

На правах рукописи



**Сарин Константин Сергеевич**

**ГИБРИДНЫЕ АЛГОРИТМЫ АНАЛИЗА ДАННЫХ НА ОСНОВЕ  
КОМПАКТНЫХ И ТОЧНЫХ НЕЧЕТКИХ СИСТЕМ ТИПА  
ТАКАГИ-СУГЕНО**

Специальность 05.13.17

Теоретические основы информатики

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени

кандидата технических наук

Томск – 2016

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

Научный руководитель – доктор технических наук профессор  
Ходашинский Илья Александрович

Официальные  
оппоненты: Пальчунов Дмитрий Евгеньевич,  
доктор физико-математических наук,  
доцент, ФГБУН «Институт математики им.  
С.Л. Соболева Сибирского отделения  
Российской Академии Наук», ведущий  
научный сотрудник лаборатории теории  
вычислимости и прикладной логики  
Силич Виктор Алексеевич,  
доктор технических наук, профессор,  
ФГАОУ ВО «Национальный  
исследовательский Томский  
политехнический университет», профессор  
кафедры оптимизации систем управления  
Института кибернетики

Ведущая организация – ФГБОУ ВО «Иркутский национальный  
исследовательский технический  
университет»

Защита состоится «01» декабря 2016 года в 15 час. 15 мин. на заседании диссертационного совета Д.212.268.05 в ТУСУРе по адресу: 634050, г. Томск, пр. Ленина, 40, ком. 201.

С диссертацией можно ознакомиться в научной библиотеке ТУСУРа по адресу: г. Томск, ул. Красноармейская, 146 и на сайте ТУСУРа по адресу: <https://storage.tusur.ru/files/50444/dissertation.pdf>.

Автореферат разослан «    »                      2016 года.

Ученый секретарь  
диссертационного совета

Евгений Юрьевич  
Костюченко

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### Актуальность работы

В связи с быстрым развитием информационных технологий, интеллектуальный анализ данных широко используется во всех сферах деятельности человека. Одна из его задач – аппроксимация – находит широкое применение в таких разделах анализа данных как машинное обучение, моделирование формирования эмпирического знания и распознавание образов. Нечеткие системы позволяют эффективно решать данную задачу.

Можно выделить следующие основные достоинства нечетких систем (НС):

1) НС типа Такаги-Сугено (далее НС Т-С) способны аппроксимировать любую гладкую нелинейную функцию (доказано В. Kosko), что позволяет моделировать широкий класс сложных нелинейных систем с заданной степенью точности;

2) в силу наличия ЕСЛИ-ТО правил НС позволяют интерпретировать результаты нечеткого моделирования;

3) в силу наличия ЕСЛИ-ТО правил НС позволяют явно учитывать экспертные знания на этапе генерации структуры, что может привести к улучшению свойства сходимости;

4) НС Т-С позволяют относительно легко применять классические и метаэвристические методы оптимизации для их построения.

НС применяются в таких сферах как повышение безопасности ядерных реакторов (разработчики компании Hitachi, Bernard, Nuclear Fuel div.), моделирование судебных процессов (разработчики Meihl Gakuin University, Nagoy University), управление скоростью линий и температурой при производстве стали (компании Kawasaki Steel, New-Nippon Steel, NKK), автоматическое управление воротами плотины на гидроэлектростанциях (компания Tokio Electric Pow.), наведение телекамер при трансляции спортивных событий (компания Omron), прогнозирование землетрясений (институт Seismology Bureau of Metrology, Japan), улучшение эффективности и оптимизация промышленных систем управления (компании Apronix, Omron, Meiden, Sha, Micom, Mitsubishi, Nisshin-Denki, Oku-Electronics) и т.д.

Решением проблем интеллектуального анализа данных с помощью НС занимались А.Н. Аверкин, А.А. Башлыков, И.З. Батыршин, Л.С. Берштейн, А.П. Еремеев, Ю.Н. Золотухин, С.М. Ковалев, Л.Г. Комарцова, Ю.И. Кудинов, А.О. Недосекин, Ф.Ф. Пашенко, Ю.П. Пытьев, В.Б. Тарасов, А.В. Язенин, Н.Г. Ярушкина, М. Akaiwa, С. Altrock, P. Angelov, L.H. Arita, R. Babuska, A. Bastian, J.C. Bezdek, J. Casillas, J.L. Castro, O. Cordon, D. Dubois, D. Filev, J. González, S.

Guillaume, F. Herrera, H. Ishibuchi, W.G. Jacoby, U. Kaymak, S. Kageyama, Y. Kitamura, B. Kosko, R. Krishnapuram, R. Kruse, E.H. Mamdani, J. M. Mendel, S. Oh, W. Pedrycz, H. Prade, H. Sugiura, M. Sugeno, T. Takagi, H. Tanaka, I.B. Turksen, R.R. Yager, T. Yasukawa, L.X. Wang, L. Zadeh.

Практические методики построения НС отстают от теоретических исследований. Анализ существующих методик показывает, что далекой от своего решения является проблема неоптимального представления знаний, связанного с большим количеством нечетких правил. Данная проблема ведет как к усложнению интерпретации полученных правил, так и к вычислительной сложности вывода на основании этих правил. Точность построенной системы входит в противоречие с количеством правил, поэтому в методике построения систем должны присутствовать механизмы, обеспечивающие компромисс между сложностью и точностью. Таким образом, задача создания методики и алгоритмов построения и оптимизации НС, позволяющих построить с одной стороны точные системы, а с другой обладающие простотой, является актуальной.

**Цели и задачи исследования.** Целью диссертационной работы является разработка алгоритмов и методики, повышающих качество анализа данных путем построения нечетких систем типа Такаги-Сугено, сопоставимых по точности и содержащих меньшее количество правил по сравнению с известными подходами, методами и алгоритмами.

Для достижения поставленной цели решены следующие основные задачи:

1) обзор существующих подходов, методов и алгоритмов анализа данных на основе НС;

2) разработка алгоритмов формирования структуры НС Т-С, характеризующихся высокой точностью при сопоставимом количестве правил по сравнению с существующими аналогами, как при нечетком разбиении входного пространства, так и при кластеризации;

3) разработка гибридного алгоритма оптимизации параметров, основанного на композиции метаэвристических алгоритмов и алгоритмов основанных на производных с целью повышения точности вывода НС Т-С;

4) разработка методики построения НС Т-С с использованием разработанных алгоритмов, осуществляющей построение на основе критериев оценивания компромисса между сложностью и точностью для получения компактных и точных НС;

5) исследование разработанных методики и алгоритмов на контрольных примерах и сравнение с аналогами.

**Объектом исследования** является процесс построения НС Т-С.

**Предметом исследования** являются алгоритмы построения НС Т-С.

**Методы исследования.** В диссертационной работе применялись методы оптимизации, интеллектуального анализа данных, теории нечетких множеств, линейной алгебры и процедурного программирования.

**Достоверность результатов** обеспечивается строгостью применения математических методов, результатами проведенных численных экспериментов, а также внедрением результатов работы в практику.

