

# How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings, Kawin Ethayarajh

Автор исследования: Хамдеева Дилара

1. Дата публикации. Конференция. В каком виде представлена.
  - Сентябрь, **2019**
  - **EMNLP-IJCNLP** (Empirical Methods in Natural Language Processing, International Joint Conference on Natural Language Processing)
  - **Oral**
2. Автор статьи.
  - Kawin Ethayarajh
  - PhD student at Stanford NLP group.
3. Область интересов.
  - Из личного блога <https://kawine.github.io>: *“My research interests are in representation learning (e.g., word embeddings, additive compositionality) and evaluation (i.e., how good is your NLP model really?).”*
  - Основная исследовательская область автора - NLP. Большинство предшествующих работ имеют довольно узкую специфику исследования: word embeddings, word relations и проч. Однако несмотря на их общую тематику, нынешняя статья не является улучшением какой-либо предыдущей работы автора. Скорее это дополнение к уже проведенным исследованиям.
4. Какие из статей в списке ссылок оказали наибольшее влияние на данную работу?

На протяжении всей статьи можно выделить две работы, на которые наиболее часто ссылается автор. А именно:

**(1) All-but-the-top: Simple and effective postprocessing for word representations. Mu et al., 2018**

К.Е. опирался на идею об изотропности, которую заметили авторы (1). В статье (1) “All-but-the-top” обнаруживается, что статические эмбединги анизотропны. Но при этом исключив анизотропность из статичных векторов можно значительно улучшить производительность на некоторых downstream tasks.

К.Е. так же как и авторы (1) попробовал устранить анизотропность в контекстных эмбедингах и пришел к выводу, что анизотропность -- это продукт контекстуализации. Если убрать анизотропность, можно увидеть насколько сильно контекстные эмбединги зависимы от контекста.

**(2) The strange geometry of skip-gram with negative sampling. Mimno and Thompson, 2017**

В данной статье исследовалась геометрия статических эмбедингов. К.Е. воспользовался геометрической интерпретации анизотропии -- вектора лежат в узком конусе. Более того, сам К.Е. пишет, что его работа наиболее схожа с исследованием (2). *“Our work in this paper is thus markedly different from most dissections of contextualized representations. It is more similar to Mimno and Thompson (2017), which studied the geometry of static word embedding spaces.”*

5. Кто цитирует данную статью? Прямые продолжения рассматриваемой статьи.



- Mu et al. (2018) показали, что для статических эмбедингов устранение анизотропности положительно сказывается на результатах. В то же время К.Е. обнаружил, что контекстные представления крайне анизотропны во всех слоях. Почему бы не попробовать сделать контекстные эмбединги более изотропными? Допустим, добавив штраф на анизотропность.
- Несмотря на то что в этом году уже публиковались исследования по решению данного вопроса, они были исключительно модельно-ориентированы, например только BERT (3, 4).
- Также автор предлагает попробовать измерять близость векторов другими метриками.

#### 8. Попробуйте найти применение статье в индустриальных приложениях.

- Конструировать статические эмбединги из контекстных. А именно, брать первые топ РС компонент на определенных слоях модели (чаще всего это нижние слои тк они менее анизотропны и менее контекстуализированы, но при этом вмещают в себя намного больше семантической информации нежели обычные статические представления).