

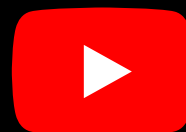
Модели вида “seq2seq”

[Ask a question](#)

Transformer models: Encoder-Decoders



Копирова...



Посмотреть на YouTube

Модели типа кодировщик-декодировщик (также называемые *sequence-to-sequence models*) используют обе части трансформера. На каждом этапе слой внимания кодировщика получает доступ ко всем словам в исходной последовательности, тогда как слой внимания декодировщика получает доступ только к тем словам, которые расположены до текущего слова.

Предобучение таких моделей может быть выполнено на задачах, используемых для предобучения моделей кодировщиков или декодировщиков, но обычно все немного сложнее. Например, модель T5 была предобучена путем замены случайных фрагментов текста (фрагменты могут содержать несколько слов) на специальную маску, цель модели - предсказать текст, который заменила маска.

Модели seq2seq лучше всего подходят для задач генерации новых предложений, зависящих от входного массива данных, например: автоматическое реферирование текста, перевод или в генеративных вопросно-ответных системах.

Представителями этого семейства являются:

- BART
- mBART
- Marian
- T5