Motivations for RL in NLG

Ki Hyun Kim

nlp.with.deep.learning@gmail.com

Image Generations

- Objective:
  - 조건에 맞는 생생한(vivid) 이미지를 생성하는 것

- MSE Loss는 특성상 blurry한 결과물을 얻게 된다.

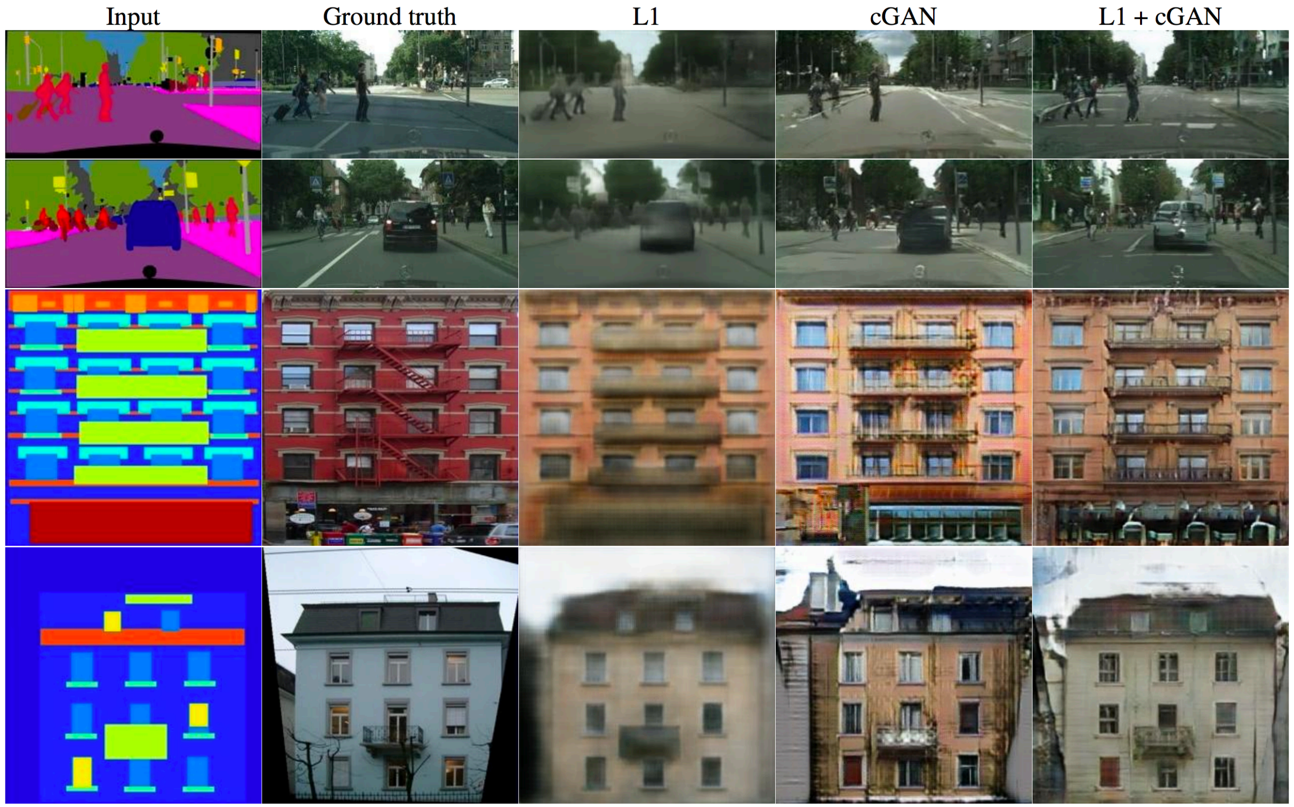


Figure 4: Different losses induce different quality of results. Each column shows results trained under a different loss. Please see `https://phillipi.github.io/pix2pix/` for additional examples.

