Abstractive Review Summarization based on Improved Attention Mechanism with Pointer Generator Network Model

J. Shobana
Assistant Professor, SRM Institute of Science and Technology, India.
E-mail: shobanaj@srmist.edu.in

M. Murali
Associate Professor, SRM Institute of Science and Technology, India.
E-mail: muralim@srmist.edu.in

Received December 02, 2020; Accepted February 10, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V18I1/WEB18028

Abstract

Nowadays online reviews play an important role by giving an helping hand to the customers to know about other customer's opinions about the product they are going to purchase. This also guides the organizations as well as government sectors to increase their quality of product and services. So automatic review summarization becomes more important rather than summarizing it manually as it saves time. The aim of this work is to produce a comprehensive summary which includes all key content from the source text. The Proposed Automatic Review Summarization model with improved attention mechanism increases the semantic knowledge and thus improves the summary's eminence. This encoder-decoder model aims to generate summary in an abstractive way. The Pointer generator mechanism solves the problem of rare words which are out-of-vocabulary and the repetition issues are overcome by coverage mechanism. Experiments were conducted on Amazon's mobile reviews dataset reveals that the proposed methodology generated more accurate abstractive review summarization when compared with existing techniques. The performance of the summary report is measured using the evaluation metric ROUGE.

Keywords

Abstractive Summarization, Pointer Generator Network, Attention Mechanism, Coverage Mechanism, ROUGE Metric.

Introduction

A large volume of data has become increasingly accessible in recent years. Digitally processed, allowing a computer to access them for research and interpretation. However, it is an costly process to manually synthesis the data by human efforts when the number of documents is significantly large. Various computerized methods have been proposed to synthesize documents automatically to provide the user with a summarized version. Automatic Review Summarization (Liu, C.L et al. 2011) is an important research study in the field of natural language processing. Text Summarization is done to extract a part or piece of content from the original content, which summarizes the main information from the original content and it is a task of compressing long text into short text without changing the meaning or inference of the whole text (Das. D & Martins, A. 2007). So, Text summary generation is a prominent tasks. The important applications for text summarization in various NLP tasks like classification of text, question and answering (Xiao. L et al., 2018), summarizing news and generating headlines. Text summarization can be included in middle of these applications to reduce the document length. Text summarization is very challenging task because, when we try to summarize a text manually, we usually read it entirely and make some points out of it. Since machines lack human knowledge and language capability, text summarization will be a difficult task. But there are various models based on machine learning to do this task. Most of the approaches in these model include sentence in the summary or not, latent semantic analysis (Campr. M & Ježek. K, 2015), sequence to sequence models (I. Sutskever et al., (2015), reinforcement learning (Wang. Q et al. 2018) and adversarial process.

Generally there are two methods (Wang. Q & Ren. J, 2020) used for automatic summarization, one is extraction and other is abstraction. In extractive approach (Sinha. A et al.2018) the sentences have been picked directly from a coherent summary. So how this method works is that first it would identify the important sections of the text cropping and it would append together to produce the final version. So this approach only depends on the extraction of sentences from the original content. Now a days most of the research is only focusing on abstractive summarization. In abstractive summarization (Rush. A et al., 2015) the approach will produce summary by using advanced natural language processing techniques, such that new shorter texts parts are generated which may not be as original content, which requires rephrasing the sentences and incorporating the information from the text to generate summaries.

The main contributions of this work are,

(i) Proposed improved attention mechanism introduced with BiLSTM encoder - LSTM decoder model enriches the semantic consistency and so the resultant summary consists of all key information of the source text.
(ii) Pointer Generator model addresses the rare words problem and Coverage mechanism overcome the repetition problem. As a result, the summary becomes concise and more accurate.

Related Work

A vast larger part of work in the previous few years has been centered around extractive Text Summarization where catchphrases or sentences from the source text are extracted for summary preparation (Erkan. G & Radev. D.R, 2004) (Wong. K.F et al. (2008) presented a survey about various summarization model and describes various issues solved by Natural Language Processing in the field of Artificial Intelligence.

