

## ОТЗЫВ

официального оппонента

на диссертацию на соискание ученой степени

кандидата физико-математических наук

Васильева Юлия Алексеевича

на тему: «Исследование и разработка методов

машинного обучения анализа выживаемости»,

по специальности 2.3.5 «Математическое и программное обеспечение

вычислительных систем, комплексов и компьютерных сетей»

### **Актуальность темы диссертации**

Существует множество прикладных задач анализа событий, при решении которых необходимо моделировать распределение вероятности времени наступления событий. Инструментарий традиционных интеллектуальных систем направлен на прогнозирование времени и вероятности событий, а для прогнозирования распределений вероятности времени событий используются методы анализа выживаемости.

При описании целевого события методы анализа выживаемости учитывают информацию о времени до наступления события и факт наступления события. В такой формулировке возможна обработка цензурированных наблюдений с неизвестным временем события, в то время как традиционные интеллектуальные методы игнорируют данные наблюдения. Более того, модели анализа выживаемости обеспечивают прогноз значений функции риска и функции выживания, что позволяет оценить вероятность наступления события в последующие моменты времени.

Рассматривая задачу построения прикладных интеллектуальных систем, нельзя не отметить сложность применения некоторых программных решений к реальным данным из-за наличия признаков различного типа (как числовых, так и строковых значений), а также отсутствия некоторых значений в силу ограниченности информации или возникновения ошибок при сборе данных. Другой проблемой является применение моделей со строгими предположениями, что может быть некорректным при их нарушении на реальных данных.

Важно отметить, что актуальным также является разработка методов построения интеллектуальных моделей, позволяющих проводить описание и

интерпретацию наблюдаемых зависимостей и достигать высокой точности в задачах прогнозирования. Безусловно удачным решением автора диссертации является выбор класса древовидных методов для решения поставленных задач, поскольку теоретически их применение не требует явной предобработки исходных данных. По мнению автора, для применения древовидных подходов необходимо разработать методы обработки разнородных данных на этапе поиска разбиения и построения листовых моделей, не используя строгие формальные предположения.

Таким образом, диссертационная работа Васильева Ю.А. направлена на решение **актуальной** задачи – разработке методов и средств построения современных интеллектуальных систем, использующих методы анализа выживаемости.

### **Основное содержание работы**

В главе 1 рассматриваются существующие стратегии сбора данных событий, а также методы построения прогнозных моделей и оценки качества. Рассматриваются подходы сбора данных, выделяя перспективные направления применения анализа выживаемости для прогнозирования целевых событий из разных прикладных областей. В частности, рассматриваются стратегии постановки исследования, возможные типы цензурирования наблюдений и особенности данных. Для каждой из возможных прогнозируемых величин (оценки вероятности и времени до события, а также функции риска и выживания) приводятся различные оценки качества и сравниваются их свойства. Для существующих методов построения интеллектуальных моделей обосновывается ограниченность их функциональности. Например, при построении прогнозов функций непрерывного времени используются статистические предположения случайности цензурирования и построения прогноза через масштабирование базовых оценок функций риска и выживания. Альтернативный подход построения моделей без строгих предположений имеет ограниченную функциональность, позволяя оценивать только точечные или дискретные риски события. Проведенное исследование подчеркивает значимость выбора класса моделей на основе деревьев решений.

В главе 2 решается задача разработки методов построения интерпретируемых деревьев выживаемости, применимых к реальным данным. Для преодоления статистических предположений, используемых при построении дерева, автором



предлагается комплекс алгоритмов для расчета взвешенного критерия разбиения, регуляризации критерия для обработки информативного цензурирования и построения оценок функций в листовых узлах дерева. Применимость методов к реальным данным обеспечивается возможностью поиска разбиения данных по числовым и категориальным переменным (отображая категории признаков на числовую шкалу) с учетом возможных пропусков.

Глава 3 посвящена исследованию применимости существующих метрик анализа выживаемости для оценки качества прогнозирования с дальнейшим проведением экспериментального исследования качества предложенного дерева выживаемости. Показано, что существующие метрики обладают рядом недостатков при оценке качества прогнозирования, определяя неравный вклад для разных наблюдений. Предложенные модификации показали большую устойчивость к особенностям реальных данных и были выбраны для оценки качества прогнозных моделей. В ходе экспериментального исследования было показано, что использование весовых схем в критерии разбиения, а также регуляризации и модифицированной непараметрической оценки приводит к повышению качества прогнозирования предложенных деревьев выживаемости, превосходя существующий подход по качеству и функциональности.

В главе 4 решается задача улучшения качества предсказаний деревьев выживания за счет ансамблирования или комбинирования прогнозов нескольких моделей. Предлагаются подходы усреднения базовых моделей и адаптивного ансамблирования с переВыборкой более сложных для прогнозирования наблюдений. В ходе экспериментального исследования было показано, что использование предложенной модификации интегральной метрики Брайера в качестве функции потерь улучшает качество прогнозирования, а результаты экспериментального сравнения качества методов говорят о превосходстве разработанных методов анализа выживаемости.

