На правах рукописи

Неделько Светлана Валерьевна

МЕТОДЫ ПОСТРОЕНИЯ ЛОГИКО-ВЕРОЯТНОСТНЫХ МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ

Специальность 05.13.17 — Теоретические основы информатики

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
кандидата технических наук

Новосибирск – 2009

Работа выполнена в Учреждении Российской академии наук Институте математики им. С.Л. Соболева Сибирского отделения РАН и в Государственном образовательном учреждении высшего профессионального образования «Новосибирский государственный технический университет»

Научный руководитель — доктор технических наук, профессор Лбов Геннадий Сергеевич

Официальные оппоненты: — доктор технических наук Золотухин Юрий Николаевич,

кандидат физико-математических наук, доцент Пестунов Игорь Алексеевич

Ведущая организация — Учреждение Российской академии наук Институт систем информатики имени А.П. Ершова Сибирского отделения РАН, г. Новосибирск

Защита состоится « 21 » января 2010 года в 14 часов на заседании диссертационного совета Д 212.173.06 при Государственном образовательном учреждении высшего профессионального образования «Новосибирский государственный технический университет» по адресу: 630092, г. Новосибирск, пр. К. Маркса, 20.

С диссертацией можно ознакомиться в библиотеке при Государственном образовательном учреждении высшего профессионального образования «Новосибирский государственный технический университет».

Автореферат разослан « 19 » декабря 2009 г.

Учёный секретарь
диссертационного совета Чубич В.М.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы

Необходимость построения математических моделей сложных объектов возникает в трудноформализуемых областях знания (медицина, геология, биология, социология, экономика и др). Описание сложного объекта включает большое число характеристик различной природы, что сопровождается также недостаточностью сведений о его структуре и взаимосвязях внутри него. Одним из видов представления эмпирической информации в естественнонаучных областях являются временные ряды.

Прогнозирование и анализ многомерных временных рядов являются известными задачами анализа данных, решению которых посвящено большое число работ. При этом имеется лишь относительно небольшое число методов, применимых для случая разнотипных переменных, в первую очередь, это методы, основанные на использовании класса логических решающих функций. Построение логико-вероятностных моделей является универсальным методом, работающим в условиях разнотипности и большого числа зависимых переменных, малого объёма выборки.

Вместе с тем, к настоящему времени остаются неизученные вопросы в решении указанных задач. Это связано, в частности, с необходимостью предсказания значений нескольких целевых переменных с учетом их взаимосвязей. Кроме того, для прикладных задач бывает оправдано построение многоальтернативных решений, например, прогнозирование целевой переменной в виде области в пространстве её значений. Также частыми особенностями прикладных задач является значительный объём эмпирических данных, требующий эффективных процедур обработки, или, наоборот, относительно небольшой объём эмпирических данных при достаточно большом числе прогнозируемых признаков.

В настоящее время не существует методов решения задач анализа и прогнозирования многомерных разнотипных временных рядов, учитывающих все указанные особенности. Поэтому разработка таких методов представляет собой актуальное направление исследований.

Цель работы

Разработка и исследование методов построения логико-вероятностных моделей многомерных разнотипных временных рядов для решения задач анализа и прогнозирования, учитывающих многомерность, разнотипность пространства целевых переменных, их зави-

симость, возможность применения многовариантного решающего правила.

Задачи исследования

Постановка задачи построения логико-вероятностных моделей многомерных разнотипных временных рядов.

Выбор и обоснование критерия качества логико-вероятностной модели временного ряда.

Разработка методов построения логико-вероятностных моделей многомерных разнотипных временных рядов

Разработка и исследование метода адаптивного поиска дерева решений.

Исследование статистической достоверности решений, получаемых разработанными методами анализа многомерных разнотипных временных рядов.

Применение разработанных методов к решению прикладных и модельных задач.

Методы исследования

Методы исследования включают в себя аппарат теории вероятностей и математической статистики, теоретической кибернетики; математическое моделирование с применением средств вычислительной техники.

Научная новизна

В диссертационной работе получены следующие научные результаты.

1. Предложен алгоритм построения логико-вероятностной модели многомерного разнотипного временного ряда на основе многовариантного решающего правила.

2. Разработан метод решения задачи прогнозирования многомерного разнотипного временного ряда путём построения информативного пространства состояний. Предложены и исследованы несколько критериев качества вероятностной модели. Предложен способ обнаружения изменения вероятностных свойств случайного процесса.

3. Исследовано качество решений, получаемых указанным алгоритмом построения логико-вероятностной модели на основе информативного пространства состояний в зависимости от длины обучающей реализации временного ряда, сложности модели (числа конечных вершин дерева), глубины предыстории. Предложен способ оценивания статистической достоверности полученных закономерностей.

4. Исследована применимость метода случайного поиска с адаптацией для построения дерева решений. Для дискретного пространства малой мощности проведено исследование связи между алгоритмом СПА поиска глобального экстремума и классом функций, решаемых данным алгоритмом.

