

УДК 004.89:004.4:681.5

На правах рукописи

A 2009
1959



САМИГУЛИНА ГАЛИНА АХМЕТОВНА

**Разработка интеллектуальных экспертных систем управления на основе
подхода искусственных иммунных систем**

05.13.18 – Математические модели, численные методы,
комплексы программ

Автореферат

**диссертации на соискание ученой степени
доктора технических наук**

Республика Казахстан
Алматы, 2009

Работа выполнена в Институте проблем информатики и управления
Министерства образования и науки Республики Казахстан

Научные консультанты: доктор физико-математических наук, профессор
Калимолдаев М.Н.

доктор физико-математических наук,
профессор Арсланов М.З.

Официальные оппоненты: доктор технических наук,
Макаренко Н.Г.

доктор технических наук,
профессор Шмыгалева Т.А.

доктор технических наук,
профессор Мурзабеков З.Н.

Ведущая организация: Казахстанско-Британский технический
университет

Защита состоится «26» июня 2009 года в «15» часов на заседании
диссертационного совета ОД 53.12.01 Института математики Министерства
образования и науки Республики Казахстан по адресу: Республика Казахстан,
050010, г. Алматы, ул. Джамбула 25, 2 этаж, конференц-зал.

С диссертацией можно ознакомиться в библиотеке Института математики
Министерства образования и науки Республики Казахстан.

Автореферат разослан « ___ » мая 2009 года

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҰЛТТЫҚ АКАДЕМИЯЛЫҚ КІТАПХАНАСЫ
№ _____

Ученый секретарь
диссертационного совета ОД 53.12.01
доктор технических наук, профессор

 А.Н.Казангапов

ВВЕДЕНИЕ

Общая характеристика работы. Одной из основных задач Искусственного Интеллекта (ИИ) является проблема моделирования биологических принципов обработки информации. Процессы, происходящие при обработке информации естественными системами, и принципы их функционирования поражают своей эффективностью, экономичностью, быстродействием и адаптационными возможностями. В настоящее время наиболее развитыми считаются генетические алгоритмы, искусственные нейронные сети (нейрокомпьютеры), клеточные автоматы и искусственные иммунные системы (иммунокомпьютеры).

Интерес к данному виду исследований постоянно возрастает. Прежде всего, притягивает способность естественных биологических систем решать многомерные задачи огромной вычислительной сложности в реальном времени. Высокопараллельные механизмы функционирования позволяют биологическим системам с легкостью справляться с задачами, которые невозможно решить с помощью последовательных ЭВМ. Решение четырехмерных задач в реальном времени, не говоря уже о многомерных нелинейных динамических задачах, неподвластно современным последовательным машинам.

Особенный интерес представляют Искусственные Иммунные Системы (ИИС). Это относительно новая область исследования. Первые попытки разработки ИИС были предприняты в 70-х годах, масштабные работы начались в 90-х годах прошлого столетия. В работах А.О. Тараканова были разработаны математические основы биологического подхода ИИС. Базовым элементом системы является пептид (белок). Биологическим прототипом ИИС является иммунная система человека и обработка информации молекулами белков на основе результатов самосборки. В основе подхода ИИС лежит идея взаимодействия между белками иммунной системы человека и чужеродными антигенами, т.е. возможность произвольного связывания (молекулярного узнавания), посредством определения минимальной энергии связи между формальными пептидами. Искусственная иммунная система обладает основными свойствами искусственного интеллекта - памятью, способностью к обучению и принятию решений в незнакомой ситуации.

Актуальность темы исследования. Подход ИИС интенсивно развивается в различных направлениях искусственного интеллекта. Большие надежды с ИИС связаны в области организации защиты от компьютерных вирусов, т.к. на сегодняшний день традиционные методы защиты уже недостаточно эффективны. Способность иммунной системы человека безошибочно распознавать чужеродный вирус может дать ключ к созданию высокоэффективных систем защиты. Актуальность данного направления подтверждают исследования американской корпорации RAND (<http://www.computerra.ru/offline/2002/430/15920/>), которая еще в 1999 году провела исследования относительно рекомендаций по выбору технологий, способных обеспечить необходимый уровень информационной безопасности

национальной информационной и оборонительной инфраструктуры Министерства обороны США. В качестве самых перспективных технологий по созданию систем защиты информации нового поколения были выбраны именно технологии формирования искусственной иммунной системы.

Действующий макет иммунной системы киберпространства (Immune System for cyberspace, ISC) был продемонстрирован компанией IBM в 1997 году на конференции в Сан-Франциско.

Почтовая служба Англии (<http://www.icsa.ac.uk/cifd>) планирует использовать в своей работе системы защиты для Internet на базе технологии ИИС (проект Computational Immunology for Fraud Detection, CIFD). Данный проект разрабатывается учеными Лондонского Королевского колледжа.

Приведенные примеры подтверждают актуальность данного направления ИИ. Диссертация посвящена разработке интеллектуальных экспертных систем управления и прогнозирования на основе ИИС для решения различных прикладных задач. Существует ряд проблем при реализации этих интеллектуальных систем. Основная трудность заключается в создании алгоритмов безошибочного распознавания образов, т.к. ошибки энергетических оценок не позволяют добиться 100% распознавания. Так же как и в искусственных нейронных сетях существует проблема создания эффективных и простых методик обучения ИИС. Необходимо уметь выделять из множества факторов самые главные, которые оказывают наибольшее влияние на процесс обработки информации, выделять информативные признаки, обучать ИИС и оценивать процесс обучения. Проблема сильно усложняется с ростом размерности системы и увеличением времени на ее обучение.

Помимо этого, ИИС в условиях реального времени для прогнозирования поведения и управления нестационарными, нелинейными, многомерными объектами с различными видами неопределенности параметров к настоящему времени практически не разработаны.

Таким образом, все вышеизложенное определяет актуальность и перспективность диссертационного исследования.

Цель исследования: построение интеллектуальных экспертных систем поддержки принятия решений, прогнозирования и управление на основе ИИС для решения различных прикладных задач.

Основные задачи исследования:

1. Анализ состояния современных проблем и основных направлений в области искусственного интеллекта и интеллектуальных систем управления. Введение в основы иммунологии, термодинамической теории строения белка.

2. Разработка интеллектуальной информационной технологии обработки, анализа и прогноза многомерных данных искусственными иммунными сетями.

3. Прогноз и анализ динамических процессов сложных нелинейных систем управления с различными видами неопределенностей (стохастических, интервально-заданных) на основе ИИС.

4. Формальное определение языка интеллектуальной технологии обработки данных на основе ИИС.

5. Разработка интеллектуальной экспертной системы дистанционного обучения в среде Internet на основе ИИС.

6. Разработка интеллектуальной информационной экспертной системы биометрического контроля доступа по рукописной подписи на основе ИИС.

7. Разработка интеллектуальной экспертной системы прогнозирования эпизоотий чумы по характеристикам штаммов возбудителя чумы.

8. Разработка пакета программ «PERTID» по обработке, анализу и прогнозу многомерных данных на основе ИИС.

Методы исследования. В диссертационной работе при решении поставленных задач были использованы: теория матриц; теория автоматического управления; основы интервального анализа; теория распознавания образов; методы статистической обработки данных, неравновесная термодинамическая теория белка; теоретические основы защиты информации и биометрии.

Научная новизна и основные положения выносимые на защиту. В диссертационной работе впервые получены следующие результаты:

1. Разработан эффективный алгоритм обучения ИИС на основе построения оптимальной структуры иммунной сети по весовым коэффициентам выделенных информативных признаков и предварительной обработке многомерных данных ИИС с помощью методов факторного анализа.

2. Разработана процедура оценки энергетических погрешностей ИИС на основе свойств гомологичных белков при решении задачи распознавания образов. Осуществлен расчет коэффициентов риска прогнозирования на основе Z - факторов.

3. Решена задача классификации решений, необходимая при распознавании образов на основе ИИС для квазирасщепленной системы управления.

