Evaluation of Neural and Non-Neural Extractive Single Document Summarization Techniques

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

Ron Fan & Aditya Saraf | NLP Capstone CSE481N Spring 2018 | University of Washington

Abstract

Current approaches to single document summarization are largely divided between neural and combinatorial approaches.

We explore the differences between the two types of models to draw conclusions about the strengths and weaknesses of each model.

Data & Experiments

Preprocessing:

- ✓ Abstract summaries → extracted summaries
- ✓ Tokenization
- ✓ Stoplist filtering
- ✗ Lemmatization
- ✗ Explicit feature tagging (POS, NER)

Data:

- CNN/DailyMail
  - Training set: ~310,000 documents with greedily generated labels
  - Test set: 10,000 documents
- Australian Law Cases
  - Additional Test set: 4,000 documents in a different domain
  - Compares domain adaptability

Results:

Model	ROUGE-1	ROUGE-1 F1	ROUGE-2	Runtime
First-K	0.410247	0.252497	0.174275	35048
Unweighted Greedy MCP	0.533134	0.237928	0.183332	38967
MCP	0.547118	0.262058	0.207809	71355
Neural	0.5397474	0.292089	0.229425	1145616
Oracle	0.68059	0.37807	0.38756	N/A

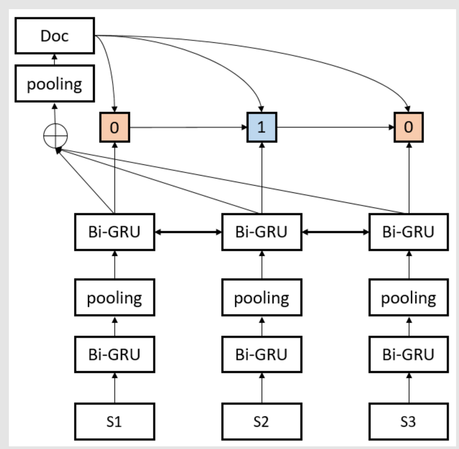
Law articles:

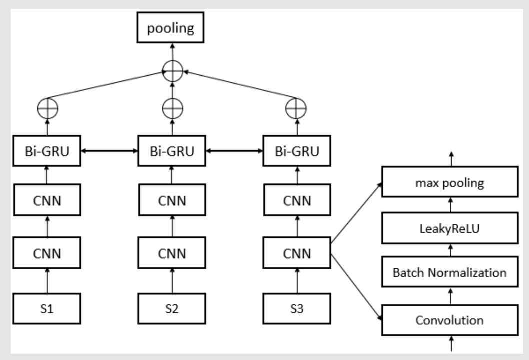
Model	ROUGE-1	ROUGE-1 F1	ROUGE-2	Runtime
MCP	0.56808	0.11932	0.17202	176708
Neural	0.54434	0.15179	0.18272	1015356
Oracle	0.66658	0.17366	0.32877	N/A

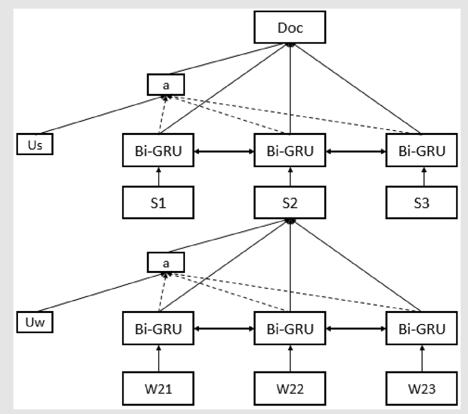
Neural Model

- SummaRuNNer derivatives
  - Multi-layer RNN
  - Hierarchical Attention Networks
  - Convolutional Networks
- Binary cross entropy loss
- Adam optimizer
- Feed-forward classification layer

$$
\begin{aligned}
P(y_j = 1|\mathbf{h}_j, \mathbf{s}_j, \mathbf{d}) = \sigma(W_c\mathbf{h}_j \quad & \texttt{\#(content)} \\
+\mathbf{h}_j^T W_s \mathbf{d} \quad & \texttt{\#(salience)} \\
-\mathbf{h}_j^T W_r \tanh(\mathbf{s_j}) \quad & \texttt{\#(novelty)} \\
+W_{ap}\mathbf{p}_j^a \quad & \texttt{\#(abs. pos. imp.)} \\
+W_{rp}\mathbf{p}_j^r \quad & \texttt{\#(rel. pos. imp.)} \\
+b), \quad & \texttt{\#(bias term)}
\end{aligned}
$$







Combinatorial Model

- Weighted Maximum Coverage Problem:

Sentences: $S = \{S_1, S_2, \ldots, S_n\}$
Summary Length: k sentences
Weighted elements (words)

→

Summary: $S' \subseteq S, |S| \leq k$
Maximize the weighted sum of unique words

- Coverage Heuristic: Assumptions
  - Including a word multiple times won't increase semantic content
  - Words are the Elementary Discourse Units of semantic information

Advantages:

