



IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: IV Month of publication: April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60113

www.ijraset.com

Call: 🕥 08813907089 🔰 E-mail ID: ijraset@gmail.com



Deep Learning Based Hybrid Technique for Improved Summarization of Abstractive Text

Lanka Shriya¹, Mekala Kamala²

¹UG Student, ²Assistant Professor, Department of Computer Science & Engineering, CMR College of Engineering & Technology, Hyderabad, India

Abstract: From various source sentences, construct summary sentences by merging facts. The process of preserving information material and overall meaning while condensing it into a shorter representation is known as abstractive text summarization, or ATS. It takes a lot of effort and time for humans to manually summarise large textual publications. In this paper, we present an ATS framework (ATSDL) based on (Long Short-Term Memory-Convolutional Neural Network) LSTM- CNN that can generate new sentences by investigating semantic phrases, which are finer- grained fragments than sentences. ATSDL, in contrast to current abstraction- based methods, consists of two primary phases: the first extracts phrases from source sentences, and the second uses deep learning to produce text summaries. Experimental results on CNN and Daily Mail datasets show that our ATSDL framework outperforms the state-of-the-art models in terms of both syntactic structure and semantics, and achieves competitive results on manual linguistic quality evaluation. In this application hybrid model is giving better performance over other state of art techniques.

Keywords: Deep learning Relation extraction, Abstractive text summarization, Text mining.

I. INTRODUCTION

A task generating a short summary the Text summarization (TS) consisting of a some sentences that capture salient ideas of a text [16]. The objectives of the paper are,

- 1) Develop a deep learning framework for abstractive text summarization that leverages the strengths of LSTM and CNN architectures.
- 2) Train the model on large-scale datasets of text documents and human-generated summaries to learn to generate informative and fluent summaries.
- 3) Conduct qualitative analysis to assess the linguistic quality, coherence, and in formativeness of the generated summaries.

In Natural language understanding overcoming this task is an important step. In a very short time also good summary and concise can help bitterly to humans comprehend the text content. Into two various categories the text summarization can be broadly classified Based on previous studies. [6] One uses traditional methods and is called Extractive Text Summarization (ETS).

Its methods for creating summaries involve selecting significant sections from the original text and putting them together to create a logical summary. The alternative is Abstractive TextSummarization (ATS), which creates summaries from scratch without limiting itself to specific phrases from the source text by producing more qualitatively human-written sentences. Initially,ETS models were suggested as a way to extract andcondense the essential semantic data included in the source text. Sequence-to-sequence models, or ATS models, have been proposed more recently as result of advances in computer performance and the growth of deep learning theory. The long short-term memory (LSTM) encoder-decoder model, which was proposed in [11], is a particularly important model within the sequence- to-sequence model framework for our purpose. It has achieved state-of-the-art performance inmachine translation, a natural language problem. Though TS and machine translation share manysimilarities, they are completely different problems. The target output sequence in machine translation is roughly the same length as the source text. Nonetheless, the output sequence in text summary usually quite brief and is not highly dependent on the length of the source text.

A major difficulty in text summarization is to best compress the source material in a lossy way while maintaining the main ideas, whereas machine translation aims for a lossless translation. While near-word-for-word alignment between source and target is widely recognised in machine translation, it is less evident in text summarization. Even though the current TS models have produced excellent results in numerous well-known datasets, not all issues have been overcome by these models. While syntactic structure and semantics are twocrucial aspects to consider when assessing TS models, each type of model concentrates on just one of them.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

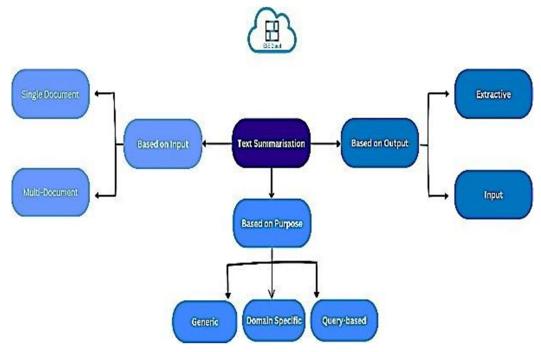


Figure 1: Different types of summarizations of text[16]

Text summarization has basic three types asmentioned below,

- 1) Type of Input data
- 2) Type of Output data
- 3) Based on the Purpose

Above three are the main categories and subtopicsare shown in figure.

The present ETS models are summaries with a few isolated sentences that are coarsely textured. One benefit of using ETS models is that the summary' sentences have to follow the syntacticstructure's rules. The potential for syntactic disarray in the summary is an inherent drawback of ETS models. This drawback arises from the fact that the summaries' neighbouring sentences aren't always adjacent in the source material. The current iterations of ATS models are detailed and incorporate semantic objects (words, sentences, etc.) in their summaries. ATS models have the benefit of inclusive semantics as, during training, they identify word collocations and produce a listof keywords based on those collocations.

We suggest an ATS framework, called ATSDL, based on LSTM-CNN to address the issues. The primary contributions of this work areas follows:

- *a)* To boost TS performance, we integrate CNN and LSTM, an LSTM model that was first created for machine translation, with summarization. With existing ATS and ETS models, ATSDL takes syntactic structure and semantics into account. The new model will produce a series of words after training. This series is a natural sentence-based summarization of the text.
- b) Using phrase location information, we address the primary issue of rare words and produce more naturally occurring sentences.
- *c)* The experiment's findings show that, on two distinct datasets, ATSDL performs better than modern abstractive and extractive summarization algorithms.

II. LITERATURE SURVEY

Distant supervision naturally aims to significantly reduce the high cost of data annotation, it is still highly constrained by the quality of training data. Based on the well-known MIML framework, we create architecture in this paper called MIML-sort (Multi-instance Multi- label Learning with Sorting Strategies). We provide three ranking-based sample selection techniques based on MIML-sort that we use to find relation extractors from a portion of the training data. The KBP (Knowledge Base Propagation) corpus, a sizable and noisy benchmark dataset for remote supervision, is used for experiment setup. Incomparison to earlier research, the suggested techniques yield significantly superior outcomes. Moreover, the combination of the three approaches produces the best F1 on the official testing set,maximising F1 from 27.3% to 29.98%. [1]



We suggest an abstraction-based multi- document summarising system that can create new sentences by examining noun/verb phrases, which are more fine-grained syntactic units than sentences. Unlike other abstraction-based methods, ours builds a pool of facts and concepts first, represented by phrases taken from the input documents. Subsequently, in order to maximisephrase salience and meet sentence construction limitations, informative phrases are chosen and combined to create new sentences. To find the global optimal solution for a summary, we simultaneously perform phrase selection and merger using integer linear optimisation. [2]

In extractive query-focused summarization, the two primary goals are query relevance rating and sentence saliency ranking. Historically, supervised summarization systems have frequently completed the two tasks independently. Nevertheless, neither of the two rankers could be trained properly since reference summaries represent the trade-off between saliency and relevance when used as supervision.

