

METL: Méthodes Statistiques de Traduction Automatique

3C124

Paola Merlo

année académique 20019-2020

Questions on Neural Machine Translation.

1. Does a NMT architecture conform to the noisy channel model? (slide 26)
 - (a) if yes, what parts of the architecture correspond to the translation model and what parts to the language models?
 - (b) if no, which equation describes the machine translation process in a NMT, (a) or (b)? Explain.
(a) $best_e = \operatorname{argmax} P(e|f)$ (b) $best_e = \operatorname{argmax} P(f|e)P(e)$.
2. The two components of a seq2seq model are the encoder and the decoder.
 - (a) They can both be RNNs but how are they different at test time? (slide 24)
 - i. Can you explain what the encoder does at test time? what is its input, what is its output overall and at each step?
 - ii. Can you explain what the decoder does at test time? what is its input, what is its output overall and at each step?
 - (b) The two components of a seq2seq model are the encoder and the decoder. They can both be RNNs. How do we train them?
 - i. Can you explain what the encoder does at training time? what is its input, what is its output overall and at each step?
 - ii. Can you explain what the decoder does at training time? what is its input, what is its output overall and at each step?
3. Why are these seq2seq systems called end-to-end? what operation is done 'end-to-end' and what are the two 'ends'?
4. Decoding (slide 30 and following)
 - (a) What is greedy decoding?
 - (b) What is beam decoding?
 - (c) Are either beam decoding or greedy decoding guaranteed to find the best (most probable) translation?
5. Advantages and disadvantages of NMT (slides 47-48)
 - (a) Choose two of each and explain.
6. Skip slides 49-50
7. Attention
 - (a) Motivation for attention : what is the bottleneck problem of the encoded-decoder architecture in NMT? (slide 58)
 - (b) How do you compute the attention scores?
 - (c) It can be claimed that attention produces a kind of soft alignment : why? (slide 68)
 - (d) What are the advantages of attention? (slide 74)