4. Разработана интеллектуальная экспертная система прогнозирования асимптотической устойчивости в среднем квадратическом стохастической системы управления на основе ИИС.

5. Получен аналог метода квазирасщепления для интервально - заданных систем управления. Разработана процедура построения операторов проектирования для интервально - заданной системы управления. Разработана математическая модель нелинейной интервально-заданной системы управления в виде квазирасщепленных подсистем. Разработана интеллектуальная экспертная система прогнозирования асимптотической устойчивости интервальной системы управления по квазирасщепленным подсистемам на основе ИИС.

6. Осуществлена группировка технологических операций при обработке информации иммунными системами. Описана технология обработки информации иммунными системами в виде формального представления нетерминального алфавита или словаря.

7. Разработаны теоретические основы, структура, алгоритмы, программное обеспечение построения ИИС дистанционного Internet – образования.

8. Разработаны теоретические основы, структура и вычислительные алгоритмы аутентификации рукописной подписи на основе биометрических технологий и подхода ИИС.

9. Разработаны теоретические основы, структура и вычислительные алгоритмы экспертной системы прогнозирования поведения чумной эпизоотической триады по выделенным информативным признакам штаммов возбудителя чумы на основе иммунных сетей.

10. Разработано программное обеспечение и получены свидетельства о государственной регистрации объекта интеллектуальной собственности на программы ЭВМ.

Обоснование и достоверность научных результатов и выводов. Все научные положения и выводы подтверждаются результатами компьютерного моделирования, их согласованностью с экспериментальными данными и экспертными оценками.

Практическая ценность и внедрение результатов работы. Разработанная интеллектуальная информационная технология обработки информации искусственными иммунными сетями может быть эффективно использована для решения широкого круга прикладных задач. Основным достоинством предложенной технологии является оценка и уменьшение энергетических погрешностей (ошибок обобщения) при обработке информации иммунными сетями, что существенно влияет на достоверность прогноза. Создание стандартных модулей оптимальных последовательностей технологических операций при обработке информации иммунными системами для различных приложений при помощи разработанного формального языка позволяет автоматизировать процесс выбора технологических цепочек и может быть применен при автоматизированном написании программ по обработке данных на основе ИИС.

Получен акт о внедрении научных результатов в учебном процессе Центрально - Азиатского Университета на кафедре «Вычислительная Техника и Информационные Технологии». Результаты исследования и разработок, включая авторские инструментальные средства, используются на лекционных, практических и лабораторных занятиях по дисциплинам кафедры: спец. курс «Искусственные Иммунные Системы», «Новые информационные технологии», «Экспертные и интеллектуальные системы», «Модели и методы управления», «Основы информационных систем», «Информационная безопасность и защита информации», «Основы компьютерного моделирования». Результаты работы использованы в учебно-методических комплексах дисциплин, курсах, в рабочих программах по вышеперечисленным дисциплинам.

Получен акт о внедрении результатов исследований в АО «Казкоммерцбанк» при анализе и прогнозе пиковых нагрузок, возникающих в платежных системах при обработке переводных операций через Казахстанский Центр Межбанковских Расчетов и по прямым корреспондентским счетам. Результаты проведенного анализа и прогноза пиковых нагрузок были использованы для эффективного распределения нагрузок между каналами

переводных операций и определения участков дальнейшей автоматизации обработки переводных операций.

Получен акт об использовании результатов исследований в АО «Банк Центр Кредит» для прогнозирования рисков максимальных отклонений рыночных показателей.

Применение результатов диссертационной работы подтверждено апробацией на Международном координационном совещании в г. Алматы по линии гранта INCO COPERNICUS «STEPICA» № ICA2-2000-10048 «Чума в Центральной Азии – эпидемиологическое исследование, основанное на пространственно - временной динамике».

Апробация работы. Основные положения и результаты работы докладывались и обсуждались на VI Международной конференции женщин-математиков «Математика. Экономика. Образование.» (Чебоксары, 1998); на XIV Международной конференции по интервальной математике (Новосибирск, 1998); на Международной научно-практической конференции «Проблемы вычислительной математики и информационных технологий» в КазНУ им. Аль-Фараби (Алматы, 1999); на VII Международной конференции женщин-математиков «Математика. Экономика. Образование.» (Ростов-на-Дону, 1999); on the Third International Conference «Differential equations and application» (Sankt-Petersburg, 2000); на Международной конференции «Проблемы управления и информатики» (Бишкек, 2000); on the International conference on systems science (Wroclaw, Poland, 2001); на международной конференции по управлению «Автоматика-2001» (Одесса, 2001); на II Всесибирском конгрессе женщин-математиков (Красноярск, 2002); на X Международной конференции женщин-математиков «Математика. Экономика. Образование». (Ростов-на-Дону, 2002); на III Международной научно-практической конференции «Энергетика, телекоммуникации и высшее образование в современных условиях» в Алматинском институте энергетике и связи (Алматы, 2002); на Международной конференции «Вычислительные технологии и математические модели в науке, технике и образовании» ВТММ-2002, которая проводилась совместно с Институтом вычислительных технологий СОРАН и КазНУ им. Аль-Фараби (Алматы, 2002); на Международной конференции «Информационные технологии и автоматизация производственных процессов» в КазНТУ им. К.И.Сатпаева (Алматы, 2002); на VIII Санкт-Петербургской Международной конференции «Региональная информатика-2002» (Санкт-Петербург, 2002); на III Международной конференции, посвященной 15-летию Международной Академии Бизнеса «Бизнес и образование: вектор развития» (Алматы, 2003); на Международной конференции по дифференциальным уравнениям, посвященной 100-летию юбилею К.П. Персидского в институте математики МОН РК (Алматы, 2003); на Международном конгрессе «Математика в XXI веке» (Новосибирск, 2003); на Международной конференции «Вычислительные информационные технологии в науке, технике и образовании» ВИТ-2003, которая проводилась совместно с Институтом вычислительных технологий СОРАН и КазНУ им. Аль-Фараби (Усть-Каменогорск, 2003); на Международной конференции «Этапы становления, современное состояние и фундаментальные

проблемы развития образования и науки Казахстана» (Караганда, Центрально-Казахстанское отделение НАН РК, 2003), on the sixth International Conference on Computational Intelligence and Natural Computation - CINC2003, (Cary, North Carolina, USA, 2003), на III Всесибирском конгрессе женщин-математиков (Красноярск, 2004); на Международной конференции «Инженерное образование и наука в XXI веке» (Алматы, КазНТУ, 2004); на республиканской научно-практической конференции «Роль науки и интеллектуальной собственности в реализации Стратегии индустриально-инновационного развития страны на 2003-2015 годы» (Астана, Евразийский национальный университет имени Л. Гумилева, 2004); на 10 Международной конференции по математике и механике (Алматы, КазНУ им.Аль-Фараби, 2004); on the International Conference «New Trends in the Computer Science Master's Curriculum» (Almaty, Kazakh National University, 2004), на IV Международной конференции «Бизнес и образование: вектор развития» Международная Академия Бизнеса (Алматы, 2004); на Международной конференции «Дифференциальные уравнения с частными производными и родственные проблемы анализа и информатики» (Ташкент, 2004); на XII Всероссийском семинаре «Нейроинформатика и ее приложения» (Красноярск, 2004); на Международной конференции «Информационные и телекоммуникационные технологии в науке, технике и образовании» ВИТ-2004, которая проводилась совместно с ИВТ СО РАН и КазНУ (Алматы, КазНУ, 2004); на Международной конференции «Инфокоммуникационные и вычислительные технологии в науке, технике и образовании» (Ташкент, 2004); на 4 Международной научно-технической конференции «Энергетика, телекоммуникации и высшее образование в современных условиях» (Алматы, АИЭИС, 2004); на 5 Международной научно-практической конференции «Методы и алгоритмы прикладной математики в технике, медицине и экономике» (Новочеркасск, 2005); на V Международной научно-практической конференции: «Бизнес и образование: вектор развития», Международная Академия Бизнеса (Алматы, 2005); на X Международной конференции «Математика. Экономика. Образование». (Ростов-на-Дону, 2005); на научно-практической конференции «Роль физико-математических наук в современном образовательном пространстве» (Атырау, 2005); на VIII Всероссийской конференции «Современные методы математического моделирования природных и антропогенных катастроф» (Кемерово, 2005); на XIV Международной конференции «Математика. Экономика. Образование.» и IV международном Симпозиуме «Ряды Фурье и их приложения» (Ростов-на-Дону, 2006); на IV Всесибирском конгрессе женщин-математиков. (Красноярск, 2006г.); на XIV Всероссийском семинаре «Нейроинформатика и ее приложения» (Красноярск, 2006); на Международной конференции «Вычислительные и информационные технологии в науке, технике и образовании. ВИТ-2006», которая проводилась совместно с Институтом вычислительных технологий СОРАН и Павлодарским Государственным Университетом им. С. Торайгырова (Павлодар, 2006); на IX Международной конференции «Интеллектуальные системы и компьютерные науки» (Москва, МГУ им. Ломоносова, 2006); на Международной конференции «Современные проблемы дифференциальных

