

Модели-кодировщики

<https://huggingface.co/learn/nlp-course/ru/chapter1/5?fw=pt>

Кодировщики используют только компонент кодировщика трансформера. На каждом этапе слой внимания может использовать все слова исходного предложения.

Эти модели часто характеризуют как имеющие двунаправленное внимание (англ. bi-directional attention), и часто называют моделями автокодировщиками.

Предварительное обучение этих моделей обычно заключается в том, чтобы как-то исказить предложение (например, путем маскировки в нем случайных слов) и поставить перед моделью задачу найти или восстановить исходное предложение.

Кодировщики лучше всего подходят для задач, требующих *понимания* всего предложения, таких как классификация предложений, распознавание именованных сущностей (и, в более общем смысле, классификация слов) и ответы на вопросы с извлечением информации из контекста (выделительные вопросно-ответные системы).

К представителям этого семейства моделей относятся:

- ALBERT
- BERT
- DistilBERT
- ELECTRA
- RoBERTa

<https://huggingface.co/learn/nlp-course/ru/chapter1/6?fw=pt>

Модели-декодировщики

Декодировщики используют только компонент декодирования трансформера.

На каждом этапе для текущего слова слой внимания может получить доступ только к словам, которые были расположены до него в предложении.

Такие модели часто называются *авторегрессионными моделями*.

Процесс предобучения декодировщиков обычно заключается в предсказании следующего слова в предложении.

2

Такие модели лучше всего подходят для задач, связанных с генерацией текста.

Представителями этого семейства моделей являются:

- CTRL
- GPT
- GPT-2
- Transformer XL

Модели вида "seq2seq"

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

Предвзятости и ограничения

Если вы намерены использовать предварительно обученную модель или ее дообученную версию в рабочей среде, имейте в виду, что, хотя эти модели являются мощными инструментами, они имеют ограничения.

Самое большое из них заключается в том, что для предварительного обучения на больших объемах данных исследователи часто очищают весь контент, который они могут найти, беря как лучшее, так и худшее из того, что доступно в Интернете.

Для иллюстрации вернемся к примеру пайплайна fill-mask с моделью BERT:

```
from transformers import pipeline

unmasker = pipeline("fill-mask", model="bert-base-uncased")

result = unmasker("This man works as a [MASK].")
print([r["token_str"] for r in result])

result = unmasker("This woman works as a [MASK].")
print([r["token_str"] for r in result])
```

```
Device set to use cpu
['carpenter', 'lawyer', 'farmer', 'businessman',
'doctor']
['nurse', 'maid', 'teacher', 'waitress', 'prostitute']
```

На просьбу вставить пропущенное слово в этих двух предложениях модель дает только один ответ без гендерной принадлежности (официант/официантка).

Другие рабочие профессии обычно ассоциируются с одним конкретным полом — и да, проститутка попала в топ-5 вариантов, которые модель ассоциирует с “женщиной” и “работой”.

Это происходит даже несмотря на то, что BERT — одна из редких моделей трансформеров, созданная не путем сбора данных со всего Интернета, а с использованием явно нейтральных данных (он обучен на английской Википедии и наборе данных BookCorpus).

Поэтому, когда вы используете эти инструменты, вам нужно помнить, что исходная модель, которую вы используете, может очень легко генерировать сексистский, расистский или гомофобный контент. **Дообучение модели на ваших данных не сможет устранить эту внутреннюю предвзятость.**

Итоги

Модель	Примеры	Задачи
Кодировщик	ALBERT, BERT, DistilBERT, ELECTRA, RoBERTa	Классификация предложений, распознавание именованных сущностей, выделительные вопросно-ответные системы
Декодировщик	CTRL, GPT, GPT-2, Transformer XL	Генерация текста
Кодировщик- декодировщик	BART, T5, Marian, mBART	Автоматическое реферирование, перевод, генеративные вопросно-ответные системы