International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 10 Oct. 2022, pp: 1020-1027 www.ijaem.net ISSN: 2395-5252

A Review Of Extractive And Abstractive Text Summarization Techniques

[1]Sai Kiran, [2]Vinod Kumar, [3]Tulasi Ram, [4]Venkata Ramana, [5]Surendra, [6]Srividya.

[1,2,3,4,5]*Student, GMR Institute Of Technology, Srikakulam, Andhra Pradesh.*
[6]*Assistant Professor, GMR InstituteOf Technology, Srikakulam, Andhra Pradesh.*

Date of Submission: 15-10-2022 Date of Acceptance: 31-10-2022

ABSTRACT: Textual content summarization robotically produces a summary containing essential sentences and consists of all relevant critical statistics from the original document. one of the foremost techniques, when considered from the precis effects, are extractive and abstractive. An extractive precis is heading in the direction of adulthood and now research has shifted towards abstractive summation and real-time summarization. although there had been so many achievements in the acquisition of datasets, techniques, and techniques published, there aren't many papers which can provide a extensive photograph of the current state of studies in this area. This paper offers a wide and systematic review of research within the area of textual content summarization posted from 2019 to 2022. There are 30 journal and conference publications which are the outcomes of the extraction of selected studies for identification and evaluation to explain research topics/traits, datasets, preprocessing, capabilities, strategies, methods, reviews, and problems on this field of studies. The results of the evaluation provide an in-depth explanation of the subjects/trends which can be the focus of their studies inside the discipline of textual content summarization; offer references to public datasets, preprocessing and features that have been used; describes the strategies and strategies which are regularly used by researchers as a contrast and way for developing strategies. at the give up of this paper, numerous guidelines for possibilities and demanding situations related to text summarization studies are noted.

KEYWORDS: Natural Language Processing, AbstractiveSummarization, BLEU, ROGUE score Extractive Summarization.

I. INTRODUCTION

NLP is the department of computer science centered on growing structures that permit computer systems to talk with human beings the use of everyday language. NLP automates the translation process among computers and people. it is a method of getting a laptop to understandably read a line of text without the laptop being fed some kind of clue or calculation. the principles of NLP lie in some of disciplines, viz. pc and data sciences, linguistics, arithmetic, electrical and electronic engineering, artificial intelligence and robotics, psychology, etc. programs of NLP include some of fields of studies, together with gadget translation, herbal language textual content processing and summarization, user interfaces, multilingual and go language statistics retrieval (CLIR), speech reputation, artificial intelligence and expert structures, and so forth. Automatic text summarization is the method, wherein a computer summarizes a textual content. A text is entered into the computer and a summarized textual content is returned, which is a non redundant extract from the authentic textual content. these days, the quantity of web pages on the net nearly doubles every year as the information is now available from a form of sources. It takes good sized amount of time to discover the applicable facts. computerized textual content Summarization will assist the users to locate the relevant information swiftly. An example of the use of summarization generation is engines like google such as Google. One of the herbal questions to ask in summarization is “What are the texts that need to be represented or stored in a summary?” The summary have to be generated with the aid of choosing the critical contents or conclusions within the authentic textual content. locating out crucial information becomes a sincerely difficult venture. currently, the want for computerized text summarization has appeared in lots of areas inclusive of information articles

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 10 Oct. 2022, pp: 1020-1027 www.ijaem.net ISSN: 2395-5252

precis, e mail summary, quick message information on mobile, and information precis for businessman, authorities officers, research, and on line engines like google to get hold of the summary of pages observed and so forth. Automatic summarization is an indispensable part of knowledge about Natural Language Processing (NLP). Automatic summation can be defined as a process of importance, while maintaining the main information content using an automatic machine with no more than less than the original text . The text summary is divided into two types, namely: a summary of man-made and a summary of machinemade. Man-made summaries are generally made by the author himself, which contains important phrases where 80% of abstract sentences made by humans are similar sentences (close sentence match) . Whereas machine made summaries are summaries made by machines with the aim of automating the summarization process performed by humans. The results obtained from machine-made summaries are generally in the form of extracts. The most popular categorization of summary into unmarried file and multi - report summarization. single record is the method of making a summary from a single textual content document. Multi - document summarization shortens a set of related files; into single precis. Usually, there are tactics to automatic summarization and of two types basically,

