Multimedia Tools and Applications (2022) 81:589–613
https://doi.org/10.1007/s11042-021-11228-w

Deep learning based Bengali question answering system using semantic textual similarity

Arijit Das[1] · Diganta Saha[1]

Received: 16 October 2020 / Revised: 28 January 2021 / Accepted: 7 July 2021 /
Published online: 14 September 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Recently, Question answering system is a major research area in language processing. Bengali isone of the most popular spoken languages in India. Still, it has faced difficulties in natural language processing.Among the semantic based systems, word mapping and keyword based approaches achieved the best results and got better attention on the user side. These systems are already implemented in various languages but not much in Indian language like Bengali. This work presents an efficient question answering system for retrieving Bengali language text. This system includes word embedding clustering and deep level feature representation for providing better grammatical similarities for retrieving the Bengali textual contents relevant to user queries. The pre-trained word embedding module is created by the help of a deep belief network. The modified density peak algorithm is employed to perform word embedding clustering.The presented work has been tested on a dataset from the Bengali corpus developed by TDIL and synthetic Bengali translated datasets accessible in English called SQuAD 2.0. This question answering system is implemented in python with NLTK tool kit and got good performance while retrieving the Bengali textual data.

Keywords NLP · Bengali question answering system · Semantic search · Deep Belief Network · Modified density peak clustering

1 Introduction

An automatic question answering (QA) model is concerned with developing systems that capableof retrieving the concise and preciseanswers for natural language questions posed by humans. In recent times, research in QA is progressivelysignificant due to the explosion of information on the internet [37]. QA is a research area that includes the fields of Information Retrieval (IR) and Natural Language Processing (NLP). In QA model, the direct access of specific textual data with computes are also not possible because of its unstructured format [34]. In such case, the similarity among the texts

✉ Arijit Das
arijit.mcse.ju@gmail.com

[1] Department of CSE, Jadavpur University, Kolkata, India

will be estimated by the information retrieval systems. The similarity measure of textual data has been processed already on different NLP application includes text classification, mining of sub topics from articles, relevance feedback, web search, disambiguation of word sense, and so on [15].

The lexical and semantic matching among the text is the standard technique for matching the similarity of text. The lexical matching based QA system has been carried out based on a dictionary of words and the lexical aspects of texts. This kind of matching is a basic one, and it never uses any grammar or Parts of Speech (POS) for analysing the textual contents [3]. As same as the lexical matching, the semantic matching also uses the semantic dictionary for textual analysis. But, additionally, it estimates the meaning of the words, grammar, and analyse the words in every sentence more deeply. It can provide more similar results than the lexical based matching [9].

The main idea behind such similarity estimation models is to retrieve the precise answers from the huge database based on the inputted query. The QA system mainly includes four different steps, indexing, creation of query, similarity estimation and comparison, and feedback [10]. The extraction of relevant answers or sentences from the large documents with key words or a query is useful in most of the NLP based applications like text summarization [8], word disambiguation [16], and text classification [24]. The keywords extraction from huge documents are also presented in some NLP works [27] with keyword extraction algorithms. For extracting such keywords, important words or sentences with the user query it is essential to identify the semantic relationship among the textual contents.

In recent times, more number of deep learning solutions has been studied to tackle the tasks of pattern recognition and multi-task information and image retrieval [22, 39, 40].The earlier machine learning approaches rely mainly on manually designed features based on expert knowledge of the domain.Feature engineering and feature extraction are key, time-consuming processes of the machine learning workflow. Following recent trends in natural language processing fields, the development of machine learning solutions for Questions Answering has started heading towards deep learning models. These solutions have replaced the aforementioned feature engineering process of the machine learning [29, 38].

In QA system, the semantic gap is a trending topic in research for describing the high level meaning of query contents and the contents of documents. It is a necessary step for selecting the relevant content based on the user query. The semantic gap among the textual contents are produced [30] by,

- Vocabulary mismatch: It represents the word with similar meaning and differed shapes.
- Granularity mismatch: It represents the differed sense and shape of the word refers to similar concepts.
- Polysemy: It represents the word that covers more senses based on its adjacent words.

The work presented in this paper focused on providing an efficient QA system for semantic search on Bengali text. There various QA techniques in this area are already done in some international languages like English; at the same time, few kinds of research are done with the Bengali language text [1]. The Bengali language is naturally more different, and it includes more vocabulary along with complex syntax. So, it provides more challenges while processing the answers retrieval process. It origins from Sanskrit language and written based on popular Indian language script Devanagari [35].

The similarities among the Bengali text can be estimated, whether in lexical or semantic. Basically, on most of the previous Bengali literature, similarity among textual contents is estimated based on the Bengali input query and the Bengali text that are available in the text documents [26]. Such similarity estimation can only describe the textual similarity among Bengali text, never describe the semantic similarity.

Example 1, *Let us keep in mind two Bengali sentences,*

- আমি গতকাল তাকে যোগাযোগ করতে ভুলে গেছি। (I forgot to contact him yesterday)
- আমি গতকাল সকালে তাকে যোগাযোগ করেছি।(I contacted him on yesterday morning)

Here, four words (আমি, গতকাল, তাকে, যোগাযোগ) are similar in both sentences textually but not semantically.

Example 2, *Let us keep in mind another two Bengali sentences*,

- আমার একটি দুইচাকা আছে । (I have a two wheeler)
- আমি সাইকেলটার মালিক। (I own the bike)

Here, both sentences are similar in a semantic manner but not textually (the single word is similar on both sentences textually).

In such a case, the document search in huge databases are more difficult because of its high volume and variety of documents [25]. The work presented in this paper provides an efficient semantic search based scheme for retrieving the relevant text or sentences from various Bengali text documents. For retrieving the relevant Bengali text, here, word embedding clustering and pre-trained word embedding modules are included. The major contributions of this paper are as follows:

➢ The major intention of this proposed Bengali QA system tries to contribute a way of finding passages within a paragraph and answering exact questions in relation to the paragraph given with relevant descriptions provided within the paragraph. As such, this model tries to save time who are looking for precise answers from any form of literature which may otherwise be very time consuming to review systematically.

➢ An effective high level feature representation based Bengali question answering system is developed using the combined features of character, pre-trained and affix level word embedding.

➢ Analysing the benefits of Bidirectional long short-term memory (Bi-LSTM) model for character level deep feature extraction to obtain highly sensitive feature level representation which extremely assists for precise answer retrieval.

➢ Development of efficient question answering model based on DBN (Deep Belief Network) for the Bengali language.

➢ We experimentally evaluate our system based on TDIL dataset which outperforming than other Bengali baseline system models and achieving promising results against the SQuAD translated Bengali language dataset. Our results report that combining three-word embedding features which enhance the entire performance of the question-answering module.

The remainder of this paper is structured as follows: Section 2 discusses the recent literature works related to QA system. Section 3 explains the proposed deep learning model based Bengali QA system. Section 4 provides the implantation outcomes of Bengali QA system. The conclusion and future direction of proposed Bengali QA system is discussed in Sect. 5.

2 Literature review

This section presents the certain recent literature works carried out to develop aquestion answering system,especially inthe Bengali language, along with other global languages.

2.1 Question answering system models in global languages

In recent times, there have been numerous studies explored for English language QA system based on the two versions of SQuAD datasets. The multilingual question answering (MQA) system which was common for both English and Hindi language presented by Gupta et al. [14] using lexico-semantic similarity of sentences guided by graph based model. Another MQA system was developed by Carrino et al [4]. for Spanish language using Multilingual-BERT model based on SQuAD translated dataset.

For low level speaking languages collecting massive datasets are often improbable due to the low frequency native language speakers, lack of expert annotators and high cost of datasets collection with precise labels. However there has been limited works explored for non-English languages, Noraset et al. [31] presented an automatic QA system for Thai language named as WabiQA, a bi-directional LSTM model was employed to read the documents and find user answers. Further, Mozannar et al. [28] presented an Arabic QA system by the aid of hierarchical TF-IDF model and pre-trained BERT (Bidirectional Encoder Representations from Transformers) model. Efimov et al. [12] presented a Russian QA system based on two machine learning based base line models and BERT model. Cui et al. [5] presented a Chinese QA system by the help of BERT based base line model. Korean question answering system developed by Lee et al. [21] where they translated English data like SQuAD 1.1 and BERT for transfer learning in training their own QA systems. Prasad et al. [33] performed the question–answer retrieval action on Bengali and Tamil language text messages based on C4.5 decision tree and Naive Bayes algorithm. At the pre-processing phase, several processes had done like, tokenization, case folding, and filtering. In the feature extraction module, a different set of features are collected. It includes TF-IDF score and word embedding features. The retrieval of answers are done by combing the obtained features. Ahmed et al. [2]. proposed an enhanced question retrieval system based on SQUAD English language dataset that can sense users intents associated with the previous question retrieval [33].

