IJRSET SEPTEMBER 2016 Volume 3, Issue 9 Pages: 18-25

ISSN 2394-739X

International Journal for Research in Science Engineering and Technology

ANALYSIS ON ADVERSE DRUG REACTION USING DIFFERENT DATA MINING TECHNIQUES

[1] Deepalakshmi. K. M, [2]Dr. M. S. Vijaya,
[1] Research Scholar, [2] Associate Professor,
[1, 2] Department of Computer Science,
[1, 2] PSGR Krishnammal College of Women,
[1, 2] Coimbatore 641 004.

ABSTRACT: Drug toxicity is a pressing health problem which is also an impediment to the development of therapeutically effective drugs. Despite many on-going efforts to determine the toxicity beforehand, computational prediction of drug side-effects remains a challenging task. An approach to predict side-effects by utilizing side-information sources for the drugs, while simultaneously comparing state-of-the-art machine learning methods to improve accuracy. Specifically, the thesis implements a data analysis pipeline for obtaining side-information that are useful for the prediction task. Then formulates the drug side-effect prediction as a machine learning problem: Given disease indications and structural features (as side-information sources) of drugs, for which some measurements of side-effect exist, predict side effect for a new drug. This paper summarizes adverse drug using mining different methods, general database analysis in data mining.

Keywords: [Drug, Side Effect, Prediction, Machine Learning, Methods]

1. INTRODUCTION

Drug side-effects/ adverse drug reactions (ADRs) is a crucial and complex challenge. The research community is concerned as drug toxicity is the fourth leading cause of death in U.S alone after cancer and heart diseases. Moreover, If the drug success rate in clinical trials increases from 25 percent to 33 percent, pharmaceutical companies can save around 200 million dollars on the drug development process and reduce one fourth of the total drug development time. Effective ADRs prediction is essential for improving patients healthcare and accelerating the drug development process. Different computational techniques have been used in recent past in order to understand the mechanism of drug reactions. The data sources used to study side effect in different studies include chemical data of both drugs and drug-targets. The major cause of drug side-effect is off-target reactions. The mechanism of action of drugs is influenced by the genomic heterogeneity of individuals and influencing chemical properties due to altering micro-environment in cellular compartments. Hence, side-effects "as clinical phenotypes" that arise in patients can assumed to be a manifestation of complex interaction of multitude of factors i-e genomic features, disease state in which drugs are administered called drug indications, chemical descriptors of drugs. Unlike drug chemical and structural properties, the genomic information of the patients is not available from public databases. A disease or a group of specific biological symptoms arises due to abnormal "interaction" or "cross-talk" of pathways. Drugs are usually administered to restore the normal state by triggering the cascade of reactions in perturbed pathways. Here genome plays a fundamental role in mechanizing biological reactions of these drugs; moreover, similar drugs are given for similar diseases. Therefore, this thesis considers that the drug indications are an approximation of missing genomic information from the patients. The basic research question this thesis aims to solve is if the chemical discriptors called as fingerprints of the drugs and drug indications are an insightful information source for the drug side-effects. Finding out these underlying relations of drugs side effect that arise due to this complex interaction is the motivation behind this study. Drug indications and chemical descriptors are used as data sources to predict ten different "common" side-effects. An analysis pipeline to predict the general side-effects is developed for this research work. A straightforward comparison of seven different machine learning algorithms and their prediction performance is presented as a result. The selected side-effects are grouped together in two categories. The categories are based on which data source is more insightful for that particular side-effect prediction. Computational analysis is performed to find a link between biological responses of patients and chemical properties of drugs. The study is constituted on three disciplines: pharmacology, chemi-informatics and machine learning. The aim is to systematically examine the relevant publicly available data for effective modelling of drug side-effects. Comparison of prediction performance of different machine learning models is made, while simultaneously co-behaving side-effects are grouped together in an attempt to better learn the pattern in interactions.