уравнений, теории операторов и космических технологий» (Алматы, ЦФМИ, 2006); на Международной конференции «Дифференциальные уравнения, проблемы анализа и алгебры» (Актюбинск, 2006), на Международной научно-технической конференции «Компьютерное моделирование и интеллектуальные системы» (Украина, Запорожье, 2007), на Международной конференции «Дифференциальные уравнения, теория функций и приложения» (Новосибирск, 2007), на XV Всероссийском семинаре «Нейроинформатика, ее приложения и анализ данных» (Красноярск, 2007); на IX Всероссийской конференции «Современные методы математического моделирования природных и антропогенных катастроф» (Барнаул, 2007); на X Всесибирском семинаре «Моделирование неравновесных систем(MNS-07)» (Красноярск, 2007); на 2 Международной конференции «Проблемы управления и информатики» (Бишкек, 2007); на Международной конференции «Информационно-коммуникационные технологии как основной фактор развития инновационного общества» (Усть-Каменогорск, ВКГТУ им. Серикбаева, 2007); на Международной конференции «Новые направления в теории динамических систем и некорректных задач» (Самарканд, 2007, <http://www.conf2007.samdu.uz>); на V Всесибирском конгрессе женщин - математиков (Красноярск, 2008, <http://www.sfukras.ru/science/congress/>); на XVI Международной конференции «Математика. Экономика. Образование». (Ростов-на-Дону, 2008); на Международной конференции «Вычислительные и информационные технологии в науке и образовании» (Новосибирск ИВТ СОРАН - Алматы КазНУ, 2008); на X Международной конференции «Проблемы эволюции открытых систем». (Алматы, 2008); на XI Всероссийском семинаре «Моделирование неравновесных систем(MNS-08)» (Красноярск, 2008); на XV Всероссийском семинаре «Нейроинформатика, ее приложения и анализ данных» (Красноярск, 2008); на Международной конференции «Современные проблемы математического моделирования и вычислительных технологий». (Красноярск, 2008, <http://math.institute.sfu-kras.ru/minct2008/>); на Международной конференции «Современные проблемы математики, информатики и управления» (Алматы, 2008); на Международной научно-практической конференции «Информационные технологии в науке и образовании» (Всемирный технологический университет ЮНЕСКО, 2009, http://www.agore.guru.ru/itno_moodle_09/); на XIV Международной конференции «Современные проблемы информатизации» (Воронеж, 2009).

Публикации. По теме диссертации опубликовано более 80 печатных работ, в том числе 2 монографии. Получено 2 авторских свидетельства.

Структура и объем диссертации. Диссертация состоит из введения и шести разделов, содержит 250 страниц, 87 рисунков и 14 таблиц, список используемых источников из 268 наименований и приложения.

ОСНОВНАЯ ЧАСТЬ

Во введении обосновывается актуальность диссертационной работы, изложены цель и задачи исследования, отмечена научная новизна и

практическая ценность работы, кратко представлена структура диссертационной работы.

Первый раздел посвящен современным проблемам и обзору основных направлений ИИ, основанных на моделировании естественных феноменов. В рамках биологического и логико-лингвистического направления рассмотрены наиболее распространенные системы искусственного интеллекта (генетические алгоритмы, нейронные сети и т.д.). Дана краткая характеристика биологических прототипов и их базовых элементов, которые определяют свойства данных систем. Представлены сферы применения, достоинства и основные недостатки этих систем ИИ.

Представлены теоретические и математические основы искусственных иммунных систем на основе алгебры кватернионов. Даны определения основных понятий ИИС.

Под искусственными иммунными системами понимаются информационные методологии, использующие понятия теоретической иммунологии для решения прикладных задач. Иммунные сети (ИС) - это адаптивные системы для обработки и анализа данных, которые представляют собой математическую структуру, имитирующую некоторые функции иммунной системы человека и обладающую свойствами, такими как способность к обучению, к прогнозированию на основе уже имеющихся временных рядов и принятию решения в незнакомой ситуации. ИС в принципе не нуждаются в заранее известной модели, а строят ее сами на основе полученной информации в виде временных рядов. Данные системы применяются при решении плохо алгоритмизируемых задач, таких как прогнозирование, классификация и управление.

Тараканов А.О. ввел математическую модель формального пептида (пептид это простейший белок, который содержит до 50 аминокислот).

Формальным пептидом (ФП) называют упорядоченную пятерку:

$$P = \langle n, U, Q, V, \nu \rangle,$$

которая включает следующие компоненты:

- количество звеньев $n > 0$;
- множество торсионных углов: $U = \{ \varphi_k, \psi_k \}, k=1, \dots, n,$

где $-\pi \leq \varphi_k \leq \pi, -\pi \leq \psi_k \leq \pi$;

- множество единичных кватернионов: $Q = \{ Q_0, Q_k \},$

где кватернионы $Q_k = Q_k(\varphi_k, \psi_k)$ и результирующий кватернион ФП Q_0 определяется как их произведение: $Q_0 = Q_1 Q_2 \dots Q_n$;

- множество коэффициентов $V = \{ v_{ij} \}, i = 1, 2, 3, 4, j \geq i$;
- функция ν (без индекса), определенная на элементах результирующего кватерниона Q_0 следующей квадратичной формой: $\nu = -\sum_{j \geq i} v_{ij} q_i q_j$.

Рассматриваются взаимодействия между белками иммунной системы человека и чужеродными антигенами, процедура молекулярного узнавания посредством определения минимальной энергии связи между формальными пептидами. Достоинствами иммунной системы являются: распределенность;

самоорганизация и эволюция; относительная простота и «легковесность», т.е. ИИС не особенно требовательна к вычислительным ресурсам; отсутствие централизованного контроля; самообучаемость; индивидуальный подход к уникальным событиям.

Разработана интеллектуальная технология обработки, анализа, прогноза данных иммунными сетями, которая нацелена на уменьшение погрешностей энергетических оценок (ошибок обобщения), повышение достоверности прогноза. Данная технология состоит из двух ключевых моментов:

- предварительной обработки данных на основе факторного анализа (метода главных компонент), позволяющей глубоко проанализировать скрытые (латентные) взаимодействия признаков, основополагающие факторы, влияющие на процесс; произвести редукцию малоинформативных признаков, несущих существенные погрешности; создание оптимальной структуры иммунной сети и сокращение времени на ее обучение;

- устранение погрешностей энергетических оценок ИИС при решении задачи распознавания образов на основе свойств гомологичных белков.

Успешное развитие ИИС напрямую зависит от создания алгоритмов, позволяющих обеспечить наиболее эффективное обучение иммунной сети за минимально короткое время. Методы предварительной обработки данных и факторный анализ применяются с целью извлечения наибольшей информации из исходных данных путем устранения избыточности.

