Automatic Summarization

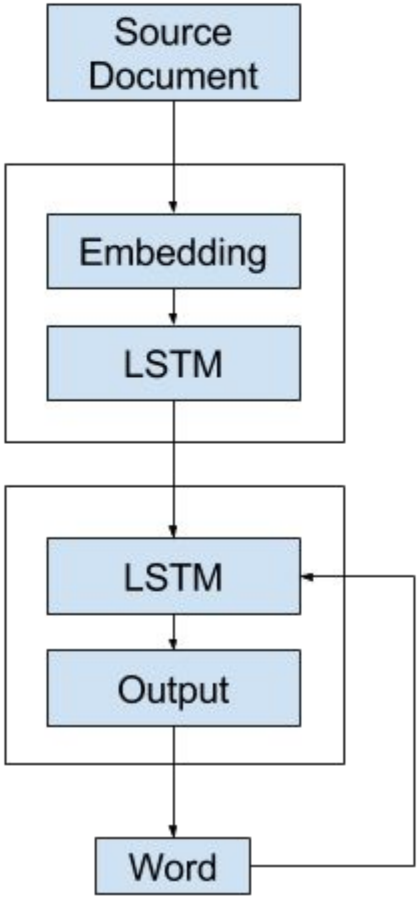
Summary Generation with Deep Learning

Motivation of Summarization

- Language is a the main vehicle of encoding human thought
- Users primarily ingest data and information through text
- Automatically distilling large bodies of text can benefit many parties:
  - Save users time reading, finding relevant information
  - Save creators time writing
  - Increase site traffic by allowing more, if shallower, interactions with articles
- Humans do this easily every day
  - Training a computer to this is another story

Goal - Deep Learning Summarizer

- Functionality
  - Learn from texts and summaries
  - Predict new summaries on unseen text
  - Dynamic to length
  - Abstractive vs Extractive
    - Abstractive,
    - Summarize based on meaning, not just statistics
- Features
  - Sequence to Sequence
  - Encoder, Decoder, and Attention
  - Inference uses previous predicted step as next input
    - Recursive

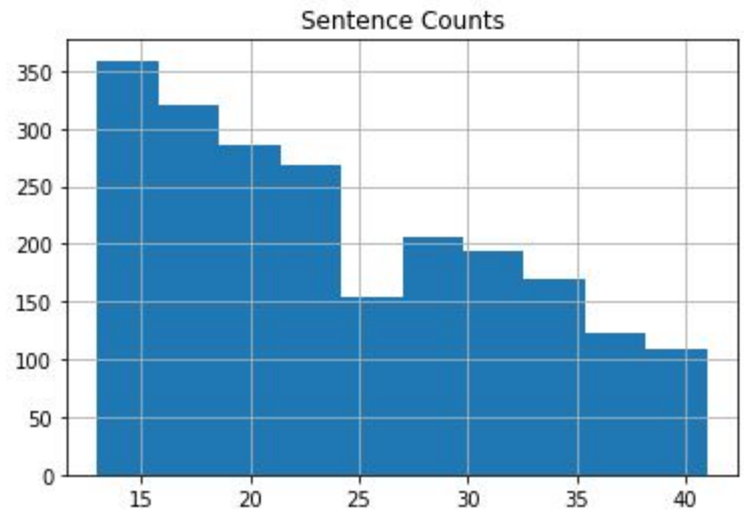


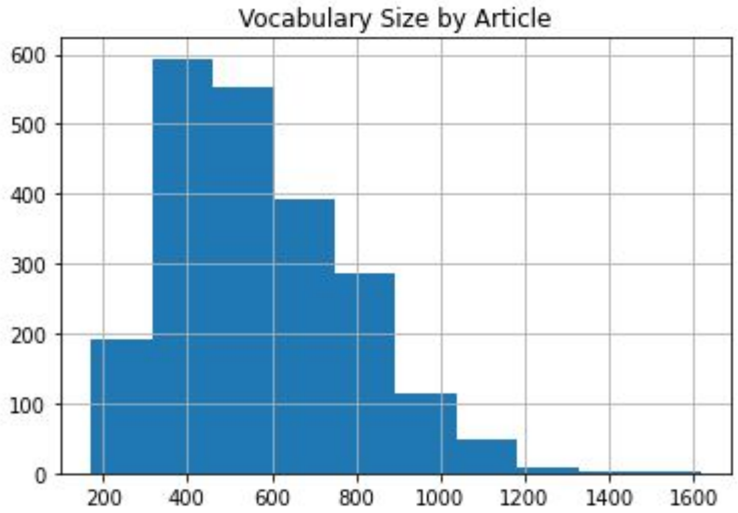
Data - Source and Generation

- News Articles from 2016-2017
  - ABC, CNN, The Huffington Post, BBC, DW News, TASS News, Al Jazeera, China Daily and RTE
  - Collected via RSS feed
- Originally 3,824 Articles
  - Filtered by sentence count, 13 - 36 sentences
  - Left 2,189 articles
- Generated Summaries
  - Gensim's TextRank, ratio summary
  - Longer articles will have longer summaries
  - Extractive Summary Technique

