Chatbot for Education System

Aishwarya Hajare, Priyanka Bhosale, Rasika Nanaware, Guruswami Hiremath
aishwaryahajare12@gmail.com, priyankaarbhosale@gmail.com, rasikananaware8@gmail.com, guru.h121@gmail.com
B.E. Computer Dept.,
AISSMS IOT, Kennedy Road, Pune

ABSTRACT
The purpose of this paper is to develop an automated system which gives reply to user query on behalf of human for education system. It can give answer to each and every query asked by end user. Existing Chatbot such as Facebook chat bot, WeChat, Natasha from Hike, Operator, etc. were giving reply from its local database. But our approach is to focus on local database as well as web database and also to make system scalable, user friendly, highly interactive. Various techniques such as neural network (RNN), NLP, pattern matching data mining algorithms are used in this paper to enhance the performance of the system

Keywords
Pattern Matching, RNN, NLP, Web mining, Deep learning, Response generation.

1. INTRODUCTION
The development in the information technology and communication has made artificial intelligent systems more complex. The AI systems are approaching human activities such as taking a decision at a particular moment, performing day to day tasks. In an artificial intelligent field, there are some hybrid methods and adaptive methods available which are making systems more complex. Not only that, but also there is a hybrid combination of natural language processing and intelligent systems. These systems can learn themselves and renew their knowledge by reading all electronic articles available on the internet Human as a user can ask to the systems like usually ask to another human. These systems are often known as internet answering engines. In addition to the internet answering engines, currently many applications are introduced such as chatter-robot or known as Chatbot which is often aimed for giving automatic reply or just for entertainment. This application's work is very simple, because the knowledge is already programmed in advance. Few of the methods used in this application are pattern-matching, natural language processing, data mining. The Chatbot would match the input sentence from the speaker or user with that pattern existed in the knowledge base. Each pattern is then compared with the knowledge of Chatbot. This knowledge has been taken from various sources. Rest of the paper is organized as follows. Section2 describes the existing systems section 3 describes proposed system, section 4 describes text Classification used in Chat Bot and section 5 describes the result followed by conclusion and future scope.

2. SURVEY DETAILS
By using artificial Intelligence field, we can develop many applications one of that is mentioned in this paper is chat bot for the education system. Though chat but has a variety of fields like marketing, medical, education, banking and finance. We will see related work in each field.

Kyo-Joong oh & Dong Kun Lee [1]
Medical field chatbot is useful for providing free counseling to the patient, suggesting various medicines, etc. Even without going to hospital or visiting with the doctor. Kyo-Joong oh & Dong Kun Lee in [1] have proposed a chat bot psychiatric counseling in mental health care service. Emotion recognition is used to implement this chatbot. Because of this, the patient is not needed at all to go to hospital. Free counseling is provided at the user’s location.

Nikita Hatwar [2]
Nikita Hatwar in [2] proposed another interesting chatbot for marketing field named as AI based chat bot. In this paper, this chat bot is used for providing guidance to the visitor of the mall, provides navigation and according to latest discount going in the shop, it was guiding recommendation system is implemented in this project according to the rating we can advise people to go to a particular shop. This feature can be extended further.

Prof. Yu Wo and Gongxiao Wang [3]
In the marketing field, the user has lots of questions related to projects whose answer is not present even in local database, to overcome this problem Prof. Yu Wo and Gongxiao Wang in [3] presented paper automatic chat bot knowledge acquisition from online via rough set and ensemble learning. The concept of ensemble learning is used here. The which constructs classification results of the learner to get the final result. Multiple rough set, classifier is constructed and trained first, then all replies are classified with these classifiers. The final results are drawn from voting to the output of these classifiers and finally it is selected as knowledge database. The disadvantage of this system is that not all replies are related to root message different forum have different styles and formats.
Naveen Kumar and Linga Chandar P.C [4]
The education system also has chat, but Naveen Kumar and Linga Chandar P.C in [4] have proposed an android based educational chat bot for visually impaired people. Voice processing, symbolic reduction and keyword detection methodologies are used. In case of noisy environment application is not recommended. Application should be scalable, so even if thousands of queries are asked at a time. Then also system is able to answer all the queries.

Juan Carlos, et al [5]
Juan Carlos, et al [5] proposed system that is chatting with an Arduino platform through the Telegram boy. The proposed system is connectingtelegram and Arduino platform using Telegram boy. This system allows people to communicate with machine user, where this machine user can operate simple and complex prototype developed by the same person. To implement this system, telegram apes are used. REST protocol is used for the integration of both.

Jacques Garcia Faust Abraham [6]
For Social networking field Jacques Garcia Faust Abraham have proposed and developed a Chatbot in a social network in [6]. This paper describes how to implement chat bot on twitter for entertainment and viral advertising using database and simple algorithms. Message reception, message passing, message processing techniques is used. With this system a person can earn more followers every day without any human intervention.

Wenzeng Zhu [7]
In social networking security is a main concern to implement this functionality. Wenzeng Zhu in [7] proposed a system that is Internet security protection for IRC based Botnet many works contain infected data which can harm our system. So in this paper, they proposed a noble paper based on IRC analysis and detection to protect our computers. IRC servers, used instead of C&C server. So the infected computers link to the IRC server. This method can protect the infected computers not to return Bots controlled by the Bot Master. The only problem with this is that we need to detect all the infected computers and help them to uninstall bot code. To return securely, but the problem is that some boot code is hard to uninstall.

