MEDICATION RECOMMENDATION SYSTEM BASED ON CLINICAL DOCUMENTS

Ayrine John
Department of Computer Science
and Engineering,
College of Engineering, Cherthala
Kerala, India-688541
ayrinesusan@gmail.com

Muhammed Ilyas H
Department of Computer Science
and Engineering,
College of Engineering, Cherthala
Kerala, India-688541
ilyashm@gmail.com

Veena Vasudevan
Department of Computer Science
and Engineering,
College of Engineering, Cherthala
Kerala, India-688541
veenavsd@gmail.com

***Abstract*—Designing medication recommendation system is a need for the fast growing world.In this fast growing world, the need for the application which recommend a medication led to a doctor friendly and hospital free atmosphere for all users all over the world.In this paper an unified extraction system with stanford parser is used for extraction of medical terms.Then K-means clustering algorithm clusters the diseases and a filtering method find out the desired medication.**

***Index Terms*—NER, NE class, SVM, FuzzyClinical text,Recommendation, Clustering, Extraction**

I. INTRODUCTION

Recommender system is a typical system that helps users to get a suggestion of items which they can use for their accurate need.The health facts provided of each individual are evaluated and extracted using recommendation systems. Unlike many other types of recommendation systems, health recommendation mainly depends on emotional, physical and psychological matters of the patients. Medication Recommendation system is a similar system that recommend the medicines for a particular symptom or disease.This system will goes through the history of the patients and the drugs they have used earlier.From the detailed investigation the system will reach to the recommendation of the medicine she/he deserves.

Due to the increased rate of information produced in the biomedical domain, and due to the crucial impact of such information act on research and upon real world applications, there is a particularly great and growing requirement for medication recommendation systems that can effectively and efficiently aid biomedical researchers and health care professionals.The data given to the medication recommendation system is clinical text.Clinical documents are free-text data sources containing valuable tablet names and symptom information, which can be analysed to improve health requirements. Clinical documents such as clinical notes contain facts about patients,such as medication conditions usually diseases, injuries, medical symptoms and responses such as diagnoses, procedures, and drugs and the time period for the medication etc.Medication Recommendation system have the ability to examine the medical terms and classify that terms to clusters on the basis of symptoms.

Two important types of facts that can be taken from a clinical note are symptoms and medications[1].Diseases,syndromes, signs, diagnose, etc. are symptom related informations, which can be used to examine diseases for patients.Alongside with that, valuable medication information is commonly embedded in unstructured text narratives which will check the multiple segments in the clinical document.

Collaborative filtering and content-based filtering are the two major methods used in recommendation of the system.

A. *Collaborative filtering*

The basic objective of collaboartive filtering[10] to go beyond the experience of an individual user profile and an alternative to use the experiences of a population or community of users. These systems outline with the assumption that a good way to find captivating content is to find people with similar tastes and to propose items they like. Typically, each user is affiliated to a set of nearest-neighbour users, comparing profiles information. With this method, objects recommendations are based on resemblance of users rather than the similarities of objects.

B. *Content-Based filtering*

The Content-Based filtering method [9] based on the textual filtering method.In content based systems, the user profile is rated. The profiles and realm documents are then used as input to the classification algorithm. The documents which are resemble (in content) to the operative user are considered interesting and are recommended to the user.

The recommendation systems available now have lot of week points.The major thing is that when a rec-

978-1-5090-1987-8/16/$31.00 ©2016 IEEE

ommendation is done,the medicine allergic to patient is not considered,only recommending a medicine for a particular disease.Next,once a medicine is recommended, the medicines cost and age of the patient ia not considered in the recommendation.Third,most of the recommendation systems are online applications,need internet access for searching the medicines.The medication recommendation system, the data are examined from clinical documents and hence online retrival of medicines and clinical related information is not needed.

