

Вопрос 1

Выполнен

Баллов: 1,00 из 1,00

Есть ли смысл дополнительно подавать в BERT последовательность токенов в обратном порядке?

Количество данных для обучения при этом увеличится в два раза (каждую последовательность будем подавать в BERT дважды: в прямом и обратном порядке).

Выберите один ответ:

- ☒ a. Нет, BERT маскирует часть токенов во входной последовательности и пытается их предсказать используя контекст, порядок (прямой или обратный) никакой роли не сыграет
- ☐ b. Да, чем больше данных, тем лучше
- ☐ c. Да, BERT сможет выделить дополнительные взаимосвязи, научившись предсказывать не только следующий токен, но и предыдущий
- ☐ d. На качество работы модели это повлияет в лучшую сторону, но это не стоит вычислительных ресурсов из-за увеличения тренировочных данных вдвое

Ваш ответ верный.

Вопрос 2

Выполнен

Баллов: 1,00 из 1,00

Выберите правильные утверждения:

Выберите один или несколько ответов:

- ☒ a. Языковое моделирование часто используют для предобучения моделей в силу наличия большого количества данных (любые тексты на нужном языке) и отсутствия необходимости в разметке
- ☒ b. Маскированная языковая модель обычно обучается в несколько раз дольше, чем стандартная однонаправленная языковая модель (предсказание следующего токена по предыдущим), т.к. маскирование только части токенов дает менее явный сигнал
- ☒ c. Одно из основных преимуществ предобученных моделей состоит в том, что мы можем использовать значительно меньше размеченных данных для дообучения, чем при тренировке модели с нуля

Ваш ответ верный.

Вопрос Инфо

Теперь давайте немного отойдём от [self-attention](#) и вспомним, как работает [LSTM](#). Именно двунаправленный LSTM и лежит в основе популярной модели, которая выдаёт хорошие контекстуализированные [эмбединги](#). Эта модель называется ELMo.^[1] [ELMo](#) расшифровывается как "Embeddings from Language Models". Модель смотрит на всё предложение, прежде чем присвоить слову какое-то [векторное представление](#). ELMo, точно так же, как и ранее рассмотренные модели, учиться решать [задачу языкового моделирования](#), но уже не с помощью архитектуры на основе [трансформера](#), а с помощью LSTM (а точнее, даже двух LSTM, которые смотрят в разные стороны). Давайте чуть подробнее про это. Одна LSTM смотрит вперёд и учиться предсказывать следующее слово при наличии контекста, то есть, всех слов, которые расположены левее того, которое нужно предсказать. А вторая LSTM смотрит назад и предсказывает предыдущее слово, зная все слова, стоящие правее от текущего слова. Такая LSTM называется двунаправленной (или, по-английски, bi-directional). Итоговое векторное представление слова получается путём конкатенации скрытых состояний из обеих частей LSTM. Можно конкатенировать эти скрытые состояния разными способами. Например, вектора скрытых состояний можно домножить на веса и, затем, [сконкатенировать](#) (или даже суммировать) в один вектор. Выбор варианта комбинации скрытых слоёв двух частей LSTM, скорее, относится к инженерным задачам, некоторым инженерным хакам, подбору эвристик. Так что сейчас мы не будем останавливаться на этом подробно.

[1] Peters M. E. et al. Deep contextualized word representations // [arXiv preprint arXiv:1802.05365](#). – 2018.

Вопрос 3

Выполнен

Баллов: 1,00 из 1,00

Сопоставьте названия нейросетей и задачи, которые используются для их предобучения.

BERT	"Маскированное" языковое моделирование и предсказание, является ли второе предложение логическим продолжением предыдущего
ELMO	Двунаправленное языковое моделирование: предсказание следующего и предыдущего токена по соответствующему контексту (токены ле
OpenAI Transformer	Однонаправленное языковое моделирование

Ваш ответ верный.

лучше за счёт решения проблемы фрагментации контекста. Той самой проблемы, когда наше предложение может разбиться на два куска, и эти кусочки отнесутся к двум разным фрагментам. Но является ли Transformer-XL наилучшей моделью на текущий момент для решения задач обработки текстов? Как оказывается, нет. State of the art (sota), на текущий момент, является сетка под названием [GPT-2](#).

[1] Mikolov T. et al. [Distributed representations of words and phrases and their compositionality](#). //Advances in neural information processing systems. – 2013. – С. 3111-3119.

[2] [Transformer-XL: Unleashing the Potential of Attention Models](#), January 29, 2019, Zhilin Yang and Quoc Le, Google AI

[3] Dai Z. et al. [Transformer-xl: Attentive language models beyond a fixed-length context](#) //arXiv preprint arXiv:1901.02860. – 2019.

[4] Young T. et al. [Recent trends in deep learning based natural language processing](#) //ieee Computational intelligence magazine. – 2018. – Т. 13. – №. 3. – С. 55-75.

[5] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. [Neural machine translation by jointly learning to align and translate](#). arXiv preprint arXiv:1409.0473 (2014).

Вопрос 4

Выполнен

Баллов: 1,00 из 1,00

Какие две проблемы есть у стандартного Трансформера, которые можно решить с помощью Transformer-XL?

Выберите один или несколько ответов:

- ☒ а. Работа с контекстом только фиксированной длины
- ☐ б. Отсутствие учета взаимного расположения токенов (не используется positional encoding)
- ☒ в. Фрагментация контекста без учета границ предложений (сильное влияние на короткие предложения)
- ☐ г. Отсутствие возможности распараллелить расчеты внутри self-attention модуля

Ваш ответ верный.

На этом мы заканчиваем обзор популярных в последнее время нейросетей и заканчиваем разговор про [transfer learning](#) в сфере обработки текстов. Мы обсудили пять популярных архитектур, составили своё представление о том, что же происходит в области обработки текстов в настоящее время и получили некоторое представление о том, какие архитектуры сейчас популярны и какие основные концепты лежат в их основе. Для более подробного понимания темы я рекомендую ознакомиться с информацией, которая доступна по ссылкам с этого слайда. Информация из обзоров под названием "illustrated transformer^[1]" и "illustrated BERT^[2]" использовалась в этой лекции, но, для того чтобы более подробно понять, что же творится под капотом [BERT](#), я рекомендую всё-таки ознакомиться с этим обзором и посмотреть на отличные иллюстрации, которые там представлены. Удачи в изучении материалов!

[1] <http://jalammar.github.io/illustrated-transformer/>

[2] <http://jalammar.github.io/illustrated-bert/>

Вопрос 5

Выполнен

Баллов: 1,00 из 1,00

Расположите названия моделей в хронологическом порядке (в порядке публикации статей - те статьи, которые вышли раньше, следует расположить выше). Сначала (вверху) должно идти название самой старой архитектуры, потом более новые (внизу).

Вот ссылки на работы:

1. ELMO: [Deep contextualized word representations](#)
2. BERT: [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#)
3. Transformer-XL: [Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context](#)
4. GPT-2: [Language Models are Unsupervised Multitask Learners](#)

ELMO

BERT

Transformer-XL

GPT-2

Ваш ответ верный.