2. ADVERSE DRUG USING MINING DIFFERENT METHODS

2.1 Refining Adverse Drug Reactions using Association Rule Mining for Electronic Healthcare Data Objective

Side effects of prescribed medications are a common occurrence. Electronic healthcare databases present the opportunity to identify new side effects efficiently but currently the methods are limited due to confounding (i.e. when an association between two variables is identified due to them both being associated to a third variable). In this paper we propose a proof of concept method that learns common associations and uses this knowledge to automatically refine side effect signals (i.e. exposure outcome associations) by removing instances of the exposure-outcome associations that are caused by confounding. This leaves the signal instances that are most likely to correspond to true side effect occurrences. We then calculate a novel measure termed the confounding-adjusted risk value, a more accurate absolute risk value of a patient experiencing the outcome within 60 days of the exposure. Tentative results suggest that the method works. For the four signals (i.e. exposure outcome associations) investigated we are able to correctly filter the majority of exposure-outcome instances that were unlikely to correspond to true side effects. The method is likely to improve when tuning the association rule mining parameters for specific health outcomes. This paper shows that it may be possible to filter signals at a patient level based on association rules learned from considering patients' medical histories. However, additional work is required to develop a way to automate the tuning of the method's parameters. In addition to the proof of concept, tentative results are presented for the automatic refinement when considering four signals that have occurred within The Health Improvement Network database for the quinolone drug family.

In this article we observed a proof of concept for a novel efficient ADR signal refinement method that filters instances of a DOI-HOI

(Drug of interest-Health outcome of interest) signal and does not require knowledge of possible confounders. The recorded history of a patient experiencing the signal is used to filter instances where the medical event can be explained by alternative causes (other than the drug). The tentative results suggest that the method has the capability to efficiently refine ADR signals but each signal may require specific tuning to determine the optimal support and confidence values to be implemented.

2.2 Biclustering of Adverse Drug Events in the FDA's Spontaneous Reporting System

we present a new pharmacovigilance data mining technique based on the biclustering paradigm, which is designed to identify drug groups that share a common set of adverse events (AEs) in the spontaneous reporting system (SRS) of the US Food and Drug Administration (FDA). Taxonomy of biclusters is developed, revealing that a significant number of bona fide adverse drug event (ADE) biclusters have been identified. Statistical tests indicate that it is extremely unlikely that the bicluster structures thus discovered, as well as their content, could have arisen by mere chance. In addition, we demonstrate the potential importance of the proposed methodology in several important aspects of pharmacovigilance such as providing insight into the etiology of ADEs, facilitating the identification of novel ADEs, suggesting methods and a rationale for aggregating terminologies, highlighting areas of focus, and providing an exploratory tool for data mining. The objective of this article is to describe a novel pharmacovigilance data mining technique designed to identify drug groups that share a common set of AEs, with which potential ADEs are analyzed and previously unrecognized ADEs may be identified. Learning Outcome In summary, the findings demonstrate the importance and utility of this biclustering methodology for many important aspects of pharmacovigilance noted in several prominent studies. Biclustering provides insight into the etiology of known ADEs and facilitates the identification of novel ADEs. It suggests methods and provides a rationale for aggregating terminologies used to describe ADEs. In addition, biclustering can be used to identify AEs of drug classes. It highlights areas of focus and provides an opportunity for enhanced targeting of novel ADEs. Finally, it provides an exploratory tool for data mining in pharmacovigilance with which the underlying large and complex database can be summarized and described in a big-picture manner, capturing important patterns as well as highlighting data quality issues, which can then be used to improve the signal-detection process.

2.3 Detecting Signals of Adverse Drug Reactions from Health Consumer Contributed Content in Social Media

Adverse drug reactions are causing a substantial amount of hospital admissions and deaths, which cannot be underestimated. Although a great effort has been put on the pre-marketing review during pharmaceutical product development, it cannot identify all possible adverse drug reactions. Many of this online discussion involve adverse drug reactions. In this work, we propose to mine the associations between drugs and adverse reactions from the user contributed content in social media. We have conducted an experiment using five drugs and five adverse drug reactions. The FDA alerts are used as the gold standard to test the performance of the proposed techniques. The result shows that the proposed technique is promising to detect the adverse drug reactions reported by FDA, such as diarrhea, heart condition, depression, and suicidal thoughts. However, adverse drug reaction such as cancer cannot be detected effectively. In this work, in order to explore the potential of detecting ADRs using online healthcare communities, we proposed to employ association rule mining to extract interesting associations of drugs and adverse reactions. When people talk about the ADRs of a specific drug, the co-occurrence of the drug and its ADR in the posts or comments of an online healthcare social media could be regarded as an association rule, and its interestingness and impressiveness can be measured by investigating such metrics as support, confidence and leverage. Association rule mining was first utilized in the field of data mining. Also, in the area of ADRs detection, this method was employed by several researchers to identify potential casual relationships between drugs and adverse reactions from electronic health data. This study is trying to initially test the effectiveness of using association rule mining to extract accurate adverse reactions caused by certain drugs from online healthcare communities.

