ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 11, Issue, 04, pp.3291-3295, April, 2019

DOI: https://doi.org/10.24941/ijcr.35214.04.2019

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

MEDICAL CHATBOT USING TOPIC – ASPECT MAPPING AND SUPPORT VECTOR MACHINE

***Stephy Joseph and Dr. Antony Viswasa Rani, A.**

Dept. of Computer Science and Engineering, Amrita College of Engineering and Technology, Erachakulam, Nagercoil, India

ARTICLE INFO

Article History:
Received 14th January, 2019
Received in revised form
17th February, 2019
Accepted 24th March, 2019
Published online 30th April, 2019

Key Words:

Medical Chatbot,
Natural language processing,
Topic modeling,
Aspect extraction,
Machine learning,
Support vector machine (SVM)

Corresponding author:* **Stephy Joseph

ABSTRACT

Usually, Users are not aware of all the details about the hospitals, doctors, treatment or symptoms regarding the particular disease. For the small problems, users have to go personally to the hospital for a check-up which is time consuming and expensive. Also handling the telephonic calls for the complaints is quite hectic. Such problems can be solved by using medical Chabot which provides proper guidance for healthy living. The proposed medical chatbot functioning based on Natural language processing, topic modeling, aspect mapping and SVM which helps users to submit their queries about health problems, concerned specialists and get the suggestion about treatment and related services. The User can ask any personal query related to health care through medical chatbot without accessing the hospital in person. The proposed approach for medical chatbot has three phase processing. The first phase is the preprocessing stage which includes Natural language processing methods like word splitting, filter out punctuations, stop word removal and finally porter stemming is done to identify the root word. The second phase includes topic modeling, Aspect extraction and topic aspect mapping. Here the identified topics and aspects are mapped together and assigned to each categorized dataset and trained a machine with SVM. The third phase of the proposed approach is the trained system identifies the aspect from the human typed sentences in medical chatbot and these identified aspects matched with the database sets. And the chatbot redirected to matched dataset and retrieved the corresponding information and displayed in chatbot. Finally, check the efficiency of proposed work with different size of the medical conversation dataset and also the efficiency of the machine learning approach.

Citation: Stephy Joseph and Dr. Antony Viswasa Rani, A. **2019.** "Medical Chatbot Using Topic – Aspect Mapping And Support Vector Machine", *International Journal of Current Research*, 11, (04), 3291-3295.

INTRODUCTION

According to Oxford Dictionary Chabot is a computer program that simulates a conversation with human users, especially over the Internet. I'd rather say: *chatbots are AI software installed on communication platforms, which can answer some *basic* questions about n a certain domain or on a certain topic.* The main aim of the medical Chatbot is to build the language gap between the user and health providers by giving immediate replies to the Questions asked by the user. Today's people are more likely addicted to the internet but they are not concerned about their personal health. They avoid going to the hospital for a small problem which may become a major disease in the future. Establishing question-answer forums is becoming a simple way to answer those queries rather than browsing through the list of potentially relevant document from the web. Many of the existing systems have some limitation such as there is no instant response given to the patients and they have to wait for expert's acknowledgment for a long time. Some of the processes may charge amount to perform live chat or telephonic communication with doctors online (Manoj Kumar, 2016).

This system allows communication between human and computer by using natural language processing (NLP). The first phase is performed with natural language i.e. identification of main linguistic relations of the completed sentences. After that, the porter stemming is used to pick out the root word and its linguistic meaning. The second phase includes topic modeling, aspect extraction, aspect-topic mapping and machine learning with SVM. A 'topic' can be defined as a recurring pattern of co-occurring words. So a 'topic' consists of a cluster of words that frequently occur together and the words in the topic make sense. The number of topics depends upon the degree of the size of the corpus and a small number of topics will provide a broad overview of the contents of the corpus. These topics are extracted using MALLET in topic modeling. The next level is aspect extraction. Aspect is a particular part or feature of a dataset or entity. It can be specifically defined as an attribute or component of an entity. An example of an aspect is 'symptoms' of a disease. The aspect extraction has become a relevant research area in the field of chatbot. After topic modeling and aspect extraction, the extracted topics and aspects are mapped together. The Topic-Aspect mapping maps the topics with the aspects identified during aspect extraction.

