Design of Automated Healthcare Chatbot Using Natural Language Processing

Dr. T. Praveen Blessington[#1], Lubna Sameer Attar[#2], Rohit Ram Bembre[#3], Komal Sanjay Bhagwat[#4], Shweta Somnath Kale[#5]

Guide, Department Of Information Technology[#1]
Students, Department Of Information Technology[#2,#3,#4,#5]
Zeal College Of Engineering And Research, Pune, Maharashtra, India

***Abstract*— Health care is very important for a good life. However, it is very difficult to get advice from a doctor for any health problem. The idea is to build a AI medical chatbot that can diagnose diseases and provide basic information about the disease before seeing a doctor. The Medical chatbot will help reduce medical costs and increase access to medical knowledge through a computer program that interacts with users using natural language. The chatbot stores the data in a database to identify keywords in the proposal, determines the request, and answers the question. Rank and similarity calculations of sentences are performed using n-grams, TF IDF, and cosine similarity. Get a score for every sentence in a given input sentence and get more similar sentences for a given query. A third party, the expert program, handles questions submitted to the bot that are not understood or not in the database.**

***Index Terms*— Chatbot, Healthcare, Virtual Assistance, Artificial Intelligence, N-gram, TFID**

I. INTRODUCTION

Computers give us information. They engage us and help us in many ways. A chatbot is a program designed to forge smart communications using text or voice. However, this article focuses only on text.

These systems can self-learn and restore knowledge with the help of people or web resources. This application is incredibly basic as knowledge is pre-stored. The system application uses a chatbot-like Q&A protocol to respond to user queries. This system is designed to reduce the user's medical costs and time since the user is not able to visit a doctor or specialist when they need it immediately.

Questions are answered based on user query and knowledge base. Important keywords are selected from sentences and responses to the sentences. If a match is found or is important, a response is given or a similar response is displayed.

The complex questions and answers in the database are viewed and answered by experts. Here, users can ask health-related questions privately, as they do not spend much time consulting with their doctor. Chat template input sentences are stored in the Relational Database Management System (RDBMS). The chatbot will reconcile the input sentence of the user question with the knowledge base. Each request is compared against the chatbot's knowledge base. Key keywords were extracted from the given input sentences and similar sentences were found. Keyword Rank and Sentence Similarity are determined using N-gram, TF-IDF, and Cosine Similarity. The interface is built autonomously using the PYTHON programming language.

II. LITERATURE REVIEW

Here, the study is based on emotion recognition classification using AI techniques. This study trains emotion classification models on many labeled data based on recurrent neural networks (RNNs), deep learning and convolutional neural networks. Verbal interaction is paramount when contacting, using natural language processing (NLP) and natural language generation (NLG) to understand user conversations. Here, a multimodal approach recognizing emotions is used. In order to study semantic information of words, corpus was collected and presented as vectors, and synonym vocabulary knowledge was collected using word vectors. [1]

This is an article that developed a chatbot for voice recognition. The question you are asking the bot is unclear. Additional processing by third-party expert systems. A webbot is created as a webfriend based on text that entertains users. If the program has a voice system as well as text, here we focus on the improved system. Step process to capture and analyze the input signal. Recognize data from server response and process information. The server used here is SOAP-based black box method. The expert system allows the to increase unlimited and autonomous intelligence.[2]

This chatbot is designed for human-machine communication. Here the system stores a knowledge base for identifying sentences and determining answers to questions. The input sentence receives a similarity score of the input sentence using a diagram. Chatbot knowledge is stored in a DBMS. [3]

The chatbot is implemented using pattern matching, which recognizes sentence order and stores response patterns. Here, the authors describe the implementation of the chatbot operating system, software, programming language, database, and how the input and output results are stored. Here the input is text() function, remaining punctuation is removed with trim() function and random() function is used to select response from database. Chatbots are used for entertainment purposes. [4]

