

На правах рукописи



Егошин Алексей Валерьевич

**АНАЛИЗ И ПРОГНОЗИРОВАНИЕ СЛОЖНЫХ
СТОХАСТИЧЕСКИХ СИГНАЛОВ НА ОСНОВЕ МЕТОДОВ
ВЫДЕЛЕНИЯ ГРАНИЦ РЕАЛИЗАЦИЙ
ДИНАМИЧЕСКИХ СИСТЕМ**

Специальность 05.13.18 – «Математическое моделирование,
численные методы и комплексы программ»

Автореферат

диссертации на соискание ученой степени
кандидата технических наук

Санкт-Петербург – 2009

Работа выполнена в государственном образовательном учреждении высшего профессионального образования «Марийский государственный технический университет».

| | |
|------------------------|--|
| Научный руководитель: | доктор технических наук, профессор Сидоркина Ирина Геннадьевна |
| Официальные оппоненты: | доктор технических наук, профессор Коробейников Анатолий Григорьевич кандидат технических наук, доцент Бондаренко Игорь Борисович |
| Ведущая организация: | Казанский государственный технический университет им. А.Н. Туполева |

Защита диссертации состоится «23» декабря 2009 г. в 15 ч. 30 мин. на заседании диссертационного совета Д 212.227.06 при Санкт-Петербургском государственном университете информационных технологий, механики и оптики по адресу: 197101, г. Санкт-Петербург, пр. Кронверкский, д.49.

С диссертацией можно ознакомиться в научной библиотеке Санкт-Петербургского государственного университета информационных технологий, механики и оптики.

Автореферат разослан «_____» _____ 2009 г.

Ученый секретарь диссертационного совета,
доктор технических наук, профессор

 Лисицына Л.С.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Современная хозяйственно-экономическая деятельность человека не возможна без принятия решений, основанных на прогнозировании развития текущей ситуации. Это справедливо для многих сфер: в экономике, в промышленном производстве при управлении технологическими процессами, здравоохранении и др.

Автоматизация сбора данных, широкое использование информационных технологий и вычислительных средств обработки числовой информации связаны с накоплением и обработкой значительных объемов наблюдений. Временные ряды (значения сигналов) при этом получают значительной длины с высокой степенью дискретизации, что позволяет рассматривать их как эволюционные. Для сложных стохастических нестационарных сигналов сам источник можно рассматривать как нелинейную динамическую систему (ДС) с неизвестными уравнениями фазовой траектории движения («черный ящик»), где сам сигнал является наблюдаемой реализацией данной системы.

Задачу прогнозирования временных рядов решали отечественные и зарубежные специалисты: Ю. И. Журавлев, Г. Г. Малинецкий, А.Г. Беляков, А.Г. Ивахненко, В.П. Боровиков, А.С. Мандель, А.А. Френкель, Е.М. Четыркин, Х. Акаике, Д. Бокс, Р. Браун, Д. Бриллинджер, Г. Дженкинс и др.

Среди основных моделей и технологий прогнозирования можно выделить: авторегрессионные модели, нейронные сети, методы классификации в контексте прогнозирования, метод группового учета аргументов, метод Гусеница, вейвлет-анализ. Данные и другие разработки применяются при построении прогнозов, однако в области применения их к стохастическим, нестационарным эволюционным рядам трудно выделить наиболее подходящий метод.

Так как все известные методы прогнозирования основываются на анализе прошлых значений сигнала, встает вопрос выбора объема ретроспективы, которая, с одной стороны, не будет учитывать явно устаревшие значения, а с другой, – будет достаточно длинной, чтобы уловить глобальную тенденцию движения сигнала. Поэтому наряду с разработкой методов прогнозирования остро стоит проблема разработки метода математически обоснованного выбора ретроспективного объема данных для использования их в прогнозирующем аппарате. Решение данного вопроса предлагается на основе задачи о разладке, теории ди-

намических систем и модифицированной парадигмы русел и джокеров Г.Г. Малинецкого. В ее рамках временной ряд рассматривается как наблюдаемые реализации ДС с неизвестными уравнениями фазовой траектории («черный ящик»), которые эволюционно (или скачками) переходят одна в другую. Тогда для построения достоверного прогноза используется ретроспектива временного ряда, являющаяся реализацией последней динамической системы на этапе эволюции глобальной ДС, так как каждая такая система проще (ее можно рассматривать как квазидетерминированную), чем их совокупность, а значит, и смоделировать ее легче. Это дает возможность более точно идентифицировать текущую динамику изменения сигнала, что, в конечном итоге, позволит строить в общем случае более точный прогноз, чем при выборе ретроспективы, исходя только из чисто эмпирических соображений. Поэтому актуальным является решение задачи определения границ реализаций ДС по наблюдаемому сигналу.

