ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Special Issue 2, January 2018

Analysis of disease from multiple health care data by using HL7 message on HDFS

K. Dhinakaran[1], Sriram N[2], Srivathsan R[3], Subash N[4]
Assistant Professor[1], Department of Computer Science and Engineering
UG Scholar[2][3][4], Department of Computer Science and Engineering
Rajalakshmi Institute of Technology, Chennai.
srirm38@gmail.com[2], srivatsancool.r@gmail.com[3], subashnarashimhan@gmail.com[4]

***Abstract*:CrowdSourcing is adistributed drawback finding techniques where very little tasks broadcasted and outsourced to crowd that were earlier completed by chosen agent .Due to the popularity and ubiquity of social networks, sentiment analysis has become an important and well covered research area. Short texts usually encounter sparse problems in representations for their limited texts length. We address this issue by clustering short texts to form a long text. Text Classification is a useful task in Text Mining. Most Researchers employ one word weight type in text classification. This application may be used for prediction of diseases from the multiple healthcare data that are received and by analyzing it with the help of HL7 preprocessor from the blood test data collected in HDFS.**

Keywords- Crowdsourcing, Data Analytics, HL7.

I. INTRODUCTION

Over the last few years, healthcare data has become more complex for the reason that large amount of data are being available lately, along with the rapid change of technologies and mobile applications and new diseases have discovered. Therefore, healthcare sectors have believed that healthcare data analytics tools are really important subject in order to manage a large amount of complex data, which can lead to improve healthcare industries and help medical practice to reach a high level of efficiency and work flow accuracy. The healthcare sector is widely considered as one of the most important industries in information technology. More and more, information technology has been considered as a practice that facilitates healthcare performance through using data and information efficiently within the healthcare sectors.

The concept of big data is not new, however the way it is defined is constantly changing. Various attempts at defining big data essentially characterize it as a collection of data elements whose size, speed, type, and/or complexity require one to seek, adopt, and invent new hardware and software mechanisms in order to successfully store, analyze, and visualize the data. Healthcare is a prime example of how the three Vs of data, velocity (speed of generation of data), variety, and volume, are an innate aspect of the data it produces. This data is spread among multiple healthcare systems, health insurers, researchers, government entities, and so forth. Furthermore, each of these data repositories is soloed and inherently incapable of providing a platformfor global datatransparency.

TEXT MINING

Text has to extract but in clear and precise way. Text mining need no interference of humans in mining. In many data mining applications, information extracted is of less precision. Text mining with clear output is used summarizing the large dataset into small dataset.

TYPES OF TEXT MINING

There are various types of text mining subfields like data mining, information retrieval, web mining and statistics. The main aim of the text mining is to practice the random textual information like text, numeric etc. In text mining analyzing of the words, grouping of words can be done in documents etc.

This classification is an important task of computational linguistic that are used in many real world problems such as spam filtering, sentiment classification,document clustering etc. Even though most approaches of text classification researches are machine learning approaches on the text classification. , the team weight is also used to filter the word features such as information gain. Most keyword list based approaches use only one type of keyword weight in order to select the keyword list member. The keyword list approach was conducted in an authority classification which is part of an automatic complaint management system in Bandung city. Bandung city government has invited a complaint channel for citizen using twitter and social websites.

TEXT CLASSIFICATION

Short texts classification has attracted increasing research interest in recent years. The emergence of solution media has given web users an accessible location for expressing and sharing their thoughts and opinions anany kinds of events and topics. The objective of sentiment analysis is to classify the sentiment of shot texts into positive, negative or sentiment classification focus on two directions: lexicon-based and supervised learning methods.

SENTIMENT ANALYSIS

Sentiment analysis has been handed as a natural language processing task. Traditional statics based methods usually perform unsatisfactorily for short texts sentiment classification due to this sparse of representations. Sentiment analysis of shot texts is considered as a much harder problem than that of conventional text such as movie review documents.

CROWDSOURCING

Crowdsourcing has been adopted in a wide variety of domains, such as design of T shirts and pharmaceutical research and development, and there are numerous crowdsourcing platforms through which customers and suppliers can find each other. Mobile Crowdsourcing (MCS) has emerged as a popular and effective method for data collection and data processing by utilizing the sensing, communication and computing capabilities of the widely available mobile devices. It combines the concepts of crowdsourcing and mobility. A MCS system is open to mobile devices to participate in any sensing and computing tasks. It allows outsourcing a complex task that is usually difficult to be completed by a single computer or a group of people to an unspecified group of mobile devices. MCS that involves human intelligence called human-assisted MCS, is an effective method to perform tasks that are easy for humans but remain difficult for machines. Human-assisted MCS can help build collaborative intelligence between human and machines. Crowdsourcing (CS) plays an important role in rapid software development. The primary theme of CS is to offer short-schedule development with parallel and micro-tasking concepts. Crowdsourced software development(CSD) uses an open-call format online to acquire a large number of workers. Crowdsourcing has also been explored to handle decision making problems in web-based systems. CrowdDB adopted human input via crowdsourcing to process queries that neither database systems nor search engines can adequately answer.

