

Summary

Авторы рассматривают проблему оптимизации размера словаря для машинного перевода. Для задачи машинного перевода выбор большего словаря не всегда дает лучшее качество, поэтому они предлагают меру качества словаря для машинного перевода которая коррелирует с итоговым BLUE полученной модели.

Также они предлагают эффективный метод подсчета данной метрики путем задачи Оптимального Транспорта

Плюсы:

- Авторы затрагивают актуальную проблему подбора размера словаря для задачи машинного перевода, так как существующий метод требует обучения модели каждый раз для замера качества
- Предложен способ и метрика для оптимизации размера словаря без построения модели и способ быстрого ее вычисления

Минусы:

- Нет четкого обоснования почему MUV хорошая метрика для замера точности модели машинного перевода. Для 30% датасетов она работает в обратную сторону. Возможно стоило попробовать придумать другую метрику которая имела бы большую корреляцию, например непонятна обратная линейная зависимость от количества токенов в формуле
- Кажется нечестным выбор BPE-30k и BPE-1k бейзлайна для всех имеющихся языков и не уточнено какая именно модель в итоге использовалась для обучения
- Показалось нечестным сравнение VOLT и MUV-Search. Насколько я понял из кода MUV-Search реализовывался с помощью построения каждый раз заново BPE и вычисления MUV, при этом не видно чтобы применялось распараллеливание. Плюс насколько я понял подсчет MUV можно реализовать онлайн в момент построение BPE-дерева, что могло бы сильно ускорить MUV-Search и сделать его равным по скорости VOLT, при этом возможно получилось бы найти более оптимальное по объему словаря решение судя по их результатам в Table 4.

Воспроизводимость:

Весь код реализации выложен в открытый доступ и можно его запустить. Есть подробный сет экспериментов и подробности реализации.

Самое непонятное в статье - почему авторы считают что MUV нужная метрика для оптимизации размера словаря, она имеет положительное влияние далеко не для всех

Оценка: 4

Confidence: 4