The works from [7, 11, 19, 23, 36] tried with hybrid classifiers for language processing. The results shows that the final classifier among the combined classifiers performs well than single classifiers. Both theoretical [6, 18] and experimental [13 32] research were performed based on the above mentioned hybrid approaches.

2.2 Question answering system models in Bengali languages

Currently, there is a significantrise in the research for the Bengali language in the area of Natural Language Processing (NLP). The QA model becoming one of the trending research topics due to the urgency of computational linguistics improvement.

Kowsher et al. [20] had developed a Bangla Informative Question Answering Model based on mathematical and statistical procedures. Here, the lexical information available on sentences were identified for improving the performance of the system. Here, the sentence alignment process is carried out with two steps, translation lexicon and word matching. Initially, the translation lexicon is created for matching the translated words by using the Google web translator plugin. In word matching phase, they conducted a various set of process includes various matching, scoring, and translation. Moreover, the accuracy of the system is enhanced by utilizing the lexical matching.

Monisha et al. [27] have developed a Question answering retrieval method for Bengali language text with latent semantic analysis (LSA). Here, the authors initially developed a document representation for processing the question answering module, and it is created with term-by-sentence matrix. Once they completed the matrix creation process, they applied the Singular Value Decomposition. It excludes the unimportant sentences by extracting the important sentences. Finally, all the sentences are ranked, and the answers were generated with LSA.

Banerjee et al. [3] proposed a Bengali question answering model based on lexical and semantic analysis. The first step is lexical analyser, it separates the input sentences into various tokens and the content sensitive grammar (CSG) were analysed. The second step is data dictionary, and it includes various POS tagging data. It tags the POS for all the Bengali tokens. The third phase is rule generation; here, grammar for the parse of tokens was created. The fourth step is a parser, and it created the parse tree by using CSG on tokens that generated with rules on the previous step. The fifth step is a semantic analyser, and it estimates the semantic similarities available on the generated parse tree. The final step is evaluator, and it evaluates the retrieved answer of the input text.

Manna and Pal [24] presented a Bengali QA system based on semantic analysis and word net. The main motive of this work is to provide a question answer retrieval method with Bengali text. This method includes two steps, features representation and classification. In the feature representation phase, the labelled data is created on the retrieval phase, and the corresponding retrieved answers are verified and ranked with SVM. For the simulation of this Bengali QA system, TDIL (Technology Development for Indian Languages) corpus is utilized.Apart from these techniques, we have presentedan intelligence question answering system of Bengali language with the aid of deep learning techniques.

2.3 Problem definition

Getting a query or question Q in natural language the system should represent in the vector space model. These vectors represents the questions. Hence, a question Qi can be denoted as follows:

$$Q_i = (wi_1, w_{i2}, ...w_i(N-1), wi_N)$$

where wi_k- number of times the system predicts k in question Qi and N-number of terms.

(a) Determine the class C of the Q. That means the system should categorize the question into any one of the eighty-six classes.
(b) Return the answer string A. A should be the Exact Answer (EA) or A should contain EA.

3 Proposed methodology

The growth of online users are increasing rapidly, for retrieving the specific contents they need an efficient search scheme. Mostly, the search schemes are available of languages like English and Chinese. The traditional search tools use keyword based algorithms for retrieving the text. It never analyse the textual information deeply with the meaning of the user query. This work presents an efficient searching scheme for retrieving the specific Bengali text content based on the user need. It analyses the text database more semantically based on the users query. The proposed semantic search model for Bengali language text is given below in Fig. 1.

In this model, it gets the input Bengali text and initiates pre-processing steps on both the inputted query and the corpus (Bengali text database). It tags the POS for the Bengali text data that already pre-processed. Based on the tagged data, the best sentences available in the corpus will be extracted with inverse filtering. Here, every sentence with a better relation with a query is obtained by global word representation. This process accelerates the performance of the system more significantly. After extracting the best sentences, similarity among user query and the extracted sentences will be estimated with knowledge based measure called Resnik similarity. It measures the similarity based on the meaning of Bengali text more deeply. It provides similarity scores among every extracted sentence and the input query. The answers ranking module compare answers to each other by placing them in order of preference. An average ranking is calculated for each answer choice which permits to quickly evaluate the most preferred answer choice. The obtained scores are ranked with page ranking method, and the top scored sentences are shown as a result. It the user is not satisfied with these results, then the DBN (Deep Belief Network) module will be enabled and the inverse filtering process will be re-initiated. The DBN module will be enabled by using fuzzy principles.

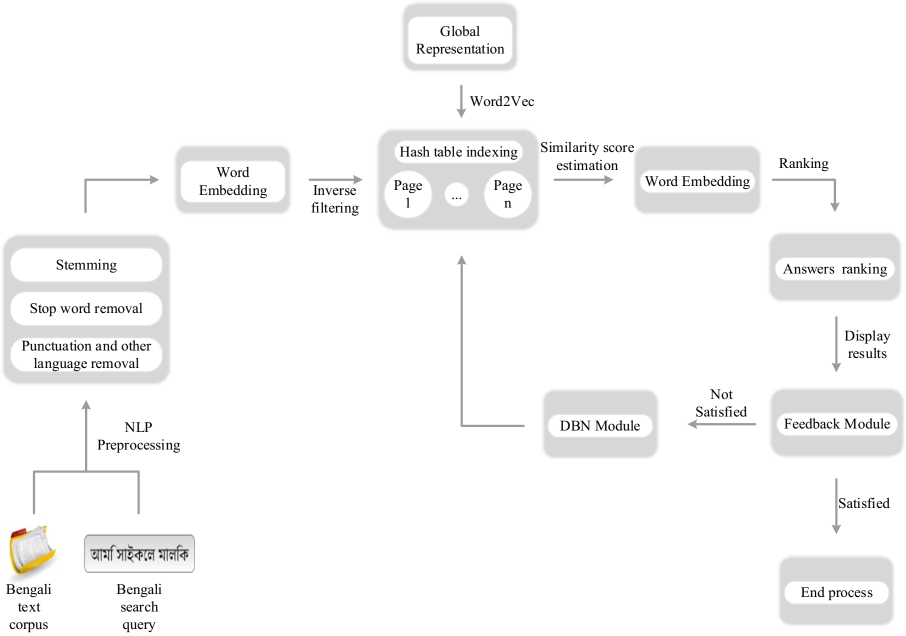


Fig. 1 Proposed Bengali semantic search based QA model

3.1 Pre-processing of Bengali text

The pre-processing of text helps to clean the corpus by eliminating the unwanted text, and it makes the system to process more accurately. Here, for cleaning the Bengali data, punctuation removal, other language word removal, stop word removal and stemming were carried out. Before initiating these steps, all sentences available in the corpus will be separated. The conducted pre-processing steps are given as follows:

3.1.1 Punctuation removal on Bengali text

In this step, different symbols that are included in the Bengali text were eliminated. Here, some of the punctuation symbols like !"#$%&’()* + ,-./:;< = > ?@[\]^_`{|} ~ were removed.

3.1.2 Other language text removal on Bengali text

In this process, other language texts that are present in the Bengali documents are eliminated. From the analysis carried out in Bengali corpus, it identified a huge amount of English language texts are available in the corpus. So, such other word text are eliminated in this step.

3.1.3 Stop word removal on Bengali text

During this process, various words that occur periodically on sentences are removed. For example, এই (this), করি (do), কি (what) and so on. This process is done with dictionary based method and the utilized stop word list includes more than 350 words. Based on this, the stop words will be removed.

Example for stop word removal on Bengali sentence

Input: ইহার ফলে কারখানার কর্মদক্ষতা বৃদ্ধি পাইবে (This will increase the efficiency of the factory).

Output: কারখানার কর্মদক্ষতা বৃদ্ধি পাইবে (Factory efficiency will increase).

Here, the difference among the inputted original text and the stop word removed output sentence is described clearly.

3.1.4 Stemming on Bengali text

At the process of stemming, various suffixes of Bengali words will be removed. In Bengali language, similar words may represent the various lexicon orders. However, because of the root of Bengali text, it never makes change a lot on the meaning of words.

Here, during this process, Bengali suffix words like ই, ছ, ত, ব, ল, ন, ক, স, ম, লা, তা, ছি, বে, তে, ছে, লে,ছি, ছে were eliminated from Bengali words in every sentence.

Example of stemming word removal on Bengali text:

Input: ইহার ফলে কারখানার কর্মদক্ষতা বৃদ্ধি পাইবে

Output: ইহা ফল কারখানা দক্ষ বৃদ্ধি পাই

The result obtained on stemming and the stemming process is clearly described in the above example.

3.2 Word embedding clustering

Normally, the adjacent of every word in the large document is semantically related to each other. Therefore, clustering techniques can be used to find semantic groups. Though the number of semantic groups is priorly unknown, and the size of word embedding vocabulary is relatively large. Especially, the word to vector representation of publically available Bengali documents comprises a massive range of words. In order to manage this issue, we integrate the modified density peak based fast algorithm to execute word embedding clustering.