II. Previous Work

Yuan Ling, Xuelian Pan, Guangrong Li, Xiaohua Hu[1]proposed, clinical document clustering based on medication/symptom names using Non-Negative Matrix factorization.The main benifit of this system include the clustering of the symptoms according to the medication for a group of similar symptoms. The main computing methods of the this system are based on the application of natural language processing (NLP) techniques to theorize the medication terms and symptoms from clinical document.The main disadvantage of the system is recommendation of medicine from the cluster is not mentioned here.

Sigfried Gold,Noemie Elhadad,James J Cimino [2] proposed a method which extracts medication events from discharge summaries.The major methods used in the system are first, the concept is defined which is to be extracted.Then the parser is build and parsing rules are generated.A test data is created to test the generated rules.It should be convey with physicians to take the final decision about medicines.Then the test data is tested and scores are defined.

Hua Xu, Shane P Stenner and et.al [3] proposed a semantic representation model for prescription type of medication findings.It mainly concerned with algorithms to find the medication names and their signatures.Two main components were discovered by analysing the clinical text.They are Med finding and Sig modifier. Med finding mainly represents the drug names and signature modifier represents strength,route,frequency etc.The major steps in the system are pre-processing,semantic tagging and parsing.

Thierry Hamon,Natalia Grabar [4] proposed a rule based approach for extraction.There are three main steps for the system.They are pre-processing,processing and post processing.

Olga Patterson,John F Hurdle [5] proposed a charactterisctic based on vocabulary and semantic type of clinical domain used in both in and out patients.Document clustering commonly used unsupervised text mining technique.The goal of document clustering is to find set of natural patterns within a bulk amount of unlabelled data inside the document and then organize similar documents.

Noha E Negm, Passent Elkafrawy and et. al [6] proposed a knowledge based medical clustering with assosiation rule mining.Assosiation rules are generated from remarkable terms that are frequently occuring.There are mainly four stages for KMDC.They are:

- online query submission
- text representation and preprocessing
- mining assossiation rules using MTHFT algorithm
- clustering pubmed abstracts

Martin Wiesner,Daniel Pfeifer[7] proposed a recommender system that supply patients a friendly information to comprehend their health status.The suggestions done froma a health recommendation system is done from a individualized health data documented in personal record.Data entries in a PHR database constitute the medical history of PHR owner.HRS will search relevant items of interest for the target user.Such items originate from health knowledge base repositories and displayed online while he/she inspects.

Sharique Hasan,George T Duncan, Daniel B Neill, Rema Padman[8] proposed a automatic detection of omissions in the medication list,identifying the drugs that the patient's are taking but not in the medication list. In this paper describes the specific drugs have been omitted from individual's medication list based on known medication of similar individuals.Different types of algorithms are used in collaborative filtering method. In popular algorithm each drug entity is searched in the list and count the remaining elements in the list and score of each drug is calculated.this is possible only for a small list of drugs.Accompaniying counting scores the entity not present in the observed partial list according to the number of times it has co-occured.Then score is calculated.In K-nearesst Neighbour method, given a partial list, there will be K training list that are closest to some distance metric.Scores for the missing entities are assigned using majority vote of the K-nearest neighbours.

III. Proposed System

For a medical recommendation system , the ultimate focus on the clinical document is the input to the system. Aim is to extract the medical entities and cluster them based on medication and symptoms and then recommend the desired.Fig 1 shows the overall architecture of proposed system
The major modules of the system are

