Genetic algorithm based Architectural framework for Natural Language Based Question Answering System

Vinutha H . Mukula Chaitanya . Nagashree N

Dr. Ambedkar Institute of Technology, Bangalore
School of Computer Science and Engineering, REVA University, Bangalore.

Received: 23 August 2022 / Revised: 21 September 2022 / Accepted: 12 October 2022

Abstract – A natural language Question Answering System (QAS) in contrast to traditional keyword search systems does not return a complete document to the user. Instead, users ask a question in natural language and receive a specific answer in return. However, most of the existing methods do not consider the profile of user who is asking the question. The availability of information in different formats, languages and different levels of granularity makes information retrieval difficult to answer the asked question. This paper proposes a framework using Genetic algorithm based optimizer in which the information about the user (asking the question) will also be considered while answering the question.

Index Terms – NLP, Question answering system, genetic algorithms, framework, machine learning.

I. INTRODUCTION

Natural Language Question Answering is the task of giving a direct (i.e. a precise answer and not a list of documents) to the natural language query/question asked by the user[1]. Natural Language Processing (NLP) is a branch of Artificial Intelligence (AI), which provides the interactions with human languages effectively. The emerging field of information retrieval (IR) based on question answering system involves the research methods from NLP, Information Extraction (IE), Database Management (DBMS) [2].

In IR process information is retrieved based on keywords from various heterogeneous resources. Because of this, the content meaning of the user's sentence in the question/query is lost. The user will receive a list of documents retrieved. However user may want a precise and comprehensive answer to his question [3].

The rest of this paper is organized as follows: Section 2 provides a survey of QAS. Section 3 proposes a methodology for QAS and section 4 concludes the paper.
Many evolutionary optimization techniques are existing in the literature. Genetic algorithm has the population represented as the set of chromosomes, fitness evaluation and genetic operators.

II. SURVEY

In [1], the author discusses the methodology for Visual Question Answering

(VQA). A GUI based VQA is implemented in the paper. In [2], the author used Efficient Net and bidirectional long short- term memory (BiLSTM) to extract more accurate image models. The author claims that the model is effective in contextual feature extraction of questions because of the BiLSTM layer that can process past and future information which works well on sequential modeling tasks.In [3], the paper provides a comprehensive detailed survey of emerging Question Answering systems and also presents the general architecture of the system.

The literature clearly showed that some hidden factors affect the performance of QASs that includes Psychology and the skill level of the user who is asking the question etc.In [4], the paper proposes a question answering system based on the BERT model that combines the co-attention mechanism and the self-attention mechanism. In [5], a semantics-guided and spatial-aware framework for natural resources Geo-analytical question answering is proposed. All the systems proposed by the above papers don't take user's information (information like profession of user, his/her skill level, etc.) into account. Next section proposes a new methodology for this reason.

III. METHODOLOGY

The general architecture will be same as in [3], but additional information about the user will also be input to the system along with the natural language question which will help the system in understanding the type of answer the user is expecting.
The proposed QAS architecture is shown in Fig. 1.

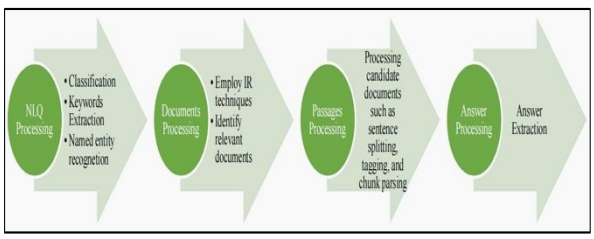

Fig. 1. The general architecture of QA system from [3].

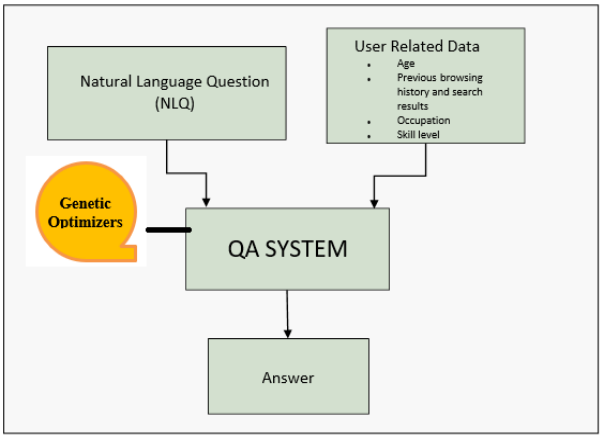

Fig. 2. Proposed system for question answering

For this to work, model should be trained using a dataset in which the user related data are present in it as attributes. The process of training takes place through genetic algorithm evolutionary technique. The steps of the proposed algorithm is as follows:

Step 1: Represent the set of chromosomes as population. In the given NLQ system the tagged sentences in the form of tokens and stemmers are considered as chromosomes. If S is a sentence and given by S = { w1,w2,…….wn} where wi being the set of words. The set of wi are called the stems.

