

Summarization of News Articles

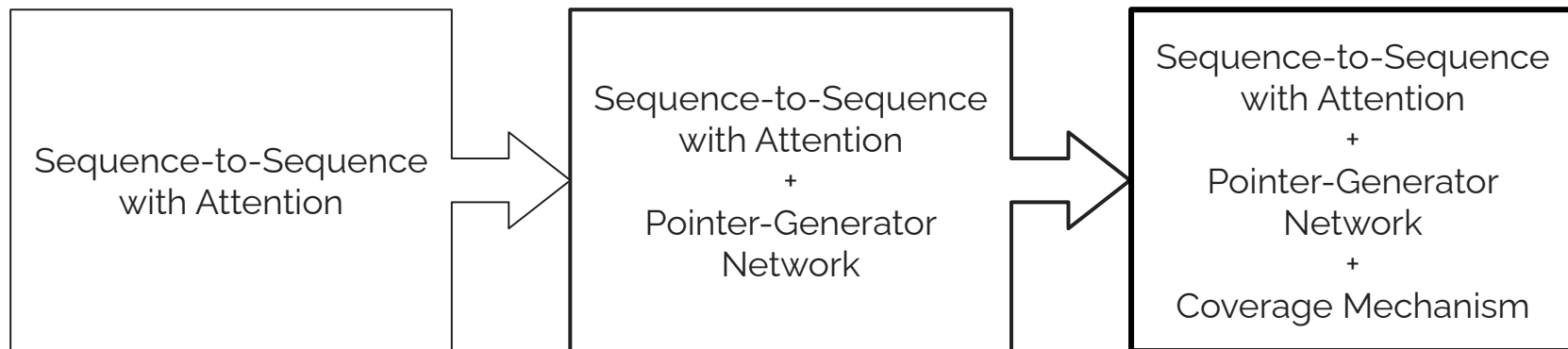
Team 29

Vasu Singhal, Tanish Lad,
Anupam Majumder, Tom Sabu
Mentor: Ramkishore Saravanan

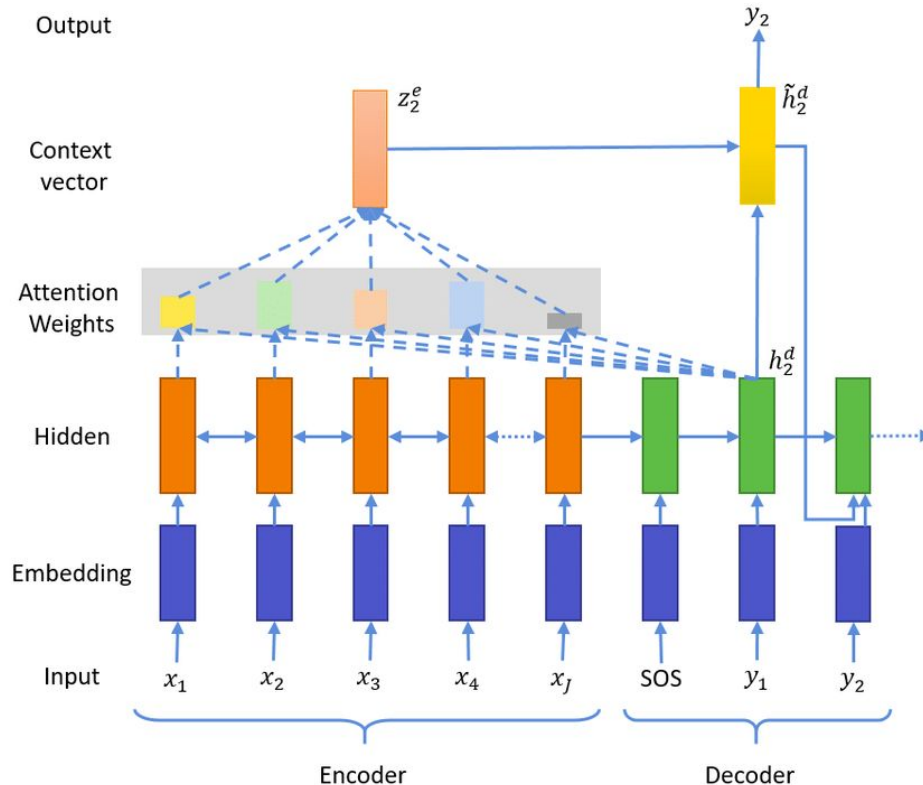
Abstractive Summarization

Abstractive summarization systems generate new phrases, possibly rephrasing or using words that were not in the original text.

