогнозирования применяется средняя относительная ошибка прогнозирования $AFER$:

$$AFER = \frac{\sum_{t=k}^n |(f(t) - d(t)) / d(t)|}{n - k + 1} \cdot 100\%; \quad AFER = \frac{\sum_{t=k+1}^n |(f(t) - d(t)) / d(t)|}{n - k} \cdot 100\%, \quad (3)$$

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

Первая из формул используется в том случае, когда ВР представлен значениями прогнозируемого фактора, вторая – если ВР представлен значениями приращений прогнозируемого фактора.

Так как любой реальный процесс подвергается воздействию множества различных факторов, то очевидно, что использование дополнительной информации позволит повысить точность прогнозирования. В связи с этим было принято решение о разработке двухфакторной модели прогнозирования, в которой используются основной (прогнозируемый) ВР $d(t)$ и вспомогательный ВР $e(t)$ ($t = \overline{0, n}$).

В качестве базовой двухфакторной модели прогнозирования на основе НМ была использована модель, предложенная Л.В. Ли, Л.Х. Вангом и С.М. Ченом. Как и в случае однофакторной модели, наряду с представлением основного и вспомогательного ВР с помощью значений факторов, было предложено использовать представление ВР с помощью значений приращений факторов. Кроме того, предложено вместо эвристической формулы, используемой в базовой двухфакторной модели и реализующей непосредственное прогнозирование на 1 шаг вперед, осуществлять прогнозирование на основе сформированных ГНЛЗ с пустыми правыми частями (как и в однофакторной модели прогнозирования). Оценка качества двухфакторной модели реализуется с использованием формул (3).

Разработка двухфакторной модели прогнозирования в целом аналогична разработке однофакторной модели. Однако в этом случае для основного (прогнозируемого) и вспомогательного ВР, описывающих некоторые факторы, определяются соответственно универсумы X и Y , количество интервалов их разбиения p и q , лингвистические термы A_1, A_2, \dots, A_p и B_1, B_2, \dots, B_q , а также фаззифицированные значения временных рядов V_1, V_2, \dots, V_p и W_1, W_2, \dots, W_q .

ГНЛЗ строятся на основе НЛЗ, формируемых для основного ВР с использованием фаззифицированных значений основного и вспомогательного ВР:

$$((V_{j_k}, W_{c_k}), (V_{j_{(k-1)}}, W_{c_{(k-1)}}), \dots, (V_{j_1}, W_{c_1})) \rightarrow V_{l_1}, V_{l_2}, \dots, V_{l_g}. \quad (4)$$

Числовые значения, формирующие вспомогательный ВР, явно не участвуют в формировании прогнозных значений основного ВР, но позволяют уточнить вид правых частей в ГНЛЗ (4). Дефаззифицированное значение величины $z(t+1)$ для $(t+1)$ -го отсчета времени ($t = \overline{k, n-1}$) определяется как

$$z(t+1) = (r_{l_1} \cdot h_{l_1} + r_{l_2} \cdot h_{l_2} + \dots + r_{l_g} \cdot h_{l_g}) / (r_{l_1} + r_{l_2} + \dots + r_{l_g}),$$

где $r_{l_1}, r_{l_2}, \dots, r_{l_g}$ – число повторов НМ $V_{l_1}, V_{l_2}, \dots, V_{l_g}$ в правых частях ГНЛЗ вида (4), а величины $h_{l_1}, h_{l_2}, \dots, h_{l_g}$ находятся в соответствии с формулой:

$$h_l = \begin{cases} (\chi_l + 0,5 \cdot \chi_{l+1}) / (1 + 0,5), & \text{если } l = 1; \\ (0,5 \cdot \chi_{l-1} + \chi_l + 0,5 \cdot \chi_{l+1}) / (0,5 + 1 + 0,5), & \text{если } 2 \leq l \leq p - 1; \\ (0,5 \cdot \chi_{l-1} + \chi_l) / (0,5 + 1), & \text{если } l = p, \end{cases}$$

где $l \in \{1, \dots, l_g\}$; χ_{l-1} , χ_l , χ_{l+1} – средние точки интервалов x_{l-1} , x_l, \dots, x_{l+1} .

