

Abstractive Text Summarization on the Gigaword Dataset

Peter Wang, Okonda Joseph, Robert Kinini
Stanford University

Abstract

Neural attentional sequence-to-sequence models hold the promise of being able to produce high quality extractive summaries. In this paper, we begin by training the standard seq2seq with attention model, extend it with ideas from Paulus et al, Nallapati et al, in order to explore the effect on model's performance. In particular, we implement intra-decoder attention to reduce repetitions. We also train a network modified to include a pointer generator network to allow the model to copy from the source sequence. We further develop a modified bottom-up abstractive summarization pipeline that is inspired by style transfer in computer vision. Finally, we train a model with hierarchical attention in order to model the source documents at both the word and sentence level.

Introduction

In our paper, we aim to produce intelligible, abstractive summaries of short news articles. We used the Gigaword dataset. The traditional deep learning approach to this task mirrors that of neural machine translation. In this approach, the problem is posed as a sequence-to-sequence problem, and an encoder/decoder network with attention is typically used to generate the summaries. We use this basic model as our baseline and then attempt to address some of its shortcomings by training/implementing (1) pointer generator network, (2) intra-decoder attention, (3) Hierarchical Attention, and (4) Bottom up summarization with modified neural style transfer

Baseline: Seq2seq with Attention

$$\text{score}(h_t^{<d>}, h_j^{<e>}) = h_t^T W_a h_j^{<e>} \quad (1)$$

$$a_t = \frac{\exp(\text{score}(h_t^{<d>}, h_j^{<e>}))}{\sum_i \exp(\text{score}(h_t^{<d>}, h_i^{<e>}))} \quad (2)$$

