

РАСШИРЕНИЕ МЕХАНИЗМОВ ВНИМАНИЯ ДЛЯ АРХИТЕКТУРЫ SEQ2SEQ.

Кузнецов Денис Павлович

Аспирант

Факультет КиБ НИЯУ МИФИ, Москва, Россия

E-mail: kuznetsov.den.p@mail.ru

За последние несколько лет значительно вырос интерес к задачам машинного перевода. Интерес был вызван успехами в этой области. Большой вклад за последнее время в машинный перевод внесли глубокие нейронные сети. Прорыв в машинном переводе был вызван изобретением принципиально новых архитектур нейронных сетей, таких как seq2seq, transformer. Одна из важнейших частей упомянутых архитектур — это, так называемые, механизмы внимания.

Механизмы внимания в нейронных сетях позволяют выделять из набора данных те, которые могут внести полезный вклад в генерацию предсказания нейронной сети.

Как было показано, в случае использования архитектуры seq2seq, использование некоторых из механизмов внимания позволяет получить результаты превышающие результаты базовой модели на восемь пунктов BLEU — одной из метрик, которая используется для определения качества перевода.

В данной работе было использовано несколько новых механизмов внимания. Один из механизмов внимания позволяет выделять участок памяти так, что с использованием рекуррентных сетей вклад в результирующие данные дает близлежащие данные. Еще один механизм внимания использует ограниченный softmax, что позволяет выделять отдельные участки данных с взаимоучитывающими распределениями внимания.

Проведенные эксперименты в лаборатории глубокого обучения и нейронных сетей МФТИ дали положительные результаты различного характера. Было выявлено, что модифицированные сети быстрее обучаются и достигают более высоких значений расчета целевых метрик, таких как BLEU и подобных.

Литература

1. Dzmitry Bahdanau, Neural Machine Translation by Jointly Learning to Align and Translate, arxiv.org, no.1409.0473, 2014.
2. Ashish Vaswani, Attention Is All You Need, arxiv.org, no.1706.03762, 2017.

3. Minh-Thang Luong, Effective Approaches to Attention-based Neural Machine Translation, arxiv.org, no.1508.04025, 2015.
4. Kishore Papineni, BLEU: a Method for Automatic Evaluation of Machine Translation, arxiv.org, no.P02-1040, 2002.