International Journal of Latest Trends in Engineering and Technology
Vol.(7)Issue(4), pp.254-258
DOI: http://dx.doi.org/10.21172/1.74.035
e-ISSN:2278-621X

REVIEW: HINDI QUESTION ANSWERING SYSTEM USING MACHINE LEARNING APPROACH

Garima Nanda[1]

Abstract: As an upshot of Natural Language Interface to Database (NLIDB), Question Answering System is relatively an Information Retrieval system which is suppose to reflect the user with the correct or closest results to the query being asked to the system in natural language. Information Retrieval and Information Extraction plays vital role in accomplishing the task of interaction between the user and the system. The paper discusses various ways and techniques of interaction between the user and the system along with different approaches. Machine Learning is one of the approaches which is preferred for the Question Answering System. Out of Supervised and unsupervised learning, supervised learning is taken priority here by going through numerous other techniques.
Keywords: Machine Learning, Classification, Question Answering System, Naïve Bayes.

I. INTRODUCTION

While typing a question in a natural language user expects to get the answer in same language. According to a native user, scenario depicted is somewhat like, a query is given by the user to the system and it replies as short and accurate answer to the user. But is it actually feasible and can happen? The answer to this question is yes! Along with the advancements from last decades, we have entered a stage where effective mechanism has become an essential part of our daily lives. With a lively growth in today's era, search engines are optimized to a great extent. When even a simple two-word query is given to the system, it provides a number of result perspectives. With growth to this extent in documental period, it is quite necessary to be effective and efficient but that too in small amount of time. When a general user asks something on the Search Engine, i.e. when a query is passed, the search engine provide the user a list of relevant possibilities of the answer to the question. Out of which the user has to select which one among them is suitable to him. What the idea of this Question Answering System depicts is when the user will query the system, he will ask the question in simple Hindi natural language and instead of getting a list of relevant documents, he will get an accurate answer that too in one word or being specific to the question.

Number of contemporary information techniques are being offered by Internet these days. As a result of this Internet is suffering from a dilemma of governing, regulating and managing the textual information available there. As a consequence of NLIDBs, Question Answering system is an idea of a user asking a question to the system and getting answer instead of list of relevant documents. The basic idea is when user will query the system in form of a question in Hindi

[1] *Department of Computer Science Chitkara University Himachal Pradesh*

language (as here the natural language considered is “Hindi”), he will be getting an answer to the question accurately and precisely in Hindi language only instead of getting a number of documents and sorting out of them.

Question Answering is one of the most popular and user friendly way of human interacting with computers and the question answering system will enable a user in accessing the information of the system naturally, by asking the question to the system and receiving output as a result. To propose a Question Answering System, many techniques and advancements of Artificial Intelligence, Human Computer Interaction, Information Retrieval and Information Extraction can be individually used or also in combination.

Earlier what used to happen be a user can ask any query and it took a lot of time to get processed and then reflecting the answer to the user. It was not that much feasible from a user’s point of view. Main problems were faced in getting to the system first, if a user was able to access the system then passing a query. Afterwards, when query was passed then it was supposed to get preprocessed and then after so long it used to give output. But recently, in last few decades, a large number of new advancements, new techniques and researches came into existence due to which numerous systems and techniques help the user to get into the same method but with different perspective.

Question Answering System is one of the ideas to be implicated taking user in consideration and that too along with the range of new proficiencies and capabilities. Classification is one the advancing method of Machine Learning Approach which let the idea of Question answering System get accomplished with any of its algorithm. Different algorithms have different results varying on different parameters.

II. RELATED WORK

Computerization of data is going on very incremental rate. Most of the data is online due to numerous advancements in technologies. To make the data accessible to everyone, alleviate then to users, whether they are technical or non technical users, some kind of interface is needed by the users. Information retrieval and Information Extraction are the two important methods being accessed for the same purpose. Number of researchers worked on some or the other kind of interfaces for the natural languages to the databases, and the ongoing research is setting another example for the same. Different platforms are being tested to make the scenario much easy for the user to let his work done. Question Answering system is an idea to make the access of the system much user friendly and efficient to use as it will be going to work in the natural language which is much more easy for the user; both from technical and non- technical background.

A framework of Question answering based on machine learning is explained which subsumes the classifier based on passage retrieval, manageable documents and other questions. It also clarifies that the question answering system is merely tact of revealing the exact answers to the questions being asked by the user over bulk of data collection (Yen, Chiehwu, Lee, Lee, Liu, 2013). A domain independent system is developed, which by using the knowledge base in Hindi natural language basically identifies a domain (Dua, Kumar, Virk, 2013).

Vector Space Model based layout of question answering system is introduced which accepts a document as a vector having a direction and magnitude. By using term vector space in VSM, researchers meant accurate answer can be retrieved out of the given question (Jovita, Lind, Hartawan, Suhartuno,2015).

Abstract of machine learning approach let number of researchers to work on Question Answering system, whether it would be Support Vector Machine, Classification or Clustering. When the research is going on, number of unfavourable situations are to be faced by the research people.

In a schematic way, merits and de-merits of using Machine Learning Approach are reflected in the chart below (Table I).

Table I. Merits and De-merits of using Machine Learning Approach	
Merits	**De-merits**
Feature Learning	*Time Consuming*
Parameter Optimization	*Avoidance of over fitting*
Understanding	*Works with Continuous Loss Functions*
Intelligent Decisions	*Large Data Requirements*
Accurate	*Error Prone*

Machine Learning is simply a tool to play with. It is just important that a user must know when and how the machine learning algorithm is going to be used. Normally, machine learning optimize over an artificial hypothesis.

