

На правах рукописи



Горбунов Иван Викторович

**АЛГОРИТМЫ И ПРОГРАММНЫЕ СРЕДСТВА ИДЕНТИФИКАЦИИ
ПАРЕТО–ОПТИМАЛЬНЫХ НЕЧЕТКИХ СИСТЕМ НА ОСНОВЕ
МЕТАЭВРИСТИЧЕСКИХ МЕТОДОВ**

Специальность 05.13.18 – Математическое моделирование, численные методы
и комплексы программ

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Томск–2014

Работа выполнена в Федеральном государственном образовательном бюджетном учреждении высшего профессионального образования «Томский государственный университет систем управления и радиоэлектроники» (ТУСУР)

Научный руководитель – доктор технических наук, профессор
Ходашинский Илья Александрович

Официальные оппоненты: Янковская Анна Ефимовна, доктор технических наук, профессор, профессор кафедры прикладной математики Томского государственного архитектурно-строительного университета;

Аксенов Сергей Владимирович, кандидат технических наук, доцент кафедры оптимизации систем управления Национального исследовательского Томского политехнического университета

Ведущая организация – Национальный исследовательский Иркутский государственный технический университет

Защита состоится 09 октября 2014 г. в 15 час. 15 мин. на заседании диссертационного совета Д 212.268.02 в Томском государственном университете систем управления по адресу:
634050, г. Томск, пр. Ленина, 40.

С диссертацией можно ознакомиться в научной библиотеке Томского государственного университета систем управления и радиоэлектроники по адресу: г. Томск, ул. Вершинина, 74 и на сайте Томского государственного университета систем управления и радиоэлектроники <http://tusur.ru/ru/science/news/diss.html>.

Автореферат разослан «__» _____ 2014 г.

Ученый секретарь
диссертационного совета



Р.В. Мещеряков

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. Нечеткие системы давно и успешно применяются в таких проблемных областях, как классификация, аппроксимация, интеллектуальный анализ данных, распознавание образов, прогнозирование и управление, в задачах принятия решений. Достоинствами нечетких систем являются невысокая стоимость разработки, гибкость, интуитивно понятная логика функционирования.

В последнее время нечеткие методы моделирования сконцентрированы на проблемах улучшения интерпретируемости нечетких систем без потери точности. Интерпретируемость является важным свойством модели, так как позволяет преобразовывать данные в знания. Поиск компромисса между точностью и интерпретируемостью является в настоящее время одним из наиболее актуальных направлений исследований в нечетком моделировании.

Основной вклад в исследование критериев интерпретируемости нечетких систем и методов ее достижения внесли R. Alcalá, J. Alonso, J.M. Andújar, R. Cannone, G. Castellano, M. Gacto, A.M. Fanelli, Gan J.Q., G. Gonzalez–Rodriguez, S. Guillaume, F. Herrera, H. Ishibuchi, Y. Jin, H. Koivisto, L. Magdalena, Mencar C., D.D. Nauck, Y. Nojima, J. V. Oliveira, W. Pedrycz, P. Pulkkinen, S. Romero, O. Sánchez, M.A. Vélez, H. Wang, T. Yamamoto, Y. Zhang, S. M. Zhou.

Целью диссертационной работы является разработка алгоритмов и программных средств идентификации Парето-оптимальных нечетких систем на основе метаэвристических методов для достижения компромисса между точностью, сложностью и интерпретируемостью нечетких систем.

Для достижения поставленной цели решены следующие основные задачи:

1) обзор существующих методик и методов построения многоцелевых Парето-оптимальных нечетких систем и критериев, используемых для оценки их интерпретируемости;

2) разработка методики построения многоцелевых Парето-оптимальных нечетких моделей на основе метаэвристических методов с применением оригинального комплексного критерия интерпретируемости нечетких систем;

3) разработка гибридного численного метода идентификации структуры и параметров нечеткой системы, основанного на группе алгоритмов пчелиной колонии и адаптивном алгоритме эволюционной стратегии;

4) разработка программных средств, реализующих предложенную методику и гибридный численный метод в многопоточном режиме, позволяющих экспортировать нечеткие системы в представление, совместимое с языком разметки прогнозного моделирования (PMML);

5) исследование разработанных алгоритмов, численных методов и предложенной методики на контрольных примерах в сравнении с аналогами.

Объектом исследований является процесс структурной и параметрической идентификации нечетких систем.

Предметом исследований являются гибридные алгоритмы и программы идентификации структуры и параметров Парето-оптимальных нечетких систем, а также критерий их интерпретируемости.

Методы исследования. В диссертационной работе применялись методы искусственного интеллекта, теории нечетких множеств, математической статистики, линейной алгебры, объектно-ориентированного и рефлексивно-ориентированного программирования.

Достоверность результатов обеспечивается строгостью применения математических методов, результатами проведенных численных экспериментов, которые сопоставимы с данными, полученными другими исследователями.

Научная новизна полученных результатов. В диссертации получены следующие новые научные результаты.

1. Разработана оригинальная трехкритериальная методика построения Парето-оптимальных нечетких моделей. Методика отличается от аналогов использованием элементарных гиперпараллелепипедов задаваемых размеров, динамически изменяющихся во время работы по каждому из критериев, и использованием однокритериальных алгоритмов идентификации параметров и структуры нечеткой системы.

2. Впервые разработан гибридный численный метод идентификации структуры и параметров нечетких систем, основанный на группе алгоритмов пчелиной колонии и адаптивном алгоритме эволюционной стратегии с применением метода наименьших квадратов и схемы островов.

3. Разработан алгоритм генерации нечетких классификаторов на основе экстремумов таблицы наблюдений. Алгоритм отличается тем, что позволяет уменьшить количество правил нечеткого классификатора не менее чем в 2 раза по сравнению с методами равномерного разбиения при сравнимой точности и повысить точность классификации более чем на 10% по сравнению с методами, основанными на нечетких s -средних, при сохранении сопоставимого количества правил.

Теоретическая значимость работы заключается в развитии технологии построения Парето-оптимальных нечетких моделей при апостериорном подходе. Разработанный гибридный метод может использоваться не только для идентификации параметров нечетких систем, но и как метод глобальной параметрической оптимизации.

