Mining Adverse Drug Reaction For Infrequent Causal Association

S.Sandhya Lakshmi[*1], S.Sheba Roshni[*2]
A.G Merlyn Beulah[# 3]
[*1,*2]U.G Student, B.E CSE, Alpha College of Engg, Chennai, T.N, India.
[#3]Assistant Professor, Dept of CSE, Alpha College of Engg, Chennai, T.N, India.

[1]sandhyabhatpsb@gmail.com

Abstract - Adverse Drug Reaction (ADR) is one of the most important issues in the assessment of drug safety. In fact, many adverse drug reactions are not discovered during limited pre-marketing clinical trials instead, they are only observed after long term post-marketing surveillance of drug usage. Recently, large numbers of adverse events and the development of data mining technology have motivated the development of statistical and data mining methods for the detection of ADRs. These stand-alone methods, with no integration into knowledge discovery systems, are tedious and inconvenient for users and the processes for exploration are time-consuming. This paper proposes an interactive system platform for the detection of ADRs. By integrating an ADR data warehouse and innovative data mining techniques, the proposed system not only supports OLAP style multidimensional analysis of ADRs, but also allows the interactive discovery of associations between drugs and symptoms, called a drug-ADR association rule, which can be further developed using other factors of interest to the user, such as demographic information. The experiments indicate that interesting and valuable drug-ADR association rules can be efficiently mined.

Index Terms—adverse drug reactions, association rules, data mining algorithms, interestingness measure, Recognition Primed
Decision mode

I. INTRODUCTION

Finding causal associations between two events or sets of events with relatively low frequency is very useful for various real-world applications. For example, a drug used at an appropriate dose may cause one or more adverse drug reactions (ADRs), although the probability is low. Discovering this kind of causal relationships can help us prevent or correct negative outcomes caused by its antecedents. However, mining these relationships is challenging due to the difficulty of capturing causality among events. In this paper, we try to employ a knowledge-based approach to capture the degree of causality of an event pair within each sequence we are going to match the data which was previously referred or suggested for treatment. . We then develop an interestingness measure that incorporates the causalities across all the sequences in a database.

ADRs represent a serious world-wide problem. They can complicate a patient's medical condition or contribute to increased morbidity, even death. Studies have shown that ADRs contribute to about 5 percent of all hospital admissions.Even though premarketing clinical trials are required for all new drugs before they are approved for marketing, these trials are necessarily limited in sample-size and duration, and thus are not capable of detecting rare ADRs. Drug safety depends heavily on post marketing surveillance that is, the monitoring of impacts of medicines once they have been made available to consumers. In the US, current post marketing surveillance methods primarily rely on the FDA's spontaneous reporting system Med Watch. Because ADR reports are filed at the discretion of the users of the system, there is gross underreporting. Systematic methods for the detection of suspected safety problems from spontaneous reports have been studied and practically implemented. For example, the FDA currently adopts a data mining algorithm called Multi-item Gamma Poisson Shrinker for detecting potential signals from its spontaneous reports. Another important signal detection strategy is known as the Bayesian Confidence Propagation Neural Network that has been used by the Uppsala Monitoring Center in routine pharmacovigilance with its World Health Organization database. Various other methods such as proportional reporting ratios empirical Bayes screening, and reporting odds ratios have been used in the spontaneous reporting centers of other nations (e.g., England and Australian). These methods have shown better performance than traditional methods. However, the performance of these techniques could be highly situation dependent due to the weaknesses and potential biases inherent in spontaneous reporting. As electronic patient records become more and more easily accessible

in various health organizations such as hospitals, medical centers, and insurance companies, they provide a new source of information that has great potential to generate ADR signals much earlier. Note that each patient case can be considered as an event sequence where various events such as drug prescription, occurrence of a symptom and lab test occur at different times. In the literature, there exist a couple of studies that attempted to find the associations between drugs and potential ADRs by mining their temporal relationships. That is, they tried to mine temporal association rules (represented as $\rightarrow y$)) where Y occurs after X within a time window of length T. These studies obtained promising results based on administrative health data. However, temporal association was the only parameter used for linking a symptom with a drug within each patient case in their work. Temporal association assumes that cause precedes effect. Other parameters such as dechallenge and rechallenge can also give direct or indirect cues of the potential causal association of a drug-symptom pair. Dechallenge is defined as the relationship between withdrawal of the drug and abatement of the adverse effect. Rechallenge describes the relationship between reintroduction of the drug followed by recurrence of the adverse event. In addition, their approaches suffer from the sharp boundary problem. On the one hand, the symptom events near the time boundaries are either ignored or overemphasized. On the other hand, two symptom events contribute equally to the interestingness measure as long as they occur within the hazard period T. That is, the length of the time duration between exposure to the drug and occurrence of the symptom has no effect on the interestingness measure. This is not true in reality because if an ADR symptom occurs within a shorter period, it is usually more likely to be caused by the drug.

