School of Engineering
Ming Hsieh Department of Electrical and Computer Engineering

End-to-End Semantic Communication Systems with Pre-Trained Language Model

Ju-Hyung Lee, Dong-Ho Lee, Eunsoo Sheen, Thomas Choi, Jay Pujara

Idea

Keyword: physical-layer (PHY) communications, link-level simulation, natural language processing (NLP), language model

- This work proposes a practical semantic communication system called seq2seq-SC, which integrates a pre-trained language model with a 5G-NR compatible link-level simulator.
- seq2seq-SC outperforms previous models in extracting semantically meaningful information, as measured by lexical similarity (BLEU) and semantic similarity (SBERT).
- This study paves the way for the integrating foundation model, such as large language model (LLM) into future wireless systems in 6G networks.

Contribution

Q 1: *How to jointly design a language model and communication system?*

Q 2: *How to measure semantic error (similarity) between transmitted and received sentences?*

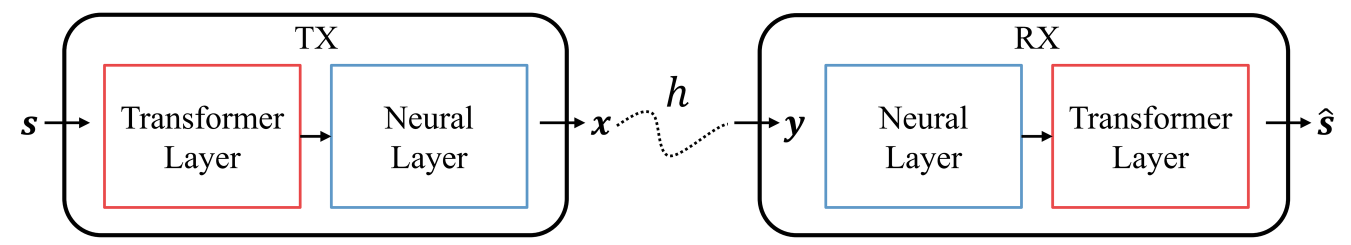
Our main contributions, which address these questions, are summarized as follows:

- We employ the pre-trained language model (`BART`) in the 5G-NR compatible link-level simulator (`NVIDIA Sionna`) to validate semantic network performance in real-world settings, answering **Q 1**.
- The proposed seq2seq-SC system efficiently extracts the semantic (meaningful) information with reduced computation effort. This network is "generalized", meaning it works with all (general) text corpus.
- To answer **Q 2** and evaluate semantic network performance in a *semantic way*, we introduce a metric called semantic similarity. The network is flexible to prioritize either perfect message delivery or semantic similarity, depending on the communication scenario.

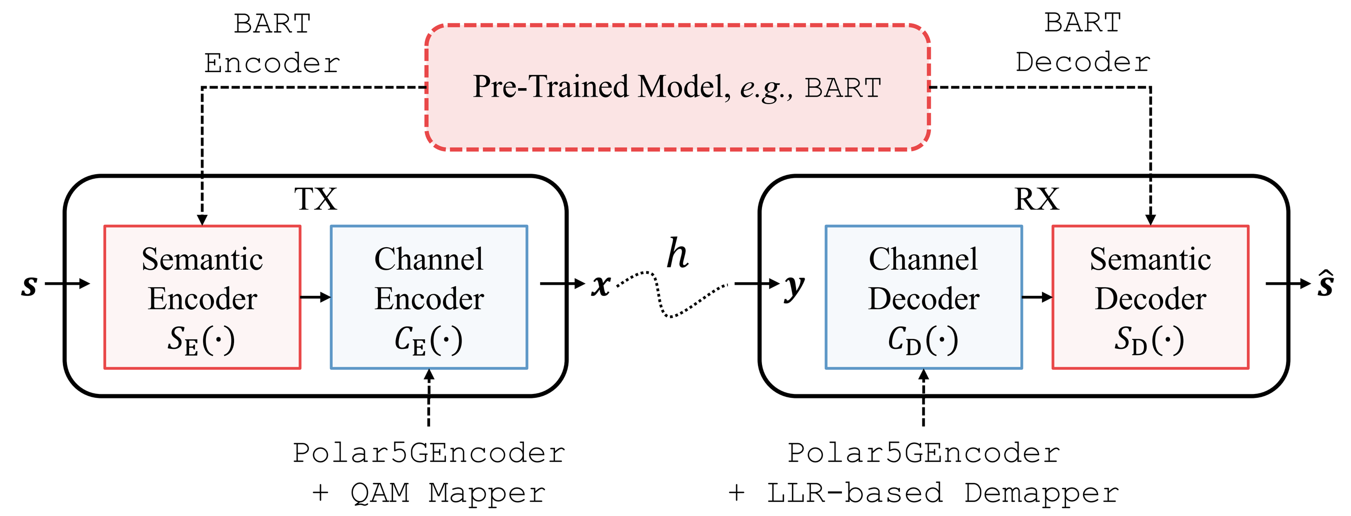
Pre-Trained Language Model

- Contextualized embeddings from pre-trained transformers are effective in providing initial representations for downstream tasks, such as semantic communication.
- Recent studies exploit transformers for text semantic communication, but training is computationally expensive and handling out-of-vocabulary (OOV) words is difficult.
- We use a pre-trained encoder-decoder transformer to initialize the parameters of the semantic encoder/decoder, reducing computational effort, improving OOV handling, and generalizing to any other text.

seq2seq-SC



(a) Conventional (DeepSC)



(b) Proposed (seq2seq-SC)