The words or its synonyms in a topic that finds a match with an aspect are identified and then the topic is mapped to that particular aspect. And each aspect is assigned with corresponding database sets having the topics. The next step is trained the system with these topics and aspects of the database. The final phase of this proposed system is the trained system is identified topics and aspect identified from the user conversations and these aspects matched with the database aspects and the matched databases are loaded in the backend of the chatbot for further communication.

LITERATURE SURVEY

A smart chatbot (Emanuela Haller *et al*., 2013) for customer care uses Software as a Service which analyzes the message of each application server. It helps the user to resolve the issue by providing a human way interactions using LUIS and cognitive services which are implemented on AWS public cloud. Admin feeds input to the machine so that machine can identify the sentences and taking a decision itself as a response to a question. The database used in the project is MySQL. The illustration and execution of SQL in the pattern matching operation is required. The conversation can be done so that it can add some knowledge to the database as it has not been modeled before. If in case the input sentences in the database did not match then it will be remodeled (Bayu Setiaji, 2016). The paper gives information regarding products which is useful for consumers to obtain what they want exactly. Question Answering (QA) systems can be identified as information accessing systems which try to answer to natural language queries by giving answers suitable answers making use of attribute available in natural language techniques (Agnese Augello, 2012). The system takes a plain text as input and answering all type of questions output by a qualified user is the output. The purpose is to provide a generic solution to this problem. This paper helps in recognizing the reality in texts and giving the past content for developing a conversation which is used in middle-school CSCL scenarios (Emanuela Haller, 2013). This paper discussed the medical chat-bots functioning depends on Natural language processing that helps users to submit their problem about health. The User can ask any personal query related to health care through the chat bot without physically available to the hospital. By Using Google API for voice-text and text voice conversion. Query is sent to chatbot and gets related answer and display answer on android app. The System's major concern behind developing this web based platform is analyzing customer's sentiments (Rashmi Dharwadkar, 2018) Paper uses artificial intelligence for predicting the diseases based on the symptoms and give the list of available treatments. It can facilitate us to figure out the problem and to validate the solution (Divya Madhu, 2017).

PROPOSED SYSTEM

In this work, medical chatbot perform with topic modelling, aspect extraction, topic-aspect mapping and pattern matching. In the first phase of this work, the dataset is collected and conversations are categorized based on subjects or areas like disease, hospitals and create the different databases and pre-processing it with NLP functions and stemming function and identified the stem words. The second phase of this work is Topic modeling, performed to extract topics from the text corpus. Here the text corpus is a collection of medical conversations. Then Topic modelling and aspect extraction is carried out for mapping the topics and aspects with different database set and the extracted topic and aspects are trained with the machine using SVM In this work the aspect extraction is performed using pattern library and aspects are tagged with database with corresponding topics. Finally the trained system are identified aspects from the human conversation and matched with db aspects and the information is retrieved and continued the conversation. The workflow of the proposed system is given below:

