A Novel Multi-Layer Sparse Regularizer based GRU Model for Consumer Reviews Summarization

Sourav Sinha[1], Dr. Revathi Sathiya Narayanan[2*] and Dr. Indrajit Mukherjee[3]

[1]Research Scholar, Department of Computer Science & Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India.
sourav_cse_phd_17@crescent.education, Orcid: https://orcid.org/0000-0002-3327-8215

[2*]Professor, Department of Computer Science & Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India. srevathi@crescent.education,
Orcid: https://orcid.org/0000-0001-9584-5089

[3]Assitant Professor, Department of Computer Science& Engineering, Birla Institute of Technology Mesra, Ranchi, India. imukherjee@bitmesra.ac.in, Orcid: https://orcid.org/0000-0001-8839-9981

Received: December 24, 2022; Accepted: January 26, 2023; Published: March 30, 2023

Abstract

Text Summarization is considered as potential study area in recent past for its well-established growing popularity in domain specific areas. The consumer dominated society approved the summarization process as a tool for communication in shortest time span with wide coverage of valued basic knowledge. The Auto Text Summarization is developed digitally with various sophisticated tools and techniques in compressed opinion building process delivering millions of consumer's verdict at one scratch. The newest methodology of deep learning neural network had been adopted to narrate the precise fact with proposed novel MSRGRU (Multilayer Stacked Regularizer GRU) model for consumer review summarization technique. The adaptation of the model by proper utilization of sparse regularization process enhances quality of predicted summaries. Eventually the accuracy level of the model is remarkable as compared to the existing state-of-the -art techniques.

Keywords: Text Summarization, Consumer Reviews, Sparse Regularization.

1 Introduction

The word "**Summarization" is** meant for shortened transformed version for a useful informative knowledge bearing article or text. The word is usually considered as antonym of comprehension where the aspect is elaborative illustration with vast documentation. The Text Summarization rather does the task of generating a brief expression for the source text protecting most of the original knowledge information.

In the past the concept of Text Summarization had developed with time as per the requirement of users. Individuals were in habit to do summarization of original text materials manually. They used to shorten the whole text into few paragraphs where each sentence might have represented the compressed paragraph matter while few words stood for sentences of original text materials. The efforts were enough

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 14, number: 1 (March), pp. 159-173. DOI: *10.58346/JOWUA.2023.I1.013*

*Corresponding author: Professor, Department of Computer Science & Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India.

to solve the task of summarization for limited source articles in those days depending on availabilities and needs. However, the summarized text might have projected more of the summary writer's knowledge on the topic rather than basic information covered by the author in original text.

In the regime of consumerism where internet and mobile technology are dominating key factors in modern lifestyle, the Text Summarization is a significant domain for knowledge transfer on daily affairs in recent review-based customer driven culture. Nowadays relevant information is communicated, rather transmitted electronically in lightning speed to the millions of individuals. Virtually the society has become a gathering of reviewers exchanging opinions on thousands of aspects comfortably. Transferring summarized write-up on salient observations with knowledge-based opinions has become part and parcel in our lifestyle with deserved ranking labels.

In other words, the Text Summarization can be compared as a distilled product of useful relevant informative details from original text source. The significance of Text Summarization depends on three major derivative results.

A. Reading time reduction
B. Rapid progress in search process of knowledge-based information.
C. Gaining more knowledge with wide coverage area in shortest time span.

Text Summarization-Components of Knowledge Transfer Process

The key objective of Text Summarization is to enhance knowledge transfer process with concentration and condensation of valuable information gathered from single or multiple documents (Abualigah L, 2020). The components in this process of transfer are explained in the following paragraphs. The components involved in structuring the processes are being diagrammatically shown in Figure-1.

Purpose– The purpose of Text Summarization is served with demonstration of concise knowledge on analytical results on different domain specific purposes pertaining to customer reviews. There are thousands of such domain specific fields where opinions are being transmitted through reviews expressing individual's sentiment or feelings. The purpose may differ with specific domains but ultimately it helps to progress in performance independently. The purpose may vary area wise with reviewed domains like product reviews, service reviews, movie reviews, literature reviews and so on.

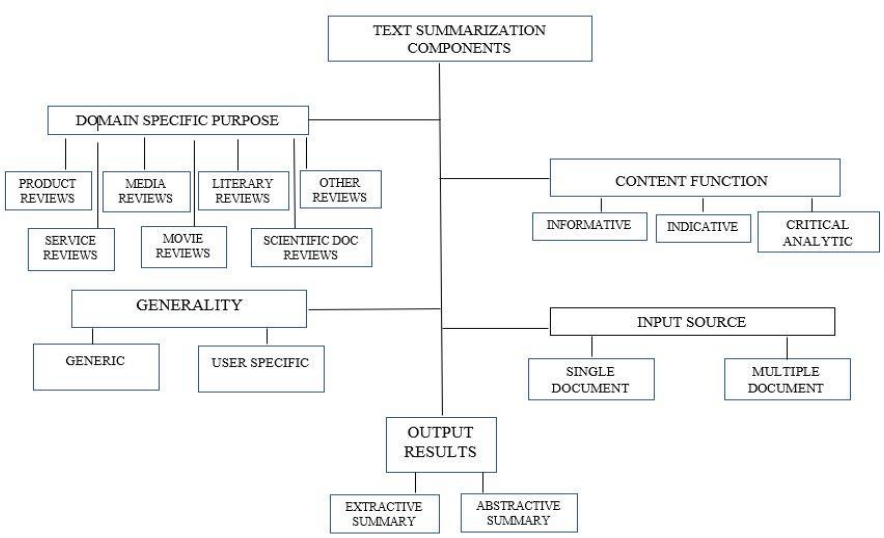


Figure 1: Components of Text Summarization

Content – The content of the summary may direct towards three different pathways for knowledge transfer for specific approach in intelligence (Hahn U, 2020). The Informative summaries perform as an alternative to source contextual material with relevant or novel information in compressed manner. The indicative summaries show the pattern of classical information retrieval approach making users alert on relevant sources. And lastly the Critical analytic summaries substantiate information gist by incorporating opinion statement and bringing expertise opinion which is not available in source document alone.

