



# Transformer Network for Abstractive Text Summarization

---

Aniruddha Bala  
15655

Chittersu Raghu  
15840

Ravi Raj Saxena  
16326

Indian Institute of Science, Bangalore

# Introduction

---

- Given an input article the task is to summarize the article with novel sentences in the output.

## Source Text

munster have signed new zealand international francis *saili* on a two-year deal . utility back *saili* , who made his all blacks debut against argentina in 2013 , will move to the province later this year after the completion of his 2015 contractual commitments . the 24-year-old currently plays for *auckland-based* super rugby side the blues and was part of the new zealand under-20 side that won the junior world championship in italy in 2011 . *saili* 's signature is something of a coup for munster and head coach anthony foley believes he will be a great addition to their backline . francis *saili* has signed a two-year deal to join munster and will link up with them later this year . ' we are really pleased that francis has committed his future to the province , ' foley told munster 's official website . ' he is a talented centre with an impressive *skill-set* and he possesses the physical attributes to excel in the northern hemisphere . ' i believe he will be a great addition to our backline and we look forward to welcoming him to munster . ' *saili* has been capped twice by new zealand and was part of the under 20 side that won the junior championship in 2011 . *saili* , who joins all black team-mates dan carter , *ma'a nonu* , conrad smith and charles *piutau* in agreeing to ply his trade in the northern hemisphere , is looking forward to a fresh challenge . he said : ' i believe this is a fantastic opportunity for me and i am fortunate to move to a club held in such high regard , with values and traditions i can relate to from my time here in the blues . ' this experience will stand to me as a player and i believe i can continue to improve and grow within the munster set-up . ' as difficult as it is to leave the blues i look forward to the exciting challenge ahead . '

## Reference summary

utility back francis *saili* will join up with munster later this year . the new zealand international has signed a two-year contract . *saili* made his debut for the all blacks against argentina in 2013 .

# Literature Survey

---

- The work by Nallapati et al provided 1<sup>st</sup> baselines for the task on CNN/DM data set. They used hierarchical RNN in their approach.
- See et al introduce a pointer generation mechanism to choose between generating a token & copying from input sequence, also keeping track of coverage to address repetition.
- **Transformers** replaces the recurrence and convolutions in neural models by self attention mechanism & are more parallelizable.
- **BERT**: learns bidirectional contextual representations.

Using Sequence-to-Sequence RNNs and Beyond (Nallapati et al., 2016)

See et al., 2017 Get to the Point: Summarization with pointer networks

Vaswani et al., 2017 Attention is all you need

Devlin et al., 2018 BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

# Contribution

---

- Explored the effectiveness of Transformer models for abstractive summarization.
- Integrated the Pointer Generator mechanism with transformer model to handle out of vocabulary words.
- Performance improvements with transfer learning by using BERT encoder.

# Datasets

---

- CNN Daily Mail dataset

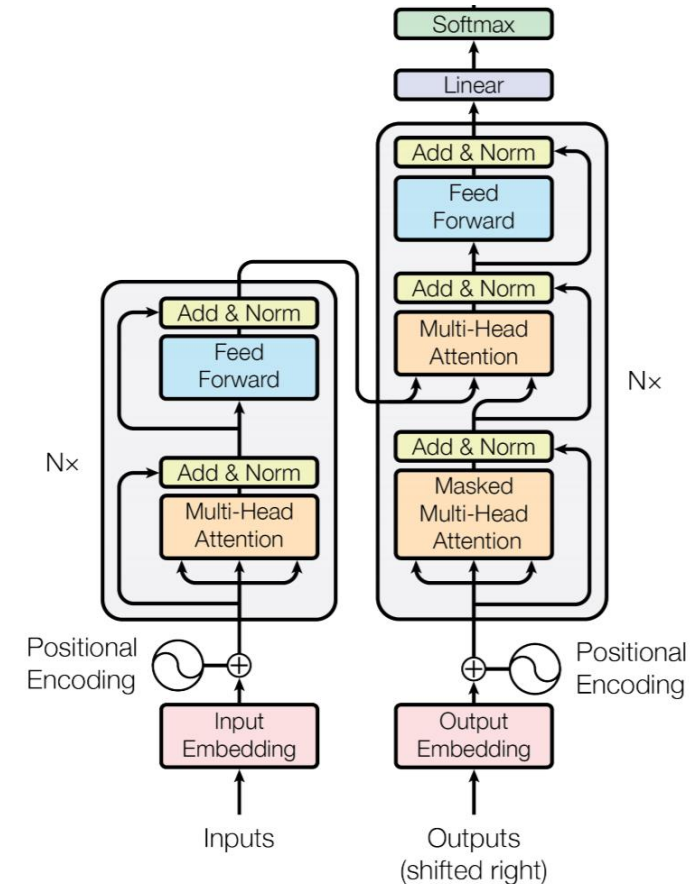
Task	Article Reference Pairs
Training	287,726
Validation	13,368
Test	11,490

