

ПРИМЕНЕНИЕ КЛАССИЧЕСКОГО МАШИННОГО ОБУЧЕНИЯ В ГЕМАТОЛОГИИ

Славинский Г.А.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Боброва Н.Л. – к. т. н., доцент, доцент кафедры информатики

Аннотация: в статье рассматриваются классы задач и методы классического машинного обучения. Представлены примеры задач гематологии, решение которых возможно с помощью классического машинного обучения. Выделены преимущества и недостатки, описаны примеры успешной реализации. Выдвинуто предположение о будущем указанного направления в медицине.

Ключевые слова: машинное обучение, персонализированная медицина, гематология, биоинформатика.

Введение. Современная медицина становится все более персонализированной и ориентированной на пациента, стремится к использованию индивидуальных профилей пациентов для разработки и выбора возможных вариантов диагностики и лечения [1].

Использование больших данных и аналитики делает возможным принятие более обоснованных решений, основанных на индивидуальных характеристиках пациента и его медицинской истории, что создает основу для интеграции методов машинного обучения.

Машинное обучение позволяет компьютерным системам обучаться и адаптироваться к новым данным без прямого программирования [2]. Данная область знаний получила широкое распространение в гистологических исследованиях, геномике и анализе медицинских данных. Таким образом, технологии искусственного интеллекта продолжают интегрироваться в исследовательскую и практическую медицину, включая гематологию [3].

Гематология является областью медицины, изучающей заболевания крови и кроветворных органов, включая анемию, лейкомию, лимфомы, тромбоцитопению. Ее значимость заключается в серьезных последствиях патологий кроветворения для здоровья пациентов, что требует комплексного подхода к диагностике и лечению заболеваний.

Основная часть. Область машинного обучения включает несколько основных направлений, к которым относятся ансамблевое обучение, нейронные сети и глубокое обучение, обучение с подкреплением и классическое машинное обучение [4]. В данной статье рассматривается классическое обучение.

Определение классического машинного обучения. Классическое машинное обучение представляет собой совокупность традиционных статистических и математических методов, что способны эффективно справляться с обработкой небольших наборов данных.

К основным особенностям классического машинного обучения относятся:

- использование классических алгоритмов;
- эффективная работа с небольшими наборами данных;
- решение формальных задач, таких как поиск и классификация;
- возможность применения технологии в различных областях.

В классическом машинном обучении выделяют два подхода к построению моделей [5]:

- обучения модели с учителем;
- обучения модели без учителя.

Обучение с учителем. Данный подход является одним из наиболее распространенных в машинном обучении, и применяется в распознавание речи, компьютерном зрении, прогнозировании. При подходе обучения с учителем, модель обучается на основе набора

размеченных данных, где каждый пример имеет соответствующую метку – желаемый выход модели, то есть установлены пары «вход-выход».

Обучение с учителем включает две подкатегории:

- классификация;
- регрессия.

Классификация направлена на предсказание категориальных меток на основе входных данных. Ее задача заключается в разделении множества объектов на группы, называемые классами, в зависимости от их характеристик, и предсказании того, к какому классу принадлежит объект [6].

В процессе классификации алгоритмы обучаются на размеченных данных, где каждому объекту уже присвоена категория. После обучения модель может предсказывать категории для новых данных.

Классификация может быть бинарной, ограничивающейся только двумя классами, или многоклассовой, что предполагает наличие более двух классов. Это зависит от задачи.

Примерами применения классификации в гематологии являются патофизиологический анализ гемограмм и оценка типовых нарушений системы крови [7]: классификация типов клеток крови (эритроцитов, лейкоцитов, тромбоцитов) на основе морфологических характеристик; классификация патологических клеток. Примером патологических клеток являются бластные клетки при лейкозах [8].

Среди основных методов классификации можно выделить [5]: логистическую регрессию; k-ближайших соседей; деревья решений; случайный лес.

