Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HIERARCHICAL ATTENTION NETWORK -ATTENTION BASED BIDIRECTIONAL CNN-RNN DEEP MODEL BASED QUESTION ANSWERING SYSTEM USING ABSTRACTIVE SUMMARIZATION

T. PRIYANKA[1], A. MARY SOWJANYA[2]

[1]Research Scholar, Department of CS&SE Andhra University College of Engineering, Visakhapatnam Andhra Pradesh, India, India.:

[2]Assistant Professor, Department of CS&SE Andhra University College of Engineering, Visakhapatnam Andhra Pradesh, India, India.

Email: [1]priyadatta24986@gmail.com, [2]sowmaa@gmail.com

ABSTRACT

Summarization aids to minimize document size while processing the meaning amidst Natural Language Processing. Summarization models based on precise sentences are adapted as they occur in inventive text and new sentences are produced with NLP models which are classified as extractive and abstractive models respectively. The complication that lies with NLP text made the abstractive summarization a complex process. An effective model is developed for abstractive summarization with a question answering system using ensemble classifiers. Review data is considered for BERT tokenization, Aspect Term extraction (ATE)is utilized to obtain aspects from review data. The sentiment rating is predicted using ensemble classifiers and fused with the majority voting model. The abstractive summarization is done using HAN-ABCDM which involves with training and testing stages. In the training stage, HAN and ABCDM are trained with summarized text, question answer pairs and sentiment ratings obtained using the majority voting method. In testing phase, the trained HAN and ABCDM are adapted to retrieve predicted answers. Considering the question answering model, the proposed HAN-ABCDM QAS offered highest valued precision of 82.1%, recall of 95.7%, and F-measure of 93.9% compared to other existing methods.
Keywords: *Abstractive Summarization, Question Answering, HAN, ABCDM, Ensemble Classifier*

1. INTRODUCTION

Text summarization involves complex mechanisms in processing the natural languages. It aims to condense a quantity of text to a small volume that comprises all essential information contained in original data. Methods related to summarization of texts are categorized into two classes, namely extractive and abstractive. Extractive summarization is to identify the most vital sentences using text by adapting statistical features and then organizing these sentences for building a summary. On the other hand, abstractive summarization is a technique wherein a meaningful summary is generated by rephrasing or utilizing new words in line of mining important sentences. In view of high strides of deep models, neural-based abstractive summarization techniques are more capable of generating smooth and understandable summaries. Nevertheless, classical neural abstractive summarization techniques are extensively susceptible to inconsistency errors which bring severe issues to usage and reliability of abstractive summarization models [1]. Examining the quality of content in a summary is a basic process in summarizing the texts. It gained immense focus by researchers over the past few years and well-known models include reference-based measures that treated a human-written reference summary as major property and scored the candidate summary on the basis of content similarity with its reference [2]. Various text processing methods are devised for helping the network users to process the textual information in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

which volume is unmanageable. Text summarization can be adapted to answer queries involving huge text corpora. The queries have a preliminary group of data which is subsequently filtered by a summarizer. It helps to build concise summaries from the document fragments considering the query terms [3].

Techniques obtained from processing the natural language and computational linguistics has basically altered how one processes, distributes, and obtains data like questions answering models and search engines. However, there exists a severe issue that haunts several text processing applications that represent how to automate and precisely assess the application quality. In some other scenarios, computation of NLP itself has led to an imperative research domain like evaluation of text summarization. The main complexity to design such computation arrives from the variety of NLP techniques and our inadequate understanding of NLP and human intelligence [4]. Annotated quantity in discussion level plays an imperative role in NLP and performs such processes like summarization of texts, question-answering models and mining the knowledge. However, the kind of discourse annotation which is pertinent varies extensively and depends on the application. Thus, an empirical assessment is essential to discover the commonalities in the alterations of the discourse phenomenon and to design general-purpose techniques for discourse assessment. Automatic mining of causality knowledge offers people with illustrations in question-answering models is a complex process [5]. The emergence of the network and the growth of contemptible huge-capacity storage devices, one is generating a tidal wave of information. This scenario makes it difficult to identify and accumulate the data one really requires. Automated summarization of text is a key technique in overwhelming this complexity [6].

Question answering represents a particular domain in retrieval of information. There exist several Question Answering System (QAS) that have their own application domain. QAS has several applications on the basis of source of answers. Like mining data from documents, learning languages, online examination models and document management are done. Major goal of QAS is to obtain relevant answers of questions instead of retrieving the whole document. There are two kinds of QAS involving open and closed domains. Open domain models are devised on the basis of the web in which there is no limitation of any era whereas closed domain models have less work domain like medicine or forecasting of weather [7]. Majority of models reported researchers depend on artificial intelligence (AI) assisted methods that combined NLP models and knowledge base and QA logics. The QA system's processing can be divided into three stages: question analysis (parsing, question classification, and query reformulation); document analysis (extracting potential documents and identifying potential answers); and retrieval analysis (extracting potential answers and ranking them). Artificial intelligence, natural language processing, statistical analysis, pattern matching, information retrieval, and information extraction are all used in the task of answering questions. Most contemporary works incorporate some or all of these approaches to generate improved systems that can manage their shortage. Additionally, the taxonomy arrived at this work for classifying the various QA systems created so far is based on the general approach used by each system in its many processing phases, which give rise to the following categorization: Three approaches: linguistic, statistical, and pattern. The information of knowledge is arranged in terms of production rules, frames, logics, templates, semantic networks and ontology's which are used throughout assessment of question-answer pairs. The Linguistic models like tagging POS, tokenization and parsing are executed to user's question to formulate it into an accurate query that mines the individual reaction from the structured dataset. Furthermore, with NLP models, some of the knowledge base QA models rely on rule-based models. After adapting NLP models, the rules are constructed to identify the question classification features. Quarc and Cqarc utilize heuristic rules which look for lexical and semantic clues for identifying the question sets. However, the classification of question set varies from one model to another [8].

The aim of the present work is to develop a model exploiting abstractive summarization for a hybrid deep model based QAS. BERT tokenization for obtaining tokens followed by aspect term extraction are usedto extract aspects. Sentiment rating is predicted with ensemble classifiers, like Deep Residual Network (DRN), Deep Quantum Neural Network (DQNN), and Deep Neuro Fuzzy Network (DNFN) and its outputs are fused with majority voting. Output generated from aspect term extraction is then considered as an input for abstractive summarization and the process is performed using HAN and ABCDM. The model comprises a training and testing phase. In the training process, HAN and ABCDM are trained with summarized text, question answer pairs and sentiments obtained using ensemble classifiers. In the testing phase, question and summary are considered as input subjected to trained HAN and ABCDM in order to retrieve relevant answers.

2. LITERATURE SURVEY

One of the major issues with the existing abstractive summarization techniques is that the produced summary can be factually inconsistent. A new model is therefore essential for effective abstractive summarization with question answering.

