Knowledge-Base Optimization to Reduce the Response Time of Bangla Chatbot

Md. Kowsher
Dept. of Applied Mathematics
Noakhali Science and Technology University
Noakhali, Bangladesh.
ga.kowsher@gmail.com

Anik Tahabilder
School of Engineering + Technology
Western Carolina University
Cullowhee, NC-28723, USA
tahabilderanik@gmail.com

Md. Zahidul Islam Sanjid
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh.
sanjidmzi53@gmail.com

Nusrat Jahan Prottasha
Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh.
nuaratjahan1234561234@gmail.com

Md. Murad Hossain Sarker
Dept. of Information and Communication Technology,
Comilla University,
Comilla, Bangladesh
mh6367828@gmail.com

***Abstract*—Chatbots have been very popular in recent years for being able to serve as a customer representative, a language learner and so forth. Long short-term memory abbreviated as LSTM is a ubiquitous artificial recurrent neural network that is frequently being used for the chatbot. Nevertheless, if a user makes the line break of sequence, then it is rare to inform the right information without the impact of the previous sequence. As a result, in case of a help desk chatbot, LSTM is not the best option for taking steps of the right information. On the other hand, mathematical and statistical procedures are prominently useful for providing the proper knowledge without having back the impact of sequence. Still, it takes more execution time to respond. The goal of this paper is to present the optimal chatbot for the lowest execution time and three mathematical and statistical strategies for Bangla Intelligence chatbot in light of information obtained from Noakhali Science and Technology University (NSTU). As the procedures, we have followed cosine similarity, Jaccard similarity, and Naive Bayes classifier. To reduce the response time, we decorated the whole path into a 3-depth tree, such as a question, topic, and answer. We have compared the performance of the selected strategies where the best accuracy was 93.22% using the cosine similarity.**

***Contribution*—This paper presents techniques to reduce the response time of statistical and mathematical Bangla Chatbot.**

Keywords— Bangla Chabot, Bangla NLP, NMF, BLTK, Bangla Corpus

I. INTRODUCTION

A chatbot is considered to be a vital form of communication for the third world countries where there are fewer people with the proper technical skillset for supporting a massive chunk of the population [1]. Recently is has been so popular that it is being used in all the major sectors like e-commerce, medical, education and so one. We have, in the meantime, applied this strategy to optimize the response time of a Bangla chatbot named Dr. apaa.

In this paper, we outlined three ideal models of a chatbot on Bangla language incorporating mathematical and statistical methods as an uninterrupted information retrieval instrument with optimal execution time. LSTM is always an excellent choice for responding to a sequence conversation; nonetheless, when information is being delivered, people always do not follow the aptitude of sequence, and wrong information can be the principal cause of a misinterpretation by the Chabot. To improve the performance of chatbot, we are presenting the Bangla chatbot based on mathematical and statistical methods. Even though it results in high response time, the methodology must search for answers over the whole conversation or information. So, we followed the response time reduction technique with the help of a tree structure using the topic modeling. Primarily, all the information and text documents were accumulated in light of Noakhali Science and Technology University, a public university in the coastal terrain Noakhali of Bangladesh. Our methodology can be divided into four major sections, i.e., data collection, data preprocessing, reducing response time and training the chatbot. We have used a handful of methods like anaphora resolution, cleaning words, stop word removal and lemmatization of Bengali lexemes for preprocessing our corpus. We accomplished the optimization purpose by modeling a 3-dept tree with the affiliation of Non-negative matrix factorization (NMF) in order to extract topic. Finally, with the incorporation of Cosine Similarity, Jaccard Similarity, and Naive Bayes classifier, we maintained the coherence and training porcedures between clients urged inquiries and the chatbot response. The summary of this research can be represented by the following points.

- Presenting three mathematical and statistical methods of Bangla chatbot.
- Focusing on optimization techniques to reduce the response time.
- Comparison of the seltective algorithms.

