



International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 10 • Number 25 • 2017

Adverse Drug Effects Analysis from Electronic Medical Records through Improved Text Mining Algorithm

Mohamed Suhail.M^a and Prakash.N^b

^aPh.D. Research Scholar, Department of Information Technology, B.S. Abdurrahman University, Chennai. Email: suhailmumu2009@gmail.com

^bAssociate Professor, Department of Information Technology, B.S. Abdurrahman University, Chennai

Abstract: Big data usually includes data set with sizes beyond the ability of commonly used traditional software tools to analyze, capture, curate, search, share, store, transfer, visualize the query, update and secure information.. The research on big data mainly relies on its 5Vs characteristics which are Volume, Variety, Velocity, Variability and Veracity. Even with all research eventually helping answer clinical realm events, according to Bennett et. al., there is more than 15 years gap between clinical research and the actual clinical care used in practice. Decisions these days are made mostly on general information that has worked before, or based on what experts have found to work in the past. Thus, the existing system doesn't have proper health care to the patients which resulted in improper health care to the patients. In India, where Electronic Medical Record (EMR) is not maintained properly. One big problem which is not noticed is Adverse Drug Reactions (ADR) of the patients. People are not having the awareness of reporting these ADR to concerned authorities to further actions. This negligence targets mostly elderly patients where they use multiple drugs. The study shows that they should be closely monitored for ADRs, to avoid clinically significant harmful consequences by maintaining proper EMR. Thus, my research is focused on providing better clinical making decisions using big data analysis on EMR through improved text mining algorithm to physicians for their patients to avoid Adverse Drug Reaction and provide quality health care services.

Keywords: Adverse Drug reactions, Electronic medical records.

1. INTRODUCTION

The healthcare industry historically has generated large amounts of data, driven by record keeping, compliance and regulatory requirements, and patient care . While most data is stored in written format, the current trend is toward rapid digitization of these large amounts of data. Driven by mandatory requirements and the capability to improve the quality of service in healthcare meanwhile reducing the costs, these massive quantities of data (known as 'big data') hold the promise of supporting a wide range of medical and healthcare functions, including among others clinical decision support, disease surveillance, and population health management [1]. Reports say data from the U.S. healthcare system alone reached, in 2011, 150 exabytes. At this rate of growth, big data

for U.S. healthcare will soon reach the zettabyte (10^{21} gigabytes) scale and, not long after, the yottabyte (10^{24} gigabytes). Kaiser Permanente, the California-based health network, which has more than 9 million members, is believed to have between 26.5 and 44 petabytes of potentially rich data from EHRs, including images and annotations. By discovering associations and understanding patterns and trends within the data, big data analytics has the potential to improve care, save lives and lower costs. When big data is synthesized and analyzed—and those aforementioned associations, patterns and trends revealed—healthcare providers and other stakeholders in the healthcare delivery system can develop more thorough and insightful diagnoses and treatments, resulting, one would expect, in higher quality care at lower costs and in better outcomes overall.

2. ADVERSE DRUG REACTION

Pharmacovigilance is an integral part of drug therapy. Still, it is not widely practiced in Indian hospitals. In various studies, adverse drug reactions have been implicated as a leading cause of considerable morbidity and mortality [2]. The incidence of adverse drug reactions (ADR) varies with studies which show incidences ranging from as low as 0.15% to as high as 30% [3]. Elderly and hospitalized patients are reported to be more susceptible to ADRs than the adult population (16.6% vs. 4.1%) [3]. Indian reports on ADR monitoring have been very few. This may be because ADR monitoring is still evolving here. After many years of hibernation, the need for an efficient pharmacovigilance programme was felt, the result of which was the institution of National Pharmacovigilance Programme in 2004. The objective of this programme is to create awareness among people on ADR monitoring and to encourage a reporting culture. This helps in better understanding on drugs and its effects. Thus the worsen effects drugs can be banned on reporting to concerned department. So the quality of service in health care can be improved.

WHY ADR Monitoring is important?

Hospital-based ADR monitoring and reporting programmes aim to identify and quantify the risks associated with the use of drugs.

It has been estimated that ADR are the 4th to 6th cause for mortality in USA. Table 1 below shows the mortality rate.