Nowadays, with the booming of online healthcare communities, more and more patients find it convenient to discuss their health conditions, treatment experience, drug they are taking and adverse reactions of them through these online social media platforms. Since these data are available and accessible to public, if we can make good use of them, ADRs might be detected much earlier and more accurately than using either spontaneous FDA reports or electronic health data. However, very few related studies have focused on social media to identify ADRs, so there is a huge potential in this research area. This study collected posts and comments data of 5 drugs from an online healthcare community – MedHelp, used as grounded truth 5 FDA alerted adverse reactions of these drugs, and employ association rule mining to detect drug adverse reaction of interest. The proposed technique is promising to detect the adverse drug reactions reported by FDA, such as diarrhea, heart condition, depression, and suicidal thoughts. However, adverse drug reaction such as cancer cannot be detected effectively. In the experiment, we calculated the values of support, confidence and leverage for each pair and the results show that our method is able to effectively detect FDA alerted adverse reactions. We also believe that our approach is promising in discovering other potential ADRs.

2.4 Fine-grained Mining of Illicit Drug Use Patterns Using Social Multimedia Data from Instagram

According to NSDUH (National Survey on Drug Use and Health), 20 million Americans consumed drugs in the past few 30 days. Combating illicit drug use is of great interest to public health and law enforcement agencies. Despite of the importance, most of the existing studies on drug uses rely on surveys. Surveys on sensitive topics such as drug use may not be answered truthfully by the people taking them. Selecting a representative sample to survey is another major challenge. In this paper, we explore the possibility of using big multimedia data, including both images and text, from social media in order to discover drug use patterns at fine granularity with respect to demographics. Instagram posts are searched and collected by drug related terms by analyzing the hashtags supplied with each post. A large and dynamic dictionary of frequent drug related slangs is used to find these posts. User demographics are extracted using robust face image analysis algorithms. These posts are then mined to find common trends with regard to the time and location they are posted, and further in terms of age and gender of the drug users. Furthermore, by studying the accounts followed by the users of drug related posts, we extract common interests shared by drug users.

Instead of using the traditional methods such as surveys, we propose to leverage social media to fetch posts from drug users with significantly less time and labor cost, while achieving decent scalability, timeliness, and accuracy. First, by using hashtags-based search to find time patterns of drug consumption, we have uncovered interesting trends in the consumption of various classes of drugs. Then, by location endpoint search, we have found interesting geographical patterns and visualized the locations by drawing bubbles on the map. Moreover, in order to keep updating our database of drug-related hashtags, we applied Frequent Itemset Mining to the hashtags in drug users' posts we found by using recent media searching. Finally, relationship endpoint gives us a way to find potential network among drug users and drug-related pages on Instagram. An important innovation of this study is in multimedia data analysis, and especially faces image analysis. We employ the state-of-the-art Face API of Project Oxford by Microsoft to study the age and gender patterns of drug users. Such fine-grained demographics at a large scale are quite consistent with the findings of National Survey on Drug Use and Health (NSDUH), demonstrating the potential of image-based data analytics for studying drug use and other risky behaviors such as underage drinking. In addition to age and gender, race related patterns can be discovered in a similar fashion, again facilitated by face image analysis. Meanwhile, we have produced potentially significant results that can be utilized in areas including linguistics and psychology.

2.5 Detecting Adverse Drug Effects Using Link Classification on Twitter Data

Adverse drug events (ADEs) are among the leading causes of death in the United States. Although many ADEs are detected during pharmaceutical drug development and the FDA approval process, all of the possible reactions cannot be identified during this period. Currently, postconsumer drug surveillance relies on voluntary reporting systems, such as the FDA's Adverse Event Reporting System (AERS). With an increase in availability of medical resources and health related data online, interest in medical data mining has grown rapidly. This information coupled with online conversations of people which involve discussions about their health provides a substantial resource for the identification of ADEs. In this work, we propose a method to identify adverse drug effects from tweets by modeling it as a link classification problem in graphs. Drug and symptom mentions are extracted from the tweet history of each user and a drug symptom graph is built, where nodes represent either drugs or symptoms and edges are labelled positive or negative, for desired or adverse drug effects respectively. A link classification model is then used to identify negative edges i.e. adverse drug effects. We test our model on 864 users using lO-fold cross validation with Sider's dataset as ground truth. Our model was able to achieve an F-Score of 0.77 compared to the best baseline model with an F -Score of 0.58.

