Abstractive Text Summarization Using Hybrid Technique of Summarization

Muhammad Irfan Liaqat
Department of Computer Science,
National Textile University, Faisalabad.
irfan704a@gmail.com

Isma Hamid *
Department of Computer Science,
National Textile University, Faisalabad.
ismahamid@ntu.edu.pk

Qamar Nawaz
Department of Computer Science,
University of Agriculture, Faisalabad.
mqamarnawaz@hotmail.com

Nida Shafique
Department of Computer Science
National Textile University, Faisalabad
nidaprincess898@gmail.com

2022 14th International Conference on Communication Software and Networks (ICCSN) | 978-1-6654-5328-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCSN55126.2022.9817599

***Abstract*—Abstractive text summarization focuses at compress the document into a shorter form while keeping the connotation intact. The extractive summary can select chunks of sentences that are very related to the document, on the other hand, an abstractive summary can generate a summary based on extracted keywords. This research proposed an abstractive text summarization model, it gets data from source data (e.g., Daily Mail/CNN) or other documents and two summaries of this are generated. one from a philologist and the other by proposed model. The summary generated by the philologist kept as a model to compare with the machine-generated summary. The proposed model increased the accuracy and the readability of the summary.**

Keywords*—*abstractive summarization, BERT2BERT, Pars-BERT, Seq-to-Seq

I. Introduction

Over time text summarization playing important role in the overflowing of web-based data. Due to the overloaded amount of information highly demand conscious and automated abstractive summarizers. On the internet, users search queries to find particular information but the user can find out information after visiting multiple web pages which is a headache and time-consuming thing.

Automatic text summarization (ATS) is a core part of natural language processing (NLP). ATS is used in many domains like newspaper headline generation, student information summary. Automatic text summarization research area of the researcher's from the 1950s [1]. There are two main categories of text summarization, Abstractive and extractive text summarization [2].Extractive text, summarization summarizes the document based on the importance of a sentence or phrase.

On the contrary, abstractive text summarization summarizes the document on the behalf of the Semantic meaning of the document and tries to summarize it like a human. It may differ from the extractive strategy. When using an abstract approach, it generates a word sequence for the generation of new summaries with almost new words, sentences, or phrases that don't belong to the original document. The length of the source document has an impact on the quality of the summary that is generated. For example, the research [3] proposed a model that can only work on short input very well but with long input, this model repeats phrases or sentences.

The best quality of ATS requires paying attention to the theme of the document and also a similarity between the source document and generated summary. Topic name extraction from the source document is a difficult task [4]. The study proposed a model that system is built on a deep neural network. The model can extract topic-related keywords and then these keywords are used as input. This model improves the sequence-to-sequence model with the semantics of the document. The sequence-to-sequence model contains of two sections layers of encoder and decoder that can generate a compressed summary of the document. There are several state-of-the-art deep learning models promising outcomes in numerous domains. As defined in paper [5] ATS can improve the recommendation system, decrease the reading time, and less error than human summaries. Now there is much improvement in ATS, which still has loopholes. Several models are proposed by many researchers semantic meaning, pointer generation, seq-to-seq, BERT, encoding-decoding dual encoding-decoding, generative adversarial network, etc.

The aim of this research is to improve the readability and correctness of the prepared summary. This research focused on the abstractive text summarization technique model it get data from source data (e.g. Daily Mail/CNN) or other documents and two summaries of this are generated. one from a philologist and the other by model. The summary generated by the philologist kept as a model to compare with the machine-generated summary. The summary can be of a single page/document or multiple documents.

The fundamental point of this research is to remove unnecessary text and generate an automated abstractive summary of the document. Existing models create multi-label configuration reductions in multiple binary classifications or multiclass classifications, allowing the use of existing loss functions (sigmide, cross-entropy, logistics, etc.). The decline in the multi-tag rank, for example, does not predict the variable number of tags, and the main disadvantages are the remote

978-1-6654-5328-8/22/$31.00 ©2022 IEEE

estimates of the performance matrix. This paper proposed a loss function, the sigmoid F1, which is an estimate of the F1 score that is

(1) smooth and manageable for stochastic gradient descent,

(2) naturally estimates a multi-label metric, and

(3) Estimates tag trends and tag counts.