Generality – The opinion expressed in Text Summarization is generic when it would address to a broad community without any target to group. User focused summaries in other sense are directed to the specific needs for any group.

Input– The input source may vary between entities to multiple documents focused on domain specific regime. The summarization of reviews for movie or literary aspects is usually made with single source text while service domain or product domain reviews need to be carried out with help from multiple source texts.

Output – This is the result component being the most significant one in the phenomenal process of Text Summarization. The outputs could have two different categories, namely Extractive and Abstractive summarization (Sarkar A, 2018).

Text Summary Categorization

The Extractive summarization is commonly prepared by selection of important parts of the original text covered with dominated phrases or words highlighting information pertaining to original concept. This is a precise version of basic text materials with the usage of original vocabulary so that the basic text sense is retained. This type of summarization does not need much to think over during preparation or in other words easy in expression category. Hence it is more popular but less innovative in the usage of new way of grammatical representation or communication through new vocabulary beyond the text.

The Abstractive summarization is more like human preferred text briefing where the concept or main idea is conveyed in shortest possible linguistic version devoid of vocabulary applied in source document. This type of summarization is more acceptable since the process of knowledge transfer is smooth and takes care comfortably without any distortion due to grammatical or semantic constrain.

Objective

The present work is emphasized on developing a competitive idea with supervised deep learning model to generate predicted abstractive summary at least on 3 sets of Amazon product review comments of electronic or household domestic items and the result of studies is found to be satisfactory when compared with existing state-of-the- art technique.

2 Literature Survey

The Twenty-first Century began with a bang having an outburst in technological development throughout all continental blocks across the world. The most memorable impact had been enrolled in applications of skill in the field of Information and Communication technology. The progressive growth in internet and mobile facilities along with developed data storage equipment introduced the major chunk of population to a newly discovered world of Communication through electronic gadgets. In recent past the step forward was so rapid that each space of technical or business areas had shown remarkable positive impact. Even societal culture got impressed with the development of various sites, forums, and social media structure for easy accessibility among thousand individuals from different countries. Last two decades the rapid growth in internet connectivity caused in the generation of voluminous text data in digital format in sudden accelerated pace. Virtually users are faced with the difficulties in handling such heavily stacked information and confused with right way of tackling those data volume for utility purpose with the help of proper sorting and extraction process. It is beyond the capacity of individual's manual effort to select the right path for relevant information at right occasion from the heavily laden text materials. Hence the need was felt to develop an affordable and acceptable tool or method to manage the data dimension wise. With vigorous attempts and sincere human efforts technocrats invented the computer-based machine learning applications to reduce the problem with optimum technical solution. Automatic Text Summarization is introduced with the idea to meet the requirement to handle the huge volume of data meticulously to extract right information at appropriate instant to generate concise representation of source text (Soumya S, 2011) (Li Z, 2020). Some of the existing text summarization work had produced with improved performance result as shown in Table 1.

Table 1: A Comparative Study of Various Review Text Summarization Approaches

Author	Year	Summarization Type	Dataset	Language	Methodology	Performance Metrics
Kumar Y J	2017	Abstractive	Document Understanding Conference	English	Adaptive Neural Fuzzy System	Precision Recall F-Score Rouge
Sarkar A	2018	Extractive	Single Text Document	Bangla	Term Frequency Semantic Similarity using wordnet	Runtime Complexity (Measured in seconds with respect to Text Size in KB)
Duan X	2019	Abstractive	Giga-word DUC2004	English	Contrastive Attention Mechanism Transformers	Rouge
Niu J	2019	Abstractive	Giga-word CNN/DM testing corpora	English	Sun Attention Mechanism (extension of Bahdanau Attention) using sequence-to-sequence model	Rouge-1 Rouge-2 Rouge-L
Madhuri J N	2019	Extractive	Multiple Text Document	English	Sentence Ranking	System Relevance
Li P	2020	Abstractive	Giga-word DUC LCSTS	Chinese	Multi-Attention Learning Supervised Unsupervised Recurrent	Rouge-1 Rouge-2 Rouge-L
Tian Y	2020	Abstractive	Amazon Products	English	Two Stage Reinforcement Learning	Rouge -1 Rouge -2 Rouge -L
Harinatha SRK	2021	Extractive	News Articles	English	Text Rank BERT	Amount of Computation Time based on number of rows
Joshi M L	2022	Extractive	Tourism and Health Corpus	Hindi	Hindi Word net Fuzzy Semantic Graph	Precision Recall F-Score Rouge