In the following section II, we have discussed the related work, and in section III, we have described our work methodology

978-1-7281-9331-1/20/$31.00 ©2020 IEEE

including data preprocessing and reducing the response time and training of the chatbot. We have described the experimental results in section IV. Finally, In section V, we have described the conclusion and future works.

II. Related Work

Kowsher, Md, et al. [2] introduced a program that utilizes corpora into chatbot systems. Their program satisfies two major targets, such as generating different versions of the chatbot in different languages and learning a very large number of categories within a short time. Dahiya M. in [3] asserted some techniques used while modeling a chatbot and applications where Chatbots could be turned to advantages. Rahman A. et al. in [4] discussed cloud-based chatbot methodology together with challenges of programming of a chatbot. Kowsher, Md, et al. [5] exhibited a computer performed scheme and program of efficient data delivering using chatbot. Khanna A. et al. [6] studied the formation and performance of a simple A.I. system chatbots along with some drawbacks of the current approach toward AI. Their works also proposed a new theory on machine learning intelligence. Nguyen Q.N. et al. manifested a model that studies the interactions between the user and a chatbot based on self-determination theory in [7]. McIntire J.P et al. [8] asserted a way to passively differentiate between humans and chatbots by studying gross communication and behavioral patterns for avoiding spam and malware. Lee, Ming-Che, et al pursued a scheme for evaluating a chatbot system using different measurement metrics [9]. Their study included training procedures of chatbots for certain user's purposes. Satu M.S et al. implemented an artificial chatting system in [10] for the e-commerce sites in order to guide the customers in handling the site as well as to help them to purchase goods according to their needs. Further, they assemble an AIML knowledge-based system to the chatbot that conducts customer queries using a pattern-matching algorithm. In [11] Anirudha P. et al. modeled a semi-supervised artificial intelligence-based chatbot system for languages not rich in (NLP). Their procedure does not require a vast number of datasets and can generate contextual responses using a language dictionary and regular customer interaction dataset. Ilić S. et al. in [12] constructed a chatbot model that originates humorous, sarcastic, and grumpy responses exploiting a seq2seq dataset of 3000 question-answer instances. Mehrjardi M.S. et al. in [13] employed self-attentional models to create end-to-end-task-oriented chatbots with more efficiency and evaluation scores than recurrence-based models. Csaky R. introduced an architecture proposal of a personalized knowledge-powered dialogue system [14] based on self-play. Unlike these works, we present three mathematical and statistical chatbot techniques and optimal ways so that users can get the correct information in the lowest execution time.

III. Methodology

In this paper, we have implemented three mathematical and statistical-based techniques to develop the Bangla chatbot alleviating the delayed response time. The chatbot, we have developed, can be used for providing frequent information via interrogation conducted by the users. The methodology is partitioned in four segments that are: data collection, data pre-processing, response time optimization and training the chatbot. We utilized NMF and SVD to reduce the response time. We have used training algorithms like Cosine Similarity, Jaccard similarity, and Naïve Bayes to obtain the consistency between the queries and the response. The chatbot needs to be trained before it can respond to any question. Questions and corresponding answers are feed into the chatbot. Then, the provided information is preprocessed to be used with the algorithm. Anaphora, Punctuation removal, stop word removal, lemmatization, synchronous word processing are the important steps of preprocessing. After preprocessing, the word is converted into number by TF-IDF. Then the using NNMF has been used to model the topic. Then the selected algorithms, SVD, Jaccard similarity, and Naive Bayes is implemented to reduce the response time. The workflow of the whole process is shown in figure 1.