In this paper, a novel summarising system named AttSum is proposed that addresses both goals simultaneously. Together with the document cluster, it automatically learns distributed representations for sentences. When a question is posed, it uses the attention mechanism to mimic thecareful observation of human behaviour. Benchmark datasets for DUC query-focused summarization are utilised for extensive experiments. In the absence of any custom features,AttSum reaches competitive results. Additionally, we note that the words identified as query-focused do address the requirement for the question. [3]

Text document representation in information retrieval, machine learning, and text mining typically uses variations of sparse Bag of Words (sBoW) vectors, such as TF-IDF [1]. Word-level synonymy and polysemy cannot be captured by sBoW style representations, despite their simplicity and intuitiveness, due to their intrinsic over- sparsity. Over fitting and decreased generalisation accuracy result from this. In this research, we present an unsupervised approach to discover enhanced sBoW document features: Dense Cohort of Terms (dCoT). By deleting and recreating random subsets of words from the unlabelled corpus, dCoT directly models absent words. In this way, the high dimensional sparse sBoW vectors are mapped into a low dimensional dense representation, and dCoT learns to rebuild frequent words from co-occurring infrequent words. We demonstrate that the reconstruction can be solved for in closed form and that the feature removal can be marginalised out. Using a number of benchmark datasets, we provide empirical evidence that dCoT features greatly increase classification accuracy for a variety of document classification tasks. [4]

Conventional methods of extractive summarization mostly depend on attributes that have been designed by humans. In this work, we offer a data-driven method based on continuous sentence characteristics and neural networks. We create a general framework consisting of an attention-based extractor and a hierarchical document encoder for single-document summarization. We are able to create many classes of word or sentence extraction summarization models thanks to this design. Using extensive datasets with hundreds of thousands of document- summary pairings, we train our models1. Tested on two summarization datasets, our models achieve performance on par with the state of the art even in the absence of language annotation. [5]

A conditional recurrent neural network (RNN) is presented here, capable of producing an overview of an input text. A unique convolutional attention-based encoder provides the conditioning, making ensuring the decoder concentrates on the right input words at every stage of generation. Our model is simple to train end-to-end on big data sets and only depends on learned features. Our tests demonstrate that the model performs competitively on the DUC-2004 shared task and achieves significant advantages over the newly proposed state-of-the-art technique on the Gigaword corpus. [6]

One of document summarization's numerous documented uses is automatic headline generation. We provide a sequence-prediction method in this work to understand how news editors title theirpieces. The approach that has been presented framesthe issue as a discrete optimisation effort within a feature-ich domain. Using dynamic programming, the global optimum in this space canbe obtained in polynomial time. We use a large corpus of financial news to train and test our model, and we assess its performance against several baselines using common measures from the document summarization area and some new metrics suggested in this study. We also use human evaluation to determine how are readable and formative the generated titles.[7]

III. PROPOSED METHOD

Text summaries can help in understanding any topic easily without reading whole content as it will give short description of complete articles. It's verydifficult to build summaries for each book manually and there are many deep learning algorithms are available while will read whole content and give summary but this summaries are not accurate as they work directly on words. Inpropose paper author using phrases (combination ofwords) from sentences to train LSTM-CNN algorithm and this phrases can help deep learning algorithm in obtaining accurate summary.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

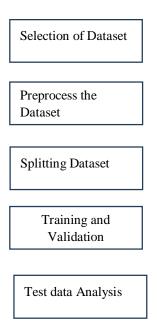


Figure 2: Proposed Block diagram of Proposed Model

There are above mentioned steps are discussed as below,

- 1) Selection of dataset : select proper dataset is key for getting improved performance
- 2) Preprocessing Dataset : to bring dataset in standard format
- 3) Splitting dataset : dataset is splitted in 80% and 20% (80% training and 20% testing)
- 4) Training the model with 80% train data from dataset and 20% data is used for validation
- 5) Test data Analysis : we can enter any text data and generate its abstractive summarization.

Abstractive Text Summary using deep learning(ATSDL) consist of two parts where first part concentrate on extracting phrases from dataset and then extracted phrases will be converted to vector and this vector will be input to LSTM-CNN algorithm to build text summary model.

User can input any text and then ATSDL will predict summary from it and this predicted summary and test data summary will be used to calculate ROUGE 1 and 2 scores. The higher the score the accurate is the summary.

In propose paper author using CNN-Daily Mail dataset which contains articles and summaries and by using this data author training LSTM-CNN (ATSDL) algorithm.

We have coded this project using JUPYTER notebook and below are the code and output screens with blue colour comments Data Collection and Pre-processing: Gather a diverse dataset of text documents and their corresponding human-written summaries from various domains such as news articles, scientific papers, and online forums. Preprocess the data by tokenization, stemming, and removing stop words.

Design a hybrid deep learning architecturethat combines LSTM and CNN components. The LSTM module captures long-term dependencies and sequential patterns in the input text, while the CNN module extracts hierarchical features and local structures.

Train the LSTM-CNN model on the pre-processed dataset using techniques such as mini-batch gradient descent and back propagation through time. Fine-tune hyper parameters, including learning rate, dropout rate, and embedding dimensions, to optimize performance.

Evaluate the performance of the trained model using standard evaluation metrics, including ROUGE-1, ROUGE-2, and ROUGE-L scores.

Compare the results against baseline models, including traditional LSTM, CNN, and seq2seq models, to assess improvements in summarization quality.

Conduct human evaluation and qualitative analysis to assess the linguistic quality, coherence, and informativeness of the generated summaries. Solicitfeedback from domain experts to validate the effectiveness of the LSTM-CNN model incapturing key information and preserving the original meaning of the text.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

IV. RESULT ANALYSIS

Performance analysis of proposed model is analysed using LSTM -CNN combination with python software. Python 3.7 is used for design and analysis with installation of relevant libraries. Many libraries are used which are opensource and make programming simple like NLTK used for text data processing, Keras used for deep learning, sk learn used for performance analysis.