На сегодняшний день получено достаточно результатов по применению искусственных иммунных сетей в различных областях, однако недостаточно исследованы вопросы выбора структуры иммунной сети. Тем не менее разработчики постоянно сталкиваются с данной проблемой при решении практических задач. Включение в модель признаков, мало влияющих на выходной параметр, затрудняет ее использование, ведет к избыточности информации и возрастанию вычислительных ресурсов. Все это приводит к ситуации, когда прогнозирование по избыточной модели менее эффективно, чем по модели из оптимального числа признаков, обладающих максимальной информативностью. Поэтому при разработке интеллектуальной информационной технологии по анализу и прогнозированию данных на основе ИИС в реальном масштабе времени особенно актуальной является задача оптимизации иммунной сети.

Введем некоторые определения, необходимые для дальнейшего изложения.

Определение 1. Под иммунной сетью связывания в подходе ИИС будем понимать любую последовательность связываний формальных пептидов.

Определение 2. Назовем гомологичным такое связывание двух n -пептидов, при котором последовательно связываются их звенья с одинаковыми номерами.

Определение 3. Под оптимальной структурой иммунной сети понимается сеть, построенная на основе весовых коэффициентов выделенных информативных признаков, наиболее полно характеризующих состояние рассматриваемой системы в зависимости от основополагающих факторов, влияющих на функционирование этой системы. Критерием является

максимальное сохранение информации при минимальном количестве признаков.

Пусть исходная совокупность данных записана в виде матрицы $A = (a_{ij}), i, j = 1, \dots, n$ размерности $(m \times n)$. Так как признаки, характеризующие состояние системы измеряются в разных единицах, то результат может существенно зависеть от выбора масштаба измерения. Поэтому необходим переход к безразмерным величинам с помощью нормирования и центрирования данных. Для этого преобразуем элементы каждого вектора, таким образом, чтобы математическое ожидание было равно нулю, а дисперсия единице.

Основной целью нормирования данных является приведение всех данных к сопоставимому виду.

Новая матрица стандартизированных переменных X записывается из элементов:

$$x_{ij} = \frac{x_{ij} - m_j}{s_j}, \quad (1)$$

где m_j - среднее значение исходных элементов j -го вектора, s_j - стандартное отклонение исходных элементов j -го вектора, которое вычисляется по

формуле:

$$s_j = \left(\frac{1}{N-1} \sum_{i=1}^n (x_{ij} - m_j)^2 \right)^{\frac{1}{2}}, \quad (2)$$

где n - общее число измерений.

Задача снижения размерности анализируемого признакового пространства и отбор наиболее информативных признаков решена с помощью метода главных компонент на основе вращения собственного вектора.

Определим базисное пространство R и проекции векторов данных на каждую из n - ортогональных осей, тогда исходную матрицу данных A , размерности $(m \times n)$ можно представить в матричной форме в следующем виде:

$$A = CV^T, \quad (3)$$

где V - матрица, столбцы которой ортогональные оси, C - матрица, строки которой координаты проекций каждого вектора данных в базисном пространстве R .

Тогда координаты новой матрицы B в матричной форме будут записаны в виде:

$$B = R^T A \quad (4)$$

Матрица преобразования R^T в двумерном пространстве имеет вид:

$$R^T = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}. \quad (5)$$

Рассчитывается корреляционная матрица:

$$C = \frac{1}{N-1} (X^T X), \quad (6)$$

где N - число столбцов в матрице X .

Пусть $Y = B^T$, $X = A^T$, тогда получим:

$$Y = XR, \quad Y^T = R^T X^T. \quad (7)$$

Необходимо найти матрицу преобразования R^T такую, чтобы, применив ее к матрице X , получить новую систему координат Y , которая удовлетворяет выражению:

$$Y^T Y = R^T X^T X R = R^T C R = \Lambda,$$

где Λ - диагональная матрица.

Необходимо, чтобы выполнялось условие:

$$C R = \lambda R, \quad (8)$$

Тогда получим:

$$(C - \lambda I) R = 0, \quad (9)$$

где λ - скалярные, диагональные элементы в матрице Λ .

Задача будет иметь решение, когда:

$$|C - \lambda I| = 0. \quad (10)$$

После нахождения решения для λ подставим их обратно в (9) и найдем матрицу преобразования R . Получив матрицу преобразования, мы можем повернуть исходные оси таким образом, чтобы дисперсии ассоциировались с новыми осями, вычислить координаты данных в новой системе координат и начертить данные.

Рассмотрим следующий пример.

Пример 1.

Допустим, что из сформированной базы данных выделены следующие признаки, характеризующие состояния рассматриваемой системы (таблица 1). Матрица входных данных A будет размерности $(m \times n)$, $m = 12$, $n = 4$.

Таблица 1 - Признаки, характеризующие рассматриваемую систему

Измерения	Признак 1	Признак 2	Признак 3	Признак 4
1	41.61	99.00	18.39	98.77
2	51.28	97.88	22.88	98.61
3	43.56	91.42	22.40	92.49
...
12	33.82	100.00	48.00	100.00

Анализ собственных значений преобразованных данных по выше предложенному алгоритму выявил собственные векторы, расположенные в порядке убывания собственных значений.

В таблице 2 приведены данные по выделению главных компонент. Было выделено четыре фактора. В таблице показаны собственные значения, соответствующие каждому фактору, процент дисперсии и накопленный процент дисперсии.

Таблица 2 - Выделение главных компонент

Компоненты	Собственные значения	% дисперсии	Накопленный % дисперсии
1	1.973	49.332	49.332
2	1.628	40.690	90.022
3	0.343	8.587	98.609
4	0.056	1.391	100.000

Коэффициенты (нагрузки), указывают относительный вклад признаков в каждом собственном векторе. Так как с помощью метода главных компонент были выделены 4 фактора, то анализ собственных значений в зависимости от числа факторов (компонент) показал, что 90% дисперсии приходится на первые два фактора. Целесообразно, для построения оптимальной структуры иммунной сети, в качестве основных, рассматривать эти два фактора, т.к. они имеют наибольшее значение для рассматриваемого процесса. Влияние третьего и четвертого факторов на систему не столь значительно, поэтому данные факторы можно не принимать во внимание. К полученному факторному решению применена процедура вращения (таблица 3).

Таблица 3 - Матрица вращения

Признаки	1 компонента	2 компонента
Признак 1	0.151	0.898
Признак 2	0.980	0.076
Признак 3	0.126	-0.905
Признак 4	0.984	-0.052

На основе произведенных преобразований можно изобразить исходные данные в новой системе, где координатные оси являются собственными векторами. Необходимо после анализа признаков отбросить те, которые лежат ближе к началу координат и являются наименес информативными.

На основе матрицы вращения (таблица 3) весовые коэффициенты признаков можно представить в виде: ω_{kp} , где k - компонента, p - признак.

Для первой компоненты весовые коэффициенты имеют вид:

$$\omega_{11} = 0,151; \omega_{12} = 0,980; \omega_{13} = 0,126; \omega_{14} = 0,984.$$

Для второй компоненты весовые коэффициенты имеют вид:

$$\omega_{21} = 0,898; \omega_{22} = 0,076; \omega_{23} = -0,905; \omega_{24} = -0,052.$$

Весовые коэффициенты признаков фактически показывают, какое количество информации несут в себе признаки. Таким образом, при построении оптимальной структуры иммунной сети, мы будем рассматривать два основополагающих фактора с учетом весовых коэффициентов информативных признаков.

Ключевым моментом в разработанной интеллектуальной технологии на основе ИИС является решение задачи распознавания образов. Для каждого класса, выделенного экспертами, формируются матрицы эталонов: $A_1, A_2, A_3, \dots, A_n$, где n - количество классов. После сингулярного разложения данных матриц получаем правые и левые сингулярные вектора $\{x_1, y_1\}, \{x_2, y_2\}$ и т.д., эталонных матриц.