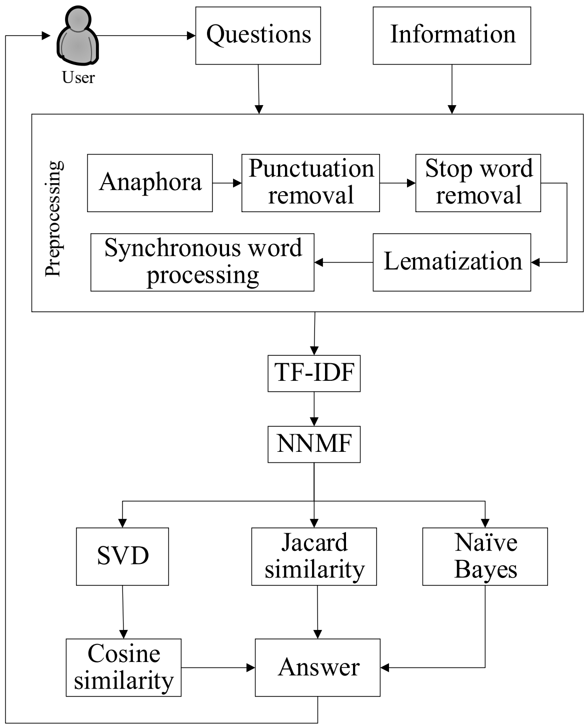

Fig.1. The workflow of response time optimization

A. Pre-Processing

The question and the answering dataset from NSTU needs to be pre-processed in order to be compatible with the selective algorithm. We have implemented several data pre-processing technique to make our collected data suitable for the selected algorithms. The first one is Anaphora [15], which replaces the pronoun to its noun because we use a pronoun instead of a noun in our sequence of questions [16].

In order to describe the preprocessing, let's consider the following questions by the user.

বাংলাদেশের সবচেয়ে বড় ছাত্রী হল কোথায় অবস্থিত? (Where is the biggest Bangladeshi female student hall situated?)

The response by the chatbot is: 'নোবিপ্রবি' (NSTU).

After getting the response of the first question, if the user asks the second question to the chatbot using the pronoun as below:

এটির অডিটোরিয়ামের নাম কি? (What is the name of its auditorium?)

Here, the pronoun 'এটি' (It) is replaced by 'নোবিপ্রবি' (NSTU).

In the next stage, we have removed the punctuation mark. In that case the question এটির অডিটোরিয়ামের নাম কি? will be turned into 'নোবিপ্রবি অডিটোরিয়ামের নাম কি' by removing all kinds of punctuation marks such as a semicolon, comma, exclamation points, question marks, etc., since they have no significant for the feature extraction, similarly the first question will be like বাংলাদেশের সবচেয়ে বড় ছাত্রী হল কোথায় অবস্থিত।

In the next stage, we have removed the stop word

A sentence consists of a rush of stop words (অথবা (or), তে (to), সাথে (with),) and these do not provide with any sentimental intuition on informative data. For this consequence, removing unwanted characters and stop words improve the performance.

In that case, the avobe two questions will be

Question-1: 'বাংলাদেশের বড় ছাত্রী হল কোথায় অবস্থিত'

Question-2: 'নোবিপ্রবি অডিটোরিয়ামের নাম'

For grammatical reasons, documents are going to use different forms of a word, for an instance ভালোবেসে (by loving), ভালোবাসবে (will love), and ভালোবাসানো (make loving). Additionally, there are families of derivationally related words with similar meanings. In many situations, it seems as if it would be useful for a search for one of these words to return documents that contain another word in the set.

The goal of both stemming and lemmatization is to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. For instance:

ভালোবেসে (by loving), ভালোবাসবে (will love), ভালোবাসানো (to make loving) => ভালোবাসা (love).

To be computerized the Bangla words and information matching, lemmatization plays a crucial role. There are very few strategies for Bangla lemmatization. After lemmatize the above two ueser questions will be transformed into 'বাংলাদেশ বড় ছাত্রী হল কোথায় অবস্থিত' and 'নোবিপ্রবি অডিটোরিয়াম নাম'

In the Bengali language, some different words or phrases can pinpoint the same meaning. When these happen, those words and phrases are depicted as the synonym of each other. Purchasers might conduct a query using various terms that are not available exactly as written in the database, but their meanings may available in the database by using some other synonyms. In this case, Bangla chatbot may not be able to answer accurately. Therefore, Synonym words processing has a far-reaching impact in Bangla chatbot as well as in Natural Language Understanding (NLU). For instance, here 'বৃহৎ (large)' is a synonym of 'বড় (big)' and 'বৃহৎ'(large) are considered to be equivalent.

