Key Question: Can we combine knowledge of article structure with RNNs to predict headlines?

Data
- Used “All the News” dataset by Andrew Thompson
- Processed 96,543 articles from 12 sources
- Example article:
  HEADLINE: Hillary Clinton: Ser Notable Moments
  ARTICLE: The name Hillary Rodham first appeared in the pages of The New York Times 67 years ago in a 1969 article about her commencement address at Wellesley the women’s college in Massachusetts...

Model
- RNN ARCHITECTURE:

<table>
<thead>
<tr>
<th>Layer Type</th>
<th>Output Shape</th>
<th>Parameters #</th>
</tr>
</thead>
<tbody>
<tr>
<td>embedding</td>
<td>(None, 50, 100)</td>
<td>4000000</td>
</tr>
<tr>
<td>lstm_1</td>
<td>(None, 50, 512)</td>
<td>1256542</td>
</tr>
<tr>
<td>dropout_1 (Dropout)</td>
<td>(None, 50, 512)</td>
<td>0</td>
</tr>
<tr>
<td>lstm_2</td>
<td>(None, 50, 512)</td>
<td>2099800</td>
</tr>
<tr>
<td>dropout_2 (Dropout)</td>
<td>(None, 50, 512)</td>
<td>0</td>
</tr>
<tr>
<td>lstm_3</td>
<td>(None, 50, 512)</td>
<td>2099800</td>
</tr>
<tr>
<td>dropout_3 (Dropout)</td>
<td>(None, 50, 512)</td>
<td>0</td>
</tr>
<tr>
<td>simplecontext_1</td>
<td>(None, 25, 944)</td>
<td>0</td>
</tr>
<tr>
<td>time_distributed_1 (TimeDist)</td>
<td>(None, 25, 40000)</td>
<td>37900000</td>
</tr>
<tr>
<td>activation_1 (Activation)</td>
<td>(None, 25, 40000)</td>
<td>0</td>
</tr>
</tbody>
</table>

Total parameters: 47,293,824
Non-trainable parameters: 0

Objective
- Based on previous study by Lopyrev
- Use encoder to process article word-by-word, decoder to generate headline word-by-word
- Lopyrev used the first 50 words of each article, can we accomplish better summarization using other parts of the article?

Model (continued)
- Hyperparameters: Adam Optimizer, first moment 0.9, second moment 0.999, epsilon 1e-8, learning rate 0.0001
- Loss function: Categorical cross entropy loss, minimized over 144 iterations of training:

```
<table>
<thead>
<tr>
<th>BLEU Score</th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.01365756221</td>
<td>0.01451416009</td>
<td>0.01220555188</td>
</tr>
<tr>
<td>25 + 25</td>
<td>0.01334964053</td>
<td>0.01323200369</td>
<td>0.0129793721</td>
</tr>
<tr>
<td>50 + 25</td>
<td>0.01362171917</td>
<td>0.01177343727</td>
<td>0.01119147869</td>
</tr>
</tbody>
</table>

Levenshtein
<table>
<thead>
<tr>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>54.77721648</td>
<td>54.52419509</td>
</tr>
<tr>
<td>25 + 25</td>
<td>54.55219124</td>
<td>55.39320007</td>
</tr>
<tr>
<td>50 + 25</td>
<td>54.92491911</td>
<td>54.19084872</td>
</tr>
</tbody>
</table>

- Where 50 corresponds to inputting the first 50 words of an article, 25 + 25 corresponds to the first 25 and last 25 words, and 50 + 25 corresponds to the first 50 and last 25 words
- Objective: maximize BLEU score, minimize Levenshtein score

Discussion
- Selected (cherry-picked) examples:
  - TRUE: Federal Judge blocks Obama Administration Protections For Transgender People*
  - PREDICTED: Federal Judge blocks Obama Administration Protections For Transgender People*
  - TRUE: Neil deGrasse Tyson and Al Gore on the future of our planet and everything else
  - PREDICTED: Neil deGrasse Tyson and Al Gore on the future of our planet and everything else
  - TRUE: Brothers share what it was like quitting their corporate jobs to sell ties on the beach and founds' Vineyard Vines, a company worth nearly $1 billion
  - PREDICTED: Brothers share what it was like quitting their corporate jobs to sell ties on the beach and workedin' Vineyard workedin' a company worth a theory

- BLEU score results match Lopyrev’s early results, but with further training he is able to attain an average BLEU score of about 0.09
- No evidence of overfitting
- Did not effectively leverage different parts of article
- Difficulty predicting <EOS> token

Future
- Train until convergence, prohibitive cost led to non-optimal results
- Further experiment with hyperparameters, architecture
- Clean data, prevalent metanoun-article text caused problems
- How to better leverage key parts of article for summarization (include parts other than exclusively introduction and conclusion)?

References
- See paper for complete list of references