

Анализ больших данных в задаче прогнозирования риска неблагоприятного клинического исхода

Научный руководитель – Рыжов Александр Павлович

Шергин И.А.¹, Фещенко Д.А.², Abdullaev A.M.³

1 - Московский государственный университет имени М.В.Ломоносова, Механико-математический факультет, Кафедра математической теории интеллектуальных систем, Москва, Россия, *E-mail: po4tashergina@yandex.ru*; 2 - ФГБУ «Национальный медицинский исследовательский центр терапии и профилактической медицины» Министерства здравоохранения Российской Федерации, Москва, Россия, *E-mail: dfeshenko@gnicpm.ru*; 3 - Дагестанская государственная медицинская академия, Республика Дагестан, Россия, *E-mail: abdullaevaslanm@mail.ru*

Под неблагоприятным клиническим исходом при выполнении плановых интервенционных вмешательств понимается такой исход в периоперационном периоде, как наступление серьезного неблагоприятного сердечно-сосудистого и церебрального события (смерть, инсульт, инфаркт миокарда и др.), а также клинически значимого кровотечения и/или специфического для каждого вида операций осложнения, приводящих к удлинению срока госпитализации и удорожанию стоимости лечения. Прогнозирование риска неблагоприятного клинического исхода имеет важное практическое значение, так как позволит обеспечить персонализированный подход в анализе рисков возможных осложнений. По данным различных литературных источников процент осложнений при выполнении плановых интервенционных вмешательств может достигать 5-10%, при этом возникновение осложнений, в среднем, увеличивает стоимость лечения вдвое по сравнению с неосложненным вариантом течения [1]. Надежный прогноз позволяет на основе рутинных методов диагностики заранее спланировать методы профилактики вероятных неблагоприятных событий.

В современных высокотехнологических клиниках благодаря внедрению медицинских информационных систем (МИС) фиксируется и накапливается большой объем данных о каждом конкретном пациенте (демографические, клинические, результаты лабораторных и инструментальных методов диагностики, характер оперативного лечения) и динамика происходящих с ним изменений.

Целью настоящей работы является исследование возможности прогноза риска неблагоприятного клинического исхода на основе данных и определение качества такого прогноза.

В качестве набора данных были использованы записи МИС клиники (5062 пациентов, 79 признаков). После предобработки количество пропусков в данных было снижено с 53% до 10%. Далее было опробовано 3 варианта заполнения недостающих значений (среднее и медианное значения, отсутствие значения).

Применялись алгоритмы decision tree, random forest, catboost, xgboost, logreg. Ввиду сильной несбалансированности классов, метрикой качества была выбрана f1-score (основную метрику precision в этой ситуации использовать не корректно [2]).

Наилучший результат продемонстрировал random forest. Качество прогноза составило 96%.

Полученные результаты позволяют разработать систему прогнозирования риска неблагоприятного клинического исхода по технологии скоринговых систем.

Работа выполнена при поддержке РФФИ (грант № 19-29-01051 “Разработка алгоритмов принятия решений для управления рисками неблагоприятных клинических событий в высокотехнологичной медицинской организации на основе технологии data mining”).

Источники и литература

- 1) Inohara T, et al. Predictors of high cost after percutaneous coronary intervention: A review from Japanese multicenter registry overlooking the influence of procedural complications. *American Heart Journal* 194: 61-72, Dec 2017.
- 2) Goutte C., Gaussier E. (2005) A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In: Losada D.E., Fernández-Luna J.M. (eds) *Advances in Information Retrieval. ECIR 2005. Lecture Notes in Computer Science*, vol 3408. Springer, Berlin, Heidelberg.