

Language GANs Falling Short

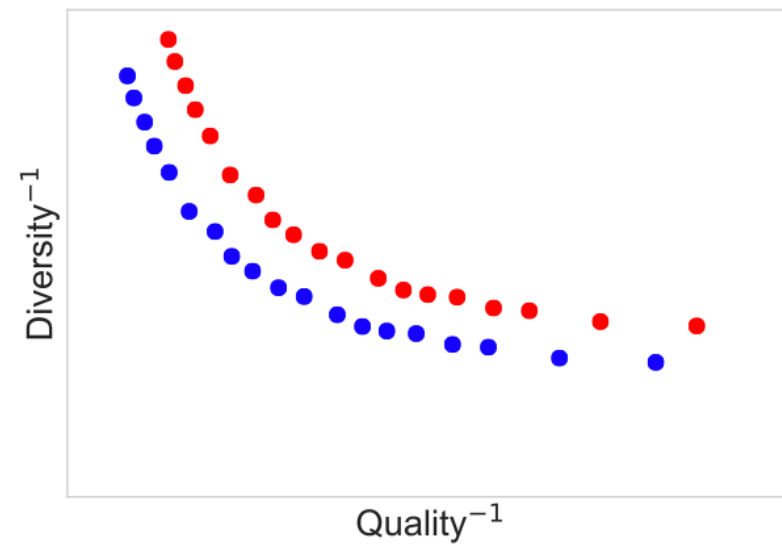
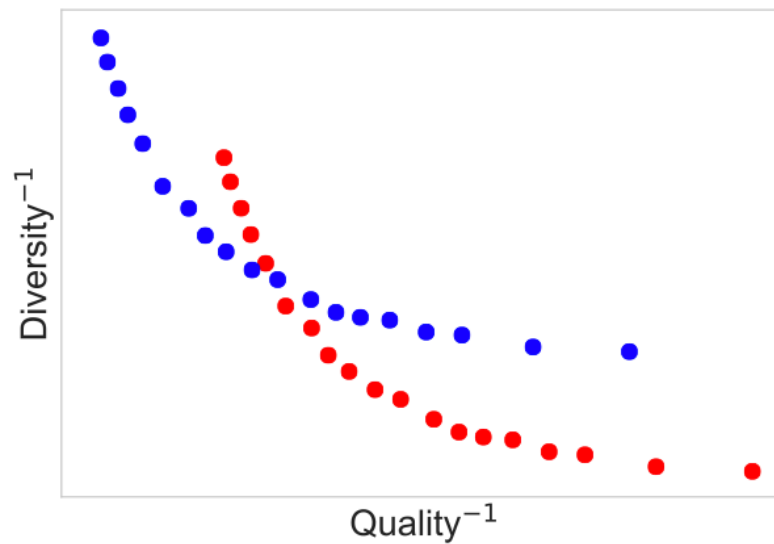
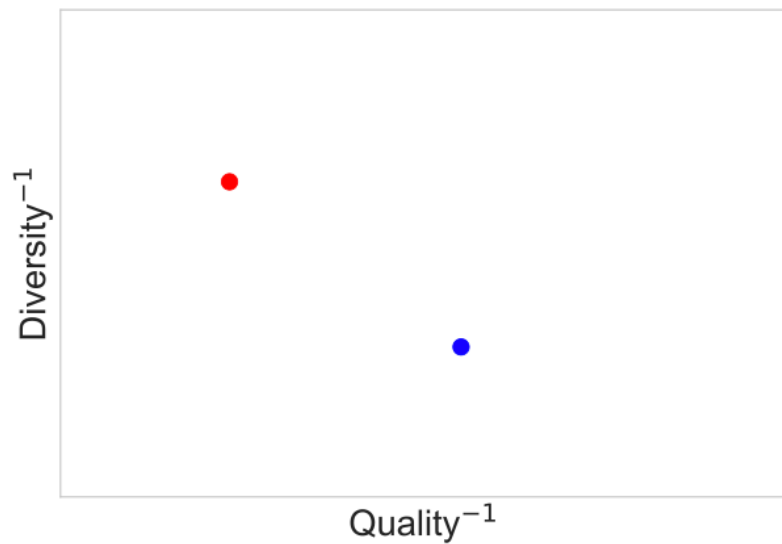
Левина Александра

