

Advanced NLP Assignment-3

Greeshma Amaraneni
2020101035

Analysis

1. Compare the model created by you and that of the original paper based on the ROUGE-L scores.

- Rouge-L score is the similarity between the candidate document and the reference similarity
- It is a metric used for evaluating automatic summarization texts generated.
- The Following the ROUGE-L scores obtained after evaluating the model,

		ROUGE-2			ROUGE-L	
Model	Recall	Precision	F1 Score	Recall	Precision	F1 Score
Pointer Generator Network	0.18	0.11	0.15	0.10	0.11	0.09
PGN+Coverage	0.21	0.19	0.15	0.11	0.17	0.12

- In the research paper “Get to the Point”, the ROUGE-L values are around 35 on average. I have implemented the Pointer Generator Network and added the coverage mechanism as well but my model doesn’t have high ROUGE-L scores like the one that is implemented, this is because number of epochs run by me are relatively less and I have used training data set smaller than the one implemented in the paper due to GPU restriction and the choice of my vocabulary is weaker than that of vocabulary used in the paper for the same reason of computational time. I have included attention only between the encoder and decoder unlike self attention included in the paper. So maybe because of these reasons my model was outperformed by the one implemented in paper.

2. Report the optimal hyperparameters.

Hyperparameters used :

Number of Epochs - 10

Loss Function - Cross Entropy Loss

Optimizer - Adam

Batch Size - 128

Learning Rate - 0.01
Embedding Size -50
Lambda (coverage loss parameter in loss function) - 2
Other trainable parameters include wh,ws,wc,b used for getting the attention weights

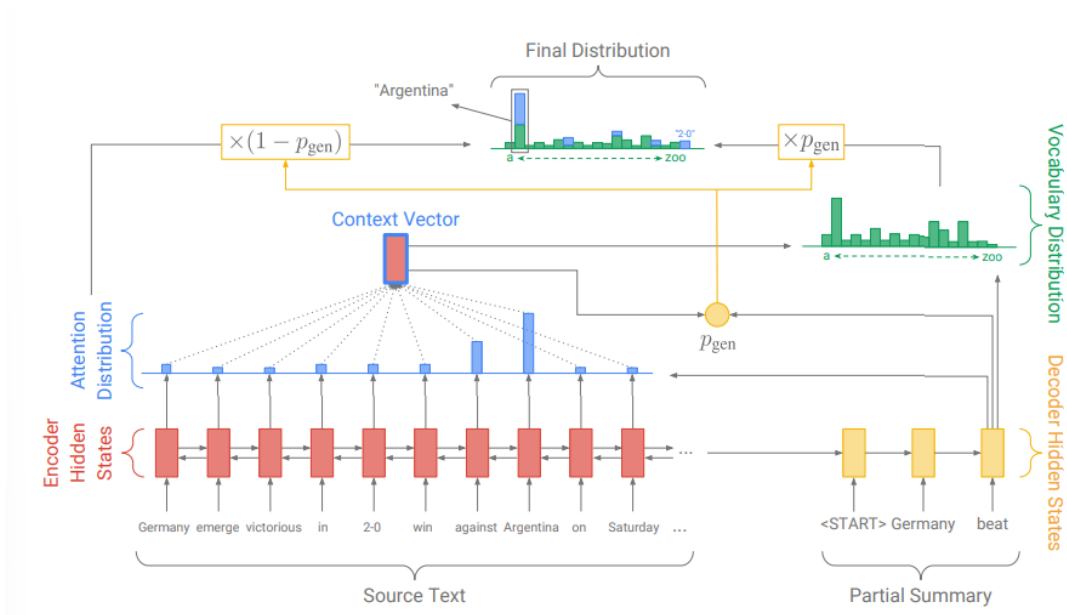
3.(a) How does a Pointer-Generator Model deal with OOV words? Describe what happens when an attention model sees OOV words.

Using the Baseline Sequence to Sequence plus the attention model makes the errors and some creates meaning less sentence. One important thing is it does not deal with the Out Of Vocabulary words (OOV). Using the sequence to sequence models in which we will use the recurrent neural networks has made the Abstraction possible but anyway these systems have the disability of dealing with OOV Words.