Nan, F *et al.*[9] reported a technique to address the factual consistencies in the process of summarization. Here, an effectual measure was devised for computing the factual consistency and then a learning model was modeled which increased the designed measure in the training model. However, this method did not include non-differentiable computation measures in model training. To consider non-differentiable computation measure, a Durmus, E *et al.*[10] devised an automatic QAS based measure for trust, that incorporates current advancements in reading conception. The obtained QA-based measure gave high correlation, particularly on extremely abstractive summaries. But the final computation depends on human annotation. To avoid human dependence, Eyal, M *et al.*[11] devised a metric, namely Answering Performance for Evaluation of Summaries (APES) for summarization. APES used current progression in reading-comprehension to enumerate capability of summary to answer questions. This method required prolonged time to accomplish a task. To minimize operating time, Dong, L*et al.*[12] devised Unified pre-trained Language Model (UNILM) which was fine-tuned for understanding NLP processes. However, this technique could not perform multi-task fine-tuning on NLU and NLG tasks. For effective fine-tuning, Chen, Y.C. and Bansal, M [13] devised a precise and quick summarization technique which first chooses relevant sentences and then rewrites them obstructively to produce a brief summary. A sentence-level policy gradient method was developed for bridging the gap amidst two neural networks. To increase speed, Narayan, S *et al.*[14] developed a training technique that globally optimized ROUGE measure with reinforcement learning. An algorithm was utilized for training neural summarization on CNN. However, this model required a long time. To minimize processing time, Su, D*et al.*[15] developed a multi-task learning model which learned the shared representation amidst various processes. This model was constructed on top of a huge pre-trained language technique, like XLNet, and then fine-tuned on several RC datasets, but it produced huge overhead. To reduce overhead, Dong, Y*et al.*[16] devised Span Fact for question answering to build corrections in system-produced summaries through span selection. This model adapted a single or multi-masking method to iteratively or auto-regressively restore entities for ensuring semantic consistency, but suffered from huge computational complexity. Chen P*et al*[4] devised an innovative semantic evaluation method for different NLP applications without manual efforts and NLP systems details. However it failed in semantic comparison of two texts and other NLP tasks. Varasai, P*et al.*[5] developed an annotation schema which depicted the relations between the part of text as well as particular textual span with the discourse relations. Although the method addressed a number of tasks in discourse level annotated corpus development, it failed to extend the annotation in QA systems.

3. PROPOSED METHODOLOGY

Advancements in generating neural text have gained huge interest in the abstractive summarization in recent years despite its progression, several issues still remain in summarization. In order to address some of the issues, an abstractive summarization with a question answering system using a hybrid deep model is developed. Initially, the review data, and products are given as an input to BERT tokenization. These tokens are given to aspect term extraction in order to extract the aspects from review data [11]. The sentiment rating is then predicted using ensemble classifiers, like DRN [12], DQNN [14], and DNFN [15] and the outputs of each model are fused with majority voting. The output obtained from aspect term extraction and BERT tokenization is used for graph generation. Using the graph, the Co-occurrence matrix is generated, which is considered as an input for abstractive summarization and the process is done using Hierarchical Attention Network (HAN) [16] and Attention-based Bidirectional CNN-RNN Deep Model (ABCDM) [13].The proposed approach consists of two phases, namely training phase and testing phase. In the training process, HAN and ABCDM are trained with input review data, summarized text, question answer pairs and fused output obtained using majority voting method. In the testing phase, question, review data and summary are considered as an input to the trained HAN and ABCDM in order to retrieve relevant answers. Figure 1 depicts the framework of abstractive summarization with a question answering system using ABCDM-HAN.

3.1. Input data

Assume a database ϑ, λ and τ for product ID, product review text and product questions, respectively with total samples count and it is expressed in the below expression.

$$\vartheta = \{\vartheta_1, \vartheta_2, \ldots, \vartheta_i, \ldots, \vartheta_m\} \tag{1}$$

where, m symbolize total product ID's in the database, and ϑ_i signifies i^{th} product ID.

$$\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_d, \ldots, \lambda_f\} \tag{2}$$

where, f symbolize total product review text in the database, and λ_d signifies d^{th} product review text.

$$\tau = \{\tau_1, \tau_2, \ldots, \tau_o, \ldots, \tau_n\} \tag{3}$$

where, n symbolize total product questions in the database, and τ_o signifies o^{th} product question.

3.2. Prediction of sentiment rate using ensemble model

The prediction of sentiment rating is executed using certain steps. Initially, the review data is accumulated, and BERT tokenization is applied to generate tokens. From each token, the ATE is applied for generating the aspects. Finally, the sentiment rating is predicted using the ensemble classifier that includes DRN, DNFN and DQNN. The steps involved are briefly described below.

a). Tokenization using BERT

After obtaining the text document λ, the tokenization is performed using BERT to split obtained documents into single word. In BERT [17], the downstream processes are effectively controlled using input or output illustration [18]. The input illustration reveals the set of sentences or single sentences into each sequence of tokens. The sentences represent a random span of adjoining text. Here, the learned embeddings are integrated using each token that express if it fit in to the sentence E or F. The outcome of BERT tokenization is termed as M, and is fed to the aspect term extraction phase.

b) Extraction of aspect terms

The BERT tokenization output M is fed as input of Aspect term extraction (ATE). The ATE helps to study characteristics of aspects and reluctantly mine aspects using the texts, and saves time and labor. ATE helps to extract and group aspects parallelly considering the seed words offered by users for various aspect classes. The synonymous aspects are located in the same class. ATE uses word embedding to reveal the concurrence word distribution, and then uses attention technique for weakening unconnected words and improves constancy of each aspect. The ATE [18] is a procedure of sequence labeling tasks that arrange input considering labels. The labels denoted as A_{asp}, C_{asp}, D indicates beginning, inside and outside of aspect terms respectively. For performing the ATE process, the input employed is "The price is reasonable although the service is poor" and it is provided by $H = \{\omega_1, \omega_2, \ldots, \omega_p\}$, and p signifies a token after tokenization, $p = 9$ indicates total tokens. The aforementioned samples are given by, $K = \{L, A_{asp}, L, L, L, L, A_{asp}, L, L\}$. The ATE output is denoted as S, and is subjected to sentiment rating prediction, performed with ensemble classifiers.

