

Development of Agriculture Chatbot using Machine Learning Techniques

Prashant Y. Niranjana, Vijay S. Rajpurohit, Rasika Malgi

Abstract- Agriculture data is a main source of country's economic growth. It is important to provide agriculture related information to all the people who are involved in agriculture activities as and when required. This meaningful information is used by people who supply services to agriculture domain and to take some correct decision related to agriculture to apply for their field. The solutions to this problem are given by the efficient interaction of computer with human. Chatbot system provides ability to extract the exact answer to the queries posed by farmers. The proposed system is called as Agriculture Chatbot system or even it is called as Question-Answering system for agriculture domain, where farmer is asking the agriculture related question which fetches the precise answers for the asked questions by farmers in natural language and processes the query using RNN (Recurrent Neural Network) deep learning algorithm to extract correct answer.

Keywords- Chatbot, Recurrent neural network, Deep learning, Natural language, precise answer.

I. INTRODUCTION

In the technological era there has been a drastic growth in the amount of data available in the web. Any type of information can be found using web. As the information increases there is a need to make users aware of the information and the techniques required to explore the data. Search engines are one of the ways of exploring the large available data. Users can fetch the information on just a click. Although this is the mostly used technique for exploring information it has disadvantage, as an answer to the queries asked it provides user a large collection of links and documents from where the reply can be found.

Agriculture is a process of producing fiber, food and other products by cultivating some plants, this practice is also called

farming. There is a significant change in this field from 20th century.

With the growing technology it is necessary for farmers to learn and adopt the changes. Hence aim is to build chatbot system that provides accurate answers to the questions. According to a large research in the field of Chatbot system, there isn't any agriculture specific system to supply exact and quick answer to the queries posed by farmers. To address this situation, the proposed system provides exact answers to the asked questions using RNN (Recurrent Neural Network) deep learning algorithm.

Chatbot system is a major research area in natural language processing, most of the recent question answering system deals with factoid questions some examples are.

What is the crop grown in belgaum? Which pesticide is need for potato?

This paper presents answer the posed questions that are syntactic and semantic with unstructured data. Semantic queries are answered which are without the keywords like What, Which, When etc.

Examples are: Plant diseases? Grown plants?

Chatbot system mainly contains three phases. Question analysis, document processing and answer extraction. Question analysis is the first step, here the user questions are presented in natural language, then processed using stemming, POS tagging and removal of key words. Document processing step fetches equivalent documents that contain particular keywords using different algorithms. In answer extraction phase it extracts the answer and verifies the answer for the correctness then supplies exact answer to the user.

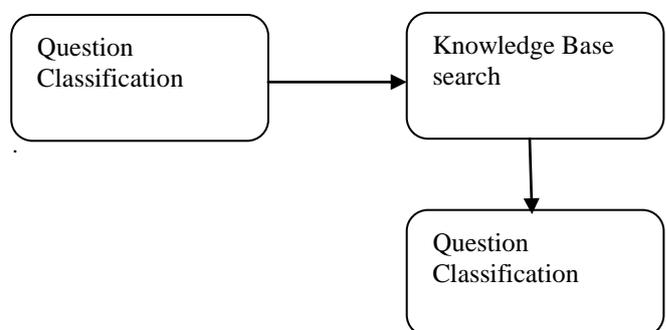


Fig. 1. Flow diagram for chatbot system

Revised Manuscript Received on December 12, 2019.

* Correspondence Author

Prashant Y Niranjana *, Computer Science , KLS Gogte Institute of Technology, Belagavi. Email: prashant053@gmail.com

Vijay S. Rajpurohit *, Computer Science , KLS Gogte Institute of Technology, Belagavi. Email: vijaysr2k@yahoo.com

Rasika Malgi, Computer Science, KLS Gogte Institute of Technology, Belagavi. Email: rasikamalgi@gmail.com

II. LITERATURE SURVEY

When we look at history of the chatbot it started in 1950 by Alan Turing who posed a question “can machines think?” It is known as Turing test. In 1964 a first chatbot was built with pattern matching technique named as ELIZA [1] for human machine interaction. Woods in 1977 gave a system LUNAR for distinguishing and analyzing the soil composition. FAQ Finder system [1] was developed on statistical and semantic similarity, it has a query list with the assistance of knowledge base questions asked are matched.