As far as Machine Learning approaches are concerned, further consider one algorithm of classification for question answering system, let's catch up the following algorithms:

• nearest neighbor

k-nearest neighbor algorithm is used for classification and regression techniques. It is also calld k-nn and is a parametric method for classification. The input is taken as the k-closest training specimen in the feature space in both the techniques i.e. regression and classification.

• ***Naïve Bayes***: It represents a supervised learning method and comes from the probabilistic classifiers as also based on the "Bayes theorem". These are highly scalable. Also this is a statistical method for classification. Here conditional probability is computed and using Bayesian probability terminologies, the simplified formulation of the probability is depicted as following:

Posterior= (prior * likelihood)/ evidence

• *Random Forest*: this is another one of the efficient method of classification which yields with high accuracy and is operated by constructing a multitude of decision trees at the time of training.

• *Decision Trees:* Classification or regression models are created in the form of tree structures in this part. The data set is broken down in small subsets and on the other side the decision tree is incrementally developed at the same time. It also reflects with good results in accuracy, precision and recall.

• *Support Vector Machines*: Another technique of classification is SVM. It configures the concept of creating decision planes that describes and defines the decision boundaries.

III. CONCLUSION

According to the discussion, the proposed system of Question Answering will enable a user to query the system in natural language (Hindi here) and the system will be replying with the concise and specific result of the query being asked. The Question Answering System uses the approach of Machine Learning. Out of numerous ways, Classification will be basically used for training the system. Along with k-means, naïve bayes, many other models and algorithms can be used for classification purposes, but Naïve Bayes which is from the family of just simple probabilistic features is preferred here. Using Naïve Bayes along with the feature of similarity measures, contribute effectively and efficiently in the whole process. Attributes like feature extraction, precision, recall, all will help in concluding with the results by giving vast idea of Question Answering system with Overall Accuracy and threshold of the system.

IV. IMPLICATIONS OF FUTURE RESEARCH

As far as future perspective is concerned, an extensive platform will be established if some other procedure will also merge in this perspective. The Question Answering System will become more inventive, fruitful and efficacious if it will be extended to multilingual by considering other native natural languages like Punjabi, Gujrati, Marathi, Telgu or so. A healthy response will be observed the system by using discrete classification techniques and more interesting results will be observed by doing comparative studies of different techniques.

V. REFERENCES

[1]. Khillare, Shelke and Mahender, C: 2000, 'Comparitive study on Question Answering Systems and Techniques', International Journal of Advanced Research in Computer Science and Software Engineering, 4, 775-778.

[2]. Jovita, Linda, Hartawan and Suhartono, 'Using Vector Space Model in Question Answering System', International Conference on Computer Science and Computational Intelligence (ICCSCI 2015), 305-311.

[3]. Abacha, Zweigenbaum, 2001, 'MEANS: A medical Question Answering System combining NLP techniques and Semantic Web Technologies', Information Processing and Management ScienceDirect, 570-594.

[4]. Liu,Yi,Chen,Song, 2016, 'A survey on frameworks and methods of Question Answering',Information Science and Control Engineering (ICISCE),2016 3rd International Conference,IEEE.

[5]. Dua, Kumar and Virk, 2013.'Hindi Language Graphical User Interface to Database Management System', International Conference on Machine Learning and Applications, IEEE.

[6]. Dwivedi and Singh, 2013, 'Research and Reviews in Question Answering System', International Conference on Computational Intelligence: Modelling Techniques and Applications, ScienceDirect,471-424.

[7]. Khalid, Jijkoun and Rijke,2007, 'Machine Learning for Question Answering from Tabular Data', 18th International Workshop on Database and Expert Systems Applications, IEEE.

[8]. Yen, ChiehWu, Yang, Lee, Lee and Liu, 2013, 'A support vector machine-based context ranking model for question answering', Information Sciences,SciencDirect, 77-87.

[9]. Bagde, Dua and Virk, 2015 'Comparison of Different Similarity Functions on Hindi QA system', ICT4SD, Springer.

[10]. Chaudhary and Battan, 2014,'Natural Language Interface to Databases-An Introduction', International Journal of Advanced Research in Computer Science and Software Engineering,7(4).

[11]. Harper, 2005, 'A review and Comparison of Classification Algorithms for medical decision making', Healthy Policy, ELSEVIER, 71(3),315-331.

[12]. Brick, Dumais, Banko, 2002,'An Analysis of the AskMSR question answering system', ACL-02 Conference on Empirical Methods in natural Language Processing, ACM, 10, 257-264.

[13]. Hirschman and Gaizauskas, 2001, 'Natural Language question Answering: the view from here', 7(4), Cambridge University Press,275-300.

[14]. Dua, Aggarwal, Kadyan and Dua, 2012,'Punjabi utomatic Spech Recognition using HTK', International Journal of Computer Science issues, ISSN: 16940814, 9(4).

[15]. Shen, Liu, Wang, Vithlani, 2016,'SocialQ&A: An Online Social Network Based Question and Answer System', IEEE Transaction on Big Data,99.

[16]. Ray and Shalaan,2016, 'A review and Future perspectives of Arabic Question Answering Systems',IEEE Transactions on Knowledge and Data Engineering, 28(12)3169-3190.

[17]. Xiang, Rong, Shen , Ouyang, Xiong, 2016,'Multidimensional Scaling Based knowledge provision for new questions in community Question Ansering Systems','Neural Network (IJCNN),2016 Joint Conference