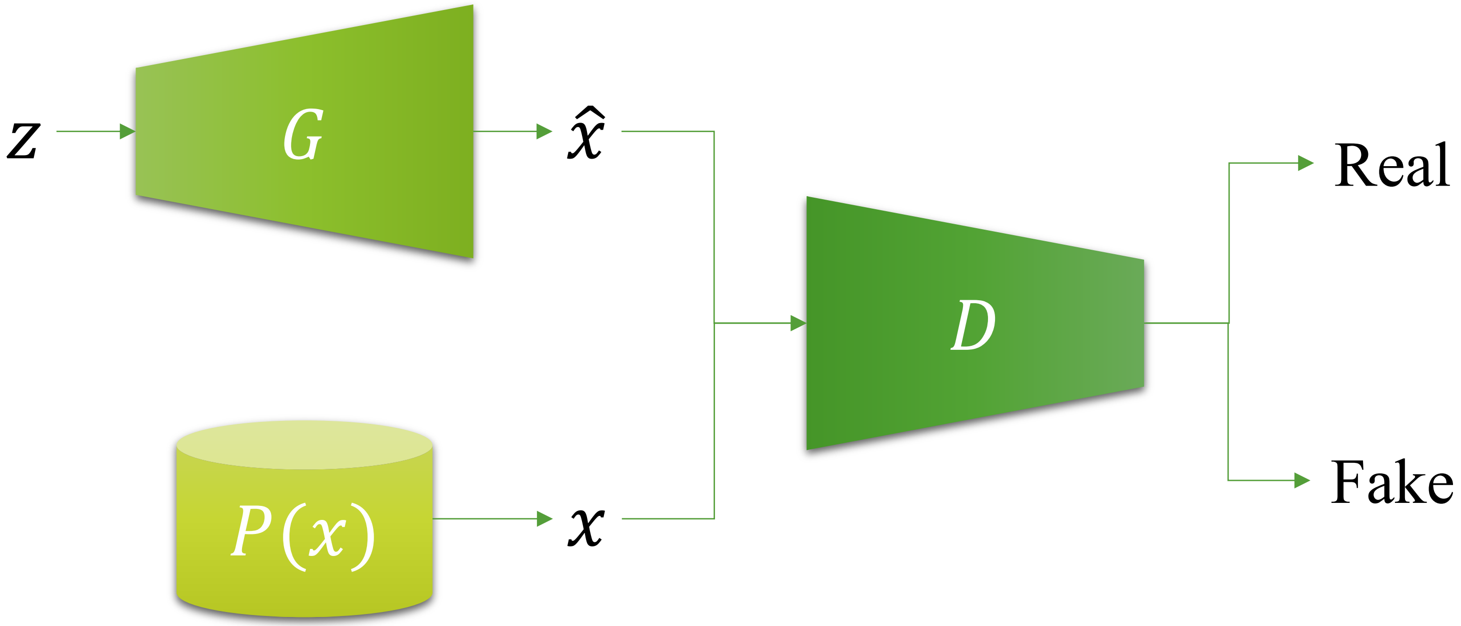
Image from [Isola et al., 2016]

Generative Adversarial Networks

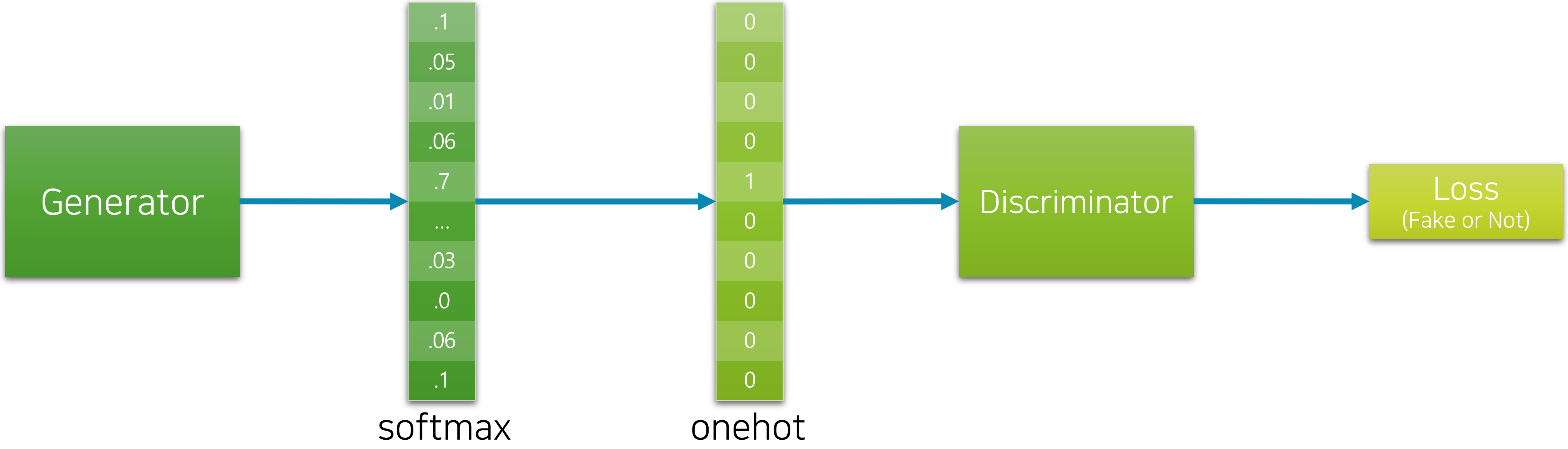
- Generator와 Discriminator가 서로 적대적으로 학습하며 발전해나감

$$\min_{G \in \mathcal{G}} \max_{D \in \mathcal{D}} \mathcal{L}(D, G) = \mathbb{E}_{x \sim p_{\text{real}}(\mathrm{x})} \left[\log D(x) \right] + \mathbb{E}_{z \sim p_z(\mathrm{z})} \left[\log \left(1 - D \circ G(z)\right) \right],$$