Цель диссертационной работы – разработка методов и алгоритмов выделения в сложном стохастическом сигнале границ перехода между динамическими системами, генерирующими сигнал и повышение на основе этого эффективности прогнозирования. Достижение поставленной цели предполагается осуществить на основе решения следующих **задач**:

- 1) провести анализ современных методов прогнозирования и обнаружения разладки в сложных стохастических сигналах;
- 2) разработать метод выделения границы реализации динамических систем на основе обнаружения изменения в локальной фрактальной размерности сигнала;
- 3) разработать методы предобработки временного ряда, позволяющие выявить временные характеристики сигнала;
- 4) разработать метод прогнозирования на основе анализа времени достижения заданного порога изменения сигнала;
- 5) разработать метод прогнозирования на основе локальных экстремумов сигнала заданного порога;
- 6) разработать программный комплекс, реализующий описанные выше методы обнаружения разладки и прогнозирования.

Объект исследования – сложный стохастический наблюдаемый сигнал.

Предмет исследования – методы обнаружения границы перехода в сложных стохастических сигналах от одной динамики к другой, позволяющие выделить для обучения предиктора фрагмент сигнала с квази-

стабильной динамикой, прогнозирующие модели на основе аппроксимации и классификации.

Методы исследования. Для решения поставленных задач использовались методы анализа фрактальности сигнала, методы теории хаоса, математической статистики и нелинейной динамики, спектральный анализ, нейронные сети.

Научные положения, выносимые на защиту:

1) метод обнаружения границы реализации динамических систем в сложном сигнале по наличию разладки в самом сигнале и локальной фрактальной размерности сигнала;

1) метод оценки степени уверенности обнаружения разладки в сигнале;

2) метод прогнозирования на основе анализа времени достижения заданного порога изменения сигнала;

3) способ повышения эффективности прогнозирования на основе анализа времени достижения заданного порога изменения сигнала.

4) метод прогнозирования на основе локальных экстремумов сигнала заданного порога.

Научные результаты. Проведенные исследования показали применимость выдвинутой гипотезы о сложном стохастическом сигнале как последовательности реализаций динамических систем, эволюционно или скачкообразно переходящих одна в другую. Получены следующие основные научные результаты:

1. Разработан метод обнаружения границы реализаций динамических систем в сложном сигнале по наличию разладки в самом сигнале и стабильности локальной фрактальности сигнала, позволяющий повысить точность обнаружения разладки в зашумленных сигналах по сравнению с известными.

2. Предложен метод оценки степени уверенности обнаружения разладки в сигнале по количеству локальных экстремумов заданного порога в решающей функции метода обнаружения, позволяющий получить численное значение степени наличия разладки в сигнале, что актуально для сложных сигналов, в которых разладку можно обнаружить и на постоянной динамике.

3. Разработан метод прогнозирования на основе преобразования временного ряда сигнала в ряд значений времени достижения заданного порога изменения, позволяющий агрегировать сигнал, отфильтровывая несущественные в практическом смысле колебания и подходить к задаче прогнозирования, как к задаче определения времени, которое потребуется для того, чтобы сигнал изменился на заданный порог.

4. Предложен способ повышения эффективности метода прогнозирования на основе преобразования временного ряда сигнала в ряд значений времени достижения заданного порога изменения.

5. Разработан метод прогнозирования на основе локальных экстремумов сигнала заданного порога, позволяющий прогнозировать существенные, в практическом отношении, изменения сигнала на основе отнесения локального экстремума к определенному классу в многомерном пространстве признаков.

Практическая ценность. Разработано программное обеспечение, реализующее методы и алгоритмы обнаружения границ реализаций динамических систем и прогнозирования. Данное обеспечение может применяться для анализа и прогнозирования сложных стохастических сигналов различного генезиса, например технологических данных и др.

Достоверность полученных результатов. Обоснованность и достоверность положений, выводов и рекомендаций подтверждаются использованием классических методов анализа и моделирования сигналов, методов математической статистики, нелинейной динамики, теории хаоса, а также вычислительными экспериментами.

Публикации и апробация результатов работы. Основные результаты настоящей диссертации опубликованы в 6 работах, в том числе одна в рекомендованном для публикации журнале по списку ВАК. Материалы диссертационной работы докладывались и обсуждались на Всероссийской конференции «Технологии Microsoft в теории и практике программирования» (Нижний Новгород, 2007); VI Международной научно-технической конференции «Информационно-вычислительные технологии и их приложения» (Пенза, 2007); Всероссийской научно-практической конференции с международным участием «Информационные технологии в профессиональной деятельности и научной работе» (Йошкар-Ола, 2007); Всероссийской научно-практической конференции с международным участием «Информационные технологии в профессиональной деятельности и научной работе» (Йошкар-Ола, 2008); Всероссийской научно-практической конференции с международным участием «Информационные технологии в профессиональной деятельности и научной работе» (Йошкар-Ола, 2009). Получено 1 свидетельство о государственной регистрации программы для ЭВМ № 2009616119.