Problem Description / Architecture

Consider a sentence $\boldsymbol{s}$ that maps to symbol stream $\boldsymbol{x}$:

$$\boldsymbol{x} = C_{\mathrm{E}}\left(S_{\mathrm{E}}\left(\boldsymbol{s}\right)\right), \tag{1}$$

where $C_{\mathrm{E}}(\cdot)$ and $S_{\mathrm{E}}(\cdot)$ represent the channel encoder and the semantic encoder, respectively. This symbol stream passes through a physical channel, h, with flat fading and noise in the RF front end of a RX; which is expressed by the received signal, $\boldsymbol{y}$:

$$\boldsymbol{y} = h\boldsymbol{x} + \boldsymbol{n}, \tag{2}$$

Here, the encoded signal by TX propagates over the Rayleigh fading channel with $\mathcal{CN}(0,1)$; RX receives the attenuated signal with $\boldsymbol{n} \sim \mathcal{CN}\left(0, \sigma_n^2\right)$. Then, $\boldsymbol{y}$ is decoded at the RX to estimate the sentence $\hat{\boldsymbol{s}}$:

$$\hat{\boldsymbol{s}} = S_{\mathrm{D}}\left(C_{\mathrm{D}}\left(\boldsymbol{y}\right)\right), \tag{3}$$

where $S_{\mathrm{D}}(\cdot)$ and $C_{\mathrm{D}}(\cdot)$ represent the semantic decoder and the channel decoder.

Goal: Minimize semantic errors while minimizing transmission overhead.

Datasets

- We train our model on the *European Parliament* dataset, with the input and output sentences being the same (*i.e.*, $\boldsymbol{s} \to \boldsymbol{s}$).
- We evaluate our model on the *Flickr* image-caption dataset, which is not seen during training, to test its generalizability and superiority. Here, the output sentence can be different from the input sentence, but the semantic meaning of the two sentences is the same (*i.e.*, $\boldsymbol{s} \to \boldsymbol{s}'$).

Experiments

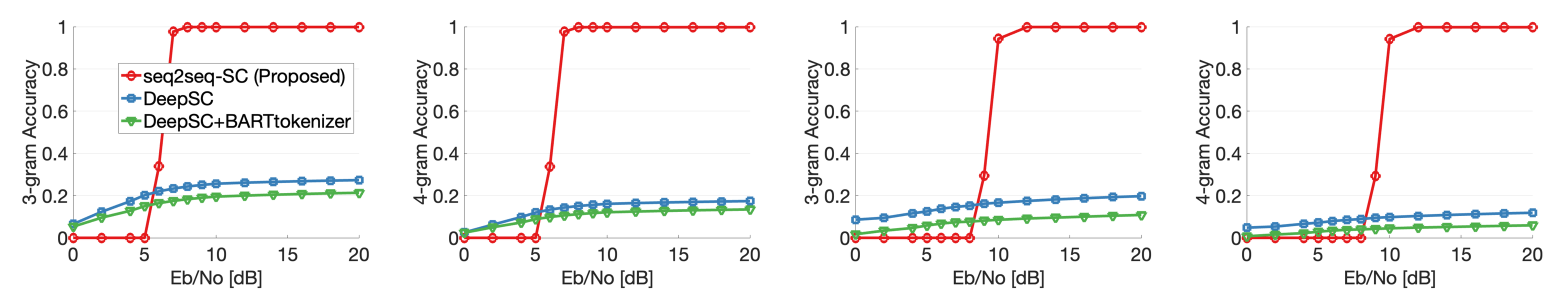
Evaluation / Setup

- Traditional communication performance measures focus on bit or symbol error rates (BER/SER), while semantic communication focuses on *meaningful* information.
- We use BLEU for lexical similarity and SBERT for semantic similarity, both of which range $0 \sim 1$, with higher scores indicating better similarity.
- BLEU computes the average of the n-gram precision scores between the input and output sentences.
- SBERT represents sentences into embeddings and computes the cosine similarity between them. For example, *"child"* and *"children"* are semantically related, but the BLEU computed lexical similarity is zero.

Type (or parameter)	Value
Channel en/decoder	Polar coding
# of information bits	512
# of codeword bits	1024
Coderate	0.5
Mapper/Demapper constellation	16-QAM
# of bit per symbol	4
Demapping method	Log-likelihood ratios
Channel	AWGN, Rayleigh fading

Simulation Setup.

BLEU score over E_b/N_o [dB]



Output Example

Transmit $\boldsymbol{s}$	An older dog and a younger one playing with a toy.		
Train	**Receive $\hat{\boldsymbol{s}}$**	**BLEU**	**SBERT**
$\boldsymbol{s} \to \boldsymbol{s}$	An older dog and a younger one playing with a toy.	1.0	1.0
$\boldsymbol{s} \to \boldsymbol{s}'$	Two dogs are playing with a toy.	0.210	0.820

Semantic Similarity

Train	Lexical Similarity	Semantic Similarity
	BLEU	SBERT
$\boldsymbol{s} \to \boldsymbol{s}$	0.993	0.999
$\boldsymbol{s} \to \boldsymbol{s}'$	0.173	0.764

Conclusions

- seq2seq-SC successfully decodes semantically meaningful message under noisy conditions and bit errors, thanks to its use of a pre-trained language model.
- seq2seq-SC can be expanded to other modalities, such as speeches and videos, for even greater communication efficiency.

Code at `github.com/abman23/seq2seq-SC`
Paper at `arxiv.org/abs/2210.15237`