This paper shows that any confusion matrix can be formulated with a soft surrogate. We reviewed the proposed damage function on text and image datasets, and with multiple metrics, in terms of the complexity of multi-tag classification evaluation. Sigmoid F1 improves one-text and two-image datasets and other damage functions on different matrixes. These results demonstrate the effectiveness of the use of speculative time measurements as abnormal rating issues such as loss functions for multi-label classification.

II. Methodology

While this research notice an expanding number of papers on the upper BERT as well as GPT models revealing empowering enhancements to Glue, SQuAD, and other comparable guidelines. It has been contended that the pre-preparing reason utilized by BERT isn't appropriate for assignments that require text unraveling, for instance, machine interpretation and the formation of restrictive text in abstracts [6]. What's the best technique to use publically accessible pre-trained outposts for hot starting generation models?

One of the critical commitments of this article is that we thoroughly try different things with a few distinct settings to consolidate BERT, GPT, and RoBERTa pre-prepared stations to send off our Transformer based model.

The models haphazardly report huge enhancements over early models that exploit pre-prepared models without management. More significantly, this straightforward procedure brings about new best in class results on machine interpretation, message synopses, sentence conveyance, and sentence combination. The proposed methodology shows in figure 1, results likewise show that a pre-prepared encoder is a fundamental part of arrangement assignments and regularly these undertakings benefit from dividing the load among the encoder and the decoder as shown in figure 2. Altogether, this research done north of 300 tests, burning through a large number of TPU v3 hours to all the more likely change the language displaying and cognizance capacities of these pre-prepared models for text age. This research sure that NLP analysts and professionals will acquire functional bits of knowledge from proposed results as they tackle the different undertakings of seq2seq.

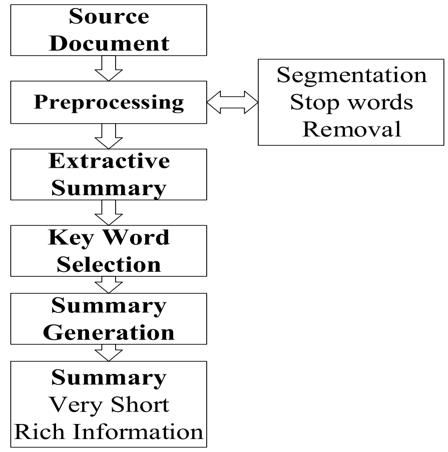

Fig. 1. Architecture model of Abstractive Text Summarization hybrid technique (Proposed Methodology).

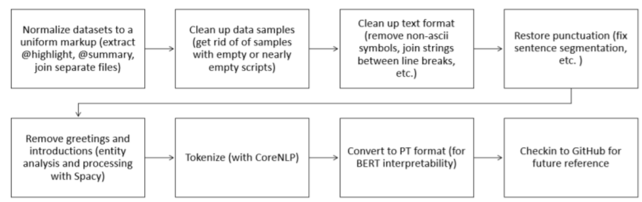

Fig. 2. Text Summarization Pipeline.

We coordinated our text with BERT's pre-prepared jargon utilizing WordPiece [7] as shown in figure 3. Contingent upon experience, we utilize one of the accompanying openly accessible designated spots: BERT-Base Cased, BERT-Base Uncased, BERT-Base Multilingual Cased [8]. The volume is around 30k. WordPress, while the multilingual designated spot has an immense jargon of 110k. BERT likewise prepares implanting positions up to 512 positions, which is the greatest info and result length in all tests.

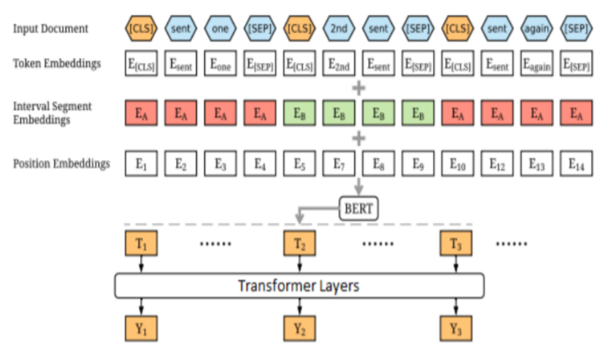

Fig. 3. BERT Arch.