Models Implemented

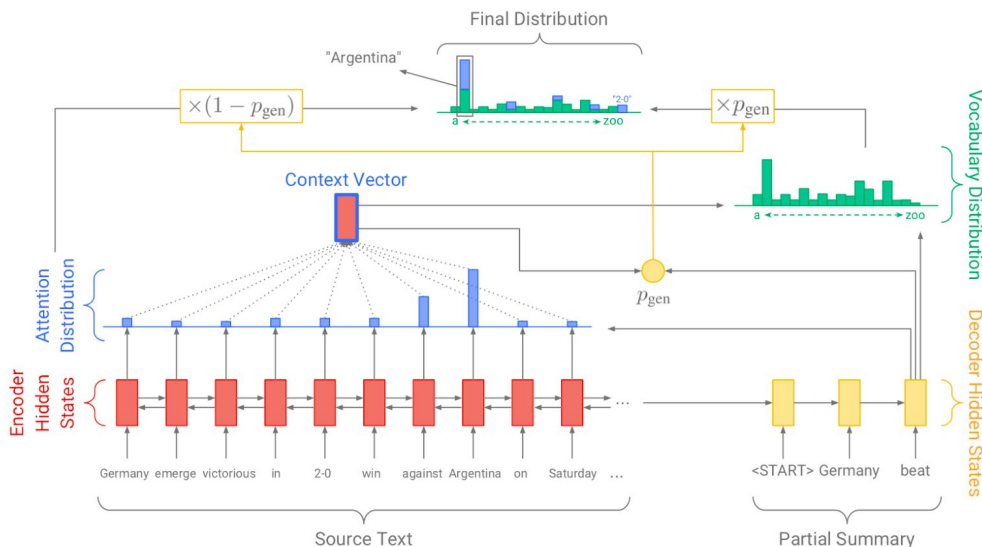


Sequence-to-Sequence with Attention



- ❑ Original article is passed through an encoder.
- ❑ The hidden states generated at each time step in the encoder are stored.
- ❑ At each decoder time step, attention weights for each encoder hidden state is calculated which is used to generate weighted sum of encoder hidden states.
- ❑ The weighted sum of encoder hidden states, previous hidden state of decoder and last generated word is used to generate the next word.
- ❑ Generated summary is compared with reference summary (cross entropy loss) to train the model.

Pointer-Generator Network and Coverage Mechanism



- ❑ Built on top of seq2seq + attention model.
- ❑ Some words are directly copied from the source article via pointing. This has two advantages:
 - ★ Less chances of producing factually incorrect summaries.
 - ★ Incorporating out of vocabulary words in generated summaries.
- ❑ Use of coverage to penalize attention over same words again and again.
 - ★ Reduces repetition of words in generated summaries.



Results

Example Article

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 contribution to arsenal 's current run . they have won eight games in a row since introducing the previously overlooked young frenchman into a more defensive midfield position . `` he was a player who was with us for seven years , from 17 , he 's now just 24 , " wenger explained . `` sometimes you have to be patient . i am very happy for him because he has shown great mental strength . " now all eyes will be on next week 's clash between arsenal and chelsea which will likely decide the title . `` they have the games in hand , " said wenger , playing down his club 's title aspirations . `` but we 'll keep going and that 's why the win was so important for us today . " relegation dogfight . meanwhile it was a good day for teams at the bottom of the league . aston villa continued their good form since appointing coach tim sherwood with a 1-0 victory over tottenham , who fired sherwood last season . belgian international christian benteke scored the only goal of the game , his eighth in six matches , to secure a vital three points to give the midlands club breathing space . another midlands club looking over their shoulder is west brom , who conceded an injury time goal to lose 3-2 against bottom club leicester city . but it was an awful day for sunderland 's former dutch international coach dick advocaat , who saw his team lose 4-1 at home against form team crystal palace . democratic republic of congo international yannick bolasie scored crystal palace 's first ever hat trick in the premier league to secure an easy victory . "

Predicted Summaries

Sequence-to-Sequence 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 .