Automatic Text Summarization reduces the text materials of varied dimensions to a condensed output with unaltered basic conceptuality as compressed. The function acts on either extracting a concise article from the original text or reconstruct a new summary with source text information. Automatic text summarization became an area prospective field for research in recent past precisely for natural language processing. Existing tools of summarization depends on the mapping of labeled usual summaries manually after feature identification from source text article. They ignore the inner structure and semantic feature information of the original document. The Automatic text summarization had experienced a great learning in solving that genuine but difficult task. (Wang Z, 2021) implements automatic abstractive text summarization with sequential encoder to decoder network using two recurrent neural networks namely word RNN and sentence RNN work together at word level and sentence level based on hybrid attention model. Author applies the proposed approach on LCSTS dataset and confirms to achieve very high ROUGE scores. Sentence Attention helps in guiding the word level attention to achieve more informative generated and readable summaries. (Hanunggul P M, 2019) develops a LSTM model with local attention mechanism to extract the abstractive text summaries using Amazon fine gained reviews dataset and also compares the result with global attention-based model. The paper explains about the better ROUGE-1 score achieved with global attention model due to more words present in actual summary whereas local attention model gives higher ROUGE-2 due to subset of words contained in the actual summary. (Gupta A, 2021) propose pre-trained transformer architecture-based models to perform the task of text-summarization. The paper evaluates and performs comparative study between the pre-trained models namely Pipeline-BART, T5, modified BART and Pegasus. According to the paper, fine tuning the pre-trained transformers language-based model can perform outstanding results and T5 model outperforms all other models. (Singal D, 2020) proposed the model on abstractive summary generation methods to train and to achieve at optimum of summarization in dialogue system. The model involves the understanding of text material and subsequent improved summarisation while maintaining original semantic quality. (Al-Sabahi K, 2018) designed a hierarchical structured self-attention mechanism-based sentence classifier using Bidirectional LSTM hidden states on CNN/Daily mail dataset to perform extractive summarization. The paper compares the proposed model with various baseline models like Lead3 model, RNN model, reinforced abstractive summarization model, graph-based approach, URANK model. It concludes saying that the proposed model outperforms other baseline models due to the hierarchical attention represents the exact document structure and self-attention results in good word embedding, hence improves the text summarization task during the automated learning process.

3 Research Work

Text Summarization is growing more significant particularly during assessment and analysis of large volume of consumer reviews data. In the current paper we propose a novel model named Multi-Layer Stacked Sparse Regularizer based GRU (MSRGRU) for review text summarization relying upon Sequence-to-Sequence based on encoder decoder architecture. During this approach, we initiate the steps like pre-processing the consumer reviews in the given dataset as input and create the Glove embeddings for each word in the review text. This Glove Embedding is being feed into the encoder and decoder layers. Subsequently we concatenate the encoder and decoder with (Bahdanau D, 2014) attention (Vaswani A, 2017) (Qiu D, 2022) mechanism. The sparse activity regularizer is finally applied on the decoder outputs in the time distributed dense layer. This decoder output is ultimately used to predict summary of consumer reviews.

Dataset

In our present research activity, we retrieved various datasets namely Amazon Fine Food Reviews and Amazon Consumer Review for Electronic Products by the process of downloading from kaggle.com website and Home Kitchen Amazon review dataset has been hosted in the GitHub website. The size used for text summarization in the deep Learning model for namely Amazon Fine Food Reviews was 40000 Review Comments and Amazon Consumer Review for Electronic Products was 20000 Review Comments respectively. The size of Home and Kitchen dataset used for text summarization was nearly 80000 Review comments

Proposed Methodology

Sequence-to-sequence models are gaining wide importance due to their success in various fields of data science. The Sequence-to-sequence model architecture helps in solving complex NLP problems like machine translation, text summarization, question answering, image captioning generation, video captioning generation respectively. In the current paper, we propose a sequence-to-sequence model with three layers Gated Recurrent Unit (GRU) as encoder and single layer Gated Recurrent Unit (GRU) as decoder in architecture. As part of our unique contribution, we try to integrate the sparse regularizer in order to calibrate the deep learning models for minimizing the loss function and overcome the over fitting and under fitting machine learning models. We also integrated the existing Bahdanau Attention mechanism to our proposed model to utilize the attention layer to target keywords during the review summarization process. The overall architecture of the proposed model has been shown in pictorial demonstration below in Figure 2.

