

ОПЫТ ПРИМЕНЕНИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ НАЗНАЧЕНИЯ ПРОЦЕДУР РЕАБИЛИТАЦИИ БОЛЬНЫХ ТУБЕРКУЛЕЗОМ

EXPERIENCE IN APPLYING MACHINE LEARNING METHODS FOR REHABILITATION OF TUBERCULOSIS PATIENTS

P. Chernavin
N. Chernavin
Yu. Chugaev
I. Cherniaev
L. Kalinina

Summary. Sanatorium-resort treatment is an important stage in the organization of medical rehabilitation of patients with tuberculosis. Help in solving this issue can be provided by the use of machine learning (ML) methods for prescribing treatment procedures. The article presents the author's approach to solving this issue using real data from a sanatorium for the rehabilitation of patients who have suffered tuberculosis. Calculations were carried out on the basis of 18 qualitative and 25 quantitative characteristics taken into account upon admission to the sanatorium. The purpose of the study was to identify the most suitable machine learning methods for prescribing 37 rehabilitation procedures.

Keywords: machine learning, precedent-based learning, mathematical programming, classification tasks, committee method, sanatorium treatment, tuberculosis.

Чернавин Павел Федорович
к.э.н., доцент, Уральский федеральный
университет (Екатеринбург)
chernavin.p.f.@gmail.com

Чернавин Николай Павлович
ассистент, Уральский федеральный
университет (Екатеринбург)
ch_k@mail.ru

Чугаев Юрий Петрович
д.м.н., профессор, Уральский государственный
медицинский университет (Екатеринбург)
doctorchugaev@mail.ru

Черняев Игорь Анатольевич
старший преподаватель, Уральский государственный
медицинский университет (Екатеринбург)
obltuborg@yandex.ru

Калинина Людмила Васильевна
заместитель директора ФГБУ «Туберкулезный
санаторий «Голубая бухта»
Минздрава России» (Геленджик)
ivanovagolubayabuhta@yandex.ru

Аннотация. Санаторно-курортное лечение является важным этапом в организации медицинской реабилитации больных туберкулезом. Помощь в решении данного вопроса может оказать применение методов машинного обучения (МО) для назначения лечебных процедур. В статье представлен авторский подход к решению этого вопроса на реальных данных санатория для реабилитации пациентов, перенесших туберкулез. Расчеты проводились на основе 18 качественных и 25 количественных признаков, учитываемых при поступлении в санаторий. Цель исследования состояла в выявлении наиболее подходящих методов машинного обучения для назначения 37 реабилитационных процедур.

Ключевые слова: машинное обучение, обучение по прецедентам, задачи классификации, метод комитетов, санаторно-курортное лечение, туберкулез.

Введение

Санаторно-курортное лечение является важным этапом в организации медицинской реабилитации больных туберкулезом [1, 2]. Физические факторы, применяемые в комплексе лечебных мероприятий при туберкулезе, принято по характеру лечебного воздействия подразделять на три группы: 1) физические факторы, обладающие преимущественно противовоспалительным, в том числе туберкулостатическим и гипосенсибилизирующим действием, способствуют повышению концентрации антибактериальных препаратов

в очаге воспаления, активации местных защитных тканевых реакций. 2) факторы, применяемые для рассасывания инфильтративных изменений, обеспечения рубцевания полостей деструкции, биостимулирующего и адаптогенного действия, с целью активации естественных защитных механизмов организма пациента: ультразвук, лазерная терапия и магнитотерапия. 3) физические факторы, способствующие минимизации остаточных туберкулёзных изменений, предотвращению избыточного образования фиброзной ткани, рассасыванию спаек и рубцов, повышению активности обмена веществ, улучшение микроциркуляции и трофики лёгочной ткани.

В настоящее время физиотерапевтические методы применяются у 89,3 % пациентов санаториев [4]. Многообразие немедикаментозных способов лечения, вошедших в клиническую практику достаточно давно и применяемых в условиях противотуберкулезных санаториев, требует оценки эффективности на основе методов доказательной медицины [2]. В условиях внедрения современной концепции здравоохранения «Медицина 4П» [5,6], составляющими которого являются персонализация (индивидуальный подход к каждому пациенту), предикция (выявление предрасположенности к развитию заболевания), превентивность (предотвращение появления заболеваний) и партисипативность (мотивированное участие пациента), особую значимость приобретает возможность построения индивидуальных траекторий реабилитации больных туберкулезом. Помощь в решении данного вопроса может оказать применение методов машинного обучения (МО) для назначения лечебных процедур. Синонимом американского термина «машинное обучение» является русскоязычное словосочетание «обучение по прецедентам», так как по сути МО является обобщением предыдущего опыта в виде решающих правил, на основе которых можно принимать решения в текущей ситуации.