Реализация и внедрение результатов работы. Результаты диссертационной работы внедрены в научно-исследовательскую деятельность «Санкт-Петербургского филиала учреждения Российской академии наук института земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова РАН» (СПбФ ИЗМИРАН), а именно: принята в эксплуатацию разработанная программная система, позволяющая

обнаруживать смену динамики в сложном сигнале и прогнозировать его дальнейшее изменение, используя информацию о границе перехода от одной динамики к другой.

Результаты работы внедрены в учебный процесс кафедры информационно-вычислительных систем Марийского государственного технического университета. Учебно-методические разработки используются студентами при изучении дисциплины «Интеллектуальные системы» специальности 230105 «Программное обеспечение вычислительной техники и автоматизированных систем».

Разработанный программный комплекс внедрен в ООО «Вятка-софт» (Санкт-Петербург) для анализа и прогнозирования посещаемости web-сервисов компании.

Имеются соответствующие акты о внедрении.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения, списка литературы и приложения. Рукопись содержит 174 страницы текста, 77 рисунков, 31 таблицу.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность темы диссертационной работы, описываются предмет и объект исследования, формулируются цель и задача исследования, научная новизна и практическая значимость результатов.

В **первой главе** дается обзор основных методов и подходов прогнозирования сложных сигналов. Оцифровка ранее не анализируемых наблюдаемых сигналов, оцифровка с более высокой частотой дискретизации привели к применению в задаче анализа и прогнозирования сложных стохастических сигналов методов теории хаоса и нелинейной динамики. Г. Г. Малинецкий ввел понятие русел и джокеров для сложных динамических систем (ДС), где под руслом понимается маломодовая (в пространстве низкой размерности) предсказуемая динамика, а под джокером – область скачкообразного изменения состояния системы, переход в точку фазового пространства, координаты которого задаются вероятностными значениями. Развитие этой модели привело к рассмотрению сложного сигнала как последовательности реализаций ДС, где каждое следующее состояние задается различным отображением в зависимости от времени:

$$\begin{aligned}
\mathbf{x}(t+1) &= \mathbf{F}_1(\mathbf{x}(t), \mathbf{x}(t-1), \dots, \mathbf{x}(t-k_1)), \quad t = 0, 1, \dots, r_1, \\
\mathbf{x}(t+1) &= \mathbf{F}_2(\mathbf{x}(t), \mathbf{x}(t-1), \dots, \mathbf{x}(t-k_2)), \quad t = r_1 + 1, \dots, r_2, \\
&\dots \\
\mathbf{x}(t+1) &= \mathbf{F}_n(\mathbf{x}(t), \mathbf{x}(t-1), \dots, \mathbf{x}(t-k_n)), \quad t = r_n, \dots, N-1,
\end{aligned} \tag{1}$$

$$\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_d(t)],$$

где t – моменты времени наблюдения; $\mathbf{x}(t)$ – d -мерный вектор состояния системы в момент времени t , состоящий из компонент $x_c(t)$, $c = 1..d$; n – число точек смены динамики; \mathbf{F}_i – вектор-функции, определяющие следующее состояние системы в различные периоды времени и имеющие k_i последних значений $\mathbf{x}(t)$ в качестве аргументов, r_i – точки смены динамики сигнала; $i = 1..n$; N – число наблюдений сигнала. Данная модель представления легла в основу анализа и прогнозирования сложных сигналов.

Общим у всех методов прогнозирования является использование ретроспективы сигнала для «обучения» аппарата прогнозирования – предиктора. Глобальные методы для обучения используют весь доступный временной ряд, локальные – только определенную часть. Применение к сложным стохастическим сигналам глобальных методов сопряжено с большими трудностями. Как показано в (1), на различных участках сигнал может описываться различными уравнениями. Поэтому предложено разделить сигнал на участки квазистационарности, с тем, чтобы анализировать и использовать для обучения предиктора последний «стабильный» по некоторой статистической мере фрагмент временного ряда.

Такая задача известна как задача обнаружения «разладки». Под разладкой понимается изменение вероятностных свойств случайного процесса – например математического ожидания или дисперсии. Как показано в работе Б. С. Дарховского, любую разладку в сигнале по статистической мере можно рассматривать как изменение математического ожидания. Однако для сложных сигналов, порождаемых нелинейными динамическими системами или сильно зашумленных, такой критерий может давать неточные оценки момента разладки. Поэтому предложено рассматривать сигнал во фрактальном представлении и искать разладку как изменение локальной фрактальности сигнала. Для этого используется индекс фрактальности, предложенный Н. В. Старченко. Под точкой разладки будем понимать момент времени смены динамики сигнала, что можно считать границей реализаций различных динамических систем.

