

A Dual Reinforcement Learning Framework for Unsupervised Text Style Transfer

Fuli Luo, Peng Li, Jie Zhou, Pengcheng Yang, Baobao Chang, Zhifang Sui, Xu Sun.

Task Description

Aims: transfer style but keep content

- Informal => formal

Gotta see both sides of the story => You have to consider both sides of the story

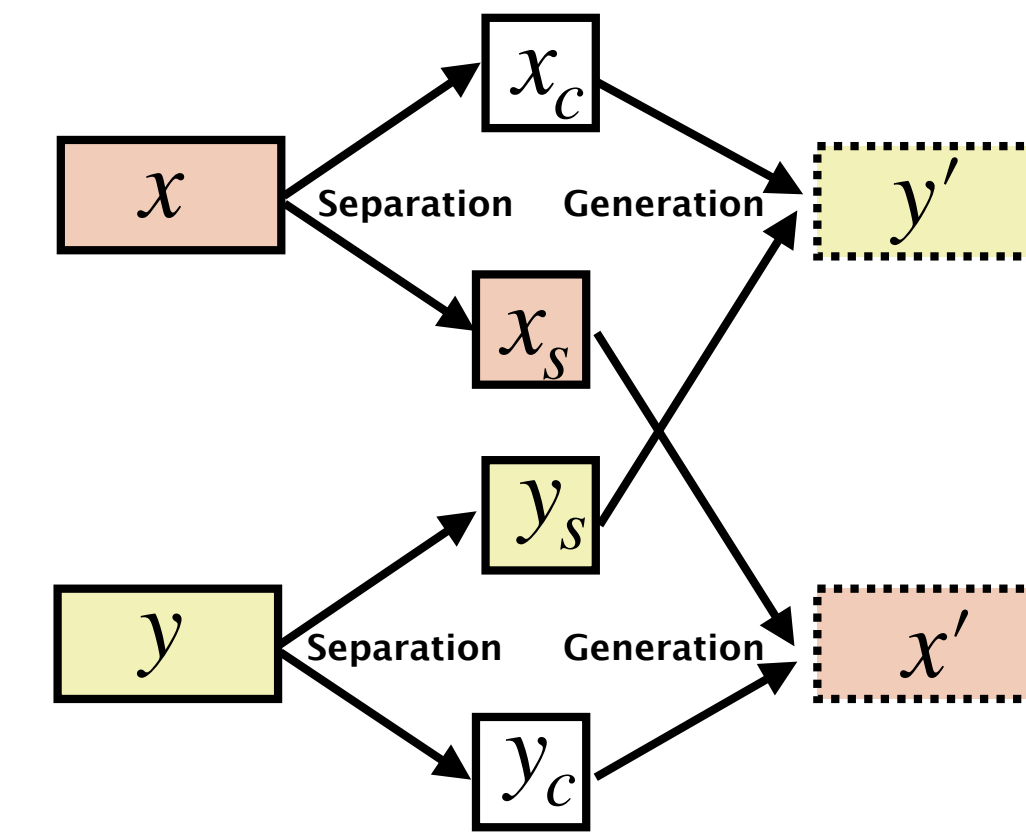
- Negative => positive

Love this restaurant and will keep coming back => Hate this restaurant and will not come back

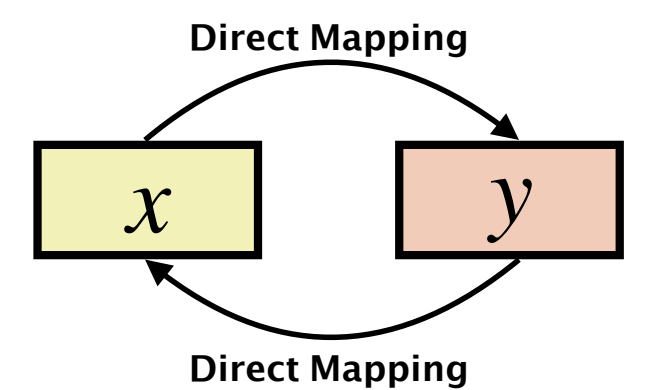
Challenge: lack parallel data

(Aligned sentences with the same content but different style)

