MEDICINE RECOMMENDATION SYSTEM BASED ON PATIENT REVIEWS

DR.KIRUBAKARAN NAMASKARAM[1], MEGHAVATH SIRISHA[2], MUDAVATH JYOTHI[3], PEDDAPADU JHANSI[4]

PROFESSOR[1], UG SCHOLAR [2, 3&4]

DEPARTMENT OF CSE, MALLA REDDY ENGINEERING COLLEGE FOR WOMEN,MAISAMMAGUDA, DHULAPALLY KOMPALLY, MEDCHAL RD, M, SECUNDERABAD, TELANGANA 500100

ABSTRACT

Most of the people tend to live a long and healthy life, where they are more conscious about their health. But many studies show that almost many people die due to the medical errors caused in terms of taking wrong medicines and these errors are caused by doctors, who prescribe medicines based on their experiences which are quite limited. As machine learning, deep learning and data mining like technologies that are emerging day by day, these technologies can help us to explore the medical history and can reduce medical errors by being doctor friendly. In this paper proposes a medicine recommendation system , which takes the patient review data and performs sentiment analysis on it to find the best medicine for a disease by using NGram model. In order to increase the accuracy, a Lightgbm model is used to perform medication analysis. The paper also discusses the advantages, disadvantages and enhancements that can be incorporated to improve the accuracy.

INTRODUCTION

One of the most concerned and searched topic on the internet is about health information. According to the Pew Internet and American Life Project, almost 60% of grownups are looking for enough health information on the web with 35% of respondents concentrating on diagnosing ailments online only[1]. Since many studies show that number of people die due to the medical errors and these errors are caused by medical practitioners, who prescribe medicines based on their experiences. As most of their experiences are limited, they often commit mistakes. This study provide a medicine recommendation system for doctors which can be used by them while prescribing medicines. A recommender framework is an ordinary framework that makes the users get a proposal of things which they can utilize for their exact need. Dissimilar to numerous different kinds of frameworks , health recommendation principally relies upon enthusiastic, physical and mental issues of the patients. A medicine recommendation system is similar system

that recommend the medicines for a particular disease based on patient reviews. This system is very essential in this fast growing technological world, which can save lives by helping doctors. In this paper, the proposed medicine recommendation system and its working is depicted, wherein it uses the current technologies like machine learning, data mining etc. to find out the interesting records hidden in the medical data and reduce the medical errors by the doctors while prescribing medicines. This system consists of following modules such as database module, data preparation, data visualization, recommendation and model evaluation module . The proposed medication recommender system uses Machine learning N-Gram and Lightgbm algorithms by using data from hospital and the best one is selected for the medicine recommendation system to attain the metrics like good accuracy, scalability and model efficiency.

RELATED WORKS Recommender frameworks point to supply clients with personalized stock and repair to alter the expanding online information over-burden drawback. Various recommender framework methods are anticipated since the mid1990s,and numerous shapes of recommender framework code were created as of late for a spread of applications. Most of the recommender advances unit of measurement connected to the e-government area[2], e-business area[3], e-commerce/e-shopping area[4], e-learning area[5] and etourism area[6] etc. The medication space incorporates uncommon recommender innovations, and this paper centers on arranging of the medication recommender framework and mining data from therapeutic case knowledge. Through on-line social organizing, the communication is monstrously progressed and totally distinctive intrigued of information is advertised on net essentially at the open pace. The total diverse information must share to chronicle highlights of potential edges and availability of utilities bits of knowledge, things, people practices and items, etc. [7]. One in all the vital areas is that the therapeutic and well being sciences is to think approximately social angles through on-line dialogs, blogs, audits, and on-line overviews, etc. [9]. The health-related substance shared through on-line feedbacks or surveys contains covered up assumption [10] designs that emerges through totally distinctive sources from medical world which offer benefits to the pharmaceutical industry[8]. Amid this, the on-line component is fantastically standard of late for online looking, diverse stock through distinctive websites like on-line buying of drugs at entryway step. Numerous websites and blogs offers clients to rate their stock with their fulfillment and quality of stock, logistics, administrations and criticism etc. , which the clients examines for a particular medicine or on quality of administrations.