Затем формируется множество матриц, рассматриваемых как образы: $B_1, B_2, B_3, B_4, \dots, B_m$, где m - количество образов.

Согласно подходу ИИС энергия связи между формальными пептидами может быть представлена в виде:

$$W_1 = -x_1^T B y_1, W_2 = -x_2^T B y_2, W_3 = -x_3^T B y_3, W_4 = -x_4^T B y_4, \dots, W_n = -x_n^T B y_n, \quad (11)$$

где T - символ транспонирования, n - количество классов.

Нативная (функциональная) укладка белковой цепи соответствует минимуму энергии связи, поэтому минимальное значение энергии связи определяет класс n , к которому принадлежит данный образ:

$$n: W_n = \min\{W_1, W_2, W_3, W_4, \dots, W_n\} \quad (12)$$

Обработка многомерной совокупности данных на основе технологии ИИС неизбежно ведет к увеличению энергетических погрешностей, которые зависят от ряда факторов, и существенно влияет на достоверность прогноза. Определение нативной укладки цепи, соответствующей минимуму энергии, затруднено из-за различных погрешностей. Погрешности ИИС возникают за счет неполноты данных, корреляции данных и ошибок измерения. Особенно эта проблема актуальна для схожих по структуре пептидов, которые имеют примерно одинаковые параметры и находятся на границах классов. Из-за энергетических погрешностей они могут быть отнесены по ошибке не в свой класс, что существенно влияет на достоверность прогноза на основе ИИС и может привести к аварийной ситуации.

Под *гомологичными белками* будем понимать белки, имеющие одинаковое происхождение и обладающие примерно одинаковыми пространственными структурами, несмотря на многочисленные мутации в аминокислотной последовательности.

Под *конформацией* будем понимать наиболее энергетически выгодные и стабильные пространственные структуры белка с его уникальной последовательностью расположения аминокислот.

Под *нативной структурой* белка понимается естественное состояние белка, находящегося в нормальной конформации. Данная конформация обуславливает специфические биологические функции белка и соответствует глобальному минимуму внутренней энергии связи.

Нативная структура и конформационные динамические свойства белковой молекулы в физиологических условиях полностью определяются аминокислотной последовательностью.

Главной характеристикой аминокислотных последовательностей, определяющей основные физические свойства белка, является повышенная стабильность нативной структуры, т.е. существование большой щели между энергией нативной структуры и минимальной энергией неверно свернутых структур (рисунок 1). Здесь E - энергия укладки белковой цепи, m - число укладок в диапазоне энергий E , E_N - энергия нативной укладки, $\langle E \rangle$ - средняя энергия укладки цепи, N - положение нативной укладки, E_c - низкоэнергетический край квазинепрерывного энергетического спектра.

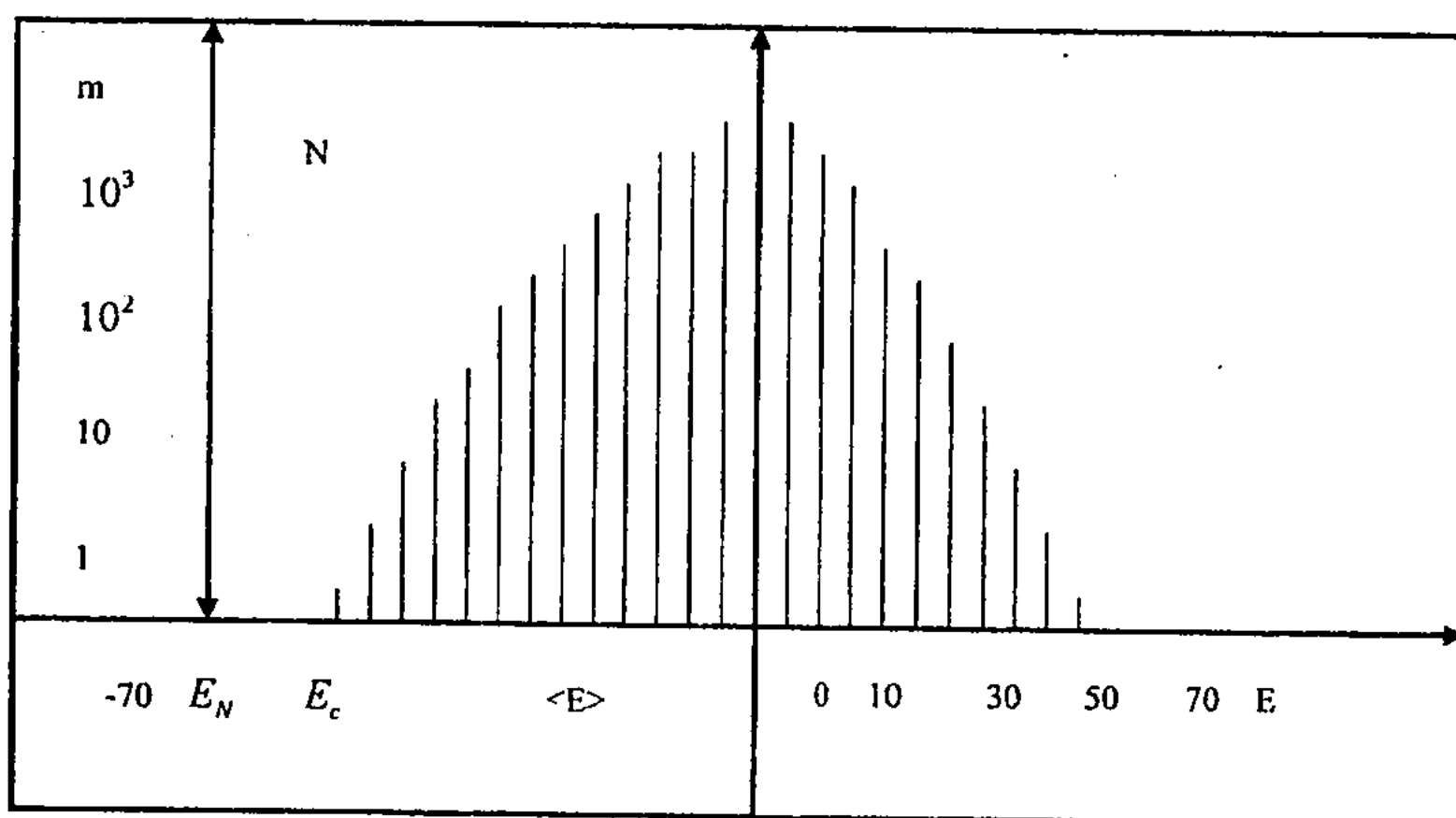


Рисунок 1 - Плотность энергетического спектра белка

Введем следующее определение.

Определение 4. Под гомологичными формальными пептидами будем понимать пептиды, которые при решении задачи распознавания образов относятся к одному классу решений.

Однако погрешности энергетических оценок могут затруднить определение нативной структуры. В искусственных иммунных системах данная ситуация возникает при решении задачи распознавания образов и, более всего это актуально для областей, находящихся на границах различных классов,

когда пептиды имеют схожее строение и наличие погрешностей ведет к большим проблемам. Определение принадлежности образов к какому - либо классу решений может быть ошибочным и повлечь за собой серьезные последствия при прогнозировании и управлении на основе ИИС, вплоть до риска аварийных ситуаций.

На рисунке 2 представлен случай, когда определение нативной структуры белка затруднено из - за различного рода погрешностей. Здесь показано, что низкоэнергетический край квазинепрерывного энергетического спектра E_c находится ниже энергии нативной укладки E_N и, поэтому определение нативной структуры невозможно.