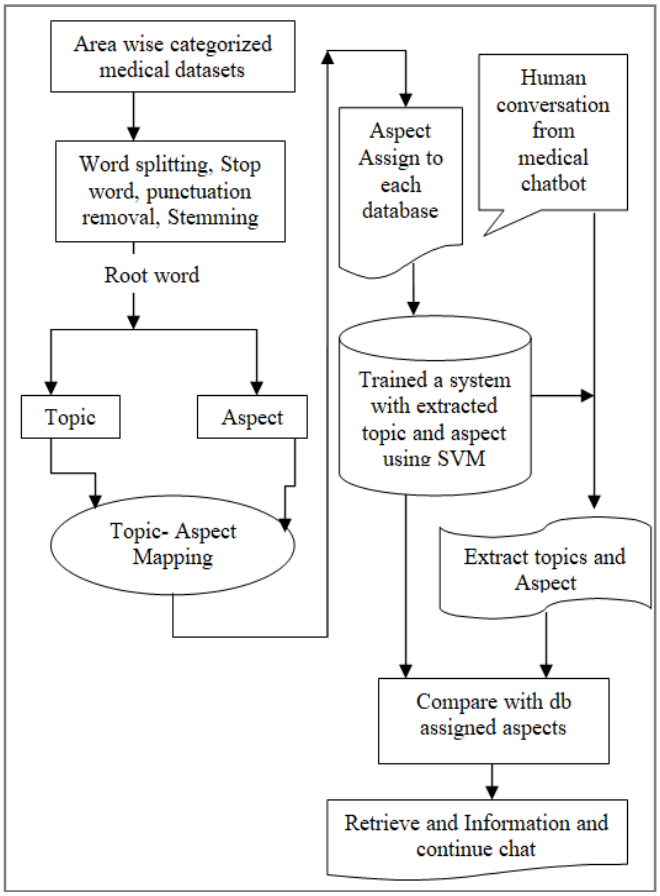

Fig. 1. Proposed system for medical chatbot

Dataset Collection: The dataset collection is the first step in this work, and here we extract the data from various sites, which consists of a large collection of conversation about the medical related topics. The 'conversation' field consists of the text of the diseases, locations, doctors, hospitals, medical shops etc and this text/conversations is separated as the individual dataset based on the different areas like disease or related hospitals etc and perform topic modelling and aspect extraction. The datasets are processed and find out which topics are included in individual documents. We need to transform them from individual text files into a single MALLET format file. MALLET can import more than one file at a time and we can import the entire directory of text files using the 'import' command.

Pre-processing: In the pre-processing level, the basic natural language processing functions are used. It is the process of converting raw text file into a well-defined sequence of linguistically meaningful units. This is an essential part of this system, since the characters, words and sentences identified at this stage are the basic elements passed to all other processing stages. The words such as and, the, but, if, that etc which obstruct analysis are thus removed. And finally the porter stemming is performed and the core/stem words are identified.

Topic Modeling: Topic modeling is the third step in the workflow and it is performed to extract topics from the text corpus. Topic models represent a family of computer programs that extract topics from texts. Topic modeling programs do not know about the meaning of words in a text, instead it considers a document as the random distribution of words which we call them as 'topics'. Here MALLET takes a text corpus and looks for patterns in the use of words and it is an attempt to inject semantic meaning into vocabulary. In the case of hundreds of review documents from an archive, topic modeling will be a good approach to understand something of what the archive contains without reading every document and this will be helpful for extracting patterns from the conversation texts. The MALLET tool runs much iteration through the entire medical conversation texts until it settles on the most likely distribution of words into topics. The MALLET command runs the topic model routine on the MALLET file using the default settings. As it runs through the routine, it tries to find the best division of words into topics. We can train MALLET to find any number of topics and here we have found 25 topics. This command finally outputs a text document and showing the topic keywords for each topic. The topic modeling extracts the thematic information behind the medical conversation corpus and some of those topics are given as follows:

Table 1. Topics extracted after topic modeling

Topic 1	Hospital life quality care location support contacts doctor specialty features insurance causality
Topic 2	Doctor location works fees op-time specialty qualification
Topic 3	Disease symptoms body parts specialty hospital doctor prescription medicine location
Topic 4	Patient near hospital issues health care condition emergency Doctor
Topic 5	Medical store location hospital prescription health value medicine availability

MALLET Natural Language Processing Toolkit: MALLET is a Java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other applications to text. It is a topic modeling package which includes an extremely fast and highly scalable implementation of Gibbs sampling, efficient methods for document-topic hyper parameter optimization and

tools for inferring topics for new documents given trained models (6). It is a topic modeling tool that is text focussed and uses command line scripts and direct Java API for topic modeling and it is based on Latent Dirichl*et al*location (6). MALLET gains its motivation from text classification and information extraction. It includes sophisticated tools for document classification, efficient routines for converting text to "features".

Topic models are useful for analysing large collections of unlabeled text. The MALLET topic modeling toolkit contains efficient, sampling-based implementations of Latent Dirichl*et al*location, Pachinko Allocation, and Hierarchical LDA. Many of the algorithms in MALLET depend on numerical optimization and it includes an efficient implementation of Limited Memory BFGS, among many other optimization methods (6). The MALLET toolkit is an Open Source Software and it involves modifying an environment variable and working with the command line. This tool is used in this work for topic modeling of product reviews to extract the thematic information behind them (McCallum, 2002).

Aspect Extraction: Aspect Extraction is performed for identifying the various aspects of mobile phones and here we have identified aspects manually from the commercial vendor websites. The some aspects, thus identified are listed below:

Table 2. Aspects

Patient	Disease	Hospital
Doctor	Medicine	Location

The aspect extraction expresses the exact medical area towards each conversation. In such cases, the overall human complicated conversation is less meaningful than area specific to each aspect. The following representation denotes the aspect specific information in the human medical conversation. In this work, we have identified 6 such aspects from the categorized medical conversational dataset.