PROPOSEDSYSTEM In this segment, the characteristics of the recommender framework are reaching to be specified, additionally the detail of our drugs recommender framework system are aiming to be presented. Recommender framework has gotten to be a profitable investigation field as the advancement of counterfeit brilliantly advances. Not at all like most current recommender frameworks that specialize in e-business, book and pictorial suggestion, our framework points at giving a virtual fully fledged specialist for unpracticed amateurs and patients in abuse right pharmaceutical. Since high accuracy and strength is vital for such an online pharmaceutical recommender framework, in this way we tend to evaluate a few information preparing approaches to induce an genuine trade-off among the precision, productivity and quantifiability. In this proposed System, the framework mainly consists of five modules, as shown in figure 1, which are (i) Database module (ii)Data preparation module (iii)Recommendation model module (iv)Model evaluation model and (v)Data visualization module. Through this process set the topic, preprocess the data to fit the objective, and create various variables to fit model. At the model part, emotion analysis using word dictionary, ngram applying deep learning, etc. were used and then accuracy of both the algorithms is evaluated.

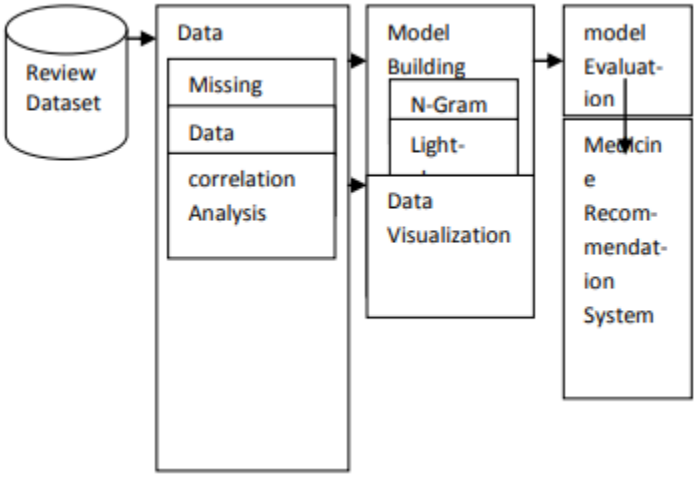

Fig :Architechture Of Medicine Recommendation System

(i) Database system module: it contains a drug review dataset with attributes like unique Id, drug name, condition(disease of patient), date , useful count, reviews and ratings given by the patients on the drugs.

(ii) Data preparation module: It comprises of information investigation and information preprocessing. The real-world information is crude information which can be fragmented, boisterous and messy. Thus, information arrangement is utilized to clean information. it comprises of missing value processing, correlation analysis and removing data redundancy

1. Data Exploration:

a) Find unique number of patient ids to check if a patient has written multiple reviews.

b) Analyze number of drugs per condition by considering condition and number of drugs.

2. Data preprocessing:

A. Find out the number of missing values for all the attributes
B. These missing values can be ignored ,removed or filled. so the rows with missing values can be removed since the dataset is very large.
C. Remove the duplicate rows from the dataset to normalize the data.
D. Delete the conditions with only one drug as only one drug may not be sufficient to recommend the best one.
E. Words like not, needn't, never etc are key parts of emotional analysis so remove them from stop words. Now clean the reviews by removing stop words.
F. Find out the correlation among attributes. It is found that reviews and useful count are highly correlated.

Model building: The entire drug review dataset is divided into two portions where 70% of the data is training data and 30% is used for testing the data. Sentiment analysis is done using N-gram deep learning model. In order to compensate the limitation of natural language processing, Lightgbm machine learning model was used and reliability was further secured through useful count.

N-Gram: N gram is a set of co-occurring words in a text. Features for supervised machine learning model such as decision tree, naïve bayes can be developed using this algorithm. Steps:

a) Find out the set of co-occurring words (gram) from the reviews starting from uni-gram.
b) 1-gram: Analyze the text with a single corpus. But it does not classify the emotion well.
c) 2-gram: it is hard to classify positive and negative reviews using bi-gram.

d) 3-gram: Tri-grams are still unable to classify the positive and negative reviews.

e) 4-gram: 4-gram classifies the emotions much better than other grams. Therefore, 4-gram is used to build the deeplearning model.

f) Now calculate the sentiment for each review where sentiment=1 when rating>5 and sentiment=0 when rating<=5.

g) Build the deep learning model using above sentiment and cleaned review.

h) Train the model by considering epochs=20 on train_data.

i) Find out the accuracy by testing the model. The accuracy is found to be 80%.

