

Повторное исследование GRAPHOS: графовые теоретические признаки не улучшают скрининг дислексии, и демографическое смешивающее влияние в наборе данных Dytective

Источник: Frontiers in Digital Health

Дата публикации: 2024-05-15

Оригинал: <https://www.frontiersin.org/articles/10.3389/fdgth.2026.1927725>

валидация моделей

диагностика

дислексия

машинное обучение

нейронауки

смещающие факторы

этика ИИ

Введение

Большинство методов машинного обучения для скрининга дислексии рассматривают показатели эффективности когнитивных задач как независимые векторы признаков. Было выдвинуто гипотезу о том, что такой подход приводит к потере информации, содержащейся во взаимозависимости между задачами, исследующими общие когнитивные процессы. В данной работе эта гипотеза проверяется, но не находит подтверждения.

Методы

Мы оцениваем GRAPHOS, графово-теоретическую основу, которая строит сети корреляций задач на основе данных геймифицированной когнитивной оценки, на наборе данных Dytective, включающем 3 644 участника (392 с дислексией). Модель, стратегия агрегации и порог принятия решений выбираются исключительно на обучающих и валидационных подвыборках, а тестовая подвыборка используется только один раз.

Результаты

По метрике AUC без порога (threshold-free AUC) GRAPHOS не превосходит сильную базовую модель с одним классификатором (0,8881 против 0,8883; $p = 0,989$ по тесту Делонга), и этот нулевой результат воспроизводится при десяти случайных разбиениях данных. Модель, не использующая признаки, полученные из графа, статистически неотличима от полной системы. Нулевая модель, основанная на перестановках с выравниванием размеров выборок, показывает, что ни одно из девяти различий сетевых показателей на популяционном уровне между группами с дислексией и без нее не выдерживает поправки на асимметрию выборки 8,3:1. Аудит подгрупп выявляет, что распространенность дислексии в данном наборе данных различается в 4,7 раза в зависимости от статуса родного языка; что четыре демографических столбца в отдельности достигают AUC 0,724 по сравнению с 0,744 для всех 192 показателей задач; и что их включение приводит к двенадцатикратной разнице в частоте ложноположительных результатов между языковыми группами.

Обсуждение

Психометрический анализ объясняет, почему графовое представление не дает преимуществ: батарея задач демонстрирует малую общую дисперсию в данной выборке (КМО 0,497, альфа Кронбаха 0,171), а разведочный факторный анализ восстанавливает 13% дисперсии, при этом 28 из 32 задач не нагружаются ни на один фактор. Мы представляем эти результаты как отрицательные, вместе с двумя контрольными процедурами, которые их выявили, и предлагаем использовать их в рутинной практике: нулевую модель с выравниванием размеров выборок везде, где сравниваются корреляционные сети между группами неравного размера, и аудит подгрупп везде, где классификатор скрининга обучается на выборке, распространенность метки в которой может быть связана с демографическим составом.