The clustering process is performed based on two parameters centroid selection and similarity measurement. Initially, the centroid will be selected from the workload randomly. After that, the Euclidean distance is calculated, which follows kernel-based similarity Measure for all data points. After calculating the distance, the local density point will be grouped and create the cluster. The local density points of data is computed based on Gaussian kernel which replaces the basic cut-off kernel.

The mathematical formulation of Gaussian kernel is follows:

$$K(\vec{y}_i, \vec{y}_j) = \exp\left(-\frac{\left\| \vec{y}_i - \vec{y}_j \right\|^2}{2\sigma^2} \right), \sigma > 0. \tag{1}$$

The distance between two data points $\vec{y}_i$ and $\vec{y}_j$ is calculated during the clustering, as follows:

$$d_{i,j} = \left\| \varphi(\vec{y}_i) - \varphi(\vec{y}_j) \right\| = \sqrt{2(1 - K(\vec{y}_i, \vec{y}_j)).} \tag{2}$$

In Eq. 1 $\vec{y}_i, \vec{y}_j$ denotes the two data points and $K(\vec{y}_i, \vec{y}_j)$ represents the kernel function of two data points, σ is a constant value. The mean value for every cluster that will be estimated and based on this mean value, the centroid in moved along the graph.

3.3 High sensitive feature representation using deep learning model based word embedding

In this phase, the major intention is to develop a deep learning model based word embedding for high sensitive feature representation for Bengali language. The conventional neural networks based techniques provided reasonable results over the past decade. But, they are failed to bagging the consecutive data because the present state is disturbed by its earlier states. Moreover, the entire inputs and outputs are independent of each other. The prior words in the sequence are must require to predict the succeeding word in a sequence. Currently, deep learning model provides an encouraging solution for succeeding word retrieval through a sequence to sequence learning process. In the Bengali language, a word frequently comprises of different morphemes owing to their clinging features. Therefore, it is not necessary to take the entire word as an input unit. Here, affixes are considered at starting and ending of the word as secondary features for semantic answers retrieval. The pre-trained fine-tuned feature representation is obtained from DBN module based on the combination of handcraft (TF-IDF) and high level deep features (character-word-affix level).

3.3.1 Global word representation

The global word representation is obtained from the combination of character based Bi–LSTM word embedding, TF-IDF based Pre-trained word embedding and similarity based affix level embedding's. The word vectors is presented in the pre-trained word embedding models for entire words available in the training data. The character-level features are obtained through Bi-LSTM network model with different kernel sizes. The pre-trained word embedding model is generated the word-level features. The final one is the affix-level feature which relates the similarity of query and text. The final word vector representation is created through the concatenation of all these features.

3.3.2 Bi_LSTM based character level word embedding

In semantic QA system, the system combined character-level features of the word is more informative than handcraft features. Hence, Bi-LSTM network is employed effectively to capture the sub word information.

Given a word $'w'$ consist of $'m'$ characters $c_1, c_2, c_3, ...c_m$, where $c_i \in V_c$ is the set of vocabulary character's. Let $C_1, C_2, C_3, ...C_m$ be the vectors that encode characters. $c_1, c_2, c_3, ...c_m$ existent in $'w'$. The embedding of characters are generated by matrix–vector formulation as below

$$C_1 = W_c V_c \tag{3}$$

where W_c is the embedding matrix, $W_c \in R_{d_c * |V_c|}$ and V_c is the one-hot vector model of a specific character. $'d_c'$ is a hyper parameter related to the dimension of the character embedding. Therefore, every word is translated into a sequence of $C_1, C_2, C_3, ...C_m$. Further, the Bi-LSTM network is executed to generate the word embedding vector.

The formula given below gives the LSTM cell process where the input gate inputs times t related to the output results h_{t-1} at the earlier state. At the current moment, the input x_t makes decision of updation. By current input data and output results of hidden layer LSTM cell at an earlier state, the currentvalue of candidate memory cell is predicted. The state value of a memory cell C_t is regulated by both the present candidate cell $\overline{C_t}$ and its individual state C_{t-1} and the forget gate and input gate in the current moment. The output gate O_t is calculated to switch the memory cell status value. Equation (4) gives the last cell's output. Character $*$ is the element wise matrix multiplication W is the weight and b bias of neuron both are obtained through training.

$$i_t = Sigmoid(W_i . [h_{t-1}, x_t] + b_i) \tag{4}$$

$$f_t = Sigmoid(W_f . [h_{t-1}, x_t] + b_f) \tag{5}$$

$$\overline{C_t} = \tanh(W_C \, . \, [h_{t-1}, x_t] + b_c) \tag{6}$$

$$C_t = f_t * C_{t-1} + i_t * \overline{c_t} \tag{7}$$

$$O_t = Sigmoid(W_O \, . \, [h_{t-1}, x_t] + b_o) \tag{8}$$

$$h_t = O_t * \tanh(c_t) \tag{9}$$

The sequence data is processed by standard LSTM cell. Meanwhile,the time series datafrequently discardsupcoming context information. Every training sequence consists of backward and forward LSTM network layers. This is the basic idea of B-LSTM. The word encoding is done by forward LSTM from beginning to end, and the backward LSTM layers encode the opposite way. Hence, at the time t the state of hidden layer Bi-LSTM is gained by addition of weights of the backward hidden layer state $\overleftarrow{h_t}$ and forward hidden layer state $\overrightarrow{h_t}$ and the detailed formula is as follows:

$$\overrightarrow{h_t} = LSTM\left(x_t, \overrightarrow{h_{t-1}}\right) \tag{10}$$

$$\overleftarrow{h_t} = LSTM\left(x_t, \overleftarrow{h_{t-1}}\right) \tag{11}$$

$$H_t = w_t \overrightarrow{h_t} + v_t \overleftarrow{h_t} + b_t \tag{12}$$

where,v_t, w_t represents the weight reacted to the backward hidden layer state $\overleftarrow{h_t}$ and forwarded hidden layer state $\overrightarrow{h_t}$ related to the Bi-LSTM hidden layer state and at the time t the bias corresponding to the hidden layer is represented as b_t. The hyper-parameter settings of Bi-LSTM are given as follows:

- Hidden layer size = 200
- Maximum number of iterations = 50
- Early stopping = 20
- Dropout = 0.25
- Optimizer = Adam
- Batch size = 50
- Initial learning = 10^{-2}

3.3.3 Pre-trained word embedding

The main idea behind this Pre-Trained word embedding is to extract the relevant sentences from Bengali corpus with the POS tagged data of sentences by using TF-IDF (term frequency-inverse document frequency). By estimating the TF-IDF score for all Bengali sentences with its tags, it exposes the grammatical similarities present among the sentences and the query. Also, it enhances the accuracy of the system with the reduced textual context. For estimating the TF-IDF score for POS tagged data, it needs to estimate the term frequency (TF) and inverse document frequency (IDF) of POS tagged data. The TF describes the frequency of tagged POS of a query in a particular sentence. The relevance of tagged POS of a query in a particular sentence is described by IDF. Based on the estimated TF-IDF score, the sentences will be indexed in the hash table. The sentences with non-zero TF-IDF score will be included in the hash table, and others are excluded. The score obtained for every sentence are used as a key to retrieve those sentences.

3.3.4 Affix level word embedding

Affix level word embedding follows knowledge based similarity estimation method called Resnik similarity measure. During this similarity estimation, it describes more knowledge based features present among the query and the hashed sentences. Rather than knowledge based measure, there are various measures also available for estimating the similarity among the texts. Some of them are text based measures, content based measures, feature based measure, and structure based measure. These measures estimate the similarity based on the text, structure, features and contents present in words. But it fails to estimate the similarity based on the meaning of the words. The knowledge based similarity measures will estimate the similarity based on the affix level of the text.

3.4 DBN module

This module makes the system to identify the best Bengali sentences from already obtained results for satisfying the users. This process will provide a more accurate result that more related to query (based on its domain). The DBN (Deep Belief Network) presented here includes multiple RBM (Restricted Boltzmann Machine) layers [24], and it is shown in below Fig. 2.

From above Fig. 3 the DBN model initially gets the TFIDF and concatenated high level features as input model it on the first hidden layer next to it. Then the modeled data will be passed to other hidden layers and the process will continue until completing all iterations (randomly based on the inputted words in Bengali sentences).