В машинном обучении, классификация выделяется эффективностью в обработке задач с несколькими установленными классами, и находит широкое применение в различных областях, включая медицинскую диагностику, фильтрацию, распознавание изображений.

Регрессия используется для предсказания непрерывных числовых значений на основе независимых признаков. Модель обучается на размеченных данных, чтобы установить связь между зависимой переменной, то есть выходом, и одной или несколькими независимыми переменными, то есть входами [6].

Способность к моделированию делает регрессию важным инструментом прогнозирования в различных областях, включая медицину. Она помогает не только в предсказании будущих значений на основе имеющихся данных, но и способствует пониманию взаимосвязей между переменными, что имеет большое значение для принятия решений.

Примеры применения регрессии в гематологии [8]:

1 Прогнозирование показателей: предсказание уровня гемоглобина на основе таких факторов, как возраст, пол, и наличие хронических заболеваний; предсказание количества клеток крови (эритроцитов, лейкоцитов, тромбоцитов) на основе определенных показателей.

2 Оценка влияния: анализ того, как различные методы лечения влияют на количество лейкоцитов или тромбоцитов в крови; предсказание ответных реакций на различные подходы к терапии, основываясь на их индивидуальных характеристиках пациента.

3 Анализ факторов риска: регрессия может быть использована для выявления факторов развития анемии на основе данных пациента и его медицинской истории; регрессионный анализ способен помочь в изучении связей между уровнями различных биомаркеров в крови и наличием определенных заболеваний, таких как лейкоemia или лимфома.

Среди основных методов регрессии можно выделить [5]: линейную регрессию; множественную регрессию; полиномиальную регрессию.

Таким образом, классификация в машинном обучении — это отнесение объекта к одному из классов на основании его признаков. В отличие от регрессии, где предсказываются непрерывные значения, классификация фокусируется на предсказании дискретных меток, таких как «да» или «нет», «болен» или «здоров».

Обучения без учителя. При данном подходе модель обучается на неразмеченных данных, без заранее известных правильных ответов или меток для входных данных. Объекты

группируются на основе их сходства. Целью обучения является выявление скрытых закономерностей, то есть попытка поиска связей в данных без внешнего руководства.

Обучение без учителя используется в тех случаях, когда нет заранее известных правильных ответов, но необходимо найти скрытые закономерности в данных. Это особенно полезно при анализе больших объемов неструктурированных данных, когда нужно выявить внутренние связи и группы.

Обучение без учителя включает следующие подкатегории:

- кластеризация;
- поиск ассоциативных правил;
- снижение размерности данных.

Кластеризация – это метод машинного обучения, который используется для группировки объектов в кластеры таким образом, чтобы объекты внутри одного кластера были скорее похожи друг на друга, чем на объекты из других кластеров. Цель кластеризации заключается в разбиении данных на группы с учетом сходства объектов [9].

Данный процесс имеет следующие задачи:

- разбиение множества объектов на группы, согласно выделенным признакам;
- выделение нетипичных объектов, не подходящих ни к одному из кластеров;
- построение иерархии множества объектов.

Методы кластеризации разбивают данные на группы таким образом, чтобы объекты внутри каждой группы были похожи друг на друга. В отличие от классификации, где объекты распределяются по заранее известным категориям, кластеризация работает без меток, и осуществляет поиск групп в данных. Это делает использование кластеризации удобным в ситуациях, когда заранее неизвестно как само количество существующих групп, так и их характеристики.

Примеры применения кластеризации в гематологии [8]:

1 Группировка клеток крови на основе их морфологических или биохимических характеристик для выявления подтипов. Это может помочь в оценке риска определенных категорий заболеваний, исходя из данных о составе крови.

2 Объединение пациентов в группы на основе их показателей, для выявления подгрупп с общими особенностями. Это может быть полезно для разработки персонализированных подходов к лечению и мониторингу состояния здоровья пациентов.

3 Кластеризация клеток с целью выявления увеличенного количество кластеров и, как следствие, заключения о клиническом проявлении заболевания.

