

# Enhance Abstractive text summarization with an extractive layer

Amit Dharmadhikari

Sakshi Gupta

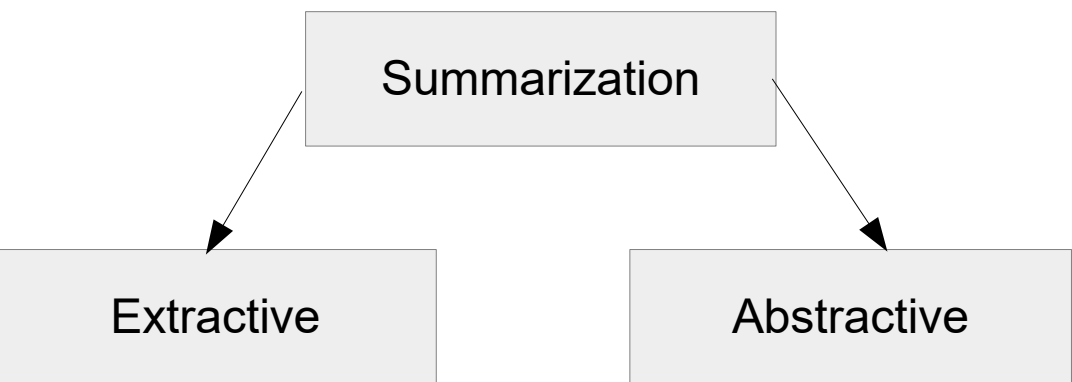
Shoaib Sheriff



## What is Summarization?

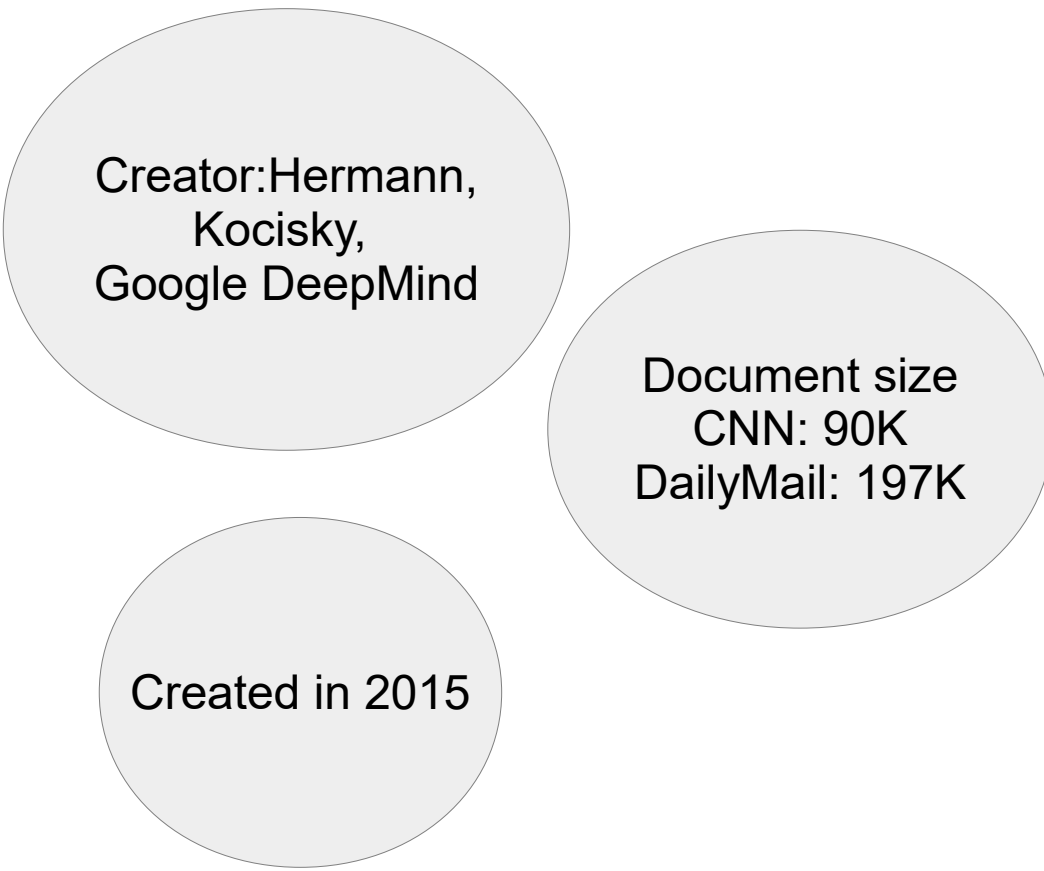
In today's world, unstructured content accounts for around 95 percent of the digital information available to us. The solution is .SUMMARIZATION.

Text Summarization is the process of extracting meaningful information which is shorter than the given text while still preserving its meaning.



- Selects a subset of sentences from the original text
- Ranks sentences by importance and combine
- Generates new sentences for the summary
- Instead of picking up directly from the original text

## Dataset Used



CNN story dataset consists of around 10000 long news articles(an average of ~800 words) with annotated summaries.