In 1996 Ask Jeeves was designed to answer every day posed questions by Garrett Gruener et.al. The system Answer bag that responds to the posed question on the ranking technique by combining the internet FAQ’s. Amazon developed A9.com in 2003, it provides online knowledge base for finding a product or advertising [1]. In 2005 Yahoo Answers was built to give responses to the asked queries.

Wolfram Alpha in 2009 was developed to answer queries related to mathematical equations. In 2010 QUORA was designed to research on any particular topic, here user can pose question, read and write answers can also discuss with other users on different topics. In the same year IBM Watson was built for a quiz show. The most common system was developed in 2011 named as Apple Siri, it provides directions and gives exact answer to the asked questions using NLP interface and speech as input. Similar to Siri Google Now was built in 2012. Wael Ghonim designed PARLIO to interact with intellects and share their opinions.

Looking at agriculture as a specific domain ADANS [2] (Agriculture Domain for question answering system) uses NLP and semantic web technologies to provide answers for asked questions. The task is done in steps where preprocessing of the natural language takes place. Here POS tagging, removal of stop words is carried out. In the next step triples are formed. Ontology is built with Protégé tool [2]. SPARQL is a language used to query the system to provide answers to the asked questions.

AGRI-QAS [4] deals with unstructured data and responds to FACTOID questions such as Which, How, What. Here there are predefined rules for pre-processing and post-processing, input is in the form of XML format [4]. Similarly there is a travel domain question answering system which focuses on usage of rule based approach for generation of SPARQL query, then calculates semantic similarity to predict the answer based on the higher similarity [3].

(A) Research Gaps

- SPARQL query is mainly suited for closed environment, it is time consuming and troublesome handling negation statements like not, don’t etc.
- NLP approach is fast as the rules are automatically defined whereas rule based approach has pre-defined rules and there are no rules defined for all the situations hence it is considered weak [1].

- Statistical approach is defined as “bag of words” it fails because it is unable to characterize each term separately and linguistic properties
- Consider machine learning algorithms like decision tree don’t provide great accuracy and efficient performance. KNN has high computational cost, Naïve Bayes has less efficiency when compared to SVM. SVM having higher efficiency when compared to previous stated algorithms, hence it outperformed [3] when compared to other classification algorithms but one drawback is it consumes large amount of system memory.
- Considering feed-forward network having better accuracy but it doesn’t have memory, hence unable to remember previous input to reproduce the output. It can be concluded that usage of deep learning (RNN) algorithm would enhance the system.

III. PROPOSED SYSTEM

- When the user provides input that is a query posed then required entities are extracted and sent for training, RNN sequence to sequence algorithm is used which considers previous output to predict the answer. Next action is to extract the exact answer from the vocabulary of words. Dataset is created assuming some questions asked by farmers, this dataset is in XML format.