Table 1
Mortality rate

Norway	11.5%
France	13%
UK	16%

With 10% of 3.63 trillion medicines popped worldwide in 2015, India is the world's third-largest medicine market. It stands to scientific reason that these drugs will have side effects

Serious effects were seen in 6.7% of patients, a 2014 study reported. Other studies have cited drug side effects as the reason for 3.4% of hospital admissions in India, 3.7% hospital readmissions, and 1.8% mortality. In the developed world, adverse reactions are believed to be the fourth-leading cause of death.

Within India, the ADR reporting rate (ADRs reported per million population) has almost doubled in the last three years to 40, but it is lower than 130, the average ADR reporting rate for high-income countries, and clearly disproportionate to the country's population and medicine consumption. In other words, India addresses the problem of adverse drug reactions by ignoring or not reporting the data. That could prove costly, said experts, if it isn't already. The side effects of a drug if found we shall determine whether the drug is to be banned or not.

A medicine labelled safe for clinical use after trials could still be found to be dangerous—as it happened with ximelagatran in 2006. Ximelagatran was generally tolerated in the trial populations, but a small proportion (5-6%) developed elevated liver enzymes level.

“ADR reporting is still too low vis-à-vis the country’s 1.28 billion population,” said Harihar Dikshit, head, Department of Pharmacology, Indira Gandhi Institute of Medical Sciences (IGIMS), Patna, one of 150 Indian ADR Monitoring Centres. Thus ADR monitoring has to be improved to a great extent for providing better quality of service in health care. Awareness has to be created on drug monitoring throughout its market life as early detection of adverse drug reactions (ADRs) can lead to alerts that prevent patient harm. Recently, electronic medical records (EMRs) have emerged as a valuable resource for pharmacovigilance. Prediction of ADR can be done successfully by analyzing the EMR (Electronic Medical Report).

3. ELECTRONIC MEDICAL RECORD

The electronic medical record (EMR) is simply the electronic format of health records. EMR stores various types of medical informations. The data ranges from medical history, prescriptions, drug allergies to the patients hospital service bills and more. Mostly, paper based systems are used which is insufficient, ineffective and involves high cost of maintenance. On the contrary, EMR has several advantages like easy data recovery, displaying ADR ,Improved diagnosis and treatments etc.

EMR helps doctors in making effective medical decisions with ease. In addition, EMR helps the service providers to effectively gather, analyse and observe patient’s medical information with the help of hospital information system (HIS). Along with managing the medical data, EMR assists in hospital order management, hospital workflow management and security of the medical data. It assists the entire healthcare delivery process in reducing cost and maximizes the profit. With help of EMR, Doctors are able to provide treatment in an effective way. The collection of clinical data enhances the efficiency and improves the data quality.

The pharmaceutical companies and regulatory authorities use EMR to improve the post marketing surveillance PMS of drugs.

EMR Adoption Rates 2012-2016

In comparison to developed nations, the adoption of EMR in India has drastically low. It is important to improve EMR standards with more health related informations like patient visit date , drug prescribed, No. of days drugs in taken, side effects if any, alternative drug prescribed, next visit date, no. of days hospitalized etc. This high standard EMR is analysed through text mining algorithm and as a result a pattern is generated which shows drugs and its worse reactions.

Table 2
EMR Adoption rates

<i>Year</i>	<i>EMR Adoption Rate</i>
2012	39.6%
2013	48.1%
2014	50.4%
2015	62.8%
2016	59%

Further annotations are prepared based on the pattern generated. This will enable the patients to be aware of drugs and its side effects. It helps in improving the quality of service in health care.

4. TEXT MINING ALGORITHM

Text mining process starts with a document collection from various resources. Text mining tool would retrieve a particular document and pre-process it by checking format and character sets. Then document would go through a text analysis phase. Text analysis is semantic analysis to derive high quality information from text. Many text analysis techniques are available; depending on goal of organization combinations of techniques could be used. Sometimes text analysis techniques are repeated until information is extracted. The resulting information can be placed in a management information system, yielding an abundant amount of knowledge for the user of that system. Text mining process is as shown in following Figure 1.

