

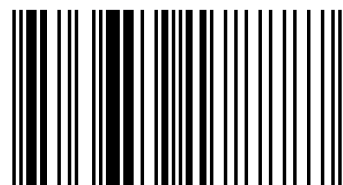
В последние годы в районах отработки полезных ископаемых участились случаи проявления горных ударов, внезапных выбросов, горнотектонических землетрясений и других геодинамических явлений. Эти процессы порождаются в массивах горных пород не только действием больших тектонических напряжений, но и увеличением глубины отработки месторождений, структурной неоднородностью массива и свойствами слагающих его горных пород. Крупные технологические взрывы, применяемые для обрушения рудных блоков на железорудных месторождениях, также сопровождаются мощными динамическими явлениями. Все это наносит значительный экономический ущерб промышленным предприятиям и гражданским объектам, а также ведет к травматизму и человеческим жертвам. Поэтому особую остроту приобрела проблема мониторинга изменения напряженно-деформированного состояния массивов горных пород. Для исключения последствий геодинамических явлений необходим надежный прогноз геологических условий ведения горных работ.

Алгоритмы обнаружения предвестников ГДС



Александр Мертвецов

к.т.н., ООО «ПС Яндекс.Деньги». Соавторы: Букреев Виктор Григорьевич, д.т.н., профессор ТПУ, научное направление: методы синтеза адаптивных систем управления сложными объектами и процессами; Колесникова Светлана Ивановна, д.т.н., профессор ТПУ, научное направление: методы исследования плохо формализуемых объектов.



978-3-659-17888-7

Мертвецов, Букреев, Колесникова

Александр Мертвецов
Виктор Букреев
Светлана Колесникова

Алгоритмы обнаружения предвестников геодинамического события

по временным рядам измерений
электромагнитных сигналов



**Александр Мертвецов
Виктор Букреев
Светлана Колесникова**

**Алгоритмы обнаружения предвестников
геодинамического события**

**Александр Мертвецов
Виктор Букреев
Светлана Колесникова**

**Алгоритмы обнаружения
предвестников
геодинамического события
по временным рядам измерений
электромагнитных сигналов**

LAP LAMBERT Academic Publishing

Impressum / Выходные данные

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Alle in diesem Buch genannten Marken und Produktnamen unterliegen warenzeichen-, marken- oder patentrechtlichem Schutz bzw. sind Warenzeichen oder eingetragene Warenzeichen der jeweiligen Inhaber. Die Wiedergabe von Marken, Produktnamen, Gebrauchsnamen, Handelsnamen, Warenbezeichnungen u.s.w. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutzgesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Библиографическая информация, изданная Немецкой Национальной Библиотекой. Немецкая Национальная Библиотека включает данную публикацию в Немецкий Книжный Каталог; с подробными библиографическими данными можно ознакомиться в Интернете по адресу <http://dnb.d-nb.de>.

Любые названия марок и брендов, упомянутые в этой книге, принадлежат торговой марке, бренду или запатентованы и являются брендами соответствующих правообладателей. Использование названий брендов, названий товаров, торговых марок, описаний товаров, общих имён, и т.д. даже без точного упоминания в этой работе не является основанием того, что данные названия можно считать незарегистрированными под каким-либо брендом и не защищены законом о брендах и их можно использовать всем без ограничений.

Coverbild / Изображение на обложке предоставлено: www.ingimage.com

Verlag / Издатель:

LAP LAMBERT Academic Publishing

ist ein Imprint der / является торговой маркой

AV Akademikerverlag GmbH & Co. KG

Heinrich-Böcking-Str. 6-8, 66121 Saarbrücken, Deutschland / Германия

Email / электронная почта: info@lap-publishing.com

Herstellung: siehe letzte Seite /

Напечатано: см. последнюю страницу

ISBN: 978-3-659-17888-7

Copyright / АВТОРСКОЕ ПРАВО © 2012 AV Akademikerverlag GmbH & Co. KG

Alle Rechte vorbehalten. / Все права защищены. Saarbrücken 2012

Рецензенты:
доктор технических наук, профессор ТГУ
Климентий Исаакович Лившиц
доктор технических наук, профессор ТУСУР
Вячеслав Михайлович Дмитриев

Обозначения и сокращения	6
ВВЕДЕНИЕ	9
Глава 1. МЕТОДЫ ОБНАРУЖЕНИЯ ПРЕДВЕСТНИКОВ ГЕОДИНАМИЧЕСКОГО СОБЫТИЯ	14
1.1. Описание предметной области	14
1.2. Обзор методов обнаружения предвестников геодинамического события	20
1.2.1. Методы, основанные на статистическом подходе, регрессионных моделях, и спектральных характеристиках временных рядов.....	20
1.2.2. Нейросетевые и нечеткие методы.....	31
1.3. Обзор программных комплексов, применяемых при обнаружении предвестников геодинамического события	39
1.4. Постановка задачи исследований.....	46
Выводы	48
Глава 2. МОДЕЛЬ ПРЕДВЕСТНИКОВ ГЕОДИНАМИЧЕСКОГО СОБЫТИЯ И ФОРАЛИЗАЦИЯ ЗАДАЧИ ИХ ОБНАРУЖЕНИЯ	48
2.1. Модель изменения характеристик ЭМС на этапе формирования очага разрушения. Постановка задачи обнаружения предвестника геодинамического события	48

2.2. Модель корректных алгоритмов, алгебраический подход к синтезу корректных алгоритмов	57
2.3. Корректирующие операции на основе модификации метода анализа иерархий	58
2.4. Описание класса алгоритмов, используемых для решения задачи обнаружения параметров предвестника геодинамического события ...	60
2.5. Задача построения оптимальной системы окрестностей	65
Выводы	67
Глава 3. АЛГОРИТМ ОБНАРУЖЕНИЯ ПРЕДВЕСТНИКОВ ГЕОДИНАМИЧЕСКОГО СОБЫТИЯ	69
3.1. Корректные алгоритмы α -разметки стохастических временных рядов измерений ЭМС	68
3.1.1. Корректный алгоритм α -разметки стохастического временного ряда ARADS	72
3.1.2. Корректный алгоритм α -разметки стохастического временного ряда SMA	74
3.1.3. Сравнительный анализ алгоритмов α -разметки	82
3.2. Корректный алгоритм β -разметки	86
3.3. Численная процедура построения оптимальной системы окрестностей	89
3.4. Алгоритм обнаружения предвестников геодинамического события	106

Выводы	117
Глава 4. ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ ИССЛЕДОВАНИЯ ГЕОФИЗИЧЕСКИХ ДАННЫХ. БАЗА ДАННЫХ ГЕОФИЗИЧЕСКОГО ПРОЦЕССА.....	119
4.1. Архитектура программного комплекса ReDSS.....	120
4.2. База данных результатов электромагнитных и акустических измерений геофизических процессов.....	126
4.3. WEB-клиент	131
4.4. Программа «ReDSS Geo» для анализа и мониторинга геофизических процессов, описание которых представлено в виде временных рядов измерений ЭМС	138
Выводы	145
ЗАКЛЮЧЕНИЕ	147
СПИСОК ЛИТЕРАТУРЫ	149
ПРИЛОЖЕНИЕ А.....	162
ПРИЛОЖЕНИЕ Б.....	166
ПРИЛОЖЕНИЕ В.....	168
ПРИЛОЖЕНИЕ Г	174

Обозначения и сокращения

АНР, АНР+ - метод анализа иерархий классический (по Т.Саати), модифицированный метод анализа иерархий (от Analytic Hierarchy Process);

ARADS – метод моделирования временного ряда (от AutoRegression, Adaptive algorithm, Difference Scheme);

AR-коэффициенты – вектор коэффициентов $\mathbf{a}^A = (a_1^A, \dots, a_p^A)$ авторегрессии;

DS-коэффициенты – вектор коэффициентов $\mathbf{a}^v = (a_1^v, \dots, a_p^v)$ разностной схемы, сопоставленной функциональной зависимости $f_v(k), v=1, \dots, h$;

SMA – алгоритм разметки временного ряда (от Symbolic, Marking, Algorithm);

SAX – алгоритм позволяющий представить временной ряд в виде символьной последовательности (от Symbolic Aggregate approXimation);

РАА – кусочное совокупное приближение (от Piecewise Aggregate Approximation);

ЭМС – электромагнитный сигнал;

СВР – стохастический временной ряд;

\mathbb{N} – множество натуральных чисел;

\mathbb{R} – поле вещественных чисел;

\mathbb{R}^n – линейное пространство над полем вещественных чисел;

$d(x, y), \rho(x, y)$ – функции расстояния;

$s^n, s = ((t_1, y_1), \dots, (t_n, y_n))$ $n \geq 1$, $t_k, y_k \in \mathbb{R}$ - вектор-объекты;

$s = (t_k, y_k) \in \mathbb{R}^2$, $t_k, y_k \in \mathbb{R}$ - точка или узел на плоскости \mathbb{R}^2 ;

\mathfrak{S} - множество стохастических вектор-объектов s ;

S^n - множество всех различных вектор-объектов длины n ;

$M = \{\mu_1, \dots, \mu_r\}$ - алфавит α -разметки;

$\mu^n = (\mu_1, \dots, \mu_n)$, $\mu_k \in M$, $k = 1, \dots, n$ - α -разметка вектор-объекта s^n ;

$X = \{\chi_1, \dots, \chi_J\}$ - алфавит β -разметки;

$\chi^n = (\chi_1, \dots, \chi_n)$, $\chi_k \in X$, $k = 1, \dots, n$ - β -разметка вектор-объекта s^n ;

\tilde{M} - множество всех α -различных разметок;

\tilde{X} - множество всех β -различных разметок;

A^α, A^β - алгоритмы α и β разметки;

$\tilde{\mathfrak{I}}_{ob}$ - пространство допустимых описаний;

I_{ob} - допустимое описание вектор-объекта s^n ;

$\Phi = \{f_v(k), v = 1, \dots, h\}$ - домен (библиотека) функции α -разметки;

$\tilde{\chi}_j = (\mu_1, \dots, \mu_m) \in \tilde{M}$ - описание параметра предвестника;

$M^X = \{\tilde{\chi}_j, j = 1, \dots, J\}$, $M^X \subset \tilde{M}$ - множество описаний параметров пред-

вестника;

$\mathfrak{M}^1, \mathfrak{M}^\circ, \mathfrak{F}$ - множества решающих правил и алгоритмических операторов, корректирующих операций, соответственно;

C, B, F - решающее правило, алгоритмический оператор, корректирующая операция, соответственно;

\mathfrak{J}_i - множество начальных информаций;

\mathfrak{J}_f - множество финальных информаций;

\mathfrak{J}_e - пространство оценок;

$\mathcal{O}(s_j, u_j) = ((t_{i_1}, y_{i_1}), \dots, (t_j, y_j), \dots, (t_{i_2}, y_{i_2}))$ - окрестность точки s_j длиной

$u_j = i_2 - i_1 + 1$, где $i_1, i_2 \in \{1, \dots, n\}$, $i_1 < i_2$, $j \in \{i_1, \dots, i_2\}$;

\mathcal{O}^n - система окрестностей соответствующая вектор-объекту s^n ;

\mathcal{L} - обучающая выборка;

$\lfloor x \rfloor$ - пол: $\max\{n \mid n \leq x, n - \text{целое}\}$;

$\lceil x \rceil$ - потолок: $\min\{n \mid n \geq x, n - \text{целое}\}$;

$\tilde{q}, \tilde{c}, \tilde{m}$ - строки, результат работы алгоритма SAX;

$\tilde{M} = \{\tilde{m}_v, v = 1, \dots, h\}$ - библиотека эталонных строк, полученная для до-

мена функций Φ ;

a - размер алфавита символов алгоритма SAX;

\mathcal{O}^n - множество всех возможных систем окрестностей \mathcal{O}^n ;

$M^Z \subseteq M^m$ - множество предварительно заданных эталонных под-
разметок;

$N(0, \sigma^2)$ - обозначения нормального вероятностного закона;

$\langle \succ \rangle$ - бинарное отношение доминирования;

ВВЕДЕНИЕ

В последние годы в районах отработки полезных ископаемых участились случаи проявления горных ударов, внезапных выбросов, горно-тектонических землетрясений и других геодинамических явлений. Эти процессы порождаются в массивах горных пород не только действием больших тектонических напряжений, но и увеличением глубины отработки месторождений, структурной неоднородностью массива и свойствами слагающих его горных пород. Крупные технологические взрывы, применяемые для обрушения рудных блоков на железорудных месторождениях, также сопровождаются мощными динамическими явлениями. Все это наносит значительный экономический ущерб промышленным предприятиям и гражданским объектам, а также ведет к травматизму и человеческим жертвам. Поэтому особую остроту приобрела проблема мониторинга изменения напряженно-деформированного состояния массивов горных пород. Для исключения последствий геодинамических явлений необходим надежный прогноз геологических условий ведения горных работ.

Одним из наиболее перспективных методов контроля изменения напряженно-деформированного состояния массивов горных пород, является бесконтактный способ, основанный на регистрации электромагнитных сигналов, возникающих при механическом воздействии на горные породы. При механоэлектрических преобразованиях в горных по-

родах параметры электромагнитных сигналов (ЭМС) несут информацию об изменении напряженно-деформированного состояния и процессах образования деструктивных зон.

В основе данного метода лежат исследования, начатые в конце 60-годов в Томском политехническом институте под руководством чл. корр. АПН, профессора А.А. Воробьева. Исследование механоэлектрических преобразований и проблемы обнаружения предвестников геодинамических явлений получило развитие в работах Гохберга М.Б., Соболева Г.А., Морозова В.Н., Курлени М.В., Мастова Ш.Р. и др.

Данные об изменении параметров ЭМС, получаемые при мониторинге изменений напряженно-деформированного состояния горных пород с использованием специальных регистраторов, представляют собой набор временных рядов. В настоящее время разработкой методов и алгоритмов обнаружения аномалий в геоинформатике (в частности, предвестников геодинамических явлений) во временных рядах занимаются А.Д. Гвишиани, С.М. Агаян, Ж. Злотники, J.-L. LeMouël, Ш.Р. Богоутдинов. Однако до настоящего времени не существуют достоверных алгоритмов обнаружения предвестников, в силу того, что аналитическое описание параметров ЭМС или характеристик электромагнитной эмиссии (ЭМЭ) неполно или невозможно в силу их нелинейности и стохастичности. Поэтому задача обнаружения предвестников геодина-

мических явлений по временным рядам измерений характеристик электромагнитных сигналов является актуальной.

Одним из направлений определения аномалий во временных рядах является направление, связанное с применением алгебраического подхода (Ю.И. Журавлев), который развивается в трудах ВЦ РАН как исследование и апробация технологии разметки временных рядов с позиций теоретико-множественного описания (К.В. Рудаков, Ю.В. Чехович).

В данной работе объектом исследования являются геодинамические явления, в процессе протекания которых возникающая электромагнитная эмиссия в горных породах служит источником электромагнитных сигналов, образующих нестационарные стохастические временные ряды. Предметом исследования являются модели предвестников геодинамических явлений и оптимальные алгоритмы их обнаружения.

Для создания алгоритмов обнаружения предвестников используются методы теории распознавания образов, теории принятия решений и многокритериального оценивания, теории нечетких множеств, формализма мультимножеств, методы теории вероятностей, информации и математической статистики. При моделировании использовались приложения Matlab, Mathcad, OriginPro, для реализации алгоритмов и программного комплекса использовались методы программирования в среде разработки Microsoft Visual Studio 2008, платформа Microsoft.NET, технологии ASP.NET, Microsoft.NET Remoting. При проектировании и

разработки базы данных использовались ERwin Data Modeler, СУБД Microsoft SQL Server Express 2005.

Достоверность результатов подтверждается корректностью постановки задачи, сопоставимостью результатов теоретических исследований с существующими положениями проблемно-ориентированной теории, корректностью постановки экспериментов и обработки экспериментальных данных, качественным и количественным соответствием теоретических исследований и экспериментальных данных, полученными как лично автором, так и другими исследователями.

Представленные в монографии математические модели, алгоритмы и программное обеспечение, использовались при выполнении грантов: РФФИ № 09-01-99014-р_офи «Разработка подходов к повышению эффективности методов и алгоритмов распознавания образов и оценки их качества», РФФИ № 11-07-00666-а «Исследование и разработка информационной системы мониторинга геодинамических процессов, основанной на электромагнитной эмиссии горных пород» и были встроены в программный комплекс «ReDSS», расположенный по адресу: <http://redss.ru>. Разработанные алгоритмы, программный комплекс и база данных используется в ПНИИ ЭДиП Национального исследовательского Томского политехнического университета (г. Томск).

В первой главе дано краткое описание способа контроля изменения напряженно-деформированного состояния, основанного на регист-

рации электромагнитной эмиссии, возникающей при механическом нагружении горных пород. Приводятся причины возникновения эмиссии и их связь характеристик с этапами деформирования горных пород. Дано описание исходных данных для исследований по обнаружению предвестников геодинамического события, проводимых в данной работе. Приводится краткий обзор методов обнаружения предвестников геодинамического события, исходными данными для которых являются временные ряды регистрируемых геофизических данных. Выполнен обзор программных комплексов применяемых при обнаружении предвестников геодинамического события. По итогам обзора формулируются основные требования к разрабатываемому программному комплексу и определяются задачи исследования.

Во второй главе рассматривается процесс изменения характеристик ЭМС в виде аддитивной смеси полезного сигнала и помехи - случайной составляющей наблюдаемого процесса, препятствующей достоверному восприятию полезного сигнала. На этапе формирования очага разрушения горных пород исходные данные представляются в виде нестационарного стохастического временного ряда. Для решения задачи обнаружения предвестника геодинамического события электромагнитный сигнал записывается в виде кусочно-заданных функций с использованием α и β - разметок временного ряда. Было показано, что задача построения α , β - разметок по стохастическому временному ряду может

быть некорректной (возможно несколько решений, удовлетворяющих заданным ограничениям), в связи с этим задача обнаружения предвестника геодинамического события разбивается на две подзадачи: построение корректных алгоритмов α и β - разметки; определение оценки принадлежности к классу предвестника по обучающей выборке. Далее, приводится описание класса корректных алгоритмов α , β - разметки. Для этого были определены: рабочая, финальная, структурная информация, локальные и универсальные ограничения, условия разрешимости и регулярности.

В третьей главе приводится описание алгоритма α -разметки ARADS и разработанного алгоритма α -разметки SMA, алгоритма β -разметки стохастических временных рядов измерений характеристик ЭМС, численной процедуры построения оптимальной системы окрестностей и алгоритма обнаружения предвестников геодинамического события. Приведен сравнительный анализ алгоритмов разметки, показана эффективность численной процедуры построения оптимальной системы окрестностей. Алгоритм обнаружения предвестников геодинамического события тестировался на временных рядах, полученных в ПНИЛ ЭДиП НИ ТПУ при разрушении лабораторных образцов отобранных с железорудного месторождения и данных, полученных во время проведения технологических взрывов на шахте Таштагольская (Кемеровская область).

В четвертой главе дано описание программного комплекса ReDSS, предназначенного для исследования и мониторинга процессов, описание которых представлено в виде временных рядов. В программном комплексе ReDSS, применяются разработанный корректный алгоритм разметки, численная процедура построения оптимальной системы окрестностей и алгоритм обнаружения предвестников геодинамического события. Представлено описание базы данных, которая предназначена для структурированного хранения временных рядов, сформированных на основе зарегистрированных специальной аппаратурой с использованием соответствующих датчиков, установленных в массиве горных пород. Далее приводится описание приложения WEB-клиента, который используется для демонстрации работы алгоритмов, реализованных в виде подключаемых модулей (которые хранятся на стороне сервера) и работы с базой данных результатов электромагнитных и акустических измерений геофизических процессов. Используя интерфейс программирования приложений (набор интерфейсов) программного комплекса ReDSS, было реализовано клиентское desktop-приложение «ReDSS Geo». Приложение «ReDSS Geo», предназначено для исследования разработанных алгоритмов и мониторинга геофизических процессов, описание которых представлено в виде нестационарных стохастических временных рядов.

Глава 1. МЕТОДЫ ОБНАРУЖЕНИЯ ПРЕДВЕСТНИКОВ ГЕОДИНАМИЧЕСКОГО СОБЫТИЯ

1.1. Описание предметной области

В современных исследованиях, связанных с прогнозированием геодинамических явлений (таких как горные удары, землетрясения), ставится задача разработки методов контроля изменения напряженно-деформированного состояния массивов горных пород и прогноза разрушения. Одним из наиболее перспективных методов, является бесконтактный метод, основанный на регистрации электромагнитной эмиссии, возникающей при механическом воздействии на горные породы. Причиной возникновения ЭМЭ является электрический заряд, появляющийся или изменяющийся вследствие развития трещин нормального отрыва при деформации и разрушении неорганических диэлектриков (какими являются основные породообразующие минералы), сопровождаемого пространственным разделением противоположных по знаку электрических зарядов¹.

В результате экспериментальных исследований установлено, что в зависимости от этапа деформирования наблюдается изменение характеристик электромагнитной эмиссии, являющейся следствием механо-электрических преобразований [1]. Причем на каждом этапе деформи-

¹ Возникновение электромагнитных сигналов связывается не только с развитием электрических разрядов, но и с изменением дипольного момента заряженной трещины, колебательным движением заряда при развитии разрушения и пр.

рования в характеристиках электромагнитной эмиссии, присутствуют отличительные особенности. Основываясь на этом, в проблемной научно-исследовательской лаборатории электроники диэлектриков и полупроводников НИ ТПУ разрабатываются методы мониторинга за изменением напряженно-деформированного состояния массивов горных пород и прогноза удароопасности, а также прочности гетерогенных материалов на ранних этапах напряженно-деформированного состояния горных пород. Проведенными исследованиями установлено [2], что характеристики ЭМЭ зависят не только от прочности, но и от структурно-текстурных особенностей, проводимости, влажности и минерального состава горных пород [3, 4]. Таким образом, в сложных гетерогенных системах, какими являются горные породы, естественно ожидать изменение интенсивности ЭМЭ от их состава и строения, а также от характера деформации.

Одним из способов выявления основных закономерностей механоэлектрических преобразований в массиве горных пород является физическое моделирование в лабораторных условиях (рис 1.1, б), которое позволяет выявить связь между напряженно-деформированным состоянием и характеристиками моделируемых механоэлектрических преобразований. Физическое моделирование механоэлектрических преобразований в лабораторных условиях проводилось на прессе одноосным сжатием образцов горных пород до разрушения.

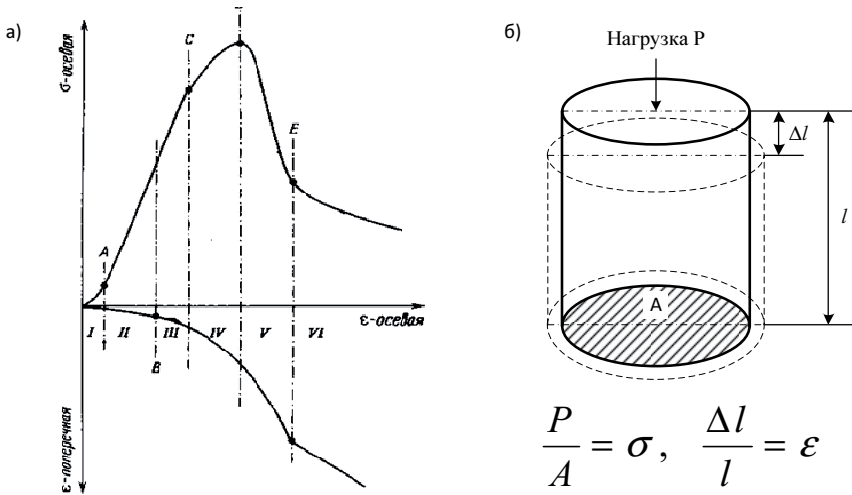


Рис. 1.1. – а) кривые осевого напряжения – осевая деформация и поперечная деформация с разделением на шесть типичных участков по характеру трещинообразования; б) схема приложения усилий на одноосное сжатие образца горной породы

При изменении механической нагрузки на горную породу изменение относительной деформации характеризует процесс трещинообразования, который по результатам исследований, выполненных рядом авторов [5-7], можно разделить на несколько типичных участков (рис 1.1, а) [33]:

- на первом участке происходит локальное разрушение при неупругом сжатии, вызванное шероховатостью поверхности торцов образца и смыканием открытых трещин;
- на следующей стадии наблюдается почти линейно-упругое состояние породы без заметного изменения структуры образца;

– третий участок характеризуется ростом существующих трещин и появлением первых микротрещин в центральной части образца. Этот участок можно охарактеризовать как формирование очага разрушения или начало разрушения;

– на четвертом участке, с приближением к пределу прочности, происходит быстрое увеличение концентрации микротрещин. Этот этап носит название – предразрушающий;

– пятый и шестой участок характеризуются образованием плоскости макротрещины. Происходит быстрое снижение несущей способности и скольжение по плоскости макротрещины с возрастающим разрушением и дроблением.

В основе исследований по обнаружению предвестников разрушения горных пород по характеристикам электромагнитной эмиссии (ЭМЭ) лежит экспериментально доказанное увеличение электромагнитной активности на этапе предразрушения [8,9]. Однако, при изменении напряженно-деформированного состояния горных пород, как для образцов в лабораторных исследованиях, так и для горных пород в массиве, процесс механоэлектрических преобразований происходит на всех этапах деформирования, начиная с момента зарождения микротрещин и заканчивая разрушением. Это объясняется тем, что сам процесс разрушения горных пород не внезапный акт, а заключительный этап процесса, характеризующегося определенной длительностью и проходящий в

своем развитии несколько стадий. Крупные трещины в горных породах образуются путем слияния многих микро- и макротрещин. Это естественное следствие неоднородности горных пород, наличия флюидов и существование трещин (блоков) различных уровней. Неоднородность строения горных пород приводит к неоднородному распределению поля механических напряжений в образце уже на самых ранних этапах механического воздействия. Возможна ситуация, когда при определенной скорости деформирования в ограниченной зоне происходит рост и накопление трещин, а в другой области происходит уплотнение с закрытием ранее существовавших трещин. Такое контрастное поведение различных частей горной породы обуславливает пространственно-временное распределение электромагнитных предвестников, вследствие высокой чувствительности механоэлектрических преобразователей, связанных со структурно-текстурными изменениями среды. Локальные напряжения, возникающие в горных породах, являются случайными функциями координат и времени. Случайной функцией координат в неоднородном теле является и локальная прочность. Микротрещины появляются там, где локальные напряжения превышают локальную прочность. Такая достаточно сложная иерархическая структура горных пород приводит к тому, что на протяжении всего процесса деформирования изменяются параметры регистрируемого ЭМС. Вследствие этого, основной задачей прогноза является оценка связи параметров ЭМС с

определенными этапами напряженно-деформированного состояния и выявление параметра, являющегося прогностическим признаком разрушения.

Методы контроля изменений напряженно-деформированного состояния горных пород и прогноза геодинамических событий на основе регистрации интенсивности ЭМЭ предполагают выполнение определенных условий. Во-первых, разрушение горных пород есть многостадийный процесс возникновения и развития трещин различного размера и формирование очага разрушения отделено во времени от самого акта разрушения. Во-вторых, активизация процесса накопления микротрещин определяет увеличение потока электромагнитной энергии непосредственно из очага готовящегося разрушения.

Исходными данными для исследований по обнаружению предвестников геодинамического события, проводимых в данной работе, являются временные ряды измерений усредненной амплитуды ЭМС на частотах 2 кГц, 15 кГц, 100 кГц и в широкой полосе от 1 до 100 кГц, которые были получены с помощью разработанного и изготовленного в ПНИЛ ЭДиП НИ ТПУ прибора РЭМС1. Время усреднения записываемых в память регистратора РЭМС1 значений в лабораторных условиях составляло 1 секунду, а в полевых условиях - 5 секунд.

1.2. Обзор методов обнаружения предвестников геодинамического события

Методы, используемые для решения задачи обнаружения предвестника геодинамического события, исходными данными которой являются временные ряды регистрируемых геофизических данных, можно разделить на две группы:

- методы, основанные на статистическом подходе, регрессионных моделях и спектральных характеристиках временных рядов;
- нейросетевые и нечеткие методы.

К первой группе относятся различные статистические критерии, методы, основанные на теории фильтрации, Фурье и вейвлет - анализе, пороговые методы. Ко второй группе относятся методы, использующие нейросетевые модели, и методы на основе теории нечетких множеств.

1.2.1. Методы, основанные на статистическом подходе, регрессионных моделях и спектральных характеристиках временных рядов

Геофизические данные, представленные в виде реальных временных рядов, как правило, содержат параметрическую неопределенность и нестационарность, их обработка требует использование различных методик и технологий. Перечислим условия применимости методов, рассматриваемых в данном параграфе.

При использовании регрессионной модели для нестационарных временных рядов и ее уточнение в зависимости значений математического ожидания, дисперсии или ковариации от времени, производится двумя способами:

- аналитическим моделированием данной зависимости;
- переходом к первым, вторым и т.д. разностям, которые выражают данную зависимость.

В авторегрессионных моделях, которые используют лаговую автокорреляцию, применяется выборочная автокорреляционная функция, усреднение производится по «окну», а его размер подбирается экспериментально. При использовании адаптивных моделей оценка текущего уровня может определяться как взвешенное среднее всех предшествующих уровней, причем веса при наблюдениях убывают по мере удаления от последнего уровня. Вопрос об оптимальном выборе сглаживающих функций и весов решается субъективно, а на практике - переборным способом.

К методам спектрального анализа обычно относят методы, использующие преобразования Фурье. В обобщенном смысле спектральным методом можно называть любое представление наблюдаемого сигнала в виде суперпозиции некоторых базисных функций (например, вейвлет-функции). При спектральном разложении сигнала имеет место «размытие» спектральных составляющих из-за конечной длительности

обрабатываемых наборов данных. Рекомендуется использовать спектральные методы в следующих случаях:

- интервал наблюдений содержит большое число колебаний с наибольшим периодом или при анализе временных рядов большой длины при условии можно пренебрежения краевых эффектов;

- форма колебаний на всех периодах не изменяется, а наибольший период и периоды гармоник стационарны на интервале наблюдения.

Для временных рядов геофизических данных, обычно, все выше приведенные условия одновременно не выполняются, поэтому попытки использования спектральных методов в этой предметной области, могут привести к неадекватным выводам. Кроме того, при неизвестной природе сигнала, когда отсутствует математическое описание шума, доказано отсутствие оптимального спектрального окна [94].

Рассмотрим методы обнаружения предвестников геодинамического события, основанные на статистическом подходе, регрессионных моделях и спектральных характеристиках временных рядов.

Метод на основе полосовых фильтров Баттерворта [32]. Для обнаружения предвестника можно использовать набор полосовых фильтров Баттерворта, охватывающих с перекрытием всю полосу частотной

характеристики предвестника. В этом методе используется трехкомпонентный поляризационный анализ.

Прогноз разрушения горных пород по величинам максимальной амплитуды и частоте спектральной плотности сигнала [10]. Данный способ прогноза разрушения горных пород в массиве заключается в анализе спектральной характеристики зарегистрированного импульса ЭМЭ (или серии импульсов), т.е. определяется максимум спектральной плотности импульса и частота, которой она соответствует. После регистрации серии сигналов, сравнивают величины амплитуд и соответствующих им частот импульсов. Затем определяют скорость изменения максимальной спектральной амплитуды во времени и по частоте, на основании которой принимается решение о начале разрушения исследуемого участка массива [11]. Этот способ позволяет определить момент времени, соответствующий началу разрушения массива. Однако, для использования данного способа необходимы специальные измерения и подготовительные работы в забое горной выработки.

Исследование процесса разрушения горных пород с помощью электромагнитной эмиссии на основе S - образной кривой в координатах "время - спектральная частота". Авторы данной методики Кулаков Г.И. и Яковицкая Г.Е. предлагают анализировать матрицы спектрально-временных амплитуд для определения уровня напряженного состояния массива горных пород и оценки развития в нем процесса

трещинообразования. Расположение точек с максимальными спектральными амплитудами на S -образной кривой подчиняется, согласно [12,13], следующей закономерности: S -образная кривая в пределах рассматриваемого участка спектрально-временной матрицы ЭМЭ начинается в области низких частот, но с возрастанием механической нагрузки сдвигается в область более высоких частот. Наибольшей частоте соответствует максимальное значение спектральной амплитуды. Авторы полагают, что появление максимального значения спектральной амплитуды характеризует образование в образце горной породы магистральной трещины.

