An Improved Attention Layer assisted Recurrent Convolutional Neural Network Model for Abstractive Text Summarization

DEEPA NAGALAVI[1], M. HANUMANTHAPPA[2], RAVIKUMAR K[3]

[1*]Department of Computer Science and Applications, Bangalore University, Bangalore, India.
2 Department of Computer Science and Applications, Bangalore University, Bangalore, India.
3 Kalinga Institute of Industrial Management(KIIT), Bhubaneshwar, India

e-mail: deepatnagalavi@bub.ernet.in, hanu6572@bub.ernet.in, ravinaac@gmail.com
*Corresponding Author: deepatnagalavi@bub.ernet.in, Tel.: +00-91-9880631087

Abstract— In the last few years text summarization has gained widespread attention across industries, especially in media and publications, research, business intelligence etc where it helps exploiting large documents to generate a new one with summarized inferences without losing the aspects. However, majority of the conventional approaches either employs extractive summarization or the abstractive summarization for single-document settings. On contrary, above stated application environments demand abstractive summarization in multiple document settings. Though, amongst the major efforts developed so far the attention based neural network methods have performed potentially; however their efficacy under multiple-documents setting and aspect-sensitive summarization has remained confined. Considering it as motive, in this research a novel and robust Improved Attention Layer assisted Recurrent Convolutional Neural Network (IA-RCNN) model is developed for Abstractive Text Summarization in multiple document settings. Unlike conventional efforts we have employed state-of-art techniques such as Sequence-to-Sequence (S2S) paradigm where the inclusion of RCNN, which is modified as Recurrent Neural Network (RNN) encoder technique for text summarization. Our proposed abstractive text summarization model encompasses semantic feature extraction, dependency parsing, semantic role labeling, semantic information etc where it exploits the structural, syntactic, and semantic information of the input text data to generate the summary. Unlike conventional Attention based Summarization, in our proposed model at first performs Clustering and Sentence Merging, which is followed by Transition-based Abstract Meaning Representation (TAMR) parsing, whose output is encoded by means of an improved Tree-LSTM RCNN model, which eventually generates single summarized sentence as output. The overall proposed model is tested with multiple text documents where simulation results affirm satisfactory performance for real-time applications.

Keywords*:* Abstractive Text Summarization; Multiple Document Summarization; Recurrent Neural Network;, Sequence-to-Sequence Paradigm; Abstractive Mean Representation.

(Received November 6st, 2018/ Accepted December 5st, 2019)

1. INTRODUCTION

Understanding intend of the different users and their opinion and finding out a cumulative (agreeable to all) statement can be vital for making an optimal decision. Undeniably, these days a firm, an agency, or an individual exploits opinions of others to make optimal decision in daily-activities. Gathering different opinions from the geographically or demographically independent users has become an inevitable practice for industries to understand audience's key perception and concerns. Undeniably, such explorative efforts have been playing decisive role in navigating a firm, allied associates, stakeholders etc. Additionally, it has been assisting non-commercial activities such as social issues assessment and problem identification, ideology formulation etc. On the other hand, the exponential rise in software computing environment and allied technologies have revitalized socio-industrial stakeholders to make optimal decision by processing raw information and the reducing the redundant one to assist computationally efficient communication. In the last few years, technologies like Internet, World Wide Web (WWW) has broadened the horizon for online communication where users can make their independent views, commends, feedback etc on certain web platforms. The reviews can be made by significantly large number of users where each participant can have its own perception, review and/or opinion. Under such gigantically huge data presence identifying the concluding statement turns out to be of great significance. However, obtaining the summarized inference (of opinion) from a gigantically huge or multiple independent reviews is highly tedious task. On contrary, absorbing the gigantically huge data and creating the opinionated information is of utmost significance. This overall process is referred as Text Summarization. In the contemporary web based applications such as social media platforms, E-commerce platforms, internet-blog spots, review platforms have exploiting independent opinion(s) of the different users (say, multiple users) and identifying a single inference or fluent opinion by text summarization is a complex task.

Numerous researches done towards text-summarization are focused for either single text document summarization or extractive measure based summarization. Noticeably, text summarization is broadly classified as extractive and abstractive methods. In extractive text summarization, sentences or phrases from the input text are included for summary formation [1, 2]. The key limitation of such approaches is that it undergoes unavoidable inclusion of redundant information that makes computation too complex and cost-consuming. Moreover, the primitive intend of summarization can't be achieved efficiently by means of extractive summarization method. Extractive method selects salient sentences from the input data or source text document without making any modification, so as to generate the summary as output. On contrary, abstractive summarization techniques can generate (summary) text beyond the original one without including texts from the input data. In other words, abstractive methods can generate more concise and coherent summaries [3-7]. Typically, abstractive text summarization method generates a short summary comprising a few sentences that captures the salient ideas of an article or a passage (say, opinion). Abstraction text summarization methods are often applied for sentence compression, syntactic reorganization and lexical paraphrasing purposes. However, based on application environment and source of input, summarization is categorized as single-document or multiple-document summarization. In practice, the information overlap amongst the multiple documents pertaining to the same topic or aspect makes the multi-document summarization highly complicate and more challenging task in comparison to the single document's summarization. Furthermore, in multiple-document summarization the input or the source documents often comprise similar information, the extractive methods might generate biased or the redundant summary and hence can't be an effective solution [8]. Stating aforesaid application environments which embody multiple source document or opinion to be transferred into single coherent one, the development of a multiple-documents summarization method can be vital [9]. It can be considered as the motivation behind the at hand study, where the emphasis is made on developing a robust and efficient multiple-document text summarization model. Though, abstractive text summarization methods have better significance towards summarization, it requires efficient and robust natural language processing (NLP) technique to generate the optimally coherent and concise summary [10]. Though, a number of researches have been done towards text summarization, the majority of the conventional efforts either employ extractive summarization or are primarily developed for single document compression or summarization. Considering multiple documents summarization, absorbing the different distinct or independent documents and their aspects, and generating a new "abstracted" paragraph is a tedious task, which requires better NLP solutions. To achieve it, though a few efforts like deep learning based methods have been proposed that maps the input data sequence into another output sequence. This process is terms as "sequence-to-Sequence (S2S)" method. In the last few years S2S has gained widespread attention across industries to solve problems like machine translation [11], speech recognition [12] and video captioning [13].

