Paper Reading

Xinwei Geng

2018-07-23

Trainable Greedy Decoding for Neural Machine Translation

- Jiatao Gu, Kyunghyun Cho and Victor O.K. Li
- a 4th year Ph.D. student in the Department of Electrical and Electronic Engineering at the University of Hong Kong
- a former visiting scholar at the CILVR lab, New York University and advised by Dr. Kyunghyun Cho
- EMNLP 2017
- Homepage: http://jtgu.me/

Related work

- Noisy Parallel Approximate Decoding for Conditional Recurrent Language Model
  - deep neural network, including a recurrent neural network, learns to stretch the input manifold and fill the hidden state space with it
  - a neighborhood in the hidden state space corresponds to a set of semantically similar configurations in the input space
  - small perturbation in the hidden space corresponds to jumping from one plausible configuration to another

$$\mathbf{h}_t = \phi\left(\mathbf{h}_{t-1} + \epsilon_t, \mathbf{E}\left[x_t\right], f(Y, t)\right),$$

$$\epsilon_t \sim \mathcal{N}(\mathbf{0}, \sigma_t^2 \mathbf{I}).$$

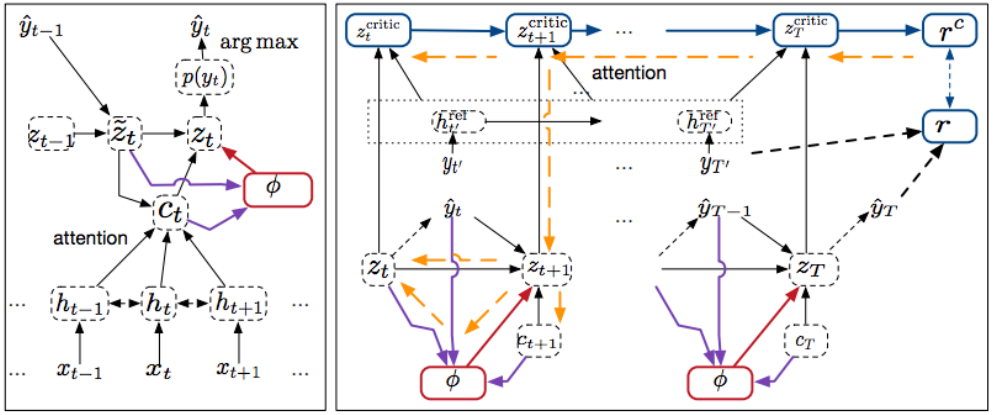
$$\sigma_t = \frac{\sigma_0}{t},$$

- There exists a decoding strategy that results in a better translation quality, and that such a better translation can be found by manipulating the hidden state of the recurrent network

Trainable Greedy Decoding

- Replacing the unstructured noise with a parametric function approximator or an agent
  - this agent takes as input the previous hidden state z_t−1, previously decoded word y_t−1 and the time-dependent context vector e_t(X; θ_e) and outputs a real-valued vectorial action a_t

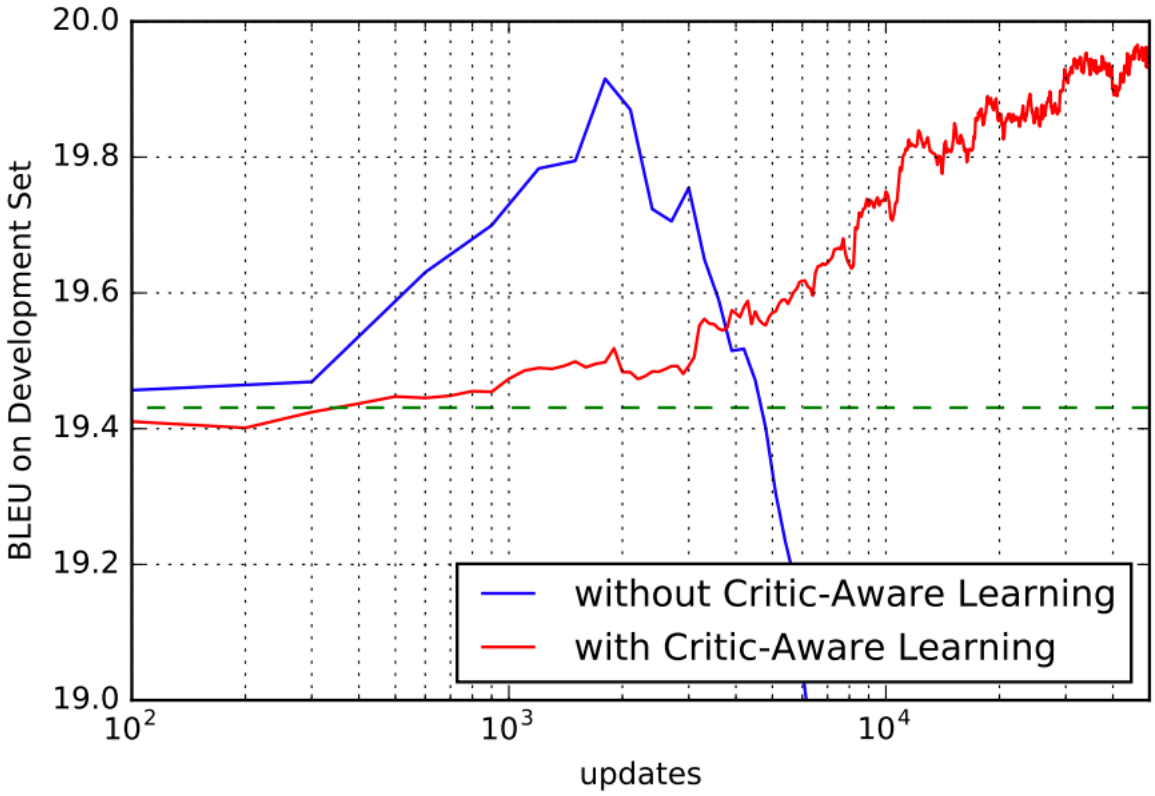
$$J^{\mathrm{A}}(\phi) = \mathbb{E}^{\hat{Y}=G_\pi(X)}_{X\sim D}\left[R(\hat{Y})\right]$$