During the modelling of DBN, it leans the model parameters θ of various RBMs and defines both the distribution through the visible vectors x and hidden vectors hi. The RBM is defined as,

$$R(x, hi|\theta) = -\sum_{i=1}^{n} bsixi - \sum_{j=1}^{m} bsjhij - \sum_{i=1}^{n}\sum_{j=1}^{m} xiwijhi_j \tag{13}$$

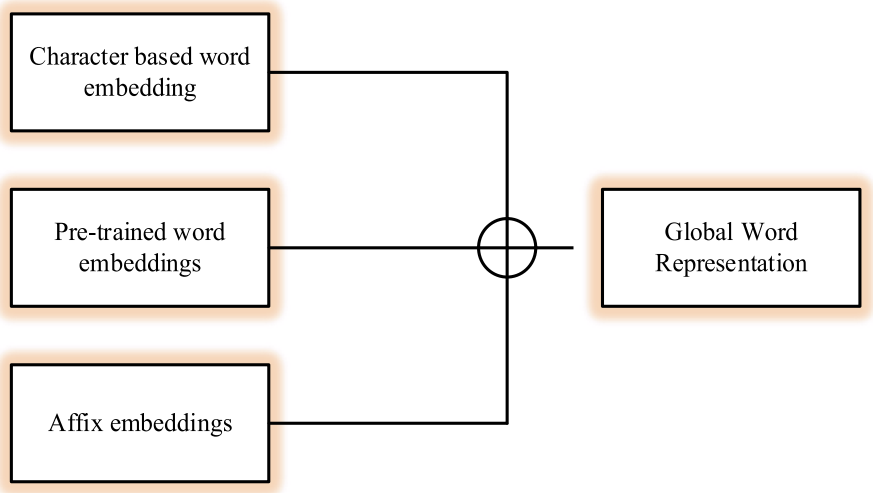

Character based word embedding
Pre-trained word embeddings
Affix embeddings
Global Word Representation

Fig. 2 Global Word Representation

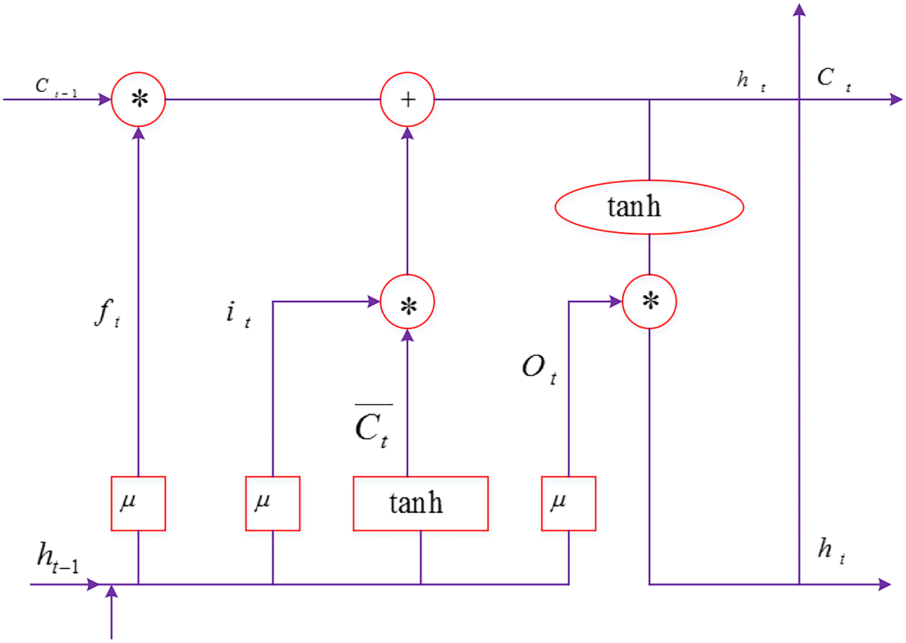


Fig. 3 LSTM structure

From Eq. 7, the model parameter θ includes bs bias and w weight. The bias of visible and hidden units are described as bs_i and bs_j. The number of elements in visible and hidden vectors are represented as n and m.

The probability distribution of visible and hidden vector are defined as,

$$P(x,hi|\theta) = \frac{e^{-R(x,hi|\theta)}}{k(\theta)} \tag{14}$$

$$k(\theta) = \sum\nolimits_{x,hi} e^{-R(x,hi|\theta)} \tag{15}$$

The normalizing factor is represented as $k(\theta)$.

The likelihood of probability distribution is given by,

$$P(x|\theta) = \frac{1}{k(\theta)} \sum_{hi} e^{-R(x,hi|\theta)} \tag{16}$$

The value of model parameter θ is estimated by increasing the logarithmic function to fit it with the input training data and it is given by,

$$\frac{\partial \log P(x|\theta)}{\partial \theta} = \sum_{t=1}^{T} \left[\left(\frac{\partial(-R(x^t, hi|\theta))}{\partial \theta} \right)_{P(hi|x^t,\theta)} - \left(\frac{\partial(-R(x, hi|\theta))}{\partial \theta} \right)_{P(x|hi,\theta)} \right] \tag{17}$$

The iterations of RBM is represented as T.

Then, back propagation procedure is originated for fine tuning the parameters of the entire network through Jaya optimization algorithm. The number of layers in the set is defined as l, and the objective function can be denoted as:

$$s\left(w_l, w_{i,k}|_{k=1}^{l-1}, b_{i,k}|_{k=1}^{l-1}\right) = \underset{w_l, w_{i,k}, b_{i,k}}{\arg\min} \frac{1}{2N} \sum_{i=1}^{N} |\, y_i - g_l(f_l(h_i^{l-1}))|^2 \tag{18}$$

where, the hidden layer activation value $(l-1)^{th}$ is represented as $h_i^{l-1} = f_{l-1}(f_{l-2}(\cdots f_1(x_i)))$, the label of x_i is represented as y_i, the weight of the final layer is represented as w_l, and the bias and the weight of the k^{th} layer is represented as $b_{i,k}$ and $w_{i,k}$. Moreover, the parameters of DBN were reconstructed through the below equations:

$$w_l := w_l + \Delta w_l = w_l - \mu\, d^l\, h^{l-1} \tag{19}$$

$$w_{1,k} := w_{1,k} + \Delta w_{1,k} = w_{1,k} - \mu\, d^k h^{k-1} \tag{20}$$

$$b_{1,k} := b_{1,k} + \Delta b_{1,k} = b_{1,k} - \mu \sum_{j=1}^{R} d^k \tag{21}$$

where, $d^l = (h^l - Y)\, h^l(1-h)$, $d^k = w_{k1}^t\, d^{k+1}(1-h^k)\,(when\, k < l)$, the learning rate is specified as μ and the size of inputted data is denoted as R. The epochs of $w_{1,k}$ and $b_{1,k}$ are accomplished up to the objective function attains the max-epoch. The necessity of training in DBN is to recognize the optimum weight parameters to minimize the objective function (18), where Jaya optimization algorithm [17] is employed to train the model. The working process with DBN is given as follows.

Algorithm for classifying the relevant Bengali sentences is given below:

Input: Top ranked Bengali sentences.

Output: Classified Bengali sentences relevant to the query.

Step 1: Initialize the no. of iterations and top ranked Bengali sentences.

Step 2: The inputted sentences are read and scaled based on the no. of input sentences and features in the form of two dimensional array,x$\{\mathrm{N_{is}}\}\{\mathrm{N_f}\}$.

Step 3: Initialize the no. of RBM (here, three RBMs were used for experimentation).

Step 4: Initialize weight w and bias b for all RBMs.

Step 5: Train the DBN.

Step 6: Update w and b while completing every iteration.

Step 7: Display the obtained relevant Bengali sentences.

Step 8: Stop the process.

3.5 Retrieval module

In the retrieval module, get the query input from the user and initially perform pre-processing as well as closed word embedding feature representation are derived as per the query data. The inputted query input for the retrieval module is expressed as B_R.

The obtained value of query input data (B_R) is matched with the trained input data (B_T).

The valuation of query matching is expressed as follows

$$B_T == B_R$$

If $B_T == B_R$ is satisfied, further move to the next stage. Else, the document is placed to the initial position $B_R = 0$ where this stage the status of the document is not found.

4 Results and discussion

This section explains the implementation done on proposed Bengali question answering system and describes the obtained results of the system. The performance evaluation of the QA system is based on three metrics namely accuracy, precision, recall and f-measure.

4.1 Dataset description

According to our knowledge, Bengali languagedoesn't have any QA corpus for research study. So, here publicly available TDIL dataset based Bengali documents are collected and human annotated questions and answers are created for performing question and answering system evaluation.

a) ***TDIL Bengali corpus***

This corpus includes 25 domain (totally 250 questions), and which are encoded in UTF-8. The domain list and no. of documents are given as follows: Accountancy (9), Agriculture (11), Anthropology (13), Astrology (2), Astronomy (2), Banking (2), Biography (31), Botany (5), BusinessMaths (11), Chemistry (8), Child Literature (60), Comp. Engineering (2), Criticism (18), Dance (3), Drawing (12), Economics (18), Education (15), Essay (40), Folk Lore (31), Games Sport (21), General Science (13), Geography (15), Geology (5), History Arts (20),and Home Science(13).

b) ***SQuAD translated Bengali corpus***

Stanford Question Answering Dataset (SQuAD) comprises a large-scale reading comprehension dataset collected in English with annotations from crowd workers.The dataset contains around 100 k QA pairs from 442 topics. For each topic there is a set of passages and for each passage QA pairs are annotated by marking the span or the part of text that answers the question.