Научные результаты, выносимые на защиту

Критерий качества логико-вероятностной модели разнотипного временного ряда.

Алгоритм построения логико-вероятностной модели многомерного разнотипного временного ряда на основе многовариантного решающего правила.

Метод решения задачи прогнозирования временного ряда путём построения информативного пространства состояний.

Способ обнаружения изменения вероятностных свойств случайного процесса с помощью разработанного алгоритма прогнозирования временного ряда.

Исследование качества прогнозирования в зависимости от длины реализации, сложности логико-вероятностной модели ряда и длины предыстории.

Практическая ценность и реализация результатов работы

Разработанные алгоритмы анализа и прогнозирования многомерных разнотипных временных рядов позволяют выявлять в массивах статистических данных закономерности более общего вида, чем предполагается в методах классификации (распознавания образов) и регрессионного анализа, при этом объём эмпирических данных может быть относительно мал. Кроме того, алгоритм построения информативного пространства состояний адаптирован к выявлению моментов изменения вероятностных свойств процесса.

Исследование статистической устойчивости логико-вероятностных моделей позволило сформулировать метод выбора эмпирически обоснованной сложности модели для заданного временного ряда. Это важно при решении прикладных задач, поскольку истинная модель процесса в них неизвестна.

Разработано программное обеспечение, реализующее предложенные методы.

Разработанные методы применены для решения задач анализа метеорологических и сейсмических данных, прогнозирования состояния ионосферы, выявления закономерностей в музыкальных произведениях, а также сравнения текстур.

Достоверность результатов

Достоверность результатов обеспечивается корректным применением математических методов и подтверждается проведенными исследованиями на модельных и прикладных задачах. Выводы, получаемые на основе статистического моделирования, обосновываются построением статистических оценок доверительных интервалов.

Апробация работы

Основные положения работы докладывались и обсуждались на следующих конференциях и семинарах.

Международные конференции «Интеллектуализация обработки информации», г. Алушта: июнь 2002 г. (ИОИ–2002), июнь 2006 г. (ИОИ–2006), июнь 2008 г. (ИОИ–2008).

Международные конференции «Knowledge-Dialogue-Solution», г. Варна, Болгария: июнь 2006 г. (KDS-2006), июнь 2007 г. (KDS-2007).

Международная конференция «Classification, Forecasting, Data Mining», г. Варна, Болгария: июнь 2009 г. (CFDM–2009).

Семинары Института математики СО РАН.

Отдельные части работы прошли экспертизу в ходе выполнения проектов, поддержанных грантами РФФИ: № 04-01-00858-а, № 07-01-00331-а.

Личный вклад

Все представленные в работе научные результаты получены соискателем лично, за исключением результатов четвертой главы, которые получены в соавторстве. В четвертой главе автором лично проведено статистическое моделирование при исследовании применимости метода случайного поиска с адаптацией для построения дерева решений и исследовании связи между алгоритмом поиска глобального экстремума и классом функций, решаемых данным алгоритмом.

Публикации

По теме диссертации опубликовано четырнадцать научных работ, в том числе три в изданиях из перечня журналов, рекомендуемых ВАК РФ, шесть в других рецензируемых журналах, пять в сборниках трудов конференций.

Структура и объем работы

Диссертационная работа состоит из введения, пяти глав, заключения, библиографического списка из 89 наименований отечественной и зарубежной литературы и трех приложений, изложенных на 149 страницах машинописного текста. Иллюстративный материал представлен 33 рисунками и 9 таблицами.

СОДЕРЖАНИЕ РАБОТЫ

Первая глава посвящена задаче построения логико-вероятностных моделей многомерного разнотипного временного ряда.

Первый параграф содержит введение в задачи прогнозирования временных рядов и краткий обзор существующих методов их решения.

Во втором параграфе рассматривается задача построения логико-вероятностной модели временного ряда, а также вопросы раздельного и совместного прогнозирования переменных ряда.

Постановка задачи прогнозирования. Имеется n-мерный временной ряд $v = \left\{ z^t \,\middle|\, t = \overline{1, N} \right\}$, $z^t = \left(z_1^t, ..., z_n^t \right)$, $z_j^t \in Z_j$. Здесь Z_j – множество допустимых значений j-й переменной ряда. $Z = \prod_{j=1}^{n} Z_j$.

Пусть ряд является реализацией случайного n-мерного процесса $z(t)$ с дискретным временем, который задается переходной (условной) вероятностной мерой $P : \Lambda \times Z^d \to [0,1]$, где Λ – σ-алгебра подмножеств из Z, а d – длина предыстории, которая определяет распределение в заданный момент.

Требуется на основе имеющихся данных v построить прогноз временного ряда в моменты времени $t > N$ в соответствии с некоторым критерием (зависит от варианта постановки задачи).

Введём обозначения $X \equiv Z^d$, $Y \equiv Z$, причём X будем использовать для обозначения пространства предысторий, а Y — для пространства значений в момент времени, для которого делается прогноз. Таким образом, Y – пространство целевых, а X – прогнозирующих переменных.