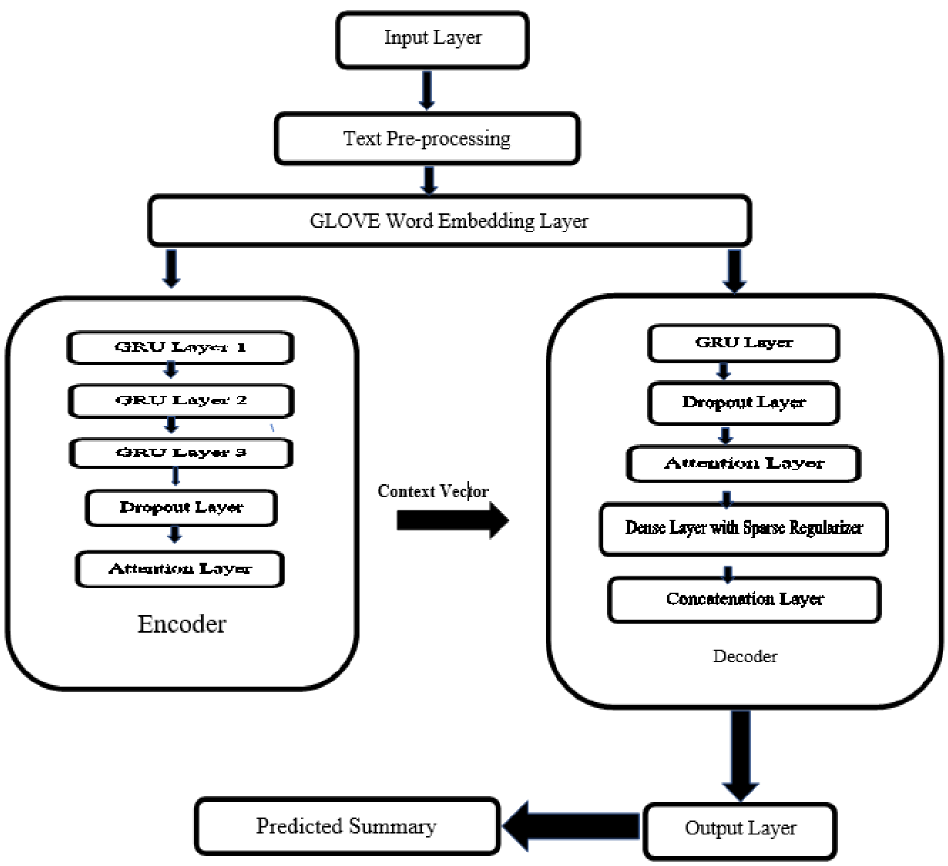


Figure 2: Architecture Diagram for Flowchart of Proposed MSRGRU Model

Input: Amazon Consumer Review Text
Output: Predicted Abstractive Summary
Performance Metric: Accuracy, Loss
Library: Pandas, Numpy, NLTK, Contractions, Keras, Tensorflow, BS4
Definition: Tokenizer (list of tokens), Stopwords (List of stopwords)

Function Text_Preprocessing (Input):
text = convert to lowercase and remove small words without meaning
tag_removal = remove the XML tags from the sentence
lowercase = change text to lowcase letters
contractionsfix = fixing the contractions
stopwords = retrieve all the stopwords from NLK package
cleantext = remove the stopwords, apostrophe symbol and punctuations from the given text
cleansummary = cleaning the summary om unnecessary contractions and short words
End Function

Function Bahdanau_Attention:/*Attention Mechanism proposed by D Bahdanau*/
Compute the Attention Weights for encoder hidden states
Calculate the Context Vector
Compute the Attention Vector
Finally calculate the Alignment score and then perform softmax
End Function

Function Sparse_Regularizer:
s_h = Average activation value of the hidden units
s_p = Sparsity parameter
w = Sparsity weight
KL = Compute the KL divergence as KL (p/h) = s_p*log(s_p/s_h) + (1-s_p)*log(1-s_p/1-s_h)
return sum(W*KL)
End Function

Function Main (Input)
max_txt_len = set the maximum allowed length of text for the consumer reviews
max_summary_len = set the maximum allowed length of the summary for consumer reviews
Text_Pre-processing (Input)
p_tr,p_val,q_tr,q_val = split the data into train and test dataset
p_tokenizer_sequence = tokenize the text associated to the respective consumer reviews into integer sequences
q_tokenizer_sequence = tokenize the summaries for respective consumer reviews into integer sequence
embedding_layer = Apply Glove embedding technique on p tokenizer sequence on encoder inputs
GRU_Enclayer1 to GRU_Enclayer3 = multi-Layer stacked GRU as encoder seq to seq model
Dec_EmbeddingLayer = Apply Glove embedding technique on p tokenizer sequence on decoder inputs
GRU_Declayer = Single Layer GRU as decoder seq to seq model
Attn_Layer = Integrate the Bahdanau Attention Mechanism (Function BahdanauAttention)

Dense_DecLayer = Apply the SparseRegularizer Function in the dense decoding layer on q tokenizer
plot_Loss = Plot the loss function against Epoch
plot_Accuracy = Plot the accuracy function against Epoch
consumer_review = Retrieve the actual review entered in consumer review
original_summary = Retrieve the original summary entered in consumer Review
predicted_summary = Predicted summary of maximum length of 10 words
End Function

Algorithm 1: Pseudo Code for Proposed MSRGRU Model for Consumer Review Text Summarization

We divide the review summarization methodology into following phases – 1) Pre-processing phase 2) Word Embedding Phase 3) Learning Phase 4) Evaluation Phase

- **Pre-Processing Phase**

Text Pre-processing is very crucial task during the text mining process before performing text summarization process. Usage of textual data without cleaning noisy elements in natural language processing tasks can make hindrance and confusion for deep learning model in predicting the summary data. Hence the preliminary step is to perform text pre-processing without deviating from the actual objective of the given problem statement. During this text pre-processing regime, we reduce the dimensions by further tokenizing the consumer review texts along with respective highlights to represent the given review text as bag of words. We further improve the data performance with stop words and punctuations deletion, elimination of unwanted XML tags, removal of unwanted symbols or special characters and essentially mapping the various contractions present in the consumer review sentences.

- **Word Embedding Phase**

Feature Extraction from texts is another important task during the text mining process. In this process meaningful representations of words present in the sentence are extracted and converted into numerical vectors which represent the features. Word embedding is used to represent the vectors for each word in the form of 2D array with rows and columns used to represent each of the vocabularies. We preferred using the GLOVE (Pennington J, 2014) word embedding in our proposed model with the following hyper-parameters as shown in Table 2. The main reason for choosing GLOVE as it captures both global and local context information of the words and builds the word co-occurrence matrix. This co-occurrence matrix helps in deriving the semantic relationships between the words using both global and local context of information.