- Extractive strategies work by using choosing a subset of present phrases, terms, or sentences in the unique textual content to shape the summary.
- Abstractive techniques build an internal semantic representation and then use herbal language generation strategies to create a precis this is towards what a human might generate. the sort of precis would possibly comprise words now not explicitly gift within the original.

The organization of the paper is as follows. section II describes approximately Literature survey and section III will talk concerning machine learning strategies used for text summarization, after fundamentals, we underlined all to be had computerized text summarization techniques and their performances in section IV. segment V affords conclusion to the work and finally, section VI provides references.

II. LITERATURE SURVEY

In this journal, the author proposed a subject-aware abstractive and extractive textual content summarization, which is primarily based on BERT. CNN/daily mail and XSum datasets reveal that the proposed model achieves new outcomes. Stacking the transformer layer inside the encoding level is able to decorate the BERT's ability to symbolize source texts, make full use brand new self-interest, and decide the significance of various components latest the sentence thru distinctive focus rankings. the 2-level extractive– abstractive model can percentage statistics and generate salient summaries, which reduces a certain diploma modern day redundancy. The ROUGE score trendy T-Bertsum is 43.85 and the version can generate summaries with exceptional consistency for the unique text however this technique has constrained processing strength for big and larger texts. [1]

In this journal, the author proposed a sentence-BERTbased similarity(SBSim) metric, which is an evaluation metric for machine translation of Indian languages, ie English to Hindi, and English to Tamil Neural Machine Translation systems. In this journal, the proposed SBSim metric, makes use of a BERT model and sentence-level embedding to evaluate Neural Machine Translation outputs.This SBSim metric is compared with the traditional string-based metrics like BLEU, and ChrF++ scores, which are widely used to evaluate MT systems. The proposed metric is also evaluated on the WMT2020 dataset and reports the highest correlation of 0.7129 with the human scores in evaluating outputs from English-to-Tamil and English-to-Hindi NMT systems. [2]

In this journal the author proposed a model for a novel hybrid model of extractive-abstractive to combine BERT (Bidirectional Encoder Representations from Transformers) phrase embedding with reinforcement mastering. The proposed version is compared with the present day popular computerized textual content summary version on the CNN/daily Mail dataset and uses the ROUGE (do not forgetorientated Understudy for Gisting evaluation) metrics as the assessment technique. within the destiny the proposed model can be prolonged with some other pre-schooling version that is more appropriate for the generative challenge and combines the high-quality-tuning pre-education version with the abstractive precis venture. This model achieved a ROGUE1 score of 37.22 and a ROGUE2 score of 15.78. [3]

In this journal the author proposed a information headline generation version. The generation version is no longer a framework with an encoder-decoder structure. This version works at the news dataset and suggests that our model achieves comparable effects within the field of news headline era. in the version with a decoder best, the present day token of the target phrases can't simplest attention on the supply tokens but additionally cognizance on the generated tokens.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 10 Oct. 2022, pp: 1020-1027 www.ijaem.net ISSN: 2395-5252

The interpreting procedure in our model is much like the human studying procedure which makes our model effective. The proposed model done a ROUGE-1 rating of 35.8. similarly studies is to improve the functionality of the feature representation and the accuracy of the word era. The dangers of this version consist of the out-of-vocabulary hassle and the word generated by using the model now and again isn't correct. [4]

