

An Optimal Approach for Mining Rare Causal Associations to Detect ADR Signal Pairs

Vikash Kumar Saw^{*}, Kumaran U^{**}

^{*} Vellore Institute of Technology
^{**} SITE Department

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. In light of this, the detection of adverse drug reactions, as early as possible, is an important topic of research for the pharmaceutical industry. 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- In this paper, we try to employ a knowledge-based approach to capture the degree of causality of an event pair within each sequence and we are going to match the data which was previously referred or suggested for treatment.

- 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

Keywords: Adverse drug reactions, association rules, casual leverage measure.

I. INTRODUCTION

The past few years have been seen a tremendous interest in area of data mining. Data mining is generally thought of as process of finding hidden, non trivial and previously unknown information in large collection of data. Association rule mining is an important component of data mining. Association rules are an important class of methods of finding regularities/patterns in data. It is perhaps the most important model invented and

extensively studied by database and data mining community. Association mining has been studied in many application domains. Support: The rule holds with support sup in T (the transaction data set) if $\text{sup}\%$ of transactions contain $X \cup Y$.

$$\square \text{ sup} = \Pr(X \cup Y).$$

Confidence: The rule holds in T with confidence conf if $\text{conf}\%$ of transactions that contain X also contain Y .

$$\square \text{ conf} = \Pr(Y | X)$$

An association rule is a pattern that states when X occurs, Y occurs with certain probability. Support count: The support count of an item set X , denoted by $X.\text{count}$, in a data set T is the number of transactions in T that contain X . Assume T has n transactions.

One of them is finding causal associations between two events or sets of events with relatively low frequency are 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 and the infrequent nature of the events of interest in these applications. In this paper, we try to employ a knowledge-based approach to capture the degree of causality of an event pair within each sequence since the determination of causality is often ultimately application or domain dependent. We then develop an interestingness measure that incorporates the causalities across all the sequences in a database. Our study was motivated by the need of discovering ADR signals in post marketing surveillance, even though the proposed framework can be applied to many different applications.

II. SYSTEM ANALYSIS IDEA

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. Motivated by the need for quantitative computer modeling and simulation of human decision process as well as computerized assistance to enhance this process, we have developed a general-purpose computational RPD model using fuzzy systems technology. Fuzzy sets, fuzzy rules, fuzzy logic, and fuzzy reasoning are used to represent, interpret, and compute

imprecise and subjective information that is commonly encountered in real-world applications. A heterogeneous similarity measure is introduced to acquire prior experiences by evaluating the extent of matching between the current situation and a past experience. Furthermore, an action evaluation strategy is developed, where fuzzy logic plays an important role. Through a simplified yet practical example of detecting adverse drug reaction, we demonstrate how the proposed computational model can be utilized to quantitatively describe and facilitate the decision process. Importantly, as a general-purpose technique, its application is beyond the medicine domain.

DRAWBACK IN EXISTING SYSTEM:

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.

PROPOSED SYSTEM:

We developed and incorporated an exclusion mechanism that can effectively reduce the undesirable effects caused by frequent events. Our new measure is named exclusive causal-leverage measure. We proposed a data mining algorithm to mine ADR signal pairs from electronic patient database based on the new measure. The algorithm’s computational complexity is analyzed. We compared our new exclusive causal-leverage measure with our previously proposed causal-leverage measure as well as two traditional measures in the literature: leverage and risk ratio. To establish the superiority of our new measure, we did extensive experiments. In our previous work, we tested the effectiveness of the causal-leverage measure using a single drug in the experiment. In this paper, we selected three drugs and evaluated the top 10 ICD-9 (International Classification of Diseases, ninth Revision) codes ranked by the exclusive causal-leverage measure for each drug. We also tested how the length of hazard period T affects the performance of the exclusive causal-leverage measure. Drugs and their associated ADRs have causal relationships. In this section, we examine how to search for potential ADR signal pairs from an electronic patient database using the above exclusive causal-leverage measure. We assume that patient data are stored in relational tables in a database and can be retrieved using database language like structured.

ADVANTAGES IN PROPOSED SYSTEM

The support count for each drug or symptom will be used to calculate the exclusive causal-leverage value for related pairs. We can find the Strength of the causality between a drug and a particular possible ADR..

