ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 11, November 2016

Framework for Smart Medicine Advisor Dictionary using Natural Language Processing

S.Pooja, Mrs.Priya Vijay

***Abstract*- Smart Medicine Advisor is an End-User support system that provides instant guidance to the patients on their health issues. Databases containing various symptoms, disease and medications are created from discharge summary using Natural Language Processing technique Named entity recognition to create medical dictionary. This system allows user to share their symptoms and issues. It then analyzes the user's symptoms to check for various diseases associated with the patient's symptoms and predict the disease and it also provides medication for the disease. It also checks for contradiction between medications.**

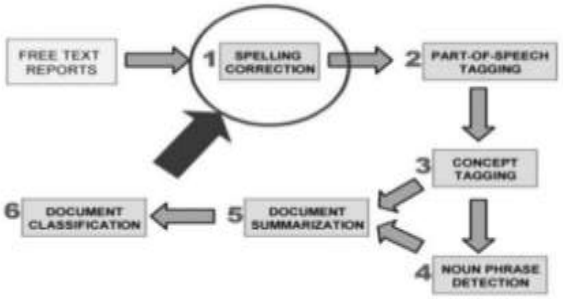
***Index Terms*: Natural language processing, Named Entity Recognition (NER).**

I. INTRODUCTION

Medical Expert systems are designed to provide advice or diagnosis based on past history automatically available without manual intervention as a software tool or web blogs. The symptoms, disease and medications can be formulated as a dictionary to provide data set for decision making based on patient's information. To create the medical dictionary discharge summary can be chunked and processed to create database using natural language processing. Then past history can be analyzed using data mining techniques to advice the patients.

Data mining has huge and complex volumes of data to be used by healthcare activities. In day to day life every traditional system are getting automated to proceed on their own with any manual intervention. In health care industry, data mining techniques can be used to mine information from patient records and to generate report or expert advice in medicines or health practices that is useful to all stakeholders. To create dataset for healthcare discharge summary can be processed to create medical dictionary. Natural language processing helps in text analysing and processing to break apart the data from the text document. In principle, natural language processing could extract the facts needed to activate many kinds of decisions rules. Named-entity recognition (NER) is a sub module of information extraction that pursues to locate and categorize named entities in text into pre-defined categories which includes the organizations, names of persons, expressions of times, locations, quantities etc.

The combination of data mining and NLP together provides the best results. The diagram [https//:slideshare.in] is depicted below



The following is the data flow that takes place in the Natural language processing. Initially The free text reports are given as input where the spelling mistakes are corrected and the concept tagging, noun phrase detection is done and further summarization of the document is done. Finally classification of documents are done.

Concept tagging and classification is used in the proposed system. Concept Tagging implicates streaming of text along with a concept dictionary , taken from an ontology that defines a particular domain which in turn provides the output as a sparser stream of concepts . They are used to find synonyms, related concepts, word sense disambiguation, clustering, summarization etc. The stream of concept is useful for relationship extraction.

Document summarization can be done using two methods, they are Extractive and abstractive summarization. Extractive is the technique in which summaries are created by extracting the relevant sentences from the documents. The documents are ranked according to the score obtained by the sentence in the document so that only relevant documents are chosen in a non-redundant manner. Abstractive summarization creates the assumption that a good summary should be constructed without using the same sentence from the document but in by generating new sentences that contain the sense and knowledge limited in the documents.

In today's high-speed world, since everything is done through online there is a huge growth in number of online health seekers. Since the mobile devices and applications

Manuscript received Oct, 2016

***S.Pooja**, PG Student, Department of Computer Science and Engg, Rajalakshmi Engg College, Chennai,India*

***Priya Vijay**, Assistant professor, Department of Computer Science and Engg, Rajalakshmi Engg College, Chennai,India*

play a vital role health seekers look for medical applications that could provide instant guidance to them without physical consultation. Since there is a lack of good expert system a system called smart medicine advisor is proposed here. In this work the advancement of expert system has been made by using natural language processing technique in which named entity recognition technique is employed for creating database. This system extracts information from discharge summaries of various patients and presents this information in structured form using a controlled vocabulary. These extracted data is used for prediction of disease and provides the medication for it. In addition this system also allows users to check the contradiction between medicine**s.**

The paper is organised as follows. Chapter II narrates literature survey. Chapter III describes the proposed model.