Unlike major conventional S2S based summarization models, in this paper we have implemented Modified Attention Layer based Recurrent Conventional Neural Network (R-CNN). Here, R-CNN model encompasses encoder and decoder models as recommended by [11], which have performed better for different machine translation (MT) purposes. However, realizing the fact that MT is different from the abstractive summarization, where abstractive summarization intends to generate very short summary without depending on the size of input texts. Unlike conventional RNN or R-CNN based encoder-decoder techniques for NLP generation (NLG), in this paper the focus is made on enhancing it by incorporating the key features such as POS tagging, semantic feature extraction, dependency parsing, semantic role labeling, semantic information based summarization. To augment the coherence and conciseness of the summarized "abstracted" text output, our proposed Improved Attention based R-CNN (IA-RCNN) model exploits the structural, syntactic, and semantic information of the input text data. Since, not much significant effort is made so far to assess whether the semantic information or allied syntactic features can optimize encoder-decoder performance, in this paper we have exploited these features to enhance overall RNN encoder-decoder for NLG purposes. Our proposed model can be stated as an enhancement of Attention Based Summarization (ABS) model proposed by [14]. Unlike conventional approaches our proposed model can be stated as an enhancement of Attention Based Summarization (ABS) model proposed by [14]. Unlike conventional approaches our proposed IA-RCNN model encodes the results retrieved from Transition based Abstract Mean Representation (TAMR) parser by attention based Improved R-CNN model with Long and Short Term Memory (LSTM) encoder. Our proposed Tree-LSTM model encompasses augmented information than the classical one [15] that make NLG more efficient, accurate and concise.

2. Related Work

This section discusses some of the key literatures pertaining to multiple text summarizations.

Considering the significance of minimal redundancy and higher relevance, authors [16, 17] applied extractive text summarization concept. Authors [18, 19] formulated text summarization as the maximum coverage problem with knapsack constraint (MCKP). Additionally, it has been formulated as the problem of sub modular function maximization [20-22] where authors recommended using greedy algorithms to achieve concise and coherent summary as output. Authors [1] at first made effort to retrieve the frequent product features based on which the summary extraction method was derived from opinion sentences and allied features. However, these conventional approaches, which are of extractive text summarization type are unable to avoid any significant inclusion of less-significant information or intends to perform eventual summarization task [17, 23-29]. Unlike extractive method, abstractive text summarization methods, especially with multiple documents and generates the final summary by understanding intend of each document or content and rewrites the output as the most relevant one. Since the beginning, numerous efforts have been made towards abstractive text summarization, such as sentence compression [30] and sentence fusion [9]. Recently, authors recommending combination of both sentence compression as well as extractive system to enhance summary output [31, 32]. Similarly, efforts were made by employing sentence fragment identification and fusion to enable important text based summarization [9, 3, 33-36]. However, it was an extractive summarization method having less efficiency as compared to the abstractive summarization approach in real- world.

Undeniably, abstractive text summarization is a tedious task that becomes even more severe in case of multiple documents. The inclusion of the sophisticated techniques such as meaning representation, content mining and organization, sentence compression, content or sentence (fragment) fusion, paraphrasing etc makes abstractive text summarization more challenging task [37-39]. Recently, authors recommended applying compressive summarization approach that in function compresses the original sentence without making any significant alteration except word-deletion. Authors [40] proposed sentence compression comprising multiple sentences, often called Multi-Sentence Compression (MSC); however it primarily depended on the syntactic parsing to constitute a dependency tree for every associated sentence in a cluster. It eventually enabled grammatical compression to perform summarization [40]. Though, a few researches intended to use syntactic parsing for text summarization, its unavailability for all languages limited its application. To alleviate such issues, authors [39] proposed graph based method as alternative that employed merely a POS tagger and list of stop words to perform summarization. In graph based methods, a directed graph was constituted where each node signifies the words and edge states the adjacency between the words in the complete sentence. In this manner, summarized sentences are generated by finding the k-shortest path in the graph. Though, this approach was a better efforts, its limitations were resolved by [41], who focused on enhancing summarization by re-ranking the fusion candidate paths as per the important phrases to generate more informative summary.

In the last few years, the exponentially rise in deep learning techniques and its robustness for NLP has broadened the horizon for academia-industries to exploit its efficacy for text summarization. Additionally, the concept of encoder-decoder neural network for end-to-end training has gained wide-spread attention across academia-industries to perform abstractive text summarization. Recently, authors [12, 46] proposed attention based encoder-decoder neural network which was primarily inherited from the field of machine translation to perform text summarization. In addition, neural sequence-to-sequence (S2S) learning concept too has gained widespread attention for the headline generation from single document [14]. However, its efficacy could not be assessed for multiple document summarizations. Since the initial efforts made by [14] achieved single sentence summarization, was later considered as text summarization by numerous authors [47-51]. However, generating a single sentence can't be universal goal of text summarization, especially with multiple text documents with large corpse size. The above stated encoder-decoder neural network models could achieve single sentence generation comprising maximum of 75 characters. It confines its applicability in major application environment. Exploring in depth, it can be found that the above stated methods generated summarized sentence by comprising grammaticality of the original sentence. Recently, authors [52] applied CNN/DailyMail corpus as a supervised training data to generate multi-sentence summary from a single document [53-57]. However, majority of these approaches generate compressed summary by deleting the words form a single source document and doesn't employ any paraphrasing concept. These approaches could not generate the new sentence form the source document words except certain morphological changes. In the last few years, authors [58, 59] identified neural network method as viable solution for the text summarization in multiple-document setting. However, [58] could perform extractive summarization only.

On contrary, authors [59] could perform compressive summary generation by means of using an ILP approach and hence could not address the issue of redundancy in the summary generated. In addition, it could not be assessed for large scale multiple-document setting text summarization. Considering the robustness of deep learning methods for

NLP purposes, [60] reframed it as a data-driven approach to perform abstractive summarization. Recently, authors [14] applied convolutional technique to encode the source and context-sensitive attention feed-forward neural network to produce the summary. As optimization of [14] work, [48] and [47] applied convolutional model as encoder, but substituted the decoder by means of RNN, which was found more efficient for summary generation. A similar effort was made by [61] who employed above stated RNN encoder decoder model for text summarization in Chinese short documents. However, this approach was found limited with English corpora. Furthermore, authors [62] applied RNN based encoder-decoder; however for extractive summarization purpose. Recently, authors found that in addition to the conventional RNN encoder-decoder based approaches, the inclusion of sequence to sequence paradigm can be vital for more cohesive and concise summarization [33][14][52]. Researchers have revealed that the strategic inclusion of attention layer based RNN with enhanced encoding and decoding and S2S paradigm can be more effective solution towards abstractive text summarization. Additionally, the inclusion of semantic features too can be helpful towards multiple text document summarizations with above stated attention based RNN encoding-decoding concept. Considering it as motivation, in this research paper the focus is made on augmenting conventional attention model to be used in conjunction with RCNN based efficient encoding-decoding for abstractive summarization in multiple-document settings.