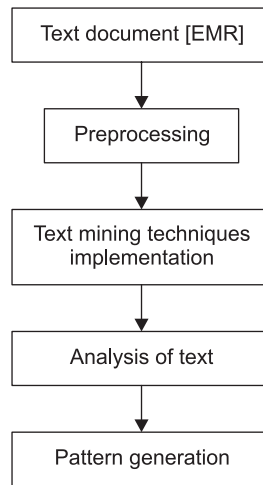


Figure 1: Text mining process

5. TEXT MINING METHOD

Pattern Taxonomy Method

In pattern taxonomy method documents are analyzed on pattern basis. Patterns can be structured into taxonomy by using is-a relation. Pattern mining has been extensively studied in data mining communities for many years. Patterns can be generated by data mining techniques like association rule mining, frequent item set mining, sequential pattern mining and closed pattern mining [4]. Use patterns in the field of text mining is difficult and ineffective, because some useful long patterns with high specificity lack in support (i.e., the low-frequency problem). Not all frequent short patterns will be useful. In research work, an effective pattern discovery technique has been proposed to overcome the low-frequency and misinterpretation problems for text mining. The pattern based technique uses two processes pattern deploying and pattern evolving [5]. This technique refines the discovered patterns in text documents. The experimental results show that pattern based model performs better than not only other pure data mining-based methods and the concept-based model, but also term-based models.

Techniques used in Text Mining

Technologies are produced by natural language processing teaches computer how to analyze , understand, and categorize text. The technologies like information extraction, summarization, categorization, clustering and information visualization, are used in the text mining process.

Information Extraction

Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents. In most of the cases this activity concerns processing human

language texts by means of natural language processing (NLP). Applying information extraction on text, is linked to the problem of text simplification in order to create a structured view of the information present in free text. The overall goal being to create a more easily machine-readable text to process the sentences. Typical subtasks of IE include:

Named entity recognition: recognition of known entity names. Eg Name of a medicine (paracetamol) and certain types of numerical expressions, 650 mg.

Coreference resolution: Detection of coreference and anaphoric links between text entities. In IE tasks, this is typically restricted to finding links between previously-extracted named entities.

General information extraction process is as shown in Figure 2.

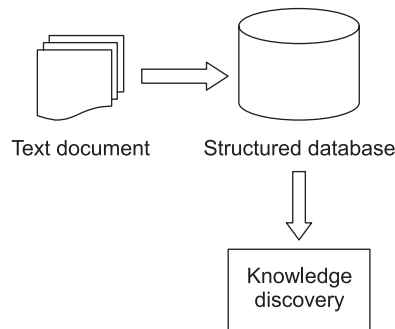


Figure 2: Information extraction process

Text Classification

Text classification is the process of classifying a document based on a predefined category. There are number of approaches for text classification. Figure 3 shows the text classification process. Document representation include representing the document in the form suitable for text mining system. As a result of this phase document is simply a collection of words.

Feature selection phase include removal of irrelevant words in the documents. This phase helps to improve classification accuracy.

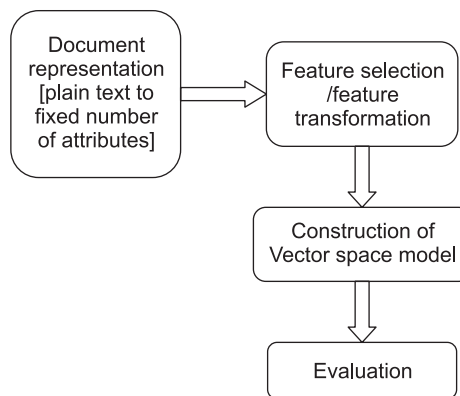


Figure 3: Text classification process

After preprocessing task , terms are weighed based on the popular method called Term Frequency Inverse Document Frequency. It is based on the comparison of frequency of a term with the rarity of a term in the complete set of documents

After this stage the performance of text classifier is evaluated.

Clustering

Clustering is the process of organizing objects into groups whose members are similar in some way. A *cluster* is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters. So, the goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. Any of the following clustering algorithms can be implemented.

