

МАТЕМАТИЧЕСКИЕ МОДЕЛИ В МАШИННОМ ОБУЧЕНИИ: ОТ РЕГРЕССИОННЫХ МОДЕЛЕЙ К ГЛУБОКИМ НЕЙРОННЫМ СЕТЯМ

Коробова В. М., Скиба И. Г.

Кафедра электронных вычислительных машин,

Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

E-mail: vikakorobova2005@gmail.com, i.skiba@bsuir.by

Машинное обучение прошло значительную эволюцию – от простых регрессионных моделей до современных глубоких нейронных сетей. Первые методы сыграли ключевую роль в формировании основ области. Со временем к ним добавились сложные подходы, которые позволили решать задачи распознавания образов, обработки речи и прогнозирования с высокой точностью. Прогресс в области машинного обучения создал новые возможности для автоматизации, что уже оказывает значительное влияние на самые разные отрасли.

ВВЕДЕНИЕ

В 1980-х годах история машинного обучения была бы сосредоточена на экспертных системах, логическом программировании и эвристическом поиске[1]. Эти подходы отражали ранние представления об искусственном интеллекте (ИИ), заложенные на конференции в Дартмуте в 1956 году, когда Джон Маккарти ввел термин «ИИ». Машины использовались для узких задач, таких как шахматы, а идеи о более сложном обучении только зарождались[1].

К началу 2000-х акцент сместился на методы опорных векторов, байесовский вывод и деревья решений. Эти алгоритмы решали задачи классификации и прогнозирования, доминируя в приложениях.

Сегодня, в 2020-х, ключевое внимание уделяется глубоким нейронным сетям (НС), особенно трансформерам. Они, разработанные для обработки естественного языка, нашли применение в распознавании изображений, управлении автономными системами и генеративных моделях. НС, вдохновленные человеческим мозгом, позволили машинам обучаться и адаптироваться, изменяя связи между "нейронами" на основе опыта.

I. ЛИНЕЙНАЯ РЕГРЕССИЯ

В 1805 году Адриен-Мари Лежандр представил то, что сегодня известно как линейная регрессия. Позже Иоганн Карл Фридрих Гаусс также был признан за свои ранние работы, выполненные около 1795 года[1]. Линейная регрессия, существующая более двух веков, состоит из входного слоя с несколькими узлами и выходного слоя, состоящего из одного узла. Каждый входной узел хранит действительное число и соединен с выходом через вес, представляющий собой действительное число.

На тот момент этот метод не назывался линейной регрессией[2]. Вместо этого использовался метод наименьших квадратов, который является математически идентичным современным подхо-

дам. Эти простые модели обеспечивают «поверхностное обучение», в отличие от более сложных моделей глубокого обучения.

Модель линейной регрессии описывается формулой, где предсказанное значение y вычисляется как сумма произведений входных значений и их соответствующих весов:

$$y = w_1x_1 + w_2x_2 + \dots + w_nx_n + b. \quad (1)$$

Градиентный спуск – один из ключевых методов оптимизации в машинном обучении, позволяющий обновлять веса и смещение модели для минимизации ошибки[2]. Процесс включает вычисление производной функции потерь и ее использование для корректировки весов:

Этот метод стал активно использоваться в 1950-х годах для решения задач минимизации. В 1980-х годах градиентный спуск стал основным инструментом для обучения нейронных сетей. Он продемонстрировал свою эффективность, особенно после публикации алгоритма обратного распространения ошибки в 1986 году.

$$w_j := w_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\mathbf{w}}(x^{(i)}) - y^{(i)}) x_j^{(i)}, \quad (2)$$

где α – скорость обучения. Этот метод стал основным инструментом для обучения нейронных сетей, продемонстрировав свою эффективность в различных областях, от компьютерного зрения до обработки естественного языка.

II. КЛАССИЧЕСКОЕ МАШИННОЕ ОБУЧЕНИЕ

Классическое машинное обучение активно развивалось после линейной регрессии, что привело к созданию ряда новых алгоритмов и методов, расширяющих возможности анализа данных.

В 1970-х годах появился метод деревьев решений, который произвел революцию в области анализа данных. Алгоритм ID3, предложенный

Россом Квинланом в 1986 году, позволял строить деревья решений, которые могут использоваться для принятия решений на основе наборов данных[3]. Деревья решений упрощают интерпретацию результатов, отходя от линейных аппроксимаций.

