

Representation learning в задачах анализа текстов

Камиль Сафин

Научный руководитель: Чехович Юрий Викторович

Московский физико-технический институт

17 апреля 2019 г.

Цель работы

Задача:

Улучшить модель вариационного автокодировщика путем добавления механизма внимания.

Мотивация:

- ▶ Добавление информации в процесс декодирования;
- ▶ Интерпретируемость скрытой переменной.

Проблема:

Добавление механизма внимания в стандартном виде ухудшает распределение скрытой переменной.

Основная литература

H. Bahuleyan, L. Mou, O. Vechtomova, P. Poupart
Variational Attention for Sequence-to-Sequence Models.
ICLM 2018.

Y. Deng, Y. Kim, J. Chiu, D. Guo, A. M. Rush
Latent Alignment and Variational Attention.
NeurIPS, 2018.

E. Dupon
Learning Disentangled Joint Continuous and Discrete Representations.
NeurIPS, 2018.

E. Jang, S. Gu, B. Poole
Categorical Reparametrization With Gumbel-Softmax.
ICLR 2017.

Постановка задачи

Дано:

$(X, Y) = \{x_i, y_i\}_{i=1}^N$ — два домена с объектами одинаковой структуры.

Каждый объект представляет разные модальности одной и той же сущности.

Примеры:

- ▶ соответствующие переводы одной фразы;
- ▶ изображения одного объекта.

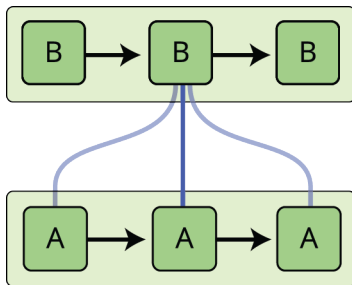
Необходимо:

Построить алгоритм, отображающий объекты одного домена в соответствующие объекты другого:

$$f : x \rightarrow y, x \in X, y \in Y, x \sim y$$

Механизм внимания

Суть: выделить часть входных данных для более детальной обработки.



На каждом шаге декодирования используется информация от каждого элемента объекта.

Вариационная оценка

Предполагаем, что объекты получены при условии некоторой скрытой переменной \mathbf{z} .

$$\log p(\mathbf{X}) \geq E_{\mathbf{z} \sim q}[\log p(\mathbf{X}|\mathbf{z})] - \mathcal{D}_{KL}[q(\mathbf{z}|\mathbf{X})||p(\mathbf{z})],$$

\mathbf{z} — скрытая переменная,

$p(\mathbf{z})$ — априорное распределение,

$q(\mathbf{z}|\mathbf{X})$ — апостериорное распределение,

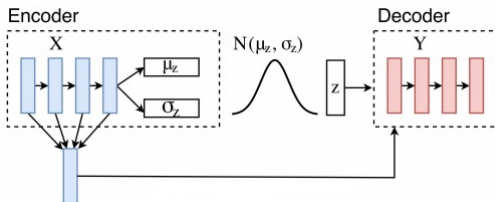
\mathcal{D}_{KL} — дивергенция Кульбака-Лейблера.

Функция ошибки:

$$J = J_{rec} + \mathcal{D}_{KL}[q(\mathbf{z}|\mathbf{X})||p(\mathbf{z})]$$

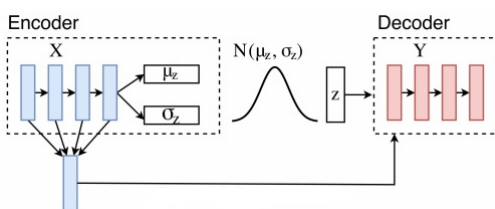
Предлагаемый подход

Добавить в модель механизм внимания.



Проблема: добавление механизма внимания в неизменном виде приводит к тому, что скрытая переменная не хранит информацию о кодируемом объекте.

Предлагаемый подход



Функция ошибки:

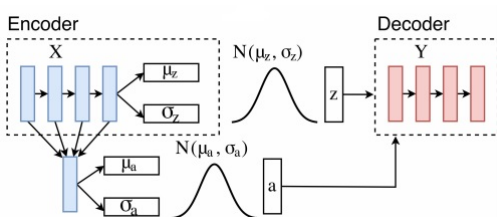
$$J = J_{rec} + \mathcal{D}_{KL}[q(z|\mathbf{X})||p(z)]$$

При данной архитектуре минимум достигается при:

- ▶ $p(z) \equiv q(z)$;
- ▶ минимизация ошибки реконструкции.