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

1. Разработана оригинальная методика построения компактных и точных НС Т-С на основании анализа экспериментальных данных. Методика отличается от аналогов использованием трех информационных статистических критериев для выбора оптимальной системы и применением оригинальных алгоритмов построения.

2. Разработан алгоритм кусочно-линейной инициализации для генерации структуры НС Т-С. Данный алгоритм основан на процедуре линейной кластеризации и позволяет построить структуру системы с функциями принадлежности гауссова типа на основе анализа экспериментальных данных.

3. Разработан алгоритм генерации структуры НС Т-С с функциями принадлежности треугольного типа, основанный на динамическом разбиении входного пространства. Количество нечетких правил определяется порогом ошибки аппроксимации. Отличительная особенность алгоритма заключается в нахождении участка данных с наибольшей ошибкой и его последующего разбиения.

4. Разработан гибридный алгоритм оптимизации параметров НС Т-С, основанный на композиции метаэвристического алгоритма “кукушкин поиск”, алгоритма градиентного спуска, рекуррентного алгоритма наименьших квадратов.

**Теоретическая значимость работы** заключается в развитии технологии построения НС интеллектуального анализа данных. Разработанный гибридный алгоритм, основанный на композиции алгоритма «кукушкин поиск», алгоритма градиентного спуска и рекуррентного алгоритма наименьших квадратов, может использоваться не только для идентификации параметров НС, но и как

алгоритм глобальной параметрической оптимизации. Кусочно-линейная инициализация может быть использована не только для генерации структур НС Т-С, но и как алгоритм кластеризации.

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

– прогнозирования плотности тока в пучке электронов форвакуумного плазменного источника. Результаты внедрены в ООО «Электронные источники».

– прогнозирования вероятности зажигания разряда в форвакуумном плазменном источнике. Результаты внедрены в ООО «Электронные источники».

Разработанные алгоритмы и методика использованы при выполнении проекта РФФИ 16-07-00034 «Методы и инструментальные средства построения самообучающихся систем, основанных на нечетких правилах» 2016–2018г.

#### **На защиту выносятся приведенные ниже положения.**

1. Разработанный алгоритм кусочно-линейной инициализации (КЛИ) позволяет построить структуры НС Т-С с функциями принадлежности гауссова типа, обладающих более высокой точностью по сравнению с общеизвестным алгоритмом генерации на основе FCRM кластеризации при сопоставимом количестве правил. Точность аппроксимации сформированных алгоритмом КЛИ систем повышается на некоторых наборах данных до 44%.

*Соответствует пункту 5 паспорта специальности: Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.*

2. Разработанный алгоритм динамического разбиения входного пространства позволяет построить структуры НС Т-С с функциями принадлежности треугольного типа, обладающих более высокой точностью по сравнению с общеизвестным алгоритмом генерации на основе равномерного разбиения входного пространства при сопоставимом количестве правил. Точность генерируемых систем повышается до 9.7% на рассмотренных наборах данных.

*Соответствует пункту 5 паспорта специальности: Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.*

3. Разработанный гибридный алгоритм оптимизации параметров НС Т-С, основанный на композиции метаэвристического алгоритма “кукушкин поиск”, алгоритма градиентного спуска и рекуррентного алгоритма наименьших квадратов, позволяет повысить точность аппроксимации НС по сравнению с использованием указанных алгоритмов по отдельности.

*Соответствует пункту 13 паспорта специальности: Применение бионических принципов, методов и моделей в информационных технологиях.*

4. Применение методики построения компактных и точных НС Т-С позволяет существенно уменьшить количество нечетких правил и повысить точность систем до 11.5% на рассмотренных наборах данных.

*Соответствует пункту 2 паспорта специальности: Исследование информационных структур, разработка и анализ моделей информационных процессов и структур.*

**Внедрение результатов диссертационного исследования.** Результаты исследовательской работы вошли в создание системы прогнозирования плотности тока в пучке электронов и прогнозирования вероятности зажигания разряда форвакуумного плазменного источника, используемой в ООО «Электронные источники».

Разработанные алгоритмы генерации структур и оптимизации параметров систем используются при изучении дисциплины «Базы данных и экспертные системы на кафедрах комплексной информационной безопасности электронно-вычислительных систем и безопасности информационных систем ТУСУР.

**Апробация работы.** Основные положения работы докладывались и обсуждались на следующих конференциях, семинарах:

- XII Международной конференция студентов и молодых ученых «Перспективы развития фундаментальных наук», г. Томск, 2015 г.;
- XI Международной IEEE Сибирской конференции по управлению и связи (SIBCON-2015), г. Омск, 2015 г.;
- Всероссийской конференции с международным участием “Знания – Онтологии – Теории” (ЗОНТ–2015), г. Новосибирск, 2015 г.;
- Международной научно-практической конференции «Электронные средства и системы управления», г. Томск, 2015 г.;
- XIII Международной конференция студентов и молодых ученых «Перспективы развития фундаментальных наук», г. Томск, 2016 г.;

– XII Международной IEEE Сибирской конференция по управлению и связи (SIBCON-2016), г. Москва, 2016 г.;

– Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР», г. Томск, 2016 г.;

– XXI Всероссийской конференции «Информационные и математические технологии в науке и управлении» ИСЭМ СО РАН, г. Иркутск, 2016 г.;

– Томском IEEE семинаре «Интеллектуальные системы моделирования, проектирования и управления» г. Томск.

**Публикации по теме диссертации.** По результатам исследований опубликовано 14 печатных работы, из которых в рекомендованных ВАК РФ периодических изданиях – 4. Четыре работы индексируются в базе научных публикаций SCOPUS. Получены 2 свидетельства о государственной регистрации программ для ЭВМ (номера свидетельств: №2015619581, №2015619582).

**Личный вклад автора.** Постановка цели и задач научного исследования и подготовка материалов к печати велась совместно с научным руководителем. Автором самостоятельно разработаны алгоритм кусочно-линейной инициализации, алгоритм генерации структуры динамическим разбиением входного пространства, алгоритмы оптимизации параметров НС и методика построения.

**Структура и объем работы.** Диссертационная работа состоит из введения, четырех глав основной части, заключения списка литературы из 96 наименований и 3 приложения. Основная часть работы содержит 138 страниц, в том числе 52 рисунка и 32 таблицы.

### **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 НС, приводится обзор алгоритмов генерации структуры и оптимизации параметров НС, а также методик построения.

**Постановка задачи.** В диссертационной работе рассматриваются НС Т-С, основная идея которых заключается в разделении пространства входных переменных на области, в которых поведение моделируемой системы можно описать с помощью простых линейных



функций. Правила нечетких систем данного типа имеют следующий вид:

ЕСЛИ  $x_1 = A_{i,1}$  И  $x_2 = A_{i,2}$  И ... И  $x_n = A_{i,n}$  ТО  $y = d_{i,0} + d_{i,1}x_1 + \dots + d_{i,n}x_n$ , где  $A_{i,j}$  – лингвистический терм, которым оценивается входная переменная  $x_j$ ; выход  $y$  задается линейной функцией от входных переменных.