Практическая значимость работы подтверждается использованием полученных в ней результатов для решения практических задач:

- создание рекомендательной системы немедикаментозного восстановительного лечения участников вооруженных конфликтов и чрезвычайных ситуаций при назначении комплексов реабилитации пациентам с посттравматическими стрессовыми расстройствами, внедренной в НИИ курортологии и физиотерапии ФМБА России (НИИ КиФ);
- разработка программного модуля, служащего для повышения качества распознавания клавиатурного почерка при двухфакторной аутентификации для предоставления доступа по протоколу RDP, внедренного в ОАО «Особая экономическая зона технико-внедренческого типа г. Томск» (ОАО ОЭЗ ТВТ).

Разработанные алгоритмы и программные средства использованы при выполнении следующих проектов:

— грант РФФИ 09–07–99008 «Исследование и разработка технологии идентификации нечетких моделей на базе метаэвристик и методов, основанных на производных» 2009–2010г.;

— грант РФФИ 12–07–00055 «Методы построения Парето-оптимальных нечетких систем на основе гибридного подхода» 2012–2014г.;

— грант РГНФ 12–06–12008 «Программный комплекс для прогнозирования эффективности реабилитации лиц опасных профессий с наиболее распространенными социально значимыми неинфекционными заболеваниями» 2012–2013г.;

— проект УМНИК 2014 «Разработка облачного менеджера паролей с двухфакторной аутентификацией на основе клавиатурного почерка».

На защиту выносятся приведенные ниже положения.

1. Трехкритериальная методика построения Парето-оптимальных нечетких моделей позволяет достичь равномерного распределения решений вдоль Парето-фронта по каждому из трех критериев: интерпретируемость, сложность и точность нечеткой системы. Кроме того, данная методика позволяет использовать однокритериальные алгоритмы идентификации параметров и структуры нечеткой системы без изменения самой методики и алгоритмов, а также позволяет регулировать соотношение скорости поиска и дистанции приближения к гипотетическому (идеальному) Парето-фронту по каждому критерию.

Соответствует пункту 1 паспорта специальности: Разработка новых математических методов моделирования объектов и явлений.

2. Численный метод идентификации структуры и параметров нечеткой системы, основанный на группе алгоритмов пчелиной колонии и адаптивном алгоритме эволюционной стратегии с применением метода наименьших квадратов и схемы островов, позволяет повысить точность вывода нечеткой системы не менее чем в 1,36 раза на обучающих выборках по сравнению с численными методами, применяемыми другими исследователями в аналогичных задачах.

Соответствует пункту 3 паспорта специальности: Разработка, обоснование и тестирование эффективных вычислительных методов с применением современных компьютерных технологий.

3. Алгоритм генерации нечетких классификаторов на основе экстремумов таблицы наблюдений.

Соответствует пункту 3 паспорта специальности: Разработка, обоснование и тестирование эффективных вычислительных методов с применением современных компьютерных технологий.

4. Многопоточная архитектура программного комплекса для построения Парето-оптимальных нечетких систем реализует предложенные алгоритмы и методику, использует оригинальное расширение формата PMML для экспорта-импорта нечетких систем с глобально определенными терминами.

Соответствует пункту 4 паспорта специальности: Реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента.

Внедрение результатов диссертационного исследования. Результаты исследовательской работы вошли в качестве программных компонентов в рекомендательную систему для немедикаментозного восстановительного лечения участников вооруженных конфликтов и чрезвычайных ситуаций при назначении комплексов реабилитации пациентам с посттравматическими стрессовыми расстройствами, используемую в НИР НИИ КиФ.

Результаты работы вошли в качестве программных компонентов в разработанный программный комплекс, предназначенный для повышения защищенности доступа по RDP с помощью внедрения дополнительного фактора аутентификации по клавиатурному почерку со сниженными по сравнению с аналогами ошибками первого и второго рода, используемый в ОАО ОЭЗ ТВТ.

Разработанные алгоритмы идентификации нечетких систем используются при изучении дисциплины «Базы данных и экспертные системы», а также в рамках группового проектного обучения на кафедре комплексной информационной безопасности электронно-вычислительных систем ТУСУР.

Апробация работы. Основные положения работы докладывались и обсуждались на следующих конференциях, семинарах, выставках: Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР», г. Томск, 2009-2014 г.; Всероссийской конференции «Проведение научных исследований в области обработки, хранения, передачи и защиты информации», г. Ульяновск, 2009 г.; Международной научной конференции «Системный анализ в медицине» (САМ 2011), г. Благовещенск, 2011 г.; Всероссийской конференции с международным участием «Знания – Онтологии – Теории» (ЗОНТ-2011), г. Новосибирск, 2011 г.; Международной заочной научно-практической конференции «Наука, образование, общество: тенденция и перспективы», г. Москва, 2013 г.; VII-ой Международной научно-практической конференции «Интегрированные модели и мягкие вычисления в искусственном интеллекте», г. Коломна, 2013 г.; Всероссийской конференции с международным участием «Современные системы искусственного интеллекта и их приложения в науке», г. Казань, 2013 г.; Международной научно-практической конференции «Электронные средства и системы управления», г. Томск, 2013 г.; XII Всероссийском совещании по проблемам управления Россия, г. Москва, ИПУ РАН, 2014 г.; Открытой выставке научных достижений молодых ученых РОСТ UP-2013, г. Томск, 2013 г.; Всероссийском конкурсе-конференции по информационной безопасности SIBINFO 2014, г. Томск; Томском IEEE семинаре «Интеллектуальные системы моделирования, проектирования и управления» г. Томск, 2012-2014 г.

Публикации по теме диссертации. По результатам исследований опубликованы 24 печатные работы, из которых в рекомендованных ВАК РФ периодических изданиях – 7. Были получены 3 свидетельства о государственной регистрации программ для ЭВМ (номера свидетельств: №2013610783, №2014610816, №2014610817).

Личный вклад автора. Постановка цели научного исследования и подготовка материалов к печати велась совместно с научным руководителем. Основ-

ные научные результаты получены лично автором. Автором самостоятельно разработан численный метод, предложенные алгоритмы, оригинальная методика и оригинальная архитектура для программного комплекса построения Парето-оптимальных нечетких систем.