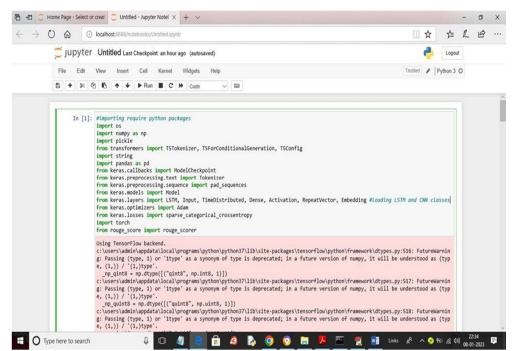
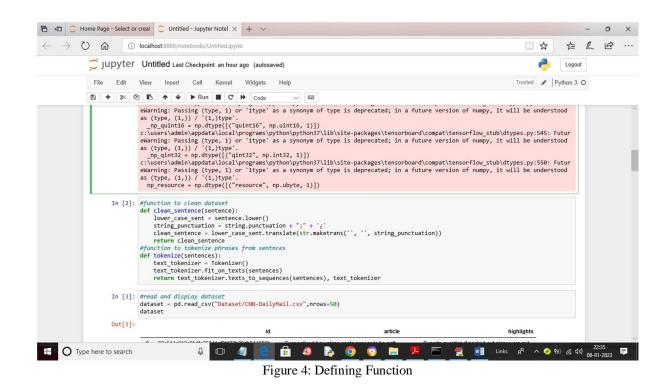


Figure 3: importing python packagesIn above screen we are importing require python packages



In above screen we are defining functions to clean text and the extract phrases from the sentences

