

На правах рукописи



Каратач Сергей Александрович

**Разработка высокопроизводительных методов
интеллектуального анализа данных на основе
нечетких систем при несинглотонной фаззификации**

Специальность 02.03.01 —
«Системный анализ, управление и обработка информации,
статистика»

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Белгород — 2025

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Белгородский государственный технологический университет им. В.Г.Шухова» (БГТУ им.В.Г.Шухова).

Научный руководитель: кандидат технических наук, профессор
Синюк Василий Григорьевич

Официальные оппоненты: **Фамилия Имя Отчество,**
доктор физико-математических наук, профессор,
Не очень длинное название для места работы,
старший научный сотрудник
Фамилия Имя Отчество,
кандидат физико-математических наук,
Основное место работы с длинным длинным
длинным длинным названием,
старший научный сотрудник

Защита состоится **DD mmmmmmmm YYYY** г. в **XX** часов на заседании диссертационного совета БелГУ.22.08 при Белгородском государственном национальном исследовательском университете (НИУ «БелГУ») по адресу: 308015, г. Белгород, ул. Победы, 85, корпус 14, каб. 1-1.

С диссертацией можно ознакомиться в библиотеке **ФГАОУ ВО «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ»).**

Отзывы на автореферат в двух экземплярах, заверенные печатью учреждения, просьба направлять по адресу: 308015, г. Белгород, ул. Победы, 85, корпус 14, каб. 1-1, ученому секретарю диссертационного совета БелГУ.22.08.

Автореферат разослан **DD mmmmmmmm**2025 года.
Электронная почта совета: BSU.22.08@bsuedu.ru; zhikharev@bsuedu.ru.

Ученый секретарь
диссертационного совета
БелГУ.22.08,
кандидат технических наук


Жихарев Александр Геннадиевич

Общая характеристика работы

Актуальность темы. Высокопроизводительный интеллектуальный анализ данных дает возможность принимать обоснованные решения на основе знаний, получаемых посредством обработки данных со скоростью близкой к реальному времени. Семейство методов мягких вычислений с применением техник высокопроизводительного анализа данных открывает возможность находить закономерности и взаимосвязи в данных, содержащих неопределенность. Одним из методов мягких вычислений предназначенным для анализа неопределенных данных являются методы нечеткого моделирования.

В описанной Л. Заде теории нечеткой логики важной проблемой остается задача нечеткого логического вывода. Распространение получили подходы опирающиеся на использование методов нечеткого вывода выработанных Э. Мамдани, П. Ларсеном, Т. Такаги, М. Сугено и Ю. Цукамато. Эти подходы, а также основанные на них производные методы, как правило, используют четкие значения входов и t -норму вместо импликации, что позволяет упростить реализацию нечеткого вывода. Однако такое упрощение приводит к несоответствию с теорией Заде, что можно выявить при рассмотрении лингвистических моделей со многими нечеткими входами, то есть когда используется несингтонный метод фаззификации.

В теории нечеткой логики нечеткий логический вывод реализуется за с помощью обобщенных нечетких правил *modus ponens* и *modus tollens* на основе *композиционного правила вывода*. При нескольких входах вычисление по данным правилом приводит к экспоненциальной зависимости вычислительной сложности от количества входов. Данное ограничение является основным препятствием для применения нечеткого логического вывода с несколькими посылками, тогда как необходимость анализа многомерных данных актуальной задачей, .

Разработку нечеткого вывода с использованием несингтонной фаззификации возродил Д. Мендель. Он продемонстрировал прирост качества нечеткого моделирования с использованием фаззификации типа non-singleton, например, в задаче прогнозирования временных рядов. Однако его исследования ограничены проработкой нечеткого вывода типа Мамдани и Такаги-Сугено. Кроме того Мендель строит формальные выкладки процедуры нечеткого вывода при использовании одной и той же t -нормы, что необоснованно сужает гибкость формул вывода.

Проблемы связанные с нечетким выводом и нечетким моделированием в России изучались и прорабатывались И. А. Ходашинским, Н. Г. Ярушкина, А. И. Аверкин.