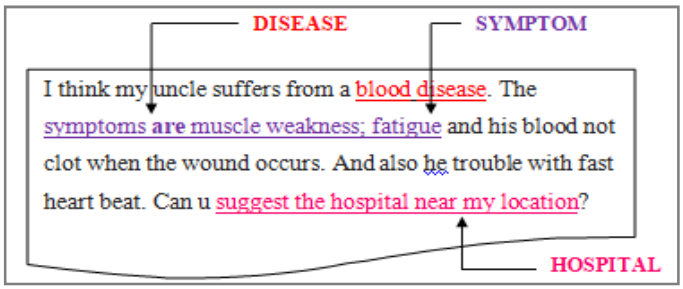

Fig. 1. Medical Conversational texts specifying various aspects

Topic – Aspect Mapping: The Topic-Aspect mapping maps the topics with the aspects identified during aspect extraction. The words or its synonyms in a topic that finds a match with an aspect are identified and then the topic is mapped to that particular aspect. For example, the topic words such as "hospital, life, quality, care, location, contacts, specialty, insurance, causality" in Topic 1 indicates the aspect 'Hospital' and the topic words such as "symptoms, disease, parts" in Topic 3 indicates the aspect 'Symptoms'. Here we have mapped the 25 topics to 6 such aspects and the following table shows the topic-aspect mapping of our work

Table 3. Topic – Aspect Mapping

Topic 1	hospital life quality care location support contacts doctor specialty features insurance causality	Hospital
Topic 2	Doctor location works fees op-time specialty qualification	Doctor
Topic 3	Disease symptoms body parts specialty hospital doctor prescription medicine location	Disease
Topic 4	Patient near hospital issues health care condition emergency Doctor	Location
Topic 5	Medical store location hospital prescription health value medicine availability	Medicine

Machine Learning with Support Vector machine: "Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges in the area of training the systems. However, it is mostly used in classification problems. Here the supervised machine learning algorithm SVM is used assign the aspects to database and trained a system with these aspect assigned database. After that the human typed sentence/conversation in the Chabot UI, the trained system identified the aspects in the sentence and based on that aspect, it redirect to assigned aspect matched dataset and retrieve the

corresponding information based on the question and displayed the related answer in the Chabot. All aspects extracted from the human conversation should match the dataset then only particular dataset will get selected and the conversation will continue. Otherwise the Chabot display the "data not found" message.

RESULTS AND EVALUATION

Through topic modeling, aspect extraction, topic aspect mapping - the key concepts in a conversational db can be identified. All aspects extracted from the human conversation should match the dataset then only particular dataset will get selected and the conversation will continue. The result of this work clearly specifies the particular topics and aspect in the database. As the first phase of this work, topic modeling is performed on the medical conversations corpus using MALLET to extract topic keywords. Here we have identified 25 such topics. The topics give a clear idea about what the entire conversation corpus is said about. In the aspect extraction step, we have identified the aspects from the dataset and in this work we have identified 6 such aspects. The topic words in each topic enable us to map each topic with a particular aspect. Thus we have mapped the 25 topics with 6 aspects as described in Table 3. The machine learning phase, the trained system identified the aspects from the user entered conversation in chatbot and based on that aspects, it redirect to matched dataset and retrieve the corresponding information based on the question and displayed the related answer in the Chabot. The below displays the screen shot of the frontend medical Chabot.

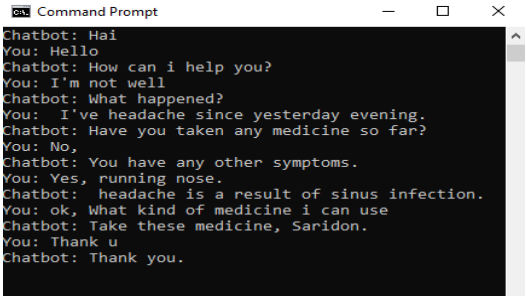

Fig. 2. Result Chabot

In the evaluation phase, to increase the dataset size, more conversation is to be added in each database. The new medical areas are also included and tried to identify more aspects and compare the efficiency of this approached method in different dataset sizes and different aspects.

The below graph represent the efficiency of this method with different conversation dataset size.