To more effectively mine infrequent causal associations, it is necessary to develop a new data mining framework. This paper is a substantial extension of our previous Work where an interestingness measure called **causal-leverage** was developed on the basis of a computational fuzzy recognition-primed decision (RPD) model.

II. EXISTING SYSTEM

The Recognition-Primed Decision (RPD) model is a primary naturalistic decision-making approach which seeks to explicitly recognize how human decision makers handle complex tasks and environment based on their experience. Recognition-Primed Decision (RPD) model is not capable of detecting rare ADR's. This works on trial and error method which is less efficient in recent days. Only single drug used in the existing system. These performances are situation dependent and reporting is not spontaneous.

III. PROPOSED SYSTEM

Effectively reduce the undesirable effects caused by frequent events. The new measure is named exclusive **Causal-leverage** measure. Here using Data mining algorithm to mine ADR signal pairs from electronic patient database based on the new measure. comparing new exclusive causal-leverage measure with previously proposed causal-leverage measure.The superiority of our new measure is established by selecting "three" drugs and evaluated the top 10 ICD-9,where as in the previous system the effectiveness was measured only using a "single" drug.

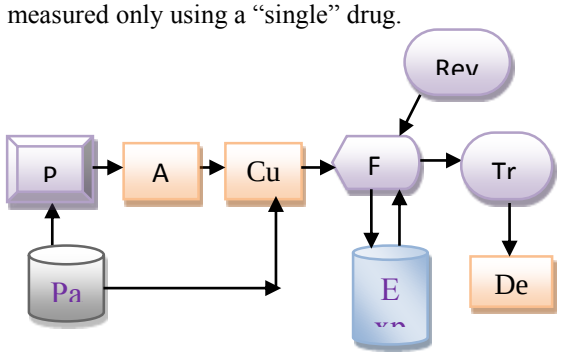
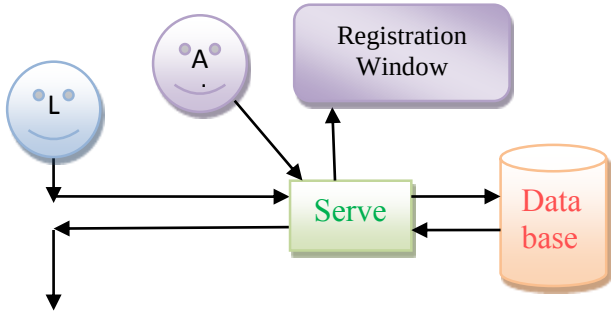


Fig .1. System Architecture

IV. MODULE DESCRIPTION

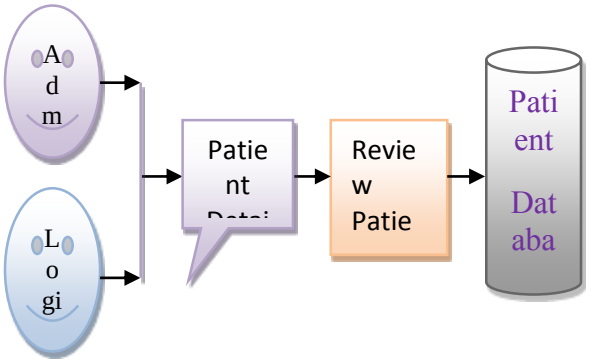
A. User or Admin Authentication Design

Plays an important role for the user to interact with login page to patient details page or admin page. Authorized users can login with their valid credentials otherwise they have to register with their details like providing username, password, mail-id, address, and phone number...Etc details. Registered details will be stored into database and will be authenticate while login time.



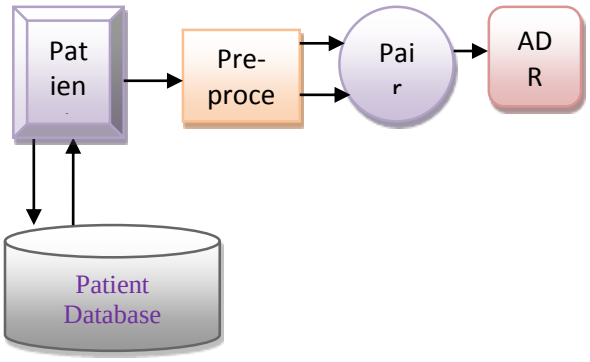
B. Patient Electronic Details

Collect the details of patient which are stored in database called "Patient Electronic Database". The details of patient reports and his symptoms will be also available for reference purpose. Determine the effect of patient and going to collect the details of each and every report of his causal effect.



