International Journal for Modern Trends in Science and Technology, 6(12): 398-401, 2020
Copyright © 2020 International Journal for Modern Trends in Science and Technology
ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST061275
Available online at: http://www.ijmtst.com/vol6issue12.html

Implementing Supervised Approach to Summarization of Research Papers

Shaguna Awasthi[1]

[1]Department of Information Technology, Delhi, India

To Cite this Article
Shaguna Awasthi, "Implementing Supervised Approach to Summarization of Research Papers", *International Journal for Modern Trends in Science and Technology*, 6(12): 398-401, 2020.

Article Info
Received on 16-November-2020, Revised on 09-December-2020, Accepted on 12-December-2020, Published on 18-December-2020.

ABSTRACT

Using automatic text summarization we can reduce a document to its main information or to what is known as crux of the document .Recent research in this zone has zeroed in on neural ways to deal with summarisation, which can be very data hungry.

This paper aims to explore a quicker way by implementing a supervised-learning based extractive summarisation system for the summarisation of research papers.

This paper also explores the possibility of any section, in a research paper being the prime section to generate summaries by utilizing ROUGE scores. An easy to implement and intuitive model is developed using glove embeddings and doc2vec to encode sentences and documents in their local and global context producing grammatically coherent summaries.

KEYWORDS: *Text Summarization, Research Papers, Abstractive model, Rouge Score, Glove*

I. INTRODUCTION

Reducing a document to its main points is known as text summarization. There are two streams of summarization approaches: extractive summarization, which copies parts of a document (often whole sentences) to form a summary, and abstractive summarization, which reads a document and then generates a summary from it, which can contain phrases not appearing in the document. [17]

Here, we are focusing on summarizing research papers. Since research papers with a specialized space of technical dominance contain ordinary and expressive language, this paper opts for extractive summarization. Despite the fact that there has been has been work on summarization research papers previously, however The fundamental issue with summarization emerges from the absence of labeled training data and the existing datasets are very small, consisting of tens of documents[2] (Kupiec et al., 1995; Visser and Wieling, 2009). Such little datasets are not adequate to train supervised summarization models.

This obstacle has been dealt with, through this paper: "A Supervised Approach to Extractive Summarization of Scientific Papers" [1]. They give us a path to mine research papers and furthermore furthermore make fundamental summaries that can be utilized as training data. Moreover, they proposed various strategies, both analytical and observational to make text summaries utilizing everything from scoring sentences dependent on word overlap to advanced methods. "A Supervised Approach to Extractive Summarization of Scientific Papers" [1] was the basis for procuring datasets for training the model, where we explored extractive methods to generate summaries

II.RELATED WORK

Early work on extractive summarisation centers solely around simple to compute statistics, for instance word frequency[3] (Luhn, 1958), location in document[4]

(Baxendale, 1958), and TF-IDF[5] (Salton et al., 1996). Exploration of more aspects such as sentence position [6](Yang et al., 2017), sentence length [7](Radev et al., 2004), words in the title, presence of formal nouns, places or things, word recurrence [8](Nenkova et al., 2006) followed.Recent methods ways deal with extractive summarisation have generally centered around neural methodologies, CBOW embeddings approaches [9](Kobayashi et al., 2015; Yogatama et al., 2015) or encoding entire papers with CNNs or potentially RNNs [10](Cheng and Lapata, 2016).

a graph-based ranking model for text processing[11](Rada Mihalcea and Paul Tarau). A stochastic graph-based method was proposed by Dragomir R. Radev [12]for the computation of relative importance of textual portions in NLP.

H. P. Luhn[3] presented how Statistical data can be derived from word recurrence and distribution can be utilized by the machine for computing a relative measure of importance.Makbule Gulcin Ozsoy,Ferda Nur Alpaslan,Ilyas Cicekli [13]proposed diverse LSA-based summarization algorithms. The calculations were assessed on Turkish and English corpus through ROUGE scores.

III. DATASET AND PROBLEM FORMATION

The dataset delivered by "Ed Collins and Isabelle Augenstein and Sebastian Riedel"[1] for extractive summarisation involved 10148 Computer Science publications[2]. Publications were acquired from ScienceDirect. Each paper in this dataset is ensured to have a title, abstract,keywords and sentences highlighted by the author. The highlighted sentences are the sentences that ought to viably pass on the fundamental takeaway of each paper and are a gold summaries, while the keyphrases are the critical subjects of the paper. This model will use the highlights or gold summaries or author generated summaries as reference summaries to train the model.

