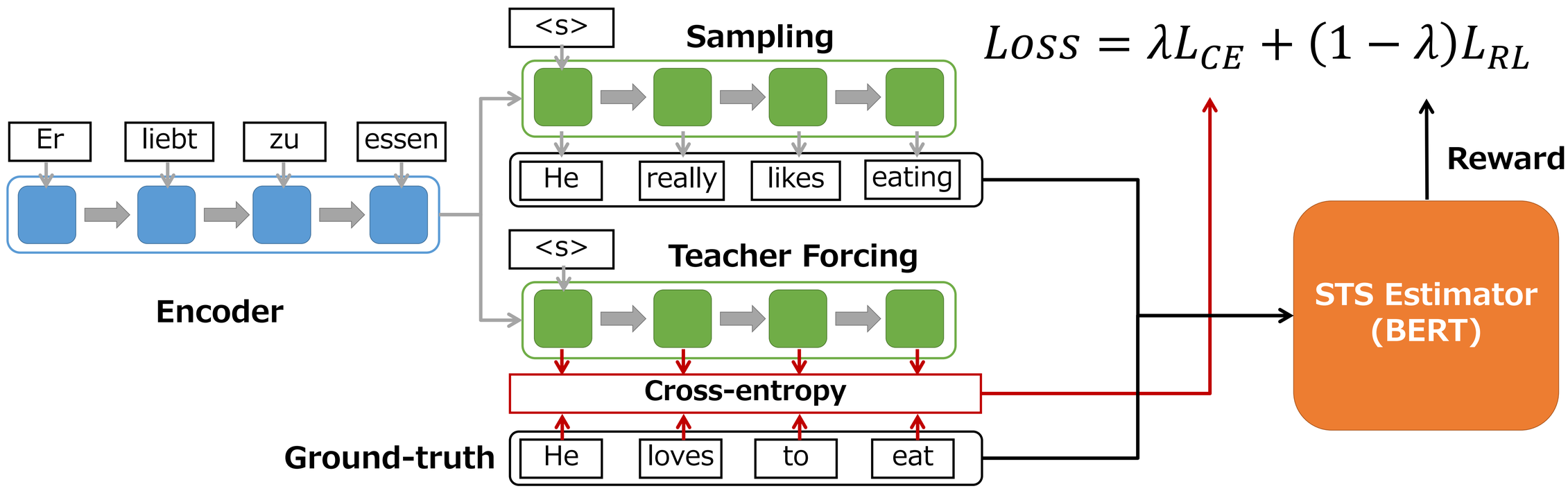
Using Semantic Similarity as Reward for Reinforcement Learning in Sentence Generation

Go Yasui[1], Yoshimasa Tsuruoka[1], Masaaki Nagata[2]
[1]The University of Tokyo, [2]NTT Communication Science Laboratories

Model Overview



Background

- Cross-entropy loss only evaluates sentences on the token-level and is unable to handle synonyms or changes in sentence structure
- Preferable to evaluate output sentences with more flexible criteria such as their Semantic Textual Similarity (STS) with ground-truth sentences
- Reinforcement Learning (RL) with estimated STS scores as reward

Related Research

Semantic Textual Similarity (STS) [Cer et al. 2017]

- Task of estimating a similarity of two given sentences on a scale of 0 (completely different) to 5 (completely equivalent)
- Estimated STS scores are continuous

Example of STS scores

Score	Sentence Pair
2.8	A man is playing a guitar. A girl is playing a guitar.
4.2	A panda bear is eating some bamboo. A panda is eating bamboo.

Sequence-level Training of RNN Models [Ranzato et al., 2016]

- Train Encoder-Decoder models using sequence-level evaluations (BLEU, ROUGE) with RL
- Sequence-level metrics contribute to the loss function of REINFORCE [Williams, 1992] as rewards

RL using STS scores

- Prepare the STS estimator by finetuning BERT [Delvin et al., 2018] to STS dataset
- Pretrain an Attention-based LSTM Encoder-Decoder model with teacher forcing
- Further train the sentence generation model with REINFORCE using estimated STS scores as rewards
- Predict reward from decoder hidden state and use the prediction to reduce reward variance

Experiment

Corpora:

- STS-B (En; 5.7k sentences)
- Multi30k-dataset (De-En; 30k sentence pairs)
- WIT3 (De-En; 200k sentence pairs)

Models:

- Baseline: trained using only cross-entropy loss
- RL-GLEU: trained with RL using GLEU scores
- RL-STS: trained with RL using estimated STS scores

Results: BLEU scores and estimated STS scores

	mscoco2017		flickr2017		TED2014		TED2015	
Model	BLEU	STS	BLEU	STS	BLEU	STS	BLEU	STS
Cross-entropy	16.44	2.76	22.22	3.03	12.54	2.63	13.43	2.80
RL-GLEU	20.13	2.93	25.83	3.15	13.97	2.71	14.59	2.89
RL-STS	18.31	2.96	24.70	3.21	13.58	2.87	14.56	2.99

Results: sample outputs

Model	Output Sentences	
Ground-truth	I'll show you what I mean.	So how do we solve?
Cross-entropy	I'll show you what I mean.	So how do we solve?
RL-GLEU	I'll show you what I mean.	So how do we solve?
RL-STS	I'm going to show you what I mean.	So how do we solve problems?

Discussion

- RL-STS has **better** BLEU scores than Cross-entropy, but is **not as good** as RL-GLEU
- Some outputs from RL-STS did not terminate; unable to account for EOS token with BERT
- Differences in outputs are **not favoured** by token-based matching and demonstrates leniency of STS evaluation
- Further evaluation is necessary (human evaluation, self-BLEU, etc.)
- Possible use of other semantic inference tasks (XNLI, paraphrasing, etc.)

References

- Cer et al. SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation. In *SemEval*, 2017.
- Ranzato et al. Sequence Level Training with Recurrent Neural Networks. In *ICLR*, 2016.
- Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs],* 2018.
- Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3), 1992.