

Применение методов причинно-следственного анализа и машинного обучения для автоматизированной интерпретации медицинских данных

Д.Б. Аллити^{1*}, И.В. Ильин¹

¹Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия
allitidb@alliti.ru

Аннотация. В условиях стремительного развития методов машинного обучения и причинно-следственных методов их интеграция в медицинские исследования приобретает особую актуальность. Внедрение подходящих методов в медицинскую область позволит проводить качественную оценку эффективности лечения на индивидуальном уровне. Цель настоящего исследования – проведение экспериментов на синтетических данных и оценка точности предсказания индивидуальных эффектов лечения с использованием методов T-learner и S-learner. В статье представлен интегрированный подход к анализу медицинских данных, сочетающий методы причинно-следственного анализа и алгоритмы машинного обучения. Впервые проведено комплексное сравнение эффективности методов T-learner и S-learner для оценки индивидуальных эффектов лечения. На основе симулированных данных экспериментально установлены условия оптимального применения этих методов в зависимости от характеристик клинических данных. По результатам проведенных экспериментов было установлено, что метод T-learner показал более высокую точность (87%) по сравнению с S-learner (84%), что делает его предпочтительным при существенных различиях между группами лечения и контроля. Однако метод S-learner показал более высокую обобщающую способность в условиях ограниченного объема данных. Для валидации предсказанных эффектов лечения применялся индекс c-for-benefit, результаты которого подтвердили высокую точность обоих методов. Полученные результаты подчеркивают потенциал интеграции методов машинного обучения и причинно-следственных методов для разработки персонализированных терапевтических стратегий и автоматизации анализа медицинских данных, позволяющих улучшить клинические исходы и качество лечения. Разработанный подход позволяет повысить точность прогнозирования результатов лечения на индивидуальном уровне и может быть интегрирован в системы поддержки принятия клинических решений. Полученные результаты предоставляют новые возможности для персонализации медицинской помощи и могут служить основой для последующих исследований в этой области.

Ключевые слова: машинное обучение, причинно-следственный анализ, оценка эффекта лечения, персонализированная медицина

Для цитирования: Аллити Д.Б., Ильин И.В. Применение методов причинно-следственного анализа и машинного обучения для автоматизированной интерпретации медицинских данных // Прикладная информатика. 2025. Т. 20. № 2. С. 65–78. DOI: 10.37791/2687-0649-2025-20-2-65-78

Application of causal inference and machine learning methods for automated interpretation of medical data

D. Alliti¹*, I. Ilin¹

¹Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, Russia
*allitidb@alliti.ru

Abstract. In the context of rapid advancements in machine learning and causal inference methodologies, their integration into medical research is of paramount importance. Implementing appropriate methods in the medical domain facilitates robust assessment of treatment efficacy at the individual level. This study aims to conduct experiments on synthetic data and evaluate the accuracy of predicting individual treatment effects using T-learner and S-learner methods. The article presents an integrated approach to medical data analysis, combining causal inference techniques with machine learning algorithms. For the first time, a comprehensive comparison of the effectiveness of T-learner and S-learner methods in assessing individual treatment effects has been conducted. Based on simulated data, the study experimentally determines the optimal application conditions for these methods, depending on the characteristics of clinical data. The experiments revealed that the T-learner method demonstrated higher accuracy (87%) compared to the S-learner (84%), making it preferable when there are significant differences between treatment and control groups. However, the S-learner method exhibited greater generalization capability in scenarios with limited data volume. The c-for-benefit index was employed to validate the predicted treatment effects, with results confirming the high accuracy of both methods. These findings underscore the potential of integrating machine learning and causal inference methods to develop personalized therapeutic strategies and automate medical data analysis, thereby improving clinical outcomes and treatment quality. The developed approach enhances the precision of predicting treatment outcomes at the individual level and can be integrated into clinical decision support systems. The presented results offer new opportunities for personalized healthcare and can serve as a foundation for subsequent research in this field.

Keywords: machine learning, causal inference, treatment effect evaluation, personalized medicine

For citation: Alliti D., Ilin I. Application of causal inference and machine learning methods for automated interpretation of medical data. *Prikladnaya informatika*=Journal of Applied Informatics, 2025, vol.20, no.2, pp.65-78 (in Russian). DOI: 10.37791/2687-0649-2025-20-2-65-78

Введение

В настоящее время развитие цифровой медицины способствует поиску новых возможностей для повышения качества клинических исследований и оптимизации лечебных вмешательств. Современные методы машинного обучения предлагают мощные инструменты для анализа больших объемов данных, а причинно-следственные методы позволяют устанавливать более точные взаимосвязи

между вмешательствами и исходами [1]. В условиях растущего объема медицинских данных и развития цифровых технологий методы машинного обучения и причинно-следственные методы становятся неотъемлемой частью современной медицины [2]. Их интеграция в медицинскую отрасль позволяет существенно улучшить точность оценки эффективности лечения на индивидуальном уровне. Одним из перспективных направлений яв-