Table 2: Hyper Parameters Configured for Proposed MSRGRU Model

Embedding Parameters	Value
Word Embedding Dimension	100 d features
Number of pre-trained word embedding	6 billion tokens
Latent Dimension	300
Number of Vocabulary	400000-word vectors

The cost functions for glove embedding of two-word vectors w_p and w_q in is represented by C:

$$C = \sum_{p,q=1}^{V} g(Ypq)\ (WpT\ Wq - \log Ypq)^2 \quad \text{(Equation 1)}$$

$$\text{Where } g(Y_{pq}) = \begin{cases} \left(\frac{y}{y_{\max}}\right)^{\propto} & if\ y < y_{\max} \\ 1 & otherwise \end{cases} \quad \text{(Equation 2)}$$

- **Learning Phase**

For given consumer review summarization y_1, y_2, y_3……...y_n are the input consumer review text sequence which are tokens obtained from the consumer reviews and z_1, z_2, z_3……z_n are the output review summaries which are generated by the proposed MSRGRU model based using deep learning based on abstractive summarization.

a) Encoder state - The primary purpose of the encoder state is to take the input consumer review text and vectorize the tokens obtained from the features extracted from the tokens in consumer reviews

y_1, y_2, y_3……...y_n are the input text sequence extracted tokens from the consumer reviews text

h_1, h_2……$h_{(t-1)}$, h_t are the sequence of hidden steps at time step t (where h represents cumulative internal hidden states)

$$h_t = W_{yh}Y_n + V_{hh}h_{t-1} \quad \text{(Equation 3)}$$

$$U_t = \sigma\,(V_{hu}h_t + W_{yu}Y_n) \quad \text{(Equation 4)}$$

Where Y_n is the hidden state of n^{th} input at given timestep t

$h_{(t-1)}$ is hidden state response at timestep t-1

U_t represents the update gate for given timestep t and σ is sigmoid activation function

$$R_t = \sigma\,(V_{hr}h_{t-1} + W_{yu}Y_n) \quad \text{(Equation 5)}$$

where R_t is reset gate for given time step t

$$h_t = h_tU_t + h_{t-1}(1 - U_t) \quad \text{(Equation 6)}$$

The stacked GRU applied in the proposed MSRGRU model consists of three GRU layers as shown in Figure 3. The GRU layer reads the word embedding in forward directions and prepares the hidden states (h_t) which generates the best representation of the input consumer reviews text.

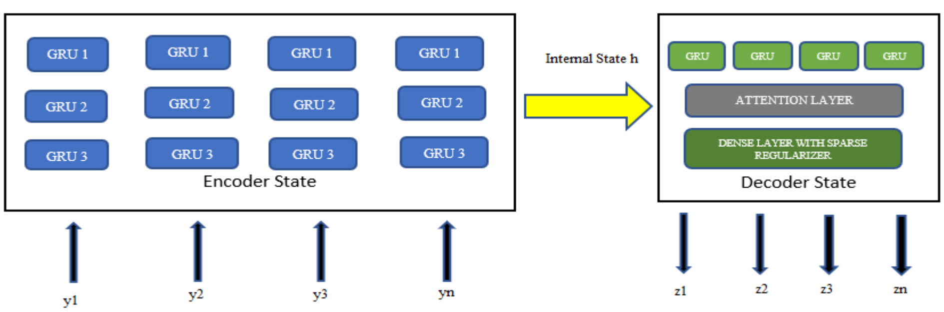


Figure 3: Encoder-Decoder State Architecture Diagram of Proposed MSRGRU Seq2Seq Model

b) Decoder state -During the decoder state, we use single GRU layer along with Bahdanau Attention Mechanism and to add novelty to the model we propose to add sparsity-based regularization technique as activity regularizer in the dense layer of the proposed MSRGRU model. The objective of adding the attention mechanism in our proposed model is to focus on specific keywords which are more important from the complete list of words which is present in the input sequence that is generated in the target sequence. In our proposed MSRGRU model, the output of the decoder layer is passed on time distributed dense layer. We configure this layer with our novel approach to integrate the sparse activity regularizer to our proposed MSRGRU model.

During the Bahdanau Attention Mechanism in the proposed MSRGRU model, the alignment score is computed using the decoder input which is used to compute the context vector

Alignment Score $e_{ij} = V_a^T \tanh(Wh_{t-1} + Uh_e)$ ………...... (Equation 7)
W = weights score for decoder and U = weights score for encoder
Attention weights $\alpha_{ij} = \exp(e_{ij}) / \sum \exp(e_{ij})$ ……............. (Equation 8)
Context vector $C_i = \sum \alpha_{ij} h_s$ where h_s = Decoder hidden state

This context vector which is computed in the above equation is then concatenated to the decoder hidden state to generate next decoder output.

$Z_n = F(Z_{n-1} - 1, SR_i, C_i, Y_i)$ ………...(Equation 9)

Z_n = nth output state of the decoder at given timestep t

SR_i-Sparse Activity Regularizer applied in the decoder state
C_i = context vector for given timestep t
Y_i = Input text sequence passed drop out

The deep neural network model is mostly prone to over fitting as it tries to learn from individual patterns rather than following a generalized approach on unknown dataset. Hence regularization plays an important role to overcome these limitations and minimize or eliminate the test errors. Activity Regularizer is often used in the deep neural network models to learn the internal representations of input data. In our proposed MSRGRU model, we propose a novel mechanism to integrate sparse activity regularizer using KL divergence for learning the sparse learned representation of the data. In our proposed MSRGRU model, we define activation function and Sparsity Regularizer cost function as derived in the following equations:

$\sigma(x) = 1/1 + e^{-x}$ where $\sigma(x)$ is activation function
β = weight of sparsity term
γ = average of activation of hidden unit
$\gamma = \text{Max}(\sum \sigma x)/n$ ………… .... (Equation 10)

Now after applying K-L divergence

$KL(\delta, \gamma) = (\sum \{ \delta \log \delta/\gamma + (1-\delta)\log(1-\delta)/(1-\gamma)$…... (Equation 11)

Now the total cost function is defined

Total Sparsity Cost (SR_i) = $\beta \sum KL(\delta/\gamma)$ ……………….................... (Equation 12)

The outputs of the decoder state are represented as $z_1, z_2, z_3 \ldots z_n$ later used in the evaluation phase for generating target sequence which is used for consumer review summary prediction.

