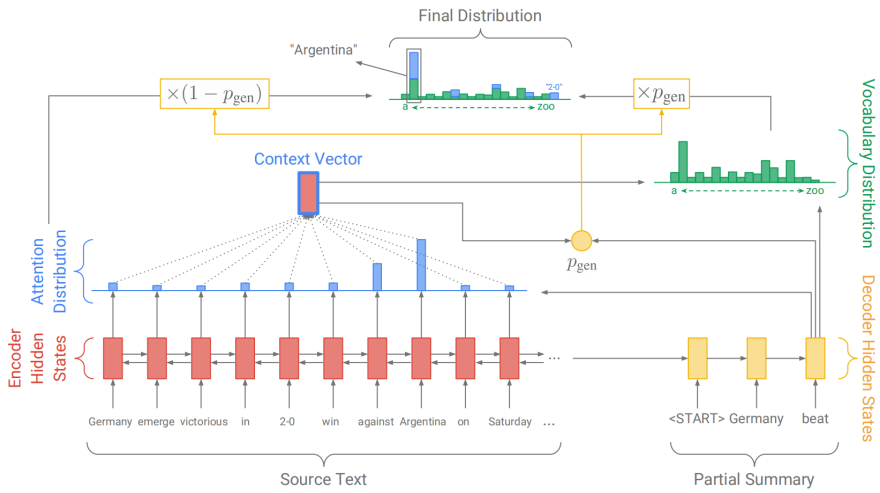
Report 2

> 2023.2.27-2023.3.5

Get To The Point: Summarization with Pointer-Generator Networks

1. Improvements to 2 shortcomes of neural sequence-tosequence models
    - Reproduce inaccurate details - A hybrid pointer-generator network
    - Repeat themselves - $courage$ to keep track of what has been summarized
2. A hybrid pointer-generator network



> 本质上是带attention的seq2seq在最后一步有一定概率选取文本里有词典里没有的词加到概率分布当中

3. Courage mechanism

    $c^t = \sum_{t'=0}^{t-1} a^{t'}$, c is a distribution over the source document words that represents the degree of courage that those words have received from the attention mechanism so far.

    $e_i^t = v^T \tanh(W_h h_i + W_s s_t + w_c c_i^t + b_{attn})$, changes the way to compute e_i^t. In attention mechanism, $e_i^t = v^T tanh(W_h h_i + W_s s_t + b_{attn})$

4. loss

    $\text{loss}_t = -\log P(w_t^*) + \lambda \sum_i \min(a_i^t, c_i^t)$, the latter part is added to penalize repeatedly atending to the same locations.

Closed-Book Training to Improve Summarization Encoder Memory

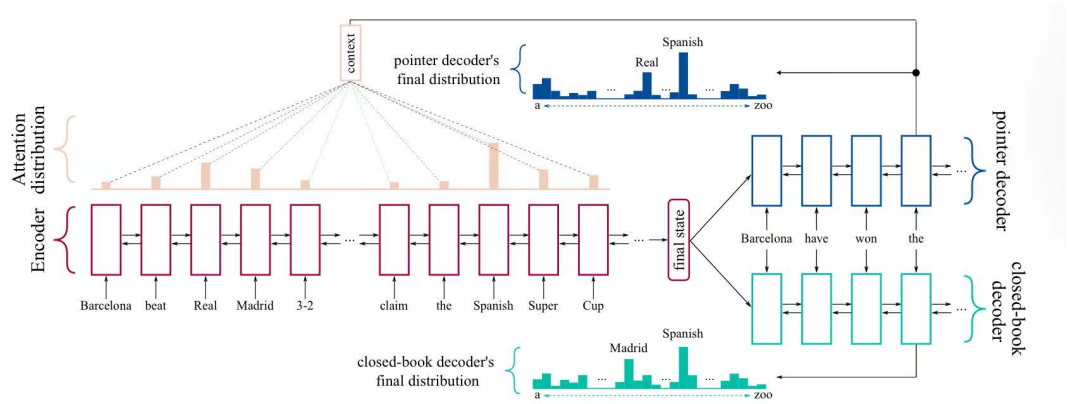


Figure 2: Our 2-decoder summarization model with a pointer decoder and a closed-book decoder, both sharing a single encoder (this is during training; next, at inference time, we only employ the memory-enhanced encoder and the pointer decoder).

> 这个模型是在上一个的基础上改进的。

1. Close-Book Decoder

   上一篇论文的问题在于c_t 可能包含太多不重要信息。本篇的改进就是加一个Close-Book Decoder to enhance encoder's memory. 相应的损失函数变成:

   $$L_{XE} = \frac{1}{T}\sum_{t=1}^{T} -((1-\gamma)\log P^t_{attn}(\omega|x_{1:t}) + \gamma\log P^t_{cbdec}(\omega|x_{1:t}))$$

   > 就是把两个结果加权一下

2. Reinforcement Learning

   套用了强化学习模板

   $$L_{RL} = \frac{1}{T}\sum_{t=1}^{T}(r(\hat{y}) - r(y^s))\log p^t_{attn}(\omega^s_{t+1}|\omega^s_{1:t})$$

   $$L_{XE+RL} = \lambda L_{RL} + (1-\lambda)L_{XE}$$

Review

seq2seq + attention NMT model

- [lstm](Understanding LSTM Networks -- colah's blog - step by step lstm walk through
- coding: cs224n a4 initial part