3.03.2021

Задача генерации текста

- MLE (различие между обучением и инференсом) → плохое качество
- GAN (нет различия) → скудное разнообразие

Компромисс между качеством и разнообразием



Генеративные состязательные сети

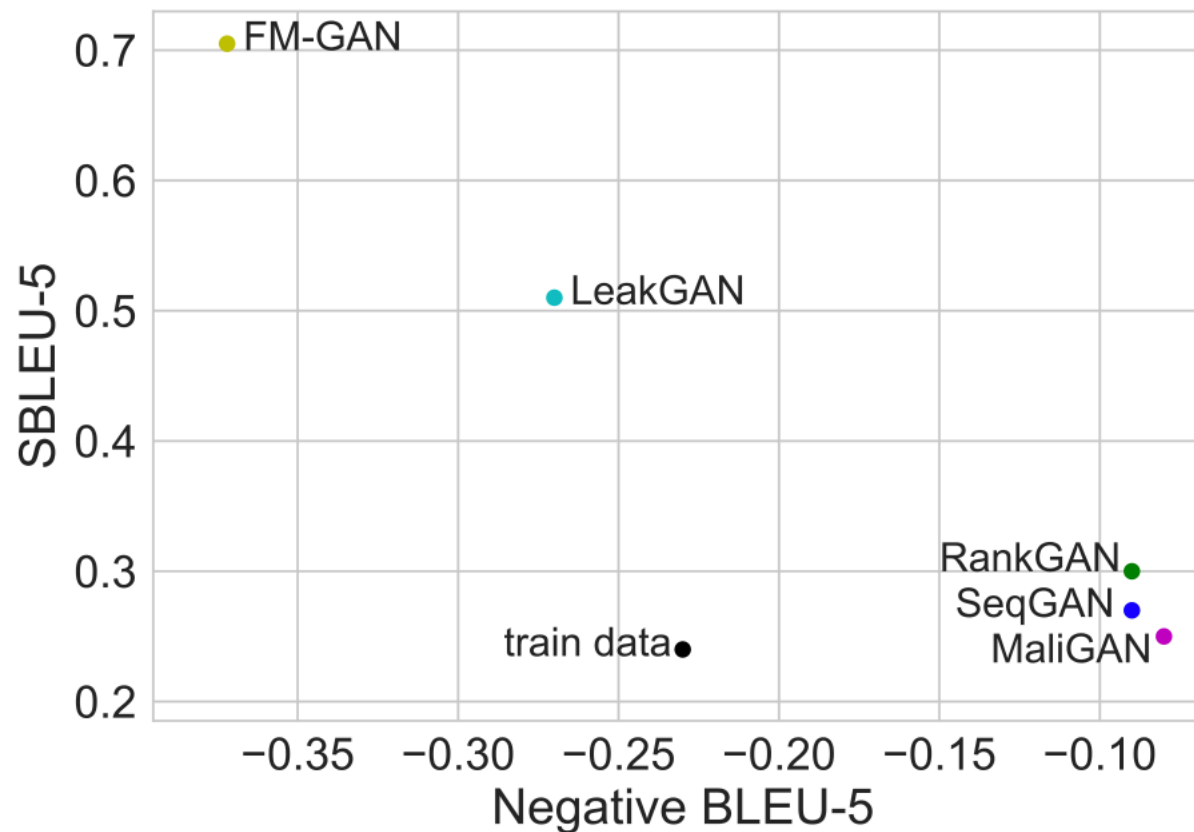
- Были предложены для непрерывных задач
- Задача генерации текста дискретная + текст генерируется последовательно токен за токеном