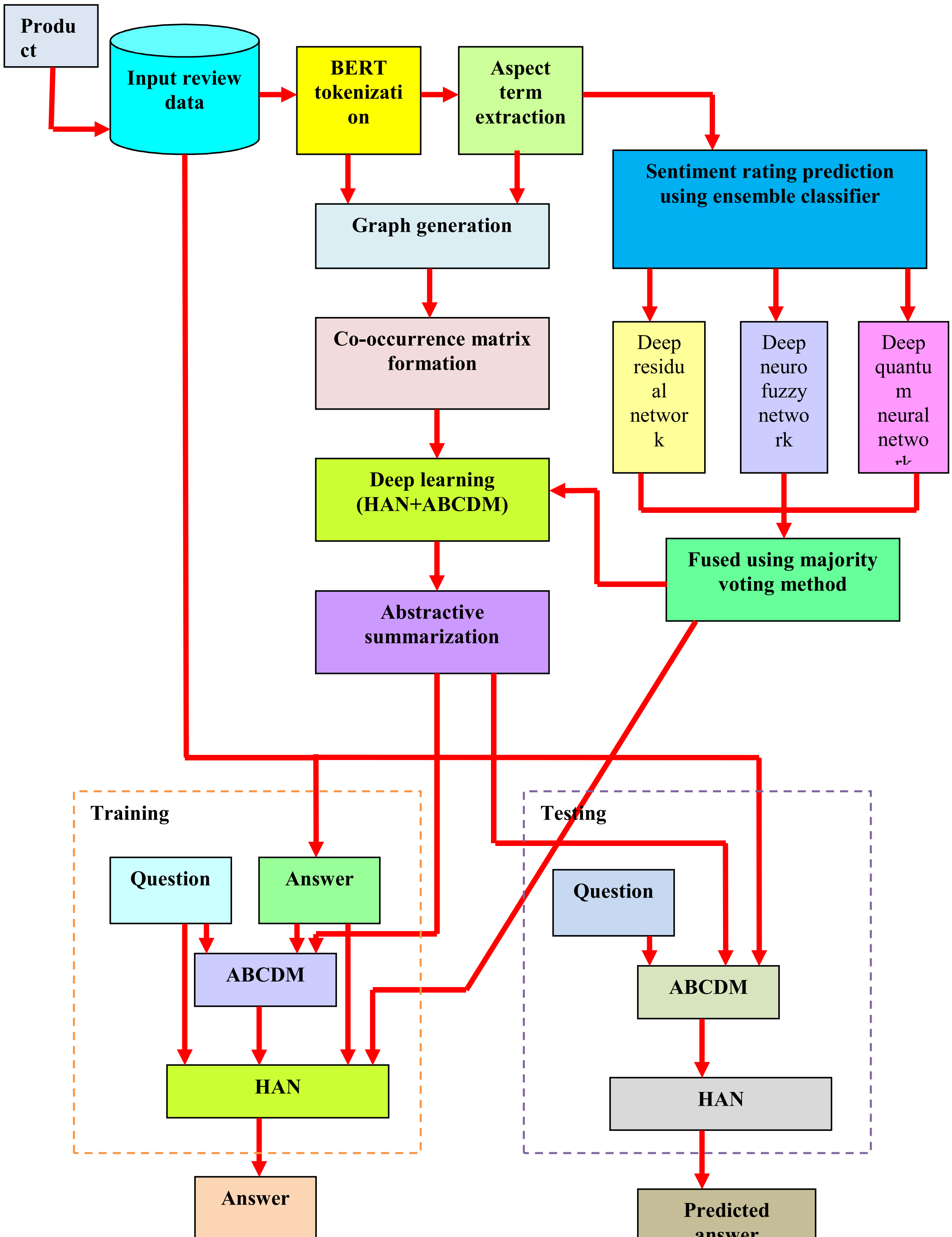


Figure 1. Structure Of Abstractive Summarization With Question Answering System Using Proposed ABCDM-HAN

c) Prediction of sentiment rating with ensemble classifier

The ATE output S is served as an input to ensemble classifier. The assessment of sentiment is important in extracting the user preferences. The sentiments are used to define the attitude of the user concerning a product. Here, the ensemble classifier is adapted to predict the sentiment rating. The ensemble contains three classifiers as described namely DRN [12], DNFN [20], and DQNN [21]and output of these classifiers is $\psi_1, \psi_2, and\ \psi_3$.

i) Deep Residual Network

The Residual Blocks idea was created to address the issue of the vanishing/exploding gradient. We apply a method known as skip connections in this network. The skip connection bypasses some levels in between to link layer activations to subsequent layers. This creates a leftover block. These leftover blocks are stacked to create resnet. The output of DRN classifiers is denoted as ψ_1

ii) Deep Neuro-Fuzzy network

The combination of deep neural networks and fuzzy logic is known as deep neuro-fuzzy. Deep neuro-fuzzy optimizer is utilized for effective cost and peak reduction optimization. It is a hybrid technique where the system objectives are computed using deep neural networks first and fuzzy logic in the second stage [42]. It consists of three layers namely, input layer, hidden layers and output layer. The input layer, a number of hidden layers for learning and validation, and the output layer make up the entire system. The input layer of a system is based on the quantity of input parameters and their fuzzification values. Rule, normalization, and defuzzification layers are the three hidden layers. The defuzzification layer is the output layer. The output of the DNFN is depicted as ψ_2

iii) Deep Quantum Neural Network

The quantum perceptron is the smallest building unit of a quantum neural network. A quantum perceptron is a unitary operator with any number of input and output qubits). The output qubits are initialized in a fiducial product state, whereas the input qubits are initialized in a potentially unknown mixed state. The output of the DNFN is given by ψ_3

d) Fusion using majority voting method

The voting is performed with estimate of each technique regarding each instance and the final prediction is done using majority votes. The outputs from the three classifiers are combined which plays a significant role in prediction. Majority voting [22] is utilized to label the output. In majority voting, highest figures of votes is needed to make decision. Outputs of DRN, DNFN and DQNN denoted as $\psi_1, \psi_2, and\ \psi_3$ are considered as an input to fusion. Consider a class that describes the classifier output U_i as depicted in d-dimensional binary vectors represented as,

$$[U_{s,1} \ldots U_{s,e}] \in \{0,1\}^e, s = 1, \cdots, V \tag{4}$$

$$\sum_{s=1}^{N} U_{s,k} = \max^c j = 1 \sum_{s=1}^{N} U_{s,j} \tag{5}$$

The output obtained by fusing ensemble classifier with majority voting method is modeled as U, that aids to forecast sentiment rating.

3.3 Abstractive summarization

The illustration of the process of abstractive generation is provided in which aspect terms S and tokenized data M are used to build summary. The abstractive summary is produced by utilizing the HAN+ABCDM. The process is elaborated below.

a) Graph generation

The graph is generated on the basis of frequency of occurrence of words. It employs the BERT tokenization output M and ATE output S as input for generating graphs. The obtained graph is expressed as ζ.

b) Cooccurance matrix generation

For producing the co-occurrence matrix, the graph ζ produced with aspect terms and tokens is adapted as an input. Here, the co-occurrence matrix is produced by employing the below expression,

$$q_{uv} = \frac{1}{2}(r_{uv} + s_{uv}) \tag{6}$$

Where, r_{uv} is the frequency of occurrence of the u^{th} and v^{th} word and s_{uv} is given by,

$$s_{uv} = \begin{cases} 1 & if \quad u \,\&\, v\ are\ AST,\ u \neq v \\ 0.5 & if \quad u\ or\ v\ is\ AST \\ 0 & if \quad u\ or\ v\ are\ not\ AST \end{cases} \tag{7}$$

The co-occurrence matrix value is produced with equation (7). The co-occurrence matrix q obtained is subjected to HAN+ABCDM, which performs abstractive summarization.

c) Abstractive summarization using HAN+ABCDM

The obtained co-occurrence matrix q is subjected as input to HAN+ABCDM, wherein the HAN+ABCDM is obtained by combining HAN [24] and ABCDM [23]. It uses hyper parameters of HAN and ABCDM. The ABCDM considers an attention model which puts more focus on various words. In addition, it minimizes the features' dimensionality. The HAN helps to categorize a text with its contextual information by collecting imperative words from the sentence vectors and imperative sentences from the document vectors. It aids to control large quantity of textual data. Moreover, it mines future and past contexts by temporal information. Thus, the combination of HAN and ABCDM improves overall efficiency for performing abstractive summarization, is performed in two phases, namely training and testing phase.

3.4. Question answering system using hybrid HAN+ABCDM architecture

The generation of question-answer is done using HAN+ABCDM. Here, the process is done using two stages, namely training and testing stage. Each of these stages is briefly examined below.

3.4.1. Training

In training phase, the question B , answer G and abstractive summarization A is provided as an input to ABCDM. The obtained result is further fed to HAN along with the question B , answer G to obtain the final answer. Figure 2 displays training phase processing in generation of question-answer.

a) ABCDM Architecture

Figure 3 displays ABCDM. In ABCDM, the embedding of GloVe word, bidirectional LSTM, bidirectional GRU, CNN and attention technique is used for improved capturing considering both local features and long-term dependencies. To generate matrix of input comment, the pre-trained GloVe embedding matrix $\varpi_g \in \aleph^{u \times v}$ with u representing total words and v express size of embedding is used to entrench comment vector $b \in \aleph^{l}$ such that l refers padding length and highest word count $w_t, t \in [1,l]$ adapted in comment as,

$$b_t = \varpi_g w_t, t \in [1,l] \tag{8}$$

The GRU and LSTM are employed to identify both short and longer sequence of data.

$$\overrightarrow{a_{t_{LSTM}}} = \overrightarrow{LSTM}(b_t), t \in [1,l] \tag{9}$$

$$\overleftarrow{a_{t_{LSTM}}} = \overleftarrow{LSTM}(b_t), t \in [l,1] \tag{10}$$

$$\overrightarrow{a_{t_{GRU}}} = \overrightarrow{GRU}(b_t), t \in [1,l] \tag{11}$$

$$\overleftarrow{a_{t_{GRU}}} = \overleftarrow{GRU}(b_t), t \in [l,1] \tag{12}$$

Using each word, w_t one can produce annotation by combining forward and backward expressed as,

$$a_{t_{LSTM}} = \left[\overrightarrow{a}_{t_{LSTM}}, \overleftarrow{a}_{t_{LSTM}} \right] \tag{13}$$

$$a_{t_{GRU}} = \left[\overrightarrow{a}_{t_{GRU}}, \overleftarrow{a}_{t_{GRU}} \right] \tag{14}$$