Самостоятельной задачей является определение оптимальных значений свободных параметров моделей, обеспечивающих максимальную точность прогнозирования в смысле получения минимального значения *AFER*.

В качестве свободных параметров в предлагаемых моделях прогнозирования рассматриваются числа корректировки границ диапазонов изменения ВР, порядок модели, а также количество интервалов разбиения универсумов, на которых определены ВР. При этом значения свободных параметров должны выбираться таким образом, чтобы выполнялось требование об отсутствии ГНЛЗ с пустыми правыми частями. Данное требование выполняется, если в каждый интервал разбиения универсума попадает хотя бы одно значение соответствующего ВР.

Для выбора параметров моделей прогнозирования предлагается использовать классический МГА, поскольку применение переборных методов поиска требует значительных временных затрат. Использование классических методов поиска экстремума, например градиентных, также априори невозможно, так как искомые параметры моделей прогнозирования явно не входят в запись целевой функции, которая в данном случае определяется по одной из формул (3). Хромосомы для однофакторной и двухфакторной моделей k -го порядка имеют соответственно вид: $S = (D_1, D_2, p, k)$ и $S = (D_1, D_2, E_1, E_2, p, q, k)$, где D_1, D_2, E_1, E_2 – положительные действительные числа, используемые при корректировке левой и правой границ универсумов основного и вспомогательного ВР; p и q – количество интервалов разбиения универсумов этих ВР соответственно; k – порядок модели. Для каждого гена в хромосоме задается диапазон возможного изменения.

При реализации МГА для исключения из популяции хромосом, не удовлетворяющих условию отсутствия ГНЛЗ с пустыми правыми частями, предложено использовать функцию соответствия следующего вида:

$$F(S) = \begin{cases} AFER, & \text{если определены все правые части} \\ & \text{в группах логических зависимостей;} \\ 100, & \text{если не определена хотя бы одна правая} \\ & \text{часть в группах логических зависимостей,} \end{cases}$$

где *AFER* вычисляется по одной из формул (3).

Хромосома, для которой в результате выполнения заданного количества поколений G МГА достигается минимальное значение функции соответствия, определяет оптимальный набор параметров модели прогнозирования.

Использование МГА для поиска оптимальных параметров моделей прогнозирования позволяет улучшить результаты прогнозирования до 2-5 раз по сравнению с результатами, получаемыми при экспертном выборе этих параметров.

Разработанный МГА имеет полиномиальную оценку вычислительной сложности: $O(G \cdot P^2 + G \cdot n^2 \cdot P)$, на которую наиболее заметное влияние оказывают размер популяции P и длина ВР n .

Третья глава диссертационной работы посвящена разработке *моделей прогнозирования ВР на основе МАКО*, реализующих краткосрочное прогнозирование на 1 шаг вперед посредством автоматического подбора аналитических зависимостей, наилучшим образом аппроксимирующих известные значения ВР, с использованием алгоритма клонального отбора.

При реализации МАКО искомая аналитическая зависимость кодируется в виде антителя, которое представляется как строка символов и должно обеспечивать распознавание антигенов, т.е. элементов прогнозируемого ВР. С целью разработки способа корректного кодирования антители предлагается использовать три предварительно заданных алфавита: *Operation* = {'+', '-', '*', '/'} - алфавит арифметических операций, включающий операции сложения, вычитания, умножения и деления; *Functional* = {'Q', 'S', 'C', 'E', 'L', '.'} - алфавит функционалов, где символы 'Q', 'S', 'C', 'E', 'L' соответствуют математическим функциям «квадратный корень», «синус», «косинус», «экспонента» и «натуральный логарифм», а символ '.' задает отсутствие функционала; *Terminal* = {'a', 'b', ..., 'z', '@'} - алфавит терминалов, в котором символы 'a', 'b', ..., 'z' определяют аргументы аналитической зависимости, а символ '@' представляет собой условное обозначение некоторой константы. Количество символов в алфавите терминалов зависит от максимально возможного порядка модели K . При этом $K \geq k$, где k – реальный порядок модели, что возможно при наличии констант и кратности вхождения некоторых аргументов в аналитическую зависимость.