3. Proposed System Architecture
4. Flowchart of the proposed system:

![System Flowchart]

Fig: System Flowchart

Working of the Proposed System:

1. Start

2. The user opens our college web page inside his browser, automatically Chatbot panel will be Pop-Up.

3. The user must have to enter mobile number and email id.

4. Entered users data is sent to OTP server and OTP is generated.

5. Generated OTP is sent to the user through email and mobile number. When the user receives the OTP, the user has to enter that OTP.

6. On success, the browser gives a refresh token server take this token for the user and call refresh token. Here verification process will end.

7. User input the queries in the he chats box, this query(string) is encrypted and sent to chat server.

8. Chat server decrypts the per user queries and look for the tags and keywords through Clustering and Reduction techniques.

9. Create tag and keywords weights for each query.

10. Match tags and keywords with their weights to find top probabilities.

11. If the keyword probabilities are low (not available in dB) then try to find answers for queries by data mining techniques.

12. If a result is found, then return () function will return the answer to user, else send the message as an admin will contact with you by email.

13. Stop.

5. Text Classification used in Chat Bot

5.1 Pattern Matching

Early Chatbot’s used pattern matching [8-10] to classify text and produce a response. This is often referred to as “brute force” as the author of the system needs to describe every pattern for which there is a response.

A standard structure for these patterns is “AIML” (artificial intelligence markup language).

```xml
<aiml version = "1.0.1" encoding = "UTF-8">  
<category>  
<pattern>How college campus is? </pattern>  
<template>Collage campus is very huge. Then also I suggest you to google it. </template>  
</category>  
</aiml>
```

The machine then produces:

Human: How college campus is?

Robot: Collage campus is very huge. Then also I suggest you to google it. The working of Pattern matching is totally based on main keywords. The keywords like placement, cut off, etc.
are extracted from the query and are given for processing.

5.2 Algorithms Used
Naive Bayes [11-14] is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, rounded, and diameter features. For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. One can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

\[ \hat{A}(c) = \frac{B_{c|b}l}{\sum_{b \in V} (B_{b|v} + 1)} = \frac{B_{c|b}l}{(\sum_{b \in V} B_{b|v}) + D} \]

An advantage of naive Bayes is that it only requires a small number of training data to estimate the parameters necessary for classification. A sample training set:
Class: Greeting
“Hi there!”
“How are you?”
“Hello”
Class: Placement
“Which companies are approaching college?”
“What is the salary range?”
“Is TCS, Tech Mahindra to the college?”
Class: Staff
“For computer department how the staff is?”
“What is the salary range?”
“Education of staff?”
Let us classify a few simple input sentences:
Input: “Hi there”
Term: “Hi” (No matches)
Term: “there” (Class: Greeting)
Classification: greeting(score=1)
Input: “Which companies are visiting to the college?”
Term: “Companies” (Class: Placement (2))
Term: “visiting” (Class: Placement (2))
Classification: Placement (Score (4))
By using an equation, we are looking for word matches given some sample sentences for each class, and we avoid having to identify every pattern.

5.3 Neural Network:
The neural network model called RNN Encoder-Decoder that consists of two recurrent neural networks (RNN). One RNN encodes a sequence of symbols into a fixed-length vector representation, and the other decodes the representation into another sequence of symbols. The encoder and decoder of the proposed model are jointly trained to maximize the conditional probability of a target sequence given a source sequence. The performance of statistical machine translation system is empirically found to improve by using the conditional probabilities of phrase pairs computed by the RNN Encoder-Decoder as an additional feature in the existing log-linear model. Qualitatively, we show that the proposed model learns a semantically and syntactically meaningful representation of linguistic phrases.

Fig. RNN

The scenario of neural network of chat both for the education system is drawn and explained in figure 1.3. As input indicates the number of inputs from the user. Consider the scenario for Input1. If the user asked about placement. It is shown with green circle. Blue circle (Hidden Layer) indicates number of classes such as placement of IT department, placement of the computer department, placement of electronics department, etc. The processing is done at hidden layer based on the classes. And the output layer generates the output.
6. Result:

(a) User credentials

(b) Generate OTP

(c) Communication
As shown in fig (a) user enters mobile number and Email id for verification, in fig (b) user enter received OTP if it is correct then communication start which is shown in fig(c). In this way system will produce output. Only the appropriate keywords are extracted from the given query and using data mining algorithm data will be mined from the knowledge based. Sentence will be formed and presented in front of user.

7. CONCLUSION AND FUTURE SCOPE:

The proposed Chabot is an amalgamation of three techniques such as pattern matching, neural network and naive bays algorithm. Because of this, end user will not come to know whether human is giving reply or a system. Queries which are not answered by Chabot are answered by human through E-mail. This system can be extended for banking sector by providing an additional layer of security. In order to decrease query processing time and serve results within a short period of time advanced query processing algorithm can be used.

8. REFERENCES

[2] Prof. Nikita Hatwar “chat bot for marketing field named as AI based chat bot.” In 2016