Во **второй главе** описывается применение локального фрактального анализа для обнаружения разладки. Фрактальный анализ достаточно широко используется для анализа сложных сигналов и представлен главным образом R/S-анализом, по которому дается оценка показателя Херста, через который вычисляется фрактальная размерность ряда. Особенностью R/S анализа является то, что для получения адекватной оценки показателя Херста требуется достаточно большое количество отсчетов сигнала – несколько тысяч. Это не позволяет использовать его для определения локальной фрактальности сигнала на практике.

Для обнаружения смены динамики сигнала предложено искать изменение локальной фрактальной размерности в статистическом смысле (по изменению среднего или дисперсии). Для локального фрактального анализа существует процедура расчета индекса фрактальности на основе минимального покрытия графика сигнала. Индекс фрактальности позволяет получать оценку фрактальной размерности сигнала по небольшому числу отсчетов – от нескольких десятков до нескольких сотен. Эта особенность позволяет давать локальную оценку фрактальной размерности сигнала по его фрагменту.

Обнаружение разладки по изменению абсолютного значения индекса фрактальности (выделение фрагментов сигнала, соответствующих различным классам по Фурье-спектру) не может применяться в сложных сигналах, так как фрагменты сигнала, имеющие различную динамику, могут иметь один и тот же класс Фурье-спектра. Поэтому предложено использовать производный сигнал индекса фрактальной размерности $\mu(t)$, построенного по принципу скользящего окна, от наблюдаемого сигнала $x(t)$ для обнаружения в нем разладки. Это позволяет определять смену динамики сигнала $x(t)$ как изменение степени стационарности его локальной фрактальной размерности.

Для обнаружения изменения однородности производного сигнала локальной фрактальной размерности использованы классические методы обнаружения разладки. Рассмотрены известные апостериорные (использующие весь доступный временной ряд) непараметрические методы, так как распределение вероятностей наблюдаемого сигнала до и после разладки, как и распределение момента возникновения разладки во времени, нам неизвестны. Проведено сравнение следующих методов: общий случай алгоритма Бродского-Дарховского, алгоритм кумулятивных сумм (АКС), метод на основе принципа минимума информационного рассогласования. В результате сравнения на модельных сигналах авторегрессии с шумом наиболее эффективным по точности и скорости обнаружения оказался АКС.

На основе исследования предложен метод обнаружения границы реализаций динамических систем в сложном сигнале, заключающийся в следующем. По исходному сигналу строится производный сигнал локальной фрактальной размерности. В этом производном и исходном сигналах методом АКС находятся моменты разладки. Так как задача обнаружения смены динамики сигнала решается для того, чтобы выделить последний фрагмент сигнала, сформированный квазистабильной ДС, не затрагивая при этом соседний фрагмент, сформированный другой ДС, предложены три способа оценки момента разладки, применяемые в зависимости от сложности и длины сигнала:

- 1) максимальный – выбирается наибольший момент, для минимизации захвата соседнего фрагмента;
- 2) средний – берется среднее значение по двум моментам;
- 3) минимаксный – берется среднее значение по среднему моментов и максимальному моменту.

Поскольку на сложном сигнале даже на постоянной динамике методы обнаружения разладки могут давать ложные обнаружения, предложен метод определения степени уверенности обнаружения разладки. Главный принцип метода – в случае истинной смены динамики график решающей функции метода обнаружения $Y(t)$ имеет один ярко выраженный глобальный экстремум $e_{\max} = |Y(t)|_{\max}$. В случае анализа хаотического сигнала с однородной динамикой, например отображения Хенона, или процесса авторегрессии первого порядка со значительной шумовой компонентой, график будет иметь несколько выраженных экстремумов $e_1, e_2, \dots, e_{\max}$ (указаны по возрастанию абсолютные значения) относительно не сильно отличающихся друг от друга по значению, e_{\max} – максимальное абсолютное значение решающей функции $Y(t)$ (положение по времени соответствует моменту разладки в сигнале). Под выраженным понимается локальный экстремум e_v , не меньше qe_{\max} , то есть $e_v \geq qe_{\max}$, $0 < q < 1$, и слева и справа существуют противоположные по знаку экстремумы e_l и e_r (или начальные, или конечные точки анализируемого сигнала), относительная разница между которыми не меньше p , т. е. $|(e_l - e_v)/e_v| \geq p$, $|(e_r - e_v)/e_v| \geq p$, $0 < p < 1$. Причем между рассматриваемым экстремумом e_v и соседними e_l и e_r нет экстремума того же знака, что и у e_v , меньшего

чем e_v . В качестве выраженных рассматриваются локальные максимумы там, где $Y(t) > 0$ и минимумы там, где $Y(t) < 0$. Поэтому степень уверенности в разладке определяется как

$$c(q, p) = 1/N_{q,p} \times 100\%, \quad (2)$$

где p и q – пороги значимости экстремума; $N_{q,p}$ – количество экстремумов, удовлетворяющих порогам q и p . Для сигналов, анализируемых в работе, использованы значения $q = 0,7$; $p = 0,5$.

