

На правах рукописи

**Шкатова Галина Ивановна**

**ИНФОРМАЦИОННАЯ СИСТЕМА ДЛЯ  
РЕШЕНИЯ ЗАДАЧ КЛАССИФИКАЦИИ  
С ИСПОЛЬЗОВАНИЕМ НЕПАРАМЕТРИЧЕСКИХ  
МЕТОДОВ И ОПЕРАТОРОВ-ПРОЕКТОВ**

Специальность 05.13.11 – Математическое и программное обеспечение  
вычислительных машин и систем

**Автореферат**

диссертации на соискание ученой степени  
кандидата технических наук

Томск-1996

Работа выполнена в Томском политехническом университете

Научный руководитель:  
кандидат химических наук,  
старший научный сотрудник Вылегжанин О.Н.

Официальные оппоненты:  
доктор технических наук, профессор Тарасенко Ф.П,  
кандидат технических наук, доцент Воловоденко В.А.

Ведущая организация: ВЦ СО РАН (г.Красноярск)

Защита диссертации состоится 11 декабря 1996г. в 15 часов на заседании  
диссертационного Совета Д 063.80.03 Томского политехнического  
университета по адресу: 634004, г.Томск, пр. Ленина,30.

С диссертацией можно ознакомиться в библиотеке института.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В настоящее время исследования в области разработки интеллектуальных систем, выполняющие роль эксперта или консультанта, лежат на магистральном направлении развития компьютерной информационной технологии. В процессе функционирования таких систем происходит накопление фактов и эвристических элементов и использование этой базы накопленных знаний вместе с механизмом "заключения" для получения логических выводов. Такие системы особенно эффективны, когда пользователь оперирует большой, постоянно увеличивающейся числовой информацией. Разработка таких систем, ориентированных на применение в различных проблемных областях, использующих различные методы обработки информации и ведущих с пользователем диалог на языке, близком к разговорному, позволяет расширить применение формализованных математических методов к решению научных и технических проблем.

Важным средством исследования сложных систем являются методы классификации и распознавания образов. Они позволяют создать представление о структуре этих систем по совокупности наблюдений, объединяя их в классы. Если в результате решения задачи классификации получены группы, однородные по характерным признакам, то в дальнейшем анализе мы можем оперировать такими группами как структурными единицами.

Известны примеры разработок интеллектуальных систем в различных проблемных областях: психодиагностике (система АСПД, Белюк Л. В.), профориентации (система "Ориентир"), (ОТЭКС, Загоруйко), (КВАЗАР, Казанцев В.С.) и др. В то же время опыт работы в области классификации и распознавания образов и участие в последних конференциях по математическим методам распознавания образов привели нас к заключению, что к настоящему времени не разработаны такие интеллектуальные системы для персональных компьютеров, которые позволяли бы им выступать в роли адаптивных экспертов при решении задач, связанных с распознаванием

образов и классификацией и давали бы возможность посредством дедукции пройти путь от решений, предоставляемых простейшими решающими правилами в ситуации хорошо разделимых классов, до сложных комбинаций таких правил, когда границы между соответствующими классами "размыты". В то же время практика показывает целесообразность разработки такого рода систем.

**Цель работы** состоит в разработке интеллектуальной интегрированной системы решения задач классификации, исследовании и практическом применении алгоритмов классификации объектов, заданных численными характеристиками. Основными задачами исследования являются:

- разработка и реализация структуры интегрированной интеллектуальной системы классификации;
- получение и исследование процедур синтеза признакового пространства, а также процедур формирования решающих правил в виде линейных функций с вещественными коэффициентами;
- разработка алгоритмов и программ решения основных задач классификации и построения линейных моделей;
- исследование разработанных алгоритмов и программ на модельных примерах и выработка рекомендаций по применению этих алгоритмов и программ;
- внедрение разработанных алгоритмов и программ в практику.

### **Научная новизна.**

1. Разработана концептуальная модель и макетный вариант интеллектуальной системы, позволяющей решать широкий набор задач классификации. В основу алгоритмов решения различных задач положен единый подход, основанный на рекуррентном псевдообращении и построении операторов-проекторов.

2. Разработан алгоритм решения задачи распознавания образов, основанный на рекуррентной процедуре, сочетающей отбор информативных

признаков и построения соответствующей линейной дискриминантной функции.