Разработанная рекурсивная процедура формирования аналитических зависимостей, основанная на свойствах бинарных деревьев, обеспечивает корректное преобразование строки символов, кодирующих антители, в аналитическую зависимость. Антители (строка символов) формируются с конца путем последовательной записи всех узлов бинарного дерева справа налево и снизу вверх, начиная с терминальных узлов. При этом терминальные узлы могут содержать только символы из алфавита терминалов, а остальные узлы, формируются из элементов алфавитов арифметических операций и функционалов. На рис. 1 показаны шаги рекурсивной процедуры для случая, когда максимальное количество возможных терминальных позиций в антители равно *Term* и определяется максимально возможным порядком модели ($Term = K$), а также приведен пример кодирования аналитической зависимости.

МАКО включает подготовительную часть, реализующую формирование начальной популяции антители, и итерационную часть, состоящую из следующих шагов: упорядочение антители по возрастанию значений аффинитета; отбор и клонирование определенной доли «лучших» антители, имеющих наименьшие значения аффинитета; гипермутация клонов антители; самоуничтожение клонов антители, «похожих» на другие клоны и антители исходной популяции; вычисление аффинитета клонов антители и формирование новой популяции; супрессия полученной популяции; генерация новых антители и добавление их к текущей популяции до получения ее исходного размера P ; проверка условия достижения заданного коли-

чества поколений G и завершение работы МАКО при выполнении этого условия, а в противном случае – повтор итерационной части. В качестве функции аффинности, являющейся аналогом функции соответствия в МГА, используется средняя относительная ошибка прогнозирования $AFER$ вида (3).

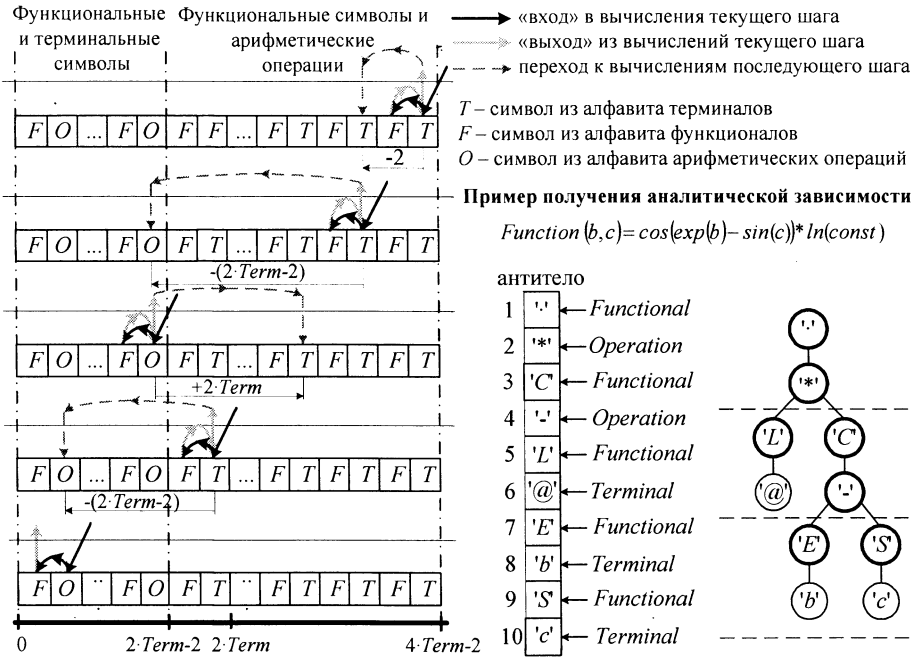


Рис. 1. Рекурсивная процедура формирования аналитических зависимостей

С использованием МАКО разработаны однофакторная и двухфакторная модели прогнозирования k -го порядка. При этом может использоваться представление ВР как непосредственно на основе значений факторов, так и с помощью значений приращений факторов. Второй вариант является более предпочтительным, как и для моделей прогнозирования на основе НМ и МГА. Главное отличие двухфакторной модели прогнозирования от однофакторной заключается в том, что в искомой аналитической зависимости присутствуют элементы как основного, так и вспомогательного ВР. Если максимально возможный порядок модели прогнозирования равен K , то алфавит терминалов должен содержать $(K+1)$ и $(2 \cdot K+1)$ символов для однофакторной и двухфакторной моделей соответственно. Длина антитела зависит от количества символов в алфавите терминалов и равна $L = 4 \cdot K - 2$.