Генеративные состязательные сети

- SeqGAN (2017) -> генератор = стохастическая стратегия обучения с подкреплением
- Corpus-level BLEU = доля n-грамм в сэмпле, которые присутствуют в референсном корпусе

Генеративные состязательные сети

- Self-BLEU
- negative BLEU



датасет EMNLP2017 News

RankGAN (2017)
MaliGAN (2017)
TextGAN (2017)
LeakGAN (2017)
IRL-GAN (2018)

CoT (2019)
FM-GAN (2020)

Температурное движение (temperature sweep)

o_t – активация генератора

W – матрица эмбедингов

α – параметр температуры

$$G_{\theta}(x_t \mid x_{1:t-1}) = \text{softmax}(o_t \cdot W / \alpha)$$

$$G_{\theta}(x_t \mid x_{1:t-1}) = \text{softmax}(o_t \cdot W/\alpha)$$

α Samples

-
- 2.0 (1) If you go at watch crucial characters putting awareness in Washington , forget there are now unique developments organized personally then why charge .
 (2) Front wants zero house blood number places than above spin 5 provide school projects which youth particularly teenager temporary dollars plenty of investors enjoy headed Japan about if federal assets own , at 41 .
-
- 1.0 (1) Researchers are expected to comment on where a scheme is sold , but it is no longer this big name at this point .
 (2) We know you ' re going to build the kind of home you ' re going to be expecting it can give us a better understanding of what ground test we ' re on this year , he explained .
-
- 0.7 (1) The other witnesses are believed to have been injured , the police said in a statement , adding that there was no immediate threat to any other witnesses .
 (2) The company ' s net income fell to 5 . 29 billion , or 2 cents per share , on the same period last year .
-
- 0.0 (1) The company ' s shares rose 1 . 5 percent to 1 . 81 percent , the highest since the end of the year .
 (2) The company ' s shares rose 1 . 5 percent to 1 . 81 percent , the highest since the end of the year .
-

Температурное движение (temperature sweep)

- Какую выбрать модель, если нужны разнообразные семплы?
- Действительно ли модель лучше остальных или она просто снижает энтропию распределения?

Эмпирические результаты

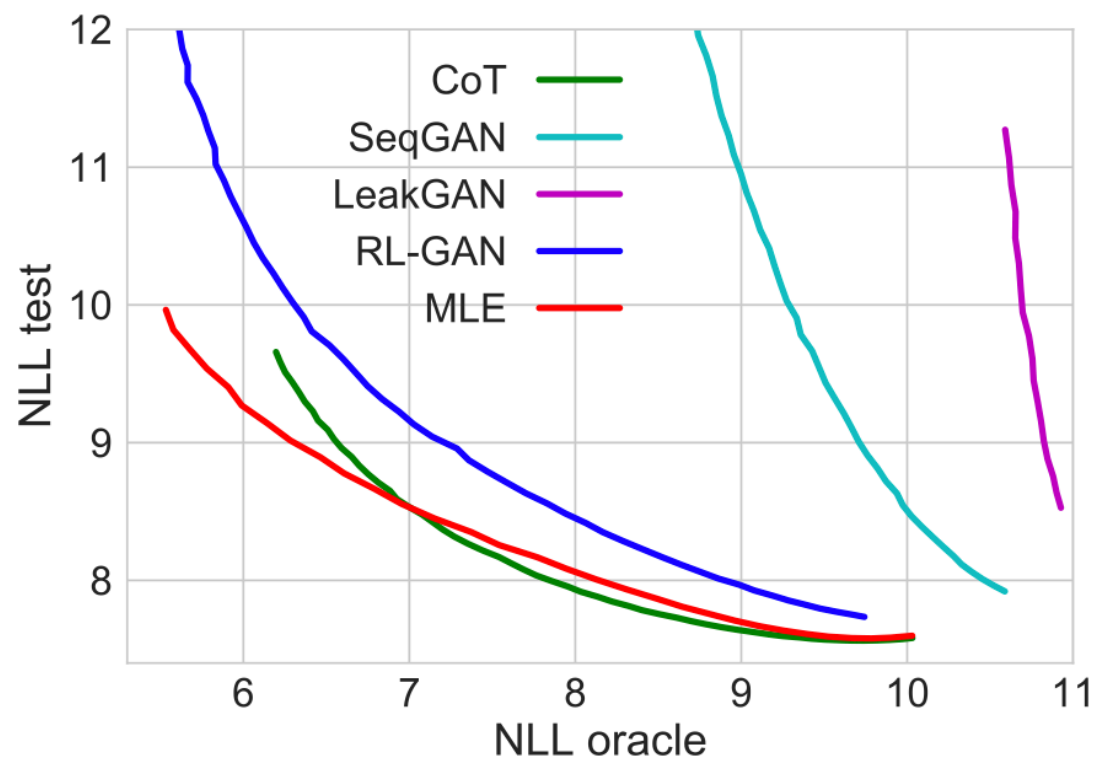
- RL-GAN:
 - SeqGAN (SeqGAN, 2017)
 - MLE предобучение генератора (SeqGAN, 2017)
 - Leaky дискриминатор (LeakGAN, 2017)
 - Step-level loss (MaskGAN, 2018)
 - Learned baseline (MaskGAN, 2018)
 - Regularizing REINFORCE with max-entropy loss (Function optimization using connectionist reinforcement learning algorithms, 1991)
- MLE модель
- Остальные модели из официальных репозиториев