- **Evaluation Phase**

During the learning phase, the proposed MSRGRU model is trained with existing dataset with review texts and original summaries. While the learning phase completes the training session, the process moves to the evaluation phase where encoder decoder GRU architecture analyses new review input text sequence and predict the target sequence for summaries by utilizing the hidden states at specific timestamp. In each timestep with new input of text sequence data the encoders and decoders are updating the internal states with timesteps progression. The encoder- decoder architecture usually performs effectively in short review text sequence, but it becomes much difficult to learn properly for larger review text sequence. Hence the attention mechanism is applied to increase the significance of specific word in the review text sequence in generating target summaries. Subsequently the novel idea to integrate sparse activity regularizer has been introduced in the dense layer to improve the quality in consumer review summarization.

4 Results and Discussion

The sequence-to-sequence analysis is the usual processing technique in deep learning arena which deals with the complex domain of data sequence summarization. Text is most common example of the sequenced data where the current observation is being linked and dependent on previous observations. Hence consumer reviews can be considered as pure example of time series of textual data. In present study, we have used three sets of Amazon product review text data being tested with greater number of NLP models in which a few of them considered to be state-of-the-art deep learning technique. However, the level of success for the model as proposed MSRGRU is being proved to be well when compared with other established methods.

Performance Evaluation Metrics

Varieties of approaches are being taken into consideration for comparison and finally achieved the optimum level in developing proposed MSRGRU model. In this process different Amazon product reviews data sets are being passed through different existing modules of summarization program. Each time the program had been allowed to run for reaching the consistency level in 50 epochs to the maximum. The review data sets in proposed module had been made to run with batch size of 256. The program was executed in Google Colab notebook in HP eleventh generation core i3 laptop.

Table 3: Performance Evaluation Chart with Amazon Reviews Datasets for Comparative Analysis

Data set	Description	Methodology	Accuracy	V-accuracy	Loss	V-loss
1	Amazon Fine food Products Reviews	Layered GRU	0.71	0.68	3.15	3.54
		Layered LSTM	0.70	0.68	1.50	1.74
		Bi LSTM	0.61	0.62	2.50	2.55
		Bi GRU	0.62	0.62	2.31	2.51
		Proposed MSRGRU	**0.72**	**0.69**	**2.85**	**3.30**
2	Amazon Electronic Products Reviews	Layered GRU	0.69	0.68	1.56	1.77
		Layered LSTM	0.70	0.69	1.41	1.70
		Bi LSTM	0.57	0.59	2.69	2.83
		Bi GRU	0.59	0.60	2.54	2.76
		Proposed MSRGRU	**0.71**	**0.70**	**2.72**	**2.97**
3	Amazon Home Kitchen Products Reviews	Layered GRU	0.74	0.67	1.28	1.89
		Layered LSTM	0.69	0.68	1.53	1.78
		Bi LSTM	0.64	0.62	2.10	2.31
		BI GRU	0.64	0.63	1.96	2.26
		Proposed MSRGRU	**0.75**	**0.68**	**1.39**	**1.99**

The Performance Evaluation Chart in Table 3 is manifesting the results on Accuracy and Loss details of the proposed MSRGRU model along with existing models considered for comparative analysis.

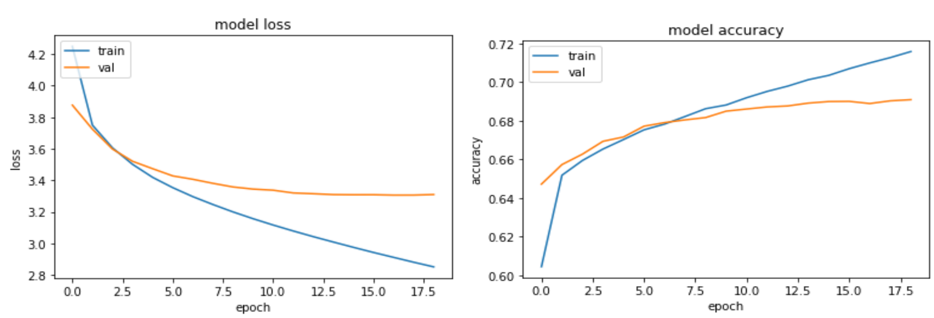
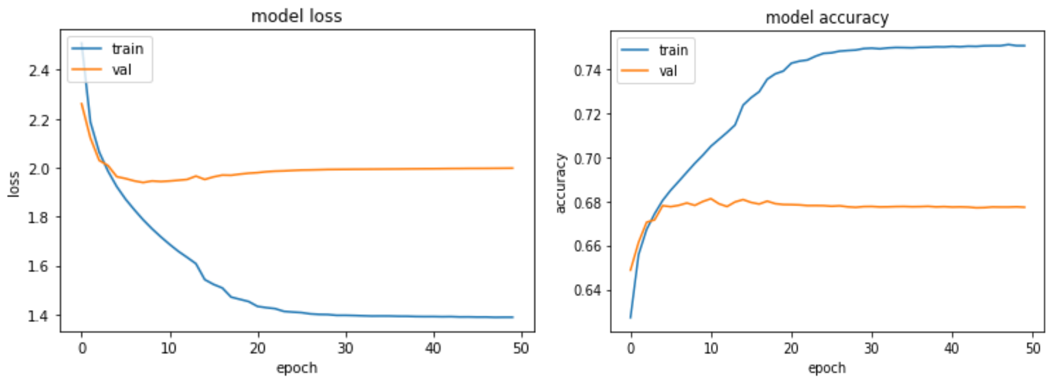