Problem Formulation

As appeared by Cao et al. (2015)[14], sentences can be acceptable in summaries, even if they are picked out of context. The majority of the highlights have this trademark, not depending on any past or ensuing sentences to make sense. Subsequently, we outline the extractive summarisation task here as a regression task where we assign each sentence a score based it's overlapping lengths with the highlighted summary. Our training data is therefore a list of sentences, encoded via glove embeddings and a score of each of them all stored in a sorted sequence in which they appear in the original paper.

Rouge Metrics

ROUGE measurements are assessment measurements for summarisation which corresponds well to humans. Recall-Oriented Understudy for Gisting Evaluation[16] technique decides the quality of the summaries by contrasting it with different summaries outlined by humans as a reference.

The instinct behind this is if a model makes a decent summary, the it must contain common sentences or words or portions with the references parcels with the human references (Hermann et al., 2015)[15] .We calculated precision, recall and F score for quantifying summarisation of scientific articles although we elect to use recall for evaluating accuracy of the model, as we will be implementing extractive model

IV. METHOD

Sentence Encoding

Word embeddings are essentially a type of word representation that connects the human comprehension of language to that of a machine. Machines have learned representation of text in a n-dimensional space where words that have a similar significance have a similar representation.

Doc2Vec

Doc2vec is a notable concept[18], used to generate vectors out of words. The objective of doc2vec is to make a numeric representation of a document, irrespective of the length of the document.

GloVe

GloVe stands for "Global Vectors". GloVe captures both global statistics and local statistics of a corpus[19], for generating the representational vectors for the words.

This paper used GloVe to generate encodings for each word while calculating the sentence score as we need the local and global context . While we used doc2vec to generate encodings for the research document.

Generating Score

Some statements in the input text contribute more to the summary of the text while some sentences are less important in accordance to the summarization task. Thus scoring sentences based on how important they are for representing the input text in summary is the next step.This score is determined by how many words are contributed by the given sentence in the author generated summary. This paper first uses glove embeddings to generate vectors for the sentences in body and highlights and then iterate through a sentence vector and the highlights vector for every common vector present the sentence score is incremented by 1 .

Preparing Training Dataset

Training dataset therefore consists of two dataframes. One of which namely X will contain vector representation of sentences and words via doc2vec stored in a sorted sequence in which they appear in the original paper. This sequencing of sentences can be later utilized for retrieving the summary lines from the original body. The second data frame namely Y, will comprise the score of corresponding sentences.

Implementing the model

This paper implements Stochastic Gradient Descent Regressor and trains the model for predicting the score of sentences with a given input of vectors. We are then using this training to predict the summary of papers whose author generated summaries are not present. The vector representation of Glove embeddings in the input document will be supplied to the model which will then predict the score of each corresponding sentence. K sentences with top scores will be then selected to represent the summary of the research paper.

V. RESULT AND ANALYSIS

To see how much each section of the input research paper adds to the gold summary, we utilize the ROUGE-L [15]score of each sentence of the sections contrasted with the gold summaries.

Title was expected to be the highest ranked as it does provide a one liner summary of the paper. The abstract being second highest is due to the fact that it is already a summary.

The assumed hypothesis was that introduction would be third highest although conclusion is third highest. The low score of introduction might be due to the fact that it's lengthy in nature.

Thus, Fig 1 suggests that summary sentences cannot be extracted from any specific section of the paper.