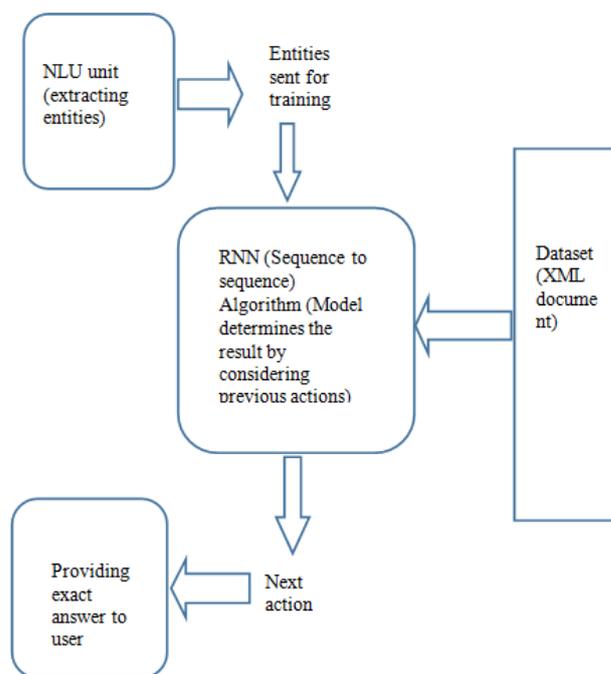


Fig. 2. System Architecture

(A) Dataset Creation

When we choose machine learning techniques to implement any task, first thing to review is the dataset that is necessary to train the model. Conversational logs are needed for sequence to sequence RNN algorithms. The RNN algorithms need to be able to understand the type of responses for every input query. Creating a dataset is a big task, some of the data is taken from the agriculture website and most of the data is collected into XML file assuming the questions that farmers can pose.

```
<category><pattern>WHAT kind of soil suitable for cultivating
dates</pattern>
<template>Sandy loam.</template>
</category>
<category><pattern>WHAT kind of soil good for cultivating dates
</pattern>
<template>Sandy loam.</template>
</category>
<category><pattern>WHAT kind of soil preferable for cultivating
dates</pattern>
<template>Sandy loam.</template>
</category>
<category><pattern>WHAT kind of soil suitable for cultivating
</pattern>
<template>What kind of plant or seed you are looking for.
</template>
</category>
<category><pattern>Types of plant diseases</pattern>
<template>Fungal, bacterial, and viral.</template>
</category>
<category><pattern>plant diseases</pattern>
<template>Genreally plants are affected by Fungal, bacterial,
and viral.</template>
```

Fig 3.2 XML document format

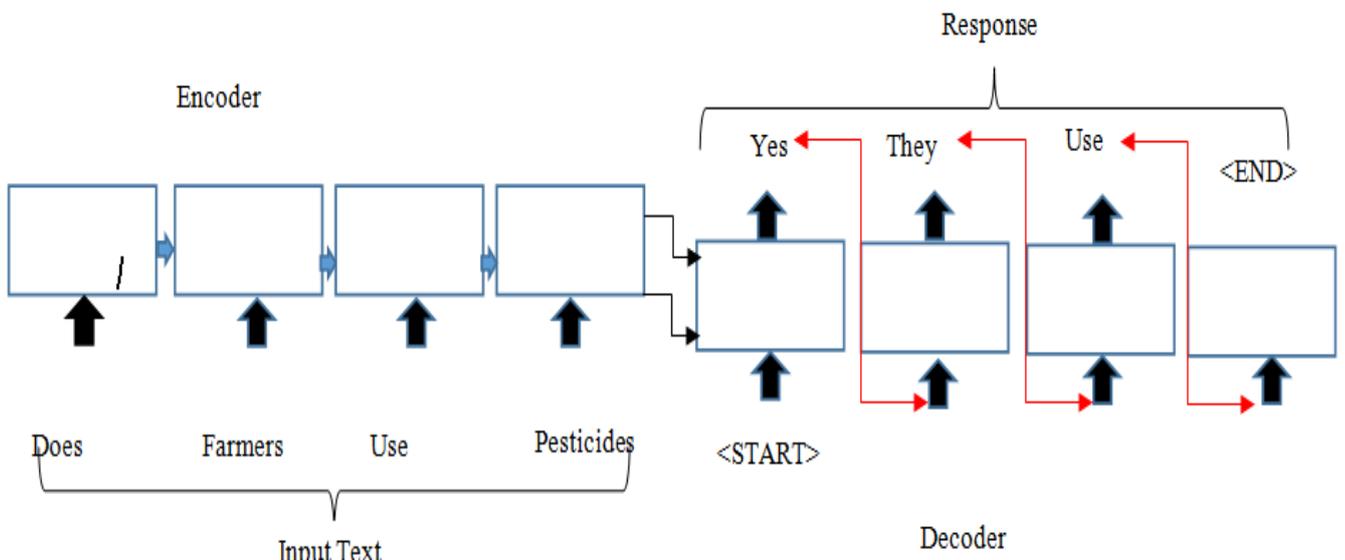


Fig. 3. Working of RNN algorithm

(B) RNN algorithm

RNN encoder and RNN decoder are two major components of sequence to sequence algorithm. Encoder encapsulates the posed input query into the fixed representation, then results into varying text length that best replies to it.

RNN has number of hidden layers, this presents the data from past time steps. For example at fourth time step hidden state vector will be the result of past four words. Finally it can be concluded that the RNN encoder can be represented as a result of whole input text. RNN decoder takes the last hidden

vector of the encoder and utilizes to forecast the words of the output response. First cell's task is to fetch the vector representation and then look into the vocabulary for the word that best suits the reply. Here probabilities for each of the words in the vocabulary are calculated and argmax values are selected. Second cells work is to combine both vector representation and output of past cell. Role of algorithm is to determine conditional probability.

$$p(o_1, o_2, \dots, o_T | i_1, i_2, \dots, i_T) = \prod_{t=1}^T p(o_t | v, o_1, o_2, \dots, o_{t-1})$$

Where i_1, \dots, i_T is the input sequence, o_1, \dots, o_T is the output sequence and v is vector representation. The left part of the equation represents probability of response sequence conditioned on the input string given and right part $p(o_T | v, o_1, o_2, \dots, o_{t-1})$ is defined as vector probabilities of vocabulary of words conditioned on vector representation and response of the past calculated time stamp. The P_i notation is the product, right side can also be represented as $p(o_1 | v) * p(o_2 | v, o_1) * p(o_3 | v, o_1, o_2)$.