TABLE I. DATA COLLECTIONS FROM VARIOUS SOURCES

#	Source	Total True Sentences
1	Chetor	166,312
2	Ted Talks	46,833
3	Persian Wikipedia	1,878,008
4	Digikala	177,357
5	Eligasht	214,328
6	BigBang Page	3,017
7	Miras-Text	35,758,281
8	Books	25,335

TABLE II. TRAINING TIME

Downstream Task	Dataset	Train Time (hh:mm:ss)
Text Classification	Digikala Magazine	00:10:40
	Persian News	00:21:15
NER	PEYMA	00:45:19
	ARMAN	00:30:57
Sentiment Analysis	Digikala sentiment	1:00:25
	Snappfood sentiment	1:00:22
	DeepSentiPers Binary	00:08:00
	DeepSentiPers Multiclass	00:15:00

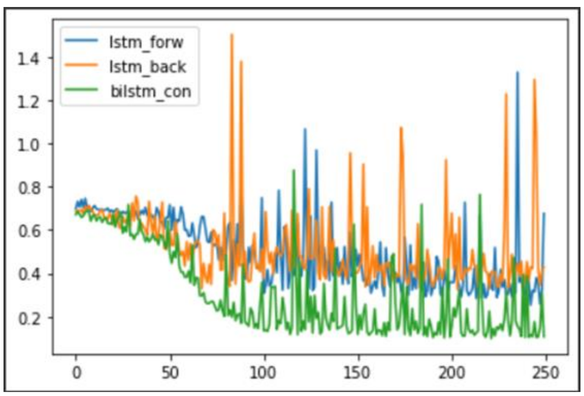

Fig. 4. LSTM performance.

III. EXPERIMENTS AND RESULTS

A. Metric

This research consider the F1 score to be a fitting measurement for checking on the proposed models as the class dispersion is unequal. Allow P to demonstrate precision and R address remembrance, then, at that point, the f1 score is determined by the weighted normal strategy which depends on the quantity of genuine marks in each class utilizing the accompanying conditions.

$$F1 - Score = 2 \times \frac{P \times R}{P + R} \quad (1)$$

B. Training Results

The consequence of the preparation stage is summed up in the accompanying table. The outcomes outline how well recently prepared organizations accomplish the destinations referenced previously. The accuracy and misfortune for both the covered language demonstrating and the forecast of the accompanying sentence are summed up in the table beneath. Moreover, the preparation misfortune charts are introduced in Figure 5.

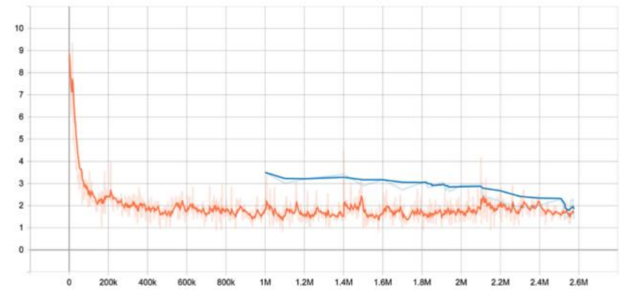

Fig. 5. Training.

TABLE III. PERFORMANCE ON DEEPSENTIPERS

Model	Multi-Class	Binary
	F_1	F_1
ParsBERT	71.31 ± 0.13	92.42 ± 0.14
CNN + FastText	66.30	80.06
CNN	66.65	91.90
BiLSTM + FastText	69.33	90.59
BiLSTM	66.50	91.98
SVM	67.62	91.31

TABLE IV. BERT MODEL PERFORMANCE

Model	Digikala		SnappFood	
	Accuracy	F1	Accuracy	F1
ParsBERT	82.68 ± 0.22	81.72 ± 0.21	87.64 ± 0.19	87.98 ± 0.17
MBERT	81.83	80.74	87.44	87.87

C. Sequence Training

This paper utilized the language model to pre-train the encoder and decoder of an RNN seq2seq model [9]. This proposed technique further developed the BLEU score on News Test 2014 by 3 and furthermore on ROUGE-L on CNN/Dailymail by 3. In any case, BLEU score on Newsquest 2014 En De is 24.7, contrasted with 30.6 in this paper, and 29.4 ROUGE-L in CNN/Dailymail, contrasted with 36.33 in the Transformer model, just as in the BERT covered language model.