Использование МАКО для поиска оптимальных аналитических зависимостей позволяет улучшить результаты прогнозирования до 4-8 раз по сравнению с результатами, получаемыми при экспертном подборе аналитических зависимостей, а также до 1,5-3 раз по сравнению с результатами прогнозирования с использованием моделей на основе НМ и МГА.

Разработанный МАКО имеет полиномиальную оценку вычислительной сложности: $O(G \cdot (n \cdot L \cdot P + P^2 + (\ln(P) \cdot L \cdot P)^2 + \ln(P) \cdot ((L \cdot P)^2 + n \cdot L \cdot P)))$, которая существенно зависит от размера популяции P и длины антитела L .

Четвертая глава, посвященная программной реализации и исследованию разработанных моделей прогнозирования, содержит описание структуры, принципов работы и примеров использования ППП «Модели прогнозирования с использованием гибридных технологий», реализующего прогнозирование на основе моделей прогнозирования, предложенных в главах 2 и 3.

ППП предназначен для решения задач краткосрочного прогнозирования (на 1 шаг вперед) процессов, представленных ВР с короткой актуальной частью. Он разработан средствами системы инженерных и научных расчетов MATLAB 7.0 и состоит из двух комплексов программ, реализующих прогнозирование с использованием моделей на основе НМ и МГА (глава 2), а также моделей прогнозирования на основе МАКО (глава 3). Комплексы программ защищены авторским правом: получены 2 свидетельства о регистрации программ для ЭВМ в РОСПАТЕНТ.

Для выявления достоинств и недостатков разработанных моделей прогнозирования был выполнен вычислительный эксперимент по краткосрочному прогнозированию факторов, характеризующих уровень безработицы, уровень занятости и уровень жизни населения в России, позволивший сделать следующие выводы:

- представление ВР с помощью значений приращений факторов, а также учет вспомогательных факторов обеспечивает для всех моделей прогнозирования повышение точности прогнозирования (в смысле достижения минимального значения средней относительной ошибки);

- применение моделей прогнозирования на основе НМ и ГА может быть рекомендовано в случае, когда необходимо получить результаты прогнозирования приемлемой точности с минимальными временными затратами;

- применение моделей прогнозирования на основе МАКО сопровождается увеличением временных затрат, однако обеспечивает при этом повышение точности прогнозирования и, кроме того, реализует формирование аналитических зависимостей, адекватно аппроксимирующих известные значения ВР прогнозируемых факторов.

На рис. 2 приведены графические зависимости для реальных и прогнозируемых значений, полученных с помощью разработанных моделей прогнозирования, для фактора «занятое население», подтверждающие приведенные выше выводы. Для двухфакторных моделей прогнозирования в качестве вспомогательного использовался фактор «экономически активное население». Реальные (известные) значения прогнозируемого и вспомогательного ВР на основе значений соответствующих факторов показаны на рис. 2, а.

В следующей таблице приведены результаты прогнозирования на 1 шаг вперед для факторов «занятое население» и «экономически активное население» с использованием классических моделей (Бурга, авторегрессионной и полиномиальной) и разработанных однофакторных моделей. «Жирным», «жирным курсивом» и «курсивом» выделены «лучшие» результаты прогнозирования, упорядоченные по возрастанию значений относительной ошибки прогнозирования *Err*.

Реальное значение	Прогноз на 1 шаг, тыс.чел./ Относительная ошибка прогноза, %				
	Модель Бурга	Авторегрессионная модель	Полиномиальная модель	Модель на основе НМ и МГА	Модель на основе МАКО
64941 тыс. чел.	64502,823/ 0,675	65597,487/ 1,011	66212,376/ 1,958	66167,79/ 1,889	65340,814/ 0,616
71502 тыс. чел.	70900,609/ 0,841	71197,104/ 0,426	72033,477/ 0,743	71049,639/ 0,633	70964,397/ 0,752

Ошибка вычисляется по формуле $Err = |f(n+1) - d(n+1)/d(n+1)| \cdot 100\%$, где $y(n+1)$ и $d(n+1)$ – предсказанное и реальное значения показателя для $(n+1)$ -го отсчета времени соответственно.