III. PROPOSED ALGORITHM

Exclusive causal-leverage measure:

Algorithm:

Searching for drugs and the support count for each drug

- 1: drugHashTable $\frac{1}{4}$ null
- 2: for each patient Pk 2 DB do
- 3: retrieve all the drugs Dk taken by the patient
- 4: for each drug dkl 2 Dk do

- 5: if (drugHashTable:containsKey(dkl) == false) do
- 6: Sigma= 1 {a new drug dkl is found and set its Support count as 1 }
- 7: else
- 8: Sigma=drugHashTable:getValue(dkl) + 1 {update Support count }
- 9: end if
- 10: drugHashTable.putValue(dkl,Sigma)
- 11: end for
- 12: end for
- 13: return (drugHashTable)

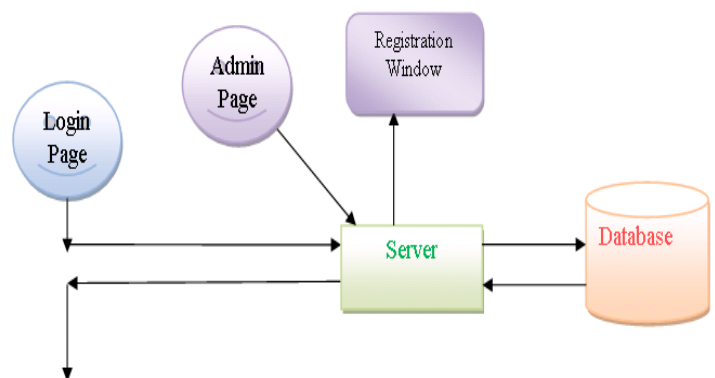
IV. METHODOLOGIES / EXPERIMENTATION

Following modules involves in Dynamic Workflow Scheduling concept. Modules:

- User or Admin Authentication Design
- Patient Electronic Details.
- Searching for ADR Signal Pairs.
- Pair Generation & Cue Abstraction.
- Searching for drugs and the support count.
- Transformed data & Association rules.

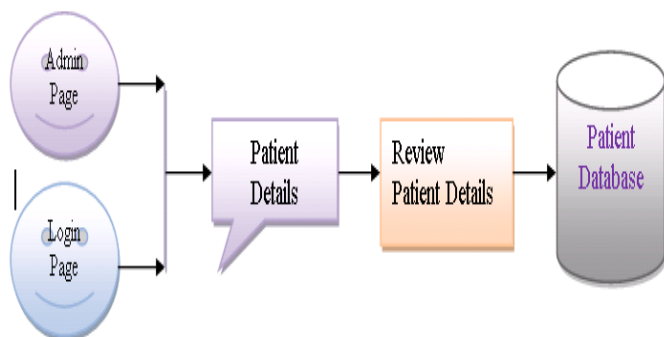
V. MODULES DESCRIPTION

User or Admin Authentication Design: This is the initial module of our project. User Authentication Login Page Design plays an important role for the user to interact with login page to patient details page or admin page. This module has been created for user authentication purpose. In this login 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. So, thereafter registered details will be stored into database and will be authenticate while login time. It will verify each and every user information details. If those details are doesn’t matches with database details then it will gives an error message and it will shows the registration page automatically. So, here we are skipping the illegal users and providing more surveillance for our application.



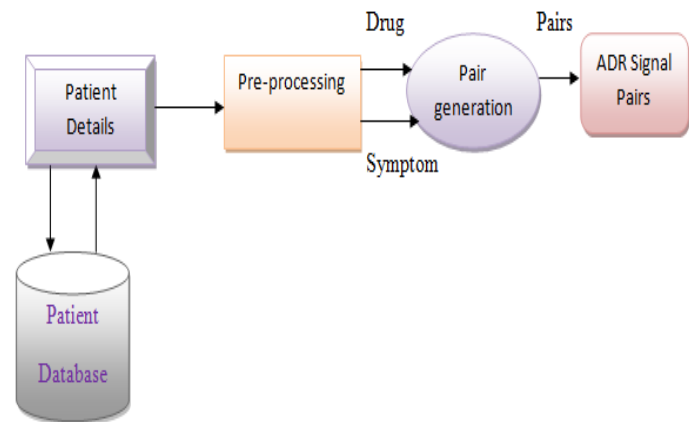
➤ Patient Electronic Details

This is the next module after our initial module. In this module, Patient electronic details will be available at this stage. Here we are going to collect the details of patient which are stored in database called “Patient Electronic Database”. In this database, the details of patient reports and his symptoms will be also available for reference purpose. Here we are going to determine the effect of patient and we are going to collect the details of each and every report of his causal effect.