After pre-processing the question will look like:

Question-1: বাংলাদেশ বৃহৎ আবাসিক ছাত্রী হল অবস্থিত [Where is the largest female hall situated in Bangladesh?]

Question-2: নোবিপ্রবি অডিটোরিয়াম নাম [What's the name of its auditorium?]

B. Reduction of Response Time

If a user ask a question, then the chatbot has to reply after searching from whole topics, since information does not appear in sequence always. It takes time to find out the desired answer, and the execution time is increased. But response time will be minimized if the chatbot finds out the most related topics based-on on questions, and this technique is called topic classification. Extracting topics is a good unsupervised data-mining technique to discover the underlying relationships between texts. There are many different approaches with the most popular probably being LDA but this work has been focused on NMF. (NMF) is an unsupervised technique so there are no labeling of topics that the model will be trained on. The way it works is that, NMF decomposes (or factorizes) high-dimensional vectors into a lower-dimensional representation. These lower-dimensional vectors are non-negative which also means their coefficients are non-negative. The whole topic modeling system helps to create a 3-depth tree, described in the figure 2.

In this paper, we have classified the information into 74 different topics related to NSTU. To avoid the complexity of calculation, let's consider two plain topics. Topic-1 includes various information on NSTU, and Topic-2 is about departmental information. Both the considered topics contain their respective influential words.

Topic-1 (NSTU): নোবিপ্রবি, অডিটোরিয়াম, ওয়েবসাইট, বিশ্ববিদ্যালয়, হল

Topic-2 (Department): নোবিপ্রবি, ডিপার্টমেন্ট, ক্লাস, চেয়ারম্যান, টিচার

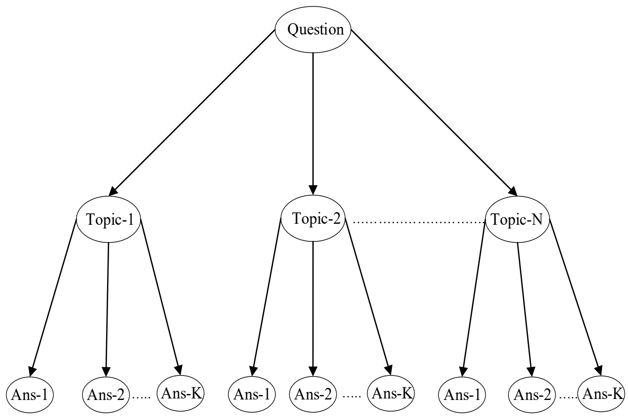

Fig.2. Data decoration

After that we have performed term frequency-inverse document frequency (TF-IDF) information retrieval on topics and question together. Table 1 shows TF-IDF calculation of topics and questions.

TABLE I. TF-IDF CALCULATION OF TOPICS AND QUESTIONS

Terms	TF*IDF		TF*IDF	
	T-1	T-2	Q-1	Q-2
নোবিপ্রবি (NSTU)	0.200	0.200	0.000	0.333
অডিটোরিয়াম (auditorium)	0.2602	0.000	0.000	0.433
ওয়েবসাইট (website)	0.2602	0.000	0.000	0.000
টিচার (teacher)	0.0000	0.2602	0.000	0.000
বিশ্ববিদ্যালয় (university)	0.2602	0.000	0.000	0.000
হল (hall)	0.2602	0.000	0.217	0.000
ডিপার্টমেন্ট (department)	0.0000	0.2602	0.000	0.000
চেয়ারম্যান (chairperson)	0.0000	0.2602	0.000	0.000
ক্লাস (class)	0.0000	0.2602	0.000	0.000