Let's look at the example, Does farmer's use pesticides? This is a query asked by user, first we need to think of how it can be answered by common people not machine. Mainly people answer it as yes, yeah and no. Once the network is trained the probability looks like given below.

$P(o_1 | v) = \{but=0.01, no=0.30, yes=0.40, yeah=0.34\}$ hence the response word from the first unit will be yes. Most likely sequence of words are given by P_i operation which we consider as an end result or output.

When compared to other machine learning techniques sequence to sequence RNN algorithm is versatile. The other approaches like SVM, logistic regression all require fixed input size and also results is fixed output size.

IV. RESULT AND DISCUSSION

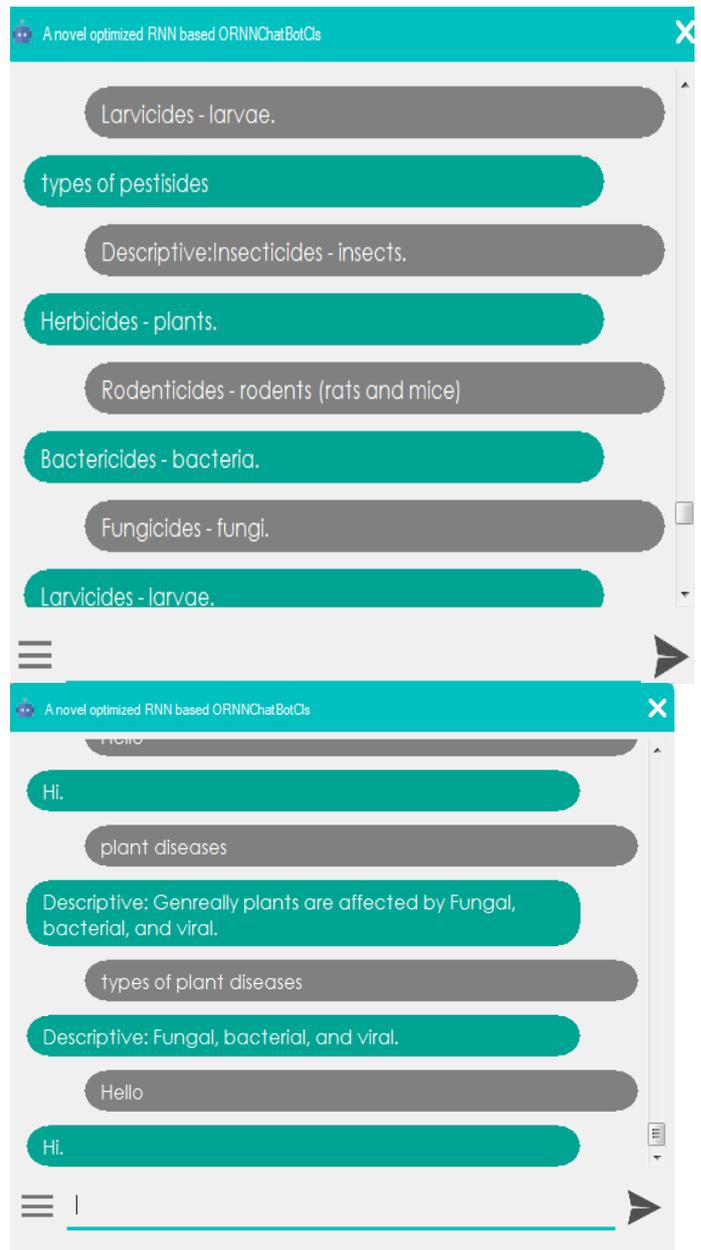
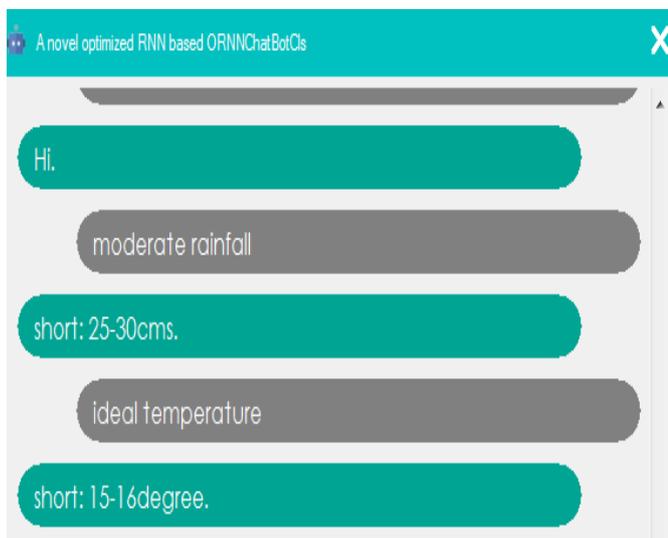