3. Research Questions

Considering overall research intends and allied implementation scope, we have defined a few key research questions, which intends to assess whether the proposed model or approach can achieve targeted goals. The research questions identified are given as follows:

RQ1: *Can the use of Improved Attention Layer assisted Recurrent Conventional Neural Network (IA-RCNN) be efficient for optimal encoding-decoding for further multiple-text summarization?*

RQ2: *Can the use of semantic and syntactic information be effective to support IA-RCNN encoder-decoder for more cohesive and concise Abstractive Multiple-text Summarization?*

RQ3: *Can the use of Word2Vec semantic feature extraction followed by IA-RCNN encoding decoding be a potential solution for S2S multiple-text abstractive summarization?*

RQ4: *Can the strategic implementation of Attention Based Summarization (ABS) and TAMR with improved Tree-LSTM encoding be effective for Abstractive Multiple-text Summarization?*

4. Our Contribution

This section primarily discusses the proposed multiple-document abstractive summarization model. As contribution in this research a novel and enhanced attention layer based RCNN model is developed, which has been implemented to generate text from multiple distinct input documents. Being an abstractive text summarization model, it intends to generate a new sentence without inheriting or making morphological changes in the text inputs. To achieve it, we have applied attention based model. Unlike conventional attention based strategy , in this research we have made enhancement so as to achieve computationally efficient and semantic also called Latent-Attention (LA) model to perform abstractive text summarization in multiple-document setup. Before discussing our proposed attention based summarization (ABS), a snippet of the attention model and its implementation is given as follows:

A. Attention based Deep Learning: Selecting the best Configuration

1. Global Vs Local Attention

In case of Global attention based neural network, all-encoder hidden states are used to characterize the attention enabled context vector to be used for each decoder, distinctly. Since, this process can be computationally complex and cost consuming; we have employed Local-Attention (LoCA) model which employs or considers merely a few hidden states which can fit into relatively smaller window. Noticeably, the window is often centered near the n-th encoder hidden state, where there can be M hidden states. In this LA configuration, the length of the window equivalent to the total number of hidden states and hence it would be $2M+1$. Such alignments can be divided into two broad types, monotonic and predictive, where in monotonic alignment the encoder hidden state n is maintained similar as the decoder position. For example, the fourth output would be n=4. In case M=2, then the attention would be merely on 3, 4, 5, 6 and 7 hidden states. On the other hand, the predictive alignment assures maintaining n as the function of the decoder state where the associated parameters h_t such as weights, etc are learnt jointly by the model. Considering the efficacy of the LoCA model with have consider it with predictive alignment scheme.

2. Hard Attention Vs Soft Attention

Typically, in Soft-Attention (SA) model, the context vector is estimated as a weight sum of the encoder hidden state. On contrary, in case of Hard-Attention (HA) instead of the weight sum and weight average the attention scores ar e considered to select a single hidden state. In fact selection of single hidden state is a complex task as in general the function named $argmax$ is applied to select the hidden state; however it functions by selecting an index pertaining to the maximum score. In such case, pushing the weight to move the score near the maximum value would not make any significant changes in index selection. In this paper we have applied hard-attention model, where the attention score is applied as the likelihood of the i-th location or text getting selected. Here, we applied a sophisticated argmax function to make the text selection. Mathematically, the hot-attention based word selection employs the following conditions (1).

$$Z_t = \sum_i s_{t,i} a_i \quad (1)$$

$$p(s_{t,i} = 1 \mid s_{j<t}, a) = \alpha_{t,i}$$

$$s_t^n \sim MultinoulliL(\{\alpha_i^n\})$$

In above equation a_i states the encoder or the input hidden state, while the parameter α states the attention score. Here, the variable $s_{t,i}$ would be on-hot variable with "1", when the i-th location is to be selected. In our paper, we have applied HA model to perform abstractive text summarization, and hence in this setup the feature vectors generated by the RCNN are the "Encoder-Hidden States".

3. Latent-Attention (LA)

As indicated in the previous sections, we intend to exploit efficacy of the latent information or the semantic features to make text summarization decision, the selection of latent-attention model can be of utmost significance. In our proposed model, we have at fist obtained the "Latent-Attention Vector (LAV)", where each dimension of the vector signifies a word and the softmax function provides a sense of relative significance across the words in the vector (obtained from the text input data). Noticeably, in this approach, each words of the input document is presented in terms of d-dimensional embedding vector and therefore for the total of J -words we retrieve $J \times D$ matrix, which is further used for training and allied processing. The use of Latent Attention (LA) model enables our approach to consider semantic features for sentence re-generation for the abstractive text summarization. Thus, observing above stated discussion and refined structural discussion, in our proposed Attention based Abstractive Text Summarization (ABAS) model we inherit the attention model with Local-Attention (LoCA), Hard-Attention (HA) and Latent Attention (LA). In this manner the attention model being employed can be defined as the function (2).

$$ABAS = f(LoCA, HA, LA) \quad (2)$$

In ABAS model, it predicts the word sequence (summary) on the basis of the NN concept in conjunction with an input sentence encoder. The detailed discussion of the ABAS model employed is given in the sub-sequent sections. Let, V be a text term or vocabulary, while x_i signifies the i-th indicator vector pertaining to the i-th word in the input document (say, input sentence). Consider that there are M words in a single input sentence of the document A. Similarly, X be the input sentence, which is signified as a sequence of indicator vectors of the length M. Mathematically,

$$x_i \in \{0,1\}^{|V|}, \text{ and} \quad (3)$$

$$X = (x_1, \dots, x_M).$$

Now, consider Y be a sequence of indicator vectors $Y = (y_1, \dots, y_L)$, whose length is L, provided $L < M$. Thus, the list of vectors comprising the sub-sequence in Y from y_{i-C+1} to y_i be the term $Y_{C,i}$. In our proposed model, we hypothesize a one-hot vector for an specific start symbol, for example "⟨S⟩", when $i < 1$. In this case, with the above stated attention model or ABAS model, the summary output $\hat{Y}$ for an input sentence X can be defined as (4).

$$\hat{Y} = \arg\max_{Y} \{\log p(Y|X)\} \quad (4)$$

$$\log p(Y|X) \approx \sum_{i=0}^{L-1} \log p(y_{i+1}|X, Y_{C,i}), \quad (5)$$

$$p(y_{i+1}|X, Y_{C,i}) \propto \exp\left(nnlm(Y_{C,i}) + enc(X, Y_{C,i})\right) \quad (6)$$