The attention technique is employed on $a_{t_{LSTM}}$ and $a_{t_{GRU}}$ to make the model able to pay more concentration to several words in review. Thus, the feature vector is obtained by extracting edifying

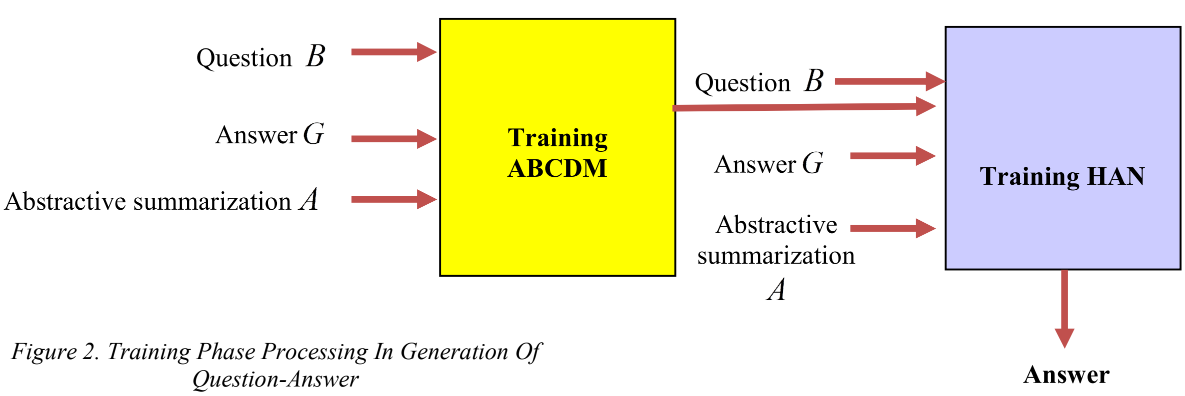


Figure 2. Training Phase Processing In Generation Of Question-Answer

words and expressed by,

$$z_{t_{LSTM}} = \tanh\left(W_{wLSTM} a_{tLSTM} + w_{wLSTM}\right) \quad (15)$$

$$z_{t_{GRU}} = \tanh\left(W_{wGRU} a_{tGRU} + w_{wGRU}\right) \quad (16)$$

$$\alpha_{tLSTM} = \frac{\exp\left(z_{t_{LSTM}}^{T} z_{w_{LSTM}}\right)}{\sum_t \exp\left(z_{t_{LSTM}}^{T} z_{w_{LSTM}}\right)} \quad (17)$$

$$\alpha_{tGRU} = \frac{\exp\left(z_{t_{GRU}}^{T} z_{w_{GRU}}\right)}{\sum_t \exp\left(z_{t_{GRU}}^{T} z_{w_{GRU}}\right)} \quad (18)$$

$$s_{LSTM} = \sum \alpha_{t_{LSTM}} a_{t_{LSTM}} \quad (19)$$

$$s_{GRU} = \sum \alpha_{t_{GRU}} a_{t_{GRU}} \quad (20)$$

where, z_t express hidden model of a_t, and z_w symbolize context vectors. The consequence of word z_t is evaluated using similarity of z_t and z_w is normalized.

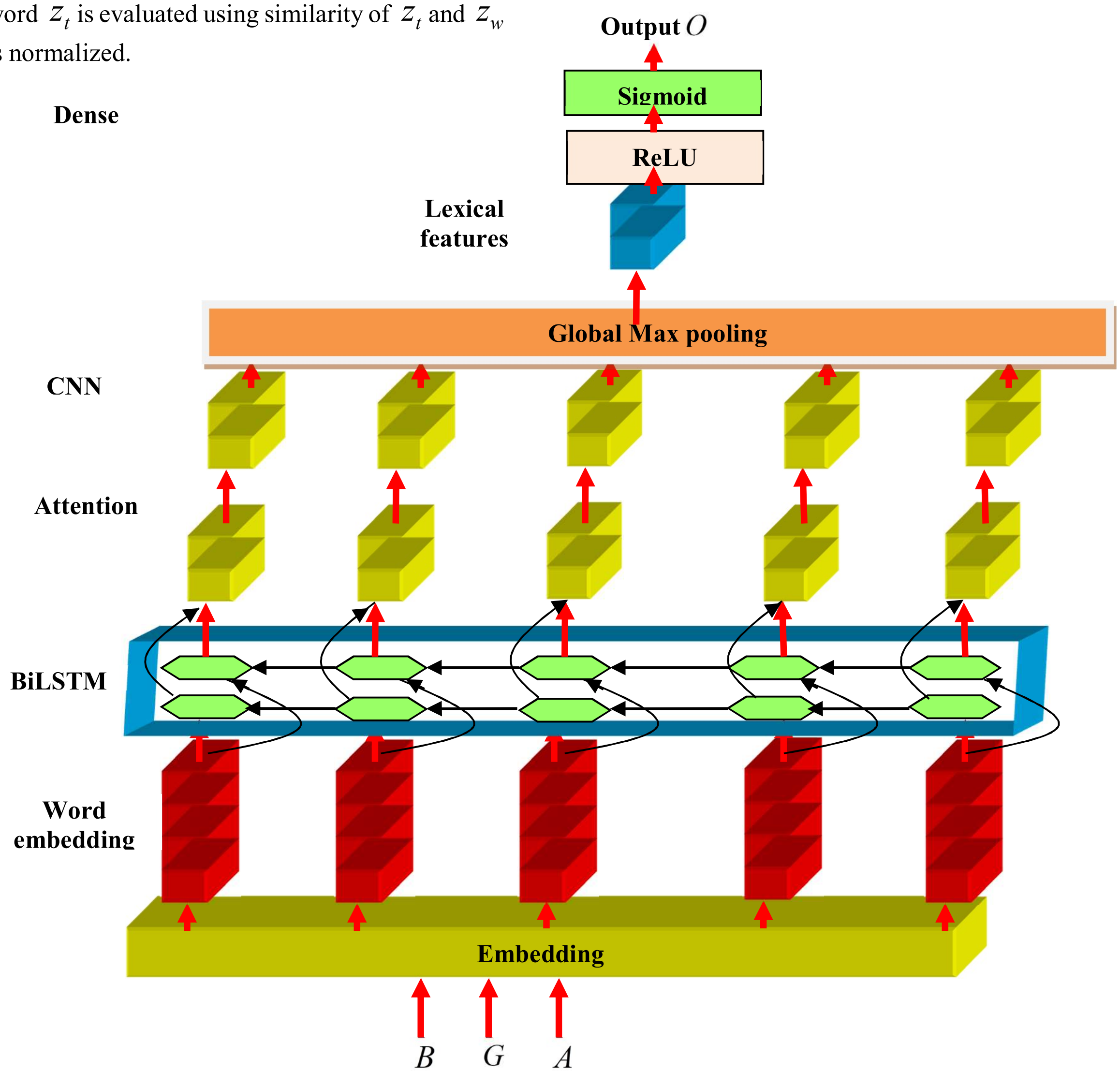


Figure 3. Architecture Of ABCDM

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Consider feature vector Lc is expressed as $Lc_i = [lc_1, lc_2, \ldots, lc_f]; i \in [1,8]$. The output of this layer is expressed as,

$$a_d = \mathrm{Relu}(\varpi_d a_p + w_d) \qquad (21)$$

where, a_p express hidden representation, and ϖ_d and w_d signifies learned parameters. The output produced by ABCDM is O.

b) HAN Architecture

HAN model covers various attributes in such a way that it includes sentence attention layer, word attention, sentence encoder, and word encoder. This model incorporates questions B, answers G and abstractive summarization A as an input. The HAN model is depicted in fig.4.