Тогда переходную меру можно записать как

$$P\left[Z / z(t-1),\, z(t-2),\, ...,\, z(t-d) \right] \equiv P\left[Y / x \right].$$

В данных обозначениях вероятность события $E_Y \in \Lambda$, $E_Y \subseteq Z$, записывается как

$$P\left(z(t) \in E_Y / z(t-1),\, z(t-2),\, ...,\, z(t-d) \right) \equiv\ P\left(E_Y / x \right) \equiv P\left[Y / x \right]\left(E_Y \right).$$

Заметим, что круглые скобки используются для указания аргумента функции, а квадратные — как часть обозначения меры.

Решающая функция обычно определяется как отображение $f : X \to Y$. Такую функцию будем называть одновариантным решающим правилом.

Многовариантное решающее правило есть $f : X \to S$, где решение $s \in S$ представляет собой множество пар $s = \left\{ \left(E_Y^k, \tilde{p}_k \right) \middle| \; k = \overline{1, M} \right\}$, $\tilde{p}_k$ – оценка условной вероятности $P\left(E_Y^k / x \right)$, $E_Y^k \subseteq Y$. В отличие от одновариантного прогноза, в данном случае имеется несколько вариантов прогноза с оценками их вероятностей.

Логико-вероятностную модель определим следующим образом:

$$f_L = \left\{ \left(E_X^l, s_l \right) \middle| \; l = \overline{1, L} \right\},$$

где $E_X^l \subseteq X$ образуют разбиение $\alpha_L = \left\{ E_X^l \;\middle|\; l = \overline{1, L} \right\}$ пространства X. Таким образом, логико-вероятностная модель представляет собой кусочно-постоянное многовариантное решающее правило.

При определении точности прогноза будем различать определенность и достоверность. Под достоверностью прогноза будем понимать некоторую меру соответствия решения распределению (например, близость оценок вероятностей). Конкретные выражения для достоверности будут приведены в дальнейшем.

Во многих случаях, используя многовариантный прогноз, можно повысить достоверность решения, за счет некоторого снижения определенности прогноза. Понятие определенности формализуем введением критерия информативности. Качество модели будем характеризовать ее информативностью и достоверностью.

В третьем параграфе вводится критерий качества прогнозирования, основанный на понятии информативности распределений.

Качество прогноза определяется степенью зависимости целевых переменных от прогнозирующих. Для численного выражения этой меры зависимости введём понятие информативности условного распределения. Под информативностью распределения понимается степень его отличия от априорного либо от равномерного распределения, которая может характеризоваться, например, расстоянием в некоторой выбранной метрике. Чем больше такое отличие, тем большей информативностью обладает распределение.

Пусть заданы пространства X и Y, вероятностные меры $P[X]$, $P[Y]$, совместное распределение $P[X, Y]$ и условная мера $P[Y/x]$, $x \in X$.

Критерий информативности определим как

$$K(P[X,Y]) = \int \rho(P[Y/x], P[Y])\, dP[X],$$

где в роли ρ выступает некоторая мера отличия распределений (имеет смысл расстояния, но выполнение всех свойств метрики не обязательно).

Использовались следующие меры различия для распределений:

$\rho_C(P_1, P_2) = \int \ln \frac{dP_1}{dP_2}\, dP_1$ — дивергенция Кульбака-Лейблера;

$\rho_E(P_1, P_2) = \left| H(P_2) - H(P_1) \right|$ - разность энтропий;

$\rho_U(P_1, P_2) = \int \left| dP_2 - dP_1 \right| = \int \left| \frac{dP_2}{d\mu} - \frac{dP_1}{d\mu} \right| d\mu$ - разность плотностей;

$\rho_M(P_1, P_2) = \max\limits_{A \in B} \left| P_1(A) - P_2(A) \right|$.

Здесь $H(P) = -\rho_C(P, \mu) = -\int \ln \frac{dP}{d\mu}\, dP$, а $B \subseteq \Lambda$ – заданный класс подмножеств из σ-алгебры Λ.

Для оценивания информативности модели введём переменную $\tilde{X}$, значения которой соответствуют областям разбиения $\alpha_L = \left\{ E_X^l \;\middle|\; l = \overline{1, L} \right\}$ для логико-вероятностной модели f_L.

Для каждого $\tilde{x} \in \tilde{X}$ введём переменную $\tilde{Y}_{\tilde{x}}$, значения которой соответствуют областям $E_Y^k \subseteq Y$ многовариантного решения $s(\tilde{x}) = \left\{ \left(E_Y^k, p_k \right) \;\middle|\; k = \overline{1, M} \right\}$, где $p_k = P\left(E_Y^k / \tilde{x} \right)$.

Критерий информативности модели:

$$K(f_L) = \int \rho\left(P[\tilde{Y}_{\tilde{x}} / \tilde{x}], P[\tilde{Y}_{\tilde{x}}] \right) dP[\tilde{X}].$$

Выборочное значение критерия получается подстановкой выборочных оценок вероятностей.