В **третьей главе** предложены методы прогнозирования на основе ряда значений времени достижения заданного порога изменения сигнала (основной) и на основе анализа локальных экстремумов заданного порога (дополнительный).

В основу первого метода положено преобразование исходного сигнала в ряд значений времени достижения порога изменения p . Преобразование к ряду времени достижения порога изменения позволяет, во-первых, агрегировать сигнал без потери существенной информации о нем, во-вторых, подойти к задаче прогнозирования с другой точки зрения. В частности, прогнозировать не следующее значение сигнала, а время, через которое изменение сигнала превысит заданный порог, т.е. исходный сигнал преобразовывается следующим образом:

$$\begin{aligned} x(t) &\equiv \{x_1, x_2, \dots, x_N\} \rightarrow x' \equiv \{x'_1, x'_2, \dots, x'_{N'}\}, \\ x_i &= x(t_i), \quad i = 1 \dots N, \quad x'_j = x(t'_j), \quad j = 1 \dots N', \\ |(x'_i - x'_{i-1}) / x'_{i-1}| &\geq p, \quad p > 0, \quad i = 2, 3, \dots, N', \\ \tau &= \{\tau_1, \tau_2, \dots, \tau_{N'-1}\}, \quad \tau_i = t'_{i+1} - t'_i, \quad i = 1 \dots N'-1, \end{aligned}$$

где N – число отсчетов в исходном сигнале; x' – преобразованный сигнал, в котором оставлены только те значения, относительная разница между которыми превышает порог p ; N' – число отсчетов в преобразованном сигнале; τ – ряд времени достижения заданного порога изменения, где каждое значение означает время, которое потребовалось сигналу для того, чтобы превысить порог изменения p . Прогнозирование осуществляется нейронной сетью (НС) типа многослойный персептрон.

Такой метод позволяет выполнять комбинированное прогнозирование, которое включает в себя:

- 1) оценку по ряду x' следующего значения – $x'_{N'+1}$;
- 2) оценку по ряду τ следующего значения – $\tau_{N'+1}$;

3) сравнение $\text{sign}(\tau_{N'+1})$ и $\text{sign}(x'_{N'+1} - x'_{N'})$.

Сопоставление результатов, выполненных по рядам x' и τ , позволяет увеличить степень надежности прогноза, так как совпадение знаков изменения прогнозируемых величин означает высокую степень его достоверности.

Второй метод прогнозирования (дополнительный) базируется на утверждении, что во многих случаях прогнозирование сигнала как задачу получения численных оценок будущей величины сигнала можно представить задачей классификации. В этом случае текущее состояние определяет некоторый класс события в будущем, например «увеличение» сигнала, «уменьшение» или «без изменения» на некоторый горизонт. Во многих прикладных случаях задачу прогнозирования можно решать в контексте определения изменения сигнала, превышающем некоторый порог, когда изменение становится значимым в практическом смысле (например, сейсмическая активность или процент выхода годных изделий на производстве). В этом случае сигнал можно представить как последовательность сменяющих друг друга локальных экстремумов (ЛЭ), между которыми изменение сигнала превышает некоторый значимый в практическом смысле порог. Формально можно записать, что исходный сигнал преобразуется следующим образом:

$$x(t) \equiv \{x_1, x_2, \dots, x_N\} \rightarrow x' \equiv \{x'_1, x'_2, \dots, x'_{N'}\},$$

$$x_i = x(t_i), \quad i = 1 \dots N, \quad x'_i = x(t'_i), \quad i = 1, 2, \dots, N',$$

$$|(x'_i - x'_{i-1}) / x'_{i-1}| \geq p, \quad p > 0, \quad i = 2, 3, \dots, N',$$

где N – число отсчетов в исходном сигнале; x' – преобразованный сигнал, где все значения являются ЛЭ в исходном сигнале; N' – число отсчетов в преобразованном сигнале; $p > 0$ – порог.

Таким образом, после прохождения заданного порога p изменения сигнала от момента фиксации последнего ЛЭ заданного порога ($x'_{N'}$) встает задача определения следующего ЛЭ, удовлетворяющего условию необходимого порога.

Для этого анализируется последний локальный экстремум подтверждающего порога p' , $0 < p' < p$ (рис. 1).

