

Transfer Learning for Abstractive Text Summarization with Pointer-Generator Networks

Rutika Banakar

School of Information, University of California, Berkeley

rutikasb@berkeley.edu

Abstract

In this work, we model abstractive text summarization with hybrid pointer-generator networks on the CNN/Daily Mail dataset and transfer the learnings to New York Times Annotated Corpus. The idea of transfer learning is to use knowledge learnt from one task or domain to apply to a new task or domain instead of starting from the scratch. Through this work, we attempt to study the transfer of learnings from neural networks trained over CNN/DailyMail dataset to the NYT corpus.

1 Introduction

Summarization is an important challenge of natural language understanding. The aim is to produce a condensed representation of an input text that captures the core meaning of the original. Most successful summarization systems utilize extractive approaches that crop out and stitch together portions of the text to produce a condensed version. In contrast, abstractive summarization attempts to produce a bottom-up summary, aspects of which may not appear as part of the original.

Transfer Learning uses the knowledge learnt by solving a problem pertaining to one task or domain (source) and apply it to a different task or domain (target). In the field of computer vision for example, models are rarely trained from scratch and features extracted from a state-of-the-art CNN pre-trained on ImageNet is used to train a new model. Similarly, we could use weights learned from a model trained over one corpus and fine tune it with a different corpus.

In this work, we train a pointer-generator network on CNN/Daily Mail corpus and then transfer the knowledge to NYT corpus. Although both the

corpora are from same domain i.e., new articles, the NYT corpus contains articles for restaurant reviews, book review, opinion articles, interview articles and other varied articles. Hence the transfer of learning is not entirely trivial. And it will also help to avoid over-fitting to the CNN/Daily Mail data.

2 Related Work

2.1 Pointer-Generator Networks

(Rush et al., 2015) were the first to apply abstractive text summarization on DUC 2004 and Gigaword datasets using feed-forward neural language model. Their approach was further improved upon with recurrent neural networks (Chopra et al., 2016), abstract meaning representations (Takase et al., 2016), hierarchical networks (Nallapati et al., 2017). (Nallapati et al., 2016) used Attention Encoder-Decoder RNNs and showed that they achieve state-of-the-art performance on the Gigaword and DUC corpora. They also released a new CNN/Daily Mail dataset with very long source documents and multi-sentence summaries, and established performance benchmarks on this dataset for further research. The same authors used hierarchical RNNs (Nallapati et al., 2017) and found that it significantly outperformed their previous model in terms of the ROUGE metric.

(See et al., 2017) used a hybrid between sequence-to-sequence attention model which is similar to the model proposed in (Nallapati et al., 2016) and a pointer network (Vinyals et al., 2015) on the CNN/Daily Mail dataset and showed that it out performed the benchmarks set by (Nallapati et al., 2016) by at least 2 ROUGE points.

2.2 CNN/Daily Mail Corpus

(Hermann et al., 2015) created two datasets using news articles for Q&A research. The datasets con-

tain 90k and 197k documents. (Nallapati et al., 2016) modified these datasets for the task of text summarization. The original authors used the human generated abstractive summary bullets from new stories in *CNN* and *Daily Mail* websites as questions and the news stories as the corresponding passages. They also released the scripts to crawl and extract pairs of passages and questions from these websites. With a simple modification to this script (Nallapati et al., 2016) were able to convert the summary bullets of each article into a multi-sentence text summary. The corpus has 286,817 training pairs, 13,368 validation pairs and 11,487 test pairs.

2.3 New York Times Annotated Corpus

(Gillick and Dunietz, 2014) released the New York Times Annotated corpus which contains 1.8 million New York Times articles. The dataset is aimed to encourage research in leveraging background information about entities (entity salience). The dataset however, also contains about 600,000 articles with hand written summaries.

3 Methods

The pointer-generator is a hybrid network that can choose to copy words from the source via pointing, while retaining the ability to generate words from the fixed vocabulary. This model makes it easy to copy words from the source text and even copy out-of-vocabulary words from the source text.

