

## **ОТЗЫВ**

### **официального оппонента на диссертацию**

**Сарина Константина Сергеевича**

**на тему «Гибридные алгоритмы анализа данных на основе компактных и точных нечетких систем типа Такаги-Сугено»**

**по специальности 05.13.17 – «Теоретические основы информатики»**

**на соискание ученой степени кандидата технических наук**

### **Актуальность избранной темы**

Быстрые темпы роста объемов данных, скорости их поступления и разнообразие требуют новых эффективных методов и аналитических инструментов обработки накопленных данных с целью извлечения полезной информации. Одним из методов анализа данных является поиск зависимостей в данных и их аппроксимация. В настоящее время наблюдается рост популярности технологий анализа данных, основанных на нечеткой логике, когда анализ может быть проведен посредством лингвистических переменных и правил, содержащих качественную оценку ситуации в виде нечетких «ЕСЛИ-ТО» правил. Такие технологии обеспечивают достаточно высокое качество аппроксимации при меньших затратах, объясняется это следующими причинами: 1) нечеткие системы являются универсальными аппроксиматорами, способными представить любую непрерывную нелинейную зависимость с любой степенью точности; 2) природа нечетких правил позволяет описать анализируемые данные в терминах причинно-следственных отношений между входными и выходными переменными; 3) при построении нечетких систем возможно объединение наблюдаемых данных и знаний эксперта. При этом к нечеткой системе, построенной на основе реальных данных, выдвигаются два основных требования: 1) система должна точно воспроизводить данные из анализируемой таблицы наблюдений и обладать высокими обобщающими способностями; 2) система должна быть представлена в формате, понятном пользователю, и помогать выявить наиболее существенные зависимости и соотношения между входными и выходными переменными, то есть нечеткие правила могут быть интерпретированы пользователем в контексте данного приложения.

Рассмотренная диссертация Сарина К.С. является развитием методов структурной и параметрической идентификации нечетких аппроксиматоров типа Такаги-Сугено с учетом компромисса между их точностью и интерпретируемостью. Однако два указанных критерия – точность и интерпретируемость – являются противоречивыми. Таким образом, можно заключить, что выбранная тема и направленность исследования научной работы являются актуальными.

### **Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации**

Обоснованность результатов диссертационного исследования подтверждена широким использованием научных работ отечественных и зарубежных авторов по теории и практике интеллектуального анализа данных, нечетких систем и

метаэвристических методов оптимизации. Автором изучены и критически проанализированы достижения и положения других авторов по вопросам проектирования нечетких аппроксиматоров. Список содержит 95 литературных источников. Достаточный уровень объективности и адекватности выводов 1) базируется на теоретических и экспериментальных исследованиях, выполненных с использованием современных методов и средств; 2) подтверждается использованием в работе последних публикаций в периодических изданиях по исследуемой тематике. Рекомендации по использованию разработанного алгоритма кусочно-линейной инициализации даны на стр. 49 и алгоритма КЛИsub – на стр. 112; рекомендации по выбору параметров алгоритма «кукушкин поиск» приведены на стр. 61; рекомендации по выбору параметра информационных статистических критериев даны на стр. 35.

### **Достоверность и новизна, полученных результатов**

Достоверность полученных результатов обеспечивается строгостью используемого математического аппарата, адекватностью и достоверностью алгоритмов, установленной путем сравнения реакции алгоритмов и известных аналогов на тестовые данные, большим количеством экспериментальных данных, подтверждающих теоретические результаты, а также внедрением разработанного комплекса алгоритмов в практику.

Новизна проведенных исследований заключается в разработке и апробации новой методики построения компактных и точных нечетких систем типа Такаги-Сугено; разработке алгоритмов структурной идентификации систем типа Такаги-Сугено, основанных на принципах кластеризации и равномерного разбиения входного пространства; разработке нового гибридного алгоритма идентификации параметров нечетких систем типа Такаги-Сугено, в состав которого включены метаэвристика «кукушкин поиск», алгоритм градиентного спуска и рекуррентный алгоритм наименьших квадратов.

### **Теоретическая и практическая значимость полученных автором результатов**

Теоретическая ценность научной работы заключается в развитии теории 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 нечетких систем, но и как алгоритм глобальной оптимизации. Алгоритм кусочно-линейной инициализации может быть применен для решения задач кластеризации.

Практическая значимость представляется разработанным комплексом программ, который может быть применен в системах аппроксимации данных. Полученные автором результаты используются для прогнозирования плотности тока в пучке электронов форвакуумного плазменного источника и для прогнозирования вероятности зажигания разряда в форвакуумном плазменном источнике 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.

### **Оценка содержания диссертации, её завершенность**

Основные результаты диссертации изложены в четырех главах, в которых достигается цель по повышению качества анализа данных путем построения нечетких аппроксиматоров типа Такаги-Сугено, точность которых сопоставима с известными аналогами, но содержащих меньшее количество правил.

