

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВОЗМОЖНОСТЕЙ ОСНОВНЫХ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОСЛЕДУЮЩЕГО ИХ ИСПОЛЬЗОВАНИЯ В ПАЦИЕНТО-ОРИЕНТИРОВАННЫХ ТЕЛЕМЕДИЦИНСКИХ СИСТЕМАХ

М.М. Баркалов, Е.А. Машина (Университет ИТМО, С.-Петербург)

**Научный руководитель – к.т.н., доцент П.В. Балакшин
(Университет ИТМО, С.-Петербург)**

В работе рассматриваются вопросы, сравнительного анализа возможностей и особенностей применения основных методов машинного обучения для выбора наилучших инструментов и технологий для реализации пациенто-ориентированных телемедицинских систем.

Цель работы – произвести обзор существующих алгоритмов машинного обучения и путем сравнительного анализа их возможностей и условий применимости выбрать наилучшие с точки зрения использования в телемедицинской области, как для получения адекватных результатов в обработке существующих документов объективного контроля состояния пациентов и их систематизации, так и для дальнейшего проведения диагностики по вводимому набору симптомов.

Описываемое исследование производится в рамках начального этапа создания пациенто-ориентированных телемедицинских систем семейства PersonalMedHelper, предполагающих, что основными их пользователями системы выступят люди, не имеющие большого опыта в диагностировании болезней, поэтому в качестве входных данных выступают симптомы и прочие характеристики человека, которые могут быть измерены в условиях отсутствия квалифицированного медицинского персонала, а так же ряд результатов общедоступных результатов объективных медицинских анализов (измерений, фактографии). А в качестве выходных данных система должна сообщить предполагаемую причину болезни и сопутствующую информацию: рекомендации по первой помощи и истории болезней с похожими случаями, которые были ранее известны системе.

Следовательно, наличие подобных ограничений на объем и характер входной-выходной информации приводит к ряду ограничений на машинное обучение медицинских информационных систем [2, 3].

Во-первых, данные, используемые для контролируемого обучения подобных систем, необходимо будет маркировать вручную, выбирая исключительно бесспорные варианты, поскольку пользователь, не обладающий необходимой медицинской квалификацией, не в силах самостоятельно учесть недостаточно коррелирующийся результаты, что делает этот процесс весьма трудоемким. Во-вторых, получить достаточно большой и полный набор данных для обучения по бесспорным вариантам может быть трудно, информация может быть «грязной», разрозненной, храниться в разных приложениях и форматах и содержать ошибки, а низкое качество данных на входе приведет к столь же низкому качеству на выходе. В-третьих, может оказаться трудно объяснить в человеческих терминах результаты, полученные с использованием больших сложных моделей, а регуляторов часто интересует, по каким критериям было принято то или иное решение, если речь идет, например, о столь непосредственно затрагивающей безопасность граждан области, как здравоохранение. В-четвертых, трансфер результатов обучения пока слабо развит и большинство алгоритмов испытывают проблемы с переносом знаний, полученных в одной ситуации, в другую. В-пятых, исходный набор данных и алгоритмы, заложенные человеком, могут быть сознательно или неосознанно искажены, что может быть особенно заметно для систем, пользователями которых могут являться люди с недостаточной квалификацией [1].

Промежуточным результатом проведенного сравнительного анализа алгоритмов машинного обучения, можно назвать алгоритмы, специализирующиеся на многоклассовой классификации для прогнозирования той или иной категории, а именно:

- логистическая регрессия (Logistic Regression) — функция вероятности принадлежности значения к классу;
- лес деревьев решений (Decision Forest) — создание множества деревьев выбора и дальнейший выбор наилучшего;
- джунгли деревьев решений (Decision Jungle) — усложненный вариант леса дерева, в котором допускается слияние деревьев;
- нейронная сеть (Neural Network) — использование набора взаимосвязанных слоёв, на которых происходит обработка данных;
- один против всех (One-vs-All) — введение меры уверенности в выборе той или иной категории.

Однако конкретная применимость тех или иных алгоритмов обработки информации в пациенто-ориентированных телемедицинских системах может быть определена лишь путем практических разработок и открыта для появления новых, более эффективных решений. При этом важно, чтобы поиск окончательного решения основывался на хорошей теоретической и фактографической базе.

Литература:

1. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2015. – 400 с.: ил.
2. Хайкин С. Нейронные сети: полный курс, 2-е изд.: Пер. с англ.-М.: ООО «И.Д. Вильямс», 2016. – 1104 С.
3. Shai Shalev-Shwartz, Shai Ben-David. Understanding Machine Learning from Theory to Algorithms. - Cambridge University Press. 2014. - P.449.