There are lots of existing methods for classifying topics. Among them, we can factorize the matrix into two non-negative matrices W and H as follows: The NMF can be obtained by using the least square method and gradient descent method. We are using the gradient descent method to factorize V into W and H (V = WH).

$$V=\begin{bmatrix} 0.384055 & 0.279394 \\ 0.499655 & 0 \\ 0.499655 & 0 \\ 0 & 0.363513 \\ 0.499655 & 0 \\ 0.499655 & 0 \\ 0 & 0.363513 \\ 0 & 0.363513 \\ 0 & 0.363513 \end{bmatrix} \begin{bmatrix} 0.520759 & 0 \\ 0 & 0.715797 \end{bmatrix}$$

Now we get using W and questions

$$Q_1 = q_1^T W = [0.10825 \quad 0 \quad]$$

and $$Q_2 = q_2^T W = [0.279788 \quad 0.093038]$$

Now we will find the Euclidean distances between user questions and topics.

Table 2 shows the TF-IDF calculation of information and questions.

TABLE II. TF-IDF CALCULATION OF INFORMATION AND QUESTIONS

Terms	Informative Questions (IQ)		Users Questions (Q)	
	IQ_1	IQ_2	Q_1	Q_2
নোবিপ্রবি (NSTU)	0.4337	0.000	0.000	0.433
অডিটোরিয়াম (Auditorium)	0.4337	0.000	0.000	0.433
নাম (Name)	0.4337	0.000	0.000	0.433
বাংলাদেশ (Bangladesh)	0.0000	0.217	0.217	0.000
বিশ্ববিদ্যালয় (University)	0.0000	0.217	0.000	0.000
বৃহৎ (Biggest)	0.0000	0.217	0.217	0.000
ছাত্রী (Student)	0.0000	0.217	0.217	0.000
হল (Hall)	0.0000	0.217	0.217	0.000
থাকা (Exists)	0.0000	0.217	0.000	0.000

$dis(Q_1,T_1) = 0.412334$ and $dis(Q_1,T_2) = 0.723962$. Since, $dis(Q_1,T_1) < dis(Q_1,T_2)$, So user question-1 is related to topic-1. Now, we will show the size or dimension reduction from the presuming questions and information using TF-IDF. Now we will use SVD in matrix C and will find U, Σ and V matrices, where $C = U\Sigma V^T$. Here, the singular value $\sigma_1 = 0.75119$ are not too different from$\sigma_2 = 0.53153$ .So, we cannot avoid σ_2. If $\sigma_1 >> \sigma_2$, then we could avoid σ_2 and it would reduce the singular value matrix size into one dimension.

Rows of V hold the eigenvector values. These are the co-ordinates of individual document vectors. Hence

$$IQ_1 = (-1,0) \text{ and } IQ_2 = (0,-1)$$

To find the new query vector coordinates, we have

$$Q_1 = q_1^T U_k \Sigma_k^{-1} = \begin{bmatrix} 0 & \dfrac{-11073727}{66442375} \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} \dfrac{-150238017}{150238000} & 0 \end{bmatrix}$$

C. Training Bangla Chatbot

In this work, we used three types of mathematical and statistical porcess, such as Cosine Similarity, Jaccard Similarity, and Naive Bayes classifier, to traing the chatbot and make the relationship with the questions and relevant answers. After we made and performance evaluation and compared their performance.

1) Cosine Similarity

After reducing the dimension of the informative questions and user questions, we get using from the last part the vectors of questions and Σ.

Now we will find the cosine similarities.

$$\cos\theta_1 = \mathrm{sim}(Q_1, IQ_1) = \frac{Q_1 \cdot IQ_1}{|Q_1||IQ_1|} = 0 \text{ and}$$

$$\cos\theta_2 = \mathrm{sim}(Q_1, IQ_2) = \frac{Q_1 \cdot IQ_2}{|Q_1||IQ_2|} = 1$$

We can see that $\mathrm{sim}(Q_1, IQ_2) > sim(Q_1, IQ_1)$. So, users question1 can be found in informative question2, i.e., IQ_2.