Прогноз разрушения массива горных пород за счет выявления в процессе регистрации ЭМИ-сигналов начала и последующего развития квазипериодических низкочастотных колебаний с нарастающими амплитудами и периодами [14]. При возникновении периодических квазирезонансных колебаний при регистрации ЭМЭ-сигналов определяются их амплитуды и периоды этих низкочастотных колебаний, считая, что их возникновение свидетельствует о начале процесса разрушения массива горных пород. Последующее измерение сигналов ЭМЭ фиксирует продолжающееся увеличение длительности периодов квазипериодических низкочастотных колебаний и их амплитуд. В случае установления момента времени, при котором начинается одновременное увели-

чение этих параметров, можно прогнозировать дальнейшее развитие процесса разрушения при условии выполнения соотношений:

$$T_2 \geq (2 \div 4)T_1, \quad (1.1)$$

где T_1 и T_2 – соответственно, длительность начального и последующего периодов квазипериодических низкочастотных колебаний сигналов, и

$$A_2 \geq 3A_1, \quad (1.2)$$

где A_1 и A_2 – соответственно, амплитуда начального и последующего периодических низкочастотных колебаний этих сигналов. Выполнение вышеприведенных соотношений дает основание считать, что наступило критическое напряженно-деформированное состояние массива, после которого последует его разрушение [15].

Данный метод позволяет диагностировать этап разрушения, но не подходит для обнаружения предвестников на этапе предразрушения горных пород, кроме того, не учитываются признаки формы ЭМЭ-сигналов.

Соотношение сигналов излучения в нагруженном и ненагруженном участках массива как прогнозный критерий разрушения. В данном методе начало разрушения исследуемого участка массива определяют по амплитудам сигналов ЭМИ на некотором интервале времени [16].

При этом интервал времени измерения разделяют на две неравные части, производя на каждый из них измерение величин амплитуд сигналов через равные промежутки времени. До нагружения исследуемого

участка массива определяют интенсивность излучения сигналов помехи, измеряя амплитуды сигналов на большей части интервала времени измерения, а начало момента разрушения исследуемого участка массива определяют по выполнению соотношения:

$$\frac{m \sum_{i=0}^{n-1} \left[y_i - \frac{1}{n} \sum_{j=0}^{n-1} y_j \right]}{(n-1) \sum_{i=0}^{m-1} \left[x_i - \frac{1}{n} \sum_{j=0}^{m-1} k_j \right]^2} \geq C_0, \quad (1.3)$$

где m – количество отсчетов амплитуд сигналов на меньшей части интервала времени измерения; n – количество отсчетов амплитуд сигналов на большей части интервала времени измерения; x – величины амплитуд сигналов на большей части интервала времени измерения; y – величины амплитуд сигналов на меньшей части интервала времени измерения; i, j – текущие координаты при отсчете амплитуд сигналов помехи и полезного сигнала; C_0 – пороговая константа, определяемая из таблиц F-распределения Фишера с $(n-1)$, $(m-1)$ степенями свободы (при заданном уровне вероятности ложного принятия решения о срабатывании).

Разделение интервала времени измерения на две неравные части и измерение амплитуд сигналов на большей из них до начала измерения позволяет оценить интенсивность излучения сигнала помехи на исследуемом участке горного массива. Последующие же измерения через равные промежутки времени амплитуд сигнала на меньшем из них по мере нагружения исследуемого участка массива дают возможность вы-

явить на фоне сигнала помехи сигнал регистрируемого полезного электромагнитного излучения [17].

Способ прогноза разрушения массива горных пород на основании уменьшения периода следования сигналов электромагнитного излучения. Данный способ прогноза предложен на основании результатов лабораторных экспериментов, заключается он в регистрации сигналов электромагнитного излучения, измерении длительности этих сигналов, дополнительном одновременном измерении длительности T_i периодов следования сигналов и нагрузки на исследуемый участок массива. Наблюдается рост нагрузки до максимального значения, а затем ее уменьшение со скоростью V_i по соотношениям:

$$T_n \leq (10^3 + 10^4) T_n, \quad (1.4)$$

$$V_n \leq (7 + 10) V_n, \quad (1.5)$$

где T_n – минимальное значение длительности следования сигналов ЭМЭ (n-й период); T_n – длительность периода следования сигналов ЭМЭ, с которого начинается ее уменьшение; V_n – максимальное значение скорости уменьшения нагрузки на исследуемом участке массива (n-й период следования сигналов ЭМИ); V_n – скорость уменьшения нагрузки на исследуемом участке массива, с которого начинается ее увеличение, судят о критическом состоянии нарушения сплошности исследуемого участка массива, после которого наступает его разрушение [15, 18].

В последние годы большое распространение получила группа методов спектрально-временного анализа осциллограмм [2], с помощью быстрого преобразования Фурье, получаемых при регистрации ЭМЭ.

Рассмотрим три существующие методики анализа информации, содержащейся в спектрально-временных матрицах.

Исследование поля изолиний спектральных амплитуд. Данная методика описана в монографии под редакцией Гохберга М.Б. [19], где показано, что к моменту разрушения образцов горных пород на спектрально-временных матрицах ЭМЭ соответствуют области повышенных значений спектральных амплитуд.

Построение трехмерного изображения спектрально-временной характеристики. Использование данной методики позволяет построить трехмерное изображение массива спектральных амплитуд, характеризующих процесс разрушения, для дальнейшего экспертного анализа [20].

Прогноз разрушения горных пород на основе особенностей спектрально-временных характеристик сигналов электромагнитного излучения [15]. В данном методе прогноз определяется в зависимости от выполнения следующих условий (при обнаружении следующих предвестников):

- одновременное увеличение максимальной спектральной амплитуды и соответствующей ей частоты по мере роста нагрузки и после-

дующее их уменьшение свидетельствует о приближении момента разделения сплошности горной породы;

- переход через нуль приращения максимальной спектральной амплитуды по мере роста нагрузки (и соответственно, времени) также может служить прогнозным признаком приближения момента разрушения;

- одновременное уменьшение скорости изменения (производной) максимальной спектральной амплитуды и ее модуля по времени в процессе роста нагрузки и производной максимальной спектральной амплитуды по частоте в зависимости от времени до значений, близких к нулю, а также последующее их возрастание по модулю служат прогнозной характеристикой приближения момента нарушения сплошности горной породы;

- уменьшение производных максимальной спектральной амплитуды по времени и частоте в зависимости от изменения спектральной частоты до минимального значения на максимальной для обеих величин частоте и последующий их рост по модулю свидетельствует о начале процесса разделения сплошности горной породы.

Подчеркнем, что описанные выше методы, основанные на спектрально-временного анализе, рассматривают поведение электромагнитного сигнала лишь на конечной стадии разрушения, т.е. рост трещин и ЭМЭ на ранних стадиях нагружения образцов горных пород, не учиты-

валась. По этой причине матрицы спектрально-временных амплитуд строились исключительно для моментов времени, предшествующих окончательному разрушению образцов.

Авторегрессионный метод обнаружения предвестников геодинамического события описан в [21]. Метод основан на использовании адаптивных и согласующих фильтров. Алгоритм такого детектора строится на основе авторегрессионного описания (АРСС-модели) предвестника и помех. Фильтрация состоит в непрерывном анализе шума. Соответственно делается прогноз характера данных, которые должны быть зарегистрированы в последующий момент. Если накопленный на текущий момент времени прогноз о шумах не сбывается, то делается предположение, что зарегистрирован предвестник. Недостатками данной группы методов является малая помехоустойчивость и большее количество ложных срабатываний, при исследовании данных ЭМЭ, полученных при разрушении образцов горной породы.

Метод обнаружения предвестников геодинамического события на основе применения многомерных модификаций спектральных методов главных компонент, канонических когерентностей и вейвлет-анализа описан в [23]. В данном методе обнаружение предвестника производится по эволюции произведения робастных покомпонентных канонических корреляций на различных уровнях детальности и частотно-временных диаграмм эволюции статистик покомпонентных канониче-

ских когерентностей, которые были построены на основе робастного метода многомерного вейвлет-анализа [22].

Статистические критерии формирования очага разрушения. В [24] экспериментально установлено существование корреляционных соотношений (1.6), (1.7) между временем развития очага T , размером области его подготовки R и величиной упругого энерговыделения при потере очаговой устойчивости E :

$$\lg T = a \lg E + b, \quad (1.6)$$

$$\lg R = c \lg E + d, \quad (1.7)$$

где a, b, c, d - постоянные величины.

В прогностические критерии не входят физико-механические параметры материалов, которые не всегда доступны на практике. Обычно используются универсальные статистические параметры, но различие в физико-механических свойствах материалов, в том числе горных пород, не может не сказываться на особенностях разрушения, в том числе на значениях коэффициентов a, b, c, d в (1.6) и (1.7), но согласно экспериментальным данным, оно не нарушает статистические зависимости процесса.

1.2.2. Нейросетевые и нечеткие методы

Нейросетевые методы. Рассмотрим использование нейросетевого подхода к обнаружению предвестников геодинамического события во временных рядах геофизических данных.

В методе, который описан в [25], используется искусственная нейронная сеть, представляющая собой, многослойный перцептрон прямого распространения (рис. 1.2).

В качестве входных параметров в нем использовались средние абсолютные значения спектральных амплитуд в девяти частотных полосах $x_1...x_9$, на выходе пять состояний $f_1...f_5$, из которых два классифицируют входной сигнал как шум, а оставшиеся три как предвестники разных видов.

Нейронная сеть обучается с «учителем» на основе коррекции ошибок. Концептуально участие учителя можно рассматривать как наличие знаний об окружающей среде, представленных в виде пар вход-выход, при этом сама среда неизвестна обучаемой нейронной сети.

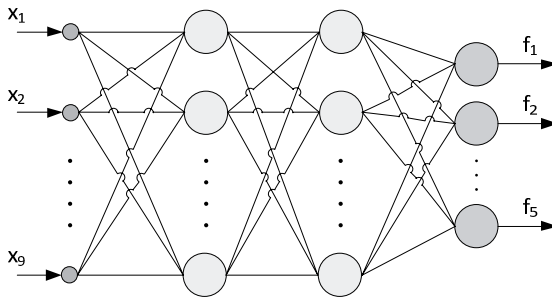


Рис. 1.2. – Многослойный перцептрон

Если предположить, что «учителю» и обучаемой сети подается общий вектор из окружающей среды, то на основе встроенных знаний «учитель» может сформировать желаемый отклик и передать обучаемой

мой сети. Этот желаемый результат представляет собой оптимальные действия, которые должна выполнить нейронная сеть. Параметры сети корректируются с учетом обучающего вектора и сигнала ошибки. Сигнал ошибки – разность между желаемым сигналом и текущим откликом нейронной сети. Корректировка параметров выполняется пошагово с целью имитации нейронной сетью поведения «учителя». Эта эмуляция в некотором статистическом смысле должна быть оптимальной. Таким образом, в процессе обучения знания «учителя» передаются в сеть в максимально полном объеме.

Общими недостатками нейросетевых методов является долгое обучение, и большой объем обучающей выборки, что часто делает применение методов данной группы невозможным. Однако применение релаксационных сетей [26], позволяет существенно уменьшить объем выборки эталонов (достаточно иметь хотя бы один эталон для каждого распознаваемого класса). Релаксационные нейронные сети, характеризуются прямым и обратным распространением информации между слоями сети. Циркуляция информации продолжается до тех пор, пока не установится состояние равновесия. Весовые коэффициенты вычисляются только один раз, до начала работы сети. Матрица весовых коэффициентов полностью определяется имеющимися в наличии эталонными образцами, которые представляют сигналы, принадлежащие различным распознаваемым классам. Вычисление указанных коэффициен-

тов может рассматриваться как обучение сети с использованием информации об эталонных представителях различных типов сигналов. В отличие от традиционных обучаемых структур, релаксационные сети запоминают образцы до ввода реальных анализируемых данных.

Нечеткие методы базируются на системе понятий теории нечетких множеств и нечеткой логики. Под нечетким множеством A понимается совокупность:

$$A = \{(x, \mu_A(x)) \mid x \in X\}, \quad (1.8)$$

где X — универсальное множество, а $\mu_A(x)$ — функция принадлежности (характеристическая функция), характеризующая степень принадлежности элемента x к нечёткому множеству A .

Функция $\mu_A(x)$ принимает значения в некотором вполне упорядоченном множестве M . Множество M называют множеством принадлежностей, часто в качестве M выбирается отрезок $[0,1]$. Если $M = \{0,1\}$, то нечёткое множество может рассматриваться как обычное, чёткое множество. В большинстве случаев предпочитают задавать треугольную функцию принадлежности.

Алгоритмы, основанные на нечеткой логике [27], моделируют рассуждения и действия человека при поиске предвестника геодинамического события. Описанные в [28, 29] алгоритмы - это попытка моделирования логики интерпретатора, распознающего «на глаз» предвестник на записи с целью ее дальнейшего автоматизированного использо-

вания для анализа больших массивов данных, не поддающихся ручной обработке. Эти алгоритмы дают оценку границ наличия искомых предвестников и осуществляют их морфологический разбор в виде начальной, центральной и конечной стадий с выделением сильных и слабых фаз в центральной стадии [30]. Достаточная «гибкость» алгоритмов обеспечивается посредством широкого набора «выпрямлений», возникающих при моделировании работы интерпретатора.

Упрощая работу интерпретатора, визуально определяющего предвестник на записи, авторы понимают ее следующим образом. Сначала он скользит взглядом по записи, оценивая для себя в терминах положительных чисел активность ее небольших фрагментов. При этом выработанные числовые оценки «мысленно» присваиваются фрагментам или центрам фрагментов записи. Так от исходной записи интерпретатор переходит к неотрицательной функции, которую естественно назвать «выпрямлением» записи.

Действительно, ее точкам, более активным с точки зрения искомых сигналов, будут соответствовать большие значения этой функции («выпрямления»). Далее, поиск интерпретатором предвестника на записи сводится к поиску возвышенностей на ее выпрямлении, отвечающим наиболее активным участкам записи. Таким образом, интерпретатор работает на двух уровнях: локальном – «выпрямлении» записи, и глобальном – поиске возвышенностей на «выпрямлении».

Рассуждения интерпретатора во многом определяются тем конкретным типом предвестника (данных), которые он рассматривает. Представляется, однако, что процесс «выпрямления» в той или иной форме оказывается «задействованным» в любом случае.

На первом этапе рассматриваемых алгоритмов происходит «выпрямление» по формуле, взятой из набора составленного по экспертным данным, этот этап называется локальным уровнем.

Собственно нечеткие методы применяются на глобальном уровне алгоритма, где данные полученные в результате процедуры «выпрямления», делятся на фоновую (спокойную) и потенциально-аномальную (неспокойную) части. Связные области в беспокойной части и будут основаниями (платформами) возвышенности. Далее на платформах определяются участки соответствующие предвестнику.

Недостатком описанного выше алгоритма является сложность определения «выпрямлений», по экспертным данным, общим же недостатком нечетких методов является экспоненциальное увеличение сложности вычислений в зависимости от количества входных переменных [31] и, как следствие, увеличение базы правил, что усложняет ее восприятие.

1.3. Обзор программных комплексов, применяемых при обнаружении предвестников геодинамического события

В данном параграфе приведен обзор программных комплексов, применяемых для обнаружения предвестников геодинамического события, реализации и тестирования методов их обнаружения.

Система SESMO1 [32], предназначенная для обнаружения короткопериодных сейсмических аномалий в режиме реального времени во временной и частотной областях. Алгоритм, используемый в системе, реализует метод обнаружения предвестников геодинамического события на основе фильтров Баттерворта. В системе для обнаружения предвестника используются восемь полосовых фильтров и трехкомпонентный поляризационный анализ.

Программный комплекс «Дыхание Земли» [34], решает задачу исследования шумов с целью обнаружения предвестников и прогноза сейсмических катастроф. Данный комплекс позволяет:

- проводить поиск периодических колебаний в заданном диапазоне периодов;
- оценивать спектр мощности методами максимальной энтропии Бурга и путем усреднения периодограмм, оценивать эволюцию логарифма спектра мощности в скользящих временных окнах заданной длины (оценка энтропии Бурга в каждом окне);

- оценивать тренд и выполняет процедуру компенсации тренда с помощью сглаживания гауссовскими ядрами с заданным масштабом усреднения или локальными полиномами с заданным порядком полинома (от 0 до 10) в скользящих окнах заданного радиуса;

- выполнять полосовую частотную фильтрацию временного ряда, пропуская гармонические сигналы с частотами в выбранной полосе;

- осуществлять ортогональное вейвлет-пакетное разложение сигнала с расщеплением каждого уровня детальности в 1 (обычное ортогональное разложение), 2, 4 или 8 раз, производить нелинейную пороговую вейвлет-фильтрацию с оптимальным вейвлетом и порогом Донохо-Джонсона или с другим вейвлетом и степенью сжатия сигнала, выбираемых пользователем, вычислять частотно-временную диаграмму логарифма квадрата модуля непрерывного вейвлет-преобразования с комплекснозначным вейвлетом Морле и логарифма квадрата модулей ортогональных вейвлет-пакетных коэффициентов с расщеплением каждого уровня детальности в 1 (обычное ортогональное разложение), 2, 4 или 8 раз («ящики Гейзенберга»);

- выделять длинные цепи скелета максимума модулей непрерывных вейвлет-преобразований с использованием 1-й и 2-й производной от гауссовского ядра;

- осуществлять подгонку для временного ряда низкочастотной гармонике с неизвестным периодом, который находится из условия минимума дисперсии остатка;

- поиск экстремальных точек, периода и амплитуды «максимального всплеска» для выделенных фрагментов.

Кроме этого, программный комплекс включает алгоритмы предварительной обработки временных рядов, переход к большему интервалу дискретизации, переход к временному ряду в приращениях, итеративное устранение больших выбросов, вычисление дисперсии исходного или преобразованного ряда.

Программно-аппаратный комплекс "АЛАРМ-СЕЙСМО 002" [35] предназначен для проведения экспериментальных исследований по проблеме прогноза землетрясений краткосрочного масштаба времени. Комплекс осуществляет измерение, сбор, компоновку, отображение, хранение и перенос геофизических данных на другие носители информации.

Комплекс построен по принципу непрерывной регистрации импульсного электромагнитного сигнала методом счета амплитудных превышений заданных порогов сигнала. Для исключения влияния помех техногенного и природного происхождения используется алгоритм формирования гистограмм. Данные измерений помещаются в базу данных.

САПР комплекса прогноза землетрясений [36]. Данная система позволяет выбирать аппаратуру, рассчитывать расстояние между датчиками, для обеспечения максимального покрытия обрабатывать данные, используя фазово-градиентный и корреляционный методы, рассчитывать фазовую скорость и градиент и определять координаты литосферных источников. Обнаружение предвестника производится пороговым методом.

Для анализа геофизических данных и решения задачи обнаружения предвестников геодинамического события, также могут использоваться следующие программные комплексы:

Проект LASIP (Local Approximations in Signal and Image Processing.) [37], реализованный в программной среде MATLAB, посвящен исследованию широкого класса эффективных адаптивных методов обработки сигналов. В проекте реализованы новые статистические методы восстановления зашумленного и размытого одномерного сигнала.

WAVELAB 850 [38], набор Matlab функции, которые были использованы авторами в различных алгоритмах вейвлет-анализа.

Система ЛОПЕГ (Богомолов В.П., Виноградов А.П., Борончихин В.А., Журавлев Ю.И. и др.) предназначена для решения задач распознавания образов и выявления логических закономерностей в знаниях, а также голосования по ним.

Система ЭКСИЛОП (А.Д. Закревский) - экспертная система логического распознавания работает с разнообразными объектами, описываемыми точками в пространстве многозначных переменных, и представлением знаний как системы закономерностей в форме секционированной булевой матрицы дизъюнктов.

Универсальная программная система интеллектуального анализа данных, распознавания и прогноза РАСПОЗНАВАНИЕ (ООО «РЕШЕНИЯ», г. Москва), предназначенная для решения задач анализа данных, прогноза, диагностики, выявления закономерностей и поддержки принятия оптимальных решений в различных областях по широкому кругу задач (в том числе и геологических).

Пакеты программ статистической обработки данных, такие как STATISTICA, SAS, SPSS, STADIA, STATGRAPHICS, SYSTAT, Олимп: СтатЭксперт Prof., Forecast Expert и др., позволяют проводить дисперсионный, линейный и нелинейный регрессионный анализ стохастических временных рядов, анализ рядов с интервенциями, экспоненциальное сглаживание, спектральный анализ одного ряда и кросс-спектральный анализ двух рядов. Реализованные в программных пакетах алгоритмы, рассчитаны на аппроксимацию имеющихся данных стационарных рядов, поскольку подбор скользящего окна остается за пределами возможностей пакетов, а методы автоматизированного «подбора» модельных трендовых функций отсутствуют. Выбор наиболее адек-

ватной модели стационарных стохастических временных рядов проводится по известным алгоритмам [39], позволяющим выбрать оптимальное число параметров в рамках дисперсионного и корреляционного анализа. Однако для нестационарных стохастических временных рядов аналитически обоснованные алгоритмы отсутствуют.

Среди систем тестирования методов обнаружения (прогноза) предвестников геодинамического события, следует отметить такие системы:

Автоматизированная система вычисления эффективности методик прогноза землетрясений [40] - разработана в лаборатории сейсмического мониторинга Камчатского филиала Геофизической Службы РАН, позволяет автоматически вычислять эффективность методики прогноза землетрясений.

В системе можно выделить следующие структурные компоненты: блок ввода данных, модуль базы данных, блок вывода отфильтрованных данных, расчетный модуль, блок вывода результатов вычислений.

Модуль базы данных включает в себя региональный сейсмический каталог, таблицы формализованных прогнозов, содержащие: заявленные автором периоды времени в которые может произойти землетрясение, диапазоны энергетических классов и магнитуд, интервалы глубин и вероятные места реализации землетрясений, которые определяются произвольными замкнутыми полигонами географических координат.

Вычисление эффективности прогнозов производится по методике, в которой за меру эффективности принято отношение вероятности сильных землетрясений в прогнозируемые периоды времени к их средней вероятности.

Полигон алгоритмов [41] предназначен для массового тестирования алгоритмов классификации на реальных задачах и представления результатов тестирования через web-интерфейс.

Система ЭЙСДОС [42] выполняет функции разработки оптимальных адаптивных методик идентификации (тестирования) и прогнозирования состояний сложных систем, разработки супертестов в самых различных предметных областях. Система позволяет проводить углубленный анализ результатов тестирования, включающий ранговый (информационный) и кластерно-конструктивный анализ эталонных описаний классов распознавания и признаков, а также анализ достоверности заполнения исходных данных.

Часть рассмотренных в обзоре программных комплексов (SESMO1, АЛАРМ-СЕИСМО 002, САПР комплекса прогноза землетрясений) являются узкоспециализированными и содержат ограниченное число реализованных методов и алгоритмов, характеризуются отсутствием или скудностью инструментария для анализа и обработки данных. Другие («Дыхание Земли», LASIP, WAVELAB, ЛОРЕГ, ЭКСИЛОП, РАСПОЗНАВАНИЕ, и др.), наоборот, содержат сравнительно большое

количество реализованных методов и алгоритмов обработки и анализа данных, являются избыточными и предназначены не только для решения задачи обнаружения предвестника геодинамического события. Стоит также отметить, что не один из рассмотренных комплексов, кроме специализированных (Автоматизированная система вычисления эффективности методик прогноза землетрясений, Полигон алгоритмов, ЭЙС-ДОС), не позволяет тестировать новые методы и алгоритмы на заданной контрольной выборке и расширять набор используемых для обнаружения методов.

Ввиду того, что существует множество методов обнаружения предвестников геодинамического события, и условия применимости каждого из них могут зависеть как от параметров исходных данных, так и от объекта наблюдения, возникает необходимость в программном комплексе, который позволял бы расширять набор используемых при обнаружении алгоритмов, их комбинировать, демонстрировать их работу используя контрольную выборку. Контрольные выборки и обучающие выборки (используются для обучаемых алгоритмов) должны формироваться по базе данных, в которой хранится история наблюдений.

1.4. Постановка задачи исследований

Под сигналом будем понимать величину, отражающую каким-либо образом состояние физической системы. Сигнал рассматривается как

результат некоторых измерений, проводимых над физической системой в процессе ее наблюдения.

В роли физической системы в нашем случае выступает горный массив и нагружаемые образцы горных пород, имеющие различные структурно-текстурные особенности.

Так как, на регистрируемый электромагнитный сигнал оказывают влияние (модель сигнала нам не известна) техногенные процессы, не связанные с изменением напряженно-деформированного состояния массива или лабораторного образца, будем считать, что временные ряды, являющиеся исходными данными (см. п. 1.1), имеют случайную составляющую и, далее, будем их называть стохастическими.

Предполагается, что изменение сигнала на этапе формирования очага разрушения (предвестника) можно описать кусочно-заданной функцией. Тогда функции из набора функциональных зависимостей, с помощью которых можно однозначно описать изменение сигнала на этапе формирования очага разрушения, будем называть параметрами предвестника.

Поэтому, для решения задачи обнаружения предвестников геодинамического события, необходимо использовать исходные данные, которые представлены в виде набора стохастических временных рядов.

Выводы

1. Так как разрушение горных пород есть многостадийный процесс возникновения и развития трещин различного размера, предвестником геодинамического события будем называть этап формирования очага разрушения, отделенный во времени от самого акта разрушения и этапа предразрушения и сопровождающийся активизацией процесса накопления микротрещин, который определяет увеличение потока электромагнитной энергии из очага готовящегося разрушения. Исходными данными для исследований по обнаружению предвестников геодинамического события, проводимых в данной работе, являются временные ряды измерений усредненной амплитуды ЭМС на частотах 2 кГц, 15 кГц, 100 кГц и в широкой полосе от 1 кГц до 100 кГц, которые были получены с помощью разработанного и изготовленного в ПНИЛ ЭДиП НИ ТПУ прибора РЕМС1.

2. Рассмотренные методы обнаружения предвестников геодинамического события, предназначены для использования на этапе предразрушения, что уменьшает время принятия решения, так как сразу за этапом предразрушения следует этап разрушения. Кроме того большинство методов основаны на спектральном анализе, и для того, чтобы получить адекватные результаты требуется наличие значительного объема исходных данных.

3. В результате обзора программных комплексов, которые могут быть использованы для обнаружения предвестников геодинамического события, показана необходимость в программном комплексе, который позволял бы расширить набор используемых при обнаружении алгоритмов, их комбинировать, демонстрировать их работу используя контрольную выборку. Контрольные выборки и обучающие выборки (используются для обучаемых алгоритмов) должны формироваться по базе данных, в которой хранится история наблюдений.

4. Предполагается, что предвестник геодинамического события может быть однозначно определен набором параметров, тогда для его обнаружения необходимо решить задачу обнаружения параметров предвестника по исходным данным, которые представлены в виде набора стохастических временных рядов.

Глава 2. МОДЕЛЬ ПРЕДВЕСТНИКОВ ГЕОДИНАМИЧЕСКОГО СОБЫТИЯ И ФОРАЛИЗАЦИЯ ЗАДАЧИ ИХ ОБНАРУЖЕНИЯ

2.1. Модель изменения характеристик ЭМС на этапе формирования очага разрушения. Постановка задачи обнаружения предвестника геодинимического события

Для моделирования регистрируемого процесса $y(t)$ изменения характеристик ЭМС используется аддитивная смесь полезного сигнала $x(t)$ и случайной составляющей в виде помехи $\eta(t)$, препятствующей достоверному восприятию сигнала:

$$y(t) = x(t) + \eta(t). \quad (2.1)$$

Изменение сигнала $x(t)$ на этапе формирования очага разрушения предлагается формировать в виде кусочно-заданной функции:

$$x(t) = \begin{cases} f_1^x(t), & t_0 < t \leq t_1 \\ f_2^x(t), & t_1 < t \leq t_2 \\ \vdots \\ f_n^x(t), & t_{n-1} < t \leq t_n \end{cases} \quad (2.2)$$

где $f_k^x(t) = \Pi(t_{k-1}, t_k) \left(f_{k-1}^x(t_{k-1}) + \phi_{l(k)} \left(\frac{t - t_{k-1}}{\Delta t} \right) \right)$, $t: N^x \rightarrow N^\phi$, $N^x = \{1, \dots, n\}$, $N^\phi = \{1, \dots, J\}$,

$k \in N^x$, $j \in N^\phi$, $\Delta t = t_k - t_{k-1}$, $\Pi(\tau_1, \tau_2) = \theta(\tau_1) - \theta(\tau_2)$, $\theta(\tau) = \begin{cases} 1, & t \geq \tau \\ 0, & t < \tau \end{cases}$ - функция Хевисайда, $\phi_j(t)$ - функция из набора $F_\omega = \{\phi_j\}$, интерпретируемая как пара-

метр предвестника, $t_0 \leq t_1 \leq \dots \leq t_n$ - моменты скачкообразного изменения вида функциональной зависимости (параметра предвестника).

Будем полагать, что регистрируемый процесс в момент времени t обязательно находится в одном из J состояний.

Выражения (2.1) и (2.2) принимаем за математическую модель этапа формирования очага разрушения (предвестника).

Поскольку регистрируемый процесс $y(t)$ подвергается дискретизации и квантованию, истинное значение дискретного сигнала целесообразно представить в виде суммы:

$$y_k = y'_k + e_k, \quad (2.3)$$

где $k > 0$, $y'_k = y(k\Delta t)$, Δt - интервал времени опроса информационных сигналов датчиков ($\Delta t = 1$ с.), e_k - погрешность квантования. Поскольку последовательность e_k носит случайный характер, то квантование можно рассматривать как случайный процесс – наложение на исходный сигнал шума, который принято называть шумом квантования [43].

Относительно шума квантования введем допущения:

1. шум квантования – стационарный случайный процесс;
2. последовательности e_k и точных значений y'_k регистрируемого процесса статистически независимы;
3. случайные величины $\{e_k\}$ представляют собой «белый» шум;

4. плотность распределения вероятности ошибки равномерна во всем диапазоне ошибок квантования; математическое ожидание m_e и дисперсия σ_e^2 шума квантования определяется следующим образом:

$$m_e = -\frac{1}{2}Q,$$

$$\sigma_e^2 = Q^2/12,$$

где, $Q = 2^{-b}$ - шаг квантования, b - число разрядов квантования.

Используя (2.1) и (2.3) можем принять, что исходные данные представлены в виде стохастического временного ряда (СВР), т.е. некоторого множества выборочных значений случайного процесса:

$$y_k = x_k + \eta_k + e_k, \quad (2.4)$$

где $x_k = x(k\Delta t)$, $\{\eta_k\}$ - белый шум с нулевым математическим ожиданием и неизвестной ограниченной дисперсией σ_η^2 .

Введем домен элементарных базисных функций $\Phi = \{f_v(k), v = 1, \dots, h\}$, где $f_v(k)$ - некоторая аналитически заданная функция.

Будем полагать, что для каждого i -го фрагмента СВР (2.4) регистрируемый процесс описывается моделью:

$$y_k^i = x_k^i + \eta_k^i + e_k^i, \quad (2.5)$$

где $i = 1, \dots, I$, $x_k^i = f(k\Delta t)$, $\Delta t > 0$, $f(t) \in \mathbb{R}$ - базисная функция из множества Φ , описывающая поведение неслучайной составляющей x_k^i в i -м фрагменте СВР.

Используя методологию, которая была предложена в работах [44,45], СВР можно представить в виде множества \mathfrak{S} стохастических вектор-объектов $s^n = ((t_1, y_1), \dots, (t_n, y_n))$, $n \geq 1$, $t_k, y_k \in \mathbb{R}$, $s_j = (t_k, y_k) \in \mathbb{R}^2$, $t_k = k\Delta t$, $\Delta t > 0$, $t_1 < \dots < t_d$, y_k удовлетворяет соотношению (2.4).

Вектор-объекты $s^n = ((t_1, y_1), \dots, (t_n, y_n))$ и $s'^n = ((t'_1, y'_1), \dots, (t'_n, y'_n))$, будем называть статистически эквивалентными, если при соотношениях $y_k = x_k + \eta_k + e_k$ и $y'_k = x'_k + \eta'_k + e'_k$ справедливо равенство $x_k = x'_k$, $k = 1, \dots, n$.

Конечное множество $M = \{\mu_1, \dots, \mu_h\}$ символов, будем называть алфавитом α -разметки. Метки (символы) μ_i соответствуют базисным функциям из множества $\Phi = \{f_v(k), v = 1, \dots, h\}$.

Любую последовательность меток $\mu^n = (\mu_1, \dots, \mu_n)$ из заданного алфавита M , назовем α -разметкой длины n .

Введем множество S^n всех различных вектор-объектов длины n , $S^n \subseteq \mathfrak{S}$, $\mathfrak{S} = \bigcup_{n=1}^{\infty} S^n$, множество M^n всех различных α -разметок длины n и множество всех α -разметок $\tilde{M} = \bigcup_{n=1}^{\infty} M^n$.

Алгоритмом α -разметки вектор-объекта s^n будем называть отображение $A^\alpha: S^n \rightarrow M^n$, $\forall n \geq 1: A^\alpha(s^n) = \mu^n$, сопоставляющее каждому вектор-объекту метку из заданного алфавита $M = \{\mu_1, \dots, \mu_h\}$.

Для описания требований к подходящим разметкам, вводится система правил разметки. Правило разметки задается бинарной функ-

цией: $v: \bigcup_{n=1}^{\infty} (S^n \times M^n) \rightarrow \{0,1\}$. Системой правил v будем называть набор функций $v = (v_1, \dots, v_h)$.