## Existing Problems

- Fails to pick important ideas when Summary has many
- The network fails to process long sequences or sentences
- It tends to pick random sentences from the original text in case of text pieces similar to blogs or opinion pieces
- Often gets tripped up on sentences with quotes. It places extra emphasis on such phrases

**Motivation :** To solve some of the above with Extractive+Abstractive Summarization

## Extractive Summarization Network

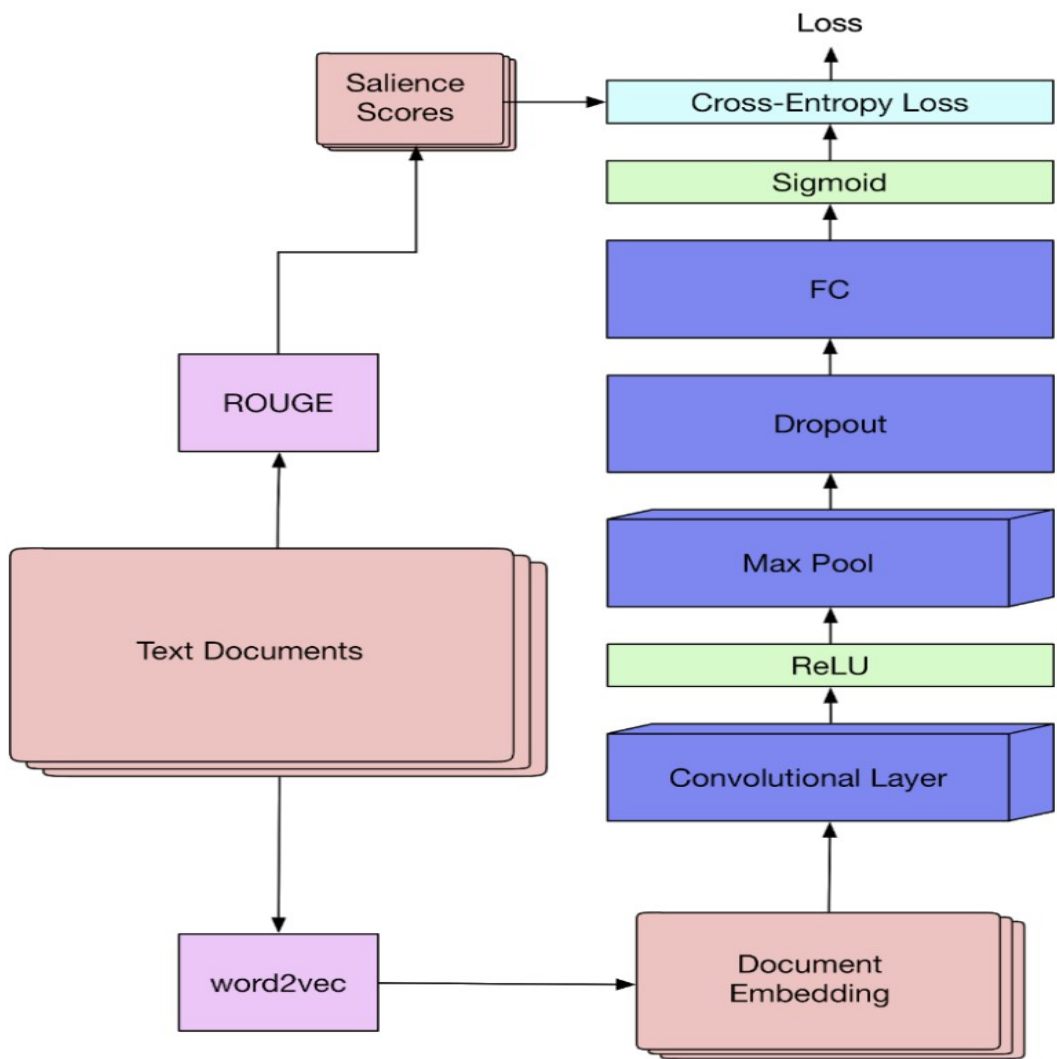
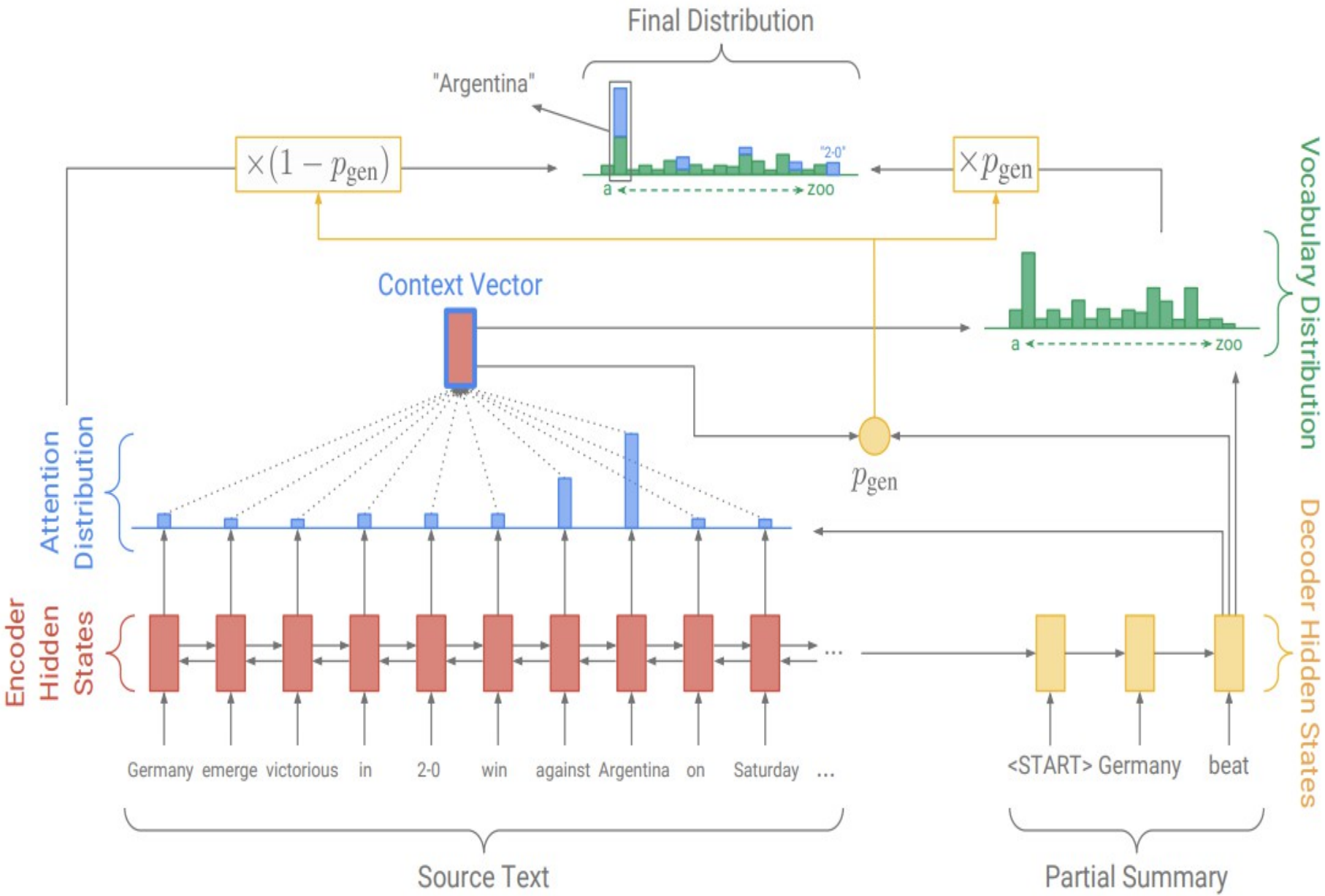