In above equation (6), the component $\mathrm{nnlm}(Y_{C,i})$ states a feed-forward NN model [63], while the second component of (6) $enc(X, Y_{C,i})$ represents the input sentence encoder with attention mechanism. In our proposed model, we employ D states the sizes (or, the dimensions) of vectors for word embedding, while H signifies the hidden layer states. Now, Let $E \in R^{D\times|V|}$ signifies the embedding matrix of the output words and consider that the parameter $U \in R^{H\times(CD)}$ and $O \in R^{|V|\times H}$ be the weight matrices of hidden and output layers, correspondingly. With above derived model, $nnlm(Y_{C,i})$ of (6) can be obtained as (7).

$$nnlm(Y_{C,i}) = Oh, \qquad h = tanh(U\tilde{y}_c) \quad (7)$$

In (7), the parameter $\tilde{y}_c$ states a concatenation of output embedding vectors from $i - C + 1$ to i. In other words, $\tilde{y}_c = (Ey_{i-C+1} \dots Ey_i)$, where it signifies a (CD) dimensional vector. Subsequently, let $F \in R^{D\times|V|}$ be embedding matrix of the input while the same for output sentence be $E' \in R^{D\times|V|}$. Additionally, $O' \in R^{|V|\times D}$ be the weight matrix for the output layer of CNN and the weight matrix to perform mapping of the embedding of C output onto the input words be $P \in R^{D\times(CD)}$. Similarly, $\tilde{X}$ be the matrix constructed form of a list of input embeddings, mathematically defined as $\tilde{X} = [\tilde{x}_1, \dots, \tilde{x}_M]$, where $\tilde{x}_i = Fx_i$. Then, the encoding function (3) $enc(X, Y_{C,i})$ can be reframed as (8).

$$enc(X, Y_{C,i}) = O'\bar{X}p \quad (8)$$

$$p \propto \exp(\tilde{X}^T P \tilde{y}_c') \quad (9)$$

In above equations, parameter $\tilde{y}_c'$ stats a concatenation of output embedding vectors from $i - C + 1$ to i, which is equivalent to the $\tilde{y}_c$. In other words,

$$\tilde{y}_c' = (E'y_{i-C+1} \dots E'y_i) \quad (10)$$

Additionally, the parameter $\bar{X}$ signifies a matrix depicting the list of mean input word embeddings within the predefined window size Q. Mathematically,

$$\bar{X} = [\bar{x}_1, \dots, \bar{x}_M] \quad (11)$$

In (11), the parameter $\bar{x}_i = \sum_{q=i-Q}^{i+Q} \frac{1}{Q} \tilde{x}_q$.

In ABAS model, equation (9) is often stated as "Attention Model (AM)" that in function encodes the relationship in between the input words and the descendent or previous output word C. For illustration, in case the previous output (i.e., C) words are hypothesized to align towards x_i, then the neighboring allied Q words $(x_{i-Q}, \dots, x_{i+Q})$ would have high weight as per (8). The encoding efficiency and suitability towards latent relationship based encoding enables AM to be used for Abstractive Text Summarization. With this motive, in this research we have applied the augmented AM, named ABAS (2) to perform abstractive text summarization in multiple document setting. Noticeably, since in this research we have applied Improved Attention Model (IA) based CNN model to be used for Abstractive Text Summarization, here onwards we pronounce it as IA-RCNN. To be noted, as NN solution, in our proposed model we have applied Recurrent Convolutional Neural Network, which is well known for its computational efficiency and accuracy for learning and eventual new text generation. The detailed discussion of the proposed IA-RCNN model for Abstractive Text Summarization is presented in the sub-sequent section.

B. ABAS Implementation

As already stated in previous section, in this research to perform abstractive text summarization we hypothesize that the well defined and strategic implementation of the syntactic and semantic characteristics of the text input or sentence can assist generating the new sentence, called summary. For instance, the key significance, subject matter, allied meanings, predicates, and objects of the generated sentence must coincide to the original (input) sentence(s). Considering this fact, in this research we have applied both semantic as well as syntactic features to generate new text, where the use of Latent Attention (LA) model has played decisive role to achieve concise and relevant sentence. Though, attention based approaches have been extensively applied for Abstract Meaning Representation; however its use of multiple inputs text summarization has not yet investigated. In majority of the classical researches, authors have directly applied input sentence for encoding. On contrary, to retain above stated syntactic as well as semantic features for text summarization (in multiple input document set up), we have at first obtained Abstract Meaning Representation, which has been fed as input to the RCNN model. The detailed discussion of the proposed Abstract Meaning Representation scheme applied in this paper is given as follows:

1. Phase-1 Abstract Meaning Representation

Typically, the concept of Abstract Meaning Representation signifies a directed and acyclic graph model which performs encoding of the signifier or the sentence meaning. In other words, this concept encodes the meaning of a sentence. Noticeably, in functional structure each node in abstract meaning representation scheme signifies the term called 'concepts', while the directed edges refer the association or the relationship in between the nodes. Here, concepts comprise English words (as we have considered input sentences in English language), while the property represents the predicates. Following the concepts proposed by [64], for edges, there used to be nearly 100 relations comprising the semantic roles on the basis of above stated property (i.e., predicators) annotations. Here, in our proposed model we applied Word2Vec concept to achieve Latent information and relationships between the words. In our proposed model, to obtain precise Abstract meaning representation of each sentences (as, we have applied multiple inputs), we have applied sequence-to-sequence (S2S) paradigm assisted Transition-based Abstract Meaning Representation (TAMR) parser. The detailed discussion of the TAMR model can be found in [65]. Now, once obtaining the TAMR values for each text inputs, the obtained values are projected to the Attention based Encoder. The detailed discussion of the proposed attention based TAMR encoding is given as follows:

2. Phase-2 Attention Assisted TAMR Encoding

Once obtaining the semantic and syntactic information of features from TAMR model, it was fed as input to a layered optimized Child-Sum Tree-LSTM based RNN model. Undeniably, majority of the classical RNN methods are time efficient and accurate; however their robustness with multiple input processing requires an optimal balance in between the computation and encoding accuracy. Considering it as motive, in this research we augmented classical RNN with lower layer selection and hidden state information parsing. In our proposed Tree based LSTM leaning model, the obtained semantic and syntactic features are encoded which are converted into a predefined Feature Embedding Vectors (FEV). To achieve computationally efficient learning and embedding, in our proposed model, we converted a Directed Acyclic Graph (DAG) model of the TAMR parser output into an equivalent Tree-structure. To achieve it, we applied the concept of Head-Nodes (also called Parent node) separation, which usually comes into existence to represent co- referential concepts, to associated or respective out-edges to the head nodes. In our proposed model, the Tree-LSTM has been further augmented to encode edge labels as TAMR facilitates both node as well as edge labels, on contrary the classical Tree-LSTM method merely encodes node labels. This enhancement enables our proposed method to achieve or retain more significant information which eventually attains more concise and cohesive summarization.

