QUESTION ANSWERING SYSTEM BASED CHATBOT FOR HEALTHCARE

Sharob Sinha[1], Harsh Patni[2]
Under the Guidance of : Prof. Anupam Mondal
1,2 Department of Computer Science & Engineering, Institute of Engineering & Management, Kolkata, India
[1]sharobsinha@gmail.com, [2]hpatni3@gmail.com

Question Answering System Based Chatbot for Healthcare

OBJECTIVE:

- To collect a dataset of all existing diseases along with their symptoms.
- To develop and implement an algorithm that can predicts diseases from the symptoms.
- To experiment on the sample training dataset.
- To compare results and efficiency for validation.

MOTIVATION:

With the increase in population we can understand that it has become very hard for doctors to assist every patient and even a patient could not get immediate help from doctors, individually.

The motivation of this project is to develop an automated chatbot in the domain of healthcare which can provide a virtual assistant. This system can recognize diseases for a set of symptoms that is provided by users and vice-versa.

INTRODUCTION

Chatbot or chatterbot term was introduced by Michael Mauldin (creator of the first Verbot, Julia) in 1994 to describe the conversational programs. The chatbot could help to improve responsiveness, increase availability, and reduce dependence man-power in today's world of automation. Responsiveness presents the quality of reacting quickly and positively at the time of multiple conversations in a particular time. So it is quite possible that a person may not be able to give an immediate response. Hence to improve responsiveness, the chatbot has been introduced by the researchers.

Fig 1: The History of Chatbots

Years ago, the landscape of healthcare was dramatically different from what it looks like today. The chatbot industry is rapidly growing while promising to cost cuts. With the increase in population it has become very hard for doctors to assist every patient and even a patient could not get immediate help from doctors individually.

	Your.MD	Sensely	Buoy Health	Infermedica	Florence
Funds raised	$17.3 MM	$11.8 MM	$9 MM	$5 MM	Unspecified
Year founded	2012	2013	2014	2012	2016
HQ location	London, UK	San Fran., CA	Boston, USA	Wroclaw, POL	Dortmund, DEU
Staff size	49	20	23	26	Unspecified
Target user	Patient, Clinician	Patient, Clinician	Patient	Patient, Clinician	Patient
Types of data processed	Chat (text)	Chat (text), image, video	Chat (text)	Chat (text), image, video	Chat (text)
Est. current users	Unspecified	Unspecified	Unspecified	Unspecified	2,000+?

Data collected from Crunchbase and LinkedIn, March 2018 techemergence

Fig 2: AI-Powered Medical Assistants

DATASET

A proper dataset is essential in building an automated chatbot application. The objective of the project was to develop a chatbot which can recognize diseases for a set of symptoms. Hence, the required dataset was all existing diseases along with their symptoms.

We have collected our dataset from the website "https://www.medicinenet.com/" since it had all the information in an organised way, which was needed for this project.

The final dataset was containing 439 diseases and 1909 symptoms. The final dataset was like figure 3 which contains diseases followed by their symptoms.

	A	B	C	D
1	alcohol abuse	blackout	dizziness	shakiness
2	keratitis	Pain in the eyes	vision disorder	blurred vision
3	infectious mononucleosis	fatigue	fever	chills
4	fungal arthritis	heat	swelling	warmth
5	scarlet fever scarlatina	peeling	roughness	red spots
6	dissociative identity disorder	narcissistic personality disorder	posttraumatic stress disor	dissociative identity disorde
7	osteoarthritis	pain in the joints	hands	hip
8	bocavirus infection	Cough	Wheezing	Fever
9	teen depression	suicide	Apathy	Complaints of pains
10	Excessive or inappropriate guilt	being late for classes	skipping school	Loss of interest in food or c
11	coxsackie virus	a poor appetite	and respiratory illness	including sore throat
12	staph infection	blister	boil	impetigo
13	dry eyes	dryness in eyes	eye discomfort	feeling of something in eye
14	salivary gland cancer	salivary gland cancer	A lump or swelling on or ne	Numbness in part of your fa
15	malaria	fever	headaches	and vomiting
16	gallbladder cancer	particularly in the upper right pc	Abdominal bloating	Fever
17	compartment syndrome	burning or cramping	Tightness in the affected li	Numbness or tingling in the
18	bunions	Swelling	redness or soreness aroun	Corns or calluses
19	hamstring injury	spasm	tightness	and tenderness
20	chickenpox	scab	ulcers	or red spots
21	herpangina	Fever	Sore throat	Small blisters and ulcers ma
22	nasopharyngeal cancer	A lump in your neck caused by a	Blood in your saliva	Bloody discharge from your
23	childhood obesity	bullying	gallstones	anorexia nervosa
24	impetigo	red rashes	blister	itching
25	multiple myeloma	Bone pain	especially in your spine or	Nausea
26	schistosomiasis	Fever	Abdominal pain (liver/splen	Bloody diarrhea or blood in
27	hyperthyroidism	Excessive sweating	Heat intolerance	Increased bowel movemen