Среди основных алгоритмов кластеризации можно выделить [5]: иерархическую кластеризацию; k-means; алгоритм DBSCAN.

Кластеризация предоставляет большие возможности для анализа данных, и способна выявлять скрытые паттерны. Эти примеры демонстрируют как сильные стороны рассматриваемых технологий, так и возможность интеграции методов машинного обучения в медицинскую практику, с целью повышения эффективности диагностики и лечения.

Уменьшение размерности – это набор техник, направленных на уменьшение числа признаков в наборе данных. Данный набор техник способен значительно улучшить производительность алгоритмов машинного обучения, упростить модели, ускорить обучение, уменьшить риск переобучения [9].

В задачах со слишком большим пространством признаков или значительным уровнем шума в информации, подобный подход может помочь уменьшить размерность данных, сохранив наиболее важную информацию. Это способствует ускорению работы алгоритмов, уменьшению уровня шума и избыточности в данных. Таким образом, цель уменьшения размерности заключается в сокращении количества признаков в данных и сохранении максимума полезной информации.

В контексте применения машинного обучения в гематологии, данный набор техник способен к определению наиболее информативных характеристик клеток, такие как размер, форма и текстура, что способно упростить последующий анализ и визуализацию данных.

Среди основных методов уменьшения размерности можно выделить [5]: анализ главных компонент (PCA), линейный дискриминантный анализ (LDA).

Таким образом, методы уменьшения размерности способны помочь в упрощении сложных наборов данных, сохраняя при этом наиболее значимую информацию. Это актуально для медицины, где данные могут содержать множество признаков, включая морфологические характеристики клеток крови, генетические маркеры и результаты лабораторных тестов.

Поиск правил в машинном обучении является процессом выявления закономерностей и правил из данных. Он часто используется для принятия решений и предсказаний. Данный подход может быть особенно полезен в медицине, где необходимо интерпретировать данные, делая выводы на основе выявленных закономерностей [9]. Целью поиска правил является выявление скрытых закономерностей и ассоциативных правил в данных.

Примеры использования ассоциативного анализа в гематологии [8]:

- выявление факторов риска развития заболеваний на основе медицинских данных;
- выявление корреляций между морфологическими признаками клеток и их типами.

Среди основных методов поиска ассоциативных правил можно выделить [5]: алгоритмы ассоциативного анализа, примером являются Apriori и FP-Growth.

Применение поиска правил способно привести к значительным улучшениям в принятии решений, повышению эффективности процессов.

Примеры успешных реализаций. Гематологическая аналитика подразумевает автоматизированный анализ крови, предоставляющий клиническую информацию о состоянии кроветворной системы [8]. Традиционные анализаторы являются дорогими и громоздкими устройствами, их доступность может быть сомнительной. С другой стороны, ряд успешных стартапов разрабатывают решения, устраняющие эти неудобства с помощью технологий искусственного интеллекта, облачных вычислений и микрофлюидики.

Примером такого стартапа является устройство, разработанное компанией Sight Diagnostics, способное выполнить общий анализ крови за несколько минут. Анализатор использует алгоритмы машинного зрения и позволяет проводить анализ прямо в кабинете врача. Другим интересным примером является устройство, разработанное PixCell Medical, способное выполнить дифференциальный анализ по капле крови за 6 минутам. В проекте применен анализ изображений на основе искусственного интеллекта.

Помимо гематологии, машинное обучение успешно оказывает влияние и на другие области медицины, включая фармакологию. Стартап ClosedLoop.ai использует прогнозирующие модели для анализа данных о здоровье пациентов и предсказания результатов лечения. Стартап BenevolentAI использует алгоритмы для анализа различных данных, стремясь к выявлению новых возможностей в терапии, что значительно сокращает время и затраты на разработку новых медикаментов.

Таким образом, использование машинного обучения в гематологии и фармакологии существенно влияет на точность и скорость обработки медицинской информации, что способно оказать положительное влияние на лечение различных заболеваний крови.