Lightgbm: It is a gradient boosting framework that uses tree based learning algorithms. It is designed to be distributed and efficient with the following advantages: Faster training speed ,higher efficiency and Lower memory usage. The Lightgbm uses XGBoost as a baseline and outperforms it in training speed and the dataset sizes it can handle. Steps:

A. Find out the correlation among the variables.

B. Calculate predict_sentiment using cleaned reviews and date.

C. Find out the correlation coefficient between predict_sentiment & rating and also between predict_sentiment & sentiment.

D. Calculate predict_sentiment2 using reviews.

E. Also find out the correlation coefficient between predict_sentiment2 & rating and also between predict_sentiment2 & sentiment.

F. Since these are not highly correlated, normalize the useful count to build the model. g) Build the machine learning model using added variables.

G. Train the model on the above features i) Find out the accuracy by testing the model. The accuracy is found to be 90%.

Model Evaluation: Model evaluation is used to evaluate all the models built by considering the metrics like model efficiency , model accuracy and model scalability and to select the one with all the higher attributes.

Steps:

- Find out the accuracy of N-gram model and the accuracy of N-gram found to be 80%.

- Find out the accuracy of lightgbm model and the accuracy of Lightgbm found to be 90%.
- Compare both the accuracies and select the model with higher accuracy. Since, the accuracy for Lightgbm is higher than the N-gram model, Lightgbm model is used for recommendation.

Visualization module: It primarily gives the visualization innovation to show a few important information behind the determination case information.

CONCLUSION The paper has proposed an idea for medicine recommendation system for medical inquiry. This approach is based on four main steps: (i) analysis of review dataset (ii) data preprocessing (iii) model building. iv) Recommending the proper medicine for a particular disease. The proposed system works as a tool for supporting the doctors in their disease diagnosis. As future work efficiency of recommendation system can be increased by including age of the person, demographic information during the training phase. Also the brand and the chemical contents available in the medicine can improve the recommended medicines.

REFERENCES:

[1] S. Fox and M. Duggan. Health online 2013. Pew Internet and American Life Project. http://pewinternet.org/Reports/2013/Healthonline.aspx, 2013.

[2] X. Guo, J. Lu, Intelligent e-government services with personalized recommendation techniques, International Journal of Intelligence Systems,2007,401–417

[3] T. Lee, J. Chun, J. Shim, S.-g. Lee, An ontology-based product recommender system for B2B marketplaces, International Journal of Electronic Commerce, 2006,125–155.

[4] J.B. Schafer, J. Konstan, J. Riedl, E-commerce recommendation applications, Applications of Data Mining to Electronic Commerce,Springer, US 2001, 115–153.

[5] O.R. Zaiane, Building a recommender agent for elearning systems, Proceedings of 2002 International Conference on Computers in Education, 2002, 55–59

[6] T.Hung-Wen,S.VonWun,A personalized restaurant recommender agent for mobile e-service, 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service. EEE, 2004, 259–262.

[7] Popescu AM, Etzioni O (2005)Extracting product features and opinions from reviews. Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, pp: 339-346

[8] Mei Q, Ling X, Wondra M, Su H, Zhai C (2007) Topic sentiment mixture: modeling facets and opinions in weblogs. Predictive Modeling of Web Users, pp: 171- 180.

[9] https://www.livewell.pk/ [Accessed on: August 2017].

[10] Aronson AR (2001) Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In: AMIA annual symposium proceedings. Washington DC: American Medical Informatics Association, pp: 17- 21.

[11] Subhash C. Pandey, ―Data Mining techniques for medical data: A Review‖, ―IEEE‖, 2016.

[12] F. O. Isinkaye, Y.O. Fola Jimi, B.A. Ojokoh,―Recmmendation systems : Principles, Methods and Evaluation‖, ―Elsevier‖, 261-273,2015.

[13] Zhang S, Zhang C, Yang Q. Data preparation for data mining[J]. Applied Artificial Intelligence, 2003, 17(5-6): 375-381.

[14] Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res.3:1289–1305

[15] Chamlertwat, W.; Bhattarakosol, P.; Rungkasiri, T.; and Haruechaiyasak, C. 2012. Discovering consumer insight from twitter viasentiment analysis. J. UCS 18(8):973– 992

[16] Lorenzo-Romero C, Constantinides E, Brunink LA (2014) Co-creation: customer integration in social media-based product and service development. Procedia Soc Behav Sci 148: 383-396.