In this journal, a singular divide-and-overcome approach for the neural summarization of long documents. The proposed approach exploits the discourse structure of the report and makes use of sentence similarity to cut up the trouble into an ensemble of smaller summarization problems. The proposed model, Dancer breaks a protracted record and its precis into more than one supply-goal pairs, which are used for education a model that learns to summarize every a part of the record separately. these partial summaries are then mixed a good way to produce a final whole summary. DANCER is a easy but effective extension that may increase the performance of various summarization fashions with minimal additional effort and assets and completed a good ROGUE-1 rating of 45.01. [5]

In this journal, a topic model-primarily based summarization method particularly the 2-tiered subject matter version is mixed with the graph-based totally TextRank method. The mixed method, referred to as TextRank stronger two-Tiered topic version, makes use of the critical sentences acquired from TextRank inside the generative method of the two-tiered version to extract better precis sentences. The proposed technique's summary effects outperform different subject matter model-based totally precis effects the use of ROUGE metrics evaluated on DUC 2005 dataset. The blended strategies TReTTM and TReETTM outperform each TTM and ETTM on Rouge-1 and Rouge-2 evaluation. in addition they outperform sentence-primarily based fashions LDCC and SenLDA-based totally summarization methods. [6]

In this journal the writer proposed an improved BERTLDA hybrid version that become built in a complex Cantonese context, concerning a combination of chinese language and English, in addition to conventional characters and emoticons. thru the collection of massive-scale text records related to the Anti-ELAB movement from a discussion board in Hong Kong, a BERT-LDA hybrid version for big-scale network public opinion analysis turned into constructed in a complex context. The evaluation and prediction of sentiment evolution of public opinion statistics, have been tried to research the laws of emotional evolution for such big-scale public opinion occasions. The progressed BERT-LDA version or sentiment class AUC fee exceeds 99.6% within the sentiment class project for the Anti-ELAB motion. [7]

This journal outlines extractive and abstractive text summarization technologies and affords a deep taxonomy of the automatic textual content summarization(ATS) area. The taxonomy presents the classical computerized textual content summarization(ATS) algorithms to fashionable deep getting to know automatic textual content summarization(ATS) architectures. on this journal, they have got also presented a systematic survey of the giant ATS domain in diverse stages: the fundamental theories with preceding studies backgrounds, dataset inspections, characteristic extraction architectures, influential text summarization algorithms, and overall performance size matrices. This journal additionally affords the contemporary boundaries and challenges of ATS strategies and algorithms, which may be further used to triumph over the restrictions in destiny studies. [8]

In this paper the author pursuits at reading the social media statistics for code-switching and transliterated to English language the usage of the unique form of recurrent neural community (RNN) known as lengthy brief-term memory (LSTM) network. The proposed version is compared with BLEU rating received for DNN technique to sequence-tosequence problems the usage of multi-layered LSTM and proved that technique now not best outperforms SMTprimarily based gadget however additionally general Recurrent Neural community (RNN) may be without difficulty trained with a greater accuracy. In future the present work can be prolonged to other social and expert media sites inclusive of fb, Instagram, LinkedIn etc and additionally it can be prolonged to perform content search related to flawed video, audio and picture content material published on social media. [9]

In this journal the writer proposed transforming text from one language to another by the usage of pc structures robotically or with little human interventions is known as machine Translation gadget (MTS). The purpose of this paper is to offer a comprehensive survey of MTS in trendy and for English, Hindi and Sanskrit languages. desired Reporting gadgets for Systematic critiques and Meta-Analyses (PRISMA) approach which includes gear and evaluation strategies as achieved in this survey especially for English, Hindi and Sanskrit languages. BLEU scores For EnglishSanskrit system translation

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 10 Oct. 2022, pp: 1020-1027 www.ijaem.net ISSN: 2395-5252

machine is 0.445 and EnglishHindi machine Translation machine is 0.75. [10]

