A NOVEL APPROACH TO PREDICT THE PERSONALITY OF THE DOCTOR USING NATURAL LANGUAGE PROCESSING

Shruthi S Shastry[1], Ajay Suresh Bhat[2], Rahul J[3], Chiranthan M[4], M A Anusuya M[5]

B E Student, Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, India[1]

B E Student, Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, India[2]

B E Student, Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, India[3]

B E Student, Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, India[4]

Associate Professor, Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, India[5]

ABSTRACT: Doctors follow a definite procedure of collecting data of various symptoms or anomalies that a patient experiences in order to analyse the situation and treat the condition. More often than not the patient has vague memories of what and when these anomalies have occurred. We aim to make accurate records of these ailments along with when it occurred to help doctors better understand the circumstance in as little time as possible. With the help of machine learning to classify the client inputs into certain pre-recorded categories and the database functionalities to store and serve the data, the medic aid app is a user-friendly experience which is compatible across various platforms.

Keywords: Personality prediction, unsupervised learning, textblob, vader sentiment analysis

INTRODUCTION

The use of social networks is increasing rapidly.Information about users and what they expressed through status updates are such important assets for research in the field of behavioural learning and human personality. The users' activities on social media provide a great platform for researchers to study and understand their online behaviours, preferences and personalities. A novel approach is proposed to predict the personality of an individual on the basis of his/her social media data in corresponding to their psychological characteristics, which will be useful for society according to their needs.

Healthcare industry generates terabytes of data every year. The medical documents maintained are a pool of information regarding patients. The task of extracting useful information or quality healthcare is tricky and important. By analyzing these voluminous data we can predict the occurrence of the people.

In this project, we aim at addressing some critical issues raised by the use of machine learning algorithms for personality prediction. Personality prediction is more and more important in social networks nowadays, as there's significant correlation between personality and real-world behavior. When we focus on computational recognition of personality, this technique is more and more promising.In the business field,personalized marketing and application design can help customers make great success in gaining more satisfaction in their service.

Objectives

The objectives of the project are :
• To develop a system which predicts the personality of the doctor by analysing their history of interactions with the several patients.
• This prediction provides important implications for improving patient satisfaction and compliance with medical treatment.
• To predict patient preferences regarding their relationships with their doctors.

PROPOSED ARCHITECTURE

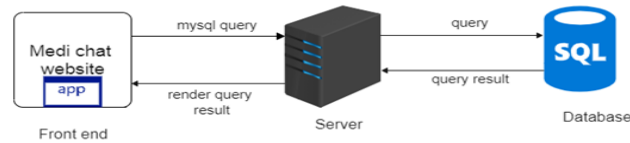The proposed architecture for the above problem statement is shown in figure

Fig 1:System design

Front-end: We use a web application supported by php and xampp by installing phpmyadmin. Web application is used as a source of data for this project. It is also the interface between doctor and patients where they can chat with each other regarding their problems. This message exchanged between doctor and patient is stored in the MySQL database as a dataset.

Back-end : For backend, we use unsupervised learning to predict the personality of the doctor. Backend retrieves the data from MySQL where the data had been stored from the front end and uses that data to predict the personality of the doctor with various sentimental analysis libraries such as TextBlob and Vader and compare the result. In TextBlob we are considering the subjective part of the result since they generally refer to personal opinion, emotion or judgment.

SYSTEM DESIGN AND IMPLEMENTATION

Dataset
The dataset comprises messages exchanged between doctor and patients. We are using the web application developed by us to collect the dataset. Hence the data is valid and not random or fraudulent. With help of this dataset we have an unsupervised machine learning model capable of detecting and assigning probabilities for different types of toxicity such as: Toxic, severe toxic, obscene, threat, insult, identity hate. This dataset is used for training the model.

Search Page
In this page the doctor or patient can search each other to start or continue a conversation. This page consists of a search tab which takes in text and performs a query and creates a session with each other to perform conversation between each other.

Chat Page
When a session is created between doctor and patient, they can chat with each other on the chat page. The Chat page consists of the name of the doctor or patient the user is sending a message. To send a message the user needs to enter the text in the text box and press enter or click on send icon.This process of sending messages for doctors remains the same.