3. Разработана процедура включения простых решающих правил типа линейных дискриминантных функций в более сложные структуры типа "коллективы" и "комитеты" решающих правил.

4. Разработан метод кластеризации, который сводит решение задачи многомерной таксономии к последовательности одномерных задач.

5. Разработан алгоритм выбора наилучших линейных регрессоров для построения линейных моделей.

**Методы исследования.** Методы и модели теории вероятностей и математической статистики, системологии, линейной алгебры, вычислительной математики и программирования.

Обоснованность и достоверность выдвигаемых и использованных в диссертации научных положений, выводов и рекомендаций подтверждается:

- теоретическим обоснованием;
- апробированием программных реализаций;
- опытом применения для решения конкретных задач;
- перспективой дальнейшего совершенствования.

**Практическая ценность.** Разработан вариант интегрированной интеллектуальной системы решения задач классификации. Разработанные в диссертационной работе алгоритмы и программы решения задач классификации и построения линейных моделей на базе рекуррентного псевдообращения реализованы в виде комплекса программ, который позволяет производить широкомасштабное исследование структуры экспериментальных данных.

**Апробация работы.** Основные положения диссертационной работы докладывались и обсуждались на: 3-й (г. Львов, 1987г.); 4-й (г. Рига, 1989г. ); 5-й (г. Москва, 1991г. ) Всесоюзных конференциях "Математические методы распознавания образов"; Всесоюзной конференции "Химическая информатика"(г. Черногловка, 1992г.).

**Публикации.** По результатам выполненных исследований было опубликовано 16 работ.

**Тезисы, представляемые к защите.**

1. Сформулирована единая математическая постановка задач классификации и построения линейных моделей.

2. В основу алгоритмов синтеза признакового пространства, построения линейных решающих функций, кластер-анализа может быть положена рекуррентная процедура построения оператора-проектора, вычисляемого с помощью псевдообратной матрицы.

3. Процедура построения линейной дискриминантной функции может быть объединена с алгоритмом выбора информативных признаков в единый рекуррентный процесс.

4. Процесс распознавания может быть реализован средствами интеллектуальной системы по принципу от "простого" к "сложному" с возможностью пересмотра стратегии на любом этапе решения.

Кроме приведенных выше тезисов к защите представлены алгоритмы и программы синтеза признакового пространства, построения линейных дискриминантных функций, построения линейных моделей, кластер-анализа, формирование коллективных методов на основе линейных дискриминантных функций.

**Структура и объем диссертации.** Диссертация состоит из введения, четырех глав и заключения, изложенных на 142 страницах машинописного текста, содержит 38 рисунков, 28 таблиц, список литературы из 90 наименований и приложения.

## СОДЕРЖАНИЕ РАБОТЫ

**Во введении** обосновывается актуальность темы диссертации, приводятся цели и задачи работы, определяются научная новизна и практическая ценность работы.

**В первой главе** рассматриваются основные проблемы, возникающие при разработке интеллектуальных систем (ИС) решения задач классификации. Они определяются наличием двух предметных областей. Первая предметная область - это область решаемой проблемной задачи, а вторая охватывает общие методы решения системных задач, к которым сводится решение проблемной задачи, или системная область.

Проблемная область ИС определена двумя классами задач:

1) разбиение множества объектов на подмножества, объединенные между собой по какому - либо критерию.

2) построение различного рода моделей, главным образом, регрессионных, которые описывают количественные связи между отдельными признаками, либо между отдельными объектами.

Показано, что при определенных условиях можно сформулировать единую математическую постановку этих задач и свести ее к задаче отыскания минимума функционала среднего риска вида:

$$I_C = \int Q(g(x,a))f(x)dx, \quad (1)$$

где  $Q(y)$  - функция потерь;  $g(x,a)$ - оцениваемая функция, в которой неизвестный параметр  $a \in A$ ,  $A$  - область допустимых значений параметров;  $f(x)$  - функция плотности вероятностей;  $x$ - вектор измерений. При построении линейных моделей  $g$  есть функция регрессии, а при РО - это решающее правило.