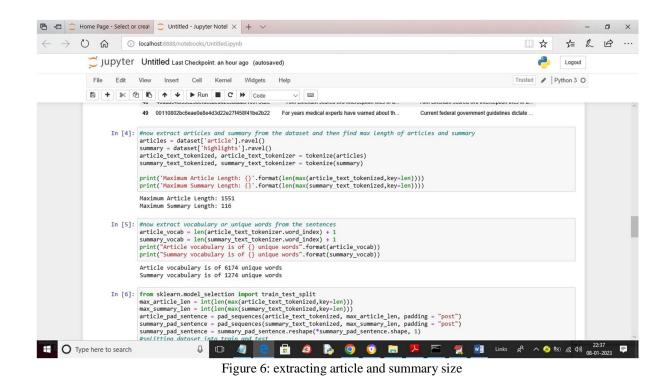


ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Control of the second sec	\rightarrow	() 俞 ()	localhost:8888/notebooks/Untitled.ipynb		C []]	☆ ☆	1	B	
File Edit View Insett Cell Kemel Widgets Help Trusted Python 3 Im 3: <i>dread and display dataset</i> dataset = pd.read_csv("Dataset/CMH-DailyHall.csv", nrows=50) dataset Im 1: <i>in fail in fail</i> <td< th=""><th>-</th><th></th><th></th><th></th><th>Carl .</th><th></th><th>~</th><th>_</th><th></th></td<>	-				Carl .		~	_	
Kernel Construction K		💭 jupyter	Untitled Last Checkpoint: an hour ago (autosa	ved)	· · · · · · · · · · · · · · · · · · ·	Logout			
In [3]: #read and display dataset dataset In [3]: #read and display dataset dataset Out[3]: Id article highlights 0 92c514c913c0bdfe25341a9fd72b29db544099b Ever noticed how plane seats appear to be gett Experts question if packed out planes are put 1 2003841c74c0e7c5b1a2489cd38d72727245a A drunk teenage boy had to be rescued by secur Drunk teenage boy climbed into lion enclosure 2 91b7d23115275c2b63a65ca98d21d9c92485149 Dougle Freedman is on the verge of agreeing a Nottingham Forest are close to extending Dougl 3 caabfbcddfe6bf1410225ba34d5ca98d821d9c92485149 Dougle Freedman is on the verge of agreeing a Nottingham Forest are close to extending Dougl 4 3 da746a7d9afcaa659088c836e6f347fe6b32a Bruce Jenner will break his silence in a two-h Tell-all interview with the reality TV star, 6 5 5 ed653bd235a8046cd3b8074a1aa51b855c8ec3 This is the moment that a crew of frefighters Giant pig fell into the svimming pool at his h 7 98be9b2d558c17d8a13597195957a7c8587dcd (CNN)So, yourd like a "Full House" reunion and Show will return with a one-hour special, foll 8 57f68638739c3a1de8d9922b389d6ded39997712 A 11120pm, former world champion Ken Doherty Reanne Evans faced Ken Doherty in World Champi		File Edit	View Insert Cell Kernel Widgets	Help	Trusted	Python 3 (С		
dataset - pd.read_csv("Dataset/CNN-DailyMail.csv", nrows-50) Out[3] id article highlights 0 92c514c913c0bdfe25341af9ld72b29db544099b Ever noliced how plane seats appear to be gett. Experts question if packed out planes are put. 1 2003841c7dc0e7c5b1a2489cd538d72727454 A drunk teenage boy had to be rescued by secur. Drunk teenage boy climbed into lion enclosure 2 91b7d2311527f5c2b63a65ca88d2149e2d88149 Dougle Freedmain to inte verge of agreeing a Nottingham Forest are dose to extending Dougl 3 caabfbcdfbieb1410295a672e633d304391bbb Liverpool target Neto is also wanted by PSG an Filorentina goalkeeper Neto has been linked wit 5 5ed5e3btd235a8046cd3b87t4ataa51b856c8e2 This is the moment that acrew of firelighters Giant pig feli into the swimming pool at his h 6 6394t751b120ceb3da5e7b53d5167fcdc80b514 The amount of time people speed listening to B Figures show that while millions still tune in 7 9bebeb2d58c17df8a135971969557a7c8887ddd (CNN)So, yourd like a "Full House" reunion and Show will return with a one-hour special, foll 8 5768683373bc3atdeded93827012 At 1120pm, former workout champion Ken Dohery Reanne Evans faced Ken Dohery in World Champi 9 61a420035blcardee51aba?F0b4400e3a5255665		8 + % 4	È Ē ↑ ↓ ► Run ■ C ⊮ Code						
Image: Provide and the second secon		In [3]:	<pre>dataset = pd.read_csv("Dataset/CNN-Daily</pre>	Mail.csv",nrows=50)					
12003841c7dc0er2c5b1a2489ccd536d72727245aA drunk teenage boy had to be rescued by securDrunk teenage boy climbed into lion enclosure291b7d231152715c2b63a65ca98d21d9c9248514aDougle Freedman is on the verge of agreeing aNotlingham Forest are close to extending Dougl3caabf9cbdf96eb1410295s673e953304391thbLiverpool target Neto is also wanted by PSG anFiorentina goalkeeper Neto has been linked wtl43da746a7d9atcaa659088c9366f43t76eb53eaBruce Jenner will break his silence in a two-hTel-Hail interview with the reality TV star, 655665a9t3d235a8046cd3b36714a1a51b956c9acThis is the moment that a crew of frefightersGiant pig fell into the swimming pool at his h66394f51b120ceb3d3e5r055d3d5167tc4cdb00541The amount of time people spend listening to BFigures show that while millinons still ture in798be9b2d558c17d8a1359719957a7c887ddd(CNNJSo, yud like a 'Full House'' reunion andShow will return with a one-hour special, foll857f8638739c3a1de8d9922b389d6ded39977121A11120pm, former wotd champion Ken DohertyReame Evans faced Ken Doherty in World Champi920778c35c19d741ct182719de33d519h4d90538c55965Billing his nails nervously, these are the firsMohammed Ali Malek, 27, has been charged with1061a240035bcca4e8fab3c79b44095126561944Englan Alastair Cook completed a mow.hAlastair Cook completed his century on the sec13391b4610e3d9doc6fol6428155c5afb288124(CNNWaren Weinstein, who appears to have beeU.S. hostage Warren Weinstein is believed to h14162c6c09093940a7		Out[3]:		article	highlights				
291b7d2311527f5c2b83a65ca88d21d9c92485149Dougle Freedman is on the verge of agreeing a.Notlingham Forest are close to extending Dougl3caabf9cbdf86b1410295a673e953d304391bbbLiverpool target Neto is also wanted by PSG anFiorentina goalkeeper Neto has been linked wit43da746a7d9afccaa659088c386ef6347fe6b53eaBruce Jenner will break his silence in a two-hTell-all interview with the reality TV star, 655ede53thd235a8046cd3b87f4a1aa51b856c8ac3This is the moment that a crew of firefightersGiant pig fell into the swimming pool at his h66394f51b120ceb3da57b53d3615f7c4cd80b51The amount of time people spend listening to BFigures show that while millions still tune in79bbe9b2d555c17df8a13597159557a7c8587dcd(KNNSo, yourd like a 'Full Houser' reunion andShow will return with a one-hour special, foll857f6863873ac3a1de8d9922b389d6dc39977012A11120pm, former world champion Ken DoheryReame Evans faced Ken Doherty in World Champi920778c35c19d741cc1822719de33a5d71e1a0b22eeAgang of six men have been jailed for a totalGang have been jailed for a total of 31 years1061a24035b6ca4e8d1ab367b54d4965132655Billing his nails nervously, these are the firsMohammed Al Malek, 27, has been charged wth126777110c19df40e5172a6b26a4e8f32d7356c561368South Korea Ambassador Lippert is now wearingLippert 42, suffered deep gashes to his hand141626c609999340a7e4b124f6ccca0ae984ecdadaAthief nicknamed the 'Black Widow of FacebookSofla Davila, 21, nicknamed the 'Black Widow of.141626c609999			0 92c514c913c0bdfe25341af9fd72b29db544099b	Ever noticed how plane seats appear to be gett	Experts question if packed out planes are put				
3 caab/95/cbdf96eb14102953673e953d304391btb Liverpool target Neto is also wanted by PSG an. Florentina goalkeeper Neto has been linked wit 4 3da746a7d9a1caa659088c8366af6347te6b53ea Bruce Jenner will break his silence in a two-h Tell-all interview with the reality TV star, 6 5 5ed593tbd235a8046c33b87t4a1aa51b856c3ea This is the moment that a crew of firefighters Giant pig fell into the swimming pool at his h 6 6394f151b120ceb3da5e7b53dd5167tcdd80b514 The amount of time people spend listening to B Figures show that while millions still tune in 7 9bbe9b2d558c17df8a135971595957a7c8587dcd (CNN)So, you'd like a 'Full House' reunion and Show will return with a one-hour special, foll 8 5768638739c3a1ded99922b389d6ded39997701 Al 1120pn, former word champion Kein Doherty Reame Evans faced Ken Doherty in World Champio 9 20778c35c19d741cc182719de336d71e1a0b229e Ag ang of six: men have been jailed for a total Mohammed All Malek, 27, has been charged with 11 1817571 telf60df3b003001441eb8449873df7 South Koree Anbassador Lipper is now wearing Lippert, 42, suffered deep gashes to his hand 12 6777110c19df40e5112a6ba92ad77356e510a4b England captain Alastair Cook completed a much Alastair Cook completed his century on the sec 13 391			1 2003841c7dc0e7c5b1a248f9cd536d727f27a45a	A drunk teenage boy had to be rescued by secur	Drunk teenage boy climbed into lion enclosure				
43da746a7d9atcaa659088c8366ef6347fe6b53eaBruce Jenner will break his silence in a two-hTel-all interview with the reality TV star, 655ed5e3fbd235a8046cd3087f4ataa51b856cBec3This is the moment that a crew of firefightersGiant pig fell into the swimming pool at his h66394151b120ceb3da567b53dd5167fc4c80b514The amount of time people spend listening to BFigures show that while millions still tune in79bbe9b2d558c17d18a13597195957a7c8587ddcd(CNN)So, you'd like a 'Full House' reunion andShow will return with a one-hour special, foll857f68633739c3a1dedd98922b3896d6cd308977012At 1120pm, former world champion Ken DohertyReane Evans faced Ken Doherty in World Champi920778c35c19d741cc182719de336d71e1a0b22eaAg ang of six men have been jailed for a totalGang have been charged with161a24035b8ca4ee6f3ba3e79b409c33e2595656Billing his nails nervously, these are the firsMohammed All Malek, 27, has been charged with118176771tel608d7b0003b0144teb84935d17South Kreve Anbassador Lippert is now wearingLippert 42, sufferd deep gashes to his had126777110c19d40e5112a6b492ad77356e510a4bEngland captain Alastair Cook completed an unchAlastair Cook complete his century on the sec13391b4610e3a9dcefd61e428155cc5afb288324cc(CNN)Warren Weinstein, who appears to have beeU.S. hostage Warren Weinstein is believed to h141626c609099340a7e4b124f6ccca0ae984eccdaA thief nickonamed the 'Black Widow of FacebookSofta Davila, 21, nicknamed the 'Black Widow o			2 91b7d2311527f5c2b63a65ca98d21d9c92485149	Dougie Freedman is on the verge of agreeing a	Nottingham Forest are close to extending Dougi				