$$c_t = \frac{T}{\sum_i} (a_t)_i h_i^{<e>} \quad (3)$$

Intra-Decoder Attention

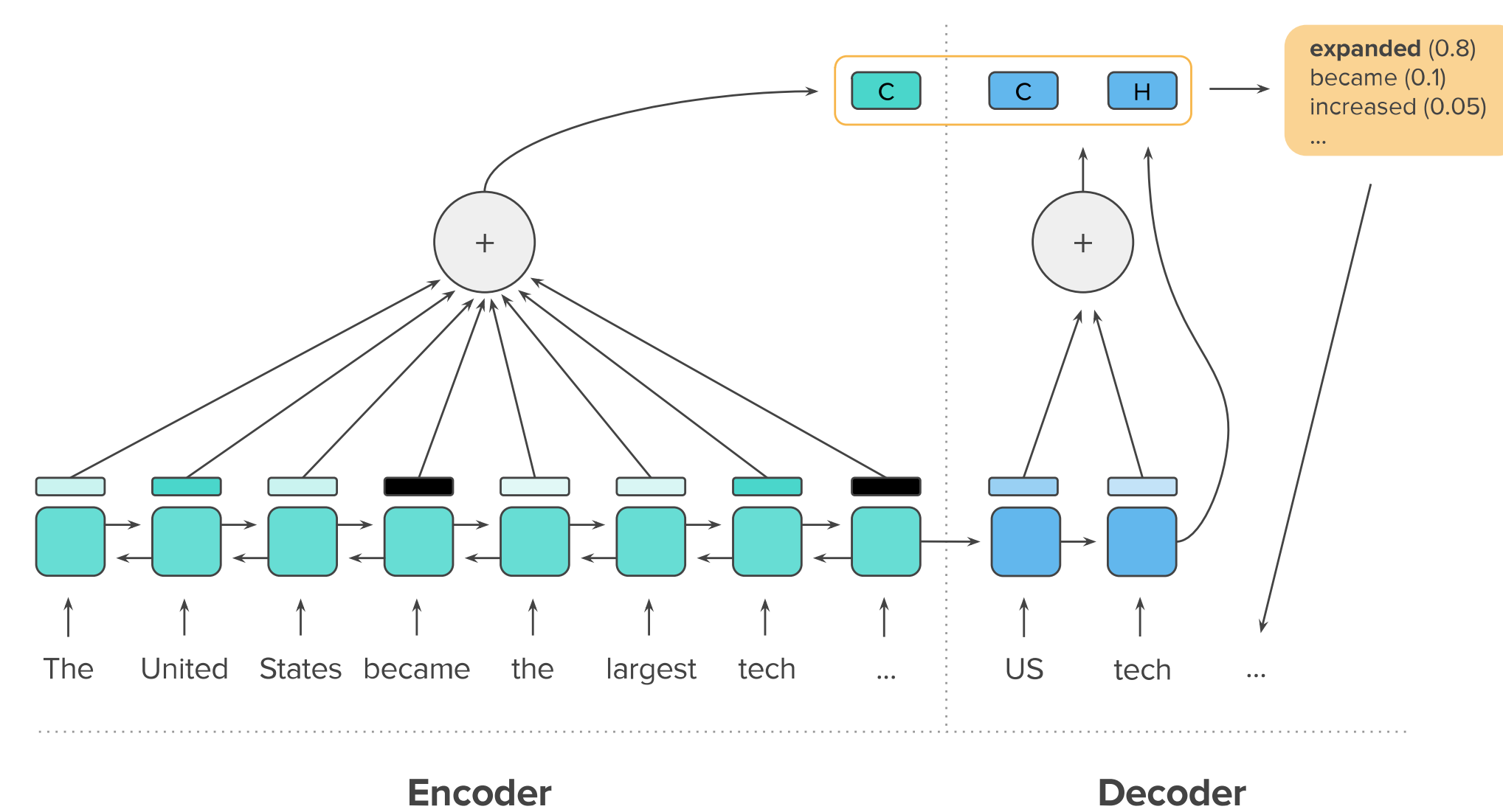


Figure 1: How intra-decoder attention fits into the baseline seq2seq model.

Pointer Generator Network

It helps solve the OOV and factual error problems. To do this, it is equipped with a switch that decides between using the generator or a pointer at every time step at the probability:

$$p_{gen} = \sigma(w_{h*}^T h_t^* + w_s^T s_t + w_x^T x_t + b_{ptr}) \quad (4)$$

where x_t is the decoder input, σ is the sigmoid function, vectors w_{h*} , w_s , w_x and scalar b_{ptr} are learnable parameters. The final distribution for the output word is:

$$p(w) = p_{gen} p_{vocab}(w) + (1 - p_{gen}) \sum_{i:w_i=w} a_i^t \quad (5)$$

For an OOV word, $p_{vocab} = 0$, so we end up pointing to source document.

