

Построение предиктивной подсистемы анализа данных обследования пациентов с субъективными и легкими когнитивными нарушениями

В.Н. Гридин¹, Б.Р. Салем¹, В.И. Солодовников¹

¹Центр информационных технологий в проектировании РАН, Маршала Бирюзова, 7а, Одинцово, Россия, 143000

Аннотация

Развитие когнитивных нарушений на начальной стадии является малоизученным и сложнодиагностируемым процессом. В данной работе предложен комплексный подход, включающий анализ исходных данных и построение предиктивной подсистемы прогнозирования дальнейшего прогрессирования осложнений у пациентов с симптомами когнитивных нарушений. Совместное использование методов статистического анализа и машинного обучения—позволили достичь 75% точности диагностирования легких когнитивных нарушений на выборке из 152 пациентов, а методы визуализации способствуют улучшению интерпретируемости создаваемых моделей.

Ключевые слова

Когнитивные нарушения, предиктивный анализ, статистический анализ, интеллектуальный анализ данных, машинное обучение

1. Введение

Риск возникновения когнитивных нарушений, способных развиться в тяжелые расстройства, включая деменцию, распространенность которой в популяции людей старше 70 лет значительна и составляет порядка 13.9%[1]. Актуальность выявления нарушений на как можно более ранней стадии определяется крайней необходимостью принять меры по замедлению прогрессирования болезни, пока патологический процесс не начал отражаться на повседневной деятельности человека. Системный анализ и визуализация данных, полученных при обследовании пациентов и прохождении ими нейропсихологических тестов, позволили выявить ряд характерных особенностей и зависимостей [2], что послужило основой для построения предиктивной подсистемы прогнозирования развития патологии.

2. Построение предиктивной модели

Предиктивный анализ представляет собой класс методов обработки исторических данных для построения математической модели с целью прогнозирования поведения исследуемой системы в будущем. Он применим в медицинской сфере для выявления особенностей течения заболевания на основе обработки массива данных, характеризующих истории развития патологического процесса у различных групп пациентов, результатов их всестороннего обследования и диагностики. Для построения предиктивной модели применим математический аппарат, включающий статистические методы, методы интеллектуального анализа, теории игр [3], где исходный массив данных используются для ее настройки с учетом ключевых характеристик и закономерностей. Полученная модель применима в повседневной практике для обработки новых случаев, поддержки принятия решений и прогнозирования. Важным заключительным этапом является регулярная верификация работы модели в процессе накопления новых данных с целью проверки ее производительности и точности.

В рамках предложенного подхода были выделены 7 этапов построения предиктивной подсистемы, представленные на рисунке 1.



Рисунок 1: Этапы построения предиктивной подсистемы

3. Используемые методы в рамках анализа данных и моделирования

В большинстве случаев данные, описывающие результаты медицинских тестирований, представляют собой множество разнотипных признаков, что затрудняет их дальнейшую обработку. В связи с этим на этапе «Анализ данных» осуществляется предобработка исходной выборки, заключающаяся в проверке, преобразовании, кодировке и нормировке. Дополнительной целью такой предобработки, является увеличение информативности примеров. Этап «Статистика» направлен на проверку гипотез и предположений о структуре выборки с использованием стандартных статистических моделей. Здесь же применимы методы понижения размерности с учетом малозначимых и сильно коррелирующих признаков. Пятый этап «Моделирование» в своей основе представляет собой итеративный процесс, благодаря которому создаётся необходимое число предиктивных моделей с использованием методов искусственного интеллекта. Так в рамках анализа медицинских данных [2] для исходной выборки среди различных вариаций классификации и кластеризации, наилучший результат продемонстрировал метод случайного леса (Randomforest) [4,5]. В свою очередь методы визуализации позволили повысить интерпретируемость создаваемых моделей.

4. Заключение

В результате проведенного исследования была разработана подсистема, в основе которой применяются методы классификации и визуализации данных. Данная подсистема позволяет определить состояние когнитивных функций пациентов с целью оценки прогноза дальнейшего развития когнитивного нарушения. Полученная подсистема достигла 75% точности диагностирования легких когнитивных нарушений на выборке из 152 пациентов. В дальнейшем планируется развитие подсистемы для увеличения точности и обеспечения возможности ее использования в диагностировании широкого спектра заболеваний.

5. Благодарности

Работа выполняется в рамках темы № 0071-2019-0001.

6. Литература

- [1] Kasper, J.D. Disability and care needs of older Americans by dementia status: An analysis of the 2011 National Health and Aging Trends Study / J.D. Kasper, V.A. Freedman, B.C. Spillman. – Washington, DC, 2014.
- [2] Gridin, V. Data visualization and clustering in the task of system analysis of the patients examination results in the initial stages of cognitive impairment / V. Gridin, B. Salem, V. Solodovnikov // International Conference on Information Technology and Nanotechnology (ITNT). IEEE Publishing. – 2020. DOI: 10.1109/ITNT49337.2020.9253186.
- [3] Finlay, S. Predictive Analytics, Data Mining and Big Data: Myths, Misconceptions and Methods. – Basingstoke: Palgrave Macmillan, 2014. – P. 237.
- [4] Breiman, L. Random Forests // Machine Learning. – 2001. – Vol. 45. – P. 5-32.
- [5] Гридина, Н.В. Использование классификаторов для целей судебно-медицинской идентификации личности (диагностики возраста) / Н.В. Гридина, Г.В. Золотенкова, А.И. Рогачев // Биомедицинская радиоэлектроника. – 2019. – Т. 22, № 5. – С. 49-54. DOI: 10.18127/j15604136-201905-06.