Step 2: Fitness evaluation function: Fitness is the major objective function in evaluating the fitness of the features which are selected earlier. The minimum edit distance method is applied as an objective function to evaluate the fitness. The edit distance formula is given by the equation (1).

$$\mathrm{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \mathrm{lev}_{a,b}(i-1,j)+1 \\ \mathrm{lev}_{a,b}(i,j-1)+1 \\ \mathrm{lev}_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases} \quad \ldots\ldots (1)$$

Step 3: Genetic operators: There are 2 genetic operators such as crossover and mutation. Crossover is the process of interchanging the feature parameters of chromosomes and

mutation involves exchanging the parameters of chromosomes.

IV. EXPERIMENTAL RESULTS

The experiment was conducted with different individuals in the dataset for demonstrating question answering systems. The system is fed through series of genetic optimizers and the result is as shown in Fig 3.

Fig 3: Question answer system

V. CONCLUSION

With the help proposed architecture the user related information (such as: Occupation, Department, skill level, etc.) can be used to estimate the type of answer he is expecting (i.e. a brief answer, one word answer, a long descriptive answer, etc.). In future we can also classify the question into different categories (factoid, yes/no, hypothetical, complex, etc. [3]) and process the answer accordingly which could satisfy needs of the customer.

REFERENCES

1. Kanakamedala, D., Veeranki, T., Bitla, R., & Vangalapudi, S. (2021, November). Visual Question Answering Using Deep Learning. In *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)* (pp. 1-5). IEEE.
2. Ahmed, S. T., Singh, D. K., Basha, S. M., Abouel Nasr, E., Kamrani, A. K., & Aboudaif, M. K. (2021). Neural network based mental depression identification and sentiments classification technique from speech signals: A COVID-19 Focused Pandemic Study. *Frontiers in public health*, 1926.
3. Gupta, R., Hooda, P., & Chikkara, N. K. (2020, May). Natural Language Processing based Visual Question Answering Efficient: an EfficientDet Approach. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 900-904). IEEE.
4. Arbaaeen, A., & Shah, A. (2020, December). Natural language processing based question answering techniques: A survey. In *2020 IEEE 7th International Conference on Engineering Technologies and Applied Sciences (ICETAS)* (pp. 1-8). IEEE.
5. Yin, J. (2022, May). Research on Question Answering System Based on BERT Model. In *2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA)* (pp. 68-71). IEEE.
6. Wang, J., Sui, H., Li, Y., & Hu, L. (2022, July). A Semantics-Guided and Spatial-Aware Framework for Natural Resources Geo-Analytical Question Answering. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium* (pp. 2430-2433). IEEE.
7. Sreedhar Kumar, S., Ahmed, S. T., & NishaBhai, V. B. Type of Supervised Text Classification System for Unstructured Text Comments using Probability Theory Technique. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(10).
8. Nagesh, N., Patil, P., Patil, S., & Kokatanur, M. (2022). An architectural framework for automatic detection of autism using deep convolution networks and genetic algorithm. *International Journal of Electrical & Computer Engineering (2088-8708)*, *12*(2).
9. Al-Shammari, N. K., Alzamil, A. A., Albadarn, M., Ahmed, S. A., Syed, M. B., Alshammari, A. S., & Gabr, A. M. (2021). Cardiac Stroke Prediction Framework using Hybrid Optimization Algorithm under DNN. *Engineering, Technology & Applied Science Research*, *11*(4), 7436-7441.
10. Nagashree, N., Patil, P., Patil, S., & Kokatanur, M. (2022). InvCos curvature patch image registration technique for accurate segmentation of autistic brain images. In *Soft Computing and Signal Processing* (pp. 659-666). Springer, Singapore.