Вывод НС Т-С описывается следующей формулой:

$$f(\mathbf{x}; \boldsymbol{\theta}, \mathbf{D}) = \frac{\sum_{i=1}^R \prod_{j=1}^n \mu_{i,j}(x_j) \cdot (d_{i,0} + d_{i,1}x_1 + \dots + d_{i,n}x_n)}{\sum_{i=1}^R \prod_{j=1}^n \mu_{i,j}(x_j)}, \quad (1)$$

или

$$f(\mathbf{x}; \boldsymbol{\theta}, \mathbf{D}) = \frac{\sum_{i=1}^R \text{MIN}_{j=1}^n \mu_{i,j}(x_j) \cdot (d_{i,0} + d_{i,1}x_1 + \dots + d_{i,n}x_n)}{\sum_{i=1}^R \text{MIN}_{j=1}^n \mu_{i,j}(x_j)}, \quad (2)$$

где  $\mathbf{x}$  – входной вектор,  $R$  – число правил;  $n$  – число входных переменных;  $\mu_{i,j}$  – функция принадлежности  $j$ -й входной переменной  $i$ -го нечеткого правила характеризующая нечеткий терм  $A_{i,j}$ ;  $\boldsymbol{\theta}$  – параметры функций принадлежности антецедентов правил;  $\mathbf{D}$  – параметры линейных функций консеквентов правил.

Целевой функцией при построении системы является среднеквадратичная ошибка аппроксимации между выходами системы и экспериментальными данными:

$$MSE(\boldsymbol{\theta}, \mathbf{D}) = \sum_{p=1}^m \left( y_p - f(\mathbf{x}_p; \boldsymbol{\theta}, \mathbf{D}) \right)^2 / m, \quad (3)$$

где  $\{(\mathbf{x}_p; y_p), p = 1, \dots, m\}$  – множество обучающих данных эксперимента (таблица наблюдений).

**Во второй главе** описывается разработанная методика построения НС Т-С, которая использует критерии оценки НС, учитывающие компромисс между сложностью и точностью: критерий *AIC* (Akaike Information Criterion), *BIC* (Bayesian Information Criterion), *HQC* (Hannan-Quinn Information Criterion).

$$AIC = \ln MSE(\boldsymbol{\theta}, \mathbf{D}) + \frac{2}{m} (m_a + m_c + cR),$$

$$BIC = \ln MSE(\boldsymbol{\theta}, \mathbf{D}) + \frac{\ln m}{m} (m_a + m_c + cR),$$

$$HQC = \ln MSE(\boldsymbol{\theta}, \mathbf{D}) + \frac{2 \ln(\ln m)}{m} (m_a + m_c + cR),$$

где  $MSE(\boldsymbol{\theta}, \mathbf{D})$  – среднеквадратичная ошибка НС на тестовых данных;  $m$  – количество наблюдений;  $m_a$  – общее число параметров antecedентов;  $m_c$  – общее количество параметров консеквентов;  $R$  – число правил;  $c$  – эмпирический коэффициент, учитывающий относительную стоимость правила.

Из множества построенных систем выбирается та, у которой значение суммы рангов по трем информационным критериям минимально.

Ниже приведено пошаговое описание предлагаемой методики.

НАЧАЛО

*Шаг 1.* Выполнить формирование исходных данных.

*Шаг 2.* Задать максимальное количество правил  $R_m$ .

*Шаг 3.* ЕСЛИ функция принадлежности гауссового типа, ТО сгенерировать множество структур алгоритмом кусочно-линейной инициализации, содержащих правил не больше  $R_m$ , перейти на шаг 4;

ИНАЧЕ ЕСЛИ функция принадлежности треугольного типа, ТО сгенерировать множество структур алгоритмом динамического разбиения входного пространства, содержащих правил не больше  $R_m$ , перейти на шаг 4;

ИНАЧЕ ВЫХОД.

*Шаг 4.* Оценить каждую полученную структуру статистическими информационными критериями  $AIC$ ,  $BIC$ ,  $HQC$ . Сформировать множество  $S$  структур с минимальным значением любого критерия среди структур с одинаковым количеством нечетких правил.

*Шаг 5.* Извлечь структуру из  $S$ ;  $j:=0$ ;  $MSEtra_0:=10^{100}$ .

*Шаг 6.*  $j:=j+1$ ;

*Шаг 7.* Оптимизировать  $\boldsymbol{\theta}$  алгоритмом «кукушкин поиск».

*Шаг 8.* Оптимизировать  $\boldsymbol{\theta}$  алгоритмом градиентного спуска.

*Шаг 9.* Оптимизировать  $\mathbf{D}$  рекуррентным алгоритмом наименьших квадратов.

*Шаг 10.* ЕСЛИ  $MSEtra_{j-1} > MSEtra_j$  ТО перейти на шаг 6.

*Шаг 11.* ЕСЛИ  $S$  не пусто ТО перейти на шаг 5.

*Шаг 12.* Оценить каждую нечеткую систему статистическими информационными критериями.

Шаг 13. ЕСЛИ система после оптимизации параметров показала результаты лучше по двум или более критериям, ТО выбрать оптимизированные системы и перейти на шаг 14.

*Шаг 14.* Выбрать нечеткую систему путем ранжирования по сумме критериев.

КОНЕЦ

Здесь  $MSEtra$  обозначает среднеквадратичную ошибку вывода нечеткой системы на обучающей части таблицы наблюдений. Следует заметить, что все оценки критериев на шагах 4 и 12 проводятся на тестовой части таблицы наблюдений, т.е. на данных, которые не входили в этапы формирования структур (шаг 3) и оптимизации параметров (шаги 7–10).

Далее в главе представлены *алгоритмы генерации структуры* НС Т-С: алгоритм кусочно-линейной инициализации (КЛИ) и алгоритм динамического разбиения входного пространства (ДРВП).

Идея алгоритма ДРВП заключается в разбиении входного пространства на нечеткие термы. На этапе инициализации каждое входное пространство равномерно разбивается на один или два нечетких терма таким образом, чтобы ошибка аппроксимации  $MSE$  полученной нечеткой системы достигла заданного порога  $\epsilon$ . Если достигнуть порога невозможно, то этап инициализации заканчивается равномерным разбиением пространства каждой переменной на два терма. Множество функций принадлежности, на которые разбито пространство переменной  $i$  ( $i=1, \dots, n$ ), обозначим  $A_i$ . Антецеденты определяются как все возможные сочетания функций принадлежности из  $\{A_1, A_2, \dots, A_n\}$ , обозначим эту процедуру `GetAntecedentParams`. Консеквенты определяются взвешенным рекуррентным алгоритмом наименьших квадратов над данными таблицы наблюдений  $T$ , обозначим эту процедуру `GetConsequentParams`.

Далее выполняется итерационный процесс, на каждом шаге которого добавляется новая функция принадлежности в одно из множеств  $A_1, \dots, A_n$  и находятся параметры  $\theta$  и  $D$ . Процесс продолжается пока ошибка аппроксимации  $MSE$  нечеткой системы больше заданного порога  $\epsilon$ .