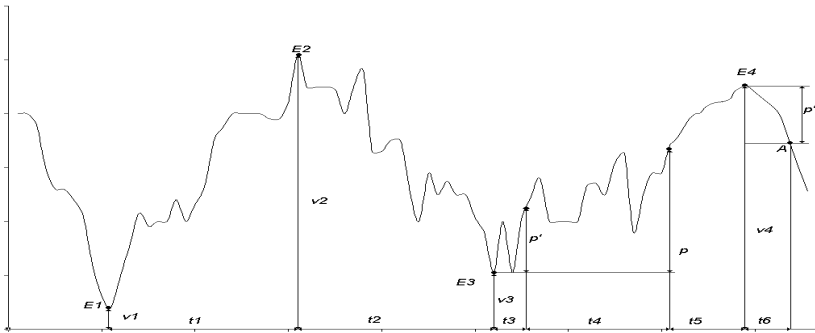


Рис.1. Сигнал с локальными экстремумами порога p и подтверждающим порогом p'

Необходимо определить, является ли анализируемый ЛЭ порога p' и ЛЭ порога p , потому что даже такое «запаздывающее» прогнозирование дает информацию о последующем изменении сигнала, а именно о том, что сигнал изменится не менее чем на заданный порог от анализируемого ЛЭ (в случае, если мы установили, что ЛЭ порога p пройден, т.е. определяем его как «истинный») или же заданный порог не будет достигнут и сигнал уйдет выше локального экстремума, если это локальный максимум, или ниже, если это локальный минимум (т.е. определяем его как «ложный»). Таким образом, необходимо отнести ЛЭ к одному из двух возможных классов – «истинному» или «ложному».

Формально задача ставится следующим образом: зная $x'_1, x'_2, \dots, x'_{N'}, x'^{p'}$, определить истинность равенства $x'^{p'} = x'_{N'+1}$, где $x'^{p'}$ – экстремум подтверждающего порога p' . На рис. 1 необходимо в точке А (в соответствующий момент времени) определить, истинным или ложным является локальный экстремум E4.

Для решения этой задачи по обучающему множеству построено многомерное пространство, в котором располагаются экстремумы. Каждый экстремум задается вектором:

$\mathbf{E}_i = [v_1, v_2, v_3, v_4, t_1, t_2, t_3, t_4, t_5, t_6]$, где i – порядковый номер экстремума. Принадлежность экстремума к классу определяется многослойным перцептроном. Методом главных компонент исходные векторы возможно преобразовывать в векторы меньшей размерности, однако, как показали эксперименты, для сложных сигналов лучше ограничиться агрегацией элементов вектора через их линейную комбинацию.

В четвертой главе рассмотрена реализация предложенных методов в виде программного комплекса, даны критерии применимости предложенного метода обнаружения разладки, метод повышения точности прогнозирования на основе анализа времени достижения порога изменения, а также результаты экспериментов на реальных сигналах.

Предложенные методы реализованы в программном комплексе «ChaosExpert». Общая последовательность прогнозирования сложного сигнала в комплексе осуществляется следующим образом: получение сигнала, получение производного от исходного сигнала ряда локальной фрактальной размерности, поиск точки разладки в исходном сигнале и в ряде локальной фрактальной размерности, получение границы между реализациями различных ДС, выделение для обучения предиктора последнего фрагмента сигнала до найденной границы (фрагмент последней сформировавшейся квазистабильной динамики сигнала), построение комбинированного прогноза, оценка его достоверности. Указанная итерация может повторяться многократно для различных масштабов представления сигнала, с тем, чтобы получить многошаговый прогноз.

Оценены критерии применимости предложенного метода обнаружения для сигналов AP(1) со значительной шумовой компонентой отдельно по изменению математического ожидания и дисперсии. Положительный эффект от применения метода достигается при изменении математического ожидания до 1,5 раз и дисперсии до 2,5 раз после момента разладки.

Для повышения точности прогнозирования предложено подавать при обучении на вход нейронной сети не только значение времени достижения заданного порога, но и соответствующее значение значимости отсчета наблюдаемого сигнала. Значимость точки $x(t_i)$ определяется как

$$I(x(t_i)) = \frac{d([x(t_{i-1}), t_{i-1}], [x(t_i), t_i]) + d([x(t_{i+1}), t_{i+1}], [x(t_i), t_i])}{N},$$

где $d(p_1, p_2)$ – манхеттеновское расстояние между точками p_1 и p_2 ; N – число отсчетов в ряде значений времени достижения заданного порога изменения. Для первой и последней точек, соответственно, левые и правые расстояния равны 0. Анализ чувствительности нейронной сети к входным переменным позволяет сделать оценку: повысит или понизит точность прогноза применение значимости значений сигнала.

Разработанные методы применялись в задаче прогнозирования числа возвратов пользователей на сайт. Использована накопленная статистика

за период от 21.05.08 до 06.06.09. Найдены точки разладки по исходному сигналу и ряду индекса фрактальности (рис. 2).