Consider, the variable n_j and e_j be the N and E dimensional embeddings for the different labels. Let these labels are assigned to the j-th node of the RNN model. Similarly, out-edge are directed to associated root node, also called parent node. Now, to perform embedding we use different weight parameters such as $W_{in}, W_{fn}, W_{on}, W_{un} \in R^{D\times N}$. Noticeably, these weight matrices are employed to perform node embeddings for n_j. Now, consider that the weights used for edge-embedding e_j be the $W_{ie}, W_{fe}, W_{oe}, W_{ue} \in R^{D\times E}$. Similarly, the weight matrices employed for output vectors connected to the child-node be the $W_{ih}, W_{fh}, W_{oh}, W_{uh} \in R^{D\times D}$ and let the total nodes or the set of nodes available in all documents be $B(j)$. Noticeably, $B(j)$ are those nodes which possess a direct edge to the j-th node in the TMAR structure. In this case, the embedding output $a_j$at node j in TAMR structure is obtained by processing our proposed Tree-LSTM model. Mathematically, it is derived as follows:

$$\tilde{h}_j = \sum_{k\in B(j)} a_k, \tag{12}$$

$$i_j = \sigma\left(W_{in}n_j + W_{ie}e_j + W_{ih}\tilde{h}_j\right) \tag{13}$$

$$f_{jk} = \sigma\left(W_{fn}n_j + W_{fe}e_j + W_{fh}a_k\right) \tag{14}$$

$$o_j = \sigma\left(W_{on}n_j + W_{oe}e_j + W_{oh}\tilde{h}_j\right) \tag{15}$$

$$u_j = tanh\left(W_{un}n_j + W_{ue}e_j + W_{uh}\tilde{h}_j\right) \tag{16}$$

$$c_j = i_j \odot u_j \sum_{k\in B(j)} f_{jk} \odot c_k \tag{17}$$

$$a_j = o_j \odot tanh\,(c_j) \tag{18}$$

Consider, J be the total nodes available in TAMR structure retrieved from the multiple documents or the provided input sentences. Then, here we have defined a matrix $A \in R^{D\times J}$ signifying the list of the hidden states a_j for all comprising nodes j. In other words let the matrix be $A = [a_1, \dots, a_J]$ and consider the weight matrix of the output layer be the $O'' \in R^{|V|\times D}$. Similarly, $S \in R^{D\times(CD)}$ signifies the weight matrix to map the context embedding of C output words onto embeddings retrieved from the nodes. In such case in our proposed model we have derived or designed an Attention assisted TAMR encoder $encTAMR(A, Y_{C,i})$. Mathematically, our proposed encoder model is defined as (19).

$$encTAMR(A, Y_{C,i}) = O''As \tag{19}$$

$$s \propto \exp(A^T S \tilde{y}'_c) \tag{20}$$

Once deriving the attention model (19), we have amalgamated it to the native model (6). Thus, the newly derived attention model for ABAS based text summarization is derived as (21).

$$p\left(y_{i+1}|X, Y_{C,i}\right) \propto \exp\left(nnlm\left(Y_{C,i}\right) + enc\left(X, Y_{C,i}\right)\right) + encTAMR(A, Y_{C,i}) \tag{21}$$

Thus, using (21), we have performed new sentence generation. Considering computational efficacy requirements, though RNN deep learning concept has always been the dominant solution, however dealing with multiple data and processing above stated mechanisms might force it to undergo computationally overburdened. To address such possible issue, unlike classical deep learning methods, we modified Tree-LSTM RCNN by changing layer structure as well as learning method. Here, we applied ADAM (Adaptive Moment Learning), an adaptive moment estimator model which assigns learning weights dynamically to perform learning or training. Unlike classical Stochastic Gradient Descent (SGD) based learning methods, which is applied in most of the deep learning method, our approach might yield more computationally efficient performance. Initially, we assigned initial learning rate as 0.0001. The implemented deep learning structure embodied two convolutional layers (CONV), Max-pooling and two Fully Connected (FC) layers, where sigmoid function was used as activation function. In addition, we applied dropout rate of 0.5 (i.e., 50% dropout), which can enhance overall computational efficiency even with large input datasets.

3. Phase-3 ABAS for Multi-Document Setup

Noticeably, in our proposed model, to perform abstractive text summarization in multiple input sentence setup (or multiple document setup), before feeding the inputs to the Tree-LSTM encoder we have processed for Sentence clustering [66], which was further processed for TAMR structure generation and RNN learning or training. In our proposed sentence clustering model we applied syntactic or Hierarchical Agglomerative Clustering (HAC) concept proposed in [67] with linking criteria. This method exploited both S2S concept as well as incremental mechanism originating with each sentence (called a cluster), and was merged the pair of each cluster iteratively after each step. To achieve it, we applied Bottom-Up concept. In our proposed model, the above stated Linking Criteria estimated the metrics used for merging. In other words it obtained the maximum distance in between a cluster (i.e., sentence) to the other. Here, we applied cosine-similarity amongst the embedding vectors obtained by Word2Vec (sentence embedding vector outputs). Assigning a threshold of 0.5, we obtained the single sentence which was processed further with TAMR and sub-sequent Tree-LSTM learning to generate the new sentence. Noticeably, though, in our proposed model the new cluster or new sentence obtained was small, however it was highly coherent as all participating sentences that a cluster contained was very similar to each other (in same cluster) and hence applying TAMR followed by Tree-LSTM RNN encoding resulted better performance. To be noted, since our proposed model at first performed TAMR, whose output was processed for encoding, even the concept of "sentence merging" [68] can also be considered to perform abstractive text summarization under multiple-documents settings.

5. Results and Discussion

This section primarily discusses the results and discussion for the at hand research.

- *Data Preparation:*

To assess the performance of the proposed Abstractive Text Summarization model, we have considered multiple documents with similar aspects. Being a multiple document text summarization model, we have ensured that the input data retains or possess similar aspect. For which we have applied two different set of input data, one from Amazon review, while another as benchmark dataset DUC 2004. In our proposed model, before processing for text summarization, at first we have processed for pre-processing where the punctuations, special characters or stopping words such as ",", ".", "!", etc are removed. To achieve it we applied Porter's Stemmer (Porter, 1999) method. In addition, the small words such as "is", "was", "have", "had" etc were removed. Furthermore, we performed contraction mapping and tokenization as supplementary pre-processing that alleviated any presence non-uniformity and achieves better computing environment for "Dynamic Programming Problem". To achieve document co-reference resolution, we applied Stanford CoreNLP. Executing for pre-processing, the sentences from each input (documents) were converted into sentence sequences. In case of Amazon review data, we considered 200 rows signifying each sentence as review. On contrary, the other dataset was applied in its native form, each with different sentences having similar aspects. Additionally as pro-processing we converted each sentence or allied words as lower case.

