

# Attention with GRU

---

EXPLORING COMPLEX REORDERING

# Motivation

---

1. In the equation  $e_{ij} = v_a^T \tanh(W_a s_{i-1} + U_a h_j)$  (Bahdanau et al., 2015), the alignment score only takes input of the previous target-side encoded vector  $s_{i-1}$  and the source-side vector  $h_j$ .
2. The previous alignment scores might also be useful, especially in a long sentence (e.g. more than 50 words).

# Attention with GRU

---

$$e_{ij} = v_a^T \tanh(v_{i,j})$$

$$v_{i,j} = (1 - z_i) \circ v_{i-1,j} + z_i \circ \bar{v}_{i,j}$$

$$\bar{v}_{i,j} = \tanh(Ws_{i-1} + U(r_i \circ v_{i-1,j}) + Qh_j)$$

$$z_{i,j} = \sigma(W_z s_{i-1} + U_z v_{i-1,j} + Q_z h_j)$$

$$r_{i,j} = \sigma(W_r s_{i-1} + U_r v_{i-1,j} + Q_r h_j)$$

*Original:*

$$e_{ij} = v_a^T \tanh(W_a s_{i-1} + U_a h_j)$$

# Attention with GRU

---

$$e_{ij} = v_a^T \tanh(v_{i,j})$$

$$v_{i,j} = (1 - z_i) \circ v_{i-1,j} + z_i \circ \bar{v}_{i,j}$$

$$\bar{v}_{i,j} = \tanh(Ws_{i-1} + U(r_i \circ v_{i-1}) + Qh_j)$$

$$z_{i,j} = \sigma(W_z s_{i-1} + U_z v_{i-1} + Q_z h_j)$$

$$r_{i,j} = \sigma(W_r s_{i-1} + U_r v_{i-1} + Q_r h_j)$$

*Original:*

$$e_{ij} = v_a^T \tanh(W_a s_{i-1} + U_a h_j)$$

# Training Using Known Alignment

---

1. In the equation  $c_i = \sum_{j=1}^{T_x} a_{ij} h_j$  (Bahdanau et al., 2015), each  $a_{ij}$  can be an output which equals to 1 if word  $i$  is aligned to word  $j$  in the known alignment.
2. The alignment model can be obtained using SMT system.

# Thanks

---

Cho, K., van Merriënboer, B., Gulcehre, C., Bougares, F., Schwenk, H., and Bengio, Y. (2014a). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the Empirical Methods in Natural Language Processing (EMNLP 2014)*. To appear.

Sutskever, I., Vinyals, O., and Le, Q. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems (NIPS 2014)*.

Bahdanau, D., Cho, K., and Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *International Conference on Learning Representations (ICLR 2015)*.