Переменная  $var\_worst$ , в пространство которой добавляется новая функция, находится путем выявления региона  $reg\_worst$  и переменной в этом регионе, которые вносят больший вклад в ошибку  $MSE$ . На пространстве  $var\_worst$  строится новая функция принадлежности с вершиной  $a$ , обозначим эту процедуру `CreateMembership`. Регион входного пространства ограничивается центрами соседних функций принадлежности. На рисунке 1 представлено входное пространство из

двух переменных. Каждая переменная разбита на три нечетких термина, что образует четыре региона  $Reg_1, \dots, Reg_4$ . Множество данных таблицы наблюдений, попавших в регион  $i$ , будем обозначать  $M_i$ .

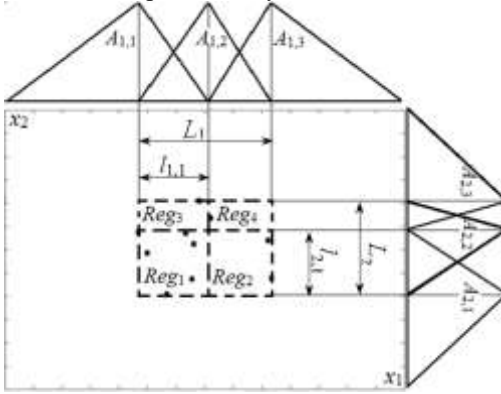


Рис. 1 Построение правил алгоритмом ДРВП

Псевдокод алгоритма приведен ниже.

**Вход:**  $\mathbf{T}$ ,  $\varepsilon$ .

**Выход:**  $\theta$ ,  $\mathbf{D}$ .

Инициализация  $\{A_1, A_2, \dots, A_n\}$ ;

$\theta := \text{GetAntecedentParams}(\{A_1, A_2, \dots, A_n\})$ ;

$\mathbf{D} := \text{GetConsequentParams}(\theta, \mathbf{T})$ ;

цикл пока ( $MSE(\theta, \mathbf{D}) > \varepsilon$ )

$$reg\_worst := \arg \max_{i \in \{1, 2, \dots, CountRegion\}} \frac{\sum_{(x,y) \in M_i} |y - y|}{|M_i|} \cdot \prod_{j=1}^n \frac{l_i^j}{L_j};$$

$$var\_worst := \arg \max_{j \in \{1, 2, \dots, n\}} \frac{\sum_{(x,y) \in M_{reg\_worst}} |y - y|}{|M_{reg\_worst}|} \cdot \frac{l_{reg\_worst}^j}{L_j};$$

$$a := \frac{\sum_{(x,y) \in M_{reg\_worst}} x_{var\_worst} \cdot |y - y|}{\sum_{(x,y) \in M_{reg\_worst}} |y - y|};$$

$\mu := \text{CreateMembership}(a, A_{var\_worst})$ ;

$A_{var\_worst} := A_{var\_worst} \cup \{\mu\}$ ;

$\theta := \text{GetAntecedentParams}(\{A_1, A_2, \dots, A_n\})$ ;

**D :=GetConsequentParams( $\theta, \mathbf{T}$ );**

конец цикла

вывод  $\theta, \mathbf{D}$ .

Здесь  $u$  и  $y$  выход таблицы наблюдений и выход НС соответственно для входного вектора  $\mathbf{x}$ .

КЛИ относится к алгоритмам кластеризации и отличается от предыдущего тем, что здесь не проводится разбиение входного пространства на нечеткие термы. Вместо этого данные таблицы наблюдений разделяются на группы, называемые кластерами, каждая такая группа будет ассоциироваться с правилом нечеткой системы. Кластеры формируются следующим образом. Задается отклонение  $e$ . Находится самая удаленная точка (из входных наблюдаемых данных) от начала координат. Относительно этой точки ведется построение кластера: в группу итерационно добавляются ближайшие к этой точке данные до тех пор, пока среднеквадратичная ошибка между выходными данными, входящими в кластер, и линейной регрессией, найденной с помощью данных в кластере рекуррентным алгоритмом наименьших квадратов, меньше  $e$ . С помощью полученного кластера строится правило.

В этом алгоритме используются функции принадлежности гауссового типа, которые характеризуются двумя параметрами:  $s$  – среднее и  $\sigma$  – отклонение. Чтобы найти эти параметры нужно воспользоваться следующими выражениями:

$$s = \frac{\sum_{k=1}^l x_k}{l}, \quad \sigma = \sqrt{\frac{2}{l} \cdot \sum_{k=1}^l (x_k - s)^2},$$

суммирование здесь ведется по всем данным, входящим в кластер, который представляет правило;  $l$  – количество данных в кластере. Консеквент правила соответствует линейной регрессии, найденной с помощью данных входящих в кластер рекуррентным алгоритмом наименьших квадратов. Построение кластеров и правил продолжается до тех пор, пока не будут пройдены все наблюдаемые данные.

Введем следующие обозначения:  $\text{Err}(C)$  – среднеквадратичная ошибка аппроксимации данных в кластере  $C$  гиперплоскостью;  $\text{FarPoint}(\mathbf{T})$  – самая удаленная точка от начала координат среди множества точек таблицы наблюдений  $\mathbf{T}$ ;  $\text{FarPoint}(P, \mathbf{T})$  – самая удаленная точка от точки  $P$  среди точек таблицы наблюдений  $\mathbf{T}$ ;  $\text{RefreshParams}(C, \theta, \mathbf{D})$  – добавление параметров antecedентов и konsekвентов правила, соответствующего кластеру  $C$ , к  $\theta$  и  $\mathbf{D}$ .

Псевдокод алгоритма приведен ниже.

Вход:  $\mathbf{T}$ ,  $e$ .

Выход:  $\mathbf{\theta}$ ,  $\mathbf{D}$ .

начало цикла ( $|\mathbf{T}|>0$ )

$C := \{\emptyset\};$

$P := \text{FarPoint}(\mathbf{T});$

$C := C \cup \{P\};$

$\mathbf{T} := \mathbf{T} \setminus \{P\};$

начало цикла ( $\text{Err}(C) \leq e$  и ( $|\mathbf{T}|>0$ ))

$p := \text{FarPoint}(P, \mathbf{T});$

$C := C \cup \{p\};$

$\mathbf{T} := \mathbf{T} \setminus \{p\};$

конец цикла

$\text{RefreshParams}(C, \mathbf{\theta}, \mathbf{D});$

конец цикла

вывод  $\mathbf{\theta}$ ,  $\mathbf{D}$ .

Далее представлены алгоритмы оптимизации параметров нечетких систем типа Такаги-Сугено: метаэвристический алгоритм кукушкин поиск, алгоритм градиентного спуска и рекуррентный алгоритм наименьших квадратов. Композиция трех алгоритмов образует гибридный алгоритм оптимизации параметров.

