NLP based Segmentation Protocol for Predicting Diseases and Finding Doctors

Aswathy K P[1], Rathi R[2], Shyam Shankar E P[3]

[1]M.E. (Computer Science and Engineering), Gnanamani College of Technology, Namakkal, Tamilnadu, India
[2]Assistant Professor, Dept. of Computer Science and Engineering, Gnanamani College of Technology, Namakkal, Tamilnadu, India
[3]Research Scholar, Dept. of Chemistry, University of Calicut, Kerala, India
---***---

Abstract - *In these decades the Web-browsing have become a habitual or imperative thing for the mankind. The people have been using many browsers, online/offline Apps and Software for their daily needs including their healthcare queries. The proper clarification of such healthcare queries and accurate disease-predictions will be very serviceable for the society. In light of this we developed a medical query system which has the ability not only to predict the disease by analyzing the user's symptoms but also to suggest doctors in user's locality if entered the locale while registering. In this paper we proposed a NLP based Unstructured Data Processing Method (NUDPM) which enables the facile segmentation and extraction of medical terms/symptoms from the user queries. The analysis of this NUDPM computed data with a predefined 'Disease data base' will predict the disease accurately. As well as the execution of Doctor-Type selection methods will choose appropriate doctors near to user's locality. Finally the system exhibits the details of the disease and nearby doctors to user.*

***Key Words*: NUDPM, NLP, Disease Prediction, Doctor Suggestion, Health Care Queries**

1. INTRODUCTION

The numbers of internet users have been incrementing very rapidly during these decades. As well as the majority of the people using the internet for clearing their healthcare queries also [1]. Today, there are thousands of internet websites and mobile apps that can be used to access health information. The people have been using these facilities for analyzing/predicting their diseases very commonly. The numbers of people depending this like online information sites for their health care queries are rapidly increasing. This facility generates a new communication channel between patients and healthcare service providers through which the patients discuss their symptoms and health information which leads to the generation of a huge amount of medical data every day. Hence, it will be very useful if we can utilize such medical data effectively. There are many common methods like Data Mining, Artificial Intelligent (AI) and Natural Language Processing (NLP) etc. for the proper extraction of significant health information from such huge medical data.

The current advanced healthcare providers are using Natural Language Processing (NLP) tools and algorithms in majority since they are not only efficient and accurate but also capable of performing more advanced functions. Since the NLP is capable of data mining and analyzing unstructured data within a specific context, it is also used in the extraction of medical data from the 'Electronic health Records (EHR) or Electronic Medical Records (EMR)' as well as the 'Biomedical Texts', where they are in the unstructured data forms [2-4].

In this paper, we proposed a new system 'NLP based Unstructured Data Processing Method (NUDPM)' based on NLP tools and 'Doctor-Type selection method' based on Google Map Application Programming Interfaces (API) to predict the diseases and to suggest the doctors near to user's locality.

2. RELATED WORK

V.Kirubha and S.Manju Priya[5], Survey on Data Mining Algorithms in Disease Prediction, focus on several number of appropriate data mining methods and tools will help to generate accurate data on disease predictions. Thus data mining is more used in medical field.

Dhanashri Gujar[6], Disease Prediction and Doctor Recommendation System, describes a method of data mining, namely, Naïve Bayes Classification algorithm used for disease prediction. It catches the patient medical records as symptoms and to predict disease. It also provides a specialist doctor for the disease.

Rida Sara Khan[7], Artificial Intelligence based Smart Doctor using Decision Tree Algorithm, developed a system using Artificial Intelligent (AI) and this system will interact with patients. The patients enter their health problems into system and system will diagnose the problem by using decision tree algorithm.

Disha Mahajan[8], Prediction System for Diseases and Suggestion of Appropriate Medicines, gives an example of online information website called MedlinePlus and it provides information on health issues. The combination of various types of functionalities of data mining and Artificial Intelligent are used effectively for prediction of disease. This website takes input data in unstructured format, extract the

symptoms and predict the disease as well as suggest suitable medicines.

Yaliang Li[9] , Extracting Medical Knowledge from Crowdsourced Question Answering Website, proposed a Medical Knowledge Extraction (MKE) system which extracts the medical knowledge from a popular website "Crowdsourced Question Answering Website". This system also suggests doctors by using Truth Discovery method.