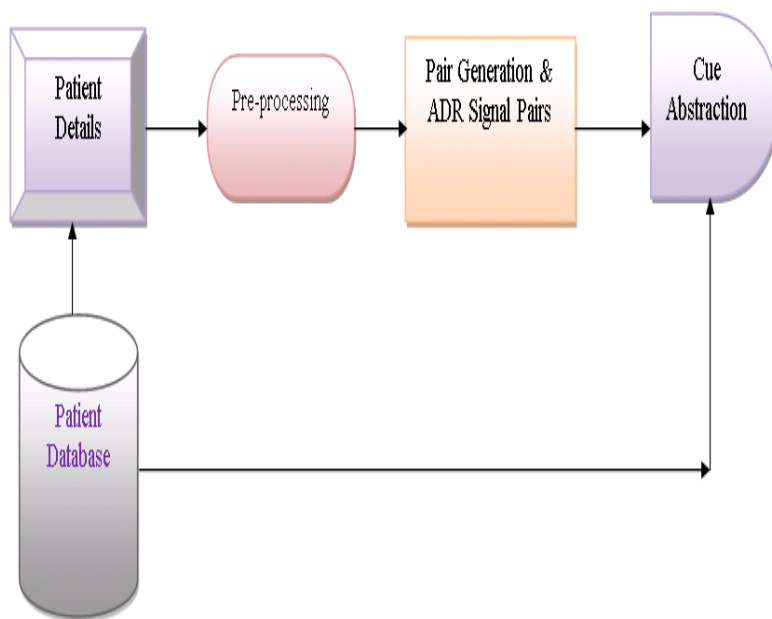
➤ Searching for ADR Signal Pairs

This is our third module section in our project. Here, we are going to collect the details of patient records from electronic database. In this section, we first introduce a new measure we recently developed. We then discuss how to mine potential ADRs from electronic patient records using this measure. We examine how to search for potential ADR signal pairs from an electronic patient database using the above exclusive causal-leverage measure. We assume that patient data are stored in relational tables in a database and can be retrieved using database language like structured query language (SQL). These tables are linked through patient identification numbers (PIDs). We also assume that the drug-related data and symptom-related data are stored in two tables called Patient Drug Table and Patient Symptom Table, respectively.



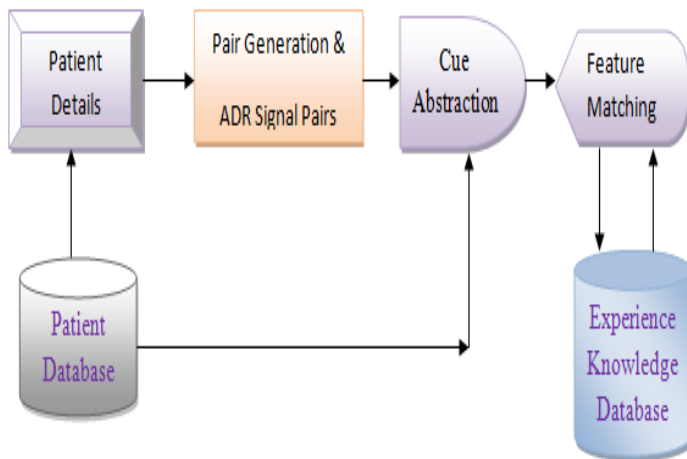
➤ Pair Generation & Cue Abstraction.

This is the fourth module in our project. In this module, we are going to generate the signal pair for each & every drug reactions of patient details which are been retrieved from database. Here, we are going to do most existing data mining methods mine all interesting association rules that combine all possible events or items in a database. Thereafter, the cue abstraction will be processed very efficiently. 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 generate by this cue abstraction will be referred for drug reaction signal pairs and its reference purpose.



➤ Searching for drugs and the support count.

This is the fifth module of our application. In this application we are going to do 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. In both cases, 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. So, here the support counts of each & every drug of signal pairs & symptoms in our process. Thereafter, it will generate the signal pairs depends upon the symptoms and after that it will generate values for reference drugs ant it’s details like very likely , probable, unlikely values for their reference purpose.



➤ Patient Electronic Details

Input:

- Patient Details
- Symptoms

Output:

- Review Patient Details.