Google cloud translation API was used to translate the context, question and answers of SQuAD 2.0 samples for 294 articles. We then randomly split the data into 70:30 split for training and testing sets. The 235 training topics consisted of 11,588 paragraphs and 73,812 question answer pairs, while 59 validation topics consisted of 17,607 questions and 2714 paragraphs. Out of 73,812 question answer pairs in the training set 36,067 questions are answerable and 37,745 questions are unanswerable. Out of 17,607 question answer pairs in the test set 8166 questions are answerable and 9441 questions are impossible to answer from the context.

The implementation of presented semantic Bengali QA system is carried out in python with NLTK toolkit.Here, the implementation is divided into two scenarios for estimating the performance of the system with a minimal amount of text documents and a higher amount of text documents. So, in the first scenario there are ten domains taken for experimentation. In the second scenario, all the 250 documents were selected and processed with

Table 1 Keyword Extraction Process

Question	Keywords
কে রনেসোঁর জনক হসিাবে ববিচেতি?	Array ([0] =>কে[1] =>রনেসোঁর [2] =>জনক
বঙ্গেল প্রসেডিন্সেি কবে প্রতষ্ঠিতি হয়?	Array ([0] =>বঙ্গেল [1] =>প্রসেডিন্সেি)
কোনবছর, সাহত্যিে নোবলে পুরস্কার দওেয়া হয়ছেলি রবীন্দ্রনাথঠাকুরকে?	Array ([0] =>কোনবছর[1] =>সাহত্যিে)
কে পশ্চমিবঙ্গ থকেে প্রথম ব্যক্তি যনিি ভারতরত্ন পুরষ্কার জতিছেলিনে?	Array ([0] =>কে [1] =>পশ্চমি [2] =>বঙ্গ [3] =>প্রথম [4] =>ব্যক্তি)
পশ্চমিবঙ্গরে আয়তনভত্তিকি বৃহত্তম জলো কোনটি?	Array ([0] =>পশ্চমিবঙ্গরে
ইংলশিচ্যানলে এ দু’বার সাঁতার কাটা প্রথম মহলিা কে?	Array ([0] =>ইংলশি)

the presented search scheme. The performance of the system is measured in terms of accuracy, precision, recall, and f-measure by without using the DBN and using the DBN.

In Table 1, the entire questions come under the field of computer domain obtained from TDIL dataset. The main purpose of keyword extraction is to form the stemming of words to discover the origin of the word.

In our QA model, the numbering system is considered a semantic feature. Here, if the question type is interconnected with a period, then the proposed QA model will give a preciseanswer. The graphical user interface designed for proposed Bengali QA system model is shown in Fig. 4.

In the question 1 present in below Table 2,while the answer 1 is correct, answer 2 matches the keywords. In question 2, one answer is obtainedcorrect. In case of question 3 first answer is right, and other answers matches the question keywords.

The process of determining precision and recall has been shown in Table 2. The formula utilized for estimating the performance are given below:

$$Accuracy : \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}+\mathrm{TN}} \tag{22}$$

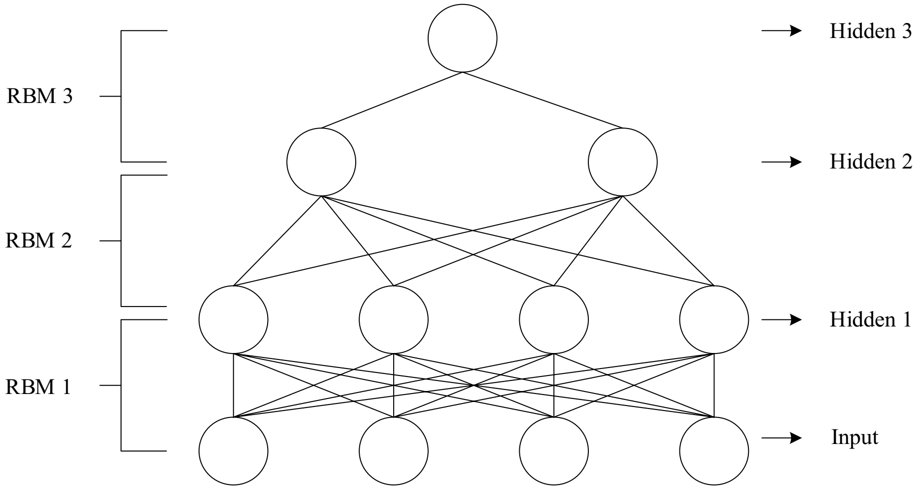


Fig. 4 Utilized DBN model

Table 2 Process for Calculating Precision and Recall

Questions	Relevant Answers Retrieved	Retrieved Answers	Relevant Answers	Precision	Recall
1) কম্পডিটাররে উত্সকী? *Transliteration: Computer er Utso Ki?* *English Translation:What is the source of computer?*	1	2	1	0.5	1
2) বশি্ব কম্পডিটার সাক্ষরতা দবিস হসিাবে কোন দনিটি পালতি হয়? *Transliteration:Biswa Computer sakkhorotadiboshiabekon din tipalito hoy?* *English Translation:Which day is celebrated as World Computer Literacy Day?*	1	1	1	1	1
3) ভারতরে হডিম্যান কম্পডিটার নামে পরচিতি কে? *Transliteration:Varoter Human Computer nam e porichitoke?* *English Translation:Who is known as the Human Computer of India?*	1	1	1	1	1
4) প্রথম কম্পডিটার গমেটি কখন প্রকাশতি হয়ছেলি? *Transliteration:prothom Computer Game tiKokhonprokashito hoy?* *English Translation:When was the first computer game released?*	1	2	1	0.5	1
5) প্রথম কীবোর্ড উদ্ভাবতি হয় কবে? *Transliteration:Prothom Keyboard udhvabito hoy kobe?* *English Translation:When was the first keyboard invented?*	1	1	1	1	1
6) কে ইন্টারনটে আবষি্কার করছেনে? *Transliteration:e internet abiskarkorechhen?* *English Translation:* *Who invented the internet?*	1	2	1	0.5	1

where, TP = No. of retrieved relevant sentences, TN = No. of non-relevant sentences not retrieved, FP = No. of retrieved non-relevant sentences, and FN = No. of not retrieved relevant sentences.

$$Precision : \frac{\text{No. of relevant answers retrived}}{\text{No. of answers retrived}} \tag{23}$$

$$Recall : \frac{\text{No. of retrived relevant answers}}{\text{No. of relevant answers}} \tag{24}$$

$$F - measure : 2 \times \frac{\text{precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{25}$$

TP = True Positive = The predicted answer and the actual answer is same.

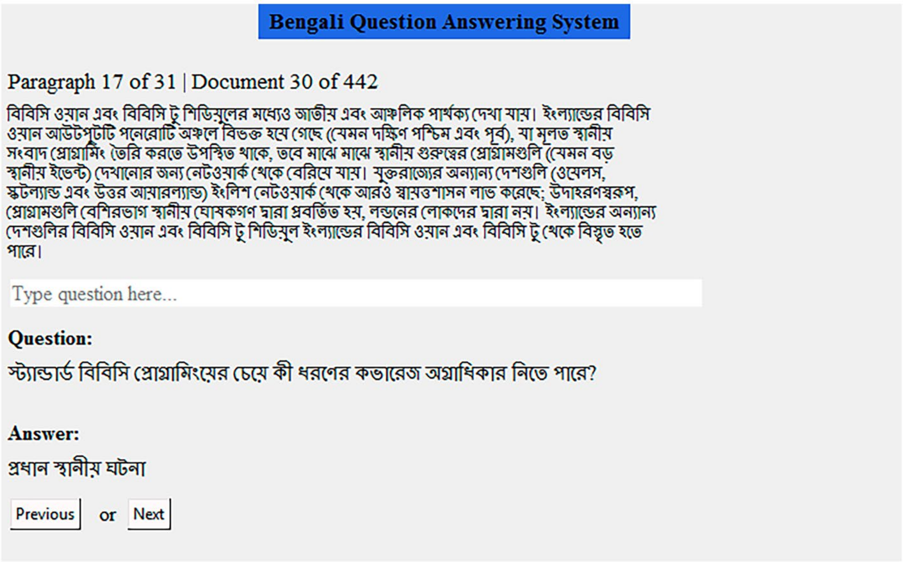
TN = True Negative = The repository does not have the answer to the question and the system predicts that the answer is not known.

FP = False Positive = The repository does not have the answer to the question, but the system predicts an answer.

