Summarization of News Articles

Project Report | Information Retrieval and Extraction
Team 29 - Vasu S., Tanish L., Tom S., Anupam M.
Mentor: Ramkishore Saravanan

1 Introduction

Automatic Document Summarization is the task of rewriting a document into a shorter form while still retaining its important content. There are important applications for text summarization in various NLP related tasks such as text classification, question answering, legal texts summarization, news summarization, and headline generation. In this project we have specifically focused on summarizing news articles.

The most popular two paradigms of text summarization are extractive approaches and abstractive approaches. **Extractive summarization** systems form summaries by copying parts of the source text through some measure of importance and then combine those part/sentences together to render a summary. Importance of sentence is based on linguistic and statistical features. **Abstractive summarization** systems generate new phrases, possibly rephrasing or using words that were not in the original text. Naturally abstractive approaches are harder. For perfect abstractive summary, the model has to first truly understand the document and then try to express that understanding in short possibly using new words and phrases. It must have complex capabilities like generalization, paraphrasing and incorporating real-world knowledge.

2 Project Details

2.1 Problem Statement

Given a large dataset of news articles, the task is to generate summaries of those articles using multiple extractive and abstractive methods and evaluate them.

2.2 Dataset

We used the non-anonymized version of **CNN/Daily Mail** Dataset [SLM17]. It contains online news articles (781 tokens on average) paired with multi-sentence summaries (3.75 sentences or 56 tokens on average). The processed version contains 287,226 training pairs, 13,368 validation pairs and 11,490 test pairs.

2.3 Approaches

We have tried out the following approaches:

- Sequence-to-Sequence Model with Attention Mechanism [Rob] [Nal+16]
- Summarization using Pointer-Generator Networks [SLM17]
- Summarization using Pointer-Generator Networks and Coverage Mechanism [SLM17]
- SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents [NZZ17]

3 Abstractive Summarization

3.1 Overview

The traditional rule-based AI did poorly on the Abstractive summarization task. But the advances in deep learning in the past few years have completely transformed the world of Abstractive summarization. Inspired by the performance of Neural Attention Model in a similar seq2seq task of Machine Translation, Rush et al. [RCW15] applied this model for Abstractive summarization and found that it already performed very well and beat the previous state-of-the-art approaches. But they applied this only to single sentences and didn't generalize. Then Nallapati et al. [Nal+16] generalized the model further and improved the results. Further improvements have been done to this baseline model by See et al. [SLM17] using Pointer Generator Networks and Coverage Mechanism and by Paulus et al. [PXS17] using Reinforcement Learning based Training procedure and intra-attention on decoder outputs as well. Both these approaches significantly improve the results.

3.2 Sequence-to-Sequence Model with Attention Mechanism

The first thing that can be done is to use a vanilla encoder-decoder model where the document is passed through the encoder and the summary is generated using the decoder. This model performs decent on short documents but it fails to scale up. There are a few problems:

1. Only the last hidden state of the encoder is passed to the decoder to predict summaries. Due to the hidden state's fixed size, it is not able to capture all the relevant information of the input sequence as the model sizes up.
2. At each generation step, only a part of the input may be relevant. But the model does not know which parts of the input to focus on during the generation of each word in the summary.

The solution to this problem is by using the attention model. Attention Mechanism calculates the importance of each input token for the current step by doing a similarity check between decoder output at this step and the input tokens. Doing this for all of the input tokens and normalizing them, we get an importance vector. We use softmax to convert them into probabilities. The context vector is then formed by a weighted sum of all input tokens and their probabilities. The equations are:

