

ОТЗЫВ

научного руководителя на диссертацию Васильева Юлия Алексеевича «Исследование и разработка методов машинного обучения анализа выживаемости», представленной на соискание ученой степени кандидата физико-математических наук по специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

Во время эпидемии COVID-19 возникла острая необходимость анализа сырых, часто противоречивых, недостоверных и низкокачественных данных, массово собираемых в процессе исследования заболеваемости населения. Традиционные методы статистического анализа, такие как метод Кокса, предполагают формирование страт пациентов по заданным предварительно правилам и с достоверными характеристиками, формальную проверку применимости моделей, корректное заполнение пропусков и построение моделей выживаемости в рамках контролируемого эксперимента.

Однако в условиях пандемии выполнение этих требований к данным было невозможно, и в то же время была необходимость разработки интеллектуальных систем на основе машинного обучения для построения прогнозных моделей и выявления ключевых факторов, влияющих на выживаемость, длительность госпитализации, время появления иммунитета и другие показатели, для которых применяются методы анализа выживаемости. Поэтому в качестве альтернативы классическим параметрическим моделям были использованы непараметрические модели. Хотя последние обладают недостатками в виде невысокой точности и слабой интерпретируемости, они послужили отправной точкой для разработки более совершенных программных инструментов анализа. В процессе решения данной прикладной задачи был использован подход на основе стратифицированных непараметрических моделей с автоматическим выбором параметров стратификации, позволяющих описывать страты пациентов, обладающих общими характеристиками, приводящими к схожему поведению функции выживаемости и риска. Развитие этого подхода привело к разработке алгоритмов типа дерева решений, которые позволяли выделить ключевые признаки и строить интерпретацию в виде правил для описаний областей в пространстве признаков пациентов. Предложенное в работе

использование взвешенных log-rank критериев для определения оптимальных границ разбиения страт в условиях цензурированной выборки позволило пренебречь предположением о пропорциональности рисков, а также контролировать важность определенных событий во времени.

В ходе работы над этими моделями решались традиционные проблемы, такие как обработка малых страт, неравномерное влияние признаков разной мощности, переобучение, информативность цензурирования и наличие пропусков, а также корректное автоматическое кодирование категориальных признаков. Для решения этих задач были предложены оригинальные алгоритмические решения, изложенные в диссертационной работе и реализованные в программной библиотеке. Однако предложенные интерпретируемые модели на основе деревьев решений не всегда обеспечивали высокое качество прогноза. В связи с этим были разработаны методы ансамблирования этих моделей. Модель бутстреп ансамбля строит множество независимых моделей на бутстреп подвыборках с возвращением, а размер ансамбля определяется по лучшему качеству прогнозирования на внешней (out-of-bag) бутстреп выборке. Предложенная модель адаптивного бустинг ансамбля с перевыборкой основана на последовательном построении базовых моделей, где каждая последующая модель обучается на наиболее сложных для прогнозирования наблюдениях от предыдущих моделей. При решении задачи выбора наилучших размеров и характеристик ансамбля было обнаружено, что не все традиционные оценки качества чувствительны к характеристикам исходной выборки. Это привело к необходимости дополнительного исследования свойств чувствительности различных метрик качества, которое также включено в диссертационную работу Васильева Ю.А.

Экспериментальная оценка качества предложенных методов в работе проводилась на шести различных медицинских наборах данных, описывающих показатели пациентов при исследовании летального исхода, рецидива и выписке. По результатам экспериментального исследования, предложенный комплекс алгоритмов превзошел существующие реализации, а предложенные методы ансамблирования превзошли существующие статистические методы и методы машинного обучения.

Все полученные алгоритмические результаты были реализованы в виде открытой программной библиотеки на языке Python, которая может использоваться широким кругом заинтересованных специалистов. Было получено свидетельство о регистрации прав на ПО (№202468193).

При выполнении диссертационной работы Ю. А. Васильев проявил себя зрелым исследователем, способным как к самостоятельному выдвиганию новых идей при решении задач в сложной области, так и к их практической программной реализации. Опубликованная в рамках настоящего исследования работа «Survival Analysis Algorithms based on Decision Trees with Weighted Log-rank Criteria» была отмечена оргкомитетом международной конференцией «International Conference on Pattern Recognition Applications and Methods» (3-5 февраля 2022, Вена, Австрия) как лучшая работа. Опубликованные работы по теме диссертации были отмечены премиями по программе «Конкурс работ, способствующих решению задач Программы развития Московского университета в 2022–2024 гг.» в номинации «Выдающиеся научные статьи». В 2024 работы Васильева Ю.А. по этой тематике получил грантовую поддержку для молодых ученых без степени от некоммерческого фонда развития науки и образования «Интеллект».

Полученные в диссертации результаты являются новыми, их достоверность подтверждается экспериментально, а также публикацией полученных результатов в рецензируемых индексируемых российских и международных научных изданиях.

Диссертационная работа выполнена на высоком уровне, представляет собой законченное научное исследование и удовлетворяет всем требованиям, предъявляемым к кандидатским диссертациям. Ее результаты получены лично автором и прошли квалифицированную апробацию. Автореферат адекватно отражает содержание диссертации. Рекомендую присудить ее автору Ю. А. Васильеву ученую степень кандидата физико-математических наук по специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Петровский Михаил Игоревич,
Кандидат физико-математических наук,
Ученый секретарь кафедры, доцент,
ИИТ ВМК МГУ имени М.В. Ломоносова
(119991, Москва, Ленинские горы, д. 1, стр. 4)
Телефон: +7 910 470 73 76
Адрес электронной почты: michael@cs.msu.su



Т.Г. Коваленко

ИИТ ВМК МГУ имени М.В. Ломоносова