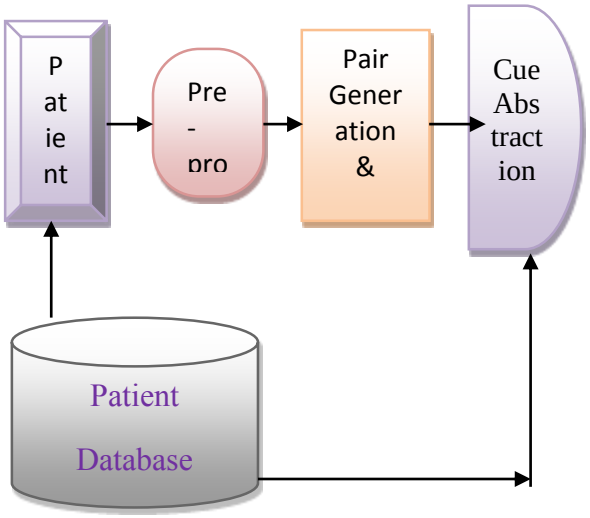
C.Searching for ADR Signal Pairs

Collect the details of patient records from electronic database. Mine potential ADRs from electronic patient records using exclusive causal-leverage measure. Patient data are stored in relational tables in a database and can be retrieved using SQL. These tables are linked through patient identification numbers (PIDs). The drug-related data and symptom related data are stored in two tables called Patient Drug Table and Patient Symptom Table.



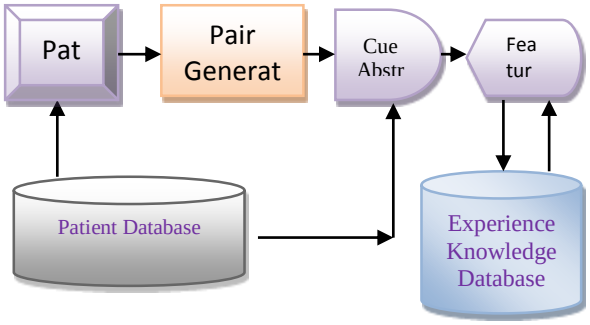
D.Pair Generation & Cue Abstraction

Generate the signal pair for each & every drug reactions of patient details which are been retrieved from database. Mine all interesting association rules that combine all possible events or items in a database. Cue abstraction will be suggests us that the symptoms of the drug reaction which was occurred in previous stage of signal pair mechanism. The symptoms which will generated by this cue abstraction will be referred for drug reaction signal pairs and its reference purpose



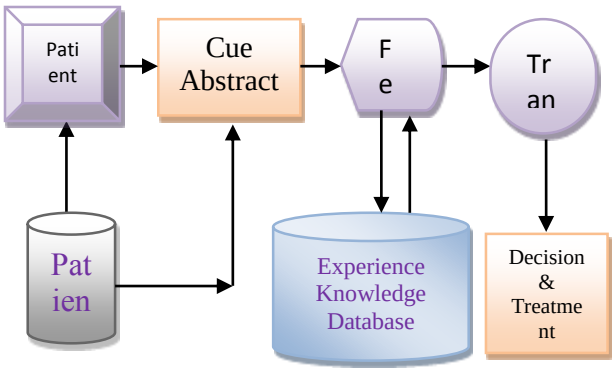
E. Searching for drugs and the support count.

If users are only interested in mining the potential ADRs of a particular drug or a couple of drugs, the users can specify the drugs of interest. Similarly, the users can also specify the list of symptoms if they want to analyze which drugs can cause the symptoms of interest. However, the Patient Drug Table and the Patient Symptom Table still need to be searched in order to get the support count for each drug or symptom. Generate the signal pairs depends upon the symptoms and after that it will generate values for reference drugs and it's details like very likely, probable, unlikely values for their reference purpose



F. Transformed data & Association rules

Verify the details like output generated by support counts details. Here is the actual formation of drug count and signal pairs of drug & symptom. Here, the formation of treatment decision will be taken in this module only after review of medical history with their signal pairs & drug symptoms they will process with their perfect decisions.