Database design:
The database consists of 3 tables. Specialization table consists of 4 columns: id, name, primary key and unique key.The name field can take the value: Cardiology, Dermatology, General Surgery, Neurology. This basically is used to map the specialization of each doctor and classify the doctors. Users table: The medicine table consists of 9 columns, they are id, role, specialization id, name, phone number, password, primary key, unique key and foreign key. The users table has a foreign key relation with the specialization table. The specialization id in the users table being the foreign key and the medicine id in the specialization table being the primary key. The relation has a cardinality ratio of 1:N. Which means a single record in the specialization table can have N records in the foreign key table (users table). Message table consists of 5 columns, they are id, sent_by, received_by, message, created at. 1 Primary key (id), and 2 foregin key(sent_by and received_by). This table is used to store messages sent by both doctors and patients. As the column name suggests, sent_by column is foreign key for the message table referencing the users table and so is the received_by column. Message columns consist of messages sent by the respective sender and receiver.

Data Preparation
To start with data preparation we query all records in the message table in the database. After we query the data from the database, we create a dataframe which is basically the replica of the SQL table messages. Currently, the data frame is raw from. We convert the raw data frame into UTF-8 format.Along with the message table also get users table from table and similarly convert the data into dataframe and then into utf-8 format. Since we need only messages sent by the doctor, we will be filtering out the messages sent by patients.Next we group the messages by doctor, irrespective of to whom he was interacting with.

Textblob prediction
Textblob is an unsupervised learning which predicts the sentiments of the text message. It provides 2 values: Polarity and Subjectivity. Here we are considering the value of subjectivity because they generally refer to personal opinion, emotion or judgment. And based on the value obtained from the above method we decide whether a doctor is excellent, good,

average and bad. Textblob uses either the Pattern Analyser library and Naive Bayes Classifier to analyse the text. By default it uses Pattern Analyser which we are using in this project.

Vader prediction

Vader is an unsupervised learning which predicts the sentiments of the text message. It provides 4 values: Positive,, Negative, Neutral and Compound. We are considering the compound value since we need the mixture of all the positive, negative and neutral values. And based on the value obtained from the above method we decide whether a doctor is excellent, good, average and bad.

FUTURE WORK

In the future we plan to enhance the model accuracy with more conversational data and once data grows, we plan to switch to supervised learning which makes the learning better. We plan to incorporate adding some other exciting features in terms of the website such as adding an upload file option in order for end users to upload x-ray, prescription etc.one conversations, an option for video calling with the doctor ,and also an option for doctor appointment . We plan on deploying our application in different OS such as IOS and Android, and make it more user-friendly and compatible.

CONCLUSION

Unstructured text classification has gained worldwide attention in the past couple of decades. It is an important step in NLP towards further analysis on unstructured text. In this work, we investigate the classification of online medical articles using linguistic (Semantic and Statistical) analysis. This work shows that incorporating domain-specific terms and keywords can effectively improve the classification accuracy of the machine learning models when it comes to specific domains.We have used two unsupervised learning libraries namely textblob and vader. Which turns out that textblob yields better results when compared to vader. Hence we will be continuing with textblob in the project. We have built a web application for the end-user to get the advantage of the model in terms of a complete product. We have hosted this website http://svachallan.com/healthassist/index.php, so that anyone can access this website and take advantage.

REFERENCE

[1] https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0217-0

[2] Jennifer Ann Golbeck, "Text-based Big Five personality score predictions generated by the Receptiviti API"-AIS Journal, Vol.2-2016.

[3] Ana Carolina E.S.Lima, Leandro Nunes de Castro, "A multi-label, semi- supervised classification approach applied to personality prediction in social media"-Elsevier, Vol 58, October 2014, Pages 122-130.

[4] Danny Azucar, Davide Marengo, Michele Settanni, "Predicting the Big 5 personality traits from digital footprints on social media" –Elsevier, Vol 124, 1 April 2018, Pages 150-159.

[5] Kusal Perera, Samantha Thelijjagoda, Dilini Sewwandi, Oshani Lakchani, "Linguistic features based personality recognition using social media data", IEEE, Vol 3-2017.

[6] Oberlander, Nowson, "The personality prediction by classification and Naive Bayes prediction model"-AIS Journal, Vol-1-2006.

[7] https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0290