Предлагаемый подход

Решение: использовать еще одно скрытое пространство векторов \mathbf{a} для механизма внимания.



Т.к. \mathbf{z} и \mathbf{a} независимы, нижняя оценка представляется в виде:

$$\mathcal{L} = E[\log p(\mathbf{X}|\mathbf{z}, \mathbf{a})] - \mathcal{D}_{KL}[q^{(\mathbf{z})}(\mathbf{z}|\mathbf{X})||p(\mathbf{z})] - \mathcal{D}_{KL}[q^{(\mathbf{z})}(\mathbf{a}|\mathbf{X})||p(\mathbf{a})].$$

Objective:

$$\mathcal{I} = \mathcal{I}_{rec} + \lambda_{KL} \left[\mathcal{D}_{KL} [q^{(\mathbf{z})}(\mathbf{z}|\mathbf{X})||p(\mathbf{z})] + \gamma_a \sum_{j=1}^{|\mathbf{X}|} \mathcal{D}_{KL} [q^{(\mathbf{a})}(\mathbf{a}_j|\mathbf{X})||p(\mathbf{a}_j)] \right]$$

Варианты распределений

Возможно задавать разные структуры пространства \mathbf{a} :

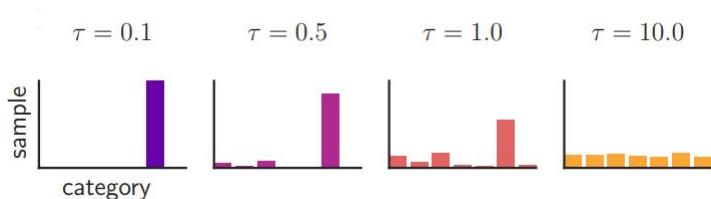
- ▶ Нормальное распределение — семплирование вектора контекста;
- ▶ Распределение по весам — семплирование весов входных элементов.

Gumbel reparametrization

Распределение Гумбеля G позволяет повторить reparametrization trick.

Основные свойства:

- ▶ $G(0, 1) = -\log(-\log U(0, 1))$;
- ▶ вероятности классов: $y_i = \frac{\exp((\log \pi_i + g_i)/\tau)}{\sum_{j=1}^k \exp((\log \pi_j + g_j)/\tau)}$, $g_i \in G(0, 1)$;
- ▶ разное поведение при разной температуре семплирования τ .



Эксперимент

Данные: EuroParl corpus

- ▶ train: 80000,
- ▶ test: 10000.

Качество: BLEU метрика.

	Без внимания	Нормальное распр.	Распр. Гумбеля
Bleu	22,5	31,2	-

Интерпретируемость компонент

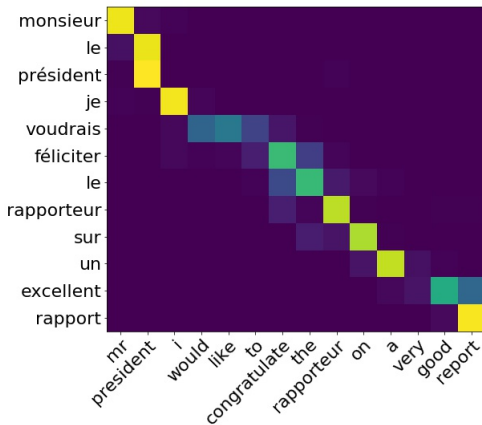


Рис.: Нормальное распределение

Интерпретируемость компонент

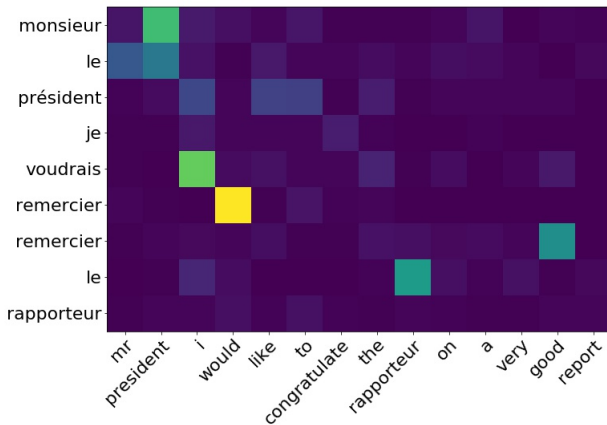


Рис.: Распределение Гумбеля

Дальнейшая работа

- ▶ Проведение эксперимента для распределения по весам;
- ▶ Тестирование модели на других задачах;
- ▶ Изучение влияния температуры семплирования на процесс обучения.

Спасибо за внимание!