Разделенный отбор признаков: локально однородный отбор признаков со структурированной группировкой для объяснимой идентификации предикторов

Источник: Frontiers in AI — Medicine

Оригинал: https://www.frontiersin.org/articles/10.3389/frai.2026.1879422

депрессия | интерпретируемость ИИ | машинное обучение | объяснимый ИИ | отбор признаков | психиатрия

Введение

Отбор признаков помогает выявлять важные признаки в высокоразмерных данных, связанных с психическим здоровьем. Однако традиционные методы отбора признаков (CFS) обычно ранжируют все признаки в едином пространстве признаков на основе общей статистической или предиктивной критерии. Этот метод глобальной оценки не учитывает тот факт, что симптомы депрессии взаимосвязаны и относятся к различным клиническим доменам. Чтобы устранить это ограничение, мы предлагаем метод отбора признаков с разделением на группы (PFS) —框架, который интегрирует группировку признаков на основе клинических доменов с отбором признаков для каждой группы.

Методы

PFS организует симптомы депрессии по клинически значимым доменам и применяет четыре дополнительных метода отбора признаков для оценки индивидуальной важности каждого признака в пределах каждого домена. Оценки важности нормализуются и объединяются, затем взвешиваются в

соответствии с ассоциацией каждого домена с общей тяжестью состояния для расчета интегрированной важности групповых признаков (IGFI). Мы оценили метод PFS на двух независимых наборах данных с многоклассовой классификацией депрессии, используя 16 классификаторов. Для выбора окончательных подмножеств признаков использовалась валидация с направляющим отбором признаков. Мы также провели анализ случайной перегруппировки в десяти испытаниях, исследовательское сравнение с использованием критерия Уилкоксона для парных выборок с методом CFS-RFE и сравнение рангов IGFI с объяснениями на основе SHAP.

Результаты

PFS выбрал 21 признак для набора данных 1 и 23 признака для набора данных 2. Используя выбранную конфигурацию опорного векторного классификатора (SVC) без настройки гиперпараметров и без применения техники синтетического оверсэмплинга меньшинства (SMOTE), PFS достиг точности тестирования 95,65% и 89,60%, а также показателей Macro-F1 95,43% и 92,37% соответственно. Анализ случайной перегруппировки показал, что предиктивный эффект клинического разделения варьировался между наборами данных. Исследовательский анализ Уилкоксона также выявил статистически значимые различия между PFS и CFS-RFE, зависящие от набора данных. Кроме того, IGFI и SHAP выявили схожие важные признаки в рамках клинических доменов.

Обсуждение

PFS обеспечивает структурированный подход к выявлению важных симптомов депрессии с сохранением информации о принадлежности к клиническим доменам. Полученные результаты указывают на то, что группировка на основе доменов может способствовать интерпретируемости, хотя ее предиктивные преимущества варьируются в зависимости от набора данных. В данном исследовании основное внимание уделяется оценке тяжести депрессии на основе анкетных данных, а не прогнозированию независимо установленного клинического диагноза.

Перевод выполнен: 03.11.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.