

Дифференцируемая нижняя граница для математического ожидания метрики BLEU

Студент: Жуков В.А. МФТИ ФИВТ, 4 курс
Научный руководитель: Кретов М.К.

25 Июня 2018

Содержание

Введение и основные понятия

Цель работы

Предлагаемый метод

Постановка задачи и используемые обозначения

Вывод нижней границы

Эксперименты и результаты

Модельная задача

Задача машинного перевод

Введение и основные понятия

Машинный перевод

Подходы к задаче машинного перевода:

- ▶ Статистический машинный перевод (SMT)
- ▶ Нейронный машинный перевод (NMT)

Оценка качества

Метрики качества в задаче машинного перевода:

- ▶ BLEU
- ▶ ROUGE
- ▶ METEOR

Sequence to sequence архитектура

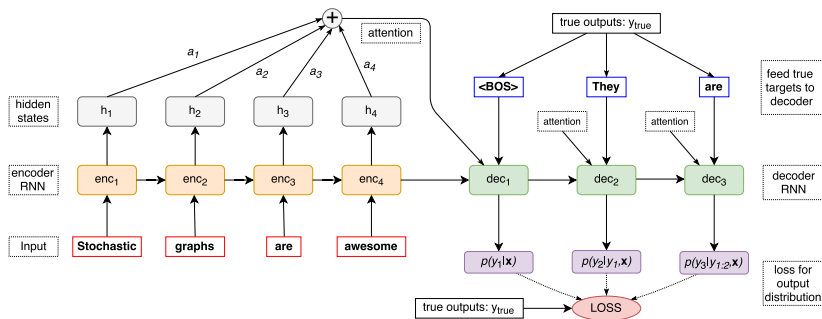


Рис. 1: Пример модели seq2seq с механизмом внимания; фаза обучения

Общий подход к оптимизации произвольной метрики

Правило REINFORCE

$$\nabla_{\theta} \mathbb{E}_{p(z; \theta)}[f(z)] = \mathbb{E}_{p(z; \theta)}[f(z) \nabla_{\theta} \log p(z; \theta)]$$

$p(z; \theta)$ – функция правдоподобия

$f(z)$ – оптимизируемая функция

Цель работы

Цель работы

Предложить более эффективный способ оптимизации (в сравнении с REINFORCE) для метрики BLEU.

$$\mathbb{E}_{\text{data}} \mathbb{E}_{\text{arch_stoch}} \mathbb{E}_{\text{output}} \text{BLEU}(x, y) \rightarrow \max \quad (1)$$

Предлагаемый метод

Метрика BLEU

$$\text{BLEU} = \text{BP} \cdot \exp \left(\sum_{n=1}^N w_n \log p_n \right) \quad (2)$$

Обычно $w_n = 1/N$, $N = 4$.

$$\text{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{1-r/c} & c \leq r \end{cases} \quad (3)$$

Нижняя граница

План вывода

- ▶ Записать метрику BLEU в матричном виде.
- ▶ Обозначить предположения необходимые для вывода.
- ▶ Вывод оценки.

Нижняя граница

Представление BLEU в матричном виде

матрица x $[\text{len}_C \times v]$

матрица y $[\text{len}_R \times v]$

$$\text{BLEU}(C, R) = \text{function}(x, y) \quad (4)$$

Нижняя граница

Необходимые предположения

- ▶ Матрица p_x получена детерминистически
- ▶ Слова R уникальны.

Нижняя граница

Итог вывода

Выход модели $p_x [seq_len \times vocab_size]$

Метки тренировочной выборки $p_y [seq_len \times vocab_size]$.

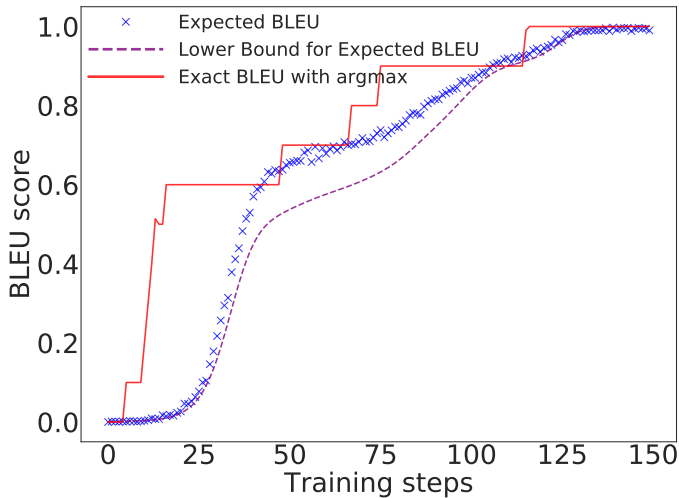
$$\mathbb{E}_{\text{output}=x \sim p_x} \text{BLEU}(x, y) \geq \text{function}(p_x, p_y) := \text{LB}$$