Структура и объем работы. Диссертационная работа состоит из введения, пяти глав, заключения, списка литературы из 140 наименований и шести приложений. Основная часть работы содержит 167 страниц, в том числе 75 рисунков и 31 таблицу.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы исследования, формулируется цель работы, излагаются полученные автором основные результаты проведенных исследований, показывается их научная новизна, теоретическая и практическая значимость, отражаются основные положения, выносимые на защиту.

В первой главе производится обзор проблемы исследования. Описываются виды нечетких систем, приводится обзор методов идентификации структуры и параметров нечеткой системы по одному и нескольким критериям.

Постановка задачи. В работе для решения задачи аппроксимации использована нечеткая система типа синглтон, j -ое правило которой имеет вид:

$$\text{ЕСЛИ } x_1 = A_{1j} \text{ И } x_2 = A_{2j} \text{ И } \dots \text{ И } x_m = A_{mj} \text{ ТО } y = a_j,$$

где $\mathbf{x} = [x_1, \dots, x_m]^T \in \mathfrak{R}^m$, a_j – действительное число, A_{ij} – нечеткая область определения i -ой входной переменной.

Вывод в нечеткой системе типа синглтон описан формулой:

$$F(\mathbf{x}) = \frac{\sum_{j=1}^n \mu_{A_{1j}}(x_1) \cdot \mu_{A_{2j}}(x_2) \cdot \dots \cdot \mu_{A_{mj}}(x_m) \cdot a_j}{\sum_{j=1}^n \mu_{A_{1j}}(x_1) \cdot \mu_{A_{2j}}(x_2) \cdot \dots \cdot \mu_{A_{mj}}(x_m)},$$

где n – количество правил; m – количество входных переменных; $\mu_{A_{ij}}$ – функция принадлежности нечеткой области A_{ij} .

Нечеткая система типа синглтон может быть представлена как $y = F(\mathbf{x}, \boldsymbol{\theta})$, где y – скалярный выход, $\boldsymbol{\theta} = \|\theta_1, \dots, \theta_M\|$ – вектор оптимизируемых параметров, $N = \sum_{i=1}^m t * b_i + n$, t – число параметров, описывающих одну функцию принадлежности, b_i – число термов, описывающих i -ую входную переменную.

В работе для решения задачи классификации использован питтсбургский тип классификатора, i -ое правило которого, описывающее принадлежность к j -ому классу, имеет следующий вид:

$$R_{ij}: \text{ЕСЛИ } x_1 = A_{1j} \text{ И } x_2 = A_{2j} \text{ И } \dots \text{ И } x_m = A_{mj} \text{ ТО } \text{class} = c_k,$$

где c_k – идентификатор k -ого класса, $k \in [1, c]$.

Выход классификатора определяется следующим образом:

$\text{class} = c_{j^*}$, $j^* = \arg \max_{1 \leq j \leq c} \beta_j$, где $\beta_j(\mathbf{x}_p) = \sum_{R_{ij}} \prod_{l=1}^m \mu_{A_{il}}(x_{pl})$ – степень принадлежности

наблюдения \mathbf{x}_p к j -ому классу.

Нечеткий классификатор описывается функцией $f: \mathfrak{R}^m \rightarrow [0,1]^c$, которая относит классифицируемый объект к каждому классу с определенной степенью принадлежности. Тогда нечеткий классификатор может быть представлен как функция: $\text{class} = f(\mathbf{x}, \boldsymbol{\theta})$, где $\boldsymbol{\theta} = \|\theta_1, \dots, \theta_N\|$ – вектор оптимизируемых параметров, $N = \sum_{i=1}^m t * b_i + \sum_{j=1}^c k_j$, t – число параметров, описывающих одну функцию принадлежности, b_i – число термов, описывающих i -ю входную переменную, k_j – число правил, относящихся к j -ому классу.

Решение проблемы построения Парето-оптимальных нечетких моделей типа синглтон и классификаторов питтсбургского типа с разными соотношениями между значениями их интерпретируемости, точности (ошибки) вывода и сложности основано на оптимизации заданных целевых функций в многомерном пространстве, координаты которого соответствуют параметрам нечетких систем:

$$\left\{ \begin{array}{ll} \min(E(\boldsymbol{\theta})) & \text{для нечеткой системы типа синглтон} \\ \max(\text{Pr}(\boldsymbol{\theta})) & \text{для питтсбургского классификатора} \\ \max(I(\boldsymbol{\theta})) & \\ \min(NR) & \end{array} \right. \quad \begin{array}{l} \text{при ограничениях:} \\ \boldsymbol{\theta}_{\min} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_{\max}, \\ E(\boldsymbol{\theta}), I(\boldsymbol{\theta}) \subseteq \mathfrak{R}, \\ NR \subseteq \{2, 3, \dots, NR_{\max}\}, \\ \forall_{x \in U} \exists_{1 \leq i \leq n} A_i(x) > 0, \end{array}$$

где $E(\boldsymbol{\theta})$ – ошибка вывода нечеткой системы типа синглтон, $\text{Pr}(\boldsymbol{\theta})$ – процент правильной классификации питтсбургского классификатора, $I(\boldsymbol{\theta})$ – значение критерия интерпретируемости нечеткой системы, NR – количество правил в нечеткой системе, NR_{\max} – максимально предусмотренное число правил; $\boldsymbol{\theta}_{\min}$, $\boldsymbol{\theta}_{\max}$ – нижняя и верхняя границы параметров решения соответственно; U – универсум, на котором определена переменная x .

Во второй главе описывается разработанная методика и алгоритм построения Парето-оптимальных нечетких моделей; комплексный критерий интерпретируемости нечетких систем; разработанные алгоритмы однокритериальной оптимизации: алгоритм пчелиной колонии для генерации правил нечеткой системы, алгоритм пчелиной колонии для идентификация параметров нечеткой системы, алгоритм адаптивной эволюционной стратегии для идентификации параметров нечеткой системы; гибридный численный метод идентификации параметров нечеткой системы, использующий алгоритмы однокритериальной оптимизации, алгоритм генерации нечетких классификаторов на основе экстремумов таблицы наблюдений.

