Text Summarization Using TensorFlow

Harsh Vora – A20445400 **Dimple Mehra - A20457225**

Research papers Used:

Paper-1: Extractive Text Summarization Using Deep Learning

Author's: Nikhil S. Shirwandkar, Samidha Kulkarni

Year of publication: 2019

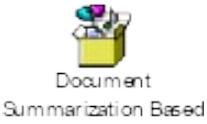
Name of publication: IEEE (2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA))

Paper-2: Extractive Document Summarization Based on Convolutional Neural Networks

Author's: Yong Zhang, Meng Joo Er and Mahardhika Pratama

Year of publication: 2016

Name of publication: IEEE (IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society



Abstract:

Text Summarization is the process of obtaining salient information from an authentic text document. In this technique, the extracted information is achieved as a summarized report and conferred as a concise summary to the user; it is challenging for the user to verve through altogether the information accessible on web.

Reduction of text is a very complex problem which, in spite of the progress in the area thus far, poses many challenges to the scientific community. It is also relevant application in today's information society given the exponential growth of textual information online and the need to promptly assess the contents of text collections. It has long been assumed that summarization presupposes to understand the input text, which means for identifying the important point of the document, explicit representation of the text must be calculated therefore, text summarization became an interesting application to test the understanding capabilities of artificial systems.

Problem Statement:

Earlier, humans used to summarize the text by their own, but today due to increasing data, it is difficult for the human beings to cope up with the huge data due to which the time required for the users to summarize and analyse the huge data is increased.

The solution to reduce reading time of the user is producing a succinct document summary.

Approach:

The objective of this project building a neural network with TensorFlow that can create relevant summaries for the dataset. To build our model we will use a two-layered bidirectional RNN with LSTMs on the input data and two layers, each with an LSTM on the target data. The deep network power is capable of

1. analysing sequences of input
2. understanding text
3. outputting sequences of output in form of summarizes

hence the name of sequence 2 sequence(seq2seq), sequence of inputs to sequence of outputs, which is the main algorithm that we use in building the model.

Datasets (We will use either of these three datasets):

We will demonstrate the model on three different datasets mentioned below:

First dataset consists of reviews of fine foods from amazon. It contains, as the name suggests, 570.000 reviews of fine foods from Amazon and summaries of those reviews. Our aim is to input a review (Text column) and automatically create a summary (Summary colum) for it.

https://www.kaggle.com/snap/amazon-fine-food-reviews/data

Second dataset is “All the news” available on Kaggle which comprises articles from different publications(New York Times, Breitbart, CNN, Business Insider, the Atlantic, Fox News, Talking Points Memo, Buzzfeed News, National Review, New York Post, the Guardian, NPR, Reuters, Vox, and the Washington Post) mainly from the beginning of 2016 to July 2017.

https://www.kaggle.com/snapcrack/all-the-news

Third dataset used is “GitHub Issues” which contains over 8 million GitHub issue titles and descriptions from 2017.

https://www.kaggle.com/davidshinn/github-issues

References:

1. https://ieeexplore.ieee.org/document/8697465
2. https://www.google.com/search?q=ieee+paper+for+text+summarizer+using+cnn&oq=ieee+paper+for+text+summarizer+using+cnn&aqs=chrome..69i57.14666j0j7&client=ms-android-google&sourceid=chrome-mobile&ie=UTF-8