➤ Searching for ADR Signal Pairs

Input:

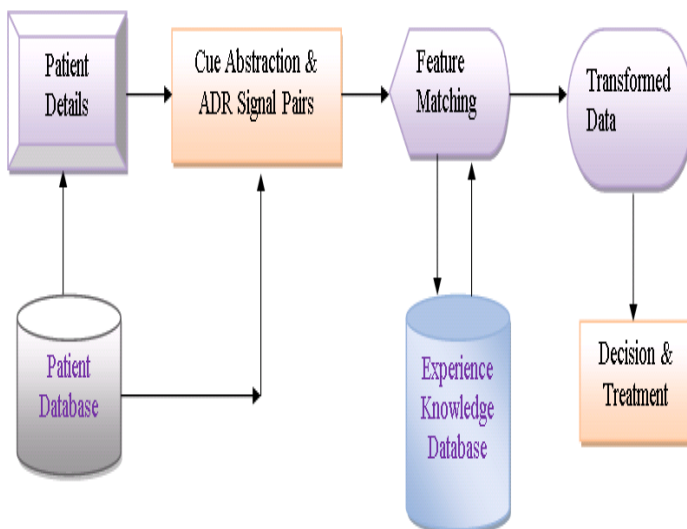
- Patient Details
- Drug symptoms

Output:

- Drug Reaction Signal Pairs.

➤ Transformed data & Association rules.

This is the last module in our project. In this section we are going to verify the details like output generated by support counts details. So, 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.



➤ Pair Generation & Cue Abstraction

Input:

- Signal Pairs
- Drugs & symptoms

Output:

- Symptom Details.

➤ Searching for drugs and the support count

Input:

- Drugs & symptoms

Output:

- Support Count of Drugs.

➤ Transformed data & Association rules

Input:

- ADR Signal pairs
- Symptom

GIVEN INPUT EXPECTED OUTPUT:

➤ User or Admin Authentication Design

Input:

- Username
- Password

Output:

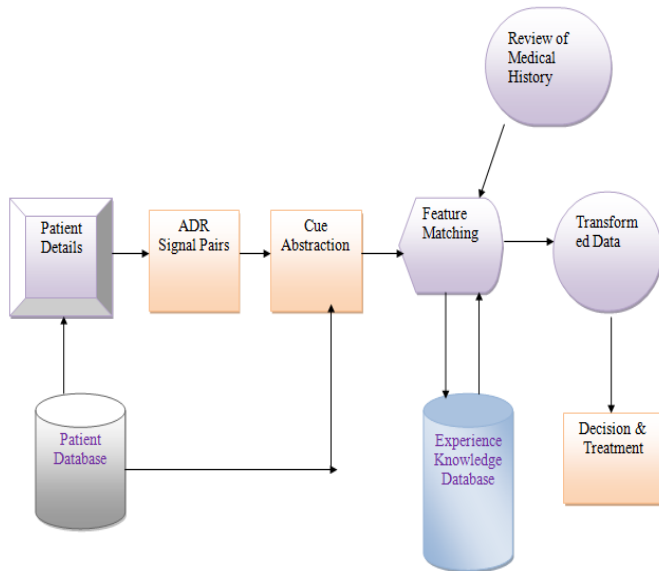
- Patient or Admin Window

Output:

- Transformed Data
- Association Treatment Rules.

DETAILED DESIGN:

System Architecture



VI. CONCLUSION

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.

VII. FUTURE ENHANCEMENTS

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. Algorithm design and efficiency analysis become more important when one studies how to efficiently mine all possible rare event sets and association rules based on minimal support. Another difference is that according to the literature, existing rare association rule mining research assumes that all the data are stored in a single table.

REFERENCES

- [1] 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.
- [2] H. Jin, J. Chen, H. He, G. Williams, C. Kelman, and C. O'Keefe, "Mining Unexpected Temporal Associations: Applications in Detecting Adverse Drug Reactions," IEEE Trans. Information Technology in Biomedicine, vol. 12, no. 4, pp. 488-500, July 2008.
- [3] 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.
- [4] 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.
- [5] A.K.H. Tung, H. Lu, J. Han, and L. Feng, "Efficient Mining of Intertransaction Association Rules," IEEE Trans. Knowledge and Data Eng., vol. 15, no. 1, pp. 43-56, Jan./Feb. 2003.
- [6] C. Marinica and F. Guillet, "Knowledge-Based Interactive Postmining of Association Rules Using Ontologies," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 6, pp. 784-797, June 2010.
- [7] L. Szathmary, A. Napoli, and P. Valtchev, "Towards Rare Itemset Mining," Proc IEEE 19th Int'l Conf. Tools with Artificial Intelligence, 2007.
- [8] H. Ying, Fuzzy Control and Modeling: Analytical Foundations and Applications. IEEE Press, 2000.

AUTHORS

First Author – Vikash Kumar Saw, MCA, Vellore Institute of Technology, Vellore.



Second Author – Kumaran U, Assistant Professor, SITE Department, Vellore Institute of Technology, Vellore.