3. PROPOSED WORK

The existing systems use various types of data mining methods to predict diseases. Data mining method serves as foundation for Artificial Intelligence (AI) and machine learning. NLP is one of the major components of Artificial Intelligence (AI) which provides the ability to perceive the human language (Unstructured data). The proposed system has mainly two sections. The first section is 'disease prediction' which uses this NLP tools to extract the symptoms from user queries and predict the diseases accordingly. The second section is 'doctor suggestion' which uses the Google Map API to suggest suitable doctors near to user's locality. The NLP can enhance the completeness and accuracy of extraction of medical data by the efficient translation of unstructured data into standardized data.
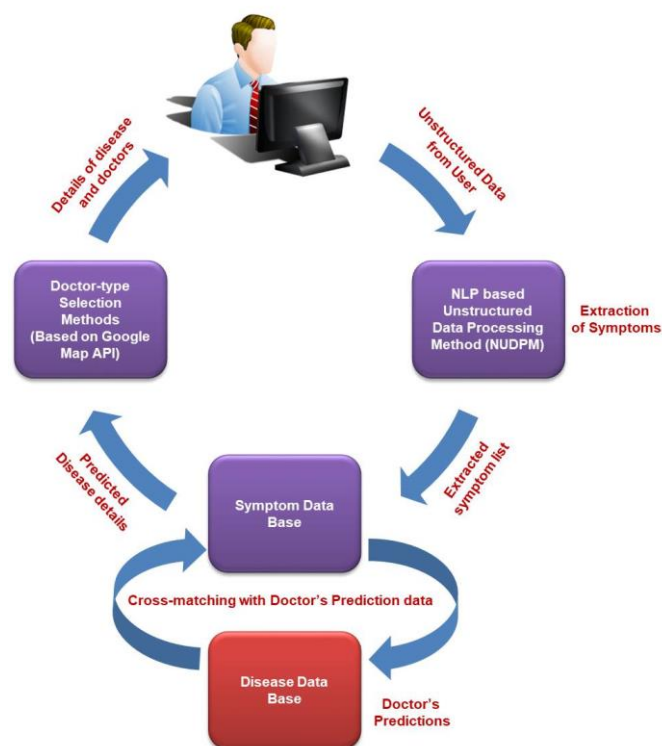
4. IMPLEMENTATION TECHNIQUES

Figure 1. System overview diagram

4.1 Disease Prediction

We have developed a medical query solution system for the exact disease predictions and doctor suggestions based on the Natural Language Processing (NLP) method. In our system, the NLP method will automatically extracts the medical data from the unstructured data (his/her medical query) entered by the users. Here the system is assigned to use various types of NLP tools to extract medical data and which is called 'extraction of symptoms'. This assigned combination of appropriate NLP tools together forms a new method called NLP based Unstructured Data Processing Method (NUDPM). The execution of this NUDPM will extract the symptoms in the query and creates a database called 'symptoms database'. At the same time, there is another database called 'disease database' which consist the list of symptoms and diseases prepared by pre-registered doctors. The system will make cross-matching between the data from both the 'symptoms database' and 'disease database'. Finally, the system will return the predicted disease.

4.1.1 NLP based Unstructured Data Processing Method (NUDPM)

Natural Language Tool Kit (NLTK) is one of the commonly used NLP libraries in Python. It is useful for all sorts of tasks like tokenization, stop word removal, part of speech tagging etc. In this paper, we consider only tokenization and segmentation parts. For extracting symptoms from user queries, we developed new method namely, 'NLP based Unstructured Data Processing Method (NUDPM). The Figure 2 shows the working diagram of NUDPM method.

The user enters their query as input to the system. The query may be a paragraph or a sentence. The query will primarily enter into the tokenization part. The Tokenization process breaks down the 'query' into smaller chunks or words. This tokenized data will process with 'stop word removal part' to filter off the stop words like 'is, this, are, a, an, the etc.' and which is considered as noise in the query.
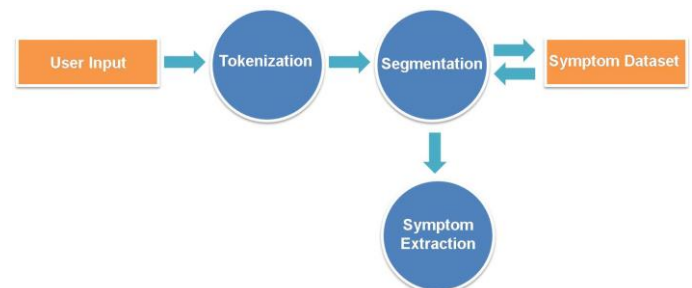

Figure 2. Working diagram of NUDPM

The next is segmentation part. It is the most important part in NUDPM method which creates meaningful phrases from the data obtained after the tokenization and stop word removal processes. And finally, the system will make cross-matching between the segmented data and symptom list in

the 'disease database'. When there is a match observed, such words will consider as symptoms in the user query. These symptoms will be used for further processes.