Целью данной работы является повышение производительности анализа неопределенных данных путем разработки математического и

программного обеспечения на основе нечетких систем при несинглтонной фаззификации.

Для достижения поставленной цели необходимо было решить следующие задачи:

1. Провести обзор проблем и предлагаемых подходов построения и реализации нечетких систем анализа данных с качественным описанием.
2. Разработать метод вывода на основе нечеткого значения истинности для системы MISO-структуры логического типа, обеспечивающий полиномиальную вычислительную сложность.
3. Выполнить программную реализацию выработанного метода нечеткого вывода с использованием технологии параллельных вычислений CUDA, обеспечив эффективность реализации. Реализовать алгоритм построения базы правил на основе данных.
4. Применить разработанный модуль нечеткого логического вывода для высокопроизводительного анализа зашумленных данных в выбранной предметной области.

Научная новизна:

1. Впервые применено нечеткое значение истинности и принцип обобщения для получения выходного значения при нескольких нечетких входах в соответствии с обобщенным нечетким правилом вывода *modus ponens* для нечетких систем логического типа, в результате чего была получена новая структура базы правил: «Если *истинно*, то B_k ».
2. Разработан метод нечеткого вывода логического типа с использованием нечеткого значения истинности, имеющий полиномиальную вычислительную сложность при многих нечетких входах.
3. Разработан метод регрессии временных рядов с нечеткими оценками измеренных значений на основе предложенного метода нечеткого вывода логического типа и алгоритм построения базы правил
....
4. Разработан параллельный алгоритм, реализующий нечеткий вывод на основе нечеткого значения истинности с применением отбора ...

Теоретическая значимость заключается в расширении класса задач анализа данных, эффективно решаемых при помощи нечеткого моделирования, соответствующего теории нечеткого вывода Л. Заде.

Практическая значимость ...

Методология и методы исследования. В работе использованы методы теории нечетких множеств, нечетких отношений, нечеткого логического вывода, принятия решений и мягких вычислений.

Основные положения, выносимые на защиту:

1. Метод вывода для нечетких систем логического типа на основе нечеткого значения истинности, имеющий полиномиальную вычислительную сложность при многих нечетких входах.
2. Метод регрессии для временных рядов с нечеткими оценками измеренных значений на основе метода нечеткого вывода логического типа.
3. Разработанный вид нечетких правил «Если *истинно*, то B_k ».
4. Разработанный параллельный алгоритм для предложенного метода вывода и его эффективная реализация на графическом процессоре с поддержкой технологии CUDA.

Соответствие диссертации научной специальности. Диссертационная работа соответствует паспорту специальности 2.3.1. «Системный анализ, управление и обработка информации, статистика» по следующим областям исследования:

- п. 10 «Методы и алгоритмы интеллектуальной поддержки при принятии управленческих решений в технических системах».

Внедрение результатов диссертационного исследования. Результаты диссертационной работы внедрены. Предложенные алгоритмы также использованы при выполнении научного проекта при поддержке РФФИ №20-07-00030 «Разработка высокопроизводительных методов интеллектуального анализа данных на основе нечеткого моделирования и создание компьютерной системы поддержки принятия решений для классификации и прогнозирования».

Достоверность полученных результатов обеспечивается корректным применением математического аппарата, экспериментальными исследованиями, апробацией на научно-практических конференциях, доказанностью выводов.

Апробация работы. Основные результаты работы докладывались на:

1. Международная конференция «Перспективные компьютерные и цифровые технологии» (ACDT 2021)», г. Белгород, 2021.
2. XV Международная научная конференция «Параллельные вычислительные технологии (ПаВТ) 2021»», г. Волгоград, 2021.
3. XI Международной научно-практической конференции «Интегрированные модели и мягкие вычисления в искусственном интеллекте (ИММВ-2022)», г. Коломна, 2022 г.
4. XX Национальная конференция по искусственному интеллекту с международным участием (КИИ-2022), г. Москва, 2022.