Fig 3: The Flowchart illustrating the steps of the working model

MODEL DEVELOPMENT

From the dataset it is observed that on an average, each disease contains around 12 symptoms. Hence, we converted the dataset into a sparse matrix. Using sparse matrices to store data that contains a large number of zero-valued elements (in our case symptoms i.e. 12 out of 1909 on an average), can both save a significant amount of memory and speed up the processing of that data. Sparse is an attribute that we can assign to any two-dimensional matrix.

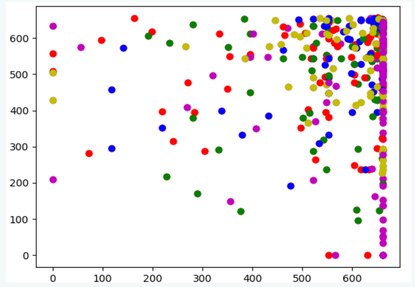

Fig 4: The Training Dataset Visualization

In order to develop the healthcare chatbot, we have employed two well-known unsupervised clustering approach namely K-Means and Mini-Batch K-Means which helps to identify groups of data, which are extracting similar information. This approach assists in learning the module with similar diseases along with their corresponding symptoms, which helps to predict a particular disease for a set of user-provided symptoms. The following algorithm illustrates the module building steps in detail.

Step 1: Initially, we have generated the sparse matrix.

Step 2: Thereafter, we have employed K-Means and Mini Batch K-Means approach on the processed sparse matrix to assign the proper groups of diseases.

Step 3: Initially, we have set the value of K as 7 and incremented the range to 12 due to the nature of our training dataset.

Step 4: On the other hand, the test dataset has been applied to the learned module to validate the proposed chatbot.

FLOWCHART

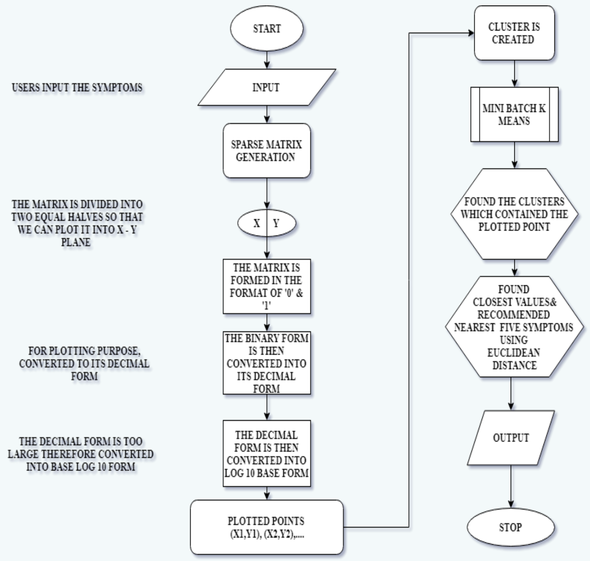

Fig 5: The Flowchart illustrating the steps of the working model

RESULT

INSTANCE	MINI BATCH K MEANS	K MEANS
7	80	74
8	83	78
9	**87**	**84**
10	81	76

Table 1: A comparative result analysis for various K-values under Mini Batch K-Means and K-Means