TABLE V. ML MODELS COMPARISON

	ELMO	GPT	BERT	UNILM
Left-to-Right LM	✓	✓		✓
Right-to-Left LM	✓			✓
Bidirectional LM			✓	✓
Sequence-to-Sequence LM				✓

To overcome the problem that codec attention is not pre-trained, [10]. The only transformative linguistic model that encodes the source and produces the destination.

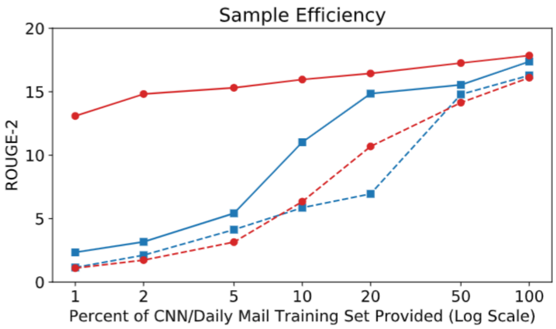

Fig. 6. LM Efficiency Graph.

TABLE VI. LANGUAGE MODELING RESULTS

Training Languages	Nepali Perplexity
Nepali + English	140.1
Nepali	157.2
Nepali + Hindi	115.6
Nepali + English + Hindi	**109.3**

TABLE VII. MODEL ACCURACY

%	**Precision**	**Recall**	**F-Measure**
ROUGE-1	28.14	30.86	27.34
ROUGE-2	07.12	08.47*	07.10
ROUGE-L	28.49	25.87	25.50

Table VI shows the different languages training on proposed model. Table VII shows the Proposed model achieved Rouge Score 8.47. Which is highest score achieved in history.

IV. CONCLUSION AND FUTURE WORK

This research focused on the generation of single paper/document-based abstractive summarization. The main purpose of this study is to develop a deep learning model, the proposed model's working is compared with two most common used machine learning algorithms: logistic regression (LR) and support vector machine (SVM) .The findings reveal that proposed deep learning model do better than the other two models in conditions of overall performance. The F1-measure approach was used to achieve this.

This research will expand the scope of the present procedure by training additional positions and bigger models at WebScaleText Corpora.

REFERENCE

[1] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 1–66, 2017.

[2] N. Moratanch and S. Chitrakala, "A survey on abstractive text summarization," in *2016 International Conference on Circuit, power and computing technologies (ICCPCT)*, 2016, pp. 1–7.

[3] R. Nallapati, B. Zhou, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence rnns and beyond," *ArXiv Prepr. ArXiv160206023*, 2016.

[4] F. You, S. Zhao, and J. Chen, "A Topic Information Fusion and Semantic Relevance for Text Summarization," *IEEE Access*, vol. 8, pp. 178946–178953, 2020.

[5] C. Zheng, K. Zhang, H. Jiannan Wang, and L. Fan, "Topic-aware abstractive text summarization," *ArXiv E-Prints*, p. arXiv-2010, 2020.

[6] M. Lewis *et al.*, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *ArXiv Prepr. ArXiv191013461*, 2019.

[7] M.-H. Su, C.-H. Wu, and H.-T. Cheng, "A Two-Stage Transformer-Based Approach for Variable-Length Abstractive Summarization," *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 2061–2072, 2020.

[8] A. Wang *et al.*, "Can you tell me how to get past sesame street? sentence-level pretraining beyond language modeling," *ArXiv Prepr. ArXiv181210860*, 2018.

[9] K. Ethayarajh, "How contextual are contextualized word representations? comparing the geometry of BERT, ELMo, and GPT-2 embeddings," *ArXiv Prepr. ArXiv190900512*, 2019.

[10] U. Khandelwal, K. Clark, D. Jurafsky, and Ł. Kaiser, "Sample Efficient Text Summarization Using a Single Pre-Trained Transformer."