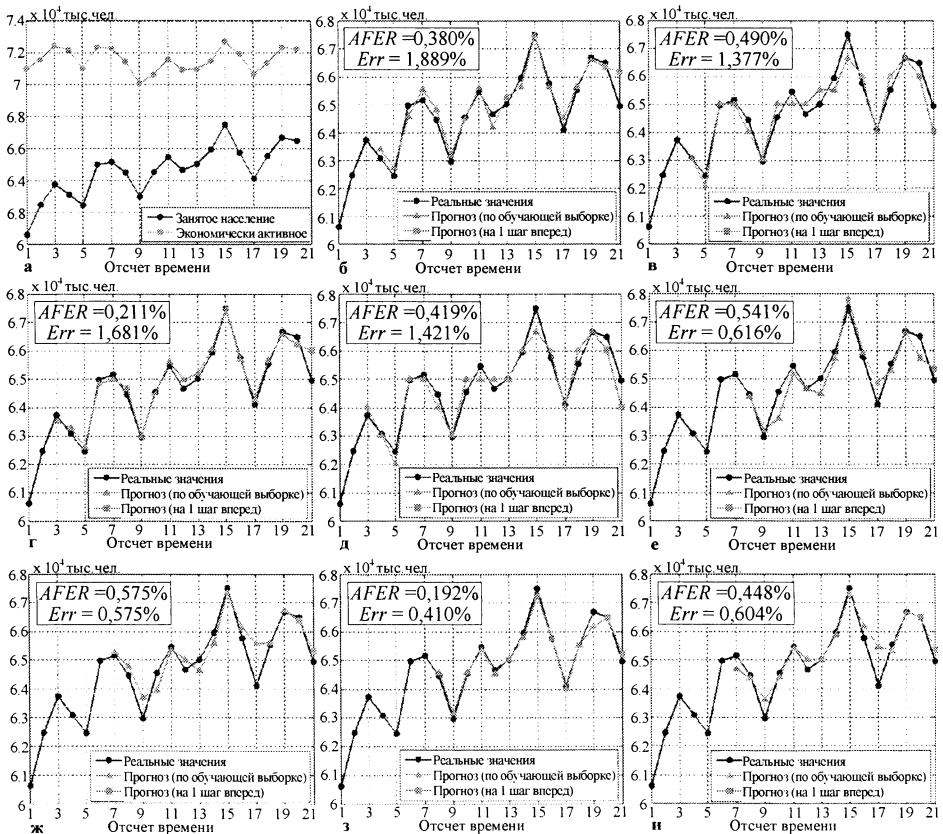


Рис. 2. Прогнозирование фактора «занятое население»: а – исходные прогнозируемый и вспомогательный ВР на основе значений соответствующих факторов; б, в, г, д – модели на основе НМ и МГА; е, ж, з, и – модели на основе МАКО; б, в, е, ж – однофакторные модели; г, д, з, и – двухфакторные модели; б, г, е, з – модели, в которых ВР представлены значениями приращений факторов; в, д, ж, и – модели, в которых ВР представлены значениями факторов (Err – относительная ошибка прогнозирования на один шаг вперед)

Следует отметить, что применение двухфакторных моделей прогнозирования позволило улучшить результаты прогнозирования для фактора «занятое население» (рис. 2).

Анализ результатов показывает, что предлагаемые модели позволяют получать результаты прогнозирования, близкие к результатам классических моделей, а в ряде случаев и лучшие, что свидетельствуют об их пригодности для выполнения краткосрочных прогнозов. При этом разработанные модели прогнозирования обеспечивают простоту реализации и наглядность представления процесса прогнозирования, не требуя от специалиста-аналитика выполнения расчета группы критериев качества модели с последующим их анализом для выбора оптимальных параметров модели, как в случае работы с классическими моделями.

Разработанные модели прогнозирования обеспечивают расширение информационно-аналитического пространства при решении задач краткосрочного прогнозирования за счет представления значений ВР с помощью НМ и формирования аналитических зависимостей с использованием МАКО, адекватно аппроксимирующих известные значения ВР прогнозируемых факторов.

Предложено использовать многоальтернативное прогнозирование, позволяющее компенсировать потенциальные ошибки применяемых моделей прогнозирования, обусловленные неполным восстановлением процессов частными прогнозирующими описаниями, посредством консолидации прогнозов.