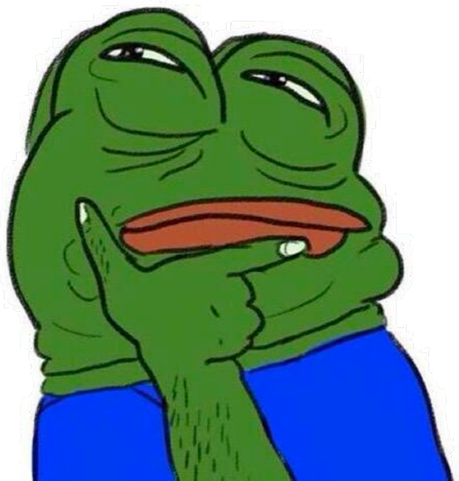
where G is generator and D is discriminator.



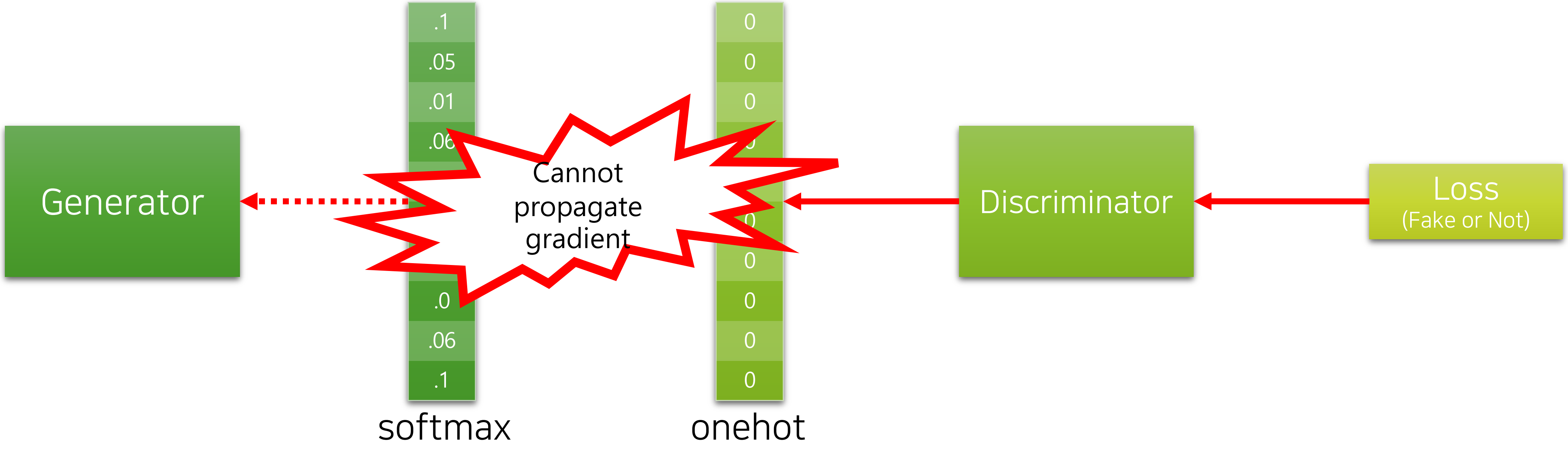
Apply GAN to Natural Language Generations



- Discriminator를 통해 더 진짜 같은 문장을 만들 수 없을까?



Why NLG cannot use GAN?



- Softmax → One-hot 과정은 gradient 전달이 불가능함

Anyway, we need another method…

- GAN propose a way to generate better image rather than using MSE.
  - MSE cannot reflect complex objective.

- NLG also needs another solution:
  - Perplexity cannot reflect NLG quality.
  - Teacher Forcing cause discrepancy between training and inference.

Perplexity cannot Reflect Generation Quality

Perplexity (Cross Entropy)

- Time-step 별 헷갈리는 평균 단어 수
  - 어순 고려 안됨

$$
\begin{aligned}
\text{PPL}(x_1, \cdots, x_n; \theta) &= P(x_1, \cdots, x_n; \theta)^{-\frac{1}{n}} \\
&= \sqrt[n]{\frac{1}{P(x_1, \cdots, x_n; \theta)}} \\
&= \sqrt[n]{\frac{1}{\prod_{i=1}^{n} P(x_i | x_{<i}; \theta)}}
\end{aligned}
$$

