

# Методы повышения разнообразия в системах машинного перевода

Михаил Солоткий

НИУ Высшая Школа Экономики  
Факультет компьютерных наук  
Базовая кафедра Яндекса

**Курсовая работа**

Научный руководитель — Бабенко Максим Александрович

Москва 2020 г.

- Обычно системы машинного перевода по одному предложению выдают одно предложение
- Цель: построить метод, генерирующий несколько разнообразных переводов
- Цель: максимизировать разнообразие и качество

- Цели пользователя неизвестны
- Помощь в изучении языка
- Генерация переводов разного стиля

# Нейросетевой машинный перевод

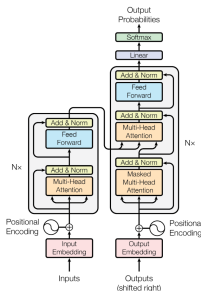


Figure 1: The Transformer - model architecture.

- $p(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^n p(y_t | \mathbf{y}_{<t}, \mathbf{x}) \rightarrow \max_{\mathbf{y}}$
- beam search выдаёт  $k$  приближённо наиболее вероятных переводов

Vaswani, A. et al (2017). Attention is all you need. Advances in Neural Information Processing Systems

- Обучение на закрытых данных
- Тестировалось на WMT'18 en-ru

- $$\text{BLEU}(out, ref) = \min \left\{ 1, \exp \left\{ \frac{\text{len}(ref)}{\text{len}(out)} \right\} \right\} \left[ \prod_{i=1}^4 \text{precision}_i \right]^{\frac{1}{4}}$$
- Не учитывает смысл перевода, а только совпадение n-грамм

---

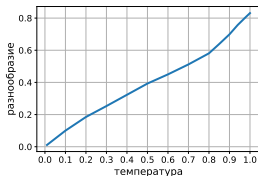
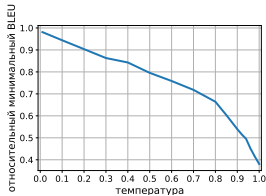
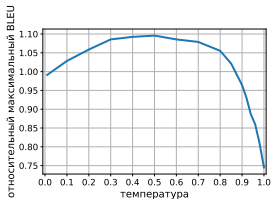
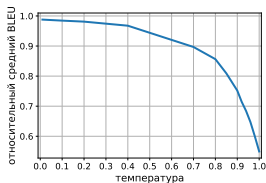
Papineni, K. et al (2002). Bleu: a method for automatic evaluation of machine translation. Association for Computational Linguistics

Sulem, E., et al (2018). BLEU is not suitable for the evaluation of text simplification. Association for Computational Linguistics.

- avg-BLEU
- max-BLEU
- min-BLEU
- $\text{self-BLEU}(\text{outs}) = \frac{1}{\text{len}(\text{outs})^2} \sum_{i=1}^m \sum_{j=1}^m \text{BLEU}(\text{outs}_i, \text{outs}_j)$
- $\text{diversity}(\text{outs}) = 1 - \text{self-BLEU}(\text{outs})$
- Некорректно сравнивать при разном числе переводов
- Фиксируем число переводов,  $k=3$

# Temperature sampling

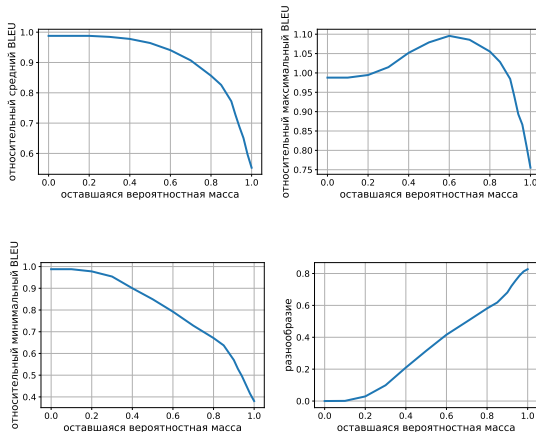
$$\text{softmax}_i(\mathbf{z}; t) = \text{softmax}_i\left(\frac{\mathbf{z}}{t}\right) = \frac{\exp(z_i / t)}{\sum_{j=1}^n \exp(z_j / t)}$$





# Nucleus sampling

- Отбросить  $p$  вероятностной массы, приходящейся на токены с самой низкой вероятностью и взять семпл



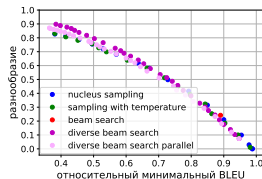
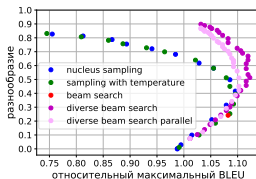
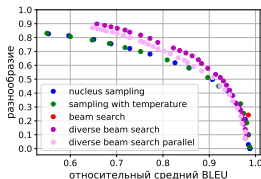
Holtzman, A. et al (2019). The curious case of neural text degeneration.

- Diverse Resampling: семплируем 10 переводов, из них выбираем наиболее разнообразные
- Модель с латентными переменными: к сети добавляется автокодировщик, энкодер и декодер в процессе перевода обуславливаются на семплируемые латентные переменные

# Diverse Beam Search

$$\ln p(y_t | \mathbf{y}_{<t}, \mathbf{x}) + \sum_{i=1}^{k-1} \lambda_i \text{div}(\mathbf{y}_{\leq t}, \mathbf{y}^i) \rightarrow \max_{y_t}$$

- $\lambda_i$  можно брать одинаковыми
- есть параллельная версия, которая чуть хуже по качеству последовательной



Vijayakumar, A. K. et al (2016). Diverse beam search: Decoding diverse solutions from neural sequence models.

# Примеры переводов

Исходное предложение: «As a coach, I would tell you it's time to run another play.»

Обычный beam search выдаёт:

- 1) Как тренер, я бы сказал вам, что пришло время запустить другую игру.
- 2) Как тренер, я бы сказал вам, что пришло время запустить еще одну игру.
- 3) Как тренер, я хотел бы сказать вам, что пришло время запустить другую игру.

Семплинг с температурой  $t = 0.5$  даёт:

- 1) Как тренер, я скажу вам, что пришло время снова сыграть.
- 2) Как тренер, я хотел бы сказать вам, что пришло время провести еще одну игру.
- 3) Как тренер, я бы сказал, что пришло время играть в другой пьесе.

Diverse beam search,  $\lambda = 10.0$  переводит так:

- 1) Как тренер, я бы сказал вам, что пришло время запустить другую игру.
- 2) Как тренер, скажу тебе, что пора начать новую игру.
- 3) Как тренер, я хочу сказать, что пришло время начать новую игру.

- Была введена формализация понятий качества и разнообразия в машинном переводе и показана их адекватность
- Реализованы и испробованы различные методы генерации разнообразных переводов, в том числе придуманы новые
- Сделано сравнение методов по качеству и разнообразию, а также с точки зрения потребления ресурсов
- Для лучшего метода – diverse beam search – предложена параллельная реализация и адаптация к специально выбранной метрике разнообразия

# Возможные вопросы

- Как меняются графики при смене домена?
- Как это использовать в production, когда нам нужен строго определённый уровень разнообразия?
- Какие минусы у diverse beam search и можно ли не использовать семплинг вообще, если есть метод лучше?
- Можно ли одновременно сделать идеальными качество и разнообразие?
- В чём принципиальная проблема с diverse resampling?
- Насколько примеры переводов отражают качество методов и не подобраны ли они специально?