

UKRAINIAN CATHOLIC UNIVERSITY

APPLIED SCIENCES FACULTY

DATA SCIENCE MASTER PROGRAMME

Generative text summarization with Recurrent Neural Networks

Machine Learning project report

Authors:

Hanna PYLIEVA

Yuriy MYKHALCHUK

Irynei BARAN

June 3, 2018



APPLIED
SCIENCES
FACULTY ●

Abstract

Invent a meaningful abstract.

In this work, we cast text summarization as a sequence-to-sequence problem and apply the attentional encoder-decoder RNN that has been shown to be successful for Neural Machine Translation [1].

1 Introduction

Invent a meaningful introduction. Per Shelpuk: Good project will be dedicated to an important problem and have a clear vision of what value it can bring to the potential users. All stages will be explored and analyzed, the approach for each of them is selected thoughtfully, compared to the alternatives and clearly explained. The results are evaluated and explained (explanation should provide additional information, not just restating the results or the code in English).

In the modern Internet age, textual data is ever increasing. According to this each Internet user will highly benefit from condensing data while preserving the information and meaning. This idea is a driver of growing interest among the research community for developing new approaches to automatically summarize the text. Automatic text summarization system generates a short summary that captures the main ideas of an input text. Since the advent of text summarization in 1950s, researchers have been trying to improve techniques for generating machine summaries which are not worse than human made summaries [2].

There are two prominent types of summarization algorithms.

- Extractive summarization copies parts of the source text through some measure of importance and then combines those part/sentences together to render a summary. Importance of sentence is based on linguistic and statistical features.
- Abstractive summarization generates new phrases, possibly rephrasing or using words that were not in the original text. Naturally abstractive approaches are harder as it involves robust natural language processing. For perfect abstractive summary, the model has to first understand the document and then express that understanding in succinct form possibly using new words and phrases. Abstractive summarization has complex capabilities like generalization, paraphrasing and incorporating real-world knowledge [3].

Majority of the work has traditionally focused on extractive approaches due to the easy of defining hard-coded rules to select important sentences than generate new ones. Also, it promises grammatically correct and coherent summary. But they often don't summarize long and complex texts well as they are very restrictive.

Abstractive methods, on the other hand, provide highly powerful and promising results. That is why in this project we implemented an algorithm for abstractive text summarization to build solid understanding if this approach and discover how it can be improved.

2 Model

2.1 Overview

Models for abstractive summarization fall under a larger deep learning category called sequence-to-sequence models, which map from an input sequence to a target sequence. We used [4] as a baseline of our work.

References

- [1] Bahdanau,D.; Cho,K.; Bengio Y. *Neural machine translation by jointly learning to align and translate*. CoRR, abs/1409.0473, 2014. <http://arxiv.org/abs/1409.0473>.
- [2] Gambhir, M. and Gupta, V. *Recent automatic text summarization techniques: a survey*. Artificial Intelligence Review, 47(1):1-66, 2017.
- [3] Singhal,S. and Bhattacharya, A. *Abstractive Text Summarization*. Department of Computer Science IIT Kanpur, 2017.
- [4] Lopyrev, K. *Generating News Headlines with Recurrent Neural Networks*. CoRR, abs/1512.01712, 2015.
- [5] Bishop, Christopher M. *Pattern Recognition and Machine Learning*. Springer, Cambridge, U.K., 2006.