# A Experimental Setup

## Data Preprocessing

We preprocess the datasets with the scripts provided by See et al. (2017), which uses Stanford CoreNLP tokenization Manning et al. (2014). We use the non-anonymized version of the CNN/DM as in previous summarization work. For the New York Times Corpus, we filter out the examples with abstracts shorter than 50 words following the criteria in (Durrett et al., 2016), yielding the NYT dataset. The statistics of the datasets are listed in Table 1. During sentence selection, we always select 3 sentences for CNN/DM and 5 sentences for NYT, which gave the best performance. For our syntactic analysis, all datasets are parsed with the constituency parser in Stanford CoreNLP (Manning et al., 2014).

## Implementation Details

We use the same pre-trained word embeddings used in (Narayan et al., 2018). The size of the sentence and document representation vectors is 200. For the compression module, we use ELMo as the contextualized encoder without fine-tuning the parameter and project the vectors back to 200 dimensions after the ELMo layer. Dropout is applied after word embedding layers and LSTM layers at a rate of 0.2. We use the Adam optimizer (Kingma and Ba, 2014) with the initial learning rate at 0.001. The model converges after 2 epochs of training. In initial experiments, we also found ELMo to be useful for sentence selection as well. However, to simplify comparisons with past work and due to scaling issues, we use it for compression only. We use ROUGE (Lin, 2004) for evaluation. During oracle construction, we use simplified unigram and bigram F1 scores as a faster approximation to the full ROUGE.

## References


---

1. Command line parameters: “-c 95 -m -n 2”

---

Table 1: Statistics of the CNN, Daily Mail, and NYT (see text) datasets. CNN features the shortest reference summaries overall, and this is where we find compression is most effective.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Train</th>
<th>#Dev</th>
<th>#Test</th>
<th>Ref len</th>
<th>Doc len</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>90266</td>
<td>1220</td>
<td>1093</td>
<td>37</td>
<td>540</td>
</tr>
<tr>
<td>DM</td>
<td>196961</td>
<td>12148</td>
<td>10397</td>
<td>61</td>
<td>593</td>
</tr>
<tr>
<td>NYT</td>
<td>137778</td>
<td>17222</td>
<td>17223</td>
<td>88</td>
<td>727</td>
</tr>
</tbody>
</table>

Table 2: Statistics of compression options in CNN. We show the top four constituency types that are compressible, along with the average length, the fraction of available compressions it accounts for, and how frequently the oracle says to compress these constituents.

<table>
<thead>
<tr>
<th>Node Type</th>
<th>Len</th>
<th>% of comps</th>
<th>Oracle comp %</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP</td>
<td>5.7</td>
<td>39%</td>
<td>67%</td>
</tr>
<tr>
<td>JJ</td>
<td>1.0</td>
<td>19%</td>
<td>84%</td>
</tr>
<tr>
<td>SBAR</td>
<td>12.1</td>
<td>11%</td>
<td>59%</td>
</tr>
<tr>
<td>ADVP</td>
<td>1.3</td>
<td>7%</td>
<td>91%</td>
</tr>
</tbody>
</table>

B Turk Instructions

Figure 1 shows the interface for Amazon turk human evaluation.

C Type Analysis

In Table 2, we show the statistics of the compression options in CNN. PP attachment and adjectives are the top 2 compression options and according to the oracle, more than half of PP and almost all of the adjectives are compressible without hurting the ROUGE.
Instructions:
In this task you will be reading text and evaluating it on the basis of how grammatical it is.
- A grammatical sentence should consist of well-formed English writing that flows naturally.
- It should not be a sentence fragment or have missing components, and it should be understandable in terms of what it is trying to communicate.
- You are asked to RANK three similar sentences based on their grammaticality.
- Note: You should not pick the same sentence as best and also second best.

Set 1

A. Hundreds of decomposed corpses were discovered buried in shallow graves in the streets of Damasak this past weekend, according to officials and a resident.

B. Kano, Nigeria (CNN) Hundreds of decomposed corpses were discovered buried in shallow graves in the streets of

C. Nigeria (CNN) Hundreds of decomposed corpses were discovered buried in shallow graves in the streets of northeastern Nigerian town Damasak this past weekend, according local officials resident.

Which sentence is the BEST one?
- [ ] A
- [ ] B
- [ ] C