Figure 4: Loss/Accuracy vs Epoch Diagrams of MSRGRU Model for Fine Food Reviews



Figures 5: Loss/Accuracy vs Epoch Diagrams of MSRGRU Model for Home Kitchen Reviews

The graphical presentations of Model loss or Model accuracy variation against epochs of iteration for different data sets are shown in Figure 4 & 5. In Figure 4 the graph in first part is showing continuous reduction of Model loss with epoch rise while in later one the trend is just opposite with increasing accuracy level with rise of epochs. However, the graphs in Figure 5 show the identical trend pattern up to 0-25 epochs after which the trends become horizontal.

Reviews Summary Generation

The proposed MSRGRU model was trained over different dataset and finally was put on to infer summary generation to the test dataset. During the process of evaluating results the automatic generated predicted summary with proposed model is initially compared with the original human summary. Also, we noticed refinement in the predicted summary takes place only after the application of the sparse regularizer which acts an integral part of denser layer for decoder state in proposed MSRGRU model. Hence below few samples of consumer review texts are attached along with original human summary, predicted summary with layered GRU model and predicted summary with proposed MSRGRU model.

Sample 1: Home Kitchen Dataset

Consumer Review Input Text 1: *These glasses are OK. After reading the reviews I thought they would be the perfect addition to my Thanksgiving table to dress it up however, I decided to use my other (more delicate and expensive crystal) beverage set instead. These glasses seem a bit rugged and have a sort of massed produced quality to them and I wanted something new to go with my very expensive China set. For the price I paid, they are sufficient for use as everyday beverage glasses. They are heavy but I was expecting something more refined I guess for a formal dinner party. But they are great for what they are. I might give them away as a gift. I'm sure they will be appreciated. These do not look like nor are they a high-quality crystal beverage set. Hope this helps in making your decision.*

Original Human summary: great everyday set

GRU Predicted summary: great glasses

MSRGRU Predicted summary: **nice set of everyday use**

In the first example (Sample 1) the consumer explains with an elaborative message on his feelings for the items he purchased recently. No doubt the message is quite lengthy but helpful. Still, it is questionable on number of individuals those having time to study the full matter. The short messages communicated in the original human summary and GRU predicted summary are also incomplete. However, the proposed MSRGRU predicted summary carry the core sentiment of the message.

Sample 2: Home Kitchen Dataset
Consumer Review Text 2: *Does everything you'd expect it to. Heavy, but isn't that the point? No quality issues and wobbling for us as well*.
Original Human summary: it rolls
GRU Predicted summary: great
MSRGRU Predicted summary: **heavy duty**

The script in next example (Sample 2) is much shorter and unable to communicate properly on the item purchased. Original human summary carry less information on its identity. But still the proposed MSRGRU model can sense the machinery item's effectiveness in its predicted summary.

Sample 3: Fine Food Reviews Dataset
Consumer Review Input Text 1: *With a consistency like a Little Debbie Brownie and a taste like the no-Pudge brownies, these will not be for everyone. However, if you are on a restricted diet (no soy, no gluten, no dairy), these bars fit the bill. They are not overly sweet, although there are 25g of carbs in each bar. The main ingredient is dates, so that is what provides the majority of the texture. The bars soft and dense, sweetened with agave nectar with no other sugar added. It does contain 5g of fibre and 7g of protein, but only has 10% iron and 4% calcium RDA.

My biggest issue is the calorie impact: at 190 calories, I want something that has a little more flavour, especially if there is little added nutrition (these can't be used as meal replacement, just as an occasional snack). The wrapper says "Your chocolate prayers have been answered." I seriously doubt anyone eating this bar would agree with that statement.*
Original Human summary: not bad for vegan gluten free bar
GRU Predicted summary: not bad
MSRGRU Machine Predicted summary: **good but not great**

The third example (Sample 3) differs from earlier samples in totality. Here the message is simple and clear about the item. All ingredients are being mentioned in the script even with health care details. The original human summary is still unable to communicate vital points while GRU predicted summary is incomplete. The MSRGRU proposed model passes exact feature details as required through its predicted summary.

The above examples argue in favour on the ability of the proposed model to generate abstractive summaries in totality. It is beyond doubt the predicted summaries generated in the proposed model are more in abstractive orientation rather than extractive nature.

5 Conclusion

In this paper we propose a multi layered stack regularizer based GRU model for consumer review summarization. The sparse activity regularizer proposed in this paper improves the performance of this automatic summarization technique; moreover, it helps in generating more precise and enlightening summaries. The use of attention mechanism helps in focusing on specific key areas to further improve the summaries. The primary objective behind adding the sparse activity regularizer is to bring on three refinements to the consumer review summaries output.

1. To address over fitting problem
2. To learn the sparse features hidden within consumer review text
3. To improve the model ability for generalising any new textual Consumer review data.

6 Future Work

The successful usage of sparse activity regularizer in dense layer at decoding state improved the level of accuracy and generating quality wise better predicted summaries for the consumer reviews. This is an area to focus on future to get more realistic features addition in the summaries and use other essential information retrieval metrics like bleu score, word error rate to measure the summarization performance. Moreover, it will be interesting just like this model if the utility of sparse regularizer can be applied on product recommendation systems for improving predictive results.

References

[1] Abualigah, L., Alfar, H.E., Shehab, M., & Hussein, A.M.A. (2020). Sentiment analysis in healthcare: a brief review. *Recent advances in NLP: the case of Arabic language*, 129-141.