We first train a pointer-gen network on CNN/Daily Mail dataset, which will be our source task/domain, resulting in a *source model*. We'll then perform a number of transfer experiments with varying number of article-summary pairs from NYT corpus which is our target task/domain.

4 Results

Acknowledgments

References

Sumit Chopra, Michael Auli, and Alexander M. Rush. 2016. [Abstractive sentence summarization with attentive recurrent neural networks](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 93–98. Association for Computational Linguistics.

Dan Gillick and Jesse Dunietz. 2014. A new entity salience task with millions of training examples. In *Proceedings of the European Association for Computational Linguistics*.

Karl Moritz Hermann, Tomáš Kociský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. [Teaching machines to read and comprehend](#). *CoRR*, abs/1506.03340.

Ramesh Nallapati, Bing Xiang, and Bowen Zhou. 2016. [Sequence-to-sequence rnns for text summarization](#). *CoRR*, abs/1602.06023.

Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. [Summarunner: A recurrent neural network based sequence model for extractive summarization of documents](#). In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 3075–3081.

Alexander M. Rush, Sumit Chopra, and Jason Weston. 2015. [A neural attention model for abstractive sentence summarization](#). *CoRR*, abs/1509.00685.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. [Get to the point: Summarization with pointer-generator networks](#). *CoRR*, abs/1704.04368.

Sho Takase, Jun Suzuki, Naoaki Okazaki, Tsutomu Hirao, and Masaaki Nagata. 2016. [Neural head-line generation on abstract meaning representation](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1054–1059. Association for Computational Linguistics.

O. Vinyals, M. Fortunato, and N. Jaitly. 2015. [Pointer Networks](#). *ArXiv e-prints*.

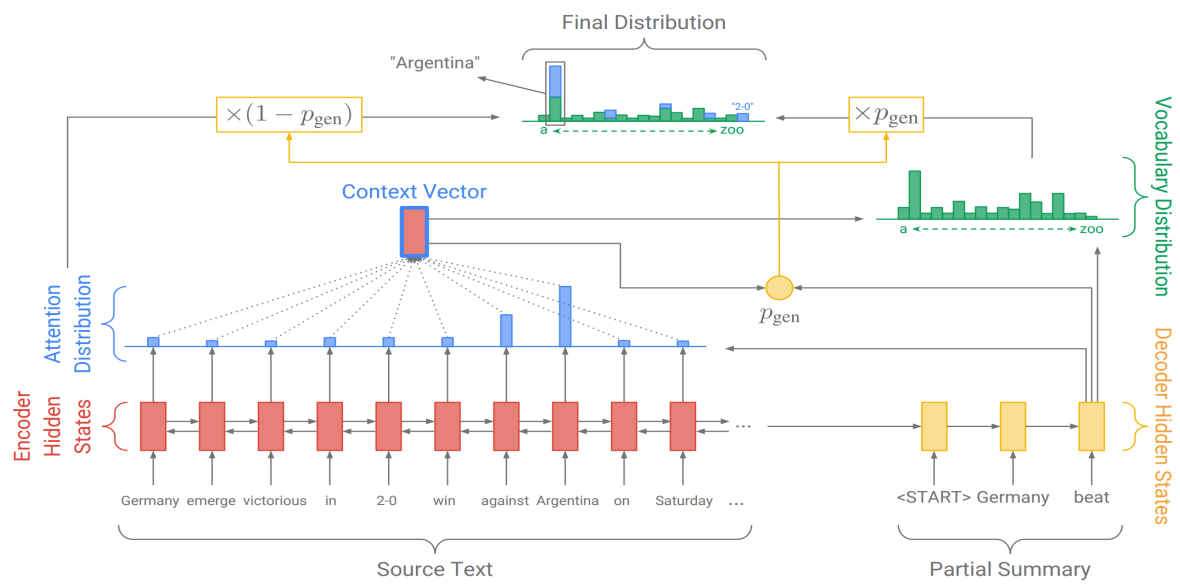


Figure 1: Pointer-generator model (See et al., 2017)