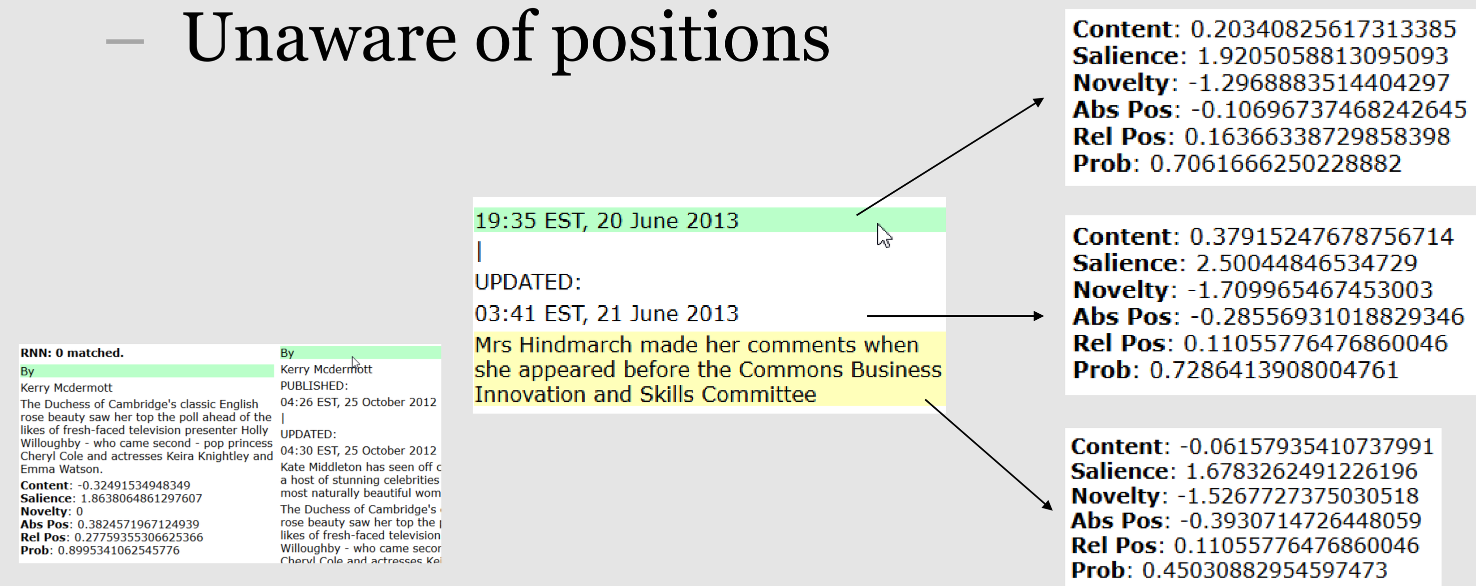
- Extremely fast: ~80 ms per article
- Competitive performance
- Intuitive

Disadvantages:

- Doesn't maximize ROUGE-1 on target domain
- Can only limit length at the sentence level
- Exponential slowdown with larger inputs

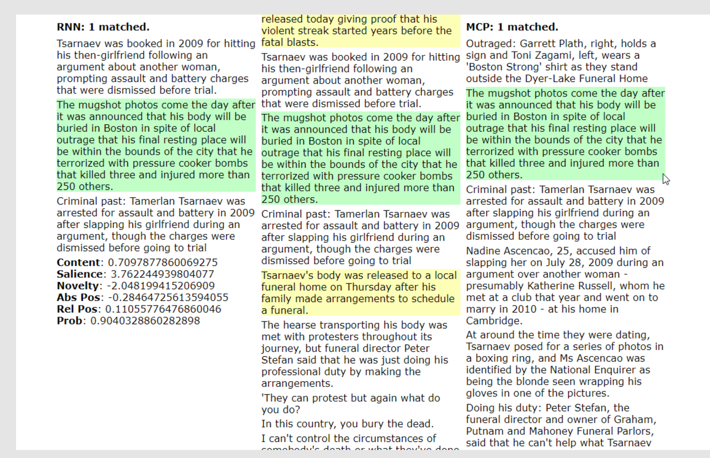
Observations and Errors

- Neural model:
  - Positional scores are overweighted
  - Definition of salience is unclear
  - Strongly prefers early sentences
- Combinatorial model:
  - Tendency to choose long sentences
  - Largely ignores semantic meaning
  - Unaware of positions



Visualizer

- Tool for easily comparing neural and non-neural understanding of documents
- Scores extracted from models



References

Diederik et al., arXiv:1412.6980 [cs.LG]
Hermann et al., arXiv:1506.03340 [cs.CL] Nallapati et al., arXiv:1611.04230 [cs.CL]
hpzhao, https://github.com/hpzhao/SummaRuNNer

Durrett, G., Berg-Kirkpatrick, T., & Klein, D. (2016). Learning-Based Single-Document Summarization with Compression and Anaphoricity Constraints. *arXiv:1603.08887 [Cs]*. Retrieved from http://arxiv.org/abs/1603.08887

Hirao, T., Yoshida, Y., Nishino, M., Yasuda, N., & Nagata, M. (2013). Single-Document Summarization as a Tree Knapsack Problem. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1515–1520). Seattle, Washington, USA: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/D13-1158

Khuller, S., Moss, A., & Naor, J. (Seffi). (1999). The Budgeted Maximum Coverage Problem. *Inf. Process. Lett.*, *70*(1), 39–45. https://doi.org/10.1016/S0020-0190(99)00031-9

Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In S. S. Marie-Francine Moens (Ed.), *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop* (pp. 74–81). Barcelona, Spain: Association for Computational Linguistics.

Nishikawa, H., Hirao, T., Makino, T., & Matsuo, Y. (2012). Text Summarization Model based on Redundancy-Constrained Knapsack Problem. Presented at the COLING.