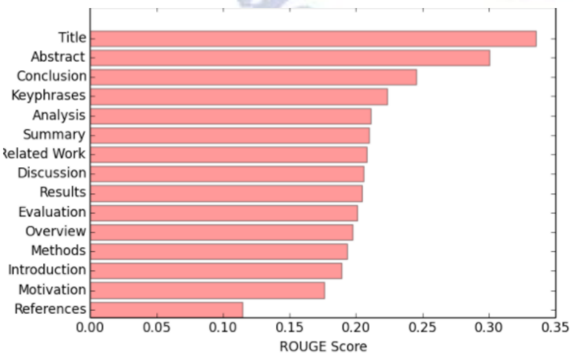

Fig 1

To quantify the results of our implemented SGD regressor model we calculate ROUGE-L scores for generated summary compared to the author generated summaries. Fig 2 shows the precision , recall and F score of the generated summaries. Precision quantifies the accuracy of how many common sentences are present between the generated summary and highlights or what % of the n-grams in the reference summary are also present in the generated summary. Recall score quantifies the accuracy for how many of those sentences were actually important for summary or what % of the n-grams in the generated summary are also present in the reference summary. This score is important to avoid generating unnecessary long summaries. As both of these scores are complementary to each other. F score is calculated as the weighted harmonic mean of the precision and recall of the test.

```
Precision is :0.23972602739726026
Recall is :0.40765391014975044
F Score is :0.30191099649433223
```

Fig 2

Thus our extractive model produced grammatically coherent and concise summaries as we have better precision as compared to recall . But the model generated irrelevant summaries as our F score and recall is poor. Although the model did not perform well, this could act as a baseline for our further research in the Supervised Approach for Extractive Summarisation. Also, more training time and computational power would definitely improve it.

VI. CONCLUSION

Through our model we have gained valuable insights into creating machine learning models for text summarization tasks. More training data, compute power, and time for training would have certainly made our results better. We have evaluated our approach in terms of its performance and the generated summaries. We also verified that there is not any specific section from which one can extract summaries.

REFERENCES

[1] A Supervised Approach to Extractive Summarisation of Scientific Papers, Ed Collins, Isabelle Augenstein, Sebastian Riedel, 2017 2

[2] Julian Kupiec, Jan Pedersen, and Francine Chen. 1995. A Trainable Document Summarizer. In Proceedings of SIGIR.

[3] Hans Peter Luhn. 1958. The Automatic Creation of Literature Abstracts. IBM Journal of research and development 2(2):159–165.

[4] Phyllis B Baxendale. 1958. Machine-Made Index for Technical LiteratureAn Experiment. IBM Journal of Research and Development 2(4):354–361.

[5] Gerard Salton, James Allan, Chris Buckley, and Amit Singhal. 1996. Automatic Analysis, Theme Generation, and Summarization of Machine-Readable

Texts. In Information retrieval and hypertext,Springer, pages 51–73.
[6] Yinfei Yang, Forrest Bao, and Ani Nenkova. 2017. Detecting (Un)Important Content for Single-Document News Summarization. In Proceedings of EACL (Short Papers).
[7] Dragomir R Radev. 2004. LexRank : Graph-based Centrality as Salience in Text Summarization. Journal of Artificial Intelligence Research 22(22):457–479.
[8] Ani Nenkova, Lucy Vanderwende, and Kathleen McKeown. 2006. A Compositional Context Sensitive Multi-document Summarizer: Exploring the Factors That Influence Summarization. In Proceedings of SIGIR.
[9] Hayato Kobayashi, Masaki Noguchi, and Taichi Yatsuka. 2015. Summarization Based on Embedding Distributions. In Proceedings of EMNLP.
[10] Jianpeng Cheng and Mirella Lapata. 2016. Neural Summarization by Extracting Sentences and Words. In Proceedings of ACL.
[11] Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into texts. Proceedings of EMNLP 85:404–411.
[12] Dragomir R Radev. 2004. LexRank : Graph-based Centrality as Salience in Text Summarization. Journal of Artificial Intelligence Research 22(22):457–479.
[13] Text summarization of turkish texts using latent semantic analysis (2010) by Makbule Gulcin Ozsoy , Ilyas Cicekli , Ferda Nur Alpaslan in Proceedings of the 23rd International Conference on Computational Linguistics, COLING ’10
[14] Ziqiang Cao, Furu Wei, Sujian Li, Wenjie Li, Ming Zhou, and Houfeng Wang. 2015. Learning Summary Prior Representation for Extractive Summarization. Proceedings of ACL .
[15] Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching Machines to Read and Comprehend. In Proceedings of NIPS.

[16] ROUGE: A Package for Automatic Evaluation of summaries by Chin-Yew Lin
[17] Text Summarization:An Overview by Samrat Babar
[18] Distributed Representations of Sentences and Documents by Quoc Le Tomas Mikolov
[19] GloVe: Global Vectors for Word Representation Jeffrey Pennington, Richard Socher, Christopher D. Manning Computer Science Department, Stanford University, Stanford, CA 94305