

Интеллектуальный анализ медицинских данных.

Колмакова Е.С.

*Научный руководитель –
канд. физ.-мат. наук, доцент
Гафаров Ф.М.*

Инфаркт – серьезное заболевание сердечно-сосудистой системы, в тяжелых случаях требующее хирургического вмешательства. По причине колоссальной важности сердца задача превентивного обнаружения инфаркта является более чем важной. Эта задача может быть решена несколькими способами: врачебный опыт и знания или компьютерное прогнозирование. Ни один из них не является абсолютот в вопросе обнаружения предрасположенности к инфаркту, но вкупе это максимально эффективный способ.

Существует множество методов компьютерного прогнозирования, один из которых – использование нейронных сетей для анализа данных пациентов, наблюдавшихся ранее, чтобы по этим данным научиться определять вероятность на успешное лечение для текущих пациентов. Для решения данной задачи потребуется затронуть множество других важных подзадач, среди которых особенно выделяются следующие:

1. Необходимость структурировать входные данные. Структурирование заключается в анализе истории болезни пациента для последующего изъятия важной информации.
2. Необходимость написания эффективного алгоритма перевода текста в формат, понятный нейронной сети, – числовой вектор. Массив входных данных крайне широк: он содержит чуть меньше десяти тысяч историй обращений пациентов, каждое из которых, в свою очередь, содержит от ста до нескольких тысяч слов. По этой причине необходимость эффективного способа векторизации крайне важна, так как в противном случае будут использованы слишком затратные ресурсы памяти и времени. Для эффективной векторизации была применена технология word2vec, позволяющая переводить текст в векторное представление быстро и результативно.

Проанализировав типичную историю болезни, была выделена структура документа и определена основная информация, которая должна составить входные данные для нейронной сети:

1. Результат лечения. Эта информация необходима в качестве источника сверки выходных данных нейронной сети.
2. Первичный осмотр. Отображает анамнез пациента, то есть все его симптомы, перенесенные заболевания, хронические заболевания, наследственность, вредные привычки и т.д.
3. Ежедневные осмотры.
4. Медицинские манипуляции, такие как ЭКГ или электрокардиограмма, коронарография и другие.

5. Записи о проведенных операциях.

Наибольшее применение в области обработки естественного языка нашли рекуррентные нейронные сети, главным отличием которых является наличие обратной связи. Реализаций у рекуррентных сетей немало, но именно LSTM (Long Short Term Memory – сети кратковременной и долговременной памяти) в настоящее время стала наиболее популярной в сравнении с остальными типами, продемонстрировав большой успех в решении многих задачах NLP (neuro-linguistic programming – нейролингвистическое программирование). LSTM сеть легко может разбирать смысл предложений, что является ключевой подзадачей при анализе медицинских документов с целью прогнозирования наступления инфаркта миокарда в определенных условиях.

Подав данные, разбитые по критерию «успешное/неуспешное лечение» нейронная сеть показала довольно хорошие результаты. Для проверки качества обучения нейронной сети было создано три тестовый набора:

- о Набор №1 - это 13% неуспешных историй на 87% успешных
- о Набор №2 - это 100% неуспешных историй на 0% успешных
- о Набор №3 - это 45% неуспешных историй на 55% успешных

Результаты по наборам получились следующие: на первом наборе процент корректности составил 89,72%, на втором наборе это 81,17%, на третьем - 87,78%. Наглядно данные представлены на Рис.1. Данное разбиение показало себя более эффективным, обучив модель с точностью в 80-90% определять успешность лечения инфаркта миокарда.

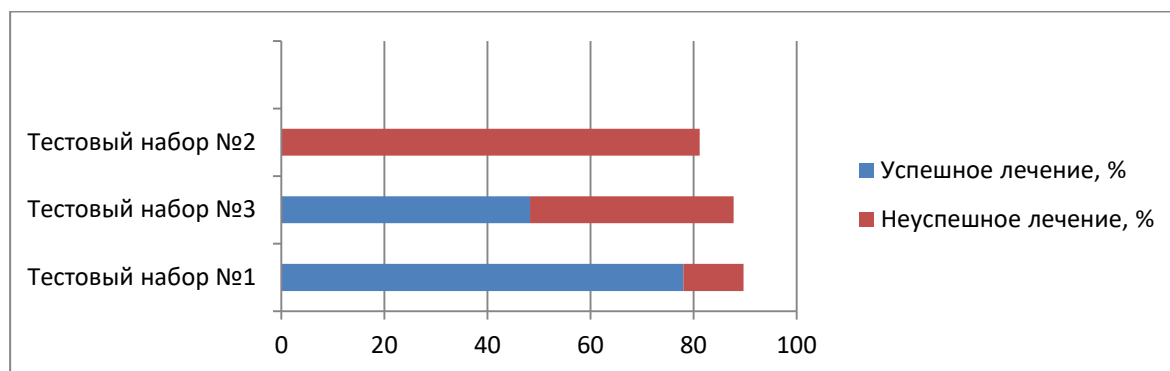


Рис 1. Результаты работы нейронной сети

ЛИТЕРАТУРА

1. *Filippo Amato* Artificial neural networks in medical diagnosis / Filippo Amato, Alberto López, Eladia María Peña-Méndez, Petr Vaňhara, Aleš Hampl, Josef Havel // Journal of Applied Biomedicine – С.48-56
2. *Sepp Hochreiter* Long Short Term Memory / Sepp Hochreiter, Jurgен Schmidhuber // Neural Computation 9(8) – 1997. – С.1735-1780

3. *Дейтел Пол Python: Искусственный интеллект, большие данные и облачные вычисления.* / Дейтел Пол, Дейтел Харви // — СПб.: Питер, 2020. — 864 с.: ил. — (Серия «Для профессионалов»).
4. *Джоши Искусственный интеллект с примерами на Python.*/ Джоши, Пратик // : Пер. с англ. - СПб. : ООО "Диалектика", 2019. - 448 с.