In this journal, the writer introduces a word-based totally automated machine translation system by using combining machine translation techniques with chinese-English phrase translation and explores the design and testing of device automated translation structures. automatic machine translation is a whole process that integrates the development of ideas, opens up the use of existing sources, and provides modules consisting of repositories, dictionaries, and so on. the main downside of this model is that it is not reliable because it does no longer have enough dependency pairs for proper translation. The proposed model achieved a BLEU rating of 13.5 and This model proposed results in a brief time with comparable BLEU rankings. [11]

In this paper, the writer proposed an implementation on the real-time synchronous translation approach, and attention on the key technology to be solved within the translation generation of actual-time synchronous translation approach. The dataset used on this paper changed into ChinaDaily dataset and take bear in mind-orientated Understudy for Gisting evaluation (ROUGE) as the evaluation index. In destiny, in phrases of time the Tri-Trophic Metapopulation Mode(TTMM) gadget has improved the interpretation results as compared to the Baseline machine, which desires to be progressed further. The experimental effects display that the blended real-time synchronous translation method and RL brings a positive diploma of optimization and done a ROGUE1 rating of 36.71. [12]

In this journal, the author proposed a conventional speechto-speech translation approach to concatenate automatic speech popularity (ASR), text-to-text system translation (MT), and textual content-to-speech synthesizer (TTS) by means of textual content records. The effects for the RNNbased model in natural speech are slightly worse as compared to the overall performance with generated speech. however, in the event that they have used the Transformer instead of RNN that is skilled using each herbal and generated speech, a excessive ASR performance changed into accomplished. This version proposed outcomes in a brief time with excessive BLEU score when in comparison to different fashions and performed a BLEU rating of 34.3. [13]

In this paper the writer proposed a multimodal neural machine translation system that uses each texts and their associated photographs to translate Korean photo captions into English. This paper makes use of records that extends the Flickr30K Entities dataset, in which the entities within the photograph are classified, every picture has its very own caption and every of it has a source sentence. The results can be analyzed thinking about 3 aspects the overall performance change similar to the Korean enter unit, the impact of image functions, and the effect of label applicants. The proposed version advanced the performance by +1.0 BLEU compared to the text-based NMT model and carried out a BLEU score of 30.7. [14]

In this paper, the author presented an analysis of seventy five BNLP studies papers and categorize them into 11 categories, specifically statistics Extraction, device Translation, Named Entity popularity, Parsing, elements of Speech Tagging, query Answering device, Sentiment evaluation, unsolicited mail and pretend Detection, text Summarization, phrase experience Disambiguation, and Speech Processing and popularity. the writer studied articles posted between 1999 to 2021, and 50% of the papers were published after 2015. This magazine gives a whole evaluation of all the natural language processing techniques which can be used to overcome future obstacles. At closing the writer mentioned demanding situations and future studies opportunities and further reviewed the characteristics and complexity important to expertise contemporary challenges on this area. [15]

In this paper, a graph-primarily based text summarization technique has been described which captures the aboutness of a textual content document. The approach has been developed using modified TextRank computed primarily based on the idea of PageRank defined for every page in the web pages. The proposed technique constructs a graph with sentences as the nodes and similarity between sentences as the weight of the brink among them. modified inverse sentence frequency-cosine similarity is used to provide exceptional weightage to special phrases within the sentence, while traditional cosine similarity treats the words similarly. The proposed technique finished a ROGUE-1 rating of 46.87. the primary drawback of the proposed algorithm is that it does no longer cope with the anaphora decision trouble. [16]

This paper proposes a version “lead3” of unsupervised extractive summarization. they have examined many ways to specific the context records, studied the relationship between sentences in the abstract. it is also proved that the context information and the relationship between the sentences are very useful to the assignment and then advanced an unmanaged summarization

device without any schooling. The dataset used in this approach is CNN/DM dataset which contains 312,000-word dependency pairs.in this paper the proposed unsupervised extractive summarization model lead3 performed a ROGUE1 rating 40 and ROGUE2 rating of 17. [17]