- Extraction module
- Clustering module
- Recommendation module

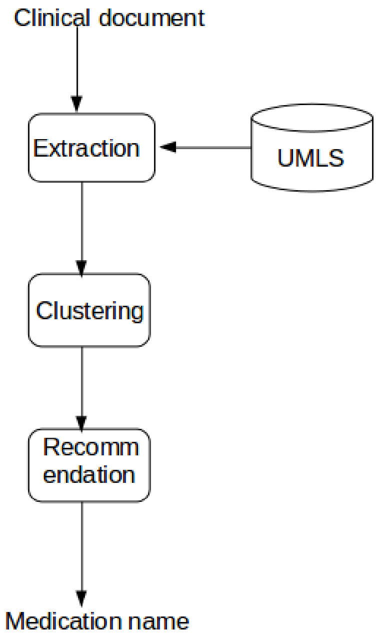

Fig. 1: Architecture of proposed system

Fig 2 shows the proposed architecture, which takes a clinical document as its input.Then after the preprocessing step, the system will produce a structured output with the fields age,date,medication,symptoms,history etc.Firstly the symptoms are extracted from the document based on context by analysing the clinical document.Based on the symptom the clinical documents are clustered.This cluster contains group of text documents which contains almost similar symptoms.From that extraction of the medication is done based on context.After extracting the medication names recommendation is done using collaborative filtering method .

A. Extraction Module

In this module, extracting the medical terms using context words.These context words are taken using a standford parser.A stanford parser[12] is a program that works out the grammatical contents in a sentence.Its helps to tag the words as noun,verb,noun phrase etc.So here analysing a text will be done by standford parser and each sentence in the text is separated into Noun,Verb,Noun Phrase,Adjective etc.That means each word is tagged as Noun,Verb,Noun Phrase etc.Mostly the medical terms will be Nouns,Noun Phrase and adjectives.These words are only taken from text and checked with UMLS database.

UMLS database [11] checking is done by taking each terms checking whether it is a medical term associated with any disease.For each term associated with a disease will retrieve a CUI(concept unique identifier) and this CUI is searched in mrconso table which will retrieve the associated diseases with the medical terms.

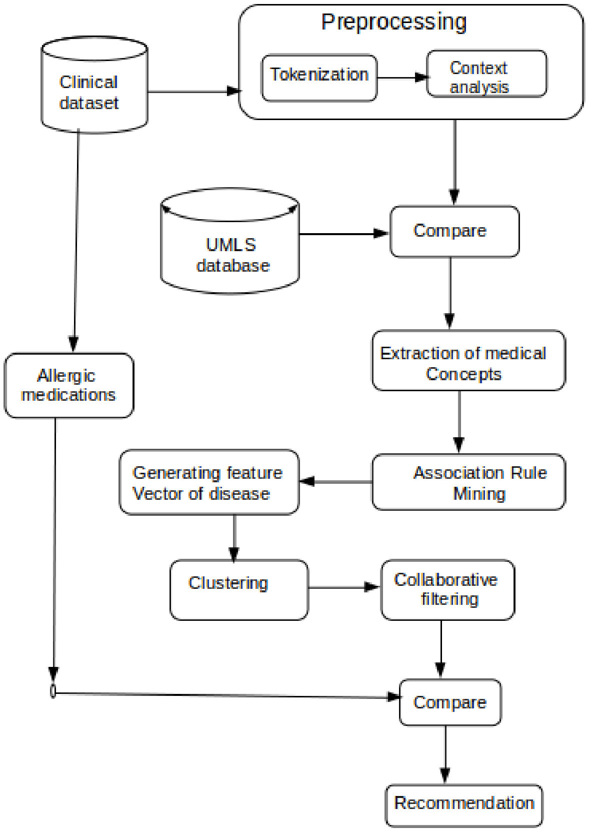

Fig. 2: Detailed Architecture

Symptoms are identified by making a symptom list and find out the frequency of occurence of symptoms in the clinical text as well as in the UMLS concepts. The most frequently occuring symptom will be the symptom of the disease associated with the clinical document.

B. Clustering Module

In this module association rule mining and clustering of medical terms are taken.Association rule mining is the method of finding interesting patterns in large databases.Apriori algorithm is used for association rule mining and which will find the minimum support count of diseases and frequency of occurence of each terms in the database.After mining process features are identified and clustering is done based on the apriori algorithm. K-means algorithm is used for clustering the clinical documents.