Here they extract words from sentences using the n-gram technique. Here n-grams are used to compare with Moro phonemes and phonemes as decision parameters and output the input as case data. A probabilistic analysis is performed on the closest match. The final expression is redirected through the expert system. [5]

Chatbot developed here for health purposes for Android apps. User sends text message or voice message using Google API. Here, the user only receives relevant responses from the chatbot. SVM algorithm is used to classify the data set. Here, Porter's algorithm is used to discard unwanted words such as suffixes or prefixes. [6]

Various documents available on the Internet, the contents of are verified by labeling the dataset using the low-dimensional demo based on n-grams, a TF-IDF matrix that generates S,U and V, and Finally, 3-matrix multiplication similarity per cosine is computed. [7]

Here, chatbots are built for customer service and acts as a public health service. The application uses N-grams, TF-IDF, and cosine similarity. The knowledge base was created to store questions and answers. Appendix uses unigrams, bigrams, and trigrams to help you answer question quickly by clearly showing the questions and keywords extracted from. [8]

The authors proposed a model more suitable for educational purposes, The idea of this study is to provide students with a more interactive way to connect to the university system.[9]

III. PROPOSED SYSTEM

Chatbots are interactive virtual assistants that automate interactions with users. The chatbot is powered by artificial intelligence that uses machine learning technology to understand natural language. Section's main motivation is to help users with non-essential medical information. When they first visit the site, users can first register and then ask the bot questions. The system uses an expert system to answer queries. The chatbot data is stored in the database in the form of template-templates. Here SQL is used to process database.

IV. SYSTEM ARCHITECTURE

Fig. 1 shows a system architecture diagram of a chatbot for a health management application. The client enters the question into the user interface as text. The UI receives the user's request and then sends a to the chatbot application.

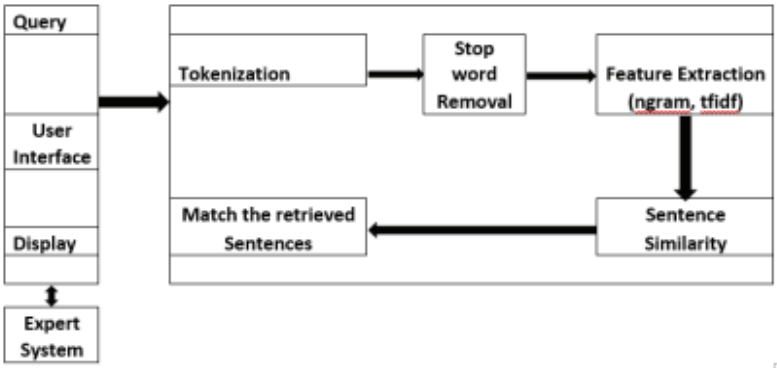

Fig. 1: System Architecture

In the chatbot application, the literary experience pre-processing step includes tokenization, at which point stopwords are removed when words are tokenized and feature extraction depends on n-gram, TF-IDF and cosine similarity . Answers to Question are stored in the knowledge database to restore receipt of Answer.

Tokenization

Tokenization is essentially the procedure of remodeling touchy statistics into insensitive statistics. The non-sensitive data that is created is called a "token" and can be used in databases or internal systems without exposing sensitive data to risk. Tokenized information can't be deciphered and reversed. There is no mathematical relationship between tokens and raw data. Tokens are essentially non-confidential placeholders stored outside of their original source.

To obtain the tokenized original sensitive data, the tokenization solution replaces the token with the original data. A token is an unrelated value, but usually retains some element of the original data, such as length or format, so it can be used for seamless business transactions. Original sensitive data is stored securely outside the organization's internal systems.

Split words or sentences word by word to increase processing power. Splits text into words whenever one of the contractions of the specified character occurs. All words are separated from sentences and punctuation marks are removed. This means the next step.

Stop words removal

Remove stop words from sentences to extract important keywords. It is mainly used to remove unnecessary things, such as words that appear too often in a sentence. It is also used to remove unimportant words or words that do not have a specific meaning, such as an, a, the. This step is used to reduce processing time or computational complexity.