- *Experiment*

As stated in above section, we considered two distinct datasets (Amazon Review and DUC, 2004) to assess efficiency of the proposed abstractive text summarization model under multiple document settings, Specifically, in proposed system the set of key attributes presents in the summary containing content coverage of summaries and linguistic quality as well as cohesive factuality of the generated summary sentences have been examined. We applied ROGUE-1 (Lin., 2004) to examine the content coverage. Noticeably, ROGUE refers the ratio of the total number of words in the predicted summary to the total number of words in the target summary. Additionally, we have performed human evaluation and assessment for linguistic quality of the newly generated sentence. In our experiment we considered DUC 2004, which embodies numerous corpora, each consisting of multiple documents. Here, for sake of simplicity for DUC 2004, we considered 2 input samples to generate the new one. Furthermore, we tuned the development parameters, especially TAMR and Tree-LSTM as per DUC 2004 with 2 distinct inputs (sentences). Similarly, for Amazon data we modified processing parameters such as number of latent information as 300, while embedding size was considered as 100. In our proposed model, Tree-LSTM parameters were tuned to achieve more efficient content coverage where pruned trees were obtained to fill the allotted summary spaces without involving any additional combination to refine final new sentence (as generated output). We measured content coverage as ROGUE score with reference to the summaries obtained for distinct inputs or datasets. In this paper, we assess statistical performance for our proposed system in terms of ROGUE-Recall and ROGUE F-Measure.

Table-1 ROGUE-1 Scores of the proposed abstractive summarization model

Dataset	R-1	F-1
DUC 2004	0.46	0.59
Amazon-Review	0.62	0.70

To examine relative efficiency of our proposed model with respect to the state-of-art existing method, such as [68] we have compared the performance in terms of ROGUE Recall (R-1) and F-Measure (F-1) (Table 2). To make comparative assessment, we have averaged the scores obtained for the two different datasets.

Table-2 Relative Performance assessment (ROGUE-1)

Dataset	R-1	F-1
[18]	0.385	-
[20]	0.39	0.38
[68]	0.38	0.38
[69]	0.39	-
Proposed	0.40	0.53

Considering above stated performance, it can easily be found that the proposed text summarization model achieves higher performance as compared to existing state-of-art abstractive (multi-document) summarization models. Considering a recent work [68], where author have applied Partial Tree-Extraction, Recombination and Linearization concept to perform abstractive text summarization in multiple document setting. In their approach, though the use of Tree-LSTM was more refined and tuned, our proposed model outperforms it [68] in terms of higher recall and F-Measure performance. In addition to the statistical outcome and allied assessment, we have examined the performance qualitatively by understanding original inputs as well as summary outputs. Results reveal that the proposed model achieves relevance and coverage, without imposing any redundant information in resulting output. The overall research inferences and future scopes are discussed in the sub-sequent section.

- *Discussion:*

Considering the overall results and allied inferences, it can be found that though a few efforts have been made by employing attention model to perform summarization, they lacked tuning their model as per input requirements. Inputting the texts document as direct input to the encoder could not exploit the semantic and syntactic features that as a result could have achieved better solution. On contrary, in our proposed model at one hand the RNN model was tuned and enhanced to support multiple input processing, on the other hand before performing word embedding we performed TAMR parsing that provided both semantic as well as syntactic information to the encoder. This can be the prime contributing factor which strengthened our model to exhibit better performance. In addition, unlike classical attention model, in this research a refined attention model with Hard-Attention, Local Attention and Latent Attention feature was applied. This as a result enhanced attention concept to yield improved performance for abstractive text summarization in multiple document setting. It affirms acceptability of the research question *RQ1*. As already stated the inclusion of TAMR enabled inheriting or accommodating both semantic as well as syntactic information (on contrary in major existing works, authors have applied graph concept or tree concept, as syntactic feature) has enhanced abstractive summarization. This as a result has achieved better (cohesive, concise and redundant free) summarization. Our proposed model generated a new sentence with low word count; however it was highly relevant, concise and of course coherent that signify its efficacy to perform under multiple large scale input-settings. It affirms acceptability of the research question *RQ2*. In our proposed model, we applied word2vec as tool to obtain the semantic features to be further encoded. Noticeably, TAMR too generated embedding output equivalent to the word2vec in multiple dimensions. The resulting output was later processes by Tree-LSTM RCNN model for summary generation. In this manner, we achieved better performance. Thus, the question RQ3 too is justified affirmatively. Considering overall functional components such as TAMR, ABAS, IA-RCNN or Tree-LSTM model, the conclusive results and allied inferences confirms that the proposed model achieves optimal performance to meet intended goal. Thus, it confirms acceptability of the RQ4, affirming that the strategic implementation of ABAS, TAMR with improved Tree-LSTM encoding can be effective for Abstractive text Summarization under multiple document setting.

6. Conclusions

In this paper, a highly robust and efficient multiple-document abstractive text summarization model is developed. Unlike classical abstractive text summarization schemes, the proposed method intended to exploit efficacy of the different enhanced NLP techniques including sequence to sequence paradigm, Improved Attention Layer assisted Recurrent Convolutional Neural Network (IA-RCNN) model for efficient encoding-decoding. Noticeably, the inclusion of AMR approach with IA-RCNN strengthened overall proposed summarization model to achieve concise and cohesive summarization without losing grammaticality of the original documents. Here, we considered semantic feature extraction, dependency parsing, semantic role labeling, semantic information etc which eventually enabled provision for the structural, syntactic, and semantic information of the input original datasets to generate the summary. Similarly, the use of Tree-LSTM augmented overall encoding process which retained aspect information and intend of overall input text. The strategic implementation of Sentence Merging and clustering, Transition based AMR (TAMR), whose output was fed as input for tree-LSTM RCNN made multiple text document summarizations more efficient. The proposed model was simulated with different set of input data, where manual quality assessment as well as statistical performance investigation revealed affirmative performance by the proposed system.

References

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’04, pages 168–177, New York, NY, USA. ACM.

[2] Lerman et al.2009] Kevin Lerman, Sasha Blair- Goldensohn, and Ryan McDonald. 2009. Sentiment summarization: Evaluating and learning user preferences. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, EACL ’09, pages 514–522, Stroudsburg, PA, USA. Association for Computational Linguistics.