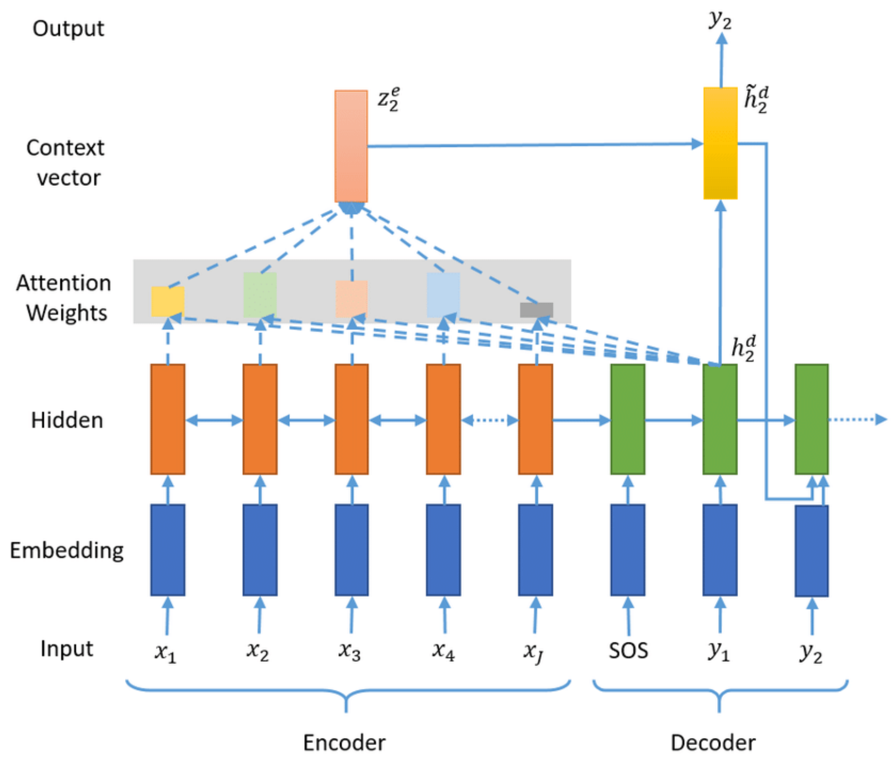

Figure 1: Sequence-to-Sequence With Attention

- $importance_{ij} = V * tanh(e(x_i)W_1 + h_j^d W_2 + b_{attn})$
- Attention Distribution $a_j = softmax(importance_{ij})$
- Context Vector $z_j^e = \sum_i e(x_i) * a_j$

In the above equations, $e(x_i)$ represented the embedding of the word x_i.
Once we get the context vector, then the process is simple and in accordance with the vanilla seq2seq network. We calculate the hidden state of the decoder, and then predict a probability distribution over the entire vocabulary for each step.

3.2.1 Problems with this Model

Even though we get decent ROUGE scores which we talk about later in the report, but still there are several problems with this model:

- They sometimes tend to reproduce factually incorrect details.
- They are repetitive and they focus on a word/phrase multiple times.
- They struggle with Out of Vocabulary (OOV) words and hence we see many UNK tokens in the predicted summary for rare words.

Some examples of the predicted summaries where we demonstrate this problems are shown in the **Appendix** section at the end of the Report.

3.3 Pointer-Generator Network

Some of the problems of the Sequence-to-Sequence with Attention Model can be solved by using the Pointer-Generator Network. Pointer-Generator Network helps to solve the challenge of OOV words and factual inaccuracies and it works better for multi-sentence summaries. The basic idea is to choose between generating a word from the fixed vocabulary or copying one from the source document at each step of the generation. It brings in the power of extractive methods by pointing towards the source words. So for OOV words, simple generation would result in UNK, but here the network will copy the OOV word from the source text. At each step we calculate