In this journal, they have proposed KeBioSum, a unique know-how infusion schooling framework, and experiment the use of some of Pre-skilled Language fashions(PLMs) as bases, for the challenge of extractive summarization of biomedical literature. a singular information-guided education framework, namely the information adapter, was used for both generative and discriminative education to aid knowledge infusion into the PLMs. to assess the effectiveness of our version, they have conducted experiments on 3 literature datasets from biomedicine: CORD19, PubMed, and S2ORC. twine-19 is an open dataset, which includes clinical papers on COVID-19. This PubMed BERT model achieved a decent ROGUE-1 score of 42.9 and a ROGUE-2 rating of 37.0. [18]

In this paper, they have proposed an automated, general, and extractive Arabic single record summarizing method aiming at producing a sufficiently informative summary. The proposed extractive technique evaluates each sentence primarily based on a aggregate of statistical and semantic functions in which a novel system is used taking into consideration sentence importance, coverage and diversity. similarly, two summarizing techniques which include scorebased totally and supervised machine gaining knowledge of were employed to supply the summary after which help in leveraging the designed features. on this paper EASC dataset was taken which contains of 153 articles. The proposed rating-primarily based strategies performed remember, precision, and F-rating of 67.0, 61.0, 64.0 respectively. [19]

In this paper the writer proposed an automatic abstractive text summarization algorithm in japanese using a neural network. in the proposed a transformer-based decoder returned the precis sentence from the output as generated by the encoder. The dataset used on this paper changed into a Livedoor information corpus together with 130,000 records factors, of which a 100,000 had been used for schooling and the accuracy of the version become 67%. The contents of the summary sentence had been repeated, the model was not able to address unknown words, and there has been a hassle with easy phrase errors. [20]

This journal proposed a version for text summarization that is BERT and this article summarization consists of three phases i.e information pre-processing phase, algorithmic processing section, and publish-processing phase. records pre-processing segment is a process of cleaning the report and the algorithmic Processing phase is the method of applying an algorithmic method and the submit-Processing section is the technique of making use of any information transformation to the goal summary. The dataset used in this technique is CNN/DM dataset which contains 312,000-word dependency pairs. The schooling time taken for a DistilBERT summarizer became round 25 minutes according to 1000 checkpoints on a Google GPU consultation and it keeps an accuracy of 98% of the BERT version. [21]

In this paper Neural system translation is a novel paradigm in device translation research. on this paper, an LSTM-based totally deep gaining knowledge of encoder-decoder model for English to Urdu translation is proposed. The parallel EnglishUrdu corpus of 1083734 tokens has been used, and out of these general tokens, 542810 had been English tokens, and 123636 had been Urdu tokens. This model has a median BLEU score of 45.83. inside the destiny, in this model, a speech recognition module may be constructed with speechto-text translation. the translation first-rate of the proposed model become excellent the word error rates have been additionally much less. the principle difficulty of this version is that it could not translate all the words into the specified language and overlooked a few words. [22]

In this journal, the writer proposed a way of textual content summarization the usage of LSA and Sentence based subject matter modeling with BERT. The consequences in extracting useful sentences from a text report that contains a beneficial amount of statistics approximately the topic on which the text document is based totally on. The proposed version achieves a ROUGE-1 0.44 score. Inside the future the use of the proposed set of rules in abstractive text summarizer where the device is generating a summary in its very own language will bring about reaching more accuracy. The proposed model generates summaries primarily based upon the semantics giving foremost consequences and acting appropriately on huge texts. [23]

In this paper, an present BERT model is used to provide extractive summarization with the aid of clustering the embeddings of sentences via ok-manner clustering but in a dynamic method to determine the variety of clusters. The dataset used for the summarization task is CNN/DailyMail. The dataset consists of CNN and every day Mail news articles. The pre-trained BERT model has been used on this journal. In destiny models, variations

of BERT should be compared and tested. The ROUGE-1(F1) score is 41.25. the principle drawback of the present model became that the complete context of the file to be summarized could not be represented in a smaller range of sentences.[24]

