

## ЛОГИЧЕСКАЯ КЛАССИФИКАЦИЯ БОЛЬШИХ ДАНЫХ

*Анисимова Диана Вадимовна*

*Студент*

*Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия*

*E-mail: diana.anisimova.01@mail.ru*

*Научный руководитель — Дюкова Елена Всеволодовна*

Рассматривается логический подход к одной из центральных задач машинного обучения — задаче классификации по прецедентам. В самых общих чертах эта задача ставится следующим образом. Исследуется множество объектов, представимое в виде объединения нескольких непересекающихся подмножеств, называемых классами. Каждый объект этого множества может быть представлен в виде набора значений целочисленных признаков (наблюдаемых характеристик объекта). Дана выборка объектов, называемых прецедентами, о которых известно, каким классам они принадлежат. Требуется уметь классифицировать объекты, не являющиеся прецедентами, то есть по признаковому описанию каждого такого объекта определять, какому классу он принадлежит.

Преимущество логического подхода заключается в возможности получения результата при отсутствии дополнительных предположений вероятностного характера и при небольшом числе прецедентов, а также в высокой интерпретируемости результата классификации. Недостаток традиционного подхода — вычислительная сложность, что естественно обусловлено применением дискретного аппарата.

В [1] предложена модель классификатора, основанная на поиске в признаковых описаниях прецедентов так называемых максимальных логических закономерностей. Главные недостатки базового варианта данной модели: низкая точность в связи с большим числом отказов от классификации и длительное время работы. Указанные недостатки устранены в новых вариантах классификатора, разработанных в [2]. Однако, несмотря на значительное увеличение скорости счета, новые алгоритмы все еще нацелены на обработку небольших объемов обучающей информации, как правило, не превышающих несколько тысяч объектов при малом числе признаков и несколько сотен признаков при небольшом числе объектов.

Для эффективного решения задач логического анализа данных актуальным направлением является исследование возможности применения нейронных сетей. Одна из первых логических нейронных

сетей была описана в работе [3]. Её главной особенностью является использование блоков, аппроксимирующих простейшие логические функции, такие как дизъюнкция, конъюнкция, отрицание. Каждый блок представляет собой два линейных слоя с функцией активации ReLU между ними. Логическая нейронная сеть использовалась в [3] для поиска максимальной совместной подсистемы системы булевых уравнений (ПМСП), левые части которых представлены в виде дизъюнктивных нормальных форм.

В настоящей работе предлагается оригинальная методика, согласно которой поиск максимальных логических закономерностей класса сводится к задаче ПМСП. Однако, в отличие от [3], задача ПМСП решается для случая, когда левые части булевых уравнений представлены конъюнктивными нормальными формами. Каждому классу соответствует своя нейронная сеть. Нейронная сеть, построенная для класса  $K$ , выдает вероятность того, что тестируемый объект не принадлежит классу  $K$ . В качестве функции потерь используется бинарная кросс-энтропия. Искомые максимальные логические закономерности формируются специальным образом в ходе обучения. На этапе классификации обученные нейронные сети применяются к тестовой выборке. Финальная оценка принадлежности объекта классу формируется аналогично известной процедуре голосования [1]. Предлагаемая схема логической нейронной сети реализована, и проведён ряд экспериментов на больших реальных данных. С этой целью использовался датасет «Connect-4» из репозитория UCI (<https://archive.ics.uci.edu/>), содержащий 67557 объектов и 126 бинарных признаков после применения процедуры One-hot кодирования. Показана практическая применимость построенных нейронных сетей для решения задачи логической классификации больших данных.

### Литература

1. Ковшов Н. В., Моисеев В. Л., Рязанов В. В. Алгоритмы поиска логических закономерностей в задачах распознавания. //Ж. вычисл. матем. и матем. физ. 2008. Т. 48, №2. С. 329–344.
2. Anisimova D., Djukova E., Djukova A. Supervised Classification Problem: Searching for Maximum Patterns // In Proceedings of the 2024 X International Conference on Information Technology and Nanotechnology (ITNT), Samara, Russian Federation, 2024, P. 1–4.
3. Shi S., Chen H., Ma W., Mao J., Zhang M., Zhang Y. Neural Logic Reasoning. 2020. <https://arxiv.org/pdf/2008.09514>