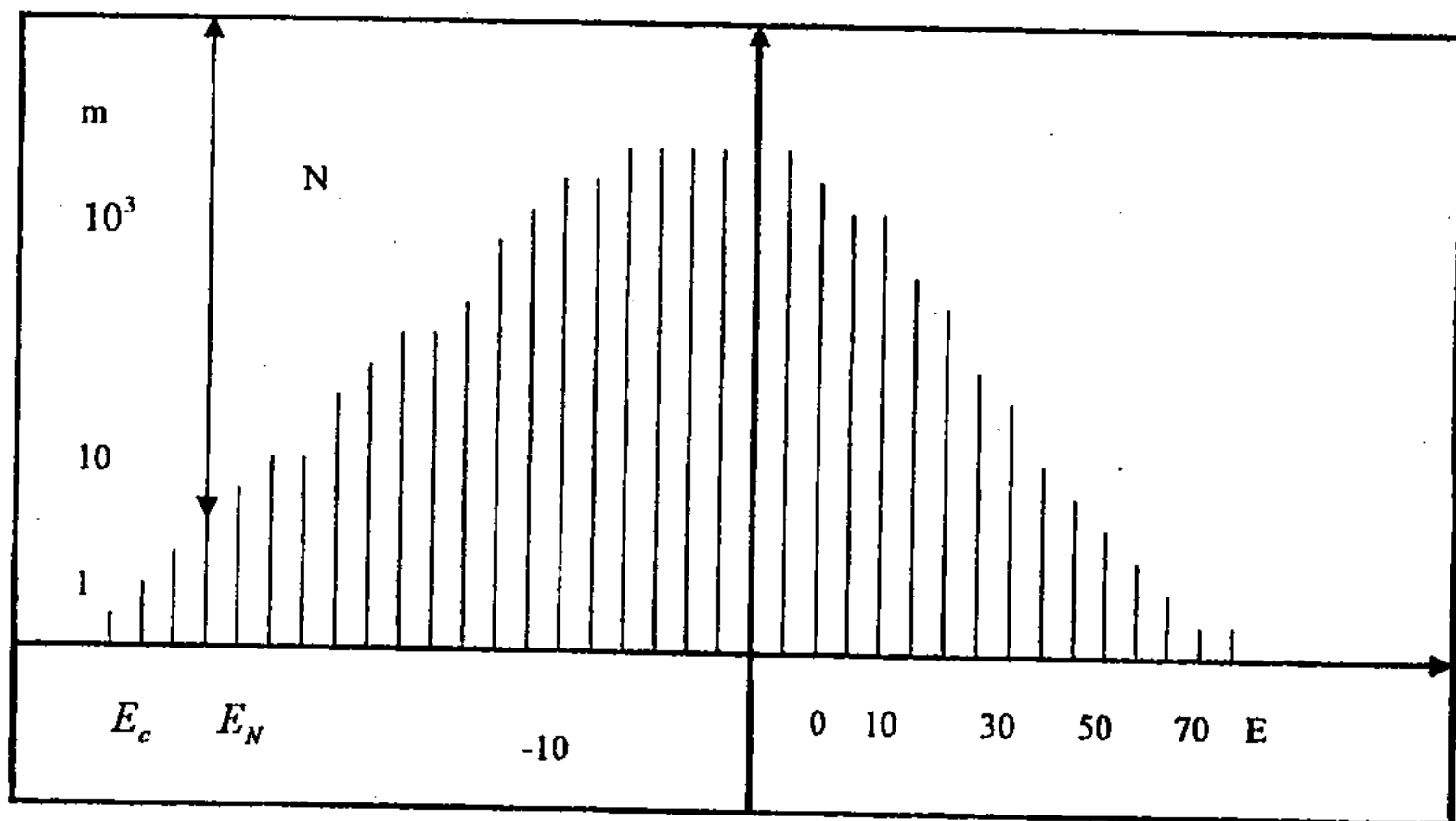


Рисунок 2 - Плотность энергетического спектра белка зашумленная ошибками

Устранение погрешностей энергетических оценок осуществляется на основе свойств гомологичных белков. Нативная (функциональная) укладка цепи N может быть найдена в процессе сворачивания цепи, где потенциал каждого взаимодействия усреднен по гомологам. Вычисленная энергия нативной структуры E'_N представляется в виде: $E'_N = E_N + \Delta E_N$, где E_N - истинное значение энергии нативной структуры, ΔE_N - погрешность энергетической оценки. Ниже приведен алгоритм оценки энергетических погрешностей ИИС.

Алгоритм:

Шаг 1. Усреднение потенциалов по гомологам:

$$\langle E_i \rangle_G = \frac{\sum^G E_i}{G}, \quad (13)$$

где символ $\langle \dots \rangle$ означает усреднение по гомологам, G - число гомологичных пептидов.

Шаг 2. Определение вычисленной энергии нативной (функциональной) структуры по усредненным гомологам:

$$\langle E_i^* \rangle_G \approx E_N. \quad (14)$$

Шаг 3. Определение Z - факторов. Значение Z - фактора определяется средним числом стандартных отклонений между энергией нативной структуры и энергией случайно выбранной укладки цепи:

$$Z = \frac{E_N - \langle E \rangle}{\sqrt{\langle (E - \langle E \rangle)^2 \rangle}}, \quad (15)$$

где $E_N \approx \langle E_i^* \rangle_G$, $\langle E \rangle$ - среднее число стандартных отклонений по гомологам, E - энергия случайно выбранной укладки цепи.

Шаг 4. Распознавание нативной структуры белков по гомологам и определение достоверности прогноза на основе ИИС в зависимости от величин Z - фактора. Рассчитываются коэффициенты риска прогнозирования:

$$K_R(G_i) = |1 - Z_i|, i = \overline{1, n}, \quad (16)$$

где n - количество гомологичных пептидов. Рассмотрим следующий пример.

Пример 2

Пусть пептиды G_1, G_2, G_3, G_4, G_5 , которые после решения задачи распознавания образов относятся к одному классу решений, являются гомологичными и имеют схожую пространственную структуру. Для удобства умножим энергии гомологичных пептидов на (-1). Тогда энергии связи гомологов для выбранного класса представлены следующим образом:

$$E_1^* = 319,456; E_2^* = 313,721; E_3^* = 377,379; E_4^* = 326,003; E_5^* = 304,987.$$

На первом шаге, согласно разработанному алгоритму, подставим значения энергии связи для гомологичных белков в (13). Так как число гомологичных белков равно пяти, то имеем следующее выражение:

$$\langle E_i^* \rangle_G = \frac{319,456 + 313,721 + 377,379 + 326,003 + 304,987}{5} = 328,309.$$

На втором шаге, подставим значение потенциалов, усредненных по гомологам в выражение (14):

$$\langle E_s \rangle_G \approx E_N \approx 328,309.$$

Будем считать, что значение энергии нативной структуры приблизительно равно среднему значению потенциалов гомологичных белков.

Затем на третьем шаге находим значения Z – факторов для различных гомологов. Для этого рассчитаем среднее число стандартных отклонений по гомологам $\langle E \rangle$. Отклонения энергий гомологов от энергии нативной структуры равны:

$$\Delta E_1 = 8,853, \Delta E_2 = 14,588, \Delta E_3 = -49,07, \Delta E_4 = 2,306, \Delta E_5 = 23,322.$$

Получим следующее выражение для среднего числа отклонений по гомологам:

$$\langle E \rangle = \frac{8,853 + 14,588 - 49,07 + 2,306 + 23,322}{5} = -0,0002$$

Подставив полученное значение в (15), рассчитаем значения Z – факторов для рассматриваемых белков:

$$Z_1 = 1,028; \quad Z_2 = 1,046; \quad Z_3 = 0,869; \quad Z_4 = 1,007; \quad Z_5 = 1,076.$$

Полученные результаты, на четвертом шаге предложенного алгоритма, сводятся в таблицу 4. На основе данной таблицы выявляются пептиды, ошибочно отнесенные к выбранному классу, и определяется степень достоверности прогноза ИИС.

Пептиды, для которых значения Z – факторов ближе к единице, содержат в себе меньше ошибок энергетических оценок, и их энергии более соответствуют нативной структуре.