In this journal, the author proposed a novel neural technique to supply dependence-based totally context for system translation. The proposed model is able to now not best encoding source lengthy-distance dependencies however also capturing useful similarities to better predict translations. The proposed version achieves widespread improvement over the baseline systems and outperforms numerous current context-improved strategies. the principle hindrance of this version is that it has to enhance its overall performance regarding word embeddings and proper translation. The proposed model gave a descent BLEU rating of 17.8.[25]

In this paper, a unique statistical method to carry out an extractive text summarization on a single document is validated. The method gives the idea of the input textual content in a brief shape this is within the shape of a meaningful summary. Sentences are ranked by using assigning weights and they're ranked based totally on their weights. exceedingly ranked sentences are extracted from the input document so it extracts vital sentences that direct to a precis of the enter report. The dataset used on this paper is the Stanford sentiment treebank which includes 10000 critiques from rotten tomatoes segregated based totally on their polarities. The version executed a respectable F1 score of 62.29. [26]

In this paper, they have offered a framework that compares selected translations (from Sanskrit to English) of the Bhagavad Gita the use of semantic and sentiment analyses. they've used a hand-labeled sentiment dataset for tuning state-of-artwork deep gaining knowledge of-primarily based language fashions called bidirectional encoder representations from transformers (BERT). The dataset that has been used on this paper is the SenWave dataset which consists of 10,000 tweets which might be hand-categorized via experts and the polarity score varies from zero-10 with appreciate to high quality. The assessment metric used in this paper is the cosine similarity which completed a decent rating of 62.0. [27]

In this journal, they've proposed KeBioSum, a novel expertise infusion training framework, and test using a number of Pre-skilled Language fashions(PLMs) as bases, for the undertaking of extractive summarization of biomedical literature. a singular understanding-guided education framework, namely the information adapter, changed into used for each generative and discriminative schooling to assist know-how infusion into the PLMs. to assess the effectiveness of our version, they have got performed experiments on 3 literature datasets from biomedicine: CORD19, PubMed, and S2ORC. cord-19 is an open dataset, which includes medical papers on COVID-19. This PubMed BERT version completed a decent ROGUE1 score of 42.9 and a ROGUE2 score of 37.0. [28]

In this paper, the applied device channels an idea called subject matter level summary. the subject level summary is a collective precis in textual content layout which consists of relevant facts on a topic in which a topic may be an idea, idea or a term user desires to understand approximately. This records in textual content format is handed directly to the abstractive summarization version which makes use of advanced NLP talents of the bidirectional encoder illustration transformers (BERT) language model to generate a subjectstage precis. The dataset used in this approach is CNN/DM dataset which includes 312,000-word dependency pairs. The noted summarization version has ROUGE ratings of 41.72, 19.39, and 38.76 for ROUGE-1, ROUGE-2, and ROUGE-L respectively. [29]

In this paper the author lays emphasis at the TextRank algorithm, a graph-based totally technique used to address the automatic article summarization problem, and proposes a variation to the similarity characteristic used to compute scores all through sentence extraction. The TextRank set of rules is primarily based in basic terms on the frequency of occurrence of phrases and does no longer require any earlier know-how of grammar. This removes the requirement of any precise tools dedicated to any precise language. The proposed model performed a median bear in mind of 1.0 and a precision of 0.49 and a fscore of 0.61. [30]

