

- и информатизация образования» DOI 10.25688/2072-9014.2020.51.1.06, Стр. 46-5
- [2] Пучков, Н. П. Интеграция компетенций в условиях цифровизации образования / Н. П. Пучков, А. И. Попов, С. И. Тормасин // Continuum. Математика. Информатика. Образование. – 2020. – № 1 (17). – С. 36 – 43. doi: 10.24888/2500-. 1957-2020-17-1-36-43.
- [3] Математика и математическое образование в эпоху цифровизации: материалы XII Всероссийской с международным участием научно-методической конференции. Красноярск, 9–10 ноября 2023 г. [Электронный ресурс]– Красноярск, 2023. -301 стр.
- [4] Шатрова, Ю. С. Возможности и угрозы при организации учебного процесса в цифровом обществе// Математическое образование в цифровом обществе: Материалы XXXVIII Междунар. науч. семинара преподавателей математики и информатики университетов и педагогических вузов, 26 – 28 сентября 2019 г., – Самара, 2019. – С. 238 – 241.
- [5] Пучков, Н. П. Информационные модели обучения в процессе организации математического образования / Н. П. Пучков, Т. Ю. Забавникова, Т. Ю. Дорохова // Фундаментальные проблемы обучения математике, информатике и информатизации образования: Материалы Междунар. конф., 25 – 27 сентября 2020 г., -Елец, 2020. – С. 148 – 150.

ГЛУБОКОЕ ОБУЧЕНИЕ И НЕЙРОННЫЕ СЕТИ

Горбунова Надежда Александровна¹, Фрицлер Никита Юрьевич², Пардабеков Азим Маратович³

^{1, 2, 3}Карагандинский университет им.Е.А. Букетова, Караганда, Казахстан

¹E-mail: ant_nadezhda@mail.ru

²E-mail: 87003863647@mail.ru

³E-mail: soulmaster09kz@gmail.com

Глубокое обучение представляет собой одно из самых значимых направлений современной информатики, изменившее подход к решению широкого круга задач: от распознавания образов до понимания естественного языка. В его основе лежат многослойные искусственные нейронные сети, способные автоматически извлекать сложные признаки из неструктурированных данных. Если классические методы машинного обучения требовали от исследователя тщательной инженерии признаков и ручного выбора алгоритмов, то глубокие архитектуры берут на себя большую часть этой работы, позволяя моделям «учиться» непосредственно из сырых входных данных.

Первые идеи об искусственных нейронных сетях появились ещё в середине XX века, однако настоящий бум исследований начался лишь в XXI веке, когда вместе с развитием аппаратного ускорения (GPU, TPU) и появлением массивных наборов данных стало возможным тренировать сети глубиной в десятки и сотни слоёв. Сегодня глубокое обучение прочно вошло во многие индустрии: медицина, автомобильная промышленность, финансовый сектор и развлекательные сервисы. В этой статье мы подробно рассмотрим, как устроены такие модели, какие архитектуры наиболее популярны и чем они отличаются друг от друга, а также какие методы применяются для их эффективного обучения и внедрения в реальные системы.

В сердце любой нейронной сети лежит искусственный нейрон – математическая конструкция, повторяющая основные принципы работы биологического аналога. На вход нейрон получает набор числовых значений, каждое из которых умножается на свой вес, суммируется с добавлением смещения и затем преобразуется через нелинейную функцию активации. Именно функция активации отвечает за способность сети моделировать сложные, не линейно-разделимые зависимости, без которых обучение глубоких моделей было бы невозможным.

Архитектурной базой большинства сетей является последовательное соединение таких нейронов в слои. Первый слой принимает на входе данные в сыром виде, скрытые слои постепенно трансформируют информацию, выявляя всё более абстрактные признаки, а последний слой выдаёт предсказание в нужном формате – будь то распределение вероятностей классов при классификации или числовое значение при регрессии. Процесс передачи сигнала от входа к выходу называется прямым распространением, а обратное распространение ошибки позволяет корректировать веса сети, минимизируя выбранную функцию потерь.