Следует отметить, что особенно важно не только предсказать данные, но и оценить точность предсказания на основе расчета коэффициентов риска прогнозирования.

Рассчитываются коэффициенты риска $K_R(G_i), i = \overline{1,5}$ для всех гомологов по формуле (16), рассматриваемых в примере. Чем больше данный коэффициент, тем больше вероятность ошибочного зачисления данного пептида к выбранному классу решений.

В нашем примере наибольший коэффициент риска $K_R(G_3) = 0,131$, поэтому, скорее всего пептид G_3 не является гомологом и может быть исключен из данного класса решений.

Таким образом, нативная структура белковой цепи, соответствующая минимуму энергии связи, является для каждого класса определенной и позволяет определить принадлежность гомологов к какому-либо классу решений. Особенно это свойство ценно для образов, которые находятся на

границах классов. Данная способность ИИС существенно уменьшает погрешности энергетических оценок, повышает достоверность прогноза поведения интеллектуальной системы.

Таблица 4 - Оценка энергетических погрешностей на основе гомологов

Гомологичные пептиды	Энергии связи E_i^* , $i = \overline{1,5}$	Отклонение энергии от нативной ΔE_i , $i = \overline{1,5}$	Z_i - факторы, $i = \overline{1,5}$	Коэффициент риска $K_R(G_i)$, $i = \overline{1,5}$.
G_1	319,456	8,853	1,028	0,028
G_2	313,721	14,588	1,046	0,046
G_3	377,379	-49,07	0,869	0,131
G_4	326,003	2,306	1,007	0,007
G_5	304,987	23,322	1,076	0,076

Для реализации предложенной технологии разработан пакет прикладных программ «PERTID» предназначенный для обработки, анализа и прогноза многомерных данных искусственными иммунными системами. Программы написаны на языке программирования DELPHI 7.0.

Во втором разделе разработаны теоретические основы для создания информационных экспертных системы прогнозирования и управления сложными нелинейными объектами в условиях априорных неопределенностей различного типа (стохастической, интервально-заданной) на основе ИИС. Рассмотрены основные трудности при создании данных систем и показана перспективность предложенного пути решения этих проблем.

Создана интеллектуальная система прогнозирования асимптотической устойчивости в среднем квадратическом стохастической системы управления на основе ИИС и подхода квазирасщепления.

Рассмотрим управляемый объект, на входе которого действуют возмущения типа «белого шума». Математическая модель представлена в векторно-матричной форме Ланжевена следующим образом:

$$\dot{X}(t) = A(t)X(t) + B(t)U(t) + D(X)Z(t), t \in I(t_0), \quad (17)$$

где множество $I(t_0) \subset [0, \infty)$, t - текущее время; $X(t) = (X_1(t), \dots, X_n(t))^T \in R^n$ - вектор состояний управляемого объекта; T - символ транспонирования; $A(t) = [a_{ij}(t)] \in R^{n \times n}$ - функциональная матрица размерности $(n \times n)$; $B(t) = (b_1(t), \dots, b_n(t))^T \in R^n$ - функциональный вектор, размерности $(n \times 1)$, соответственно, удовлетворяющие соотношениям:

$$A^- \leq A(t) \leq A^+, B^- \leq B(t) \leq B^+. \quad (18)$$

Неравенства (18) понимаются в покомпонентном смысле, где A^+, A^-, B^+, B^- - известные константные матрицы и векторы соответствующих размерностей; $U(t)$ - скалярное управление; $Z(t) \in R^n$ - аддитивное векторное возмущение типа «белого шума», которое действует на входе объекта управления: $Z(t) = (Z_1(t), \dots, Z_n(t))^T$.

Статистические характеристики компонент «белого шума» задаются:

$$M\{Z_i(t)\} = 0, \quad i = \overline{1, n}, \quad M\{Z_i(t)Z_j(\rho)\} = r_{ij} \delta(t - \rho), \quad i, j = \overline{1, n}, \quad (19)$$

где M - символ математического ожидания; r_{ij} - заданная интенсивность соответствующих компонент белого шума; $R = \{r_{ij}, i, j = \overline{1, n}\}$ - ковариационная матрица; $\delta(t - \rho)$ - дельта-функция Дирака; $D(X)$ - матрица шумов, размерности $(n \times n)$, зависящая от переменных состояний объекта управления: $D(X) \in R^{n \times n}$;

$$D(X) = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ -X_1 & -X_2 & \dots & -X_n \end{bmatrix}$$

Управление $U(t)$ выбирается такое, чтобы обеспечить желаемую динамику в замкнутой системе:

$$U(t) = U(X(t), t). \quad (20)$$

Желаемая динамика замкнутой системы управления задается в виде:

$$\sigma(t) = C^T X = 0, \quad (21)$$

где $C \in R^n - Const, C^T = [C_1, \dots, C_n], C_n = 1$.

Соотношения (17) - (21) определяют математическую модель, далее называемую S -системой.

Задача исследования формулируется следующим образом: разработать интеллектуальную экспертную систему исследования асимптотической устойчивости в среднем квадратическом стохастической системы управления (S -системы) на основе метода квазирасщепления и биологического подхода Искусственных Иммунных Систем с целью прогнозирования поведения сложной нелинейной системы и оперативного управления текущей ситуацией в реальном масштабе времени.

Недостатком математических моделей в форме Ланжевена является наличие возмущения типа «белого шума», что ведет к разрывности траекторий движения. Более удобной формой представления математических моделей

являются стохастические квазирасщепленные дифференциальные уравнения типа Ито.

После применения процедуры квазирасщепления имеем две подсистемы: \overline{S}_1 и \overline{S}_2 , записанных относительно функций $x'(t)$ и $\sigma(t)$ в виде стохастических дифференциальных уравнений типа Ито. При этом сохраняется кинематическое подобие исходной S - системе.

Применение подхода Искусственных Иммуных Систем для создания интеллектуальных квазирасщепленных систем управления обеспечивает выполнение основных принципов организации интеллектуальных систем управления и позволяет осуществить достоверный прогноз поведения интеллектуальной системы, выбор предпочтительных управлений и корректировку текущего управления в реальном масштабе времени.

Для того чтобы осуществить прогноз поведения интеллектуальной системы на основе ИИС необходимо решить задачу классификации решений.

Пусть управление $U(t)$ будет таким, что для момента времени $t, \geq t_0$ для подсистем \overline{S}_1 и \overline{S}_2 решения $x'(t), \sigma(t)$ удовлетворяют следующему неравенству:

$$\|\sigma(t)\| \leq \delta \|x'(t)\| + \eta, \quad (22)$$

где $\delta, \eta - \text{Const.}, \delta \geq 0, \eta \geq 0$.

В данном случае асимптотика решений S - системы определяется асимптотикой решений \overline{S}_1 - подсистемы.

В случае, когда выполняется обратное неравенство, то есть:

$$\|\sigma(t)\| > \delta \|x'(t)\| + \eta, \quad (23)$$

где $t \in I(t_1)$, то асимптотика решений определяется подсистемой \overline{S}_2 .

В пространстве состояний выделим множества конусного типа следующего вида:

$$G_{\delta, \eta} = \{x \in R^n : \|\sigma(x)\| \leq \delta \|x'(x)\| + \eta\}, \text{ пусть } \eta=0. \quad (24)$$

$$\overline{G}_{\delta, \eta} = \{x \in R^n : \|\sigma(x)\| > \delta \|x'(x)\| + \eta\}, \text{ пусть } \eta=0. \quad (25)$$

Таким образом, заданные множества порождают в пространстве R^n два класса решений:

$$\{x(t)\}_1, \{x(t)\}_2 \text{ для } t \in I(t_0).$$

Первый класс решений относится к области $G_{\delta, \eta}$ (подсистема \overline{S}_1). Вторым класс решений относится к области $\overline{G}_{\delta, \eta}$ (подсистема \overline{S}_2).

