

Состязательное устранение смещения для справедливого прогнозирования диабета по возрасту: компромисс между производительностью и справедливостью и зависимость от разбиения данных в машинном обучении

Источник: Frontiers in Digital Health

Оригинал: <https://www.frontiersin.org/articles/10.3389/fdgth.2026.1816806>

дебиасинг

диабет

машинное обучение

прогностическая аналитика

справедливость алгоритмов

Введение

Модели машинного обучения, используемые для прогнозирования риска развития диабета, могут содержать связанные с возрастом систематические ошибки (biases), которые снижают точность диагностики для определенных демографических групп. **Состязательное устранение смещения** (adversarial debiasing) с использованием **слоя реверсии градиента** (gradient reversal layer, GRL) предлагает теоретически обоснованный подход к обучению представлениям, которые инвариантны к защищенному признаку; однако его практическая эффективность в реалистичных условиях дисбаланса подгрупп в наборах данных здравоохранения не была полностью охарактеризована.

Исследовательский вопрос

Улучшает ли состязательное устранение смещения с использованием GRL справедливость прогнозирования диабета в отношении возраста, и как эффекты справедливости варьируются в зависимости от различных разделов данных?

Методы

Метод состязательного устранения смещения оценивался для смягчения возрастного смещения при прогнозировании диабета с использованием общедоступной базы данных диабета индейцев Пима (Pima Indians Diabetes Database, $n = 768$). Были использованы все восемь предикторов набора данных; для оценки справедливости на основе возраста были выделены три возрастные группы (<30 , $30-50$ и >50 лет). Состязательная нейронная модель со слоем реверсии градиента сравнивалась с базовой моделью логистической регрессии. Признаки были стандартизированы с использованием масштабировщика (scaler), обученного только на тренировочных данных. Разделение на обучающую и тестовую выборки было стратифицировано по результату наличия диабета. Общие показатели эффективности (**точность** [accuracy], **полнота** [recall], **ROC-AUC**) и разрыв в равенстве полноты (recall parity gap) между возрастными группами рассчитывались на основной размеченной тестовой выборке ($n = 154$); устойчивость оценивалась на пяти независимых случайных реализациях (random seeds, 0–4).

Результаты

На основной тестовой выборке состязательная модель улучшила полноту (recall) для самой малочисленной возрастной группы [>50 лет: $0,5556 \rightarrow 0,7778$, +22,22 процентных пункта (pp)], сохраняя при этом сопоставимую общую дискриминационную способность (ROC-AUC: $0,7852 \rightarrow 0,7896$, +0,45 pp). Однако разрыв в равенстве полноты увеличился с $0,0996$ до $0,2153$ (+11,57 pp), что отражает одновременное снижение полноты для группы <30 лет ($-6,25$ pp). При использовании пяти случайных реализаций средний разрыв в равенстве полноты показал незначительное среднее снижение ($0,3282 \rightarrow 0,3033$, $-2,49$ pp), но с высокой вариативностью (стандартное отклонение [SD] $> 0,27$), превышающей среднюю разницу. Состязательная модель уменьшила разрыв в справедливости в трех из пяти реализаций, увеличила его в одной и не внесла изменений в одной.

Заключение

Состязательное устранение смещения может улучшить прогностическую полноту для недостаточно представленных демографических подгрупп, но оно не гарантирует стабильного улучшения справедливости в различных разделах данных, особенно когда размеры выборок подгрупп малы. Оценка с использованием нескольких случайных реализаций (multi-seed evaluation) необходима для надежной оценки справедливости; единичного разделения на обучающую и тестовую выборки недостаточно.

Перевод выполнен: 17.07.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.