When 'K' instance was taken as '9'

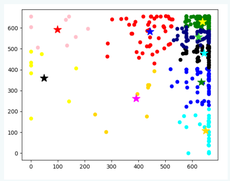

Fig 6: Mini Batch K Means

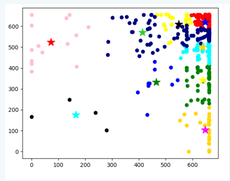

Fig 7: K Means

We have observed that among all the values of K, when the value of K=9, the chatbot provides a better output with respect to different K values.

Taking the K value – 2 to 6, the accuracy didn't get any such significant values. Taking instances K 7 – 12**,** it was found that:

•Using Mini Batch K means, at K – 9 as the instance, the best cluster formation was found to be **(87%)** accurate based on our dataset.

•Using K means, K – 9 as the instance, the best cluster formation was found to be **(84%)** accurate based on our dataset.

It is observed that the chatbot is unable to predict a disease with unknown symptoms. Hence the result and this score motivates to enhance the range of our experimental dataset as well as features in future experiments. The proposed chatbot presents as a baseline system to develop an advanced chatbot in the healthcare domain.

DEPLOYMENT

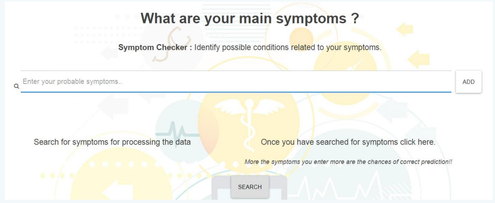

Fig 8: The main user interface of the healthcare chatbot

OUTPUT

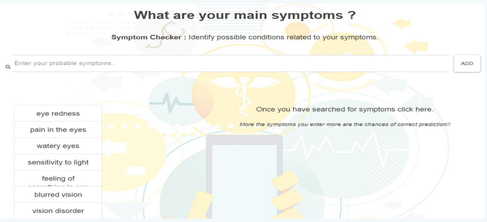

Fig 9: The Symptoms of the disease Iritis is given as input

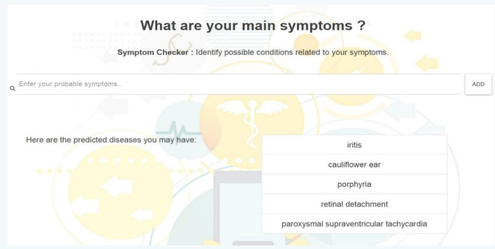

Fig 10: The Result Predicted by the system have the disease Iritis

ACKNOWLEDGMENT

We are sincerely grateful to our advisor and mentor Prof. Anupam Mondal of the Computer Science & Engineering, IEM Kolkata, for his constant support, significant insights and for generating in us a profound interest for this subject that kept us motivated during the entire duration of this project.

We would also like to express our sincere gratitude to Prof. Dr. Satyajit Chakrabarti (Director, IEM), Prof. Dr. Amlan Kusum Nayak (Principal, IEM) and Prof. Dr. Himadri Nath Saha, HOD of Computer Science & Engineering and other faculties of Institute of Engineering & Management, for their assistance and encouragement.

Last but not the least, we would like to extend our warm regards to our families and peers who have kept supporting us and always had faith in our work.

REFERENCES

1. Turing A.M. (2009) Computing Machinery and Intelligence. In: Epstein R., Roberts G., Beber G. (eds) Parsing the Turing Test. Springer, Dordrecht

2. Weizenbaum, Joseph. "ELIZA---a computer program for the study of natural language communication between man and machine." Communications of the ACM 9.1 (1966): 36-45.

3. Meskó, Bertalan, Gergely Hetényi, and Zsuzsanna Győrffy. "Will artificial intelligence solve the human resource crisis in healthcare?." BMC health services research 18, no. 1 (2018): 545.

4. Ng, Andrew. "CS229 Lecture notes." CS229 Lecture notes 1.1 (2000): 1-3.