Related Work

Taranath N.L,Shantha Kumar BPatil,Premajyothi Patil,C.K.Subbaraya [1]had developed a medical decision support system that provides a medical decision support for doctors, nurses and other medical staffs for the patient with missing data. Which had disadvantage of medical contradiction which is addressed in the proposed system

Liqiang Nie, Meng Wang, Luming Zhang[2], Shuicheng Yan had proposed a system that studies the information needs of health seekers in terms of queries and then select those that ask for possible diseases of their manifested symptoms for further analytics. Using Learning techniques it also finds the possible related diseases to the symptom. It provides only community based service.

Em Tekin, Onur Atan, Mihaela Vander Schaar[3] had discovered a expert selection system that prescribe the patient through online based on the context of the patient.A new class of algorithms discovers the most relevant context and discovered the best clinic and expert to get advice to make a Decission with patient's contexts. It diagnosed the disease and suggested medical experts

Oana Frunza, Dina Inkpen, Thomas Tran [4] described a Machine Learning constructed methodology for building an application to classify and publicize diseases. Authors extracted data from various existing medical reference papers and identified the semantic relationship between treatments and diseases. It identifies the relationships.

Gayle McElvain, Anand Madhavan [5] developed a system that identified and disseminated health care information. It also extracted the sentences that has mentioned treatments and disease and identifies the semantic relation between them. It annotates the medical terms in discharge summary

Carol Friedman, Xiao Zeng, Lyudmila Shagina [6] discovered a system that annotates the medical terms in discharge summary. When Submitting a Medical text, it in turn processes and provides the structural text back.It Extracts medical terms from the medical data.

Sune Pletscher-Frankild , Albert Palleja, Kalliopi Tsafou , Janos X. Binder and Lars Juhl Jensen [7] presented a system for disease gene association system for extracting disease–gene relations from biomedical abstracts, which had a

Dictionary efficient with named entity tagger for recognition of human genes and diseases. It is combined with a score pattern that considered co-occurrences between sentences and the disease recognition. They have developed a Named entity recognition tool to create dictionary for disease ontology and combined it with a co-occurrence scoring scheme to efficiently and precisely extract disease–gene relations from Medline

P.Shirisha [8] proposed a work to show what Machine Learning (ML) and Natural Language Processing (NLP) techniques used for depiction of information and what classification algorithms are apt for detecting and categorizing relevant medical information as short words. This paper narrated the machine learning and Natural language processing techniques to extract knowledge from medical documents.

Abdur Rehman, Haroon.A.Babri, Mehreen saeed [9], developed a system to extract the relation between medical terms automatically using a dictionary of medical terms which is aided by sentence classification of the medical report. The semantic parser is used to parse the sentences automatically. Author used four classifications techniques viz. Adaptive Decission tree, SVM, Naive Bayes, and CNB to mine words. Authors depicted the relationship between two medical terms only but they doesn't mine medical data.

Aida Bchir, Wahiba Ben Abdessalem Karaa [10] suggested techniques for extracting relationship between drugs and diseases . Author''s first step is to employ Natural Language Processing techniques to preprocess the abstracts and a second step is to extract a set of features from the preprocessed summaries which Lastly extracted a disease-drug relation using machine learning classifier

In the existing systems to match a document against the dictionary they have used tagging algorithm in c++ in which the tagging speed and the memory efficiency is less.So,In proposed system the primary learning algorithm SVM is implemented. Perceptron Algorithm with Uneven Margins is also encompassed to rival SVM with much reduced time.

II. PROPOSED ARCHITECTURE

In this proposed work NLP technique is used to extract the the disease symptom and medicine relationship through a technique called named entity recognition.

A. DICTIONARY CONSTRUCTION:

To construct a dictionary of diseases, medicines and symptoms for use in Named Entity Recognition (NER), we extract all names and synonyms from the Disease Ontology. This dictionary is used for the comparison of symptoms given by the patients for prediction.

Dictionary creation is a subtask in Named entity recognition in which we create a medical dictionary in which all the

medical ontological terms are been fed.This dictionary is given as a input in the program to identify and tag the medical terms in the document.The tagger tags the name of disease ,symptoms and drugs which are moved to the database for further classification. In the proposed work SVM is used to do named entity recognition. The algorithm is given below as figure 2

Algorithm 1 Simple SVM

```
candidateSV = { closest pair from opposite classes }
while there are violating points do
  Find a violator
  candidateSV = candidateSV ⋃ violator
  if any α_p < 0 due to addition of c to S then
    candidateSV = candidateSV \ p
    repeat till all such points are pruned
  end if
end while
```

Fig 2: SVM algorithm

B. INFORMATION EXTRACTION:
After dictionary creation the next task is to extract information on associations between symptoms, diseases and medicine. For this NER is used and the information extracted is stored in the database.