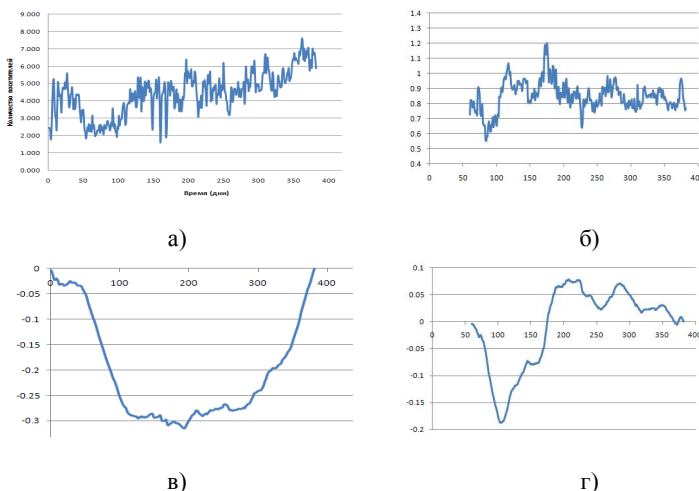


Рис.2. Возвраты пользователей на сайт (а), производный ряд индекса фрактальности (б), решающая функция АКС по исходному сигналу, абсолютный максимум на 192 отсчете (в), решающая функция АКС по ряду индекса фрактальности, абсолютный максимум на 105 отсчете (г)

Степень уверенности обнаруженной разладки в обоих случаях равна 100%. Сравним точность обнаружения разладки прогнозом на шаг вперед по 10 примерам, выполненным многослойным персептроном (один скрытый слой, 3 нейрона с сигмоидальной функцией активации), в зависимости от объема обучающей выборки (без использования границы смены динамики, с разладкой в исходном сигнале, с разладкой в индексе фрактальности сигнала). Результаты представлены в табл. 1.

Таблица 1

| E_{all} | E_i | E_a |
|-----------|-------|-------|
| 6,43 | 5,81 | 5,88 |

Использованы следующие обозначения: E_{all} – относительная ошибка прогноза при обучении на всем доступном множестве примеров; E_i – обучающее множество – выбирается до границы, найденной по ряду индекса фрактальности для прогнозируемого сигнала; E_a – обучающее

множество – выбирается до границы, найденной по исходному сигналу методом АКС. В результате сопоставления ошибок прогнозирования можно сделать вывод, что при использовании фрагмента сигнала для обучения предиктора, найденного по ряду оценки локальной фрактальной размерности, прогноз получается в среднем более точный.

Далее построен ряд значений времени достижения порога изменения в 10%. Результаты его прогнозирования представлены в табл. 2. Для обучения выбран фрагмент до момента разладки, оцененного по индексу фрактальности.

Таблица 2

| Способ / № примера (значение) | 1 (1) | 2 (3) | 3 (-5) | Еср |
|--------------------------------------|--------|--------|----------|---------------|
| Время достижения | 83,97% | 52,76% | -127,40% | 88,04% |
| Время достижения и значимость | 95,15% | 0,88% | 62,08% | 52,7% |
| APCC(2,2) | -126% | 38% | 34,6% | 66,2% |

В скобках у номера примера указано прогнозируемое значение (время, которое потребуется сигналу для изменения на 10%). Использованы следующие обозначения: время достижения – на вход НС подавалось только время достижения заданного порога изменения; время достижения и значимость – на вход НС подавалось время достижения заданного порога изменения и значимость данного отсчета; APCC(2,2) – результаты прогнозирования моделью авторегрессии (порядок 2) – скользящего среднего (порядок 2). Отрицательная ошибка свидетельствует о том, что неверно определен знак (а для ряда значений времени порога достижения это имеет ключевое значение). E_{cp} – средняя относительная ошибка прогноза (по абсолютным значениям). Прогнозирование предложенным методом дает более точный прогноз по сравнению с известным в среднем на 13%.

Прогнозирование на основе ЛЭ порога $p = 15\%$ и подтверждающего порога $p' = 5\%$ дало следующий результат: из 4-х «истинных» ЛЭ верно распознано 4, из 5 «ложных» – верно распознан 1. Такой результат можно объяснить тем, что при данном пороге изменению соответствует прогноз на 3-5 отсчетов вперед. Применение меньших порогов затруднено из-за низкой частоты дискретизации.

Рассмотрен пример прогнозирования магнитуды зафиксированных землетрясений. Использовался ряд значений по наблюдениям с 1960 по 2007 гг. Степень уверенности в разладке по исходному сигналу составила 0,5; по индексу фрактальности – 0,33. Это говорит о том, что, возможно, разладки в данном сигнале нет. Результаты прогнозирования частично подтвердили данное предположение – точность прогноза су-

щественно не изменилась при изменении объема обучающего множества. Средняя относительная ошибка по 9 прогнозам с верно определенным знаком изменения сигнала для порога в 20% по множеству с учетом разладки составила 90%, для одного неверно определенного – 111%. Без учета разладки (обучение по всему множеству) – 105% для восьми и 111% для двух, соответственно.