We observed a methodology to derive an ADE detection model to detect ADEs from publicly available twitter data. This model also provides a way to build a medical profile of users from their tweet history. We have shown that building a medical profile of a user helps to detect ADEs more accurately than the baseline model. However, the main problem with extracting medical information from social media messages is that most of the users are not domain experts and often use general terminology to describe their medical conditions. This leads to ambiguity and inaccurate representation of information. This problem can be alleviated by generating a better lexicon or by mapping medical ontologies to better map text terms to formal concepts used the medical literature. Another possible extension to this work is to use three labels: normal interaction, ADE and no interaction. This may also open a way to study the long-term effects of ADEs that are labelled as normal interactions within the current time duration.

2.6 Effective Algorithm For Mining Adverse Drug Reactions

One of the most important issues in the assessment of drug safety is Adverse Drug Reactions (ADR). In premarketing clinical trials, most of the ADRs are not discovered because of the limitations in size. But that are discovered in post-marketing surveillance that is, the impacts of medicines are monitored when they have been delivered to the user. Nowadays, many data mining techniques and methodologies have been developed to motivate the mining and detection of ADRs. These methods are inconvenient and inefficient for users and time consuming. This problem can be alleviated by generating a better lexicon or by mapping medical ontologies to better map text terms to formal concepts used the medical literature. We proposed a combined system platform for the detection of ADRs. This system platform proposed a new data mining algorithm, named as NM-Algorithm. It is based on the genetic algorithm with supervised learning. This proposed algorithm is completely different from the association rule mining. This proposed algorithm covers both similarity and non similarity between the elements so that it is much more efficient than others. In this proposed system, we try to employ an interactive approach to capture the adverse drug reactions between drugs and their reactions. To capture the relationship drugs and symptoms, first we concentrated on generating the initial population named as medical datasets. Based on genetic algorithm with supervised learning, we generate the training tuple. The fitness function that is NM ratio is calculated by comparing the initial population with the training instances. From the fitness function results, the adverse drug reactions are effectively mined.

A system framework has been developed to achieve better post marketing surveillance. In this framework, we have developed NM algorithm and it is based on the genetic algorithm. This provides the information that can help people to discover the causality of a type of events and avoid its potential adverse effects. Users can interact this platform to examine various forms of ADR signals from different viewpoints, by selecting and readjusting parameters of measures of interest. One of the main problems is that the pharmocovigilance using computer systems is a lack of standard measures for signal detection. This paper presents a preliminary development of ADR detection and analysis and there is much scope for extending research, such as this system only discovers drug-ADR and multiADRs. The algorithms will be improved to consider about unsupervised learning methods. It is planned to include more iterations after finding the fitness functions because iterations reduces some unwanted measure from the results.

3. GENERAL ANALYSIS DATABASES

The data sources used to predict the side-effects are the known drug-disease associations and fingerprints/ chemical descriptors of the drugs. Ten different but common side-effects which were used for this study are namely Headache, Dizziness, Weakness, Abdominal Pain, Nausea, Chronic Fatigue, Diarrhea, Rashes, Dermatitis, Vomiting. These side-effects with highest variance were selected for this study. Precisely, data from therapeutic indications of drugs along with their chemical properties (chemical descriptors) are used to predict clinical phenotypes (side-effect) of drugs.

3.1 Data acquisition and dataset construction

Data was collected from public data repositories CHEMBL and SIDER. These databases are freely accessible bioinformatics resource of information. ChEMBL is a bioactivity database which contains information that ranges from drug indications, Targets,e-t-c (Gaulton et al., 2011). SIDER is a public database that contains side-effects of drugs which include side effect frequency, drug side effect classifications and drug–target relations (Kuhn, Letunic, Jensen, & Bork, 2015).

3.1.1 Drug indications data

Drug-Disease relationship is known as drug indications. This drug-disease connectivity map data is retrieved from CHEMBL 22 database with MeSH ids (Medical Subject Headings (MeSH) vocabulary).