Разметка называется подходящей (соответствующий алгоритм A подходящим), если выполняется условие:

$$\sum_{i=1}^h v_i \geq e_\alpha,$$

где v_i - правило разметки из системы правил $v = (v_1, \dots, v_h)$, e_α - некоторое пороговое значение.

Ввиду того, что аналитическое описание функциональных зависимостей из $F_\omega = \{\phi_j\}$, соответствующих j -тому состоянию наблюдаемого процесса неизвестно, экспертом задается их описание $\tilde{\chi}_j = (\mu_1, \dots, \mu_m) \in M^X$, $M^X = \{\tilde{\chi}_j\}$, $M^X \subset \tilde{M}$, $\mu_i \in M$, $j = 1, \dots, J$, $i = 1, \dots, m$, на языке M .

Конечное множество $X = \{\chi_1, \dots, \chi_J\}$ символов, будем называть алфавитом β -разметки. Метки (символы) χ_i соответствуют описаниям $\tilde{\chi}_j = (\mu_1, \dots, \mu_m) \in M^X$ функциональных зависимостей из множества $F_\omega = \{\phi_j\}$.

Любую последовательность меток $\chi^n = (\chi_1, \dots, \chi_n)$ из заданного алфавита X , назовем β -разметкой длины n .

Введем множество X^n всех различных β -разметок длины n и множество всех β -разметок $\tilde{X} = \bigcup_{n=1}^{\infty} X^n$.

Алгоритмом β -разметки будем называть отображение $A^\beta: M^n \rightarrow X^n$, $\forall n \geq 1: A^\beta(\mu^n) = \chi^n$, сопоставляющее α -разметке μ^n каждого вектор-объекта s^n метку из заданного алфавита $X = \{\chi_1, \dots, \chi_J\}$.

Исходя из (2.1), (2.2), и предположения, что регистрируемый процесс в момент времени t обязательно находится в одном из J состояний, каждому вектор-объекту s^n соответствует некоторая β -разметка χ^n .

Каждому вектор-объекту s^n поставим в соответствие его допустимое описание (множество признаков) $I_{ob} = \{p_1, \dots, p_J\}$, $I_{ob} \in \mathfrak{J}_{ob} \subseteq \mathbb{R}^J$, где p_j - частота вхождения метки χ_j из алфавита X в β -разметку χ^n вектор-объекта s^n , \mathfrak{J}_{ob} - множество всех допустимых описаний.

Все допустимые описания I_{ob} вектор-объектов, которые могут быть интерпретированы как предвестник геодинамического события, представим в виде нечеткого множества²:

$$I_{\text{ЭФОР}} = \{(I_{ob}, g(I_{ob})) \mid I_{ob} \in \mathfrak{J}_{ob}\}, \quad (2.6)$$

где $g: \mathfrak{J}_{ob} \rightarrow [0,1]$ - функция принадлежности, характеризующая степень принадлежности допустимого описания I_{ob} вектор-объекта s^n нечеткому множеству $I_{\text{ЭФОР}}$.

Замечание. Обоснованием однозначного сопоставления двух динамических объектов, отвечающих за динамику возникновения микротрещин и за динамику изменения формы ЭМС (2.2), является показан-

² Zadeh L.A. Fuzzy sets. — Information and Control, 1965, vol.8, N 3, pp.338-353.

ная в работах [2,57-59] взаимосвязь изменения формы ЭМС и процесса трещинообразования в горных породах.

Постановка задачи. Пусть изменение сигнала $x(t)$ на этапе формирования очага разрушения описывается моделью (2.2), исходные данные представлены в виде множества \mathfrak{S} стохастических вектор-объектов s^n , а класс предвестников геодинамического события задан нечетким множеством (2.6).

Будем предполагать, что по результатам лабораторных испытаний (одноосное сжатие образцов горных пород), экспертами была сформирована обучающая выборка \mathfrak{L} , элементами которой являются вектор-объекты $s^m = ((t_1, y_1), \dots, (t_m, y_m))$, соответствующие этапу формирования очага разрушения для различных образцов горных пород.

При зафиксированном наборе базисных функций $\Phi = \{f_v(k), v=1, \dots, h\}$ и, задаваемых экспертом описаний $M^X = \{\tilde{\chi}_j\}$ функциональных зависимостей $F_\omega = \{\phi_j\}$, ставятся задачи α , β разметки стохастического временного ряда.

Поскольку, возможно различие α -разметок μ^n , μ'^n и, соответственно, различие β -разметок χ^n , χ'^n , статистически эквивалентных вектор-объектов s^n и s'^n (исходя из (2.1) и (2.4)) при зафиксированном алфавите M , задача построения α , β разметок некорректна (существует несколько решений, удовлетворяющих заданным ограничениям). В свя-

зи с этим, возникает необходимость в построении корректных алгоритмов для решения задачи α , β разметки.

Таким образом, для решения задачи обнаружения предвестников геодинамического события требуется:

- 1) построить корректный алгоритм α , β разметки;
- 2) по обучающей выборке \mathcal{L} построить оценку характеристической функции $g: \mathcal{J}_{об} \rightarrow [0,1]$.

2.2. Модель корректных алгоритмов, алгебраический подход к синтезу корректных алгоритмов

Для решения задачи построения корректных алгоритмов α , β разметки используется теория синтеза эффективных алгоритмов и алгебраический подход к синтезу корректных алгоритмов [46-49], теория универсальных и локальных ограничений (К.В. Рудаков).

Согласно теории синтеза эффективных алгоритмов, структура модели корректного алгоритма, представлена на рисунке 2.1.

С помощью рабочей информации, финальной информации и структурной информации, описывается предметная область моделируемых алгоритмов:

- под рабочей информацией понимается описание исходных данных используемых алгоритмом;
- финальная информация - это описание информации формируемой в результате работы алгоритма;

- структурная информация - это описание предварительно заданных параметров, используемых алгоритмом.

Требования к моделируемым алгоритмам задаются в виде локальных, универсальных ограничений и условия разрешимости и регулярности:

- согласно локальным ограничениям, для заданного набора исходных данных, в результате работы алгоритма, должен быть сформирован набор финальных информации, совпадающий с предварительно заданными (обучающая выборка) значениями;

- универсальные ограничения - это дополнительные ограничения формируемые в зависимости от решаемой моделируемые алгоритмами задачи;

- выполнение условия разрешимости и регулярности гарантирует наличие и единственность решения для алгоритма, удовлетворяющего локальным и универсальным ограничениям.

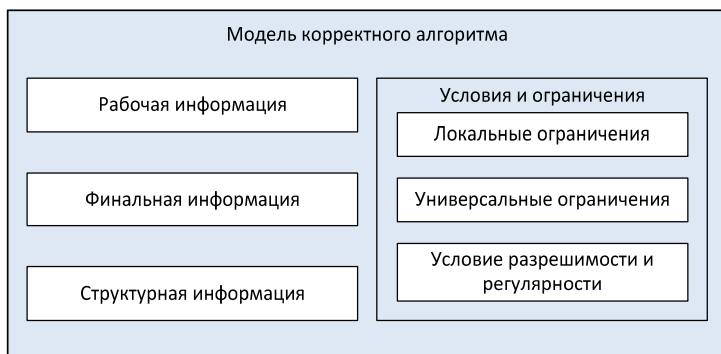


Рис. 2.1. Структура модели корректного алгоритма

Для построения алгоритма из класса заданного с использованием модели корректного алгоритма (рис. 2.1), применяется алгебраический подход к синтезу корректных алгоритмов [50].

Пусть имеется множество начальных информации \mathfrak{I}_i и множество финальных информации \mathfrak{I}_f . Тогда для решения задачи синтеза корректного алгоритма требуется построить алгоритм, реализующий отображение из \mathfrak{I}_i в \mathfrak{I}_f , удовлетворяющий локальным и универсальным ограничениям. В алгебраическом подходе к синтезу корректных алгоритмов, наряду с множествами \mathfrak{I}_i и \mathfrak{I}_f , вводится пространство оценок \mathfrak{I}_e .

Запишем эвристическую информационную модель алгоритмов: $\mathfrak{M} \subseteq \{A|A: \mathfrak{I}_i \rightarrow \mathfrak{I}_f\}$, которая представляет собой семейство суперпозиций:

$$\mathfrak{M} = \mathfrak{M}^1 \circ \mathfrak{M}^0 = \{C \circ B | C \in \mathfrak{M}^1, B \in \mathfrak{M}^0\}, \quad (2.7)$$

где $\mathfrak{M}^0 \subseteq \{B|B: \mathfrak{I}_{ob} \rightarrow \mathfrak{I}_e\}$ - семейство алгоритмических операторов;

$\mathfrak{M}^1 \subseteq \bigcup_{p=0}^{\infty} \{C|C: \mathfrak{I}_e^p \rightarrow \mathfrak{I}_f\}$ - семейство решающих правил.

Введем, семейство корректирующих операций \mathfrak{F} :

$$\mathfrak{F} \subseteq \bigcup_{p=0}^{\infty} \{F|F: \mathfrak{I}_e^p \rightarrow \mathfrak{I}_e\}.$$

Конструкция (2.7) позволяет реализовать идею о совместном использовании нескольких алгоритмов при решении одной задачи путем применения корректирующих операций.

Используя корректирующие операции, запишем расширенную эвристическую информационную модель алгоритмов:

$$\mathfrak{F}(\mathfrak{M}) = \{F(A_1, \dots, A_p) \mid F \in \mathfrak{F}, (A_1, \dots, A_p)\},$$

То есть, основная идея алгебраического подхода состоит в том, чтобы расширить исходную модель алгоритмов с помощью корректирующих операций, для этого строится некоторое количество алгоритмов и корректирующая операция над ними организуется с таким расчетом, чтобы результирующий алгоритм гарантированно удовлетворял локальным ограничениям.

2.3. **Корректирующие операции на основе модификации метода анализа иерархий**

Модификация метода анализа иерархий. Корректная модификация метода анализа иерархий (Analytic Hierarchy Process, АНР, [51]), была предложена в [44]. Данная модификация (далее АНР+) отличается от классического АНР тем, что гарантирует сохранение отношения предпочтения при изменении множества сравниваемых альтернатив и обеспечивает выбор наилучшего решения по всей совокупности критериев на основе максимизации введенной функции относительного сходства.

Обозначим матрицу парных сравнений (МПС) $A^k = \|a_{ij}^k\|_{g \times g}$, $a_{ij}^k = \frac{w_i^k}{w_j^k}$ альтернатив $Z = \{z_1, z_2, \dots, z_g\}$, где w_i^k и w_j^k компоненты вектора

$\mathbf{w} = (w_1^k, w_2^k, \dots, w_g^k)^T$ весов k -го критерия, соответствующие альтернативам z_i и z_j , $k=1, \dots, K$, K - количество критериев выбора альтернативы z_c , $c=1, \dots, g$, $g=|Z|$.

Матрица $A^k = \|a_{ij}\|_{g \times g}$ должна обладать свойствами:

$$a_{ij} > 0, \quad a_{ij} = a_{ji}^{-1}, \quad a_{ij} = a_{im} a_{mj}, \quad A \cdot \mathbf{w} = \lambda \cdot \mathbf{w},$$

где λ - максимальное собственное значение матрицы A , \mathbf{w} - нормированный собственный вектор совместной матрицы A , соответствующий ее максимальному собственному значению [52].

Согласно методу АНР+ корректная функция скаляризации критериев имеет следующий вид:

$$F_{\Sigma}(i) = v_i / \sum_{l=1}^g v_l, \quad v_i = \sum_{l=1}^g w_{il}(i), \quad w_{ij}(i) = \sum_{k=1}^3 \xi_k w_{ij}^k(i), \quad w_{ij}^k(i) = \frac{w_i^k}{w_i^k + w_j^k}, \quad (2.8)$$

где $w_i^k = w_i'^k / \sum_{l=1}^g w_l'^k$, $w_i'^k = \left(\prod_{l=1}^g a_{il}^k \right)^{1/g}$, $i=1, \dots, g$, ξ_k , $\sum_{k=1}^3 \xi_k = 1$ - весовые коэффициенты критериев.

Корректирующие операции. Запишем в общем виде композицию $A(t)$ некоторых базовых алгоритмов разметки $A_c(t)$:

$$A(t) = \sum_{c=1}^g \omega_c(t) A_c(t), \quad \omega_c(t) = \begin{cases} 1, & c = l(t) \\ 0 & c \neq l(t) \end{cases}, \quad l(t) = \arg \max_{c=1, \dots, g} w_c(t), \quad (2.9)$$

где $w_c(t)$ - весовой коэффициент c -го алгоритма в момент времени t ;
 $\omega_c(t)$ - индикаторная функция алгоритма $A_c(t)$, такая что

$$\sum_{c=1}^g \omega_c(t) = 1, \quad \forall t = 1, \dots, n.$$

Весовые критерии $w_c(t)$ определяются, используя метод АНР+. Для этого, в качестве набора альтернатив примактся множество базовых алгоритмов $\mathfrak{A} = \{A_1, \dots, A_g\}, \mathfrak{A} \subseteq \mathfrak{M}$. Каждому базовому алгоритму разметки $A_c(t)$, соответствует K критериев, где K – равно мощности используемого алфавита разметки. По обучающей выборке для каждого алгоритма и каждого критерия определяются весовые коэффициенты $w_c^k, k=1, \dots, K, c=1, \dots, g$. Далее используя формулу(2.8), определяются весовые коэффициенты $w_c(t)$ алгоритмов $A_c(t), c=1, \dots, g$.

2.4. Описание класса алгоритмов, используемых для решения задачи обнаружения параметров предвестника геодинимического события

Используя модель корректного алгоритма, для задачи обнаружения параметров предвестника определим класс корректных алгоритмов α, β разметки.

Так как задача α -разметки стохастического временного ряда является промежуточной для решения задачи (β -разметки) разбиения исходного множества вектор-объектов \mathfrak{E} на J классов (состояний), то приведенные далее ограничения, условия и описание предметной области отражают суть этих двух информационно взаимосвязанных задач.

В соответствии с моделью корректного алгоритма (рис. 2.1) определим для задачи α, β разметки стохастического временного ряда

структурную, рабочую, финальную информации, локальные и универсальные ограничения, условия разрешимости и регулярности.

Рабочая информация (объект исследования): временные ряды - как результаты измерений ЭМС при изменении напряженно-деформированного состояния горного массива или нагружении лабораторных образцов в последовательные моменты времени с определенным шагом дискретизации, представленные в виде множества \mathfrak{S} стохастических вектор-объектов $s^n = ((t_1, y_1), \dots, (t_n, y_n))$, $n \geq 1$, $t_k, y_k \in \mathbb{R}$, $s_j = (t_k, y_k) \in \mathbb{R}^2$, $t_k = k\Delta t$, $\Delta t > 0$, $t_1 < \dots < t_d$, y_k и удовлетворяющих соотношению (2.6).

Финальная информация (множество финальных информаций): совокупность α -разметок стохастических вектор-объектов, пар (s^n, μ^n) , $\mu^n \in M^n$, $\mu^n = (\mu_1, \dots, \mu_n)$, $\mu_i \in M$, совокупность α и β разметок стохастических вектор-объектов, троек (s^n, μ^n, χ^n) , $\chi^n \in X^n$, $\chi^n = (\chi_1, \dots, \chi_n)$, $\chi_i \in X$.

Структурная информация (множество начальных информаций): алфавит α -разметки – конечное множество $M = \{\mu_1, \dots, \mu_h\}$ меток функций из набора $\Phi = \{f_v(k), v = 1, \dots, h\}$; алфавит β -разметки $X = \{\chi_1, \dots, \chi_j\}$ меток описаний $\tilde{\chi}_j = (\mu_1, \dots, \mu_m) \in M^x$ функциональных зависимостей из $F_\omega = \{\phi_j\}$. Контрольная выборка – множество $S_L = \{(s_j^n, \mu_j^n, \chi_j^n)\}$, $s_j^n \in S^n$, $\mu_j^n \in M^n$, $\chi_j^n \in X^n$, $j = 1, \dots, N_{S_L}$.

Локальные ограничения: искомые алгоритмы A^α и A^β должны удовлетворять условиям: $A^\alpha(s_j^n) = \mu_j^n$ и $A^\beta(s_j^n) = \chi_j^n$, где тройка $(s_j^n, \mu_j^n, \chi_j^n) \in S_L$, $j = 1, \dots, N_{S_L}$.

Универсальные ограничения:

1. разметки μ^n и χ^n вектор-объекта s^n должны быть статистически согласованы. Для этого должно выполняться условие:

$$P(\mu_i, \chi_j) = P(\chi_j)P(\mu_i / \chi_j), \quad i = 1, \dots, m, \quad P(\mu_i) = \sum_{j \in 1, \dots, j} P(\chi_j)P(\mu_i / \chi_j), \quad \forall i = 1, \dots, h,$$

где χ_j - метка $\tilde{\chi}_j = (\mu_1, \dots, \mu_m) \in M^X$;

2. синтезируемый алгоритм β -разметки не может присвоить точке СВР с меткой μ_i соответствующую β -метку χ_j , если истинная α -метка μ_i не входит в последовательность $\tilde{\chi}_j = (\mu_1, \dots, \mu_m) \in M^X$, соответствующую β -метке χ_j : $P(\chi_j / \mu_i) = 0$, $\mu_i \notin \tilde{\chi}_j$;

3. вероятность присвоения β -метки χ_j , если истинная α -метка μ_i входит в последовательность $\tilde{\chi}_j = (\mu_1, \dots, \mu_m) \in M^X$ соответствующую β -метке χ_j , отлична от нуля: $P(\chi_j / \mu_i) > 0$, $\mu_i \in \tilde{\chi}_j$;

4. вероятность правильной β -разметки в случае $\tilde{\chi}_j = (\mu_i)$, отлична от нуля: $P(\chi_j / \mu_i) > 0$.

Условие разрешимости и регулярности задачи α, β разметки стохастического временного ряда: исходя из (2.1) и (2.4), возможно различие α -разметок μ^n и μ'^n , полученных в результате работы алгоритма

A^α , для статистически эквивалентных вектор-объектов s^n и s'^n при зафиксированном алфавите M .

Пусть алгоритм α -разметки стохастического временного ряда принимает решение о разметке каждой точки на основе анализа ее окрестности [45].

Окрестностью $o(s_j, u_j)$ точки $s_j = (t_j, y_j)$ вектор-объекта s^n будем называть вектор-объект: $o(s_j, u_j) = ((t_{i_1}, y_{i_1}), \dots, (t_j, y_j), \dots, (t_{i_2}, y_{i_2}))$, $i_1, i_2 \in \{1, \dots, n\}$, $i_1 < i_2$, $j \in \{i_1, \dots, i_2\}$, где величина $u_j = i_2 - i_1 + 1$ состоит из суммы размеров $i_2 - i_1$ окрестности и точки $s_j = (t_j, y_j)$. Точку $s_j = (t_j, y_j)$ будем называть опорной точкой окрестности.

В связи с этим, разрешимость и регулярность задачи α, β разметки стохастического временного ряда будет определяться выбором окрестности $o(s_j, u_j)$ для каждой точки $s_j = (t_j, y_j)$ размечаемого стохастического временного ряда, представленного в виде вектор объекта s^n .

2.5. Задача построения оптимальной системы окрестностей

При разметке вектор-объекта s^n , алгоритм α -разметки A^α ставит в соответствие каждой точки $s_j = (t_j, y_j)$ метку из алфавита M на основании анализа ее окрестности $o(s_j, u_j)$.

Системой окрестностей O^n вектор-объекта s^n будем называть набор окрестностей $o(s_j, u_j), u_j > 0$, соответствующих каждой точки s_j вектор-объекта s^n .

В связи с существованием множества α -разметок для статистически эквивалентных вектор-объектов или окрестностей при фиксированном алфавите M и фиксированной системе правил α для обеспечения разрешимости и регулярности задачи α , β разметки (например, при нулевом размере окрестности, задача разметки неразрешима) ставится задача построения оптимальной системы окрестностей O^* .

Задача построения оптимальной системы окрестностей O^* , заключается в поиске такой системы окрестностей, которая удовлетворяла бы следующим требованиям:

- минимальность неопределенности α -разметки вектор-объекта:

$$H_1(P^\mu, O^*) = -\sum_{l=1}^h p_l^\mu \log_2(p_l^\mu) \rightarrow \min, \quad P^\mu = (p_1^\mu, \dots, p_h^\mu), \quad (2.10)$$

где P^μ - вероятностное распределение меток из алфавита $M = \{\mu_1, \dots, \mu_h\}$ α -разметки;

- минимальная неопределенность β -разметки, согласно допустимой α -разметке стохастического временного ряда:

$$H_2(P^X, O^*) = -\sum_{l \in \{1, \dots, J\}} p_l^X \log_2(p_l^X) \rightarrow \min, \quad P^X = (p_1^X, \dots, p_J^X), \quad (2.11)$$

где P^X - вероятностное распределение меток из алфавита $X = \{\chi_1, \dots, \chi_J\}$ β -разметки;

- отсутствие или минимальное количество ошибок на прецедентах для алгоритмов α , β разметки.

Кроме того, существует ряд рекомендаций [53,54,55] по способу выбора оптимальной системы окрестностей:

1. для авторегрессионных моделей приемлемые результаты по точности дает выборка значений в 3–4 раза большая порядка авторегрессии [39];

2. при $n=2p+1$ реализуется точная «подгонка» по заданным парам (x_j, y_j) в смысле среднеквадратического критерия;

3. характер изменения окна и величины его сдвига зависит от выборочной дисперсии вектор-объекта (для каждого диапазона дисперсий – оптимальная пара значений «размер окна - сдвиг окна»).

Выводы

1. Предложена математическая модель этапа формирования очага разрушения (предвестника). Изменение сигнала $x(t)$ на этапе формирования очага разрушения предлагается представлять в виде кусочно-заданной функции (2.2).

2. Дана формализация задачи обнаружения предвестников геодинамического события. Задача поиска параметров предвестника, рассматривается как задача разметки стохастического временного ряда. В результате определено, что для решения задачи обнаружения предвестников геодинамического события требуется построить корректный алгоритм α , β разметки и по обучающей выборке и выполнить оценку характеристической функции нечеткого множества (2.6).

3. Дано описание класса алгоритмов, предназначенных для решения задачи обнаружения параметров предвестника геодинамического события. В соответствии с моделью корректного алгоритма (рис. 2.1) для задачи α , β разметки стохастического временного ряда была определена: структурная, рабочая и финальная информация; локальные и универсальные ограничения; условие разрешимости и регулярности.

4. В качестве корректирующих операций, при построении корректных алгоритмов α , β разметки предлагается использовать корректирующие операции на основе модификации метода анализа иерархий.

5. Для обеспечения разрешимости и регулярности задачи α , β разметки была поставлена задача построения оптимальной системы окрестностей. Критерии оптимальности системы окрестностей: минимальность неопределенности α -разметки вектор-объекта; минимальная неопределенность β -разметки согласно допустимой α -разметке стохастического временного ряда; отсутствие или минимальное количество ошибок на прецедентах для алгоритмов α , β разметки.

Глава 3. АЛГОРИТМ ОБНАРУЖЕНИЯ ПРЕДВЕСТНИКОВ ГЕОДИНАМИЧЕСКОГО СОБЫТИЯ

В данной главе приведен алгоритм, позволяющий решить задачу обнаружения предвестников геодинамического события, которые описываются математической моделью (2.2). Для решения этой задачи в соответствии с алгебраическим подходом к синтезу корректных алгоритмов (см. п. 2.2) были разработаны корректные алгоритмы α , β разметки и численная процедура построения оптимальной системы окрестностей.

Алгоритм обнаружения предвестников геодинамического события позволяет построить оценку характеристической функции \hat{g} по обучающей выборке $\mathcal{L} = \{L_1, \dots, L_q\}$, где $L_i = (s_{2kHz}^n, s_{15kHz}^n, s_{100kHz}^n, s_{BBkHz}^n)$.

Вектор-объекты $s_{2kHz}^n, s_{15kHz}^n, s_{100kHz}^n, s_{BBkHz}^n$, входящие в обучающую выборку соответствуют участку временных рядов, полученных при лабораторном разрушении образцов горных пород, границы которого определяются экспертом (считаем, что этот участок соответствует третьему этапу процесса трещинообразования, т.е. этапу формирования очага разрушения).

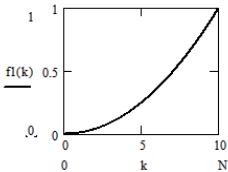
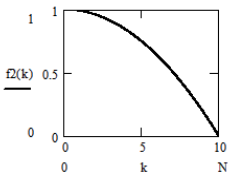
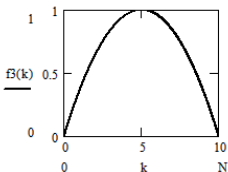
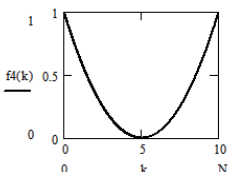
3.1. Корректные алгоритмы α -разметки стохастических временных рядов измерений ЭМС

Зафиксируем домен функций $\Phi = \{f_v(k), v=1, \dots, h\}$ и сформируем алфавит M как совокупность меток функций из множества Φ . В качестве домена выбраны четыре ($h=4$) функции (см. табл. 3.1).

Так как в соответствие каждой функции $f_v(k)$, $v=1, \dots, h$ ставится уникальная метка, то зафиксируем алфавит разметки $M = \{m1, m2, m3, m4\}$.

Для α -разметки стохастических временных рядов измерений ЭМС, в соответствии с алгебраическим подходом (см. п. 2.2), был разработан корректный алгоритм α -разметки стохастических временных рядов SMA (от Symbolic, Marking, Algorithm) основанный на идее представления временного ряда длиной n в виде символьной последовательности произвольной длины w , ($w < n$, как правило $w \ll n$), высказанной и реализованной в виде алгоритма SAX в работах Lin, J., Keogh, E., Lonardi. [60,61]. Близость строки, полученной с использованием алгоритма SAX и функции из $\Phi = \{f_v(k), v=1, \dots, h\}$ заданной в виде символьном виде по заданной метрике, и определяет соответствующую метку [62].

Таблица 3.1 Библиотека функций с аналитическим описанием и метками алфавита разметки, N – максимальное значение параметра нормирования k (по оси Ox)

N	График функции	Аналитический вид функции	Метка
1		$f_1(k) = \left(\frac{k}{N}\right)^2$	$m1$
2		$f_2(k) = 1 - \left(\frac{k}{N}\right)^2$	$m2$
3		$f_3(k) = 1 - \left(\frac{2k}{N} - 1\right)^2$	$m3$
4		$f_4(k) = \left(\frac{2k}{N} - 1\right)^2$	$m4$

Эффективность алгоритма SMA в смысле точности разметки стохастического вектор-объекта показана в сравнении с корректным алгоритмом разметки ARADS (от AutoRegression, Adaptive algorithm, Difference Scheme) [44], основанном на существовании однозначного соответствия между функциями $\Phi = \{f_v(k), v=1, \dots, h\}$ с дробно-рациональным Z -преобразованием соответствующих им последовательностей и значениями векторов коэффициентов разностных схем

(DS-коэффициентов), полученных посредством Z -преобразования. Близость по норме векторов коэффициентов авторегрессионной зависимости (AR-коэффициентов) и DS-коэффициентов определяет тип функции [56].

3.1.1. Корректный алгоритм α -разметки стохастического временного ряда ARADS

На рисунке 3.1 приведена блок-схема алгоритма ARADS. Запишем входные данные алгоритма:

– окрестность $\alpha(s_j, u_j) = ((t_i, y_i), \dots, (t_j, y_j), \dots, (t_{i_2}, y_{i_2}))$ точки $s_j = (t_j, y_j)$ и соответствующий ей участок временного ряда $y_i, \dots, y_j, \dots, y_{i_2}$;

– алфавит разметки $M = \{m1, m2, m3, m4\}$, составленный из меток функций $\Phi = \{f_v(k), v=1, \dots, h\}$. Обучающая выборка представляет набор пар $S_{LS} = \{(\mu_1, Y_1), \dots, (\mu_h, Y_h)\}$, где $\mu_v \in M$, $Y_v = \{f_v(t_1), f_v(t_2), \dots, f_v(t_n)\}$ - ряд дискретных во времени значений функции $f_v(k)$, $n = \max_v n_v$, $n_v = 2p_v + 1$, p_v - минимальный допустимый размер окрестности f_v , который должен быть больше порядка авторегрессии в 3-4 раза [39].

В результате работы алгоритма получаем метку $\mu \in M$ точки $s_j = (t_j, y_j)$ при фиксированной окрестности $\alpha(s_j, u_j) = ((t_i, y_i), \dots, (t_j, y_j), \dots, (t_{i_2}, y_{i_2}))$.

На первом шаге функциям $\Phi = \{f, (k), v = 1, \dots, h\}$ сопоставляется набор DS-коэффициентов $\mathbf{a}^v = (a_1^v, \dots, a_p^v)$, где $p = \max_v p_v$ и $a_j^v = 0, \forall j = p_v + 1, \dots, p$. Применяя эвристический прием, AR-коэффициенты, найденные для окрестности $\sigma(s_j, u_j)$, сравниваются с модельными AR-коэффициентами \mathbf{a}^A , вместо DS-коэффициентов. Поэтому для каждого ряда Y_v рассчитывается (блоки 3-5, рис. 3.3) вектор $\mathbf{a}^A = (a_1^A, \dots, a_p^A)$ (в качестве начальных условий используются DS-коэффициенты \mathbf{a}^v).

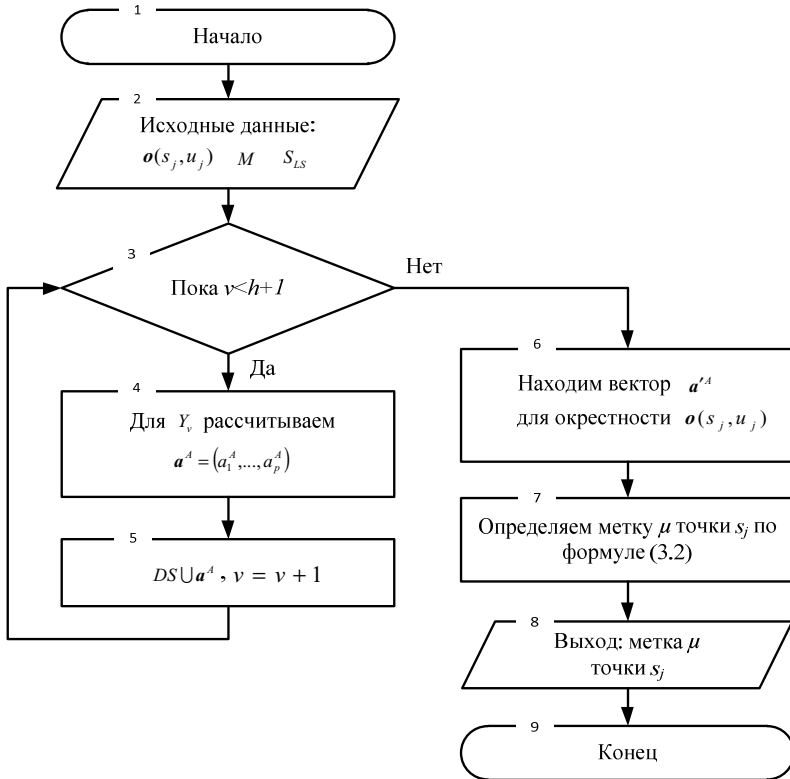


Рис. 3.1. Блок-схема алгоритма ARADS

Далее находим векторы \mathbf{a}'^A AR-коэффициентов, соответствующие окрестности $\mathcal{O}(s_j, u_j)$:

$$\mathbf{a}'^A = (a'_{1^A}, \dots, a'_{p_v^A}) = \arg \min_{a'_o \in R} \sum_{l \in \{i_1, \dots, j_1, \dots, i_2\}} \phi^* \left(\left| y_l - \sum_{o=1}^{p_v} a'_o y_{l-o} \right| \right), \quad (3.1)$$

где ϕ^* - монотонно возрастающая на $(0, \infty)$ функция потерь, которая обладает свойствами: $\phi^*(0) = 0$, $\ddot{\phi}^*(x) > 0$ для любого $x > 0$, например, можно использовать следующие функции: $\phi^*(x) = \frac{|x|}{|x|+1}$, $\phi^*(x) = \arctg(|x|)$, $\phi^*(x) = 1 - \exp(-|x|)$.

В итоге определяется метка μ функции (метка точки $s_j = (t_j, y_j)$), используя одну из заданных метрик ρ :

$$\mu = \arg \min_{v=1, \dots, J} \rho(\mathbf{a}^A, \mathbf{a}'^A), \quad (3.2)$$

Запишем для алгоритма ARADS систему правил разметки $v^1 = \{v_v^1\}$. Аксиома $v_v^1(\mathcal{O}(s_j, u_j), \mu, \mathbf{a}'^A)$, это бинарная функция, задаваемая в одной точке $s_j = (t_j, y_j)$ с окрестностью $\mathcal{O}(s_j, u_j) = ((t_{i_1}, y_{i_1}), \dots, (t_j, y_j), \dots, (t_{i_2}, y_{i_2}))$ по правилу: точка $s_j = (t_j, y_j)$ является точкой с меткой μ , если имеют место (3.1), (3.2) при фиксированных функции ϕ^* и метрике ρ .