FN = False Negative = The repository has the answer to the question, but the system predicts that the answer is not known.

4.2 Results on the first scenario

In this scenario, text documents of first 10 domains are trained and classified the Bengali sentences based on the user's query. The obtained results on this scenario on different performance measures are plotted as a graph, and it is described below in Fig. 5.



Bengali Question Answering System

Paragraph 17 of 31 | Document 30 of 442

বিবিসি ওয়ান এবং বিবিসি টু শিডিয়ুলের মধ্যেও জাতীয় এবং আঞ্চলিক পার্থক্য দেখা যায়। ইংল্যান্ডের বিবিসি ওয়ান আউটপুটটি পনেরোটি অঞ্চলে বিভক্ত হয়ে গেছে (যেমন দক্ষিণ পশ্চিম এবং পূর্ব), যা মূলত স্থানীয় সংবাদ প্রোগ্রামিং তৈরি করতে উপস্থিত থাকে, তবে মাঝে মাঝে স্থানীয় গুরুত্বের প্রোগ্রামগুলি (যেমন বড় স্থানীয় ইভেন্ট) দেখানোর জন্য নেটওয়ার্ক থেকে বেরিয়ে যায়। যুক্তরাজ্যের অন্যান্য দেশগুলি (ওয়েলস, স্কটল্যান্ড এবং উত্তর আয়ারল্যান্ড) ইংলিশ নেটওয়ার্ক থেকে আরও স্বায়ত্তশাসন লাভ করেছে; উদাহরণস্বরূপ, প্রোগ্রামগুলি বেশিরভাগ স্থানীয় ঘোষকগণ দ্বারা প্রবর্তিত হয়, লন্ডনের লোকদের দ্বারা নয়। ইংল্যান্ডের অন্যান্য দেশগুলির বিবিসি ওয়ান এবং বিবিসি টু শিডিয়ুল ইংল্যান্ডের বিবিসি ওয়ান এবং বিবিসি টু থেকে বিস্তৃত হতে পারে।

Type question here...

Question:

স্ট্যান্ডার্ড বিবিসি প্রোগ্রামিংয়ের চেয়ে কী ধরণের কভারেজ অগ্রাধিকার নিতে পারে?

Answer:

প্রধান স্থানীয় ঘটনা

Fig. 5 Graphical User interface for Bengali QA system evaluation

The results of this scenario showing that the presented Bengali search scheme earned better results and especially with the use of DBN the system performed very well on retrieving the relevant result for the input query.

4.3 Results on the second scenario

In this scenario, all the documents from 25 domains are trained and Bengali sentences related to the user's query will be retrieved. The results obtained in this scenario is described below in Fig. 6.

From the results with 25 domains, the performance of the system is slightly low when compared with the results obtained with ten domains. Because of more contents, some of the irrelevant results are also displayed with the relevant results. The overall accuracy obtained on every domainwere described in below Fig. 7.

The accuracy of all domains also showing that it can earn better results with minimal textual contents. It is mainly because of the similar grammatical similarity retrieving from different irrelevant sentences. However, there only minimal researches are done with Bengali textual contents; in such case the performance of the presented semantic based Bengali search system is better. Moreover, the performance evaluation of both TDIL and SQuAD translated dataset is shown in Figs. 8 and 9 based on the evaluation measures of accuracy, precision, recall and f-measure.

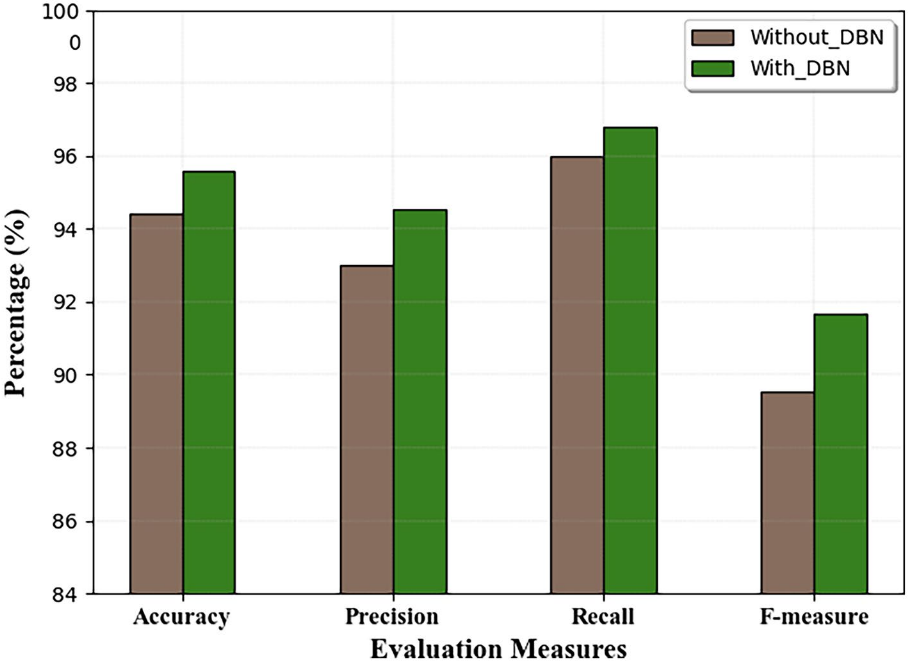


Fig. 6 The results obtained for TDIL (10 domains)

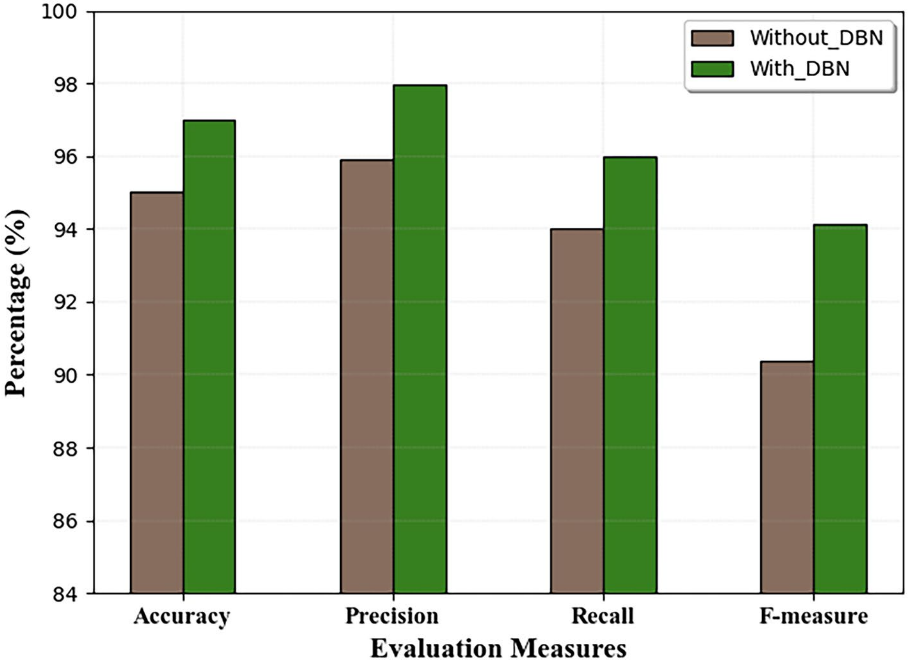


Fig. 7 The results obtained for TDIL (25 domains)