*Метаэвристический алгоритм «кукушкин поиск»* построен на основе имитации способа размножения кукушки, когда она находит недавно построенные гнезда и подкладывает в них свои яйца (заменяет своими), которые в итоге могут быть выкинуты хозяином гнезда. В основе алгоритма лежат три следующих правила: 1) кукушка откладывает по одному яйцу в случайно выбранное гнездо, которое представляет собой решение; 2) часть лучших решений будет перенесена в следующее поколение; 3) количество гнезд фиксировано и есть вероятность того, что хозяин может обнаружить чужое яйцо; в этом случае хозяин может выбросить яйцо из гнезда или вовсе отказаться от гнезда и построит новое на новом месте. Применяется алгоритм для оптимизации параметров antecedentных частей правил  $\mathbf{\theta}$ . Целевая функция оптимизации определена в (3).

Пошаговая реализация алгоритма представлена ниже.

*Шаг 1.* Инициализация исходной популяции.

Задается  $S$  – размер популяции  $\Theta = \{\theta^1, \theta^2, \dots, \theta^s\}$ , «начальное положение кукушки»  $\theta^{\text{cur}}$  устанавливается в значение  $\theta$  после построения структуры или после предыдущих этапов оптимизации параметров НС. Задается  $p$  – вероятность, с которой гнездо может быть «покинуто» хозяином, т.е. вероятность удаления векторов из множества  $\Theta$ . Задается количество итераций  $N$  в качестве критерия остановки и  $m$  – число «худших» решений. Количество генерируемых векторов  $l$  полагаем равным  $S$ .

*Шаг 2.* Случайным образом генерируется  $l$  векторов решений  $\theta$ .

*Шаг 3.* Генерация нового решения на основе полетов Леви.

Выполняется «случайное перемещение кукушки», которое выражено изменением  $\theta^{\text{cur}}$  по закону Леви  $\theta^{\text{cur}} = \theta^{\text{cur}} + \text{Levi}$ , где  $\text{Levi}$  – вектор случайных прыжков полета Леви, элементы которого вычисляются по правилу  $\text{Levi} = \gamma \cdot u / |v|^{1/\beta}$ , где  $\gamma$  – коэффициент прыжка полета Леви; параметр  $\beta$  принимает значения из интервала  $[1, 2]$ ;  $u, v$  – нормально распределенные величины  $v \sim N(0; \sigma_v^2)$ ,

$$u \sim N(0; \sigma_u^2), \sigma_v^2 = 1, \sigma_u^2 = \left\{ \frac{\Gamma(1+\beta) \sin(\pi\beta/2)}{\Gamma[(1+\beta)/2] \cdot 2^{(\beta-1)/2}} \right\}^{1/\beta}, \Gamma(x) - \text{гамма-функция.}$$

Случайным образом выбирается решение  $\theta^{\text{new}}$  из популяции  $\Theta$ .

*Шаг 4.* Оценка качества решения.

Если  $MSE(\theta^{\text{cur}}) > MSE(\theta^{\text{new}})$ , то  $\theta^{\text{cur}} = \theta^{\text{new}}$ .

*Шаг 5.* Удаление плохих решений.

ЕСЛИ не выполнено заданное количество итераций  $N$ , ТО выбирается заранее заданное количество  $m$  «худших» решений; для каждого из них генерируется случайное число  $h$  в диапазоне  $[0, 1]$ , и если значение  $h$  для решения оказывается больше значения заданной вероятности  $p$ , гнездо-решение удаляется; вместо удаленного решения формируется новое;  $l$  полагается равным количеству уничтоженных векторов-решений, переход на шаг 2.

ИНАЧЕ выбирается лучшее решение с минимальным значением  $MSE(\theta)$  и выход из алгоритма.

*Алгоритм градиентного спуска* применяется для оптимизации параметров antecedentes частей правил  $\theta$ , задачей оптимизации является минимизация целевой функции (3). Алгоритм является итерационным, на каждом шаге которого вектор параметров  $\theta$  преобразовывается согласно следующему выражению:

$$\theta = \theta - h \cdot \text{grad}(MSE(\theta)),$$

где  $h$  – длина шага в направлении спуска,  $\text{grad}(MSE(\boldsymbol{\theta}))$  – градиент, вектор, элементы которого показывают направление в сторону наискорейшего возрастания функции  $MSE(\boldsymbol{\theta})$ . В диссертационной работе найдены значения элементов градиента для параметров функций принадлежности гауссового и треугольного типов. Далее будут показаны формулы для нахождения элементов градиента параметров гауссовых функций. Если в нечеткой системе используется вывод (1), то элемент градиента для среднего  $s_{i,j}$   $i$ -го правила  $j$ -й переменной равен:

$$\text{grad}_k = \sum_{p=1}^m \left( f(\mathbf{x}_p, \boldsymbol{\theta}) - y_p \right) \cdot \frac{\left( y_p^i - f(\mathbf{x}_p, \boldsymbol{\theta}) \right)}{\sum_{i=1}^R \prod_{j=1}^n e^{-\left( \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}} \right)^2}} \cdot 2 \cdot \frac{(x_{p,j} - s_{i,j})^2}{\sigma_{i,j}^3} \cdot \prod_{j=1}^n e^{-\left( \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}} \right)^2} \right),$$

и отклонения  $\sigma_{i,j}$ :

$$\text{grad}_k = \sum_{p=1}^m \left( f(\mathbf{x}_p, \boldsymbol{\theta}) - y_p \right) \cdot \frac{\left( y_p^i - f(\mathbf{x}_p, \boldsymbol{\theta}) \right)}{\sum_{i=1}^R \prod_{j=1}^n e^{-\left( \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}} \right)^2}} \cdot 2 \cdot \left( \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}^2} \right) \cdot \prod_{j=1}^n e^{-\left( \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}} \right)^2} \right),$$

где  $y_p^i$  – значение выхода  $i$ -го правила для входного вектора  $\mathbf{x}_p$ .

Если в нечеткой системе используется вывод (2), то элемент градиента для среднего  $s_{i,j}$  принимает следующий вид:

$$\text{grad}_k = \sum_{p=1}^m \left( f(\mathbf{x}_p, \boldsymbol{\theta}) - y_p \right) \cdot \frac{\left( y_p^i - f(\mathbf{x}_p, \boldsymbol{\theta}) \right)}{\sum_{i=1}^R \text{MIN}_{j=1}^n e^{-\left( \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}} \right)^2}} \cdot \frac{\partial u_{i,p}}{\partial s_{i,j}},$$

$$\text{где } \frac{\partial u_{i,p}}{\partial s_{i,j}} = \begin{cases} 2 \cdot \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}^2} \cdot e^{-\left( \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}} \right)^2} & \text{если } \text{MIN}_{j=1}^n e^{-\left( \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}} \right)^2} = e^{-\left( \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}} \right)^2} \\ 0 & \text{если } \text{MIN}_{j=1}^n e^{-\left( \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}} \right)^2} \neq e^{-\left( \frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}} \right)^2} \end{cases}.$$



Элемент градиента для отклонения  $\sigma_{i,j}$ :

$$\text{grad}_k = \sum_{p=1}^m \left( f(\mathbf{x}_p, \boldsymbol{\theta}) - y_p \right) \cdot \frac{(y_p^i - f(\mathbf{x}_p, \boldsymbol{\theta}))}{\sum_{i=1}^R \text{MIN}_{j=1}^n e^{-\left(\frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}}\right)^2}} \cdot \frac{\partial u_{i,p}}{\partial \sigma_{i,j}},$$

$$\text{где } \frac{\partial u_{i,p}}{\partial \sigma_{i,j}} = \begin{cases} 2 \cdot \frac{(x_{p,j} - s_{i,j})^2}{\sigma_{i,j}^3} \cdot e^{-\left(\frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}}\right)^2} & \text{если } \text{MIN}_{j=1}^n e^{-\left(\frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}}\right)^2} = e^{-\left(\frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}}\right)^2} \\ 0 & \text{если } \text{MIN}_{j=1}^n e^{-\left(\frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}}\right)^2} \neq e^{-\left(\frac{x_{p,j} - s_{i,j}}{\sigma_{i,j}}\right)^2} \end{cases}$$

*Рекуррентный алгоритм наименьших квадратов* применяется для оптимизации параметров консеквентных частей правил  $\mathbf{D}$ . В диссертационной работе задача нахождения параметров  $\mathbf{D}$ , минимизирующих (3), сводится к нахождению линейной регрессии.