Since, $\mathrm{sim}(q_2, IQ_1) > sim(Q_2, IQ_2)$, so users question-2 can be found in informative question1, i.e., IQ_1.

2) Jaccard Similarity

From the pre-processed question and information, we can urge the Jaccard similarity to make the best relationship. Now, let an example using set notation and Venn-diagram

P= {বাংলাদেশ (In Bangladesh), বড় (Biggest), আবাসিক (Residential), ছাত্রী (Student), হল (Hall), অবস্থিত (Situated)}

Q= {বাংলাদেশ (Bangladesh), বিশ্ববিদ্যালয় (University), বৃহৎ (Biggest), ছাত্রী (Student), হল (Hall), থাকা (Exists)}

$$J_{index}(P, Q) = \frac{|P \cap Q|}{|P \cup Q|} = 0.33$$

In this way we find, sim (Q_1, IQ_2) $= \frac{1}{3}$ and sim (Q_1, IQ_1) = 0. So, the response of question-1 can be found the information of questin-2. Again, sim (Q_2, IQ_1) = 1 and sim (Q_2, IQ_2) = 0. So, the response of question-2 can be found the information of questin-1.

3) Naive Bayes Experiments

Having done the pre-processing of the data, we have converted every word to the term-document matrix and calculated the probability as

TABLE III. NAIVE BAYES CLASSIFICATION

	Doc	Questions	Class
Training	1	নোবিপ্রবি অডিটোরিয়ামের নাম (The name of NSTU audotorium)	IQ_1
	2	বাংলাদেশের বৃহৎ ছাত্রী হল থাকা (The existance of the biggest female student hall of Bangladesh)	IQ_2
Test	3	বাংলাদেশ বৃহৎ ছাত্রী হল অবস্থিত (The biggest female student hall exists)	Q_1

$$P(\text{বাংলাদেশ (Bangladesh)} \mid IQ_1) = \frac{|0+1|}{|3+10|} = \frac{1}{13}$$

$$P(\text{বাংলাদেশ (Bangladesh)} \mid IQ_2) = \frac{|1+1|}{|6+10|} = \frac{1}{8}$$

$$P(\text{বৃহৎ (Biggest)} \mid IQ_1) = \frac{|0+1|}{|3+10|} = \frac{1}{13}$$

$$P(\text{বৃহৎ (Biggest)} \mid IQ_2) = \frac{|1+1|}{|6+10|} = \frac{1}{8}$$

$$P(\text{ছাত্রী } (Student) \mid IQ_1) = \frac{|0+1|}{|3+10|} = \frac{1}{13}$$

$$P(\text{ছাত্রী } (Student) \mid IQ_2) = \frac{|1+1|}{|6+10|} = \frac{1}{8}$$

$$P(\text{হল (Hall)} \mid IQ_1) = \frac{|0+1|}{|3+10|} = \frac{1}{13}$$

$$P(\text{হল (Hall)} \mid IQ_2) = \frac{|1+1|}{|6+10|} = \frac{1}{8}$$

$$P(\text{অবস্থিত (Situated)} \mid IQ_1) = \frac{|0+1|}{|3+10|} = \frac{1}{13}$$

$$P(\text{অবস্থিত (Situated)} \mid IQ_2) = \frac{|0+1|}{|6+10|} = \frac{1}{16}$$

Now Probability, $P(IQ_1|Q_1) = \frac{1}{2} * \left(\frac{1}{13}\right)^5 = \frac{1}{742586}$

And $P(IQ_2|Q_1) = \frac{1}{2} * \left(\frac{1}{8}\right)^4 * \frac{1}{16} = \frac{1}{131072}$

Since $P(IQ_2|Q_1) > P(IQ_1|Q_1)$ so the replay of question Q_1 can be found in IQ_2. The response of Q_2 can be found in a similar way.

