A Survey on Chat-Bot system for Agriculture Domain

Prashant Y. Niranjan
Dept. Computer Science and Engg.
KLS's Gogte Institute of Technology
Belagavi, India
prashant053@gmail.com

Vijay S. Rajpurohit
Dept. Computer Science and Engg.
KLS's Gogte Institute of Technology
Belagavi, India
vijaysr2k@yahoo.com

Rasika Malgi
Dept. Computer Science and Engg.
KLS's Gogte Institute of Technology
Belagavi, India
rasikamalgi@gmail.com

Abstract - **Chat-Bot system is a part of natural language processing, where it requires system to be trained as per the human language, so that it can satisfy the needs of the user. Agriculture domain is leading source of countries growth. At present farmers are not much aware about recent technologies and practices being used in agriculture field. Extraction of meaningful answer by machine learning techniques is a problem, that has been studied by many machine learning experts as well as advanced machine learning techniques are introduced. These techniques are applied to extract the accurate answer. We can call this as an Agriculture Question Answering System, where the farmer can query the system and the system understands the query and responds to a given query. In this paper, we have reviewed extracting a precise answer for a given question by mainly focusing on machine learning techniques. We have made suggestions and provided the comparative analysis.**

Keywords: Question answering system, Machine learning, precise.

I. INTRODUCTION

Agriculture is the art and science of soil cultivation, it is the key to development of the country. Today modern way of plant breeding, agrochemicals like pesticides, fertilizers and due to increased technological developments our farmers are left behind, so they should take a step ahead. Making them aware regarding latest techniques is necessary.

Chat-Bot is a question answering system which can also be called as human-machine or human-computer interaction. The user queries the machine it should respond with the accurate answers.

Significant research is made on question answering system in various domains such as medicine and travel. Considering agriculture as a specific domain there are systems which are capable of exploring the web data, during this process, retrieving exact answer is not possible. Since the website also leads to certain unrelated answers to the asked queries. To get the exact response we make use of deep learning (RNN) technique. Hence, study is to help farmers query regarding the crops, raw materials used, plants grown in particular area, usage of pesticides and fertilizers etc.

Chatbot system mainly contains three phases. Question analysis, document processing and answer extraction. Question analysis is the initial step, here the user query in natural language is processed using POS tagging, stemming and removal of key words. Document processing step fetches equivalent documents that contain particular keywords using different algorithms. In answer extraction phase it retrieves the answer and tests the answer for the correctness and provides exact answer to the user. This paper is organized as: Section 1 presents the brief introduction about the paper; Section 2 shows table for history of chatbot, section 3 discusses the literature survey of the existing chatbot methods and techniques used. Section 4 provides research gaps involved in the current system and this paper is concluded in Section 4.
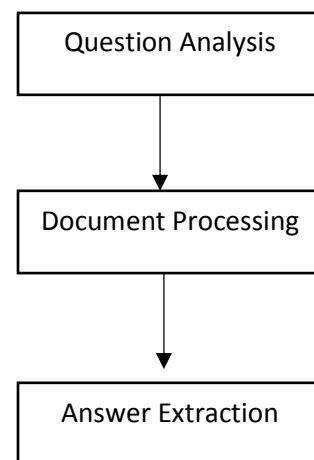

Fig.1 Flow of chat-Bot system

II. HISTORY OF CHATBOT

Development of the question answering system was started in 1950, when Alan Turing threw down a test for machine and proposed a question, "Can Machines Think?" He gave a Turing test which can test if it is a human or a machine. Here human talks to machine using teletype and questions it then the judgments are based on the response given.

System	Author	Technique Used	Year
BaseBall	Geen et al.,	In American league used to respond to queries regarding dates, locations etc., based on pattern matching	1993
ELIZA	Joseph Weizenbaum	Using pattern matching technique build a first chatbot. Used for man-machine communication	1964
LUNAR	woods	Used for comparing and analyzing soil composition and lunar rocks. Mainly for natural language questions.	1977
GUS	Bobrow	To fetch information about airlines	1977
CHATBOT	Michel Mauldin	Used for providing service to customers mainly in dialogue systems	1994
FAQ Finder	Burke et al.	Using statistical and semantic similarity, with the help of knowledge base that contains question list, this KB is matched to the questions been asked.	1995
ASK Jeeves	Garett Gruener et.al.	Response to the everyday asked questions.	1996
START	Katz	Creating a Knowledge automatically	1997
		from internet using unstructured data. Use's queries are answered using the created KB. Latter natural language queries were also answered	
JAVELIN	Nyberg et.al.	With the usage of text corpus it extracts closely related responses to the asked questions.	2002
AnswerBag	Wired Ivy,	Using rating many answers where provided to the asked queries by consolidating internet FAQ's	2003
A9.com	Amazon	The online is Knowledge base for searching a product and advertising	2003
MIT's Jupter System	Zue et al.	To extract worldwide weather data telephonic interaction was developed	2005
Yahoo! Answers	Yahoo	Answers the queries asked by uses and also pose questions to get response.	2005
Blurtit	Blurtit Limited	Group of people answer to the queries submitted by other people using their knowledge	2006
Evi	William Tunstall Pedoe	This was acquired by Amazon. Using smart phones directly responds to the queries asked in natural English language.	2007
Wofram Alpha	Wolfram Research	Mostly for mathematical equations,	2009

		responses are given by formulating replies from external source to the asked questions.	
IBM Watson	David Ferrucci, Thomas J Watson	This was modeled to give responses to the asked question on quiz show.	2010
QUORA	Adam D'Angelo and Charli Cheever	Developed to help students to research on any topic. Users can ask questions, write answers and discuss on various issues.	2010
Apple siri	Allple Inc	Provides response to the asked questions, gives directions and requests are allocated to a group of internet services to perform some actions. This is done using speech as input and NLP interface.	2011
Google Now	Google	Similar to Siri was developed to perform actions, answer queries and make recommendations.	2012
PARLIO	Wael Ghonim	Allowed people to share their opinions and communicate with intellectuals and public figures.	2014

III. LITERATURE SURVEY

Literature survey has been discussed in this section to highlight the work carried out till now in chatbot system. It works mainly under three phases Question Identification, Knowledge Base searching and providing exact answers.