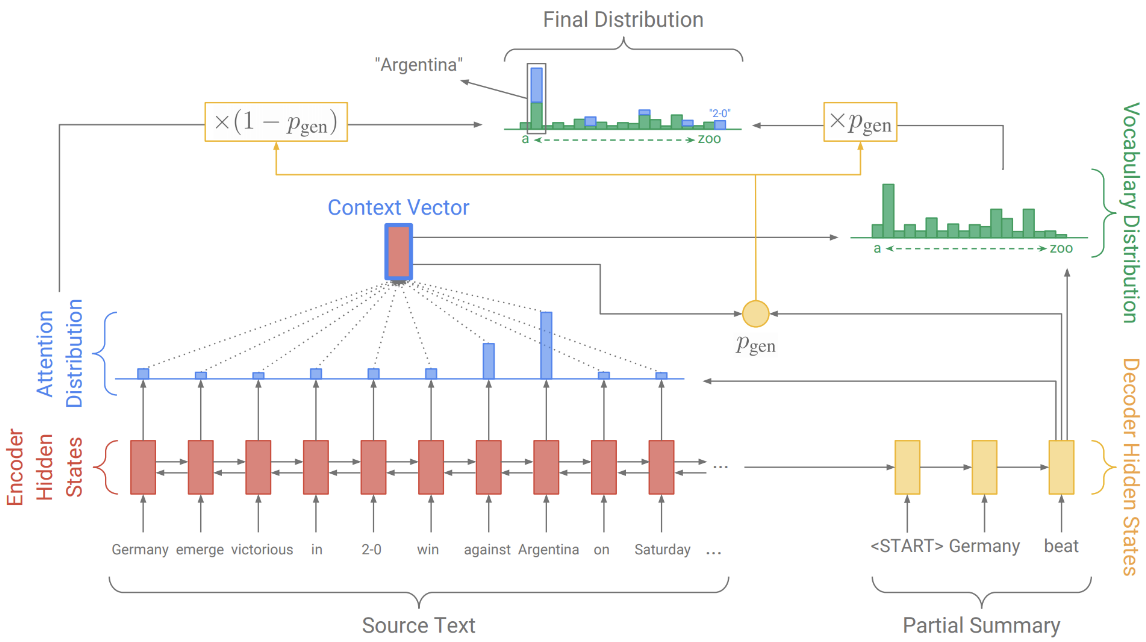

Figure 2: Pointer-Generator Network with Coverage Mechanism

generation probability p_{gen} as

$p_{gen} = \sigma(w_h^T * h_j^* + w_s^T * h_j + w_x^T * x_j + b_{ptr})$

where $x_j = y_{j-1}$, the input at j_{th} step in the decoder. We use this p_{gen} as a switch. To get the final distribution, we use

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

Note that we need to align the indexes so that the same words are present at indexes in both the distributions. If it is an OOV word, then $P_{vocab} = 0$, and hence we end up pointing.

3.3.1 Problem with this Model

Although the problem of OOV words and factually incorrect words has been solved, but there is still one major issue. Because the pointer pays attention to same words/phrases again and again, some words/phrases in the summaries are repeated and it is a major problem.

3.4 Coverage Mechanism

The cause of repetitiveness of the model can be accounted for by increased and continuous attention to a particular word. Hence, we form a coverage vector as follows:

$c_j = \sum_{j'=0}^{j-1} a^{j'}$

Intuitively, by summing the attention at all steps we are keeping track of how much coverage each encoding, $e(x_i)$ has received. Now, give this as input to attention mechanism. The updated formula for calculating importance is,

$importance_{ij} = V * tanh(e(x_i)W_1 + h_j W_2 + W_c c_i^j + b_{cov})$

Now to make the network learn to not focus on things that have already been covered we penalize attending to things that have already been covered.

$loss_{cov} = \sum_i min(a_i^j, c_i^j)$

The above loss penalizes overlap between attention at this step and coverage till now. The total loss is,

$loss_j = -logP(w_j^{optimal}) + \lambda loss_{cov}$

3.5 Generation of Summaries

Once we have the probability distributions to predict the word at the j_{th} time step, we can use various approaches to generate summaries:

- Take the word that has the maximum probability
- **Beam Search:** Choose the top k most likely target words and then feed them all into the next decoder input. So at each time-step j the decoder gets k different possible inputs. It then computes the top k most likely target words for each of these different inputs. Among these, we keep only the top-k out of k^2 and reject the rest. This process continues till the last time step. This ensures that each likely target word gets a fair shot at generating the summary.

