News Article Summarizer

-by

Gautam shanbag

Aspril Tauro

**Abstract-**

News article Summarizer was created using 4 models based on a technique called as Extractive summarizer. Every model gave summary of ¼ size of the article reducing the size of article while keeping the context intact. The report gives detail explanation of the use case, techniques used, deployment, reference’s and time lines.

**Use case-**

In today’s fast paced, hustling and busy life it’s difficult to stay updated with the news and going through the whole article to have better depth of understanding is perilous task in itself.

And our team wanted to build a product which would allow user the user to have the context of news without going through the full article.

**It is said a person with half the knowledge is more dangerous than a person with no knowledge**, and this was our challenge as it was important that the summary of the news convey all the major plots that article aims to convey.

**Techniques used-**

There are 2 types of Summarizer:

1. Extractive Summarizer
2. Abstractive Summarizer

The summarizer we created is based on extractive summarizer.

The methods used were,

1. Word Counts.
   1. The sentences are scored based on the number of lemmatized words contained.
   2. The scored sentences are arranged in descending order based on their scores.
   3. The summary output contains the number of sentences with top scores.
2. Entity recognition and dependency parsing
   1. In any article, named entities are very important.
   2. The words dependent on the named entity tell us what action the named entity is going to take.
   3. Therefore Based on this principle using dependency parsing and named entities assigned scores to sentences and choosing the best N sentences for the summary.
3. TextRank
   1. TextRank works on similar lines of a PageRank.
   2. TextRank instead of webpages we use sentences.
   3. Similarity between any 2 sentence’s is used as equivalent to the webpage transition probability.
4. Restricted Boltzman machine.
   1. Restricted Boltzman machine is based on deep belief network called as Boltzman machine neural network
   2. In Boltzman Machine every node is connected with every other node.
   3. And this poses as its limitation as it is very complicated to implement such a huge network.
   4. Hence certain rules imposing restriction over Boltzman machine is done, which restricts the node in a layer to connect to nodes of different layer but not to nodes of the same layers. Hence it is called as Restricted Boltzman machine.
   5. The main analogy of using RBM is to get the latent factors between sentence’s to form a summary.

**Deployment-**

The project is deployed locally using Flask frame work. We aim to deploy the project on the server using Amazon web services.

**References-**

* Generating News Headlines with Recurrent Neural Networks-<https://arxiv.org/pdf/1512.01712.pdf>
* Restricted Boltzmann Machine <https://arxiv.org/pdf/1708.04439.pdf>
* A Neural Attention Model for Abstractive Sentence Summarization - <https://arxiv.org/pdf/1509.00685.pdf>
* Experiments in Automatic Text Summarization Using Deep Neural Networks - <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.8775&rep=rep1&type=pdf>

**Deliverables-**

|  |  |
| --- | --- |
| **Milestones** |  |
| Obtaining Data | Completed |
| Data exploration and pre processing | Completed |
| Exploring various techniques and checking the effectiveness | Completed |
| Model creation | Completed |
| Issues and evaluation | Completed |
| Deployment and GUI | Deployed locally |