Фоменко Артём Викторович
студент
Баранов Илья Викторович
студент
Сергеев Александр Эдуардович
канд. физ.-мат. наук, доцент

ФГБОУ ВО «Кубанский государственный аграрный
университет им. И.Т. Трубилина»
г. Краснодар, Краснодарский край

БУЛЕВЫ ФУНКЦИИ И ИХ ПРИЛОЖЕНИЯ В МАШИННОМ ОБУЧЕНИИ: ОТ ЛОГИЧЕСКИХ ПРАВИЛ К ОБЪЯСНИМОМУ ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ

Аннотация: *в работе анализируются теоретические и прикладные аспекты использования булевых (логических) функций в современных задачах машинного обучения и объяснимого искусственного интеллекта. Показано, что модели, представленные в виде логических конструкций, обладают высокой степенью прозрачности, что критично для регулируемых отраслей: медицины, финансовой сферы и образования. Рассмотрены классические и новые методы синтеза, минимизации и извлечения логических правил из данных, включая алгоритм RIPPER, дистилляцию нейросетей и комбинированные подходы. Особый акцент сделан на исследованиях российской научной школы в области дискретной математики, логико-правилевых моделей и их использования в анализе образовательных процессов.*

Ключевые слова: *логические функции, объяснимый ИИ, машинное обучение, прозрачные модели, правила классификации, ДНФ, интерпретируемость, алгоритм RIPPER, аналитика в образовании, дискретная математика, российские разработки в XAI.*

Современные технологии искусственного интеллекта и машинного обучения активно внедряются в различные сферы человеческой деятельности: от медицины и финансов до транспорта и обучения. Глубокие нейронные сети, включая свёрточные и рекуррентные архитектуры, показывают выдающиеся результаты в задачах классификации и прогнозирования, однако их внутренняя работа часто остаётся непонятной для пользователя, превращаясь в своеобразный «чёрный ящик». Недостаток прозрачности снижает уровень доверия, затрудняет выявление смещений в данных и усложняет соблюдение нормативных требований, например, в области обработки персональных данных. Объяснимый искусственный интеллект (XAI) возник как направление, призванное сделать модели машинного обучения интерпретируемыми без существенной потери их эффективности. В регуляторных документах и аналитических отчётах отмечается, что в ближайшие годы организации будут всё чаще требовать от ИИ-систем не только точности, но и понятности принимаемых решений. В таких чувствительных областях, как здравоохранение, правоприменение и педагогика, важно не только получить результат, но и понять логику, по которой он был сформирован. Это позволяет экспертам проверять корректность выводов, исправлять ошибки и формировать обоснованное доверие к автоматизированным системам. Булевы функции, имеющие глубокие корни в дискретной математике и логике, предлагают естественный формализм для создания интерпретируемых моделей в виде набора логических условий. Они позволяют описывать зависимости с помощью двоичных операций (И, ИЛИ, НЕ), которые легко переводятся в понятные правила, обеспечивая «прозрачность» модели в отличие от сложных нейросетевых ансамблей. В данной статье рассматривается взаимосвязь булевых функций с машинным обучением, их место в методологии объяснимого ИИ, современные исследовательские тренды и практические примеры использования, в том числе в сфере образования.

Булева функция – это фундаментальное понятие дискретной математики, определяемое как отображение $f: \{0,1\}^n \rightarrow \{0,1\}$, где n – число входных переменных (признаков). Она описывает логические взаимосвязи с использованием