C.SYMPTOM-DISEASE-DRUG DATABASE:
Several existing databases contains disease– drug associations, mainly obtained through manual curation of medical literature.Here we create our own disease –drug association database from discharge summary of patients.

There are two parts one is information extraction and the other is user interface part. In Information extraction (IE) the discharge summary is given as input the given discharge summary undergoes text processing. A named entity recognition task is performed and the disease, symptoms and drugs are extracted from each discharge summary and the results are stored in databases.

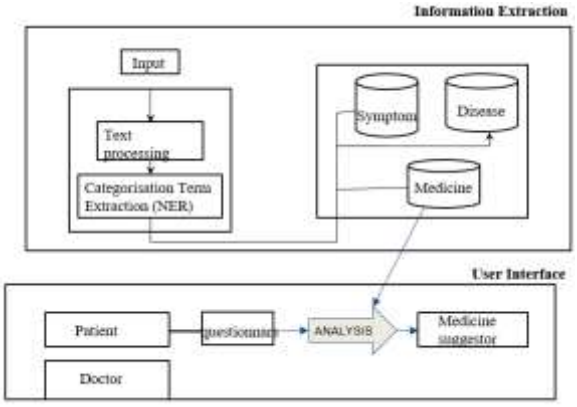


Fig3.Proposed Architecture

Named entity recognition (NER) involves identifying the limitations of the name in the text and understanding it's meaning, often through mapping the entity to a exclusive concept identifier in an appropriate ontology. Dictionary based NER is used in the proposed work. As name itself suggests, for Dictionary based NER, resource document with list of names for all entity type as dictionary is needed for recognition of the text. So dictionary is created.The user interface part uses clustering and other data mining techniques to predict the disease according to the symptom provided by the patients and suggest necessary medicines to them. In addition it also checks for contradiction between medications.

III. CONCLUSION

This approach is very useful for constructing e-health data dictionary for customers as it provides advice online through an application which in turn saves their time. It also has an added advantage of contradiction checking. The data dictionary can be further used by data mining techniques to provide expert system for medication. The future work of the system is to develop a tool to provide customization for the patients along with the advice about the medication for the patients.

REFERENCES

1] . NL Taranath,Shantakumar B Patil , Premajyothi Patil; C. K. Subbaraya ,"Medical decision support system for the missing data using data mining" ,IEE 2015CEM
[2] Liqiang Nie, Meng Wang, Luming Zhang, Shuicheng Yan, Bo Zhang, Senior and Tat- Seng Chua,"Disease Inference from Health-Related Questions via Sparse Deep Learning" IEEE"2014
[3] Tekin, Onur Atan, and Mihaela van der schaar, "Discover the Expert: Context-Adaptive Expert Selection for Medical Diagnosis",June 2015
[4] M. M. Abbasi, S. Kashiyarndi, "Clinical Decision Support Systems: A discussion on different methodologies used in Health Care," International Journal of Computer Science and Information Security, Vol. 8, No. 4, 2010.
[5] OanaFrunza, Diana Inkpen, and Thomas Tran, "A Machine Learning Approach for Identifying Disease-Treatment Relations in Short Texts," IEEE, VOL. 23, NO. 6, JUNE 2011.
[6]Friedman and Clayton PD. Natural language processing in an operational clinical information system. Journal of Natural Language Engineering, 2011.
[7] Sune Pletscher-Frankild , Albert Palleja, Kalliopi Tsafou , Janos X. Binder,Lars Juhl Jensen ; "DISEASES: Text mining and data integration of disease–gene associations" ,ELSEVIER, December 2014.
[8] P.Shirisha; "ML-Based Approach for Identifying Disease-Treatment Relations in Short Texts" International journal and magazine of Engineering, Technology, Management and Research. Volume No:2, Issue No 11(November 2015)
[9] Abdur Rehman, Haroon.A.Babri, Mehreen saeed,"Feature Extraction Algorithm For Classification Of Text Document" December 2011.
[10]C. Giuliano, L. Alberto, and R. Lorenza, "Exploiting Shallow Linguistic Information for Relation Extractionfrom BiomedicalLiterature," Proc. 11th Conf. European Chapter of the Assoc. for Computational Linguistics, 2006.