TYPES OF CROWDSOURCING

CROWD VOTING

Crowdvoting occurs when a website gathers a large group's opinions and judgments on a certain topic. The Iowa Electronic Market is a prediction market that gathers crowds' views on politics and tries to ensure accuracy by having participants pay money to buy and sell contracts based on political outcomes.

CROWD FUNDING

Crowdfunding is the process of funding projects by a multitude of people contributing a small amount to attain a certain monetary goal, typically via the Internet.[93] Crowdfunding has been used for both commercial and charitable purposes.[94] The crowdfunding model that has been around the longest is rewards-based crowdfunding. This model is where people can prepurchase products, buy experiences, or simply donate.

CLASSIFICATION

Naive Bayes has been studied extensively since the 1950s. It was introduced under a different name into the text retrieval community in the early 1960s and remains a popular (baseline) method for text categorization, the problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features.

II. EXISTING SYSTEM

In this existing system the real-time transactional healthcare data together with the accumulated historical EHR data in a hospital are critical for clinicians and other care providers to make the right medical decisions that lead to the best care or services received by each patient. As the business in a hospital continues to grow, the healthcare data volume and complexity becomes larger and larger, and quickly reaches a point where traditional RDBMS-based EHR system fail to handle, which likely results in the construction of an integrated clinical care and/or EDW system. The major problem existing in a RDBMS-based EHR system or its derivative – integrated clinical care systems is the difficulty to use and present all the historical data to its clients.

III. PROPOSED SYSTEM

The clinical data from the daily operations of MC hospitals and clinics are made up of a variety of document types – many in the format of HL7 V2 messages. Creating a new patient record or updating an existing patient record may result in the creation of one or more HL7 messages. Each document type is managed or generated by a single source system at MC. We develop the application for deep analyze in the healthcare data HL7 message format. Proposed work not process only inside the hospital this system mainly focused on disease prediction for various strategies, this source data started from medical laboratory normally users are going to check their health condition in diagnostic center. they transfer all the patient disease information to healthcare care officer, the user also can provide their medical info to the health care officer, they collect all the diagnostic center information all well as the patient each and every area this medical report forward to the government medical research center. Medical reports convert to the HL7 message by using HAPI HL7 v2 format stored in the HDFS. Then the research center can analyze the healthcare information from various area and city.

Modules

- Diagnostic Center Report Submission
- District Healthcare Officer and User
- HL7 Message Preprocessing
- Medical Research Center Analysis

Diagnostic Center Report Submission:

Every hospital having diagnostic center some places diagnostic center working separately. Doctors will suggest diagnostic center to the patient they want to know about patient health condition and disease. After taking medical checkup in the diagnostic center this data will be transferred to the District Healthcare officer this same process will be applicable for the every diagnostic center. But the transfer medical report not contains any personal information about the patient.

District Healthcare Officer and User:

The healthcare officer getting information from every diagnostic center, all the healthcare care report will be combined into a single file. The user can update their disease information directly to health care officer, after submission the user will get notification about the disease and precautions. Healthcare officer transferred entire detail to the medical research center.

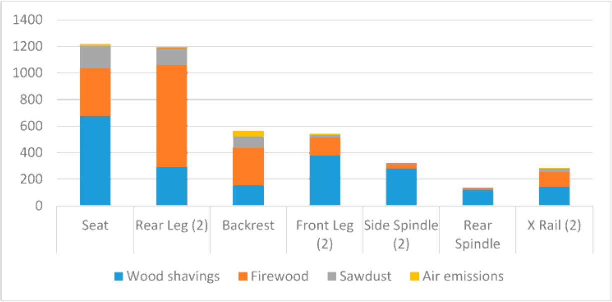
HL7 Message Preprocessing:

Healthcare officer report will be converted into HL7 message format and stored into the HDFS (Hadoop Distributed File System).The preprocessing is a kind job to be process the HL7 messages, where the level of process happening in different procedure the report contains all medical information it will categorize by age, gender, symptoms and so on, this preprocessing trigger some time automatically in our application. So the medical analysis information dynamically changed every time.