операций конъюнкции (∧, И), дизъюнкции (∨, ИЛИ) и отрицания (¬, НЕ). Например, функция f(x, y) = x ∧ y истинна только тогда, когда оба аргумента истинны. Подобные функции составляют основу цифровой логики, применяемой при проектировании микросхем, в криптографии и других областях. Одним из наиболее распространённых способов представления булевых функций является дизъюнктивная нормальная форма (ДНФ), в которой функция записывается как дизъюнкция конъюнкций литералов. Например, f(x, y, z) = (x ∧ y ∧ ¬z) ∨ (¬x ∧ y ∧ z). ДНФ обладает высокой интерпретируемостью, поскольку каждый конъюнктивный член соответствует определённому правилу, покрывающему часть положительных примеров. Конъюнктивная нормальная форма (КНФ) представляет функцию в виде конъюнкции дизъюнкций и часто используется в задачах проверки выполнимости булевых формул (SAT-задачи). Минимизация булевых функций – важная практическая задача, направленная на удаление избыточных компонентов для получения компактных ДНФ или КНФ; она тесно связана с NP-полной задачей о минимальном покрытии множества. Для её решения применяются алгоритм Квайна-МакКласки и эвристические методы, такие как Espresso, позволяющие находить близкие к минимальным формы для реальных данных. В монографии Я. Крама и П. Хаммера «Boolean Functions: Theory, Algorithms, and Applications» подробно изложены теоретические основы и алгоритмы работы с булевыми функциями, а в российском учебном пособии «Булевы функции и построение логических схем» приведены примеры минимизации и синтеза схем. При моделировании знаний булевы функции интерпретируются как наборы логических правил: каждый литерал соответствует бинарному признаку (например, условие «возраст > 30» кодируется как 1 или 0). Свойства функций, такие как монотонность или линейность (возможность представления через полиномы Жегалкина), используются для анализа структуры зависимостей. В российской математической традиции развиты спектральные методы и аппарат полиномов Жегалкина для исследования дискретных функций, что отражено в учебных курсах по теории булевых функций. Расширение классического булевого подхода на многозначные и нечёткие логики позволяет учитывать неопределённость и

неточность в данных. В современных работах по XAI рассматриваются и более сложные логические конструкции, например, пороговые операторы («истинно, если выполнено не менее k условий»), что повышает выразительность моделей при сохранении их интерпретируемости.

Связь булевых функций с машинным обучением особенно заметна в задачах обучения по прецедентам, где требуется вывести логическую формулу, приближающую целевую зависимость. Фактически, данные могут рассматриваться как частичная таблица истинности, и задача сводится к индукции булевых правил. Известным алгоритмом индукции правил является RIPPER, предложенный У. Коэном в 1995 году. Он использует стратегию последовательного покрытия: строит правила в форме ДНФ, а затем проводит их сокращение для минимизации ошибки. Получаемые правила вида «ЕСЛИ (признак A = 1) И (признак B = 0), ТО класс = 1» непосредственно соответствуют булевым выражениям и легко воспринимаются человеком. Деревья решений (ID3, C4.5 и др.) также тесно связаны с булевым представлением: каждый путь от корня к листу задаёт конъюнкцию условий, а всё дерево в целом может быть представлено как ДНФ. Булевы функции используются для анализа и упрощения деревьев, например, при ограничении глубины или числа листьев в целях повышения интерпретируемости. В современных работах по интерпретируемому ИИ (например, в статье С. Рудин 2019 года) отмечается, что в задачах с высокими рисками предпочтительнее изначально строить прозрачные модели, а не пытаться объяснять сложные «чёрные ящики» постфактум. Булевы функции также применяются при дистилляции сложных моделей: из нейронных сетей извлекаются глобальные или локальные логические правила, аппроксимирующие их поведение. В исследовании Ф. Мерани и Дж. Хоу (2019) рассматриваются точные и приближённые методы извлечения правил из нейросетей с бинарными признаками, основанные на выборке входных данных. Методы локальной интерпретации, такие как LIME и SHAP, могут дополняться глобальными логическими моделями, что обеспечивает сочетание локальной и глобальной объяснимости. Для данных с бинарными признаками булевы функции особенно естественны: непрерывные признаки могут быть

бинаризованы через пороговые преобразования, после чего модель представляется в виде системы логических правил. В работе М. Николау и соавторов (2020) анализируется обучаемость булевых функций глубокими нейронными сетями и обсуждаются структурные ограничения таких схем. В ансамблевых методах, например в случайном лесе, каждое дерево может рассматриваться как булева подфункция, а весь ансамбль – как сложная комбинация логических условий, что открывает возможности для последующего упрощения и извлечения более компактных правил. В российской научной литературе логико-правилевые модели и нечёткая логика применяются для поддержки принятия решений в медицине, технической диагностике и образовании, где важна понятность выводов. В публикациях по объяснимому ИИ в медицинской диагностике рассматриваются системы, в которых решения сопровождаются пояснениями на основе логических правил и значимости признаков. В исследованиях по образно-логическим нейронным сетям подчёркивается роль логических рассуждений и интерпретируемых структур в архитектуре ИИ-систем.

