



Искусственный интеллект и машинное обучение в здравоохранении

Махлышева А, Баккеволл ПА, Нуришетта АТ, Линстад Л

Широкое использование информационных систем в здравоохранении и хранение информации о здоровье пациента в электронном виде генерируют огромные объемы данных. Данные хранятся в разных системах в различных форматах. Около 80 процентов медицинских данных не структурировано (см. Рис.1).

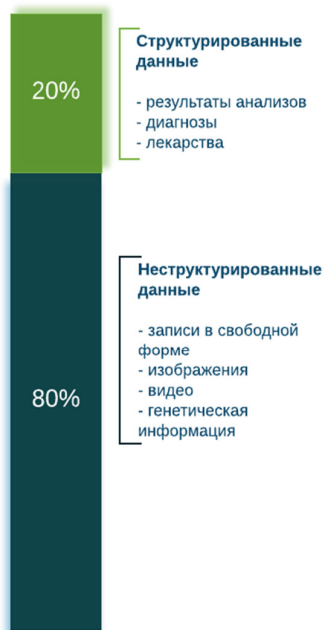


Рис.1: Типы медицинских данных

Данные имеют высокий потенциал для предоставления медицинских услуг и улучшения их качества, научных исследований, национального здравоохранения, планирования и управления системой здравоохранения.

Сочетание большого объема данных разного типа, возросших вычислительных мощностей и улучшенных методов анализа способствует быстрому автоматизированному созданию алгоритмов машинного обучения (МО), способных анализировать комплексные данные и предоставлять точные результаты.

Четыре типа машинного обучения

- При обучении с учителем алгоритмы учатся путем сравнения собственных прогнозов с данными, которые проверены и маркированы экспертом.
 - При обучении без учителя цель состоит в выделении базовых структур или закономерностей в данных. При этом данные заранее экспертом не маркируются.
- Полуавтоматическое обучение сочетает использование небольшого количества маркированных данных с большим количеством немаркированных данных.
- При обучении с подкреплением алгоритм вознаграждается при принятии правильного решения в данной ситуации.

Трудности при внедрении машинного обучения

Алгоритм становится «чёрным ящиком», когда процесс, стоящий за полученными результатами, слишком сложен для понимания.



Обработка персональных медицинских данных должна гарантировать *конфиденциальность*. Согласно Общему регламенту по защите данных (GDPR), данные должны быть подходящими, актуальными и ограниченными целью, для которой они обрабатываются. Пациенты должны быть информированы о целях обработки данных и иметь возможность выбора, разрешать ли использование своих персональных данных.

Алгоритм может делать *неверные прогнозы*. Часто причинами этого являются переобучение и недообучение. В случае переобучения алгоритм показывает хорошие результаты на тестовых данных, но плохо работает с новыми данными, поскольку вместо нахождения общей закономерности, объясняющей вариацию, находятся случайные закономерности в тестовых данных. При недообучении алгоритм не полностью использует предоставленные ему данные для обучения, что имеет место при выборе недостаточно сложной модели. Переобучение и недообучение связаны между собой: при уменьшении одного показателя другой возрастает. Важно найти баланс между нами - тогда общая ошибка прогнозирования станет наименьшей.

Что необходимо для успешного внедрения машинного обучения?

Системы МО отличаются от традиционных систем программного обеспечения. МО использует самообучающиеся, непрерывно совершенствующиеся алгоритмы. Перед внедрением в реальные условия система должна пройти обучение с использованием местных данных. У каждого алгоритма есть свои сильные и слабые стороны. Важно опробовать несколько алгоритмов, чтобы выявить наиболее подходящий для решения данной задачи. Стратегия разработки пилотной (пробной) системы МО для здравоохранения может заключаться в оценке клинических потребностей и технологических возможностей; успешные пилотные проекты готовы к внедрению.

МО показывает многообещающие результаты во многих исследовательских проектах. Для использования МО в поддержке принятия клинических решений, кроме настроенной модели МО, необходим доступ к данным пациентов. Это возможно при интеграции с системой электронных данных пациентов или другими системами хранения медицинских данных. Данная интеграция возможна 1) напрямую с системой электронных данных пациентов, 2) как облачное решение, предоставляемое третьей стороной, или 3) как облачное решение собственной разработки.

Где машинное обучение имеет наибольший потенциал?

МО преобразует данные в знания. Использование МО станет переломным¹ по крайней мере в трёх областях медицины при разных сроках реализации:

¹ Переломная технология - технология, способная коренным образом изменить установленные структуры и бизнес-модели. Например, если большая часть работы по анализу медицинских изображений в радиологии перейдёт к машинным алгоритмам, это можно считать переломом.



1. *Интерпретация медицинских изображений.* Эта область наиболее развита. Многие проекты доказали свою эффективность, показав результаты на том же или даже лучшем уровне врачей-специалистов. В ближайшие годы область будет стремительно развиваться.
2. *Прогнозирование.* Алгоритмы прогнозирования не настолько развиты. Пройдет около пяти лет, прежде чем эта область станет достаточно зрелой.
3. *Диагностика.* Это самая сложная из трёх областей. Потребуется около десяти лет, прежде чем такие решения будут готовы к использованию на практике.

Технология искусственного интеллекта может сделать здравоохранение проактивным. Однако для реализации огромного потенциала технологии необходимы дополнительные исследования, также учитывающие вопросы этики, конфиденциальности и информационной безопасности.

Кратко по теме

Алгоритмы машинного обучения могут анализировать большое количество разнообразных данных с точными результатами. Три переломных области машинного обучения в области здравоохранения:

- *Интерпретация медицинских изображений (глазные заболевания, радиология, патология)*
- *Прогнозирование (деменция, метастатический рак, инсульт)*
- *Диагностика (онкология, патология, редкие заболевания)*

Здравоохранение становится более проактивным с использованием ИИ, но для того, чтобы потенциал мог быть полностью реализован, требуется больше исследований и разработок.

Список литературы

1. Obermeyer, Z. and E.J. Emanuel, Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. The New England Journal of Medicine, 2016. 375(13): p. 1216-1219.