V. SCOPE OF THE PROJECT

It is majorly used for Immediate Treatment for patients. However, mining the relationships between Drug and its Signal Reaction will be treated by In-Experienced Physician's also.

VI. CONCLUSION & FUTURE WORK

Mining the causal association between two events is very important and useful in many real applications. It can help people discover the causality of a type of events and avoid its potential adverse effects. However, mining these associations is very difficult especially when events of interest occur infrequently. We have developed a new interestingness measure, exclusive causal-leverage, based on an experience-based fuzzy RPD model. The Experimental results of ADR Signal pairs which are based on Definitive ADRs, Possible ADRs & Probable ADRs are placed in cloud database environment which can be effectively used through entire global network. Experimental results showed that our algorithm could effectively make known ADRs rank high among all the symptoms in the database.

VII.REFERENCE

[1] A. Szarfman, J.M. Tonning, and P.M. Doraiswamy,"Pharmacovigilancein the 21st Century: New Systematic Tools for an Old Problem, "Pharmacotherapy, vol. 24, pp. 1099-1104, 2004.

[2] L. Hazell and S.A.W. Shakir, "Under-Reporting of Adverse Drug Reactions - A Systematic Review," Drug Safety, vol. 29, pp. 385- 396, 2006.

[3] S.J. Evans, P.C. Waller, and S. Davis, "Use of Proportional Reporting Ratios (PRRs) for Signal Generation from Spontaneous Adverse Drug Reaction Reports," Pharmacoepidemiology and Drug Safety, vol. 10, pp. 483-486, 2001.

[4] Y. Ji, H. Ying, M.S. Farber, J. Yen, P. Dews, R.E. Miller, and R.M. Massanari, "A Distributed, Collaborative Intelligent Agent System Approach for Proactive Postmarketing Drug Safety Surveillance," IEEE Trans. Information Technology in Biomedicine, vol. 14, no. 3, pp. 826-837, Dec. 2010.

[5]H. Yun, D. Ha, B. Hwang, and K.H. Ryu, "Mining Association Rules on Significant Rare Data Using Relative Support," J. Systems Software, vol. 67, pp. 181-191, 2003.

[6] L. Geng and H.J. Hamilton, "Interestingness Measures for Data Mining: A Survey," ACM Computing Surverys, vol. 38, no. 3,article 9, 2006.

[7] K. Hartmann, A.K. Doser, and M. Kuhn, "Postmarketing Safety Information: How Useful are Spontaneous Reports?," Pharmacoepidemiology and Drug Safety, vol. 8, pp. 65-71, 1999.

[8] D. Heckerman, "Bayesian Networks for Data Mining," Data Mining Knowledge Discovery, vol. 1, pp. 79-119, 1997.

[9] L. Szathmary, P. Valtchev, and A. Napoli, "Finding Minimal Rare Itemsets and Rare Association Rules," Proc. Fourth Int'l Conf Knowledge Science, Eng. and Management, pp. 16-27, 2010.

[10] Y. Ji, H. Ying, P. Dews, M.S. Farber, A. Mansour, J. Tran, R.E. Miller, and R.M. Massanari, "A Fuzzy Recognition-Primed Decision Model-Based Causal Association Mining Algorithm for Detecting Adverse Drug Reactions in Postmarketing Surveillance,"Proc. IEEE Int'l Conf. Fuzzy Systems, 2010.

[11] Y. Ji, H. Ying, P. Dews, A. Mansour, J. Tran, R.E. Miller, and R.M. Massanari, "A Potential Causal Association Mining Algorithm for Screening Adverse Drug Reactions in Postmarketing Surveillance," IEEE Trans. Information Technology in Biomedicine, vol. 15, no. 32, pp. 428-437, May 2011

[12] Y. Ji, R.M. Massanari, J. Ager, J. Yen, R.E. Miller, and H. Ying, "A Fuzzy Logic-Based Computational Recognition-Primed Decision Model," Information Science, vol. 177, pp. 4338-4353, 2007.

[13] J. Lazarou, B.H. Pomeranz, and P.N. Corey, "Incidence of Adverse Drug Reactions in Hospitalized Patients: A Meta-Analysis of Prospective Studies," J. Am. Medical Assoc., vol. 279, pp. 1200-1205, Apr. 1998.

[14] W. DuMouchel, "Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System," Am. Statistician, vol. 53, pp. 177-190, 1999.

[15] P.N. Tan and V. Kumar, "Interestingness Measures for Association Patterns : A Perspective," Department of Computer Science, Univ. of Minnesota, 2000