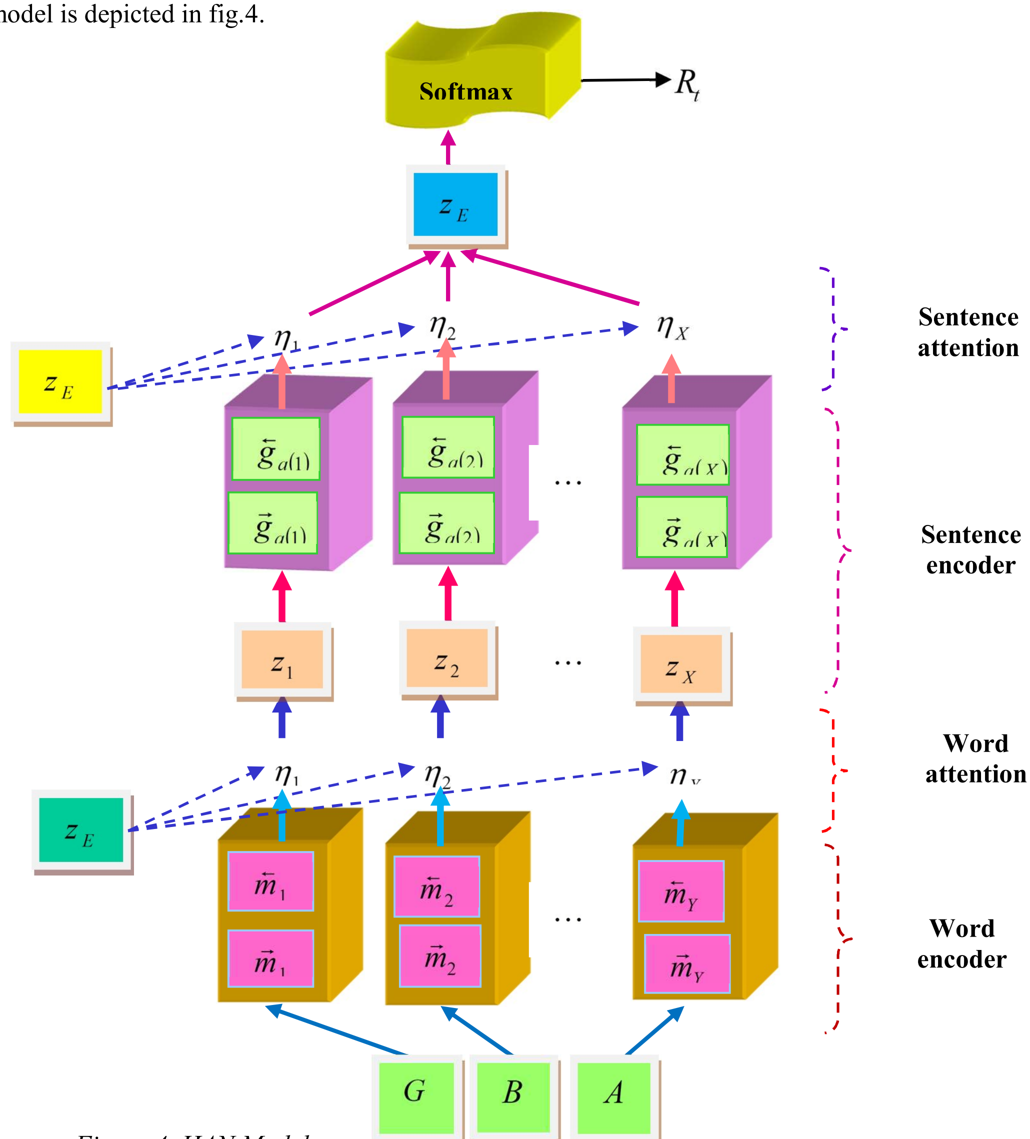


Figure 4. HAN Model

i) Word encoder

The input questions B, answers G and abstractive summarization A are used to vectors considering embedding matrix W. In addition, the bidirectional GRU is adapted to produce word annotations.

$$E = FD_g.;\, D_g \in [1,...,\mathrm{Y}] \tag{22}$$

$$\vec{t} = \vec{\Re}(E), \alpha_\ell \in [1, \mathrm{Y}] \tag{23}$$

$$\overleftarrow{t} = \overleftarrow{\Re}(E), \alpha_\ell \in [\mathrm{Y}, 1] \tag{24}$$

The annotation is attained for mined feature by combining forward hidden state $\vec{t}$ and backward hidden state $\overleftarrow{t}$ in such a way that, $t = \left[\vec{t}, \overleftarrow{t}\right]$.

ii) Word attention

Thus, all words do not constantly give explanation of sentence meaning. Meanwhile, the attention factor is modeled to mine words and is noteworthy in sentences. The illustration of informative features is united to generate the sentence vector.

$$x_{\mathrm{P}} = \tanh(J.t + o) \tag{25}$$

$$\rho = \frac{\exp\left(x_{\mathrm{P}}^J\, x_E\right)}{\sum_Q \exp\left(x_{\mathrm{P}}^J\, x_E\right)} \tag{26}$$

$$E = \rho.t \tag{27}$$

Initially, m that refers annotation of word is subjected to one-layer MLP for generating x_{P} that indicates the hidden representation of t. In addition, x_{P} that indicates word importance is computed using word level context vector x_E and attain normalized weight ρ based on the SoftMax function.

iii) Sentence encoder

It adapted and devised bidirectional GRU to encode sentences using sentence vector E.

$$\overrightarrow{t_f} = \overrightarrow{\Re}(E) \tag{28}$$

$$\overleftarrow{t_f} = \overleftarrow{\Re}(E) \tag{29}$$

Thus, annotation of sentence is acquired by combining $\overrightarrow{t_f}$ and $\overleftarrow{t_f}$, wherein $t_f \left[{}_f t = \overrightarrow{t_f}, \overleftarrow{t_f}\right]$ review adjacent sentences.

iv) Sentence attention

The sentence level context vector M is utilized to discover sentence consequences.

$$z_f = \tanh\left(F_I t_f + o_I\right) \tag{30}$$

$$\eta_f = \frac{\exp\left(z_f^J V\right)}{\sum_f \exp\left(z_f^J V\right)} \tag{31}$$

$$Q = \sum_f \eta_f . t_f \tag{32}$$

where, Q expose document vector. The output is denoted as R_t.

3.4.2. Testing

In testing phase, the question B, input review data J and abstractive summarization A are provided as an input to trained HAN. The obtained result is further fed to ABCDM to obtain the final answer. Figure 5 displays schematic representation of testing phase in generation of question-answer.

4. RESULTS AND DISCUSSION

The efficiency of ABCDM+HAN is inspected by evaluating techniques with different measures by varying review percentage.

4.1. Setup of experiments

Both ABCDM and HAN are scripted in PYTHON.

4.2. Dataset description

Amazon question/answer data [25] and Amazon product data [26] are used as dataset for the present study.

a) Amazon question/answer data

This Amazon question/answer data [25] comprises Question and Answer data acquired from the Amazon with total 1.4 million answered questions. It is combined with Amazon product review data by matching product ID in the Q/A dataset. This review data involves product metadata. It involves certain parameters like ASIN, question Type, answer Type and so on.

b) Amazon product data

This Amazon product data [26] comprises reviews of products and metadata collected from Amazon that involves 143.7 million reviews that arrives from May 1996 to July 2014. It contains certain attributes like reviewer ID, ASIN, summary and so on.