В 1995 году был представлен алгоритм опорных векторов (SVM), разработанный Владом Сапорву и его коллегами[4]. SVM использует гиперплоскости для разделения классов[4] и стал универсальным инструментом для классификации и регрессии, особенно в высокоразмерных данных. Его способность работать с нелинейными данными благодаря ядровым функциям делает его особенно эффективным.

Случайный лес, предложенный в 1996 году Джереми Куном, использует ансамблевый подход, строя множество деревьев решений и комбинируя их результаты. Этот метод оказался особенно полезным для повышения точности и устойчивости к переобучению[5].

На протяжении 2000-х годов распространение получили вероятностные модели: наивный байесовский классификатор, который делает предположение о независимости признаков и эффективно используется для классификации текстов и спам-фильтрации. Методы ансамблирования, такие как AdaBoost, также стали популярными, объединяя слабые классификаторы для улучшения производительности.

III. ГЛУБОКОЕ ОБУЧЕНИЕ

История глубокого обучения начинается с первых моделей, имитирующих работу человеческого мозга[6]. В 1943 году Уоррен Маккаллок и Уолтер Питтс предложили первую математическую модель нейрона, положившую начало исследованиям нейронных сетей. В 1958 году Фрэнк Розенблатт представил перцептрон — простейшую нейронную сеть для бинарной классификации, которая продемонстрировала возможность обучения на данных.

После периода затишья в 1960-х и 1970-х годах в 1986 году Джеффри Хинтон и его коллеги возродили интерес к нейронным сетям с помощью алгоритма обратного распространения ошибки, который позволил эффективно обучать многослойные нейронные сети[1].

С начала 2000-х годов, с ростом вычислительных мощностей и объемов данных, глубокое обучение снова стало популярным. В 2012 году команда Алекс Кризевского представила архитектуру AlexNet на конкурсе ImageNet, продемонстрировав, как сверточные нейронные сети (CNN) могут достигать выдающихся результатов в распознавании изображений.

Глубокое обучение начало активно применяться в обработке естественного языка (NLP). В 2013 году была представлена модель Word2Vec, использующая нейронные сети для векторизации слов.

В 2014 году архитектура VGG показала, что увеличение глубины сети и использование малых сверток улучшают производительность, а в 2015 году ResNet ввела остаточные связи, позволяя создавать очень глубокие сети и избегая проблем с исчезающими градиентами.

Ключевым моментом стало появление архитектуры трансформеров в 2017 году, предложенной Васвани и др. Эта архитектура, основанная на механизме внимания, кардинально изменила подходы к NLP и привела к разработке таких моделей, как BERT и GPT, которые продемонстрировали выдающиеся результаты в ряде задач[1].

ЗАКЛЮЧЕНИЕ

Машинное обучение, начиная с простых регрессионных моделей и заканчивая глубокими нейронными сетями, претерпело значительные изменения. Эти изменения открывают новые возможности в различных отраслях и создают фундамент для дальнейших исследований и разработок. Трансформеры продолжают расширять горизонты машинного обучения, меняя наш подход к решению сложных задач и повышая эффективность в различных приложениях. В дальнейшем стоит ожидать появления новых архитектур и методов, которые еще больше ускорят развитие этой области.

IV. СПИСОК ЛИТЕРАТУРЫ

1. J. Schmidhuber, Annotated history of modern AI and deep learning / J. Schmidhuber // arXiv:2212.11279. 2022. 75 p.
2. article4 J. H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, The Annals of Statistics*, vol. 29, no. 5, pp. 1189-1232, 2001.
3. article5 R. Quinlan, Induction of Decision Trees, Machine Learning, vol. 1, no. 1, pp. 81-106, 1986.
4. article6 V. Vapnik, The Nature of Statistical Learning Theory, Springer, 1995.
5. article7 L. Breiman, Random Forests, Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
6. article8 Скиба, И. Г. Эффективное использование нейронных сетей в управлении персоналом для повышения производительности = Effective use of neural networks in personnel management to improve productivity / И. Г. Скиба, С. Н. Нестеренков, Д. А. Жалейко // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сборник научных статей X Международной научно-практической конференции, Минск, 13 марта 2024 г. : в 2 ч. Ч. 1 / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В. А. Богущ [и др.]. – Минск, 2024. – С. 363–367.