Эксперименты и результаты

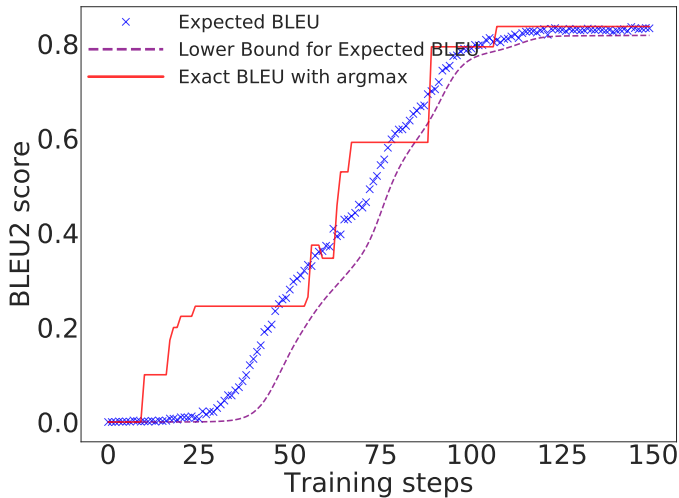
Модельная задача

- ▶ Размер словаря равен 10000
- ▶ Последовательности длины 10
- ▶ Матрица p_x сгенерированная случайно
- ▶ Функция потерь LB
- ▶ Оптимизируется матрица p_x

Модельная задача



Модельная задача



Результаты на задаче машинного перевода

	CE	RF	LB
1	26.60 ± 0.14	27.16 ± 0.18	27.83 ± 0.17
5	27.56 ± 0.17	27.85 ± 0.11	28.10 ± 0.17

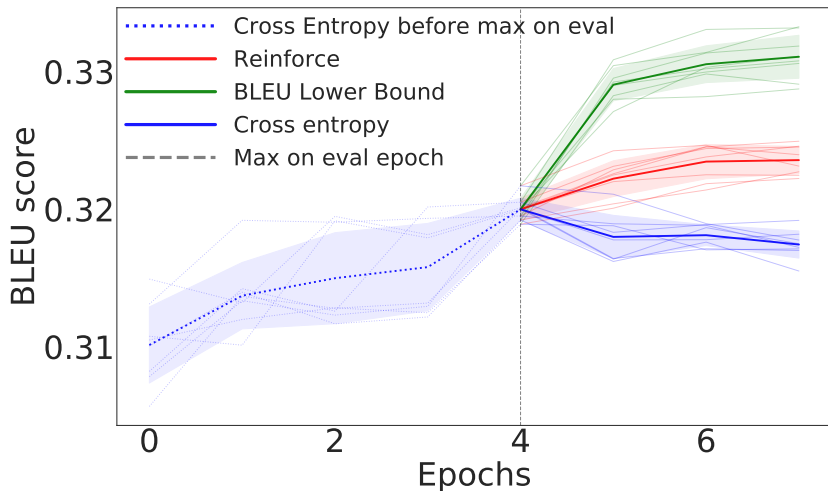
Таблица 1: Результат на задаче перевода для датасета IWSLT'14.
DE \rightarrow EN

Результаты на задаче перевода

	CE	LB
1	17.12 ± 0.2	18.65 ± 0.13
5	19.23 ± 0.19	19.03 ± 0.15

Таблица 2: Результат на задаче перевода для датасета WMT'14.
EN \rightarrow DE

Результаты на задаче перевода



Выводы

- ▶ Эффективная, дифференцируемая нижняя оценка для математического ожидания метрики BLEU.
- ▶ Улучшение итоговой метрики на задаче перевода.

Возможное дальнейшее развитие

- ▶ Другие метрики:

$$\mathbb{E}_{x \sim p_x, y \sim p_y} [M(x, y)] = f(p_x, p_y)$$

Аппроксимация математического ожидания произвольной метрики при помощи нейросетей.

- ▶ Добавление возможности оптимизации через стохастичность внутри сети:

$$\mathbb{E}_{\text{data}} \mathbb{E}_{\text{arch_stoch}} \mathbb{E}_{\text{output}} \text{BLEU}(x, y) \rightarrow \max$$

Вопросы

Вопросы?