Оценка функционала среднего риска м.б. получена с помощью функционала эмпирического риска:

$$I_{\mathfrak{G}}(a) = \sum_{i=1}^n Q_i(g, a), \quad (2)$$

При этом функция потерь для задачи (РО) с обучающей выборкой задается в виде:

$$Q_i(g, a) = (\varpi_i - g(x_i, a))^2, \quad (3)$$

где  $\omega_i$  - реакция учителя, а при построении линейных моделей в виде:

$$Q_i(g, a) = (y_i - g(x_i, a))^2, \quad (4)$$

здесь  $y_i$  - наблюдаемое значение функции регрессии. Функция  $p$ , представляющая собой верхнюю оценку уклонения функционала эмпирического риска от функционала среднего риска:

$$P(\sup(I_C(a) - I_{\mathfrak{G}}(a)) \leq \rho) = \eta, \quad (5)$$

где  $\eta$  - доверительная вероятность,  $\rho$  определяется формулой:

$$\rho = 2\sqrt{\frac{\ln(m^2) - \ln(\eta)}{2l}}, \quad (6)$$

здесь  $l$  - объем выборки;  $m^s$ - функция роста, определяющая максимальное число способов разделения выборки.

Утверждается, что поскольку в описываемой системе используются либо линейные дискриминантные функции (ДДФ), либо линейные модели, то минимум функционала эмпирического риска достигается при значении параметра  $a$ , равном:

$$a = G^+(x) \cdot y, \quad (7)$$

где  $G^+(x)$ - матрица, псевдообратная к матрице  $G$ , элементы которой суть значения функции  $g(x)$  в точках  $x$ , а  $y$  - значения решающей функции в этих точках. Этот вывод позволил вести разработку алгоритмов на единой основе.

Показано, что для оценки качества получаемых решений м.б. использованы обычные статистические критерии.

Показано, что в качестве меры информативности признака, характеризующего классифицируемые объекты, можно взять дисперсию компоненты вектора этого признака, ортогональной к множеству остальных векторов-признаков:

$$D(x) = \frac{1}{N} \cdot x_i^T \cdot R \cdot x_i, \quad (8)$$

где  $x$  - значение вектора-признака,  $R$  - оператор-проектор:

$$R = 1 - X^\dagger \cdot X^{+\dagger}, \quad (9)$$

где  $X^{+\dagger}$ - матрица, псевдообратная к матрице  $X^\dagger$  составленной из векторов-столбцов остальных признаков.

А мерой информативности всей совокупности признаков для данной обучающей выборки может служить величина:

$$S = \frac{1}{m} \cdot \text{tr}(X \cdot X^T) , \quad (10)$$

где  $X$  - матрица из всех векторов.

Проведен анализ предметной области с позиций построения системы. Определены требования к организации и функционированию систем, основанные на двух логических схемах:

- логика действия пользователя, который на каждом шаге делает выборы, ведущие по его предположению к достижению глобальной цели;
- логика действий системы, которая на каждом шаге предлагает пользователю список альтернатив, достижимых посредством имеющихся ресурсов.

Соответственно осуществляется разделение компетенции ИС и пользователя, как Лица Принимающего Решение.

Построена концептуальная схема ИС. Для этого разработаны типы задач, определенных проблемной областью, типы задач, связанных с математическим моделированием, и типы системных требований. Сформулированы задачи, составляющие проблемную область:

1. Синтез пространства распознавания:

- а) выбор исходного признакового пространства;
- б) масштабирование;
- в) выбор признаков, используемых для классификации.

2. Построение решающих правил:

- а) линейных дискриминантных функций (ЛДФ);
- б) коллектива ЛДФ (выбор функций для коллектива, обучение и синтез коллектива);

в) комитета ЛДФ (отбор функций для комитета, настройка параметров комитета).

3. Кластерный анализ.

4. Построение линейных моделей:

- а) выбор наилучших линейных регрессоров;
- б) оценка коэффициентов линейных моделей;
- в) проверка адекватности;

Требования к системе определяются необходимостью обеспечить пользователя следующими возможностями: выбором режимов и параметров обработки, прокладкой маршрута на графе обработки, отображением и оценкой качества промежуточных решений по выбранным критериям, дискриминацией ложных решений по формализованным и неформализованным критериям, предоставлением возможности возврата к уже пройденным вершинам графа обработки или оперативного изменения маршрута обработки, выполнением операции обмена с архивом, выбором формы представления конечных результатов.

Для обеспечения системных требований в систему должны быть заложены следующие компоненты: совокупность априорных знаний об изучаемом объекте, система целей и задач, в том числе набор статистических гипотез и критериев их принятия или дискриминации, система выбора стратегий пользователя, сценариев и схем обработки, система статистического анализа, система интерпретации результатов, помогающая делать выводы и формулировать очередные локальные и глобальные цели.