5. XVII Международная научная конференция «Параллельные вычислительные технологии (ПаВТ) 2021», г. Санкт-Петербург, 2023.
6. XXI Национальная конференция по искусственному интеллекту с международным участием (КИИ-2023), г. Смоленск, 2023.
7. X Всероссийская научно-техническая конференция «Информационные технологии в науке, образовании и производстве» (ИТ-НОП-2025), г. Орел, 2025.

Личный вклад. Все изложенные в диссертации результаты исследования получены либо соискателем лично, либо при его непосредственном участии.

Публикации. Основные результаты по теме диссертации изложены в 9 печатных изданиях, 2 из которых изданы в журналах, рекомендованных ВАК, 2 — в периодических научных журналах, индексируемых Web of Science и Scopus, 4 — в тезисах докладов. Зарегистрированы 2 программы для ЭВМ.

Содержание работы

Введение обосновывает актуальность анализа зашумленных и неопределённых данных, подчёркивает важность мягких вычислений и нечетких систем для обработки информации в условиях неполноты и противоречивости. Указывается, что существующие методы канонического нечеткого логического вывода неэффективны при несинглтонной фаззификации из-за экспоненциальной сложности. Цель — разработка метода нечеткого вывода на основе нечеткого значения истинности (НЗИ), его ускоренная реализация и применение к прогнозированию временных рядов.

Первая глава содержит анализ методов Мамдани, Такаги–Сугено и логического подхода Заде. Подтверждается, что при несинглтонной фаззификации классические схемы приводят к экспоненциальному росту сложности (вывод: необходим новый подход). Вводится формальное определение НЗИ:

$$\mu_{CP(A,A')}(v) = \sup_{\mu_A(x)=v} \mu_{A'}(x)$$

и показано, что для гауссовых функций принадлежности оно вычисляется аналитически:

$$\mu_{CP(A,A')}(v) = \max \left\{ e^{-\frac{((a-c)-b\sqrt{-2\ln v})^2}{2d^2}}, e^{-\frac{((a-c)+b\sqrt{-2\ln v})^2}{2d^2}} \right\}$$

(см. формулу (1.23)). На рисунках ?? и ?? показано, как ширина входного множества влияет на выход, что подтверждает вывод о необходимости учёта фаззификации. Приведены аксиомы истинности, ложности, абсолютной

истинности и квазиистинности (см. рис. ??), что подтверждает универсальность введённого понятия НЗИ.

Вторая глава посвящена построению метода нечеткого вывода на основе НЗИ. Подтверждается, что переход к пространству истинности позволяет снизить сложность с экспоненциальной до полиномиальной:

$$\mu_{B'_k}(y|\mathbf{x}') = \sup_{v \in [0,1]} \{ \tau_{\mathbf{A}_k|\mathbf{A}'}(v) \star I(v, \mu_{B_k}(y)) \}$$

(формула (2.14)), где $\tau_{\mathbf{A}_k|\mathbf{A}'}(v)$ — свёртка НЗИ по входам (формула (2.13)). Приведены конкретные схемы дефаззификации (формулы (2.21), (2.24)), а также показано, что для S- и R-импликаций вычисления упрощаются до поиска максимума НЗИ. В разделе приведён пример применения к авто-регрессионной задаче:

$$\hat{y}_{t+1} = f(y_{t-p}, \dots, y_t)$$

и построения базы правил по окнам временного ряда с фаззификацией через гауссовы функции. Экспериментально подтверждено (см. рис. ??), что новая схема обеспечивает выигрыш по времени и памяти.

Третья глава содержит подробное описание реализации на CUDA/Kokkos. Подтверждается, что организация данных в разделяемой памяти CUDA-блоков (рис. ??) позволяет обрабатывать большие пакеты данных с минимальными задержками. Приведён и реализован алгоритм динамического программирования для свёртки НЗИ (алгоритм ??), что подтверждает снижение сложности до $O(nD_{ftv})$. Для ускорения вывода реализована фильтрация ближайших правил по мере совместимости гауссовых функций:

$$\rho(A_1, A_2) = \sqrt{\int_{-\infty}^{\infty} (g_1(x) - g_2(x))^2 dx}$$

и показано, что это позволяет существенно уменьшить число активных правил без потери качества (см. результаты на рис. ??). Для интеграции с Python реализована обёртка на Cython, что подтверждается успешным использованием модели в экспериментах.