В четвертом параграфе описываются класс логических решающих функций и алгоритм LRP. Предлагается алгоритм построения логико-вероятностной модели временного ряда на основе модифицированного алгоритма поиска логических закономерностей. Приводятся иллюстрации работы алгоритма.

Во ***второй главе*** представлен метод построения логико–вероятностных моделей многомерного разнотипного временного ряда путём адаптивного выбора пространства состояний случайного процесса, описывающего временной ряд.

В первом параграфе изложены идеи метода построения информативного пространства состояний временного ряда, предложены различные критерии качества эмпирической модели. Обсуждаются критерии близости логико-вероятностной модели к истинному случайному процессу.

Метод построения эмпирической модели состоит в следующем.

Пусть $\lambda = \left\{ E^{\omega} \subseteq Z \;\middle|\; \omega = \overline{1,k} \right\}$, $\bigcup_{\omega=1}^{k} E^{\omega} = Z$, $\omega \neq \varpi \Rightarrow E^{\omega} \cap E^{\varpi} = \varnothing$ - некоторое разбиение пространства *Z*. Тогда исходному многомерному ряду v можно сопоставить одномерную символьную последовательность $w = \left\{ \omega^t \;\middle|\; z^t \in E^{\omega^t},\ t = \overline{1,N} \right\}$.

Случайному процессу *z*(*t*) будет соответствовать процесс $\omega(t)$, переходные вероятности для которого обозначим

$$p_{\omega_0|\omega_1,\omega_2,\ldots,\omega_d} = P\left(\omega(t) = \omega_0 / \omega(t-1) = \omega_1,\ \ldots,\ \omega(t-d) = \omega_d\right).$$

Совместная вероятность есть

$$p_{\omega_0 \ldots \omega_d} = \mathrm{P}\left(\bigwedge_{\tau=0}^{d} \left(\omega(t-\tau) = \omega_\tau\right) \right) = \mathrm{P}\left(\bigwedge_{\tau=0}^{d} \left(Z(t-\tau) \in E^{\omega_\tau}\right) \right).$$

Критерий информативности $K_U = \int \Delta_U dP[X,Y]$ для разбиений принимает вид

$$K_U(\lambda) = \sum_{\omega_0=1}^{k} \ldots \sum_{\omega_d=1}^{k} \left| \; p_{\omega_0} - p_{\omega_0|\omega_1,\omega_2,\ldots,\omega_d} \; \right| \cdot p_{\omega_1\omega_2\ldots\omega_d}$$

или в эквивалентной форме

$$K_U(\lambda) = \sum_{\omega_0=1}^{k} \ldots \sum_{\omega_d=1}^{k} \left| \; p_{\omega_0\omega_1\ldots\omega_d} - p_{\omega_0}\, p_{\omega_1\ldots\omega_d} \right|,$$

где $p_{\omega_1\ldots\omega_d} = \sum_{\omega_0=1}^{k} p_{\omega_0\omega_1\ldots\omega_d}$ - вероятность появления предыстории длины d, $p_{\omega_0} = \sum_{\omega_1=1}^{k} \ldots \sum_{\omega_d=1}^{k} p_{\omega_0\omega_1\ldots\omega_d}$.

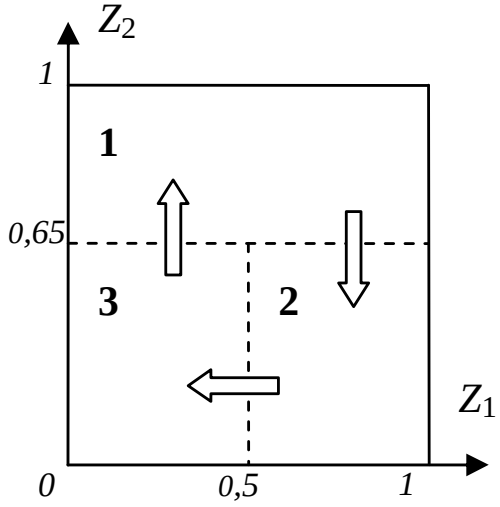
При *d* = 1 получаем

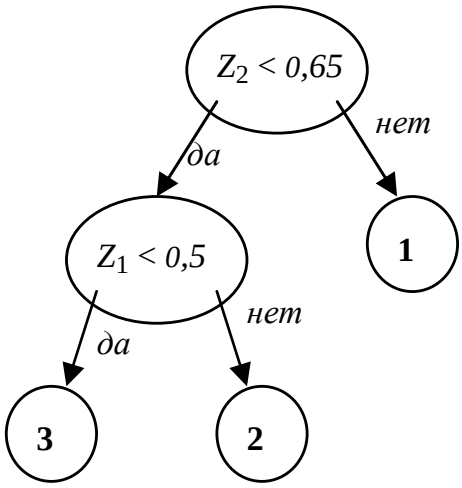
$$K_U(\lambda) = \sum_{\omega_0=1}^{k} \sum_{\omega_1=1}^{k} \left| p_{\omega_0\omega_1} - p_{\omega_0} p_{\omega_1} \right|.$$

Другие критерии качества эмпирической модели также легко вычисляются для процесса $\omega(t)$, в частности, критерий на основе дивергенции примет вид

$$K_C(\lambda) = \sum_{\omega_0=1}^{k} \ldots \sum_{\omega_d=1}^{k} p_{\omega_0\omega_1\ldots\omega_d} \ln \frac{p_{\omega_0\omega_1\ldots\omega_d}}{p_{\omega_0} p_{\omega_1\ldots\omega_d}}.$$

При построении модели по выборочной реализации входящие в выражения вероятности оцениваются соответствующими частотами. Исследована связь между критерием, основанным на дивергенции, и критерием максимального правдоподобия.