4.2 Doctor Suggestion

The final output will be more beneficial if the system can suggest doctors near to user's location while predicting their diseases. We integrated this feature to our disease prediction system using the Doctor-Type selection method. This method comprised of a coordinated execution between user/doctor registration part and Google Map API system as shown in Figure 3. The Google API system is an application programming interface provided by the Google which allow us to incorporate the Google map data into our system [10-12]. The registration part will collects the location details of both users and doctors on their registration. These location details will be processed with the Doctor-Type selection method and provide the details of appropriate doctors to the users with the disease prediction.
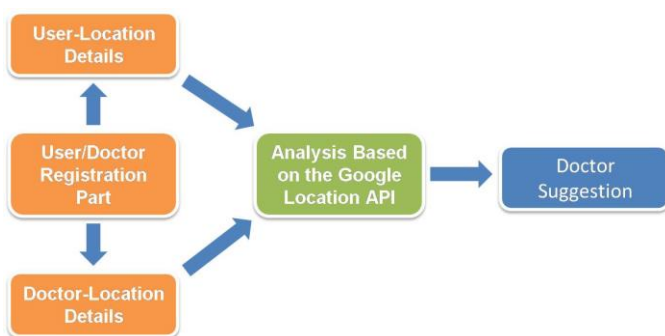

Figure 3. Working diagram of Doctor-Type selection method

5. CONCLUSION

We have developed a perfect system for medical query solutions based on NLP. Our system enables the users to find their diseases and an appropriate doctor near to their own location simultaneously. For enhancing its accuracy and performance we proposed and implemented a method NUDPM successfully. Hence it is sure that our system can serve the users in a better manner than many of the currently available systems.

REFERENCE

1. A. Spink et al., "A study of medical and health queries to web search engines", Health Information and Libraries Journal, vol. 21, 2004, pp. 44–51

2. Y. Wang et al., "Clinical information extraction applications: A literature review", Journal of Biomedical Informatics, vol. 77, Nov 2017, pp. 34-49

3. G. Dhole and N. Uke," Medical Information Extraction Using Natural Language Interpretation", Advances in Vision Computing: An International Journal (AVC), vol.1, March 2014, pp. 19-25

4. A. A. Thomas, "Extracting data from electronic medical records: validation of a natural language processing program to assess prostate biopsy results", World Journal of Urology, vol. 32, Feb 2014, pp. 99-103

5. V. Kirubha and S. M. Priya, "Survey on Data Mining Algorithms in Disease Prediction", IJCTT, vol. 38, Aug 2016, pp. 124-128

6. D. Gujar, R. Biyani, T. Bramhane, S. Bhosale, T. P. Vaidya," Disease Prediction and Doctor Recommendation System", IRJET, vol. 5, Mar 2018, pp. 3207-3209

7. R. S. Khan, A. A. Zardar, Z. Bhatti, "Artificial Intelligence based Smart Doctor using Decision Tree Algorithm", JICT, vol. 11, Dec 2017, pp. 1-5

8. D. Mahajan, M. Phalak, S. Pankore, S. Pathan, D. Abin, "Prediction System for Diseases and Suggestion of Appropriate Medicines", IJIACS, vol. 6, Nov 2017, pp. 348-350

9. Y. Li, C. Liu, N. Du, W. Fan, Q. Li, J. Gao, C. Zhang, H. Wu, "Extracting Medical Knowledge from Crowd sourced Question Answering website", IEEE, 2016, DOI: 10.1109/TBDATA.2016.2612236

10. Maps JavaScript API-Google Developers. http://developers.google.com/maps/documentation/javascript/tutorial

11. J. Cui, X. Wang, "Research on Google map algorithm and implementation", Journal of Information and Computational Science, vol. 5, May 2008, pp. 1191-1200

12. S. Kobayashi, "A Geographical Information System Using the Google Map API for Guidance to Referral Hospitals", Journal of Medical Systems, vol. 34, Dec 2010, pp. 1157-1160