Свёрточные нейронные сети (CNN) стали настоящей революцией в задачах обработки изображений. Их ключевым отличием от «обычных» полносвязных моделей является использование операций свёртки. Фильтры небольшого размера сканируют входную картину, улавливая локальные паттерны – края, текстуры, простые геометрические формы. По мере углубления в архитектуру фильтры обнаруживают всё более высокоуровневые признаки: от контуров отдельных объектов до целых сцен.

Дополняющие слои подвыборки (например, max-pooling) постепенно снижают пространственные размеры карт признаков, сохраняя лишь наиболее информативные отклики и тем самым уменьшая объём вычислений. Такой подход не только повышает устойчивость к смещениям и деформациям изображений, но и позволяет эффективно обучать глубинные сети, в которых десятки и сотни свёрточных слоёв последовательно выделяют признаки разной сложности. Примеры успешных CNN-архитектур включают классические модели типа AlexNet, ResNet и современные лёгкие сети для мобильных устройств, где важны как точность, так и скорость вывода.

В задачах, где данные обладают последовательной структурой – например, при работе с текстом, временными рядами или аудиосигналами – высокую эффективность демонстрируют рекуррентные нейронные сети (RNN). Главная идея заключается в сохранении «состояния» (hidden state), которое аккумулирует информацию о предыдущих элементах последовательности. Это позволяет модели учитывать контекст при обработке текущего шага и формировать предсказание, ориентируясь на исторические данные.

Однако при попытке обучить классические RNN на длинных последовательностях исследователи столкнулись с проблемой исчезающих или взрывающихся градиентов, что делало сложным захват долгосрочных зависимостей. Для преодоления этого ограничения были разработаны специальные механизмы, такие как LSTM (Long Short-Term Memory) и GRU (Gated Recurrent Unit). Они вводят дополнительные «гейты», которые динамически контролируют процесс запоминания и забывания информации, существенно улучшая способность сети сохранять и извлекать важные факты из далёкого прошлого.

В 2017 году в научной работе «Attention Is All You Need» была предложена архитектура трансформера, которая в очередной раз изменила ландшафт глубокого обучения, особенно в области обработки естественного языка. В отличие от RNN, трансформер полностью

отказывается от рекуррентности, полагаясь на механизм внимания (attention), который вычисляет, какие части входной последовательности наиболее релевантны для получения представления конкретного элемента.

Self-attention позволяет каждому слову «смотреть» на все остальные, что обеспечивает эффективное моделирование как локальных, так и глобальных зависимостей. Многоголовое внимание расширяет эту идею, обучая несколько независимых attention-процессов, каждый из которых фокусируется на разных аспектах текста. Кодировщик–декодировщик трансформера превращает входной текст в скрытые векторы, а затем на их основе генерирует выходную последовательность, будь то перевод, аннотация или ответ в диалоге. Именно трансформеры лежат в основе моделей BERT и GPT, продемонстрировавших выдающиеся результаты в широком спектре NLP-задач.

Эффективная тренировка глубоких сетей требует не только мощного железа, но и грамотного выбора функций потерь, оптимизаторов и приёмов регуляризации. Для классификации обычно используется кросс-энтропия, а для регрессии – среднеквадратичная ошибка. Классический стохастический градиентный спуск (SGD) обновляет веса по направлению антиградиента, однако адаптивные методы, такие как Adam, учитывают моменты градиентов и позволяют ускорять сходимость при более гибком подборе скорости обучения.

С параллельным ростом глубины сетей усилились и риски переобучения. Dropout, нормализация батчей (batch normalization) и искусственное расширение данных (data augmentation) становятся не просто дополнительными опциями, а необходимыми этапами в процессе обучения. Dropout случайным образом «отключает» часть нейронов во время каждой итерации, вынуждая модель искать более общие представления, а батчнормализованные слои стабилизируют распределение активаций, что ускоряет и стабилизирует процесс оптимизации.