Journal of Theoretical and Applied Information Technology
31st March 2023. Vol.101. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

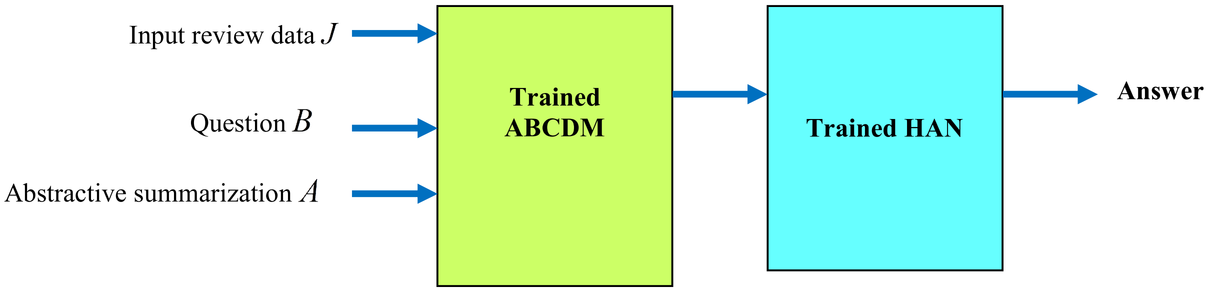


Figure 5. Schematic Representation Testing Phase In Generation Of Question-Answer

4.3. Performance measures
The developed technique is assessed with certain measures for computing efficiency as described below.

4.3.1 Precision: It refers closeness degree of more than two sizes amongst each other, and is formulated by,

$$\Pr = \frac{\partial_p}{\partial_p + \hbar_p} \qquad (33)$$

where, ∂_p signifies true positive, and $\hbar_p$ is false positive.
4.3.2. Recall: It computes entire actual positives wherein model accumulates with label of true positive, and it is expressed by,

$$\mathrm{Re} = \frac{\partial_p}{\partial_p + \hbar_f} \qquad (34)$$

where, $\hbar_f$ is false negative.

4.3.3. F Measure: It refers weighted harmonic mean of recall and precision, and is modeled by,

$$FM = 2 \times \left(\frac{\Pr * \mathrm{Re}}{\Pr + \mathrm{Re}} \right) \qquad (35)$$

4.3.4. Rouge: It is used to evaluate automated text summarization and translation of machines. It performs by comparing and producing the summaries automatically and translates a set of reference summaries. It is formulated as,

$$Rouge = \frac{N_o}{T} \qquad (36)$$

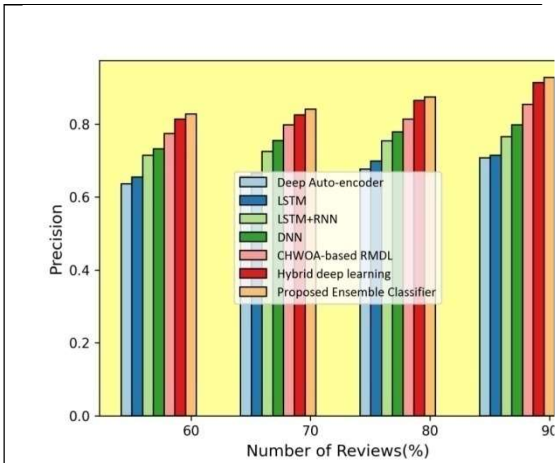
where, N_o signifies overlapping words count, T symbolize total words in reference summary.

4.4. Comparison of Proposed Method with other Existing Methods

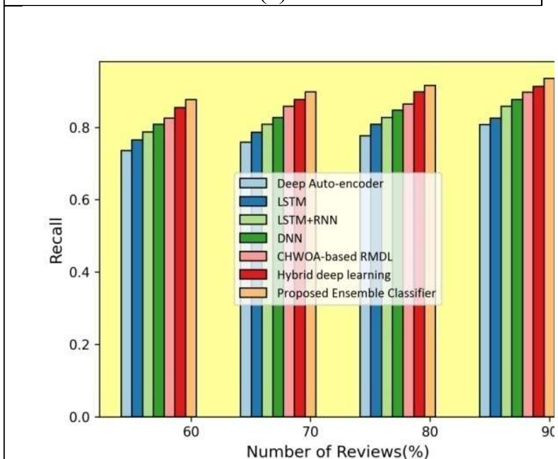
The sentiment rating methods considered for comparison with the proposed method includes Deep-Autoencoder [27], LSTM [28], LSTM+RNN [29], DNN [30], CHWOA-based RMDL[31], hybrid deep learning[32].
The abstractive summarization methods considered for comparison with the proposed method include sEndorsement model [33], RNN+LSTM [34], RARS [35], HITS based attention model [36] and CNN-Transfer learning[37].
The question and answer methods considered for comparison with the proposed method includes TransTQA[38], SemBioNLQA[39], XLNet,[40] Multi-KB QA[41] andQAInfomax[43].
The performance of the designed ensemble classifier is assessed by using different measures as a function of number of reviews. The discussion regarding the sentiment rating, abstractive summarization and Question and answering analysis is described below.

4.4.1. Assessment of sentiment rating
Figure 6 displays the assessment of sentiment rating with different metrics. The data indicates that efficiency of Sentiment Rating in terms of Precision, Recall and F-measure vary as a function for number of reviews ranging from 60 – 90%, compared with the existing methods the performance is Deep Auto-encoder < LSTM << LSTM+RNN < DNN < ChWoA-based RMDL < Hybrid Deep Learning < Proposed Ensemble Classifier.

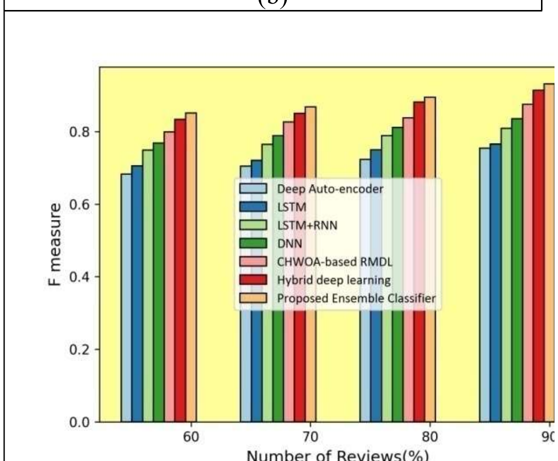
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(a)



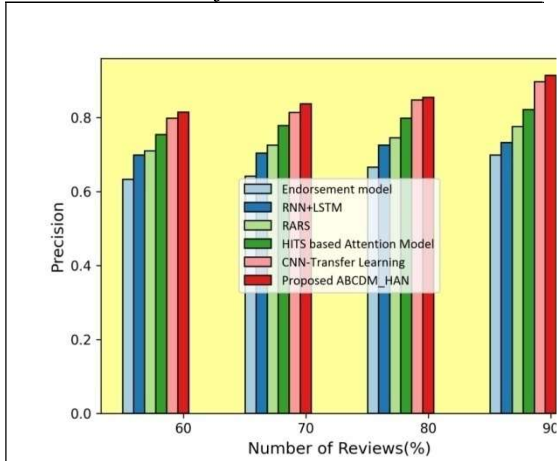
(b)



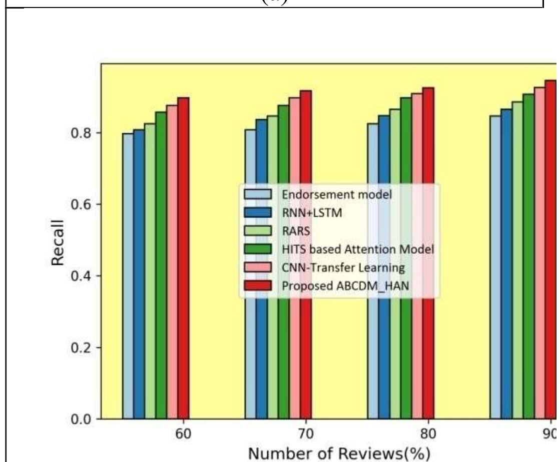
(c)

Figure 6. *Assessment Of Sentiment Rating With*
A) Precision Of Ensemble Classifier
B) Recall Of Ensemble Classifier
C) F Measure Of Ensemble Classifier