На основании концептуальной модели сформированы функциональные возможности системы.

**Вторая глава** представляет разработанные графы подсистем. В соответствии с концептуальной моделью в каждый из моментов взаимодействия пользователя с системой, ИС находится в определенном информационном состоянии. Сеанс общения пользователя с системой можно представить себе как процесс перехода из одного информационного состояния в другое пока

не будет достигнута глобальная цель пользователя. Объединение множества возможных состояний и возможных путей перехода из одного состояния в другое образуют граф. Последовательность шагов решения конкретной задачи составляет маршрут на этом графе. Стратегия - правила выбора маршрута, позволяющего перейти из исходного информационного состояния в другое, соответствующее поставленной цели.

Последовательность операций, обеспечивающая взаимодействие методологических средств, ресурсов базы данных и базы знаний, и экспертной поддержки для удовлетворения потребностей пользователя в обработке конкретной концептуальной модели, определяется соответствующим графом задачи. Рассматриваются графы подготовки данных, распознавания образов для задач, которые реализованы в данной системе.

**В третьей главе** дано обоснование и изложение алгоритмов задач, используемых в системе. Выбор алгоритмов определялся стремлением к унификации используемого математического аппарата. Решаемые задачи представлены следующими группами:

- 1) предварительная подготовка данных;
- 2) построение решающих правил;
- 3) построение линейных моделей;
- 4) кластерный анализ.

Содержание предварительной обработки заключается в масштабировании и выделении информативных признаков.

1. В основу метода отбора информативных признаков легли: обеспечение максимальной полноты сохранения информации и устойчивость операторов отображения, используемых в процедурах классификации.

Пусть  $X(N \times M)$  - матрица, описывающая  $N$  объектов, охарактеризованных  $M$  признаками, ранг матрицы  $rk X = k$ . Предлагается метод отбора признаков, основанный на формировании матрицы  $X_k$  содержащей выбранные из  $X$   $k$  столбцов, образующих набор максимального объема. Устойчивость обеспечивает полнота ранга матрицы  $X$ . Критерием выбора

очередного вектора  $x$  из матрицы  $X$  является длина  $t$  его компоненты, ортогональной к линейному пространству, натянутому на  $X$ :

$$t = x_{K+1}^T \cdot R_K \cdot x_{K-1} \geq x_i^T \cdot R_K \cdot x_i, i \in I \quad (11)$$

где  $I$  - множество индексов выбираемых столбцов матрицы  $X$ ,  $R$  - оператор-проектор вида:

$$R_K = 1 - X_K \cdot X_K^+ , \quad (12)$$

здесь  $X_K^+$ - матрица, псевдообратная к матрице  $X_K$ ,  $K$  - количество отобранных столбцов. Процесс повторяется пока  $t \geq d$  - некоторого порога, определяемого погрешностью измерения элементов матрицы  $X$ .

Показана возможность поэтапной обработки, когда матрица  $X$  разбивается на  $l_1$  групп  $X_1, X_2, \dots, X_{l_1}$  по  $n_j$  столбцов в каждой группе, где  $j=1, \dots, l_1$ . Процедура выбора применяется к каждой из матриц  $X_j$ .

2. В настоящей работе в качестве решающих правил используется линейные дискриминантные функции, а также их комбинации в коллективах и комитетах. Использование линейных дискриминантных функций в качестве базовых основано на том, что качество распознавания определяется не только величиной функционала эмпирического риска  $I_9$ , но и значением функции  $p$ , которое тем меньше, чем меньше размерность пространства распознавания и "проще" класс решающего правила.

Метод построения ЛДФ опирается на известную матричную формулу:

$$D(Z) = \left( Z - \frac{\bar{X}_1 - \bar{X}_2}{2} \right) \cdot S^{-1} \cdot (\bar{X}_1 - \bar{X}_2)^T , \quad (13)$$

где  $Z$  - вектор значений признаков классифицируемого объекта,  $\bar{X}_1$  - вектор средних значений признаков объектов, принадлежащих к классу 1,  $\bar{X}_2$  - вектор средних значений признаков объектов, принадлежащих к классу 2,  $S$  - ковариационная матрица вида :

$$S = \frac{1}{FN} \cdot (X - \bar{X})^T \cdot (X - \bar{X}), \quad (14)$$

где  $X$  - матрица общих средних значений,  $FN = (N - 2)$  число степеней свободы. В работе доказано, что ковариационная матрица  $S^{-1}$  может быть представлена в виде:

$$S^{-1} = X^{+} \cdot (X^{+})^T \cdot FN, \quad (15)$$

где  $X^{\perp} = X - \bar{X}$ , а  $X^{+}$  - матрица, псевдообратная к  $X^{\perp}$ . На этом основании построен рекуррентный алгоритм построения ЛДФ, сочетающий отбор информативных признаков с использованием операторов-проекторов (12).