ADANS (An Agriculture Domain for question answering system) represents a response to the question provided in natural language [2]. NLP (Natural Language Processing) and semantic web is used technologies are used. SPARQL is formulated by the system, from the questions represented in natural language [2]. Question Answering system for the travel domain [3] presents research in travel domain that classifies the questions semantically into EAT's (Expected answer type). EAT's found using different machine learning algorithms. AGRI-QAS [4] focuses on processing unstructured data and provides response for FACTOID queries such as 'which', 'what', 'who', 'where'.

(A)Question Identification

In ADANS [2] the work is done in in steps. It initiates with preprocessing of queries where tokenization, stop words removal and POS tagging takes place [2]. After the completion of preprocessing, next step is to form triples, using Stanford dependency tree. Dependency tree provides relationship between words in a sentence. To get the desired triple edges of dependency tree are removed and merged [2].

QA system for travel domain classifies question using four aspects firstly taxonomy, question base, classifies and features [3]. Taxonomy is created with 63 fine classes and 7 coarse classes that matches to an EA and information regarding hotels, flights, restaurants is collected from Trip advisor [3]. Classifier used was SVM with a linear kernel function. Numerical features are only accepted by SVM hence, non-numeric features are represented as bipmap during implementation. Experiment was conducted with Naïve-Bayes, random forest and decision trees. However SVM performed the best among all these techniques [3].

Question processing in AGRI-QAS [4] is defined by specific rules for pre-processing and post-processing. Adding hyphen between consecutive names, replacing two words with a similar word are some pre-processing rules.

(B) Knowledge Base

ADANS [2] uses ontological Knowledge base. Building ontology is done using a tool Protégé, this is widely used for ontology construction. Firstly domain is decided, then different entities and their relationships are identified.

QA system for travel domain generates query by searching in KB. Proprietary ontology is used as KB, this is created using the taxonomy and is still under

development [3]. Out of user query it is essential to derive SPARQL, using rule based approach RDF triples are created. Along with EAT created in the first step, SPARQL question pattern is matched to triples this triple generation works only with statements hence, question to statement conversion is necessary. QuestiontoStatementTranslator is used for this process.

AGRI-QAS [4] takes input as XML documents and grammar parser, POS tagging, entity recognizer are implemented. Instead of tagging words as part of speech, indexing the documents according to domain specific terms is carried out using domain specific named entity recognizer [4].

(C) Answer Extraction

The ADANS system uses SPARQL is a language used for querying data in RDF format [2].

SPARQL query is generated using rule-based technique, stop words are removed from the triples. Then relation list from ontology is extracted using EAT and subject. QA system for travel domain [3] calculates semantic similarity of elements in relation list and highest similarity is considered for answer extraction.

IV. RESEARCH GAPS

According to the survey SPARQL query is used to extract answers [3] [4].Major drawback is, it can't hold dynamic data that gets updated frequently. SPARQL works well in closed environment. It is time consuming and difficult to handle negation statements.

Rule based and NLP technique

Rule based approach [3] is used for text related system and generally used for semi-structured and structured documents. This approached contains predefined rules hence, learning is reduced [1]. There are no rules defined for every situation, so it is also time consuming. In NLP approach automatically rules are defined and this technique has lower precision and high recall.

Supervised machine learning

Various supervised machine learning algorithms are used as classifier [4]. Decision tree is a simple algorithm used for classification, issue is this technique doesn't produce efficient performance and accuracy. K-nearest neighbor has high computational cost and based on the properties of the new training sample that are assumed the developed model is considered same. When compared to SVM, Naïve

Bayes has low efficiency and research is been carried out to rise the productivity. SVM is a supervised machine learning algorithm which is mainly used for classification of the data. Since, it is non-linear and multidimensional it works better in automatically classifying the questions. Hence SVM performed better when compared to all supervised learning classification techniques [4]. One disadvantages of SVM is most of the system memory is been consumed and complexity of the algorithm. Due to this it can be suggested to use neural networks.

Statistical Approach

This technique is called "bag of words". It plays important role in web data and online platforms. Document contains various words that are identified by group of keywords and based on the frequency weight is assigned to each word [1]. Drawback is every term is treated separately and it has failed to characterize linguistic properties for group of phrases and terms.

Suggested Solution

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

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

V. CONCLUSION

Modern technology and large amount of data has made difficult for farmers to receive the exact information online in required time. There are various chatbot or question answering system developed but very few provides correct and efficient answers. Challenge is to give exact answer for the already built chatbot system. A well-developed system has to fetch the user query in form of natural language, using various techniques such as removal of stop words, POS tagging, stemming and using efficient classifier it has to classify user questions to give correct and precise answer. In this paper we have studied various approaches and identified drawbacks. Hence, we can suggest to implement

RNN (Recurrent neural network) approach, this technique contains internal memory and can remember the previous input to predict the next output. Due to its efficiency farmer queries can be answered accurately.

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 Answeing System using Ontologies". Internaltional 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 (MERCon) 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 Raghuwanshi, "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.