National conference on Engineering Innovations and Solutions (NCEIS – 2018)
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 4 | Issue 6 | ISSN : 2456-3307

Opinion Based Learning Model in Medical Sector

Divya Y A
Department of Information Science and Engineering GSSS Institute of Engineering and Technology for Women, Mysuru, Karnataka, India

ABSTRACT

Online health communities continue to offer huge variety of medical information useful for medical practitioners, system administrators and patients alike. In this work we collect real time health posts from reputed websites, where patients express their views, including their experiences and side-effects on drugs used by them. We propose to perform Summarization of user posts per drug, and come out with useful conclusions for medical fraternity as well as patient community at a glance. Further, we propose to classify the users based on their `emotional state of mind'. Also, we shall perform knowledge discovery from user posts, whereby useful `patterns' about the triad `drugs-symptoms-medicine' is done by Association Rule Mining.

Keywords: Association, Data Mining, Summarization, Classification

I. INTRODUCTION

Online health communities offer huge variety of medical information useful for medical practitioners, system administrators and patients alike. For knowledge mining of the health posts, we propose to apply different important operations like - Association Rule Mining, Summarization and sentiment analysis on data obtained from the health forum.

Summarization is nothing but taking information from the source, extracting content from it, and presenting the most useful content to the user in a condensed form and suitable to the user's application needs.

Association rule mining is a popular and widely-known data mining task. It is used to find out interesting relations between variables in large database. Rules generated by association have two disjoint set of items having form LHS (Left Hand Side) => RHS (Right Hand Side). The rule says that RHS is likely to occur whenever the LHS set occurs. After the rules have been obtained, they are extracted and post processed.

The extracted rules from the health community dataset could take one or more of the following form

1. symptoms - disease
2. disease - disease
3. medicine - disease
4. disease – medicines

Finally Sentiment Analysis or Opinion Mining is task of finding sentiments from text. These sentiments take different forms like – opinions from people, attitudes and emotions toward an entity.

II. MOTIVATION

The doctor should give the proper treatment for the patient's disease. As multiple drugs are available for the particular disease, there is need of identifying the

popular drug. As symptoms are related to disease and diseases are related to drugs, there is a need for the system, which discovers the relationship between symptoms-diseases-drugs. In the proposed system, we are achieving this based on the patient opinions.

III. PROBLEM DEFINATION

Health communities' just collects real time health posts, where patients express their views, including their experiences and side effects on drugs used by them. These systems just collect the data, stores in database and retrieve the same in future but no extraction of useful information is done. The proposed system is performing summarization of user posts per drug, and come out with useful conclusions for medical fraternity and prediction of disease based on symptoms.

IV. SYSTEM ANALYSIS

A. Existing System

Health communities just collect real time health posts, where patients express their views, including their experiences and side-effects on drugs used by them. These systems just collect the data, stores in database and retrieve the same in future, but no extraction of useful information is done.

B. Proposed System

Proposed system is a medical sector application. It collects the posts from the users (medical practitioners, patients) related to side effects on drug and summarizes all the user posts and come out with useful conclusions. The outcome is to discover the patterns by Association Rules and also finding sentiments from the opinions.

V. METHODOLOGY

We have identified different modules like - Keyword Extraction, Association, Summarization and Classification. Details of these modules are described below.

- **Keyword Extraction**

In this module, input data is taken as user posts obtained from website. It removes the irrelevant words and retaining keywords.

- **Summarization**

We are performing on summarization of topmost drug family only. We are using WordNet dictionary to detect correct sense of word. It would help to generate better summary results.

- **Association**

We propose to determine the different types of association. These associations could be among disease, drug and symptoms.

- **Classification**

We are taking output of summarization as input for this module. We classify the users into classes like depressed and satisfied.

Apriori Algorithm

Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions. This algorithm is used to extract useful information from large amounts of data.

Association rule learning is a popular method for discovering relations between variables in large databases. An example of association rule learning are the rule {fever} -> {calphol} found in the health data in the health forum. The rule indicates that the patient who has fever likely to take calphol drug. Association rules are used to show the relationships between data items.

The steps are:

Step 1: Scan the opinion data set and determine the support(s) of each item.

Step 2: Generate L1 (Frequent one item set).

Step 3: Use Lk-1, join Lk-1 to generate the set of candidate k - item set.

Step 4: Scan the candidate k item set and generates the support of each candidate k – item set.

Step 5: Add to frequent item set, until C=Null Set.

Step 6: For each item in the frequent item set generate all non empty subsets.
Step 7: For each non empty subset determine the confidence. If confidence is greater than or equal to this specified confidence .Then add to Strong Association Rule.

VI. RESULT AND DISCUSSIONS

The results obtained from the given dataset are classified into two classes' i.e, "satisfied" and "depressed".

Below is a summary of results. We have performed experiments using Apriori algorithm. A confusion matrix is a table that is used to describe the performance of a classification model (or classifier) on a set of test data for which the true values are known the classifier made 200 predictions (e.g., 200 patients were being tested). Out of those 200 cases, the classifier predicted "satisfied" 100 times and "depressed" 100 times. In reality, 100 patients in the sample are belongs to the class satisfied, and 100 patients are belongs to depressed class.

Results:

Table 1

		Predicted Class	
		Depressed	Satisfied
Actual Class	Depressed	100	0
	Satisfied	100	0

VII. CONCLUSION

We proposed an approach to analyze user posts from health communities for knowledge discovery. This will help patients to find out association among different drugs, diseases and symptoms. It will help doctors to find out side-effects of different drugs so they can prescribe better drugs to other patients with similar disease. Pharmaceutical companies will be benefited as we are classifying users of particular drug into different classes like depressed and satisfied. This will be indirect input to companies to decide which drug is popular and whether to produce alternate drug. Thus our work benefit to all three parties–medical fraternity, patient community and pharmaceutical companies.

VIII. REFERENCES

[1] G. Vinodhini and RM. Chandrasekaran: Sentiment Analysis and Opinion Mining: A Survey International Journal of Advanced Research in Computer Science and Software Engineering, 2014.

[2] Alok Ranjan Pal and Diganta Saha: An Approach to Automatic Text Summarization using WordNet, IEEE International Advance Computing Conference (IACC), 2014.

[3] Lakshmi K.S and G. Santhosh Kumar: Association Rule Extraction from Medical Transcripts of Diabetic Patients, IEEE, 2014.

[4] Nikfarjam A and Gonzalez GH : Pattern mining for extraction of mentions of Adverse Drug Reactions from user comment, AMIA Annu Symp Proc. 2011; 2011: 1019–1026, 2011 Oct 22

[5] Shailesh Kumar Yadav : Sentiment Analysis and Classification: A Survey, International Journal of Advanced Research in Computer Science and Management Studies, volume 3, Issue 3, March 2015.

[6] JesminNahar, Tasadduq Imam, Kevin S. Tickle, Yi-Ping Phoebe Chen, "Association rule mining to detect factors which contribute to heart disease in males and females", J. Nahar et al. Expert Systems with Applications 40 (2013) 1086–1093, Elsevier, 2012.

[7] Rafael Ferreira, FredericoFreitas, Luciano de Souza Cabral, Rafael DueireLins, Rinaldo Lima, Gabriel Franca, Steven J. Simske, and Luciano Favaro, "A Context Based Text Summarization System",11th IAPR International Workshop on Document Analysis Systems,pp 66-70, 2014.

[8] Khairullah Khan, BaharumBaharudin, Aurnagzeb Khan, Ashraf Ullah, "Mining opinion components from unstructured reviews: A review", Journal of King Saud University – Computer and Information Sciences (2014), Elsevier, 2014.