Трехкритериальная методика построения Парето-оптимальных нечетких моделей. При формировании методики построения Парето-оптимальных нечетких моделей приняты соглашения и допущения: 1) каждому решению (нечеткой модели) ставится в соответствие точка в пространстве изменения значений

критериев; 2) множество новых решений генерируется из одного выбранного решения-предка; 3) выбор текущего предка осуществляется на основе меры, значение которой меняется от одной точки пространства к другой и рассчитывается как алгебраическая сумма приращений количества элементарных гиперпараллелепипедов (ЭГ) по каждому критерию; 4) формирование новых решений выполняется однокритериальными алгоритмами генерации структуры и оптимизации параметров нечетких моделей.

Пошаговое описание методики приведено ниже.

Вход. Максимальное заполнение ЭГ; максимальное количество выборов решения в качестве предка; начальное значение размера ЭГ по каждому критерию; количество итераций.

Выход. Парето-фронт построенных нечетких систем.

Шаг 1. Сформировать ЭГ в пространстве изменения значений критериев.

Шаг 2. Инициализировать решение-предок.

Шаг 3. Создать каждым алгоритмом генерации структуры и оптимизации параметров нечеткой системы множество решений и поместить их в архив.

Шаг 4. Сформировать список кандидатов на роль текущего предка, выбрав из архива решения, находящиеся в тех ЭГ, которые не заполнены до максимального значения, и у которых не превышен порог количества выборов в качестве предка.

Шаг 5. Если список кандидатов пуст, то уменьшить размеры ЭГ; обнулить все счетчики, отвечающие за подсчет количества выборов решений в качестве предка; перейти на *шаг 4*.

Шаг 6. Выбрать текущего предка по максимальному значению меры суммы приращений количества ЭГ по каждому критерию; увеличить на единицу счетчик числа выборов у выбранного предка.

Шаг 7. Если не выполнено заданное число итераций, то перейти на *шаг 3*.

Шаг 8. Сформировать из архива найденных решений Парето-фронт.

В методике используется оригинальный *комплексный критерий интерпретируемости*, представленный формулой:

$$I(\theta) = \frac{1}{n} \sum_{j=1}^n \left[\frac{Ctf(v(j))}{v(j) \cdot (v(j) - 1) / 2} \sum_{k=1}^{v(j)-1} \sum_{l=k+1}^{v(j)} [GI3(A_{kj}, A_{lj}) \cdot Linds(A_{kj}, A_{lj})] \right],$$

где $v(j)$ – количество термов, описывающих j -ю входную переменную;

$$Ctf(v(j)) = \begin{cases} 1, & \text{если } v(j) \leq 5 \\ \frac{1}{1 + \left[\frac{v(j) - 5}{4} \right]^2}, & \text{если } v(j) > 5 \end{cases} \text{ – индекс понятности, основанный на коли-}$$

честве термов;

$$GI3(A_{kj}, A_{lj}) = \sqrt[3]{GIS(A_{kj}, A_{lj}) \times GIC(A_{kj}, A_{lj}) \times GIB(A_{kj}, A_{lj})} \text{ – индекс геометрической различимости термов } A_{kj} \text{ и } A_{lj}$$

$$Lindis(A_{kj}, A_{lj}) = \left\{ \begin{array}{l} 1, \text{ если } A_{kj} \cap A_{lj} = \emptyset \\ 1 - \frac{abs(k-l+1)}{v(j)}, \text{ иначе} \end{array} \right\} - \text{штраф за пересечения «семантиче-}$$

ски» далеких термов, индексы k и l «семантически» упорядочены.

В $GIZ(\underline{A}_{kj}, A_{lj})$ используются следующие индексы: индекс различимости термов на основе площади пересечения $GIS(\underline{A}_{kj}, A_{lj})$; индекс $GIC(\underline{A}_{kj}, A_{lj})$, показывающий различимость термов на основе расстояния между их пиками; индекс $GIB(\underline{A}_{kj}, A_{lj})$, определяющий различимость термов на основе расстояния взаимного пересечения термов относительно расстояния между их пиками.

Гибридный численный метод идентификации структуры.

Вход: алгоритм пчелиной колонии для генерации правил ($algorithms_1$), алгоритм пчелиной колонии для идентификации параметров ($algorithms_2$), алгоритм адаптивной эволюционной стратегии ($algorithms_3$), метод наименьших квадратов ($algorithms_4$), $count_recv_{alg}$ – количество принимаемых внешних лучших баз правил алгоритмом alg , $count_sent_{alg}$ – количество отправляемых лучших баз правил алгоритмом alg , θ – начальная база правил нечеткого аппроксиматора.

Выход: θ^* – оптимизированная база правил нечеткого аппроксиматора; **Ocean** – упорядоченное множество баз правил; $maxsize_Ocean$ – максимальная мощность множества **Ocean**.

Шаг 1. Инициализация $maxsize_Ocean = \max_{1 \leq i \leq |algorithms|} (count_recv_{algorithms_i})$.

Шаг 2. Запуск на исполнение каждого алгоритма из **algorithms** с копией θ в качестве параметра.

Шаг 3. Пока запущен хотя бы один алгоритм из **algorithms** *делать:*

Шаг 3.1. Если $algorithms_i$ *готов* отправить базы правил **To Ocean = Ocean** \cup $algorithms_i.sent(count_sent_{algorithms_i})$.

Шаг 3.2. Если $|Ocean| > maxsize_Ocean$ **To** из **Ocean** исключаются $|Ocean| - maxsize_Ocean$ баз правил с наибольшей ошибкой аппроксимации.

Шаг 3.3. Если $algorithms_i$ *готов* принимать внешние базы правил **To** $algorithms_i.recv(Ocean, count_recv_{algorithms_i})$.

End *делать.*

Шаг 4. $\theta^* = \arg \max_{1 \leq i \leq |algorithms|} (algorithms_i.getBest\theta; Ocean)$.

Алгоритм генерации нечетких классификаторов на основе экстремумов таблицы наблюдения.