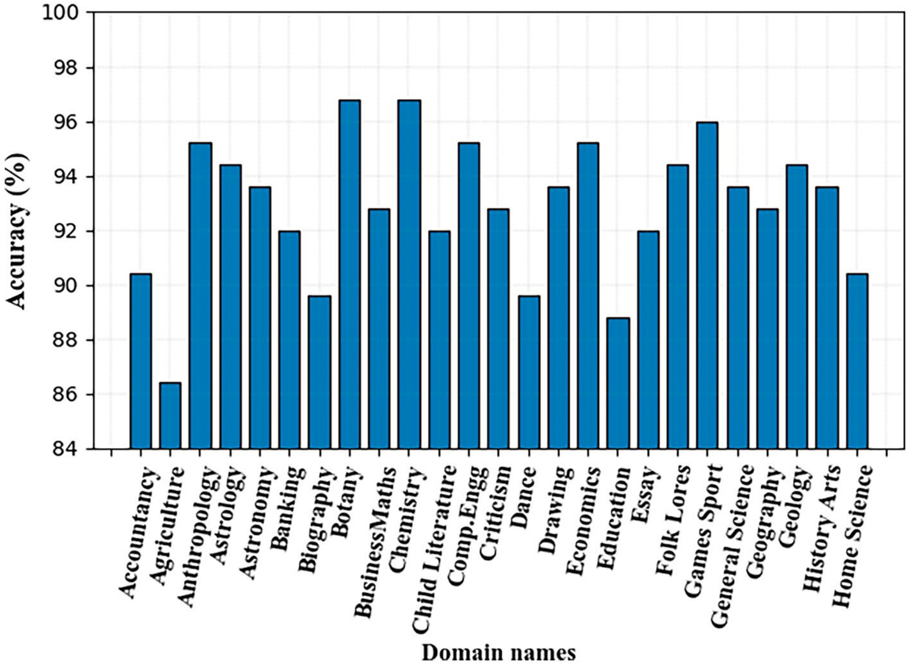


Fig. 8 Domain wise accuracy of TDIL dataset

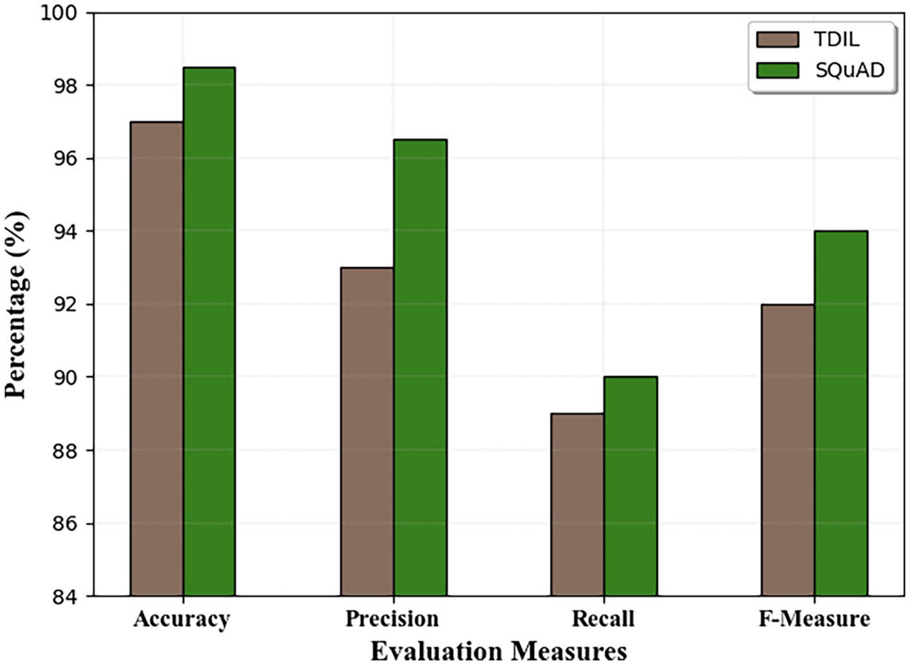


Fig. 9 Performance evaluation of both TDIL and SQuAD dataset

4.4 Impact of high level deep feature based word representation

In order to evaluate the performance of the high level deep feature based word representation, the comparison has been made on various distinct word representation. Concatenating the character level word vector to pre-trained word vector leads to extract the word-level features effectively. Table 3 shows the impact of various word representation on DBN model. This comprises the effectiveness of the Fast Text model on distinct word representation and concatenated word representation.The character based word representation provides supreme accuracy performance compared to the other word representation, which denotes the importance of character-based word representations.

Table 3 Significance of various word representations on DBN Model (TDIL Bengali Corpus) in terms of accuracy

Feature Combination	Accountancy	Banking	Chemistry	Dance	Economics	Geography
FastText	**92.5**	**93.4**	**96.1**	**90.4**	**95.8**	**94**
FastText + char_Bi-LSTM	**95.6**	**94.9**	**96.5**	**91.5**	**96.2**	**94.8**
FastText + char_Bi-LSTM + Affix	**97**	**95.5**	**97**	**93.7**	**96.9**	**95.5**

Table 4 Evaluating the effectiveness of various word representations on question answering tasks by changing the training data size

Training Data Size	**Feature Combination**	**TDIL**	**SQuAD translated**
Complete (100%)	FastText	95.3	96
	FastText + char_Bi-LSTM	96.2	97.2
	FastText + char_Bi-LSTM + Affix	97	98.5
80%	FastText	93	95
	FastText + char_Bi-LSTM	93.8	94.2
	FastText + char_Bi-LSTM + Affix	95.2	96
60%	FastText	91.5	93.2
	FastText + char_Bi-LSTM	92.2	94
	FastText + char_Bi-LSTM + Affix	93	93.7
40%	FastText	90.3	92.4
	FastText + char_Bi-LSTM	91.2	91.8
	FastText + char_Bi-LSTM + Affix	92.5	92

4.5 Significance of training data size

In semantic question answering models, training data size gains the significant importance. Simulations have been performed with various training data size. The effectiveness of the model gets maximizes while the maximization of training data size. Table 4 shows the effectiveness of different word representations on various sizes of the training data. Here initially, an entire training data (80% of the total data) is used for training using various word representations. Further, 20% reduction is made to each of the training data files and repeated the same simulations.

4.6 Comparison analysis with other methods

The comparative analysis has been evaluated based on the Bengali question answering system with respect to other methods and datasets. Table 5 shows the performance comparison of other methods with our proposed method based on accuracy measure.

The comparative analysis has been evaluated based on SQuAD data set along with our SQuAD translated dataset is shown in Tables 6. The proposed Bengali search scheme achieved good accuracy compared to other approaches.

Table 5 Comparison with Bengali Dataset

Author	Language	Dataset	Accuracy
Kowsher et al. [20]	Bengali	Own Created Dataset	93.22%
Monisha et al. [27]	Bengali	Own Created Dataset	90.6%
Banerjee et al. [3]	Bengali	BCSTAT.COM Dataset	87.79%
Manna and Pal [24]	Bengali	TDIL Dataset	80%
Proposed	Bengali	TDIL Dataset	97%

Table 6 Comparison with SQuAD Dataset

Author	Language	Dataset	F1
Lan et al. [41]	English	SQuAD	91.2
Zhang et al. [42]	English	SQuAD	93.2
Arijit Das	Bengali translated	SQuAD	98.5

4.7 Statistical validation: analysis of variance (ANOVA)

The proposed retrieval model has been statistically evaluated based on two important metrics (retrieval accuracy and retrieval error) which are contradicted each other. To validate the statistical significance of the proposed DBN model the well-known statistical evaluation method called analysis of variance (ANOVA) test is used. The ANOVA test is performed on two parameters which is retrieval accuracy value. Then, the outcome of the hypothesis test has compared with the DNN and ANN based models to prove the statistical significance of the proposed model based on input parameters of retrieval accuracy. Normally, an ANOVA test can deliver the insight that whether the null hypothesis (*Hnull*) which states that mean of two or more models for the selected group of samples are similar and therefore null hypothesis statement should be rejected. The parameters utilized in ANOVA test deliver the results in the form of F-statistic. The *Hnull* will be rejected only when the given two statements aresatisfied.

(i) The *p-value* should be less than the significancelevel.
(ii) The value of *F-statistic* must be higher than the *F-critical*value.

Also, the alternative of hypothesis *Halt* can define as Eq. (27) to counter the null hypothesis *Hnull.*

$$Hnull: \ \mu Proposed = \mu DNN = \mu ANN \tag{26}$$

$$Halt: \ \mu Proposed \neq \ \mu DNN \neq \ \mu ANN \tag{27}$$

To conduct ANOVA test, the number of a trails taken five times for validation based on varying the size of training data for all retrieval models. Furthermore, additional critical metrics such as significance level value α = 0.05 and confidence interval (CI) = 95%. The Table 4 have display the input given to perform the ANOVA test for retrieval accuracy output value of the test. Further, with confidence interval = 95%, the output of ANOVA test is listed in view of f-ratio and p-value. After the evaluation of the test outcome listed in Table 7, it can be proved that the difference in the mean value of error has accepted are statistically significant, hence the null hypothesis H_{null} is rejected and accept the alternative hypothesis H_{alt}. Furthermore, in ANOVA test for retrieval accuracy, the f-ratio value is

Table 7 Summary of input Retrieval Accuracy

Models	Trials(N)	Mean	Std. Dev	Std. Error
Proposed	5	94.084	3.055	1.3662
DNN	5	90.984	3.4746	1.5539
ANN	5	88.972	3.1061	1.3891

Table 8 Summary of output Retrieval Accuracy

Source	Degree of Freedom	Sum of Squares	Mean Square
Between Groups	2	66.3178	33.1589
Within Groups	12	124.2149	10.3512
Total	14	190.5327	

3.2034. The p-value is 0.0768. Hence, the ANOVA test result is not significant at $p < 0.05$ and significant at $p < 0.10$. It also gives us the future roadmap to make the retrieval model efficient in term of accuracy and error measure (Table 8).

The *f*-ratio value is 3.2034. The *p*-value is 0.0768. The result is not-significant at $p < 0.05$ and significant at $p < 0.10$.