III. CONCLUSION

With records increasing at a excessive pace. computerized Sumz is certainly a completely important topic. It intends to keep a remarkable quantity of time, power and resources. research remains being conducted as frequently the results produced are beside the point to the authentic textual content however with time increasingly more green techniques are being proposed. The proposed studies has analyzed diverse algorithms and techniques proposed within the final three years. these techniques together or in my view give marvelous consequences. Their accuracy scores are

in comparison on the premise of ROGUE rating. There isn't any precise model with the quality consequences but some works decently in unique circumstances. it is determined that, greater research is being carried out by incorporating deep learning on this field, which additionally guarantees greater green solutions to the hassle. This paper explains each the extractive and abstractive methods along the strategies applied, its exhibition carried out. content summarization has its importance in each businesses simply as research & enhancing network. Abstractive summarization wishes to be centered more, because it greater advances and bring the summaries by way of studying of information. hence, it is a piece complex then extractive methodology, but it provides regularly crucial and right precis as evaluate to extractive methodology. textual content summarization as part of NLP may be helpful in each day basis work development. need of the hour is simply to searching for the capability of the methodologies which can be certainly implemented to get the preferred statistics.

IV. FUTURE SCOPE

For future scope, the above techniques may additionally transform for multilingual so that this expertise isn't limited to just English. Research works need to be carried out so that it will shift single files to multi document Sumz. furthermore GANs and transfer studying may be used to make those fashions greater efficient and effective. Extra works need to be executed that allows you to prepare datasets with greater reliable floor reality. New metrics can be investigated and can be used in an automatic assessment environment to degree the general high-quality including grammar, fluency, clarity, prominence and relatedness. The above mentioned automatic summarization techniques can also be improvised and used for multiple document summarization. The above mentioned extractive and abstractive techniques can also be combined together inorder to perform hybrid text summarization on any form of text.

REFERENCES

[1] Ma, T., Pan, Q., Rong, H., Qian, Y., Tian, Y., & Al-Nabhan, N. (2021). T-bertsum: Topic-aware text summarization based on bert. IEEE Transactions on Computational Social Systems, 9(3), 879-890.

[2] Mrinalini, K., Vijayalakshmi, P., & Nagarajan, T. (2022). SBSim: A Sentence-BERT Similarity-Based Evaluation Metric for Indian Language Neural Machine Translation Systems. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30, 1396-1406.

[3] Wang, Q., Liu, P., Zhu, Z., Yin, H., Zhang, Q., & Zhang, L. (2019). A text abstraction summary model based on BERT word embedding and reinforcement learning. Applied Sciences, 9(21), 4701.

[4] Li, P., Yu, J., Chen, J., & Guo, B. (2021). HG-News: News Headline Generation Based on a Generative Pre-Training Model. IEEE Access, 9, 110039-110046.

[5] Gidiotis, A., & Tsoumakas, G. (2020). A divide-and-conquer approach to the summarization of long documents. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 3029-3040.

[6] Akhtar, N., Beg, M. S., & Javed, H. (2019, August). TextRank enhanced topic model for query focussed text summarization. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-6). IEEE.

[7] Tan, X., Zhuang, M., Lu, X., & Mao, T. (2021). An analysis of the emotional evolution of large-scale internet public opinion events based on the BERT-LDA hybrid model. IEEE Access, 9, 15860-15871.

[8] Mridha, M. F., Lima, A. A., Nur, K., Das, S. C., Hasan, M., & Kabir, M. M. (2021). A survey of automatic text summarization: Progress, process and challenges. IEEE Access, 9, 156043-156070.

[9] Vathsala, M. K., & Holi, G. (2020). RNN based machine translation and transliteration for Twitter data. International Journal of Speech Technology, 23(3), 499-504.

[10] Bawa, S., & Kumar, M. (2021). A comprehensive survey on machine translation for English, Hindi and Sanskrit languages. Journal of Ambient Intelligence and Humanized Computing, 1-34.

[11] Ning, J., & Ban, H. (2021). Design and Testing of Automatic Machine Translation System Based on Chinese-English Phrase Translation. Mobile Information Systems, 2021.

[12] Ke, X. (2022). English synchronous real-time translation method based on reinforcement learning. Wireless Networks, 1-13.