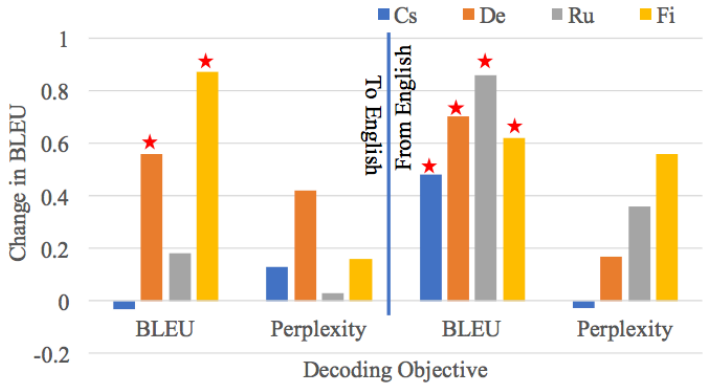
$$J^{\mathrm{C}}(\psi) = \mathbb{E}^{\hat{Y}=G_\pi(X)}_{X\sim D}\left[R^c_\psi(z_{1:T}) - R(\hat{Y})\right]^2.$$

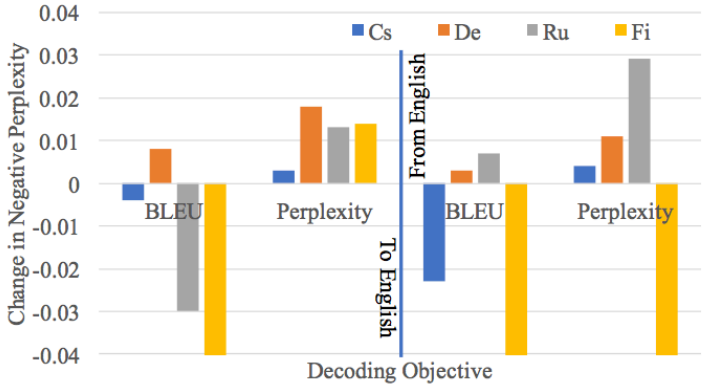
$$\hat{J}^{\mathrm{A}}(\phi) = \mathbb{E}^{\hat{Y}=G_\pi(X)}_{X\sim D}\left[R^{\mathrm{C}}(\hat{Y})\right].$$

Learning Curvels

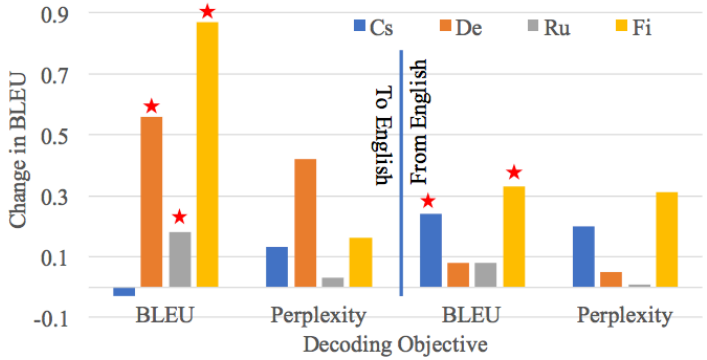


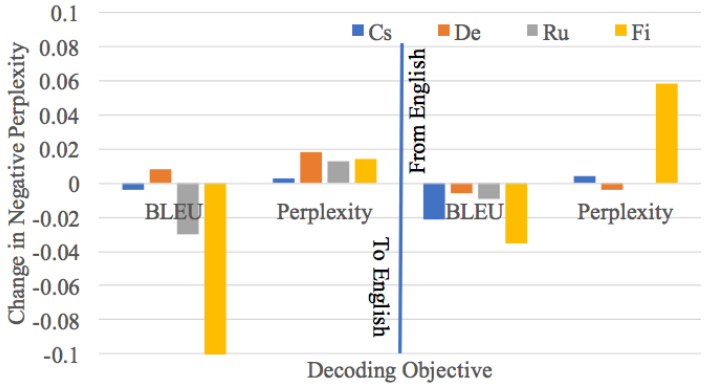
Experimental results





(a) Trainable Greedy Decoding





(b) Beam Search + Trainable Greedy Decoding

Examples

- the trainable greedy decoder focuses on fixing prepositions and removing any unnecessary symbol generation

```
(a) S:  Главное зеркало инфракрасного космического телескопа имеет диаметр 6,5 метров
    T:  The primary mirror of the infrared space telescope has a diameter of 6.5 metres .
    G:  The main mirror of the infrared spaceboard has a diameter 6.5 m .
    A:  The main mirror of the infrared space-type telescope has a diameter of 6.5 meters .

(b) S:  Еще один пункт - это дать им понять , что они должны вести себя онлайн так же , как делают это оффлайн .
    T:  Another point is to make them see that they must behave online as they do offline .
    G:  Another option is to give them a chance to behave online as well as do this offline .
    A:  Another option is to give them to know that they must behave online as well as offline .

(c) S:  Возможен ли долговременный мир между арабами и израильтянами на Ближнем Востоке ?
    T:  Can there ever be a lasting peace between Arabs and Jews in the Middle East ?
    G:  Can the Long-term Peace be Out of the Middle East ?
    A:  Can the Long-term Peace be between Arabs and Israelis in the Middle East ?
```

Thanks & QA