Переходные вероятности

	1	**2**	**3**
1	0	1	0
2	0	0	1
3	1	0	0

Рис. 1 . Для тестового примера изображены соответствующие состояниям области в пространстве значений переменных, переходные вероятности и дерево решений, построенное предложенным алгоритмом

Во втором параграфе рассматриваются варианты алгоритма прогнозирования многомерного разнотипного временного ряда с использованием различных классов разбиений пространства состояний:

разбиения на интервалы и в классе линейных функций. Эффективность работы метода продемонстрирована решением модельной задачи.

На *рис. 1* для модельной задачи приведены области, соответствующие состояниям случайного процесса, дерево решений, задающее разбиение на указанные области, и матрица переходных вероятностей $p_{\varpi|\omega} = p_{\omega\varpi} / p_{\omega}$. Значения в таблице отражают вероятность перехода из состояния, задаваемого номером строки, в состояние с номером столбца. В соответствии с заданной моделью генерировались реализации случайного процесса, по которым предложенный алгоритм практически точно (границы областей оценивались с среднем с погрешностью $\frac{1}{N}$) восстанавливал заложенные закономерности.

Приведём пример реализации данного процесса в виде последовательности пар значений переменных Z_1 и Z_2. Имеем двумерный ряд: ((0,5;0,71), (0,73;0,45), (0,08;0,28), (0,04;0,85), (0,69;0), (0,49;0,55), (0,2;0,88), (0,83;0,39), (0,5;0,47), (0,87;0,81), (0,61;0,51), (0,39;0,23), (0,8;0,75), (0,7;0,32), (0,19;0,34), (0,74;0,95), (0,62;0,16), (0,33;0,45), (0,38;0,89), (0,85;0,25), (0,19;0,05), (0,65;0,68), (0,78;0,45), (0,46;0,04), …). Несмотря на простоту модели, при визуальном анализе приведённой реализации процесса закономерности неочевидны.

Алгоритм позволяет восстанавливать закономерности также при наличии «шума», когда нулевые переходные вероятности заменялись на δ, а единичные на 1–2δ, при $\delta < 0{,}2$, $N \approx 100$.

Третий параграф посвящён задаче обнаружения изменения вероятностных свойств случайного процесса, описывающего временной ряд. Предложен метод решения указанной задачи с помощью представленного алгоритма прогнозирования. Проведено сравнение способности различных критериев качества эмпирической модели к выявлению моментов изменения модели ряда, что проиллюстрировано тестовым примером.

Содержание ***третьей главы*** составляет исследование статистической достоверности полученных решений.

В первом параграфе рассматривается гипотеза о том, что полученные алгоритмом закономерности в виде матрицы переходных вероятностей могут быть получены для случайной последовательности событий. Предлагается проверять указанную гипотезу методом перестановок по времени значений исследуемого ряда.

Во втором параграфе проведено статистическое моделирование в задаче классификации. Исследовано поведение алгоритма LRP в зависимости от объёма выборки, числа конечных вершин дерева.

Исследование качества прогнозирования на основе адаптивного выбора пространства состояний временного ряда приводится в третьем параграфе. Поведение предложенного алгоритма исследуется в зависимости от длины обучающей реализации ряда, сложности модели ряда и длины предыстории. Полученные результаты сравниваются с аналогичными для задачи классификации. Также излагается практический критерий выбора сложности модели временного ряда.