Data - Lexical Features

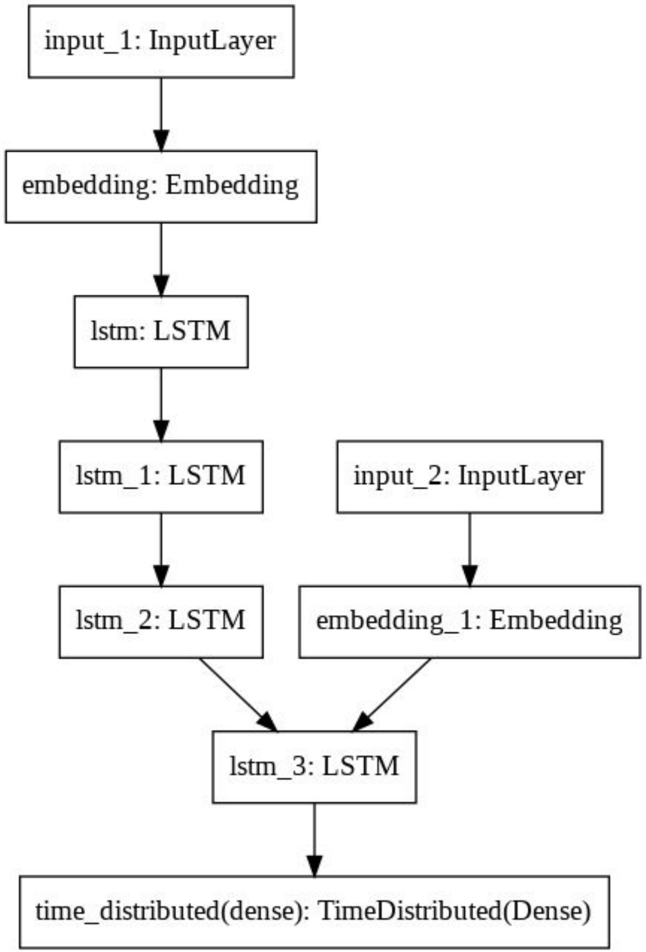
- Source Vocabulary: 78,680 words
- Target Vocabulary: 22,663 words





Neural Network 1:
Simple Encoder/Decoder

- Encoder:
    - Reads in source one word at a time
    - Learns context of sequence
    - Final state initializes decoder
- Decoder:
    - Encoder, but for target minus 1
    - Trained to predict next step in target
- Inference:
    - Cannot use target on test data
    - Instead, encode source, initialize decoder, generate one step, then use predicted word as next input
        - Continues until end token or max length reached



Encoder/Decoder Results

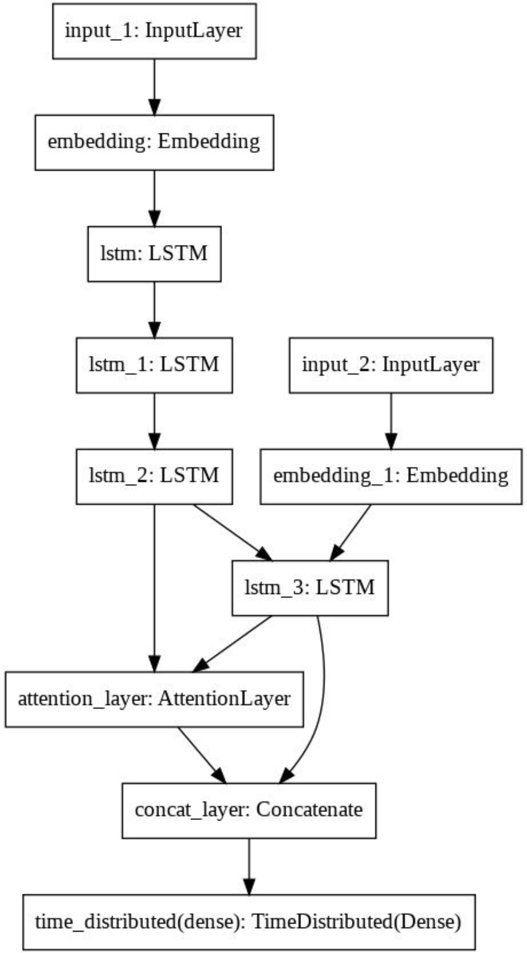
Generated Summary: executive we public she into into while into while about while according according according are russia is is first us first us percent have us and first first foreign on on other all general public all government when were was general was general was u the what police had had that trump's had the had wednesday so had so will thursday u white white police police had trump's had trump's trump's had trump's had the had she out out than than south south south south south south between i i i i when by united of united of she national was there million u more including their south south south south between i i i by united all our our that international no no no group their under their that minister national was russian police police police including time police time police time a a a a a a a when a a a when a a when year year year on on other public report when united she report rights over before before before told before told told during while while while we while we public public his public all according according mr mr mr mr mr our white white white week court he he their their that that minister chinese minister report president his public being state up up police police police time police time a a a a a a a a when a a a when a a when year year year on on other public report when united she report rights over before before before told before told told during while while while we while we public public his public all according according mr mr mr mr mr our white white white week court he he their their that that minister chinese minister report president his public being state up up police

