

Построение классификаторов в задаче прогнозирования риска неблагоприятного клинического исхода на основе клинико-демографических показателей пациента ¹

Б. Э. Горный², А. П. Рыжов³, А. С. Строгалов⁴, А. А. Хусаенов⁵,
И. А. Шергин⁶, Д. А. Фещенко⁷, А. М. Абдуллаев⁸, А. В. Концевая⁹

¹ 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» и Междисциплинарной научно-образовательной школы Московского университета «Мозг, когнитивные системы, искусственный интеллект»

² *Горный Борис Эмануилович* — к.м.н, ведущий научный сотрудник отдела первичной профилактики ХНИЗ в системе здравоохранения ФГБУ «НМИЦТПМ», e-mail: bgornyy@gnicpm.ru

Gornyi Boris Emanuilovich — Ph.D. in medical science, leading reseacher of department of Primary Prevention of Chronic Non-Communicable Diseases in the Healthcare System, National Medical Research Center for Therapy and Preventive Medicine

³ *Рыжов Александр Павлович* — д.т.н., к.ф.-м.н., MBA, профессор кафедры математической теории интеллектуальных систем мех.-мат. ф-та МГУ; e-mail: alexander.ryjov@gmail.com

Ryjev Alexander Pavlovich — Sc.D. in engineering, Ph.D. in mathematics, MBA, professor, Moscow State University, Faculty of Mechanics and Mathematics, Mathematical Theory of Intelligent Systems department

⁴ *Строгалов Александр Сергеевич* — к.ф.-м.н., доцент кафедры математической теории интеллектуальных систем мех.-мат. ф-та МГУ; e-mail: strogalov@mail.ru

Strogalov Alexander Sergeevich — Ph.D. in mathematics, associate professor, Moscow State University, faculty of Mechanics and Mathematics, Mathematical Theory of Intelligent Systems department

⁵ *Хусаенов Артем Азатович* — аспирант кафедры математической теории интеллектуальных систем мех.-мат. ф-та МГУ; e-mail: a.khusaenov@mail.ru

Khusaenov Artem Azatovich — postgraduate student, Moscow State University, faculty of Mechanics and Mathematics, Mathematical Theory of Intelligent Systems department

⁶ *Шергин Иван Андреевич* — студент 6 курса кафедры математической теории интеллектуальных систем мех.-мат. ф-та МГУ; e-mail: i.a.shergin@gmail.com

Shergin Ivan Andreevich — 6 year student, Moscow State University, faculty of Mechanics and Mathematics, Mathematical Theory of Intelligent Systems department

⁷ *Фещенко Дарья Анатольевна* — заведующая операционным блоком ФГБУ «НМИЦТПМ»; e-mail: dfeshenko@gnicpm.ru

Feshchenko Daria Anatol'evna — the head of surgical block of National Medical Research Center for Therapy and Preventive Medicine, Moscow

⁸ *Абдуллаев Алсан Мурадovich* — аспирант отдела нарушений сердечного ритма и проводимости ФГБУ «НМИЦТПМ»; e-mail: abdullaevaslanm@mail.ru

Abdullaev Aslan Muradovich, postgraduate student, National Medical Research Center for Therapy and Preventive Medicine, Moscow

В докладе представлены результаты исследования, выполненного совместно сотрудниками Национального медицинского исследовательского центра терапии и профилактической медицины МЗ РФ и механико-математического факультета МГУ имени М.В. Ломоносова, показывающего применимость методов анализа данных в решении задачи прогнозирования риска неблагоприятного клинического исхода на основе информации о пациенте, доступной в клинике (медицинская информационная система «Медиалог»).

Ключевые слова: профилактическая медицина, неблагоприятный клинический исход, углубленный анализ данных.

1. Введение

Возникновение неблагоприятных событий в процессе оказания медицинской помощи возникает у 10-15% госпитализированных пациентов [1]. Снижение даже на несколько процентов возникновения таких событий позволит сохранить тысячи жизней. Одним из путей решения этой важнейшей проблемы является использование интеллектуальных информационных технологий, позволяющих прогнозировать риск возникновения неблагоприятного клинического исхода у пациентов.

В качестве данных для анализа были использованы 79 клинико-демографических и лабораторных параметров из МИС «Медиалог» НМИЦТМПМ о 5062 пациентах, которым были выполнены высокотехнологичные эндоваскулярные (60%) и интервенционные аритмологические вмешательства (40%) [2]. В группу рентгенэндоваскулярных операций входили как диагностические (25% - коронарография, ангиография аорты и ее ветвей, ангиография брахиоцефальных артерий), так и лечебные процедуры (75% - ангиопластика и стентирование коронарных (77%), каротидных (7%) и периферических артерий (12,7%), транскатетерное протезирование аортального клапана (0,3%), ренальная денервация 3%). В среднем 15% операций проводились по экстренным показаниям.