Для вычисления результирующего прогноза на основе оставшихся после экспертной отбраковки прогнозов предлагается использовать два варианта консолидации, реализующие вычисление центра тяжести для одноточечных множеств и

взвешенного среднегеометрического: $Final_1 = \sum_{i=1}^M \alpha_i \cdot F_i$ и $Final_2 = \sum_{i=1}^M F_i^{\alpha_i}$ соот-

ветственно, где M – количество моделей прогнозирования, используемых для построения альтернативного прогноза; F_i – значение прогноза по i -й альтернативной модели прогнозирования; $\alpha_i = AFER_{I(i)} / S$ – весовой коэффициент i -й модели прогнозирования; $I(i)$ – массив номеров моделей прогнозирования, упорядоченных по возрастанию значений средней относительной ошибки прогнозирования; $AFER_{I(i)}$ – значение средней относительной ошибки $I(i)$ -й модели прогнозирования;

$S = \sum_{i=1}^M AFER_i$; $AFER_i$ – значение средней относительной ошибки прогно-

зирования по i -й модели прогнозирования ($i = \overline{1, M}$). При этом модели прогнозирования с меньшим значением средней относительной ошибки прогнозирования ставится в соответствие весовой коэффициент с большим значением.

Анализ результатов консолидированных прогнозов для различных процессов показал, что ни один из вариантов консолидации не обладает явным преимуществом. Следовательно, может быть использован любой из вариантов или оба одновременно для определения интервала принадлежности значения консолидированного прогноза.

В заключении приводится обобщение основных результатов диссертации.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Проведенные исследования позволяют сформулировать следующие основные результаты.

1. Выполнен анализ проблемы прогнозирования процессов, представленных временными рядами с короткой актуальной частью, выявлены достоинства и недостатки известных традиционных методов, моделей и алгоритмов прогнозирования временных рядов с короткой актуальной частью, определены перспективные направления исследований. Исследована возможность использования инструментария ТНМ, ГА и ИИС при разработке моделей краткосрочного прогнозирования (на 1 шаг вперед) временных рядов с короткой актуальной частью.

2. Разработаны и исследованы однофакторные и двухфакторные модели краткосрочного прогнозирования k -го порядка, которые работают с приемлемыми временными затратами и обеспечивают повышение точности прогноза за счет:

- применения групп нечетких логических зависимостей для описания поведения временных рядов и рационального выбора параметров моделей с помощью разработанного модифицированного генетического алгоритма;

- автоматического формирования аналитической зависимости, адекватно описывающей известные значения временного ряда, с использованием разработанного модифицированного алгоритма клонального отбора.

3. Разработана рекурсивная процедура формирования аналитических зависимостей, реализующая способ корректного кодирования антител в модифицированном алгоритме клонального отбора с помощью бинарных деревьев.

4. Разработана методика обобщения прогнозов, позволяющая компенсировать потенциальные ошибки применяемых моделей прогнозирования, обусловленные неполным восстановлением процессов частными прогнозирующими описаниями, посредством консолидации прогнозов.

5. Разработан пакет прикладных программ ППП для краткосрочного прогнозирования временных рядов с короткой актуальной частью на основе разработанных моделей прогнозирования.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Демидова Л.А., Скворцова Т.С. Применение генетических алгоритмов для прогнозирования нечетких временных рядов // Вестник Рязан. гос. радиотехн. ун-та. – 2010. – № 1 (вып. 31). – С. 95-98.

2. 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.

3. 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. – № 4.1(42). – С. 131-136.

4. Скворцова Т.С. Прогнозирование значений временных рядов с использованием аппарата теории нечетких множеств и генетического алгоритма / Свидетельство о гос. регистрации программы для ЭВМ в РОСПАТЕНТ, № 2009614678 от 01.09.2009.

5. Демидова Л.А., Скворцова Т.С. Прогнозирование значений временных рядов с использованием модифицированного алгоритма клонального отбора / Свидетельство о гос. регистрации программы для ЭВМ в РОСПАТЕНТ, № 2011613458 от 03.05.2011.