Выбор диапазона температур

- На обеих осях есть границы информативности
- По оси разнообразия нет смысла увеличивать энтропию MLE модели → снижение температуры остальных моделей до уровня разнообразия MLE модели
- По оси качества снижение температуры по тех пор, пока разница с GAN не была очевидна/пока не было достигнуто качество Reverse LM

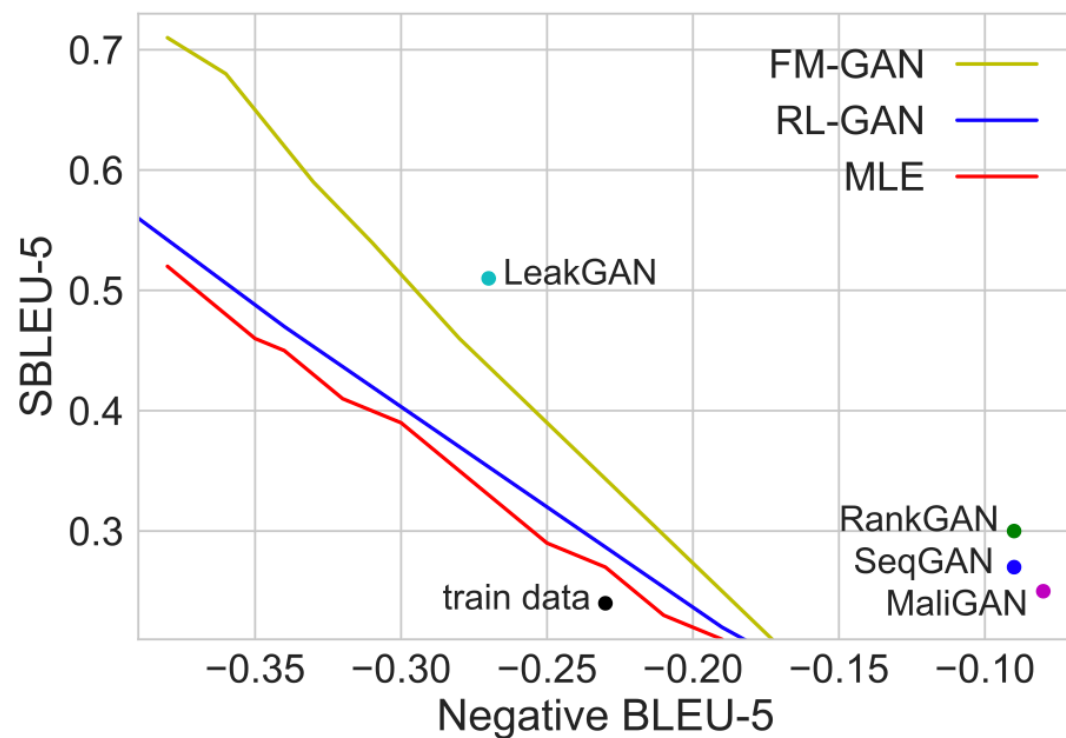
Генерация синтетических данных

- LSTM оракул со скрытым слоем размера 32

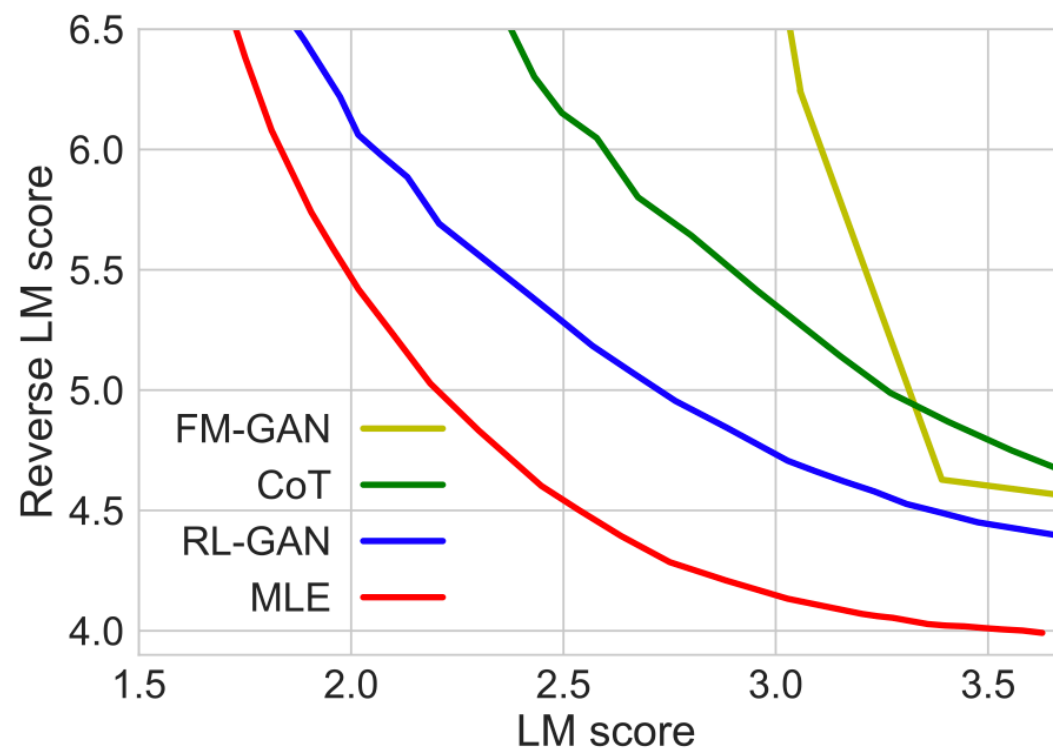


Model	NLL_{oracle}
SeqGAN (Yu et al., 2017)	8.74
RankGAN (Lin et al., 2017)	8.25
LeakGAN (Guo et al., 2017)	7.04
IRL (Shi et al., 2018)	6.91
MLE ($\alpha = 1.0$)	9.40
MLE ($\alpha = 0.4$)	5.50
MLE ($\alpha = 0.001$)	4.58