[3] Lidong Bing, Piji Li, Yi Liao, Wai Lam, Weiwei Guo, and Rebecca Passonneau. 2015. Abstractive multi-document summarization via phrase selection and merging. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 1587–1597, Beijing, China, July. Association for Computational Linguistics.

[4] Lu Wang and Claire Cardie. 2013. Domain-independent abstract generation for focused meeting summarization. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1395–1405, Sofia, Bulgaria, August. Association for Computational Linguistics.

[5] Kavita Ganesan, ChengXiang Zhai, and Jiawei Han. 2010. Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions. In Proceedings of the 23rd international conference on computational linguistics, pages 340–348. Association for Computational Linguistics.

[6] Giuseppe Di Fabbrizio, Amanda J Stent, and Robert Gaizauskas. 2014. A hybrid approach to multi-document summarization of opinions in reviews. INLG 2014, page 54.

[7] Shima Gerani, Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng, and Bita Nejat. 2014. Abstractive summarization of product reviews using discourse structure. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1602–1613, Doha, Qatar, October. Association for Computational Linguistics.

[8] Mir Tafseer Nayeem and Yllias Chali. 2017a. Extract with order for coherent multi-document summarization. In Proceedings of TextGraphs@ACL 2017: the 11th Workshop on Graph-based Methods for Natural Language Processing, Vancouver, Canada, August 3, 2017, pages 51–56.

[9] Regina Barzilay and Kathleen R. McKeown. 2005. Sentence fusion for multi-document news summarization. Comput. Linguist., 31(3):297–328, September.

[10] Yllias Chali, Moin Tanvee, and Mir Tafseer Nayeem. 2017. Towards abstractive multi-document summarization using submodular function-based framework, sentence compression and merging. In Proceedings of the Eighth International Joint Conference on Natural Language Processing, IJCNLP 2017, Taipei, Taiwan, November 27 - December 1, 2017, Volume 2: Short Papers, pages 418–424.

[11] Dzmitry Bahdanau, Kyunghyu Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. CoRR, abs/1409.0473.

[12] Dzmitry Bahdanau, Jan Chorowski, Dmitriy Serdyuk, Philemon Brakel, and Yoshua Bengio. 2015. End-to-end attentionbased large vocabulary speech recognition. CoRR, abs/1508.04395.

[13] Venugopalan et al.2015] Subhashini Venugopalan, Marcus Rohrbach, Jeff Donahue, Raymond J. Mooney, Trevor Darrell, and Kate Saenko. 2015. Sequence to sequence - video to text. CoRR, abs/1505.00487.

[14] Alexander M. Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. CoRR, abs/1509.00685.

[15] Kai Sheng Tai, Richard Socher, and Christopher D. Manning. 2015. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015), pages 1556– 1566.

[16] J. Carbonell and J. Goldstein. 1998. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, pages 335–336. ACM.

[17] G¨unes Erkan and Dragomir R. Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. J. Artif. Int. Res., 22(1):457–479, December.

[18] H. Takamura and M. Okumura. 2009. Text summarization model based on maximum coverage problem and its variant. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, pages 781–789. Association for Computational Linguistics.

[19] H. Morita, T. Sakai, and M. Okumura. 2011. Query snowball: a co-occurrence-based approach to multi-document summarization for question answering. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2, pages 223–229. Association for Computational Linguistics.

[20] H. Lin and J. Bilmes. 2011. A class of submodular functions for document summarization. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics.

[21] H. Lin and J. Bilmes. 2010. Multi-document summarization via budgeted maximization of sub-modular functions. In Proceedings of the 2010 Conference of the North American Chapter of the Association for Computational Linguistics, pages 912–920. Association for Computational Linguistics.

[22] A. Dasgupta, R. Kumar, and S. Ravi. 2013. Summarization through submodularity and dispersion. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics ACL, pages 1014–1022.

[23] Joel Larocca Neto, Alex Alves Freitas, and Celso A. A. Kaestner. 2002. Automatic text summarization using a machine learning approach. In Proceedings of the 16th Brazilian Symposium on Artificial Intelligence: Advances in Artificial Intelligence, pages 205–215.

[24] Kam-Fai Wong, Mingli Wu, and Wenjie Li. 2008a. Extractive summarization using supervised and semi-supervised learning. In Proceedings of the 22Nd International Conference on Computational Linguistics - Volume 1, pages 985– 992.

[25] Katja Filippova and Yasemin Altun. 2013. Overcoming the lack of parallel data in sentence compression. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1481–1491.

[26] Carlos A. Colmenares, Marina Litvak, Amin Mantrach, and Fabrizio Silvestri. 2015. Heads: Headline generation as sequence prediction using an abstract feature-rich space. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 133–142.

[27] M. Litvak and M. Last. 2008. Graph-based keyword extraction for single document summarization. In Coling 2008, pages 17–24.

[28] B. Favre K. Riedhammer and D. Hakkani-Tur. 2010. Long story short â˘A ¸S global unsupervised models for keyphrase based meeting summarization. In Speech Communication, pages 801–815.

[29] David Martins de Matos JoÃ˘co P. Neto Anatole Gershman Jaime Carbonell Ricardo Ribeiro, LuÃ s Marujo. 2013. Self reinforcement for important passage retrieval. In 36th international ACM SIGIR conference on Research and development in information retrieval, pages 845– 848.

[30] K. Knight and D. Marcu. 2000. Statistics-based summarization - step one: Sentence compression. In Proceedings of the 17th National Conference on Artificial Intelligence, Austin.

[31] T. Berg-Kirkpatrick, D. Gillick, and D. Klein. 2011. Jointly learning to extract and compress. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pages 481–490. Association for Computational Linguistics.

[32] A. Martins and N. A. Smith. 2009. Summarization with a joint model for sentence extraction and compression. In Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing, pages 1–9. Association for Computational Linguistics.

[33] K. Filippova, E. Alfonseca, C. A. Colmenares, L. Kaiser, and O. Vinyals. 2015. Sentence compression by deletion with lstms. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 360–368.

[34] K. Ganesan, C. Zhai, and J. Han. 2010. Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions. In Proceedings of the 23rd international conference on computational linguistics, pages 340–348. Association for Computational Linguistics.

[35] K. Thadani and K. McKeown. 2013. Supervised sentence fusion with single-stage inference. In Proceedings of the International Joint Conference on Natural Language Processing, pages 1410–1418.

[36] J. C. Cheung and G. Penn. 2014. Unsupervised sentence enhancement for automatic summarization. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 775–786. Association for Computational Linguistics.

[37] James Clarke and Mirella Lapata. 2006. Models for sentence compression: A comparison across domains, training requirements and evaluation measures. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics, ACL-44, pages 377–384, Stroudsburg, PA, USA. Association for Computational Linguistics.