BLEU

- 각 n-gram 별 precision의 가중 평균

$$
\text{BLEU}(\hat{y}, y) = \text{brevity\_penalty}(\hat{y}, y) \times \prod_{n=1}^{N} p_n^{w_n},
$$

$$
\text{where brevity\_penalty}(\hat{y}, y) = \min\left(1, \frac{|\hat{y}|}{|y|}\right)
$$

$$
\text{and } p_n^{w_n} \text{ is precision of } n\text{-gram with weight } w_n = \frac{1}{2^n}.
$$

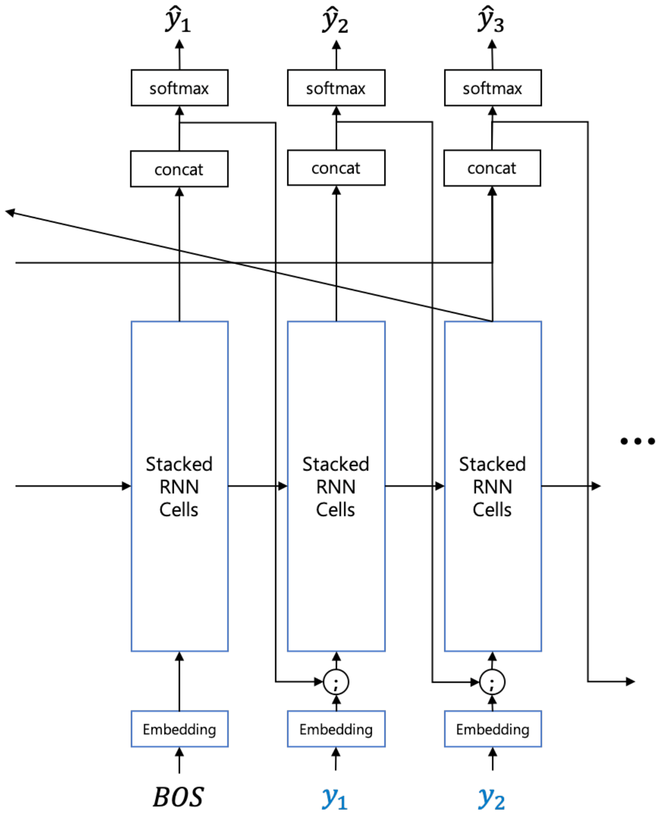
Discrepancy caused by Teacher Forcing

Teacher Forcing

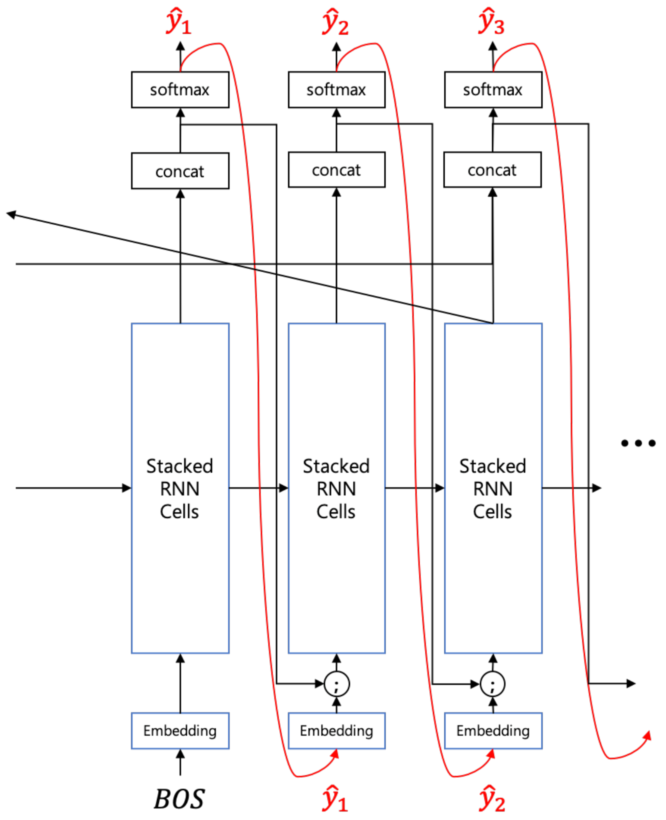
- 실제 정답을 다음 time-step의 입력으로 넣어줌

Sampling based Generations

- 추론시에는 예측값을 다음 time-step의 입력으로 넣어줌



Teacher-Forcing



Sampling

Summary

- 생생한 이미지를 만들어내고자 하는 복잡한 objective는 MSE로 표현할 수 없음
  - GAN의 discriminator가 복잡한 objective를 대체

- 번역 품질과 같은 복잡한 objective는 Cross Entropy (PPL)로 표현할 수 없음
  - BLEU는 미분이 불가능하므로, loss함수로 부적합

- RL을 통해 BLEU에 대해 미분 없이 학습 가능
  - 또한 sampling 기반 방식이므로, teacher forcing 없이 학습 가능