Bottom up Summarization

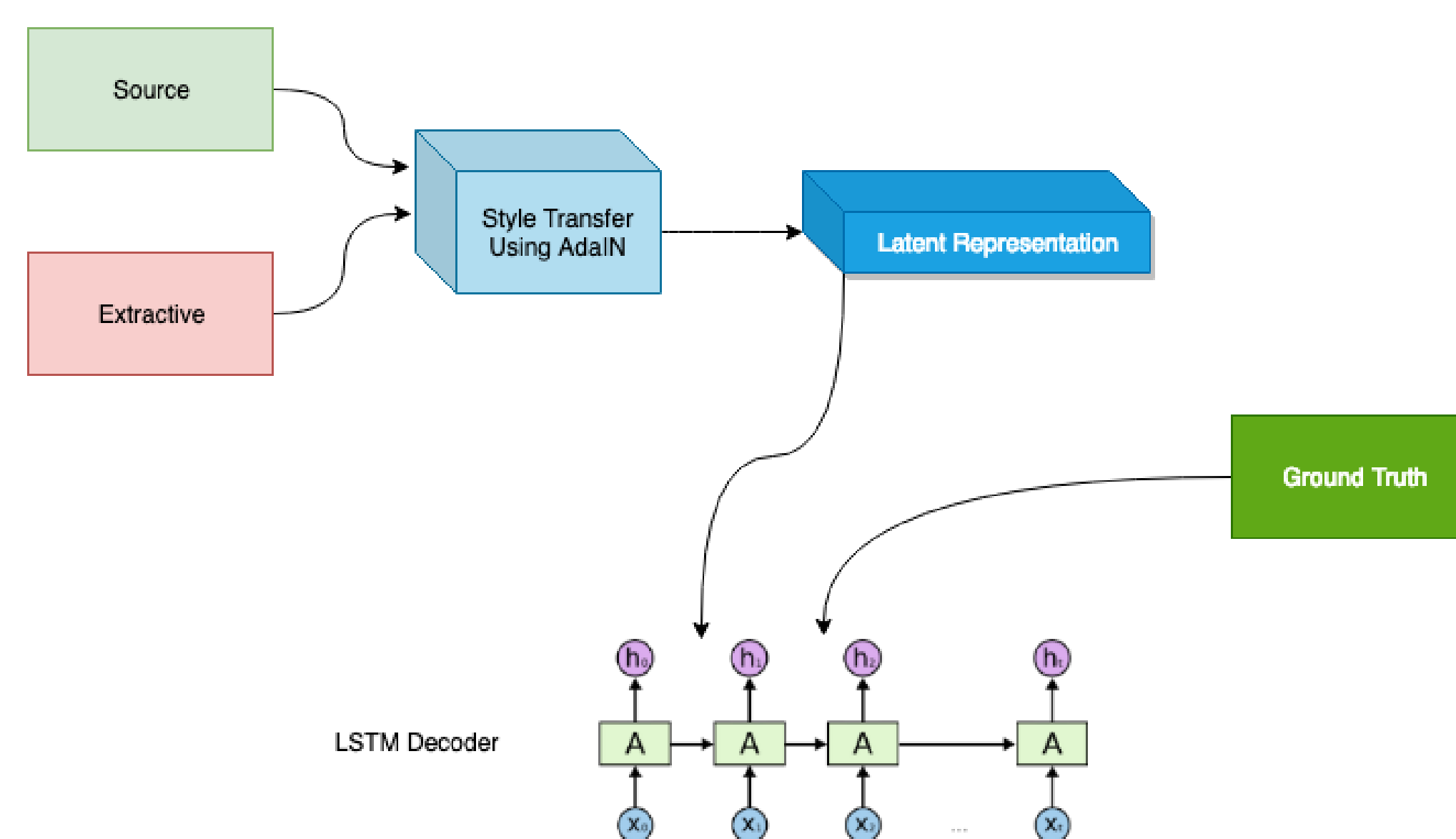


Figure 2: A flow Diagram Showing how our Bottom Up model works

Hierarchical Attention

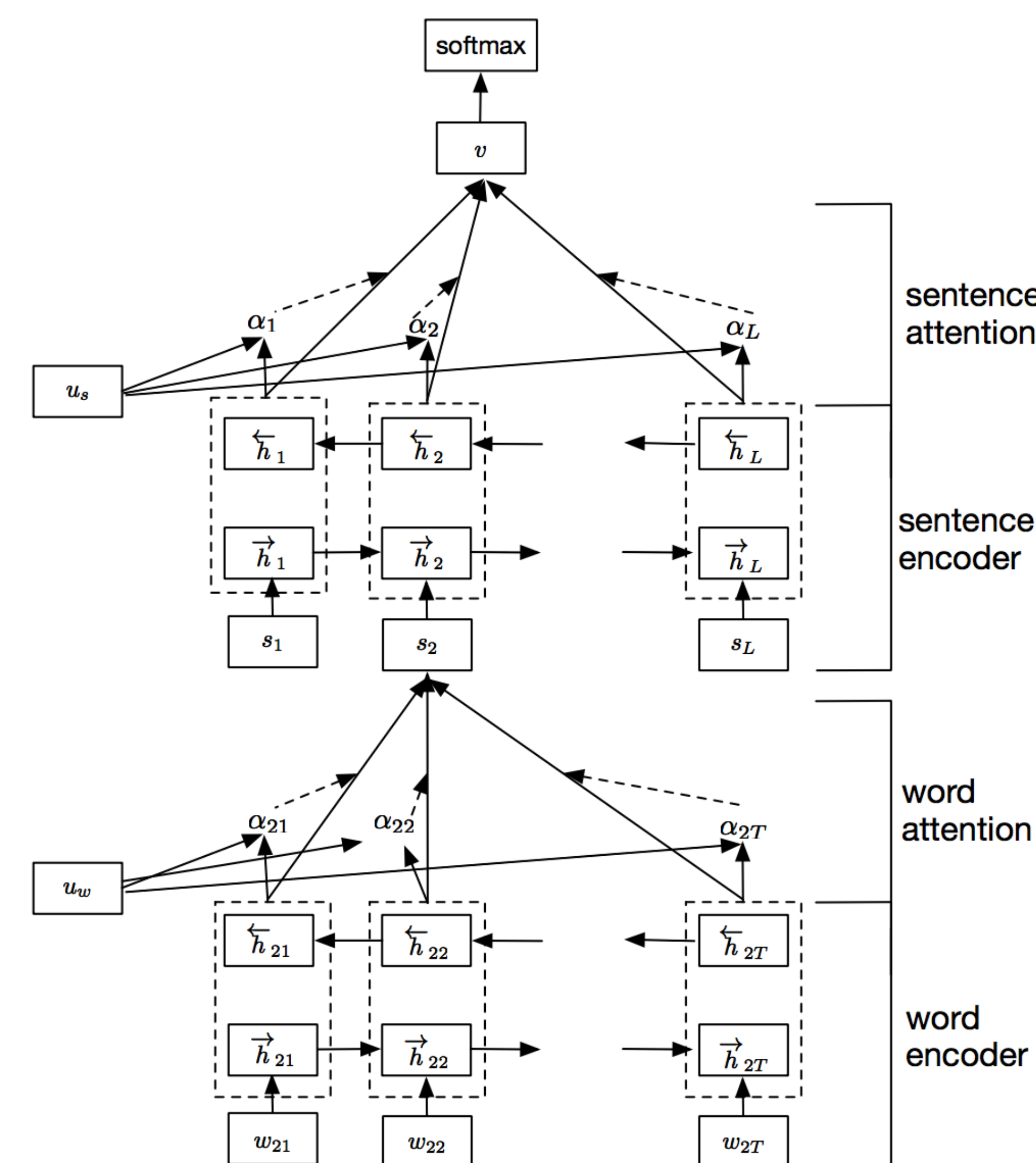


Figure 3: A schematic of how hierarchical attention works.

It calculates word-level attention using the first encoder, and sentence-level attention using second encoder. Word level attention is then re-weighted by corresponding sentence level attention as follows:

$$P^a(j) = \frac{p_a^w(j) p_s^a(s(j))}{\sum_{k=1}^{N^d} p_w^a p_s^a(s(k))} \quad (6)$$

Results

The table below shows ROUGE scores from our experiments so far.

Method	R-1	R-2	R-L
Baseline [6]	35.80	16.00	36.63
Baseline + Intra-Attention [10]	38.68	18.32	37.28
Baseline + Pointer Generator [14]	35.68	17.71	32.57
Baseline + Hierarchical Attention	16.42	4.774	14.94
Bottom-up Summarization	10.31	3.77	12.33

Table 1: ROUGE F1 scores on the Gigaword dataset.

Figure 4: ROUGE F1 scores

Analysis

High Quality Summary Output

Pointer Generator

S: charles kennedy , leader of britain 's third-ranked liberal democrats , announced saturday he was quitting with immediate effect and would not stand in a new leadership election . us president george w. bush on saturday called for extending tax cuts adopted in his first term , which he said had bolstered economic growth .

T: leader of britain 's third political party quits

P: charles kennedy announces resignation

Low Quality Summary Output

Bottom-up Summarization

S: at least two people have tested positive for the bird flu virus in eastern turkey , health minister recep akdag told a news conference wednesday .

T: two test positive for bird flu virus in Turkey....

P: new zealand tests positive for bird flu

Acknowledgements

- Microsoft for Azure credits
- Abigail See for insightful feedback
- Chris Manning and staff for an incredible class.

References

- [1] Caglar Gulcehre Bing Xiang Ramesh Nallapati, Bowen Zhou. Abstractive text summarization using sequence-to-sequence rnns and beyond. In *NIPS*, 2016.
- [2] Romain Paulus, Caiming Xiong, and Richard Socher. A deep reinforced model for abstractive summarization. *CoRR*, abs/1705.04304, 2017.