Следует отметить, что, конечно, научно-исследовательские работы в этом направлении ведутся различными авторами [8,9]. Целью данного исследования было: на основе реальных данных за предшествующие периоды действующего санатория по реабилитации больных туберкулезом создать компьютерную систему (КС), которая бы подсказывала на основе предыдущего опыта какие процедуры лечения следует назначить конкретному пациенту исходя из его анамнеза и текущего состояния здоровья. При этом, у медицинского работника существует право как согласиться с данными предложениями, так и отказаться от всех или части предлагаемых процедур и назначить свои. Процедуры, предлагаемые компьютером, и их корректировка медицинским работником сохраняются в базе данных и на основе нового опыта производится дообучение КС. Далее мы опишем последовательность создания КС и результаты ее применения.

Речь в данном случае идет не о замене медицинских работников электронно-вычислительными машинами. Такой проблемы, в настоящий момент, просто не существует. Наоборот, санаторно-курортный комплекс испытывает острый дефицит опытных кадров, что приводит к снижению качества обслуживания [7]. Поэтому созданию эффективных систем поддержки принятия клинических решений на основе обобщения опыта предшествующих поколений с целью помощи действующим сотрудникам медицинских организаций является наиважнейшей необходимостью.

Структура исходных данных

Наше исследование проводилось на основании данных когорты 450 пациентов, проходивших лечение в учреждении здравоохранения, оказывающего специализированную медицинскую помощь по профилю «фтизиатрия» в условиях санатория. За единицу наблюдения был принят случай лечения и для каждого пациента была оформлена карта исследования собственной разработки, которая включала 211 формализованных характеристик пациента, в том числе: пол, возраст и другие медико-социальные характеристики, жалобы, данные анамнеза заболевания и анамнеза жизни, результаты осмотра, объективные данные методов лабораторной, лучевой и функциональной диагностики, оценки психоэмоционального состояния до и после проведенного лечения. Данные карт вносились в компьютерную базу данных исследования. В процессе лечения каждому пациенту назначается комплекс из 6–12 процедур, из общего списка состоящего из 104 доступных методов лечения. После отсева слабо заполненных, дублирующих друг друга и низко дисперсных признаков для дальнейшего анализа были оставлены 44 входных признаков. После исключения специфических и редко используемых процедур были оставлены 38 наиболее массовых процедур. Отметим, что массовые не означает, что назначаются всем без учета характеристик пациента. В нашем случае каждая из данных процедур назначалась строго в соответствии с показаниями от 5 до 45 % пациентов.

На всех этапах построения КС мы регулярно учитывали мнение практических специалистов. При назначении процедур специалисты обычно учитывают историю болезни (анамнез) и текущее состояние пациента. Анамнез обычно описывается качественными признаками, которые обычно принято называть категориальными (пол, группа диспансерного учета, наличие вредных условий труда в прошлом, перенесенные заболевания, алкоголизация и т.п.). Состояние пациента при поступлении практически всегда описывается количественными признаками (возраст, давление, пульс, показатели биохимии крови и т.п.). Все эти признаки измеряемы и имеют свои единицы измерения. Поэтому далее признаки были разбиты на 18 категориальных и 25 количественных. Такое разбиение является удобным и с вычислительной точки зрения так как работа с категориальными признаками имеет свою специфику. Все категориальные признаки были преобразованы в числовой формат. Кодировка категориальных признаков имеет важное значение так как может влиять на результаты решения. Желательно, чтобы она носила не случайный, а осмысленный характер и изменение величины категориального признака было общепринятым (например, группа диспансерного учета) или соответствовала степени восприятия пациента медицинским персоналом. Поэтому кодировка категориальных признаков тщательно согласовывалась

с практическими медицинскими работниками, имеющими большой опыт.

Методы поиска решающих правил

По каждой лечебной процедуре решалась отдельная задача классификаций, то есть множество наблюдений по каждой процедуре разбивалось на два класса: те, кому была назначена данная процедура и те, кому она не была назначена. То есть требовалось решить 74 задачи классификации (2 вида признаков, 37 процедур).

Так как заранее было не известно какой из методов МО даст наилучшие результаты, то использовались основные классические методы МО. Для категориальных признаков использовались категориальный бустинг, дерево решений, случайный лес, наивный байесовский метод, линейное разделение множеств. Для количественных признаков линейное разделение, метод ближайших соседей, наивный байесовский метод, дерево решений, случайный лес. Генеральная выборка разбивалась на обучающую и тестовую в пропорции 80 / 20 %. Так как классы были несбалансированными, то оценка качества решения проводилась по метрике площадь под ROC кривой (AUC ROC). Требовалось определить какие методы и для каких процедур будут иметь метрику AUC ROC более 0.75 одновременно на обучающей и на тестовой выборках.

При назначении процедур на основе категориальных признаков наивный байесовский метод показал наилучшие результаты по 8 процедурам, случайный лес по 7, категориальный бустинг по 4, дерево решений по 1, линейное разделение по 11.

При назначении процедур на основе количественных признаков наивный байесовский метод показал наилучшие результаты по 2 процедурам, случайный лес по 3, линейное разделение по 29. Дерево решений и метод ближайших соседей не показали хороших результатов ни по одной процедуре.