3.1.2 Drug descriptors and targets

The 2D fingerprints represent the structural properties of drugs like number of bonds and atoms (non-H atoms and rotatable bonds), functional groups, C-chains, Ring structure and size. MACCS fingerprints are one of the conventional examples of 2D fingerprints that represent drugs with a set of 166 fragments. These descriptors are used for investigating bio-activity of drugs.

3.1.3 Dataset statistics and Features description

Following dataset statistics summarize the constructed dataset for analysis. 2634 unique Drugs (CHEMBL compound Ids) were retrieved from CHEMBL along with 412 drug-target ids (tids), ATC codes. ID Cross-references with other databases namely STITCH database, PubChem database, Drugbank database are also retrieved using AWK and bash scripting. For the number of features, there are 725 mesh ids (drug indications) and 166 ”MACCS” fingerprints calculated from the SMARTS codes retrieved from the CHEMBL 22 database and calculated using rcdk R package. From the 1434 marketed drugs available, 2966 side-effects were retrieved from SIDER 4 with Medra Ids. which is a quantitative measurement recorded as side-effect frequency from 30 patients with recorded medical history. Overall side effect data 9 has in total 97.5 percent data as 0, .1 percent of data is 1 and 2.3 percent of the data matrix contains continuous values within 0 and 1. However, the final dataset that was used after preprocessing were complete cases with available side-effects is summarized.

CONCLUSION

Drug reaction also called substance use disorder, is a dependence on a legal drug or medication. In this paper, we have explained about the problems of adverse drug reaction. In future work we will discuss about other related to drugs and its adverse effects, this analysis identifies the converse effects of drugs and predicts the side effect for a drug available in the market.

REFERENCES

[1] Reps, Jenna M., Uwe Aickelin, Jiangang Ma, and Yanchun Zhang. "Refining adverse drug reactions using association rule mining for electronic healthcare data." In Data Mining Workshop (ICDMW), 2014 IEEE International Conference on, pp. 763-770. IEEE, 2014.

[2] Harpaz, Rave, Hector Perez, Herbert S. Chase, Raul Rabadan, George Hripcsak, and Carol Friedman. "Biclustering of adverse drug events in the FDA's spontaneous reporting system." Clinical Pharmacology & Therapeutics 89, no. 2 (2011): 243- 250.

[3] Yang, Christopher C., Ling Jiang, Haodong Yang, and Xuning Tang. "Detecting signals of adverse drug reactions from health consumer contributed content in social media." In Proceedings of ACM SIGKDD Workshop on Health Informatics. 2012.

[4] Yiheng Zhou, Numair Sani and Jiebo Luo. “Finegrained Mining of Illicit Drug Use Patterns Using Social Multimedia Data from Instagram.” In proceedings of IEEE International Conference on Data Mining Workshop. IEEE 2014.

[5] Satya Katragadda Harika Karnati Murali Pusala, Vijay Raghavan and Ryan Benton. “Detecting Adverse Drug Effects Using Link Classification on Twitter Data”. In proceedings of IEEE International Conference on Bioinformatics and Biomedicine (BTBM) 2015.

[6] H.Sankara Vadivu, E.Manohar And R.Ravi. “Effective Algorithm For Mining Adverse Drug Reactions” International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014

[7] Deepa A, et al., “Causal Association Mining for Detection of Adverse Drug Reactions”, IEEE: International Conference on Computing Communication Control and Automation, 2013.

[8] DuMouchel W, “Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system”, Am Stat, Volume 53(3) PP: 170–190, 1999.

[9] Griffiths T.L. and Steyvers M, “Finding Scientific Topics”, In: Proceedings of the National Academy of Sciences of the United States of America, Volume 101, PP: 5228-5235, 2004.

[10] Hakobyan L, Haaijer-Ruskamp, de Zeeuw D, and P. Denig, "A review of methods used in assessing non-serious adverse drug events in observational studies among type 2 diabetes mellitus patients", Health Qual Life Outcomes, vol. 9, PP:83, 2011.

[11] Haodong Y, Christopher C, “Harnessing Social Media for Drug-Drug Interactions Detection”, IEEE International Conference on Healthcare Informatics, 2013.

[12] Harvey J, Murff, Vimla L, Patel, George H, and David W. Bates “Detecting adverse events for patient safety research: a review of current methodologies”, Journal of Biomedical Informatics PP: 131–143, 2003.