Направление объяснимого ИИ и логических моделей активно развивается, особенно с конца 2010-х годов. В статье С. Рудин (2019) обосновывается необходимость перехода от сложных непрозрачных моделей к интерпретируемым, в частности к моделям на основе логических правил, которые можно формально анализировать и верифицировать. В современных работах по выразительным булевым моделям (например, в статье Х. Ибарса и соавторов 2023 года) предлагается расширять классические булевы формулы дополнительными операторами, позволяющими описывать сложные комбинации признаков. Такие подходы позволяют строить интерпретируемые логические модели, по точности сопоставимые с более сложными методами машинного обучения на табличных данных. Открытые программные библиотеки, реализующие методы построения булевых формул ограниченной сложности, показывают, что логические модели могут успешно применяться в финансах, здравоохранении и других областях, где важна прозрачность. Активные исследования направлены на поиск баланса между сложностью формул (и, следовательно, их понятностью) и точностью прогнозов.

Обзорные работы по XAI отмечают значимую роль логико-правилевых моделей, включая булевы функции и их расширения, особенно в здравоохранении и анализе сложных систем. В российском контексте вопросы этики и прозрачности ИИ обсуждаются в специализированных публикациях, где подчёркивается необходимость того, чтобы система могла пояснить свои решения человеку-эксперту.

В образовательной аналитике объяснимый ИИ на основе булевых функций применяется для прогнозирования успеваемости студентов и выявления факторов риска. В качестве исходных данных могут использоваться посещаемость, активность в системах управления обучением, результаты тестирований и другие показатели учебной деятельности. Набор данных UCI Student Performance (П. Кортез и А. Силва, 2008) содержит разнообразные учебные и социально-демографические признаки, что делает его удобным для построения интерпретируемых моделей. Используя алгоритмы типа RIPPER или другие логико-правилевые подходы, можно получить правила вида: «ЕСЛИ посещаемость высокая И активность в курсе выше среднего И средний балл по тестам высок, ТО вероятность успешного завершения курса велика». Подобные правила предоставляют преподавателю и студенту понятные критерии, показывающие, какие аспекты учебной деятельности наиболее важны для успеха. В работах по объяснимому прогнозированию успеваемости (например, в статье Цз. Яна и соавторов 2024 года) используются базы правил и методы учёта неопределённости, демонстрирующие высокую точность классификации на учебных выборках. В исследовании З. Озполата и соавторов (2021) рассматриваются методы XAI для формирования автоматизированной обратной связи студентам и рекомендаций по самоорганизации обучения. Логико-правилевые модели позволяют автоматически выявлять типичные паттерны поведения обучающихся и формировать индивидуальные рекомендации, при этом сами правила остаются понятными для педагогов и администраторов. Российские исследования в области профессионального и медицинского образования подчёркивают важность использования данных об успеваемости и активности студентов для оценки качества подготовки, что создаёт благоприятные условия для внедрения объяснимых аналитических систем. В таких

условиях булевы и нечёткие правила могут служить понятным инструментом для диагностики образовательных результатов и планирования индивидуальных траекторий обучения. Преимущества подобных подходов включают повышение прозрачности и доверия: студенты видят, какие действия улучшат их результаты, а преподаватели получают информацию о том, какие элементы курса нуждаются в корректировке. Ограничения связаны с необходимостью бинаризации или дискретизации признаков, а также с ростом сложности логических формул при увеличении числа факторов, однако нечёткие расширения и методы минимизации булевых функций позволяют частично смягчить эти проблемы.

Несмотря на достоинства, булевы функции имеют ряд ограничений. Основная теоретическая трудность – экспоненциальный рост числа возможных комбинаций входных переменных с увеличением их количества, что осложняет прямое построение и анализ функций для задач большой размерности. Кроме того, точные методы минимизации булевых форм вычислительно трудоёмки, а модели могут быть чувствительны к зашумлённости данных. В контексте объяснимого ИИ эти проблемы частично решаются с помощью приближённых методов построения правил, эвристических алгоритмов минимизации и комбинации логических моделей с другими подходами машинного обучения. Перспективными направлениями считаются применение методов булевой оптимизации и гибридных алгоритмов для поиска компактных формул, а также интеграция логических моделей с графовыми нейронными сетями для анализа структурированных данных. Ожидается, что по мере ужесточения требований к прозрачности и подотчётности ИИ-систем спрос на интерпретируемые решения на основе логических и булевых моделей будет возрастать, в том числе в здравоохранении и образовании. Это стимулирует разработку новых алгоритмов, сочетающих теоретические достижения дискретной математики и практические потребности прикладных областей.