С целью повышения качества принимаемого решения при сложных границах между классами, разработаны процедуры объединения построенных ЛДФ в коллективы и комитеты решающих правил.

3. Разработка алгоритма построения моделей основана на утверждении, что близость к минимуму функционала среднего риска  $I_c(a)$  обеспечивается при минимизации  $I_c(a)$  на достаточно узком классе функций. В частности, таким классом функций может быть класс  $F(x, a)$  функций, линейных по параметрам :

$$F(x, a) = a_0 + \sum_{j=1}^k a_j \cdot \varphi_j(x), k = 0, 1, 2, \dots \quad (16)$$

где  $a_j$  - вещественные числа,  $\phi_j$ -вещественные функции,  $j = 1, \dots, k$ . Показано, что для системы наблюдений оценка вектора  $\hat{a} = \{a_0, \dots, a_k\}$ , доставляющего минимум функционалу  $I_c$ , может быть получена в виде:

$$\hat{a} = \Phi^+ \cdot y, \quad (17)$$

где  $\Phi^+$  - матрица, псевдообратная к матрице  $\Phi$  вида:

$$\Phi = \begin{pmatrix} 1 & \varphi_1(x_1) & \dots & \varphi_k(x_1) \\ 1 & \varphi_1(x_2) & \dots & \varphi_k(x_2) \\ \dots & \dots & \dots & \dots \\ 1 & \varphi_1(x_n) & \dots & \varphi_k(x_n) \end{pmatrix}, \quad (18)$$

для которой уже разработаны расчетные формулы. Т. о., в основу всех алгоритмов положена единая вычислительная процедура - операция псевдообращения.

4. В работе предлагается алгоритм кластерного анализа, который переводит процедуру решения многомерной задачи к последовательности одномерных. Утверждается, что если классы линейно разделимы, то метод приводит к выделению таксонов. В основу метода положено две рекуррентно выполняемые операции: выбор осей, вычисление проекций.

На каждом шаге рекурсии выбирается ось, проходящая через два объекта выборки, наиболее удаленных друг от друга в пространстве, перпендикулярном пространству, в котором классификация уже была осуществлена, и вычисляются проекции всех классифицируемых объектов на эту ось. Для построения очередной оси выбирается два объекта: первый  $X_k^1$ -для которого строка признаков имеет наибольшую самоковариацию:

$$X_k^1 = \max_{i \in J} (X_i \cdot X_i^T), \quad (19)$$

где  $J$ - множество индексов, а второй объект  $X_k^2$  - через который проходит ось - это будет объект с наибольшей проекцией на вектор  $OX_k^1$ :

$$X_k^2 = \max_{i \in J} (-X_k^{1+} \cdot X_i), \quad (20)$$

где  $X_k^{1+}$  - столбец, псевдообратный к строке  $X_k^1$ :

$$X_k^{1+} = \frac{X_k^{1T}}{X_k^{1T} \cdot X_k^1}. \quad (21)$$

Проекции всех классифицируемых объектов на ось  $X_k^1 \cdot X_k^2$  вычисляются по формуле:

$$P_j = (X_k^1 - X_k^2)^+ \cdot X_j \quad (22)$$

где;  $j=1, N-2$  - порядковый номер объекта,  $N$ - число объектов. Разделение объектов на кластеры производится на основании сравнения расстояний между проекциями объектов на ось:

$$D_j = P_{j+1} - P_j, \quad (23)$$

где  $j = 1, \dots, N-1$ .