As proposed by (Yao et.al 2018) Unique in relation to extractive techniques duplicating units from the source article straightforwardly, abstractive summarization utilizes the coherent language for human to sum up key content from the source text. To generate a summary, convolutional encoder to encoder based neural network model is used in this work (Rush. A et al.2015). The model achieves good results over the datasets Gigaword and DUC-20041. BERT model for summarizing the bio-medical text and attains better results which is described by (Moradi. M et al.2020). The author (Hernández-Castañeda et al, 2020) proposed a language and domain independent extractive text summary model using the Lexical-semantic structure. (Alzuhair. A., & Al-Dhelaan. M. 2019) explained a Graph-based summarization methodology by using four multiple weighing schemas combination. DUC 2003 and DUC 2004 datasets are used for evaluation of the model. (Sun, X., & Zhuge, H. 2018) proposed the Semantic link based scientific paper summarization method is proposed. The widely used Reinforced ranking algorithm is utilized to generate a organized extractive summary. Accordingly, abstractive methodologies (Su, M.H et al, 2020) can create substantially more assorted and more extravagant summary reports generated from two-stage transformer model. The recent interesting research area is abstractive text summarization.

A multi head pointer network model (Quian et al. 2015) suggested to enrich the semantics of the summary being generated. Experiments conducted on CNN/Daily mail dataset. Summary generation based on semantic enriched attention is proposed by (Guo, Q et al. (2019). Bi-recursive sequence network with dual-encoder with gain-benefit gate is used to overcome semantic loss in this model. To handle this problem, a pointer mechanism (PM) is proposed to use a new decoder network to point back to out of Vocabulary words and phrases in the input text and copy them into the output. Another copying mechanism derives the representations of OOV words from their corresponding context in the input text (Ding, J et al. 2020).

(Johannes et al. 2018) proposed multidimensional representation based extractive summary model which shows good performance on news article dataset. Multi-Objective algorithm (Sanchez-Gomez et al. 2019) based multi-document extraction model is addressed. (Liang. Z et al. 2020) proposed automatic text summarization encoder- decoder model. soft attention and self-attention mechanisms are utilized to improve the performance of the model. In (Li et al.2020), Multi-sentence summarization model is discussed. This model utilized double attention with copy generator network for solving the repetition problem. Topic Information (Zeng. W, et al. 2016) combined with semantic information summarization model focus on document's core content and by finding the similarity between the key content and the original text, it produces high quality summary.

Proposed Architecture

As depicted in Fig. 1, the proposed model composed of an encoder and an improved attention mechanism, Pointer Network and enhanced Coverage mechanism.

1) A Bidirectional LSTM (Liang. Z et al. 2020) is the encoder which encodes the input text.
2) The decoder is unidirectional LSTM (Sanchez-Gomez et al. 2019) which will output the summary.
3) A improved attention mechanism with pointer network is utilized in this model. This attention mechanism (Ding, J et al. 2020) increases the semantic understanding of the text and Pointer generator network (Yao. K et al.2018) solves the problem of out of vocabulary and rare words.
4) The Coverage mechanism (See. A et al 2017) is used to address the repetition problem in the summary which makes the summary more readable and accurate.

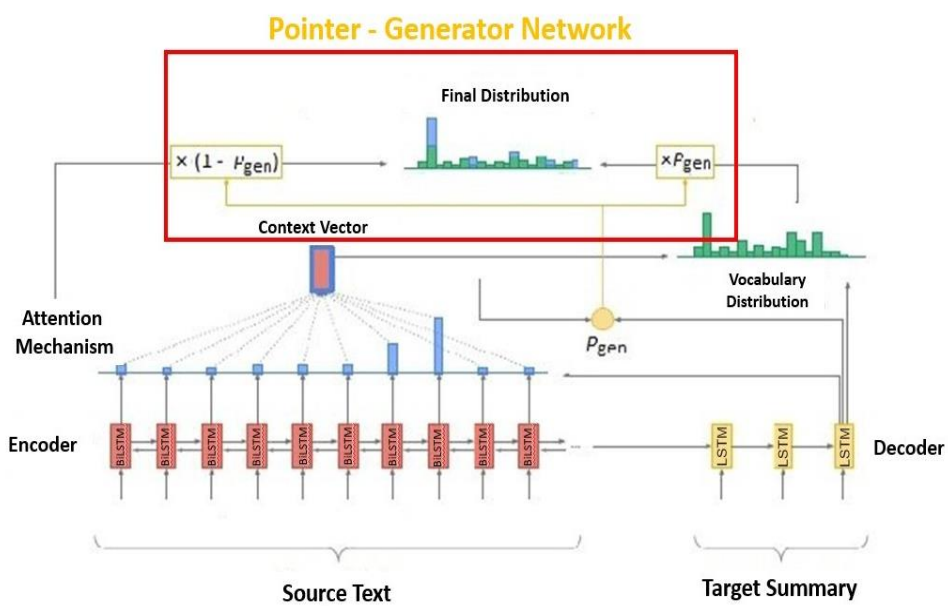

Figure 1 The structure of the proposed pointer generator model with improved attention mechanism

Encoder

In this model, a bidirectional LSTM encoder (Liang. Z et al.2020) is used for processing the text. The encoder converts the sequence of input into word embedding S=$\{s_1, s_2 \ldots\ldots s_n\}$ and the hidden state collection is denoted as HS.