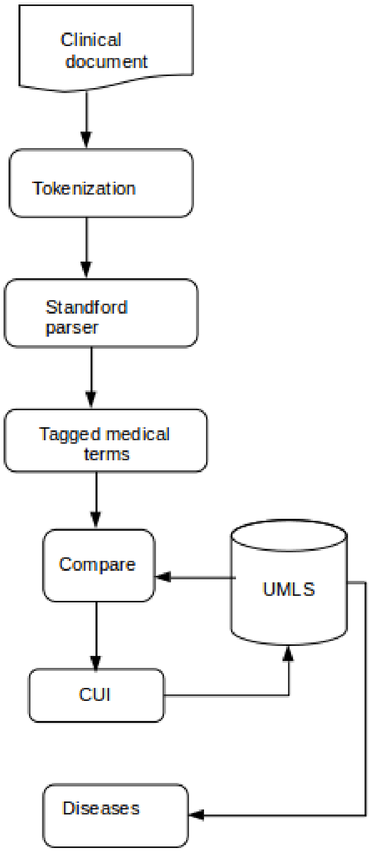

Fig. 3: Architecture of proposed system

Algorithm 1 Algorithm to find symptoms

```
Input: Clinical text
IF clinical doc is a text
    Split the clinical doc using delimiters(sdoc)
    IF sdoc "NN","VP","NP","JJ"
        add the next possible context words(medical terms) to array list (AL1)
  Check AL1 with database values to get associated medical terms and diseases
  Return symptoms
  Else
      Return empty set
Else terminate
```

Algorithm 2 Algorithm for association rule mining

```
Input: Candidates itemset of size k ($C_k$)
Frequent itemset of size k ($L_k$)
$L_1$:frequent items
For $k = 1; L_k! = \phi; k + +$
    do begin
     $C_{k+1}$ =candidates generated from $L_k$
    for each transaction in database do
        increment the count of all candidates in $C_{k+1}$ that are contained in t
        $L_{k+1}$ =candidates in $C_{k+1}$ with minimum support
    end
return$U_k L_k$
```

The removal of multiple entries from the feature matrix is performed. There are chances for multiple entries of same data in the data sets. So those features which were already added into the feature matrix on future encounters are removed. The medical information cannot be subjected to stemming. It cannot remove stopwords either. This is due to the reason that certain medical terms have similarities with these data and hence applying stemming or stopwords will result in the loss of such medical terms from the data set. This will therefore cause the dataset to become imperfect. So as a measure to avoid such losses the processes steming and stopwords are not performed in the medical data.

If the result of the comparison is true then such a medical term which might be a disease or a feature(or symptom) of a disease is stored as medical features. Such extracted medical features are then subjected to Association rule mining [13]. This is done to find the support and confidence of each disease corresponding to its features and also corresponding to other disease. If a disease is found to provide confidence to another disease then that disease is added as a feature of the other disease. Thus a matrix is formed from the informations obtained from the Association rule mining. This matrix is frequently subjected to updations to make feature matrix more accurate. On the completion of mining the feature matrix information is used to make clusters of correlating data.

Algorithm 3 Algorithm to cluster using K-Means

```
Input: Clinical document
For $i = 0; i < Vectorspace.size()$
     Sim=0
     For clusterkeyset
         csim=probability of each candidate of vectorspace.get(i)
        IF$csim > sim$
         sim=csim
         centre=csim
Add cluster centres
Else
Return empty set
```

C. Recommendation Module

n this module recommendation of the medication is done. Collaboartive filtering is the technique used for recommendation. For that each document cluster is compared with the other cluster and the similarities between the diseases is analysed and find out the best recommendation from the filtering process.The work flow of a collaborative filtering is discussed below.
1.A user express his/her intention of knowing the tablet for a particular disease in the application.
2.The system matches this users ratings against the knowledge database and finds the disease he/she searches.
3. Similarly that disease will be searched in the repository and retrived the relevant tablet he/she needs.
4.Also if any allergic tablets are found in the medication list, then the application will automatically intimate the user about the caution of allergic disease.

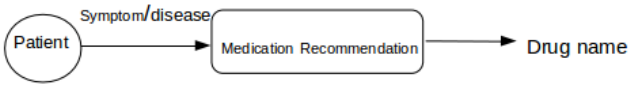

Fig. 4: Recommendation

IV. Experiments and Analysis

Experiments are done on different transcriptions.The result found by analysing the contents are done using confusion matrix,precision,recall,accuracy.