Вектор параметров консеквентов представим следующим образом:

$$\mathbf{D} = [d_{1,0}, \dots, d_{R,0}, d_{1,1}, \dots, d_{R,1}, \dots, d_{1,n}, \dots, d_{R,n}]^T.$$

Введем матрицу  $\mathbf{Z}$ :

$$\mathbf{Z} = \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{R,1} & x_{1,1} \cdot \beta_{1,1} & \cdots & x_{1,1} \cdot \beta_{R,1} & x_{1,2} \cdot \beta_{1,1} & \cdots & x_{1,n} \cdot \beta_{R,1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \beta_{1,m} & \cdots & \beta_{R,m} & x_{m,1} \cdot \beta_{1,m} & \cdots & x_{m,1} \cdot \beta_{R,m} & x_{m,2} \cdot \beta_{1,m} & \cdots & x_{m,n} \cdot \beta_{R,m} \end{bmatrix},$$

$$\text{где } \beta_{i,p}(\mathbf{x}) = \frac{\prod_{j=1}^n \mu_{i,j}(x_{p,j})}{\sum_{i=1}^R \prod_{j=1}^n \mu_{i,j}(x_{p,j})}, \text{ если используется вывод (1),}$$

$$\text{и } \beta_{i,p}(\mathbf{x}) = \frac{\text{MIN}_{j=1}^n \mu_{i,j}(x_{p,j})}{\sum_{i=1}^R \text{MIN}_{j=1}^n \mu_{i,j}(x_{p,j})}, \text{ если используется вывод (2).}$$

Поскольку этот алгоритм, является итерационным, то  $\mathbf{D}(k)$  будет обозначать значение  $\mathbf{D}$  на  $k$ -шаге. Согласно алгоритму  $\mathbf{D} = \mathbf{D}(k)$ , где  $\mathbf{D}(k)$  вычисляется по следующей схеме:

$$\mathbf{D}(k) = \mathbf{D}(k-1) + \mathbf{P}(k-1) \cdot \mathbf{Z}_k^T \cdot (y_k - \mathbf{Z}_k \mathbf{D}(k-1)),$$

$$\mathbf{P}(k) = \mathbf{P}(k-1) - \frac{\mathbf{P}(k-1) \cdot \mathbf{Z}_k^T \cdot \mathbf{Z}_k \cdot \mathbf{P}(k-1)}{1 + \mathbf{Z}_k \cdot \mathbf{P}(k-1) \cdot \mathbf{Z}_k^T},$$

где  $k = 1, 2, \dots, m$ ;  $\mathbf{P}$  – матрица размерностью  $R(n+1) \times R(n+1)$  и  $\mathbf{P}(0) = \alpha \mathbf{I}$ ,  $\alpha$  – большое число (в работе использовалось  $10^{100}$ ),  $\mathbf{I}$  – единичная матрица;  $\mathbf{D}(0)$  – значение исходных параметров консеквентов нечеткой системы до начала оптимизации.

В третьей главе содержится описание тестовых данных, результаты и анализ проведенных экспериментов генерации структуры, оптимизации параметров и построения НС алгоритмами и методикой, разработанными в работе. Приводится сравнение полученных результатов с алгоритмами и методиками, предложенными другими авторами.

Тестирование алгоритма КЛИ происходило в сравнении с общеизвестным алгоритмом на основе FCRM кластеризации по схеме пятикратной кросс-валидации. Реальные данные были взяты из репозитория *KEEL* (<http://keel.es>). Усредненные значения количества правил ( $R$ ), среднеквадратичной ошибки на обучающей ( $MSE_{tra}$ ) и тестовой выборке ( $MSE_{tst}$ ) приведены в таблице 1.

Таблица 1. Результаты тестов алгоритмов генерации структур КЛИ и алгоритма на основе FCRM кластеризации

Набор данных	Алгоритм на основе FCRM кластеризации			КЛИ		
	$R$	$MSE_{tra}$	$MSE_{tst}$	$R$	$MSE_{tra}$	$MSE_{tst}$
ELE1	2	$3.952 \cdot 10^5$	<b><math>3.994 \cdot 10^5</math></b>	1.2	$4.179 \cdot 10^5$	$4.154 \cdot 10^5$
PLA	1.8	2.349	<b>2.328</b>	1.5	2.332	2.344
QUA	1	0.036	0.036	1	0.036	0.036
ELE2	2.4	$2.443 \cdot 10^4$	$2.446 \cdot 10^4$	3.8	$1.87 \cdot 10^4$	<b><math>1.883 \cdot 10^4</math></b>
FRIE	4	4.857	4.946	5.8	4.573	<b>4.838</b>
MPG6	3.2	9.923	10.209	4.6	8.077	<b>9.057</b>
DELA1L	1.4	$2.95 \cdot 10^{-8}$	$2.96 \cdot 10^{-8}$	1	$2.95 \cdot 10^{-8}$	<b><math>2.95 \cdot 10^{-8}</math></b>
DEE	2.8	0.161	0.166	4.6	0.141	<b>0.154</b>
DELELV	1	$2.09 \cdot 10^{-6}$	$2.09 \cdot 10^{-6}$	1	$2.09 \cdot 10^{-6}$	$2.09 \cdot 10^{-6}$
ANA	2.4	0.09	<b>0.092</b>	1.8	0.17	0.168
MPG8	2.6	9.031	9.513	2.8	9.182	<b>9.477</b>
ABA	1.6	4.802	4.797	1.8	4.681	<b>4.698</b>
CONC	3	82.237	83.833	2.6	71.468	<b>71.651</b>
STP	4.4	3.393	3.307	4.2	1.756	<b>1.841</b>
WAN	3.8	2.985	2.714	3.4	1.707	<b>1.759</b>
WIZ	2.8	1.879	2.025	4.2	1.353	<b>1.419</b>

Результаты показали, что на большинстве из участвующих в эксперименте реальных данных, алгоритм КЛИ при статистически неразличимом по критерию знаков ( $p\text{-value}=0.017$ ) количестве правил, генерирует нечеткие системы, обладающие более высокой точностью по сравнению с алгоритмом на основе FCRM кластеризации (колонка *MSE<sub>st</sub>*). Улучшение ошибки варьировалось в пределах от 2.1% до 44%.