HS is represented as HS= $\{h_1,h_2 \ldots\ldots h_n\}$.

The encoder needs to focus on the core information for producing better summary report so the improved attention mechanism helps the encoder to match itself with the key information from the text collected dynamically. The generated hidden state h by combining before and after will contribute towards the calculation of the vector y, the key information.

Improved Attention Mechanism

The important content of the text will be missed from the focus and the generation of summary is affected during decoding if the encoder encodes an excessive amount of useless data. So in such case, it becomes essential work to highlight the important features of the source text. Self-attention mechanism is needed to match the encoder with itself to dynamically collect salient features in the text. Sequence of the source text will be converted to word embeddings and the bidirectional LSTM encoder is used for processing. so sequence will be obtained from the hidden states of the encoder. The design of the improved attention (Ding, J., et al 2020) is expressed as,

$$HS\{h_1,h_2 \ldots\ldots h_n\}= BiLSTM\{s_1, s_2 \ldots s_n\} \quad (1)$$

Where, $\{s_1, s_2 \ldots\ldots s_n\}$ denotes the source text sequence and HS denotes hidden states collection of $\{h_1,h_2 \ldots\ldots h_n\}$. The encoder Bi-LSTM (Liang. Z et al.2020) mapped the source text sequence to the vector of the hidden state.

$$x_i^j\left(h_i,h_j\right)=V^T \tanh\left(w_1 h_i+w_2 h_j\right)+l_{att} \qquad (2)$$

Where w_1,w_2 denotes the optimizing parameter and V_T is also an optimization parameter. l_{att} represents the learnable parameter whereas tanh is a non-linear activation function. The importance of the optimization Parameter is the mapping of the semantic representation [21] of the current state with the hidden state of encoder-decoder. x_j^i denotes the i^{th} hidden state's similarity with jth hidden state.

$$e_i = \sum_n x_i^n \qquad (3)$$

$$b' = \text{softmax}(e) \qquad (4)$$

Where the essence of i^{th} hidden state is denoted by e^i and the attention weights (Liang. Z et al. 2020) of all hidden state is calculated at b' and key information is related to weights can be collected from b'. Softmax is the activation function which choose the value of [0,1]. The significance of i^{th} hidden state is expressed by the equation (3).

$$Y = \sum b' HS^{T} \quad (5)$$

The normalization of the probability distribution attention [22] by the softmax function is done to make b' more better. b' denotes the weight vector and it sum up the attentional weights of all hidden states. The key information of vector y is obtained from the probability distribution of b' weight vector.

Pointer Generator Network

Pointer generator network model (You, F et al. 2020) is introduced to extract the original document's text which will be used to improve the accuracy of the summary. So as to solve the trouble of out of vocabulary which will generate text precise, the pointer network model (O. Vinyals et al, 2015) is added on the idea version with self- interest mechanism. The context vector of the encoder and the output state of the hidden layer inside the decoder are enter into two linear connection layers and generating a opportunity distribution on the dictionary to predicting new words, the calculation system (Guo, Q et al, 2019) is as follows:

$$e_l^t = v_i^T \tanh(w_3 h_i + w_4 d_t + l_{att}) \quad (6)$$

where, w3 and w4 are optimization parameters and l_{att} denotes the learnable parameter. d_t denotes the decoding state. The attention in the pointer network gives the information about the important word at each moment which will be useful for the decoder's prediction. At certain moment, while generating the weight distribution, the attention is to added with the decoding state d_t.

$$b^t = soft\max(e^t) \quad (7)$$

b_t is the current moment's attention distribution. Softmax function is used for normalization.

$$s_t = \sum_i b^t h_i \quad (8)$$

The sum of the weights of the encoder's hidden state is calculated as s_t(Guo, Q et al,2019). S_t is called as context vector. The encoder 's context vector and the decoder's hidden layer output state are the inputs of the connection layer which will generate the probability distribution (Yao, K et al, 2018) of the vocabulary P_{Vacob}.

$$P_{vacob} = soft\max(c'(c[d_t, s_t] + a) + a') \quad (9)$$

The learnable parameters are a, a', c, c'. The context vector is S_t.The Probability distribution is denoted as P_{Vacob} [13]. p_w is the probability of predicting the word w in the vocabulary. The formula used for the calculation (is

$$P(w) = P_{vocab}(w) \quad (10)$$

To solve the out-of-vocabulary words, the model through pointer network need to copy words from the source directly or need to generate.

