

### Vocabulary Learning via Optimal Transport for Neural Machine Translation

1. Запишите формулу MUV (Marginal Utility of Vocabularization) с обозначениями.
2. Опишите постановку задачи Optimal Transport Problem.
3. Опишите, как в методе VOLT на каком-то определенном временном шаге  $t$  происходит переход от оптимизации верхней оценки MUV для множества словарей  $v(t)$  к транспортной задаче.

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

1. ... скоро будут...

### The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks

1. ... скоро будут ...

### Sharpness-Aware Minimization for Efficiently Improving Generalization

1. Опишите шаги алгоритма обучения с помощью SAM. Для нормы epsilon используется  $l_2$  норма ( $p = 2$ ).
2. Опишите, что такое m-sharpness в SAM.
3. Что такое SAM optimization problem? Покажите, как выводятся формулы для подсчета градиента SAM loss-a.

### Momentum Residual Neural Networks

1. Сформулируйте формулу для прямого прохода momentum модели. Как ее обратить?
2. Приведите и поясните оценки потребляемой памяти для обычного резнета и momentum резнета.
3. Перечислите основные свойства моментум резнета.

### DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort

1. Написать формулу AdaIN, рассказать, что за переменные и для чего используется в StyleGAN.
2. Описать процесс обучения DatasetGAN.
3. Описать процесс поиска плохих изображений при генерации датасета.