Во введении обоснована актуальность темы диссертации, сформулированы цели и задачи работы, указаны научная новизна и практическая значимость полученных результатов, представлена структура диссертации.

В главе 1 автором проведен подробный анализ состояния проблемы исследования, в котором указано место нечетких систем в интеллектуальном анализе данных, обосновано применение нечетких систем типа Такаги-Сугено для решения задач аппроксимации, определены критерии построения нечетких систем, выделены два основных этапа построения систем: генерация структуры и оптимизации параметров, сформулированы требования к алгоритмам, реализующим эти этапы.

В главе 2, основываясь на проведенном литературном обзоре, для соблюдения компромисса «точность-сложность» автор предлагает использовать три информационных статистических критерия, которые положены в основу разработанной четырехэтапной методики построения нечетких систем. На этапе формирования множества структур с помощью алгоритмов кусочно-линейной инициализации и динамического разбиения входного пространства генерируются базы нечетких правил с функциями либо гауссова, либо треугольного типа. Структуры, оценённые статистическими информационными критериями как лучшие, поступают на этап оптимизации параметров. На этом этапе выполняется поиск оптимальных параметров antecedentных и consequentных частей нечетких ЕСЛИ-ТО правил. Положительным моментом данной части работы является применение двух разных типов алгоритмов для оптимизации параметров antecedentов, сначала метаэвристическим алгоритмом кукушкин поиск, потом алгоритмом градиентного спуска. Оптимизация consequentных частей правил выполняется рекуррентным алгоритмом наименьших квадратов. На следующем этапе, после того как множество оптимизированных нечетких систем оценено статистическими информационными критериями, выбирается оптимальная система, в которой соблюден компромисс между точностью и сложностью.

Разработанный автором алгоритм кусочно-линейной инициализации (КЛИ) нечетких систем типа Такаги-Сугено относится к алгоритмам кластеризации данных. Автором представлено два варианта данного алгоритма: добавлением точки в кластер либо удалением точки из кластера. Положительным моментом данной части работы являются рекомендации по применению данных вариантов алгоритма

в зависимости от зашумленности исходных данных и их объема. Алгоритм КЛИ работает с функциями принадлежности гауссова типа. Алгоритм динамического разбиения входного пространства (ДРВП) находит переменные и участки нечеткого разбиения, которые оказывают наибольшее влияние на точность аппроксимации, и разбивает эти участки на более мелкие отрезки, формируя при этом функции принадлежности треугольного типа. Положительным моментом данной части работы являются возможность задавать в качестве параметров алгоритма, как порог точности, так и максимальное количество правил.

Для оптимизации параметров antecedentных частей правил автор впервые предлагает использовать комбинацию: метаэвристика «кукушкин поиск» и алгоритм градиентного спуска, сохраняя тем самым высокие способности глобального поиска, присущие метаэвристике, и не менее высокие способности локального поиска, присущие градиентному алгоритму.

В главе 3 работоспособность предложенной методики и алгоритмов подтверждена многочисленными экспериментами как на синтетических данных, так и на наборах реальных данных из репозитория KEEL. Результаты экспериментов позволили автору сделать вывод о том, что на большинстве участвующих в эксперименте реальных данных, алгоритм КЛИ при сопоставимом количестве правил, генерирует нечеткие аппроксиматоры, обладающие более высокой точностью по сравнению с алгоритмом на основе FCRM кластеризации, а алгоритм ДРВП строит нечеткие аппроксиматоры, обладающие более высокой точностью по сравнению с алгоритмом на основе равномерного разбиения входного пространства. Результаты применения оригинальной методики построения нечетких аппроксиматоров сравнивались с результатами известных методик ANFIS-SUB, TSK-IRL, LEL-TSK, METSK-HD. Сравнительный анализ показал, что при существенно меньшем количестве нечетких правил в системах, на большинстве тестируемых реальных данных точность аппроксимации превосходит точность аппроксиматоров-аналогов.

В главе 4 рассмотрено применение предложенных методики и алгоритмов для решения двух задач: прогнозирования вероятности зажигания разряда от параметров поджигающего импульса в форвакуумном диапазоне давлений при использовании разряда по поверхности диэлектрика и нахождения зависимости плотности тока по сечению пучка от параметров импульса в форвакуумном источнике, расстояния от экстрактора источника и расстояния от центра пучка вдоль его сечения. Ошибка аппроксимации при решении обеих задач не превышает допустимый инженерный уровень, что позволило автору сделать вывод о практической применимости описанного подхода для аппроксимации и прогнозирования.

В приложении приведены копии свидетельств о государственной регистрации программ для ЭВМ( №2015619581 – «Программа структурной идентификации нечетких систем типа Такаги-Сугено методом динамического разбиения входного пространства», №2015619582 – «Программа структурной идентификации нечетких систем типа Такаги-Сугено методом кусочно-линейной инициализации») и акты о внедрении результатов диссертационной работы: в ООО «Электронные

источники»; в Томском государственном университете систем управления и радиоэлектроники.