The pointer probability generation [13] based on the context vector S_t, the input x_t at the time t and the output of the decoder O_t is calculated as follows.

$$P_{gen} = \sigma\left(u_s s_t + u_0 o_t + u_x x_t + p_{ptr}\right) \quad (11)$$

Where, p_{gen} falls in the range [0,1] and U_s, U_o, U_X and p_{ptr} are called as learnable parameters. The sigmoid function is denoted as σ which is used for activation. To copy word from the source text or to generate the word is to be decided by p_{gen} finally.

The formula (See, A et al 2017) for calculating all words from the original text and the probability distribution of the dictionary obtained word W is

$$P(w) = P_{gen} P_{vacob}(w) + \left(1 - P_{gen}\right) \sum_i : w_i = w^{a_i^j} \quad (12)$$

The above equation denotes that if the word w is not chosen from the source text then $\sum_i$ $:w_i =$ $w^{a_i^j}$ = 0. p_{gen} denotes pointer generator and is determined from equation (11) and P_{Vacob} refers to the probability distribution. If the word w is out of vocabulary then the value P_{Vacob} (w)=0. The cross-entropy loss function is utilized by the model for calculating the loss,

$$loss_t = -\log P\left(z_t^*\right) \quad (13)$$

For every timestamp t, z*t is the target predicted word. The entire sequence's overall loss is calculated as,

$$loss = \mathrm{I/T} \sum_t^T = 0^{loss}{}_t \quad (14)$$

The total training steps are denoted as T. The overall sequence loss is called at time step t. The overall sequence loss is called at time step t.

Coverage Mechanism

Repetition is a not unusual trouble in lots of natural language processing duties and is a common trouble for all model (See, A., et al. 2017), mainly in neural system translation obligations, which regularly use sequence-to-sequence as a benchmark structure. Due to the fact the neural machine translation challenge is much like the text summarization mission and turned into developed in advance, the neural device translation optimization version is used in many text summarization optimization schemes.

The essential model, the coverage mechanism (Tu, et al. 2016) is used to address the repetition issues. The coverage mechanism focus the attention version to awareness on non-repeating words by using covering the vector. The coverage vector is the sum of the attention distributions computed by using all previous prediction steps. The version has already paid attention to the phrases of the unique textual content. simultaneously, the loss feature is used to punish repeated attention to reduce repetition. consequently, a truncation mechanism is used to enhance the coverage mechanism, making the loss feature more accurate, reducing the range of repetitions and growing reliability.

In coverage mechanism (You, F et al. 2020), the coverage vector M^v is defined. The vector M^v is the total of all distributions of the attention over the previous time stamps of the decoder

$$M_i^v = \sum_{i=0}^{i-1} a_i^{'} \quad (15)$$

Where the coverage vector $M_0^v = 0$ as the first step in decoding does not cover any text in the source. As the attention mechanism uses the coverage vector as a extra information, the updated formula [13][26] is given as follows,

$$e_i^t = V_i^T \tanh\left(w_3 h_i + w_4 h_4 + w_m M_i^v + l_{att}\right) \quad (16)$$

Where, the learning parameter w_k is at the same length as v_i. The main aim of the coverage mechanism (You, F et al. 2020) is to help the model to avoid repeated attention to the same words. An additional coverage loss function is defined to avoid or penalize repetitively attention to same parts. The loss function (Yao, K et al.2018) defined already in equation (14) is rewritten as,

$$loss = 1/T \sum_{i=0}^{T} \left(loss_i + \lambda \sum_i \min\left(a_i^t, M_i^t\right)\right) \quad (17)$$

Where lambda λ is a hyper- parameter and i denotes the decoding time-step. The new-loss function is generated from the old loss calculation through lambda λ hyper-parameter. The goal of coverage mechanism is to avoid the decoders attention over some parts of the source text with the help of the previous attentional weights. Thus, the coverage mechanism act as a repetition problem avoider from the decoder and the encoder.

Decoder

Unidirectional LSTM is utilized as decoder for generating summary. The goal of the decoder is to produce better summary without semantic –loss of core content. Semantic -loss in the major problem in the process of decoding. The word prediction accuracy of the decoder decreases gradually if the length of the summary increases. The proposed method solves this problem with the help of improved attention mechanism and pointer generated network and thus generated the summary with high accuracy.

