



🏠 [Эмбединги слов и параграфов](#)

Эмбединги слов и параграфов

В этом разделе вы изучите классические способы представления слов текста в виде вещественных векторов фиксированной длины (эмбедингов). Это важно для нейросетевой обработки текста, поскольку нейросетевые модели не умеют работать с категориальными объектами (такими, как слова), зато отлично справляются с обработкой вещественных векторов. Изученные методы применимы не только для представления слов в тексте, но и для представления любых дискретных объектов в последовательностях (нуклеотидов в последовательности ДНК, действий пользователя на сайте, товаров, купленных в интернет-магазине и т.д.)

[□ Эмбединги слов](#)

[Введение, что такое эмбединг слов, символов и любых дискретных объектов простыми словами.](#)

[□ Совстречаемость слов](#)

[Обучение эмбедингов слов с помощью счётчиков совстречаемости слов и используя меру PPMI.](#)

[□ Латентный семантический анализ](#)

[Снижение размерности эмбедингов слов и документов с помощью метода Latent Semantic Analysis \(LSA\).](#)

[□ Word2vec](#)

[Модели Word2vec \(CBOW и SkipGram\) и модель fastText для построения эмбедингов слов текста.](#)

[□ Оптимизация Skip-Gram](#)

[Вычислительно эффективные способы настройки модели Word2Vec на примере SkipGram, используя Hierarchical SoftMax и негативное сэмплирование \(negati...](#)

[□ Эмбединги параграфов](#)

[Базовые методы для построения эмбедингов предложений, параграфов и документов целиком.](#)

[□ Вопросы](#)

[Предыдущая страница](#)
[«YOLACT](#)

[Следующая страница](#)
[Эмбединги слов »](#)

© 2023-25 [Виктор Китов](#). [Новости проекта](#).