INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 3 - V7I3-1590)
Available online at: https://www.ijariit.com

A novel approach to improve Chat Bot accuracy using Natural Language Processing

Karan Heranjal
kheranjal@gmail.com
Jain (Deemed-to-be University), Bengaluru, Karnataka

Haripriya V.
author2@gmail.com
Jain (Deemed-to-be University), Bengaluru, Karnataka

Dr. Anilkumar T. V.
author3@gmail.com
LEAD College of Management, Palakkad, Kerala

ABSTRACT

Chat bots are finding increasing usage these days in applications ranging from simple automated FAQ's to sophisticated e-commerce buying recommendations. The technology powering these bots has undergone tremendous improvements thanks to new age sciences such as Machine Learning, AI and Natural Language Processing. Although chat bots have come quite far in effectiveness and imitating real humans, there's still lapses and areas where they can be improved in order to make them much more powerful and closer to human interaction. This is the main aim of the proposed work where one such limitation has been considered and attempt to address it by utilizing Natural Language Processing.

Keywords: *Artificial Intelligence, Machine Learning, Natural Language Processing, Chatbot*

1. INTRODUCTION

When it comes to the ever-evolving digital world, chatbots have gained importance and have many advantages over humans when it comes to providing assistance. Chatbots works much faster because they are written in software and calculate responses almost instantly also it can store a lot of information and huge knowledge base. Any amount of data can be handled thereby provides scalability. Since it is automated there is no cost involved such as salary payment to the employee. However, while all these advantages may hold true, the chatbots today still face many issues, some trivial and some complex. They don't behave like humans as much as we want them to. Various issues are existing with the chatbots like it cannot respond like humans do and will not be able to retain customers and persuade or market services to them intelligently. In addition, chatbots match keywords and thus don't consider the meaning of the question being asked, which often leads to incoherent responses. In order to find the solution for the lastly mentioned disadvantage, the proposed work has been implemented a new chatbot system.

2. OBJECTIVES

To build a normal chatbot that works by means of keyword matching between question and responses. The semantic weakness of this chatbot has been projected, hence bringing us to the aforementioned limitation. Finally, a new chatbot has been developed powered by NLP and harnessing semantic similarity to match responses based on meaning.

3. PROPOSED IMPROVEMENTS

As opposed to keyword matching, we are going to introduce synonym formulas to perform semantic similarity matching, Stopword removal to reduce noise in meaning matching, and reduce the load on both memory and processing of the server. Also, we make this chatbot cross platform, and have both standalone as well as an add-on feature, so companies won't have to migrate from existing infrastructure and can easily upgrade their existing bot with our features. This will also cater to many different query asking styles, since the meaning is processed and not the words themselves.

4. PROBLEM ANALYSIS

We can boil down the essence of the techniques to improve our chat bot and make it more human-like, into two major phenomena.

4.1. Understanding

Humans have fuzzy logic and can interpret queries asked differently, partially and even with grammatical errors. This is because they don't pay attention to each word, their order etc. but can generalize the meaning of the question asked as a whole based on its semantic features. This is what we try to achieve using NLP in our chatbot.

4.2. Meaning Matching: In order to give a response, even if they don't know the exact answer, humans manage to give a response that is almost accurate to the query being asked. This is thanks to a matching based not on word exactness, word order, size or even grammar correctness, but due to finding a response that is most semantically similar to the generalized meaning of the query being asked.

5. METHODOLOGY

Python, Django and SQLite have been used for the backend implementation of this project, along with Bootstrap for a responsive frontend that accepts input. The algorithm used in this work is based on the sum of the common synonyms of each word from the compared sentences, and then selecting the most similar sentence from this result. The data taken is a simple admissions FAQ consisting of Question Answer pairs from the Jain university website. The detailed methodology applied is illustrated in the below sections.

Let Stopwords = S, the no-duplicates set of english stopwords from NLTK

S = {"and", "a", "the", "is", and so on}

Eq 1: Stopwords definition

Let User query = Q, which we tokenize into a set of words, remove duplicates and call it the User Set, US.

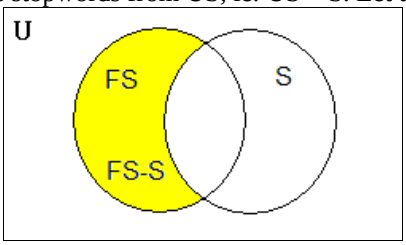
Q = "Word1 Word2 Word 2 Word3 Word4"

Tokens = ["Word1", "Word2", "Word2", "Word3", "Word4"]

US = {"Word1", "Word2", "Word3", "Word4"}

Eq 2: Tokenization and removing duplicates

Perform Set Difference to remove stopwords from US, ie: US – S. Let the final query set be called as FS, such that FS = US – S



FS = US - S

Eq 3: Removing Stopwords

Now FS contains words W1, W2, W3 and so on.