Четвёртая глава содержит экспериментальные подтверждения на реальных данных. Для задачи TWSE показано, что оптимальные параметры ($p = 5$, $\phi = 0.1$, $N = 50$) обеспечивают наилучшее качество (табл. ??), а графики (рис. ??) демонстрируют влияние гиперпараметров. Для задачи эквайринга приведены схемы кластеризации (рис. ??), анализ распределения (рис. ??), а также графики зависимости времени вывода и качества от размера окна (рис. ??, ??). Итоговые таблицы (табл. ??) показывают, что модель превосходит классические и нейросетевые подходы по MAE и

NDEI. Это прямо подтверждает выводы о практической применимости и эффективности метода.

В заключении перечислены основные научные результаты, каждый из которых подтверждён приведёнными формулами, схемами, экспериментальными графиками и таблицами: (1) формализация и аналитическая реализация метода на основе НЗИ; (2) снижение вычислительной сложности и эффективная параллельная реализация; (3) экспериментальная оценка на реальных задачах; (4) создание программной системы с поддержкой GPU и интеграцией в Python.

В **заключении** приведены основные результаты работы, которые заключаются в следующем:

1. Проведен анализ современных методов нечеткого вывода для анализа зашумленных и неопределённых данных. Показано, что существующие подходы либо не обеспечивают канонический логический вывод, либо неэффективны при большом числе входов и использовании несинглтонной фаззификации.
2. Разработан и теоретически обоснован новый метод нечеткого логического вывода на основе нечеткого значения истинности (НЗИ), позволяющий реализовать канонический логический вывод Заде с поддержкой несинглтонной фаззификации и полиномиальной вычислительной сложностью по числу входов. Предложены эффективные алгоритмы вычисления НЗИ и их свертки для многомерных входов.
3. Разработаны и реализованы алгоритмы дефаззификации для различных категорий нечетких импликаций, включая упрощённые схемы для S-, R- и Q-импликаций, что позволило повысить вычислительную эффективность и расширить область применимости метода.
4. Выполнена высокопроизводительная программная реализация предложенного метода с использованием технологий параллельных вычислений на GPU (CUDA, Kokkos). Разработаны оптимизации по организации памяти и вычислений, обеспечивающие масштабируемость и эффективность при обработке больших массивов данных.
5. Проведена экспериментальная проверка разработанной модели на задачах прогнозирования временных рядов в экономике и финансах. Показано, что предложенная модель обеспечивает сопоставимое или лучшее качество прогнозирования по сравнению с классическими и нейросетевыми подходами, а также обладает высокой вычислительной производительностью.
6. Практическая значимость работы подтверждена реализацией программного комплекса с удобным интерфейсом для интеграции в современные системы анализа данных, а также возможностью

применения в задачах поддержки принятия решений, прогнозирования и интеллектуального анализа данных.

7. Перспективы дальнейших исследований связаны с развитием методов автоматического построения базы правил, адаптивной фаззификации, расширением класса функций принадлежности, а также применением разработанного подхода к задачам с многомерными и потоковыми данными.

При использовании пакета `biblatex` список публикаций автора по теме диссертации формируется в разделе «Публикации.» файла `common/characteristic.tex` при помощи команды `\nocite`

Каратач Сергей Александрович

Разработка высокопроизводительных методов интеллектуального анализа
данных на основе нечетких систем при несинглтонной фаззификации

Автореф. дис. на соискание ученой степени канд. техн. наук

Подписано в печать _____._____._____. Заказ № _____

Формат 60×90/16. Усл. печ. л. 1. Тираж 100 экз.

Типография _____