

## Вопрос 1

Выполнен

Баллов: 1,00 из 1,00

Зачем нужен механизм внимания в задачах Seq2Seq?

Выберите все варианты (один или несколько), соответствующие применениям механизма внимания в Seq2Seq.

Выберите один или несколько ответов:

- ☒ a. Механизм внимания позволяет учитывать далёкие зависимости.
- ☐ b. Механизм внимания обязательно уменьшает количество параметров в нейросети.
- ☒ c. Механизм внимания усиливает поток информации между энкодером и декодером (больше информации о входном тексте доходит до декодера и генерируемый текст сильнее зависит от входного текста).
- ☒ d. Механизм внимания позволяет избежать "забывания" начала входного текста при использовании рекуррентного энкодера.

Ваш ответ верный.

## Вопрос 2

Выполнен

Баллов: 1,00 из 1,00

Почему имеет смысл декодировать выходы seq2seq модели с помощью лучевого поиска?

Выберите один или несколько вариантов.

Выберите один или несколько ответов:

- ☐ a. Лучевой поиск всегда находит оптимальное решение задачи дискретной оптимизации  $\arg \max_{y_1, \dots, y_m} P(y_1, \dots, y_m)$
- ☒ b. Перебор, пусть и ограниченный, позволяет частично исправлять ошибки (когда модель считает неуместный токен наиболее вероятным)
- ☒ c. Перебор позволяет найти несколько правдоподобных вариантов выходной последовательности.
- ☐ d. Жадный алгоритм гораздо медленнее лучевого поиска

Ваш ответ верный.

## Вопрос 3

Выполнен

Баллов: 1,00 из 1,00

Допустим, мы генерируем последовательности токенов с помощью модели и лучевого поиска.

В какой-то момент в списке частичных решений *Beam* появилась последовательность "a b c d" со следующей оценкой правдоподобия

$$P(abcd) = P(a)P(b|a)P(c|ab)P(d|abc) = 0.3 \cdot 0.5 \cdot 0.7 \cdot 0.2 = 0.021$$

Выберите другие частичные решения, которые могут привести к вытеснению этой последовательности из списка частичных решений.

Выберите один или несколько ответов:

- ☒ a.  $P(zxywv) = P(z)P(x|z)P(y|zx)P(w|zxy)P(v|zxyw) = 0.4 \cdot 0.5 \cdot 0.7 \cdot 0.3 \cdot 0.7 = 0.029$
- ☒ b.  $P(dg) = P(d)P(g|d) = 0.2 \cdot 0.2 = 0.04$
- ☐ c.  $P(abcdf) = P(a)P(b|a)P(c|ab)P(d|abc)P(f|abcd) = 0.3 \cdot 0.5 \cdot 0.7 \cdot 0.2 \cdot 0.99 = 0.0208$
- ☒ d.  $P(ae) = P(a)P(e|a) = 0.3 \cdot 0.1 = 0.03$

Ваш ответ верный.

В этом видео мы поговорили о задаче преобразования последовательностей. В англоязычной литературе такие задачи обычно называют термином [seq2seq](#). Это очень популярная задача, которая, в первую очередь, находит применение в машинном переводе в, диалоговых системах, в суммаризации (упрощённом пересказе) и многих других. В общем виде, архитектура состоит из [энкодера](#) и декодера. Однако, чтобы информация от энкодера лучше передавалась в декодер и вся схема лучше обучалась, часто между энкодером и декодером добавляют механизм внимания. В разное время были предложены разные архитектуры для энкодера и декодера, основанные на [рекуррентных нейросетях](#), на свёрточных, а также на механизме [внутреннего внимания](#) (то есть, [трансформер](#)). Для декодирования последовательностей часто используется алгоритм [лучевого поиска](#), который позволяет найти баланс между полностью жадным решением и полным перебором. Для того, чтобы модель генерировала разнообразные ответы, необходимо ограничивать её уверенность. Для этого используются разнообразные регуляризаторы или альтернативные функции потерь.

**Вопрос 4**

Выполнен

Баллов: 1,00 из 1,00

Выберите характерные проблемы, с которыми приходится бороться именно при разработке seq2seq моделей.

Выберите один или несколько ответов:

- ☒ a. Проблема низкого разнообразия генерируемых последовательностей
- ☒ b. Учет всего контекста в энкодере
- ☐ c. Проблема выбора гиперпараметров обучения
- ☐ d. Проблема нехватки печенок
- ☒ e. Слабая связь энкодера и декодера

Ваш ответ верный.