Тестирование алгоритма ДРВП происходило в сравнении с общеизвестным алгоритмом равномерного разбиения входного пространства. Результаты тестов представлены в таблице 2. Результаты показали, что при сопоставимом количестве правил, алгоритм ДРВП генерирует нечеткие системы, обладающие более высокой точностью по сравнению с алгоритмом на основе равномерного разбиения входного пространства. Улучшение ошибки варьировалось в пределах от 0.7% до 9.7%.

Таблица 2. Результаты тестов алгоритмов генерации структур ДРВП и равномерного разбиения входного пространства

Набор данных	Алгоритм равномерного разбиения			Алгоритм ДРВП		
	<i>R</i>	<i>MSE<sub>tra</sub></i>	<i>MSE<sub>st</sub></i>	<i>R</i>	<i>MSE<sub>tra</sub></i>	<i>MSE<sub>st</sub></i>
ELE1	8.8	$3.541 \cdot 10^5$	<b><math>3.704 \cdot 10^5</math></b>	7.2	$3.612 \cdot 10^5$	$3.785 \cdot 10^5$
PLA	9.2	2.287	2.304	8	2.266	<b>2.289</b>
QUA	14.4	0.036	0.036	2.4	0.036	0.036
ELE2	24	$2.379 \cdot 10^4$	$2.36 \cdot 10^4$	22.6	$2.234 \cdot 10^4$	<b><math>2.235 \cdot 10^4</math></b>
FRIE	48	4.478	4.845	48	4.214	<b>4.377</b>
MPG6	48	9.876	10.67	48	9.265	<b>9.863</b>

Результаты применения методики с алгоритмами генерации и оптимизации параметров проводились в сравнении другими методиками построения систем типа Такаги-Сугено.<sup>1</sup>

<sup>1</sup> METSK-HD<sup>e</sup>: A multiobjective evolutionary algorithm to learn accurate TSK-fuzzy systems in high-dimensional and large-scale regression problems / M.J. Gacto, M. Galende, R. Alcalá, F. Herrera // Information Sciences. – 2014. – Vol. 276. – P. 63–79.

Таблица 3. Результаты работы методик построения НС Т-С на разных наборах данных

Набор данных	ANFIS-SUB			TSK-IRL			Linear-LMS		LEL-TSK			METSCK-HD			Методика данной работы		
	R	$\frac{MSEtra}{2}$	$\frac{MSEst}{2}$	R	$\frac{MSEtra}{2}$	$\frac{MSEst}{2}$	$\frac{MSEtra}{2}$	$\frac{MSEst}{2}$	R	$\frac{MSEtra}{2}$	$\frac{MSEst}{2}$	R	$\frac{MSEtra}{2}$	$\frac{MSEst}{2}$	R	$\frac{MSEtra}{2}$	$\frac{MSEst}{2}$
ELE1	27.8	1.513	2.150	19.2	1.414	2.074	1.993	2.093	27	1.190	2.402	11.4	1.350	2.022	<b>2</b>	1.728	<b>1.79</b>
PLA	114	1.011	1.504	21	1.090	1.146	1.166	1.172	66	1.032	1.188	19.2	1.057	1.136	<b>3.2</b>	1.097	<b>1.123</b>
QUA	40.4	0.015	0.155	102	0.0164	0.023	0.0178	0.0179	127	0.0151	0.0308	18.3	0.0171	0.0181	<b>2</b>	0.0176	<b>0.0177</b>
ELE2	<b>2</b>	8208	8525	262	17024	19786	13361	13541	44.8	2928	3752	36.9	2270	<b>3192</b>	3.8	5031	5617
FRIED	53.8	0.085	3.158	3055	0.433	1.419	3.612	3.653	435	0.322	<b>1.070</b>	66	1.075	1.888	<b>3.6</b>	1.4868	1.604
MPG6	299.6	0.002	8.079	785.8	1.338	5.029	5.780	6.084	79	1.473	6.357	83.6	1.082	4.478	<b>2</b>	3.612	<b>4.298</b>
DELAAIL	57.2	0.973	1.484	233.2	1.321	1.419	1.478	1.480	105.2	1.693	1.760	26.8	1.190	<b>1.402</b>	<b>3</b>	1.380	1.404
DEE	290.6	3087	2083	3054	0.545	882.02	0.081	0.085	57.8	0.662	0.682	50.6	0.080	0.103	<b>3</b>	0.0616	<b>0.079</b>
DELELV	<b>2</b>	1.010	1.020	727.4	1.005	1.345	1.048	1.049	219.9	0.9642	2.788	39.1	0.972	1.031	2.4	1.0081	<b>1.0159</b>
MPG8	13.8	1.191	26.136	819	1.784	4.338	5.317	5.678	129	1.574	7.111	64.2	1.154	5.391	<b>3.2</b>	2.8897	<b>3.878</b>
ABA	9	2.008	2.733	2077	2.581	2.642	2.443	2.472	107	2.040	2.412	23.1	2.205	2.392	<b>2.8</b>	2.1637	<b>2.1952</b>
CON	20.6	12.286	188.29	2754	13.167	<b>19.151</b>	53.475	54.735	325	10.692	31.430	53.7	15.054	23.885	<b>5</b>	18.883	23.054

В таблице 3 представлены результаты тестов методик по схеме пятикратной кросс-валидации. На большинстве тестируемых реальных данных с помощью методики и алгоритмов диссертационной работы были построены системы по точности превосходящие результаты рассмотренных аналогов от 0.4% до 11.5%, при этом количество нечетких правил в системах оказалось существенно меньше. Результаты ошибок в таблицах должны быть умножены на  $10^5$ ,  $10^{-8}$ ,  $10^{-6}$  для наборов данных Ele1, Delail, Delelv соответственно.

**В четвертой главе** рассмотрено применение разработанных алгоритмов и методики построения для прогнозирования плотности тока в пучке электронов и прогнозирования вероятности зажигания разряда в плазменном источнике, внедренных в ООО «Электронные источники» г. Томск.

Поставлена задача определения зависимости плотности тока и вероятности зажигания разряда от параметров плазменного источника. Поскольку малоизученными остаются процессы функционирования дуги с катодным пятном в рабочем диапазоне давлений форвакуумных плазменных источников электронов, затруднено построение точных аналитических моделей указанных процессов. Зависимость была определена с помощью построения НС Т-С на основе экспериментальных данных. Приемлемый уровень ошибки позволил применять построенные системы в качестве прогнозирования плотности тока в пучке электронов и вероятности зажигания разряда в плазменном источнике.

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

### **ЗАКЛЮЧЕНИЕ**

В диссертационной работе решена важная задача разработки алгоритмов анализа данных на основе компактных и точных нечетких систем, имеющая существенное значение для развития теории и практики машинного обучения и обнаружения новых знаний.