В **заключении** приведены основные результаты, полученные в ходе диссертационного исследования. В работе решена задача прогнозирования сложных стохастических сигналов на основе обнаружения в ретроспективе сигнала границы между реализациями различных динамических систем (разладки). Найденная граница используется для обучения предиктора на строго заданном множестве последних значений сигнала, по которому моделируется последняя квазистабильная динамика сигнала. Комбинированное прогнозирование осуществляется на основе ряда значений времени достижения заданного порога изменения.

Предложенные методы, алгоритмы и программный комплекс могут найти свое применение в областях, где ведется работа со сложными эволюционными сигналами: в медицине, биологии, контроле технологических процессов, финансово-экономической деятельности, энергетике.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИОННОЙ РАБОТЫ

1. Предложен метод обнаружения границы реализации динамических систем в сложном сигнале на основе обнаружения разладки в ряде локальной фрактальной размерности сигнала для обоснованного выбора обучающего множества для предиктора.

2. Предложен оригинальный метод прогнозирования на основе анализа времени достижения заданного порога изменения сигнала, позволяющий прогнозировать не значение сигнала, а время, через которое сигнал изменится на заданный порог.

3. Предложен способ повышения точности метода прогнозирования на основе анализа времени достижения заданного порога изменения сигнала.

4. Разработан метод прогнозирования на основе локальных экстремумов сигнала заданного порога.

5. Разработан метод оценки степени уверенности обнаружения разладки в сигнале по количеству выраженных локальных экстремумов заданного порога, позволяющий получить численные оценки наличия смены динамики в сигнале.

6. Определен наиболее эффективный по точности апостериорный непараметрический метод обнаружения разладки в сложном сигнале.

7. Разработанные методы реализованы в программном комплексе, позволяющем выделять границы реализаций динамических систем в сигнале и строить более точный прогноз.

ОСНОВНЫЕ ПОЛОЖЕНИЯ ДИССЕРТАЦИОННОЙ РАБОТЫ ИЗЛОЖЕНЫ В ПУБЛИКАЦИЯХ

*Статьи в журналах, рекомендованных ВАК для публикации
результатов диссертации на соискание ученой степени доктора
и кандидата наук*

1. Егошин, А. В. Исследование возможности выявления динамических систем по наблюдаемому временному ряду в задаче прогнозирования /А. В. Егошин, И. Г. Сидоркина // Вестник Чувашского университета. – Чувашский государственный университет им. И. Н. Ульянова. – 2008. – № 2. – С. 211-216.

Другие издания

2. Егошин, А. В. Выявление границ реализаций динамических систем в модельных стохастических сигналах / А. В. Егошин // Информационные технологии в профессиональной деятельности и научной работе: сборник материалов всероссийской научно-практической конференции с международным участием: в 2 ч. – Йошкар-Ола: Марийский государственный технический университет, 2009. – Ч.1 – С. 91-95.

3. Егошин А. В. Анализ времени достижения сигналом порога изменения в задаче нейросетевого прогнозирования временных рядов / А. В. Егошин // Информационно-вычислительные технологии и их приложения: сборник статей VI Международной научно-технической конференции. – Пенза: РИО ПГСХА, 2007. – С. 74-76.

4. Егошин А. В. Постановка задачи прогнозирования временного ряда порождаемого динамической системой / А. В. Егошин // Информационные технологии в профессиональной деятельности и научной работе: сборник материалов всероссийской научно-практической конференции с международным участием. – Йошкар-Ола: Марийский государственный технический университет, 2007. – С. 136-140.

5. Егошин А. В., Сидоркина И. Г. Рекурсивное прогнозирование эволюционных временных рядов с использованием многослойного персептрона / А. В. Егошин, И. Г. Сидоркина // Информационные техноло-

гии в профессиональной деятельности и научной работе: сборник материалов всероссийской научно-практической конференции с международным участием: в 2 ч. – Йошкар-Ола: Марийский государственный технический университет, 2008. – Ч.2 – С. 46-48.

6. Егошин А. В. Исследование возможности создания самообучаемой нейронной сети для решения задачи прогнозирования финансовых временных рядов / А. В. Егошин // Технологии Microsoft в теории и практике программирования: материалы конференции / под ред. проф. Р. Г. Стронгина. – Нижний Новгород: Изд-во Нижегородского госуниверситета, 2007. – С. 79-83.

Усл. п. л. 1,0. Тираж 100 экз. Заказ № 4238.

Редакционно-издательский центр
Марийского государственного технического университета
424006 Йошкар-Ола, ул. Панфилова, 17