Точность распознавания функций из множества Φ алгоритмом ARADS при фиксированной зависимости ϕ от вида ρ , приведена на рисунке 3.2.

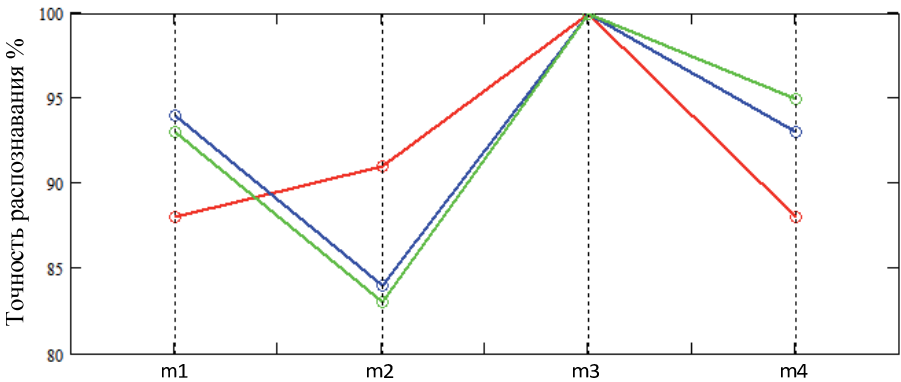


Рис. 3.2. Графики точности классификации по каждой метке из $M = \{m1, m2, m3, m4\}$ для трех метрик: Чебышевское расстояние (красный цвет); линейная метрика (зеленый цвет); Евклидова метрика (синий цвет)

Для построения графика изображенного на рисунке 3.2 использовалась выборка объемом $N_c = 100$, состоящая из модельных временных рядов (полученных путем дискретизации функций из $\Phi = \{f_v(k), v = 1, \dots, h\}$) длиной в 21 отсчет с наложенным гауссовским шумом $N(0, 2)$.

Наличие на рисунке 3.2 участков «хуже», «лучше», «соизмеримо» для алгоритмов ARADS из одного и того же класса алгоритмов $A^a(\Phi, \phi, \rho)$ являются основанием для построения композиций алгоритмов $F(A_1^a, \dots, A_r^a)$ где F - корректирующая операция. Используем корректирующие операции на основе модификации метода анализа иерархий (см. п. 2.3).

Весовые коэффициенты алгоритмов из формулы (2.8), рассчитывались по формуле (2.9), используя весовые коэффициенты $w_c^k, k = 1, \dots, 4$

(так как $|M|=4$), $c=1, \dots, 3$ (при использовании трех видов метрик), выбираемых из числа ошибочных классификаций k -й метки c -м алгоритмом (с использованием c -ой метрики) на контрольной выборке:

$$w_c^k = N_C - \frac{1}{4}(w_{k,c}^{N(0,1)} + w_{k,c}^{N(0,2)} + w_{k,c}^{N(0,3)} + w_{k,c}^{N(0,5)}),$$

где N_C - объем выборки; $w_{k,c}^{N(0,1)}$, $w_{k,c}^{N(0,2)}$, $w_{k,c}^{N(0,3)}$, $w_{k,c}^{N(0,5)}$ - количество ошибочных классификаций k -й метки для рядов зашумленных гауссовским шумом с соответствующими параметрами.

При определении весов использовалась контрольная выборка, состоящая из временных рядов длиной в 21 отсчет (т.к. длина окрестности $u_j = i_2 - i_1 + 1$), состоящих из дискретизированных значений функции $\Phi = \{f_v(k), v=1, \dots, h\}$ с наложенным гауссовским шумом, параметры которого определяются: $N(0, 1)$, $N(0, 2)$, $N(0, 3)$, $N(0, 5)$ с объемом выборки $N_C = 100$.

3.1.2. Корректный алгоритм α -разметки стохастического временного ряда SMA

Для представления временного ряда в виде символьной последовательности (далее строки) используется алгоритм SAX (Symbolic Aggregate appRoXimation) [61].

Алгоритм SAX позволяет представить n - размерный временной ряд строкой, длина которой определяется w , ($w < n$, как правило, $w \ll n$). Строка состоит из символов некоторого алфавита мощностью $\alpha > 2$.

Алгоритм отличается от прочих [63], своей процедурой дискретизации, в которой используется промежуточное представление временного ряда. Данные сначала преобразуются в некоторое промежуточное представление, используя кусочное приближение (далее РАА, от Piecewise Aggregate Approximation) [64], а затем уже в символьную последовательность.

С помощью РАА временной ряд длиной n можно представить в w -мерном пространстве вектором $C = (c_1, \dots, c_w)$, где i -й элемент вектора c вычисляется по следующей формуле:

$$c_i = \frac{w}{n} \sum_{j=n^{w(i-1)+1}}^{n^w} y_j, \quad (3.3)$$

где $n^w = \left\lfloor \frac{n}{w} \right\rfloor$, т.е. $\frac{n}{w}$ округляется к меньшему целому числу.

Временной ряд предварительно нормируется, а затем разделяется на w равных по размеру участков, для каждого участка вычисляется среднее значение. Из множества полученных средних значений формируется вектор c .

Процедура квантования заключается в выборе уровней λ_i в зависимости от размера алфавита символов, являющимися точками, через которые проходят параллельные прямые разделяющие пространство под кривой Гаусса [65] на участки равной площади (рис. 3.3). Уровни квантования λ_i для алфавита размером от 3 до 8 приведены в табл.3.2.

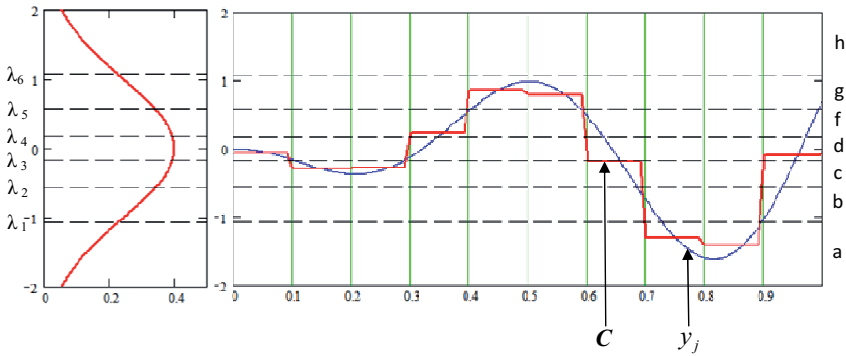


Рис. 3.3. Иллюстрация процедуры дискретизации алгоритма SAX

Таблица 3.2 Уровни квантования для алфавита символов размером от 3 до 8

Уровень квантования	Размер алфавита α					
	3	4	5	6	7	8
λ_1	-0,43	-0,67	-0,84	-0,97	-1,07	-1,15
λ_2	0,43	0	-0,25	-0,43	-0,57	-0,67
λ_3		0,67	0,25	0	-0,18	-0,32
λ_4			0,84	0,43	0,18	0
λ_5				0,97	0,57	0,32
λ_6					1,07	0,67
λ_7						1,15

Полученную в результате дискретизации строку обозначим $\tilde{q} = (\tilde{q}_1, \dots, \tilde{q}_w)$, где w – длина строки.

Используя алгоритм SAX, формируются эталонные строки (табл. 3.3), как символьное описание функций из $\Phi = \{f_v(k), v=1, \dots, h\}$, предварительно заданной длины w из символов алфавита α .

Таблица 3.3 Эталонные строки для алфавита размером $\alpha=4$ и различных значений длины строки w

Метка	w=4	w=6	w=8	w=10
$m1$	abcd	aabccdd	aabbccddd	aaabbccddd
$m2$	dcba	ddcbaa	ddccbbaa	dddccbbaaa
$m3$	adda	abddba	abcdccbba	abcddddcbba
$m4$	daad	dcaacd	dcbaabcd	dcbaaaabcd

Блок-схема алгоритма SMA приведена на рисунке 3.4. Входные данные алгоритма:

- окрестность $\sigma(s_j, u_j) = ((t_i, y_i), \dots, (t_j, y_j), \dots, (t_{i_2}, y_{i_2}))$ точки $s_j = (t_j, y_j)$ и соответствующий ей участок временного ряда $y_i, \dots, y_j, \dots, y_{i_2}$;
- алфавит разметки $M = \{m1, m2, m3, m4\}$, составленный из меток функций $\Phi = \{f_v(k), v=1, \dots, h\}$;
- размер алфавита символов α и количество символов в строке w .

В результате работы алгоритма получаем метку $\mu \in M$ точки $s_j = (t_j, y_j)$.

В цикле (блок 3-4), используя алгоритм SAX, формируются набор $\tilde{M} = \{\tilde{m}_v, v=1, \dots, h\}$ эталонных строк, как символьное описание функций из множества $\Phi = \{f_v(k), v=1, \dots, h\}$.

После нахождения символьного описания \tilde{c} окрестности $\sigma(s_j, u_j) = ((t_i, y_i), \dots, (t_j, y_j), \dots, (t_{i_2}, y_{i_2}))$ точки $s_j = (t_j, y_j)$ вычисляется расстояние до каждой эталонной строки \tilde{m}_v . Далее, определяется метка μ по формуле:

$$\mu = \arg \min_{v=1, \dots, h} \rho(\tilde{m}_v, \tilde{c}). \quad (3.4)$$

Для вычисления расстояния $\rho(\tilde{m}, \tilde{c})$ между символьными описаниями временных рядов используются аналоги линейной, Евклидовой метрик и Чебышевского расстояния, определенные с использованием функции (введена в [61]) расстояния между символами алфавита α :

$$d(m, c) = \begin{cases} 0, & \text{если } |m - c| \leq 1 \\ \lambda_{\max(m, c) - 1} - \lambda_{\min(m, c)}, & \text{в противном случае} \end{cases}$$

где, m, c - номера символов в используемом алфавите α ; λ_j - определяется по таблице 3.2.

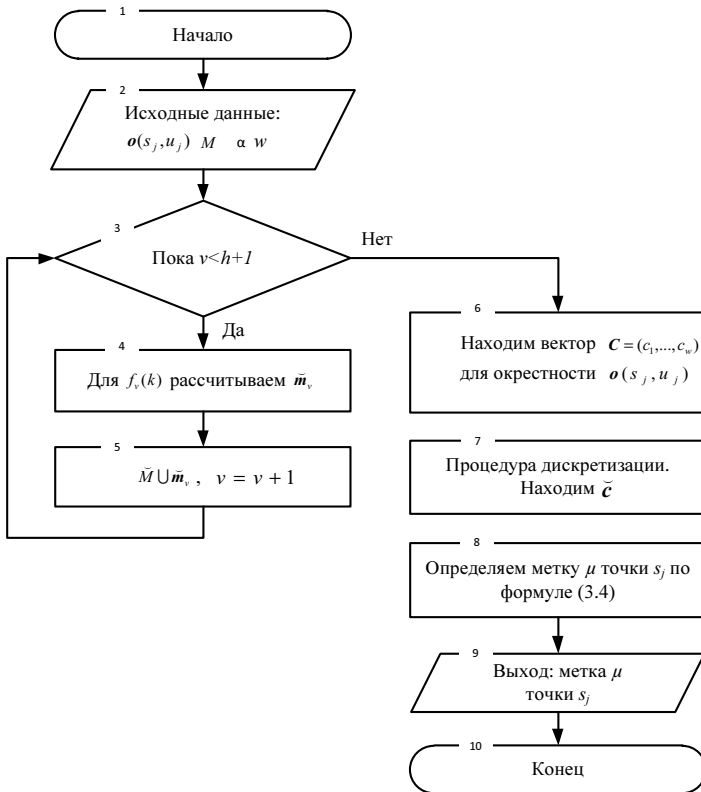


Рис. 3.4. Блок-схема алгоритма SMA

Для примера запишем аналог Евклидовой метрики:

$$\rho(\tilde{\mathbf{m}}, \tilde{\mathbf{c}}) = \sqrt{\frac{n}{w} \sum_{i=1}^w d(|\tilde{m}_i|, |\tilde{c}_i|)}, \quad (3.5)$$

где, $|\tilde{m}_i|, |\tilde{c}_i|$ - номера символов \tilde{m}_i и \tilde{c}_i в используемом алфавите.

Запишем для алгоритма SMA систему правил разметки $\mathbf{v}^2 = \{v_v^2\}$.

Аксиома $v_v^2(\mathbf{o}(s_j, u_j), \mu, \tilde{\mathbf{c}})$, это бинарная функция, задаваемая в одной точке $s_j = (t_j, y_j)$ с окрестностью $\mathbf{o}(s_j, u_j) = ((t_{i_1}, y_{i_1}), \dots, (t_j, y_j), \dots, (t_{i_2}, y_{i_2}))$ по правилу: точка $s_j = (t_j, y_j)$ является точкой с меткой μ , если $\mu = \arg \min_{v=1, \dots, h} \rho(\tilde{\mathbf{m}}_v, \tilde{\mathbf{c}})$.

Для класса алгоритмов α -разметки SMA $A^\alpha(\Phi, \alpha, w, \rho)$, точность распознавания функций из Φ при фиксированных α и w , в зависимости от вида ρ , представлена на рисунке 3.5.

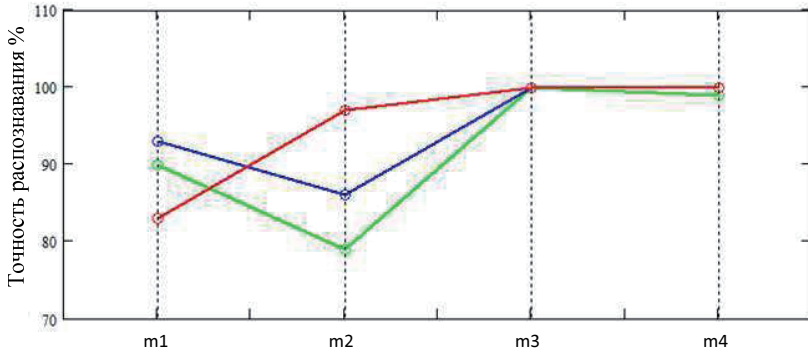


Рис. 3.5. Графики точности классификации по каждой метке из $M = \{m1, m2, m3, m4\}$ и трех метрик (аналогов): Чебышевское расстояние (красный цвет); линейная метрика (зеленый цвет); Евклидова метрика (синий цвет)

Для построения графика, показанного на рисунке 3.5, использовалась та же выборка, что и в п. 3.1.1.

В виду отсутствия в классе алгоритмов $A^\alpha(\Phi, \alpha, w, \rho)$ такого алгоритма, который обеспечивал бы наилучшую точность распознавания каждой функции из множества Φ , возникает необходимость в построении алгоритмической композиции $F(A_1^\alpha, \dots, A_r^\alpha)$, где F - корректирующая операция. По аналогии с алгоритмом ARADS (см. п. 3.1.1) используются корректирующие операции на основе модификации метода анализа иерархий (см. п. 2.3). Аналогично п.3.1.1 и по той же контрольной выборке, определяются весовые коэффициенты алгоритмов.

3.1.3. Сравнительный анализ алгоритмов α -разметки

Для тестирования алгоритмов были сформированы две контрольные выборки S_C^1 и S_C^2 . При формировании первой выборки S_C^1 , использовался временной ряд измерений ЭМС на частоте 100 кГц при лабораторном сжатии образца горной породы №16 (приложение А). График изменения ЭМС при одноосном сжатии образца представлен на рисунке 3.6, j -ой точке ряда ставилась в соответствие окрестность $\theta(s_j, u_i)$ размером $u_1 = 4, u_2 = 8, u_3 = 12, u_4 = 16, u_5 = 20$, где $j = 1, \dots, N_C$, объем выборки $N_C = 100$. Для каждой окрестности $\theta(s_j, u_i)$, $i = 1, \dots, 5$ точки s_j эксперт определил метку из алфавита $M = \{m1, m2, m3, m4\}$. Пример разметки временного ряда алгоритмом SMA при размере окрестности $u = 20$ приведен на рисунке 3.7.

Контрольная выборка S_C^2 состоит из временных рядов длиной в $u_1 = 4$, $u_2 = 8$, $u_3 = 12$, $u_4 = 16$, $u_5 = 20$ отсчет, состоящих из дискретных значений функции из множества $\Phi = \{f_v(k), v=1, \dots, h\}$ с наложенным гауссовским шумом, параметры которого: $N(0, 0.01)$, $N(0, 0.02)$, $N(0, 0.03)$, $N(0, 0.05)$, $N(0, 0.1)$, $N(0, 0.2)$, объем выборки $N_C = 100$.

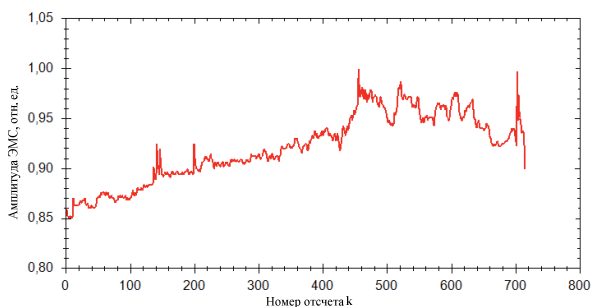


Рис. 3.6. Временной ряд лабораторных измерений ЭМС на частоте 100 кГц для образца №16

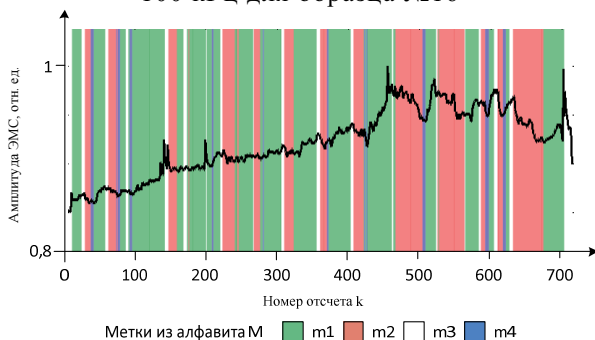


Рис. 3.7. Пример разметки временного ряда лабораторных измерений ЭМС на частоте 100 кГц для образца №16, алгоритмом SMA при $u = 20$

Качество разметки соответствующее разным метрикам, определенное по контрольной выборке S_C^1 (сформированной по реальным данным) для алгоритмов ARADS и SMA, представлено в таблицах 3.4 и 3.5.

Из таблицы 3.4 видно, что применение корректирующей операции для алгоритма ARADS дает выигрыш 12 – 23 %, а для алгоритма SMA 1-7 %. Из таблиц 3.4 и 3.5 можно сделать заключение, что наименьшее количество ошибочных разметок для всех $u_i, i=1, \dots, 5$, было получено при использовании корректного алгоритма SMA (при использовании корректирующей операции).

Таблица 3.4 Сравнение качества разметки алгоритмом ARADS выборки S_C^1 по трем метрикам и композиции алгоритмов

Используемая метрика ρ	Количество ошибочных разметок (при $n=100$)				
	$u_1 = 4$	$u_2 = 8$	$u_3 = 12$	$u_4 = 16$	$u_5 = 20$
Евклидова метрика	27	39	40	37	33
Линейная метрика	28	41	41	40	34
Чебышевское расстояние	39	42	44	40	39
Композиция алгоритмов	25	22	21	20	21

Таблица 3.5 Сравнение качества разметки алгоритмом SMA выборки S_C^1 по трем метрикам и композиции алгоритмов

Используемая метрика ρ	Количество ошибочных разметок (при $n=100$)				
	$u_1 = 4$	$u_2 = 8$	$u_3 = 12$	$u_4 = 16$	$u_5 = 20$
Евклидова метрика	19	26	30	29	33
Линейная метрика	18	28	36	35	45
Чебышевское расстояние	18	30	33	31	37
Композиция алгоритмов	17	20	24	24	26

Контрольная выборка S_C^2 , полученная путем зашумления эталонных временных рядов, используется для оценки чувствительности алгоритмов разметки к шуму.

В таблицах 3.6 и 3.7 переведено количество ошибочных разметок в зависимости от параметров зашумления и размера $u_i, i=1, \dots, 5$, для алгоритмов ARADS и SMA по контрольной выборке S_c^2 .

Таблица 3.6 Количество ошибочных разметок в зависимости от параметров зашумления и размера u_i , для корректного алгоритма α -разметки ARADS

Композиция алгоритмов					
Параметры зашумления	Количество ошибочных разметок (при $n=100$)				
	$u_1 = 4$	$u_2 = 8$	$u_3 = 12$	$u_4 = 16$	$u_5 = 20$
N(0, 0.01)	11	2	1	0	0
N(0, 0.02)	11	6	2	1	1
N(0, 0.03)	13	6	4	3	2
N(0, 0.05)	22	13	10	11	8
N(0, 0.1)	30	21	19	19	15
N(0, 0.2)	39	29	25	25	22

При зашумлении гауссовским шумом с параметрами N(0, 1), для получения безошибочной разметки при использовании алгоритма ARADS, необходимо увеличить размер окрестности до $u_4=17$ (табл. 3.6), тогда как при использовании алгоритма SMA ряд размечается безошибочно (табл. 3.7) при всех значениях $u_i, i=1, \dots, 5$.

Таблица 3.7 Количество ошибочных разметок в зависимости от параметров зашумления и размера u_i , для корректного алгоритма α -разметки SMA

Композиция алгоритмов					
Параметры зашумления	Количество ошибочных разметок (при $n=100$)				
	$u_1 = 4$	$u_2 = 8$	$u_3 = 12$	$u_4 = 16$	$u_5 = 20$
N(0, 0.01)	0	0	0	0	0
N(0, 0.02)	1	0	0	0	0
N(0, 0.03)	5	0	0	0	0
N(0, 0.05)	9	2	0	0	0
N(0, 0.1)	16	5	1	0	0
N(0, 0.2)	35	18	4	3	1

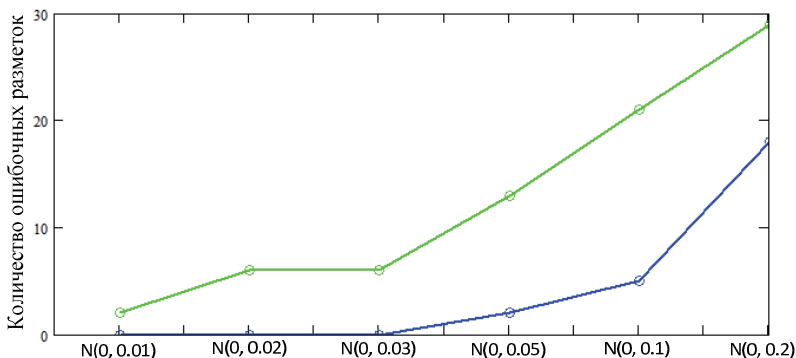


Рис. 3.8. Зависимость количества ошибочных разметок от параметров зашумления (при $u_2 = 8$), для корректных алгоритмов α -разметки: ARADS – зеленым цветом, SMA – синим цветом

По данным из таблиц 3.6 и 3.7 и рисунку 3.8 видно, что алгоритм ARADS более чувствителен к шуму, то есть обладает меньшей обобщающей способностью [66], чем алгоритм SMA.

3.2. Корректный алгоритм β -разметки

Экспертами в данной отрасли знаний по результатам лабораторных исследований разрушения образцов горной породы были выявлены характерные закономерности или признаки (параметры предвестника), которые указывают на предвестник геодинамического события (этап формирования очага разрушения) [2, 57-59]: частое скачкообразное изменение средней амплитуды ЭМС и учащение импульсов ЭМС.

Для их обнаружения запишем эти параметры предвестника на языке $M = \{m_1, m_2, m_3, m_4\}$ и сформируем множество $M^z = \{\tilde{\chi}_j\}$. Множество

M^X задает алфавит β -разметки $X = \{\chi_1, \dots, \chi_J\}$, где метки χ_j соответствуют описаниям параметров предвестника $\tilde{\chi}_j$ (табл. 3.11).

Входными данными алгоритма β -разметки, являются:

- α -разметка μ^m , где $m = \max_{j=1, \dots, J} |\tilde{\chi}_j|$, $\tilde{\chi}_j \in M^X$;
- алфавит β -разметки $X = \{\chi_1, \dots, \chi_J\}$, составленный из меток описаний параметров предвестника $M^X = \{\tilde{\chi}_j\}$;

В результате работы алгоритма получаем метку $\chi \in X$ соответствующую α -разметке μ^m .

Метка χ определяется по формуле:

$$\chi = \arg \min_{j=1, \dots, J} \rho(\mu^m, \tilde{\chi}_j). \quad (3.6)$$

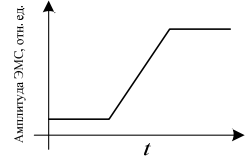
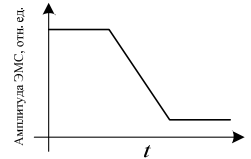
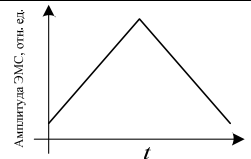
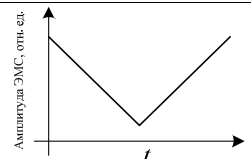
Расстояние между эталонным описанием параметра предвестника $\tilde{\chi}_j$ и α -разметкой μ^m , вычисляется на основе расстояния Дамерау-Левенштейна [67], метрики Хемминга или суммы разностей частот вхождения всех n -грамм [68] (n -граммой η^n называется подразметка длиной n) $\eta^n \in N_n(\tilde{\chi}_j)$ эталона и частот n -грамм α -разметки μ^m , при $n=1, \dots, g-1$ (в случае, если α -разметку μ^m и эталон $\tilde{\chi}_j$ интерпретировать как строки символов из алфавита M):

$$d^N(\mu^m, \tilde{\chi}_j) = \frac{1}{m-1} \sum_{n=1}^{m-1} \frac{1}{h_n} \sum_{\eta^n \in N_n^U} |p(\eta^n, \mu^m) - p(\eta^n, \tilde{\chi}_j)|,$$

где m - длина разметки μ^m и эталона $\tilde{\chi}_j$; $h_n = 2(g-n)+2$ - коэффициент нормирования; $N_n^U = N_n(\tilde{\chi}_j) \cup N_n(\mu^m)$, $N_n(\tilde{\chi}_j), N_n(\mu^m)$ - множества всех n -

грамм эталона $\tilde{\chi}_j$ и α -разметки μ^m соответственно; $p(\eta^n, \mu^m)$ - частота вхождения n -граммы η^n в разметку μ^m . Расстояние d^N (далее, метод N-грамм) определено на интервале $[0, 1]$.

Таблица 3.11 Описания параметров предвестника $\tilde{\chi}_j$ и соответствующие им метки из алфавита X

№, k		$\tilde{\chi}_j$	Метка
1		m4m1m1m1m3	x1
2		m3m2m2m2m4	x2
3		m1m1m3m2m2	x3
4		m2m2m4m1m1	x4

Точность распознавания описаний параметров предвестника $M^X = \{\tilde{\chi}_j\}$ алгоритмом β -разметки, зависимости от вида ρ приведена на рисунке 3.9.

Для построения графика, изображенного на рисунке 3.9, использовалась выборка объемом $N_C = 100$, состоящая из разметок μ^m длиной

$m = 5$. Каждая разметка μ^m формировалась, используя корректный алгоритм α -разметки SMA, по набору из пяти модельных временных рядов (получены путем дискретизации функций из $\Phi = \{f_v(k), v=1, \dots, h\}$) длиной в 21 отсчет, с наложенным гауссовским шумом $N(0, 0.02)$.

Так как ни один из трех способов вычисления расстояния $\rho(\mu^m, \tilde{\chi}_j)$ не обеспечивает максимума точности распознавания каждого эталонного описания параметра предвестника $\tilde{\chi}_j \in M^x$, то возникает необходимость в построении алгоритмической композиции $F(A_1^\beta, \dots, A_r^\beta)$, где F - корректирующая операция. По аналогии с алгоритмом ARADS, SMA (см. п. 3.1.1) используются корректирующие операции на основе модификации метода анализа иерархий (см. п. 2.3).

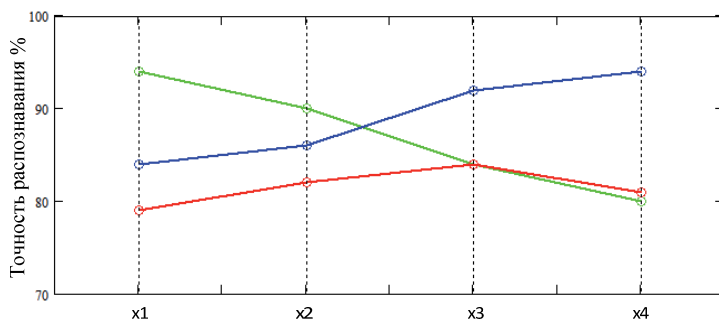


Рис. 3.9. Графики точности классификации по каждой метке из $X = \{x1, x2, x3, x4\}$ (табл. 3.11) и для трех способов вычисления расстояния $\rho(\mu^m, \tilde{\chi}_j)$: метрика Хемминга – **красным цветом**; метод N-грамм – **зеленым цветом**; расстояние Дамерау-Левенштейна – **синим цветом**

Для определения весовых коэффициентов использовалась контрольная выборка объемом $N_C = 100$, состоящая из разметок μ^m , длиной

$m = 5$. Для определения разметок μ^m , по комбинации модельных рядов, зашумленных гауссовым шумом с параметрами: $N(0, 0.01)$, $N(0, 0.02)$, $N(0, 0.03)$, $N(0, 0.05)$.

3.3. Численная процедура построения оптимальной системы окрестностей

В главе 2. п. 2.4 дано определение системы окрестностей \mathcal{O}^n и приведена общая постановка задачи поиска оптимальной системы окрестности.

Обозначим \mathcal{O}^n множество всех возможных систем окрестностей $\mathcal{O}^n = \{\mathcal{O}_1^n(s_1, u_c), \dots, \mathcal{O}_n^n(s_n, u_c)\}$ с размером $u^{\min} < u_c < u^{\max}$, заданных для вектор-объекта s^n , где u^{\min} и u^{\max} выбираются исходя из рекомендаций, приведенных в [39, 53, 54, 55].

Определим критерии оптимальности исходя из требований, выдвигаемых к оптимальной системе окрестностей:

1) минимальная неопределенность разметки μ^n вектор-объекта s^n , т.е. минимум энтропийного критерия (2.10): $H_1(P^\mu, \mathcal{O}^n) \rightarrow \min$, где $P^\mu = (p_1^\mu, \dots, p_n^\mu)$ - вероятностное распределение меток из алфавита $M = \{\mu_1, \dots, \mu_h\}$ α -разметки;

2) точность разметки, в смысле расстояния $\rho(\gamma_v, \gamma'_i)$, где γ_v, γ'_i - признаки эталонных временных рядов полученных дискретизацией значений функций из $\Phi = \{f_v(k), v=1, \dots, h\}$ и признаки временного ряда соот-

ветствующего окрестности $\mathbf{o}_i^n(s_i, u_c), \mathbf{o}_i^n \in \mathbf{O}^n$, определяемые используемым алгоритмом α -разметки A^α для $i=1, \dots, n$;

3) оптимальная система окрестностей \mathbf{O}^n должна иметь минимально возможный размер u_c .

Предварительный выбор систем окрестностей из множества \mathfrak{D}^n осуществляется исходя из минимума неопределенности β -разметки (2.11):

$$H_2(\hat{P}^x, \mathbf{O}^n) \rightarrow \min, \hat{P}^x = (\hat{p}_1^x, \dots, \hat{p}_j^x),$$

где P^x - оценка вероятностного распределения меток из алфавита $X = \{\chi_1, \dots, \chi_j\}$ β -разметки.

Рассмотрим задачу многокритериального выбора оптимальной системы окрестностей. Задача задается множеством допустимых решений $\mathfrak{X} \subseteq \mathfrak{D}^n$ и частными критериями $f_1(X_c), f_1 \in Y_1, f_2(X_c), f_2 \in Y_2, f_3(X_c), f_3 \in Y_3$, где $X_c = \mathbf{O}_c^n, \mathbf{O}_c^n \in \mathfrak{D}^n, X_c \in \mathfrak{X}, c=1, \dots, |\mathfrak{X}|$ образующими векторный критерий $f = (f_1, f_2, f_3)$, который определен на множестве \mathfrak{X} непустым множеством векторов оценок $Y = f(\mathfrak{X}) = Y_1 \times Y_2 \times Y_3 \in R^3$, бинарным отношением доминирования $<>$ заданным на пространстве R^3 .

Считается что, альтернатива X_c предпочтительнее ($X_c \succ X_b$) альтернативы X_b по отношению к векторному критерию $f = (f_1, f_2, f_3)$, если $\forall j \in [1,3], f_j(X_c) \leq f_j(X_b)$, и $\exists j \in [1,3], f_j(X_c) < f_j(X_b)$.

Альтернатива X_c , т.е. система окрестностей O_c^n , называется парето-оптимальной (оптимальной по Парето) [69] по отношению к векторному критерию $f = (f_1, f_2, f_3)$, если не существует такой альтернативы X_b (окрестности O_b^n), что $X_b \succ X_c$. Все парето-оптимальные альтернативы образуют множество Парето $P(X)$.