6. Демидова Л.А., Сковрцова Т.С. Разработка двухфакторной модели прогнозирования временных рядов с использованием генетического алгоритма // Математическое и программное обеспечение вычислительных систем: Межвуз. сб. – М.: Горячая линия – Телеком, 2008. – С. 99-108.

7. Сковрцова Т.С. Оценка сложности генетического алгоритма оптимизации параметров модели прогнозирования // Математическое и программное обеспечение вычислительных систем: Межвуз. сб. – М.: Горячая линия – Телеком, 2009. – С. 77-80.

8. Сковрцова Т.С. Разработка двухфакторной модели прогнозирования временных рядов с использованием алгоритма клонального отбора // Математическое и программное обеспечение вычислительных систем: Межвуз. сб. / Рязан. гос. радиотехн. ун-т. – Рязань, 2011. – С. 29–35.

9. Демидова Л.А., Сковрцова Т.С. Программный комплекс прогнозирования значений временных рядов с использованием гибридных технологий // Математическое и программное обеспечение вычислительных систем: Межвуз. сб. / Рязан. гос. радиотехн. ун-т. – Рязань, 2011. – С. 57-61.

10. Сковрцова Т.С. Анализ задачи прогнозирования временных рядов на основе двухфакторной модели с использованием генетического алгоритма // Проблемы передачи и обработки информации в сетях и системах телекоммуникаций: материалы 15-ой международной научно-технической конференции. Часть 1. – Рязань: РГРТУ, 2008. – С. 37-39.

11. Сковрцова Т.С. Анализ и разработка нечеткой модели на основе генетического алгоритма для прогнозирования временных рядов // Проблемы информатики в образовании, управлении, экономике и технике: сборник статей VIII Всероссийской научно-технической конференции. – Пенза: Приволжский Дом знаний, 2008. – С. 252-254.

12. Сковрцова Т.С. Использование генетического алгоритма для настройки параметров нечеткой модели при прогнозировании временных рядов // Новые информационные технологии в научных исследованиях и в образовании: материалы XIII Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. Часть 1. – Рязань: РГРТУ, 2008. – С. 117-119.

13. Сковрцова Т.С. Особенности реализации генетического алгоритма для прогнозирования нечетких временных рядов // Методы и алгоритмы принятия эффективных решений: материалы международной научной конференции. Часть 2. – Таганрог: Изд-во ТТИ ЮФУ, 2009. – С. 81-83.

14. Сковрцова Т.С. Сравнительный анализ методов прогнозирования временных рядов с короткой длиной актуальной части // Новые информационные технологии в научных исследованиях и в образовании: материалы XIV Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. – Рязань: РГРТУ, 2009. – С. 218-220.

15. Сковрцова Т.С. Особенности применения генетического алгоритма для анализа временных рядов // Тенденции развития современных информационных технологий, моделей экономических, правовых и управленческих систем: сб. статей V международной научно-практической конференции. – Рязань: Ряз. филиал МЭСИ, 2010. – С. 190-192.

16. Сковрцова Т.С. Разработка модели прогнозирования временных рядов с использованием алгоритма клонального отбора // Математические методы и информационные технологии в экономике, социологии и образовании: сборник статей XXVI Международной научно-технической конференции. – Пенза: Приволжский Дом знаний, 2010. – С. 107-109.

17. Демидова Л.А., Сковрцова Т.С. Использование инструментария искусственных иммунных систем в задачах прогнозирования значений временных рядов // Современные проблемы информатизации в экономике и обеспечении безопасности: сб. трудов Международной открытой научной конференции. Вып. 16. – Воронеж: Научная книга, 2011. – С. 84-87.

Скворцова Татьяна Сергеевна

**МОДЕЛИ ПРОГНОЗИРОВАНИЯ ПРОЦЕССОВ,
ПРЕДСТАВЛЕННЫХ ВРЕМЕННЫМИ РЯДАМИ
С КОРОТКОЙ АКТУАЛЬНОЙ ЧАСТЬЮ**

А в т о р е ф е р а т
диссертации на соискание учёной степени
кандидата технических наук

Подписано в печать 03.11.2011 г. Формат бумаги 60×84 1/16.
Усл. печ. л. 1,0. Уч.-изд. л. 1.0. Тираж 100 экз.
Рязанский государственный радиотехнический университет.
390005, г. Рязань, ул. Гагарина, д. 59/1.