Глава 5 содержит описание разработанной интеллектуальной системы анализа событий в виде библиотеки анализа выживаемости. Приводится описание архитектуры, модулей и особенностей программной реализации. В главе представлены основные сценарии использования библиотеки, такие как: сбор и подготовка данных, построение непараметрических моделей, построение деревьев выживаемости и интерпретация зависимостей, построение ансамблей деревьев

выживаемости, оценка качества прогнозирования. Глава завершается экспериментальной оценкой вычислительных затрат предложенных моделей.

**Научная новизна, обоснованность и достоверность научных положений, выводов и рекомендаций, сформулированных в диссертации**

Научная новизна результатов диссертации Васильева Ю.А. заключается в разработке комплекса новых алгоритмов для построения моделей анализа выживаемости. Алгоритмы осуществляют построение моделей путем выявления закономерностей в данных без статистических предположений, работая с реальными данными различного рода (числовые и категориальные значения с возможными пропусками). В частности, предложенный метод построения деревьев выживаемости обеспечивает получение интерпретируемых прогнозов, а предложенные ансамбли деревьев демонстрируют точные прогнозы функций выживания и риска при решении задач анализа реальных медицинских данных.

Практическая значимость результатов диссертации Васильева Ю.А. заключается в том, что предложенные алгоритмы, реализованные в виде программной библиотеки анализа выживаемости и могут быть использованы для решения различных прикладных задач анализа событий и разработки альтернативных современных интеллектуальных систем.

Таким образом, результаты диссертации являются новыми, имеют теоретическое и практическое значение для исследования и построения интеллектуальных систем анализа выживаемости.

Все сформулированные в диссертации утверждения, их обоснования и заключительные выводы корректны. Они согласуются с проведенными вычислительными экспериментами и известными результатами в области решения задач анализа выживаемости. Цитирование работ других авторов полное. Диссертация содержит необходимые библиографические ссылки, имеющие отношения к теме исследования.

Основные результаты и положения достаточно полно изложены в 4 работах, опубликованных в изданиях, рекомендованных для защиты в диссертационном



совете МГУ им. М.В. Ломоносова по специальности 2.3.5. Зарегистрированы права на ПО для ЭВМ.

### **Замечания**

Недостаточно полно проведена экспериментальная оценка вычислительных затрат предложенных алгоритмов.

Дополнительно следовало бы провести оценку вычислительной сложности предложенных алгоритмов.

Объемы используемых наборов данных не позволяют сделать выводы о масштабируемости предложенных решений для прикладных задач на больших объемах данных.

В тексте диссертации встречаются опечатки.

Вместе с тем, указанные замечания не умаляют значимости диссертационного исследования.

### **Заключение**

Диссертация Ю.А. Васильева является завершенной научно-квалификационной работой, выполненной автором самостоятельно на высоком научном уровне, и вносит значимый вклад в теорию и практику разработки интеллектуальных систем анализа выживаемости.

Работа актуальна, имеет теоретическую и практическую значимость. Полученные автором результаты достоверны, выводы и заключения обоснованы. По каждой главе и работе в целом сделаны выводы. Автореферат диссертации правильно и полно отражает её содержание и основные положения, выносимые на защиту.

Диссертация отвечает требованиям, установленным Московским государственным университетом имени М. В. Ломоносова. Содержание диссертации соответствует специальности **2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»** (по физико-математическим наукам), а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова, а также оформлена согласно требованиям Положения о совете по защите диссертаций на соискание ученой степени кандидата наук Московского государственного университета имени М.В. Ломоносова.

Таким образом, соискатель **Васильев Юлий Алексеевич** заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Официальный оппонент:

доктор физико-математических наук, профессор,  
главный научный сотрудник Отделения суперкомпьютерных систем и параллельных вычислений Федерального государственного бюджетного учреждения «Национальный исследовательский центр «Курчатовский институт»

Сотников Александр Николаевич

Дата: 10.12.2024

Контактные данные:

Тел.: +7 (985) 767 43 57, email: [alexandersotnikov@yandex.ru](mailto:alexandersotnikov@yandex.ru),

Специальность, по которой официальным оппонентом защищена диссертация:

05.13.16 «Применение вычислительной техники, математического моделирования и математических методов в научных исследованиях»

Адрес места работы: 119334, Москва, Ленинский проспект, д. 32А,

Федеральное государственное бюджетное учреждение «Национальный

исследовательский центр «Курчатовский институт», отделение суперкомпьютерных систем и параллельных вычислений

Тел.: +7 (495) 938 18 75, email: [asotnikov@jsec.ru](mailto:asotnikov@jsec.ru)

Подпись А.Н.Сотникова заверяю

Адрес: Москва, Ленинский проспект, д. 32А

Первый заместитель

Главного ученого секретаря

К.Е. Борисов