Процедуру построения оптимальной системы окрестностей можно условно разделить на три этапа: предварительный выбор систем окрестностей из множества Ω^n (этап предобработки); расчет критериев для каждой выбранной системы (этап расчета критериев); формирование множества Парето $P(X)$ и выбор оптимальной системы окрестностей (этап принятия решения). Рассмотрим каждый этап процедуры более подробно.

Этап предобработки. Блок-схема этапа предобработки процедуры построения оптимальной системы окрестностей представлена на рисунке 3.10.

Входными параметрами этапа и всей процедуры в целом (рис. 3.10, блок 2) являются: s^n - исходный вектор-объект, для которого необходимо найти оптимальную систему окрестностей O_s^n , u^{\min} и u^{\max} - минимальный и максимальный размер окрестности, ε^X - пороговое значение энтропии.

В цикле (рис. 3.10, блок 3-6) по $u_c = u^{\min}, \dots, u^{\max}$ для вектор-объекта s^n и системы окрестностей O_c^n размером u_c , используя алгоритм A^α ,

находим α -разметку μ^n и далее, на основе алгоритма A^β определяется β -разметка χ^n . В свою очередь, для β -разметки χ^n вычисляется оценка вероятностного распределения $\hat{P}^{\chi} = (\hat{p}_1^{\chi}, \dots, \hat{p}_j^{\chi})$.

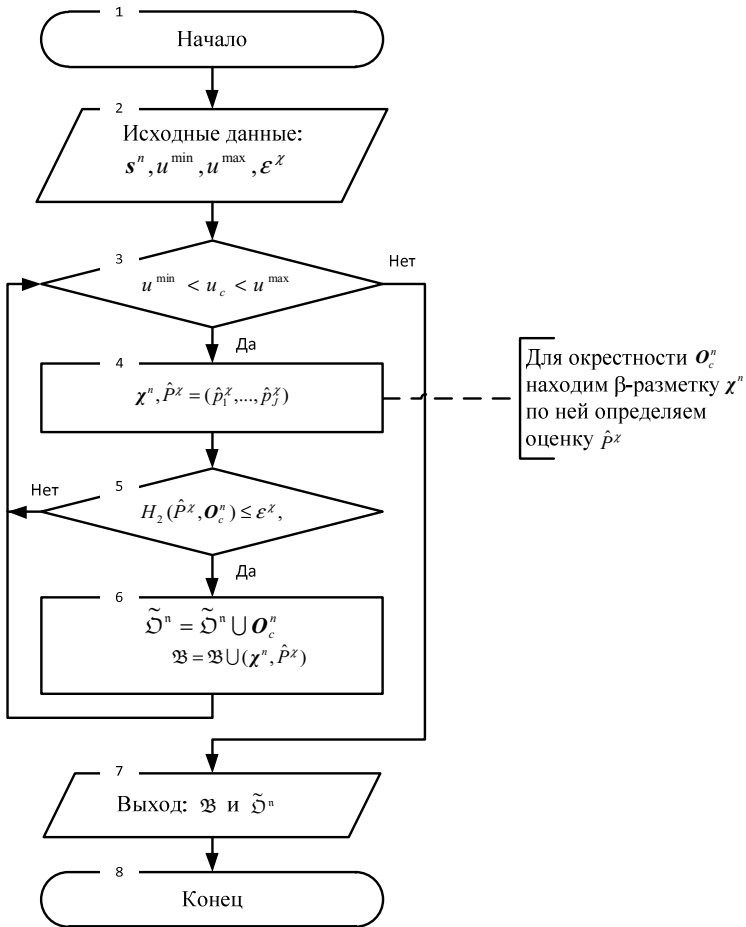


Рис. 3.10. Блок-схема этапа предобработки процедуры построения оптимальной системы окрестностей

Система окрестностей O_c^n размером u_c и пара (χ^n, \hat{P}^{χ}) добавляется во множество $\tilde{\mathfrak{S}}^n$ и \mathfrak{Z} соответственно, при условии:

$$H_2(\hat{P}^x, O_c^n) \leq \varepsilon^x, \quad (3.5)$$

где $\hat{P}^x = (\hat{p}_1^x, \dots, \hat{p}_j^x)$ - вектор оценок вероятностей меток из алфавита $X = \{\chi_1, \dots, \chi_j\}$, определяемый по β -разметке χ^n для системы окрестностей O_c^n . В результате на выходе получаем два множества $\tilde{\Omega}^n$ и \mathfrak{B} .

Этап расчета критериев. Для расчета критерия №1 рассмотрим задачу нахождения оценки вероятностного распределения $\hat{P}^\mu = (\hat{p}_1^\mu, \dots, \hat{p}_h^\mu)$ для всех меток из алфавита $M = \{\mu_1, \dots, \mu_h\}$ по распределению $P^x = (p_1^x, \dots, p_j^x)$ при наличии меток из алфавита $X = \{\chi_1, \dots, \chi_j\}$ с условием, что для вектор-объекта s^n существует истинная разметка μ^n с вероятностным распределением $P^\mu = (p_1^\mu, \dots, p_h^\mu)$.

На этапе преобработки определяется оценка вероятностного распределения $\hat{P}^x = (\hat{p}_1^x, \dots, \hat{p}_j^x)$ (используя формулу полной вероятности), которая записывается в виде:

$$p_j^x = P(\tilde{\chi}_j) = \sum_{i=1}^m p_i^\mu P(\tilde{\chi}_j | \mu_i), \tilde{\chi}_j \in M^x, \mu_i \in \tilde{\chi}_j, m = |\tilde{\chi}_j|,$$

где p_i^μ из последовательности вероятностного распределения $P^\mu = (p_1^\mu, \dots, p_h^\mu)$ символов разметки μ^n .

Запишем оценку максимального правдоподобия β -разметки χ^n , для этого будем считать, что появление меток χ_j из X в ней статистически независимо:

$$P_n(P^\mu, P^\chi) = \prod_{j=1}^J \left(\sum_{i=1}^m p_i^\mu P(\tilde{\chi}_j | \mu_i) \right)^{n_j}, \quad (3.6)$$

где n_j - число появления метки χ_j в β -разметке χ^n .

Параметры распределения $P^\mu = (p_1^\mu, \dots, p_h^\mu)$ определяются из условия максимума $P_n(P^\mu, P^\chi)$ или, что эквивалентно минимизации величины $-\log_2 P_n(P^\mu, P^\chi)$ (логарифмическая функция правдоподобия):

$$-\log_2 P_n(P^\mu, P^\chi) = -n \sum_{j=1}^J \frac{n_j}{n} \log_2 \sum_{i=1}^m p_i^\mu P(\tilde{\chi}_j | \mu_i) = -n \sum_{j=1}^J \hat{p}_j^\chi \log_2 \sum_{i=1}^m p_i^\mu P(\tilde{\chi}_j | \mu_i),$$

где $\hat{p}_j^\chi = \frac{n_j}{n}$. Обозначим $H(P^\mu, P^\chi) = -\frac{\log_2 P_n(P^\mu, P^\chi)}{n}$.

Как показано в [44], если для каждого $l=1, \dots, J$ имеет место соотношение:

$$\sum_{j=1}^J \sum_{i=1}^m \hat{p}_j^\chi P(\tilde{\chi}_j | \mu_i) = \sum_{i=1}^m p_i^\mu P(\tilde{\chi}_l | \mu_i), \quad (3.7)$$

и не все компоненты $P^\mu = (p_1^\mu, \dots, p_h^\mu)$ одновременно равны нулю, то существует единственная α -разметка, набор вероятностей которой P^μ является единственным набором, максимизирующим вероятность β -разметки $P_n(P^\mu, P^\chi)$.

Под точностью разметки (критерий №2) будем понимать величину дисперсии ошибок разметки:

$$D_c^\varepsilon | O_c^n = \frac{1}{n} \sum_{i=1}^n (\varepsilon_i^\mu - \bar{\varepsilon}^\mu)^2, \quad (3.8)$$

где $\bar{\varepsilon}^\mu = \frac{1}{n} \sum_{j=1}^n \bar{\varepsilon}_j^\mu$, $\varepsilon_i^\mu = \min_{v=1,\dots,h} \rho(\gamma_v, \gamma'_i)$, γ_v, γ'_i - признаки эталонных временных рядов полученных дискретизацией значений функций из $\Phi = \{f_v(k), v=1,\dots,h\}$ и признаки окрестности $\mathbf{o}_i^n(s_i, u_c), \mathbf{o}_i^n \in \mathbf{O}_c^n, \mathbf{O}_c^n \in \tilde{\mathfrak{D}}^n$, соответственно, определяемые используемым алгоритмом α -разметки A^α .

Блок схема этапа расчета коэффициентов процедуры построения оптимальной системы окрестностей представлена на рисунке 3.11.

Исходными данными для второго этапа (рис. 3.11, блок 2) будет множество систем окрестностей $\tilde{\mathfrak{D}}^n$ и множество \mathfrak{X} пар (χ^n, \hat{p}^χ) , сформированные на первом этапе процедуры, а так же точность ε с которой вычисляются оценки $\hat{P}_c^\mu | \mathbf{O}_c^n = (\hat{p}_1^\mu, \dots, \hat{p}_h^\mu)$, $\mathbf{O}_c^n \in \tilde{\mathfrak{D}}^n$.

Так как уравнение (3.6) аналитически неразрешимо, оценки $\hat{P}^\mu = (\hat{p}_1^\mu, \dots, \hat{p}_h^\mu)$ вектора вероятностей $P^\mu = (p_1^\mu, \dots, p_h^\mu)$ для каждой системы окрестностей из $\tilde{\mathfrak{D}}^n$ определяются итеративно с заданной точностью ε (используются результаты, полученные в [70]).

На первом шаге (рис. 3.11, блок 3) вычисляется начальное приближение вероятностного распределения $\hat{P}_c^\mu = (\hat{p}_1^\mu, \dots, \hat{p}_h^\mu)$ разметки μ^n для каждой системы окрестностей $\mathbf{O}_c^n \in \tilde{\mathfrak{D}}^n$, где n - размер вектор-объекта s^n .

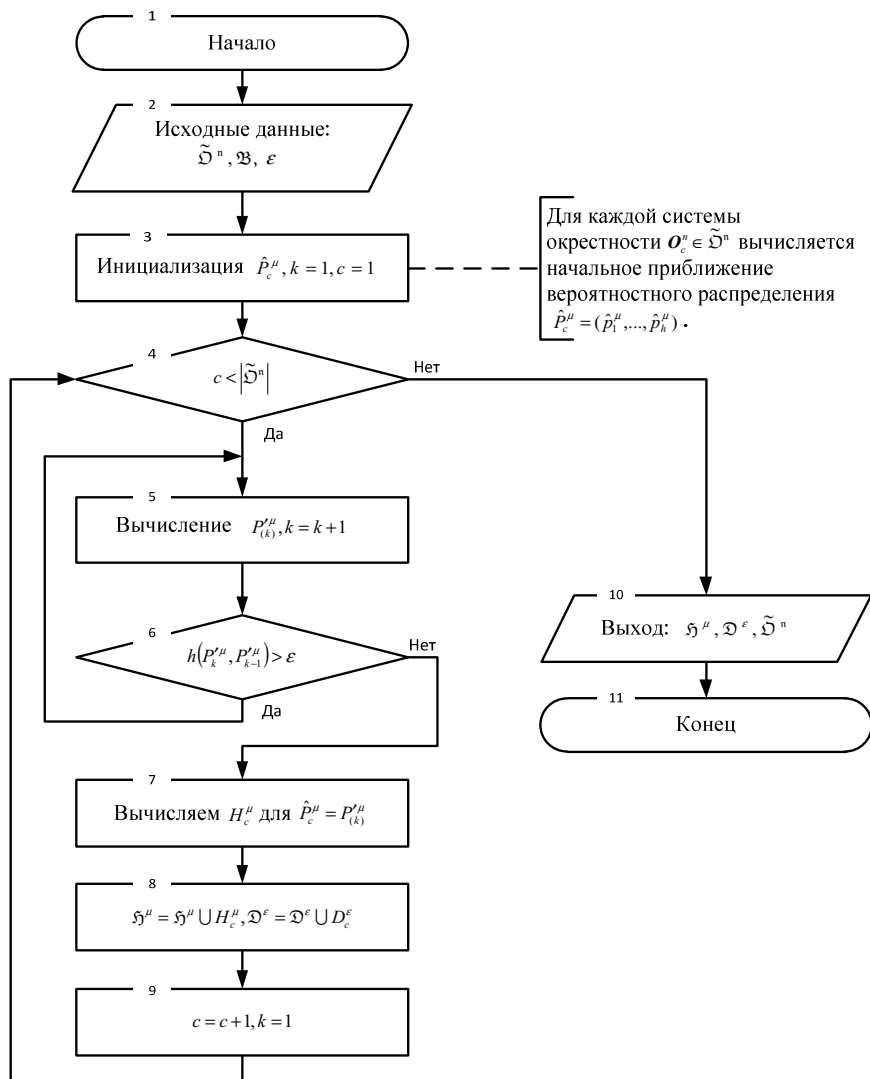


Рис. 3.11. Блок-схема этапа расчета критериев

После этого, в цикле (рис. 3.11, блок 4-8) для каждого элемента множества \tilde{D}^n и соответствующей пары из множества \mathfrak{B} , итеративно вычисляются значения $P_{(k)}^\mu$, до тех пор, пока не выполнится неравенство

$h(P_{(k)}^{\mu}, P_{(k-1)}^{\mu}) > \varepsilon$, где $h(P_{(k)}^{\mu}, P_{(k-1)}^{\mu}) = |H(P_{(k)}^{\mu}, \hat{P}_{\chi}^{\mu}) - H(P_{(k-1)}^{\mu}, \hat{P}_{\chi}^{\mu})|$, k - номер итерации, значения P^{μ} определяются с учетом преобразования [44, 71]: $P^{\mu} = \varphi(\hat{P}^{\mu}) = (p_1^{\mu}, \dots, p_h^{\mu})$.

Используя значения $P_{(k-1)}^{\mu} = (p_{1,(k-1)}^{\mu}, \dots, p_{h,(k-1)}^{\mu})$, вычисленные на $k-1$ шаге, определяются значения вектора $P_{(k)}^{\mu} = (p_{1,(k)}^{\mu}, \dots, p_{h,(k)}^{\mu})$ на k -том шаге, по формуле:

$$p_{i,(k)}^{\mu} | \mathcal{O}_c^n = \frac{\sum_{j=1}^J \hat{p}_j^{\chi} p_{i,(k)}^{\mu} P(\tilde{\chi}_j^m | \mu_i)}{\sum_{i=1}^m p_i^{\mu} P(\tilde{\chi}_j^m | \mu_i)}, \mathcal{O}_c^n \in \tilde{\mathfrak{D}}^n.$$

После того как $h(P_{(k)}^{\mu}, P_{(k-1)}^{\mu})$ становится меньше ε , полагается $P_{(k)}^{\mu} = \hat{P}_c^{\mu}$ и рассчитывается показатель неопределенности разметки μ^n , т.е. информационная энтропия:

$$H^{\mu}(\hat{P}_c^{\mu}) | \mathcal{O}_c^n = - \sum_{i=1}^h \hat{p}_i^{\mu} \log_2(\hat{p}_i^{\mu}) \rightarrow \min, \mathcal{O}_c^n \in \tilde{\mathfrak{D}}^n;$$

Энтропия $H^{\mu}(\hat{P}_c^{\mu}) = H_c^{\mu}$ включается во множество \mathfrak{H}^{μ} , далее рассчитывается точность разметки D_c^{ε} по формуле (3.8) для системы окрестностей $\mathcal{O}_c^n \in \tilde{\mathfrak{D}}^n$, которая включается во множество $\mathfrak{D}^{\varepsilon}$.

В результате на втором этапе процедуры для множества систем окрестностей $\tilde{\mathfrak{D}}^n$ формируются множества критериев \mathfrak{H}^{μ} и $\mathfrak{D}^{\varepsilon}$ (критерий №1 и №2).

Этап принятия решения. Для формирования парето-оптимального множества систем окрестностей $\Pi(\mathfrak{X})$, определим крите-

при $f_1(X_c) = \{H_c^\mu | X_c\}$, $H_c^\mu \in \mathfrak{H}^\mu$ и $f_2(X_c) = \{D_c^\varepsilon | X_c\}$, $D_c^\varepsilon \in \mathfrak{D}^\varepsilon$, $f_3(X_c) = \{u_c | X_c\}$, при условии что, $X_c = \mathbf{O}_c^n$, $X_c \in \mathfrak{X}$, $\mathfrak{X} = \tilde{\mathfrak{D}}^n$.

Для выбора оптимальной системы окрестностей из множества $\Pi(\mathfrak{X})$, применяется модификация метода анализа иерархий АНР+ (см. п. 2.3).

Определим матрицу парных сравнений в виде $A^k = \|a_{ij}^k\|_{g \times g}$, где $a_{ij}^k = \frac{w_i^k}{w_j^k}$ находится для альтернатив $\Pi(\mathfrak{X}) = \{X_1, X_2, \dots, X_g\}$, w_i^k и w_j^k - компоненты вектора $\mathbf{w} = (w_1^k, w_2^k, \dots, w_g^k)^T$, $w_c^k = f_k(X_c)$, $k = 1, \dots, 3$, $c = 1, \dots, g$, $g = |\Pi(\mathfrak{X})|$.

Для каждой системы окрестностей из множества $\Pi(\mathfrak{X})$ обобщенный критерий F_{Σ} рассчитывается по формуле (2.8).

Блок схема этапа принятия решений процедуры построения оптимальной системы окрестностей представлена на рисунке 3.12.

Входными параметрами этапа (рис. 3.12, блок 2) являются: $\tilde{\mathfrak{D}}^n$ - множества систем окрестностей; \mathfrak{H}^μ , \mathfrak{D}^ε - множества критериев полученных на втором этапе процедуры; $\Xi = \{\xi_1, \xi_2, \xi_3\}$ - весовые коэффициенты критериев.

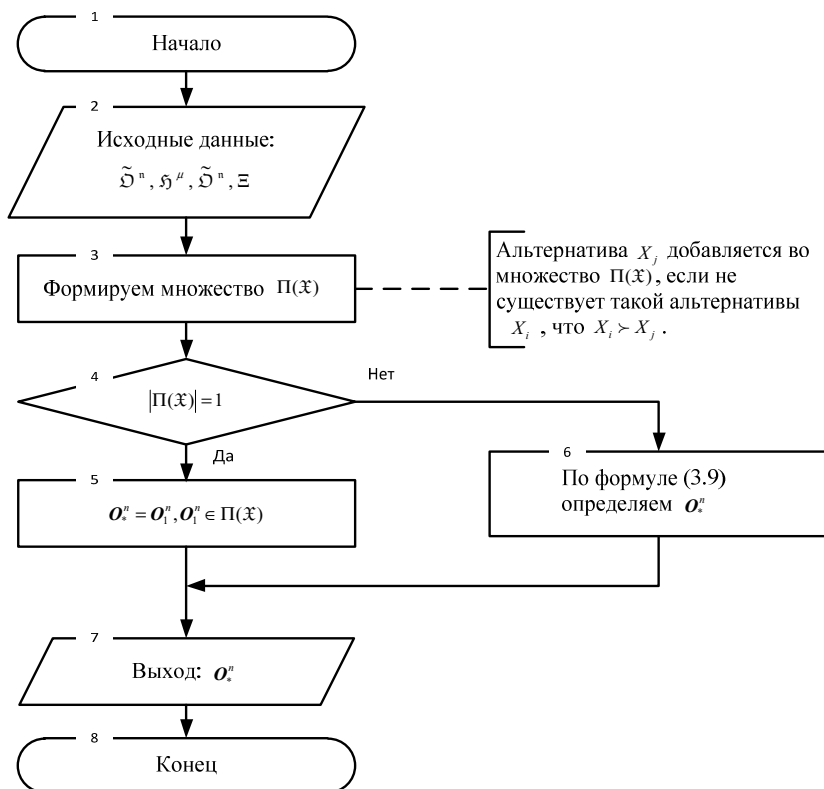


Рис. 3.12. Блок-схема этапа принятия решений

Формируется множество (рис. 3.12, блок 3) парето-оптимальных систем окрестностей $\Pi(\tilde{x})$ и после анализа мощности множества $\Pi(\tilde{x})$ (рис. 3.12, блок 4) оптимальная система окрестностей $\mathcal{O}^n \in \Pi(\tilde{x})$ считается найденной (рис. 3.12, блок 5) или из множества $\Pi(\tilde{x})$ выбирается лучшая система окрестностей \mathcal{O}^n (рис. 3.12, блок 6) в смысле минимума обобщенного критерия F_Σ :

$$\mathcal{O}^n = \mathcal{O}_c^n, \mathcal{O}_c^n \in \Pi(\tilde{x}), c = \arg \min_{i=1, \dots, g} F_\Sigma(i), \quad (3.9)$$

где F_{Σ} - обобщенный критерий, который рассчитывается на основе множества весовых коэффициентов критериев Ξ по формуле (2.8).

Оценка времени сходимости процедуры построения оптимальной системы окрестностей. Для оценки скорости сходимости процедуры расчета вероятностного распределения $\hat{P}^{\mu} = (\hat{p}_1^{\mu}, \dots, \hat{p}_h^{\mu})$ по распределению $P^{\chi} = (p_1^{\chi}, \dots, p_j^{\chi})$ в результате численного моделирования была получена оценка времени работы t (по 20 независимым запускам, $t = \lceil \bar{t} \rceil$, $\bar{t} = \sum_{j=1}^{20} t_j / 20$, где t_j - время работы на j -том запуске), в миллисекундах и итерациях, для различных значений точности ε (табл. 3.8). Расчет проводился при фиксированном размере алфавита α -разметки $|M|=4$ и фиксированном размере алфавита β -разметки X , мощность которого равна $|X|=4$.

Таблица 3.8 Количество итераций и время работы процедуры в зависимости от различных значений точности ε

№	Точность, ε	Количество итераций	Время в миллисекундах
1	0,1	2	4
2	0,05	3	4
3	0,01	11	20
4	0,005	14	29
5	0,001	23	50
6	0.0005	27	59
7	0.0001	35	70
8	0.00001	63	132

Согласно методике предложенной в [72], характерное время релаксации T_R и время выхода на стационарное значение T_S определяется

из зависимости $\varepsilon(t)$ (рис 3.13) точности ε расчета оценки $\hat{P}^\mu = (\hat{p}_1^\mu, \dots, \hat{p}_h^\mu)$, от времени работы процедуры t (количества итераций). Зависимость $\varepsilon(t)$ получена аппроксимацией значений из таблицы 3.8, экспоненциальной функцией.

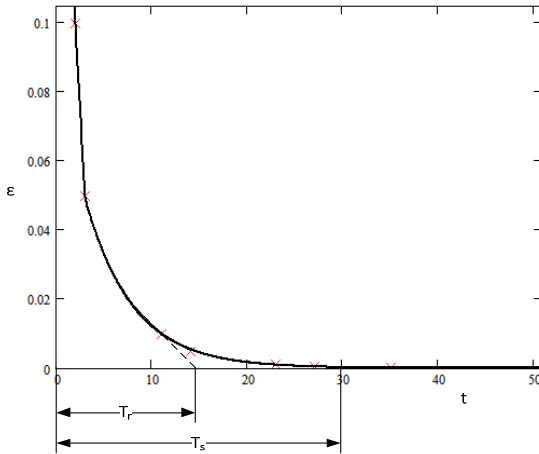


Рис. 3.13. Зависимость $\varepsilon(t)$

Время релаксации T_R равно значению t' , которое отвечает точке пересечения прямой, угол наклона которой соответствует начальному наклону кривой $\varepsilon(t)$ и оси Ox .

Для данных представленных в таблице 3.4, $T_R \approx 14$ итераций, $T_S \approx 30$ итераций, т.е. характерное время сходимости процедуры равно $T \approx T_S \approx 30$ итерациям.

Оценка эффективности процедуры построения оптимальной системы окрестностей. Покажем эффективность процедуры построения оптимальной системы окрестностей, для этого зафиксируем эта-

лонный временной ряд и соответствующий ему вектор-объект s_i^n , в данном случае временной ряд лабораторных измерений ЭМС на частоте 100 кГц, полученный при разрушении образца №16 (п. 3.1.3, рис. 3.6).

Будем использовать алфавит α -разметки $M = \{m1, m2, m3, m4\}$ (см. табл. 3.1) и алфавит β -разметки $X = \{x1, x2, x3, x4\}$ с соответствующими описаниями параметров предвестника (см. табл. 3.11) $M^Z = \{\tilde{\chi}_j\}, j=1, \dots, 4$, $\tilde{\chi}_1 = (m2, m3, m1, m1)$, $\tilde{\chi}_2 = (m1, m1, m3, m2)$, $\tilde{\chi}_3 = (m1, m3, m3, m2)$, $\tilde{\chi}_4 = (m2, m4, m4, m1)$. Зададим минимальный и максимальный размер системы окрестностей $u_{\min} = 4$, $u_{\max} = 32$ и шаг их перебора $u_{step} = 4$, точность $\varepsilon = 0.01$ вычисления оценок $\hat{P}^\mu = (\hat{p}_1^\mu, \dots, \hat{p}_h^\mu)$, множество весовых коэффициентов критериев $\Xi = \{0.4, 0.5, 0.1\}$. В качестве алгоритма A^α будем использовать корректный алгоритм α -разметки SMA.

Для оценки эффективности процедуры построения оптимальной системы окрестности была сформирована выборка $\mathcal{S}_N = \{s_i^n\}$, путем зашумления эталонного временного ряда гауссовским шумом с параметрами $N(0, 0.001)$, $N(0, 0.005)$, $N(0, 0.01)$, $N(0, 0.05)$, $N(0, 0.1)$, $N(0, 0.2)$, $N(0, 0.3)$.

Для эталонного вектор-объекта (стохастического временного ряда) s_i^n и каждого вектор-объекта из выборки $\mathcal{S}_N = \{s_i^n\}$, с помощью численной процедуры построения оптимальной системы окрестностей, была определена оптимальная система окрестностей O_i^n , $O_i'^n$, $i=1, \dots, 7$, соот-

ответственно (множества Парето, полученные для эталонного вектор-объекта и 4-х первых вектор-объектов выборки приведены в табл. 3.9, 3.10).

Таблица 3.9 Множество $\Pi(x)$ парето-оптимальных систем окрестностей для эталонного вектор-объекта S_1 (оптимальная система окрестностей выделена серым цветом)

№	F_{Σ}	u	H^{μ}	D^{ϵ}
1	0,3461	4	1,5283	0,2573
2	0,3262	8	1,5575	0,1745
3	0,3277	12	1,5628	0,1641

Таблица 3.10 Множество $\Pi(x)$ парето-оптимальных систем окрестностей для зашумленных вектор-объектов S_i (оптимальная система окрестностей выделена серым цветом)

N(0,0.1)				
№	F_{Σ}	u	H^{μ}	D^{ϵ}
1	0,3408	4	0,4742	0,2738
2	0,3297	8	0,5948	0,1741
3	0,3293	12	0,6194	0,1552
N(0,0.5)				
1	0,3414516	4	0,4746887	0,3085934
2	0,3430594	8	0,5939235	0,2291687
3	0,3154889	12	0,6027253	0,1487736
N(0,1)				
1	0,2579508	4	0,4744502	0,3638493
2	0,2564969	8	0,587065	0,2609968
3	0,2379696	12	0,610581	0,1720572
4	0,2475827	16	0,7893965	0,1547488
N(0,5)				
1	0,1654993	16	0,5724539	0,2994559
2	0,158999	20	0,5888996	0,2364285
3	0,1606447	24	0,7796125	0,187399
4	0,1715296	32	0,7596385	0,231921

Используя полученные оптимальные системы окрестностей O_i^n , O_i^n , $i=1,\dots,7$, с помощью алгоритмов A^{α} , A^{β} были получены α и β размет-

ки μ_i^n , χ_i^n (для эталонного вектор-объекта) и μ_i^n , χ_i^n , $i=1,\dots,7$ (для вектор-объектов из выборки).

Введем функцию, с помощью которой будем определять различие между двумя β -разметками χ_a^n и χ_b^n :

$$Q^x(\chi_a^n, \chi_b^n) = \frac{d_H(\chi_a^n, \chi_b^n)}{n} \cdot 100\%,$$

где n – длина β -разметок χ_a^n и χ_b^n , $d_H(\chi_a^n, \chi_b^n)$ – расстояние Хэмминга.

Тогда относительным числом ошибок β -разметки χ_i^n , полученной для вектор-объекта s_i^n из S_N , используя систему окрестностей O^n размером u , будем называть величину:

$$Q_i^N(u) = Q^x(\chi_i^n, \chi_i^n).$$

Зависимости относительного числа ошибок $Q_i^N(u)$, от размера $u=4,\dots,32$ (с шагом 4), для 4-х зашумленных вектор-объектов $s_i^n \in S_N$, $i=1,\dots,4$ представлены на рисунке 3.14, по оси Oy значения функции $Q_i^N(u)$, по оси Ox размер u системы окрестностей O_i .

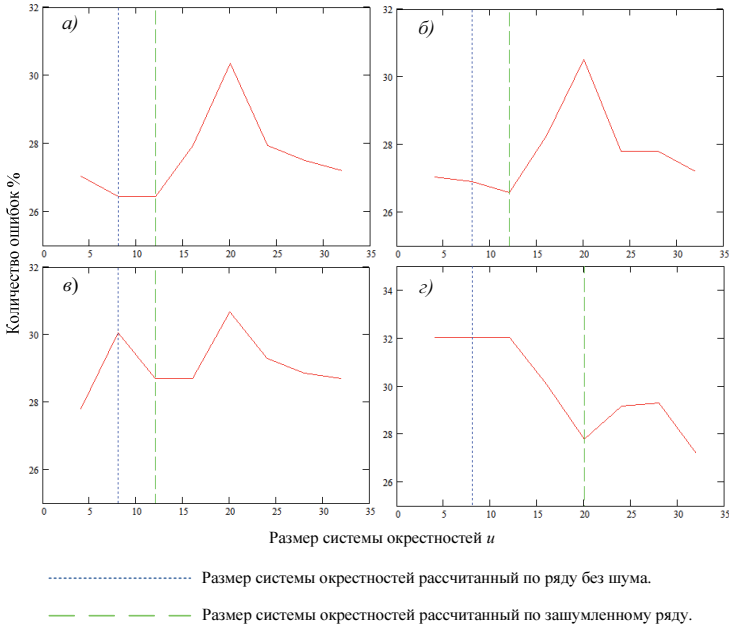


Рис. 3.14. Зависимости $Q_i^N(u)$ для: а) вектор-объекта s_1^n , шум $N(0, 0.001)$; б) вектор-объекта s_2^n , шум $N(0, 0.005)$; в) вектор-объекта s_3^n , шум $N(0, 0.01)$; г) вектор-объекта s_4^n , шум $N(0, 0.05)$

Из графиков зависимостей $Q_i^N(u)$ (рис. 3.14) видно, что размеру оптимальной системы окрестностей определенной для зашумленного вектор-объекта $s_i^n \in \mathcal{S}_N$, соответствует локальный минимум относительного числа ошибок.

Оценим эффективность численной процедуры построения оптимальной системы окрестностей, в зависимости от уровня шума. Под эффективностью будем понимать разницу:

$$\delta_i = Q_i^N(u_1) - Q_i^N(u_i),$$

где u_i - размер оптимальной системы окрестностей, рассчитанный для эталонного ряда (без шума); u_i - размер оптимальной системы окрестностей, рассчитанный для вектор-объекта $s_i^n \in \mathcal{S}_N$ (с шумом).

То есть, δ_i - это выигрыш в качестве распознавания описаний параметров предвестника $M^X = \{\tilde{\chi}_1, \tilde{\chi}_2, \tilde{\chi}_3, \tilde{\chi}_4\}$ по α -разметке μ_i^n вектор-объекта s_i^n , полученной с использованием оптимальной системы окрестностей O_i^n размером u_i , вместо O_i^n размером u_i . График зависимости δ_i от уровня зашумления представлен на рисунке 3.15.

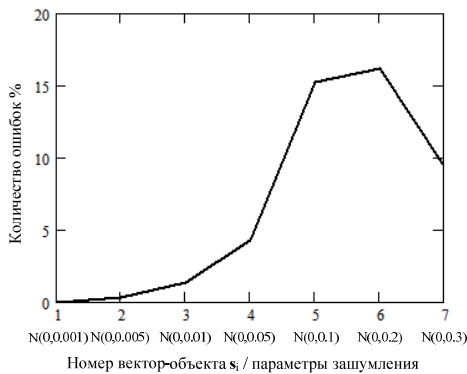


Рис. 3.15. Эффективность δ_i процедуры построения оптимальной системы окрестностей в зависимости от уровня, накладываемого на вектор-объект s_i^n шума

Из графика (рис. 3.15) видно, что использование численной процедуры построения оптимальной системы окрестностей позволяет повысить качество распознавания описаний параметров предвестника $M^X = \{\tilde{\chi}_1, \tilde{\chi}_2, \tilde{\chi}_3, \tilde{\chi}_4\}$ (в смысле близости β -разметки зашумленного вектор-объекта s_i^n к β -разметке не зашумленного вектор-объекта s_i^n) до 15 %.

