

ПРИНЦИПЫ РАБОТЫ АЛГОРИТМОВ ГЛУБОКОГО МАШИННОГО ОБУЧЕНИЯ И ПОСЛЕДУЮЩЕЕ ИХ ИСПОЛЬЗОВАНИЕ В АВТОМАТИЗИРОВАННЫХ ПАЦИЕНТО-ОРИЕНТИРОВАННЫХ СИСТЕМАХ.

Яганова М.М. (МБОУ СОШ №17, Тверь)

Научный руководитель - Е.А. Машина (Университет ИТМО, С.-Петербург)

В работе рассматриваются вопросы, касающиеся особенностей применения и принципов работы наилучших методов машинного обучения для применения их в качестве инструментов для создания и использования автоматизированных пациентно-ориентированных систем.

Цель работы – проанализировать существующие методы и алгоритмы глубокого машинного обучения, провести сравнительную оценку их способностей, требований и условий, учитывая преимущества и недостатки каждого из вариантов, и, по итогу, выбрать наилучшие варианты для создания концепции одной из систем с точки зрения удобного и комфортного использования пациентом при проведении диагностики по вводимому набору симптомов.

Данное исследование проводится в рамках проекта PersonalMedHelper и система «Family doctor» является прототипом этого семейства телемедицинских систем. Целевой аудиторией и основные пользователи – люди, не обладающие специальными навыками для диагностирования болезней, поэтому информация, указываемая в систему, должна быть измерима в условиях отсутствия квалифицированного персонала и специального оборудования. В качестве результата анализа введенных данных на выходе пациент получает четкий и понятный предполагаемый диагноз, а также рекомендации о лечении и профилактики того или иного заболевания, истории болезней с похожими случаями, контактную информацию о близлежащих поликлиниках.

В связи с такими ограничениями на указываемую информацию и объем данных, существует ряд ограничений на используемые инструменты машинного обучения [2, 3].

При выборе подходящего метода машинного обучения следует учитывать следующие факторы:

1. Используемые данные для контролируемого обучения нейронных сетей должны быть подвергнуты сортировке и маркировке, так как пациент, не обладающий соответствующими знаниями в области медицины, не сможет самостоятельно учесть недостаточно коррелирующийся результаты, что приводит к замешательству пользователя.
2. Большие объемы данных, используемые и обрабатываемые нейронными сетями, чаще всего хаотично находятся в системе в разных папках, форматах и приложениях, что приводит к ошибкам в анализе и неточности выдаваемого результата. В связи с этим, входные данные имеют достаточно низкое качество, что приводит к низкому качеству выходных данных.
3. Для увеличения точности результата, стоит использовать самообучаемые алгоритмы машинного обучения, которая могла бы анализировать данные всех пациентов, использующих систему, и делать выводы, поправки в процессе утверждения диагноза.
4. Алгоритмы машинного обучения должны быть способны анализировать каждый объем данных пациентов индивидуально, не используя усредненные значения и результаты.
5. Достаточно сложно описать человеческими понятиями результаты анализа больших объемов данных, а так как пациенту очень часто интересно и важно знать, по каким критериям вынесено то или иное решение, поэтому стоит использовать алгоритмы, умеющие описать получение данных выводов.

б. Учитывая погрешности при занесении данных, а также округление значений и других изменений данных, происходит искажение информации, что очень критично для анализа нейронными сетями [1].

По итогу исследования алгоритмов машинного обучения, учитывая выявленные требования, стоит выделить, что наиболее подходящими являются алгоритмы, специализирующиеся на многоклассовой классификации, то есть способны учитывать обширный объем параметров, таких как: логистическая регрессия, лес деревьев решений, джунгли деревьев решений, улучшенное дерево принятия решений, нейронная сеть, метод опорных векторов и другие [1].

Окончательный конкретный выбор подходящих алгоритмов обработки объемов данных в автоматизированных пациенто-ориентированных системах должен быть выявлен на практике и тестовом анализе различных заболеваний. Возможно расширение списка возможных методов машинного обучения и появление новых, более эффективных и удобных способов.

Литература:

1. Ясницкий Л.Н. Введение в искусственный интеллект: учеб.пособие для студ.высш.учеб.заведений, 3-е изд.,стер. - М.:Издательский центр «Академия», 2010. - 176 с.
2. I.H. Witten, E. Frank Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). — Morgan Kaufmann, 2005 ISBN 0-12-088407-0
3. Люгер Дж. Ф. Искусственный интеллект: стратегии и методы решения сложных проблем = Artificial Intelligence: Structures and Strategies for Complex Problem Solving / Под ред. Н. Н. Куссуль. -- 4-е изд.. -- М.:Вильямс, 2005.

Автор:

Научный руководитель: