# Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation

## Link

## Summary

- This paper introduces the RNN Encoder-Decoder neural network architecture. Here an encoder network maps variable length input sequence to a fixed length vector representation and a decoder network maps the vector back to a variable length target sequence. The two networks are trained jointly to maximize the conditional probability of the target sequence given a source sequence. The two sequence length may be different.

- A trained model can be used in two ways: a) use the model to generate a target sequence given a source sequence and b) score a given pair of source and target sequence i.e., calculate $p(target|source)$.

- They introduce a new type of hidden unit(GRU). It consists of two gates: update and reset gate. Reset gate is computed as

$$r = \sigma(W_r x + U_r h_{t-1})$$

  Update gate is computed as:

$$z = \sigma(W_z x + U_z h_{t-1})$$

  Final hidden state is then

$$h_t = z h_{t-1} + (1 - z)\tilde{h}_t$$

  where

$$\tilde{h}_t = \phi(W x + U(r \odot h_{t-1}))$$

  $tanh$ is used as $\phi$ in the paper.

  When reset gate is close to zero $h_t$ depends only on the current input. So it allows the network to drop information that is found to irrelevant later in the future. Update gate controls how much information will carry over from previous hidden state. If its one previous hidden state will be simply passed to current hidden state with no contribution from current input. When $r = 1$ and $z = 0$ $h = \phi(W x + U h_{t-1})$ which is just the vanilla RNN. The gates act on individual elements of hidden units so individual hidden units might have different values for update and reset gates. Units that capture short term dependencies will have reset gate more active, units that learn to capture longer term dependencies will have update gate more active. GRU is a simpler specialization of LSTM.

- The encoder decoder network was used to score phrase pairs for machine translation task (English to French) and used as an extra feature in statistical machine translation(SMT) system. This improved the translation performance in terms of BLEU score.

- The RNN Encoder-Decoder can capture both semantic and syntactic structures of the phrases. When used to generate phrase, it can generate well formed phrase that are not in the source phrase table.