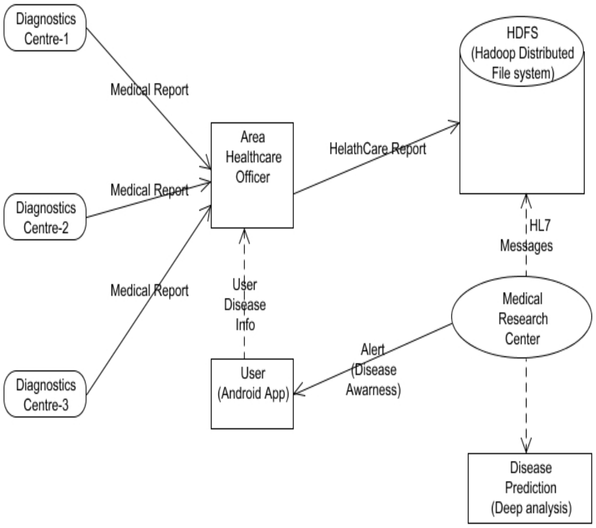
Medical Research Center Analysis:

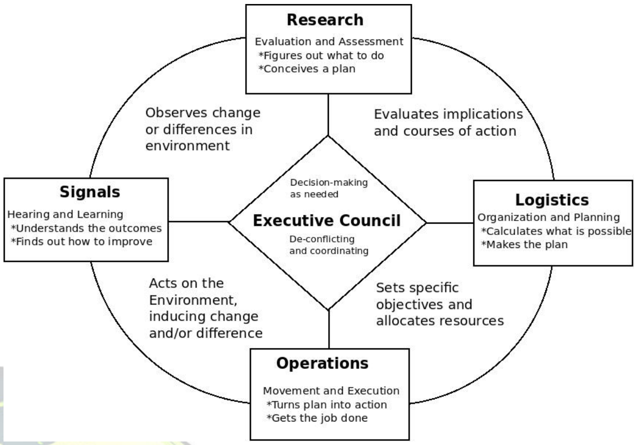
This is the Government medical research center analyze the medical report in various category. They can find out which disease affected most of the area and many people, if suppose result displayed disease dengue system should give detail about the disease. Which aged peoples are affected by the dengue, which symptoms peoples are affected by this disease which area people affected by this disease all these details will be analyzed in this application. Result showing any district or area the precautions will be delivered to those users.

IV. BLOCK DIAGRAM REPRESENTATION



SYSTEM ARCHITECTURE:




Research
Evaluation and Assessment
*Figures out what to do
*Conceives a plan
Observes change or differences in environment
Evaluates implications and courses of action
Signals
Hearing and Learning
*Understands the outcomes
*Finds out how to improve
Decision-making as needed
Executive Council
De-conflicting and coordinating
Logistics
Organization and Planning
*Calculates what is possible
*Makes the plan
Acts on the Environment, inducing change and/or difference
Sets specific objectives and allocates resources
Operations
Movement and Execution
*Turns plan into action
*Gets the job done

V. LITERATURE SURVEY

Jinpeng Chen et.al proposed a evaluating theory on the basis of Crowdsourced ranking functions for querying the spatial keyword in the year 2017 ,In this author has been evaluated about the ranking functions for spatial keyword[1]. Chen Wenliang et.al proposed an Automatic word clustering using Global Information for Text Categorization with a minimal cost algorithm copyrighted by ACM 2004[8]. FabrizioSebastiani proposed that Automated Text Categorisation on machine learning with the help of processed crowdsourcing in the year 2002[10]. Agarwal . a et.al proposed that twitter data sentiment analysis in the proceedings of languages in social media workshop association for Computational Linguistics by text mining and data analysis in the year 2011 june [7].

Dino Ienco Rosa Meo said that exploration and reduction by Hierarchical Clustering of the feature space using the method classification and clustering algorithm in Dipartimento di Informatica, Universit`a di Torino,Italy[9]. Jinpeng chen et. al evaluated the ranking function using crowdsourcing based process for special keyword querying with the help of text processing of the keyword[1.1]. Go, et.al proposed a system for sentiment classification for twitter using distant supervision by help of text classification and data analytics by keyword in the year 2009[6]. Reynold Cheng et.al proposed a Comet: A crowdsourced multi-label task system using mobile crowdsourcing in the year 2017[4].

HamedNilforoshan et.al developed a system called Precog: Improving Crowdsourced data quality not after acquisition using a concept called as Crowd methodology in the year 2017[2]. Luan Tran et.al constructed a system for Real time Framework for Task Assignment in Hyperlocal spatial crowdsourcing by the help of technology called crowd funding in the year 2017[3].