556ed5e3tbd23588046cd3887/441aa518856c8ec3This is the moment that a crew of firefightersGiant pig fell into the swimming pool at his h66394151b120ceb3da5e7b53dd5167fcdc80b014The amount of time people spend listening to BFigures show that while millions still tune in798be9b2d558c17d8a13597195957a7c8587ddcd(CNN)So, yourd like a "Full House" reunion andShow will return with a one-hour special, foll857f68638739c3a1de8d9922b389d6dcd39977012At 1120pm, former world champion Ken DohertyReanne Evans faced Ken Doherty in World Champi92077Rd35c19d741c182719de33d67t1e1a0b22e8Ag ang of sk men have benjailed for a totalGang have been pialed for a total of 31 years1061a24035b8ca4eef3ab3e79b440e33e5259656Biting is nails nervously, thes are the firsMohammed Ali Malek, 27, has been charged with1118179711ef08d40b6112a6b4928d717356e51044tEngland captain Alastair Cook completed a muchAlastair Cook completed his century on the sec133914610e3a9dcefd61e428155cc5afb268104ze(CNN)Warren Weinstein, who appears to have beeU.S. hostage Warren Weinstein is believed to h141626cc6096993940a7e4b124f6ccca0ae984ecdadA thief nicknamed the 'Black Widow of FacebookSofia Davila, 21, nicknamed the 'Black Widow of			3 caabf9cbdf96eb1410295a673e953d304391bfbb	Liverpool target Neto is also wanted by PSG an	Fiorentina goalkeeper Neto has been linked wit				
6 6394151b120ceb3da567b53dd5167fc4cf80b514 The amount of time people spend listening to B Figures show that while millions still tune in 7 98be9b2d558c17df8a13597195957a7c8587dcd (CNN)So, you'd like a "Full House" reunion and Show will return with a one-hour special, foll 8 57f68638739c3a1de8d9922b389d6ded39977012 At 11:20pm, former world champion Ken Doherty Reanne Evans faced Ken Doherty in World Champi 9 20778c35c19d741cc182719de33d6711e1ab22ee A gang of sk men have been jaled for a total Gang have been jaled for a total of 31 years 10 61a24035b8cae4ee5fab3e79b440e33a5259565 Billing his nails nervously, these are the firs Mohammed Ali Malek, 27, has been charged with 11 18179711ef08dd7b030b001441eb88495f38d17 South Korea Ambassador Lippert is now wearing Lippert, 42, suffered deep gashes to his hand 12 6777110c19df40e5f12a6b922a77356e51044F England captain Alastair Cook completed a much Alastair Cook completed his century on the sec 13 391b4610e3a9dcefd61e428155cc5afb2683124c (CNN)Warren Weinstein, who appears to have bee U.S. hostage Warren Weinstein is believed to h 14 1626c60969993940a7e4b124f6ccca@ae984ecdad A thief nicknamed the 'Black Widow of Facebook Sofia Davila, 21, nicknamed the 'Black Widow o			4 3da746a7d9afcaa659088c8366ef6347fe6b53ea	Bruce Jenner will break his silence in a two-h	Tell-all interview with the reality TV star, 6				
7 9bbe9b2d558c17df8a13597195957a7c8587ddcd (CNN)So, you'd like a 'Full House' reunion and Show will return with a one-hour special, foll 8 57f68638739c3a1de8d9922b389d6ded39977012 Al 1120pm, former world champion Ken Dohery Reanse Evans faced Ken Doherty in World Champio 9 20778c35c190741cc182719de33d711e10b228e A gang of six men have been jailed for a total Gang have been jailed for a total. Mohammed Ali Malek, 27, has been charged with 10 61a24035b8cae4ee5fab3e79b4409c33e525956 Bitling his nails envously, these are the firs Mohammed Ali Malek, 27, has been charged with 11 18179711ef060df3b003b001441eb68495f38017 South Korea Ambassador Lippert is now wearing Lippert, 42, suffered deep gashes to his hand 12 6777110c13df40e5172a6ba22ad7356565124c CNNWarren Weinstein, who appears to have bee U.S. hostage Warren Weinstein is believed to h 14 1626c6096993940a7e4b124f6ccca0ae984ecdad A thief nicknamed the 'Black Widow of Facebook Softa Davila, 21, nicknamed the 'Black Widow o			5 5ed5e3fbd235a8046cd3b87f4a1aa51b856c8ec3	This is the moment that a crew of firefighters	Giant pig fell into the swimming pool at his h				
8 57/686337393a1de8d9922b389d6ded39977012 Al 1120pm, former world champion Ken Doherty Reanne Evans faced Ken Doherty in World Champi 9 20778c35c19d741cct82719de336d71e1a0b222e A gang of six men have been jailed for a total Gang have been jailed for a total 10 61a24035b8cae4ee5fab3e73b4409c33e5259565 Bitling his nails nervously, these are the firs Mohammed Ali Malek, 27, has been charged with 11 18179711ef608d3b003b001441eb68495f3ad17 South Korea Ambassador Lippert is now wearing Lippert, 42, suffered deep gashes to his hand 12 67/7110c19d140e5112a6ba92ad773565e10a4b England captain Alastair Cook completed a much Alastair Cook completed his century on the sec 13 391b4610e3a9docfd61e428155cc5afb268312ct (CNN)Warren Weinstein, who appears to have bee U.S. hostage Warren Weinstein is believed to h 14 1626c6090993940a7e4b124f6ccca@ae984ecdad A thief nicknamed the 'Black Widdow of Facebook Sofia Davila, 21, nicknamed the 'Black Widow or			6 6394f51b120ceb3da5e7b53dd5167fc4cf80b514	The amount of time people spend listening to B	Figures show that while millions still tune in				
9 20778c35c19d741cc182719de336d71e1a0b228e A gang of six men have been jailed for a total. Gang have been jailed for a total of 31 years 10 61a24035b8cae4ee5tab3e79b4409c33e529565 Bitling his nails nervously, these are the firs. Mohammed Ali Malek, 27, has been charged with 11 18179711ef608d13b003b001441eb6849513d17 South Korea Ambassador Lippert is now wearing Lippert, 42, suffered deep gashes to his hand 12 67/7110c19d140e5112a6ba92ad77356e510a4bt England captain Alastair Cook completed a much Alastair Cook completed his century on the sec 13 391b4610e3a9docfd61e428155cc5afb268312ct (CNN)Warren Weinstein, who appears to have bee U.S. hostage Warren Weinstein is believed to h 14 1626c6090993940a7e4b124f6ccca@ae984ecdad A thief nicknamed the 'Black Widow of Facebook Sofia Davila, 21, nicknamed the 'Black Widow o			7 98be9b2d558c17df8a13597195957a7c8587ddcd	(CNN)So, you'd like a "Full House" reunion and	Show will return with a one-hour special, foll				
10 61a24035b8cae4ee5fab3e79b4409c33e5259565 Billing his nails nervously, these are the firs Mohammed Ali Malek, 27, has been charged with 11 f8179711ef608d3b003b001441eb68495f38d17 South Korea Ambassador Lippert is now wearing Lippert, 42, suffered deep gashes to his hand 12 6777110c19df40e5112a6ba92ad77356e510a4b England captain Alastair Cook completed a much Alastair Cook completed his century on the sec 13 391b4610e3a9docdf61e428155cc5afb268312ct (CNN)Warren Weinstein, who appears to have bee U.S. hostage Warren Weinstein is believed to h 14 f626c6096993940a7e4b124f6ccca0ae984ecdad A thief nicknamed the 'Black Widow of Facebook Sofia Davila, 21, nicknamed the 'Black Widow o			8 57f68638739c3a1de8d9922b389d6ded39977012	At 11:20pm, former world champion Ken Doherty	Reanne Evans faced Ken Doherty in World Champi				
11 18179711el608df3b003b001441eb68495f38d17 South Korea Ambassador Lippert is now wearing Lippert, 42, suffered deep gashes to his hand 12 6717110c19df40e51f2a6ba92ad77356e510a4b England captain Alastair Cook completed a much Alastair Cook completed his century on the sec 13 391b4610e329dcafd61e428155cc5afb268312ct (CNN)Warren Weinstein, who appears to have bee U.S. hostage Warren Weinstein is believed to h 14 1626c6096993940a7e4b124f6ccca0ae984ecdda A thief nicknamed the 'Black Widow of Facebook Sofia Davila, 21, nicknamed the 'Black Widow o			9 20778c35c19d741cc182719de336d71e1a0b228e	A gang of six men have been jailed for a total	Gang have been jailed for a total of 31 years				
12 67/7110c19df40e51f2a6ba92ad77356e510a4b England captain Alastair Cook completed a much Alastair Cook completed his century on the sec 13 391b4610e3a9dcefd61e428155cc5afb2683124c (CNN)Warren Weinstein, who appears to have bee U.S. hostage Warren Weinstein is believed to h 14 1626c6096993940a7e4b124f6ccca0ae984ecdda A thief nicknamed the 'Black Widow of Facebook Sofia Davila, 21, nicknamed the 'Black Widow o			10 61a24035b8cae4ee5fab3e79b4409c33e5259565	Biting his nails nervously, these are the firs	Mohammed Ali Malek, 27, has been charged with				
13 391b4610e3a9dcefd61e428155cc5afb2683124c (CNN)Warren Weinstein, who appears to have bee U.S. hostage Warren Weinstein is believed to h 14 f626c6096993940a7e4b124f6ccca0ae984ecdda A thief nicknamed the 'Black Widow of Facebook Sofia Davila, 21, nicknamed the 'Black Widow o			11 f8179711ef608df3b003b001441eb68495f38d17	South Korea Ambassador Lippert is now wearing	Lippert, 42, suffered deep gashes to his hand				
14 1626c6096993940a7e4b12416ccca0ae984ecdda A thiel nicknamed the 'Black Widow of Facebook Sofia Davila, 21, nicknamed the 'Black Widow o			12 67f7110c19df40e51f2a6ba92ad77356e510a4fb	England captain Alastair Cook completed a much	Alastair Cook completed his century on the sec				
			13 391b4610e3a9dcefd61e428155cc5afb2683124c	(CNN)Warren Weinstein, who appears to have bee	U.S. hostage Warren Weinstein is believed to h				
15 9c6cc6f9b1c273e8c58cd6c0ca574d2aa9b1007b As Australian Fashion Week comes to a close, a Australian Fashion Report revealed the Austral			14 f626c6096993940a7e4b124f6ccca0ae984ecdda	A thief nicknamed the 'Black Widow of Facebook	Sofia Davila, 21, nicknamed the 'Black Widow o				
			15 9c6cc6f9b1c273e8c58cd6c0ca574d2aa9bf007b	As Australian Fashion Week comes to a close, a	Australian Fashion Report revealed the Austral				