Experiments and Analysis

The dataset details, the experimental setup, evaluation metric and the analysis of results are explained in this section**.**

Dataset Details

Amazon mobile review dataset is used for conducting experiments. This dataset consists of 8000 reviews about the product mobile.

Experiment Setup

The experiment is conducted on Pytorch framework with NVIDIA 1050Ti GPU. The encoder is the bidirectional LSTM with two layers and unidirectional LSTM with two layers is used as the decoder for this work. The embedding size and hidden units are set as 512.The learning rate is initially set as 1* 10^{-3}. The batch size is fixed as 64 and the model has been trained using Adam Optimizer (Liang, Z et al.2020) with cross entropy loss. The gradients are clipped at 10.0.

Evaluation Metric

The performance of the proposed model is evaluated using ROUGE evaluation metric. ROUGE (Lin, C.Y. Lin 2004) stands for Recall-Oriented Understudy of Gisting Evaluation which is the standard metric for evaluating the effectiveness of the automatic generated summaries against the human-generated summary. By calculating the overlapping lexical units, the quality of the summarization is measured by ROUGE tool kit (Lin, C.Y. Lin 2004). ROUGE-1 score is computed by finding out the overlapping of unigrams between System – generated and human-generated summaries. ROUGE-2 score is computed by finding out the overlapping of bigrams between System generated and human-generated summaries. Rouge-L measures the longest common sequences. Experiments conducted on the dataset Amazon mobile reviews using metrics ROUGE-1, ROUGE-2 and ROUGE-L. The results of three evaluation metrics are tabulated in table1 as follows.

Baseline Summarization Models

The widely used three base-line summarization methods are RNN, Seq-Seq model with attention, Pointer generator + attention mechanism. The results of the proposed model is compared with the results obtained from these models. The brief description of these models are:

- **RNN** (I. Sutskever et al. 2015): This model represents recurrent neural network architecture for generating summaries. RNN architecture takes source document as input and generate the summary as output.
- **Seq2Seq+attention** (Rush, A.M et al. 2015)**:** Represents standard seq2seq neural network model based on baseline attention mechanism to improve the semantics of the text in the summaries. Attention mechanism is employed in many NLP applications like machine translation etc.
- **Pointer Generator + attention** (Li. Zhexin et al. 2020)**:** Introduces pointer generator network model with attention mechanism. This model extracts words from the original text to produce accurate summaries. This address the problem of sparse words.

Results and Discussion

In this section, the results of the proposed model and other three widely used base-line models are tabulated in Table 1. The comparison between the proposed model and existing classical model are analyzed based on amazon's mobile review dataset.

Table 1 Results of the Proposed model and base-line summarization models on Amazon's mobile review dataset

Summarization Model	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**
RNN	31.6	16.24	27.32
Seq2Seq +attention	37.33	18.31	33.15
Pointer Generator +attention mechanism	39.8	21.2	36.4
Proposed Model (Pointer Generator + Improved attention mechanism + Coverage mechanism)	41.2	22.4	37.1

From the comparative analysis done on amazon's mobile review dataset, it is clear that our proposed model (Pointer Generator +improved attention mechanism + Coverage mechanism) yield better results. The proposed system achieves higher score of 1.4% on ROUGE-1, 1.2% on ROUGE-2 and 0.7% on ROUGE-L than the Pointer Generator + attention mechanism. The proposed model attains 3.87% ROUGE-1 score, 4.09% on ROUGE-2 and 3.95% on ROUGE-L which are higher than the Seq2Seq +attention model. The RNN model achieves 31.6% ROUGE-1 score, 16.24% on ROUGE-2 and 27.32% on ROUGE-L.