3.6 Evaluation

We need to evaluate the quality and coherence of the summary. By quality, we mean how well does the summary capture the source document and by coherence, we mean if it's grammatically correct and is human readable. It is better to use metrics like ROUGE and METEOR. They are essentially string matching metrics. ROUGE is the most popular and has various variants like ROUGE-N and ROUGE-L. ROUGE-N measures the overlap of N-grams between the system and reference summary. A bigger N implies more fluency in the summary, because matching with a bigger portion of the reference summary which is more fluent, implies more fluency. ROUGE-L is based on longest common subsequence, thereby also taking into account sentence level similarity. Given in the table below are the ROUGE scores we calculated for our models:

ROUGE Scores			
Model	ROUGE-1	ROUGE-2	ROUGE-L
seq2seq with attention	27.039	9.483	24.691
pointer-generator	34.237	14.392	**31.216**
pointer-generator + coverage	**34.263**	**14.528**	31.016

3.6.1 Issues with the metric

ROUGE as a Metric is deficient. As pointed out by the Reinforcement learning method of Paulus et al. [PXS17], it is possible to achieve a very high ROUGE score, without the summary being human readable. This clearly means that the way ROUGE measures summary is different from how humans evaluate a summary.

3.6.2 Issues with the Dataset

We used the CNN/Daily Mail Dataset. A peculiar characteristic of it and other news datasets is that one can come up with a pretty good summary only by looking at the top few sentences because the crux of the news is present in the very first few sentences. We need a richer dataset for multi-sentence text summarization. Also, another peculiarity of these models as with any other Deep Learning model is that they require huge amounts of data and computational power.

4 Extractive Summarization

4.1 Overview

Extractive methods aim to select salient snippets, sentences or passages from documents, while abstractive summarization techniques aim to concisely paraphrase the information content in the documents.

4.2 Recurrent Neural Network based Sequence Model

In this model, we treat extractive summarization as a sequence classification problem wherein, each sentence is visited sequentially in the original document order and a binary decision is made (taking into account previous decisions made) in terms of whether or not it should be included in the summary. We use a GRU based Recurrent Neural Net-work [Chu+14] as the basic building block of our sequence classifier. A GRU-RNN is a recurrent network with two gates,**u** called the update gate and **r**, the reset gate, and can be described by the following equations:

$$u_j = \sigma(W_{ux}x_j + W_{uh}h_{j-1} + b_u)$$

$$r_j = \sigma(W_{rx}x_j + W_{rh}h_{j-1} + b_r)$$

$$h'_j = tanh(W_{hx}x_j + W_{hh}(r_j \odot h_{j-1}) + b_h)$$

$$h_j = (1 - u_j) \odot h'_j + u_j \odot h_{j-1}$$

Our model consists of a two-layer bi-directional GRU-RNN, whose graphical representation is presented in Figure 3.

The first layer of the RNN runs at the word level, and computes hidden state representations at each word position sequentially, based on the current word embeddings and the previous hidden state. We are using pretrained GloVe embeddings [PSM14] to embed each word of the