Таким образом, наилучшие результаты показал метод линейного разделения. По семи процедурам при анализе на основе категориальных признаков ни по одному из методов МО метрика AUC ROC не превысила планку 0.75. При анализе на основе количественных признаков аналогичная ситуация была по четырем процедурам. При построении компьютерной системы по этим процедурам использовался ансамблевый подход. Если большинство методов, имело метрику AUC ROC более 0.7, то процедура назначалась и не назначалась в противном случае. То есть, применялась логика большинства.

Отдельно отметим, что все применяемые методы легко объяснимы практическим медицинским работникам. Наивный байесовский метод — это, просто чисто статистический подход, метод ближайших соседей и линейное разделение имеет четкую геометрическую интерпретацию, дерево решений визуально можно изобразить на бумаге, а случайный лес и категориальный бустинг — это ансамбль из деревьев.

На основе найденных РП была построена КС в виде взаимосвязанных таблиц Excel. Выбор Excel определялся тем, что большинство персонала обладало навыками работы в этой среде. Поэтому персоналу просто требовалось ввести категориальные и количественные признаки пациента при поступлении в санаторий.

После ввода данных пациента КС рекомендует какие лечебные процедуры следует назначить конкретному пациенту исходя из его категориальных и количественных признаков на момент поступления. Однако, у врача есть право как отменить назначение, так и назначить дополнительные процедуры.

Назначения КС и врача автоматически сохраняются в базе данных. Это сделано для дальнейшего анализа результатов назначений. Опытная эксплуатация проводилась в ФГБУ «Туберкулезный санаторий «Голубая бухта» Минздрава России. По мнению практических специалистов КС работоспособна и является хорошим помощником при назначении процедур. Случаи несовпадения назначений КС и врачей анализируются отдельно раз в полугодие. На основе фактов несовпадения и новых данных корректируются РП и соответствующие изменения вносятся в КС.

Выводы

Предлагаемый авторами подход обладает следующими достоинствами:

1. Результаты КС хорошо интерпретируются и удобны для восприятия практикующими медицинскими работниками так как они активно участвовали в ее создании, а все используемые методы МО имеют хорошие визуальные интерпретации, которые демонстрировались и пояснялись практическим специалистам.
2. КС проста в использовании и обслуживании.
3. Последовательность и методы анализа данных не привязаны к профилю конкретного санатория. Поэтому подход тиражируем, требуются только незначительная настройка системы на исходных данных конкретного санатория, определяемые спецификой профиля медицинского учреждения.

ЛИТЕРАТУРА

1. Клинические рекомендации «Туберкулез у взрослых» /Общероссийская общественная организация «Российское общество фтизиатров», Национальная ассоциация некоммерческих организаций фтизиатров «Ассоциация фтизиатров» 2022 151с. 62–63. Доступ по ссылке https://cr.minzdrav.gov.ru/recomend/16_2. Ссылка активна на 24.06.2023 г.
2. Баласаняц Г.С. Концепция развития фтизиатрической санаторной помощи больным туберкулезом в Российской Федерации. Медицинский альянс научно-практический журнал. — СПб, 2013. — №4. — стр.79–83.
3. Григорьев Ю.Г. Противовиброзная электромагнитная и лазерная терапия во фтизиатрии//Туберкулёз и социально значимые заболевания № 4_2018 С.68–74.
4. Ломаченков В.Д., Стрелис А.К. Физиотерапия при туберкулезе легких Москва «Медицина», 2000, 64с. с. 2
5. Пальцев М.А., Белушкина Н.Н., Чабан Е.А. «4П-медицина как новая модель здравоохранения в Российской Федерации»// ОРГЗДРАВ: Новости. Мнения. Обучение. Вестник ВШОУЗ, № 2 (2), 2015, с. 48–54.
6. Hood L., Balling R., Auffray C. Revolutionizing medicine in the 21st century through systems approaches // Biotechnol. J. 2012. Vol. 7 (8). P. 992–1001, doi: 10.1002/biot.201100306
7. Парфенов А.А., Датий А.В., Лимонов В.И., и др. Проблемы развития санаторно-курортного комплекса России. Вестник восстановительной медицины, № 4(92), 2019, с. 67–69.
8. Санаторий с искусственным интеллектом. Что придумали ученые в Крыму?/ Пресс-служба Корпорации развития Республики Крым 04 января 2022. Доступ по ссылке <https://invest-in-crimea.ru/news/sanatoriy-s-iskusstvennym-intellektom-chto-pridumali-ucheny-v-krymu>. Ссылка активна: 24.06.2023 г.
9. Тополь Э. Искусственный интеллект в медицине. М.; Альпина Пабли, 2022. — 398 с.

© Чернавин Павел Федорович (chernavin.p.f@gmail.com); Чернавин Николай Павлович (ch_k@mail.ru);
Чугаев Юрий Петрович (doctorchugaev@mail.ru); Черняев Игорь Анатольевич (obltuborg@yandex.ru);
Калинина Людмила Васильевна (ivanovagolubayabuhta@yandex.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»