[2] Al-Sabahi, K., Zuping, Z., & Nadher, M. (2018). A hierarchical structured self-attentive model for extractive document summarization (HSSAS). *IEEE Access*, *6*, 24205-24212.

[3] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate.

[4] Duan, X., Yu, H., Yin, M., Zhang, M., Luo, W., & Zhang, Y. (2019). Contrastive attention mechanism for abstractive sentence summarization.

[5] Gupta, A., Chugh, D., & Katarya, R. (2022). Automated news summarization using transformers. In *Sustainable Advanced Computing: Select Proceedings of ICSAC 2021*, 249-259. Singapore: Springer Singapore.

[6] Hahn, U., & Mani, I. (2000). The challenges of automatic summarization. *Computer*, *33*(11), 29-36.

[7] Hanunggul, P.M., & Suyanto, S. (2019). The impact of local attention in lstm for abstractive text summarization. *In IEEE International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 54-57.

[8] Harinatha S.R.K., Tasava B.T., Qomarriyah M.M. (2021) Estimating Extractive Summarization Techniques on News Articles. *International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, Indonesia, 2021*, 88-94.

[9] Joshi, M.L., Mittal, N., & Joshi, N. (2022). Improving the performance of semantic graph-based keyword extraction and text summarization using fuzzy relations in Hindi Wordnet. *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1-18.

[10] Kumar, Y.J., Kang, F.J., Goh, O.S., & Khan, A. (2017). Text summarization based on classification using ANFIS. *Advanced Topics in Intelligent Information and Database Systems 9*, 405-417.

[11] Li, P., Bing, L., Wei, Z., & Lam, W. (2020). Salience estimation with multi-attention learning for abstractive text summarization.

[12] 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.

[13] Madhuri, J.N., & Kumar, R.G. (2019). Extractive text summarization using sentence ranking. *In IEEE international conference on data science and communication (IconDSC)*, 1-3.

[14] Niu, J., Sun, M., Rodrigues, J.J., & Liu, X. (2019). A novel attention mechanism considering decoder input for abstractive text summarization. *In ICC IEEE International Conference on Communications (ICC)*, 1-7.

[15] Pennington, J., Socher, R., & Manning, C.D. (2014). Glove: Global vectors for word representation. *In Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, 1532-1543.

[16] Paul, A., Wang, L., Selvi, S.S.D., & Rangan, C.P. (2020). Non-transferability in proxy re-encryption revisited. *Journal of Internet Services and Information Security*, *10*(3), 1-30.

[17] Qiu, D., & Yang, B. (2022). Text summarization based on multi-head self-attention mechanism and pointer network. *Complex & Intelligent Systems*, 1-13.
[18] Sarkar, A., & Hossen, M.S. (2018). Automatic bangla text summarization using term frequency and semantic similarity approach. *In IEEE 21st International Conference of Computer and Information Technology (ICCIT)*, 1-6.
[19] Singhal, D., Khatter, K., Tejaswini, A., & Jayashree, R. (2020). Abstractive Summarization of Meeting Conversations. *In IEEE International Conference for Innovation in Technology (INOCON)*, 1-4.
[20] Soumya, S., Kumar, G.S., Naseem, R., & Mohan, S. (2011). Automatic text summarization. In *Computational Intelligence and Information Technology: First International Conference, CIIT 2011, Pune, India. Proceedings*, 787-789. Springer Berlin Heidelberg.
[21] Tian, Y., Yu, J., & Jiang, J. (2019). Aspect and opinion aware abstractive review summarization with reinforced hard typed decoder. *In Proceedings of the 28th ACM International Conference on Information and Knowledge Management* 2061-2064.
[22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*.
[23] Wang, Z. (2021). An Automatic Abstractive Text Summarization Model based on Hybrid Attention Mechanism. *In Journal of Physics: Conference Series, 1848*(1), 1-6. IOP Publishing.
[24] Dataset: Electronic Products - https://www.kaggle.com/datasets/lievgarcia/amazon-reviews
[25] Fine Food Review - https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews
[26] Home Kitchen - https://nijianmo.github.io/amazon/index.html

Authors Biography

Sourav Sinha received his B.E degree in Information Technology and M.E degree with distinction in Software Engineering from Birla Institute of Technology, Mesra, Ranchi, India in 2009 and 2011 respectively. He is currently pursuing towards his PhD degree at the B.S. Abdur Rahman Institute of Science and Technology Chennai, India. His areas of interest are Machine Learning, Artificial Intelligence, Information Retrieval and Text Mining.

Dr. S. Revathi obtained her B.E.(CSE) in 1994, M.E.(CSE) in 2000 and PhD in 2013. She is currently working as Professor in the Department of Computer Science and Engineering at B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India. She has published more than 50 peer reviewed journal and conference papers in the domain of Mobile Ad hoc Network, Internet of Things, Network Security, Data mining etc.

Dr. Indrajit Mukherjee received M.Sc., Electronics degree from the University of Ranchi, India in 1995, MCA degree and Ph.D. (Computer Science) from Birla Institute of Technology, Mesra, Ranchi, India in 2001 and 2013 respectively. Currently, he is an Assistant Professor in the Department of Computer Science & Engineering, Birla Institute of Technology, Mesra, Ranchi, India. His research interests include Web-Based learning, Data Mining, Big Data Handling, Web Service Applications, and Soft Computing. He has more than 20 research papers and 3 Ph.D. awarded to his credit. He is a lifetime member of the Indian Society for Technical Education (ISTE), India.