Figure 5: Reading and Displaying Dataset

In above screen reading and displaying dataset values. Dataset with multiple rows and columns is displayed in above tabular format.



In above screen extracting article and summary sizeand then extracting total articles and summary unique phrases words.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

🖻 🖷 🙄 Home Page - Select or creat 🙄 Untitled - Jupyter Notel 🗙 🕂	×		-	٥	×
\leftrightarrow \rightarrow \circlearrowright \bigcirc localhost:8888/notebooks/Untitled.ipynb	1	☆ ☆	h	ß	
Jupyter Untitled Last Checkpoint: an hour ago	(autosaved)	n Logou	rt		
File Edit View Insert Cell Kernel Wid	dgets Help Trust	d 🖋 Python 3	0		
E + ≫ 42 K + ↓ ► Run E C ⇒	Code				
<pre>lstml = LSTW(32, return, sequences. r, vec = RepeatVector(max_summary_] lstm2 = LSTW(32, return, sequences, logits = TimeDistributed(Dense(sun cm_model = Nodel(linput_sequence, cm_model.compile(loss-sparse_cate model_check_point = Nodel(check hist = cn_model.fit(article_p f = open("model/history, pckl', pickle.dump(hist.history, f) f.close() else: cnm_model.load_weights("model/ print(cnm_model.summary()) cmm_model = TSForConditionalGenerat toKenizer = TSToKenizer.from_pretr device = torch.device('cpu') print("Model_loaded") </pre>	<pre>article_len,)) rticle_vocab, output_dim=128,)(input_sequence) -False)(embedding) #defining lstm input model len)(lstm1) -True, dropout=0.2)(r_vec)#defining lstm output model mmary_vocab))(lstm2) Activation('softmax')(logits)) #now create CNW model by using lstm input and outp egorical_crossentropy, optimizer=Adam(1e=3), metrics=['accuracy']) lghts.hdf5') == False: kpoint(filepath='model_model_weights.hdf5', verbose = 1, save_best_only = True) pad_sentence, summary_pad_sentence, batch_size=8, epochs=5000, validation_data=(X, 'wb') /model_weights.hdf5')</pre>				
input_1 (InputLayer) (None	e, 1551) 0				~
Type here to search	C 🗄 🥝 🕞 🧿 💽 🔚 🖾 🧖 Links 🤻	ላ 📀 🖾 🌾 ሳ	。) 22:: 08-01-	38 2023	₽.