Вход: Число классов m , таблица наблюдений $\{x_q, t_q\}$, *Type* – тип функции принадлежности.

Выход: Начальная база правил классификатора θ .

Шаг 1. $\theta = \emptyset$.

Шаг 2. Для каждого k -го класса *делать:*

Шаг 2.1. Для каждого j -го признака *делать:*

Шаг 2.1.1. Поиск $\min_{kj} class = \min_q(x_{qj})$.

Шаг 2.1.2. Поиск $\max class_{kj} = \max_q(x_{qj})$.

Шаг 2.1.3. Создание терма A_{kj} типа *Type*, накрывающего интервал $[minclass_{kj}, maxclass_{kj}]$.

End делатъ (j).

Шаг 2.2. Создание правила R_k : ЕСЛИ $x_1=A_{k1}$ И $x_2=A_{k2}$ И $x_3=A_{k3}$ И ... И $x_n=A_{kn}$ ТО $class=c_k$, $w=1$.

Шаг 2.3. $\theta := \theta \cup R_k$.

End делатъ (k).

Шаг 3. По каждому j -го признаку **делатъ**:

Шаг 3.1. Проверка на наличие ненакрытых термом областей определения переменной.

Шаг 3.2. **Если** ненакрытые места найдены **То**

Шаг 3.2.1. Ближайший терм слева передвигает свою правую границу на размер незаполненного расстояния.

Шаг 3.2.2. Ближайший терм справа передвигает свою левую границу на размер незаполненного расстояния.

End Если.

End делатъ (j).

В третьей главе представлены следующие разделы: UML-описание области исследования; построение многопоточной архитектуры; реализация программного комплекса построения Парето-оптимальных нечетких моделей; классы программного комплекса, схема их взаимодействия и параллельного выполнения. Описано PMML-представление нечеткой системы в виде XML документа.

Предложена многопоточная архитектура разработанного программного комплекса с широким набором схем параллельного выполнения каждого реализованного алгоритма и диспетчер, отвечающий за подбор оптимального по времени варианта исполнения под используемый процессор. Правила, которые строит и использует диспетчер многопоточных вычислений, имеют вид:

IF TaskType= «тип задачи» **AND** ProcessorType= «семейство и производитель процессора» **THEN** PF=ValuePF, PBS= ValuePBS, PBP= ValuePBS, PES= ValuePES, PPO= ValuePPO;

где «тип задачи» – это одно из значений {sequence; hybride; pareto}. Значение sequence обозначает последовательный запуск алгоритмов; hybride указывает на запуск алгоритмов идентификации структуры и параметров в гибридном численном методе; значение pareto используется для алгоритмов идентификации структуры и параметров в составе методики; PF, PBS, PBP, PES, PPO – обозначение алгоритмов для идентификации структуры и параметров.

Выражение PBS= ValuePBS означает, что в алгоритме пчелиной колонии для генерации правил следует использовать схему параллельного выполнения с номером ValuePBS.

Предложено описание нечетких систем в стандарте PMML 4.2. Пример использования предложенного стандарта:

Описание нечеткого термина	Описание правила нечеткой системы
<pre><FuzzySet name ="Inhabit- antsT1"> <TriangeMemberFunction lower="-40" pick="1.6284395819225612" upper="360"/> </FuzzySet></pre>	<pre><SingletonRule score="80"> <CompoundPredicate booleanOperator="and"> <SimplePredicate field="Inhabitants" opera- tor="Equal" value="InhabitantsT1"/> <SimplePredicate field="Distance" opera- tor="Equal" value="DistanceT1"/> </CompoundPredicate> </SingletonRule></pre>

Четвертая глава содержит описание тестовых данных, результаты и анализ проведенных экспериментов по генерации нечетких систем и параметрической идентификации; даются рекомендации по использованию параметров разработанных алгоритмов; приводятся результаты сравнения разработанного гибридного численного метода с алгоритмами, предложенными другими авторами.

Пример построенных трехкритериальных Парето-оптимальных нечетких моделей на данных Plastic из репозитория KEEL представлен на рисунках 1, 2 и 3.

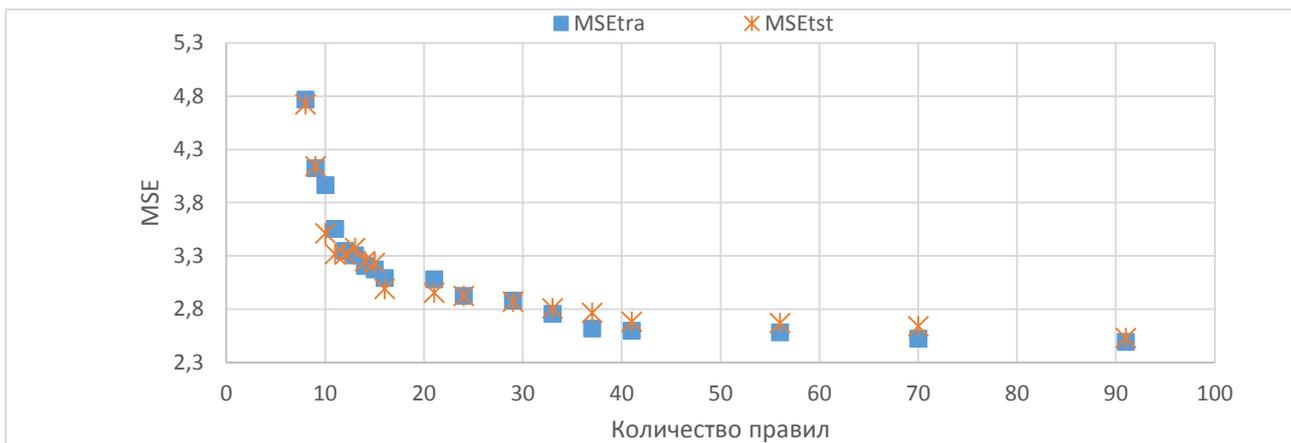


Рис. 1 – Проекция трехкритериального Парето-фронта на оси MSE и «Количество правил»