- ROUGE Metric

ROUGE-1, ROUGE-2, ROUGE-L

# Model-1

- **Key Features:**
- Input positional encodings
  - $PE_{pos,2i} = \sin(pos/10000^{2i/d_{model}})$
  - $PE_{pos,2i+1} = \cos(pos/10000^{2i/d_{model}})$
- Multi headed self attention
  - $Q = W_Q X$
  - $K = W_K X$
  - $V = W_V X$
- Layer norm and Residual connections
- Position wise feed forward network
  - $FFN(X) = \max(0, XW_1 + b_1) W_2 + b_2$
- **Advantage:** No recurrence , one can leverage parallelism



# Model-2

Generation probability

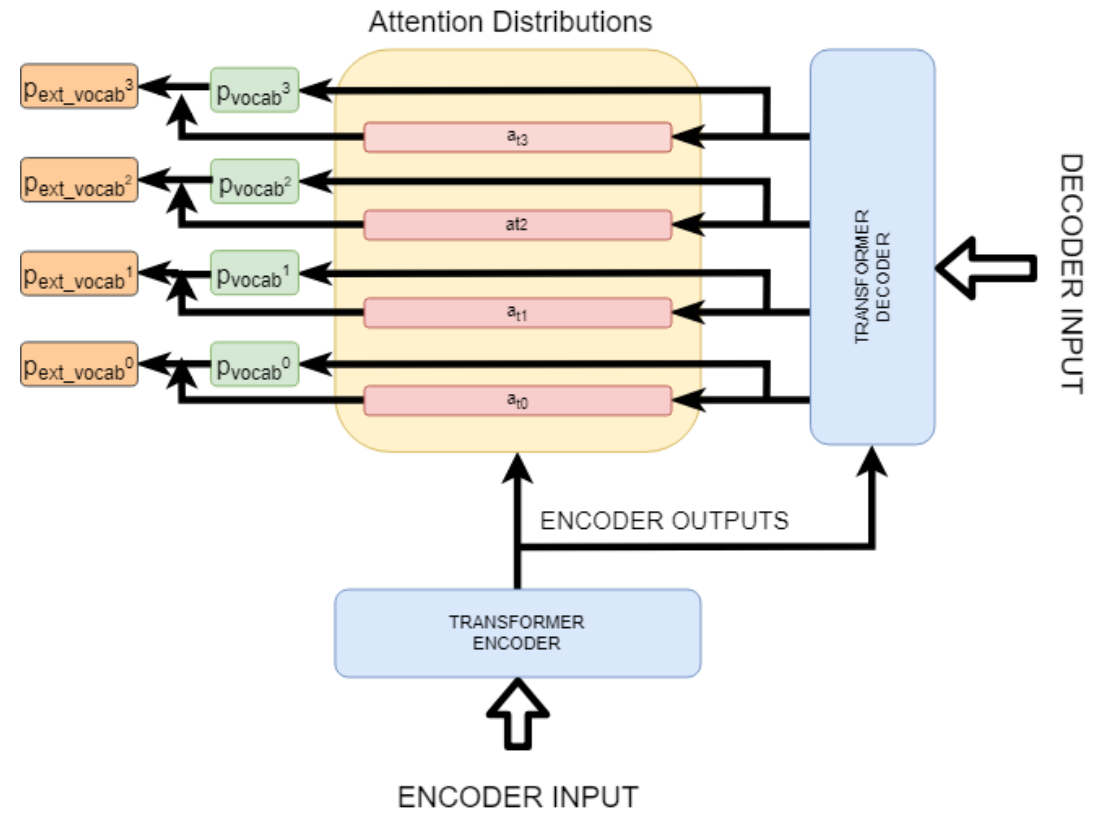
$$P_{gen} = \text{sigmoid}(W_{gen}[X_d, X_c] + b_{gen})$$

Vocab probability

$$P_{vocab} = \text{softmax}(W_{vocab}[X_d] + b_{vocab})$$

Final Extended vocabulary

$$P_{ext} = P_{gen} * P_{vocab} + (1 - P_{gen}) * P_{attn}$$



# Model-3

---

- In this we use the BERT base uncased pretrained PYTORCH model
- We replace our encoder with BERT encoder.
- We change our model hidden dimension to match BERT's hidden dimension which is 768
- As we use BERT embedding we also use the BERT provided vocabulary which is of dimension 30,524



# Experimental Setup

---

Hyperparameter	Value
Batch size	8
Input sequence length	400
Output sequence length	100
Hidden dimension	512 , 768 (for BERT)
FFN projection dimension	1024
Number of layers	6
Vocab size	50,000, 30,524 (for BERT)