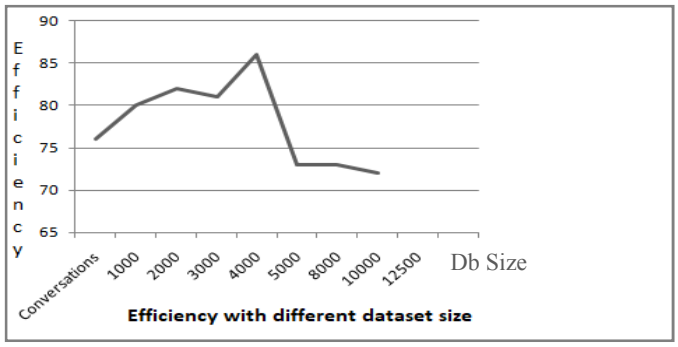

Fig. 3. Efficiency graph with different size dataset

When the dataset size increased the efficiency level increased up to a point and decreased from that particular. Compared with other algorithms without aspect mapping, the proposed system performed with high efficiency.

The below graph represent the efficiency of this proposed system with different methods.

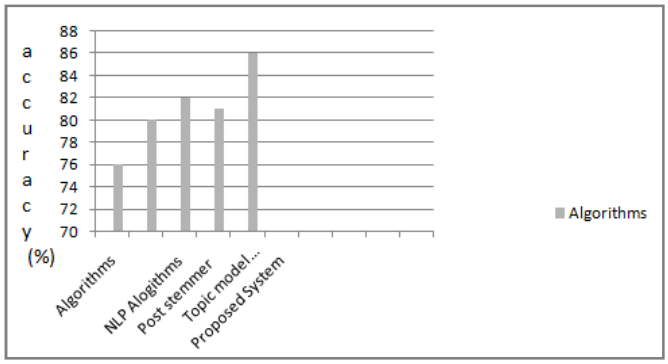

Fig. 4. Accuracy graph with different algorithms

And also check the efficiency of the SVM with this proposed medical chabot system. In this below table shows the Confusion Matrix of SVM with proposed model. The size of dataset is 5000.

Table 4. Confusion Matrix of SVM

True Positive (TP) 3423	False Positive (FP) 561
False Negative (FN) 392	True Negative(TN) 624

From the above confusion tables, we have calculated the values for True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN). Based on these values we can calculate the Accuracy, Precision and Recall of the Sentiment Analysis system described in this work. Accuracy, Precision and Recall are defined as follows:

Accuracy=((TP + TN) / TP+TN+FP+FN)
Precision= TP / (TP + FP)
Recall= TP / (TP + FN)

The Accuracy, Precision and Recall values of SVM with this proposed model is 80, 72, 50 respectively. The future work is to propose a new approach to improve the efficiency of this medical chatbot and apply different machine learning approaches apply for this proposed model and make comparison and find which machine learning approach is more fit for this proposed work.

REFERENCES

Agnese Augello, Giovanni Pilato, Alberto Machi' ICAR Istituto di Calcolo e Reti ad Alte Prestazioni CNR - Consiglio Nazionale delle Ricerche Viale delle Scienze , 978-0-7695- 4859-3/12 $26.00 © 2012 IEEE . "An Approach to Enhance Chatbot Semantic Power and Maintainability: Experiences within the FRASI Project".

Bayu Setiaji, Ferry Wahyu Wibowo, 2016. Department of Informatics Engineering STMIK AMIKOM Yogyakarta, Yogyakarta, Indonesia, 2166-0670/16 $31.00 © IEEE "Chatbot Using A Knowledge in Database-Human-to-Machine Conversation Modeling".

Madhu D., C. J. N. Jain, E. Sebastain, S. Shaji and A. Ajayakumar, "A novel approach for medical assistance using trained chatbot," 2017*, International Conference on Inventive Communication and Computational Technologies (ICICCT),* Coimbatore, 2017, pp. 243-246.

Emanuela Haller, 2013. Traian Rebedea Faculty of Automatic Control and Computers university Politehnica of Bucharest, 978-0-7695-4980-4/13 $26.00 © 2013 IEEE. "Designing a Chat-bot that Simulates a Historical Figure".

McCallum, Andrew K. 2002. "{MALLET: A Machine Learning for Language Toolkit}.

Mrs. Rashmi Dharwadkar, Dr.Mrs. Neeta A. Deshpande, Department of Computer Engineering, D.Y. Patil College, of Engineering, Akurdi, Pune,India, A Medical ChatBot", International Journal of Computer Trends and Technology (IJCTT) – Volume 60 Issue 1- June 2018"

V.Manoj Kumar 2016. "Sanative Chatbot For Health Seekers",JECS Volume 05 Issue 3 March Page No.16022-16025.

****