Figure 7: Defining LSTM-CNN model

In above screen defining LSTM-CNN model and then training this model to get below output

	localhost:8888/notebooks/Untitled.ipy	nb			0	\$ 7	≦ <i>l</i> ~	ß
💭 jupyter	Untitled Last Checkpoint: an hou	r ago (autosaved)				e 14	ogout	
File Edit	View Insert Cell Kernel	Widgets Help			Trusted		n 3 O	
🖹 🕇 🗶 ć	è 🗈 ♠ ♦ ▶ Run 🔳 C	➢ Code ∨ □						
	<pre>Model: "model_1"</pre>						>	
	Layer (type)	Output Shape	Param #					
	input_1 (InputLayer)	(None, 1551)	0					
	embedding_1 (Embedding)	(None, 1551, 128)	790272					
	lstm_1 (LSTM)	(None, 32)	20608					
	repeat_vector_1 (RepeatVecto	(None, 116, 32)	0					
	lstm_2 (LSTM)	(None, 116, 32)	8320					
	time_distributed_1 (TimeDist	(None, 116, 1274)	42042					
	activation_1 (Activation)	(None, 116, 1274)	0					
	Total params: 861,242 Trainable params: 861,242 Non-trainable params: 0							
	None Model Loaded							
In [8]:	<pre>#now calculate ROUGE SCORE 1 #create phrase tokenization tokenizedText = tokenizer.en #now predict summary using CD</pre>	code(articles[0], returr		ax_length=512, truncati	on=True).to(device)			
Type here to search		🧧 🔒 🧕) 📄 📜 🖂	🛒 💽 Links g ^R 🗸		2	2:39

In above screen we can see LSTM-CNN modeldetails with different layers and then model is loaded

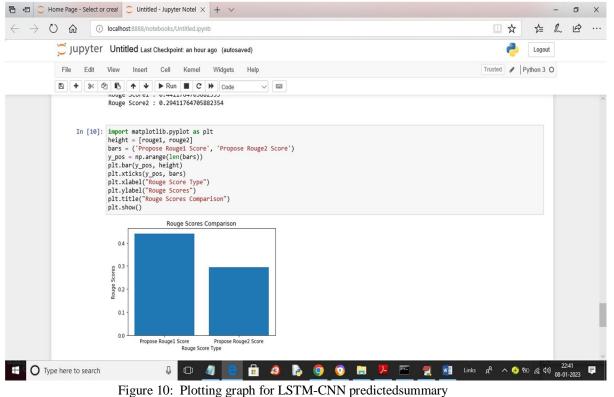


ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

🖥 🖅 🔿 Home Page - Select or creat 🔿 Untitled - Jupyter Notel X + 🗸				-	٥	×
\leftrightarrow \rightarrow \circlearrowright for the state of t		☆	∽≦	h	Ø	
C JUPYTER Untitled Last Checkpoint: an hour ago (autosaved)		0	Logout			
File Edit View Insert Cell Kernel Widgets Help	Trusted	@ F	ython 3 🕻	>		
B + M 2 F + V + Run ■ C H Code V						
Total params: 861,242 Trainable params: 861,242 Non-trainable params: 0 None Model Loaded						^
<pre>In [8]: #now calculate ROUGE SCORE 1 and 2 on test data #create phrase tokenization tokenizedText = tokenizer.encode(articles[0], return_tensors='pt', max_length=512, truncation=True).to(device) #now predict summary using CNW MODEL on given phrases summaryIds = cnn_model.generate(tokenizedText, min_length=30, max_length=120) #now extract summary from predicted array predict = tokenizer.decode(summaryIds[0], skip_special_tokens=True) #calculate rouge scores not est data scorer = rouge_scorer.RougeScorer(['rouge1', 'rougeL'], use_stemmer=True) scores = scorer.score(summary[0]0], prdict) rouge1 = np.amax(scores.get('rouge1')) rouge2 = np.amax(scores.get('rouge1')) print() print('Propose ATSDL Rouge Scores') print('Rouge Score1 : "+str(rouge1)) print() </pre>						
Propose ATSDL Rouge Scores Rouge Scorel : 0.4411764705882353 Rouge Score2 : 0.29411764705882354						~
🐨 🔿 Type here to search 🛛 🖉 💷 🧟 📴 🛱 🧔 🕒 🤦 📄 💆 🗁 🛒 📖 Links	x ^R ⁄	^ 📀 የ	s <i>(ii</i> : 4»)	22: 08-01-	9 2023	Ş.

Figure 9: Applying test article on LSTM-CNNtrained model

In above screen we are applying test article on LSTM-CNN trained model and then calculate Rouge 1 and 2 score on predicted summary and we got Rouge 1 score as 0.44 and Rouge Score as 0.29 which is closer and little higher which is given in base paper.



In above screen we are plotting graph for LSTM- CNN predicted summary rouge scores where x-axisrepresents rouge 1 and 2 and y-axis represents scores



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Home Page - Select or creal Untitled - Jupyter Notel × + ×				-	٥	×
\leftarrow \rightarrow \circlearrowright \bigcirc localhost/8888/notebooks/Untitled.ipynb		☆	₹	h	B	
📁 JUPYter Untitled Last Checkpoint: an hour ago (unsaved changes)		2	Logout			
File Edit View Insert Cell Kernel Widgets Help	Trusted	Pyt	hon 3	•		
<pre> B + 3 < 2 K + Run C > Code B Propose Rouge1 Score Propose Rouge2 Score Rouge Score Type In [*]: #now predict summary from user input text print("Enter Your Text to predict Summary") input_text = input() tokenizedText = tokenizer.encode(input_text, return_tensors='pt', max_length=512, truncation=True).to(device) summaryIds = cnn_model.generate(tokenizedText, min_length=30, max_length=120) predict_summary = tokenizer.decode(summaryIds[0], skip_special_tokens=True) print() print("Predicted Summary : "spredict_summary) </pre>						^
Enter Your Text to predict Summary						
In []:						
In []:						
Type here to search	_к р 🖸	< 📀 🖭	<i>(</i> (; 句))	22:- 08-01-	42 •2023	₽

Figure 11: Input Text Field for Abstractive Summarization

In above screen after executing block you will get text field and then enter some TEXT in that text field and press enter key toget summary output