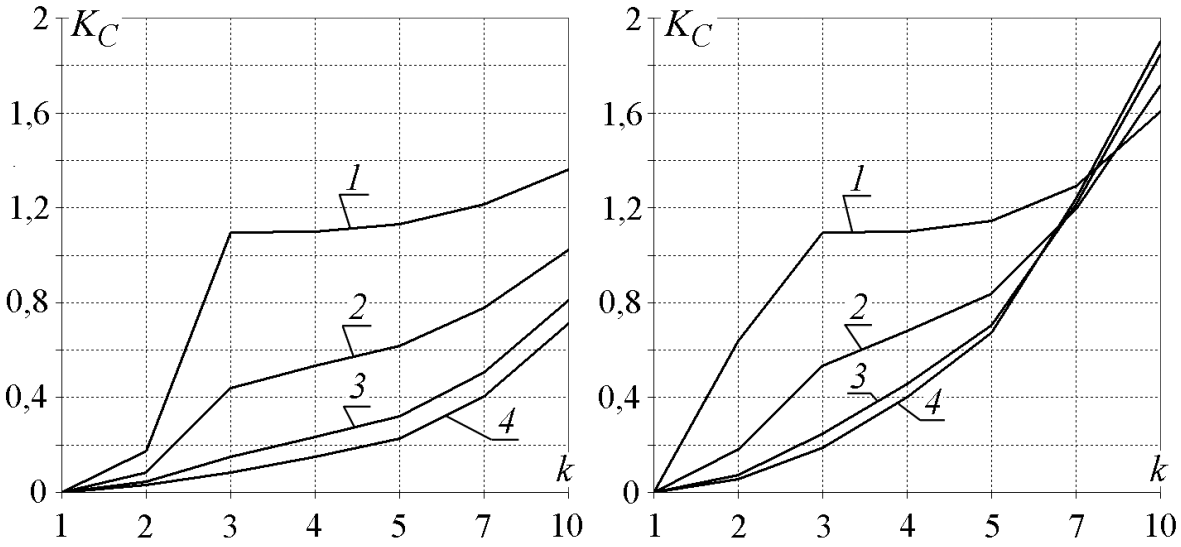


Рис. 2. Зависимость среднего критерия информативности эмпирической аппроксимации случайного процесса от числа областей разбиения k при $N = 100$ и различных δ: 1) δ = 0; 2) δ = 0,1; 3) δ = 0,2; 4) δ = 0,33. Для левой диаграммы $d = 1$, для правой $d = 2$

Для измерения достоверности построенной модели, т.е. близости построенной модели к истинной вероятностной модели процесса, можно использовать критерии, основанные на расстояниях между распределениями:

$$R'\left(P[X,Y], P'[X,Y]\right) = \int \rho\left(P'[Y/x], P[Y/x]\right) dP[X]$$

или

$$R''\left(P[X,Y], P'[X,Y]\right) = \rho\left(P'[X,Y], P[X,Y]\right),$$

где $P'[X,Y]$ есть выборочная оценка вероятностной меры. При этом наиболее подходящим расстоянием представляется ρ_U.

В работе методом статистического моделирования исследована зависимость критерия $\rho_U(P'[X,Y],P[X,Y])$ от длины реализации и сложности модели k.

В реальной ситуации мы не знаем истинной модели процесса, поэтому выбор оптимальной сложности решения приходится делать на основе получаемых значений критерия на обучающей последовательности. Зависимость среднего выборочного значения критерия информативности (использовалась дивергенция) от числа областей приведена на *рис. 2*. Можно заметить, что значение критерия в точке $k = 3$, соответствующей истинной сложности модели процесса, достаточно хорошо выделяется на графиках.

Практическим критерием выбора сложности может служить разность между получаемым значением информативности и значением информативности, полученным на такой же длины реализации «равномерного процесса» (этому соответствует кривая при $\delta = 0{,}33$).

В четвёртом параграфе проведено сравнение получаемых значений смещения эмпирического критерия качества алгоритма прогнозирования временного ряда при использовании линейных разделяющих функций со смещением эмпирического риска для задачи классификации двух образов в дискретном пространстве переменных.

В ***четвёртой главе*** рассматривается задача адаптивного поиска решающей функции.

В первом параграфе приводится обзор методов поиска логической решающей функции. Нахождение оптимальной логической решающей функции представляет собой задачу многоэкстремальной оптимизации, для решения которой существует метод СПА.

Задача поиска глобального экстремума и алгоритм СПА для её решения рассматриваются во втором параграфе.

В третьем параграфе излагается алгоритм адаптивного поиска дерева решений, использующий такую меру близости деревьев решений, как энтропийная метрика на разбиениях. Также исследова-

Таблица 1

Значения переходных вероятностей для состояний процесса, оцененные для событий с магнитудой не менее 5

p_i	i	**1**	**2**	**3**	**4**
0,39	**1**	0,48	0,32	0,1	0,1
0,33	**2**	0,37	0,41	0,12	0,1
0,12	**3**	0,34	0,28	0,23	0,15
0,16	**4**	0,24	0,24	0,1	0,42

но поведение указанного алгоритма в задаче классификации для модельных примеров.

Четвёртый параграф посвящён исследованию связи между алгоритмом поиска глобального экстремума и классом функций, который решается этим алгоритмом. Находится класс функций, решаемых алгоритмом СПА.

Пятая глава посвящена проблемам практического применения методов, изложенных во второй главе.

Первый параграф содержит краткое описание структуры разработанного программного обеспечения.

Во втором параграфе приводится обзор задач из коллекции задач машинного обучения UCI, которые можно решать предложенными методами.

В третьем параграфе решается задача выявления общих закономерностей в хоралах И.С. Баха.

Четвёртый параграф посвящён задаче прогнозирования состояния ионосферы.

Анализ метеорологических данных представлен в пятом параграфе.

Задача поиска закономерностей в сейсмических данных решена в шестом параграфе.