Limitations of Encoder Decoder

- Using the entire source sequence to generate output
- Longer sources lose meaning as they are stored in fixed-length vectors
- “Attention” is a special mechanism that can learn which parts of the sequence are most important
    - Allows for longer inputs to be used
    - Emphasizes different parts of the input sequence that are more relevant to the target sequence

Neural Network 2:
Encoder/Decoder + Attention

- Global Attention Layer
  - Uses all hidden states of encoder (considers the whole input)
  - Learns what parts of source are relevant to target
- Trains similarly to the Simple Model
  - Encoder generates states on source, decoder learns to predict target steps
  - Attention Layer and Decoder are combined to learn what is important in source
  - The emphasized source and decoder are used to predict the next step in target
- Inference is the same, using the source to generate the first prediction, and then feeding the prediction back in.



Encoder/Decoder + Attention Results

Generated Summary: with with officials officials officials thursday public russia minister so minister minister out been been could as officials officials thursday been could officials thursday two officials officials thursday made russia minister minister most been him called most city he most election city he election city group but called also from while more i i to to to to to to before he an he he city more more i i what he their to to to he before an he he city he more election an state state state year year year i by international year told by year told also from south minister minister minister minister minister minister been him donald called police chinese was them national national order national order order national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order betweennational order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order between order

Results Comparison

- Similarities
  - Nonsensical
  - Full, padded target length
  - Both models only generate 1 summary
- Model 1
  - Vocabulary of 77 unique words
  - Many long, single-word sequences
  - More weight on preceding word, less on source (especially long sources)
- Model 2
  - Vocabulary of 45 words
  - Single word sequences are much shorter
  - Three word pattern in back-half of summary

Generated Summary: executive we public she into into while into while about while according according according are russia is is first us first us percent have us and first first foreign on on other all general public all government when were was general was general was u the what police had had that trump's had the had wednesday so had so will thursday u white white police police had trump's had trump's trump's had trump's had the had she out out than than south south south south south south between i i i i when by united of united of she national was there million u more including their south south south south between i i i by united all our our that international no no no group their under their that minister national was russian police police police including time police time police time a a a a a a a when a a a when a a when year year year on on other public report when united she report rights over before before before told before told told during while while while we while we public public his public all according according mr mr mr mr mr our white white white week court he he their their that that minister chinese minister report president his public being state up up police police police time police time a a a a a a a a when a a a when a a when year year year on on other public report when united she report rights over before before before told before told told during while while while we while we public public his public all according according mr mr mr mr mr our white white white week court he he their their that that minister chinese minister report president his public being state up up police

Generated Summary: with with officials officials officials thursday public russia minister so minister minister out been been could as officials officials thursday been could officials thursday two officials officials thursday made russia minister minister most been him called most city he most election city he election city group but called also from while more i i to to to to to to before he an he he city more more i i what he their to to to he before an he he city he more election an state state state year year year i by international year told by year told also from south minister minister minister minister minister minister been him donald called police chinese was them national national order national order order national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order betweennational order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order national order between national order between order national order between order

Interpretations and Notes

- Small sample size limits learning
  - ~1,500 training articles
  - News from around the same time will be about similar topics
- Models cannot differentiate between sources, but do learn important vocabulary
- Models are not learning the “start” and “stop” tokens
- There is a desirable change by adding the attention layer

Improvements

- Data
  - Text data can always be cleaner
  - Shorter inputs, organic targets
  - Identify how “start” and “stop” tokens are being lost
  - More data would demonstrate effects more clearly
  - Experiment with stop words
- Models
  - More layers may learn longer sequences better
  - Experiment with different attention mechanisms
  - Pretrained Embeddings (BERT)
  - Tensorflow debugging

Use Cases

- Online Publications
  - Articles tend to be “top heavy”
  - Train on first few paragraphs and existing summaries
  - Automate summarization writing for writers and editors
- News aggregating services
  - May not have editors
  - More removed from authors
- Online stores
  - Many reviews already have user summaries
  - Generate premade summaries, dropdowns

Questions?

Thank You!

Appendix - Links and Resources

- Data Source: Harvard
  - https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GMFCTR
- Machine Learning Mastery Resources:
  - Encoder/Decoders: https://machinelearningmastery.com/encoder-decoder-long-short-term-memory-networks/
  - Encoder/Decoder for Text Summarization https://machinelearningmastery.com/encoder-decoder-models-text-summarization-keras/
- AnalyticsVidhya https://www.analyticsvidhya.com/blog/2019/06/comprehensive-guide-text-summarization-using-deep-learning-python/
- Github Repo: https://github.com/DJGump/text_summary_tensorflow