Решение об отнесении данного расстояния к межкластерному принимается на основании значения критерия:

$$P_n(d) = 1 - d \cdot \sum_{i=1}^{[n \cdot (1-d)]} \binom{n}{i} \left(1 - d - \frac{i}{n}\right)^{n-i} \cdot \left(d + \frac{i}{n}\right)^{i-1}, \quad (24)$$

где  $d$  вычисляется по формуле:

$$d = \frac{r_{\min}}{\rho_{\max}} - 1, \quad (25)$$

$r_{\min}$  - это минимальное из рассматриваемых межкластерных расстояний, а  $\rho_{\max}$  - максимальное из внутрикластерных расстояний, попавших в интервал рассмотрения;  $[n(1-d)]$  - целая часть от числа  $n(1-d)$ ,  $n$  - число интервалов, заключенных между рассматриваемым расстоянием и, сначала, ближайшим левым межкластерным расстоянием, а затем, ближайшим правым;  $\binom{n}{i}$  - число сочетаний из  $n$  по  $i$ .  $P(d) = g$ , где  $g$  берется равным 0.9, 0.95, 0.99 или др. в зависимости от желания пользователя. Если вычисленное  $P(d)$  получается меньше заданного значения, то расстояние относится к межкластерным, иначе к внутрикластерным. Рекуррентный процесс заканчивается, когда длина очередной выбранной оси становится меньше некоторого порога.

**Четвертая глава** содержит описание структуры информационной системы. Представлена архитектурная модель, описывается система подготовки данных и формирование базы знаний. Излагается структура интегрированной среды. Описывается комплекс программ системы и экспертная поддержка. Приводится содержательная интерпретация процесса достижения целей задач подготовки данных и распознавания образов.

**В пятой главе** приведены результаты исследования с помощью разработанной системы данных клиноортостатической пробы (КОП) группы новорожденных. КОП заключается в измерении показателей сердечнососудистой деятельности человека в трех состояниях: покоя, воздействия гравитационной нагрузки типа "ступенька" и снятия этой нагрузки. В дальнейшем для краткости будем называть эти три состояния: "покой", "нагрузка", "сброс нагрузки". Ставились следующие задачи:

1. Сформировать совокупность признаков для проведения классификации.
2. Определить их сравнительную информативность.
3. Разработать методику отнесения классифицируемых объектов к определенному классу по состояниям: здоров, средней тяжести, тяжелый, крайне тяжелый.
4. Установить возможность прогноза состояния объекта.

На основании анализа исходных данных произведена классификация объектов в соответствии с медицинскими показателями по: степени доношенности; апгару; биологическому анамнезу матери; акушерскому анамнезу; гинекологическому анамнезу; течению беременности; родам; диагнозу ребенка.

Проведен анализ статистической значимости изменения различных показателей при проведении КОП, и показано, что только изменения частоты сердечных сокращений (ЧСС) являются статистически значимыми.

Сформирован набор из 12 признаков, характеризующих объект в состоянии "покоя", а также отражающих изменение состояния объекта при "нагрузке" и "сбросе нагрузки":

- 1 - среднее значение систолического артериального давления (САД) в "покое";
- 2 - среднее значение диастолического артериального давления (ДАД) в "покое";
- 3 - среднее значение давления (СРАД) в "покое";
- 4 - среднее значение частоты сердечных сокращений (ЧСС) в "покое";
- 5 - среднее значение показателя Кердо в "покое";
- 6 - среднее значение отклонения ЧСС при "нагрузке" от среднего значения ЧСС "в покое";
- 7 - тангенс угла наклона прямой, аппроксимирующей кривую изменения ЧСС при "нагрузке";

- 8 - дисперсия ЧСС при "нагрузке";
- 9 - среднее значение отклонения ЧСС при "сбросе нагрузки" от среднего ЧСС в "покое";
- 10 - тангенс угла наклона прямой, аппроксимирующей кривую изменения ЧСС при "сбросе нагрузки";
- 11 - дисперсия ЧСС при "сбросе нагрузки";
- 12 - разница между средним ЧСС и последним его значением при "сбросе".

Проведен анализ взаимной коррелированности и с помощью алгоритма отбора признаков, описанного в главе 3, выделен набор из 8 информативных признаков с номерами: 5, 10, 8, 4, 6, 1, 3, 11, 12, из которых первые 4 признака дают 80% вклада в общую изменчивость.

Было проведено сопоставление изменения показателей состояния для объектов с синдромами общего угнетения и возбуждения, доношенных и недоношенных и показано, что наличие возбуждения, угнетения, а также фактор доношенности не являются значимыми для формирования класса.

Были построены решающие правила для отнесения классифицируемых объектов к классам: "здоров", "средней тяжести", "тяжелый", "крайне тяжелый". Качество распознавания на обучающей выборке по построенным ЛДФ не хуже, чем 90%.