IV. EXPERIMENT & RESULTS

We designed Bangla chatbot hinged on several mathematical and statistical procedures. Our framework and procedures require to be evaluated by a range of experiments. Hence, we elucidated several experiments to compute the accuracy of our model. In this segment, we put forward the questions that we intend to clarify to the experiments and characterize the experimental setup/formation/scheme. Further, we examined the performance and outcome of our proposed work.

A. Corpus

Primarily, we embodied five types of Bengali written texts for the construction of Bangla chatbot. Our first corpus includes 28,324 Bengali root words. With this corpus, we looked forward to the lemmatization of Bengali words. The purpose of our second corpus is the withdrawal of stop words from the incorporated/interpolated documents and questions. There are 382 stop words amassed together to satisfy this goal. Our third corpus contains questions and relevant answers constructed form informational documents we compiled presciently. The Informational documents consist of 74 topics, including hall information, department information, teacher information, library, NSTU nature, bus schedules, etc. of Noakhali Science and Technology University (NSTU). The total number of questions and corrosponding responses on our training dataset was 31027 that covers all 74 topics. For test purpose, we have used almost 6000 questions related to our topics from general people. We uploaded the chatbot in a web based platform and general people participated to test the chatbot performance by asking relevant questions.

B. Experimental Setup and Training

The implementation of our proposed work was performed on Anaconda, a open-source python distribution with Python 3.7. In our task, Anaconda has been used as the apportionment of Python. Anaconda has plenty of open source packages useful for data science and machine learning. Since there is no toolkit to process the Bangla langauge, we are developing a toolkit name 'Ekushey Bangla Lanauge Toolkit'. We also implement the whole procedures of this work in our toolkit.

This toolkit has most of the popular pre-processing steps, Morever, it is occupied with TF-IDF, SVD, cosine similarity, Jaccard similarity, and Naive Bayes algorithm. Ekushey's training process involves loading example dialouge or data into the chatbot's database. For our implementation, we used the Bot Trainer Class of Ekushey to train the bot. At first, we have created a Bangla corpus in the data folder of the Ekushey in the predefined JSON or txt format. A screencopy of the training procedure using the 'Ekushey Bangla Lanauge Toolkit'is shown in the figure 3.

```
from ekushey.bot import trainer
chat = trainer(RemovPunc = 'yes',
               RemovStW = 'yes',
               lemma = 'yes')

chat.training('data.txt')
chat.reply('তুমি কেমন আছ?')
```

Fig.3. Training using Ekushey toolkit

C. Result

As mentioned earlier, we divided the information into 74 different categories. First, the topics with the related questions are trained in order to categorize the topic aiming to reduce the response time then we trained the model using the dataset of 31027questions and corresponding answers related to NSTU. Every training model is separated related to its topic. We fed the chatbot with almost 6000 questions as testing data from selected 74 topics related to Noakhali Science and Technology University (NSTU). For a question, the bot determines the topic, then it goes to a training model related to its topic. For specific the question-answer related to the topic, we obtained 95.32% accuracy in cosine Similarity, 93.46% in Naive Bayes classifier, and 86.91% in Jaccard Similarity. In the action of the topic, the topic recognition got an accuracy of 96.72% for the testing questions. Overall, considering the whole methodology, in response to the question-answer, we obtained 93.22% accuracy in cosine Similarity, 91.34% in Naive Bayes classifier and 82.64% in Jaccard Similarity.

V. CONCLUSION & FUTURE WORKS

Executing Bangla chatbot with minimum response time for necessary information retrieval was one of the principal challenges of this project. The theoretical and experimental methodology of our propounded work has been manifested with a decent accuracy estimation. In this paper, we have demonstrated three procedures for implementing a Bangla Bot mechanism using machine learning, mathematics, and statistics instead of LSTM cause of having some drawbacks for recovering correct information. In addition, we have developed a Bangla chatbot that provides the correct information in the shortest possible time. To establish the whole framework, we sought after a couple of procedures like pre-processing, response time reduction, and setting up the association among information and questions.