Таким образом, содержание диссертации последовательное и логичное. Тема и содержание диссертации соответствуют паспорту специальности 05.13.17. Диссертация является законченным научным трудом и выполнена автором самостоятельно на достаточном научном уровне. Диссертация содержит достаточное количество исходных данных, имеет пояснения, рисунки, графики, подробные примеры и расчеты. Написана технически квалифицированно и аккуратно оформлена. По каждой главе и работе в целом имеются выводы.

### **Достоинство и недостатки**

Общей положительной стороной диссертации является то, что для решения слабо структурированных проблем автор использует методологию системного анализа, в котором: осуществляет формулировку проблемной ситуации; определяет цель; определяет критерии достижения цели; проводит построение алгоритмов для обоснования решения; проводит поиск допустимых вариантов решения; согласовывает решение; подготавливает решение к реализации; управляет ходом реализации решения; проверяет эффективность решения. Для решения хорошо структурированных количественно выражаемых проблем автор использует известную методологию исследования операций, которая состоит в построении методики и применении методов для отыскания оптимальной стратегии управления целенаправленными действиями: проводит абстрагирование, формализацию и конкретизацию; проводит анализ и синтез, использует методы индукции и дедукции; проводит композицию и декомпозицию, структурирование и реструктурирование; осуществляет макетирование; реализует алгоритмизацию; осуществляет моделирование и эксперимент; проводит экспертное оценивание и тестирование.

### **Замечания по диссертационной работе в целом**

1. В первой главе следовало бы больше уделить внимания используемым терминам «кластеризация» и «регрессия», поскольку они присутствуют в последующих разделах диссертации.
2. К исследованиям, проведенным в третьей главе, следовало бы добавить результаты применения статистической обработки.
3. Отсутствует обоснование применения функций принадлежности в алгоритмах кусочно-линейной инициализации и динамического разбиения входного пространства.
4. Отсутствуют рекомендации по выбору параметров алгоритма КЛИ-add в зависимости от аппроксимируемых данных.

Отмеченные недостатки снижают качество исследований, но они не влияют на главные теоретические и практические результаты проведенного диссертационного исследования.

## Соответствие автореферата основному содержанию диссертации

Основные этапы научной работы, выводы и результаты представлены в автореферате. Автореферат соответствует основному содержанию диссертации.

## Соответствие диссертации и автореферата требованиям ГОСТ Р 7.0.11-2011

В целом оформление диссертации и автореферата соответствует требованиям ГОСТ Р 7.0.11-2011, однако библиографические записи документов в списке литературы диссертации и автореферата оформлены с отступлением от ГОСТ Р 7.0.11-2011.

## Заключение

Диссертация Константина Сергеевича Сарина является законченным научно-исследовательским трудом, выполненным автором самостоятельно на высоком научном уровне и содержащим новые научные результаты. Рассмотренная в диссертационной работе проблема анализа данных на основе компактных и точных нечетких систем типа Такаги-Сугено соответствует специальности 05.13.17 – «Теоретические основы информатики». В работе приведены научные результаты, позволяющие их квалифицировать как, имеющие существенное значение для развития методов машинного обучения и обнаружения знаний. Полученные автором результаты достоверны, выводы и заключения обоснованы. Работа базируется на достаточном числе исходных данных, примеров и расчетов. Она написана доходчиво, грамотно и аккуратно оформлена. По каждой главе и работе в целом сделаны четкие выводы. Основные научные результаты диссертации опубликованы в 14 печатных работах, из которых в рекомендованных ВАК РФ периодических изданиях – четыре. Четыре работы индексируются в базе научных публикаций SCOPUS. Автореферат соответствует основному содержанию диссертации.

Диссертационная работа отвечает требованиям п. 9 «Положения присуждения ученых степеней», а ее автор Сарин Константин Сергеевич заслуживает присуждения ученой степени кандидата наук по специальности 05.13.17 – «Теоретические основы информатики».

Официальный оппонент,  
доктор технических наук, профессор,  
профессор кафедры оптимизации  
систем управления Института кибернетики  
Национального исследовательского  
Томского политехнического университета  
634050, г.Томск, пр. Ленина, 30.  
(3822) 60-61-32.

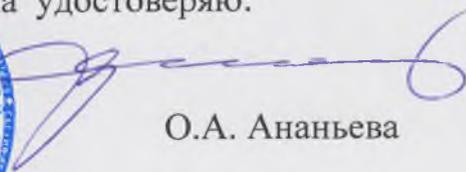
[vas@tpu.ru](mailto:vas@tpu.ru)

Подпись профессора Силича Виктора Алексеевича удостоверяю:  
Ученый секретарь ученого совета  
Национального исследовательского  
Томского политехнического университета



31.10.16

В.А. Силич



О.А. Ананьева