FS = {W1, W2, W3, and so on}

Eq 4: Final User Set Definition

Using Wordnet, Generate synonym set Syn1 for each word W1. Repeat for all User Words in Set FS, to get Syn sets for each word.

Syn1 = SynonymSet(W1)

Syn(N) = SynonymSet(W(N))

Eq 5: Synonym Set Definition

Combine all Syn Sets above and create a no-duplicate set USyn containing all synonyms of all words in User query word set FS.

USyn = Syn1 + Syn2 + Syn3 + so on

Eq 6: Making a Superset of Synonym Sets

Repeat Steps 2 to 8, for each Response in the database R, to get final no-duplicate sets RSyn1, RSyn2, RSyn3 and so on, for each R.

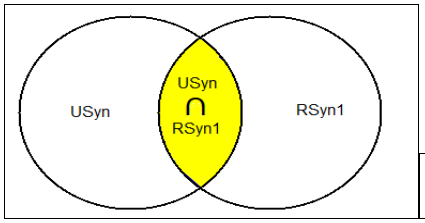
RSyn1 = R1W1Syn1 + R1W2Syn2 + so on

RSyn2 = R2W1Syn1 + R2W2Syn2 + so on

Where R is Response, W is each word.

Eq 7: Response Synonym Set definition

Lets take the first response RSyn1, and Perform Set Intersection to get number of common synonyms between USyn and RSyn1, lets call this I1 (Intersection 1)



$$I1 = USyn \cap RSyn1$$

Eq 8: User Synonym Set Intersection of Response Synonym Set

Repeat Steps 9 and 10, taking USyn and each of the remaining response sets RSyn2, RSyn3, RSyn4 and so on to get I2, I3, I4 and so on. Make a list of all the I values.

$$IList = [I1, I2, I3, \text{and so on}]$$

Eq 9: Creating List of all Similarities

Take the highest I value from the IList containing [I1, I2, I3 and so on], which denotes the highest semantic similarity, then get the list index of the same and finally return the Response R from Response list at that index, that gave us the highest I value.

$$IMax = Max(IList)$$

$$MaxIndex = IList.index(IMax)$$

$$FinalResponse = RL(MaxIndex)$$

Where RL is the database response list containing R1, R2 and so on.

Since each I corresponds one-to-one with R, we can take the max index of IList and get the response R with that index.

Eq 10: Equation to calculate most similar response

6. RESULTS

Here, the final result of the proposed work is depicted with the comparisons between a normal chatbot and the proposed chatbot, in order to note the benefits and increased effectiveness.

