

УДК 004.8

*Кирилова Є. В., магістрант,
Шушура О. М., д.т.н.,
Соломаха С. А., к.е.н.*

Державний університет інформаційно-комунікаційних технологій

АНАЛІТИЧНИЙ ОГЛЯД СУЧАСНИХ АРХІТЕКТУР DNN, CNN, RNN ТА TRANSFORMER У ЗАДАЧАХ РЕГРЕСІЇ ТА КЛАСИФІКАЦІЇ

Зростання складності реальних даних, їх багатовимірність та нелінійність зумовлюють потребу у моделях, здатних забезпечувати високу точність апроксимації залежностей у задачах класифікації та регресії. Сучасні архітектури глибоких нейронних мереж (Deep Neural Networks, DNN) демонструють значні переваги у моделюванні складних функцій та роботі з великими наборами даних. Особливої актуальності набувають такі архітектури, як згорткові нейронні мережі (Convolutional Neural Networks, CNN), рекурентні мережі (Recurrent Neural Networks, RNN), мережі LSTM, а також моделі на основі механізму самоуваги (Transformers) [1]. Необхідність визначення ефективності цих архітектур у регресійних і класифікаційних задачах різних типів мотивує виконання комплексного аналітичного огляду.

Метою дослідження є порівняльний аналіз сучасних архітектур нейронних мереж, визначення їх внутрішньої структури, обчислювальних властивостей, здатності до узагальнення, ефективності у задачах з різними типами даних, а також виокремлення перспектив розвитку та застосування.

Глибокі нейронні мережі (DNN) — це багатошарові моделі прямого поширення сигналу, що містять кілька прихованих шарів з нелінійними функціями активації та здатні апроксимувати складні функції завдяки властивості універсального наближення.

Згорткові нейронні мережі (CNN) — архітектури, створені для обробки просторово структурованих даних. Основою CNN є згорткові шари, що виконують фільтрацію з використанням спільних ваг, що значно зменшує кількість параметрів та підвищує стійкість до варіацій у вхідних даних. CNN забезпечують автоматичне виділення ознак різного рівня, що дозволяє ефективно розв'язувати задачі комп'ютерного зору та обробки сигналів [2].

Рекурентні нейронні мережі (RNN) моделюють залежності у послідовних даних за допомогою внутрішнього стану, який оновлюється покроково. Основним недоліком класичних RNN є

проблема зникання градієнтів, що ускладнює їх застосування для довгих послідовностей.

LSTM (Long Short-Term Memory) покращують базову архітектуру RNN завдяки гейтовим механізмам, що дозволяють зберігати релевантну інформацію протягом тривалих інтервалів. Це забезпечує високу точність у задачах прогнозування часових рядів, обробки текстів та визначення трендів [3].

Трансформерні архітектури (Transformers) замінили рекурентність механізмом самоуваги (self-attention). Структура Encoder–Decoder дозволяє моделювати глобальні залежності між елементами послідовності та паралелізувати обчислення. Механізм Multi-Head Self-Attention забезпечує аналіз даних у кількох підпросторах ознак, що робить трансформери ефективними у задачах класифікації й регресії, особливо на великих масивах даних.

Сучасні модифікації, такі як ResNet, дозволяють стабілізувати навчання глибоких CNN, тоді як оптимізатори AdamW покращують збіжність DNN. Використання Batch Normalization та Dropout знижує ризик перенавчання. У регресійних задачах найвищу точність демонструють трансформери, а у класифікаційних задачах зображень — CNN. У задачах з вираженими часовими залежностями LSTM зберігають значну перевагу.

Аналіз показує, що кожна архітектура нейронних мереж має власні оптимальні області застосування. CNN переважно використовуються для просторових даних, LSTM — для послідовностей, DNN — для табличних даних, а Transformers — для задач з глобальними залежностями та великими обсягами інформації. Перспективи подальших досліджень пов'язані з оптимізацією обчислювальної складності трансформерних моделей, створенням гібридних архітектур CNN–Transformer та впровадженням методів pruning, quantization і distillation для зменшення ресурсних витрат.

Список використаних джерел:

1. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016. 800 p..
2. LeCun Y., Bengio Y., Hinton G. Deep learning. Nature. 2015. Vol. 521, No. 7553. P. 436–444.
3. Hochreiter S., Schmidhuber J. Long short-term memory. Neural Computation. 1997. Vol. 9, No. 8. P. 1735–1780.