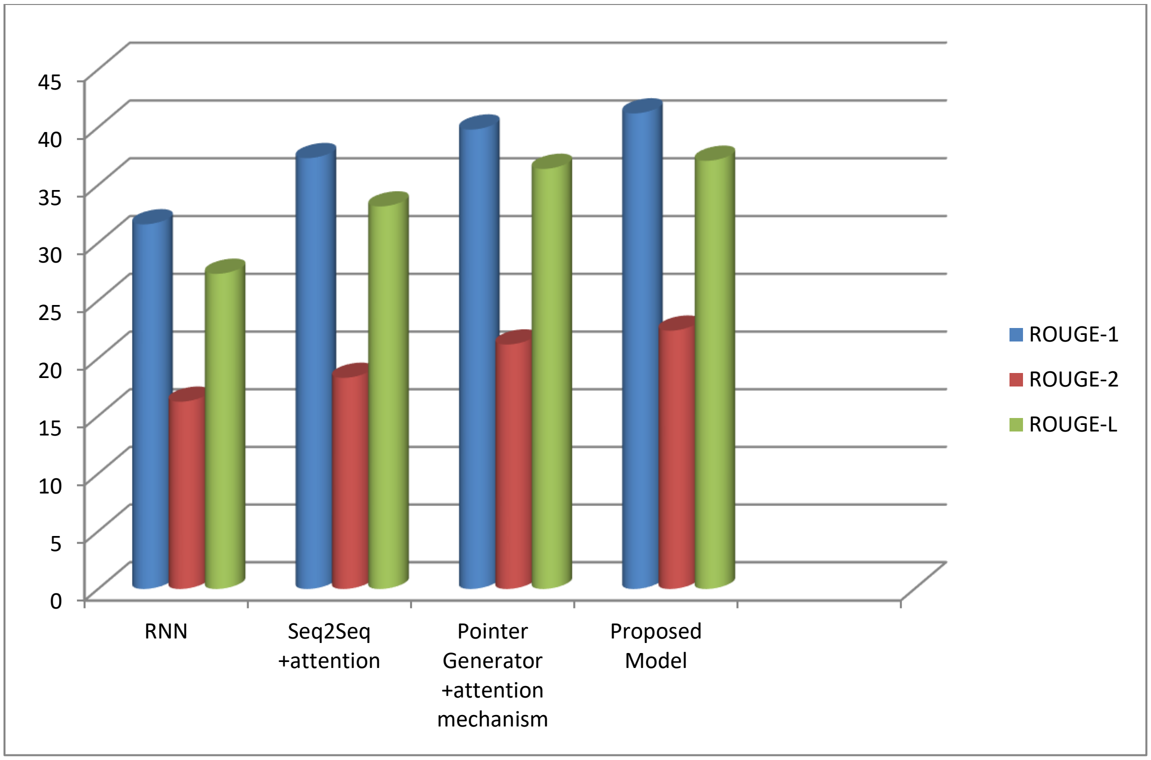

Figure 2 Comparative analysis of the proposed model with three baseline methods

In this work, the comparative performance analysis is done between the proposed system and three popular existing methods namely RNN, seq-seq model with attention, pointer generator with attention mechanism and the results are shown in fig 2.

Analysis of Review Summarization

The review summary generated by the proposed system is compared with the existing system namely Pointer Generated with Coverage Mechanism in order to analysis the quality.

Table 2 Review summarization from sample Reviews by existing and Proposed models

Review-1
In addition to being smaller and lighter than the iPhone 11, it has flat sides and a flat screen. We've been living with curved edges on iPhones for six years, since 2014's iPhone 6. So part of my affection might just be that the iPhone 12 feels new. But it's also that it feels like a throwback to the iPhone 4 and 5 models, which were the last iPhones whose design I truly loved. Despite all of those flat edges, the seams and the corners are beveled just enough to make it comfortable to hold. Overpriced and without charger too. This is the best iPhone yes. The Max model gives you the better screen and overall is a camera beast. I would highly recommend this to anyone. Regarding value for money, its sad that we live in a country where Falling Rupee, Import Duty and High GST are the reasons for such high poricing. A rich country like USA pays only $1099 (Rs. 81,500) whereas its priced a whopping Rs. 48,400 more in India. That's almost 60% more. Shame!. It is a ceramic glass wouldn't hurt to buy a screen guard. If u can afford it u can totally go 4 it.

Review -2
The rails on the iPhone 12 are matte finish aluminum, and I prefer them to the glossy steel on the Pro models. Unfortunately, the rear glass is super glossy, super prone to picking up fingerprints, and as susceptible to picking up tiny little micro-abrasions as ever. Most people will put a case on their phone anyway. This phone along with iPhone 12 pro has 6 gb ram. I do not game on the phone so unfortunately I am not in a position to answer this query. But looking at the processor speed as well as the ram in this phone it should not be an issue. Please note that this phone has 60hz refresh rate screen compared to phones equipped with 90hz, 120hz. or more refresh rate screens which would make a difference while gaming on the phone.

Review -3
Not recommended. It is a waste of money. Be logical enough to make a call on this. Impressed! I've tested it with iphone11 and other iphones, it's remarkable. It has unquestionably better video quality and touch sensitivity. The screen is bigger than I thought. I loved the Gold variant. But if u have a 11 max or a X max I wouldn't suggest for an upgradation. I feel that the next iphone will have way more cooler features so wouldn't hurt to wait. My previous iphone was 6, so in my case am totally peachy with this phone.