3.4. Алгоритм обнаружения предвестников геодинамического события

Алгоритм обнаружения предвестников можно разделить на два этапа:

- этап обучения, на котором по сформированной экспертами обучающей выборке $\mathcal{L} = \{L_1, \dots, L_q\}$ для каждого вектор-объекта из набора $L_i = (s_{2kHz}^n, s_{15kHz}^n, s_{100kHz}^n, s_{BBkHz}^n)$ с помощью алгоритмов α - β разметки определяются интервальные оценки частоты вхождения параметров предвестника, заданные описаниями M^x ;

- этап обнаружения, на котором, используя рассчитанные на этапе обучения интервальные оценки частоты вхождения параметров предвестника, алгоритмы α - β разметки и численную процедуру построения оптимальной системы окрестностей для каждого вектор-объекта из набора $(s_{2kHz}^n, s_{15kHz}^n, s_{100kHz}^n, s_{BBkHz}^n)$ оценивается значение функции $g(t)$, характеризующая степень принадлежности участка временного ряда к классу предвестников, заданному нечетким множеством (2.6).

Далее приведено подробное описание каждого этапа алгоритма обнаружения предвестников геодинамического события и приведены экспериментальные результаты.

Этап обучения. Блок-схема этапа обучения представлена на рисунке 3.16.

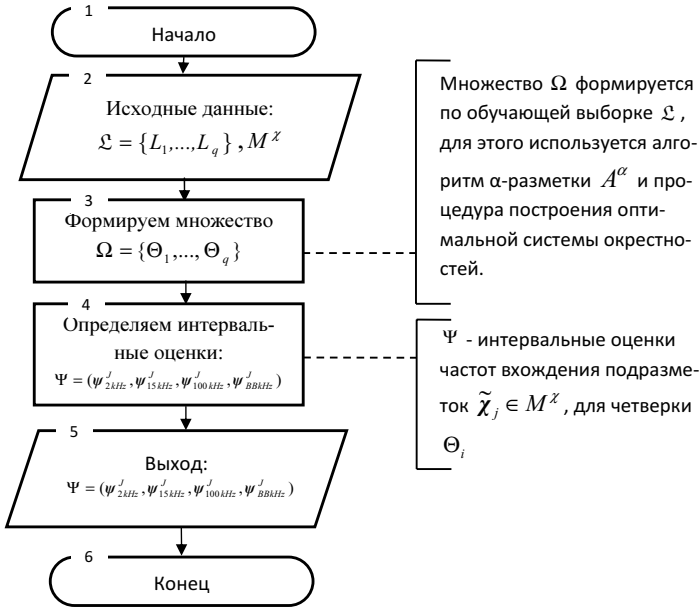


Рис. 3.16. Блок-схема этапа обучения

Используя алгоритм α -разметки A^α и системы окрестностей $(O_{2kHz}^n, O_{15kHz}^n, O_{100kHz}^n, O_{BBkHz}^n)$, которые определяются с помощью численной процедуры построения оптимальной системы окрестностей (см. п. 3.2.), по обучающей выборке $\mathcal{L} = \{L_1, \dots, L_q\}$, где $L_i = (s_{2kHz}^n, s_{15kHz}^n, s_{100kHz}^n, s_{BBkHz}^n)$ формируется множество $\Omega = \{\Theta_1, \dots, \Theta_q\}$, где $\Theta_i = (\mu_{2kHz}^n, \mu_{15kHz}^n, \mu_{100kHz}^n, \mu_{BBkHz}^n)$ (рис. 3.16, блок 2).

По множеству $\Omega = \{\Theta_1, \dots, \Theta_q\}$, используя алгоритм β -разметки A^β , определяются интервальные оценки, частоты вхождения описаний параметров предвестника $\tilde{\chi}_j \in M^X$ в каждую из разметок четверки $\Theta_i = (\mu_{2kHz}^n, \mu_{15kHz}^n, \mu_{100kHz}^n, \mu_{BBkHz}^n)$:

$$\Psi = (\psi_{2kHz}^J, \psi_{15kHz}^J, \psi_{100kHz}^J, \psi_{BBkHz}^J), \quad \Psi^J = \{\psi_1, \dots, \psi_J\}, \quad \psi_j = \{(p, \gamma_j(p)) | p \in \{p_k^{\min}, \dots, p_k^{\max}\}\},$$

где $\gamma_j(p)$ - функция принадлежности, характеризующая степень принадлежности частоты вхождения p описания параметра предвестника $\tilde{\chi}_j$ к множеству ψ_j .

Функция принадлежности [73] $\gamma(p)$ принимает значения в линейно упорядоченном множестве $\Gamma=[0,1]$ и определяется экспертом.

Этап обнаружения. Блок схема этапа обнаружения представлена на рисунке 3.17.

Исходными данными алгоритма (рис. 3.17, блок 1) являются:

- четверка $S_T = (s_{2kHz}^n, s_{15kHz}^n, s_{100kHz}^n, s_{BBkHz}^n)$ которая формируется за время T наблюдения геодинамического процесса;
- интервальные оценки $\Psi = (\psi_{2kHz}^J, \psi_{15kHz}^J, \psi_{100kHz}^J, \psi_{BBkHz}^J)$, полученные на этапе обучения;
- множество описаний параметров предвестника M^X , которое использовалось на этапе обучения.

По четверке $S_T = (s_{2kHz}^n, s_{15kHz}^n, s_{100kHz}^n, s_{BBkHz}^n)$, используя алгоритм разметки A^n и численную процедуру построения оптимальной системы окрестностей, получаем $\Theta = (\mu_{2kHz}^n, \mu_{15kHz}^n, \mu_{100kHz}^n, \mu_{BBkHz}^n)$.

Далее, на основе алгоритма β -разметки определяем частоты вхождения описаний параметров предвестника $\tilde{\chi}_j \in M^X$ в каждую разметку четверки $\Theta = (\mu_{2kHz}^n, \mu_{15kHz}^n, \mu_{100kHz}^n, \mu_{BBkHz}^n)$. В результате получаем набор

$I_T = (I_{2kHz}^J, I_{15kHz}^J, I_{100kHz}^J, I_{BBkHz}^J)$ допустимых описаний $I^J = (p_1, \dots, p_J)$ вектор-объектов из набора $S_T = (s_{2kHz}^n, s_{15kHz}^n, s_{100kHz}^n, s_{BBkHz}^n)$, где p_j - частота вхождения описания параметра предвестника $\tilde{\chi}_j$ в α -разметку μ^n вектор-объекта s^n



Рис. 3.17. Блок-схема этапа обнаружения

В итоге определяем принадлежность $I_T = (I_{2kHz}^J, I_{15kHz}^J, I_{100kHz}^J, I_{BBkHz}^J)$ к интервальной оценке $\Psi = (\psi_{2kHz}^J, \psi_{15kHz}^J, \psi_{100kHz}^J, \psi_{BBkHz}^J)$. Для этого рассчитываем по каждому элементу четверок I_T и Ψ :

$$g(I^J) = \begin{cases} 0, & \text{если } \gamma_k^\Sigma(I^J) = 0 \\ \frac{1}{d} \gamma_k^\Sigma(I^J), & \text{в противном случае} \end{cases}, \quad \gamma_k^\Sigma(I^J) = \sum_{j=1}^J \gamma_j(\rho_j),$$

где $\gamma_j(p_j)$ - функция принадлежности p_j к множеству $\psi_j \in \Psi^j$.

На выходе алгоритма получаем четверку $G = (g_{2kHz}, g_{15kHz}, g_{100kHz}, g_{BBkHz})$ и среднее значение \hat{g} элементов четверки G . Величину \hat{g} интерпретируем как оценку близости наблюдений геодинамического процесса S_T к объектам из обучающей выборки $\mathcal{L} = \{L_1, \dots, L_q\}$. Значение \hat{g} , будем называть оценкой значения функции $g(t)$, рассчитанного для четверки $S_T = (s_{2kHz}^n, s_{15kHz}^n, s_{100kHz}^n, s_{BBkHz}^n)$ (формируется по предыстории наблюдения) в момент времени t наблюдения геодинамического процесса.

Экспериментальные результаты. Обучающая выборка $\mathcal{L} = \{L_1, \dots, L_6\}$ (фрагмент обучающей выборки приведен в приложении Б) формировалась по временным рядам, полученным при разрушении шести образцов горной породы: № 3,4,6,8,9,10 (приложение А), в качестве функции принадлежности использовался гауссиан. Для получения α -разметки временных рядов использовался корректный алгоритм SMA.

Полученные на этапе обучения интервальные оценки $\Psi = (\psi_{2kHz}^j, \psi_{15kHz}^j, \psi_{100kHz}^j, \psi_{BBkHz}^j)$ частоты вхождения описаний параметров предвестника $\tilde{\chi}_j \in M^z$ в каждую разметку четверки $\Theta_i = (\mu_{2kHz}^n, \mu_{15kHz}^n, \mu_{100kHz}^n, \mu_{BBkHz}^n)$ представлены на рисунке 3.18. Для каждой подразметки построена карта поверхности, где по оси Ox отложена частота, на которой проводились измерения (каждой частоте соответствует

временной ряд), по оси O_y частота вхождения подразметки. Цветовой гаммой показано изменение значения функции принадлежности γ_k .

Численные значения интервалов частоты вхождения параметров предвестника $\tilde{\chi}_j \in M^x$ приведены в таблице 3.12.

Алгоритм обнаружения тестировался на временных рядах, полученных в ПНИЛ ЭДиП НИ ТПУ при разрушении лабораторных образцов, отобранных с железорудного месторождения, и данных, полученных во время проведения технологических взрывов на шахте Таштагольская.

На графиках 3.19 и 3.20 показано изменение оценки \hat{g} значения характеристической функции $g(t)$, построенное для временных рядов, полученных при разрушении образцов №3 и №9. Участки, входящие в обучающую выборку $\mathcal{L}_a = \{L_1, \dots, L_6\}$, по которой были получены интервальные оценки $\Psi = (\psi_{2kHz}^J, \psi_{15kHz}^J, \psi_{100kHz}^J, \psi_{BBkHz}^J)$, выделены на графиках синей штриховой линией.

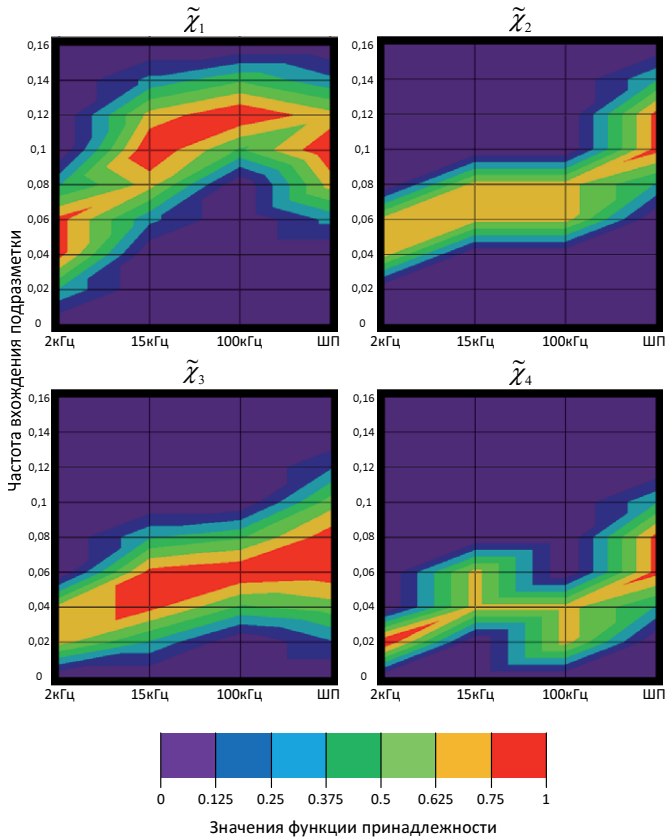


Рис. 3.18. Значения функции принадлежности γ_j для каждого параметра предвестника $\tilde{\chi}_j \in M^x$ в зависимости от их частоты вхождения и частоты измерения ЭМС

Таблица 3.12 Интервальные оценки частоты вхождения параметров предвестника $\tilde{\chi}_j \in M^x$ в каждую разметку из набора $(\mu_{2\text{kHz}}^n, \mu_{15\text{kHz}}^n, \mu_{100\text{kHz}}^n, \mu_{\text{ВВkHz}}^n)$

Частота измерения ЭМС	Номер параметра предвестника $\tilde{\chi}_j \in M^x$			
	1	2	3	4
2 кГц	[0,0.1]	[0.04,0.1]	[0, 0.06]	[0, 0.04]
15 кГц	[0.04,0.16]	[0.06,0.12]	[0,0.1]	[0.02,0.08]
100 кГц	[0.1,0.16]	[0.06,0.12]	[0.02,0.1]	[0, 0.06]
ШП	[0.06,0.16]	[0.08,0.16]	[0,0.14]	[0.02,0.12]

Из графиков 3.19 и 3.20 видно, что максимум оценки значений характеристической функции соответствует участкам временных рядов входящими в обучающую выборку.

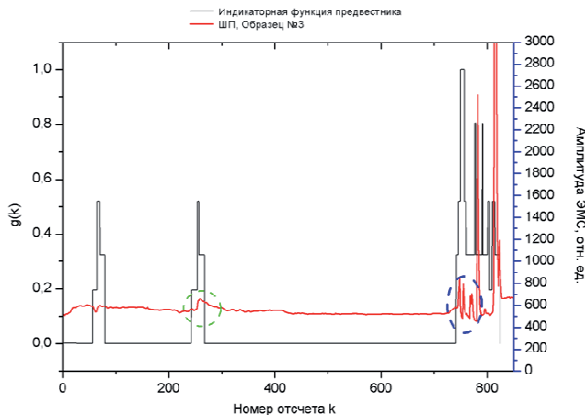


Рис. 3.19. Характеристическая функция $g(k)$, построенная для временных рядов измерений ЭМС, полученных при лабораторном разрушении образца №3

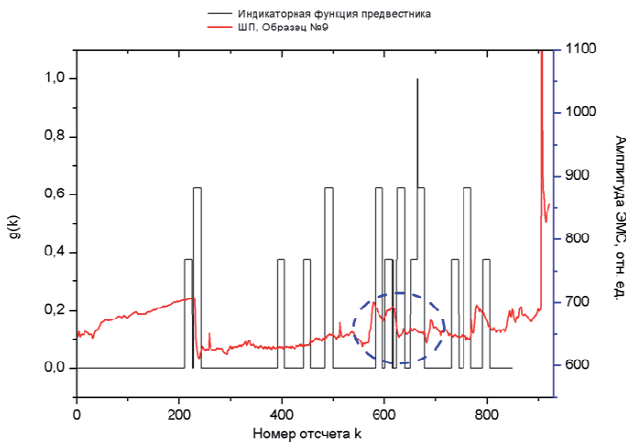


Рис. 3.20. Характеристическая функция $g(k)$, построенная для временных рядов измерений ЭМС, полученных при лабораторном разрушении образца №9

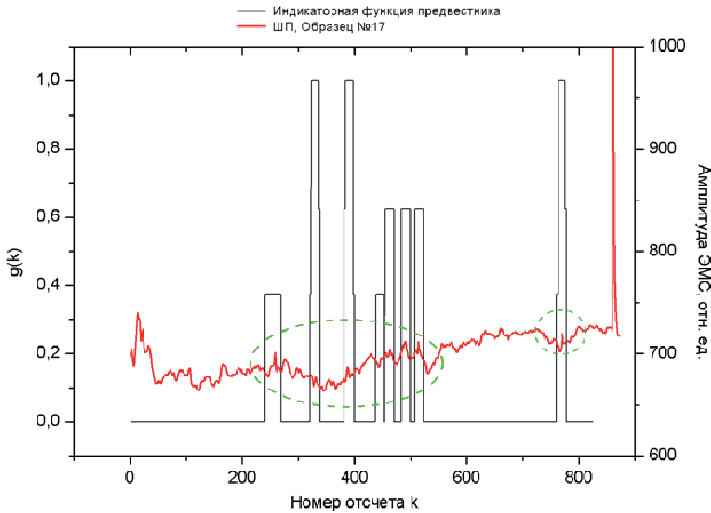


Рис. 3.22. Характеристическая функция $g(k)$, построенная для временных рядов измерений ЭМС, полученных при лабораторном разрушении образца №17

Для образца №17 (приложение А) характеристическая функция имеет два максимума (рис. 3.21), которые соответствуют зонам с возможным формированием очага разрушения (выделены зеленой штриховой линией).

Алгоритм был опробован на реальных данных, полученных во время проведения технологических взрывов на шахте Таштагольская (рис. 3.22, 3.23).

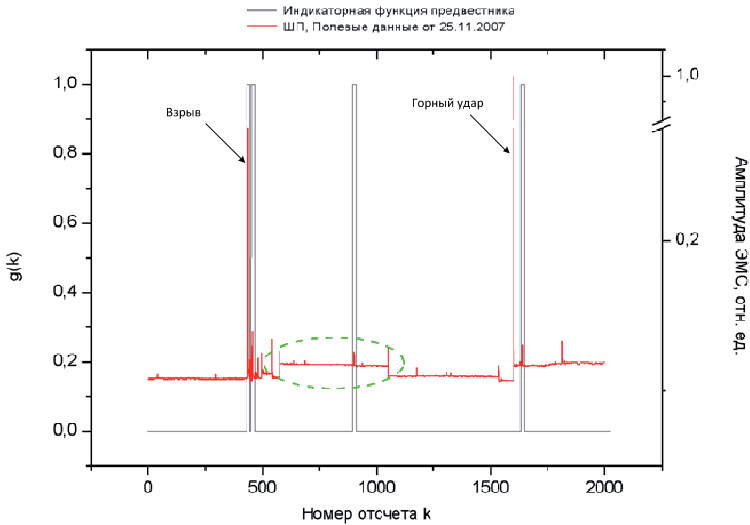


Рис. 3.22. Характеристическая функция $g(k)$, построенная для временных рядов измерений ЭМС, полученных при полевых измерениях в ноябре 2007 года (технологических взрыв и последовавший за ним горный удар)

Как было показано в [2], при обработке рудных месторождений происходит перераспределение напряженно-деформированного состояния горного массива. Использование при этом взрывной технологии приводит к возникновению динамических проявлений горного давления различного энергетического класса. После массовых взрывов в шахте Таштагольская существенно изменяется интенсивность электромагнитного сигнала. Возрастание интенсивности потока электромагнитной эмиссии после взрыва объясняется тем, что в зоне наблюдения после мощного динамического воздействия происходит резкое изменение напряженно-деформированного состояния: подвижки по тектоническим нарушениям, разгрузка напряженных зон, попадающих в зону действия

взрыва. Это сопровождается динамическими явлениями, в виде горных ударов, толчков.

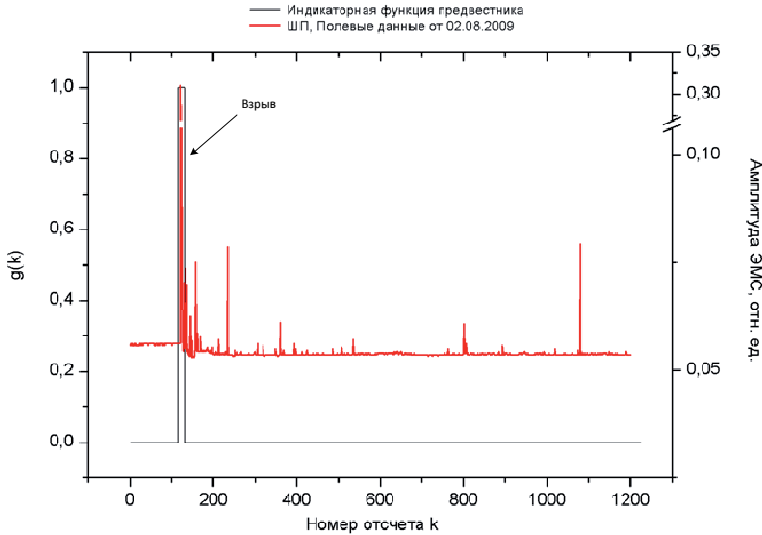


Рис. 3.23. Характеристическая функция $g(k)$, построенная для временных рядов измерений ЭМС, полученных при полевых измерениях в августе 2009 года (технологических взрыв, горного удара не последовало)

Особый интерес представляют фрагменты изменения ЭМС в момент массовых взрывов. Для тестирования были выбраны данные полученные в ноябре 2007 года (взрыв и последовавший за ним горный удар, рис. 3.22) и августе 2009 года (горного удара не было, рис. 3.23).

На графике 3.22 в области выделенной зеленой штриховой линией происходит изменение уровня ЭМС, согласно установленной в [2,57-59] взаимосвязи быстрых и медленных изменений напряженно-деформированного состояния массива горных пород с вариациями ин-

тенсивности ЭМС, данную область можно интерпретировать как предвестник геодинамического события (в данном случае горного удара).

Характеристическая функция $g(k)$ имеет максимум в зоне предвестника предшествовавшей горному удару (рис. 3.22) и равна нулю после технологического взрыва для данных (рис. 3.23), где горного удара не последовало.

Выводы

1. Предложен корректный алгоритм SMA разметки стохастического временного ряда. Применение корректирующей операции позволяет улучшить качество разметки (уменьшить количество ошибочных разметок) алгоритма SMA на 1-7 %. Алгоритм SMA, по сравнению с алгоритмом ARADS в зависимости от параметров зашумления и размера окрестности, обеспечивает уменьшение количества ошибочных разметок до 22 %. В результате сравнительного анализа алгоритмов SMA и ARADS было показано, что при увеличении уровня зашумления алгоритм ARADS совершает большее количество ошибочных разметок при том же уровне шума. Этот факт позволяет заключить, что алгоритм SMA обладает большей обобщающей способностью.

2. Для обеспечения разрешимости и регулярности задачи α , β разметки и уменьшения числа ошибочных разметок стохастического временного ряда предложена численная процедура построения оптимальной системы окрестностей и проведена оценка времени ее сходи-

мости. В результате, определено характерное время сходимости равное 30 итерациям. При испытаний, проведенных на тестовой выборке, было показано, что использование численной процедуры построения оптимальной системы окрестностей позволяет повысить качество распознавания описаний параметров предвестника (в смысле близости β -разметки зашумленного вектор-объекта к β -разметке не зашумленного вектор-объекта) до 15 %.

3. Используя корректный алгоритм α -разметки SMA, алгоритм β -разметки и численную процедуру построения оптимальной системы окрестностей, предложен алгоритм обнаружения предвестников геодинамического события позволяющий для исходных данных в виде набора вектор-объектов оценить значение характеристической функции, которая характеризует принадлежность исходного набора вектор-объектов к классу предвестника. Работоспособность (в смысле соответствия расположения максимума оценки значения характеристической функции экспертным оценкам границы этапа формирования очага разрушения) алгоритма была показана на данных, полученных при лабораторном разрушении образцов горной породы, и данных изменения ЭМС в момент проведения технологических взрывов на шахте Таштагольская.

Глава 4. ПРОГРАММНЫЙ КОМПЛЕКС ДЛЯ ИССЛЕДОВАНИЯ ГЕОФИЗИЧЕСКИХ ДАННЫХ. БАЗА ДАННЫХ ГЕОФИЗИЧЕСКОГО ПРОЦЕССА

Разработанные в главе 3 алгоритмы применяются в новом программном комплексе (ПК) ReDSS [78,79], который предназначен для исследования и мониторинга процессов представленных в виде временных рядов. Основными отличиями ПК ReDSS от подобных комплексов, описание которых приведено в главе 1, является распределенная архитектура и независимость от сторонних пакетов программ (сторонние пакеты программ, например MATLAB, могут быть использованы для создания подключаемых библиотек, используемых сервером).

Распределенная архитектура ПК ReDSS позволяет оперировать большими объемами данных (например, обучающие выборки в виде набора многомерных временных рядов), использовать набор разнообразных алгоритмов, предназначенных для удаленного анализа и обработки временных рядов.

Для структурированного хранения временных рядов, сформированных на основе зарегистрированных специальной аппаратурой, установленной в массиве горных пород, была разработана база данных. База данных используется программным комплексом ReDSS при формировании обучающей выборки и тестировании алгоритмов при определении предвестников горного удара.

Разработанная база данных может применяться в различных интеллектуальных системах сбора и обработки информации для исследования геофизических процессов.

Программный комплекс разработан для операционной системы MS Windows, как наиболее распространенных ОС. Для успешного запуска ПК ReDSS необходим компьютер с ОС Windows XP, Vista 7 с установленными Microsoft .NET Framework 2.0 и СУБД Microsoft SQL Server Express 2005.

Главной причиной выбора платформы Microsoft.NET является наличие объектно-ориентированной архитектуры для поддержки распределенных приложений Microsoft.NET Remoting [82].

Серверным доменом приложения выбрана служба Windows, поскольку она является наиболее подходящим способом реализации постоянно исполняющегося сервера (они способны запускаться автоматически и не требуют входа пользователя в систему на компьютере-сервере).

Для создания WEB-интерфейсов базы данных и ПК ReDSS использовалась технология ASP.NET, так как она является составной частью платформы Microsoft.NET.

4.1. Архитектура программного комплекса ReDSS

Программный комплекс имеет распределенную архитектуру, которая предполагает, что на стороне сервера развертывается база данных

(исходные данные, тестовые выборки) и набор алгоритмов включенных в состав комплекса (в виде подключаемых модулей). Для настройки сервера, а также для демонстрации его работы и просмотра списка подключенных модулей, в состав комплекса входит WEB-клиент. Кроме того, к серверу можно подключаться из сторонних приложений, для этого доступны программные сборки с интерфейсами 1 и 2 (рис. 4.1).

Любой пользователь комплекса может добавить свой модуль в каталог алгоритмов, используя сборки с интерфейсами 1, 3 (рис. 4.1). Каждый модуль представляет собой отдельную библиотеку, подключаемую к ПК.

Описание основных программных интерфейсов 1,2,3 (рис. 4.1) приведено в приложении В.

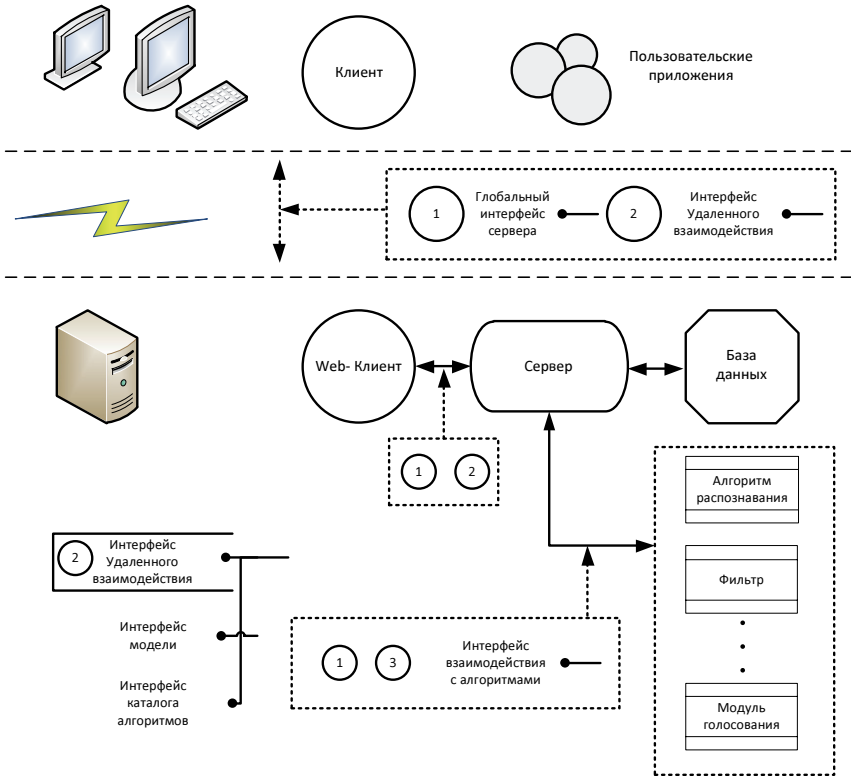


Рис. 4.1. Блок-схема программного комплекса

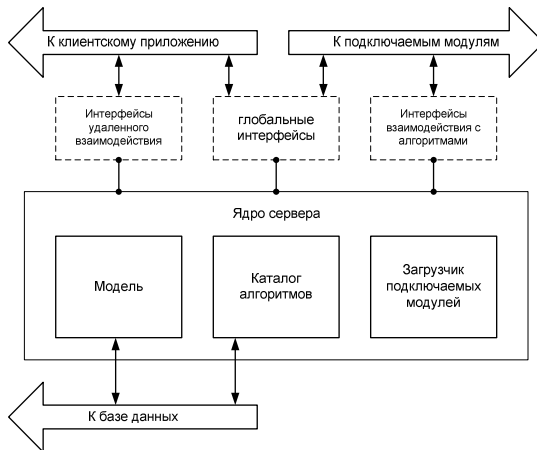


Рис. 4.2. Блок-схема сервера

Статический класс `PluginLoader` предназначен для загрузки алгоритмов, реализованных в виде библиотек. Он включает в себя четыре списка (рис. 4.3) которые формируются из описаний доступных библиотек (класс `UIDPlugin`, рис 4.3):

- `Filters` – содержит различные алгоритмы и процедуры обработки временных рядов (фильтры, нормализация значений ряда, вычитание математического ожидания и др.);
- `Generators` – включает в себя библиотеки организующие доступ к различным источникам исходных данных (доступ к базе данных на стороне сервера или на стороне клиента, математические модели и др.);
- `Monitors` – вывод результатов работы алгоритмов и их предварительная обработка (запись в файл, подготовка к печати и выводу на график)
- `Recognizeers` – алгоритмы разметки, принятия решений, распознавания.

Далее рассмотрим блок-схему сервера, показанную на рисунке 4.2 и диаграмму классов ядра сервера (рис. 4.3).

Блок модели входящий в состав сервера, содержит параметры используемых клиентским приложением модулей и описание их взаимодействия. В каталоге алгоритмов, хранится описание всех модулей, которые в данный момент может использовать клиент.

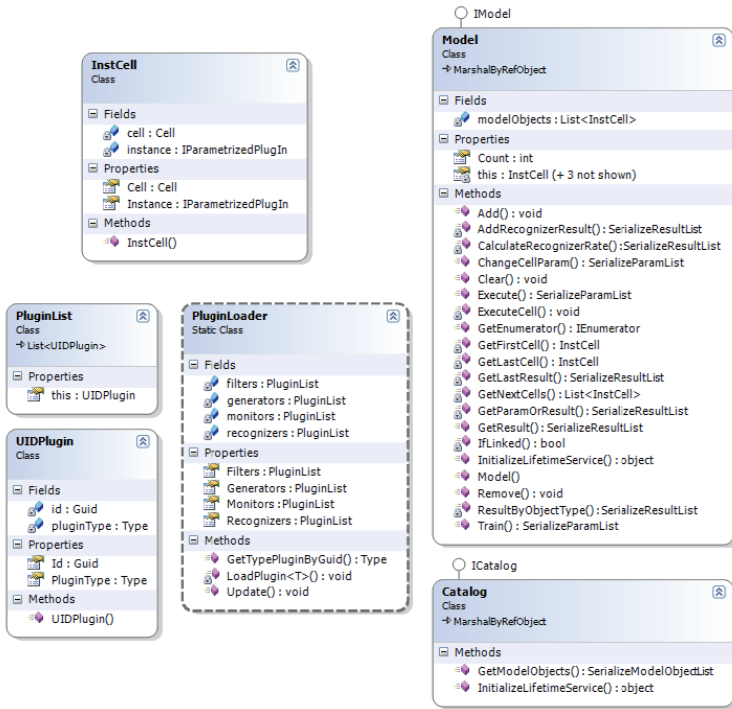


Рис. 4.3. Диаграмма классов ядра сервера

С помощью метода `GetModelObjects` класса `Catalog` клиентское приложение получает список доступных описаний модулей (см. интерфейс `IModelObject` приложение В), схема взаимодействия модулей из этого списка, выстроенная клиентским приложением обрабатывается классом `Model`, для этого создаются экземпляры каждого модуля (класс `InstCell`) из клиентской схемы.

Принцип построения клиентских приложений можно проиллюстрировать с помощью схемы, показанной на рисунке 4.4.