Factually Incorrect

Repetition

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 .

Factually correct now

Repetition is still there

Pointer-Generator Network + Coverage Mechanism

Factually correct and no repetition

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 .

Example Article

b"-lrb- cnn -rrb- 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 . he was inducted into the rock and roll hall of fame in 1987 , and has 30 grammy nominations . king , 89 , has used various models of gibson guitars over the years , and named each one of them lucille . in the 1980s , gibson officially dropped the model number on the guitar he used last and most . it became a custom-made signature model named lucille , manufactured exclusively for the `` king of the blues . " some of his hits include `` the thrill is gone , " which won him his first grammy in 1970 , `` there must be a better world somewhere " and `` when love comes to town , " a collaboration with u2 . last year , the bluesman suffered from dehydration and exhaustion after a show in chicago , forcing him to cancel the remainder of his tour . cnn 's greg botelho and sonya hamasaki contributed to this report . "

Predicted Summaries

Sequence-to-Sequence 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 .

Out of vocab words

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 .

Directly copied from source
article using Pointer Network

Sequence-to-Sequence isn't always bad!

Article

b"-lrb- cnn -rrb- buckle up for another mercedes battle in the formula one world championship . lewis hamilton took victory for the silver arrows at the chinese grand prix -- only for his teammate nico rosberg to later claim he had ruined his race by driving slowly . the reigning world champion captured his 35th grand prix win with a peerless pole-to-flag win under shanghai 's sunny skies . `` it was great to have a smooth weekend , " the british racer said before spraying the champagne on the podium . `` the team have done a fantastic job to up our pace and improve after our last race where we struggled a bit . " f1 championship standings . hamilton had lost the malaysia grand prix in a strategic battle with the ferrari of sebastian vettel but it was business.....vettel is just 13 points behind hamilton in the world championship and four points ahead of rosberg , after the first three races of the year . rosberg unhappy . it was a disheartening weekend for rosberg -- the son of 1982 world.....

Predicted Summary

lewis hamilton took victory for the silver arrows at the chinese grand prix .
he had lost the malaysian grand prix in a german battle with the ferrari of sebastian vettel .
it was the first three races of the year .

Is Pointer-Generator with Coverage really abstractive?

Article

b"-lrb- cnn -rrb- 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

Predicted Summary

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 .

ROUGE Scores

	Rouge-1 (F1 score)	Rouge-2 (F1 score)	Rouge-L (F1 score)
Sequence-to-Sequence with Attention	0.27039	0.09483	0.24691
Pointer-Generator Network	0.34237	0.14392	0.31216
Pointer-Generator Network + Coverage Mechanism	0.34263	0.14528	0.31016

Extractive Summarization

Identifying important sections of the text and generating them verbatim
producing a subset of sentences from the original.

