

УДК 004.8

Шабля Р. - ст. гр. СНм-61

Тернопільський національний технічний університет імені Івана Пулюя

ПРОБЛЕМА НЕОДНОРІДНОСТІ ЕЛЕКТРОННОЇ МЕДИЧНОЇ ДОКУМЕНТАЦІЇ У ЗАДАЧАХ АНАЛІЗУ ЗАХВОРИЮВАНЬ

Shablia R.

Ternopil Ivan Puluj National Technical University

THE PROBLEM OF HETEROGENEITY OF ELECTRONIC MEDICAL RECORDS IN DISEASE ANALYSIS TASK

У сучасних системах охорони здоров'я електронна медична документація є основним джерелом даних для побудови моделей прогнозування захворювань та підтримки клінічних рішень. Активне впровадження інформаційних систем у медичну практику призвело до накопичення значних обсягів даних про пацієнтів, що відкриває нові можливості для застосування методів машинного навчання. Проте ефективність таких моделей суттєво обмежується якістю, структурою та узгодженістю вхідних даних.

Ключовою проблемою є неоднорідність медичних даних. Інформація про пацієнта формується з різних джерел, включаючи лабораторні дослідження, результати інструментальних обстежень, текстові висновки лікарів, історії лікування та дані моніторингу стану пацієнта. Ці дані відрізняються за форматом (структуровані, напівструктуровані, неструктуровані), частотою оновлення, точністю вимірювань та рівнем деталізації. У результаті виникають труднощі при їх інтеграції в єдину аналітичну систему, що негативно впливає на якість побудованих моделей. Додатковою складністю є те, що електронна медична документація часто містить пропуски, дублювання записів, суперечливу інформацію та шум. Причинами цього можуть бути людський фактор, різні стандарти ведення документації, технічні обмеження медичних інформаційних систем, а також нерегулярність спостережень за пацієнтом. У таких умовах традиційні підходи до аналізу даних, які передбачають їх повноту та узгодженість, втрачають ефективність.

Особливо критичною є проблема обробки неструктурованих текстових медичних записів, які містять значну частину клінічно важливої інформації, зокрема опис симптомів, висновки лікарів та рекомендації щодо лікування. Такі дані не можуть бути безпосередньо використані в моделях машинного навчання без попередньої обробки. Відсутність або обмежене використання методів обробки природної мови призводить до втрати значної частини інформації, що знижує якість аналізу та точність прогнозування.

Таким чином, підвищення ефективності прогнозних моделей потребує розробки комплексних підходів до попередньої обробки та уніфікації медичних даних. До таких підходів належать методи очищення даних, заповнення пропусків, нормалізації показників, інтеграції даних із різних джерел, а також використання алгоритмів для роботи з неструктурованою інформацією. Важливим напрямом є також застосування моделей, здатних працювати з неоднорідними та часовими даними, включаючи гібридні підходи, що поєднують різні типи інформації. Неоднорідність електронної медичної документації є одним із ключових факторів, що обмежує точність та надійність моделей прогнозування захворювань. Подолання цієї проблеми потребує комплексного підходу до обробки, інтеграції та аналізу даних, а також використання сучасних методів машинного навчання, здатних ефективно працювати з різномірною медичною інформацією.