Исследование динамических свойств исходной S - системы сводится к исследованию либо подсистемы \overline{S}_1 , либо подсистемы \overline{S}_2 .

Так как при квазирасщеплении считается, что пространство R^n представляется прямой суммой подпространств L_1 и L_2 : $L_1 \subseteq R^n, L_2 \subseteq R^n$, то

область $G_{\delta,\eta}^1$ будет дополнением множества $G_{\delta,\eta}$ до подпространства L_1 , а область $\overline{G_{\delta,\eta}^2}$ будет дополнением множества $\overline{G_{\delta,\eta}}$ до подпространства L_2 .

Область $G_{\delta,\eta}^1$ будет относиться к третьему классу решений, а область $\overline{G_{\delta,\eta}^2}$ к четвертому классу.

На рисунке 3 представлена классификация решений квазирасщепленных подсистем.

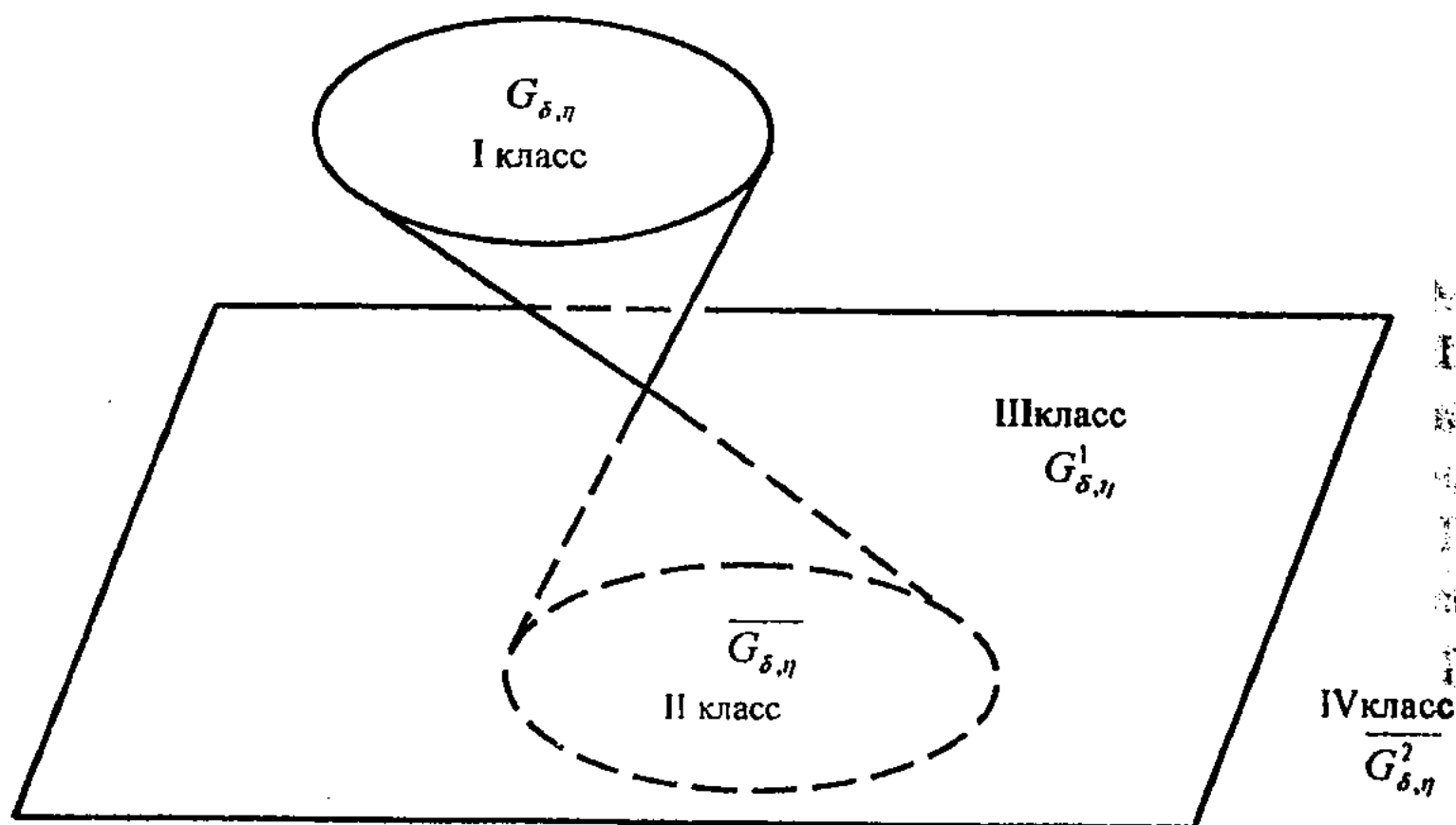


Рисунок 3 - Классификация решений квазирасщепленных подсистем.

Данная классификация решений необходима для исследования асимптотической устойчивости в среднем квадратическом квазирасщепленных подсистем на основе подхода ИИС. Каждой квазирасщепленной подсистеме соответствует определенная область решений, которая принадлежит к одному из четырех классов.

Для удобства вышеприведенные сведения помещены в таблицу 5.

Таблица 5 - Классификация областей решений

Подсистема	Область решения	Класс
$\overline{S_1}$	$G_{\delta,\eta}$	I
$\overline{S_2}$	$\overline{G_{\delta,\eta}}$	II
$\overline{S_1}$	$G_{\delta,\eta}^1$	III
$\overline{S_2}$	$\overline{G_{\delta,\eta}^2}$	IV

Согласно идеологии ИИС формируются эталонные матрицы управления $U_1, U_2, \bar{U}_1, \bar{U}_2$ для каждого из 4 классов.

Классы выделяются экспертами в зависимости от областей функционирования и определяются технологическими процессами, параметрами квазирасщепленных подсистем \bar{S}_1, \bar{S}_2 .

С использованием сингулярного разложения матриц разработаны процедуры обучения с учителем, классификации исходной информации и распознавания образов для исследования асимптотической устойчивости стохастических квазирасщепленных подсистем с целью качественного прогноза поведения интеллектуальной системы, выбора предпочтительных управлений и корректировки текущего управления.

Ниже приведен алгоритм, который состоит из 14 шагов.

Алгоритм:

Шаг 1. Получение выражений операторов проектирования P_1 и P_2 для стохастической системы управления (S - системы) в форме дифференциальных уравнений Ланжевена.

Шаг 2. Получение квазирасщепленных подсистем S_1' и S_2' на основе операторов проектирования P_1 и P_2 в форме Ланжевена.

Шаг 3. Получение стохастических квазирасщепленных подсистем \bar{S}_1 и \bar{S}_2 в форме дифференциальных уравнений типа Ито.

Шаг 4. Классификация областей решений квазирасщепленных подсистем \bar{S}_1 и \bar{S}_2 .

Шаг 5. Нормирование и центрирование входных признаков согласно формулам (1) - (2).

Шаг 6. Выделение информативных признаков на основе факторного анализа (метода главных компонент) и снижение размерности анализируемого пространства признаков.

Шаг 7. Создание с помощью экспертов матриц управления (эталонов) для каждого класса, сворачивание их в квадратные матрицы для улучшения специфичности узнавания, сингулярное разложение данных матриц и определение правых и левых сингулярных векторов. Матрицы эталонов рассматриваются как антигены.

Шаг 8. Процедуры обучения Искусственной Иммунной Системы с учителем.

Шаг 9. Создание матриц управления (образов), сворачивание их в квадратную матрицу для улучшения специфичности узнавания. Матрицы образов рассматриваются как антитела.

Шаг 10. Определение минимальной энергии связи между формальными пептидами (антителами и антигенами) и решение задачи распознавания образов. Минимальное значение энергии связи указывает на класс, к которому принадлежит данный образ.

Шаг 11. Оценка энергетических ошибок Искусственной Иммунной Системы на основе свойств гомологичных белков по формулам (13-15).