Fig. 4. Results

Fig. 4. Shows the chatbot which classifies the question as short or descriptive type for a given question and also retrieves the required answer for agiven query.

V. CONCLUSION

Agriculture chatbot plays a very important role in agriculture sector to help all farmers and who are involved in agriculture activities which analyses the asked questions and provides the meaning information to the farmer to take correct decisions within time.

REFERENCES

1. Deepa Yogish, Manjunath T N, Ravindra S Hegadi. "Survey on Trends and Methodsof an Itelligent Answering System". International Conference on Electrical, Electronics,



Communication, Computer and optimization Techniques pp 346-353 2017.

2. Manmita Devi, Mohit Dua, "ADANS:An Agriculture Domain Question Answering System using Ontologies". International Conference on Computing, Communication and Automation (ICCCA) pp 122-127 2017.
3. Hasangi Kuhaduwa, Dilshan Pathirana, Pathum Liyana Arachchi, Vishma Dias, Surangika Ranathunga, Upali Kohomban. "Question Answering System for the Travel Domain" Moratuwa Engineering Research Conference (MERCCon) pp 449-454 2017.
4. Sharvari Gaikwad, Rohan Asodekar, Sunny Gadia, Vahida Z. Attar. "AGRI-QAS Question-Answering System for Agriculture Domain". International Conference on Advances in Computing, Communication and Informatics (ICACCI) pp 1474-1478 2017.
5. Sweta P. Lende and M M Raghuvanshi, "Question Answering System on Education Acts using NLP Techniques", IEEE sponsored word conference on futuristic trends in Research and Innovation for Social Welfare 2016.
6. Garima Nanda, Mohit Dua, Krishma Singla, "A Hindi Question Answering System Using Machine Learning Approach" ICCTICT 2016
7. Sanglap Sarka R, Venkatesh Rao, Baala Mithra S M, Subrahmanya V R K Roa, "NLP algorithm Based Question and Answering System", IEEE 2015.
8. A.B Abacha, P Zweigenbaum",MEANS: A medical question answering system combining NLP techniques and semantic web technologies", Information Processing and Management, vol 51, Issue. 5, pp 570-594, September 2015.

AUTHORS PROFILE



Prashant Y. Niranjana, Working as assistant professor in computer Science department at KLS gogte Institute of Technology Belagavi. He Completed his M.Tech & B.E from Visvesaraya Technological university Belagavi, Karnataka India.



Dr. Vijay S. Rajpurohit is a Professor in the Dept. of Comp Science and Engineering, KLS Gogte Institute of Technology, Belgavi, Karnataka, India. He completed his B.E in Computer Science & Engg. from Karnataka University Dharwad in 1997, M. Tech. in CSE from NITK, Surathkal in 2003 and Ph. D. in Robotics and Vision from Manipal University in 2009. He is life member of SSI, ISTE, ISC, CSI, IACSIT and IAENG.

From 1998 to 2010, he was on the faculty at Manipal Institute of Technology, Manipal India and from 2010 to till date he is working as a professor at Gogte Institute of Technology, Belgaum, India. He was a reviewer/technical committee member for 9+ international conferences and journals. He has chaired a session for International Conference on Automation Robotics Control and Vision-2010, at Nanyang Technological University, Singapore in 2010 for which 100% travel grant was sanctioned for him by AICTE Govt. of India.

Rasika Malgi is a master student in the Dept. of Comp Science and Engineering, KLS Gogte Institute of Technology, Belagavi, Karnataka, India.