Генерация связанных с популяцией сценариев пациентов из общенациональных данных триажной оценки: исследование разработки и валидации фреймворка

Источник: JMIR Medical Informatics

Дата публикации: 2024-05-01

Оригинал: https://medinform.jmir.org/2026/1/e104801

NLP | валидация ИИ | генерация данных | триаж | эпидемиология

Введение

Цифровые инструменты поддержки принятия решений, такие как системы сортировки пациентов и чекеры симптомов, поддерживают миллионы решений, связанных со здоровьем, каждый год. Их качество и безопасность обычно оцениваются с использованием текстовых случаев пациентов, известных как клинические сценарии (case vignettes). Однако наборы клинических сценариев, составленные экспертами, охватывают лишь ограниченный спектр реальных клинических проявлений и не имеют популяционных весов, что позволило бы экстраполировать результаты оценки на базовую популяцию пациентов.

Цель

Настоящее исследование было направлено на разработку основанного на данных фреймворка для автоматической генерации управляемого человеком набора клинических сценариев из национальных данных сортировки, которые отражают широкое разнообразие клинических проявлений и связывают каждый сценарий с количественным весом, отражающим количество лежащих в его основе оценок пациентов.

Методы

Из 3,2 миллиона оценок сортировки, проведенных в течение 1 года с использованием структурированного программного обеспечения для сортировки в немецкой службе медицинской помощи вне больничных учреждений (телефонная сортировка и онлайн-самосортировка) и в совместных контактных пунктах службы амбулаторной неотложной помощи и стационарных отделений неотложной помощи, мы случайным образом отобрали 50 000 случаев. Анкеты сортировки были преобразованы в семантические эмбеддинги с использованием модели немецкого Sentence Transformer и сгруппированы с помощью агломеративной кластеризации. Для кластеров, содержащих достаточное количество оценок, мы сгенерировали один репрезентативный случай с помощью двухэтапной оптимизации методом имитации отжига, которая минимизировала расстояние от центроида кластера, одновременно максимизируя количество отвеченных вопросов сортировки. Каждому репрезентативному случаю был присвоен вес, основанный на размере его исходного кластера. Анализ чувствительности на основе сходства оценивал, сохраняются ли эти веса в полной популяции за 1 год. Наконец, пары «вопрос-ответ» репрезентативных случаев были преобразованы в текстовые клинические сценарии с помощью контролируемого запроса (prompting) большой языковой модели.

Результаты

Кластерный анализ дал 514 включенных кластеров, охватывающих 96,8% отобранных 50 000 оценок. Сгенерированные репрезентативные случаи показали сильное согласие с рекомендацией по срочности лечения, полученной большинством в их исходном кластере (коэффициент корреляции Спирмана $\rho=0,78$; $p<0,001$) и содержали в среднем на 4,3 отвеченных вопроса сортировки больше, чем исходные случаи в пределах их кластеров. При взвешивании по размеру кластера репрезентативные случаи аппроксимировали выборочные распределения по срочности лечения, демографическим показателям и симптомам, хотя некоторые систематические отклонения сохранились, наиболее заметным из которых было избыточное представительство случаев у женщин (13,5%), пациентов в возрасте 14–49 лет (8,0%) и категории срочности «Как можно скорее» (6,6%). Из 121 зафиксированного симптома 101 (83,5%) были покрыты репрезентативными случаями; остальные встречались менее чем в 0,5% выборки. Веса клинических сценариев на основе кластеров были сильно

коррелированы с весами популяции на основе сходства (коэффициент корреляции Спирмана ρ=0,77; p<0,001), и 90,1% оценок в полной популяции за 1 год совпали хотя бы с одним сценарием.

Выводы

Мы представляем основанный на данных фреймворк для получения управляемого набора клинических сценариев с популяционными весами из национальных данных сортировки. Полученные сценарии отражали широкое разнообразие клинических проявлений, аппроксимировали ключевые характеристики выборки и обеспечивали явную количественную связь с количеством лежащих в их основе оценок. После медицинского экспертного обзора и доработки эти сценарии могут поддержать более учитывающую популяцию оценку и обеспечение качества цифровых инструментов поддержки принятия решений.

Перевод выполнен: 04.11.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.