Fig. 2: Full training procedure

## Abstractive Summarization Network



## Block Diagram



## Results

GROUND TRUTH	BASELINE MODEL	PROPOSED MODEL
kim kardashian is compiling a 352-page book of selfies titled `` selfish " pepper schwartz : normal people would find this extreme , but not the queen of narcissism . she says kim k is a canny self-promoter ; it 's when you just love yourself so much 24/7 . schwartz : why are young women so entranced with kardashian ?	kim k is the queen of narcissism , the queen of narcissism . she said she was assembling sexy pictures of herself from a trip to southeast asia for a book to give to her newish husband , kanye west	kim k is one of the people who is `` famous for being famous " the queen of narcissism , most of us have a more ambivalent relationship to our own image . if it 's a flattering picture ,
at least five killed in oklahoma . american red cross opens up at least two shelters . truck stop on interstate 40 east of oklahoma city destroyed by twister . storms damaged homes , overturned cars , downed trees in norman , oklahoma .	at least five people were killed in oklahoma on monday . a severe storm system spawned multiple tornadoes across the state . the location in oklahoma of the fifth person died was unknown late monday night .	at least five people were killed in oklahoma on monday . a severe storm system spawned multiple tornadoes across the state . a truck stop east of oklahoma city was demolished , taking a direct hit from one of the

## Evaluation Metrics

### Cosine Similarity

way of quantifying the similarity between two documents in the vector space

### BLEU

Measures precision: how much the words in the output appeared in the human annotated summaries

### ROUGE-I

Measures recall: how much the words in the human annotated summaries appeared in the output

## Accomplishments and Conclusion

- Developed a method to augment benefits of extractive summarization with state-of-the-art abstractive summarization methods
- Identified relative importance of entities and used it to generate a new probability distribution mechanism in the decoder
- Performance
  - BLEU:** Around 40% (2463/7054) performed better
  - Cosine Similarity:** Around 35% (2823/7054) did better
  - ROUGE-I:** Around 34% (2396/7054) did better