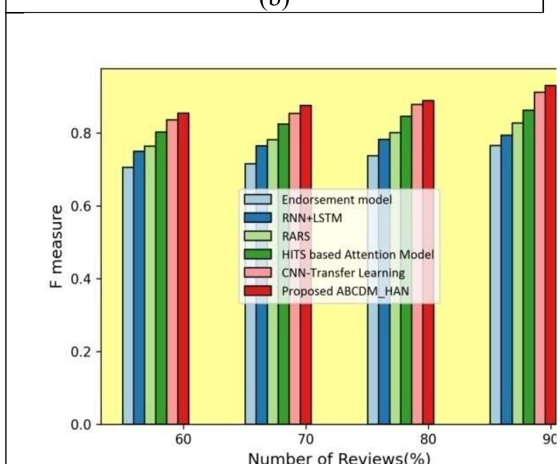
4.4.2. Assessment of abstract summarization



(a)

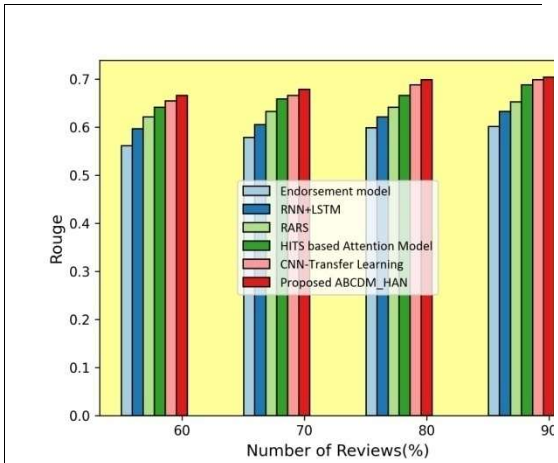


(b)



(c)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

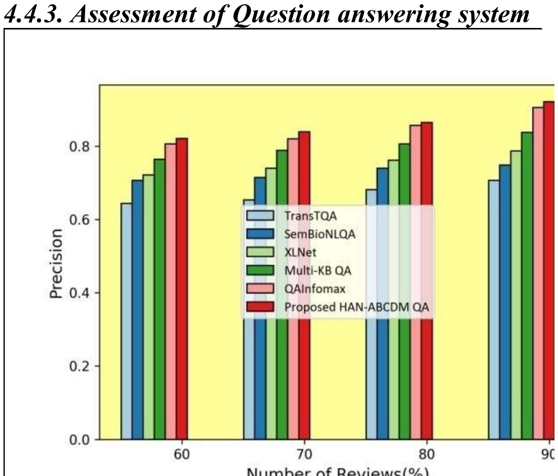


(d)

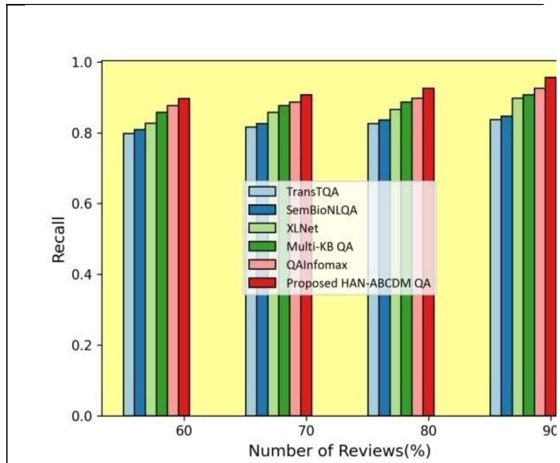
Figure 7. *Assessment Of Abstract Summarization Rating Using*
A) Precision Of ABCDM-HAN
B) Recall Ofabcdm-HAN
C) F Measure Of ABCDM-HAN
D) Rouge Of ABCDM-HAN

Figure 7 depicts analysis of abstractive summarization with different metrics as a function of % of reviews. From the experimental results, it can be seen that the abstractive summarization values vary as a function of review % varying from 60 to 90%. The above data indicates that performance of abstractive summarization in terms of Precision, Recall and F-measure vary as a function for number of reviews ranging from 60 – 90%, compared with the existing methods the performance is Endorsement model < RNN + LSTM < RARS < HITS based attention < CNN-Transfer Learning < Proposed HAN+ABCDM.

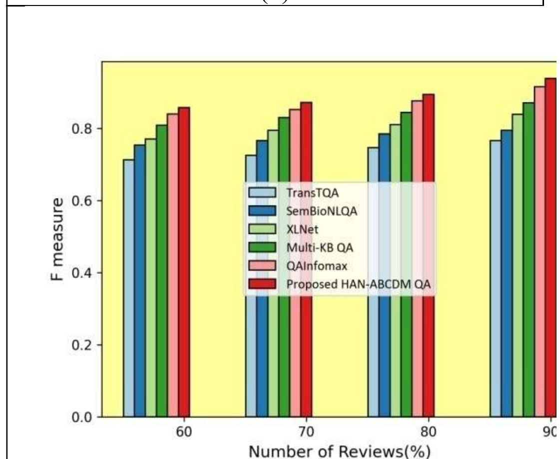
4.4.3. Assessment of Question answering system



(a)



(b)



(c)

Figure 8. *Assessment Of Question And Answering System Using*
A) Precision O Fhan-Abcdm Qa
B) Recall Of Han-Abcdm Qa
C) F Measure Of Han-Abcdm Qa

Figure 8 depicts analysis of Question and answering system with some metrics. The assessment graph is explicated in figure 8. The above data indicates that efficiency of Question Answering System in terms of Precision, Recall and F-measure vary as a function for number of reviews ranging from 60 – 90%, compared with the existing methods the performance is TransTQA<SemBioNLQA<XLNet< Multi-KB QA <QAInfomax< Proposed HAN-ABCDM QA.

5. CONCLUSION AND FUTURE WORK

Based on the above results and discussion, it ca be concluded that abstractive summarization enhances the efficiency of HAN-ABCDM based Question Answering System, key steps involved in Hybrid Deep Learning model are described below.The review data is first subjected to BERT tokenization to generate tokens and then aspect term extraction is used to extract the aspects from review data. Further, the sentiment rating is predicted using ensemble classifiers, like DRN, DQNN, and DNFN and its obtained outputs are fused with the majority voting. The output obtained from aspect term extraction is considered as an input for abstractive summarization and the process is done using HAN and ABCDM with training and testing stages. In the training stage, HAN and ABCDM are trained with summarized text question answer pairs and fused output obtained from ensemble classifier. In testing stage, question and summary are considered as an input subjected to the trained HAN and ABCDM in order to retrieve relevant answers. The proposed method is effective in enhancing quality of summaries. Considering question answering model, the proposed HAN-ABCDM QA offered the highest precision of 82.1%, recall of 95.7%, and F-measure of 93.9% which are higher than the corresponding values obtained for other exiting methods.

REFERENCES

[1]Huang, Y., Feng, X., Feng, X. and Qin, B, "The factual inconsistency problem in abstractive text summarization: A survey," arXiv preprint arXiv:2104.14839, 2021.

[2]Deutsch, D., Bedrax-Weiss, T. and Roth, D, "Towards question-answering as an automatic metric for evaluating the content quality of a summary," Transactions of the Association for Computational Linguistics, vol.9, pp.774-789, 2021.

[3]Chali, Y., Matwin, S. and Szpakowicz, S, "Query-biased text summarization as a question-answering technique," In Proceedings of the AAAI Fall Symposium on Question Answering Systems, pp. 52-56, 1999.

[4].Chen, P., Wu, F., Wang, T. and Ding, W, "A semantic qa-based approach for text summarization evaluation," In Proceedings of the AAAI Conference on Artificial Intelligence, vol.32, no.1, 2018.