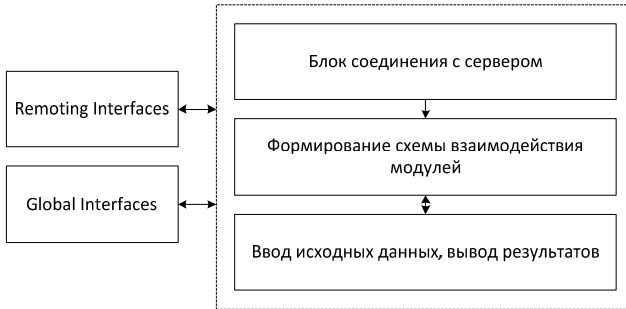


Рис. 4.4. Общая схема клиентского приложения

Клиентское приложение обычно состоит из трех блоков (рис. 4.4):

- блок соединения с сервером, где клиент по заданному адресу соединяется с сервером, получает список модулей и на сервере создается экземпляр класса Model;

- блок формирования схемы взаимодействия модулей, где согласно задаче решаемой клиентским приложением, из модулей входящих в полученный от сервера список, формируется схема их взаимодействия;

- ввод исходных данных, вывод результатов – используя экземпляр класса Model и схему взаимодействия модулей, клиентское приложение получает от сервера результаты обработки исходных данных.

Обмен данных с сервером осуществляется с использованием глобального интерфейса сервера (Global Interfaces, приложение В) и интерфейса удаленного взаимодействия (Remoting Interfaces, приложение В).

4.2. База данных результатов электромагнитных и акустических измерений геофизических процессов

База данных предназначена для структурирования и хранения временных рядов, сформированных на основе зарегистрированных специальной аппаратурой с использованием соответствующих датчиков, установленных в массиве горных пород, а также временных рядов полученных при разрушении образцов горных пород в лабораторных условиях [82]. База данных применяется для формирования обучающих и тестовых выборок при решении задачи обнаружения предвестника геодинамического события.

Опишем логическую модель базы данных. Основными объектами предметной области являются:

- временной ряд, полученный при разрушении образца горной породы (т.е. при лабораторном эксперименте, далее «лабораторный ряд»);
- временной ряд, сформированных по результатам измерений проводимых в массиве горных пород (далее «ряд полевых измерений»).

На уровне реализации к этим объектам добавляется - пользователь системы. Введение последнего продиктовано сложной структурой организации санкционированного доступа к двум основным объектам.

Общими параметрами лабораторных рядов и рядов полевых измерений являются:

- тип использованного для измерений датчика;
- календарная дата измерений, в результате которых был получен данный временной ряд;
- частотный диапазон, в котором измерялись сигналы;
- измеряемая величина (ЭМС, акустика и др.).

Для лабораторных временных рядов к этому списку добавляется описание разрушаемого образца и вид воздействия на образец, а для рядов полевых измерений – координаты места установки регистратора, временной интервал измерений и события, произошедшие на этом интервале.

Тип датчика, частотный диапазон, вид измеряемой величины, описание образцов, события образуют вспомогательные сущности.

Описание образцов объединяет в себе информацию о горной породе, из которой состоит образец, внешнем виде образца после разрушения и о присвоенном ему номере. Информация о виде измеряемой величины, частотном диапазоне объединена, в сущности «тип данных». Сущность «события» хранит информацию о типе события (технологический взрыв, горный удар и др.) времени, когда оно произошло и координатах места.

Так как, необходимо иметь ссылку на оригинальный файл временного ряда, помимо того что хранится в базе, вводятся сущности: «ори-

гинальный файл - лабораторный ряд», «оригинальный файл - ряд полевых измерений» и «оригинальный файл».

Для того, чтобы указать принадлежность того или иного ряда к определенной выборке, вводятся сущности: «тэг - лабораторный ряд», «тэг – ряд полевых измерений» и сущность «тэг».

Для облегчения управления доступом к базе данных вводятся роли пользователей, с которыми сопоставляются с одной стороны - комплект прав доступа, необходимых для выполнения конкретных функций (добавление, изменение, удаление, просмотр данных), с другой стороны – подмножество пользователей, которые должны иметь эти права. Сущность «пользователь-роль» содержит информацию о роли пользователя. Пользователь базы данных может входить в одну из трех ролей:

- «гости» - разрешено только просматривать данные, но не разрешено их скачивать (если доступ к базе осуществляется через WEB-интерфейс);

- «пользователи» - могут просматривать, добавлять, удалять данные, если пользователь работает с базой через WEB-интерфейс, то он может скачивать данные;

- «администраторы» - разрешено все действия пользователя, плюс - возможность удаления, редактирования данных, добавления новых пользователей.

Внутримашинная база данных состоит из совокупности бинарных файлов и базы данных под управлением СУБД MS SQL Server. Использование файловой системы и бинарных файлов обусловлено наличием набора данных, которые могут быть добавлены пользователем через WEB-интерфейс.

Структура таблиц базы данных, отображающих основные сущности, представлена в таблицах 4.1 (лабораторный временной ряд), 4.2 (временной ряд полевых измерений).

Таблица 4.1 Структура таблицы *LaboratoryMeasurements* (лабораторный временной ряд)

Название поля	Тип данных	Комментарии
ID_LaboratoryMeasurements	int	Первичный ключ.
ID_User	int	Вторичный ключ, указывающий пользователя, который внес последние изменения.
ID_SensorType	int	Вторичный ключ для связи с таблицей типов используемых для измерений датчиков.
ID_DataType	int	Вторичный ключ для связи с таблицей типов данных (измеряемая величина, частотный диапазон).
ID_Sample	int	Вторичный ключ для связи с таблицей образцов
Date	date	Дата проведения эксперимента.
Data	varhar	Временной ряд.
OriginFile	varhar	Ссылка на файл вре-

		менного ряда.
ImpactOnSample	varhar	Тип воздействия на образец.
Coments	varhar	Комментарии.
ChangeDate	datetime	Дата и время последнего изменения.

Таблица 4.2 Структура таблицы *FieldData* (временной ряд полевых измерений)

Название поля	Тип данных	Комментарии
ID_FieldData	int	Первичный ключ.
ID_User	int	Вторичный ключ, указывающий пользователя, который внес последние изменения.
ID_SensorType	int	Вторичный ключ для связи с таблицей типов используемых для измерений датчиков.
ID_DataType	int	Вторичный ключ для связи с таблицей типов данных (измеряемая величина, частотный диапазон).
Date	date	Дата проведения эксперимента.
TimeInterval	varhar	Временной интервал измерений.
Data	varhar	Временной ряд.
OriginFile	varhar	Ссылка на файл временного ряда.
DetectorLocation	varhar	Координаты места установки регистратора
Comments	varhar	Комментарии.
ChangeDate	datetime	Дата и время последнего изменения.

Полная физическая схема базы данных отражающая ее структуру приведена в приложении Г.

4.3. WEB-клиент

Для демонстрации работы алгоритмов, реализованных в виде подключаемых модулей (которые хранятся на стороне сервера) и работы с базой данных результатов электромагнитных и акустических измерений геофизических процессов, разработано WEB-приложение, главная страница которого показана на рисунке 4.5.

ReDSS
информационная интеллектуальная программная система

Алгоритмы
О проекте
Авторы

[Прогнозирование на основе авторегрессионной модели](#)
[Фильтрация на основе авторегрессионной модели](#)
[Морфологическая фильтрация](#)
[Алгоритм разметки ARADS](#)
[Математическая модель контура смещения теплоула...](#)

База данных результатов электромагнитных и акустических измерений геофизических процессов.

Логин/Пароль: _____
Вход

Демо Описание

Алгоритм разметки временных рядов ARADS

y
 1,02
1,00
0,98
0,96
0,94
0,92
0,90
0,88
 x
 0 10 20 30 40 50 60 70 80 90

Выбрать тестовый временной ряд: (2) Профелирование
 Загрузить временной ряд: (2) Выберите файл | файл не выбран
 Величина скользящего окна: (2) 8
 Размер алфавита: (2) 4

Старт Обработанный ряд График (PNG)

Все права защищены © 2009 Axid Studio

Рис. 4.5. Главная страница web-клиента

На главной странице (рис. 4.5) можно выделить три основных блока: блок меню алгоритмов; блок описания и демонстрации работы алгоритмов; блок доступа к базе данных.

Меню алгоритмов формируется по списку модулей, полученному от сервера (см. п. 4.1), используя интерфейс ICatalog (приложение В). Для выбранного в меню, алгоритма, в зависимости от активной вкладки (вкладки «Демо» и «Описание», рис. 4.5), отображается его описание, либо функционал, с помощью которого можно его протестировать.

Во вкладке «Демо» пользователь может протестировать выбранный алгоритм, используя временной ряд из предлагаемого набора или загрузить свой, предварительно выставив необходимые параметры. Обработка данных, введенных пользователем, начинается по нажатию кнопки «старт».

Результат работы алгоритма показан в виде графика (для алгоритмов разметки разноцветные участки фона соответствуют различным меткам из алфавита разметки), кроме этого, пользователь может загрузить полученный на выходе алгоритма ряд в виде текстового файла или сохранить изображение полученного графика в формате PNG (рис. 4.5).

Используя форму ввода логина и пароля, находящуюся в правом верхнем углу главной страницы, пользователь получает доступ к управлению базой данных результатов электромагнитных и акустических измерений геофизических процессов.

Опишем основные функции WEB-клиента базы данных. Интерфейс WEB-страницы, которая открывается после входа в базу данных, показан на рисунке 4.8.

Панель управления находящаяся в верхней области страницы, содержит: панель управления аккаунтами со ссылкой на аккаунт менеджер (доступна только для администратора) и информацию о пользователе (рис. 4.6); блок навигации, состоящий из меню таблиц базы данных (рис. 4.7, а) и панели быстрого поиска по базе данных (рис. 4.7, б).



Рис. 4.6. Панель управления аккаунтами

Используя аккаунт менеджер, администратор базы данных может регистрировать новых пользователей в системе, изменять права доступа или удалять уже существующих.



Рис. 4.7. Блок навигации (а – меню таблиц базы данных, б – панель быстрого поиска по базе данных)

Используя меню таблиц (рис. 4.7, а) пользователь выбирает таблицу для ее просмотра и редактирования. Панель быстрого поиска позволяет по введенной комбинации символов найти все строки в базе данных (за исключением таблиц управления доступом и служебных

таблиц). Поиск ведется по всем редактируемым (т.е. по полям, значения которых вводятся пользователем) строковым полям.



Рис. 4.8. WEB-страница, на которую происходит переадресация после аутентификации пользователя базы данных

В верхней части блока просмотра и редактирования содержимого таблицы, расположена панель фильтров (рис. 4.9), предназначенная для упрощения поиска групп данных в базе.

Слева от панели фильтров расположены кнопки добавления и удаления строк таблицы. Добавление новой строки в таблицу осуществляется путем заполнения формы (форма добавления нового образца показана на рисунке 4.10), переадресация происходит после нажатия кнопки «добавить» (рис. 4.9).

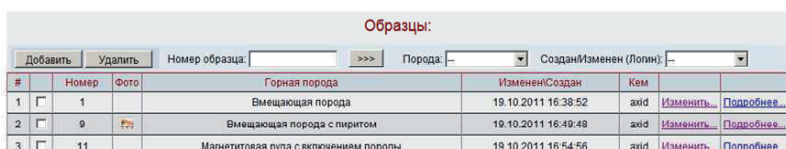


Рис. 4.9. Фрагмент таблицы образцов и ее панель фильтров

Рис. 4.10. Форма добавления нового образца

Для того чтобы удалить строку или строки из таблицы, галочкой выделяются подлежащие удалению строки и нажимается кнопка «удалить». Подобным образом организовано добавление и удаление строк для всех таблиц базы данных (изменяется только заполняемая пользователем форма).

Страница для просмотра данных соответствующих образцу №9 показана на рисунке 4.11 (для получения доступа к странице переходим по ссылке «подробнее...» строка №2, рис. 4.9).

#	Тип данных	Дата эксперимента	Изменить	Просмотреть
1	Смет(шт)	25.05.2008	Изменить	Просмотреть
2	Амплитуда ЭМС(1шт)	25.05.2008	Изменить	Просмотреть
3	Амплитуда ЭМС(100 кгц)	25.05.2008	Изменить	Просмотреть
4	Амплитуда ЭМС(10 кгц)	25.05.2008	Изменить	Просмотреть
5	Амплитуда ЭМС(2 кгц)	25.05.2008	Изменить	Просмотреть

Рис. 4.11. Просмотр данных по образцу №9

Для конкретного образца формируется таблица всех ассоциированных с ним временных рядов, выводится его фотография после разрушения, тип горной породы и его описание.

Страницы просмотра информации о полевых и лабораторных временных рядах представлены на рисунках 4.12 и 4.13, соответственно.

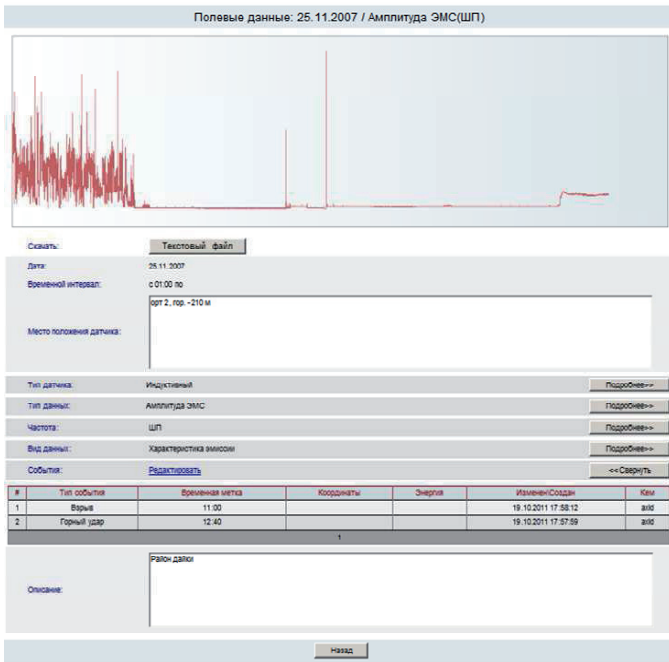


Рис. 4.12. Страница просмотра полевых данных

При просмотре полевых данных генерируется график временного ряда и все его параметры. Если за указанный интервал времени наблюдения произошли события (технологический взрыв, горный удар и др.), и эти события были добавлены в базу, то при подробном просмотре полевых данных выводится таблица событий (рис. 4.12).

Страница просмотра информации о временных рядах, ассоциированных с каким либо из образцов (рис. 4.13), отличается наличием информации о нем и отсутствием таблицы событий.

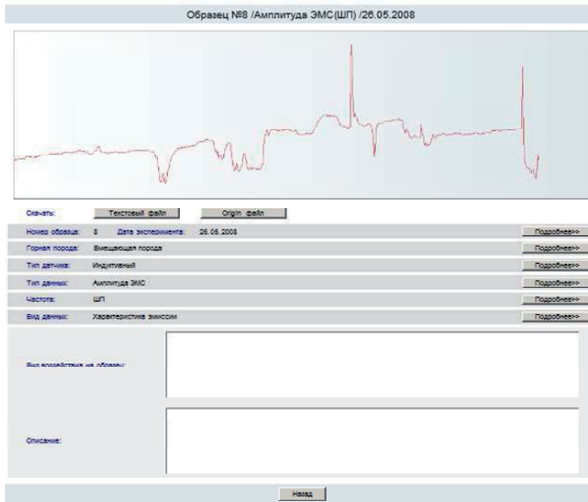


Рис. 4.13. Страница просмотра информации об измерениях, полученных в результате лабораторного эксперимента

Пользователь может скачать (кроме пользователей входящих в группу «гостей») любой временной ряд в текстовом формате и в исходном формате (если файл в исходном формате задан для данного временного ряда).

4.4. Программа «ReDSS Geo» для анализа и мониторинга геофизических процессов, описание которых представлено в виде временных рядов измерений ЭМС

Алгоритмы разметки временного ряда SMA, ARADS [80], численная процедура построения оптимальной системы окрестностей и алго-

ритм обнаружения предвестника геодинамического события (гл. 3), используя интерфейс IParametrizedPlugIn (приложение В), реализованы в виде подключаемых модулей и размещены на сервере.

Для исследования разработанных алгоритмов и мониторинга геофизических процессов, описание которых представлено в виде временных рядов, было разработано клиентское desktop-приложение «ReDSS Geo».

Программа «ReDSS Geo» позволяет:

- определять оптимальный размер скользящего окна (то есть вычислять оптимальный размер системы окрестностей), используемый при α -разметки временного ряда;
- производить разметку временных рядов геофизических данных, используя один из двух алгоритмов (SMA или ARADS);
- вычислять характеристическую функцию для набора временных рядов геофизических данных.

Далее кратко описаны основные принципы работы пользователя с программой и функционал ее графического интерфейса.

В результате запуска программы «ReDSS Geo» открывается главное окно программы, показанное на рисунке 4.14. В главном окне программы располагаются основные элементы управления, позволяющие пользователю выбрать и настроить алгоритм разметки, найти опти-

мальные системы окрестностей для загруженных временных рядов и построить характеристическую функцию.

Для загрузки временных рядов используется пункт меню «Файл» > «Открыть». Пользователь указывает путь к папке с файлами временных рядов, названия файлов указываются в настройках «Настройки» > «Формат файлов» (рис. 4.15).

На главном окне программы расположены три вкладки: «алгоритм разметки»; «размер системы окрестностей»; «характеристическая функция».

Во вкладке «алгоритм разметки» (рис. 4.16) пользователь может выбрать и настроить один из двух алгоритмов разметки и загрузить библиотеку функций.

Библиотека функций загружается из текстового файла (для файлов библиотек функций зарезервировано расширение «flib»), содержащего в каждой строке временной ряд, полученный дискретизацией базисной функции.

Для настройки алгоритма SMA, пользователю необходимо выбрать одну из предлагаемых метрик (аналоги Евклидовой, линейной метрик или расстояния Чебышева) или использовать их комбинацию, указать размер алфавита символов и количество символов в строке на выходе алгоритма.

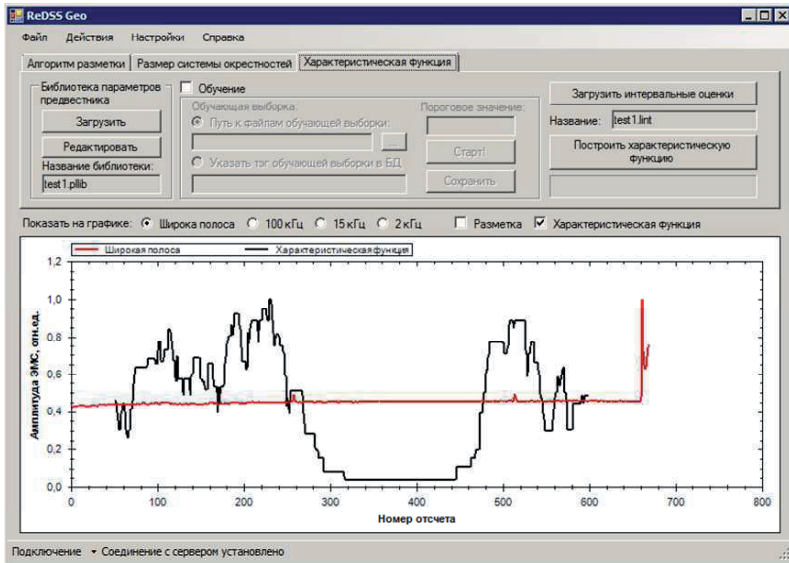


Рис. 4.14. Главное окно программы «ReDSS Geo»

Если пользователь в качестве алгоритма α -разметки выбрал алгоритм ARADS, ему необходимо выбрать метрику (Чебышевское расстояние, линейная метрика или Евклидова метрика) или использовать их комбинацию (корректирующая операция).

При выборе комбинированной метрики необходимо указать веса метрик для каждой библиотечной функции или загрузить ранее сохраненный набор.

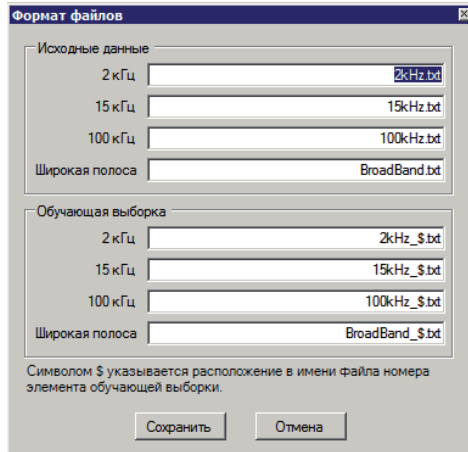


Рис. 4.15. Окно настройки формата имен файлов

На вкладке «размер системы окрестностей» (рис. 4.17), пользователь может вручную задать размер системы окрестностей (имеется в виду размер u_c см. п. 3.2), для каждого временного ряда или определить его при помощи численной процедуры построения оптимальной системы окрестностей, нажатием кнопки «Рассчитать».

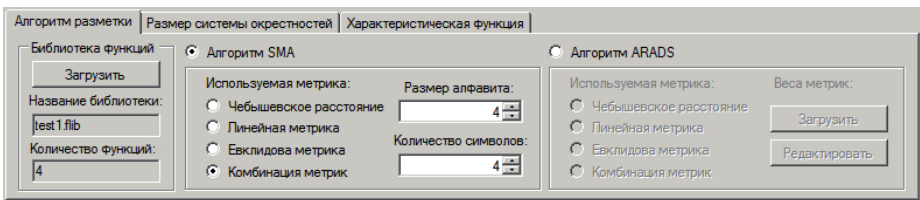


Рис. 4.16. Вкладка выбора и настройки алгоритма α -разметки

Для настройки процедуры пользователю необходимо указать весовые коэффициенты критериев (энтропийного критерия, точности разметки и размера окрестности), сумма весов должна быть равна единицы, ввести пороговое значение энтропии (чтобы отключить предвари-

тельный выбор систем окрестностей по пороговому значению энтропии, необходимо ввести «-1») и задать точность расчета энтропийного критерия.

Параметры процедуры могут быть сохранены в текстовом файле, и при следующем запуске программы загружены из него.

Рис. 4.17. Вкладка расчета размера систем окрестностей

Выбрав вкладку «характеристическая функция» (рис. 4.18) пользователь может построить характеристическую функцию, используя библиотеку параметров предвестника (которая задается экспертом), для выбранных временных рядов. При этом используется выбранный алгоритм α -разметки с введенными настройками и рассчитанные для загруженных временных рядов размеры систем окрестностей.

Рис. 4.18. Вкладка расчета характеристической функции

Перед тем как нажать кнопку - «построить характеристическую функцию», необходимо загрузить интервальные оценки или рассчитать их, используя панель обучения.

После активизации панели обучения необходимо указать путь к файлам обучающей выборки или ввести тег, используемой в базе данных для пометки рядов входящих в обучающую выборку и указать пороговое значение, используемое при поиске параметров предвестника. После выполнения данных действий необходимо нажать кнопку «Старт!». Интервальные оценки, полученные в результате обучения, можно сохранить и использовать их при следующем запуске программы. Вместе с интервальными оценками сохраняются параметры выбранного алгоритма и размеры систем окрестностей, при загрузке интервальных оценок они тоже восстанавливаются.

Для сохранения и загрузки интервальных оценок файла, размеров систем окрестностей и параметров алгоритма разметки (для файлов интервальных оценок зарезервировано расширение «`lint`»), используется xml парсер.

Параметры предвестника представляют собой комбинацию меток функций из библиотеки, которая загружаются на вкладке «алгоритм разметки». Для файлов библиотек параметров предвестника зарезервировано расширение «`plib`».

Полученные значения характеристической функции можно вывести на график или сохранить на диск в виде текстового файла («Файл» > «Сохранить»).

Выводы

1. Приведено описание архитектуры разработанного программного комплекса ReDSS, который позволяет: тестировать новые методы и алгоритмы, предназначенные для обработки и анализа временных рядов; проводить исследование и мониторинг процессов, описание которых представлено в виде временных рядов. Программный комплекс ReDSS имеет распределенную архитектуру, позволяющая оперировать большими объемами данных (например, обучающие выборки в виде набора многомерных временных рядов), использовать набор разнообразных алгоритмов предназначенных для удаленного анализа и обработки временных рядов.

2. Представлено описание структуры базы данных результатов электромагнитных и акустических измерений геофизических процессов, разработанной и применяемой для структурирования и хранения временных рядов, сформированных на основе зарегистрированных специальной аппаратурой с использованием соответствующих датчиков, установленных в массиве горных пород, а также временных рядов полученных при разрушении образцов горных пород в лабораторных условиях. База данных используется в программном комплексе ReDSS для

формирования обучающих, тестовых выборок при решении задачи обнаружения предвестника геодинамического события. Кроме этого база данных может быть применена в различных интеллектуальных системах сбора и обработки информации при исследовании геофизических процессов.

3. Приведено описание WEB-интерфейсов программного комплекса ReDSS и базы данных результатов электромагнитных и акустических измерений геофизических процессов. Разработанный WEB-интерфейс позволяет демонстрировать работу алгоритмов используемых в программном комплексе и управлять базой данных.

4. Используя интерфейс программирования приложений (набор интерфейсов) программного комплекса ReDSS, было реализовано клиентское desktop-приложение «ReDSS Geo». Алгоритмы разметки временного ряда SMA, ARADS, численная процедура построения оптимальной системы окрестностей и алгоритм обнаружения предвестника геодинамического события реализованы в виде подключаемых модулей и размещены в каталоге алгоритмов сервера программного комплекса ReDSS. Приложение «ReDSS Geo» предназначено для исследования разработанных алгоритмов и мониторинга геофизических процессов, описание которых представлено в виде временных рядов.

Результаты, представленные в 4-й главе, опубликованы в работах [74-81].

ЗАКЛЮЧЕНИЕ

В заключении отметим основные теоретические и практические результаты:

1. предложена математическая модель этапа формирования очага разрушения (предвестника), позволяющая представить изменение электромагнитного сигнала комбинацией функциональных зависимостей (параметров предвестника).

2. разработан корректный алгоритм разметки стохастического временного ряда SMA. Алгоритм SMA по сравнению с алгоритмом ARADS, в зависимости от параметров зашумления и размера окрестности, обеспечивает уменьшение количества ошибочных разметок, примерно, до 22 %.

3. для обеспечения разрешимости и регулярности задачи разметки и уменьшения числа ошибочных разметок стохастического временного ряда, разработана численная процедура построения оптимальной системы окрестностей, позволяющая снизить количество ошибок обнаружения параметра предвестника, примерно, на 15%.

4. создан алгоритм обнаружения предвестников геодинамического события, позволяющий оценить степень принадлежности стохастического временного ряда к классу предвестников. Алгоритм обнаружения был исследован на временных рядах, полученных в ПНИЛ ЭДиП ТПУ при разрушении лабораторных образцов отобранных с железоруд-

ного месторождения и данных, полученных при проведении технологических взрывов на шахте Таштагольская.

5. разработан программный комплекс ReDSS, имеющий распределенную архитектуру. В программном комплексе используются разработанные алгоритмы, реализованные в виде подключаемых модулей. Используя интерфейс программирования приложений (набор интерфейсов) программного комплекса ReDSS, реализовано клиентское desktop-приложение «ReDSS Geo», предназначенное для исследования разработанных алгоритмов и мониторинга геофизических процессов, описание которых представлено в виде временных рядов.

6. разработана база данных результатов электромагнитных и акустических измерений геофизических процессов, предназначенная для структурированного хранения временных рядов, сформированных на основе зарегистрированных специальной аппаратурой с использованием соответствующих датчиков, установленных в массиве горных пород. База данных используется в ПК ReDSS, как источник исходных данных или при формировании обучающих, контрольных и тестовых выборок.

СПИСОК ЛИТЕРАТУРЫ

1. Потураев В.Н., Булат А.Ф., Хохолев В.К. Об особенностях комплексной регистрации электромагнитного и акустического излучения при разрушении горных пород//ДАН СССР. - 1989. - Т. 308. - N 6. - С.1351-1354.

2. Яворович Л.В. Взаимосвязь параметров электромагнитных сигналов с изменением напряженно-деформированного состояния горных пород: Дис. канд. техн. наук. Томск, 2005. 196 с.

3. Иванов В.В., Егоров П.В., Колпакова Л.А., Пимонов А.Г. Динамика трещин и электромагнитное излучение нагруженных горных пород // ФТПРПИ. - 1988. - N 5. - С. 20 - 27.

4. Гордеев В.Ф., Малышков Ю.П., Чахлов В.Л и др. Электромагнитная эмиссия диэлектрических материалов при статическом и динамическом воздействии //ЖТФ.- 1994. - Т.64, в. 4. - С.57-67.

5. Bieniawski Z.T.: Mechanism of brittle fracture of rock – CSIR Report, MEG580, Pretoria.-224, 1967.

6. Cook N.G.W.: Failure of rock. – J. Rock Mech. Min. Sc.-1965.- №2.- P.389-403.

7. Wawersik W.R. and Brace W.F.: Post – failure behaviour of a granite and diabase. – Rock Mech., 1971.-№3.-P 61-85.

8. Маслов Ш.Р., Гольд Р.М., Яворович Л.В. Влияние масштабного фактора на электромагнитную эмиссию/Томск. политех. институт.-

Томск, 1987.- 12 с.- Деп. в ВИНТИ 19.11.87, N 8167- В87.

9. Егоров П.В., Иванов В.В., Колпакова Л.А. О некоторых закономерностях импульсного электромагнитного излучения щелочногалоидных кристаллов и горных пород//ФТПРПИ. - 1988.- N 1. - С. 67-70.

10. А.с. 1562449 СССР. Способ прогноза разрушения массива горных пород / М.В. Курленя, В.Н. Опарин, Г.Е. Яковицкая. – Оpubл. в Б.И. – 1990. - №17.

11. М.В. Курленя, А.Г. Вострецов, Г.И. Кулаков. Г.Е. Яковицкая. Регистрация и обработка сигналов электромагнитного излучения горных пород. – Новосибирск: Издательство СО РАН. 2000 г, 231 с.

12. Курленя М.В., Кулаков Г.И., Яковицкая Г.Е. Спектрально-временной анализ электромагнитной эмиссии при трещинообразовании образцов горных пород//ФТПРПИ. - 1993. - N 1. - С. 3 - 13.

13. Кулаков Г.И., Яковицкая Г.Е. Особенности изменения спектра частот электромагнитного излучения при разрушении образцов горных пород//ПМТФ. - 1994. - Т.35, N 5 (207). - С. 160 - 165.

14. Кулаков Г.И., Кривецкий А.В., Бритков Н.А. и др. Электромагнитное излучение при разрушении стекол // Стекло и керамика. – 1998. - № 4. – С. 7 – 10.

15. Яковицкая Г.Е. Методы и технические средства диагностики критических состояний горных пород на основе электромагнитной эмиссии – Новосибирск: Параллель, 2008. – 315 с.

16. Патент 2006884 РФ. Устройство для регистрации электромагнитного излучения, возникающего при трещинообразовании горных пород/ М.В. Курленя, Г.И. Кулаков, В.А. Марков, Г.Е. Яковицкая. – Оpubл. в Б.И. – 1994. - №2.

17. М.В. Курленя, А.Г. Вострецов, Г.И. Кулаков. Г.Е. Яковицкая. Регистрация и обработка сигналов электромагнитного излучения горных пород. – Новосибирск: Издательство СО РАН. 2000 г, 231 с.

18. Пат. 2338065 РФ, МПК Е 21 С 39/00. Способ прогноза разрушения горных пород / В.Н. Опарин, А.Г. Вострецов, В.Е. Петров, Г.Е. Яковицкая. – Оpubл. в БИ, 2008, № 35.

19. Поиск электромагнитных предвестников землетрясений / Под ред. М.Б. Гохберга. - М.: ИФЗ АН СССР, 1988. - 167 с.

20. Гуфельд И.Л., Рожной А.А., Никифорова Н.Н. и др. Спектрально-временные характеристики электромагнитного излучения и токов в образце на стадии его разрушения//Напряженно-деформированное состояние массива горных пород.- Новосибирск: изд. ИГД СО АН СССР, 1988.

21. Кушнир А.Ф. Мостовой С.В. Статистический анализ геофизических полей, Киев, Наукова думка, 1990, 270с.

22. Любушин А.А. (мл), Копылова Г.Н. Многомерный вейвлет-анализ временных рядов электротеллурических наблюдений на Камчатке // Физика Земли. 2004. № 2. С. 82-96.

23. Копылова Г.Н. Поиск предвестников землетрясений по данным электротеллурических наблюдений на Камчатке // Солнечно-земные связи и электромагнитные предвестники землетрясений : III междунар. конф., с. Паратунка Камч. обл., 16–21 авг. 2004 г. : сб. докл. – Петропавловск-Камч. : ИКИР ДВО РАН, 2004. – 456 с.

24. Томилин Н.Г., Дамаскинская Е.Е., Павлов П.И. Статистическая кинетика разрушения горных пород и прогноз сейсмических явлений //Физика твердого тела. – 2005. – Т.47, вып. 5. – С. 955-959.

25. Romeo G. Seismic signal detection and classification using artificial neural networks. Special issue on the workshop “Planning and procedures for GSETT-3”, Erice, November 10-14, 1993. Annali di Geofisica. Vol. XXXVII, N. 3, 1994, p. 343-353.

26. Куравский Л. С., Баранов С. Н., Буланова О. Е., Кравчук Т. Е. Нейросетевая технология диагностики патологических состояний по аномалиям электроэнцефалограмм // Нейрокомпьютеры: разработка и применение.2007. № 4.