But where as the Pointer-generator-networks gives the facility to the model to copy the Words from the Source text by using pointing. This type of model is hybrid network between the baseline and the pointer network because it allows words copying from vocabulary and using pointing. Here the attention distribution is,

$$e_i^t = v^T \tanh(W_h h_i + W_s s_t + b_{\text{attn}})$$
$$a^t = \text{softmax}(e^t)$$

Here v,Wh,Ws and batten are the learnable parameters and the attention distribution for the pointer generator network is,



Also her P_{gen} is calculated using the context vector h_t^* and decoder state s_t and decoder input x_t . Then it is used to generate word from vocabulary by sampling from P_{vocab} or else by

Copying it from input sequence by sampling the attention distribution and we get the probability distribution as

$$P(w) = p_{\text{gen}} P_{\text{vocab}}(w) + (1 - p_{\text{gen}}) \sum_{i:w_i=w} a_i^t$$

Using the attention mechanism when it sees the Out of vocabulary word that is (w) then

$P_{\text{vocab}}(w) = 0$ then $p_{\text{gen}} P_{\text{vocab}}(w)$ will be 0.

And if w does not appear in the source document then right part that is , $\sum_{i:w_i=w} a_i^t$ will be zero. In this way it will distribute the probabilities for the out of vocabulary words when the attention model sees the OOV Words. In this way it has the ability to produce out of Vocabulary words.

3)b) What is the repetition problem in Seq-2-Seq models? How is it solved?

Because of using the baseline sequence to sequence plus attention mechanism model we got the factual errors and got the meaningless sentence and it does not deal with the OOV words and repetition. So We introduced the pointer-generator model which deals with the problem of OOV words and it is accurate enough in dealing with that and it produces meaning full sentences but the only drawback is that, there is a repetition problem in it that is It repeats itself. For example if we look at the example below.

(The below is considered from Get to the Point Research Paper)

<p>Original Text (truncated): lagos, nigeria (cnn) a day after winning nigeria's presidency, <i>muhammadu buhari</i> told cnn's christiane amannpour that he plans to aggressively fight corruption that has long plagued nigeria and go after the root of the nation's unrest. <i>buhari</i> said he'll "rapidly give attention" to curbing violence in the northeast part of nigeria, where the terrorist group boko haram operates. by cooperating with neighboring nations chad, cameroon and niger, he said his administration is confident it will be able to thwart criminals and others contributing to nigeria's instability. for the first time in nigeria's history, the opposition defeated the ruling party in democratic elections. <i>buhari</i> defeated incumbent goodluck jonathan by about 2 million votes, according to nigeria's independent national electoral commission. the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.</p>
<p>Baseline Seq2Seq + Attention: UNK UNK says his administration is confident it will be able to destabilize nigeria's economy. UNK says his administration is confident it will be able to thwart criminals and other nigerians. he says the country has long nigeria and nigeria's economy.</p>
<p>Pointer-Gen: muhammadu buhari says he plans to aggressively fight corruption in the northeast part of nigeria. he says he'll "rapidly give attention" to curbing violence in the northeast part of nigeria. he says his administration is confident it will be able to thwart criminals.</p>
<p>Pointer-Gen + Coverage: muhammadu buhari says he plans to aggressively fight corruption that has long plagued nigeria. he says his administration is confident it will be able to thwart criminals. the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.</p>

Using the pointer-generator network you can see the repetition of sentences in the summarization from the source text ,for this problem we use the Coverage mechanism that is we use coverage to keep track of what has been summarized, which discourages repetition.

In our coverage model, we maintain a coverage vector c^t (eq1), which is the sum of attention distributions over all previous decoder timesteps. c^t is a (unnormalized) distribution over the words in the source material that shows how much attention has been paid to those words thus far. The attention mechanism also receives information from the coverage vector. This makes sure that a recall of the attention mechanism's past decisions informs its present decision(eq2)—where to focus its attention next. This should make it simpler for the attention mechanism to stop attending to the same places repeatedly and stop producing repetitive text as a result. Since the summarization should not require uniform coverage, we only penalize the overlap between each attention distribution and coverage up to the point where repeated attention is prevented, which is why we also define a coverage loss to discourage repeatedly attending to the same locations.The primary loss function is then combined with the coverage loss, which has been reweighted by a hyperparameter, to create a new composite loss function(eq3).

eq(1)

$$c^t = \sum_{t'=0}^{t-1} a^{t'}$$

eq(2)

$$e_i^t = v^T \tanh(W_h h_i + W_s s_t + w_c c_i^t + b_{\text{attn}})$$

eq(3)

$$\text{loss}_t = -\log P(w_t^*) + \lambda \sum_i \min(a_i^t, c_i^t)$$