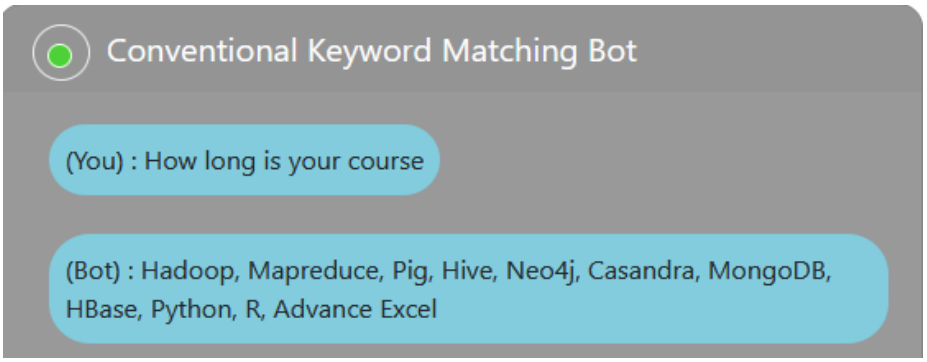


Fig 1: Screenshot 1

Fig 2: Screenshot 2

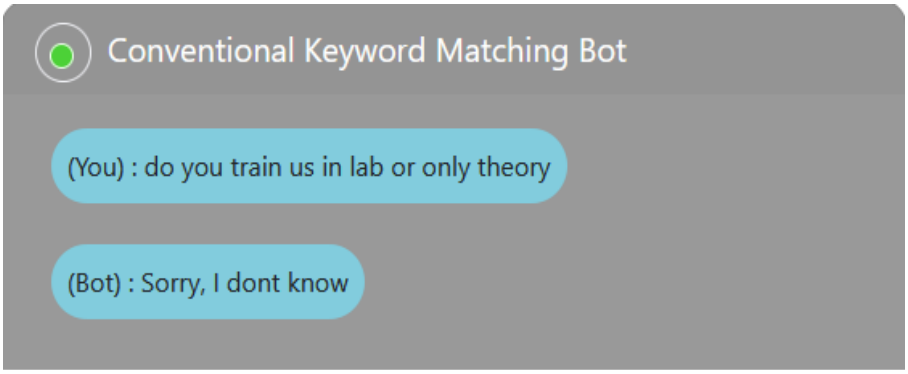


Fig 3: Screenshot 3

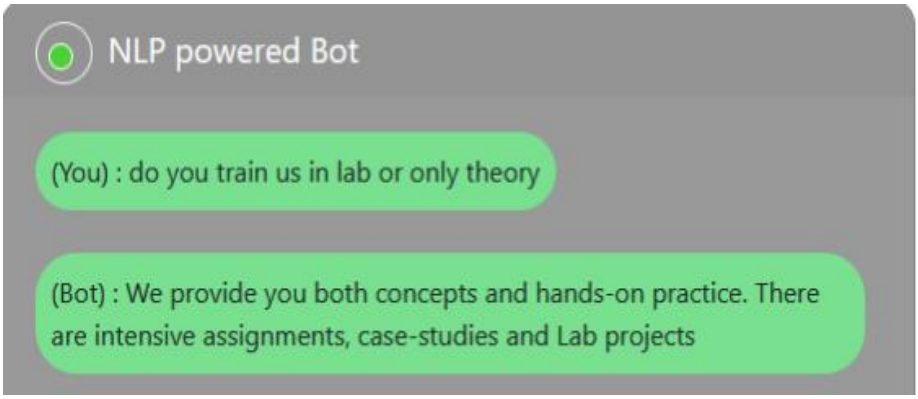


Fig 4: Screenshot 4

7. CONCLUSION

Chatbot is a software that simulates human –like conversation with users through online via text message. Thereby it helps the users to know more about particular website. The proposed work mainly concentrates on the meaning of the words rather than the word itself. The Chatbot selects the most similar responses from the database and it uses semantic similarity to provide more accurate result. The Chatbot is independent of the length of the query and caters to numerous different queries and asking styles. As a result we the present work provides high accuracy and correctness of the response as opposed to the conventional Chatbots. As any system the limitation of this work is the database considered for the work is static and it cannot differentiate between Nouns, Verbs and Adjectives. Large dataset can be used in the future also can make Chatbot friendlier and empathetic and less robotic and it is suggested to include an option in the future work to escalate queries to humans when required.

REFERENCES:

[1] Setiaji, B., & Wibowo, F. W. (2016, January). Chatbot using a knowledge in database: human-to-machine conversation modeling. In *2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)* (pp. 72-77). IEEE.

[2] Rahman, A. M., Al Mamun, A., & Islam, A. (2017, December). Programming challenges of chatbot: Current and future prospective. In 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 75-78). IEEE.

[3] Anghelescu, P., & Nicolaescu, S. V. (2018, June). Chatbot Application using Search Engines and Teaching Methods. In *2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (pp. 1-6). IEEE.

[4] Hristidis, V. (2018, September). Chatbot technologies and challenges. In 2018 First International Conference on Artificial Intelligence for Industries (AI4I) (pp. 126-126). IEEE.

[5] Khin, N. N., & Soe, K. M. (2020, February). University Chatbot using Artificial Intelligence Markup Language. In 2020 IEEE Conference on Computer Applications (ICCA) (pp. 1-5). IEEE.

[6] Khanna, A., Pandey, B., Vashishta, K., Kalia, K., Pradeepkumar, B., & Das, T. (2015). A study of today's ai through chatbots and rediscovery of machine intelligence. International Journal of u-and e-Service, Science and Technology, 8(7), 277-284.

[7] Nuruzzaman, M., & Hussain, O. K. (2018, October). A survey on chatbot implementation in customer service industry through deep neural networks. In *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)* (pp. 54-61). IEEE.

[8] Ahsaee, M. G., Naghibzadeh, M., & Yasrebi, S. E. (2010, December). Using WordNet to determine semantic similarity of words. In *2010 5th International Symposium on Telecommunications* (pp. 1019-1027). IEEE.

[9] Abdul-Kader, S. A., & Woods, J. (2017, September). Question answer system for online feedable new born Chatbot. In 2017 Intelligent Systems Conference (IntelliSys) (pp. 863-869). IEEE.

[10] Das, S., & Kumar, E. (2018, December). Determining accuracy of chatbot by applying algorithm design and defined process. In 2018 4th International Conference on Computing Communication and Automation (ICCCA) (pp. 1-6). IEEE.

[11] Hristidis, V. (2018, September). Chatbot technologies and challenges. In 2018 First International Conference on Artificial Intelligence for Industries (AI4I) (pp. 126-126). IEEE.

[12] Kucherbaev, P., Bozzon, A., & Houben, G. J. (2018). Human-aided bots. IEEE Internet Computing, 22(6), 36-43.

[13] Rahman, A. M., Al Mamun, A., & Islam, A. (2017, December). Programming challenges of chatbot: Current and future prospective. In 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 75-78). IEEE.

[14] Lobur, M., Romanyuk, A., & Romanyshyn, M. (2011, February). Using NLTK for educational and scientific purposes. In *2011 11th international conference the experience of designing and application of CAD systems in microelectronics (CADSM)* (pp. 426-428). IEEE.

[15] Raj, S., Raj, & Karkal. (2019). Building chatbots with Python. Apress.