A true positive(TP)[14] indicate assigning of two similar documents to the same cluster, a true negative(TN) indicates assigning two dissimilar documents to different clusters.There are two types of errors.A (FP) decision assigns two dissimilar documents to the same cluster.A (FN) indicates assigning of two similar documents to different clusters. Here investigation is done based on pecision and recall. How much tablets are correctly retrived and how much tablets are wrongly retrived.
precision=$TA/(TA+FA)$
recall=$TA/(TA+FB)$
accuracy=$(TA+TB)/(TA+TB+FA+FB)$

TABLE I: Relevant and Retrieved Tablets

	Tablets retrieved	Relevant tablets
Diseases given	40	35
Symptom given	38	30

Based on analysis 50 diseases are given. Out of the 50 diseases 40 are retrieved relevant. Hence precision will be 80% and recall will be 87.5 % since retrived ones are all relevant.So accuracy will be 80 %.Precision and accuracy increases with increase in training data.While giving symptoms 35 are retrieved out of that 30 are relevant.So precision will be 76 % and recall will be 78 %.

V. Summary and Conclusions

In this approach present an idea for medication recommendation system for medical inquiry. This approach is based on four main steps: (i) analysis of clinical document (ii) retrieval of relevant medical terms and (iii) Clustering the medication and symptoms. iv) Recommend the proper medication. The proposed system can also work as a tool for supporting the doctors in their disease diagnosis.As future work efficency of recommendation system can be increased by including age of the person,demographic informations during the training phase.Also the brand and the chemical contents available in the medicine can improve the recommended medicine.

References

[1] Yuan Ling, Xuelian Pan, Guangrong Li, Xiaohua Hu, *Clinical Document Clustering Based on Medication/Symptom names using Multi-view Non-negative Matrix Factorization*, IEEE Transactions on NanoBioScience, 2015.

[2] Sigfried Gold,Noemie Elhadad,James J Cimino, *Extracting Structured Medication Event Information from Discharge Summaries,* JAMIA symposium proceedings, 2008.

[3] Hua Xu, Shane P Stenner,Son Doan,Kevin B Johnson, Lemuel R Waitman, Joshua C Denny, *Medex:a medication information extraction system for clinical narratives,* JAMIA symposium proceedings, 2009.

[4] Thierry Hamon,Natalia Grabar, *Linguistic approach for identification of medication names and related information in clinical narratives,* JAMIA Symposium proceedings, 2010.

[5] AOlga Patterson,John F Hurdle, *Document Clustering of clinical Narratives:A systematic Study of clinical Sublanguages,* JAMIA Symposium proceedings, 2011.

[6] Noha E Negm, Passent Elkafrawy,Mohamed Amin,Abdel-Badeeh M Salem, *KMDC:Knowledge based Medical Document Clustering System using Association Rules Mining,* International Journal of Bio-Medical Informatics and e-Health,vol 1 August 2013.

[7] Martin Wiesner,Daniel Pfeife, *Adapting Recommender Systems to the Requirements of Personal Health Record Systems,* ACM 978-1-4503-0030-8/10/11, November 2010.

[8] Sharique Hasan,George T Duncan, Daniel B Neill, Rema Padman, *Towards a collaborative Filtering Approach to Medication Reconciliation,* JAMIA Symposium proceedings,2008.

[9] http://recommender-systems.org/content-based-filtering,

[10] http://recommender-systems.org/collaborative-filtering/

[11] UMLS Reference Manual Bethesda (MD): National Library of Medicine (US)-2015AB Release Information http://www.nlm.nih.gov/research/umls/

[12] Standford parser:http://nlp.stanford.edu/software/lex-parser.shtml

[13] Association rule mining: http://searchbusinessanalytics.techtarget.com/.../

[14] http://www.kdnuggets.com/faq/precision-recall.html