Review -4
Great display, battery, cameras, usability, ease of access &a whole lot more. Running out of Kidneys to sell to buy iPhones in India. Just no words, what a beauty. Upgrading from iPhone X, this feels so big and heavy. But do not has much discomfort while using, you will get used it of. Make sure to get a decent Screen guard and clear case just to safeguard this beauty. Had pre booked this iPhone 12 Pro Max. Amazon de livered it promptly within four days of launch. Kudos to this lightening, fast delivery. Now about device, Since I upgraded from 11, I could notice the difference between A13 and A14 speed. Miraculously faster. Performance wise this indeed is a notch up.

Review -5
Build quality is significantly improved. This design which Apple brought back from iPhone 5 series is absolutely lovely! These two factors give it a classy and stylish look combined with solid and durable finish. The phone does feels premium when one holds it. All the more reasons to go ga-ga about this pretty beast iPhone is indeed pricey but if one can shell these bucks; you should not blink your eye before buying. The upgrade does provide you with enhanced durability, performance and looks!

Reference Summary (Human-Generated)
This phone is has come out of shells!!!...It has good video and audio quality. A good camera. It has solid finish which comes with multiple colors. It has high resolution and clarity and price is high.
Existing System Generated Summary (Pointer Generated + Attention mechanism)
It has classy and stylish look combined with solid and durable finish. & a whole lot more. This is the best iPhone yes. This model gives the best camera and resolution. It is highly recommended to everyone. Regarding value for money, its sad that we live in a country where Falling Rupee, Import Duty and High GST are the reasons for such high pricing. !It looks so good and it has a good enhanced version of camera quality.

Proposed System Generated Summary (Pointer Generated + Improved Attention mechanism + Coverage mechanism)
iphone 12 has a flat screen, great display, battery, good camera, usability, ease of access. It is overpriced and it comes without charger. quality is significantly improved. It has classy and stylish look combined with solid and durable finish.

Some sample mobile reviews from amazon's mobile review dataset about the product iPhone 11 are taken for analysis. The sample reviews, reference summary, Base-line model's summary and summary of proposed model are shown in the table.2. The reference summary is the human generated summary. The proposed model generated summary is compared with the existing classical model namely Pointer generator +Attention mechanism model generated summary. The proposed model shows superior performance in summarizing all key information with proper grammatical structure than the base-line model.

Conclusion

Text summarization will reduce the document text size by giving a brief summary with core context and the highlights of the salient features in the original text. A concise review summary will be valuable for the customer in getting other user's opinions expressed in thousands of reviews. The Proposed system is evaluated using the dataset amazon's mobile reviews. The generated summary shows improvement in following up the grammatical structure. The quality of the generated summary is assessed by the ROUGE evaluation metric. The performance of the proposed methodology increases significantly when compared with the baseline methods.

References

Liu, C.L., Hsaio, W.H., Lee, C.H., Lu, G.C., & Jou, E. (2011). Movie rating and review summarization in mobile environment. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(3), 397-407.

Das, D., & Martins, A. (2007). A survey on automatic text summarization. literature survey for language and statistics. *II Course at CMU*, 4, 192–195, 57.

Xiao, L., Wang, N., & Yang, G. (2018). A reading comprehension style question answering model based on attention mechanism. *In IEEE 29th International Conference on Application-specific Systems, Architectures and Processors (ASAP)*, 1-4.

Campr, M., & Ježek, K. (2015). Comparing semantic models for evaluating automatic document summarization. *In International Conference on Text, Speech, and Dialogue*, Springer, Cham, 252-260.

Sutskever, O.V., & Le, Q.V. (2015). Sequence to sequence learning with neural networks. *In Conference on Advances in Neural Information Processing Systems*, Montreal, QC, Canada, 3104–3112.

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. https://doi.org/10.3390/app9214701

Wang, Q., & Ren, J. (2020). Summary-aware attention for social media short text abstractive summarization. *Neurocomputing*, 425, 290-299. https://doi.org/10.1016/j.neucom.2020.04.136

Sinha, A., Yadav, A., & Gahlot, A. (2018). Extractive text summarization using neural networks, arXiv:1802.10137. http://arxiv.org/abs/1802.10137

Rush, A.M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization, *In Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2015, 379–389.