Для изучения вопроса о разделении анализируемых объектов на классы на основании формальных методов проведена процедура таксономии. В результате чего выделены классы, объединяющие объекты по тяжести состояния, что хорошо согласуется с результатами классификации по обучающей выборке.

Для выявления взаимосвязи между объектами, а также вычисляемым параметрам, были построены различные линейные модели, которые могут быть в дальнейшем интерпретированы специалистами, как причинно-следственные.

Для оценки изменения измеряемых показателей, используемых для классификации, была проверена гипотеза, что реакция ребенка в произвольный день может быть представлена в виде линейной комбинации такой реакции в некоторые фиксированные - информативные дни жизни.

Оказалось, что для здоровых детей таким информативным днем является один день и он приходится на 3, 4 или 5 дни жизни, а для больных детей соответственно количество информативных дней равно трем. Судя по полученным результатам, можно оценить динамику изменения состояния объекта по дням жизни относительно фиксированных дней, что представляется очень перспективным для изучения закономерности развития, как здоровых детей, так и прогнозирования состояния здоровья детей с различными патологиями.

Приведены графики, отображающие изменения измеряемых показателей по выбранному базовому дню для "здоровых" объектов, объектов с состоянием "средней тяжести" с состояниями "возбуждения" и "угнетения".

## **ВЫВОДЫ**

Основным результатом диссертационной работы является разработка макетного варианта интегрированной интеллектуальной системы, использующей комплекс алгоритмов решения задач восстановления зависимостей. Сформулируем основные выводы.

1. Определен перечень задач, составляющих предметную область разработанной интеллектуальной системы. Показано единство математической постановки задач распознавания образов и оценивания регрессий, что позволило унифицировать алгоритмы выбранного перечня задач.

2. На основании семантического анализа и анализа структуры предметной области определены требования к функциональным возможностям системы, разработана концептуальная модель интеллектуальной системы и макетный вариант интеллектуальной системы, позволяющей решать широкий набор задач классификации.

3. На базе концептуальной модели разработаны графы решения задач классификации. Показано, что решение конкретной задачи может быть представлено в виде маршрута, прокладываемого на графе. При решении конкретной задачи Пользователь, выбирая в ответ на предложения системы соответствующие его выбору альтернативы, прокладывает маршрут, на общем графе решения задач исходя из имеющихся ресурсов, значений оценочных функций, специфики полученных результатов.

4. Разработана функциональная структура интеллектуальной системы.

5. Для решения задач, составляющих предметную область системы, разработаны алгоритмы и программы, позволяющие решать широкий набор задач классификации: синтеза признакового пространства, построения линейных дискриминантных функций, построения линейных моделей, кластер-анализа, формирования коллективных методов на основе линейных дискриминантных функций. В основу алгоритмов решения различных задач положен единый подход, основанный на рекуррентном псевдообращении и построении операторов-проекторов.

6. Разработан алгоритм решения задачи распознавания образов, основанный на рекуррентной процедуре, сочетающей отбор информативных признаков и построения соответствующей линейной дискриминантной функции.

7. Разработан механизм включения простых решающих правил типа линейных дискриминантных функций в более сложные структуры типа "коллективы" и "комитеты" решающих правил.

8. Разработан метод кластеризации, который сводит решение задачи многомерной таксономии к последовательности одномерных задач.

9. Разработан алгоритм выбора наилучших линейных регрессоров для построения линейных моделей.

10. Разработаны архитектура и структура информационной системы, включающие подсистемы подготовки БД и БЗ, блок управления, блок экспертной поддержки и методологические средства.

11. Работоспособность разработанных алгоритмов и системы в целом демонстрируется на примерах решения задач классификации в областях геохимии, химии и медицине.

12. Приведены результаты обработки данных клиноортостатической пробы группы новорожденных.

Основное содержание диссертации опубликовано в следующих работах:

1. Берестнева О.Г., Васильев Н.В., Кочегуров В.А., Константинова Л.И., Берестнева О.Г., Удут И.В., Шкатова Г.И. Системный анализ и разработка методологии исследования процессов адаптации биосистем к измененным климато-географическим условиям. // Проблемы солнечно-биосферных связей. - Новосибирск., 1982. С. 124-126.