Преимущества и недостатки применения классического обучения в гематологии. Машинное обучение обеспечивает новые возможности для обработки стремительно растущих объемов информации. Алгоритмы позволяют исследовать генетические вариации и их связь с заболеваниями, что способствует пониманию механизмов заболеваний и разработке целевой терапии.

Классическое машинное обучение предоставляет мощные инструменты для эффективной классификации различных типов заболеваний, прогнозирования вероятности рецидива или ответа на лечение, определение индивидуализированного подхода к лечению, выявления скрытых паттернов и связей, что не заметно при традиционных методах анализа.

Преимущества использования классического обучения в гематологии:

- высокая точность диагностики заболеваний;
- способность к более глубокому пониманию патофизиологий.

Недостатки классического обучения в гематологии:

- необходимость в больших объемах данных;
- зависимость от качества данных.

Таким образом, классическое обучение в гематологии имеет свои сильные и слабые стороны, которые важно учитывать при выборе методов диагностики и лечения.

Заключение. Благодаря своей способности обрабатывать и анализировать большие объемы данных, классическое машинное обучение имеет значительный потенциал как в гематологии, так и в разработке новых лекарственных препаратов. Оно способно улучшить диагностику состояния здоровья пациентов, персонализировать лечение и оптимизировать клинические процессы.

Большое количество задач, при решении которых вполне обоснованно применение методов классического машинного обучения, отражает актуальность их использования в гематологии. Предполагается, что необходимость использования методов машинного обучения будет расти вместе с объемами информации. Повсеместная цифровизация отраслей только подчеркивает актуальность использования современных методов обработки информации в медицине.

Список литературы

1. Blobel, B. *Translational medicine meets new technologies for enabling* / B. Blobel // *Studies in Health Technology and Informatics*. – 2013. – №189 – P. 8–23.
2. Глубокое обучение в биологии и медицине / Рамсундар Б. [и др.]. – Москва : ДМК, 2020. – 200 с.
3. Лучинин, А. С. Искусственный интеллект в гематологии / А. С. Лучинин // *Клиническая онкогематология*. – 2022. – №15. – С.16-27.
4. Mitchell, T. M. *Machine learning* / T. M. Mitchell – New York ; McGraw-Hill Education, 1997. – 432 p.
5. Чернавин, П. Ф. Практический курс классического машинного обучения с использованием моделей математического программирования: учебное пособие. / П. Ф. Чернавин, Н. П. Чернавин, Ф. П. Чернавин. – Екатеринбург : Издательство Уральского университета, 2023. – 124 с.
6. Кугаевских, А. В. Классические методы машинного обучения: учебное пособие. / А. В. Кугаевских, Д. И. Муромцев, О. В. Кирсанова. – Санкт-Петербург : ИТМО, 2022. – 53 с.
7. Черных, Е. М., Компьютерная система классификации лейкоцитов на изображениях клеток крови / Е. М. Черных, В. М. Михелев // *Научные результаты. Информационные технологии*. – Белгород : БелГУ, 2019. – №3 – С. 25–32.
8. Висмонт, Ф. И. Патологический анализ гемограмм и оценка типовых нарушений системы крови: учебное пособие. / Ф. И. Висмонт, Л. С. Лемешонок, Д. М. Попутников. – Минск : БГМУ, 2011. – 79 с.
9. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. / П. Флах – Москва : ДМК Пресс, 2015. – 387 с.

UDC 004.9:616.15

APPLICATION OF CLASSICAL MACHINE LEARNING IN HEMATOLOGY

Slavinsky G.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Boborova N.L. – Cand. of Sci., associate professor, associate professor of the department of Computer Science

Annotation. The article discusses the classes of problems and methods of classical machine learning. It presents examples of hematology tasks that can be solved using classical machine learning. The advantages and disadvantages are highlighted, along with examples of successful implementations. A hypothesis about the future of this direction in medicine is proposed.

Keywords. Machine learning, personalized medicine, hematology, bioinformatics.