[13] Kano, T., Sakti, S., & Nakamura, S. (2020). End-to-end speech translation with transcoding by multi-task learning for distant language pairs. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 1342-1355.

[14] Heo, Y., Kang, S., & Yoo, D. (2019). Multimodal neural machine translation with weakly labeled images. IEEE Access, 7, 54042-54053.

[15] Sen, O., Fuad, M., Islam, M. N., Rabbi, J., Masud, M., Hasan, M. K., ... & Iftee, M. A. R. (2022). Bangla Natural Language Processing: A Comprehensive Analysis of Classical, Machine Learning, and Deep Learning Based Methods. IEEE Access.

[16] Mallick, C., Das, A. K., Dutta, M., Das, A. K., & Sarkar, A. (2019). Graph-based text summarization using modified TextRank. In Soft computing in data analytics (pp. 137-146). Springer, Singapore.

[17] Zeng, H., & Chen, G. (2020, December). Unsupervised extractive summarization based on context information. In 2020 IEEE 6th International Conference on Computer and Communications (ICCC) (pp. 1651-1655). IEEE.

[18] Xie, Q., Bishop, J. A., Tiwari, P., & Ananiadou, S. (2022). Pre-trained language models with domain knowledge for biomedical extractive summarization. Knowledge-Based Systems, 109460.

[19] Qaroush, A., Farha, I. A., Ghanem, W., Washaha, M., & Maali, E. (2021). An efficient single document Arabic text summarization using a combination of statistical and semantic features. Journal of King Saud University-Computer and Information Sciences, 33(6), 677-692.

[20] Iwasaki, Y., Yamashita, A., Konno, Y., & Matsubayashi, K. (2019, November). Japanese abstractive text summarization using BERT. In 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI) (pp. 1-5). IEEE

[21] Abdel-Salam, S., & Rafea, A. (2022). Performance Study on Extractive Text Summarization Using BERT Models. Information, 13(2), 67.

[23] Andrabi, S. A. B., & Wahid, A. (2022). Machine translation system using deep learning for English to Urdu. Computational Intelligence and Neuroscience, 2022.

[23] Gupta, H., & Patel, M. (2021, March). Method Of Text Summarization Using Lsa And Sentence Based Topic Modelling With Bert. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 511-517). IEEE.

[24] Srikanth, A, Umasankar, A. S., Thanu, S., & Nirmala, S. J. (2020, October). Extractive text summarization using dynamic clustering and co-reference on BERT. In 2020 5th International Conference on Computing, Communication and Security (ICCCS) (pp. 1-5). IEEE.

[25] Chen, K., Zhao, T., Yang, M., Liu, L., Tamura, A., Wang, R., ... & Sumita, E. (2019). A neural approach to source dependence based context model for statistical machine translation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(2), 266-280.

[26] Madhuri, J. N., & Kumar, R. G. (2019, March). Extractive text summarization using sentence ranking. In 2019 International Conference on Data Science and Communication (IconDSC) (pp. 1-3). IEEE.

[27] Chandra, R., & Kulkarni, V. (2022). Semantic and sentiment analysis of selected bhagavad gita translations using BERT-based language framework. IEEE Access, 10, 21291-21315.

[28] Xie, Q., Bishop, J. A., Tiwari, P., & Ananiadou, S. (2022). Pre-trained language models with domain knowledge for biomedical extractive summarization. Knowledge-Based Systems, 109460.

[29] Ramina, M., Darnay, N., Ludbe, C., & Dhruv, A. (2020, May). Topic level summary generation using BERT induced Abstractive Summarization Model. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 747-752). IEEE.

[30] Sehgal, S., Kumar, B., Rampal, L., & Chaliya, A. (2019). A modification to graph based approach for extraction based automatic text summarization. In Progress in advanced computing and intelligent engineering (pp. 373-378). Springer, Singapore.