Для поиска закономерностей использованы данные по землетрясениям из общедоступной базы ftp://www.ncedc.org/pub/catalogs/cnss/. Были выбраны сейсмические события за период с 1970 по 2005 гг.

Первый исследованный ряд был составлен из событий с магнитудой не менее 5, произо-

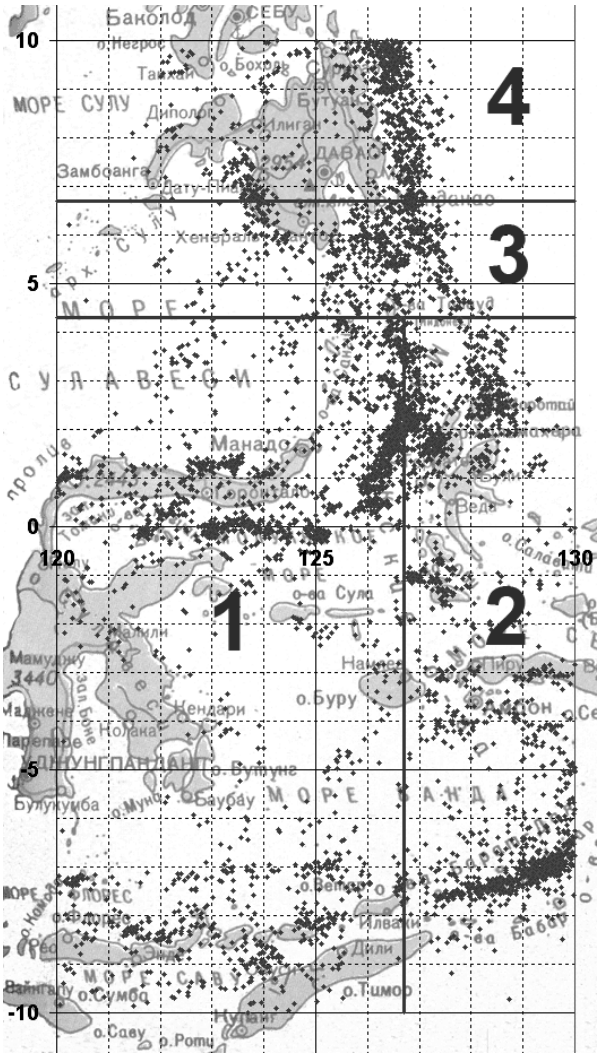


Рис. 3. Пространство состояний процесса для событий с магнитудой не менее 5

шедших между -10° и 10° географической широты и между 120° и 130° долготы (часть территорий Индонезии и Филиппин). Широта соответствует переменной Z_1, долгота — Z_2. Отрицательные значения соответствуют южной широте и западной долготе.

На указанном массиве данных было построено описанным выше алгоритмом дерево решений с числом конечных вершин 4. Глубина предыстории взята 1. Значение критерия K_U на данном дереве решений составило 0,072.

Полученное значение критерия невелико, однако природа сейсмических процессов такова, что сильные закономерности в такого рода данных отсутствуют. При этом найденные закономерности имеют высокую достоверность, поскольку на случайных перестановках вероятность получить значение критерия, превосходящее 0,072, имеет порядок 10^{-6}.

На *рис. 3* изображены выделенные состояния процесса, в *табл. 1* приведены оценки переходных вероятностей.

Седьмой параграф посвящён решению задачи сравнения текстур с использованием логико-вероятностных моделей временных рядов.

Приложения включают описание разработанных программных средств, а также справку об использовании результатов работы.

В ***заключении*** приводятся основные результаты работы, задаются направления продолжения проведенных исследований.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Основные результаты диссертационной работы состоят в следующем.

1. Дана математическая постановка задачи построения логико-вероятностных моделей многомерных разнотипных временных рядов как задачи информативной аппроксимации случайного процесса.
2. Предложен и обоснован критерий качества логико-вероятностной модели временного ряда, основанный на понятии информативности модели, понимаемой как отличие условной вероятностной меры от априорной.
3. Разработан алгоритм построения логико-вероятностной модели многомерного разнотипного временного ряда на основе многовариант-

ного решающего правила. Работоспособность методов проиллюстрирована на модельных задачах.

4. Предложен метод решения задачи анализа и прогнозирования временного ряда путём построения информативного пространства состояний. Метод, в частности, даёт возможность обнаружения изменения вероятностных свойств случайного процесса.

5. Предложен метод исследования статистической достоверности решений, получаемых разработанными методами анализа многомерных разнотипных временных рядов. Исследована зависимость качества прогнозирования от длины реализации, сложности логико-вероятностной модели ряда и длины предыстории. Оценена ёмкость алгоритма прогнозирования временного ряда при использовании линейных разделяющих функций путём сравнения полученных значений смещения эмпирического критерия качества с результатами для задачи классификации двух образов в дискретном пространстве переменных.