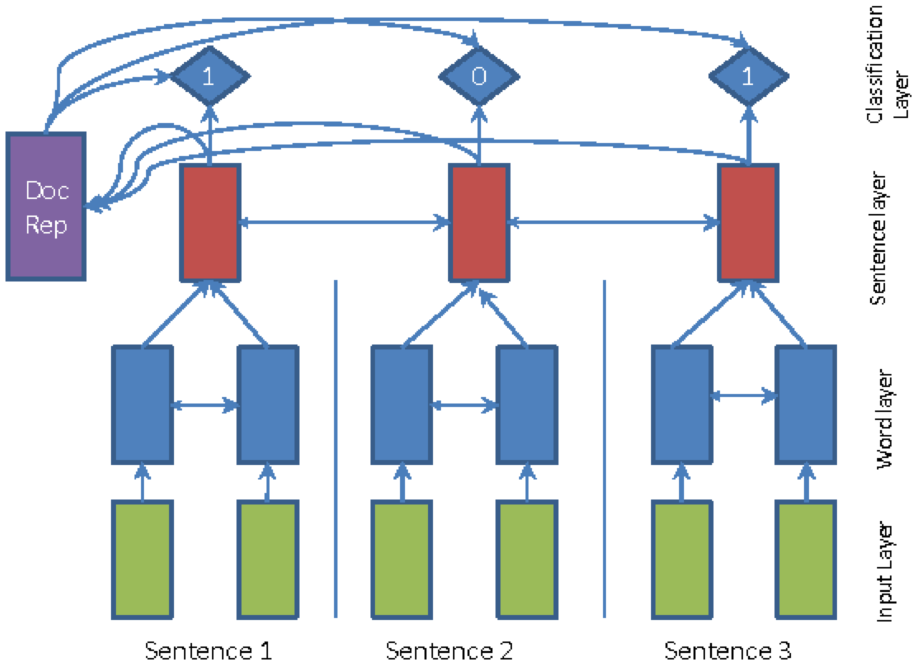

Figure 3: SummaRuNNer

sentence. We also use another RNN at the word level that runs backwards from the last word to the first, and we refer to the pair of forward and backward RNNs as a bi-directional RNN. The model also consists of a second layer of bi-directional RNN that runs at the sentence-level and accepts the average-pooled, concatenated hidden states of the bidirectional word-level RNNs as input. The hidden states of the second layer RNN encode the representations of the sentences in the document. The representation of the entire document is then modeled as a non-linear transformation of the average pooling of the concatenated hidden states of the bidirectional sentence-level RNN, as shown below.

$$d = tanh(W_d \frac{1}{N_d} \sum_{j=1}^{N^d} [h_j^f, h_j^b] + b) \tag{1}$$

where h_j^f and h_j^b are the hidden states corresponding to the j^{th} sentence of the forward and backward sentence-level RNNs respectively, N_d is the number of sentences in the document.

For classification, each sentence is revisited sequentially in a second pass, where a logistic layer makes a binary decision as to whether that sentence belongs to the summary, as shown below.

$P(y_j = 1|hj, sj, d) = \sigma(W_c hj + h_j^T W_s d + h_j^T W_r tanh(S_j) + W_{ap} p_j^a + W_{rp} p_j^r + b)$

here, the term $W_c h_j$ represents the information content of the j^{th} sentence, $h_j^T W_s d$ denotes the salience of the sentence with respect to the document, $h_j^T W_r tanh(s_j)$ captures the redundancy of the sentence with respect to the current state of the summary. We try to squash the summary representation using the *tanh* operation so that the magnitude of summary remains the same for all time-steps. The next two terms model the notion of the importance of the absolute and relative position of the sentence with respect to the document. We consider p_j^a and p_j^r, the absolute and relative positional embeddings respectively, as model parameters as well.

We minimize the negative log-likelihood of the observed labels at training time.

$$l(W, b) = -\sum_{d=1}^{N} \sum_{j}^{N_d} (y_j^d log P(y_j^d = 1 | h_j^d, s_j^d, d_d)) + (1 - y_j^d log(1 - P(y_j^d = 1 | h_j^d, s_j^d, d_d)))$$

where x is the document representation andy is the vector of its binary summary labels. At test time, the model emits probability of summary membership $P(y_j)$ at each sentence sequentially, which is used as the model's soft prediction of the extractive summary

4.3 Evaluation Metrics

After training the model on the dataset, we obtained the following ROUGE scores as the evaluation metrics.

```
Evaluation with Avg
        rouge-1:        P: 23.58        R: 50.55        F1: 31.80
        rouge-2:        P:  8.65        R: 19.05        F1: 11.75
        rouge-3:        P:  4.49        R: 10.04        F1:  6.12
        rouge-4:        P:  2.71        R:  6.12        F1:  3.70
        rouge-l:        P: 20.01        R: 38.08        F1: 26.01
        rouge-w:        P: 11.51        R: 11.90        F1: 11.47
```

Figure 4: ROUGE Scores for the model

5 Future Work

We can also try some of the following ways to further improve our model.

5.1 Intra-Attention on Decoder

Traditional approaches attend only on the encoder states. But the current word being generated also depends upon what previous words were generated. So we can use Intra-Attention on Decoder outputs. This approach helps avoids repeating particular word. The decoder context vector c_j can be generated in a similar way to encoder attention. Finally, c_j can be passed on to generate $P_{vocab}(w)$.