George Forman et.al proposed a barely scratched surface theory on the basis of Feature Selection: We've barely scratched the surface was Published in IEEE Intelligent Systems in the year November 2005[11]. Huan Liu et.al proposed a text selection on the basis of Evolving Feature text Selection was Published by the IEEE Computer Society in the year 2003[14]. Huan Liu et.al proposed a theory based on Classification and Clustering Algorithms for Toward Integrating Feature Selection and was published in Department of Computer Science and Engineering in the year 2005[13]. Hisham Al-Mubaid et.ai proposed theory on the basis of Distributional Clustering and Learning by a Hisham Al-Mubaid Technique and was published in IEEE in the year 2006[12]. Martin H.C. Law and Mario proposed a theory on the basis of Feature Selection in Mixture Models in Simultaneous and was published in IEEE in the year 2004[16]. Jinxiu Chen1 Donghong Ji1 Chew Lim Tan

Et.al proposed a theory on the basis of Selection Relation on Unsupervised and was published in Institute for Infocomm Research in the year 2002[15].

Tao Liu and Shengping Liu et.al propose a theory on the basis of Text Clustering in An Evaluation on Feature Selection for Text Clustering Proceedings of the Twentieth International Conference on Machine Learning and was published in Washington DC on the year 2003[17]. C. Chen, Y. Huang et.al proposed a theory on Crowdsourcing in Interactive Crowdsourcing to Spontaneous Reporting of Adverse Drug Reactions and was published in Proc. IEEE International Conf. Communications (ICC'14)in the year 2014[20]. K. Stol, T. LaToza et.ai proposed a theory on Crowdsourcing software engineering and was published in IEEE Software, vol. 34 in the year 2017[19].

VI. REFERENCE

1. "Crowdsourcing-Based Evaluation of Ranking Functions for Spatial Keyword Querying" by Jinpeng Chen et. al (2017)

2. "PreCog: Improving Crowdsourced Data Quality Before Acquisition" by HamedNilforoshan et.al (2017)

3. "A Real-Time Framework for Task Assignment in Hyper-local Spatial Crowdsourcing" by Luan Tran et.al (2017)

4. "Comet: A Crowdsourcing Multi-Label Task System" by Reynold Cheng et.al (2017)

5. Benkler, Y. 2002. Coase's penguin,or Linux and the nature of the firm. Yale Law Journal 112, 367--445.

6. Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1, 12.

7. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., &Passonneau, R. (2011,June). Sentiment analysis of twitter data. In Proceedings of the Workshop on Languages in Social Media (pp. 30-38). Association for Computational Linguistics.

8. Chen Wenliang, Chang Xingzhi, and Wang Huizhen, "Automatic Word Clustering for TextCategorization Using Global Information"Copyright ACM 2004.

9. Dino Ienco Rosa Meo "Exploration and Reductionof the Feature Space by

Hierarchical Clustering"Dipartimento di Informatica, Universit`a di Torino,Italy.

10. FabrizioSebastiani "Machine Learning in Automated Text Categorization" ACM ComputingSurveys, Vol. 34, No. 1, March 2002.

11. George Forman "Feature Selection: We've barelyscratched the surface" Published in IEEE IntelligentSystems, November 2005.

12. Hisham Al-Mubaid and Syed A. Umair "A Hisham Al-Mubaid Technique Using DistributionalClustering and Learning Logic" IEEE 2006.

13. Huan Liu and Lei Yu "Toward Integrating FeatureSelection Algorithms for Classification and Clustering" Department of Computer Science andEngineering,2005.

14. Huan Liu, "Evolving Feature text Selection" Published by the IEEE Computer Society 2003.

15. Jinxiu Chen1 Donghong Ji1 Chew Lim Tan "Unsupervised Feature Selection for Relation Extraction" Institute for Infocomm Research 2002.

16. Martin H.C. Law and Mario A.T. Figueiredo "Simultaneous Feature Selection Using Mixture Models" ii IEEE 2004.

17. Tao Liu and Shengping Liu "An Evaluation onFeature Selection for Text Clustering" Proceedingsof the Twentieth International Conference onMachine Learning (ICML-2003), Washington DC,2003.

18. Yanjun Li Congnan Luo, "Text Clustering withFeature Selection by Using Statistical Data" IEE 2008.

19. K. Stol, T. LaToza, and C. Bird, "Crowdsourcing for softwareengineering," IEEE Softw, vol. 34, no. 2, 2017

20. C. Chen, Y. Huang, Y. Lou, C. Liu, L. Meng, Y. Sun, K. Bian, A.Huang, X. Duan, and B. Jiao, "Interactive Crowdsourcing toSpontaneous Reporting of Adverse Drug Reactions," Proc. IEEEInternational Conf. Communications (ICC'14), pp. 4275-4280, 2014,doi: 10.1109/ICC.2014.6883992