**Отчёт по результатам декодирования (Greedy, Beam, Beam+LM, Beam+LM Rescore)**  
  
Данный эксперимент проводился с целью реализовать и сравнить качество распознавания речи (ASR) четырьмя методами:

1. Greedy (жадное CTC-декодирование)
2. Beam Search (без LM)
3. Beam Search + Language Model (shallow fusion)
4. Beam Search + LM Rescore (дополнительный двухшаговый проход)

В качестве модели использовалась Wav2Vec2 (facebook/wav2vec2-base-960h), а для LM – трёхграммная KenLM-модель.

Мы использовали **Character-level Levenshtein distance**, показывающую число операций (вставка, удаление, замена символов), необходимых для преобразования расшифровки в эталонную транскрипцию. Чем меньше это число, тем лучше качество распознавания.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Сэмпл** | **Greedy** | **Beam** | **Beam+LM** | **Beam+LM Rescore** | **Beam+LM (модель без прунинга)** | **Beam+LM Rescore**  **(модель без прунинга)** |
| 1 | 8 | 10 | 9 | 11 | 15 | 11 |
| 2 | 0 | 5 | 6 | 5 | 22 | 5 |
| 3 | 1 | 5 | 6 | 5 | 15 | 5 |
| 4 | 4 | 8 | 8 | 8 | 17 | 8 |
| 5 | 1 | 2 | 2 | 2 | 5 | 2 |
| 6 | 15 | 16 | 18 | 16 | 18 | 16 |
| 7 | 17 | 19 | 19 | 19 | 20 | 19 |
| 8 | 14 | 15 | 16 | 15 | 19 | 15 |

Наблюдения

* Greedy: во многих тестовых примерах Greedy выдаёт результат лучше (или не хуже) остальных методов. В одном из случаев (пример №2) жадный декодер и вовсе ошибается на 0 символов, то есть идеально совпадает с эталоном.
* Beam Search: даёт немного более длинные гипотезы, но не всегда это приводит к улучшению. Иногда, наоборот, он добавляет мелкие ошибки или «подвисает» на CTC-повторах, что в итоге даёт Levenshtein distance больше, чем у Greedy.
* Beam Search + LM (shallow fusion): при больших α,β (например, 1.0, 1.0) мы наблюдали резкое ухудшение результатов из-за чрезмерного «вмешательства» LM. При уменьшении α до 0.01 и β=0 LM перестаёт «портить» выход, но и не сильно помогает. Например, в первом примере становится на 1 символ лучше, чем Beam без LM, но всё равно не догоняет Greedy.
* Beam Search + LM Rescore: результаты похожи на обычный Beam. Если LM не даёт сильного положительного вклада, то и финальный рескоринг не улучшает ситуацию заметно.

**Интересные моменты:** изначально были α=1.0, β=1.0, но LM «ломал» расшифровку (добавлял десятки лишних символов). Постепенно уменьшая α (до 0.4, 0.1, 0.01) и обнуляя β, мы пришли к балансу, при котором LM уже не вредит, но заметного выигрыша в метриках тоже не даёт. Возможно LM обучен на другом стиле данных или акустические вероятности и так достаточно точны. Большая модель не дает преимуществ, даже ровно наоборот - дает ухудшение. Также изменение бима не дало улучшений при данных вводных.

**Итог:** greedy лучшие результаты показывает, что наводит на мысль, что facebook/wav2vec2-base-960h отлично обучен - потому и гриди показывает лучший результат.