5.2 Reinforcement Learning

The training in our baseline model is simply word-level Supervised Training. The model's aim is to output the reference summary, so we define a cross entropy loss between the target and the produced word. But this approach is fundamentally flawed. There are various ways in which the document can be effectively summarized. The reference summary is just one of those possible ways. Hence, the model's aim shouldn't be just restricted to outputting only the reference summary. There should be some scope for variations in the summary. This is the essential idea behind the Reinforcement Learning based training approach.

In this approach, during training, we can first let the model generate a summary using its own decoder outputs as inputs. This is essentially sampling as described above. After the model produces its own summary, we can evaluate the summary in comparison to the reference summary using the ROUGE metric. We then define a loss based on this score. If the score is high that means the summary is good and hence the loss should be less and vice-versa.
Both the above ideas were described by Paulus et al. [PXS17]

5.3 Abstractive approach for label extraction

To eliminate the need to generate approximate extractive labels, we could train SummaRuNNer abstractively. This can be done by coupling it with an RNN decoder that models the generation of abstractive summaries at training time only. Instead of optimizing the log-likelihood of the extractive ground truth, we minimize the negative log-likelihood of the words in the reference summary. In addition, one could pre-train the extractive model using abstractive training. Further, a joint extractive-abstractive model could be implemented where the predictions of our extractive component form stochastic intermediate units would be consumed by the abstractive component.

But the issues with the metric and lack of a good dataset are still a great challenge and the problems in generalizing to multi-sentence summarization still remain.

6 Future Work

We have so far looked at various abstractive and extractive summarization models. This report shows that Deep Learning based approaches are promising and give some hope in solving text summarization with more and more accuracy.

References

[Chu+14] Junyoung Chung, Çaglar Gülçehre, KyungHyun Cho, and Yoshua Bengio. "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling". In: *CoRR* abs/1412.3555 (2014). arXiv: 1412.3555. http://arxiv.org/abs/1412.3555.

[Nal+16] Ramesh Nallapati, Bowen Zhou, Cicero Nogueira dos santos, Caglar Gulcehre, and Bing Xiang. *Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond*. 2016. arXiv: 1602.06023 [cs.CL].

[NZZ17] Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. "SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents". In: *ArXiv* abs/1611.04230 (2017).

[PSM14] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. "GloVe: Global Vectors for Word Representation". In: *Empirical Methods in Natural Language Processing (EMNLP)*. 2014, pages 1532–1543. http://www.aclweb.org/anthology/D14-1162.

[PXS17] Romain Paulus, Caiming Xiong, and Richard Socher. *A Deep Reinforced Model for Abstractive Summarization*. 2017. arXiv: `1705.04304 [cs.CL]`.

[RCW15] Alexander M. Rush, Sumit Chopra, and Jason Weston. *A Neural Attention Model for Abstractive Sentence Summarization*. 2015. arXiv: `1509.00685 [cs.CL]`.

[Rob] Sean Robertson. *NLP From Scratch: Translation with a Sequence to Sequence Network and Attention.*

[SLM17] Abigail See, Peter J. Liu, and Christopher D. Manning. *Get To The Point: Summarization with Pointer-Generator Networks*. 2017. arXiv: `1704.04368 [cs.CL]`.

7 Appendix

Example summaries for the abstractive models are given below:

Article (truncated): arsenal kept their slim hopes of winning this season 's english premier league title alive by beating relegation threatened burnley 1-0 at turf moor . a first half goal from welsh international aaron ramsey was enough to separate the two sides and secure arsenal 's hold on second place . more importantly it took the north london club to within four points of first placed chelsea , with the two clubs to play next week . but chelsea have two games in hand and play lowly queens park rangers on sunday , a team who are themselves struggling against relegation . good form . arsenal have been in superb form since the start of the year , transforming what looked to be another mediocre season struggling to secure fourth place – and with it champions league qualification – into one where they at least have a shot at winning the title . after going ahead , arsenal rarely looked in any danger of conceding , showing more of the midfield pragmatism epitomized by the likes of francis coquelin , who also played a crucial role in the goal . " he has been absolutely consistent in the quality of his defensive work , " arsenal coach arsene wenger told sky sports after the game when asked about coquelin 's ...