Feature extraction based on N-gram TFIDF

Feature extraction is the manner of decreasing features from a document. Rank properties in line with documentation. By doing this step it upgrades the speed and adequacy of the document. It is used to extract the set of keywords and frequency of the keywords in the document.

TF-IDF

Term frequency and Inverse document frequency is used to calculate the weight of each term in the sentence.

The term frequency is used to check how many times the term as been occurred in a particular sentence using the formula below.

tf = tfi

Note: tf stands for term frequency

IDF is used to calculate the weight of unusual words in all reports of a document.The words that appear in a while in the document have a high IDF score. It is given by the condition underneath

idf(t) = log(N/(df + 1))

Note: idf stands for inverse document frequency.

tf and idf are combined to get the weight of a term or word in a document.. The tf and idf values are multiplied to obtain the weight of each term in the document.

tf-idf(t, d) = tf(t, d) * log(N/(df + 1))

N-gram

N-gram is an attempt to extend the N-gram model with a variable length mechanism. A sequence can be a grouping of words, word class, grammatical feature or whatever a succession of something that the modeller thinks bearing significant language structure data. In this system, N-gram is used for text compression or reduce the data space in the document, to extract the relevant keywords from the database.

Sentence Similarity

Cosine similarity is used to test the similarity between two sentences. The similarity between the query and the document is directly proportional to the number of query weights. Since the frequency of member cannot be negative, the result of calculating the similarity of two documents is between 0 and 1. The formula for calculating the similarity of cosine is:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

V.CONCLUSION

Chatbots are remarkable communication tools. Here, the app is designed to provide quality responses in a short amount of time. Ease the burden on response providers by using an expert system to deliver responses directly to users. This project is designed to help users save time consulting doctors or specialists for medical solutions. Here, we developed an application to extract keywords from user queries using N-grams, TF-IDF.

A weight is applied to each keyword to get the correct answer for your query. The web interface is designed for users entering queries. Security and performance updates enhance the app to keep you and your characters safe and get answers.

REFERENCES

1. K. Oh, D. Lee, B. Ko and H. Choi, "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation," 2017 18th IEEE International Conference on Mobile Data Management (MDM), Daejeon, 2017, pp. 371-375. doi: 10.1109/MDM.2017.64
2. Du Preez, S.J. & Lall, Manoj & Sinha, S. (2009). An intelligent web based voice chat bot. 386-391.10.1109/EURON.2009.5167660
3. Bayu Setiaji, Ferry Wahyu Wibowo, "Chatbot Using a Knowledge in Database: Human-to- Machine Conversation Modeling", Intelligent Systems Modelling and Simulation (ISMS) 2016 7th International Conference on, pp. 72-77, 2016.
4. Dahiya, Menal. (2017). A device of verbal exchange: Chatbot. INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING. 5. 158-161.2017. C.P. SHABARIRAM, V.Srinath, C.S
5. C.P. Shabariram, V. Srinath, C.S. Indhuja, Vidhya (2017).
6. Ratatta: Chatbot Application Using Expert System, International Journal of Advanced Research in Computer Science and Software Engineering,2017
7. Mrs Rashmi Dharwadkar1, Dr.Mrs. Neeta A. Deshpande,
8. A Medical Chatbot, International Journal of Computer Trends and Technology (IJCTT)-Volume 60 Issue 1- June 2018
9. Farheen Naaz, Farheen Siddiqui, modified n-gram based model for identifying and filtering near-duplicate documents detection, International Journal of Advanced Computational Engineering and Networking, ISSN: 2320- 2106, Volume-5, Issue-10, Oct.-2017

10. N-gram Accuracy Analysis in the Method of Chatbot Response, International Journal of Engineering & Technology. (2018) Shukla, V.K, Verma, A, "Enhancing LMS Experience through AIML Base and Retrieval Base Chatbot using R Language", 2019 International Conference on Automation, Computational and Technology Management (ICACTM)