# Results

---

Model	ROUGE-1	ROUGE-2	ROUGE-L
Seq2seq + attn baseline (150k vocab)	30.49	11.17	28.08
Seq2seq + attn baseline (50k vocab)	31.33	11.89	28.83
Pointer Generator	36.44	15.66	33.42
Transformer	12.79	1.18	11.72
Transformer + Pointer Generator	28.52	7.02	26.14
BERT + Pointer Generator	30.98	12.25	28.93

See et al., 2017 Get to the Point: Summarization with pointer networks

# Generated Summaries

---

**Article:** a kentucky man has been arrested after police say he was found under the influence while riding a horse on us 23 . michael kimmel , 40 , was taken into custody by kentucky state police on monday evening after they received a 911 call about an intoxicated horse rider . trooper j. gabbard 's report says that kimmel was ordered to stop , but instead dismounted and ran away wearing only a brown hat , jeans and boots . trooper gababrd launched a manhunt for kimmel and later found him on horseback again and according to the floydcountytimes , he resisted arrest , saying , ' i did n't do s \*\*\* , i was just riding my horse . ' according to the arrest report , kimmel would not take a sobriety test and refused a breath and blood alcohol test . however , officers said he had slurred speech , smelt of alcohol and was unsteady on his feet . “ subject made threats to ‘ -lsb- expletive -rsb- -lsb- expletive -rsb- up driving drunk in a car next time and he would give me something to worry about , ” the arrest citation quotes kimmel as saying . kimmel , who has used the alias “ mike bicycle , ” is currently on probation for a prior conviction for burglary . he is currently + in the floyd county jail on \$ 5000 cash bond . kimmel faces dui , fleeing or evading police , and other possible charges .

**Reference Summary:** michael kimmel was taken into custody wearing only boots , jeans and a cowboy hat .

**Predicted Summary:** michael kimmel , 40 , was taken into custody on monday . police received 911 call police call about an intoxicated horse rider . he was arrested after they received a 911 call about an intoxicated horse rider .

# Generated Summaries

---

**Article:** a mother-of-three was shot and killed by her husband who then turned the gun on himself , according to police in tulare , california . the couple have been named as georgina rojas-medina , 41 , and her common-law husband , gerardo tovar , 44 . neighbors say they were alerted to the bodies by the couple 's 4-year-old daughter just after midnight on saturday . mother-of-three georgina rojas-medina , 41 , was shot and killed by her common-law husband , gerardo tovar , 44 , who then turned the gun on himself on saturday night , according to police in tulare , california . bloody footprints show the path the 4-year-old girl had to make after finding her mother shot to death and a father who 'd killed himself . bloody footprints show the path the 4-year-old girl had to make after finding her mother shot to death and a father who 'd killed himself , reports abc30 . ' the 4-year-old who was present on scene at the time of the incident was not injured , ' said sgt. andrew garcia of the tulare police department .

**Reference Summary:** mother-of-three georgina rojas-medina , 41 , was shot and killed by her common-law husband on saturday night . gerardo tovar , 44 , then turned the gun on himself in tulare , california . the couple 's bodies were discovered by their 4-year-old daughter who walked out of the house and asked a neighbor to call the police . the 4-year-old girl and two other siblings , ages 10 and 12 , were not hurt in the shooting .

**Predicted Summary:** mother-of-three georgina rojas-medina , 41 , was shot and killed by her husband , gerardo , 44 , 44 , on saturday night . her husband , gerardo , 41 , then turned the gun on himself .

# Analysis

---

- Transformer model performs poorly
  - learnt probability distributions over the vocabulary is not good enough
- Integration of Pointer Generator significantly improves performance
  - eliminates UNK's from the output sequence with direct copy
  - increases the probability mass of the words in the vocab distribution which receive more attention in the input sequence
- With BERT encoder we are able to beat our baselines
  - better language representations learned from unsupervised pre-training helps

# Conclusions

---

- Pointer Generator mechanism is a useful technique to augment and enhance the output probability distribution
- Transfer learning can ease the model's workload of training end to end and help in achieving faster convergence with better performance.
- **Pros:**
  - Able to generate multi-sentence summaries
  - Faster training
- **Cons:**
  - Repetitive sentences in the output

# Future Work

---

- Further work should be focused on integrating the coverage mechanism in Transformer models which is not trivial
- Recurrence formulation of the coverage formula restricts it's use in Transformers
- Our proposed approach:
  - Use two layered attention
  - Use a lower triangular mask of ones and multiply with the attention probabilities obtained from the previous layer
- Future work can also be focused at investigating ways to improve the Transformer model without Pointer Generator

**THANK YOU**