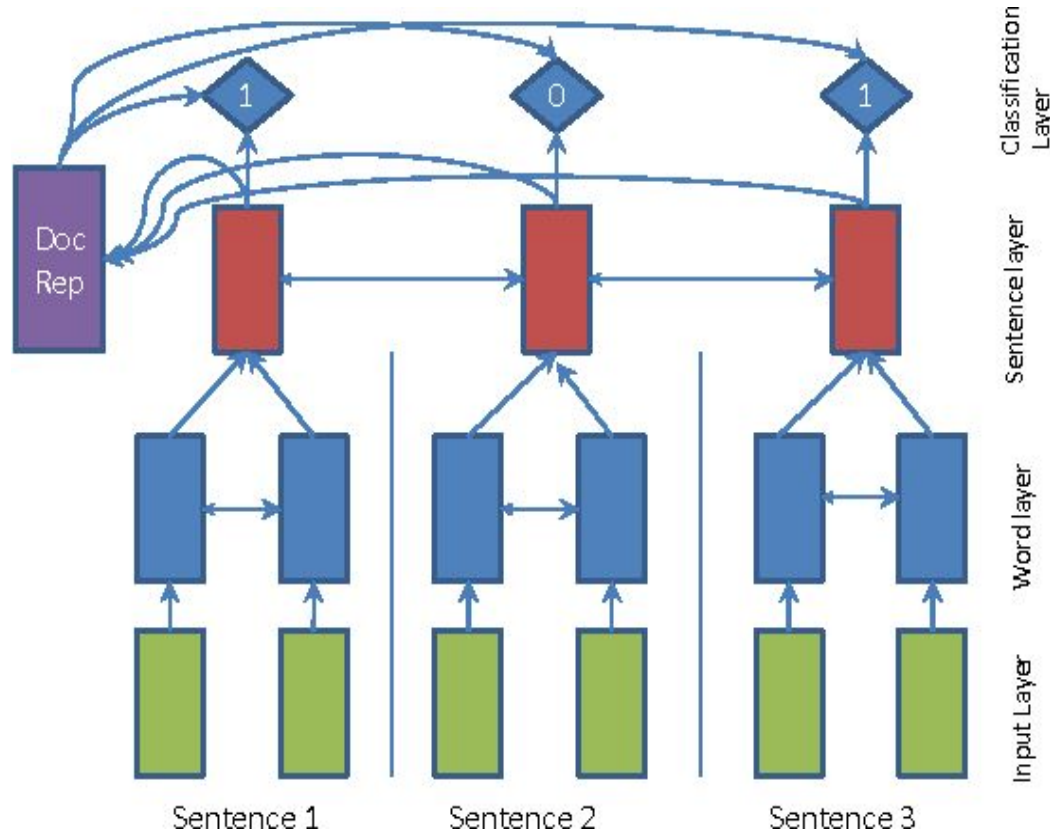
SummaRuNNer

Recurrent Neural Network (RNN) based sequence model for extractive summarization of documents.

Extractive features include : information content, salience and novelty, absolute and relative positional embeddings.

GloVe (6B,50) was used as pre-trained embeddings.

SummaRuNNer



- ❑ A two-layer bi-directional RNN based sequence classifier.
- ❑ First layer runs at the word level whereas the second layer operates at the sentence level.
- ❑ Sigmoid activation based classification layer decides whether or not each sentence belongs to the summary.

SummaRuNNer - Summary Comparison

Actual Summary	Predicted Summary
<p>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</p>	<p>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</p>
<p>new : vice president guy scott will take over until elections , cabinet secretary says sata took office in 2011 after the incumbent president tearfully conceded he was nicknamed " king cobra " because of his fiery tongue speculation over his health has intensified since he took office</p>	<p>zambian president michael sata , who was nicknamed " king cobra " for his fiery comebacks and larger-than-life personality , has died sata took office in september 2011 after the incumbent president tearfully conceded in a televised speech , a rarity in a continent known for volatile elections and leaders fighting their defeat tooth-and-nail sata had traveled to london for unspecified medical treatment last week , and died at a hospital there tuesday evening he is ineligible to be elected president because his parents were not born in zambia , msiska said scott , who is of scottish descent , will be the first white president in sub-saharan africa since apartheid</p>

SummaRuNNer - Rouge Scores

	Precision	Recall	F1-score
Rouge-1	0.24	0.50	0.32
Rouge-2	0.09	0.19	0.12
Rouge-3	0.04	0.10	0.06
Rouge-4	0.03	0.6	0.3
Rouge-l	0.28	0.51	0.36
Rouge-w	0.16	0.20	0.17

Thanks!