5 Conclusion

The question answering system presented in this work utilized a different set of process and methods for retrieving the textual contents that are relevant to the query. This system contains different pre-processing steps, POS tagging, inverse filtering, semantic similarity estimation, ranking, and DBN. The Bengali POS tag helps the system to retrieve more grammatical similarity based contents with the inverse filtering method. The obtained best grammatical Bengali contents are utilized for measuring the semantic similarity with input user query. The Bengali textual contents with more semantic similarity will be ranked and displayed as the results of the user's search. Then if the user is not satisfied with the provided results, the ranked sentences will be passed to the DBN module. It classifies the most relevant Bengali textual contents and displays it as the search results. The experimentation is conducted with both minimal and maximum amount of Bengali textual contents. The system acquired up to 95% and 97%, with a minimal amount of contentsby without utilizing the DBN and utilizing the DBN, also acquired up to 94.3% and 95.6% with a maximum amount of contents by without utilizing the DBN and utilizing the DBN. The only accessible QA dataset for Bengali comprises only 250 questions. In future, we plan to increase the size of dataset. Another future direction of this work istrying different models like transfer learning, zero-shot learning, expand our work to cross lingual question answering system.

References

1. Ahmad A, Md. Amin R, Chowdhury F (2018) Bengali document clustering using word movers distance. International Conference on Bangla Speech and Language Processing (Icbslp) 2018:1–6
2. Ahmed R, Al Hasan M, Selim MR (2018) Aligning Sentences In English-Bengali Corpora 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (Ic4me2) 1–5
3. Banerjee S, Naskar SK, Rosso P, Bndyopadhyay S (2019) Classifier combination approach for question classification for Bengali question answering system. Sadhana 44(12). https://doi.org/10.1007/s12046-019-1224-8
4. Carrino CP, Costa-jussà MR, Fonollosa JA (2019) Automatic Spanish translation of the SQuAD dataset for multilingual question answering. arXiv preprint arXiv:1912.05200
5. Cui Y, Liu T, Che W, Xiao L, Chen Z, Ma W, Wang S (2018) A span-extraction dataset for chinese machine reading comprehension. arXiv preprint arXiv:1810.07366

6. Choudhary L (2012) Role of ranking algorithms for information retrieval. Int J Artif Intell Applications 3(4):203–220
7. Chowdhury SR, Sarkar K, Dam S (2017) An approach to generic Bengali text summarization using latent semantic analysis. International Conference On Information Technology (Icit) 11–16
8. Das A, Halder T, Saha D (2017) Automatic extraction of Bengali root verbs using Paninian grammar" in Proc. 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), Bangalore, India, p 953–956. https://doi.org/10.1109/RTEICT.2017.8256739
9. Das A, Mandal J, Danial Z, Pal AR, Saha D (2019) A Novel Approach for Automatic Bengali Question Answering System using Semantic Similarity Analysis. arXiv preprint. arXiv:1910.10758
10. Das A, Saha D (2017) Improvement of electronic governance and mobile governance in multilingual countries with digital etymology using sanskrit grammar, in Proc. IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, Bangladesh, p 502–505. https://doi.org/10.1109/R10-HTC.2017.8289008
11. Dhar A, Dash NS, Roy K (2018) Categorization of bangla web text documents based on Tf-Idf-Icf text analysis scheme social transformation – digital way 477–484
12. Efimov P, Chertok A, Boytsov L, Braslavski P (2020) SberQuAD - Russian reading comprehension dataset: Description and analysis. Experimental IR Meets multilinguality, multimodality, and interaction. CLEF 2020 (Vol. 12260). Springer
13. Grave E, Bojanowski P, Gupta P, Joulin A, Mikolov T (2018) Learning word vectors for 157 languages. In Proceedings of the International Conference on Language Resources and Evaluation (LREC)
14. Gupta D, Ekbal A, Bhattacharyya P (2019) A deep neural network framework for englishhindi question answering. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 19(2):1–22
15. Gupta D, Kumari S, Ekbal A, Bhattacharyya P (2018) MMQA: A multi-domain multi-lingual question-answering framework for English and Hindi. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC)
16. Islam MA, Kabir MF, Abdullah-Al-Mamun K, Huda MN (2016) Word/phrase based answer type classification for bengali question answering system. In 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV) IEEE, p 445–448. https://doi.org/10.1109/ICIEV.2016.7760043
17. Jaya RR (2016) A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. Int J Ind Eng Comput 7(1):19–34
18. Jiang M, Liang Y, Feng X, Fan X, Pei Z, Xue Y, Guan R (2018) Text classification based on deep belief network and softmax regression. Neural Comput Appl 29(1):61–70
19. Khatun S, Hoque MM (2018) Semantic analysis of bengali sentences. International Conference on Bangla Speech and Language Processing (Icbslp) 2018:1–6
20. Kowsher M, Rahman MM, Ahmed SS, Prottasha NJ (2019) Bangla Intelligence Question Answering System Based on Mathematics and Statistics. In 2019 22nd International Conference on Computer and Information Technology (ICCIT) IEEE 1–6
21. Lee K, Yoon K, Park S, Hwang S (2018) Semi-supervised training data generation for multilingual question answering. LREC
22. Ling H, Wu J, Huang J, Chen J, Li P (2020) Attention-based convolutional neural network for deep face recognition. Multimed Tools Appl 79(9):5595–5616
23. Mahmudand A, Khan M (2007) Research report on Bengla tagset. Brac University
24. Manna PP, Pal AR (2019) Question Answering System in Bengali Using Semantic Search. In 2019 International Conference on Applied Machine Learning (ICAML) IEEE, p 175–179. https://doi.org/10.1109/ICAML48257.2019.00041
25. Marcos-Pablos S, García-Peñalvo FJ (2018) Information retrieval methodology for aiding scientific database search. Soft Computing 1–10
26. MdShajalal, Aono M (2018) Semantic textual similarity in bengali text. International Conference on Bangla Speech and Language Processing (Icbslp) 1–5
27. Monisha STA, Sarker S, Nahid MMH (2019) "Classification of Bengali Questions Towards a Factoid Question Answering System," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019, pp. 1-5. https://doi.org/10.1109/ICASERT.2019.8934567
28. Mozannar H, Hajal KE, Maamary E, Hajj H (2019) Neural arabic question answering. arXiv preprint arXiv:1906.05394

29. Nie L, Zhao YL, Wang X, Shen J, Chua TS (2014) Learning to recommend descriptive tags for questions in social forums. ACM Trans Inf Syst (TOIS) 32(1):1–23
30. Nguyen G-H, Tamine L, Soulier L, Souf N (2018) A tri-partite neural document language model for semantic information retrieval. The Semantic Web 445–461
31. Noraset T, Lowphansirikul L, Tuarob S, Wabi QA (2021) A wikipedia-based thai question-answering system. Inf Process Manag 58(1):102431. https://www.sciencedirect.com/science/article/pii/S0306457320309249
32. Pandey HM (2016) Jaya a novel optimization algorithm: What, how and why? In: 2016 IEEE 6th International Conference-Cloud System and Big Data Engineering (Confluence), IEEE 728–730
33. Prasad SS, Kumar J, Prabhakar DK, Tripathi S (2016) Sentiment Mining: An Approach for Bengali and Tamil Tweets, 2016 Ninth International Conference on Contemporary Computing (Ic3) 1–4
34. Pundge AM, Khillare SA, Mahender CN (2016) Question answering system, approaches and techniques: a review. Int J Comput Appl 141(3):0975–8887
35. Purkaystha B, Datta T, Md. Islam S, Marium-E-Jannat (2018) Layered representation of bengali texts in reduced dimension using deep feed forward neural network for categorization, 2018 21st International Conference of Computer and Information Technology (Iccit) 1–5
36. Sarkar K, Chatterjee S (2017) Bengali-to-english forward and backward machine transliteration using support vector machines. Commun Comput Inf Sci 552–566
37. Soares MA, Parreiras FS (2020) A literature review on question answering techniques, paradigms and systems. J King Saud Univ-Comput Inf Sci 32(6):635–46
38. Srba I, Bielikova M (2016) A comprehensive survey and classification of approaches for community question answering. ACM Trans Web (TWEB) 10(3):1–63
39. Wang L, Qian X, Zhang Y, Shen J, Cao X (2019) Enhancing sketch-based image retrieval by cnn semantic re-ranking. IEEE Trans Cybern 50(7):3330–3342
40. Zhou S, Jia J, Yin Y, Li X, Yao Y, Zhang Y, Ye Z, Lei K, Huang Y, Shen J (2019) Understanding the teaching styles by an attention based multi-task cross-media dimensional modeling. In Proceedings of the 27th ACM International Conference on Multimedia 1322–1330
41. Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R. Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942. 2019 Sep 26. https://arxiv.org/abs/1909.11942
42. Zhang Z, Yang J, Zhao H. Retrospective reader for machine reading comprehension. arXiv preprint arXiv:2001.09694. 2020 Jan 27. https://arxiv.org/abs/2001.09694

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.