1. Разработана методика построения компактных и точных нечетких систем типа Такаги-Сугено на основе экспериментальных данных. Данная методика использует алгоритмы построения, разработанные в диссертационной работе, и учитывает компромисс между точностью и сложностей построенных систем благодаря использованию трех информационных статистических критериев: Akaike Information Criterion, Bayesian Information Criterion, Hannan-Quinn Information Criterion. Оптимальная система выявляется на основе суммы рангов по значению трех критериев.

2. Разработан алгоритм кусочно-линейной инициализации, позволяющий генерировать структуры нечетких систем типа Такаги-Сугено. Алгоритм основан на кластеризации данных и генерирует нечеткие системы с функциями принадлежности гауссова типа. Сформированные системы имеют более высокую точность по сравнению с системами, сгенерированными общеизвестным алгоритмом на основе FCRM кластеризации при сопоставимом количестве правил.

3. Разработан алгоритм генерации структуры нечетких систем типа Такаги-Сугено с функциями принадлежности треугольного типа, основанный на динамическом разбиении входного пространства. Особенностью данного алгоритма является то, что количество правил системы определяется алгоритмом, в зависимости от выставленного порога точности. Построенные системы имеют более высокую точность по сравнению с системами, сгенерированными общеизвестным алгоритмом равномерного разбиения входного пространства при сопоставимом количестве правил.

4. Разработан гибридный алгоритм оптимизации нечетких систем типа Такаги-Сугено. Данный алгоритм позволяет повысить точность систем за счет подбора параметров антецедентных и консеквентных частей правил. Особенностью алгоритма является то, что он основан на композиции метаэвристического алгоритма «кукушкин поиск», алгоритма градиентного спуска и рекуррентного алгоритма наименьших квадратов.

5. Разработанные алгоритмы и программы внедрены в ООО «Электронные источники» для создания системы прогнозирования вероятности зажигания разряда и прогнозирования плотности тока в пучке электронов форвакуумного плазменного источника.

#### **СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ**

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

1. Ходашинский И.А., Минина Д.Ю., **Сарин К.С.** Идентификация параметров нечетких аппроксиматоров и классификаторов на основе алгоритма «кукушкин поиск» // Автометрия. – 2015. – Том 51, №3. – С.27-34.

2. **Сарин К.С.**, Медовник А.В., Ходашинский И.А. Нечеткий аппроксиматор значения плотности тока в пучке электронов форвакуумного плазменного источника // Доклады ТУСУР. – 2015. – № 4 (38). – С. 161–165.

3. Ходашинский И. А., **Сарин К.С.**, Черепанов С. А. Об одном методе инициализации нечетких систем типа Такаги-Сугено // Автометрия. – 2016. – Том 52, №2. – С.61-70.

4. Ходашинский И.А., **Сарин К.С.** Методика построения компактных и точных нечетких систем типа Такаги–Сугено // Доклады ТУСУР. – 2016. –Т. 19, № 1. – С. 50–56.

*Другие работы, опубликованные по теме диссертации:*

5. **Сарин К.С.**, Ходашинский И.А. Метод кусочно-линейной инициализации нечетких систем типа Такаги–Сугено // Робототехника и искусственный интеллект: матер. VI Всерос. науч.-техн. конф. с международным участием. Сиб. федер. ун-т. – Красноярск: Центр информации, ЦНИ «Монография». – 2014. – С. 147–152.

6. Hodashinsky I.A., **Sarin K.S.**, Zykov D.D. Takagi-Sugeno fuzzy systems structure identification based on piecewise linear initialization // Proceedings of the 2015 International Siberian Conference on Control and Communications (SIBCON). Omsk: Omsk State Technical University. Russia, Omsk, May 21–23, 2015. – P. 1-4. IEEE Catalog Number: CFP15794-CDR. ISBN: 978-1-4799-7102-2. DOI: 10.1109/SIBCON.2015.7147261. (SCOPUS)

7. **Сарин К.С.**, Ходашинский И.А. Три алгоритма генерации структуры нечеткой системы типа Такаги-Сугено // Материалы докладов Всероссийской конференции с международным участием Знания-Онтологии-Теории (ЗОНТ-2015). Российская академия наук, Сибирское отделение; Институт математики им. С.Л. Соболева. – 2015. – С. 124-132.

8. Hodashinsky I., Minina D. Yu., **Sarin K. S.** Identification of the parameters of fuzzy approximators and classifiers based on the cuckoo search algorithm // Optoelectronics, Instrumentation and Data Processing. – 2015. – V. 51, №. 3. – P. 234-240. ISSN: 8756-6990. DOI 10.3103/S8756699015030048 (SCOPUS)

9. **Сарин К.С.**, Медовник А.В., Зыков Д.Д. Нечеткая система типа Такаги-Сугено прогнозирования вероятности зажигания разряда в форвакуумном диапазоне давлений // Сборник научных трудов XII Международной конференции студентов и молодых ученых. Национальный исследовательский Томский политехнический университет. – Томск, 2015. – С. 1533-1535.

10. Ходашинский И.А. Алгоритмы структурной идентификации компактных и точных нечетких систем / И.А. Ходашинский, И.В. Горбунов, **К.С. Сарин**, С.Р. Субханкулова // Информационные и математические технологии в науке и управлении. – 2016. – Т. 27, №1 – С. 82-93.

11. Hodashinsky I.A. Takagi–Sugeno Fuzzy System to Predict Discharge Ignition Probability in the Foreline Pressure Rang / I.A. Hodashinsky, A.V. Medovnic, **K.S. Sarin**, D.D. Zykov // Key Engineering Materials. – 2016. – Vol. 683. – P. 576-582. (SCOPUS)

12. **Сарин К.С.**, Медовник А.В. Нечеткая имитационная система выбора оптимальных значений плотности тока в пучке электронов форвакуумного плазменного источника // Сборник научных трудов XIII Международной конференции студентов и молодых ученых «Перспективы развития фундаментальных наук». Национальный исследовательский Томский политехнический университет. Томск. – 2016. – Том 7. – С. 120-122.

13. **Сарин К.С.** Гибридная технология построения компактных и точных нечетких систем типа Такаги-Сугено // Материалы международной научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР–2016», Томск, 25–27 мая 2016 г. – Томск: В-Спектр, 2015: – Ч. 3. – С. 233-236.

14. Hodashinsky, I.A., **Sarin, K.S.**, Cherepanov, S.A. Initialization method for fuzzy Takagi–Sugeno systems // Optoelectronics, Instrumentation and Data Processing. – 2016. – Vol. 52, № 3. – P. 266-273. doi:10.3103/S8756699016030080 (SCOPUS)

*Свидетельства о регистрации программы:*

1. Программа структурной идентификации нечетких систем типа Такаги-Сугено методом динамического разбиения входного пространства: свидетельство № 2015619581 Рос. Федерация: / **Сарин К.С.** Ходашинский И.А.; правообладатель ФГБОУВПО ТУСУР. – 2015.

2. Программа структурной идентификации нечетких систем типа Такаги-Сугено методом кусочно-линейной инициализации: свидетельство № 2015619582 Рос. Федерация: / **Сарин К.С.** Ходашинский И.А.; правообладатель ФГБОУВПО ТУСУР. – 2015.