Gold Standard Summary:
arsenal beat burnley 1-0 in the epl .
a goal from aaron ramsey secured all three points .
"win cuts chelsea 's epl lead to four points .

Seq2Seq with Attention:
arsenal have two games in hand and play in the premier league.
arsenal have won eight games in a row since the start of the year .
arsenal have won eight games in a row since the start of the year .

Pointer-Generator Network:
arsenal have been in superb form since the start of the year.
chelsea have two games in hand and play lowly queens park rangers on sunday.
chelsea have two games in hand and play lowly queens park rangers on sunday.

Pointer-Generator Network + Coverage Mechanism:
arsenal kept their slim hopes of winning this season 's english premier league title alive by beating relegation threatened burnley 1-0 at turf moor .
a first half goal from welsh international aaron ramsey was enough to separate the two sides and secure arsenal 's hold on second place .
but chelsea have two games in hand and play lowly queens park rangers on sunday , a team who are themselves struggling against relegation .

Comments: Normal seq2seq with attention produces factually incorrect summary (Chelsea and not Arsenal, has two games in hand) and the sentences in the summaries are getting repeated. Adding Pointer generator produces factually correct summary but repetion still persists. Adding Coverage removes repetion also.

Article (truncated): blues legend b.b. king was hospitalized for dehydration , though the ailment did n't keep him out for long . king 's dehydration was caused by his type ii diabetes , but he " is much better , " his daughter , claudette king , told the los angeles times . the legendary guitarist and vocalist released a statement thanking those who have expressed their concerns . " i 'm feeling much better and am leaving the hospital today , " king said in a message tuesday . angela moore , a publicist for claudette king , said later in the day that he was back home resting and enjoying time with his grandchildren . " he was struggling before , and he is a trouper , " moore said . " he was n't going to let his fans down . " no more information on king 's condition or where he was hospitalized was immediately available . b.b. is short for blues boy , part of the name he used as a memphis disc jockey , the beale street blues boy ...

Gold Standard Summary:
b.b king is now out of the hospital and back at home .
bluesman suffered from dehydration and exhaustion after a 2014 show in chicago .
b.b. is short for blues boy , part of the name he used as a memphis disc jockey .

Seq2Seq with Attention:
[UNK] king was hospitalized for dehydration , but he " is much better , " his daughter , [UNK] king says .
" he was struggling before , and he is a [UNK] , " he says .
" he was struggling before , and he is a [UNK] , " he says .

Pointer-Generator Network:
b.b. king was hospitalized for dehydration , though the ailment did n't keep him out for long .
he was inducted into the rock and roll hall of fame in 1987 , and has 30 grammy nominations .

Pointer-Generator Network + Coverage Mechanism:
blues legend b.b. king was hospitalized for dehydration .
king 's dehydration was caused by his type ii diabetes .
king 's dehydration was caused by his type ii diabetes , but he " is much better , " his daughter , claudette king , told the los angeles times .

Comments: Normal seq2seq model is unable to generate out of vocabulary words and thus it replaces them by unknown token. This is not a problem with pointer generator as there it directly copies the tokens from source sentence.

Article (truncated): kanye west has settled a lawsuit with a paparazzi photographer he assaulted – and the two have shaken on it . the photographer , daniel ramos , had filed the civil suit against west after the hip-hop star attacked him and tried to wrestle his camera from him in july 2013 at los angeles international airport . west pleaded no contest last year to a misdemeanor count of battery over the scuffle . a judge sentenced him to two years ' probation , as well as anger management sessions and community service . ramos and his lawyer , gloria allred , sought general and punitive damages in the civil suit , saying that west had interfered with the photographer 's rights to pursue a lawful occupation . the case had been set for trial next week , but allred issued a statement tuesday night saying ramos ' side had filed a dismissal " because the case was settled to the satisfaction of the parties . " she did n't disclose the details of the settlement other than saying that " one important aspect of it was an apology by kanye west to our client , daniel ramos . " her statement included a picture of west and ramos shaking hands , which she said happened after the apology . the original incident was caught on video , including the following exchange . " kanye ! kanye ! talk to me , kanye ! " ramos shouts outside a terminal at the los angeles airport on the night of july 19 , 2013 . " what 's ' going on ? why ca n't we talk to you ? i mean , why ? " he asks as west moves through a group of paparazzi ...