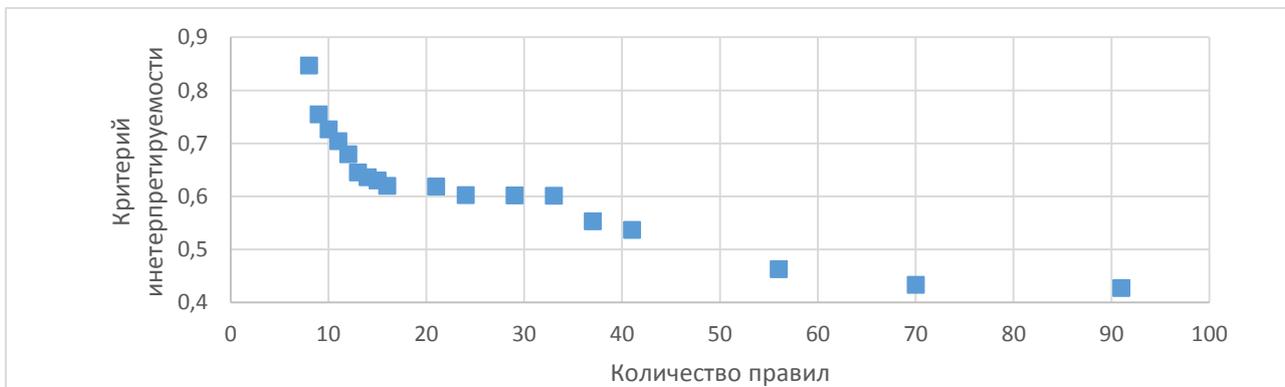


Рис. 2 – Проекция трехкритериального Парето-фронта на оси «Критерий интерпретируемости» и «Количество правил»

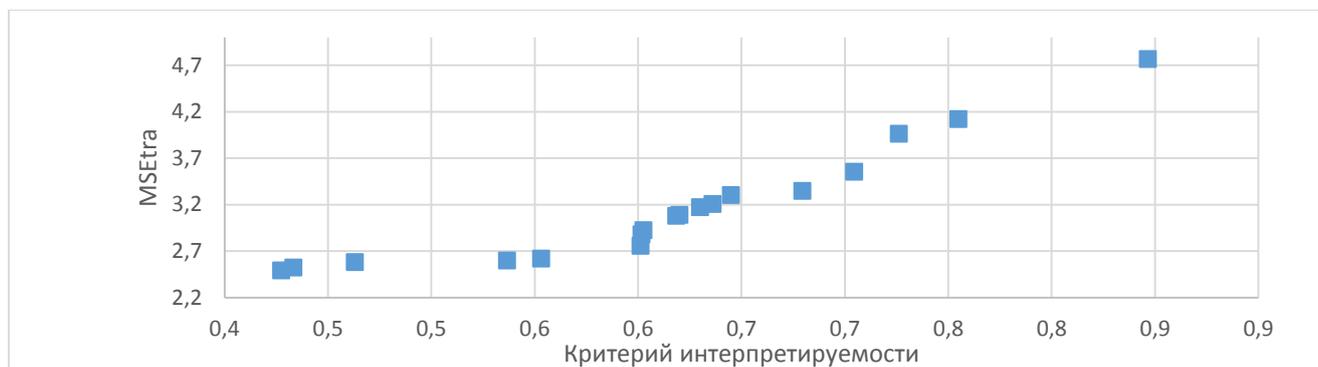


Рис. 3 – Проекция трехкритериального Парето-фронта на оси MSE и «Критерий интерпретируемости».

Пример результатов, полученных гибридным численным методом на данных Ele-2 при сравнении с аналогами, представлен в таблице 1.

Таблица 1. Результаты работы гибридного численного метода идентификации параметров нечеткой системы в сравнении с аналогами

Алгоритм	Обучающая выборка $MSEtra/2$	Тестовая выборка $MSEtst/2$
<i>Гибридный численный метод</i>	6326	7612
TS-NSGA-II	14488	18419
TS-SPEA2	13272	17533
Pulkkinen	9366	10429
TS-SP2-St	17619	22099
Wang–Mendel	56135	56359
COR-BWAS	51332	51370
Thrift	73153	84236
Pittsburgh	105359	132565
Fuzzy-GAP	139583	145031
Pitts-DNF min	101472	106009
Pitts-DNF med	43465	49655
Pitts-DNF max	35104	44009

Примеры результатов инициализации, полученных алгоритмом генерации нечетких классификаторов на основе экстремумов таблицы наблюдений на данных *Iris* в сравнении с аналогами, представлены в таблице 2.

Таблица 2. Результаты инициализации, полученные алгоритмом генерации нечетких классификаторов в сравнении с аналогами

Алгоритм генерации	Процент правильной классификации		Количество правил	Количество термов
	Обучающая	Тестовая		
<i>алгоритм генерации нечетких классификаторов</i>	95,16	90,66	3	12
Равномерное 2x2x2x2	66,67	66,67	16	8
Равномерное 2x3x2x2	82,17	80,67	24	9
Равномерное 2x2x3x3	95,5	97,33	36	10
Равномерное 3x3x3x3	89,33	89,16	81	12

Равномерное 3x3x4x3	95,67	93,33	108	13
Равномерное 4x3x4x3	96,83	94,67	144	14
Fuzzy-C-Means	64,32	64,66	4	16
Gustafson-Kessel	56,83	62,66	4,6	20
Gath-Geva	74,66	76,66	3,4	13,6

Проведено исследование временной эффективности каждой схемы параллельного выполнения алгоритмов и методики.

В таблице 3 представлен пример ускорения алгоритма пчелиной колонии для генерации правил за счет применения многопоточной архитектуры при различных схемах параллельного выполнения.