6. Разработанный метод прогнозирования временного ряда путём построения информативного пространства состояний применён для решения нескольких прикладных задач: задачи анализа трехмерного временного ряда, представленного метеорологическими данными (температура воздуха, объём осадков, объём водостока), задачи поиска закономерностей в сейсмических данных, а также двух задач из репозитория UCI. Оценена статистическая достоверность полученных закономерностей, проведено исследование поведения алгоритма на случайных перестановках значений временного ряда.

7. Исследована применимость метода случайного поиска с адаптацией для построения дерева решений. Для дискретного пространства малой мощности проведено исследование связи между алгоритмом поиска глобального экстремума и классом функций, решаемых данным алгоритмом.

Список публикаций автора по теме диссертации

1. Миренкова С. В. Метод прогнозирования многомерного разнотипного временного ряда в классе логических решающих функций / С. В. Миренкова // Искусственный интеллект. – 2002. – № 2. – С. 197–201.
2. Неделько С. В. Критерий информативности матрицы переходов и прогнозирование разнотипного временного ряда / С. В. Неделько // Искусственный интеллект. – 2004. – № 2. – С. 145–149.

3. Nedel’ko S. V. Extreme Situations Prediction by Multidimensional Heterogeneous Time Series Using Logical Decision Functions / S.V. Nedel’ko // Proc. of XI-th Intern. Conf. «Knowledge-Dialogue-Solution», Varna, Bulgaria, 20-30 June 2005. – Vol. 1. – P. 84–87. [Прогнозирование экстремальных ситуаций на основе многомерных разнотипных временных рядов с использованием логических решающих функций].
4. Nedel'ko S. V. Extreme situations prediction by multidimensional heterogeneous time series using logical decision functions / S. V. Nedel’ko // Information Theories & Applications. – 2006. – Vol. 13, № 3. – P. 290-294. [Прогнозирование экстремальных ситуаций на основе многомерных разнотипных временных рядов с использованием логических решающих функций].
5. Неделько С. В. Исследование связи характеристик алгоритма поиска глобального экстремума и класса функций в дискретном пространстве малой мощности / С. В. Неделько, В. М. Неделько // Искусственный интеллект. – 2006. – № 2. – С. 201–205.
6. Nedel'ko S. V. On relationship between a search algorithm and a class of functions on discrete space / V. M. Nedel'ko, S. V. Nedel'ko // Proc. of XII-th Intern. Conf. «Knowledge-Dialogue-Solution», Varna, Bulgaria, 20-25 June 2006. – Vol. 1. – P. 287–291. [О связи между алгоритмом оптимизации и классом функций в дискретном пространстве].
7. Nedel'ko S. V. Finding the relationship between a search algorithm and a class of functions on discrete space by exhaustive search / V. M. Nedel'ko, S. V. Nedel'ko // Information Theories & Applications. – 2007. – Vol. 14. – P. 339–343. [Исследование связи между алгоритмом оптимизации и классом функций в дискретном пространстве с использованием полного перебора].
8. Неделько С. В. Построение логико-вероятностных моделей временного ряда при анализе сейсмических данных / С. В. Неделько, Т. А. Ступина // Научный вестник НГТУ. – 2007. – № 4 (29). – С. 33–42.
9. Неделько С. В. Прогнозирование разнотипного временного ряда методом адаптивного формирования пространства состояний в классе логических решающих функций / С. В. Неделько // Proc. of XIII-th Intern. Conf. «Knowledge-Dialogue-Solution», Varna, Bulgaria, 18-24 June 2007. – Vol. 1. – P. 118–122.

10. Неделько С. В. Адаптивное прогнозирование многомерного временного ряда / С. В. Неделько // Таврический вестник информатики и математики. – 2008. – № 2. – С. 104–110.
11. Неделько С. В. Исследование статистической устойчивости логико-вероятностных моделей временного ряда / С. В. Неделько // Научный вестник НГТУ. – 2008. – № 4 (33). – С. 43-52.
12. Неделько С. В. Построение логико-вероятностных моделей временных рядов с использованием цепей Маркова / С. В. Неделько // Classification, Forecasting, Data Mining: Int. book series «Information Science and Computing», № 8. – Supplement to the Int. J. «Information Technologies and Knowledge». – ITA, FOI ITHEA, Sofia, 2009. – Vol. 3. – P. 83–90.
13. Неделько С. В. Адаптивный поиск глобального экстремума на множестве деревьев решений / Г. С. Лбов, В. М. Неделько, С. В. Неделько // Труды Конгресса по интеллектуальным системам и информационным технологиям «AIS-IT'09». – М.: Физматлит, 2009. – Т. 1. – С. 182–190.
14. Неделько С. В. Метод адаптивного поиска логической решающей функции / Г. С. Лбов, В. М. Неделько, С. В. Неделько // Сибирский журнал индустриальной математики. – 2009. – Т. XII, № 3 (39). – С. 66–74.

Отпечатано в типографии Новосибирского
государственного технического университета
630092, г. Новосибирск, пр. К. Маркса, 20,
тел./факс (383) 346-08-57
формат 60 х 84 1/16 объём 1,25 п.л. тираж 110 экз.
заказ № 124 подписано в печать 17.12.09 г .