- K-means
- Fuzzy C-means
- Hierarchical clustering
- Mixture of Gaussians

Visualization in Text Mining

One of the most common strategies used in text mining is to identify important entities within the text and attempt to show connections among those entities.

Following Figure 4 shows steps involved in visualization process.

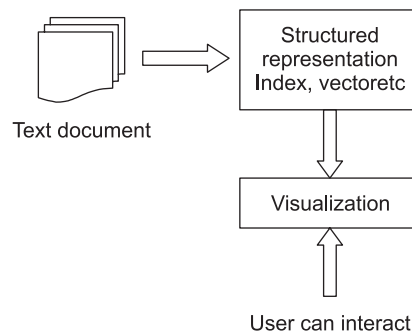


Figure 4: Visualization

The goal of information visualization divided into three steps: (1) Data preparation step includes deciding and obtaining original data of visualization and form original data space. (2) The process of analyzing and extracting visualization data needed from original data and to form visualization data space is known as Data analysis and extraction. (3) Visualization mapping step employ certain mapping algorithm to map visualization data space to visualization target.

6. SUMMARIZATION

The main idea of summarization is to find a representative subset of the data, which contains the information of the entire set. Text summarization is to reduce the length and detail of a document while retaining most important points and general meaning. Text summarization is helpful for to figure out whether or not a lengthy document meets the user’s needs and is worth reading for further information hence summary can replace the set of documents. In the time taken by the user to read the first paragraph text summarization software processes and summarizes the large text document. It is difficult to teach software to analyze semantics and to interpret meaning of text document even though computers are able to identify people, places, and time. Humans first reads entire text section to summarize then try to develop a full understanding, and then finally write a summary, highlighting its main points. Summarization process include following steps: (1) Pre-processing obtain a structured representation of the original text. (2) To transform summary structure from text structure algorithm is applied in next processing step. (3) In the invention step the final summary is obtained from the summary structure.

7. CONCLUSION

This paper tells the importance of high standard EMR which enhances effective analysis of drug reactions. The improved text mining algorithm used is expected to generate accurate patterns on drug –AE relationship.

REFERENCES

- [1] Miller K, Big Data Analytics in Biomedical Research, Biomedical Computation Review. (Winter 2011/2012): 14-21. http://biomedicalcomputationreview.org/sites/default/files/w12_fl-big_data.pdf. (Accessed December 4, 2013)
- [2] Belton KJ. Attitude survey of adverse drug-reaction reporting by health care professionals across the European Union. The European Pharmacovigilance Research Group. *Eur J Clin Pharmacol.* 1997; 52:423–427. [PubMed]
- [3] Belton KJ, Lewis SC, Payne S, Rawlins MD, Wood SM. Attitudinal survey of adverse drug reaction reporting by medical practitioners in the United Kingdom. *Br. J. Clin Pharmacol.* 1995; 39:223–226. [PMC free article] [PubMed]
- [4] S.-T. Wu, Y. Li, Y. Xu, B. Pham, and P. Chen, “Automatic Pattern- Taxonomy Extraction for Web Mining,” *Proc. IEEE/WIC/ACM Int’l Conf. Web Intelligence (WI ’04)*, pp. 242-248, 2004.
- [5] S.-T. Wu, Y. Li, and Y. Xu, “Deploying Approaches for Pattern Refinement in Text Mining,” *Proc. IEEE Sixth Int’l Conf. Data Mining (ICDM ’06)*, pp. 1157-1161, 2006.
- [6] W. Lam, M.E. Ruiz, and P. Srinivasan, “Automatic Text Categorization and Its Application to Text Retrieval,” *IEEE Trans. Knowledge and Data Eng.*, Vol. 11, No. 6, pp. 865-879, Nov./Dec. 1999.
- [7] H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins, “Text Classification Using String Kernels,” *J. Machine Learning Research*, Vol. 2, pp. 419-444, 2002.
- [8] S. Shehata, F. Karray, and M. Kamel, “Enhancing Text Clustering Using Concept-Based Mining Model,” *Proc. IEEE Sixth Int’l Conf. Data Mining (ICDM ’06)*, pp. 1043-1048, 2006.
- [9] <http://www.practicefusion.com/blog/ehr-adoption-rates/>