Таблица 3. Сравнение временных затрат на генерацию 10 правил алгоритмом пчелиной колонии для генерации правил

Номер тестового стенда	Intel i5 2410		Intel i7 3770		AMD Opteon 6272	
	Время выполнения	Ускорение	Время выполнения	Ускорение	Время выполнения	Ускорение
PBS0	31,78	1,00	18,16	1,00	74,52	1,00
PBS1	25,11	1,27	7,61	2,39	36,63	2,03
PBS2	24,55	1,29	7,27	2,50	33,11	2,25
PBS3	23,81	1,33	6,91	2,63	19,29	3,86
PBS4	17,67	1,80	6,78	2,68	20,17	3,69
PBS5	18,87	1,68	7,09	2,56	18,98	3,93

Эксперименты показали, что во многих случаях тестовые стенды одинаково положительно реагируют на одинаковые схемы параллельного выполнения, но в двух случаях эффективность выполнения на тестовом стенде с процессором AMD Opteon 6272 отличается от стендов с процессорами Intel i5 2410 и Intel i7 3770 ввиду отличия в процессоре количества ядер, архитектуры ядер и серверному назначению, которое отражается в дополнительной оптимизации микрокода предсказателя ветвлений.

Исследование показало, что, применяя лучшие схемы параллельного исполнения, можно уменьшить время выполнения алгоритмов идентификации параметров от 1,27 раза до 3,93 раза.

В пятой главе рассмотрено применение разработанных гибридных алгоритмов для назначения комплексов реабилитации пациентам с посттравматическими стрессовыми расстройствами, внедренных в НИИ КиФ, и особенности использования и модификации нечеткого классификатора.

В данном оздоровительном учреждении используется пять комплексов реабилитации. Полученные результаты представлены в таблице 4.

Таблица 4. – Результаты применения разработанных алгоритмов

Номер комплекса	1		2		3		4		5	
	Среднее	СКО								
Набор данных										
Обучающая выборка	83,87	2,42	84,05	3,41	86,04	2,38	84,21	2,44	91,25	2,43
Тестовая выборка	81,25	6,84	76,48	5,22	75,79	5,49	78,57	6,93	78,88	8,15

Предложенный алгоритм для повышения точности распознавания клавиатурного почерка на основе алгоритма идентификации параметров классификаторов и гибридного численного метода внедрен в ОАО ОЭЗ ТВТ.

После недели использования были получены следующие результаты: 9,72% – ошибка первого рода и 4,36% – ошибка второго рода на обучающей выборке; на тестовой выборке: 14,41% – ошибка первого рода и 7,64% – ошибка второго рода. После этого алгоритм использовался в режиме адаптации еще месяц. Результаты после месяца адаптации: 2,79% – ошибка первого рода и 0,69% – ошибка второго рода на обучающей выборке; на тестовой выборке: 4,58% – ошибка первого рода и 1,73% – ошибка второго рода.

В **заключении** сформулированы основные научные и практические результаты.

ЗАКЛЮЧЕНИЕ

1. Предложена методика для построения Парето-оптимальных нечетких моделей по трем критериям. Данная методика позволяет получать нечеткие модели с различным количеством правил, термов и расположением термов. Методика использует элементарные гиперпараллелипеды с динамически изменяемыми размерами, что позволяет добиться равномерного распределения получаемых решений по каждому из критериев. Методика предусматривает возможность дополнения новыми методами идентификации нечетких систем. Использование методики позволяет эксперту выбрать подходящую нечеткую модель из решений, расположенных на Парето-фронте.

2. Предложен оригинальный комплексный критерий, учитывающий геометрическую и семантическую интерпретируемость нечетких систем.

3. Предложен гибридный численный метод, основанный на схеме островов. Метод имеет следующие преимущества: а) гибкость и устойчивость к «застреванию» в локальных экстремумах, полученные за счет применения метода пчелиной колонии; б) скорость сходимости, присущую алгоритму эволюционной стратегии с адаптацией параметров на основе матрицы ковариации; в) точность, определяемую методом наименьших квадратов.

4. Предложен алгоритм генерации нечетких классификаторов на основе экстремумов таблицы наблюдений, позволяющий генерировать более точные не-

четкие классификаторы, чем часто применяемые методы на основе кластеризации *c*-средних и его модификаций Gustafson-Kessel, Gath-Geva, при сравнимом или меньшем количестве правил. Предложенный алгоритм позволяет получить классификаторы, сравнимой точности с методом равномерного разбиения, при меньшем количестве правил.

5. Разработан программный комплекс с оригинальной многопоточной архитектурой, основанной на диспетчеризации параллельных вычислений, позволяющей добиться минимизации времени генерации и оптимизации нечетких систем. Архитектура является расширяемой за счет возможности подключения дополнительных алгоритмов идентификации структуры и параметров.

6. Предложено описание нечетких систем типа синглтон и питтсбургских классификаторов, совместимое со стандартом RMMML 4.2.

7. Внедренная рекомендательная система и результаты проведенного исследования с применением разработанных алгоритмов используются при выполнении НИР «Немедикаментозное восстановительное лечение участников вооруженных конфликтов и чрезвычайных ситуаций» и для назначения комплексов реабилитации пациентам с посттравматическими стрессовыми расстройствами в НИИ КиФ.

8. Проведены исследования эффективности предложенных алгоритмов в задаче распознавания клавиатурного почерка. Осуществлено внедрение в ОАО ОЭЗ ТВТ.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

Работы, опубликованные автором в ведущих рецензируемых научных журналах, рекомендованных ВАК Министерства Образования и Науки Российской Федерации:

1. **Горбунов И.В.** Алгоритмы муравьиной и пчелиной колонии для обучения нечетких систем / И.В. Горбунов, И.А. Ходашинский, П.А. Дудин // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2009. – № 2(20). – С. 157–161.
2. **Горбунов И.В.** Технология усиленной аутентификации пользователей информационных процессов / И.В. Горбунов, Р.В. Мещеряков, И.А. Ходашинский, М.В. Савчук // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2011. – № 2(24). – С. 236–248.
3. **Горбунов И.В.** Методы нечеткого извлечения знаний в задачах обнаружения вторжения / И.В. Горбунов, И.А. Ходашинский, Р.В. Мещеряков // Вопросы защиты информации. – 2012. – № 1. – С. 45–49.
4. **Горбунов И.В.** Оптимизация параметров нечетких систем на основе модифицированного алгоритма пчелиной колонии / И.В. Горбунов, И.А. Ходашинский // Мехатроника, автоматизация, управление. – 2012. – №10. – С. 15–20
5. Построение нечетких систем прогнозирования эффективности немедикаментозного лечения / И.А. Ходашинский, **И.В. Горбунов**, П.А. Дудин, Д.С. Синьков,

А.А. Зайцев // Информатика и системы управления. – 2012. – №3(33). – С. 140–150.