[5]Varasai, P., Pechsiri, C., Sukvaree, T., Satayamas, V. and Kawtrakul, A, "Building an Annotated Corpus for Text Summarization and Question Answering," In LREC, 2008.

[6]Hirao, T., Sasaki, Y. and Isozaki, H, "An extrinsic evaluation for question-biased text summarization on QA tasks," In Proc. of the NAACL Workshop on Automatic Summarization, pp. 61-68, 2001.

[7]Pundge, A.M., Khillare, S.A. and Mahender, C.N, "Question answering system, approaches and techniques: a review," International Journal of Computer Applications, vol.141, no.3, pp.0975-8887, 2016.

[8]Dwivedi, S.K. and Singh, V, "Research and reviews in question answering system," Procedia Technology, vol.10, pp.417-424, 2013.

[9]Nan, F., Santos, C.N.D., Zhu, H., Ng, P., McKeown, K., Nallapati, R., Zhang, D., Wang, Z., Arnold, A.O. and Xiang, B, "Improving factual consistency of abstractive summarization via question answering," arXiv preprint arXiv:2105.04623, 2021.

[10]Durmus, E., He, H. and Diab, M, "FEQA: A question answering evaluation framework for faithfulness assessment in abstractive summarization," arXiv preprint arXiv:2005.03754, 2020.

[11]Eyal, M., Baumel, T. and Elhadad, M, "Question answering as an automatic evaluation metric for news article summarization," arXiv preprint arXiv:1906.00318, 2019.

[12]Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., Gao, J., Zhou, M. and Hon, H.W, "Unified language model pre-training for natural language understanding and generation," Advances in Neural Information Processing Systems, vol.32, 2019.

[13]Chen, Y.C. and Bansal, M, "Fast abstractive summarization with reinforce-selected sentence rewriting," arXiv preprint arXiv:1805.11080, 2018.

[14]Narayan, S., Cohen, S.B. and Lapata, M, "Ranking sentences for extractive summarization with reinforcement learning," arXiv preprint arXiv:1802.08636, 2018.

[15]Su, D., Xu, Y., Winata, G.I., Xu, P., Kim, H., Liu, Z. and Fung, P, "Generalizing question answering system with pre-trained language model fine-tuning," In Proceedings of the 2nd Workshop on Machine Reading for Question Answering, pp. 203-211, 2019.

[16] Dong, Y., Wang, S., Gan, Z., Cheng, Y., Cheung, J.C.K. and Liu, J., "Multi-fact correction in abstractive text summarization," arXiv preprint arXiv:2010.02443, 2020.

[17] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", arXiv preprint

arXiv:1810.04805, 2018.
[18] Heng Yanga, Biqing Zeng, JianHao Yang, Youwei Song, and RuyangXua, "A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction", Neurocomputing, vol.419, pp.344-356, 2021.
[19] Zhicong Chen, Yixiang Chen, Lijun Wu, Shuying Cheng, and Peijie Lin, "Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions", Energy Conversion and Management, vol.198, pp.111793, 2019.
[20] SakeenaJavaid, Muhammad Abdullah, Nadeem Javaid, Tanzeela Sultana, Jawad Ahmed, and Norin Abdul Sattar, "Towards buildings energy management: Using seasonal schedules under time of use pricing tariff via deep neuro-fuzzy optimizer", In 1th International Wireless Communications & Mobile Computing Conference (IWCMC), pp. 1594-1599, June 2019.
[21] Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J. Osborne, Robert Salzmann, Daniel Scheiermann, and Ramona Wolf, Training deep quantum neural networks. Nature communications, vol.11, no.1, pp.1-6, 2020.
[22]Javid, I., Alsaedi, A.K.Z. and Ghazali, R, "Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method," International Journal of Advanced Computer Science and Applications, vol.11, no.3, 2020.
[23] Mohammad Ehsan Basiri, ShahlaNemati, MoloudAbdar, Erik Cambria, and U. Rajendra Acharya, "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis", Future Generation Computer Systems, vol.115, pp.279-294, 2021.
[24] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy, "Hierarchical attention networks for document classification", In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pp. 1480-1489, June 2016.
[25] Amazon question/answer data, "https://jmcauley.ucsd.edu/data/amazon/qa/," Accessed on october 2022.
[26]Amazonproduct data, "https://snap.stanford.edu/data/amazon/productGraph/", Accessed on october 2022.
[27]Socher, R., Pennington, J., Huang, E.H., Ng, A.Y. and Manning, C.D., "Semi-supervised recursive autoencoders for predicting sentiment distributions," In Proceedings of the conference on empirical methods in natural language processing, pp. 151-161, 2011.
[28]Trueman, T.E. and Cambria, E., "A convolutional stacked bidirectional lstm with a multiplicative attention mechanism for aspect category and sentiment detection," Cognitive Computation, vol.13, no.6, pp.1423-1432, 2021.
[29]Abdelgwad, M.M., Soliman, T.H.A., Taloba, A.I. and Farghaly, M.F., "Arabic aspect based sentiment analysis using bidirectional GRU based models," arXiv preprint arXiv:2101.10539, 2021.
[30]Kane, B., Jrad, A., Essebbar, A., Guinaudeau, O., Chiesa, V., Quénel, I. and Chau, S., "CNN-LSTM-CRF for Aspect-Based Sentiment Analysis: A Joint Method Applied to French Reviews," In ICAART, vol.1, pp. 498-505, 2021.
[31]Chen, Z., Zhang, K., Chan, T. H., Li, X., & Zhao, S. (2022). A Novel Hybrid Whale-Chimp Optimization Algorithm for Structural Damage Detection. *Applied Sciences*, *12*(18), 9036.
[32]Dang, Cach N., María N. Moreno-García, and Fernando De la Prieta. "Hybrid deep learning models for sentiment analysis." *Complexity* 2021 (2021).
[33] Xu, H., Liu, H., Zhang, W., Jiao, P. and Wang, W., "Rating-boosted abstractive review summarization with neural personalized generation", Knowledge-Based Systems, vol.218, pp.106858, 2021.
[34]Sheela, J. and Janet, B., "An abstractive summary generation system for customer reviews and news article using deep learning", Journal of Ambient Intelligence and Humanized Computing, vol.12, no.7, pp.7363-7373, 2021.
[35]Lebanoff, L., Wang, B., Feng, Z. and Liu, F., "Modeling Endorsement for Multi-Document Abstractive Summarization", arXiv preprint arXiv:2110.07844, 2021.
[36] Cai, X., Shi, K., Jiang, Y., Yang, L. and Liu, S., "HITS-based attentional neural model for abstractive summarization", Knowledge-Based Systems, vol.222, pp.106996, 2021.

[37]Song, Shengli, Haitao Huang, and TongxiaoRuan. "Abstractive text summarization using LSTM-CNN based deep learning." *Multimedia Tools and Applications* 78.1 (2019): 857-875.

[38]Yu, W., Wu, L., Deng, Y., Mahindru, R., Zeng, Q., Guven, S., & Jiang, M. (2020, October). A technical question answering system with transfer learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 92-99).

[39]Sarrouti, Mourad, and Said Ouatik El Alaoui. "SemBioNLQA: A semantic biomedical question answering system for retrieving exact and ideal answers to natural language questions." *Artificial intelligence in medicine* 102 (2020): 101767.

[40]Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV. Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems. 2019;32.

[41]Cheng L, Xie F, Ren J. KB-QA based on multi-task learning and negative sample generation. Information Sciences. 2021 Oct 1;574:349-62.

[42] Kavitha, C., Rao, M.B., Srikanth, B. et al. Zero shot image classification system using an optimized generalized adversarial network. Wireless Netw 29, 697–712 (2023). https://doi.org/10.1007/s11276-022-03166-8

[43]Yeh YT, Chen YN. QAInfomax: Learning robust question answering system by mutual information maximization. arXiv preprint arXiv:1909.00215. 2019 Aug 31.