Sooner, Bangla chatbot can be exploited in personal, industrial and health sectors in both auditory or textual methods. Bangla bot can be an effective tool for providing education in Bangladesh, where almost 42 million people (26%) are still illiterate [17]. Bangla bot can also be a personal cooking, fitness and fashion instructor and also could be used for entertainment purposes. According to an estimation of the World, Bank Bangladesh is suffering from a deficit of proper healthcare services. There is only one physician for 2,000 individuals. In this situation, an adequately trained Bangla chatbot can alleviate such need by providing primary healthcare information. In the future, our scheme could be made more advanced by incorporating Deep Learning algorithms and train the bot with more sophisticated data.

REFERENCES

[1] Kowsher, Md, et al. "Doly: Bengali Chatbot for Bengali Education." 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). IEEE, 2019.

[2] Kowsher, Md, et al. "Bangla Intelligence Question Answering System Based on Mathematics and Statistics." 2019 22nd International Conference on Computer and Information Technology (ICCIT). IEEE, 2019.

[3] M. Dahiya and S. Gill, "Detection of Rogue Access Point in WLAN using Hopfield Neural Network," International Journal of Electrical and Computer Engineering (IJECE), vol. 7, no. 2, p. 1060, Jan. 2017.

[4] A. M. Rahman, A. A. Mamun, and A. Islam, "Programming challenges of chatbot: Current and future prospective," 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2017.

[5] Kowsher, Md, et al. "Bengali Informative Chatbot." 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2). IEEE, 2019.

[6] A. Khanna, B. Pandey, K. Vashishta, K. Kalia, B. Pradeepkumar, and T. Das, "A Study of Todays A.I. through Chatbots and Rediscovery of Machine Intelligence," International Journal of u- and e-Service, Science and Technology, vol. 8, no. 7, pp. 277–284, 2015.

[7] Q. N. Nguyen and A. Sidorova, "Understanding user interactions with a chatbot: A self-determination theory approach," Am. Conf. Inf. Syst. 2018 Digit. Disruption, AMCIS 2018, pp. 1–5, 2018.

[8] J. P. Mcintire, L. K. Mcintire, and P. R. Havig, "Methods for chatbot detection in distributed text-based communications," 2010 International Symposium on Collaborative Technologies and Systems, 2010.

[9] Lee, Ming-Che, et al. "Study on emotion recognition and companion Chatbot using deep neural network." MULTIMEDIA TOOLS AND APPLICATIONS (2020).

[10] D. Gordon, A. Kembhavi, M. Rastegari, J. Redmon, D. Fox, and A. Farhadi, "Iqa: Visual question answering in interactive environments," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4089–4098.

[11] A. Paul, A. Haque Latif, F. Amin Adnan, and R. M. Rahman, "Focused domain contextual AI chatbot framework for resource poor languages," J. Inf. Telecommun., vol. 3, no. 2, pp. 248–269, 2019.

[12] S. Ilić, R. Nakano, and I. Hajnal, "Designing dialogue systems: A mean, grumpy, sarcastic chatbot in the browser," arXiv Prepr. arXiv1909.09531, 2019.

[13] M. S. Mehrjardi, A. Trabelsi, and O. R. Zaiane, "Self-Attentional Models Application in Task-Oriented Dialogue Generation Systems," arXiv Prepr. arXiv1909.05246, 2019.

[14] R. Csaky, "Proposal Towards a Personalized Knowledge-powered Self-play Based Ensemble Dialog System," arXiv Prepr. arXiv1909.05016, 2019.

[15] R. Mitkov, Anaphora resolution: the state of the art. Citeseer, 1999.

[16] I. Q. Sayed, "Issues in anaphora resolution," Stanford, 2003.

[17] V. Sharma, N. Kulkarni, S. Pranavi, G. Bayomi, E. Nyberg, and T. Mitamura, "BioAMA: towards an end to end biomedical question answering system," in Proceedings of the BioNLP 2018 workshop, 2018, pp. 109–117.