6. **Горбунов И.В.** Алгоритмы генерации структур двухкритериальных Парето-оптимальных нечетких аппроксиматоров / И.В. Горбунов, И.А. Ходашинский, Д.С. Синьков // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2013. – № 1(27). – С. 135–142.

7. **Горбунов И.В.** Алгоритмы поиска компромисса между точностью и сложностью при построении нечетких аппроксиматоров / И.В. Горбунов, И.А. Ходашинский // Автометрия. – Новосибирск: Издательство СО РАН – 2013. – №6(49). – С. 51–61.

Другие работы, опубликованные автором по теме диссертации:

8. **Горбунов И.В.** Особенности построения нечетких классификаторов на основе алгоритма пчелиной колонии // Материалы докладов Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР–2009». – Ч. 2. – Томск: Изд-во «В–Спектр». – 2009. – С. 104–107.

9. Оптимизация параметров нечетких моделей методами роевого интеллекта / **И.В. Горбунов**, И.А. Ходашинский, П.А. Дудин, Д.С. Синьков // Всероссийская конференция с элементами научной школы для молодежи «Проведение научных исследований в области обработки, хранения, передачи и защиты информации»: сборник научных трудов. В 4 т., т. 2. – Ульяновск: УлГТУ. – 2009. – С. 74–82.

10. **Горбунов И.В.** Модифицированный алгоритм пчелиной колонии для тонкой настройки правил нечеткого классификатора // Материалы докладов Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР–2010». Ч. 2. – Томск: Изд-во «В–Спектр», 2010. – С. 109–112.

11. **Горбунов И.В.** Унифицированное представление параметров нечеткой системы / И.В. Горбунов, П.А. Дудин, А.В. Боровков // Материалы докладов Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР – 2011». – Ч. 2. – Томск: Изд-во «В–Спектр», 2011. – С. 168–170.

12. **Горбунов И.В.** Применение нечетких систем для управления температурным полем длинного стального стержня шестигранного сечения / И.В. Горбунов, А.В. Гладков // Материалы докладов Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР–2011». – Ч. 2. – Томск: Изд-во «В–Спектр», 2011. – С. 171–173.

13. Методы вычислительного интеллекта в прогнозировании эффективности немедикаментозного лечения / **И.В. Горбунов**, А.А. Зайцев, И.А. Ходашинский,

П.А. Дудин, Д.С. Синьков // *Материалы V Международной научной конференции (заочной) «Системный анализ в медицине» (САМ 2011)*. – Благовещенск, 2011. – С. 25–28.

14. **Горбунов И.В.** Построения нечетких классификаторов на основе алгоритма пчелиной колонии / И.В. Горбунов, И.А. Ходашинский // *Труды Всероссийской конференции «Знания–Онтологии–Теории», т. 2*. – 2011. – С.117–126.

15. **Горбунов И.В.** Эффект переобучения при использовании модифицированного алгоритма пчелиной колонии для нечеткого аппроксиматора // *Материалы докладов Всероссийской научно–технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР–2012»*. – Ч. 3. – Томск: Изд–во «В–Спектр». – С. 28–33.

16. **Горбунов И.В.** Алгоритмы генерации компактных баз правил для нечеткого аппроксиматора // *Материалы международной заочной научно-практической конференции «Наука, образование, общество: тенденция и перспективы»*. – Ч. 2. — Москва – 2013. – С. 98–104.

17. **Горбунов И.В.** Оценка эффективности генерации баз правил нечеткого аппроксиматора модификациями алгоритма с-средние для задачи Парето оптимизации // *Материалы докладов Всероссийской научно–технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР–2013»*. – Ч. 4. – Томск: Изд–во «В–Спектр». – С. 27–31.

18. **Горбунов И.В.** Генерация структуры двухкритериальных Парето–оптимальных нечетких аппроксиматоров / И.В. Горбунов, И.А. Ходашинский, Д.С. Синьков // *Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник научных трудов VII–й Международной научно–технической конференции*. В 3–х томах. Т.1. – М.: Физматлит, 2013. – С. 458–469.

19. **Горбунов И.В.** Алгоритм параметрической идентификации базы правил нечеткого классификатора на основе эволюционной стратегии / И.В. Горбунов, А.Ц. Гунгаев // *Сборник трудов Всероссийской конференции с международным участием «Современные системы искусственного интеллекта и их приложения в науке»*. – Казань, 2013. – С. 111–118.

20. **Gorbunov I.V.** Algorithms of the Tradeoff between Accuracy and Complexity in the Design of Fuzzy Approximators/ I.V. Gorbunov, I.A. Hodashinsky // *Optoelectronics, Instrumentation and Data Processing*. – 2013. – Vol. 49, № 6. – P. 569–577.

21. **Горбунов И.В.** Отбор информативных признаков для назначения комплекса терапевтического лечения // *Сборник трудов Международной научно-практической конференции «Электронные средства и системы управления»*. – Ч. 1. – Томск, 2013. – С. 132–136.

22. **Горбунов И.В.** Особенности использования нечеткого классификатора и алгоритмов машинного обучения для аутентификации по клавиатурному почерку

// Сборник трудов Международной научно-практической конференции «Электронные средства и системы управления». Ч. 2. – Томск, 2013. – С. 13–18.

23. **Горбунов И.В.** Оценка эффективности параллельной реализации алгоритма пчелиной колонии для идентификации параметров нечеткой системы // Материалы докладов Всероссийской научно–технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР–2014». – Ч. 4 – Томск: Изд–во «В–Спектр» 2014. – С. 34–36.

24. **Горбунов И.В.** Алгоритмы идентификации интерпретируемых и точных нечетких классификаторов / И.В. Горбунов, И.А. Ходашинский // Материалы XII Всероссийского совещания по проблемам управления. Россия, Москва, ИПУ РАН. – 2014 – С. 3269–3280.