Генерация длинных текстов

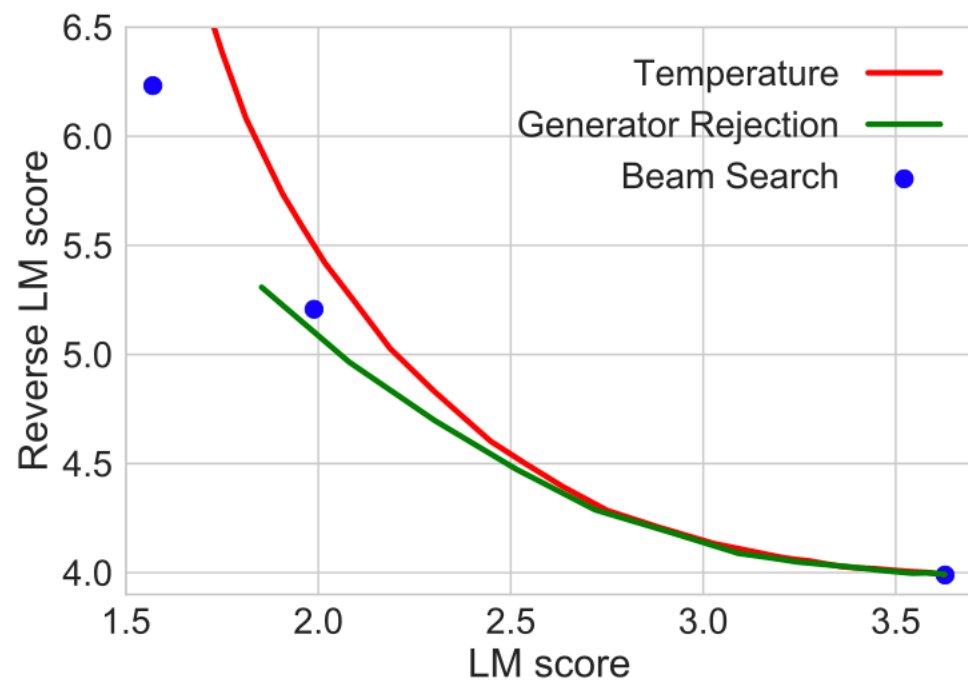


(a) Local metrics

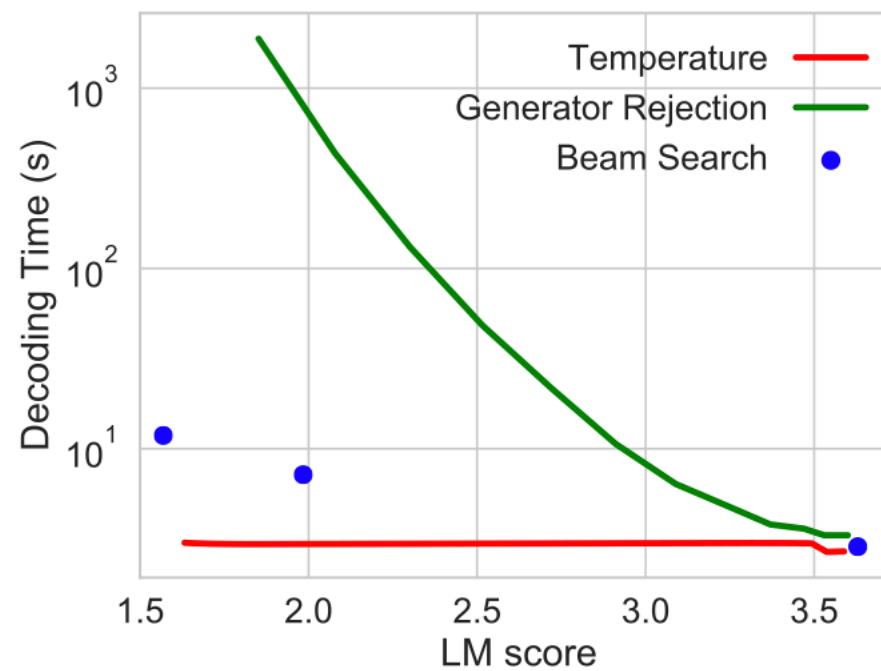


(b) Global metrics

Сравнение механизмов декодинга

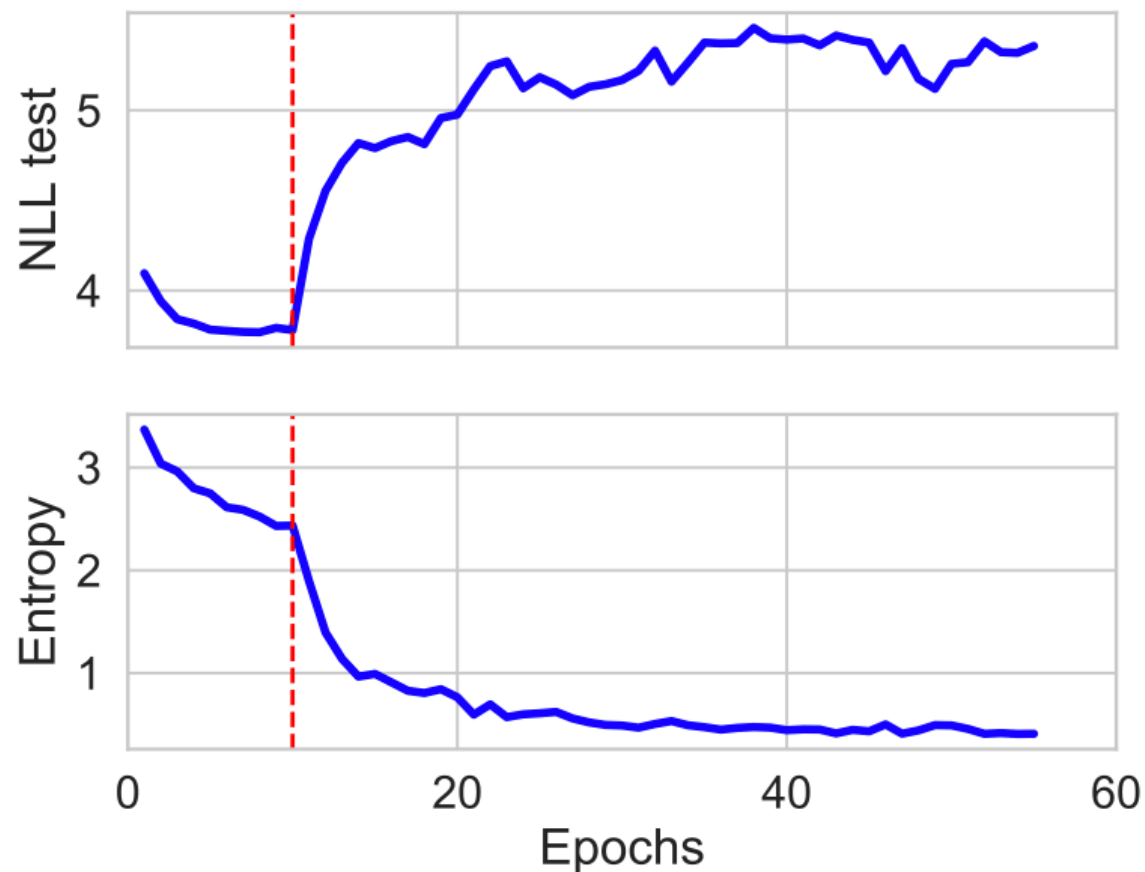


(a) Diversity and quality



(b) Decoding time and quality

Энтропия распределения генератора



Итоги

- Хорошо настроенная языковая модель является сильным бейзлайном для задачи генерации текста
- Метод температурного движения предоставляет возможность охарактеризовать результативность модели в терминах метрики качества-разнообразия
- Состязательные модели эффективны на реальных текстах, но все еще существует много трудностей в связи с дискретной задачей, которые предстоит преодолеть

Вопросы

1. Какую метрику исследовали авторы статьи? С какой частью данной метрики лучше у GAN-моделей, а с какой у MLE-моделей? Почему?
2. В чем заключается метод temperature sweep?
3. Опишите эксперименты из статьи и скажите, у какой модели (GAN или MLE) были лучше результаты в каждом эксперименте.