Gold Standard Summary:
the rapper assaulted the photographer at los angeles international airport in 2013 .'
"west apologized as part of the settlement , the photographer 's lawyer says ."

Seq2Seq with Attention:
kanye west has settled a lawsuit with a clown photographer .
the photographer , daniel ramos , had filed the civil suit against west .
the [UNK] was caught on video , including the following exchange .

Pointer-Generator Network:
kanye west pleaded no contest last year to a misdemeanor count of battery over the scuffle .
west pleaded no contest last year to a misdemeanor count of battery over the scuffle . west pleaded no contest last year to a misdemeanor count of battery over the scuffle .

Pointer-Generator Network + Coverage Mechanism:
kanye west has settled a lawsuit with a paparazzi photographer he assaulted .
the photographer , daniel ramos , had filed the civil suit against west after the hip-hop star attacked him and tried to wrestle his camera from him in july 2013 at los angeles international airport .
west pleaded no contest last year to a misdemeanor count of battery over the scuffle .

Comments: Pointer generator model along with coverage mechanism becomes too extractive in some cases which is not desired from an abstractive model.

Example summaries for Extractive model.

Article (truncated): washington (cnn) – nearly 16 years after a fuel tank explosion destroyed twa flight 800 , killing all 230 aboard , the federal aviation administration on friday proposed to fine boeing co. $ 13.57 million for failing to meet a deadline intended to prevent similar catastrophes . the faa said boeing failed to meet a 2010 deadline to give airlines information on how to reduce fuel tank flammability , missing the deadline by 301 days for its b-747 aircraft , and by 406 days for its b-757 planes . because of the missed deadline , airlines have asked the faa for extensions to make necessary fixes , the faa said . the faa said it is considering extending a deadline requiring airlines to retrofit half of their aircraft by 2014 , but will not extend a 2017 deadline to retrofit all impacted aircraft . some 383 boeing aircraft in the united states are affected by the delays , it said . " we take this matter very seriously , " said acting faa administrator michael huerta . " we have issued hundreds of directives to eliminate fuel ignition sources over the past 16 years , and this step will add another layer of safety . " in a two-page letter to boeing on friday , the faa proposed the fine of $ 13,574,400 . miles kotay , a boeing spokesman , said friday in a statement that the company is " committed " to continuing efforts to provide a solution to the problem . " boeing has since provided the service instructions to the faa concerning the out-of-production aircraft that are the subject of the proposed penalty , " the statement said . " in compliance with the rule changes , boeing has already included a flammability reduction system in the basic design on the 747-8 and 787 . the system is being installed on all boeing airplanes currently in production (737 , 747-8 , 767 , 777 and 787) and is available for retrofit on all other out-of-production models . the system is currently in service on 1,805 boeing airplanes around the world . " the july 17 , 1996 , explosion of twa 800 , a boeing 747 , was one of ...

Gold Standard Summary:

the faa says boeing has failed to meet a deadline to prevent similar crashes as twa flight 800
the flight crashed in july 1996 after an explosion in the central fuel tank
boeing has n't yet given airlines information on how to reduce fuel tank flammability , faa says

---

Predicted:

nearly 16 years after a fuel tank explosion destroyed twa flight 800 , killing all 230 aboard , the federal aviation administration on friday proposed to fine boeing co $ 1357 million for failing to meet a deadline intended to prevent similar catastrophes
the faa said boeing failed to meet a 2010 deadline to give airlines information on how to reduce fuel tank flammability , missing the deadline by 301 days for its b-747 aircraft , and by 406 days for its b-757 planes
some 383 boeing aircraft in the united states are affected by the delays , it said because of the missed deadline , airlines have asked the faa for extensions to make necessary fixes , the faa said
the july 17 , 1996 , explosion of twa 800 , a boeing 747 , was one of the deadliest accidents in aviation history , and was among the most difficult to solve

---

Comments: Predicted summaries are longer much longer than gold standard summaries