

## Рецензия

Статья посвящена анализу контекстной зависимости контекстозависимых векторных представлений. Для чего вводятся 3 метрики, после чего производится анализ геометрического представления для контекстуализированных моделей/ сравнительный анализ трех таких моделей (ELMo, BERT и GPT2) и сравнительный анализ статичных и контекстуализированных моделей на основе данных метрик. В результате авторы обнаружили, что слоям рассмотренных моделей почти не свойственна изотропия (с исключением одного слоя в ELMo, который, однако, нельзя назвать контекстно зависимым), хотя при этом для статичных представлений изотропия наоборот, полезна. Также были выделены группы слов, обладающих наиболее выраженной контекстной специфичностью, а также показана невозможность качественно представить контекстозависимое представление через конечную комбинацию статичных.

Сильные стороны статьи:

- Все выводы, сделанные в статье, обоснованы и явно вытекают из экспериментов, подробное описание которых присутствует в тексте. Необходимые для интерпретации величины средних значений для введенных метрик также выведены конкретно под эту задачу.
- Рассматриваемые контекстные представления достаточно молоды, и на момент написания статьи проведенные исследования, изучающие их поведение фокусировались лишь на изучении наличия влияния контекста на модель.
- Метод же описанный в первой части работы впервые был применен по отношению к контекстным векторным представлениям (до этого подобные исследования проводились только со статичными представлениями)

Слабые стороны статьи:

- Хотелось бы понять логику выбора именно этих статических эмбедингов для сравнения. Нам вкратце описывается что это - одни из лучших существующих моделей. Однако, лучше было бы также указать, как они ведут себя на выбранных наборах данных относительно других статических векторных представлений. Может быть под конкретно эту задачу кто-то из прямых конкурентов работает лучше.
- В работе вводятся метрики, для понимания численных значений которых требуется сравнение с некоторым бейзлайном. Однако, для метрики MEV корректность выполнения сравнения с анизотропичным бейзлайном не описывается

“ Though not visible in Figure 4, the raw MEV of many words is actually below the anisotropy baseline”

Текст статьи:

- В целом, статья написана понятным, легко воспринимаемым языком. Логические связи внутри текста не нарушены. А так же все выводы делаются пошагово с объяснением нюансов.
- Единственным крупным замечанием можно выделить, что в обзоре литературы упоминается статья, которая использует подобный подход, однако ни здесь ни далее в тексте не уточняется в чем заключается степень схожести.

“It is more similar to Mimno and Thompson (2017), which studied the geometry of static word embedding spaces.”

Воспроизводимость статьи:

- Эксперименты описаны четко, нет опущенных шагов.
- Указаны версии предобученных моделей, которые были использованы:

“We use the pretrained models provided in an earlier version of the PyTorch-Transformers library” (с гиперссылкой на библиотеку)

- Также указаны, какие наборы данных были использованы:

“Our input data come from the SemEval Semantic Textual Similarity tasks from years 2012 - 2016 (Agirre et al., 2012, 2013, 2014, 2015)”

- Необходимые для экспериментов метрики, введенные в статье, подробно расписаны.

Итоговая оценка

Критерий НИПС

9: Top 15% of accepted NeurIPS papers. An excellent submission; a strong accept.  
I will fight for accepting this submission.

Критерии уверенности

4: You are confident in your assessment, but not absolutely certain.

It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.