Motivation

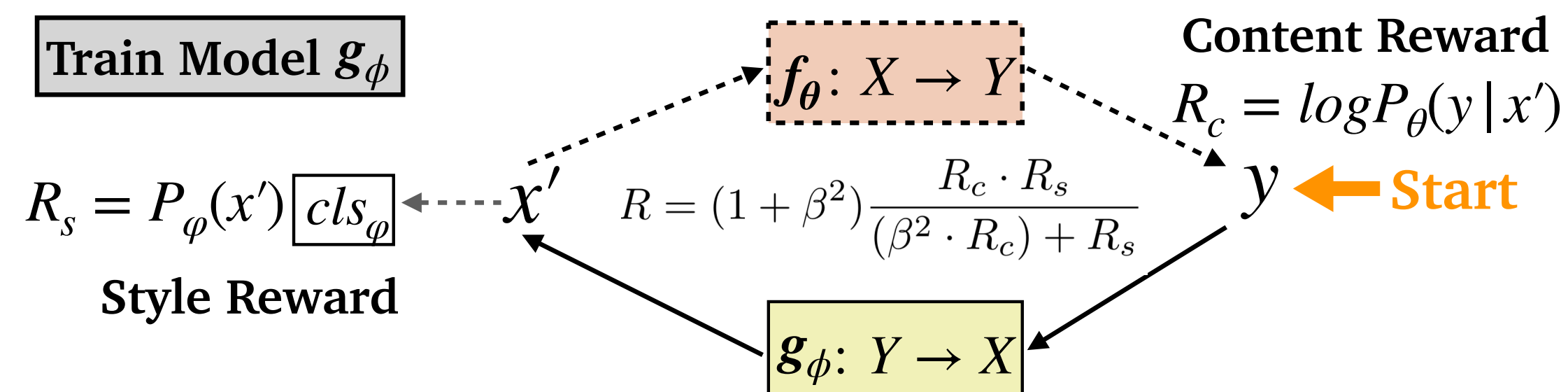
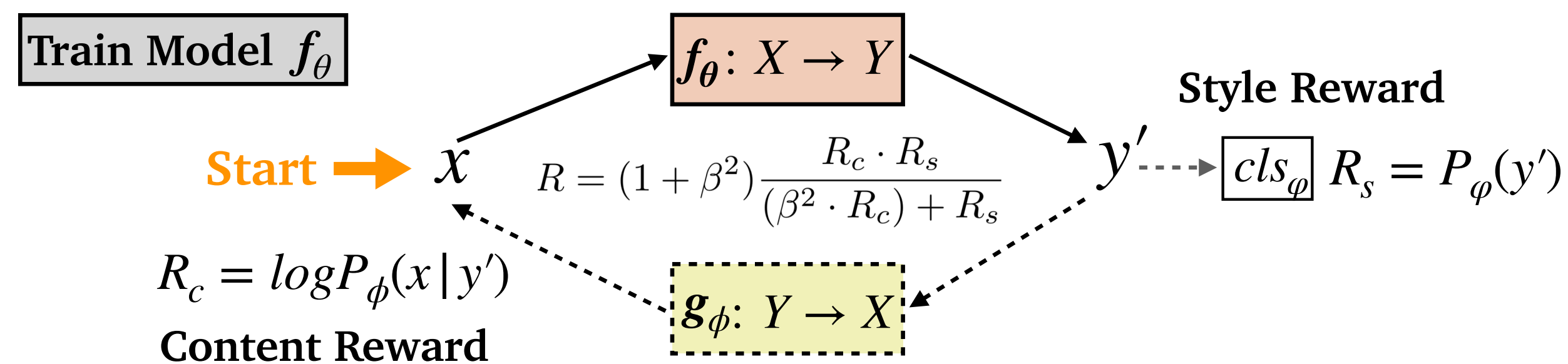


Previous Work



This Work

Model: Dual Reinforcement Learning (DualRL)



1. RL training: Policy Gradient

$$\begin{aligned} \nabla_{\theta} \mathbb{E}[R] &= \nabla_{\theta} \sum_k P(\mathbf{y}'_k | \mathbf{x}; \theta) R_k \\ &= \sum_k P(\mathbf{y}'_k | \mathbf{x}; \theta) R_k \nabla_{\theta} \log(P(\mathbf{y}'_k | \mathbf{x}; \theta)) \\ &\simeq \frac{1}{K} \sum_{k=1}^K R_k \nabla_{\theta} \log(P(\mathbf{y}'_k | \mathbf{x}; \theta)) \end{aligned}$$

2. Teacher-Forcing: Back-translation

$$\nabla_{\theta} \mathcal{J}(\theta) = \nabla_{\theta} \log(p(\mathbf{x} | S(\mathbf{x}, v_y; \theta), v_x; \theta))$$

$$\text{Annealing: } p = \min(p_0 \times r^{\frac{i}{d}}, p_{max})$$