Image: Construction of the set of t		7 5≜	h l	Ŕ
JUPYTET Untitled Last Checkpoint: an hour ago (autosaved)	e 1997	Logout		
File Edit View Insert Cell Kernel Widgets Help	Trusted	Python 3	•	
01-				
0.0 Propose Rouge1 Score Propose Rouge2 Score				
Rouge Score Type				
In [*]: #now predict summary from user input text				
print("Enter Your Text to predict Summary")				
<pre>input_text = input()</pre>				
<pre>tokenizedText = tokenizer.encode(input_text, return_tensors='pt', max_length=512, truncat summaryIds = cnn_model.generate(tokenizedText, min_length=30, max_length=120)</pre>	tion=True).to(device)			
<pre>predict_summary = tokenizer.decode(summaryIds[0], skip_special_tokens=True)</pre>				
<pre>print() print("Predicted Summary : "+predict summary)</pre>				
<pre>print("Predicted Summary : "+predict_summary)</pre>				
print("Predicted Summary : "+predict_summary) Enter Your Text to predict Summary Ind acquired immunodeficiency syndrome (AIDS). ×				
print("Predicted Summary : "+predict_summary) Enter Your Text to predict Summary				
print("Predicted Summary : "+predict_summary) Enter Your Text to predict Summary and acquired immunodeficiency syndrome (AIDS). ×				
print("Predicted Summary : "+predict_summary) Enter Your Text to predict Summary ind acquired immunodeficiency syndrome (AIDS). × In []:				
<pre>print("Predicted Summary : "+predict_summary) Enter Your Text to predict Summary</pre>				
<pre>print("Predicted Summary : "+predict_summary) Enter Your Text to predict Summary</pre>	Einks gR 🗐 🔿	 • • • • • • • • • • • • • • • • • • •	22:43 08-01-20	223

In above screen in text field I entered some TEXT and press enter key to get below output



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

<pre>Suppyter Untitled Last Checkpoint: an hour ago (unsaved changes)</pre>
<pre>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>></pre>
<pre>print(Enter Four Field & predict Summary) input_text = input() tokenizedText = tokenizer.encode(input_text, return_tensors='pt', max_length=512, truncation=True).to(device) summaryIds = cnn_model.generate(tokenizedText, min_length=30, max_length=120) predict_summary = tokenizer.decode(summaryIds[0], skip_special_tokens=True) print() print("Predicted Summary : "spredict_summary) Enter Your Text to predict Summary The International Journal of Maternal and Child Health (MCH) and AIDS (13MA) is a multidisciplinary, peer-reviewed, global health (MCH) and human immunodeficiency virus (N y) and acquired immunodeficiency syndrome (AIDS). Predicted Summary : (13MA) is a multidisciplinary, peer-reviewed, global health, open access journal that publishes original research and child health (MCH) and human immunodeficiency virus (N y) and acquired immunodeficiency syndrome (AIDS). Predicted Summary : (13MA) is a multidisciplinary, peer-reviewed, global health, open access journal that publishes original research articles.</pre>
areas of maternal and child health (MCH) and human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS). In []:
In []:

In above screen first we are displaying entered TEXT and in blue colour displaying predicted shortsummary from given TEXT. Similarly you can input any text and get predicted summary

V. CONCLUSION

In this study, we construct a unique LSTM- CNN-based ATSDL model in the field of TS, which addresses several importantissues. Recent ETS models are more concerned with syntactic structure, whereas recent ATS models concentrate more on semantics. The advantages of both summarization approaches are combined in our model. The new ATSDL model learns the collocation of phrases after first extractingimportant phrases from the source text using a phrase extraction technique called MOSP. Lastly, we carry out thorough tests on two distinct datasets, and the outcome demonstrates that our model performs better than the most advanced methods in terms of both syntactic structure and semantics. So LSTM-CNN is used for performance analysis and to give abstractive text summarization.

REFERENCES

- [1] Angeli G, Tibshirani J, Wu J et al (2014) Combining distant and partial supervision for relation extraction[C]. EMNLP, pp 1556–1567
- [2] Bing L, Li P, Liao Y et al (2015) Abstractive multi- document summarization via phrase selection and merging[J]. arXiv preprint arXiv:1506.01597Buss D.M., Shakelford T.K. Human aggression in evolutionarypsychological perspective. Clinical Psychology Review, 17, 1997. P. 605–619. DICTIONARIES:
- [3] Cao Z, Li W, Li S et al (2016) Attsum: joint learning of focusing and summarization withneural attention[J]. arXivpreprint arXiv:1604.00125LITERATURE:
- [4] Chen M, Weinberger KQ, Sha F (2013) An alternative text representation to TF-IDF and Bag-of-Words[J]. arXiv preprint arXiv:1301.6770
- [5] Cheng J, Lapata M (2016) Neural summarization by extracting sentences and words[J]. arXiv preprint arXiv:1603.07252
- [6] Chopra S, Auli M, Rush AM (2016) Abstractive sentence summarization with attentive recurrent neural networks[C]. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human LanguageTechnologies, pp 93–98
- [7] Colmenares CA, Litvak M, Mantrach A et al (2015) HEADS: headline generation as sequence prediction using an abstract feature-rich space[C]. HLT-NAACL,pp 133-142
- [8] rkan G, Radev DR (2004) Lexrank: graph-based lexical centrality as salience in text summarization[J]. J Artif Intell Res 22:457–479
- [9] Filippova K, Altun Y (2013) Overcoming the lack of parallel data in sentence compression[C]. EMNLP, pp 1481–1491
- [10] Gu J, Lu Z, Li H et al (2016) Incorporating copying mechanism in sequence-to-sequence learning[J]. arXiv preprint arXiv:1603.06393
- [11] Hu B, Chen Q, Zhu F (2015) Lcsts: a large scale chinese short text summarization dataset[J]. arXiv preprint arXiv:1506.05865
- [12] Li J, Luong MT, Jurafsky D (2015) A hierarchical neural autoencoder for paragraphs and documents[J]. arXiv preprint arXiv:1506.01057
- [13] Lin CY, Hovy E (2003) Automatic evaluation of summaries using n-gram co-occurrence statistics[C]. Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1. Association for Computational Linguistics, pp 71–78 1
- [14] Litvak M, Last M (2008) Graph-based keyword extraction for single-document summarization[C]. Proceedings of the workshop on Multi-source Multilingual Information Extraction and Summarization. Association for Computational Linguistics, pp 17–24
- [15] Lopyrev K (2015) Generating news headlines with recurrent neural networks[J]. arXiv preprint arXiv: 1512.01712
- [16] https://www.e2enetworks.com/blog/an- introduction-to-text-summarization-with-nlp







10.22214/IJRASET

45.98



IMPACT FACTOR: 7.129







INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 🕓 (24*7 Support on Whatsapp)