27. А.Д. Гвишиани, С.М. Агаян, Ш.Р. Богоутдинов, А.В. Леденев, Ж. Злотники, Ж. Боннин. Математические методы геоинформатики. II. Алгоритмы нечеткой логики в задачах выделения аномалий на временных рядах. Киев: Кибернетика и системный анализ, 2003, № 4, с.103-111.

28. А.Д. Гвишиани, С.М. Агаян, Ш.Р. Богоутдинов, Ж. Злотники.

Алгоритмы нечеткой логики в задачах выделения аномалий на временных рядах. В кн.: «Очерки геофизических исследований. К 75-летию Объединенного института физики Земли им. О.Ю. Шмидта». М.: ОИФЗ РАН, 2003, с.257-262.

29. A.D. Gvishiani, S.M. Agayan, Sh.R. Bogoutdinov, S.A. Tikhotsky, J.Hinderer, J.Bonnin, M.Diament. Algorithm FLARS and recognition of time series anomalies. System Research & Information Technologies. 2004, no. 3, 7-16.

30. J. Zlotnicki, J.-L. LeMouel, A. Gvishiani, S. Agayan, V. Mikhailov, Sh. Bogoutdinov. Automatic fuzzy-logic recognition of anomalous activity on long geophysical records. Application to electric signals associated with the volcanic activity of la Fournaise volcano (Reunion Island). Earth and Planetary Science Letters, vol. 234, 2005, p.261-278.

31. Нечеткие множества в моделях управления и искусственного интеллекта / Под ред. Д.А.Поспелова. - М.: Наука, Гл. ред. физ.-мат. лит., 1986 – 321 с.

32. Кедров О.К., Пермякова В.Е., Стеблов Г.М. Методы обнаружения слабых сейсмических явлений в пределах платформ. М., ОИФЗ РАН, 2000, 101с.

33. Введение в механику скальных пород: Пер. с англ. / Под ред. Х. Бока. – М.: Мир, 1983. – 276 с.

34. Проект: Методы, программное обеспечение, распределенные

вычисления, экспериментальные данные и распределенные вычисления сейсмических шумов Земли («Дыхание Земли») [Электронный ресурс]/ Институт физики Земли им. О.Ю.Шмидта РАН; чл.корр. Г.А. Соболев, д.ф.м.н. А.А. Любушин. – Электрон.дан. – М.: Электронная земля. Информационный портал «Геофизика», [2011]. – Режим доступа: http://earth.wdcb.ru/sob_r.html, свободный. – Загл. с экрана.

35. Степанов М. В. Изучение краткосрочных предвестников землетрясений методом электромагнитной эмиссии: Дис. канд. физ.-мат. наук. М., 2004.

36. Семенов Н.А. Методы автоматизированного проектирования системы прогнозирования землетрясений: Дис. канд. техн. наук. Санкт-Петербург, 2007. 136с.

37. Katkovnik, V., Egiazarian K., Astola J., Local Approximation Techniques in Signal and Image Processing, SPIE Press, Monograph Vol. PM157, September 2006. Hardcover, 576 pages.

38. Wavelab 850 [Электронный ресурс]/ Stanford University; Donoho D., Maleki A., Morteza S. – Электрон.дан. – Stanford: Department of Statistics, [2011]. – Режим доступа: <http://www-stat.stanford.edu/~wavelab/>, свободный. – Загл. с экрана.

39. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Части 1 и 2. – М.: «Мир», 1974. – 406 с.

40. Воропаев П.В. Разработка автоматизированной системы вы-

числения эффективности методик прогнозов землетрясений // Проблемы комплексного геофизического мониторинга Дальнего Востока России. Первая региональная научно-техническая конференция. Петропавловск-Камчатский. 11-17 ноября 2007 г. тезисы докладов. Петропавловск-Камчатский: ГС РАН, 2007. 118 с.

41. Полигон алгоритмов [Электронный ресурс]: система для массового тестирования алгоритмов классификации на реальных задачах. – Электрон.дан. –М., [2011]. – Режим доступа:

<http://poligon.machinelearning.ru/>, свободный. – Загл. с экрана.

42. Луценко Е.В. Универсальная автоматизированная система распознавания образов "Эйдос" (версия 4.1).-Краснодар: КЮИ МВД РФ, 1995.- 76с..

43. Гольденберг Л.М., Матюшкин Б.Д., Поляк М.Н. Цифровая обработка сигналов. Справочник – М.: Радио и связь, 1985. 312 с.

44. Колесникова С.И. Проблемно-ориентированные модели распознавания и оценивания состояний сложных объектов. Авторефер.дис. докт. техн. наук. Томск, 2011.

45. Чехович Ю.В. Элементы алгебраической теории синтеза обучаемых алгоритмов выделения трендов: Дис. канд. физ.-мат. наук. М., 2003.

46. Журавлев Ю.И. Корректные алгебры над множествами некорректных (эвристических) алгоритмов. I-III // Кибернетика. - 1977. - № 4.

- С. 5-17, 1977. - № 6. - С. 21-27, 1978. - № 2. - С. 35-43.

47. Журавлёв Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации // Проблемы кибернетики. - 1979. -Т. 33. - С. 5-68.

48. Журавлёв Ю. И, Рудаков К. В. Об алгебраической коррекции процедур обработки (преобразования) информации // Проблемы прикладной математики и информатики. - 1987. - С. 187-198.

49. Журавлев Ю.И. Об одном классе алгоритмов над конечными множествами, ДАН СССР. – Т. 151-5. - 1963. - С. 1025-1028.

50. Рудаков К.В., Чехович Ю.В. Алгебраический подход к проблеме синтеза обучаемых алгоритмов выделения трендов // Доклады РАН. - 2003. -Т. 388. - № 1.- С. 33-36.

51. Саати Т. Принятие решений. Метод анализа иерархий: Пер. с англ.- М.: "Радио и связь", 1993.- 320 с.

52. Саати Т.Л. Принятие решений при зависимостях и обратных связях: Аналитические сети. – М.: Издательство ЛКИ, 2008. – 360 с.

53. Feigenbaum M.J. Universal Behavior in Nonlinear Systems. - Los Alamos Science, 1980. – V.1. – № 1. – P. 4–27.

54. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование. – М.: Финансы и статистика, 2001. – 228 с

55. Безручко Б.П., Смирнов Д.А. Математическое моделирование и хаотические временные ряды. – Саратов: ГосУНЦ «Колледж», 2005. –

320 с.

56. Мертвецов А.Н., Колесникова С.И. Метод разметки стохастического временного ряда // Труды Института Системного Анализа РАН. 2011. Т.61. № 1. С. 48-59.

57. Беспалько А.А., Яворович Л.В., Колесникова С.И., Букреев В.Г., Мертвецов А.Н., Федотов П.И. Исследование изменений характеристик электромагнитных сигналов при одноосном сжатии образцов горных пород Таштагольского рудника // Изв. вузов. Физика. – 2011. – Т.54. -№ 1/2. - С. 78-85.

58. Беспалько А.А., Яворович Л.В., Колесникова С.И., Букреев В.Г., Мертвецов А.Н., Федотов П.И. Деформирование и разрушение материалов с дефектами и динамически явления в горных породах и выработках // Материалы Юбилейной XX Международ. науч. Школы, Симферополь: Таврич. нац. ун-т, 2010. – С. 58-70.

59. Беспалько А.А., Букреев В.Г., Яворович Л.В., Мертвецов А.Н., Виитман Е.В., Колесникова С.И. Исследование процесса образования деструктивных зон в образцах горных пород по электромагнитной эмиссии // Материалы Международного симпозиума «Наноматериалы для защиты промышленных и подземных конструкций» и XI Международной конференции «Физика твердого тела» (ФТТ-XI), 9-12 июня 2010 г., Усть-Каменогорск, 2010. - С. 177-181.

60. Patel, P., Keogh, E., Lin, J., Lonardi, S. Mining Motifs in Massive

Time Series Databases. In proceedings of the 2002 IEEE International Conference on Data Mining. Maebashi City, Japan. Dec 9-12.

61. Lin, J., Keogh, E., Lonardi, S., Chiu, B. A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego, CA. June 13.

62. Мертвецов А.Н. Алгоритм разметки временного ряда для построения модели предвестника геодинамического события // XVII Международная научно-практическая конференция студентов, аспирантов и молодых ученых «Современные техника и технологии» / Сборник трудов в 3-х томах. Т.2. – Томск: Изд-во Томского политехнического университета, 2011.- С. 386-387.

63. Keogh, E., Chakrabarti, K., Pazzani, M. & Mehrotra, S. Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases. In proceedings of ACM SIGMOD Conference on Management of Data. Santa Barbara, CA, 2001. P. 151-162.

64. Yi, B, K., & Faloutsos, C.. Fast Time Sequence Indexing for Arbitrary Lp Norms. In proceedings of the 26st Int'l Conference on Very Large Databases., Cairo, Egypt. 2000. pp. 385-394.

65. Larsen, R. J. & Marx, M. L. An Introduction to Mathematical Statistics and Its Applications. Prentice all, Englewood, Cliffs, N.J. 2nd Edition. 2000. pp. 97-104.

66. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. – М.: Наука, 1974. – 416 с.
67. Левенштейн В. Двоичные коды с исправлением выпадений, вставок и замещений символов // Доклады Академий Наук СССР 163.4:845-848. 1965.
68. Ukkonen, E., Approximate string-matching with q-grams and maximal matches // Theoretical Computer Science. 1992. P. 191-211.
69. Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. – Н.: Наука. – 2007. – 255 с.
70. Шоломов Л.А. О собственной информации нечетких текстов // Нелинейная динамика и управление. Вып. 6. – М.: Физматлит, 2008. – 340 с.
71. Галлагер Р. Теория информации и надежная связь. – М., «Советское радио», 1974. – 720 с.
72. Редько В.Г., Цой Ю.Р. Оценка эффективности эволюционных алгоритмов // Доклады АН, 2005. Т. 404. N. 3. С. 312-315.
73. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. - М.: Мир, 1976. - 168 с.
74. Мертвецов А.Н., Колесникова С.И., Букреев В.Г., Цой Ю.Р. Информационная система для распознавания состояний стохастической системы // Программные продукты и системы. – 2010. – № 4.– С. 128-132.

75. Мертвецов А.Н., Колесникова С.И., Цой Ю.Р. Информационно-аналитическая поддержка обработки зашумленных временных рядов // Компьютерные науки и технологии: сб. трудов Первой Междунар. научно-практ. конф. – Белгород: ГиК, 2009. – Ч. 1:– С. 186–190.

76. Мертвецов А.Н., Колесникова С.И., Букреев В.Г. Программный комплекс для обнаружения закономерностей во временных рядах в условиях неопределенности // Кибернетика и высокие технологии XXI века 13-15 мая 2009 г. Воронеж . Россия. НПФ «Саквее» ООО. 2009.– С. 251-259.

77. Мертвецов А.Н., Колесникова С.И., Букреев В.Г., Цой Ю.Р. Информационно-диагностическое программное средство для распознавания состояний зашумленного динамического объекта // Материалы XXXVI Международной конференции и дискуссионного научного клуба «Информационные технологии в науке, образовании, телекоммуникации, бизнесе (IT+S&E`09)» – Украина, Крым, Ялта–Гурзуф: Приложение к журналу «Открытое образование», 2009. – С. 92-94.

78. Мертвецов А.Н. Программный комплекс для обработки и анализа временных рядов // VIII Всероссийская научно-практическая конференция с международным участием «Современные информационные технологии в науке, образовании и практике» (2009 г., Оренбург, Россия): Труды конференции - Оренбург: ИПК ГОУ ОГУ, 2009. – С. 72-80

79. Свидетельство об официальной регистрации программы ЭВМ

№2010610441. Интеллектуальная модель распознавания состояний динамических систем «IReDSS» / А.Н. Мертвецов, С.И. Колесникова, В.Г. Букреев, Ю.Р. Цой, В.С. Лаходьнов. Заявка 2009616425. Дата поступл. 16.11.09. Зарегистр. 11.01.2010.

80. Свидетельство о государственной регистрации программ для ЭВМ №2011613749. Автоматическая разметка временного ряда с нелинейным трендом / Мертвецов А.Н., Колесникова С.И., Букреев В.Г. Заявка 2011611917. Дата поступл. 22.03.11. Зарегистр. 13.05.2011.

81. Свидетельство об официальной регистрации программы ЭВМ № 2010620277. База данных результатов электромагнитных и акустических измерений геофизических процессов / А.А. Беспалько, В.Г. Букреев, С.И. Колесникова, А.Н. Мертвецов, Л.В. Яворович. Заявка 2010620117. Дата поступл. 19.03.10. Зарегистр. 14.05.2010.

82. Маклин С., Нафтел Дж, Уильяме К. M15 Microsoft .NET Remoting / Пер. с англ. М.: Издательско-торговый дом «Русская Редакция», 2003. – 384 с.: ил.

ПРИЛОЖЕНИЕ А

Используемые лабораторные данные

При определении параметров предвестника геодинамического события использовались лабораторные данные при одноосном сжатии до разрушения 16-ти образцов с железорудного месторождения Таштагольское (табл. П.1).

Таблица П.1 Тип горной породы и номера используемых образцов

№	№ образца	Тип горной породы (вмещающая порода, руда)
1	1	Вмещающая порода
2	3	Магнетитовая руда с кальцитом
3	4	Магнетитовая руда с включением породы
4	5	Магнетитовая руда с включением породы
5	6	Магнетитовая руда с включением породы
6	8	Вмещающая порода
7	9	Вмещающая порода с пиритом
8	10	Магнетитовая руда с кальцитом
9	11	Магнетитовая руда с включением породы
10	12	Магнетитовая руда с включением породы
11	14	Магнетитовая руда с включением породы
12	16	Вмещающая порода
13	17	Магнетитовая руда с включением породы
14	18	Вмещающая порода
15	20	Кварцит
16	21	Кварцит

Для примера приведем изменение усредненной амплитуды ЭМС в частотных полосах с центральными частотами 2 кГц, 15 кГц, 100 кГц и в широкой полосе от 1 до 100 кГц, измеренное, для образцов №16 и №11, при одноосном сжатии до разрушения (рис. П.1, П.2.).

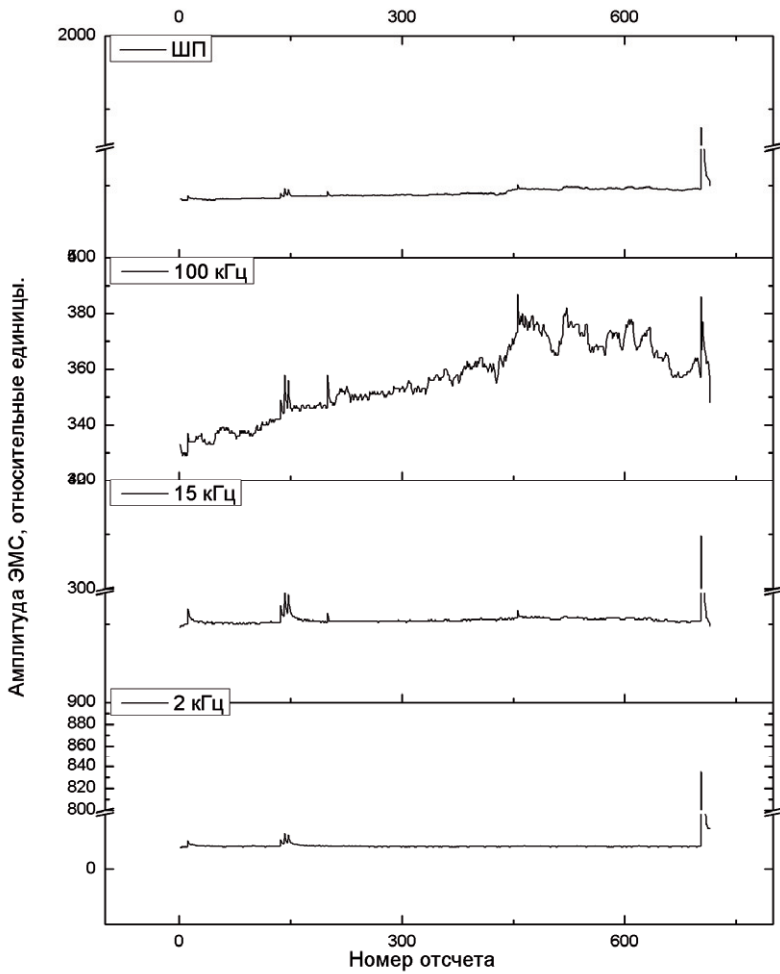


Рис. П.1. Изменение усредненной амплитуды ЭМС в частотных полосах с центральными частотами 2 кГц, 15 кГц, 100 кГц и в широкой полосе от 1 до 100 кГц, для образца № 16

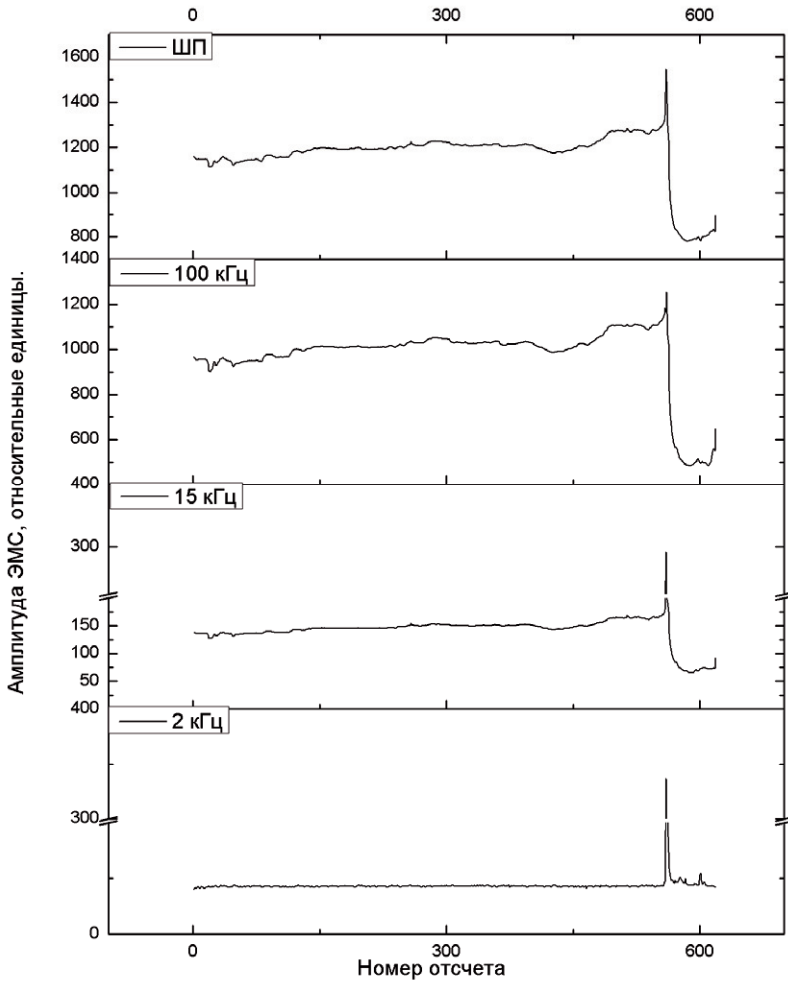


Рис. П.2. Изменение усредненной амплитуды ЭМС в частотных полосах с центральными частотами 2 кГц, 15 кГц, 100 кГц и в широкой полосе от 1 до 100 кГц, для образца № 11

ПРИЛОЖЕНИЕ Б

Фрагмент обучающей выборки

Обучающая выборка формировалась по временным рядам, полученным при разрушении шести образцов горной породы: № 3,4,6,8,9,10. Временные ряды, входящие в обучающую выборку, соответствуют участкам временных рядов, которые были получены при лабораторном разрушении образцов горных пород, границы участка определяются экспертом (считаем, что этот участок соответствует этапу формирования очага разрушения).

Ниже приведены графики временных рядов, входящих в обучающую выборку, полученные по временным рядам изменение усредненной амплитуды ЭМС на частоте 100 кГц для образцов № 3,4,8.

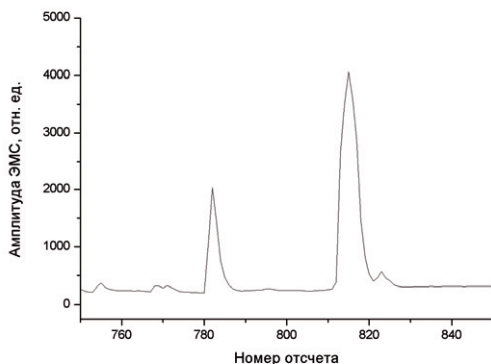


Рис. П.3. График временного ряда, входящего в обучающую выборку (соответствует образцу №3)

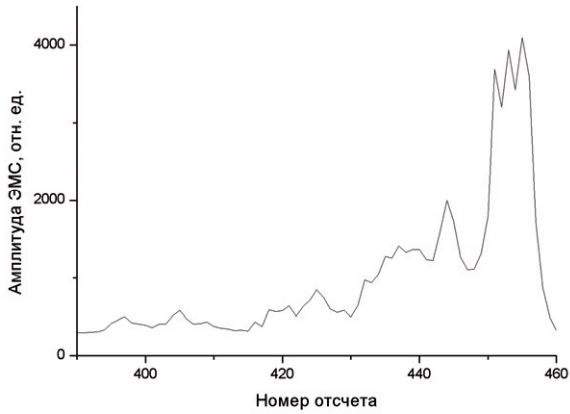


Рис. П.4. График временного ряда, входящего в обучающую выборку (соответствует образцу №4)

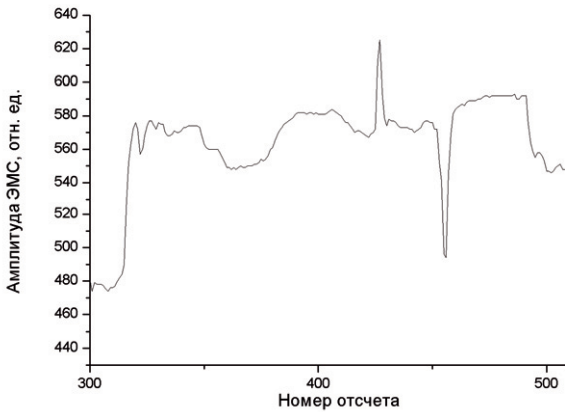


Рис. П.5. График временного ряда, входящего в обучающую выборку (соответствует образцу №8)

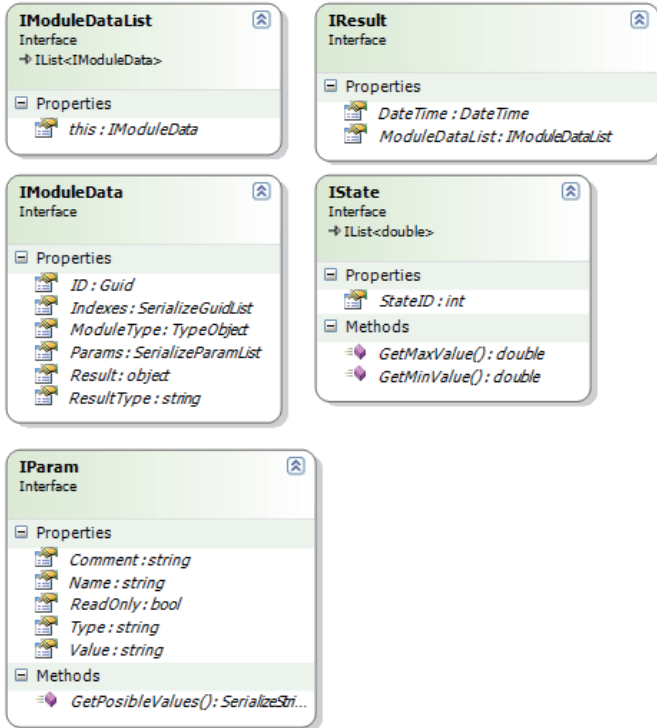
Глобальные интерфейсы сервера

Рис. П.6. Диаграмма глобальных интерфейсов сервера

На рисунке П.6 приведена диаграмма глобальных интерфейсов сервера, описание интерфейсов их свойств и методов представлено в таблице П.2.

Таблица П.2 Описание глобальных интерфейсов сервера, их свойств и методов

Название	Тип	Комментарий
IModuleData		Интерфейс предназначен для описания используемого модуля (алгоритма). Используется клиентским приложением.
ID	Guid	Уникальный идентификатор модуля.
Indexes	SerializeGuidList	Связи с другими модулями.
ModuleType	TypeObject	Тип модуля.
Params	SerializeParamList	Список используемых параметров.
Result	object	Результат работы модуля.
ResultType	string	Тип данных, в котором предоставляются результаты работы модуля.
IModuleDataList		Список модулей используемых клиентским приложением.
this[Guid id]	IModuleData	Индексатор по уникальному идентификатору модуля.
IParam		Описание параметра модуля.
Comment	string	Комментарии.
Name	string	Название.
ReadOnly	bool	Флаг, возможности редактирования значения параметра.
Type	string	Тип данных параметра.
Value	string	Значение параметра.
IResult		Описание результатов работы используемых модулей.
DateTime	DateTime	Время получения результата (указывается сервером)
ModuleDataList	IModuleDataList	Список модулей используемых клиентским приложением.
IState		Описание объекта обучающей выборки.
StateID	int	Идентификатор класса.
GetMaxValue() GetMinValue()	double	Методы получения максимального/минимального значения элементов обучающей выборки.

Интерфейсы удаленного взаимодействия

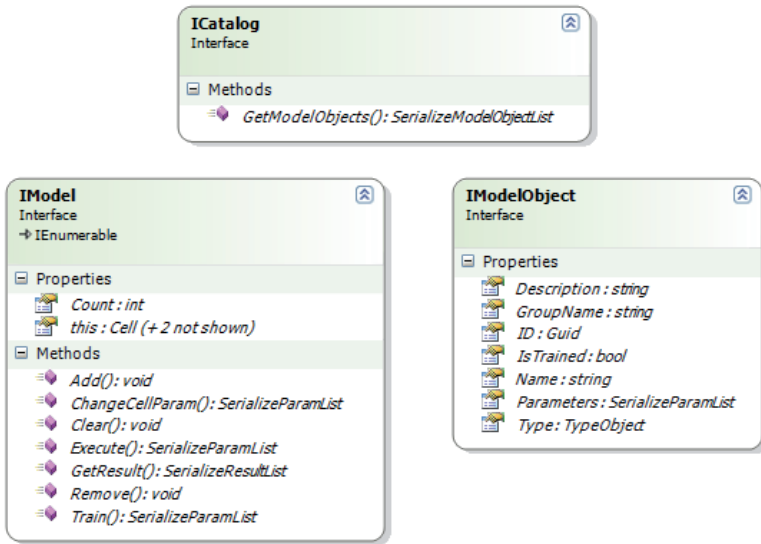


Рис. П.7. Диаграмма глобальных интерфейсов сервера

На рисунке П.7 приведена диаграмма интерфейсов удаленного взаимодействия, описание интерфейсов их свойств и методов представлено в таблице П.3.

Таблица П.3 Описание интерфейсов удаленного взаимодействия их свойств и методов

Название	Тип	Комментарий
IModel		Интерфейс предназначен для формирования используемой клиентской частью комбинации алгоритмов, для дальнейшего расчета на сервере.
Count	int	Количество используемых модулей (алгоритмов)
this[int]	Cell	Индексатор, предназначенный для получения доступа к используемому модулю по индексу.
Add()	void	Добавление модуля.
ChangeCellParam()	SerializeParamList	Изменение параметров модуля.
Clear()	void	Очистка списка используемых модулей.
Execute()	SerializeParamList	Расчет результатов работы заданной комбинации алгоритмов.
GetResult()	SerializeResultList	Получить результат.
Remove()	void	Удалить модуль.
Train()	SerializeParamList	Обучить модуль, используя заданную обучающую выборку.
ICatalog		Предназначен для доступа к каталогу алгоритмов на сервере.
this[Guid id]	IModuleData	Список доступных клиенту модулей (алгоритмов).
IModuleObject		Описание модуля хранящегося в каталоге алгоритмов на сервере.
Description	string	Описание модуля.
GroupName	string	Название группы, куда он входит.
ID	Guid	Уникальный идентификатор.
IsTrained	bool	Флаг, информирующий о необходимости обучения.
Name	string	Название
Parameters	SerializeParamList	Параметры модуля
Type	TypeObject	Тип модуля.

Интерфейсы взаимосвязи с алгоритмами

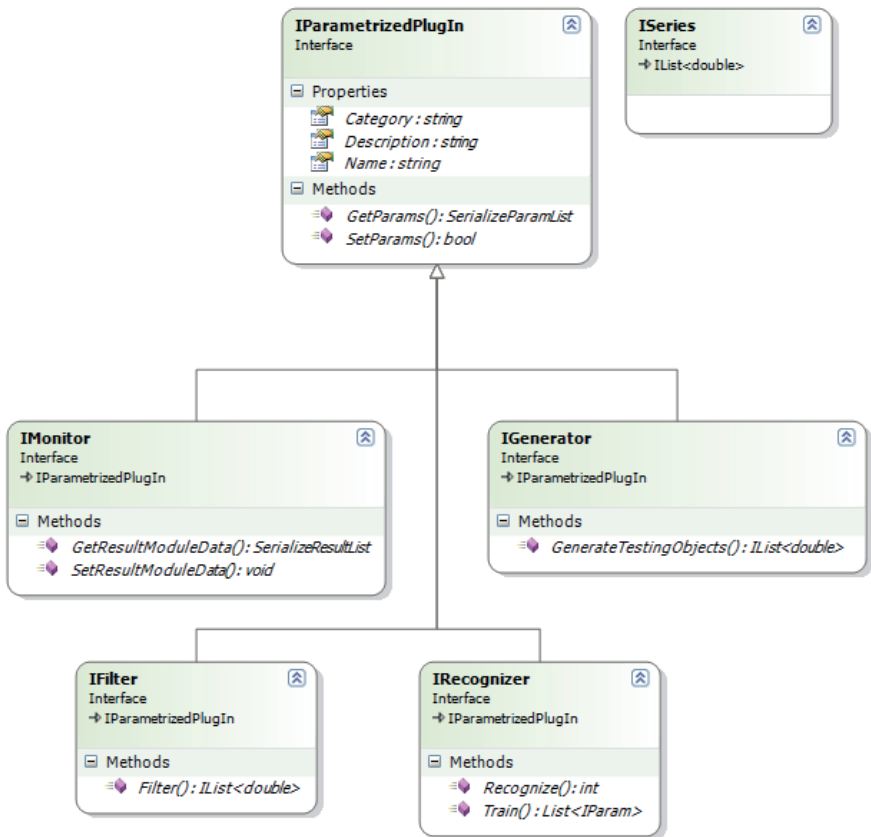


Рис. П.8. Диаграмма интерфейсов взаимосвязи с алгоритмами

На рисунке П.8 приведена диаграмма интерфейсов взаимосвязи с алгоритмами, которые используются сервером для взаимодействия с подключаемыми модулями (библиотеками), описание интерфейсов их свойств и методов представлено в таблице П.4.

Таблица П.4 Описание интерфейсов взаимодействия с алгоритмами их свойств и методов

Название	Тип	Комментарий
IParametrizedPlugIn		Общий интерфейс подключаемого модуля.
Category	string	Категория в каталоге алгоритмов
Description	string	Описание
Name	string	Название
GetParams()	SerializeParamList	Получить список параметров
SetParams()	bool	Установить параметры модуля.
IMonitor		Интерфейс подключаемых модулей предназначенных для вывода результатов работы и их предварительной обработки.
GetResultModuleData()	SerializeResultList	Получить данные хранящиеся в мониторе.
SetResultModuleData()	void	Записать данные в монитор.
IFilter		Интерфейс для описания различных алгоритмов и процедур обработки исходных данных.
Filter()	List<double>	Выполнить процедуру обработки.
IRecognizer		Интерфейс для описания алгоритмов разметки, принятия решений, распознавания.
Recognize()	int	Выполнить алгоритм.
Train()	List<IParam>	Обучить алгоритм.
IGenerator		Интерфейс для описания библиотеки организующие доступ к различным источникам исходных данных
GenerateTestingObjects()	List<double>	Получить исходные данные.
ISeries		Интерфейс для описания исходных данных.



MoreBooks!
publishing



yes i want morebooks!

Покупайте Ваши книги быстро и без посредников он-лайн – в одном из самых быстрорастущих книжных он-лайн магазинов! окружающей среде благодаря технологии Печати-на-Заказ.

Покупайте Ваши книги на
www.more-books.ru

Buy your books fast and straightforward online - at one of world's fastest growing online book stores! Environmentally sound due to Print-on-Demand technologies.

Buy your books online at
www.get-morebooks.com



VDM Verlagsservicegesellschaft mbH

Heinrich-Böcking-Str. 6-8
D - 66121 Saarbrücken

Telefon: +49 681 3720 174
Telefax: +49 681 3720 1749

info@vdm-vsg.de
www.vdm-vsg.de