В докладе приводится описание формальных свойств используемого набора данных, процедур предобработки данных, и решение задачи прогнозирования риска неблагоприятного клинического исхода различными алгоритмами анализа данных.

⁹ *Концевая Анна Васильевна* — д.м.н, заместитель директора ФГБУ «НМИЦТМПМ»; e-mail: AKontsevaya@gnicpm.ru

Kontsevaya Anna Vasilievna, Doctor of Medicine, deputy director of National Medical Research Center for Therapy and Preventive Medicine, Moscow

2. Основные результаты

2.1. Случайный лес

Для построения классификатора был выбран алгоритм построения дерева решений (Decision tree), так как требовалась хорошая интерпретируемость модели для медицинских специалистов. Для увеличения качества итогового дерева на исходных данных сначала использовался алгоритм случайного леса (Random Forest). Далее были выбраны 15 признаков с наибольшим вкладом в предсказание случайного леса, и итоговое дерево решений строилось на выборке, состоящей только из них.

Использовались меры качества алгоритма классификации F1-score и ROC AUC, так как они хорошо подходят для несбалансированных данных (значения целевой переменной относятся как 3079/67). Во избежание переобучения алгоритмов Decision tree и Random forest были введены различные ограничения на их параметры. Использовалась кросс-валидация (cross-validation). Результаты: случайный лес - F1-score = 0.97, ROC AUC = 0.85; дерево решений - F1-score = 0.96, ROC AUC = 0.64.

2.2. Нейронная сеть

Неблагоприятные события составляли 2% от доступных данных. Нейронная сеть была обучена выявлять отклонение от благоприятного события [3]. Обучающее множество содержало 2826 благоприятных событий, тестовое содержало 66 неблагоприятных событий (все доступные) и 66 благоприятных событий. Результаты: на основе 23 существенных признаков достигнуто 100% точности тестирования. При добавлении 12 несущественных признаков достигнута точность 96,96%. Ошибка 3% пришлась на ошибку второго рода.

3. Выводы

Итоги проведенного исследования показали, что методы машинного обучения можно применять для анализа медицинских данных. Показано, что можно предсказывать неблагоприятный клинический исход пациента при поступлении в клинику, причем с хорошей точностью.

В ходе исследования были выявлены показатели, которые имели наибольшую прогностическую ценность. Полученные результаты могут быть использованы в качестве рекомендации для медицинских работников при первичном осмотре поступивших пациентов, а также позволяют разработать систему прогнозирования риска неблагоприятного клинического исхода по технологии скоринговых систем. Получен вклад наиболее важных переменных в предсказание риска неблагоприятного клини-

ческого исхода. Имея значения этих признаков, медицинский работник в том числе может самостоятельно оценить риск неблагоприятного клинического исхода.

Список литературы

- [1] Baker G.R. et al, “The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada”, *Can. Med. Assoc. J. Can Med Assoc*, **170**:11 (2004), 1678–1686.
- [2] Горный Б.Э., Рыжов А.П., Строгалов А.С., Журавлев А.Д., Хусаенов А.А., Шергин И.А., Фещенко Д.А., Абдуллаев А.М., Концевая А.В., “Оценка риска неблагоприятного клинического исхода методами углубленного анализа данных”, *Интеллектуальные Системы. Теория и приложения*, **25**:2 (2021), 23–45.
- [3] Хусаенов А.А., “Автоассоциативные нейронные сети в задаче классификации с усеченным множеством”, *Интеллектуальные Системы. Теория и приложения*, **25**:4 (2021).

The adverse clinical outcome risk classifiers constructing on clinical and demographic patient data

Gornyi B.E., Ryjov A.P., Strogalov A.S., Khusaenov A.A., Shergin I.A., Feshchenko D.A., Abdullaev A.M., Kontsevaya A.V.

The report presents the result of joint research carried out by researchers of the National Research Center for Therapy and Preventive Medicine of the Ministry of Health of the Russian Federation and by researchers of Faculty of Mechanics and Mathematics of the Lomonosov Moscow State University, showing the data analysis methods applicability in the problem of an unfavorable clinical outcome risk predicting based on clinic patient information (medical information system «Medialog»).

Keywords preventive medicine, adverse clinical outcome, in-depth data analysis.

References

- [1] Baker G.R. et al, “The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada”, *Can. Med. Assoc. J. Can Med Assoc*, **170**:11 (2004), 1678–1686.
- [2] Gornyi B.E., Ryjov A.P., Strogalov A.S., Zhuravlev A.D., Khusaenov A.A., Shergin I.A., Feshchenko D.A., Abdullaev A.M., Kontsevaya A.V., “The adverse clinical outcome risk assessment by deep data analysis methods”, *Intelligent Systems. Theory and Applications*, **25**:2 (2021), 23–45 (In Russian).
- [3] Khusaenov A.A., “Autoassociative neural networks in a classification problem with truncated dataset”, *Intelligent Systems. Theory and Applications*, **25**:4 (2021) (In Russian).