[38] James Clarke and Mirella Lapata. 2008. Global inference for sentence compression: An integer linear programming approach. Journal of Artificial Intelligence Research, 31:399–429.

[39] Katja Filippova. Multi-Sentence Compression: Finding Shortest Paths in Word Graphs. In Proc. of the 23rd International Conference on Computational Linguistics (Coling 2010), pages 322–330, 2010.

[40] Katja Filippova and Michael Strube. 2008. Sentence fusion via dependency graph compression. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP ’08, pages 177–185, Stroudsburg, PA, USA. Association for Computational Linguistics.

[41] Florian Boudin and Emmanuel Morin. 2013. Keyphrase extraction for n-best reranking in multi-sentence compression. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 298–305, Atlanta, Georgia, June. Association for Computational Linguistics.

[42] Mir Tafseer Nayeem and Yllias Chali. 2017b. Paraphrastic fusion for abstractive multi-sentence compression generation. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017, pages 2223–2226.

[43] Siddhartha Banerjee, Prasenjit Mitra, and Kazunari Sugiyama. 2015. Multi-document abstractive summarization using ilp based multi-sentence compression. In Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI’15, pages 1208–1214. AAAI Press.

[44] Dung Tran Tuan, Nam Van Chi, and Minh-Quoc Nghiem, 2017. Multi-sentence Compression Usin Word Graph and Integer Linear Programming, pages 367–377. Springer International Publishing, Cham.

[45] Elaheh ShafieiBavani, Mohammad Ebrahimi, Raymond K. Wong, and Fang Chen. 2016. An efficient approach for multi-sentence compression. In Proceedings of The 8th Asian Conference on Machine Learning, volume 63of Proceedings of Machine Learning Research, pages 414–429, The University of Waikato, Hamilton, New Zealand, 16–18 Nov. PMLR.

[46] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In Empirical Methods in Natural Language Processing (EMNLP), pages 1412– 1421, Lisbon, Portugal, September. Association for Computational Linguistics.

[47] Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, C¸ a glar Gulc¸ehre, and Bing Xiang. 2016. Abstractive text summarization using sequence-to-sequence rnns and beyond. CoNLL 2016, page 280.

[48] Sumit Chopra, Michael Auli, and Alexander M. Rush. 2016. Abstractive sentence summarization with attentive recurrent neural networks. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 93–98, San Diego, California, June. Association for Computational Linguistics.

[49] Jun Suzuki and Masaaki Nagata. 2017. Cutting-off redundant repeating generations for neural abstractive summarization. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, pages 291–297, Valencia, Spain, April. Association for Computational Linguistics.

[50] Qingyu Zhou, Nan Yang, Furu Wei, and Ming Zhou. 2017. Selective encoding for abstractive sentence summarization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1095–1104, Vancouver, Canada, July Association for Computational Linguistics.

[51] Shuming Ma, Xu Sun, Jingjing Xu, HoufengWang,Wenjie Li, and Qi Su. 2017. Improving semantic relevance for sequence-to-sequence learning of chinese social media text summarization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 635–640, Vancouver, Canada, July. Association for Computational Linguistics.

[52] Karl Moritz Hermann, Tom´aˇs Koˇcisk´y, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, NIPS'15, pages 1693–1701, Cambridge, MA, USA. MIT Press.

[53] Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer generator networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1073–1083, Vancouver, Canada, July. Association for Computational Linguistics.

[54] Piji Li, Wai Lam, Lidong Bing, and Zihao Wang. 2017b. Deep recurrent generative decoder for abstractive text summarization. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing,pages 2091–2100, Copenhagen, Denmark, September. Association for Computational Linguistics.

[55] Romain Paulus, Caiming Xiong, and Richard Socher. 2017. A deep reinforced model for abstractive summarization. CoRR, abs/1705.04304.

[56] Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018a. Ranking Sentences for Extractive Summarization with Reinforcement Learning. In Proceedings of the NAACL 2018 - Conference of the North American Chapter of the Association for Computational Linguistics.

[57] Shashi Narayan, Nikos Papasarantopoulos, Shay B. Cohen, and Mirella Lapata. 2018b. Neural extractive summarization with side information. In Proceedings of the Thirty-Second AAA Conference on Artificial Intelligence, February.

[58] Michihiro Yasunaga, Rui Zhang, Kshitijh Meelu, Ayush Pareek, Krishnan Srinivasan, and Dragomir Radev. 2017. Graph-based neural multi-document summarization. In Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017), pages 452–462, Vancouver, Canada, August. Association for Computational Linguistics.

[59] Piji Li, Wai Lam, Lidong Bing, Weiwei Guo, and Hang Li. 2017a. Cascaded attention based unsupervised information distillation for compressive summarization. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 2081–2090, Copenhagen, Denmark, September. Association for Computational Linguistics.

[60] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel P. Kuksa. 2011. Natural language processing (almost) from scratch. CoRR, abs/1103.0398.

[61] Baotian Hu, Qingcai Chen, and Fangze Zhu. 2015. Lcsts: A large scale chinese short text summarization dataset. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1967–1972, Lisbon, Portugal, September. Association for Computational Linguistics.

[62] Jianpeng Cheng and Mirella Lapata. 2016. Neural summarization by extracting sentences and words. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.

[63] Yoshua Bengio, R´ejean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A Neural Probabilistic Language Model. The Journal of Machine Learning Research, 3:1137–1155.

[64] Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for Sembanking. In Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse, pages 178–186.

[65] Chuan Wang, Nianwen Xue, and Sameer Pradhan. 2015. A Transition-based Algorithm for AMR Parsing. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2015), pages 366–375.

[66] Mir Tafseer Nayeem, Tanvir Ahmed Fuad, Yllias Chali. Abstractive Unsupervised Multi-Document Summarization using Paraphrastic Sentence Fusion. Proceedings of the 27th International Conference on Computational Linguistics, pages 1191–120 Santa Fe, New Mexico, USA, August 20-26, 2018.

[67] Fionn Murtagh and Pierre Legendre. 2014. Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? J. Classif., 31(3):274–295, October.

[68] Yllias Chali, Moin Tanvee, Mir Tafseer Nayeem. Towards Abstractive Multi-Document Summarization using submodular Function-Based Framework, Sentence Compression and Merging. Proceedings of the The 8th International Joint Conference on Natural Language Processing, pages 418–424, Taipei, Taiwan, November 27 – December 1, 2017c 2017.

[69] D. Wang, S. Zhu, T. Li, and Y. Gong. 2009. Multi-document summarization using sentence based topic models. In Proceedings of the ACLIJCNLP 2009 Conference, pages 297–300. Association for Computational Linguistics.