2. Берестнева О. Г., Кочегуров В. А., Константинова, Пеккер Я. С., Шумилов Б. М., Шкатова Г. И. Алгоритмы автоматизированной обработки электрокардиосигналов при массовых обследованиях рабочих промышленных предприятий в условиях Крайнего Севера// Проблемы создания технических средств для диагностики и лечения сердечнососудистой системы; Тез. докл. Всесоюз. конф. - Москва, 1983. С.51-52.

3. Берестнева О. Г., Иглакова Е. Ф. , Шкатова Г. И. , Шумилов Б. М. Программное обеспечение автоматизированного комплекса для массовых профилактических осмотров населения. //Автоматизированные электронные системы для массовых профилактических осмотров населения. -Л. ,1985, С. 62-63.

4. Берестнева О.Г., Шкатова Г.И., Шумилов Б.М. Алгоритмы автоматизированной обработки электрофизиологических сигналов при массовых обследованиях. // Применение математических методов обработки ме-

дико-биологических данных в ЭВМ и медицинской технике. -М., 1984, С. 65-67.

5. Берестнева О.Г., Пеккер Я. С., Шкатова Г.И. Программное обеспечение автоматизированного комплекса для массовых обследований. // Вопросы медицинской электроники., -Таганрог., 1986, п. -6, - С. 26-27.

6. Берестнева О.Г., Кочегуров В.А., Константинова Л.И., Удут И.В., Шкатова Г.И. Математическое моделирование процессов адаптации на разных уровнях организации биосистем // Отчет. / Томский политехнический институт, п гр 78045087, -Томск. ,1981, 97с.

7. Берестнева О. Г., Кочегуров В. А., Константинова Л. И., Шкатова Г. И., Шумилов Б.М. Создание аппарата для автоматизированной обработки медико-биологических данных при массовых обследованиях рабочих промышленных предприятий в условиях Крайнего Севера. //Отчет. /Томский политехнический институт, игр 0182. 2018779,-Томск., 1982, 79с.

8. Вылегжанин О.Н., Шкатова Г.И. Выбор информативных признаков описания объектов для задачи распознавания образов // Математические методы распознавания образов: Тез. докл. НИ Всесоюз. конф. 10-11 ноября 1987.- Львов, 1987.4.1, С.139-140.

9. Вылегжанин О.Н. , Шкатова Г.И. Сравнительная оценка двух методов выбора наилучших линейных регрессоров// Применение математических методов и ЭВМ в медико - биологических исследованиях. - Томск: Изд-во Томск, политехн. ин-та, 1988. С. 18-22.

10. Вылегжанин О.Н., Шкатова Г.И. Отбор признаков для решения задачи распознавания образов по данным физиологических измерений. Томск, 1989. Деп. в ВИНТИ 21.03.89 \_ 1747-В89.

11. Вылегжанин О.Н., Шкатова Г.И. Построение линейных дискриминантных функций с использованием псевдообратной матрицы. Томск, 1990. Деп. в ВИНТИ

12. Вылегжанин О.Н., Шкатова Г.И. Рекуррентный метод построения линейной дискриминанты о и функции, минимизирующий верхнюю

оценку функционала среднего риска // Математические методы распознавания образов: Тез.докл.У Всесоюз. конф. ноябрь 1991. -Москва, 1991. Ч.1.С.25-26.

13. Вылегжанин О.Н., Шкатова Г.И., Пеккер Я.С., Рауш А.Я. "РАСКОЛ" - пакет прикладных программ для распознавания образов //Математические методы распознавания образов: Тез. докл. V Всесоюз. конф. ноябрь 1991. - Москва, 1991. Ч.1.С.27.

14. Вылегжанин О.Н., Шкатова Г.И. Алгоритмы построения линейных моделей//Химическая информатика: Тез. докл. всесоюз. конф. 1992,-Черноголовка, Ч.1.С.290.

15. Вылегжанин О.Н., Шкатова Г.И. Поэтапный отбор признаков для решения задачи распознавания образов //Математическая теория и методология распознавания: Тез. докл. IV Всесоюз. конф. 1989,-Рига, 1989. Ч.1.С.31-32.

16. Шумилов Б.М., Шкатова Г.И. Алгоритмы и программы автоматизированной обработки кардиологической информации при массовых обследованиях // Вопросы разработки и внедрения радиоэлектронных средств при диагностики сердечно-сосудистых заболеваний. -М.: Радио и связь, 1987. С. 139-140.