Булевы функции, исторически сложившиеся как фундаментальный объект дискретной математики и алгебры логики, приобретают всё большее значение в области объяснимого искусственного интеллекта. Представление моделей в виде логических формул и правил обеспечивает прозрачность и возможность

формального анализа, что особенно важно для задач с высокими рисками, таких как медицинская диагностика и образовательная аналитика. От классических нормальных форм до современных методов построения логических моделей, ориентированных на XAI, логические представления демонстрируют способность сочетать интерпретируемость с приемлемым качеством предсказаний. В образовании и других прикладных областях это позволяет создавать системы поддержки принятия решений, которые понятны специалистам и поддаются проверке. По мере развития ИИ и усиления требований к объяснимости можно ожидать, что булевы функции и основанные на них модели останутся одним из ключевых подходов к построению интерпретируемых систем машинного обучения. Российская школа дискретной математики и исследования в области объяснимого ИИ создают научную базу для дальнейших теоретических и прикладных разработок в этой сфере.

Список литературы

1. Крама И. Булевы функции: теория, алгоритмы и приложения / И. Крама, П.Л. Хаммер. – Кембридж: Издательство Кембриджского университета, 2011. – 712 с.

2. Рудин К. Прекратите объяснять чёрные ящики моделей машинного обучения для решений с высокими ставками – используйте интерпретируемые модели / К. Рудин // Природа машинного интеллекта. – 2019. – Т. 1. №5. – С. 206–215.

3. Коэн У.У. Быстрый эффективный метод индукции правил / У.У. Коэн // Труды Двенадцатой международной конференции по машинному обучению (ICML‘95). – 1995. – С. 115–123.

4. Ибарз Х. Объяснимый ИИ с использованием выразительных булевых формул / Х. Ибарз // Машинное обучение и извлечение знаний. – 2023. – Т. 5. №4. – С. 1760–1795.

5. Мериани Ф.А. Точные и приближённые методы извлечения правил из нейронных сетей с булевыми признаками / Ф.А. Мериани, Дж.М. Хоу // Труды 11-й Международной объединённой конференции по вычислительному интеллекту. – 2019. – С. 424–433.

6. Николау М. Понимание обучаемости булевых функций глубокими нейронными сетями / М. Николау. – 2020.

7. Ян Цз. Метод прогнозирования успеваемости студентов на основе двухуровневой прогрессивной классификационной базы вероятностных правил / Цз. Ян // Электроника. – 2024. – Т. 13. №22. – С. 4358.

8. Озполат З. Объяснимый ИИ для автоматизированной обратной связи и интеллектуальных рекомендаций по поддержке саморегуляции учащихся / З. Озполат // Рубежи в области искусственного интеллекта. – 2021. – Т. 4. №723447.

9. Гартнер. Прогнозы 2023: подотчётность и управление в сфере ИИ: исслед. отчет. – Gartner, 2023.

10. Квинлан Дж.Р. C4.5: программы для машинного обучения / Дж.Р. Квинлан. – Сан-Франциско: Morgan Kaufmann, 1993. – 302 с.

11. Кортеш П. Применение методов интеллектуального анализа данных для прогнозирования успеваемости учащихся средней школы / П. Кортеш, А. Силва // Труды 5-й Конференции по технологиям будущего в бизнесе. – 2008. – С. 5–12.

12. Каро М.К. Квантовое обучение булевых линейных функций относительно произвольных распределений / М.К. Каро // Обработка квантовой информации. – 2020. – Т. 19. №172.

13. Булевы функции и построение логических схем: учеб. пособие для студентов технических вузов. – М.: МИРЭА.

14. Кочергин В.В. Булевы функции: лекционные материалы по теории дискретных функций / В.В. Кочергин. – М.: МГУ, кафедра дискретной математики.