Для реализации глубоких моделей сегодня доступны мощные фреймворки: TensorFlow с высокоуровневым API Keras и гибкая библиотека PyTorch, предлагающая динамическое построение графа вычислений. Они интегрированы с CUDA и другими системами ускорения на GPU и TPU, что позволяет существенно сокращать время обучения. Распределённое обучение с использованием библиотек Horovod или встроенных средств фреймворков даёт возможность масштабировать эксперименты на больших кластерах.

Переход от исследовательского прототипа к промышленной эксплуатации моделей (MLOps) включает контроль версий данных и артефактов моделей, настройку конвейеров CI/CD и мониторинг качества предсказаний в реальном времени. Инструменты вроде MLflow, Kubeflow и DVC упрощают эти задачи, обеспечивая воспроизводимость экспериментов и ускоряя выход продукта на рынок.

Глубокие нейронные сети нашли применение в самых разных сферах: от автономного вождения, где модели распознают дорожные знаки и пешеходов, до медицины, где на основе снимков МРТ строятся прогнозы развития заболеваний. В рекомендационных системах стриминговых и e-commerce сервисов глубокие подходы анализируют поведение пользователей и предлагают наиболее релевантный контент. В финансовой отрасли сети обучаются предсказывать ценовые тренды и выявлять аномалии в транзакциях.

Сегодня перед сообществом стоят новые вызовы: как повысить интерпретируемость «чёрных ящиков» глубоких моделей, как сократить их энергетические затраты и расширить способность работать с ограниченными данными (few-shot и zero-shot learning).

Развитие лёгких архитектур, квантование и исследования в области искусственного общего интеллекта (AGI) определяют траекторию дальнейших исследований и практических внедрений.

Глубокое обучение и нейронные сети за последние годы проложили путь к решению задач, которые ранее казались недостижимыми. Их сила кроется в способности автоматически извлекать сложные признаки из разнообразных данных при помощи последовательных преобразований. Понимание фундаментальных принципов работы нейронных элементов, механизмов обучения и современных архитектур позволяет исследователю и практику осуществлять грамотный выбор моделей и инструментов. В будущем развитие таких направлений, как интерпретируемость, энергоэффективность и обучение при малом количестве данных, будет определять, насколько глубоко и широко эти технологии проникнут в новые отрасли. Для дальнейшего углубления рекомендуются классические монографии и текущие публикации ведущих конференций NeurIPS, ICML и CVPR.

Список литературы

- [1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
- [2] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the ACL*, 4171–4186
- [3] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56), 1929–1958

DEEP LEARNING: ТЕРЕҢ НЕЙРОНДЫҚ ЖЕЛІЛЕРДІҢ МҮМКІНДІКТЕРІ МЕН ТӘУЕКЕЛДЕРІ

Датқабаета Мөлдір Асқарқызы¹, Дәкәрім Гүлдерай Қанатқызы², Елеусіз Маржан Ержанқызы³, Хасанова Айдана Болатқызы⁴

^{1,2,3,4}Е.А. Бөкетов атындағы Қарағанды университеті, Қарағанды қ., Қазақстан

Аннотация.

Бұл мақалада заманауи жасанды интеллекттің негізгі бағыты болып саналатын Deep Learning технологиясы жан-жақты талданады. Терең нейрондық желілердің теориялық негіздері, құрылымы және оқыту тәсілдері түсіндіріліп, олардың компьютерлік көру, табиғи тілді өңдеу, медицина, көлік және мазмұн генерациясы секілді түрлі салалардағы қолдану мүмкіндіктері көрсетіледі. Сонымен қатар, Deep Learning жүйелерінің шектеулері мен қоғамға тигізетін ықтимал тәуекелдері – атап айтқанда, «қара жәшік» проблемасы, қате шешімдер, deepfake технологиясы, әлеуметтік әділетсіздік және экологиялық салдарлар