Erkan, G., & Radev, D.R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization, CoRR, *Journal of Artificial Intelligence Research*, *22*, 457-479. http://arxiv.org/abs/1109.2128

Wong, K.F., Wu, M., & Li, W. (2008). Extractive summarization using supervised and semi-supervised learning. *In Proceedings of the 22nd international conference on computational linguistics*, 985-992.

El-Kassas, W.S., Salama, C.R., Rafea, A.A., & Mohamed, H.K. (2020). Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*, 113679. https://doi.org/10.1016/j.eswa.2020.113679

Yao, K., Zhang, L., Du, D., Luo, T., Tao, L., & Wu, Y. (2018). Dual encoding for abstractive text summarization. *IEEE transactions on cybernetics*, *50*(3), 985-996. https://doi.org/10.1109/TCYB.2018.2876317

Moradi, M., Dorffner, G., & Samwald, M. (2020). Deep contextualized embeddings for quantifying the informative content in biomedical text summarization. *Computer methods and programs in biomedicine*, *184*, 105117.

Hernández-Castañeda, Á., García-Hernández, R.A., Ledeneva, Y., & Millán-Hernández, C.E. (2020). Extractive automatic text summarization based on lexical-semantic keywords. *IEEE Access*, *8*, 49896-49907. https://doi.org/10.1109/ACCESS.2020.2980226

Alzuhair, A., & Al-Dhelaan, M. (2019). An Approach for Combining Multiple Weighting Schemes and Ranking Methods in Graph-Based Multi-Document Summarization. *IEEE Access*, *7*, 120375-120386. https://doi.org/10.1109/ACCESS.2019.2936832

Sun, X., & Zhuge, H. (2018). Summarization of scientific paper through reinforcement ranking on Semantic Link Network. *IEEE Access*, *6*, 40611-40625.

Su, M.H., Wu, C.H., & Cheng, H.T. (2020). A Two-Stage Transformer-Based Approach for Variable-Length Abstractive Summarization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *28*, 2061-2072. https://doi.org/10.1109/Taslp.2020.3006731

Vinyals, O., Fortunato, M., & Jaitly, N. (2015). Pointer networks, *In Proceeding on Advances in Neural Information Processing Systems*, 2692–2700.

Guo, Q., Huang, J., Xiong, N., & Wang, P. (2019). MS-pointer network: Abstractive text summary based on multi-head self-attention. *IEEE Access*, *7*, 138603-138613. https://doi.org/10.1109/ACCESS.2019.2941964

Ding, J., Li, Y., Ni, H., & Yang, Z. (2020). Generative Text Summary Based on Enhanced Semantic Attention and Gain-Benefit Gate. *IEEE Access*, *8*, 92659-92668. https://ieeexplore.ieee.org/document/9091541

Zenkert, J., Klahold, A., & Fathi, M. (2018). Towards extractive text summarization using multidimensional knowledge representation. *In IEEE International Conference on Electro/Information Technology* (EIT), 0826-0831.

Sanchez-Gomez, J.M., Vega-Rodríguez, M.A., & Pérez, C.J. (2019). An Indicator-based Multi-Objective Optimization Approach Applied to Extractive Multi-Document Text Summarization. *IEEE Latin America Transactions*, *17*(08), 1291-1299.

Liang, Z., Du, J., & Li, C. (2020). Abstractive social media text summarization using selective reinforced Seq2Seq attention model. *Neurocomputing*, *410*, 432-440.

Li, Z., Peng, Z., Tang, S., Zhang, C., & Ma, H. (2020). Text summarization method based on double attention pointer network. *IEEE Access*, *8*, 11279-11288. https://doi.org/10.1109/ACCESS.2020.2965575

Zeng, W., Luo, W., Fidler, S., & Urtasun, R. (2016). Efficient summarization with read-again and copy mechanism, CoRR, vol. abs/1611.03382, 2016. http://arxiv.org/abs/1611.03382

You, F., Zhao, S., & Chen, J. (2020). A Topic Information Fusion and Semantic Relevance for Text Summarization. *IEEE Access*, *8*, 178946-178953. https://doi.org/10.1109/ACCESS.2020.2999665

See, A., Liu, P.J., & Manning, C.D. (2017). Get to the point: Summarization with pointer-generator networks, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, *1*, 1073–1083.

Tu, Z., Lu, Z., Liu, Y., Liu, X., & Li, H. (2016). Modeling coverage for neural machine translation, *In Proceedings of 54th ACL*, Berlin, Germany, 76-85.

Lin, C.Y.L. (2004). Rouge: A package for automatic evaluation of summaries. *In Text summarization branches out: Proceedings of the ACL2004 workshop*, 1095–1104.